D31.2

Risk-aware deployment and intermediate report on status of legislative developments in data protection

Project number:	609611
Project acronym:	**PRACTICE**
Project title:	Privacy-Preserving Computation in the Cloud
Project Start Date:	1 November, 2013
Duration:	36 months
Programme:	FP7/2007-2013
Deliverable Type:	Report
Reference Number:	ICT-609611 / D31.2 / 1.0
Activity and WP:	Activity 3 / WP31.2
Due Date:	31 October 2015 - M24
Actual Submission Date:	10th November, 2015
Responsible Organisation:	UMIL
Editor:	Ernesto Damiani
Dissemination Level:	PU
Revision:	2.00
Abstract:	This deliverable reports an overview of the current legal framework regulating storage and processing the data on the cloud and develops a methodology to analyze the business risks associated with outsourcing data, supported by a web-based tool.
Keywords:	Legal Framework, Secure computation, Data protection directive, Risk assessment methodology

This project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 609611.

Editor

Ernesto Damiani (UMIL)

Contributors (ordered according to beneficiary numbers)

Ernesto Damiani (UMIL)
Valerio Bellandi (UMIL)
Stelvio Cimato (UMIL)
Gabriele Gianini (UMIL)
Gerald Spindler (UGOE)
Matthis Grenzer (UGOE)
Niklas Heitmüller (UGOE)
Philipp Schmechel (UGOE)

Executive Summary

The goal of PRACTICE's WP31 is twofold: (i) reporting on the current legal framework regulating the protection of data stored and processed on the cloud and (ii) developing a risk assessment methodology for data sharing in cloud-based services.
Extending the report already provided in the previous deliverable, the first part of D31.2 is devoted to a complete overview of the current legal framework regulating data protection in the European Union, discussing the EU Data Protection Directive currently in force and the proposals for a General Data Protection Regulation (GDPR), highlighting their relevance to the processing of personal data on the cloud.
The second part of the deliverable is devoted to the description of the methodology supporting the risk-aware deployment of secure computation, aiming to provide the analysis and the quantitative evaluation of risks related to privacy and confidentiality breaches during the execution of a multi-party business process on the cloud. We discuss in detail how to estimate the probability of information disclosure among colluding partners, introducing a new possibilistic approach, based on the analysis of the micro-economics underlying the business process and the information's value. We also present a web-based tool that can be used model and to simulate the business processes executed on the cloud.

Contents

List of Figures

Chapter 1

Introduction

WP31's objectives are from one side the evaluation of the legal aspects related to the outsourcing of data and of computation to the cloud, and on the other side the development of models and techniques to quantify the business risks associated with data sharing in collaborative services. More precisely, long term's goals of WP31 are (i) the clarification of the legal framework regulating the placing and the processing of sensitive data in locations where different privacy regulations hold, possibly establishing a set of guidelines compliant to international legal frameworks and (ii) the development of a quantitative risk assessment methodology for the deployment of secure computation protocols on the cloud.

The first part of the deliverable is devoted to a complete overview of the current legal framework regulating data protection in the European Union. In particular, we discuss the EU Data Protection Directive currently in force and the proposals for a General Data Protection Regulation (GDPR), highlighting their relevance to the processing of personal data on the cloud.

Specifically, Chapter 2 provides a detailed analysis of the directive, focusing on the distinction of roles and responsibilities from the legal point of view between data controller and data processor, and outlining the requirements for the upcoming new Data Protection Regulation. The proposal of the Council of the European Union for a GDPR has been analyzed and included in the document, especially regarding personal data and encryption. Furthermore, all relevant legal developments concerning the project have been examined and included in the text.

In turn, Chapter 3 reports on some case studies where security controls implementing Secure Multiparty Computation (SMC) techniques have been deployed and discusses their compliance with the current legal framework and with the proposed GDPR.

The second part of the deliverable reports the development of the methodology introduced in D31.1. Chapter 4 describes the iterative methodology for risk-aware deployment of security controls enabling the analysis and the management of risks in cloud business processes. Chapter 5 discusses a new possibilistic approach for likelihood estimation, discussing when and where it should be used extending the traditional probabilistic approach. Some case studies, taken from D24.2, are analyzed using the proposed methodology in Chapter 6. Finally, in Chapter 7, an open source web tool that supports our methodology for process-based risk assessment is introduced. The tool can be used both to model and to simulate the business processes executed on the cloud.

Part I

Part I - Legal Status on Data Protection

Chapter 2

Cloud Computing under the European Data Protection Law

When Cloud Computing is used, legal problems might arise for every party involved. It can particularly be hard to achieve compliance especially with the data protection law. Chapter 2 of this deliverable provides an analysis of the European data Compliance especially with the data protection law can be hard to achieve. Chapter 2 of this deliverable provides an analysis of the European data protection law *de lege lata* and *de lege ferenda*.

2.1 Legal Framework

Prior to analysing specific problems for Cloud Computing arising from the data protection law, the basic functioning and key principles of the current and the upcoming state of law shall be outlined. The causes of and possible solutions for the legal difficulties can only be accurately explained and understood, provided that there is a general understanding of the underlying legal framework.

2.1.1 The Data Protection Directive 95/46/EC (DPD)

The Directive 95/46/EC on the protection of individuals with regard to the processing of personal data, and on the free movement of such data, was adopted in 1995 by the European Community to protect the privacy of individuals with regard to the processing of personal data. [1] EU Directives lay down certain end results that must be achieved in every Member State. National authorities have to adapt their laws to meet these goals and to implement the directives into their national law, but are free to decide how to do so. Nevertheless, directives have to be implemented in such a way that the best result is achieved ("effet utile"). Article 288 of the Treaty on the Functioning of the European Union defines how the Unions competences can be exercised. [2]

> "Article 288 (ex Article 249 TEC): [...] A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods.[...]"

Each Directive specifies the date by which the national implementing laws must be adopted. A directive is addressed to the Member States, not to the citizens. Citizens may claim those rights directly

[1] *Hon/Millard/Walden*, Who is Responsible for 'personal data' in Cloud Computing?, The Cloud of Unknowing, Part 2, p. 3.

[2] Treaty on the Functioning of the European Union, Official Journal C 326 of 26/10/2012, 0001 0390, available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=EN.

only if directives state rights for citizens *and* if they are not implemented in due time by national authorities.

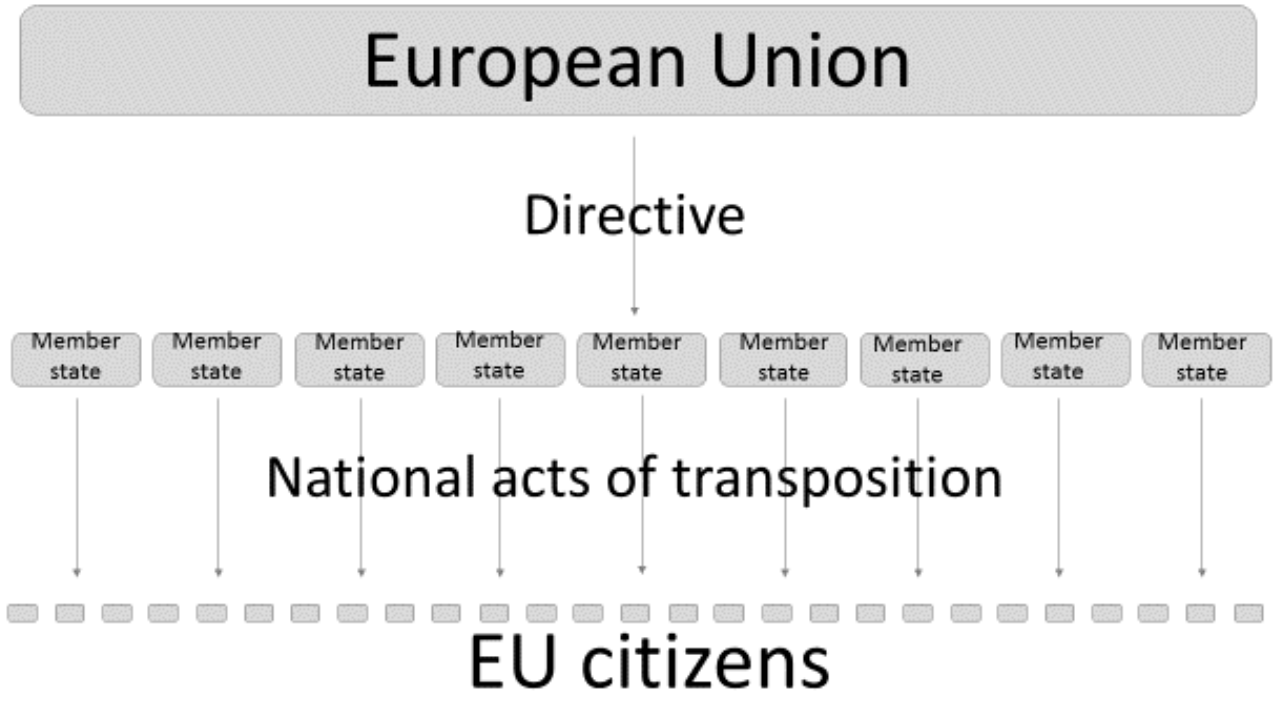

Figure 2.1: EU Directives

Directives are used to harmonize different national laws in order to create and foster the internal European market (e.g. product safety standards). [3]

Directives may differ concerning the grade of harmonization; for example, a *de minimis* harmonization allows Member States some leeway to pass laws which go beyond that level, but it a full harmonization effectively preventing Member States from surpassing the directive.

Concerning the Directive 95/46/EC the European Court of Justice (ECJ) passed a judgment in which it stated that the directive fully harmonizes the data protection law. This means the Member States are not allowed to provide a lower level of protection than the directive demands, nor are they allowed to go beyond it. [4] Directive 95/46/EC imposes complete harmonization of national laws. [5] Directive 95/46/EC is intended to ensure free movement of personal data while guaranteeing a high level of protection for the rights and interests of data subjects, equivalent in all Member States. Consequently, Article 7 of Directive 95/46/EC sets out an exhaustive and restrictive list of cases in which the processing of personal data can be regarded as lawful. That interpretation is corroborated by the term "may be processed only if" which demonstrates the exhaustive and restrictive nature of the list appearing in that article. Thus, the Member States can neither add new principles relating to the lawfulness of processing, nor impose additional requirements. [6]

[3] http://ec.europa.eu/eu_law/introduction/what_directive_en.htm.

[4] *ECJ*, decision of 24/11/2011 - C-J046/10.

[5] *ECJ*, decision of 24/11/2011- C468/10; *Kühling*, EuZW 2012, 281 (282).

[6] *ECJ*, decision of 04/10/2001 - C-450/00, Commission of Luxembourg, available at:http://ec.europa.eu/anti_fraud/documents/data-protection/dpo/ecj_decisions_relating_data_protection_en.pdf.

Exempted from the scope of the directive (Article 3 Par. 2 of Directive) are areas related to the second and third so-called pillars of the European Union, i.e. the common foreign and security policy, police, and judicial cooperation in criminal matters.
The Directive generally prohibits the processing of personal data unless the person concerned has expressly consented to the processing of sensitive data or the processing is necessary to "keep the dissolution of the rights and obligations of the data controller in the field of employment law." In addition, the Directive allows Member States to provide for exceptions for reasons of substantial public interest.
In telecommunications, the Data Protection Directive is complemented by the regulation adopted in the 2002 Directive 2002/58/EC (Directive on privacy and electronic communications).

2.1.1.1 Territorial Scope of the DPD

Since there might be a various number of parties (entities from all over the world) involved in cloud computing solutions, the important issue of international jurisdiction has to be addressed.

The DPD states that each Member State shall apply its data protection law when a "controller" carries out data processing by an establishment on the territory of a Member State. An exception to this principle is provided if the processor does not have an establishment in a Member State but uses equipment situated on the territory of a Member State for the purposes of processing. In this case the European data protection law is applicable to the activities of the processor as well. Even an end-users machine could be considered 'equipment situated on the territory of a Member State' if it is used for storing a cookie or collecting data with java scripts. [7]

In contrast, if a webpage is accessible from the EU but hosted by a server in a third country, no equipment situated inside the EU is used. For the territorial scope of the Directive it is not concerned with where a service is aimed at, but rather, where the resources used for providing this service are located (this principle will change with the upcoming Data Protection Regulation, see 4.1.1). [8] A cloud server in Europe would qualify as 'equipment' in the sense of the DPD. [9]

If a controller is established on the territory of several Member States, they have to ensure compliance with each of the applicable national laws , Article 4 DPD.
Even though recital 19 of the DPD states that an establishment on the territory of a Member State "implies the effective and real exercise of activity through stable arrangements", there is no legal definition of 'establishment' in the DPD. The ECJ clarified in a recent decision that the words "in the context of the activities of an establishment" cannot be interpreted restrictively. [10] According to the ECJ, the concept of "establishment" has to be defined versatile as it, "which departs from a formalistic approach whereby undertakings are established solely in the place where they are registered. Accordingly (...) both the degree of stability of the arrangements and the effective exercise of activities in that other Member State must be interpreted in the light of the specific nature of the economic

[7] As for example stated by the German court KG Berlin in its ruling from 24/01/2014, 5 U 42/12, 28 f., available at:http://www.berlin.de/imperia/md/content/senatsverwaltungen/justiz/kammergericht/presse/5_u_42_12_urteil_vom_24.1.2014_kammergericht_anonymisiert.pdf?start&start&ts=1392399485&file=5_u_42_12_urteil_vom_24.1.2014_kammergericht_anonymisiert.pdf.

[8] *Hon/Hörnle/Millard*, Data Protection Jurisdiction and Cloud Computing - When are Cloud Users and Providers Subject to EU Data Protection Law?, The Cloud of Unknowing, Part 3, p. 7 ff.; *Wieczorek*, DuD 2013, 644 (646); *Gabel*, in: Taeger/Gabel, BDSG, par. 1, recital 59.

[9] *Giedke*, Cloud Computing, p. 205 ff.

[10] *ECJ*, decision of 01/10/2015, Case C-230/14 Weltimmo s. r. o. v Nemzeti Adatvédelmi és Információszabadság Hatóság, recital 25.

activities and the provision of services concerned. This is particularly true for undertakings offering services exclusively over the Internet" (recital 29). On the other hand, it is generally not necessary that the establishment is independent from the controller in order to be considered as a controller itself (for the definition of 'data controller' see 2.3). [11]

Another ongoing case of the ECJ concerns a request for a preliminary ruling by the Supreme Court of Austria (OGH). The case, Verein für Konsumenteninformation v Amazon EU Sárl, refers to an action of the Austrian consumer association against Amazon (which has its European seat in Luxemburg). The case is concerned with Amazons clauses in its standard terms and conditions according to which data from Austrian users might be exchanged with credit-risk assessment and financial services companies in Germany and Switzerland. The association argued that Austrian data protection law should apply to this case. The question of the Austrian court to the ECJ is:

> "4.2. Is the processing of personal data by an undertaking that in the course of electronic commerce concludes contracts with consumers resident in other Member States, in accordance with Article 4(1)(a) of Directive 95/46/EC (...) and regardless of the law that otherwise applies, governed exclusively by the law of the Member State in which the establishment of the undertaking is situated in whose framework the processing takes place or must the undertaking also comply with the data protection rules of those Member States to which its commercial activities are directed?" [12]

The Austrian OGH advocated that the data law of the Member State in which the defendant company has its headquarters should be applicable. [13] The upcoming decision of the ECJ will thus certainly have an impact on the correct interpretation of the territorial scope of the DPD within the Member States.
One of the cases decided by the ECJ highlights the difficulties to handle the notion of establishment in the DPD in practice. [14] The arguments brought forward by the General Advocate in the case of Google vs. Spain are worth being cited literally in order to emphasize the spectrum of interpretation concerning the notion of "establishment":

> "In my opinion the Court should approach the question of territorial applicability from the perspective of the business model of internet search engine service providers. This, as I have mentioned, normally relies on keyword advertising which is the source of income and, as such, the economic raison d'łtre for the provision of a free information location tool in the form of a search engine. The entity in charge of keyword advertising (called 'referencing service provider' in the Court's case-law) is linked to the internet search engine. This entity needs presence on national advertising markets. For this reason Google

[11] The German court Oberverwaltungsgericht (OVG = circuit court in administrative affairs) Schleswig-Holstein had to decide whether or not European data protection law was applicable for the data processing of Facebook, also in which European country Facebooks respective establishment is acting. The court ruled that even though the US-American parent company Facebook Inc. is the only shareholder of the Irish subsidiary Facebook Ltd., the Irish company can be qualified as an establishment within the EU as Facebook Ireland obviously handled some of the data processing, OVG Schleswig Holstein, decision of 22/04/2013; however, another German court (Kammergericht (KG) Berlin (circuit court in civil law issue) in its ruling from 24/01/2014) contradicted that perspective that since the parent group Facebook Inc. is responsible for all decisions concerning data processing in the end, the Irish subsidiary Facebook Ltd. is not an establishment in the sense of the directive. This interpretation of 'establishment' does not comply with the directive's distinction between 'controller' and 'establishment'

[12] Request for a preliminary ruling of the ECJ from the Oberster Gerichtshof (Austria) lodged on 27 April 2015, Case C-191/15 Verein für Konsumenteninformation v Amazon EU Sárl.

[13] *OGH*, decision of 09/04/2015 2 Ob 204/14k = GRUR Int. 2015, 722 (725).

[14] *ECJ*, decision of 13/05/2014 C-131/12 Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez.

> has established subsidiaries in many Member States which clearly constitute establishments within the meaning of Article 4(1)(a) of the Directive. It also provides national web domains such as google.es or google.fi. The activity of the search engine takes this national diversification into account in various ways relating to the display of the search results because the normal financing model of keyword advertising follows the pay-per-click principle.
> 65. For these reasons I would adhere to the Article 29 Working Party's conclusion to the effect that the business model of an internet search engine service provider must be taken into account in the sense that its establishment plays a relevant role in the processing of personal data if it is linked to a service involved in selling targeted advertisement to inhabitants of that Member State.
> 66. Moreover, even if Article 4 of the Directive is based on a single concept of controller as regards its substantive provisions, I think that for the purposes of deciding on the preliminary issue of territorial applicability, an economic operator must be considered as a single unit, and thus, at this stage of analysis, not be dissected on the basis of its individual activities relating to processing of personal data or different groups of data subjects to which its activities relate.
> 67. In conclusion, processing of personal data takes place within the context of a controller's establishment if that establishment acts as the bridge for the referencing service to the advertising market of that Member State, even if the technical data processing operations are situated in other Member States or third countries." [15]

In the final judgment, the ECJ followed the Advocate General's opinion:

> "55 In the light of that objective of Directive 95/46 and of the wording of Article 4(1)(a), it must be held that the processing of personal data for the purposes of the service of a search engine such as Google Search, which is operated by an undertaking that has its seat in a third State but has an establishment in a Member State, is carried out 'in the context of the activities' of that establishment if the latter is intended to promote and sell, in that Member State, advertising space offered by the search engine which serves to make the service offered by that engine profitable." [16]

In the case of cloud computing the DPD would be applicable if the cloud is an entity established within the jurisdiction of a Member State, if data processing is carried out within the context of the activities of this establishment. Every instance of processing a provider tends to carry out would then be covered, including the transfer of data to a non EU country. In general, the directive is applicable if the cloud provider processes the data on a server within a Member State. If the provider is processing data using a machine physically in a certain Member State, this state's law is applicable as long as the provider is not established in another EU-Member State. However, according to the above mentioned ECJ decision "Google Spain" it is already sufficient that the Directive can be applied due to the fact that tthe cloud provider was established and is active in an EU Member State. It is not necessary that this establishment is directly involved in processing the data or has any particular responsibility

[15] Opinion of Advocate General Jääskinen, delivered on 25/06/2013, Case C 131/12 Google Spain SL/AEPD, recital 67

[16] *ECJ*, decision of 13/05/2014, Case C-131/12 Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez, rec. 55; c.f. the decision of the District Court of Heidelberg LG Heidelberg, decision of 09/12/2014 2 O 162/113, recital 32 = MMR 2015, 348 (349), which decided that a link displayed by the google search engine had an unambiguous reference to Germany and that thus German law was applicable.

concerning the processing; it is sufficient from an economic perspective that the establishment supports the activities of the cloud provider, such as in the Google Spain Case regarding the selling of advertisement etc. This however does not mean that the subsidiary company can be sued, as the actual processor remains responsible. Hence, it is sufficient to apply the DPD provided that an establishment operates the funding for the cloud provider.

2.1.1.2 Material Scope and Fundamentals of the DPD

The focus of the Directive concerning "protection of individuals with regard to the processing of personal data and on the free movement of such data" (informal: "Data Protection Directive") is mentioned in Article 1:
'Object of the Directive
(1) In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons and in particular their right to privacy with respect to the processing of personal data.
(2) Member States shall neither restrict nor prohibit the free flow of personal data between Member States for reasons connected with the protection afforded under paragraph 1. [17]
Thus, the Directive again clarifies the two goals of fostering the internal market and guaranteeing basic rights for individuals concerning the protection of their personal data (privacy).
The Directive regulates the processing of personal data regardless of whether such processing is automated or not.

2.1.1.2.1 Material Scope of the Data Protection Directive

The Directive only protects the personal data of individuals, whilst corporate entities are excluded from the scope of the Directive. "Personal data" is any information relating to an identified or identifiable person, regardless of which aspects of the person the information may affect. Some examples are privacy issues, such as in private or job-related spheres, characteristics, skills of an employee, psychological characteristics or elements of someone's biography. [18]
Whilst the Directive applies in general for all kinds of processing data there are still a number of distinctions made by the Directive. In the case of non-automatic processing the Directive only addresses data processing stored in a (physical) dossier. However, in this report we will deal solely with requirements for automatic processing of data due to the character of cloud computing. Moreover, Article 3 Par. 2 refers to some exceptions: [19]
One of the exceptions relevant for internet services (and users) refers to the exemption of exclusive personal and familiar activities. In other words, all activities on social networks etc. (user-generated content) which remain in the social and private sphere are not affected by the Data Protection Directive. However, this exception does not alter the obligations of the operator of a social network.
Finally, the European Court clarified that processing for public safety and prosecution purposes is not within the scope of this Data Protection Directive.

2.1.1.2.2 Fundamentals of the Data Directive

The main principle is that personal data should not be processed at all unless the data processing operator complies with certain requirements. These refer to: transparency, legitimate purpose, and proportionality.

[17] Directive 95/46/EC of the European Parliament and of the Council, Official Journal L 281 of 23/11/1995, 0031 0050.
[18] *Dammann*, in: *Dammann/Simitis*, EG-Datenschutzrichtlinie, Art. 2, p.109
[19] *Ehmann/Helfrich*, EG-Datenschutzrichtlinie, Art. 3, recital 16.

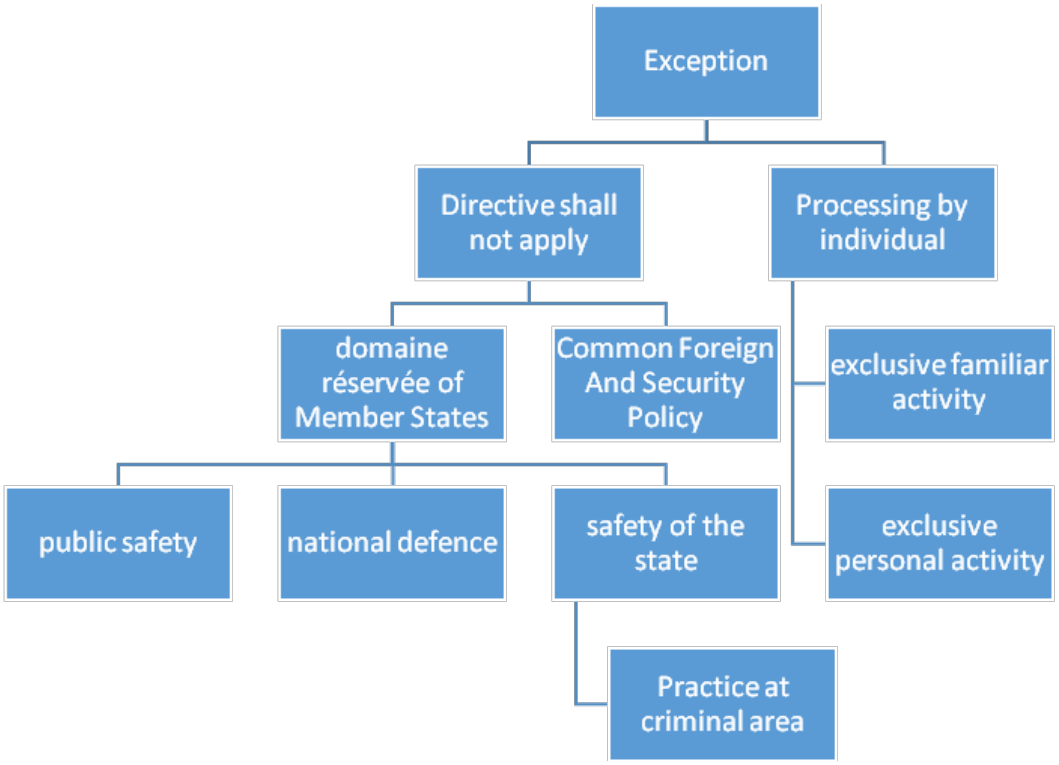

Figure 2.2: Applicability of the Data Protection Directive

Transparency The individual has the right to be informed should his personal data be processed, Article 10 and 11. Before starting the processing the controller has to provide information about his identity (name and address), the purpose of processing, the recipient of the data and, if necessary, further information to guarantee fair processing in respect of the data subject. [20]
Personal data can only be processed if the controller complies with the requirements stated in Article 7 and 12. Thus, explicit consent of the data subject is indispensable for the performance of contractual obligations or the entering into a contract.

Legitimate Purpose Personal data shall only be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes", Article 6 (b).

Proportionality Personal data may only be processed if the processing is "adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed", Article 6. This processing has to be carried out "fairly and lawfully". Furthermore, the collected data must be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified".
Moreover, the Directive demands the controller to "keep [the data] in a form which permits identification of data subjects for no longer than is necessary for the purposes of which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use", Article 6 Par. 1e.
Finally, the directive tightens the requirements for specific sensitive personal data regarding "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and [...]

[20] Data subject is the official notion used by the Data Protection Directive, referring to the individual being affected by data processing.

data concerning health or sex life". The processing of this kind of data may only be justified if the requirements stated in Article 8 Par. 2 are fulfilled such as a specific consent or protecting the vital interests of the data subject.

2.1.2 The Proposal for a General Data Protection Regulation

The General Data Protection Regulation (GDPR) is a proposed Regulation of the European Union, which aims to harmonize the rules regarding the processing of personal data mainly by private companies across the EU - Whether the Regulation should also touch other sectors such as tax authorities, social security, etc., is currently subject to heavy debate. The outdated [21] Data Protection Directive 95/46/EC from 1995 was intended to encourage the free movement of personal data within Europe by harmonizing national provisions on data protection. [22] However, the scope of implementation of the Directive led to different interpretations of the national data protection laws and to a minimum standard. [23] Hence, the proposed Regulation is to ensure a uniform standard of data protection [24], the protection of personal data and the free movement of such data within the European Union. On October 21st 2013, the European Parliaments LIBE Committee (Committee for Civil Liberties, Justice and Home Affairs) adopted a number of proposed changes to the General Data Protection Regulation published by the EU Commission on January 25th 2012. [25] On October 22nd 2013, the Home Affairs Committee of the European Parliament launched o the start of negotiations with the European Commission, and the Council of the European Union the so-called trilogue. On March 12th 2014 the European Parliament adopted a legislative resolution based on the proposal after the first reading in the parliament, adopting the LIBE Committees changes to the original proposal. [26]
On June 15th 2015, the Council of the European Union presented a new proposal for the General Data Protection Regulation with several changes and amendments, which led to the beginning of a series of trilogue-negotiations. [27] The European Union expects to enact the Regulation by the end of 2015. [28]
The proposal for a General Data Protection Regulation maintains the main principle of the directive 95/46/EC to generally prohibit the processing of personal data, unless the person affected has given

[21] *Tene*, International Data Privacy Law 2011, 15 (15); *Hon/Millard*, Data Export in Cloud Computing How can Personal Data be Transferred outside the EEA?, The Cloud of Unknowing, Part 4, p. 2; *Sartor*, International Data Privacy Law 2013, 3 (3).

[22] *Hon/Millard/Walden*, The Problem of 'Personal Data' in Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part. 1, p. 4;*Leonard*, International Data Privacy Law, 2014, 53 (53).

[23] *Klar*, ZD 2013, 109 (109 ff.); While one could have understood the *Lindqvist- decision* of the ECJ (of 06/11/2003 C-101/91) in the way, that the Directive 95/46/EC requires only minimum standards of the Member States, it is obviously after the *ASNEF-decision* (24/11/2011- C-468/10), that the conditions of admissibility of the data handling are already largely fully harmonized.

[24] *Eckhardt/Kramer/Mester*, DuD 2013, 623 (630).

[25] Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) in the version adopted by the European Parliament after the LIBE-Committee's vote, available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A7-2013-0402+0+DOC+XML+V0//EN; *Heinemeyer*, Verfahrensstand-Anzeiger; *Härting*, CR 2013, 715 (715 ff.).

[26] European Parliament, legislative resolution of 12/03/2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 0011-C7-0025/2012-2012/0011 (COD) (Ordinary legislative procedure: first reading), available at: http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7-TA-2014-0212&language=EN.

[27] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) of 15/06/2015 ST 9565 2015 INIT, available at: http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.

[28] A timetable for the reform is available at: http://www.eppgroup.eu/de/news/Data-protection-reform-timetable.

their consent or legal permission.

2.1.2.1 Difference between a Directive and a Regulation

Regulations are passed either jointly by the EU Council and European Parliament, or by the Commission alone [29] and are the most direct form of EU law - as soon as they are passed, they have binding legal force throughout every Member State, with the same effects as national laws and eventually overruling them. National governments do not have to take action themselves to implement EU regulations. Art 288 of the Treaty on the Functioning of the European Union defines a regulation as

> "Article 288 (ex Article 249 TEC): [...] A regulation shall have general application. It shall be binding in its entirety and directly applicable in all Member States. [...]"

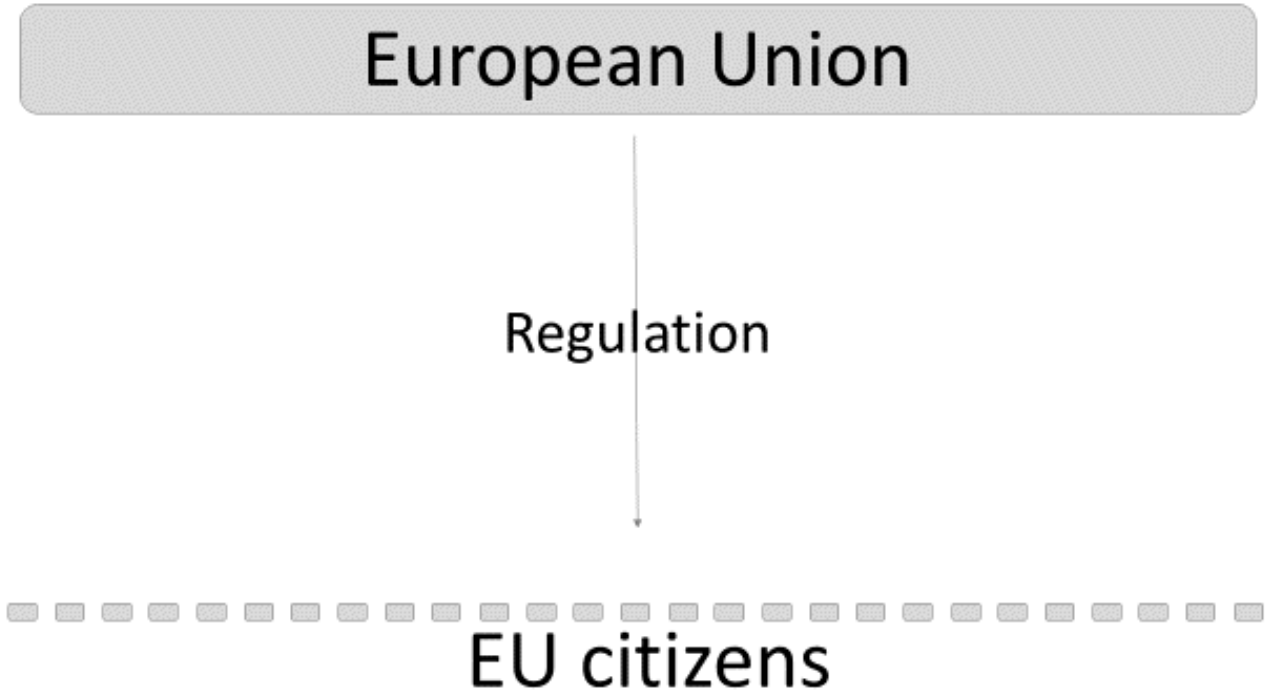

Figure 2.3: Applicability of the Data Protection Directive

The proposed data protection Regulation will therefore be directly binding without being in need of a national act of transposition. This is an important difference between the current Directive and the proposed Regulation, since the directive had to be implemented into the national laws by the governments of Member States.

2.1.2.2 Territorial Scope of the GDPR

The territorial scope of the Regulation is specified in three cases, Article 3 Par. 1 3 [30]:

> "1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union.[31]

[29] *Wieczorek*, DuD 2013, 644 (646).
[30] *Wieczorek*, DuD 2013, 644 (646).
[31] The LIBE Proposal furthermore includes the clause: "whether the processing takes place in the Union or not".

> 2. This Regulation applies to the processing of personal data of data subjects in the Union by a controller or processor not established in the Union, where the processing activities are related to:
> (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
> (b) the monitoring of such data subjects. [32]
> 3. This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where the national law of a Member State applies by virtue of public international law."

Hence, many data processing operations by providers of services outside the European Union would thus fall into the scope of the European data protection law. The (proposed) recitals 19 and 20 highlight these intentions: [33]

> "(19) Any processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union should be carried out in accordance with this Regulation, regardless of whether the processing itself takes place within the Union or not. Establishment implies the effective and real exercise of activity through stable arrangements. The legal form of such arrangements, whether through a branch or a subsidiary with a legal personality, is not the determining factor in this respect.
>
> (20) In order to ensure that individuals are not deprived of the protection to which they are entitled under this Regulation, the processing of personal data of data subjects residing in the Union by a controller not established in the Union should be subject to this Regulation where the processing activities are related to the offering of goods or services to such data subjects, or to the monitoring of the behaviour of such data subjects." (Wording of the proposal of the Commission) The LIBE-Proposal adds to recital 20 that the Regulation should be applicable "where the processing activities are related to the offering of goods or services, irrespective of whether connected to a payment or not, to such data subjects, or to the monitoring of such data subjects" and states that "in order to determine whether such a controller is offering goods or services to such data subjects in the Union, it should be ascertained whether it is apparent that the controller is envisaging the offering of services to data subjects residing in one or more Member States in the Union."

The Proposal of the Council adds several requirements to recital 20 in order to determine whether a controller is offering goods or services to data subjects in the Union:

> "In order to ensure that individuals are not deprived of the protection to which they are entitled under this Regulation, the processing of personal data of data subjects residing in the Union by a controller not established in the Union should be subject to this Regulation where the processing activities are related to the offering of goods or services to such data subjects irrespective of whether connected to a payment or not, which takes place in the Union. In order to determine whether such a controller is offering goods or services to such data subjects in the Union, it should be ascertained whether it is apparent that the controller is envisaging doing business with data subjects residing in one or more Member States in the Union. Whereas the mere accessibility of the controller's or an intermediary's website in the Union or of an email address and of other contact details or the

[32] The Proposal of the Council adds: "as far as their behaviour takes place within the European Union".

[33] Cf. LIBE-Proposal, available at: http://www.janalbrecht.eu/fileadmin/material/Dokumente/DPR-Regulation-inofficial-consolidated-LIBE.pdf.

> use of a language generally used in the third country where the controller is established, is insufficient to ascertain such intention, **factors such as the use of a language or a currency generally used in one or more Member States with the possibility of ordering goods and services in that other language, and/or the mentioning of customers or users residing in the Union, may make it apparent that the controller envisages offering goods or services to such data subjects in the Union."**

2 The concept of services is governed by Article 57 TFEU (freedom to provide services) or by Article 4 No. 1 of the Services Directive 2006/123/EC. [34] Services are all activities covered under Article 57 TFEU, which are normally provided for remuneration, insofar as they are not subject to the rules on free movement of goods, capital and on the free movement of the person. By making it clear in the definition of the regulation, the service does not have to be paid for, commercial and non-commercial websites are covered.
The definition of goods is governed by Article 28 Par. 2 TFEU. Regardless of the nature of the transactions, this is a set of objects which can be, in respect of commercial transactions, brought across a boundary. [35] These goods do not need to be physical, but have a market value.
When the processing operation of the observation of the behavior of the person affected occurs, according to recital 21 the Article 3 Par. 2 (b) applies.

> "(21) The processing of personal data of data subjects residing in the Union by a controller not established in the Union should also be subject to this Regulation when it is related to the monitoring of their behaviour taking place within the European Union. In order to determine whether a processing activity can be considered to 'monitor' data subjects, it should be ascertained whether individuals are tracked, regardless of the origins of the data, or if other data about them is collected, including from public registers and announcements in the Union that are accessible from outside of the Union, including with the intention to use, or potential of subsequent use of data processing techniques which consist of applying a 'profile', particularly in order to take decisions concerning her or him or for analyzing or predicting her or his personal preferences, behaviors and attitudes." (wording of the Proposal of the Council)

For example, when Internet activities are tracked by means of data processing techniques by which a person is assigned to a profile, tracking-tools which operate on the use of cookies [36] for targeted advertising are particularly affected.[37] Due to the altered wording of 'monitoring' in Article 3 par 2 (b), a selective observation is not covered.
The Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where the national law of a Member State applies by virtue of public international law according to Article 3 Par. 3. Pursuant to recital 22 this affects diplomatic places such as embassies or consulates. [38]
Hence, the former territorial principle of Article 4 of the Data protection Directive 95/46/EG has been abandoned in favor of a more market- and user-orientated model. [39] This very broad territorial scope

[34] *Wieczorek*, DuD 2013, 644 (647); Klar, ZD 2013, 109 (113); Treaty on the Functioning of the European Union, available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=EN; Directive 2006/123/EC of the European Parliament and of the Council of 12/12/2006 on services in the internal market, available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0123&from=EN.

[35] *ECJ*, decision of 09/07/1992 C-2/90, recital 26.

[36] *Art. 29-Working Party*, Opinion 04/2012, WP 194, 1 ff.

[37] *Peifer*, K&R 2011, 543 (543 ff.); *Rammos*, K&R 2011, 692 (692 ff.); *Klar*, ZD 2013, 109 (113).

[38] *Art. 29-Working Party*, Opinion 08/2010, WP 179, 22 ff.; *Wieczorek*, DuD 2013, 644 (648).

[39] *Härting*, BB 2012, 459 (462); *Piltz*, Datenschutzreform: aktueller Stand der Verhandlungen im Rat, 20/01/2014.

of the proposal has the potential to strengthen the protection of European citizens' rights, since the provider of services or goods is bound to European data protection law irrespective where they are established. For Cloud Computing this might lead two different outcomes, depending on how many parties are involved. If a cloud provider from a non-EU/EEA country offers their services directly to the data subjects inside the EU/EEA in a business-to-consumer-relationship, they will be governed by European data protection law. However, if the cloud provider offers their cloud services to a company in a business-to-business-relationship which uses the services to 'process' its customers data, the cloud provider is not considered to offer services or goods directly to the data subjects (the company's customers). [40]
Due to the GDPRs broad claim of applicability and the fact that citizens no longer have to consider the location of the processors servers[41], the affected person might be more successful in regards to asserting his or her rights. [42]
However, this approach may go significantly beyond what could be considered realistically enforceable: A researcher established outside the Union could monitor - among others - EU-citizens internet activities (even if their website is not even supposed to target EU-citizens), and therefore be governed by European data protection law without even being aware of it. [43] Moreover, it is improbable that the EU could enforce data protection standards to providers situated outside the Union or who do not have any business within the EU. European supervisory authorities are not able to act outside the Union. Article 51 GDPR only states that:

> "1. Each supervisory authority shall be competent to perform the duties and to exercise the powers conferred on it in accordance with this regulation on the territory of its own Member State [. . .]"

Thus afr, there is no solution to this problem. [44] Although Article 25 GDPR states that a controller outside the Union that is affected by its data protection law shall designate a representative in the Union, there are no possibilities for sanctions or measures against such controllers in the GDPR; this has been criticized by the former German Federal Minister of Justice Sabine Leutheusser-Schnarrenberger. [45]

2.1.2.3 Material Scope of the GDPR

The proposed General Data Protection Regulation does not abandon the basic principles of the DPD. For example, the principle of prohibition with reservation of authorisation in the data protection law has not been weakened in the proposal; [46] on the contrary, it has been enhanced pursuant to Article 6 GDPR. [47] The processing of personal data shall be, as regulated in the DPD, lawful only if the data subject has given consent in accordance with Article 7 GDPR (see 2.4.2.3.2) to the processing of their personal data or if after consideration the processing is necessary for legal purposes. The permissions contained in Article 6 GDPR are more general and unspecific [48] which, as a consequence, leads to an increased importance of consent. [49] The scope of the proposal for the GDPR is, as well as in the DPD, defined by the processing of personal data (Article 2 Par. 1 in conjunction with Article 4 Par. 2,

[40] *Hornung/Sädtler*, CR 2012, 638 (640).
[41] *Roßnagel/Richter/Nebel*, ZD 2013, 103 (104).
[42] *Nebel/Richter*, ZD 2012, 407 (410).
[43] *Spindler*, GRUR 2013, 996 (1003); Spindler, GRUR-Beilage 2014, 101 (107).
[44] *Hornung/Sädtler*, CR 2012, 638 (640).
[45] *Leutheusser-Schnarrenberger*, MMR 2012, 709 (710).
[46] *Taeger* in: Taeger/Gabel, BDSG, par. 4a, Recital 4.
[47] *Härting*, CR 2013, 715 (717).
[48] *Roßnagel/Richter/Nebel*, ZD 2013, 103 (104).
[49] *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, Recital 13.

Article 1 GDPR, see in detail 2.2.3). Moreover, according to Article 2 Par. 2 lit.d GDPR, in general the applicability remains restricted to processing of personal data outside the private sphere. Thus the Regulation 'does not apply to the processing of personal data by a natural person within the context of an activity that is exclusively personal or in the domain of the household', which implies that the actual cohabitation of the persons is affected and not their relationship in terms of family law matters. [50]

2.1.2.4 Fundamentals of the GDPR

As seen in 2.1.1.2.2 for the DPD, the proposals for a GDPR also require the controller to comply with the provisions regarding transparency, legitimate purpose and proportionality.

2.1.2.4.1 Transparency

Article 14 Par. 1 GDPR provides additional information regaridng the data subjects in comparison to the transparency provisions of Art. 10 and 11 DPD. [51] It includes absolute obligations to inform the data subject about, e.g. the identity of the data protection officer, the period for which the personal data will be stored, the existence of the right to request from the controller access to and rectification or erasure of the personal data as well as the right to lodge a complaint to the supervisory authority (see in detail 2.4.2.3.2). Moreover, the LIBE-Proposal introduces in Article 13a new standardised information policies and the GDPR extends the general information obligations in Article 32 to a specific communication to the data subject in case of a personal data breach (see 2.5.5). Furthermore, according to Article 28 each controller and processor shall maintain documentation of all processing operations under its responsibility.

2.1.2.4.2 Legitimate Purpose and Proportionality

Article 5 lit. (b) of the Commission's proposal and of the LIBE-Proposal includes provisions regarding the purpose limitation, according to which personal data shall be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes." The proposal of the Council adds that "further processing of personal data for archiving purposes in the public interest or scientific, statistical or historical purposes shall in accordance with Article 83 not be considered incompatible with the initial purposes." The Commission's and the LIBE-Proposal state in Article 5 Par. 1 lit. c that processing of personal data shall be "limited to the minimum necessary", whereas similar to the DPD's provisions, the proposal of the Council stipulates that personal data must not be processed excessively in relation to the purposes for which they are processed. [52] Furthermore, Article 6 Par. 4 Sentence 2 of the Council's proposal relaxes the principle of purpose limitation by stating that "further processing by the same controller for incompatible purposes on grounds of legitimate interests of that controller or a third party shall be lawful if these interests override the interests of the data subject." At first sight, this could be interpreted as a relief for companies that run big data analyses, however, according to Article 14 Par. 1b and Article 14a Par. 3a this would be very difficult in practice as the controller has to inform the data subject prior to further processing with information on other legitimate purposes as well as any additional relevant information.
Moreover, Article 2 par. 2 lit. d includes exceptions for natural persons in the course of personal

[50] *Dammann*, in: Simitis, BDSG, par. 1, Recital 243; moreover, the LIBE-Proposal expands this exception in sentence 2 of Article 2 Par. 2 lit.d GDPR to "publication of personal data where it can be reasonably expected that it will be only accessed by a limited number of persons."

[51] *Härting*, Internetrecht, Recital 369.

[52] *Leucker*, PinG 2015, 195 (198).

or household activities, which however, according to recital 15, should not apply to controllers or processors which provide the means for processing personal data for those activities.
'Profiling' is regulated in Article 20 of the proposals for a GDPR. According to the proposal of the Commission "every natural person shall have the right not to be subject to a measure which produces legal effects concerning this natural person or significantly affects this natural person, and which is based solely on automated processing intended to evaluate certain personal aspects relating to this natural person or to analyse or predict in particular the natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour." According to Par. 2 it shall only be lawful if the processing is carried out in the context of a contract, if it is expressly authorized by a Union or Member State law or if it is based on the data subject's consent. The LIBE-Proposal states that every natural person shall have the right to object to profiling and that the data subject shall be informed about the right to object to profiling in a highly visible manner. Similar to the Commission's proposal, the proposal of the Council requires that the profiling "produces legal effects concerning him or her or significantly affects him or her." Nevertheless, it will be difficult to comply with the GDPR in terms of profiling and scoring. [53]
Finally, recital 38a of the proposal of the Council provides an intra group exemption for companies by stating that "controllers that are part of a group of undertakings or institution affiliated to a central body may have a legitimate interest to transmit personal data within the group of undertakings for internal administrative purposes.

2.2 Personal Data and Encryption

The European data protection law only applies if 'personal data' is processed. Because of that it is very important to understand what data qualifies as personal data. Depending on how 'personal data' is defined, the effect a valid encryption of this data takes, might be different.

2.2.1 Personal Data and Encryption under the DPD

In the following we assume that a data controller, i.e. cloud-computing client, holds information about data subjects and wants this information to be stored in a cloud computing environment.
In the center of any consideration concerning cloud-based information processing is the definition of 'personal data' provided by the Data Protection Directive (DPD). Information that is not, or ceases to be, 'personal data', may be processed, in the cloud or otherwise, and is therefore not affected by data protection law requirements.
Thus, if the information held by the data controller is considered to be personal data "in the cloud" in terms of data protection, such cloud-computing operations (e.g. storing and processing in the cloud) would normally fall under the respective national data protection acts or within the scope of the DPD. [54] In cloud computing, the 'personal data' definitional issue is crucial with respect to anonymized, pseudonymized and encrypted data. Concerning encrypted data be it encrypted while in transmission, storage or computations – one of the issues refers to the problem of whether it still qualifies as personal data.

[53] C.f. *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 24; see also recitals 58 and 58a.

[54] Directive 95/46/EC of the European Parliament and of the Council of 24/10/1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML; German Federal Data Protection Act, available at: http://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.html#p0061.

2.2.1.1 Personal Data and Cloud Computing

As already outlined, the characteristics of personal data is crucial for the application of the Data Protection Directive. Hence, we have to take a closer look on the criteria for assessing these characteristics:

2.2.1.2 Article 2 (a) Data Protection Directive

According to Article 2 (a) of the Data Protection Directive 'personal data' shall mean any information relating to an identified or identifiable person ('data subject') [55]; an identifiable person is one who can be identified, directly or indirectly, in particular by referencing an identification number, or to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity. Whether or not information qualifies as 'personal data', it depends on the circumstances in each individual case. For instance, a common family name may not single someone out within a country but is likely to identify a pupil in a classroom. Moreover, if the data processing controller is able to combine information with other data in order to identify individuals, then the information that was originally considered 'personal data' may change.

2.2.1.3 Recital 26 Data Protection Directive

Recital 26 of the Directive renders more precisely the notion of 'personal data':

> *"Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible."*

Thus, leaving aside apparently 'non-personal' information, indeed, recital 26 of the DPD explicitly recognizes that information constituting 'personal data' may be rendered 'anonymous'. Therefore the data can be used freely by data controllers/operators such as cloud computing operators, if it is being anonymized. Moreover, the transmission of data may fall outside of the scope of the DPD if the data no longer qualifies as personal data; otherwise, the data subject's consent is needed (see 2.4.2). This raises an important legal question: Is the cloud-computing provider considered to be a data 'processor' *who processes personal data* on behalf of the controller (see 2.3), i.e. the cloud-computing client? Unfortunately, recital 26 of the DPD is prone to various interpretations and thus there is no clear answer. [56]

2.2.1.4 Relative or Absolute Identifiability of Persons

The criteria concerning the *identifiability* of persons required by Article 2 (a) DPD are still subject to debate, in particular, whether or not a so-called *absolute* or *relative approach* has to be the basis for

[55] *Kokott/Sobotta*, International Data Privacy Law 2013, 222 (223); CJEU, Joined Cases C92/09 and C93/09 Volker und Markus Schecke and Eifert [2010] ECR I-11063, par. 52, 53 and 87.

[56] *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing What Information is Regulated? - The Cloud of Unknowing, Part 1, p. 13

assessing controller's abilities to identify a person. [57]
In short, the "absolute approach" determines all possibilities and chances in which the data controller would be able to identify the data subject individually. Thus, all ways and means for a data controller without any regard to expenses etc. are taken into account. Even theoretical chances of combining data so that the individual is identifiable are included. If identifiability is assessed *absolutely*, then it is sufficient for the application of personal data acts if anyone in the world is able to decrypt or decode the encrypted data. [58] Applied to cloud computing, as long as anyone in the world is able to decrypt the data set, the operations of the cloud computing provider are subject to data protection legislation, even if the cloud computing provider does not possess the key for decryption. Based on this approach data protection legislation is applicable, regardless of the applied encryption technique, as long as one entity holds the key for decoding.
In contrast, the "relative approach" considers the relevance of the necessary effort required by the data controller to identify the data subject. [59] Therefore, only realistic chances of combining data in order to identify an individual are taken into account. With regards to encryption issues, data protection legislation is only applicable if the data controller is able to decrypt a certain data set [60] or, at least, has reasonable chances of obtaining the decrypting key.
In the case law of some courts, the trend is beginning to lean towards favoring a relative understanding . [61] In contrast, some national authorities - for instance, the so-called *Düsseldorfer Kreis* (a committee consisting of all German federal and regional supervisory authorities) support the absolute approach [62] , as well as some other academics. [63] Despite its enormous practical impact, this aspect has not been clarified yet, neither by the ECJ [64] nor by the European Commission. However, in October 2014 the German Federal Court of Justice (BGH) requested the ECJ [65] for a preliminary ruling in accordance with Article 267 TFEU on the interpretation of the dispute regarding whether a dynamic IP address can be considered as personal data. [66] The German Federal Court of Justice states in its request to the ECJ that a relative approach could be in accordance with recital 26 of the DPD, according to which "to determine whether a person is identifiable, account should be taken of all the

[57] *Bergt*, ZD 2015, 365 (365 ff.) provides an up-to-date summary of the different opinions.

[58] *Art. 29-Working Party*, Opinion 04/2007, 7; OLG Hamburg, MMR 2008, 687 (688); *Nink/Pohle*, MMR 2015, 563 (565) which criticize that consequently this approach would lead to the result that there would virtually be no more anonymous data; *Pahlen-Brandt*, DuD 2008, 34 (38).

[59] *Dammann* in: Simitis, BDSG, par. 3, recital 32; *Gola/Klug/Körfferq*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 10; *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 115, 116; *Schulz* in: Beck'scher Kommentar zum Recht der Telemediendienste, par. 11 TMG, recital 24; *Roßnagel/Scholz*, MMR 2000, 721 (723); *Meyerdierks*, MMR 2009, 8 (8 ff.); *Eckhardt*, K&R 2007, 601 (603); *Voigt*, MMR 2009, 377 (377); *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 46.

[60] *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 115, 116.

[61] *England and Wales High Court (Administrative Court)*, [2011] EWHC 1430 (Admin), Case No. CO/12544/2009, recital 51 f.; *Upper Tribunal (Administrative Appeals Chamber)*, [2011] UKUT 153 (AAC), Appeal Number: GI/150/2011, GI/151/2011, GI/152/2011, recital 128; *House of Lords*, [2008] UKHL 47, recital 27; *AG München*, ZUM-RD 2009, 413 (414) = BeckRS 2008, 23037; *OLG Hamburg*, MMR 2011, 281; *LG Wuppertal*, MMR 2011, 65 (66); *LG Berlin*, CR 2013, 471; different point of view *AG Berlin-Mitte*, ZUM 2008, 83 = K&R 2007, 600 (601); *VG Wiesbaden*, MMR 2009, 428 (432).

[62] http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DuesseldorferKreis/Nov09Reichweitenmessung.pdf?__blob=publicationFile.

[63] *Brennscheidt*, Cloud Computing, p. 51; *Kuner*, European Data Protection Law, p. 92; *Marnau/Schlehahn*, Cloud-Computing: Legal Analysis, TClouds (D 1.2.2), p. 26 f.; *Pahlen-Brandt*, DuD 2008, 34 ff; *Weichert*, in: Däubler/Klebe/Wedde/Weichert, BDSG, par. 3, recital 13, 15.

[64] *Kühling/Klar*, NJW 2013, 3611 (3614).

[65] *ECJ*, Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland.

[66] *German Federal Court of Justice (BGH)*, decision of 28/10/2014 - VI ZR 135/13 = MMR 2015, 131.

means likely reasonably to be used to identify the said person" (recital 25).[67] The court also states in recital 28 that the "wording of the provision of the directive appears to be ambiguous" and argues that even if means shall be taken into account that can be used by a third party to identify the person, a relative approach of the identifiability of the person affected would be possible if means are only taken into account that could realistically be used.
However, the general opinion of the European Commission and of several Member States regarding this case tend to veer towards an absolute approach. [68] Thus, the ECJ will have to resolve the dispute between an absolute or relative approach regarding IP-addresses by interpreting Article 2 (a) DPD and especially recital 26. [69] Its decision will certainly have a major influence on the handling of data on the internet and will have a huge impact on the general interpretation of defining 'identifiability'. [70]

The position of the Article 29 Data Protection Working Party [71] describes its stance concerning Art. 2 (a) DPD as follows:

> *" 'Anonymous data' in the sense of the Directive can be defined as any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking account of all the means likely reasonably to be used either by the controller or by any other person to identify that individual. 'Anonymized data' would therefore be anonymous data that previously referred to an identifiable person, but where that identification is no longer possible. Recital 26 also refers to this concept when it reads that 'the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer 'identifiable'. Again, the assessment of whether the data allows identification of an individual, and whether the information can be considered as anonymous or not depends on the circumstances, and a case-by-case analysis should be carried out with particular reference to the extent that the means are likely reasonably to be used for identification as described in Recital 26. This is particularly relevant in the case of statistical information, where despite the fact that the information may be presented as aggregated data, the original sample is not sufficiently large and other pieces of information may enable the identification of individuals."* [72]

This opinion of the Working Party is interpreted by some authors as being cryptic as the Working Party has used indirectly similar notions for other cases (mentioned in the opinion) which are implicitly favorable to the relative approach. [73] Others argue that the opinion includes a rather absolute stance. [74]

Indeed, if one takes a closer look at the Working Party's statement, it should be noted that it considers not only the means potentially used by the controller to identify the data subject but also the means that might be used by third parties. [75] This instance can be regarded as an indication for an absolute

[67] C.f. *Brink/Eckhardt*, ZD 2015, 205 (209).
[68] *Bergt*, IP-Adressen: EU-Kommission gibt BGH Nachhilfe in Sachen Grundrechte, available at: http://www.cr-online.de/blog/2015/09/13/ip-adressen-eu-kommission-gibt-bgh-nachhilfe-in-sachen-grundrechte/.
[69] *German Federal Court of Justice (BGH)*, decision of 28/10/2014 - VI ZR 135/13 = MMR 2015, 131 (132 f.), recitals 27, 29 ff.
[70] *Bär*, MMR 2015, 134 (135 f.); *Nink/Pohle*, MMR 2015, 563 (564).
[71] http://ec.europa.eu/justice/data-protection/article-29/index_en.htm.
[72] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 21; see also *Leonard*, International Data Privacy Law, 2014, 53.
[73] Cf. criticism of *Kühling/Klar*, NJW 2013, 3611 (3614); *Pahlen-Brandt*, DuD 2008, 34 f.
[74] Cf. *Eckhardt*, CR 2011, 339 (341, 343); *Stimerling/Hartung*, CR 2012, 60 (63).
[75] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 18 f.

approach. However, the Working Party seemingly recognizes situations in which a set of data should be regarded as *personal* data with respect to one entity but not with respect to another one [76], which, in turn, implies a relative approach. The reason for this apparent contradiction is that the Working Party puts emphasis on the circumstances of the particular situation of the processing action rather than on the personal perspective (thus *whose* capacities have to be considered: Only the ones of the controller or of any other person in the world?). Hence, on the one hand, the assessment of the data has to take into account means for identification that can be used by the controller or any other third party [77] but, on the other hand, these means are limited to those which are reasonably likely to be used *in a concrete situation*. Simply theoretical chances of identification are insufficient to constitute the personal-characteristic of the data. [78] The results of this opinion should, in many cases especially with respect to encryption technologies be similar to the relative approach, since both consider only realistic chances to identify the data subject.
Moreover, singular indications of a relative approach can be found in the legislation of some EU-Member States (in particular, Great Britain and Austria). The British Data Protection Act of 1998 expressly focusses in Part I, 1 on information that is or is likely to come in the possession of the data controller in order to assess the identifiability: [79]

> *" 'personal data' means data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the* **data controller***, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual"*

This definition clearly differs from the provided formulations in Art. 2 (a) DPD and recital 26 of the DPD by taking (expressly) only the perspective of the controller. [80] One may note this instance while assessing British court decisions (such as the ones referred to above). Hence, the risk of inconformity with a Directive arises out of the provision in case the absolute approach prevails. Furthermore, within the EU Member States, it does not appear to be the 'usual' practice to implement the DPD requirements of the term 'personal data' by expressly focusing on the controller's perspective only (which can be interpreted as a sign for a relative understanding). [81] Therefore, a general stance of national legislators in the EU that are in favor of a relative approach to interpret the term 'personal data' within the DPD cannot be determined from those single provisions.

A remarkable gradation was stated in the Austrian data protection law in par. 4 No. 1 DSG 2000: [82]

> " 'Data' ('Personal Data"): Information relating to data subjects (sub-par. 3) who are identified or identifiable; Data are "only **indirectly personal**" for a controller (sub-par. 4), a processor (sub-par. 5) or recipient of a transmission (sub-par. 12) when the Data relate to the subject in such a manner that **the controller**, processor or recipient of a transmission cannot establish the **identity** of the data subject by legal means"

[76] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 15 f.
[77] Cf. also *Bygrave*, Data Privacy Law, p. 132
[78] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 15
[79] Cf. *Kuner*, European Data Protection Law, p. 95 f.
[80] Cf. *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 19, recital 97.
[81] Cf. List of provision formulations in *Kuner*, European Data Protection Law, p. 95 f.
[82] Austrian data protection act from 2000, BGBl. I Nr. 165/1999, last amendment 23/05/2013, BGBl. I Nr. 165/1999, english version available at: https://www.dsb.gv.at/DocView.axd?CobId=41936.

The Austrian law ostensibly combines the relative and a rather absolute approach. With respect to the cited provision, data generally has to be rendered "personal" if the controller or any other person is capable of identifying the data subject (which indicates an absolute understanding). [83] However, whenever the controllers themselves cannot identify the data subject by using lawful and reasonable means, all processing actions done by them are privileged in many provisions. [84] This special category is called "indirectly personal" data by Austrian law.

In other words: As long as identifiability can be denied on the basis of a relative approach, the Austrian data protection act is applicable, but with less strict requirements (with respect to the particular controller). For instance, the transmission of such "indirectly personal" data into third countries does not require a permission by the data protection authority (par. 12 section 2 No. 2 of the Austrian data protection act). Nevertheless it should be stressed that the DPD does not provide such a subcategory within the category of personal data; there is no differentiation between data that allow a direct identification of the data subject and those indirectly doing so. Both cases expressly constitute (one category of) personal data (see Art. 2 (a) DPD). [85]
In order to avoid conflicts with the DPD (and constitutional law), there are trends to reduce the scope of the category of "indirectly personal" data in Austria by using a very restrictive interpretation of that term. [86]
Regarding a cloud computing scenario where encryption technology is used, the Austrian law could consider the processed information "indirectly personal" data relating to the controller, if the encryption has a sufficient level of security and the controller has at least no realistic chance to obtain the decryption-key (by lawful means). So the data would not fall outside the scope of data protection law, completely, but only a reduced level of data protection provisions would be imposed by the controller. However, it is argued by some Austrian authors that even (securely) encrypted data does not render them "indirectly personal" even though encryption might be a typical example for data from which a controller - who does not hold the decryption key - cannot identify the data subject. [87] As a consequence, even this kind of data would be (directly) personal data and hence the data protection act would apply comprehensively without any privilege. In sum, one should not generalize the Austrian law approach, since it is on the one hand based upon a unique interpretation of the DPD assuming differences between a direct and an indirect identifiability and is on the other hand subject to a controversy regarding the actual scope of the term "indirectly personal" data.

In this respect the second dispute concerns the technical demands to the level of encryption. Thus, the question is: which technical level of encryption has to be reached in order to assume that the *reconstruction/decryption* and *de-anonymization* of personal information/data is impossible do we need *absolute (theoretical) security* or is *state-of-the-art security* sufficient? [88] To put it simply, the current question in the legal debate is: what level of *encryption* or *anonymization* must be achieved to avoid the applicability of the data protection law?

2.2.1.4.1 The Impact of the Absolute Approach upon Cloud Computing and Encryption

The absolute approach is a radical perspective that widens scope of Data Protection. The DPD would not be applicable only if there is no (*theoretical*) chance for the cloud-computing provider to re-

[83] *Pollirer/Weiss/Knyrim*, Datenschutzgesetz 2000, par. 4, recital relating to Z 1, p. 20.
[84] *Pollirer/Weiss/Knyrim*, Datenschutzgesetz 2000, par. 4, recital relating to Z 1, p. 20. f.
[85] Cf. *Bergauer*, Jahrbuch Datenschutzrecht 2011, 55 (60).
[86] Cf. *Bergauer*, Jahrbuch Datenschutzrecht 2011, 55 (57 f.).
[87] Cf. *Bergauer*, Jahrbuch Datenschutzrecht 2011, 55 (62).
[88] This distinction is made, for instance, in nuclear law and other laws referring to 'dangerous' technologies.

combine the data in order to identify the data subject. In particular, encryption of data would not change the basic character of personal data, itself, only render it more difficult for unauthorized people to access. [89] Hence, from a radical stance, encryption makes it more difficult to identify and "read" the personal data; however, it does not exclude the *theoretical* outcome of obtaining a key and access to the data. Thus, from this perspective, encryption is considered more of a *technical security measure* to ensure that data is not accessible to unauthorized persons rather than changing the quality of data (in contrast to anonymizing it). Even if the encrypted data is being used, for instance in calculations, and if the data does not lose (in the decrypted version) the personal references, the DPD would be applicable, as the cloud computing provider would still (*theoretically* able to decrypt it. [90]
Alternatively, supposing that the data controller has key-coded/encrypted the original personal data (changed names to code numbers, with a 'key' showing which number corresponds to which name), destroyed the original personal data, but still possesses the key, then individuals can be identified from the key-coded data when used in combination with the key. If encryption were applied to a data set, the whole data set would be transformed – not just names within the data set. However, where the data controller possesses the decryption key, encrypted personal data might be revealed in a similar way to key-coded data. If so, it would still be considered as 'personal data'.

2.2.1.4.2 The Impact of the Relative Approach on Cloud Computing and Encryption

As outlined above, the relative approach concentrates on the reasonable means for a data controller to identify the data subject and to get access to the personal data. If neither the cloud provider nor the cloud computing client (the data controller) keeps a master key to the respective data or data set of the customer, no personal data or information could be considered as processed or transferred abroad, since no one could decrypt this data except the key holder. [91] Only the data subject who exclusively has the key could decode the data. Hence, consent to such processing or transfer is not required because no personal information would be implied during the process neither at the very beginning, when the data is being transferred to the cloud computing client, nor afterwards, when transferred to the cloud computing provider. Furthermore, if only the cloud user holds the decryption key and not the provider the data cannot be rendered "personal", since the provider is not capable of decrypting the data and identifying the data subject. The user on the other side still processes personal data, since the assessment can vary depending on the particular person in question. [92]
In other words, the relative approach focuses on reasonable terms by which a provider may identify the data subject, particularly if it would be economically (and legally [93]) feasible. [94] As possibilities/capacities of providers and their economic interest may vary widely, in general terms what qualifies as a reasonable effort to de-anonymize cannot be assessed. Thus, data is not "personal data" anymore in the sense of the DPD if the reference to individuals can, at least, under regular condi-

[89] Orientierungshilfe Cloud Computing, Version 2.0, p. 12, which states that '*regularly* data does not lose its character as personal data by encryption'; although, '*regularly*' means that there are special cases where encryption can have the effect, that no personal reference exists, c.f. *Eckhardt*, DuD 2015, 176 (179 f.).

[90] C.f. *Brennscheidt*, Cloud Computing, p. 52 f; *Nink/Pohle*, MMR 2015, 563 (566).

[91] *Hon/Millard/Walden*, The Problem of 'Personal Data' in Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part. 1, p. 28; *Frauenhofer Institut*, Cloud-Computing für die öffentliche Verwaltung, p. 116; *Spies*, MMR-Aktuell 2011, 313727

[92] Cf. *Hon/Millard/Walden*, The Problem of 'Personal Data' in Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part. 1, p. 25.

[93] *Nink/Pohle*, MMR 2015, 563 (565); *Spindler/Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, par. 11 TMG recital 8, 11; different opinion: *Bergt*, ZD 2015, 83 (85). For instance, non-disclosure provisions or secrecy legislation may impede any re-combination of data between different providers. The supporters of the absolute approach negate these barriers.

[94] *Gola/Klug/Körffer* , in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 44.

tions, no longer be reconstructed, i.e. a decryption or de-anonymization is almost impossible. [95] If, under usual conditions, de-anonymization can be regarded as impossible, then the (anonymized) data cannot be qualified as 'personal'. [96]
Of course, the DPD can be applied if the encryption still implies personal information [97]; and in light of this one ought to bear in mind that any personal identifier (even an IP-address [98]) may be qualified as personal data.
Offering no more than utility infrastructure services IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) providers (and certain Software as a Service providers) may not even know whether the information being processed while using their services is really 'personal data'. Hence, some authors argue that it may even seem inappropriate to apply the DPD to such cloud infrastructure providers as the processing of personal data depends upon their customer's choices. [99]
With regard to future decryption tools, the relative approach concentrates on the actual available technologies not on tools that will be available in the future. However, the actual technological capacities may change by that time, meaning that "identifiability" may change, as well. [100] Therefore, representatives of the relative approach [101] tend to apply the DPD to encrypted data because eventually, technical tools may facilitate decryption, like the encryption of DVDs. [102] Thus, foreseeable technical developments should be taken into account when assessing the current quality of personal data. [103]
In addition, the uncertainty of when decryption can be done reasonably should not be borne by the protected individual given the uncertainty of security levels provided by encryption. [104]
Nonetheless, in order to check if one falls into the scope of the DPD (if a new technology arises which had been unknown before) the data controller has to verify available technologies continuously; hence, a dynamic obligation is imposed upon the controller in order to regularly evaluate the used technologies. Concerning encryption technology as a means to change the character of "personal" data and render it "impersonal," the encryption operators have to continuously check the state-of-the-art encryption technology. [105]

2.2.1.4.3 Conclusion

As the absolute approach extends the scope of DPD to nearly all kind of data processing, [106] from the perspective of the authors of this report (and from the perspective of the majority of authors),

[95] *Gola/Klug/Körffer* , in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 44; Kroschwald, ZD 2014, 75, (78).

[96] *Kühling/Klar*, NJW 2013, 3611, (3613); *Dammann*, in: Simitis, BDSG, par. 3, recital 32; *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 10; *Polenz* in: Kilian/Heusser, Computerrechts-Handbuch, Part 13, recital 68.

[97] *Spies* , MMR-Aktuell 2011, 313727

[98] C.f.*Bergt* , ZD 2015, 365 (370 f.); *Spindler/Nink* , in: Spindler/Schuster, Recht der elektronischen Medien, par. 11 TMG recital 11 ff; *ECJ*, Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland (see 2.2.1.3).

[99] *Hon/Millard/Walden*, The Problem of 'Personal Data' in Cloud Computing - What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 26.

[100] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 15; *Kroschwald*, ZD 2014, 75 (78).

[101] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 7; *LG Frankenthal*, MMR 2008, 687 (689); *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 14.

[102] *Kroschwald*, ZD 2014, 75 (79).

[103] *Spies*, MMR-Aktuell 2011, 313727.

[104] *Stadler*, Datenschutz: IP-Adressen als personenbezogene Daten

[105] *Jotzo*, Der Schutz personenbezogener Daten in der Cloud, p. 68; *Kroschwald*, ZD 2014, 75 (78 f.); *Art. 29-Working Party*, Opinion 04/2007, WP 136, 15; Cf. also *Roßnagel/Scholz*, MMR 2000, 721 (723).

[106] *Meyerdierks*, MMR 2009, 8 (10); *Peifer*, K&R 2011, 543 (544); *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 115.

the stringer arguments tend to favor the relative approach. [107] Based on the absolute approach data controllers (or data processors) cannot accurately assess if the DPD is applicable, since the DPD would be extended to an omni-present law without any real boundaries. [108] Furthermore, it should be considered that the specific purpose of the Directive is in particular the protection of the right to privacy of natural persons (see Article 2 No. 1 DPD). In scenarios where no realistic or "reasonable" chances to identify the data subject exist, with respect to the concrete situation of the processing actions, the purpose of the DPD is not affected at all. Therefore, it does not seem necessary to apply restricting data protection laws under those circumstances. [109]
Yet, we have to note that, even on the grounds of the relative approach, re-combinability of "harmless data" and creating profiles out of these data (Big data) do fall under the scope of the DPD. [110] Even if at the beginning of data processing the data was not personal, we have to keep in mind, that every data processor has to check if the data they used is already 'personal data' or not. [111] Furthermore, data which is related to things ("Internet of things") can turn out to be personal data if the data can be brought with reasonable effort [112] into a direct relationship with a person. [113]

2.2.2 Summary

As the previous paragraphs have illustrated, the technical requirements set forth by data protection laws concerning cloud computing and encryption in particular, the standards are still not fully settled. In a nutshell, based upon the required expenses, such as time and labor, encryption technologies must be in a way sophisticated that efforts to attribute information to persons (to decrypt) must be realistic. According to the relative approach the perspective of the data processor is relevant in order to assess the (un)reasonable efforts to decrypt the data, thus forgoing an objective point of view that would consider if anyone in the world would be able to decrypt it. Whilst there is evidence of strong support for the relative approach, even in the newly proposed General Data Protection Regulation, we have to emphasize that there is currently no judgment by a higher court (but the decision of the ECJ, as mentioned above, has to be awaited) on this matter that confirms the relative approach. Moreover, the above-mentioned decision of the ECJ is yet to be determined.

2.2.3 Personal Data and Encryption under the GDPR

The three proposals for a GDPR describe personal data differently. Thus, to determine whether encrypted data will be treated as personal data under the GDPR or not, it is necessary to scrutinize and compare each proposal.

[107] *Dammann* in: Simitis, BDSG, par. 3, recital 32; *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 10; *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 115, 116; *Schulz* in: Beck'scher Kommentar zum Recht der Telemediendienste, par. 11 TMG, recital 24; *Roßnagel/Scholz*, MMR 2000, 721 (723); *Meyerdierks*, MMR 2009, 8 (8 ff.); *Eckhardt*, K&R 2007, 601 (603); *Voigt*, MMR 2009, 377; *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 46.

[108] *Meyerdierks*, MMR 2009, 8 (10).

[109] Cf. *Eckhard*, CR 2011, 339 (342); *Härting*, ITRB 2009, 35 (37); *Maisch*, ITRB 2011, 13 (14).

[110] Proposal for a Regulation of the European Parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012/0011 (COD) of 25/01/2012, recital 24: online identifiers combined with other information, available at: http://www.ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_eu.pdf.

[111] *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 116.

[112] *Gerlach*, CR 2013, 478 (479); *Spindler*, Verhandlungen des 69. Deutschen Juristentages, Band I, Gutachten, 2012, F 121; *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 10.

[113] *Art. 29-Working Party*, Opinion 04/2007, WP 136, 19 ff.

2.2.3.1 The Proposal of the Commission

The Proposal of the Commission defines 'personal data' in Article 4 Par. 2 as (2) "any information relating to a data subject." The Problem of 'Data subject' moreover is defined in Par. 1 as:

> "(...) an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person."

By depicting that means can be used by "any other person" represents a rather objective approach. [114] This assumption is furthermore clarified by recital 23:

> "The principles of protection should apply to any information concerning an identified or identifiable person. To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual."

Nevertheless, the term "means reasonably likely to be used" can also be interpreted as a relative component. [115] Additionally, recital 24 establishes that:

> "When using online services, individuals may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses or cookie identifiers. This may leave traces which, combined with unique identifiers and other information received by the servers, may be used to create profiles of the individuals and identify them. It follows that identification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances."

Sentence three, which states that identification numbers, etc. are not necessarily to be considered as personal data, is another indication towards for a relative approach of the proposal, because it dissents with how authors who favor an absolute approach normally classify dynamic IP-addresses. [116] Thus, the Commission's proposal tends towards the objective approach, yet still maintains several relative elements. A legally secure distinction between the two approaches is difficult to ascertain. [117] Therefore, it is conceivable to describe the approach of the Commission can be described as both objective and relative. [118] According to recital 23 Sentence 3 of the Commission, the GDPR will moreover not be applicable to anonymous data:

> "The principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable."

Considering the absolute approach, this would suggest that it is irrelevant whether the data has been encrypted or not: every piece of information that can be associated to a person must, therefore, to be considered as personal data [119] this would greatly extend the scope of the regulation on the European

[114] *Brink/Eckhardt*, DuD 2015, 205 (209); *Härting*, BB 2012, 459 (463); *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 12.
[115] *Brink/Eckhardt*, DuD 2015, 205 (209).
[116] *Brink/Eckhardt*, DuD 2015, 205 (209).
[117] *Härting*, BB 2012, 459 (463); *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 12
[118] *Brink/Eckhardt*, ZD 2015, 205 (209); *Karg*, DuD 2015, 520 (525).
[119] *Härting*, CR 2013, 715 (718).

level. [120] Hence, cloud services which store users' information would more likely fall under the scope of the regulation.
On the other hand, it can be argued that the extent of effort required to obtain the link between the affected person and the data should be greatly considered.
Consequently, with a relative approach it can still be identified that the recital may account for the means used by the respective controller **and** a third person - but only if those means are reasonably likely to be used. [121] If the data is not reasonably likely to be decrypted, the data could be considered non-personal (i.e. anonymous data) because the affected person would not be identifiable. Nonetheless, the proposal of the Commission creates legal uncertainty regarding the question, whether encryption is a suitable method to anonymize personal data.

2.2.3.2 The Proposal of the European Parliament (LIBE-Proposal)

With the LIBE-Proposal two new definitions have been added to the GDPR: It will provide precise definitions of 'pseudonymous data' and 'encrypted data' in Article 4 Par 2a and 2b:

> "(2a) 'pseudonymous data' means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution;"

Thus, pseudonymous data still falls within the scope of the Regulation and is considered personal data.

> "(2b) 'encrypted data' means personal data, which through technological protection measures is rendered unintelligible to any person who is not authorised to access it"

Nevertheless, in general, the LIBE-Definition of 'encrypted data' does not exclude encrypted data from the applicability of the GDPR, in general, since the definition concerns 'personal data' that has been altered to be unintelligible. From a legal perspective as implied by the GDPR, the direct effect encryption of data is relatively small. According to Article 32 Par. 3 GDPR, if data has been encrypted, the controller is not required to communicate a data breach to the data subject. [122] The notification requirements in Article 13 and 13a GDPR provide for an indication of whether or not the data processed will be encrypted (but not anymore included in the proposal of the Council). An indirect effect (not explicitly mentioned in the GDPR) that encryption might have on the processing of personal data could be the strengthening of the legitimate interests pursued by the controller during the balancing of interests required for an explicit legal permission to process data according to Article 6 Lit. f GDPR (see 2.4.2.2).
The fact that there are regulations concerning encrypted data within the GDPR could be interpreted to mean that encryption does not prevent the applicability of the European data protection law: If encrypted data would not fall under the scope of the GDPR, regulations concerning encrypted data within the GDPR would make no sense, entirely. This interpretation would support an absolute approach (see 2.2.1). However, it does not take into account that the qualification of data as personal or non-personal depends on the respective controller's evaluation.
According to this approach (the relative approach, as described above, see 2.2.1), for the party able to decrypt the data, this data has to be considered personal; whereas, for the party not able to decrypt

[120] *Hullen*, PinG 2015, 210 (211). However, note that this extension depends on the former practice in Member States. Germany already used a wider notion of personal data, even according to the so-called 'relative approach', see 2.2.1.
[121] *Lang*, K&R 2012, 145 (146).
[122] The proposal of the Council maintains this exemption in Article 32 Par. 3a GDPR and adds in comparison to the other two proposals the term 'encryption'. See further 2.5.5.

it, the data is being considered anonymous. Hence, the norms of the GDPR that concern encrypted data are interpreted merely as establishing rules for the controller that is able to decrypt the data and how they should process it. In other terms, the norms do not mean that encrypted data always has to be considered personal data for every party. The GDPR's acknowledgment of encryption technologies and the benefit granted by Article 32 Par. 3 to the controller who encrypts data can offer an incentive to controllers to encrypt the affected person's data before processing it. However, it does not answer the question of whether or not encrypted data is considered personal data for a party that is unable to decrypt it. This remains dependent on the approach taken to define "identifiability" (see the following). Yet, the proposals seem to assume that the processing of encrypted data is less dangerous for the affected person's privacy than the processing of un-encrypted data (because the controller does not have to report a data breach to the data subject if the data was encrypted). The LIBE-Proposal combines the proposed elements of the Commission and defines personal data in Article 4 Par. 2 as follows:

> "Personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, unique identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social or gender identity of that person;"

Recital 23 furthermore states that:

> "The principles of data protection should apply to any information concerning an identified or identifiable natural person. To determine whether a person is **identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify** or single out the individual **directly or indirectly**. To ascertain whether means are reasonable likely to be used to identify the individual, account should be taken of **all objective factors**, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development. (...)

Thus, as in the Commission's proposal, again the knowledge of a third party can be used to determine whether a person is identifiable. To limit this wide scope of the proposal, "objective factors" shall be taken into account to determine whether means are reasonably likely to be used to identify the individual. However, this does not reduce the risk that any third person could be able to re-identify the data with means that are no great effort for him or her. [123] Moreover, recital 24 of the LIBE-Proposal is another hint for the absolute approach of the proposal:

> "When using identifiers provided by devices, applications, tools and protocols, such as **Internet Protocol addresses**, cookie identifiers and Radio Frequency Identification tags, this Regulation **should be applicable to processing involving such data**, unless those identifiers do not relate to an identified or identifiable natural person."

The Parliament's and the Commission's version of the GDPR will, as well as the Commission's, not be applicable to anonymous data: Recital 23 clarifies that the data protection legislation does not apply to anonymous data: [124]

[123] C.f. *Hullen*, PinG 2015, 2010 (211).
[124] *Härting*, CR 2013, 715 (718).

> "The principles of data protection should therefore not apply to anonymous data, which is information that does not relate to an identified or identifiable natural person. This Regulation does therefore not concern the processing of such anonymous data, including for statistical and research purposes."

Moreover, as non-identifiable data is not covered by the privacy regulation, anonymity is also mentioned within the context of health data in Article 81. Hence, an exact definition of when data becomes anonymized is not provided by the Regulation but rather than described by Recital 23. Unfortunately, this "definition" does not resolve the aforementioned dispute between the different approaches (relative vs. absolute) to define *anonymization*. Therefore the same problems as mentioned persist such as upcoming techniques to decrypt or to identify data subjects by combining different pieces of information. [125] Techniques such as removing or scrambling direct identifiers or even indirect identifiers, cannot irreversibly anonymize the data irreversible virtually. [126] Therefore, according to the absolute approach almost all data should be considered 'personal data'. The LIBE version of the proposal has furthermore been provided with an explanation, written by draftsperson Jan Albrecht, Member of the European Parliament. It allows insight into the motives behind, at least, the LIBE version of the GDPR. In this explanation, it is stated that the GDPRs purpose is to protect the fundamental rights of the affected persons. In light of this, a limitation of the 'personal data' definition's scope is rejected. [127] According to the explanation, all objective factors should be taken into account when determining if data is 'personal data'. This is clearly a vote for an absolute approach, although it can be criticized for the same reasons as described above (see 2.2.1.4.3).

2.2.3.3 The Proposal of the Council

Finally, the proposal of the Council seems to follow the proposal of the Commission [128] and defines 'personal data' in Article 4 as follows:

> "(1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person."

Moreover, its recital 23 tries to specify when exactly a person is identifiable:

> "The principles of data protection should apply to any information concerning an identified or identifiable natural person. Data including pseudonymised data, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person. To determine whether a person is identifiable, account should be taken of all the means to be used either by the controller or by any other person to identify the individual directly or indirectly."

[125] *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing – What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 22

[126] *Hon/Millard/Walden*, The Problem of 'Personal Data' In Cloud Computing – What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 22

[127] *Albrecht*, Draft Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 0011-C7-0025/2012-2012/0011 (COD) of 16/01/2013, 212, available at: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/pr/924/924343/924343en.pdf.

[128] Karg, DuD 2015, 520 (521).

Similar to the Commission's proposal, means reasonably likely to be used by a third person have to be taken into account, which again tends towards an absolute approach, however the term 'reasonably likely' suggests limitation through relative elements. The recital continues as follows:

> "To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development. "

Similar to the LIBE-Proposal, this illustrates a further attempt to limit the broad absolute elements of the proposal. Finally, recital 24 can, in contrast to the LIBE-Proposal, be read in favor of a relative approach. Nevertheless, this also leaves room for interpretations:

> "When using online services, individuals may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses or cookie identifiers. This may leave traces which, when combined with unique identifiers and other information received by the servers, may be used to create profiles of the individuals and identify them. Identification numbers, location data, online identifiers or other specific factors as such should not be considered as personal data if they do not identify an individual or make an individual identifiable."

Thus, similar to the Commission's proposal, the proposal of the Council includes both objective and relative elements.
As well as the other two proposals, the Council's proposal states in recital 23 that

> "the principles of data protection should therefore not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable. This Regulation does therefore not concern the processing of such anonymous information, including for statistical and research purposes."

Again, the proposal unfortunately does not resolve the issue. When personal data is encrypted and consequently considered anonymous with solid legal certainty, this may in turn aid the development of an encryption technology that excludes personal data from the scope of the Regulation.
In contrast to the LIBE-Proposal, the Council does not further include a definition of 'encryption' in its proposal. Pseudonymisation however is defined in Article 4 Par. 3b:

> "pseudonymisation means the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person."

Recital 23a moreover states that:

> "The application of pseudonymisation to personal data can reduce the risks for the data subjects concerned and help controllers and processors meet their data protection obligations. The explicit introduction of 'pseudonymisation' through the articles of this Regulation is thus not intended to preclude any other measures of data protection."

Thus, as in the case in recital 23, pseudonymised data should be considered as personal data, because the connection between the person and the information on the basis of a corresponding rule remains.

[129] Nevertheless, pseudonymisation is an instrument to help controllers and processors comply with their data protection obligations, which are mentioned in several Articles. For example in Article 23 concerning 'Privacy by Design and by Default', in Article 30 as a technical measure to ensure the security of processing or in Article 83 regarding the processing of personal data for archiving purposes in the public interest or for scientific, statistical and historical purposes to minimize the processing of personal data.

2.2.3.4 Summary and Impact on Cloud Computing and Encryption

In sum, the proposals of the Commission and the Council include both absolute and relative elements, whereas the LIBE-Proposal is inclined towards an absolute approach. Thus, the awaited final text of the GDPR could potentially be influenced by the decision of the ECJ mentioned above. With a solely relative approach, the consequence for cloud computing is that data protection law would not be applicable to the cloud provider if encryption technologies are used. For the controller the data nevertheless remains personal data (see in detail 2.3.2), thus the relative approach leads to a partial non-applicability of data protection law. With the absolute relative approaches of the Commission and the Council, supplementary knowledge of third persons should be considered, but only if they are reasonably likely to link the data to an individual. Therefore, encryption of personal data could be a way to anonymize personal data, provided that it is on an adequate level and state-of-the-art and provided that it is not reasonably likely that a third person could re-identify the data to the data subject. The risks of re-identification nevertheless still exist and the controller has to ensure an adequate level of encryption (c.f. the chapter of the DPD, 2.2.1). With an absolute approach which the LIBE-Proposal ostensibly favors, encryption would never be sufficient to anonymize personal data, and data protection law would always be applicable. Still, the anticipated final draft of the GDPR is yet to give binding answers to the question whether encrypted data has to be considered as personal data or not.

2.3 The Responsible Party (the Controller) and Processing on behalf of the Controller

2.3.1 The Responsible Party (the Controller) and Processing on behalf of the Controller under the DPD

2.3.1.1 Relevance

The data protection law addresses the consequence of being a controller. All requirements needed to fulfill compliance with the data protection law have to be ensured by the controller, and possible fees and court rulings will apply to them. Concerning cloud computing, there can be a lot of entities involved in the whole process of storing and using data in the cloud. For a legal evaluation it is crucial to determine the respective controller. Whereas the cloud user might have clients whose data they are working with, the cloud provider might have subcontractors whose resources they are using when their own capabilities are limited. [130] One should distinguish between "single" controllers, joint controllers, processors and third parties.

[129] *Karg*, DuD 2015, 520 (522).
[130] *Brennscheidt*, Cloud Computing, p. 59.

2.3.1.2 The Controller

Defined as the "natural or legal person that is alone, or jointly with others, responsible for the processing of data," a "data controller" determines the purposes and means of the processing, Art. 2 (d) DPD. It is not necessary for the controller, themselves, to process the data (see 2.3.1.4). It is necessary to describe in detail two important elements included in this definition. First, the controller is the determining element the one who makes the decisions - with respect to the specific data processing action. Second, the subjects left to the controller's determination are the purposes and means of the processing. The element of determination is a matter mainly based upon factual control, which arises out of the circumstances of the concrete situation. Assessing those circumstances, the controlling-capacity might be derived from explicit legal competence, if one entity is either explicitly appointed as a controller or is imposed with particular data processing duties by legal provisions. It might also be indicated by traditional roles, which usually involve certain data responsibilities, e.g. the collection of specific information about employees by the employer. Finally, the factual influence has to be assessed. For this purpose the contractual relations between the parties can be analyzed. An improtant indication could be whether the role of the controller is assigned to one party, or whether this party can be considered dominant relating to data issues altogether, this might be an important indication. However, contractual provisions are not decisive in every case - especially if they do not reflect the factual circumstances. Where doubts occur, the actual control of the parties has to be measured and assessed, taking into consideration the degree of influence actually exercised and the reasonable expectations of the data subjects concerned . [131]

2.3.1.3 Joint Controlling

In a more simplified data processing situation, there might only be one party held responsible when relating to the processing action, as a controller. Nevertheless, the definition provided within Art. 2 (d) expressly includes "control jointly executed by more than one entity." In scenarios where many parties are involved, it is conceivable that various entities can take on the role of joint controllers. As a consequence, each of these parties are bound to the provisions stated within the DPD, with respect to the entire processing action. [132] The general criteria to assess this form of controlling are, in principle, the same as for "normal" controlling of only one party (see 2.3.1.2). [133] In other words, two or more parties are joint controllers if they determine the essential means and the purposes of the data processing solely together. [134] However, in practice, the line between joint controlling, on the one hand, and order processing (see 2.3.1.4.2) of data, on the other hand, is blurred and often leads often to quarrels with supervisory authorities.
The entities do not need to have a close relationship to each other for instance, a civil partnership or similar close contractual relations. The parties can generally choose any legal form to establish their relationship though, this does not affect the responsibility imposed by data protection law. [135] However, contractual agreements can contain important indications for assessing joint controlling (as well as for "single" controlling, see 2.3.1.2) in many cases. Nevertheless, a complete assessment of all specific circumstances is required in order to decide whether parties should take the decisions jointly, or if only one party has to be regarded a ("single") controller. [136]Therefore, it is not important

[131] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 8 ff.
[132] *Wolff/Brink*, Datenschutz in Bund und Ländern, par. 3, recital 112.
[133] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 18.
[134] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 18; *Funke/Wittmann*, ZD 2013, 211 f.; see also: *Alich/Nolte*, CR 2011, 741, (743 f).
[135] *Dammann*, in: Simitis, BDSG, par. 3, recital 226.
[136] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 18; see also 2.3.1.2.

who has the formal right to decide what happens with the data, rather it is crucial who has the actual competence to determine the purposes and means of the processing.[137]
The legal assessment is unambiguous regarding where the different parties jointly determine both the purposes and the means of one particular processing action. However, the Art. 29 Working Party's opinion includes a broader approach to define the scope of joint controlling. According to this opinion, it should be noted, that joint controllers do not need to share the same purposes of the processing - they might differ. Depending on the situation, it either suffices if they only set up an infrastructure of data processing and determine the essential elements of the means to be used or if they share the same purpose without jointly deciding on the means. [138]
Furthermore, as the Art. 29 Working Party argues, the question of joint controlling is not a matter of one particular data processing action. As Art. 2 (b) DPD states, the term "processing" is not limited to one single action but also includes a "set of operations" (see 2.1.4). [139] Especially in the context of IT-infrastructures, there can be many parties involved in different data processing operations of a particular set of personal data. A distinction has to be made if those parties are either "single" controllers that are independent from each other or if they are joint controllers (or if it is a case of order processing, see 2.3.1.4). It is possible that the involved parties divide different tasks and processing operations in a way so that each single action appears to be independent and executed by only one controller. However, by taking into consideration the whole set of operations the "macro-level' – the entities can also be regarded as joint controllers'. This result can be derived from mutually determined purposes and a cooperatively set framework that determines the essential means or if the decisions relating to both questions are taken together. [140] Again, the question of joint controlling is as with respect to "single" controlling a matter of the specific circumstances if the parties factually determine the purposes and/or essential means together. Though many different scenarios with different legal assessment can occur, one example may illustrate the issue: [141] An airline, a hotel chain and a travel agency establish a platform provided through the internet that allows enhanced collaborative travel reservation management between them. They jointly state which data are to be stored on the platform, how reservations are managed and confirmed, to whom access to the data shall be granted, etc. Here, all three parties are joint controllers, with respect to the processing executed by using the common internet-platform, since they decided, at least, about the essential means of the processing. However, one should bear in mind, that the Art. 29-Working Party opinions have no binding statements (see Art. 29 section 1 DPD). In particular, it may be subject to further discussion if such a broad understanding of joint controlling can generally be accepted. The ECJs recent Google Spain judgment seems to embrace such an understanding. A joint controllership was assumed without the controllers intending to cooperate or jointly deciding on the purpose of the data processing. [142] Simply the fact that both parties were able to control the processing had been sufficient for the ECJ to assume joint controllership. [143]
In a usual cloud computing scenario, the cloud-provider does not determine the means and purposes of the data processing, and there is usually no controller, at all (see 2.3.2.4.2). Hence, joint controlling might occur with respect to cases in which more than one user controls the processing action by taking these decisions jointly.

[137] *Jandt/Roßnagel*, ZD 2011, 160, *Jotzo*, MMR 2009, 232 f.
[138] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 19 f.
[139] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 18.
[140] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 20.
[141] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 20.
[142] *ECJ*, Judgment from the 13th May 2014 in Casec-131/12 – Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez, rec. 40.
[143] Cf. *Spindler*, JZ 2014, 981 (983).

2.3.1.4 Processing on behalf of the Controller

2.3.1.4.1 The Processor

As mentioned above, the controller does not necessarily have to be the entity actually processing the data. On the contrary, companies whose main business is outside the IT-sector tend to outsource data processing. According to the law, a "processor" is any legal entity processing the data on behalf of the controller (Article 2 (d) of the DPD) the outsourcing company. All data processing the processor does, is considered as processing done by the controller (the outsourcing company) whose responsibility relating to these processing actions is not affected. As a consequence, all given consent and all legal permissions that the controller has are valid to permit the processor's actions regarding personal data. The processor is treated as if they belonged to the controller's entity. Therefore, no permission is needed for data transfers between the controller and the processor. Sometimes this scenario is also called "order processing".
Acting "on behalf" of the controller contains two basic elements: on the one hand, a processor acts in the controller's interests and not for their own purposes. On the other hand, they are bound to the controller's instructions (see Article 16 DPD), at least with respect to the purposes of the processing and the essential means that are used. In this context, the purpose is the "anticipated outcome that is intended or that guides your planned actions" and the means can be defined as "how a result is obtained or an end is achieved.[144] Furthermore, only an entity legally separated from the controller is in general able to act as a processor. [145]

2.3.1.4.2 Distinction between Processor and Controller

Whenever one entity processes (personal) data for another one, the question that arises is whether or not the one actually processing has to be considered a controller or a processor. The distinction between these two roles should be carried out on the basis of the potential control of the party in question. That means, that whoever fulfills the described conditions of being a controller is regarded as a controller and not as a processor (and of course - neither as a third party). [146] Thus, if one determines the purposes and essential means (at least by giving instructions) he is a controller. [147]
In this context, it is crucial to specify which particular decisions can be delegated to the processor, and in contrast how much leeway or discretion is assigned to the processing party so that it can be already considered as a controller rather than a mere processing party, due to the freedom to decide upon specific means of data processing etc. The possible decisions that are subject to delegation can be divided into two categories requiring different legal assessment: Decisions concerning the purpose of the processing cannot be delegated and are reserved for the controller's authority only. [148] As a consequence, the cloud service provider will be considered a controller, themselves, if they collect their users' personal data for their own purposes. [149]
Decisions that concern the means of the processing, such as which software should be used may on the other hand be delegated to the processor. However, this does not encompass every technical or organizational question. Some are deeply linked to the lawfulness of the processing and, therefore, essential in a way that they can only be answered by the controller. In particular, this is especially

[144] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 13 f., 25.
[145] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 25.
[146] *Brennscheidt*, Cloud Computing und Datenschutz, p. 67; *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 11, recital 9.
[147] Cf. *Hilber*, Handbuch Cloud Computing, p. 350.
[148] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 15 f.
[149] *Giedke*, Cloud Computing, p. 202; *Art. 29-Working Party*, Opinion 08/2010, WP 179, 27; *Art. 29-Working Party*, Opinion 05/2012 WP 196, 10.

relevant t aspects such as the duration of the processing, granting access to third persons, and the choice of which data should be processed. [150]

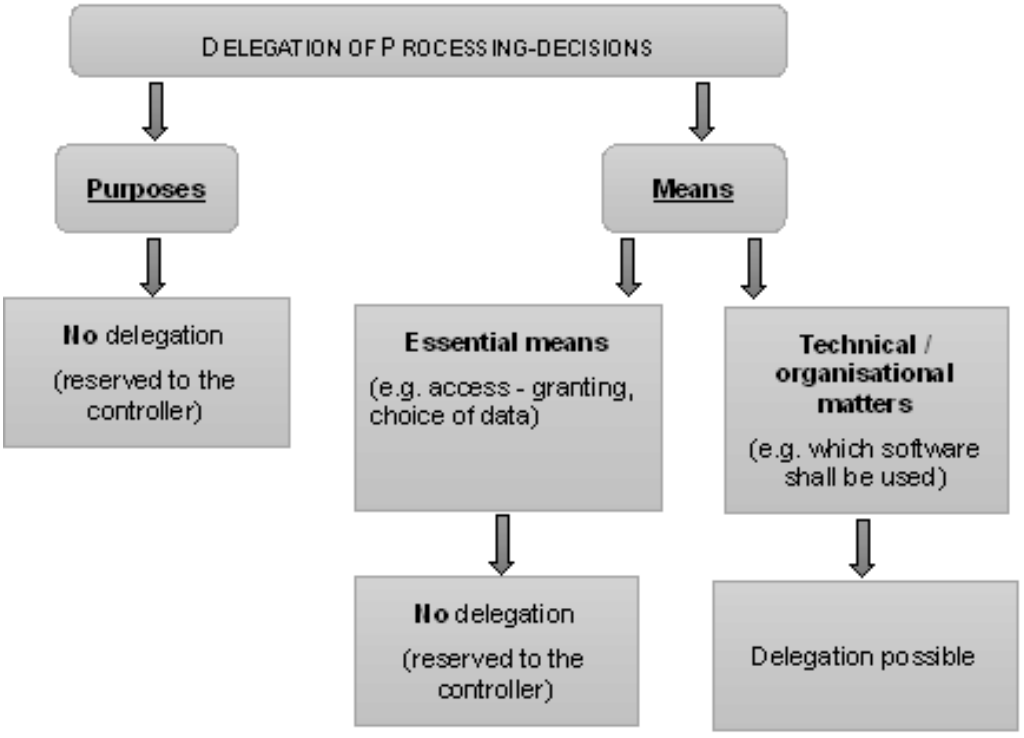

Figure 2.4: Order-processing - Delegation of decision

In a typical cloud computing scenario, the provider only supplies the controller with the technical framework. The latter is the one determining the purposes of the processing. Usually, the controller decides which data are processed and how long the processing will take and, therefore, governs the (essential) means. However the cloud provider only computes the data, as they are bound by the contract concluded with the cloud user, thus possessing little discretionary power that, normally, does not lead to a controllership. [151]

Even though cloud computing can, therefore, typically be regarded as processing on behalf of the controller, in terms of Article 16 DPD, [152] it is deliberated whether or not there can be scenarios in which the provider acts neither as a processor nor as a controller. It is possible that the cloud user does not give any instructions to the cloud service provider on how to handle the data. One might only use the provider's software in a SaaS solution to compute over self-processed input and receive the results. The provider does not exercise any data processing but only establishes and maintains the technology to support data processing that is completely initiated and conducted by the controller, themselves. In such cases, it is argued that one does not "process" on behalf of another but is only indirectly concerned with the data processing and, thus, cannot be considered a processor. [153] Others argue that, under those circumstances, the provisions for data processors apply, as well, since the risks for the personal data do not differ significantly when compared to a situation in which the processor

[150] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 14.

[151] *Brennscheidt*, Cloud Computing, p. 67 f.; *Hennrich*, CR 2011, 546 (548); cf. also *Wolff/Brink*, Datenschutz in Bund und Ländern, par. 3 BDSG, recital 111; *Niemann/Paul*, Praxishandbuch Rechtsfragen des Cloud Computing, chapter D, recital 31 ff.

[152] *Brennscheidt*, Cloud Computing, p. 67 f; apparently assumed in: *Art. 29-Working Party*, Opinion 05/2012, WP 196.

[153] *Hon/Millard/Walden*, Who is Responsible for "personal data" in Cloud Computing? - The Cloud of Unknowing, Part 2, p. 17; *Spindler/Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, par. 11 BDSG, recital 8.

directly processes the data. [154] At the very least, the provider's mere physical control over the data requires the implementation of sufficient safeguards to sustain data security in those cases (assumed one shares that approach), for instance measures to prevent data from accidental loss. [155]
However, this discussion should not be overrated. It is important to bear in mind that whenever a cloud-service includes any form of data storage (on the provider's servers) which goes beyond a mere temporary caching, then this storage constitutes a relevant act of data processing. Accordingly, the provider has to be considered a processor. [156] This applies to an even greater extent if the provider fulfills monitoring tasks with respect to the personal data, e.g. concerning the access or use. [157]
However, there can be situations in which the provider fulfills the requirements of controlling and therefore acts as a controller, and not as a processor. A few examples shall be emphasized. In one instance, a former processor starts processing data for their own purposes, or others', other than those originally determined by the (former) controller. For example, if the "processor" starts to use stored customer data in order to provide commercial advertising in a manner not intended by the user, with respect to this new processing action, they are a controller, since they set a new purpose. [158] The same might apply if they exceed other competences, such as granting data access to unauthorized third parties. [159] Furthermore, the provider could be assigned not only with providing the technical framework but also with completing the complete task that leads to the processing action. Whenever the provider is empowered with the competences to decide the essential means and purposes with respect to that task, they are a controller even though the involved parties may consider them, rather, as a processor. [160] The outsourcing of a company's accountancy is a typical example, for this respect. [161]

2.3.1.4.3 Legal Requirements

There are certain legal requirements to fulfill before (order) processing takes place 'on behalf of the controller' (Article 17 Par. 3 DPD), for example the carrying out of the processing must be governed by a contract or legal act binding the processor to the controller. The processor must be bound to instructions from the controller, and it must be guaranteed that technical and organizational measures are provided to protect personal data against leaks. The main aim is to oblige the processor to follow the controller's instructions, similar to an employee's obligation. For the purposes of verification, the sections of the contract or the legal act relating to data protection and the requirements concerning the technical and organizational measures shall be in writing or in another equivalent form. [162] One may note that users, especially small cloud users, usually do not have a considerable influence on the contractual clauses often provided in a standardized form by the provider. However, it remains part of the controller's responsibility to only enter into processing-contracts which are in complete

[154] Cf. *Schneider*, Handbuch des EDV-Rechts, chapter B, recital 266 f.

[155] *Hon/Millard/Walden*, Who is Responsible for "personal data" in Cloud Computing?, The Cloud of Unknowing, Part 2, p. 22.

[156] *Pohle/Ammann*, K&R 2009, 625 (630); *Spindler/Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, par. 11 BDSG, recital 8; see also more differentiated if the provider has a mere passive role: *Hon/Millard/Walden*, Who is Responsible for "personal data" in Cloud Computing? The Cloud of Unknowing, Part 2, p. 18 ff.

[157] *Hon/Millard/Walden*, Who is Responsible for "personal data" in Cloud Computing?, The Cloud of Unknowing, Part 2, p. 17.

[158] *Art. 29-Working Party*, Opinion 05/2012, WP 196, 14.

[159] *Hon/Millard/Walden*, Who is Responsible for "personal data" in Cloud Computing?, The Cloud of Unknowing, Part 2, p. 20.

[160] Cf. *Brennscheidt*, Cloud Computing und Datenschutz, p. 67; *Funke/Wittmann*, ZD 2013, 221 (223); *Hoeren*, in: Roßnagel, Handbuch Datenschutzrecht, chapter 4.6, recital 97.

[161] *Petri*, in: Simitis, BDSG, par. 11, recital 28.

[162] *Art. 29-Working Party*, Opinion 05/2012, WP 196, 12.

compliance with the respective legal data protection provision. A lack of actual power does not justify concluding an unlawful processing-contract. [163]
The EU's Art. 29-Working Party recommends certain issues to be covered in a contract between the cloud provider and the user. For example they provide:

- for details concerning the client's instructions to be issued to the provider and
- relevant penalties, including potential actions against the provider, in case of non-compliance,
- specification of the security measures the provider must comply with,
- subject and time frame of the cloud service to be provided,
- a confidentiality clause,
- the controller's rights to monitor,
- the cloud provider's obligation to cooperate,
- a list of locations in which the data may be processed, and
- the prohibition of communicating data to third parties or subcontractors not mentioned in the contract. [164]

From a critical point of view, these requirements are difficult to fulfil in practice. On the one hand, it is highly unlikely that big global players in the cloud computing business will actually be bound and controlled by mid-sized or small companies concerning cloud computation (for instance, referring to inspections on the spot). On the other hand, a company not operating in the IT-sector might not even be interested in or be able to provide this kind of control. [165] Since the data might be stored not in one but in many different locations visiting the provider's data centres for an on-site audit seems to be implasible for the cloud user. In addition, it might even be hard to tell where exactly the data will be stored due to the scalability of cloud services. [166] Besides the difficulties for a cloud user to visit and audit all data centres his provider is using, it would constitute a data-security risk for the provider to let (all of) their users inspect all their data centres. This model of control is based upon the classic outsourcing model, with only one data centre to be controlled that might not be located in another country. However, other options to fulfil the cloud user's legal obligations to control his provider have been proposed: As the directive does not require the controller to ensure the processor's compliance by themselves, they could rely on a qualified third party to control the processor (third-party auditing model). [167] On the other hand, the cloud user would still have to pay forthe services of this third party, something that might be impractical even for private individuals. The controller could demand inspection reports from the processor recording his processing activities, but this would not ensure the processor's actual compliance, since those reports would be made by the processor, themselves. [168] An effective, yet practical, way to ensure compliance is data protection certification. [169] Here,

[163] *Art. 29-Working Party*, Opinion 01/2010, WP 169, 26; Hartung/Storm, in: Hilber, Handbuch Cloud Computing, p. 357.

[164] The whole list of recommendations has 14 items and can be found in Art. 29-Working Party, Opinion 05/2012, WP 196, 12 f.

[165] *Heidrich/Wegener*, MMR 2010,803 (806).

[166] *Brennscheidt*, Cloud Computing, p. 102.

[167] *German Federal Office for Information Security Technology*, Safety Recommendation for Cloud Computing Providers, p. 63

[168] *Brennscheidt*, Cloud Computing, p. 105.

[169] *Art. 29-Working Party*, Opinion 05/2012, WP 196, 22.

a third party provides the necessary assessment of the cloud provider. Compared to the third party audit-model mentioned before, the difference is that not every client of the provider has to hire the third party individually. The certification costs are initially covered by the cloud provider and then redistributed to all possible clients by the provider making it possible to professionally control every data centre, and affordable even for private customers. Being certified might provide a competitive advantage for big, global players since this advertises a high standard of data protection to possible clients. The directive does not mention such certificates explicitly. Nevertheless, they could be used by a controller to ensure the compliance of the processing done on their behalf. [170]

2.3.1.4.4 By Processor Outside the EU/ EEA

If the processor does not fall under the jurisdiction of an EU/EEA member-state, data transmission between the controller and the processor generally have to comply with the described conditions. In addition, the requirements of data transfer to third countries have to be met (for more details see 2.4.3); under no circumstances shall personal data be transferred to a third country that is not providing an adequate level of protection without the described requirements (see 2.4.3) . Nevertheless, the contract binding of the processor to the controller can be used to ensure necessary safeguards. Thus, in sum, only if either an adequate level of protection is provided within the third country or other sufficient safeguards are ensured will the DPD allow it to constitute an order processing, including the legal privileges described in 2.3.1.4.1. [171]

2.3.2 The Responsible Party (the Controller) and Processing on behalf of the Controller under the GDPR

In the same way, the GDPR distinguishes between the entity responsible and the entity actually processing the data. Nevertheless, there will be changes in the particular responsibilities of those entities and new ways for the controller to make sure his processor complies with the law. It is essential that order processing under the GDPR meets all prerequisites described below.
It has been criticized that there is no regulation within the GDPR that explicitly states that transfers from a controller to the processor are allowed if 'order processing' takes place.[172] Yet, this critique does not take into account that the legitimation for such transfers lies in the model of "order processing," itself. Without this legitimation, all provisions regarding processing on behalf of the controller would be meaningless. [173]
In general, the processing of data by the processor (such as computation of cloud-stored data in the cloud) is permitted if the controller would be allowed to do it himself (be it by consent or be it other explicit legal permissions). Hence, data processing is permitted under same circumstances and requirements as for the controller. The processing is done on behalf of the controller, i.e. the law treats the processing as if the controller would do it himself. Therefore, the controller needs to be the party deciding why and how the processing is done (see 2.3.2.1 et seq.).

[170] For a detailed description of data protection seals see *Brennscheidt*, Cloud Computing, p. 105 ff.

[171] *Brennscheidt*, Cloud Computing, p. 76.

[172] *Nebel/Richter*, ZD 2012, 407 (411); *Roßnagel/Nebel/Richter*, ZD 2013, 103 (105); c.f. *Koòs/Englisch*, ZD 2014, 276 (284), who see the legitimation in Article 6 lit. f GDPR, if data transfers between the controller and the processor will be considered as necessary for the purposes of the legitimate interests pursued by the controller and not overridden by the interests of the data subject (see 2.4.2.2) and therefore be based on a express legal permission.

[173] C.f. regarding the DPD, but with the same problem: *Drews/Montreal*, PinG 2014, 143.

2.3.2.1 Rules for the Controller

The controller is defined in Article 4 Par. 5 GDPR as

> "the natural or legal person, public authority, agency or any other body which alone, or jointly with others, determines the purposes, conditions and means of the processing of personal data; [...]."

In comparison to the DPD, there will be no significant changes to the definition of 'controller,' for cloud computing. The cloud user as the entity determining the purpose and the means of the data processing will still be considered the controller (for several controllers see 2.3.2.2). The user (= the controller) is thus responsible for the data processing and will be accountable if legal requirements are not met. The controller's main duties are regulated in Article 22 of the GDPR.

> "Article 22 Par. 1: "The controller shall adopt appropriate policies and implement appropriate and demonstrable technical and organizational measures to ensure and be able to demonstrate in a transparent manner that the processing of personal data is performed in compliance with this regulation, i.e. having regard to the state of the art, the nature of personal data processing, the context, scope and purposes of the processing, the risks for the rights and freedoms of the data subjects and the type of the organization, both at the time of the determination of the means for processing and at the time of the processing itself." (Wording of the LIBE-Proposal)

The wording of the proposal of the Council is different to the LIBE-Proposal and excludes the examples given in the proposal of the Parliament. The Council's proposal is a rather risk-based approach, according to which the scope of the technical and organizational measures shall be related to the risk to the personal rights of the data subjects affected: [174]

> Article 22 Par. 1: "Taking into account the nature, scope context and purposes of the processing as well as the likelihood and severity of risk for the rights and freedoms of individuals, the controller shall implement appropriate measures and be able to demonstrate that the processing of personal data is performed in compliance with this Regulation."

Simplified, this means that it is the controller's (the cloud user's) job to ensure that the GDPR's requirements are fulfilled when they initiate data processing. To reach that goal, the controller has to implement technical and organizational measures (see 2.4.4.2) and adopt appropriate policies. To determine if these measures are valid to ensure compliance with the data protection law and the data subjects' privacy, Article 22 Par. 1 provides certain criteria (at least in the LIBE-Proposal).
Specifically, the criteria of the 'state of the art' and 'the risks for the rights and freedoms of the data subjects' can be addressed efficiently by using privacy preserving cloud computing technologies developed by PRACTICE. (see 2.4.4.2)
Nevertheless, the LIBE-Proposal and the Council's proposal omit several paragraphs out, that the proposal of the European Commission introduced to the GDPR in 2012, e.g. measures such as "designating a data protection officer pursuant to Article 35(1)".
Aside from the obligation to ensure compliance and to provide policies that respect the data subject's free choices (Article 22 Par. 1a, LIBE-Proposal), the controller also has to be able to demonstrate the adequacy and effectiveness of those measures and policies. To achieve this, Recital 60 of the LIBE-Proposal recommends independent internal or external auditors (see 2.3.2.4). The Council's proposal furthermore states in Recital 60 that "these measures should take into account the nature,

[174] *Petri*, ZD 2015, 305 (308).

scope, context and purposes of the processing and the risk for the rights and freedoms of individuals" and provides in Recital 60a) several examples of risks and sets in Recital 60b) that risks "should be evaluated on an objective assessment". Article 28 requires documentation (or in the words of the Council: "records of categories of personal data processing activities") of the data processing by the controller (as wel as the processor). They must cooperate with the supervisory authority of Article 29 (no longer included in the Council's proposal); take technical and organizational measures to ensure the security of processing, Article 30 [175]; alert and inform clients about data breach, according to Article 31 Par. 2; conduct a privacy impact assessment under certain conditions of Articles 32a (only in the LIBE-Proposal), 33 Par. 1 or seek a prior authorization in accordance with Article 34 Par. 1 (only in the proposal of the European Commission); appoint a data protection officer, as requested in Article 35 Par. 1 [176]; as well as comply with rules for transfers to third countries, as mentioned in Article 40 ff. The powers of regulators may be expressly addressed to the processors, according to Article 53 Par 1 (a).

2.3.2.2 Joint Controllers

The GDPR's definition of 'controller' allows several entities to be considered as 'joint controllers'. Since the GDPRs definition of 'controller' has only been slightly changed in comparison to the DPDs definition, the distinction between one 'controller' or several 'joint controllers' is still the same (see 2.3.1.4.2) under the DPD. In the case of a 'joint controllers' scenario, it might be difficult to determine the specific responsibilities of each controller. Article 24 GDPR binds joint controllers to come to an arrangement that clarifies each controllers' duties. According to Recital 62 GDPR, the arrangement should reflect the controllers' roles and relationships. The essence of the arrangement has to be made available to the data subject. This is important, since it is necessary that the arrangement determines which controller is responsible for the procedures and mechanisms involved in exercising the rights of the data subject. The reason behind this is that the joint controllers might not be equally capable of negotiating a contract. Additionally, one controller could have a direct relationship to the data subject, whereas another one might not; moreover, they may not be able to control the type and amount of data. [177] The arrangement the GDPR demands should be viewed as a useful tool for cloud participants when they are considered joint controllers. Determining the cloud service's details between the cloud provider and the cloud user(s) may be included in the contract. If the respective responsibilities are not clear to the data subject, all joint controllers are liable, together or separately. In this specific case, the rationale is to provide the data subject with more protection. [178]
This underlines Article 24 Par. 2 of the proposal of the Council, which states that "irrespective of the terms of the arrangement the data subject may exercise his or her rights (...) in respect of and against each of the controllers", unless he or she "has been informed in a transparent and unequivocal manner which of the joint controllers is responsible" (Par. 3).

[175] The Council's proposal gives "pseudonymisation" of personal data to ensure a level of security appropriate to the risk" as an example.

[176] In contrast to the first two proposals, an **obligation** to appoint a data protection officer is **not anymore included** in the Council's proposal Article 35 Par. 1 states that "The controller or the processor may, or where required by Union or Member State law shall, designate a data protection officer".

[177] *Kelly*, ITRE Committee Opinion on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 0011-C7- 0025/2012-2012/0011 (COD) of 26/02/2013, 102, available at: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/itre/ad/927/927816/927816en.pdf.

[178] *Comi*, IMCO Committee Opinion on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 0011 -C7-0025/2012-2012/0011 (COD) of 28/01/2013, 79, available at: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/imco/ad/924/924645/924645en.pdf.

2.3.2.3 Rules regarding the Processor

In the relationship between the cloud provider and the cloud user, the cloud provider usually acts as the processor as defined in Article 4 Par. 6 (also often called "order processing").

> "(6): 'processor' means a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller;"

The cloud user remains accountable to the person (his client) concerned; [179] one should bear in mind that the cloud user often offers services to their clients, thus has to be qualified as a controller (Client cloud user cloud provider). The controller's duties regarding the processor commence before the processing on their behalf takes place, and in addition they have to choose a processor who will comply with the GDPR's requirements:

> "Article 26 Par. 1: Where processing is to be carried out on behalf of a controller, the controller shall choose a processor providing sufficient guarantees to implement appropriate technical and organizational measures and procedures in such a way that the processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject, in particular in respect of the technical security measures and organizational measures governing the processing to be carried out and shall ensure compliance with those measures."

Article 30 GDPR clarifies what is meant by 'technical and organisational measures' (see 2.4.4.2). Those measures shall, among other things, at least "protect stored or transmitted personal data against accidental or unlawful destruction, or accidental loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure" (Article 30 Par. 2 lit. b LIBE-Proposal and Article 30 Par. 1 (a) proposal of the Council). This is possible with encryption technology.
If the cloud provider uses privacy-preserving technologies like encryption, such as those developed by PRACTICE, it can be assumed the provider (at least partially) fulfils its duties regarding aforementioned technical security measures, e.g. with the help of certifications or privacy seals (see 2.3.2.4 and also 2.4.4.2 regarding technical and organizational measures).
Of course, the cloud providers still have to take organizational measures to fulfill all of their duties regulated in Article 30 GDPR.
The controller not only has to choose a sufficient processor. The Regulation sticks to the former approach of the DPD and requires the controller to ensure they have control over the data processing (determining the means of the processing, the required organisational and technical measures, processing only on their instructions, their inspection rights, etc.) by contractual obligations of the processor; Article 26 Par. 2 defines a set of rules that must, in practice, be endorsed in the contract:

[179] *Kroschwald*, ZD 2014, 75 (78); *Weichert*, DuD 2010, 679 (682).

Proposal of the Commission	Proposal of the Parliament (LIBE)	Proposal of the Council
(a) act only on instructions from the controller, in particular, where the transfer of the personal data used is prohibited;	(a) process personal data only on instructions from the controller, unless otherwise required by Union law or Member State law;	(a) process the personal data only on instructions from the controller, unless required to do so by Union or Member State law to which the processor is subject; in such a case, the processor shall inform the controller of that legal requirement before processing the data, unless that law prohibits such information on important grounds of public interest;
(b) employ only staff who have committed themselves to confidentiality or are under a statutory obligation of confidentiality;	(b) employ only staff who have committed themselves to confidentiality or are under a statutory obligation of confidentiality;	(b) (. . .)
(c) take all required measures pursuant to Article 30;	(c) take all required measures pursuant to Article 30;	(c) take all measures required pursuant to Article 30;
(d) enlist another processor only with the prior permission of the controller;	(d) determine the conditions for enlisting another processor only with the prior permission of the controller, unless otherwise determined.	(d) respect the conditions for enlisting another processor, such as a requirement of specific prior permission of the controller;
(e) insofar as this is possible given the nature of the processing, create in agreement with the controller the necessary technical and organisational requirements for the fulfilment of the controller's obligation to respond to requests for exercising the data subject's rights laid down in Chapter III;	(e) insofar as this is possible given the nature of the processing, create in agreement with the controller the appropriate and relevant technical and organisational requirements for the fulfilment of the controller's obligation to respond to requests for exercising the data subject's rights laid down in Chapter III;	(e) taking into account the nature of the processing, assist the controller in responding to requests for exercising the data subject's rights laid down in Chapter III;
(f) assist the controller in ensuring compliance with the obligations pursuant to Articles 30 to 34;	(f) assist the controller in ensuring compliance with the obligations pursuant to Articles 30 to 34, taking into account the nature of processing and the information available to the processor;	(f) assist the controller in ensuring compliance with the obligations pursuant to Articles 30 to 34;

(g) hand over all results to the controller after the end of the processing and not process the personal data otherwise;	(g) return all results to the controller after the end of the processing, not process the personal data otherwise and delete existing copies unless Union or Member State law requires storage of the data;	(g) return or delete, at the choice of the controller, the personal data upon the termination of the provision of data processing services specified in the contract or other legal act, unless there is a requirement to store the data under Union or Member State law to which the processor is subject;
(h) make available to the controller and the supervisory authority all information necessary to control compliance with the obligations laid down in this Article.	(h) make available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow on-site inspections;	(h) make available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow for and contribute to audits conducted by the controller. The processor shall immediately inform the controller if, in his opinion, an instruction breaches this Regulation or Union or Member State data protection provisions.

Although the data protection law will be renewed, the practical problems will still be the same. The cloud user, or/and the controller, might not be in the position to determine contractual clauses but might have to agree to whatever the much stronger processor (the cloud provider) dictates (see 2.3.1.2.1). It may also be impossible for the cloud user to do on-site inspections for the reasons described above. This problem has been addressed by the GDPR, since it is now possible for the controller to rely on data protection seals and third party audits (see 2.3.2.4).
In contrast to the DPD, the legal consequence of a breach of this agreement is explicitly regulated. Thus, Article 26 Par. 4 states:

> "If a processor processes personal data other than as instructed by the controller or if they become the determining party in relation to the purposes and means of data processing, the processor shall be considered to be a controller in respect of that processing, and shall be subject to the rules on joint controllers laid down in Article 24." [180]

The shift of the processor's role (from mere processing to determining and controlling any data processing) thus leads thus to a re-qualification of the processor now as a data controller with all obligations and duties.
Important for cloud computing services is the allowance stated in Article 26 Par. 2 (d) that the processor may use the services of other processors. [181] The trilogue negotiations will have to clarify Par. 2 (d), concerning the question whether another processor can only be enlisted with the prior permission of the controller, as provided by the proposal of the Commission. The LIBE-Proposal on the

[180] Not anymore included in the Council's proposal.
[181] *Brennscheidt*, Cloud Computing, p. 116.

other hand declares this permission to be required only if no different procedure has been determined within the order. The Council however provides in Article 26 Par. 1a. a mediatory solution according to which "the prior specific or general written consent of the controller" is needed. Therefore, the Council's proposal permits more possibilities for sub-cloud-providers. [182]
Thus, a cloud provider may mandate other sub-contractors (sub-cloud providers, etc.) to process the data. However, the data controller is still in charge of controlling the whole process, so that he or she has to assure that his or her inspection rights, etc., are also enforceable in the relationship with the third-party processor (sub cloud provider). In the proposal of the Council, Article 26 Par. 2a. stipulates these principles as follows: "Where a processor enlists another processor for carrying out specific processing activities on behalf of the controller, the same data protection obligations as set out in the contract or other legal act between the controller and the processor (...) shall be imposed on that other processor (...). Where that other processor fails to fulfil its data protection obligations, the initial processor shall remain fully liable to the controller for the performance of that other processor's obligations." Par. 4 of the Commission's proposal and of the LIBE proposal as described above has similar consequences for the processor. Moreover, according to the proposals of the Commission and the Parliament, the requirement of Article 17 Par. 4 DPD in written form will not be included in the future Regulation. In Article 26 Par. 3 GDPR there is just an obligation to "document" the controller's instructions and the processor's obligations. [183] The proposal of the Council however demands that "the contract or the other legal act (...) shall be in writing, including in an electronic form."

2.3.2.4 Privacy Seals and Certification

According to Article 39 GDPR, the data protection authority can act as a certification authority. Each controller and data processor has the right to apply for a certification procedure as mentioned in Article 39. LIBE-Proposal:

> "(1a) Any controller or processor may request any supervisory authority in the Union for a reasonable fee, taking into account the administrative costs, to certify that the processing of personal data is performed in compliance with this Regulation in particular with the principles set out in Article 5, 23 and 30, the obligations of the controller and the processor, and the data subject's rights," [184]
>
> (1b) The certification shall be voluntary, affordable, and available via a process that is transparent and not unduly burdensome."

The certification procedure, however, may turn out to be, in practice, one of the most important tools for data controllers to present evidence required by Article 26 Par. 1, concerning the selection of processors with sufficient guarantees for data protection, particularly appropriate technical and organizational measures. [185] This could potentially be a solution for the dilemma arising from the disparity of power between the cloud computing participants as: the cloud provider will be able to request a certification that the cloud user can rely on. However, there is no obligation for certification. [186] Moreover, Article 39 Par. 1d provides for third party certification procedures if the data protection authority has accredited them. [187]

[182] C.f.*Petri*, ZD 2015, 305 (309).
[183] C.f. *Petri*, ZD 2015, 305 (308).
[184] Wording of the proposal of the Council: "(...) for the purpose of demonstrating compliance with this Regulation of processing operations carried out by controllers and processors. The specific needs of micro, small and medium - sized enterprises shall be taken into account."
[185] *Brennscheidt*, Cloud Computing, p. 116
[186] *Härting*, CR 2013, 715 (720).
[187] *Härting*, CR 2013, 715 (720).

> "(1d) During the certification procedure, the supervisory authority may accredit specialised third-party auditors to carry out the auditing of the controller or the processor on their behalf. [...]"

However, the Council's proposal affirms in Article 39 Par. 2 that:

> "A certification pursuant to this Article does not reduce the responsibility of the controller or the processor for compliance with this Regulation and is without prejudice to the tasks and powers of the supervisory authority which is competent pursuant to Article 51 or 51a."

Nevertheless, the proposal of the Council settles in Article 39 Par. 1 that:

> "(...) the establishment of data protection certification mechanisms (...) **for the purpose of demonstrating compliance with this Regulation of processing operations carried out by controllers and processors.** "

The Council specifies this in Article 22 Par. 2b. by accepting certification mechanisms "as an element to demonstrate compliance with the obligations of the controller." Regarding processors, Article 26 Par. 2 (aa) of the Council furthermore states that:

> "Adherence of the processor to an approved code of conduct pursuant to Article 38 or an approved certification mechanism pursuant to Article 39 may be used as an element to demonstrate sufficient guarantees referred to in paragraphs 1 and 2a."

The LIBE-Proposal provides in Par. 3a. similar wording:

> "The sufficient guarantees referred to in paragraph 1 may be demonstrated by adherence to codes of conduct or certification mechanisms pursuant to Articles 38 or 39 of this Regulation."

Recital 63a of the Council's proposal underlines that:

> "(...) an approved certification mechanism may be used as an element to demonstrate compliance with the obligations of the controller."

In accordance with these two proposals a certificate of the processor (issued by an accredited third party) may thus be considered sufficient evidence to verify compliance with these obligations. [188] Nevertheless, the extent to which certifications have a relieving effect for cloud computing remains unclear.[189] Not only can the processor request a certification, but the controller might have an interest in getting certified, as well. A cloud user (as the controller) may be able to prove to his clients that he uses a cloud service that is compliant with data protection law, that especially provides sufficient technical and organizational safeguards. According to Art 39 Par. 3 GDPR, the Commission is empowered to adopt delegated acts (see 2.5.2) to further specify the criteria and requirements for the certification mechanisms. Although the certification of the processor can make it more straightforward for the controller to present evidence required by Article 23 (1), a certificate will expire after five years.

> (1g) "Notwithstanding paragraph 1f, the certification shall be valid for maximum five years." [190]

[188] *Petri*, ZD 2015, 305 (308).

[189] *Brennscheidt*, Cloud Computing und Datenschutz, p. 116; *Heckmann*, in: jurisPK-Internetrecht, recital 695.

[190] It only lasts three years in the Council's proposal, see Article 39 Par. 4.

Prior to relying on a certificate, the processor will, therefore, at least be obliged to validate if it has expired or not. [191] Nevertheless, the problem that arises from the fact that the cloud user has to ensure by contract that has control over the provider if an 'order processing' shall take place is not solved by a certification. [192] A cloud provider (even if he is they are a big global player) ought to make it easier for their client the cloud user to fulfill their obligations by providing standard contracts for their cloud services that involve the requirements of Art. 26 (2). This way, a lawful use of the cloud service would be possible for the cloud user, as a controller, if the user wants to compute personal data in the cloud. If lawful usage of a cloud service for the computation of personal data is not possible due to the cloud provider denying a contract with their clients (the cloud users, controllers) that, in a way, binds the cloud provider compliant to Art. 26 (2) due to the provider's more powerful position. This in turn might lead to a disadvantage on the European market once the GDPR comes into effect. Finally, the new Article 39a of the Council's proposal introduces several procedures and requirements for certifications and the certification bodies.

2.3.2.5 Liability

According to Art. 77 GDPR, if data has been processed unlawfully, the data subject has the right to claim compensation, even for non-pecuniary damages. Unlike the DPD, it is not only the controller who is liable for such damages. If an 'order processing' takes place, the processor faces liability, too:

> "Article 77 (1): Any person who has suffered damage, including non-pecuniary damage, as a result of an unlawful processing operation or of an action incompatible with this Regulation shall have the right to claim compensation from the controller or the processor for the damage suffered."

This could potentially have a huge impact on cloud providers (usually processors on behalf of the controller, the cloud user), as it could be more promising to hold the solvent provider liable for the affected person (usually the cloud user's client) than holding the cloud user liable. The GDPR incorporates the possibility to avoid liability for damages.

> "(3) The controller or the processor may be exempted from this liability, in whole or in part, if the controller or the processor proves that they are not responsible for the event giving rise to the damage."

Since it is the processor or the controller that needs to prove that they are not responsible for the damage, according to Art. 28, they are both obliged to take the technical and organizational measures that the GDPR demands and fulfill their duty to document the processing. This works to the advantage of the affected person claiming compensation for damages as it is up to the responsibility of the processing parties to prove that they are not liable. On the other hand, the affected person still has to provide evidence for the causation of the unlawful processing for the damages. It has been criticized that this might not be possible for the affected person because he will not have insight into, or be able to document, the controller's or the processor's internal procedures. [193] If several processors or joint controllers caused the damages, this would serve as a further advantage for the affected person as illustrated below:

> "(2) Where more than one controller or processor is involved in the processing, each of those controllers or processors shall be jointly and severally liable for the entire amount

[191] *Sydow/Kring*, ZD 2014, 271 (275).
[192] *Sydow/Kring*, ZD 2014, 271 (275).
[193] *Roßnagel/Richter/Nebel*, ZD 2013, 103 (108).

> of the damage, unless they have an appropriate written agreement determining the responsibilities pursuant to Article 24"

The possibility of joint liability for joint controllers demonstrates the importance of both parties coming to an agreement that fully reflects their responsibilities in the context of data processing. This way, only the respective controller will be liable for possible damages caused by his actions.
However, the Council's proposal grants a privilege to processors who are not responsible for the damage caused by the processing of a controller:

> Article 77 Par. 2: "A processor shall be liable for the damage caused by the processing only where it has not complied with obligations of this Regulation specifically directed to processors or acted outside or contrary to lawful instructions of the controller."

In contrast to the strict regulations of the LIBE-Proposal, the exception above is a positive provision for cloud computing providers who act as processors. The processor shall be exempted from liability if they are able to prove that they are not in any way responsible for the damage (Par. 3).
If a controller or processor is liable for the damage, they can claim back parts of the compensation from the other responsible party in accordance to par. 5 of the Council's proposal.

> (Par. 5): "Where a controller or processor has (...) paid full compensation for the damage suffered, that controller or processor shall be entitled to claim back from the other controllers or processors involved in the same processing that part of the compensation corresponding to their part of responsibility for the damage in accordance with the conditions set out in paragraph 2."

2.3.2.6 Commissioned Data Processing in Third Countries

Cloud computing service, even those based outside Europe, may become subject to the European Data Protection Regulation (see 2.1.2.2). According to Article 3 Par. 1., the scope of the proposed Regulation is not restricted to companies that have an establishment in the EU. [194] Article 3 Par. 2 further extends the scope of the regulation to providers outside of the European Union that process the data of European Citizens (Article 3 Par. 2 (a) and (b)). [195] Even pure targeting and gathering data of EU-citizens by companies outside the EU is now covered by Article 3 Par 2 (a). [196]
As further outlined in 2.4.3.2, a data transfer to a processor in a third country has to be evaluated in two steps according to the DPD. First, a data transfer needs permission, which can be provided if the 'order processing' meets all requirements described above. Second, the transfer into the third country has to be legal, which will be further elaborated in 2.4.3.2.

2.4 Requirements for Legal Data Processing

2.4.1 The Definition of 'Processing'

The Data Protection Directive defines the "processing of data" as:

> "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation

[194] *Härting*, BB 2012, 459 (462); *Wieczorek*, DuD 2013, 644 (648).
[195] *Hon/Millard/Walden*, The Problem 'personal data' in Cloud Computing – What Information is Regulated?, The Cloud of Unknowing, Part 1, p. 3; *Klar*, ZD 2013, 109 (112).
[196] *Härting*, BB 2012, 459 (462); *Wieczorek*, DuD 2013, 644 (648); *Klar*, ZD 2013, 109 (112).

> or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction', Art. 2 (b). "

The exceedingly broad definition of 'processing' leads to the applicability of the DPD and, in tun, to the general prohibition of processing the data unless the DPD allows for it. From the moment the data is collected by the data subject to the very last use of that data, every single step in between has to be either explicitly allowed by law or needs the data subject's consent.
Thus, data controllers can only avoid the applicability of the DPD by rendering the data "not personal". Otherwise, they can have a duty to comply with the requirements - asking the user for explicit consent or presenting reasons that fall under the justifications provided by the DPD.
If personal data is anonymized, this may, technically, mean that it gets altered, but, for the purposes of the Data Protection Directive, 'alteration' means changing the content of the information, not its appearance. [197]
Since the anonymization of personal data eliminates the connection to a person, the encryption of data is one of few possibilities to anonymize personal data and, therefore, the process of encrypting data does not fall under the data protection law, either.

2.4.2 Informed Consent or Explicit Legal Permission

2.4.2.1 Legal Permissions in the DPD

Article 7 DPD enumerates the possible legal grounds for data processing. The first possibility is the affected person's informed consent (see 2.4.2.3). If the controller has not gained the data subject's consent, a lawful processing is possible on the grounds of one of the legal permissions stated in Article 7 (b) to (f) DPD:

> "[. . .]
>
> (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or
>
> (c) processing is necessary for compliance with a legal obligation to which the controller is subject; or
>
> (d) processing is necessary in order to protect the vital interests of the data subject; or
>
> (e) processing is necessary for the performance of a task carried out in the public interest, or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed; or
>
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party, or parties, to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject, which require protection under Article 1 (1)."

Those permissions transferred by the Member States into national law are conclusive, meaning that they are not simply examples among other possible legal grounds but also the only lawful means to process data without the data subjects consent. They permit processing only when it is necessary for certain purposes and not beyond that, corresponding with the DPDs fundamental principle

[197] See also *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 3, recital 30 f.

of proportionality laid down in Article 6 DPD (see 2.1.1.2.2.3). Whereas lit (b) to (e) are applicable only to specific purposes, lit (f) allows the Member States to provide a legal grounding with a larger scope. Nonetheless, heeding Article 6 DPDs main principles, processing on the grounds of a permission based on lit (f) always requires a proportionality test. This indicates that balance has to be struck between the data subjects' and the controllers' interests. Only when the controllers' interests in processing the data without consent outweigh the data subjects' interests in having to consent to the processing can the processing can be lawfully done on the grounds of lit (f). Since gaining informed consent can be difficult for data processing with cloud computing (see 2.4.2.3), it would theoretically be useful if data could be processed without consent. Cloud Computing liberates the cloud user from providing their own physical resources required for carrying out data processing. Instead, the user is able to utilise scalable cloud resources on demand when making use of a cloud provider's infrastructure or even software. This may lead to financial advantages for the cloud user. It is questionable, though, if financial advantages are sufficient to outweigh the data subject's interest concerning comprehensive data protection. [198] This interest is based on a European fundamental right, Article 7 and 8 CFR. [199]

In a recent decision, the ECJ was required to evaluate a similar conflict of interests: a Spanish citizen demanded that Google was obliged to remove personal data within search results and cease from making the relevant search results available to the public. [200] The Spanish citizen argued that Google could no longer base the processing of the data on the legal grounds of Article 7 lit (f). The ECJ emphasized the strong position of the data subject stating that, in this case, merely the economic interest of Google would not be able to justify the processing. [201] The court held that as a rule the rights of the data subject override the economic interests of the operator of the search engine. [202] Although the decision was not concerned with a balance of interests between a cloud user and an affected data subject, but rather with a data subject who was facing a disadvantageous Google search result (linking information to him), the ECJs reasoning can also be applied to Cloud Computing, as well. Therefore, processing on the grounds of Article 7 lit (f) should not only be justifiable through reasons of financial advantage for the cloud user.

2.4.2.2 Legal Permissions in the GDPR

The proposal for a GDPR includes permissions for the processing of personal data that function as exceptions from the general prohibition referred to in 2.1.2. Besides the previous given consent by the affected person (see 2.4.2.3), Article 6 GDPR mentions five other exceptions (wording of the LIBE-Proposal):

> "[. . .]
>
> (b) processing is necessary for the performance of a contract to which the data subject is

[198] Against a saving of costs as a justification for processing *Nägele/Jacobs*, ZUM 2010, 281 (290); *Niemann/Paul*, K&R 2009, 444 (449) on the other hand recognizing a saving of costs as a justification; principally recognizing financial aspects as an possible justification, but only if it would be unreasonable for the controller to waive the processing *Hoeren*, in: Roßnagel, Handbuch Datenschutzrecht, Chapter 4.6, rec. 31, see also *Spindler/Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, par. 28 BDSG, rec. 6

[199] Charter of Fundamental Rights of the European Union, Official Journal of the European Communities, 2000/C 364/01 of the 18.12.2000, available at: http://www.jura.uni-wuerzburg.de/fileadmin/02120300/_temp_/Abbreviations.pdf.

[200] *ECJ*, Judgment from the 13th May 2014 in Casec-131/12 – Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez.

[201] *ECJ*, Judgment from the 13th May 2014 in Casec-131/12 – Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez. par. 81

[202] *ECJ*, Judgment from the 13th May 2014 in Casec-131/12 – Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez. par. 97

party, or in order to take steps at the request of the data subject prior to entering into a contract;

(c) processing is necessary for compliance with a legal obligation to which the controller is subject;

(d) processing is necessary in order to protect the vital interests of the data subject;

(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller, or in case of disclosure by the third party to whom the data is disclosed, and which meet the reasonable expectations of the data subject, based on his or her relationship with the controller, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject, which require protection of personal data. This shall not apply to processing carried out by public authorities in the performance of their tasks."

The proposal of the Council includes two amendments to Article 6 GDPR:

"(d) processing is necessary in order to protect the vital interests of the data subject or of another person;

(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child."

Essentially, the DPD allows for data processing only if the affected party consents to it or if the data processing is necessary for legitimate purposes pertaining to the processor. For Cloud Computing, the same problems as described in 2.4.2.1 exist for a legally admitted processing without consent. It has been criticized that lit. (b) only covers contractual claims and does not include statutory claims. [203] Nevertheless, lit. b includes the "performance of a contract", without a restriction to claims. Moreover, lit. (f) covers all legitimate interests, in case they are not overridden by the data subjects interests; thus, data processing in order to enforce a statutory claim could be lawful without consent of the affected person under certain circumstances. Although, if the processing for direct marketing purposes is permitted according to lit. (f), the data subject is able to object to the processing at any time, free of charge, and without any further justification, Article 19 Par. 2 GDPR. According to Article 19 Par. 1 GDPR the data subject has the right to object to the processing of their data with the result that "the controller shall no longer process the personal data [204] unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defense of legal claims" (wording of the Council's proposal). This extensive right to object does not exist if the processing is based on lit (b), which might in turn cause lit (b) to be the more reliable rationale for processing.

[203] *Berg*, PinG 2013, 69 (70).

[204] In the proposal of the Commission: "which is based on points (d) and (e) of Article 6(1)"; in the LIBE-Proposal: "which is based on points (d), (e) and (f) of Article 6(1); and in the Council's proposal: "which is based on points (e) or (f) of Article 6(1), the first sentence of Article 6(4) in conjunction with point (e) of Article 6(1) or the second sentence of Article 6(4).

2.4.2.3 Informed Consent and Cloud Computing

2.4.2.3.1 Data Protection Directive

In case the DPD is applicable, the safest way to ensure compliance with the Data Protection Directive's national acts of implementation is to ask the data subject for his or her explicit consent. Cited by the Directive in Article 7, consent is the first out of seven legal grounds for personal data processing.
According to Article 8 of the Directive, consent needs to be given explicitly for processing special categories of data. Some Member States view consent as a preferred ground for lawfulness, whereas others view it as one of six options. Every other legal basis for data processing requires a necessity-test. In contrast, the data subject's consent allows the data processor to go beyond what is necessary for their purposes; [205] in other words, the data processor is not bound by a strict proportionality test under these circumstances. [206] However, as noted already, the DPD treats specific sensitive data in an intensified manner, such as data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and data concerning health or sex life.
Consent in the sense of the law is only effective when it is informed and given freely, and unambiguously. Informed consent implies that the data subject has been given certain information before data is processed, including the recipients or categories of recipients of the data (Article 10 (c) Data Protection Directive). [207] It also should additionally be made clear to the data subject whether data will be transferred to a non-EU-state. [208]
These requirements of ex-ante information and transparency lead to difficulties. For example: in case data is to be computed in a cloud, it might be hard to tell when the data will be transferred to a server (to which server?) and in which state this server will be operated. [209] Due to the scalability of cloud computing, the method of storage and the "division of labour" amongst the different servers might be 'decided' by automated programs and could change within seconds. [210] A distinction has to be made between two scenarios. In the first one, the cloud user is the data subject himself (e.g. the user of an online e-mail service, like Gmail). In the other scenario, the cloud user is outsourcing data of a third party (e.g. a company handles its clients' data with a cloud solution). In the first scenario, only one data subject has to give consent to the data processing, which can be done before the user subscribes to the cloud service. In the other scenario, the user would need the consent of every single one of his clients (cascade of consent). This can be done when a new client and the cloud user make their first contractual agreement regarding whatever service the user is offering, but it becomes nearly impossible for old clients, as every one of those must be contacted, given time to react, and give his consent. Implementing the declaration of consent in the general contract terms and conditions might be valid if the client has to accept them actively, but changing existing terms and conditions and informing old clients does not provide a given consent. If consent is given by accepting terms and conditions, legal requirements regarding consumer protection law have to be met, as well. [211]
To ensure compliance with the data protection law, consent would not be the best solution in such a case. [212] For a cloud provider, it would be useful to make the storage of data scalable by location,

[205] *Art. 29-Working Party*, Opinion 15/2011, WP 187, 7f.
[206] *Nägele/Jacobs*, ZUM 2010, 281 (290); *Rath/Rothe*, K&R 2013, 623 (624).
[207] *Taeger*, in: Taeger/Gabel, BDSG, par. 4a, recital 30; *Nord/Manzel*, NJW 2010, 3756 (3757).
[208] *Simitis*, in: Simitis, Bundesdatenschutzgesetz, par. 4a, recitals 70 ff.
[209] *Nägele/Jacobs*, ZUM 2010, 281; *Schultze-Melling*, in: Taeger/Gabel, BDSG, par. 9, recital 104.
[210] *Millard*, Cloud Computing, Chapter 1.1, 1.2; *Funke/Wittmann*, ZD 2013, 221 (222).
[211] *Spindler*, GRUR-Beilage 2014, 101.
[212] *BITKOM*, Leitfaden Cloud Computing, p. 51; *German Federal Office for Information Security Technology*, Safety Recommendation for Cloud Computing Providers, p. 73; *Brennscheidt*, Cloud Computing, p. 152; *Art. 29-Working Party*, Opinion 15/2011, WP 187, 12.

so that the user can choose certain servers to be used for their computation. [213] This way, the data subject can be informed specifically about the location of their data. The same principle applies to the provider's sub-contractors: the user should choose which subcontractor they will use for the specific computation, thus informed consent is safeguarded. Nevertheless, this might be impossible for cloud computing services using resources of other cloud providers. For instance, SaaS provider, Dropbox, builds their service on IaaS by Amazon's S3 (a double 'layer'). Even an SaaS built on a PaaS lying on an IaaS, itself, is possible (e.g. Facebook apps on Heroku on Amazon). [214] In such cases, cloud users will not necessarily know in which data centers, or even countries, their data is stored or with whom their provider has a subcontractor relationship. In fact, the providers may not even be aware themselves. [215]

2.4.2.3.2 Informed Consent and Obligation of Transparency under the GDPR

Article 14 GDPR extends the approach of the DPD concerning transparency for data subjects (and also goes beyond existing national laws, like in Germany) [216] by specifying to the data subject the information that has to be supplied prior to the collection of data. For the purposes of cloud computing, these obligations to inform raise a lot of issues. In order to get an idea of the upcoming problems, we have to take a closer look at the required information of Article 14 Par. 1 GDPR (wording of the LIBE-Proposal):

> "(. . .) the controller shall provide the data subject with at least the following information:
>
> (a) the identity and the contact details of the controller and, if any, of the controller's representative, of the data protection officer;
>
> (b) the purposes of the processing for which the personal data are intended, as well as information regarding the security of the processing of personal data, including the contract terms and general conditions where the processing is based on point (b) of Article 6(1) and, where applicable, information on how they implement and meet the requirements of point (f) of Article 6(1);
>
> (c) the period for which the personal data will be stored or, if this is not possible, the criteria used to determine this period;
>
> (d) the existence of the right to request from the controller access to and rectification or erasure of the personal data concerning the data subject, to object to the processing of such personal data, or to obtain data;
>
> (f) the recipients or categories of recipients of the personal data;
>
> (g) where applicable, that the controller intends to transfer the data to a third country or international organization and on the existence or absence of an adequacy decision by the Commission, or in case of transfers referred to in Article 42, Article 43, or point (h) of Article 44(1), reference to the appropriate safeguards and the means to obtain a copy of them;

[213] For an example such a service is provided by Amazon Web Services, available at: http://aws.amazon.com/de/ec2/pricing/effective-april-2014/

[214] *Millard*, Cloud Computing, Chapter 3.2.

[215] *ComputerWorldUK Cloud Vision blog*, Cloud computing and EU data protection law, Part one: Understanding the international issues, available at: http://blogs.computerworlduk.com/cloud-vision/2011/09/cloud-computing-and-eu-data-protection-law/index.htm.

[216] *Jaspers*, DuD 2012, 571 (572).

(ga) where applicable, information about the existence of profiling, of measures based on profiling, and the envisaged effects of profiling on the data subject;

(gb) meaningful information about the logic involved in any automated processing;

(h) any further information which is necessary to guarantee fair processing in respect of the data subject, having regard to the specific circumstances in which the personal data are collected or processed, in particular the existence of certain processing activities and operations for which a personal data impact assessment has indicated that there may be a high risk; (ha) where applicable, information whether personal data was provided to public authorities during the last consecutive 12-month period."

The proposal of the Council changed several of these paragraphs by shifting or deleting them or by introducing new provisions to Article 14:

"1. Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with the following information:

(a) the identity and the contact details of the controller and, if any, of the controller's representative; the controller shall also include the contact details of the data protection officer, if any;

(b) the purposes of the processing for which the personal data are intended (. . .) as well as the legal basis of the processing.

(. . .)

1a. In addition to the information referred to in paragraph 1, the controller shall at the time when personal data are obtained provide the data subject with such further information that is necessary to ensure fair and transparent processing (. . .), having regard to the specific circumstances and context in which the personal data are processed:

(a) (. . .);

(b) **where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party**;

(c) the recipients or categories of recipients of the personal data;

(d) where applicable, that the controller intends to transfer personal data to a recipient in a third country or international organisation;

(e) the existence of the right to request from the controller access to and rectification or erasure of the personal data or restriction of processing of personal data concerning the data subject and to object to the processing of such personal data (. . .) as well as the right to data portability;

(ea) **where the processing is based on point (a) of Article 6(1) or point (a) of Article 9(2), the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal**;

(f) the right to lodge a complaint to a supervisory authority (. . .);

(g)**whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the data and of the possible consequences of failure to provide such data**;

> (h) **the existence of automated decision making** including profiling referred to in Article 20(1) and (3) and information concerning (...) the logic involved , as well as the significance and the envisaged consequences of such processing for the data subject."

In addition to this, the proposal of the Council introduced a new Article 14a to the Regulation within which the controller shall provide the data subject with more information in cases where the data has not been obtained from the data subject.
It is evident that due to the variety of data processing procedures and sub-providers in the cloud it is nearly impossible to provide sufficient information to the data subject. For instance the recipients of the personal data cannot be realistically be identified in a cloud in advance as the storing and processing depends upon the (global and dispersed) available capacities. The same is true for the transfer of data to a third country which cannot be easily assessed in advance (see 2.4.2.3). If a controller intends to collect data by using a cloud service in order to process the data, it will be even more important for them that the data is not considered 'personal data' under the GDPR.
According to Article 14 GDPR, the controller has to provide the information before personal data is collected - which also includes the collecting of data based on an explicit legal permission, as provided by Article 6. In contrast, the DPD simply requires the controller to provide such information to gain informed consent. The GDPR, on the other hand, requires that such information is also provided before data is collected, on the grounds of legal permission.
The cloud user (controller) can comply with these obligations by choosing a cloud provider who enables them to determine which servers in what country will be used to offer the cloud service. [217]
Whereas the DPD only requires a freely given consent, particularly an informed consent, the GDPR demands far more requirements from a controller.
Article 4 Par. 8 of the LIBE-Proposal intensifies the requirements for a valid consent by demanding an "explicit" consent (in contrast to Article 7 lit. a DPD, which considers it to be sufficient if "the data subject has unambiguously given his consent") and moreover "a statement or a clear affirmative action." Thus, the possibility of hiding statements in terms and conditions or using implied consent would no longer be possible. [218] The users will normally have to 'opt-in'. However, the Council removed the word "explicit" from the definition in its proposal and re-established the possibility of giving implied consent, which can be used for non-sensitive data processing.
In accordance with Article 6 Par. 1 (a) GDPR, the processing of personal data shall be legitimated through unambiguous consent only if this consent refers to specific and defined purposes. Nevertheless, the proposal of the Council introduced some exceptions to the principle of strict purposes, e.g. by stating in Article 6 Par. 4 Sentence 2 that "further processing by the same controller for incompatible purposes on grounds of legitimate interests of that controller or a third party shall be lawful if these interests override the interests of the data subject." Additionally, "further processing of personal data for archiving purposes in the public interest or scientific, statistical or historical purposes shall in accordance with Article 83 not be considered incompatible with the initial purposes."
Furthermore, Article 7 Par. 4 of the LIBE-Proposal regulates that "consent shall be purpose-limited and shall lose its validity when the purpose ceases to exist or as soon as the processing of personal data is no longer necessary for carrying out the purpose for which they were originally collected" this paragraph has also not been included in the Council's proposal. Moreover, a provision which would have meant the invalidity of a consent "where there is a significant imbalance between the position of the data subject and the controller" is no longer included both in the LIBE and the Council's proposals.

[217] The determination of the servers used is possible e.g. if Amazon is chosen as a cloud provider, see http://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/.
[218] *Hartung/Storm*, in: Hilber, Handbuch Cloud Computing, Teil 4, recital 66.

[219] With regard to Cloud Computing, an imbalance between the data subject and the controller could occur if e.g. an employer runs e-mail accounts of his or her employees in the cloud or processes personal data of his or her employees in the cloud. According to the Commission's proposal, it will be difficult for employers to receive valid consent from their employees due to this imbalance and a lack of voluntary consent. [220]
Article 7 Par. 3 provides the data subject with the right to withdraw his or her consent at any time without any further requirements and this right has to be articulated prior to the data subject giving consent. Furthermore, Article 8 Par. 1 Sentence 1 GDPR states that "the processing of personal data of a child below the age of 13 years shall only be lawful if and to the extent that consent is given or authorised by the child's parent." Sentence 2 states that "the controller shall make reasonable efforts to verify such consent, taking into consideration available technology without causing otherwise unnecessary processing of personal data."
Moreover, Article 7 of the GDPR does not demand a written form for a consent, thus signifying that electronic consent is sufficient.
These obligations to inform are flanked by the provision in Article 13a of the LIBE-Proposal for the Regulation (which is not included in the other two proposals) that requires the data controller to provide standardized and easily legible information (what is, in detail, prescribed by the annex of the proposed Regulation). Concerning the information of Article 14, the information has to be specified according to the individual circumstances of the data subject, for instance, information about the national competent supervising authority or about options to file a complaint.

2.4.3 Data transfer to third Countries

2.4.3.1 The DPD

Unless the data subject consents or the provisions of the DPD expressly permit it, transferring data to a 'third country' (a state not within the EU or the EEA) is principally forbidden. [221] The same problems mentioned above can also occur in connection with the data subject's consent to data processing in a cloud when the data subject has to agree to a transfer in an unsafe country. [222] Hence, there is a difference between the legal permission to process data and the legal permission to transfer the data into a third country. The data transfer is only rendered if both requirements are fulfilled separately.
One of the main exceptions refers to the 'adequate level of protection' in the third country, Article 25 of the Directive. [223] An adequate level of protection assumes that the data protection standards in the respective country are comparable to European standards. This has to be officially acknowledged by the European Commission as has been the case for the following countries: Andorra, Argentina, Australia, Canada, Switzerland, Faeroe Islands, Guernsey, the State of Israel, the Isle of Man, Jersey, New Zealand, and Eastern Republic of Uruguay. [224] Being of particular relevance for Cloud Computing, the USA has not been acknowledged. [225] However, the Safe Harbor agreement between the

[219] *Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, 4a BDSG, recital 8; the former provision of the Commission respective an imbalance has been heavily criticised, see *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 14; *Roßnagel/Kroschwald*, ZD 2014, 495 (500); however, recital 34 of the proposal of the Council states again that a consent is not freely-given "where there is a clear imbalance between the data subject and the controller".

[220] C.f. *Hartung/Storm*, in: Hilber, Handbuch Cloud Computing, Teil 4, recital 70.

[221] In detail *Art. 29-Working Party*, Working Document on Transfers of personal data to third countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, WP 74.

[222] See also *Brennscheidt*, Cloud Computing, p. 175.

[223] *Hon/Millard*, Data Export in Cloud Computing How can Personal Data be Transferred outside the EEA?, p. 5.

[224] All decisions by the European Commission regarding the acknowledgment of third countries are available at: http://ec.europa.eu/justice/data-protection/document/international-transfers/adequacy/index_en.htm.

[225] No such decision has been made by the commission; *Gabel*, in: Taeger/Gabel, BDSG, par. 4b, recital 23; *BITKOM*,

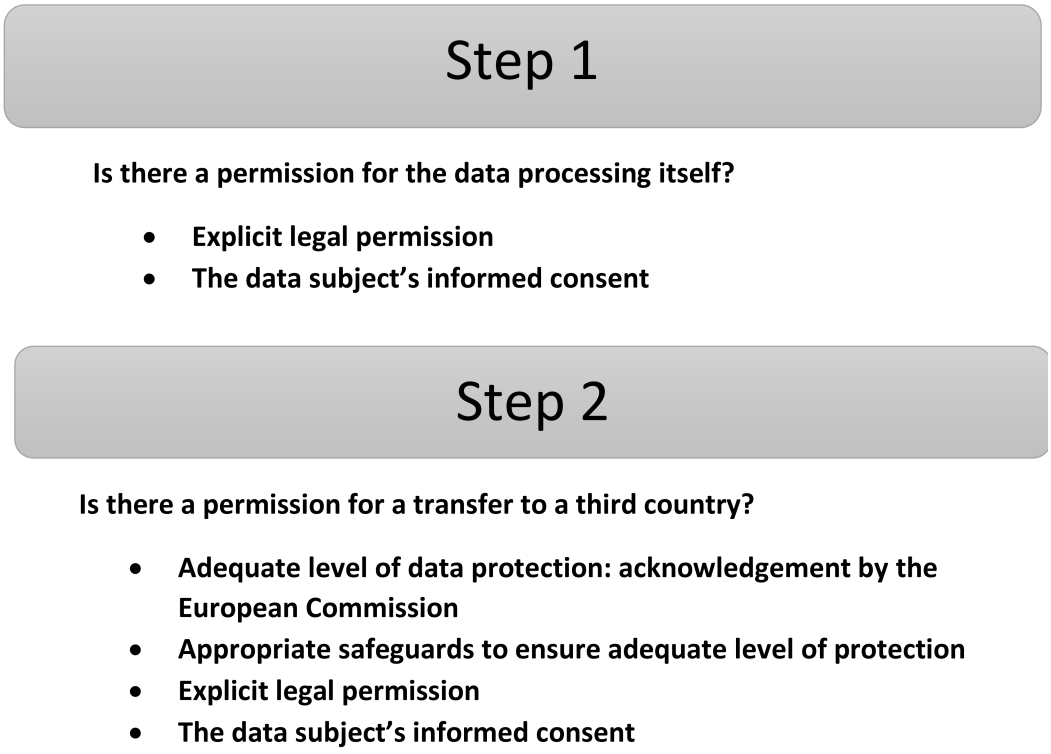

Figure 2.5: Data transfers to third countries

EU and the USA provides a possibility for American companies to comply with the DPD when transferring personal data to the USA. [226] The companies can certify themselves and opt into the program by signing a declaration of accession and by publishing a privacy statement of the US Department of Commerce. [227] The USA does not provide an adequate level of data protection in accordance with Article 25 Par. 6 DPD, therefore Safe Harbor has been negotiated outside the scope of Article 26 DPD as an international treaty between the EU and the USA, the principles of the agreement are based on Article 25 Par. 1 and 2 DPD. [228]
Nevertheless, this changed crucially since the European Court of Justice decided as a result os the case Maximillian Schrems versus Data Protection Commissioner that the decision of the Commission regarding the adequate level of data protection in the US is invalid. [229]
According to the ECJ a system of self-certification does not contradict the requirement laid down in Article 25 Par. 6 DPD. However, the reliability of such a system is at stake, in particular that the concept of adequacy is (also) essentially founded on the establishment of effective detection and supervision mechanisms concerning any infringements upon the rules, thus ensuring the protection of fundamental rights (recital 81). The Safe Harbor principles are applicable solely to self-certified US organizations whereas US public authorities are not required to comply with them (recital 28). The court states that "national security, public interest, or law enforcement requirements have primacy

Leitfaden Cloud Computing, p. 53

[226] The Safe Harbor Principles are available at: http://export.gov/safeharbor/; Hon/Millard, Data Export in Cloud Computing How can Personal Data be Transferred outside the EEA?, p. 15; a summary of the essential rules of the Safe Harbor Principles is provided by *Wisskirchen*, CR 2004, 862 (864 f.).

[227] *Hartung/Storm*, in: Hilber, Handbuch Cloud Computing, Teil 4, recital 237.

[228] *Savin*, EU Internet Law, p. 204; v. d. *Bussche*, in: Plath, BDSG, par. 4b, recital 30. At the moment, there are still ongoing negotiations between the EC and the USA about a new agreement.

[229] *ECJ*, decision of 06/10/2015, Case C362/14 Request for a preliminary ruling of the ECJ from the High Court (Ireland), Maximillian Schrems v Data Protection Commissioner and Digital Rights Ireland Ltd.

over the Safe Harbor principles, primacy pursuant to which self-certified United States organisations receiving personal data from the European Union are bound to disregard those principles without limitation where they conflict with those requirements and therefore prove incompatible with them" (recital 86). Thus, the Safe Harbor Agreement enables the US authorities to infringe in European fundamental rights (c.f. recital 87). Furthermore, in recital 89 the court criticizes the lack of effective legal protection against interferences of that kind and refers to the opinion of European Court of Justice's advocate general Yves Bot [230] in its recital 204 according to which "the private dispute resolution mechanisms and the FTC, owing to its role limited to commercial disputes, are not means of challenging access by the US intelligence services to personal data transferred from the European Union." The ECJ states that "legislation not providing for any possibility for an individual to pursue legal remedies in order to have access to personal data relating to him, or to obtain the rectification or erasure of such data, does not respect the essence of the fundamental right to effective judicial protection, as enshrined in Article 47 of the Charter which requires everyone whose rights and freedoms guaranteed by the law of the European Union are violated to have the right to an effective remedy before a tribunal in compliance with the conditions laid down in that article" (recital 95). Nevertheless, the Commission did not state in the Safe Harbor Agreement that the USA ensures an adequate level of protection by reason of its domestic law or its international commitments (recital 97). Thus, "it is to be concluded that Article 1 of Decision 2000/520 fails to comply with the requirements laid down in Article 25(6) of Directive 95/46, read in the light of the Charter, and that it is accordingly invalid" (recital 98). Furthermore, the agreement shall be invalid, "because the national supervisory authorities must be able to examine, with complete independence, any claim concerning the protection of a person's rights and freedoms in regard to the processing of personal data relating to him. That is in particular the case where, in bringing such a claim, that person raises questions regarding the compatibility of a Commission decision adopted pursuant to Article 25(6) (...) (recital 99). However, Article 3 Par. 1 of the Safe Harbor decision lays down specific rules regarding the powers available to the national supervisory authorities in light of a Commission finding relating to an adequate level of protection (recital 100) and is consequently invalid, as well as the decision in its entirety (recital 105). Apart from this case, the agreement has recently also been criticized for not living up to the requirements of European Data Protection law, however, American companies who joined the Safe Harbor agreement and are following its principles were still considered to be providing an adequate level of protection. [231] Furthermore, some national authorities, like the German supervisory authorities, required "on-the-spot" examination of the validity of the US-company's claim to obey the Safe Harbor agreement. [232] Due to their inability to access their cloud provider's data processor, smaller cloud users may face severe problems in terms of passing these tests. A cloud user's obligation to monitor the cloud provider also contradicts the basic idea of cloud computing the possibility to outsource the data processing without having to control it anymore. Therefore, compliance by means of the Safe Harbor Agreement is impractical for cloud solutions, at least given the actual practice of some supervisory authorities and the decision of the ECJ. [233]

Besides the officially acknowledged countries mentioned above (where no explicit consent from the user is needed), the data controller who wishes to transfer data in other countries may use other forms

[230] Opinion of Advocate General *Yves Bot*, delivered on 23/09/2015, Case C362/14 Request for a preliminary ruling of the ECJ from the High Court (Ireland), Maximillian Schrems v Data Protection Commissioner.

[231] The Safe Harbor Principles are available at: http://export.gov/safeharbor/; *Hon/Millard*, Data Export in Cloud Computing How can Personal Data be Transferred outside the EEA?, p. 15.

[232] *Düsseldorfer Kreis*, Decision of 28th/29th April 2010, available at: http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DuesseldorferKreis/290410_SafeHarbor.pdf?__blob=publicationFile; *Haag*, in: Leupold/Glossner MAH IT-Recht, Teil 4, recital 40; *Marnau/Schlehahn*, DuD 2011, 311 (315).

[233] *Brennscheidt*, Cloud Computing, p. 166; *Heidrich/Wegener*, MMR 2010, 803 (806); *Giedke*, Cloud Computing, p. 233.

of justification provided by the DPD. In general, this is possible, if the controller adduces evidence of adequate safeguards with respect to the protection of the data subject's rights, Article 26 Par. 2 of the Directive. Those safeguards can be based upon appropriate standard contractual clauses which the EU Commissions has acknowledged regarding processors in third countries [234] between the controller and the entity receiving the data, ensuring an adequate level of protection. Those clauses are used to establish rules for the third country party that displays the same standard of protection as the EU does for data subject's rights. Yet, the benefit of a lawful transfer to the third country only exists if the clauses acknowledged by the Commission are used exactly how the Commission provided them and without alteration. [235] For a cloud user who wishes to transfer data to a cloud provider within a third country, the standard contractual clauses would then only be useful if the cloud provider agrees to those exact clauses. It seems unlikely that a cloud provider in contractual agreement with many cloud users would alter these existing contracts , but would rather agree to the standard contractual clauses. Another way of providing adequate safeguards are the so-called Binding Corporate Rules (BCR). Other than the standard contractual clauses, BCR are not mentioned explicitly in the Directive. Nevertheless, Article 26 Par. 2 is not exhaustive, which suggests that appropriate safeguards might be other measures than the explicitly mentioned standard contractual clauses; they are only an example among other possible safeguards. [236] BCR are supposed to ensure that there is an adequate level of data protection for data transfers within a corporation, regardless of the countries the corporation might be seated in.

> "The rules must apply generally throughout the corporate group, irrespective of the place of establishment of the members, or the nationality of the data subjects whose personal data is being processed, or any other criteria or consideration." [237]

The BCR have to be binding or legally enforceable and should be regarded as "sufficient safeguards" within the context of Article 26 Par. 2 DPD. They are meant to be used by multinational companies to allow international data transfers. [238] There are no model BCR provided by the Art. 29-Working Group or the Commission, such as in the case of standard contractual clauses. However, the Art. 29-Working Group proposed crucial elements of BCR and how these rules might be structured in a single document. [239] BCR have to be acknowledged by a supervisory authority in an EU Member State. In case of such an acknowledgement, authorities of most EU-Member States acknowledge BCR automatically, thus creating some form of European passport (notwithstanding the fact that the

[234] See also *Art. 29-Working Party*, Opinion 03/2009, WP 161; Standard Contractual Clauses I, Commission Decision of 15th June 2001 on standard contractual clauses for the transfer of personal data to third countries, under Directive 95/46/EC, C(2001) 1539 (2001/497/EC), available at: https://www.datatilsynet.no/Global/04_skjema_maler/EUs%20standardkontrakter1_ENG.pdf; Standard Contractual Clauses II, Commission Decision of 27th December 2004 amending Decision 2001/497/EC as regards the introduction of an alternative set of standard contractual clauses for the transfer of personal data to third countries, C(2004) 5271 (2004/915/EC), available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:385:0074:0084:en:PDF; Commission Decision 2010/87/EU of 05.02.2010 on Standard Contractual Clauses for Data Processors established in Third Countries, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:039:0005:0018:EN:PDF.

[235] *Gola/Klug/Körffer*, in: Gola/Schomerus, Bundesdatenschutzgesetz, par. 4c, recital 14; *Spindler/Nink*, in: Spindler/Schuster, Recht der elektronischen Medien, par. 4c BDSG, recital 20.

[236] *Art. 29-Working Party*, Working Document on Transfers of personal data to third countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, WP 74, p. 6.

[237] *Art. 29-Working Party*, Working Document on Transfers of personal data to third countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, WP 74, p. 8.

[238] *Art. 29-Working Party*, Working Document on Transfers of personal data to third countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, WP 74, p. 8.

[239] C.f. *Art. 29-Working Party*, Working Document Setting up a framework for the structure of Binding Corporate Rules, WP 154.

DPD does not contain such a procedure). [240] In some specific cases, BCR may be used for cloud computing-related data transfers, however, they will be restricted to internal data transfers across borders. [241] On the other hand, most cloud related data transfers to third countries will not be within a corporation but rather effectuated in a cloud, thus transferring data from a cloud user to a cloud provider. Therefore, BCR do not provide a general solution for cloud computing related to third-country transfers. [242] Nevertheless, following the decision of the ECJ regarding Safe Harbor it is doubtful whether standard contractual clauses or BCR can be seen as a lawful way to transfer data to the USA, because they also don't prevent USA authorities to access personal data transferred to the US. One potential solution would involve acquiring the consent of the data subjects for the data transfer, which however in most of the cases would not be possible. Thus, the new agreement between the Commission and the USA, which is currently being negotiated, has to be awaited to gain legal certainty.

2.4.3.2 The GDPR

Concerning the transfer of data to companies/data processors located outside the EU, the Regulation adheres to the former approach of the DPD. [243] The GDPR also demands the two steps necessary for a lawful transfer, as mentioned above (2.3.2.6). The first step refers to the permission to process the personal data. The second one concerns the transfer of data to a third country, thus safeguarding an adequate level of protection (comparable to the European level), which is crucial for any transmission. The instruments which a data processor can use to comply with these requirements remain basically the same: The transmission can be based on:

- an acknowledgement of adequacy by the EU Commission (Article 41 GDPR), or
- 'binding corporate rules' (Article 42 Par. 2 a and Article 43 GDPR), or
- a European Data Protection seal (Article 42 Par. 2 aa GDPR) [244] (see 2.3.2.4), or
- standard data protection clauses adopted by the Commission (Article 42 Par. 2 b [245] or contract clauses (Art. 42 Par. 2 c GDPR), or
- contract clauses approved by a supervisory authority (Art. 42 Par. 2 d GDPR) [246]

Concerning the benchmarks and relevant criteria which the EU Commission will use for acknowledgement of an adequate level, Article 41 Par. 2 in the wording of the LIBE-Proposal requires the Commission to

> "... give consideration to the following elements:
>
> (a) the rule of law, relevant legislation in force, both general and sectorial, including concerning public security, defense, national security and criminal law as well as the implementation of this legislation, the professional rules and security measures which

[240] *Brennscheidt*, Cloud Computing, p. 173.
[241] *Niemann/Paul*, K&R 2009, 444 (449).
[242] *Brennscheidt*, Cloud Computing, p. 174.
[243] *Nebel/Richter*, ZD 2012, 407 (412).
[244] Only included in the LIBE-Proposal; the Council's proposal nevertheless includes "an approved certification mechanism pursuant to Article 39 together with binding and enforceable commitments of the controller or processor" (Article 42 Par. 2 (e)).
[245] Not included in the LIBE-Proposal.
[246] Article 42 Par. 2a (a) in the Council's proposal.

> are complied with in that country or by that international organization, jurisprudential precedents, as well as effective and enforceable rights including effective administrative and judicial redress for data subjects, in particular for those data subjects residing in the Union whose personal data are being transferred; [247]
>
> (b) the existence and effective functioning of one or more independent supervisory authorities in the third country or international organization in question responsible for ensuring compliance with the data protection rules, including sufficient sanctioning powers, for assisting and advising the data subjects in exercising their rights and for co-operation with the supervisory authorities of the Union and of Member States; and
>
> (3) the international commitments the third country or international organization in question has entered into, in particular any legally binding conventions or instruments with respect to the protection of personal data."

In addition recital 81 of the GDPR states that:

> "In line with the fundamental values on which the Union is founded, particularly the protection of human rights, the Commission should, in its assessment of the third country, take into account how a given third country respects the rule of law".

Moreover, the Council's proposal adds inter alia:

> "The third country should offer guarantees that ensure an adequate level of protection in particular when data are processed in one or several specific sectors. In particular, the third country should ensure effective data protection supervision and should provide for cooperation mechanisms with the European data protection authorities, and the data subjects should be provided with effective and enforceable rights and effective administrative and judicial redress." [248]

In sum, the Commission ought to balance all of these elements and compare the level of data protection in the third country to the one in Europe.
Transfers by the way of BCR are specified in Article 43 GDPR. BCR have to fulfil certain criteria in order to make a data transfer to a third country lawful. They have to ensure all essential principles and enforceable rights of the GDPR to be considered an appropriate safeguard for third country transfers. Their purpose is to enable corporate groups to transfer data to entities within the same corporate group (recital 85 GDPR). BCR will be approved by the Commission if they fulfil Article 43's criteria. BCR have indeed been generally accepted by Article 26 Par. 2 DPD as adequate safeguards, however they have not yet been mentioned explicitly in the DPD. [249] The LIBE committee of the EU Parliament significantly changed the former version, with regards to cloud computing. BCR shall only be approved by the Commission if they categorically have a binding character:

[247] The Council's proposal includes some different elements in Article 41 Par. 2 (a): "the rule of law, respect for human rights and fundamental freedoms, relevant legislation, both general and sectoral, data protection rules and security measures, including rules for onward transfer of personal data to another third country or international organisation, which are complied with in that third country or international organisation, as well as the existence of effective and enforceable data subject rights and effective administrative and judicial redress for data subjects whose personal data are being transferred".

[248] But not included in the LIBE-Proposal.

[249] *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 55.

Article 43 Par. 1 (a):

> "BCR are legally binding and apply to and are enforced by every member within the controllers' group of undertakings and those external subcontractors that are covered by the scope of the binding corporate rules, and include their employees"

Hence, BCR have to bind not only members of the controller's group but also subcontractors this is an amendment which is aimed specifically at cloud computing services. [250]
Moreover, the European data protection seal (Article 39) can be used by the controller to provide evidence of a processors' compliance with the GDPR in the case of processing on their behalf (see 2.3.2.4). Moreover, the seal can provide evidence for appropriate safeguards concerning the level of data protection in order to permit a transfer to a third country. Consequently, the data protection seal can be important for both steps needed to render a third-country transfer lawful.
Appropriate safeguards can also be provided by means of standard contractual clauses or contract clauses approved by a supervisory authority. In each model the contract has to be concluded between the controller transferring the data and the party receiving the data in the third country. Thus, the receiving party shall be bound to European data protection principles. According to Article 14 GDPR, although the person whose data is being processed is not part of the contract, this person has to be provided with information. Whereas standard contract clauses which are acknowledged by the EU Commission(Article 42 (2 c) benefit from a general validity (Article 62 Par. 1 (b) GDPR) [251] individual contract clauses of a controller need prior authorization from the competent supervisory authority (not the commission), Article 42 Par. 4 GDPR. [252] Standard contract clauses are already acknowledged in Article 26 Par. 4 DPD and shall remain valid according to Article 41 Par. 8 of the Commission's and of the LIBE-Proposal and according to Par. 4a. of the Council's proposal.
Standard data protection clauses according to Article 42 Par. 2b of the Commission's and of the Council's proposals are not provided in the LIBE-Proposal, they shall according to the Parliament only be accepted by national supervisory authorities or according to Article 57 GDPR. [253]
The Safe Harbor agreement (see 2.4.3.1) will not be affected by the GDPR. [254] Recital 79 states that:

> "This Regulation is without prejudice to international agreements concluded between the Union and third countries regulating the transfer of personal data, including appropriate safeguards for the data subjects- ensuring an adequate level of protection for the fundamental rights of citizens"

The Council's proposal adds:

> "Member States may conclude international agreements which involve the transfer of personal data to third countries or international organisations, as far as such agreements do not affect this Regulation or any other provisions of EU law and include safeguards to protect the rights of the data subjects."

[250] *Kelly*, ITRE Committee Opinion on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 C7- 0025/2012 2012/0011(COD)) 26th of February 2013, p. 140, available at: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/itre/ad/927/927816/927816en.pdf.

[251] Not anymore included in the Council's proposal, but in accordance with Article 42 Par. 2c the clauses have to be adopted by a supervisory authority and the Commission pursuant to the examination procedure referred to in Article 87 Par. 2.

[252] Article 42 Par. 2a in the Council's proposal.

[253] *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 55.

[254] *Nebel/Richter*, ZD 2012, 407 (412).

According to Article 41 Par. 8 of the Commission's proposal decisions adopted by the Commission on the basis of Article 25 Par. 6 or Article 26 Par. 4 DPD shall remain in force, until amended, replaced or repealed by the Commission. The LIBE-Proposal alters this into a stricter regulation that states that the decisions "shall remain in force until five years after the entry into force of this Regulation unless amended, replaced or repealed by the Commission before the end of this period". The proposal of the Council only demands that the Commission monitor the functioning of the decisions (Par. 4a.).
Data transfers to a controller or processor within the USA will, therefore, still be possible if the receiving party follows the Safe Harbor principles. This does not solve the problems arising from the rather self-regulatory character of Safe Harbor described above and the potential invalidity of the agreement (2.4.3.1). Furthermore, sentence two of recital 79 added by the Council will not make it easier to achieve international agreements compliant with this regulation.
If appropriate safeguards have not been taken to guarantee an adequate level of data protection, the transfer of personal data to a third country can only be carried out if Article 44 GDPRs requirements are met. Thus, either the data subject has to give his consent (causing the same problems as described above, see 2.4.2.3) [255] to the transfer or one of the legal permissions in Article 44 (b) to (g) should be applicable. Those permissions are similar to Article 6 GDPR's legal permissions (see 2.4.2.2) for processing personal data. Note that Article 44 takes effect on the second step (if the transfer to a third country is lawful) and not on the first step (if the processing, itself, is lawful). Thus, when comparing the provisions of the GDPR proposals to the DPD, no significant changes can be found in the proposals for a GDPR.

2.4.4 Technical and Organizational Measures

2.4.4.1 Under the DPD

Appropriate technical and organizational measures have to be provided in order to avoid data leaks, data loss and illegal forms of personal data processing, Article 17 Par. 1 Data Protection Directive. The core security objectives are availability, confidentiality, and integrity; in addition, transparency, accountability and portability also have to be taken into account. [256] As the DPD does not specify exactly which measures have to be taken, data controllers are, to some extent (and depending upon the practice of national supervisory authorities), flexible to adopt the appropriate measures. Existing ISO/IEC standards can be adopted and applied by data processing entities to ensure providing appropriate technical and organizational measures. They can be used as a general guide for initiating and implementing the IT security management process. [257]
In order to achieve the goal of enhancing (or guaranteeing) the safety of personal data, one of the crucial elements is the isolation of every client's computing on every level of the cloud computing stack. In other words, computing processes have to be protected from other parties who want to access it, so that personal data is literally "safe".
Moreover, the IT-infrastructure (networks, IT-systems, applications) has to be secure, even including physical resources, like buildings and employees. [258] Providing availability of data means ensuring timely and reliable access to personal data. Integrity implies that data is authentic and has not been maliciously or accidentally altered during processing, storage or transmission. Thus, remote

[255] And in the Council's proposal this consent even has to be given 'explicitly'.
[256] See also *Art. 29-Working Party*, Opinion 05/2012, WP 196, 14
[257] For a list of ISO standards with further explanation see *German Federal Office for Information Security Technology*, BSI-Standard 100-1 Information Security Management Systems, p. 8.
[258] *German Federal Office for Information Security Technology*, Safety Recommendation for Cloud Computing Providers, p. 28 ff.

administration of a cloud platform should only take place via a secure communication channel. [259]
Article 17 DPD states that the measures taken have to protect the personal data against unauthorized disclosure or access. The state of the art has to be respected in order to assess which measures are appropriate.
One of the technical means to ensure confidentiality is encryption which can protect data against illegal access, disclosure or alteration during its storage and transfer. Encryption can anonymize data or at least pseudonymised data (which complies with the principle of data minimization and data reduction) and thus protect personal data against misuse, especially against attacks from the Internet. [260]

Hence, for PRACTICE, one important tool to ensure confidentiality is encryption.

2.4.4.2 Under the GDPR

The GDPR will change the specification of technical and organizational measures. Article 30 GDPR regulates the controller's and processor's duties, regarding the detailed measures to be taken. Nevertheless, the core principles set out in Article 30 are similar to those developed by the DPD. The GDPR includes them explicitly in the norm.

> "Article 30 Par. 1a. GDPR (wording of the LIBE-Proposal)
>
> Having regard to the state of the art and the cost of implementation, such a security policy shall include:
>
> (a) The ability to ensure that the integrity of the personal data is validated;
>
> (b) The ability to ensure the ongoing confidentiality, integrity, availability and resilience of systems and services processing personal data;
>
> (c) The ability to restore the availability and access to data in a timely manner in the event of a physical or technical incident that impacts the availability, integrity and confidentiality of information systems and services"

To determine the state of the art the European Data Protection Board [261] will be entrusted to issue guidelines, recommendations and best practices (Article 30 Par. 3 GDPR, but not included in the Council's proposal).
Encryption of data will still be a very useful tool to accomplish the task of ensuring integrity and confidentiality, set by Article 30 GDPR. The measures shall at least,

> "[. . .] protect personal data stored or transmitted against accidental or unlawful destruction, or accidental loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure;" (Article 30 Par. 2 lit. b only of the LIBE-Proposal)

As mentioned above, encryption-technologies are developed to prevent unauthorized access to data. This shall be accomplished by having

> "[. . .] regard to the state of the art and the costs of their implementation" (Article 30 Par. 1 of the proposal of the Commission and of the LIBE-Proposal).

[259] *Art. 29-Working Party*, Opinion 05/2012, WP 196, p. 14 f.
[260] *Hartung/Storm*, in: Hilber, Handbuch Cloud Computing, Teil 4 recital 117; furthermore, sentence 3 of the attachment to par. 9 BDSG explicitly mentions encryption as a possible technical and organizational measure.
[261] A board composed of the heads of the supervisory authorities of the Member States and the European Data Protection Supervisor - similar to the Art. 29 Working Party Articles 64 GDPR.

The proposal of the Council adds in its Article 30 Par. 1:

> "Having regard to available technology and the costs of implementation and taking into account the nature, scope, context and purposes of the processing as well as the likelihood and severity of the risk for the rights and freedoms of individuals, the controller and the processor shall implement appropriate technical and organisational measures, such as pseudonymisation of personal data to ensure a level of security appropriate to the risk."

In accordance with this, encryption has the potential to be an appropriate technical measure as it means even more security for personal data than the above mentioned pseudonymisation of personal data (at least according to the relative approach). Hence, it will be necessary to monitor the Data Protection Board's publications and always use encryption that is considered as 'State-of-the-art'.
Developing technologies using state-of-the-art encryption to enable privacy-preserving cloud computation is the main goal of PRACTICE. Cloud users and providers will be able to take technical measures as demanded by Article 30 GDPR when using PRACTICE technologies.

2.5 Other reforms by the GDPR

2.5.1 Third Country Actions against Data Controllers

Regarding the requirement of a controller or processor to disclose personal data, Article 43a Par. 1 (only) of the LIBE-Proposal provides a verdict on enforceability and "reject-ability" of judgments of a court, a tribunal or a decision of an administrative authority of a third country [262] This negative clause is clearly aimed at activities of third countries obliging providers (data controllers, processors) to disclose personal data following the NSA scandals and revelations of Edward Snowden. Whereas the first unofficial draft of the Regulation by the Commission in late November 2011 (which had been leaked to the public) contained a similar provision in its Article 42, the official proposal in January 2012 omitted that provision. [263] The EU Parliament discernibly reinstated this article as a reaction to the monitoring activities of foreign intelligence agencies.
Moreover, a US-Court recently obliged a cloud provider to disclose data not only stored in the USA, but also stored on server based in Ireland. [264] The court denied to quash a warrant after Microsoft filed an instant motion against it. The warrant based on the US Stored Communications Act obliged Microsoft to disclose information to the US government, and was specifically referring to an e-mail account hosted in Dublin, and, therefore, stored within the EU. Microsoft argued that the Stored Communications Act cannot be applied extraterritorially, however the court refused to accept this argument and extended the application of the SCA to third countries.
To protect persons within the EU from having their personal data transferred to a third country based on a third-country ruling which is not compliant with the European data protection law, Recital 90 GDPR only of the LIBE-Proposal states that

[262] *Stadler*, Der Datenschutz bietet keine Handhabe gegen die überwachungspraxis der Geheimdienste; *Klinger*, jurisPR-ITR 6/2014 annotation 2.

[263] *Logemann*, LIBE-Ausschuss bestätigt Gesetzentwurf zur EU-Datenschutz-Grundverordnung; Bergemann, EU-Datenschutzverordnung darf nicht Merkels NSA-Feigenblatt werden; Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Version 56 (29/11/2011), available at: http://statewatch.org/news/2011/dec/eu-com-draft-dp-reg-inter-service-consultation.pdf.

[264] In re Warrant to Search a Certain E-Mail Account Controlled and Maintained by Microsoft Corp., F. Supp. 2d., No. 13 Mag. 2814, 2014 WL 1661004, at *11 (S.D.N.Y. Apr. 25, 2014), CRi 2014, 91 and available at: http://www.nysd.uscourts.gov/cases/show.php?db=special&id=398.

> "In cases where controllers or processors are confronted with conflicting compliance requirements between the jurisdiction of the EU, on the one hand, and that of a third country, on the other, the Commission should ensure that EU law takes precedence at all times"

In case of an order issued by a third-country court or supervisory authority, etc., the controller or processor and, if existing, the controller's representative, shall notify the supervisory authority of the request without undue delay and must obtain prior authorization for the transfer or disclosure by the supervisory authority (Article 43a Par. 2). In principal, no judgment of a court or tribunal of an administrative authority in a third country will be recognized in the EU if a controller or processor is forced to disclose personal data (Article 43a Par. 1). The supervisory authority has to assess the compliance of the requested disclosure with the Regulation and, in particular, if the disclosure is necessary and legally required in accordance with Article 44 Par. 1 d and Article 44 Par. 5. Without prejudice to Article 21, the controller or processor must also inform the data subjects of the request and of the authorization by the supervisory authority and, where applicable, inform the data subject whether personal data was provided to public authorities during the last consecutive 12-month period, pursuant to point of Article 14 Par. 1.
Thus, the European Data Protection law can require data controllers and processors to break a third country's law in order to comply with Article 43 (a). According to the GDPR, If the supervisory authority does not acknowledge the data transfer required by a third-country authority, the controller or processor will be in a collision of obligations. [265] This difficult situation is addressed only in Recital 90 GDPR of the LIBE-Proposal:

> "The Commission should provide guidance and assistance to the controller and processor, and it should seek to resolve the jurisdictional conflict with the third country in question"

Nevertheless, this declaration of will does not really provide a clear solution for the dilemma of a cloud provider under the control of European law and the law of a third country.

2.5.2 Privacy by Design and by Default

One of the main innovations of the proposed Regulation refers to Privacy by Design, Article 23. The Privacy by Design principle requires all producers, data controllers, etc., to respect data protection issues whilst developing or implementing new IT-systems or products. [266] Thus, any privacy issues shall already be addressed during the development of new technologies in order to find solutions from scratch. Data Protection by Design must, particularly, take into account the entire lifecycle management of personal data from collection to processing to deletion, systematically focusing on comprehensive procedural safeguards regarding accuracy, confidentiality, integrity, physical security and deletion of personal data. However, the GDPR obliges the controller and the processor to respect the principle of Privacy by Design; in contrast, the developer of a data processing technology is not directly addressed by Art. 23 GDPR. The proposals of the Commission and of the Council only address the controller(s) and not the processor.
From a policy perspective, this solution appears to omitthe relevant aspects of Privacy by Design, as a controller or a processor is often not able to influence the development of a technology. In particular, a cloud user in the role of the responsible data controller is not developing the technology the cloud provider is using. Even the cloud provider may just present their service based on technologies offered by third parties (such as software developers etc.). [267]

[265] *Plath*, Datenherausgabepflicht für Cloud-Anbieter nach US-Recht vs. EU-Datenschutzrecht.

[266] *Decker*, Die neue europäische Datenschutzgrundverordnung welche änderungen sind für deutsche Unternehmen zu erwarten?; *Schaar*, Privacy by design; Krempl, EU-Datenschützer fordert Einbau von Datenschutz in die Technik.

[267] *Roßnagel/Richter/Nebel*, ZD 2013, 103 (105).

To demonstrate compliance with the requirements of Privacy by Design, an approved certification mechanism pursuant to Article 39 may be used, but only provided by the proposal of the Council in Par. 2a. The proposal of the Commission empowers the Commission in Par. 3 "to adopt delegated acts in accordance with Article 86 for the purpose of specifying any further criteria and requirements for appropriate measures and mechanisms" and in Par. 4 to "lay down technical standards", however, this could collide with the German theory of legislative reservation ('Wesentlichkeitstheorie') [268] and as a consequence has not been included in the proposals of the Parliament and of the Council.
Concerning Privacy by Design, encryption technologies, as well as other technologies which guarantee privacy by technological means, such as data minimisation and pseudonymisation, explicitly mentioned by the proposal of the Council, are one of the crucial enhancements to which the GDPR refers.
As PRACTICE develops privacy-preserving cloud computing technologies, users are able to use cloud computing for computation over encrypted data. As a benefit from the encryption, the data will be secured against unauthorized access and alteration. The technologies are developed to improve privacy and prevent third parties from learning confidential information from the start. Thus, the principle of Privacy by Design is fulfilled par excellence. With those privacy-preserving cloud technologies, controllers and processors can live up to Art. 23 GDPR's requirements.

2.5.3 'Right to erasure'

Whereas the (heavily-debated) 'right to be forgotten and to erase' seriously affected providers by obliging them to ensure that data would also be deleted on third-party caches and servers, the LIBE-Proposal of the GDPR provides only for a 'right to delete' or erasure in Article 17. [269] Although the term has changed, the original proposal's content continues to exist. The data controllers should, therefore, be obliged to provide information about a deletion request of an interested party to third parties to whom data has been passed on. However, a lot of details still remain unresolved: for example, the balance between the right of the public to be informed by archives and historical information, and the right of the individual to have the data deleted. This clash between the fundamental right to privacy and the right of the public to be informed could have been dealt with efficiently in the ECJs recent Google Spain decision. Unfortunately, the ECJ stated very briefly that as a general rule the data subject's rights, override the interest of the public. [270] However, the ECJ also conceded that, depending on the sensitivity of information stored, there are specific cases in which the interest of the person whose data is being processed may be outweighed by the publics' interest in accessing that information (for instance, in cases of persons of public interest).
The Council's proposal re-introduced the term "right to be forgotten" in Article 17. Personal data shall be erased "without undue delay" and the data subject in this proposal only has to object to the processing of personal data. [271] This is in contrast to the LIBE-Proposal, which demands a final court judgement until the data has to be erased.
Furthermore, the Council has regulated in Par. 2a that controllers who made the data public are obliged to take all reasonable steps to inform other controllers which are processing the data that the data subject has requested erasure. The test of reasonability refers to available technology as well as the cost of implementation.

[268] *Schulz*, CR 2012, 204 (206).
[269] *Fazlioglu*, International Data Privacy Law, 2013, p. 149; *Sartor*, International Data Privacy Law 2013, 3 (9); *Krügel*, ZD-Aktuell 2014, 03870; *Roßnagel/Richter/Nebel*, ZD 2013, 103 (107).
[270] *ECJ*, Judgment of 13th May 2014 in Case C-131/12 Google Spain SL/Google Inc. v AEPD/Mario Costeja Gonzalez, rec. 81.
[271] And if there are no overriding legitimate grounds for the processing

However, no data shall be erased in accordance with Par. 3, for example when processing of the personal data is necessary (1) for exercising the right of freedom of expression and information, or (2) for compliance with a legal obligation which requires processing of personal data by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or for archiving purposes in the public interest or for scientific, statistical and historical purposes.
Thus, the Council's proposal for a right to erasure is less stringent than the other two proposals of the Commission or of the European Parliament.
For Cloud Computing it is, after all, important that the cloud user is able to compel the cloud provider to delete personal data. Hence, if the provider is processing data on behalf of the user, then the provider should be bound by such a contractual obligation to delete data - within the general contract framework needed for 'order processing' (see 2.3.2).

2.5.4 Significant Increase of Fines

The lack of enforcement is one of the most important concerns of the current data protection legislation. One of the actions taken by the EU Commission (and the Parliament) to overcome these deficits is an increase of fines in Article 79 to a maximum of 2% [272] of the global annual turnover of an infringing company which is comparable only to antitrust fines.
Even more, Article 79 Par. 2a.c) of the LIBE-Proposal provides for fines of up to 5% of the annual worldwide turnover of a company, or up to 100 million Euros if severe breaches of data protection duties should occur. [273] The fines shall in each individual case be effective, proportionate and dissuasive.
Finally, Article 63 Par. 1 GDPR of the Commission's proposal and of the LIBE-Proposal envisages to strengthen enforcement across borders:

> "For the purposes of this Regulation, an enforceable measure of the supervisory authority of one Member State shall be enforced in all Member States concerned."

Thus, the provision introduces some form of mutual acknowledgement of enforceable orders in Europe - which, however, is not included in the Council's proposal.

2.5.5 Report of Data Breach

In case of data breaches, the controller has to inform the supervisory authority without undue delay. Moreover, the processor of data has to inform the controller about any data breaches without excessive delay. [274] The obligation to report data breaches covers all kinds of personal data. Even unauthorized access to data within the controller's company or agency is considered to be a data breach, and thus has to be notified to the supervisory authority.
Article 31 Par. 3 GDPR lists the minimum requirements that a notification has to meet. The notification has to include the number of data subjects and data records concerned. It might be hard to tell how many data subjects or data records have been lost if a server that is used for cloud computing has been compromised, due to the scalability of cloud computing and the fast transfer of data sets. [275]
According to Article 32 Par. 3a of the proposal of the Council, the communication to the data subject shall not be required if the he or she "has implemented appropriate technological and organisational

[272] See Par. 6 of the Commission's proposal and Par. 3 of the new introduced Article 79a of the Council's proposal.
[273] *Natz/Wolters* , LMuR 2014, 3 (4); *Krügel*, ZD-Aktuell 2014, 03870.
[274] Regarding the difference between controller and processor see 2.3.2.
[275] For the same reason it might be hard to gain informed consent for processing data in a cloud see 2.4.2.3 and 2.4.2.4.

protection measures and those measures were applied to the data affected by the personal data breach, in particular those that render the data unintelligible to any person who is not authorised to access it, such as encryption." [276] However, since the time-period of the data breach is relevant for the assessment of the security of the encryption technology used, and the standard for the safe encryption continuously changes, regulations concerning the question which level of technical security is adequate can unfortunately not be found in the GDPR's wording. [277] The Data Protection Authority is obligated to keep a public register of the types of breaches notified. Article 31 refers to all kind of data breaches, making no difference between third-party attacks (hacker etc.), employees, etc.
Still unresolved and implicitly left to the Member States is the issue of civil liability for data breaches, particularly if omitted breach notifications may constitute grounds for civil action.

2.5.6 Right to Data Portability

The GDPR introduces in Article 18 of the Commission's proposal the new right to data portability, according to which "the data subject shall have the right, where personal data are processed by electronic means and in a structured and commonly used format, to obtain from the controller a copy of data undergoing processing in an electronic and structured format which is commonly used and allows for further use by the data subject." The data subject shall have the right to transmit in an electronic format personal data retained by an automated processing system into another one. This is similar to the provisions the proposal of the Council provides in its Article 18. The LIBE-Proposal does not include Article 18 but provides in Article 15 similar provisions regarding a right to data portability and states in Par. 2a that "the data shall be transferred directly from controller to controller at the request of the data subject where technically feasible and available." Nevertheless, recital 55 of the proposal of the Council states that the right "should not apply where processing is based on another legal ground other than consent or contract." Thus, cloud providers will have to install a technical environment with which they can provide the data subjects with their personal data if obtained, furthermore, they will have to be able to transfer the data to other controllers.

2.5.7 One-Stop-Shop

Another innovation of the GDPR is the introduction of a 'one-stop-shop' for the European data protection supervisory authorities. [278] According to Article 51 Par. 2 of the proposal of the Commission the supervisory authority of the main establishment of the controller or processor shall be competent to supervise the processing activities of the controller or the processor in all Member States if the controller or processor is established in more than one Member State. Article 4 Par. 13 defines 'main establishment' regarding the controller as "the place of its establishment in the Union where the main decisions as to the purposes, conditions and means of the processing of personal data are taken; if no decisions as to the purposes, conditions and means of the processing of personal data are taken in the Union, the main establishment is the place where the main processing activities in the context of the activities of an establishment of a controller in the Union take place." Regarding the processor, it means the place of its central administration in the Union. The LIBE-Proposal does not include Article 51 Par. 2, but introduces a 'Lead Authority' in Article 54, "the supervisory authority of the main establishment of the controller or processor shall act as the lead authority responsible for the supervision of the processing activities." The proposal of the Council establishes this system

[276] The other two proposals also include this relief in their Article 32 Par. 3, nevertheless, the term 'encryption' is added by the Council for the first time, moreover, recital 68a of the Council's proposal mentions it explicitly.
[277] *Marschall*, DuD 2015, 183 (189).
[278] C.f. *Hullen*, in: v. d. Bussche/Voigt, Konzerndatenschutz, Teil 8, recital 39 ff.

in Article 51a, according to which "the supervisory authority of the main establishment (...) of the controller or processor shall be competent to act as lead supervisory authority for the transnational processing of this controller or processor." 'Transnational processing' means in accordance with Article 4 Par. 19b of the Council's proposal "processing which takes place in the context of the activities of establishments in more than one Member State of a controller or a processor in the Union and the controller or processor is established in more than one Member State or processing which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State." Moreover, Article 54a regulates the cooperation between the lead supervisory authority and other concerned supervisory authorities. Hence, under the GDPR companies with establishments in several Member States will only have to deal with the supervisory authority of the Member State in which their main establishment is seated.

Chapter 3

Legal Case Studies

3.1 Encrypted Databases Encrypted HANA

3.1.1 Functioning

Stealing private information by collecting data is a significant problem, especially for online applications. One of the solutions is to encrypt sensitive data, in order to make sure that no unauthorized person is able to use the sensitive data. However, if personal data is encrypted, some applications and programs may not be able to handle and further process that encrypted data. However, Encrypted HANA is based on CryptDB and is able to solve this problem. Encrypted HANA addresses two threats: The first refers to a "curious" database administrator who tries to learn and, in the worst case, spy on personal data by snooping on the database management system (DBMS) [1] , see Figure 1. The second threat concerns an external attacker who gains complete control of an application and database management system. [2] To avoid these two threats, Encrypted HANA minimizes the amount of confidential information revealed to the database management system server whilst providing a variety of queries over the encrypted data, thus enabling further processing of data.

3.1.1.1 Three Main Ideas of Encrypted HANA

Encrypted HANA tries to solve this problem by using three main ideas: [3]

3.1.1.1.1 Execution of SQL- Queries Over Encrypted Data

The main idea of Encrypted HANA is based upon an SQL-aware encryption strategy. [4] SQL-queries consist of a well-defined set of primitive operators, such as equality checks, order comparisons, aggregates, and joins. By adapting known encryption schemes and using new privacy-preserving cryptographic method for joins, HANA encrypts each data item in a way that allows the DBMS to compute the transformed data. Encrypted HANA is efficient because it mostly uses symmetric-key encryption,

[1] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 1 f.

[2] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2 f.

[3] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 1 ff.

[4] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p.1 f.; Popa /Zeldovich/Balakrishman CryptDB: A Practical Encrypted Relational DBMS, p. 3.

avoids fully homomorphic encryption, and runs on unmodified DBMS software (by using user-defined functions).

3.1.1.1.2 Adjustable Query-Based Encryption

The encodings differ in their encryption and safety. [5] Some methods of encryption are more easily decrypted than others but have to be used for certain queries over the information stored on the DBMS. To prevent all possible disclosures of the encrypted data, Encrypted HANA changes the encryption scheme to some specific data elements of the SQL-queries, depending on the queries observed at run-time. For the efficient implementation of these adjustments, Encrypted HANA uses multiple layers of encryption.

3.1.1.1.3 Chain Encryption Keys to User Passwords

One of the principle ideas of Encrypted HANA refers to chaining the encryption. [6] With this method, each data item on the database proxy server can be decrypted only via a chain of keys rooted in the password of one of the users with access to that specific data. If the user is not logged into the application and if the administrator or an external attacker does not know the password, the description of the data cannot be overruled. To create that chain of keys, Encrypted HANA allows the developer to provide policy annotations to the application's SQL schema, specifying which users have access.

3.1.1.2 Benefits from Encrypted HANA

Encrypted HANA ensures that the sensitive data is never available in plaintext at the DBMS (Database Management System) server. The information sent to the DBMS depends on the classes of computation required by the application's queries; thus, the DBMS cannot compute the encrypted results that involve computation classes not requested by the application.

3.1.1.3 Encrypted HANA's Architecture

Encrypted HANA's architecture consists of two parts: a database proxy and an unmodified Database Management System server. Encrypted HANA uses user-defined functions to perform cryptographic operations in the database management. [7]
For many years, the problem in using encryption was that the programs could not handle strongly encrypted files. Encrypted HANA avoids these problems by intercepting all SQL queries in a database proxy which rewrites queries to execute on encrypted data. The system allows the users to send queries to an encrypted set of data and get the answer they need from it without ever self-decrypting the stored information. [8] The database proxy encrypts and decrypts all data, and changes some query operators while preserving the semantics of the query. The unmodified DBMS never receives decryption keys

[5] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

[6] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

[7] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 3.

[8] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

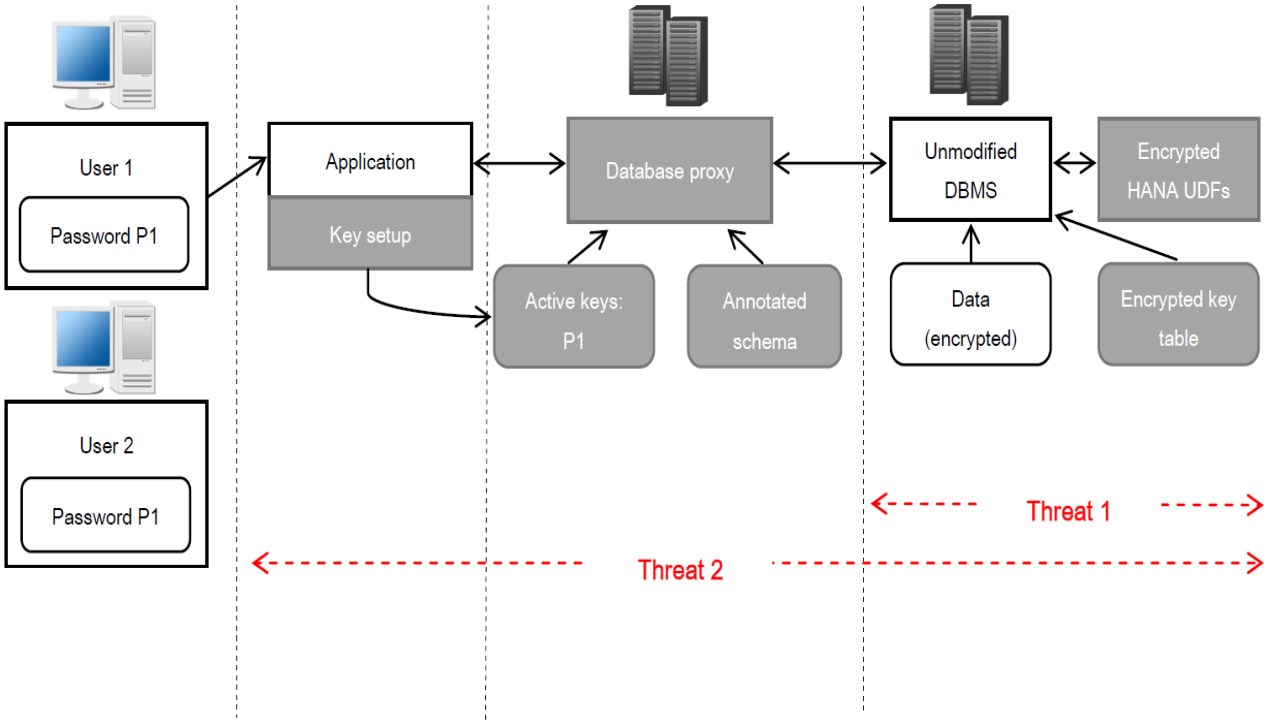

Figure 3.1: Architecture of Encrypted HANA

of the plaintext; thus, the administrator or an external attacker never sees sensitive data. Hence, no one can access the private data without authorization (threat 1). [9]
To protect and shield against the application, the database proxy and the unmodified DBMS would enter into an agreement (threat 2). [10] The developers annotate their SQL schema to define different principals whose keys will allow access to decrypt the different parts of the database. They also make a small change to their applications to provide encryption keys to the proxy. Then the database proxy determines which parts of the database should be encrypted with which key. As a result, Encrypted HANA can guarantee the confidentiality of the data belonging to users who are not logged in the database proxy server during the compromise and who do not log in until the compromise is detected and fixed by the administrator. [11]
Encrypted HANA can protect data confidentiality, however, it cannot ensure integrity or completeness of results. Therefore, it is possible that a malicious administrator or external attacker may gain access to the application, the database proxy or the DBMS, and delete the existing data.

3.1.1.4 Queries over Encrypted Data

As mentioned already, Encrypted HANA guarantees security of personal data. It enables the execution of SQL queries on encrypted data without any need to change the existing applications to work with Encrypted HANA. The DBMS's query plan for an encrypted query is exactly the same as for the original query. Only the operators comprising the query, such as selections, projections, joins,

[9] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

[10] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

[11] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 2.

aggregates and orderings, are performed in ciphertexts and use modified operators in some cases. Encrypted HANA proxy accumulates a secret master key (key), the database scheme, and the current encryption layers of all columns. The DBMS only gets access to an anonymized scheme, encrypted user data, and some auxiliary tables used by Encrypted HANA. Encrypted HANA also supplies the server with Encrypted HANA-specific, user-defined functions that enable the server to calculate on ciphertexts for certain operations. [12] Processing a query in Encrypted HANA takes four steps:

1. At first the applications distribute a query, which the proxy obstructs and rewrites: it anonymizes each table and column name, and, by using the master key, encrypts each constant in the query with an encryption scheme appropriate for the required operation [13]

2. Subsequently the proxy examine if the DBMS should be given keys to adjust encryption layers before carrying out the query, and if so, issues an UPDATE query at the DBMS that summon a UDF to adjust the encryption layer of the appropriate columns [14]

3. In the third step, the proxy returns the encrypted query to the DBMS which it carryies out using standard SQL [15]

4. In the end, the DBMS sends the encrypted query result back, which the proxy decrypts and delivers to the application. [16]

3.1.1.5 End User Applications with CryptDB as an Underlying Technology

Google recently deployed a system for performing SQL-like queries over an encrypted database following the CryptDB design. [17] Their service is able to use the encryption building blocks from CryptDB, rewrite queries and annotate the schema, as in CryptDB. [18] Lincoln Labs also started working with CryptDB and added its design on top for their D4M Accumulono-SQL engine. [19]

3.1.2 Legal Evaluation and Risk Assessment

3.1.2.1 Introduction: Legal Classification of the Involved Parties and the Data Processing Activities

There are two steps needed to run Encrypted HANA. Before queries can be run over encrypted data on the DBMS-server, the data has to be stored on the server. If we assume that the original (plaintext) data provided by the data subject (the affected person) is personal data, then the storage on the server has to be qualified as "processing," according to the DPD. During this first step, the entity which transfers the data is then simultaneously qualified as the controller as well as the processor. The storage of data is expressly mentioned in Article 2 (b) of the DPD as an action considered processing.

[12] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 4.

[13] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 4.

[14] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 4.

[15] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 4.

[16] *Popa/Redfield/Zeldovich/Balakrishman*, CryptDB: Protection Confidentiality with Encrypted Query Processing, p. 4.

[17] See http://css.csail.mit.edu/cryptdb/.

[18] *Popa*, Research Statement, p. 3.

[19] *Popa*, Research Statement, p. 3.

During the second step, when the data has already been stored on the DBMS-server, the DPD will cover the computation of the data, itself. According to Article 2 (b) of the DPD, running the queries means using, aligning or combining, and consulting the data. The controller would still be the client and the processor would now be considered to be the provider of Encrypted HANA, as they would do the actual computation.
Therefore, we will have to carefully assess in the following whether the DPD is applicable, and in particular if personal data is really being affected or sufficiently anonymized by means of encryption:

3.1.2.2 Applicability of the DPD

As the data which is about to be transferred to the DBMS-server will be encrypted and only stored as ciphertext, it may fall out of the scope of the DPD, as it is no longer "personal data" from the perspective of DBMS, etc. As shown above, the effect encryption has on personal data is considered controversial.
According to the absolute approach, an anonymization and, therefore, the elimination of the data's connection to a data subject is only achieved when absolutely no one is able to decrypt the data. The client using Encrypted HANA is able to use the key and decrypt the data stored on the DBMS. Hence, regardless of the encryption, personal data is processed by the DBMS provider on behalf of the user of Encrypted HANA and, therefore, the Directive would be applicable. The absolute approach does not distinguish between those who are able to decipher the data and those who are not. Thus, even though the encrypted data is not readable for the controller (without considerable effort) it is still to be considered personal data as, at least theoretically, one person could access the personal data.
In contrast, the relative approach distinguishes between persons able to decrypt the cipher text (using reasonable efforts) and those who are not. The client using Encrypted HANA to run queries over their data is able to decipher the data they want to store on the DBMSserver. [20] The provider running the DBMSserver, on the other hand, is thus unable to decrypt the data. Moreover, if data is transferred to an entity unable to relate the data to certain persons (i.e. if the encryption standard is sufficient), the Directive is not applicable on this transfer as due to the encryption, personal data is not affected any more. In other words, the applicability of the data protection law depends on the party receiving the data, not the one sending it. [21] It is not important if the party giving away the data is able to relate the data to a data subject as long as the receiving party is unable to.
Furthermore, the legal assessment depends on the standards required for encryption. Following the approach that a state-of-the-art encryption is sufficient, the encryption used by Encrypted HANA should be adequate. If an absolute encryption is demanded, Encrypted HANA's level of encryption might not be considered to be sufficient, especially concerning the insecure ways of encryption referred to in 3.1.1.1.2 of this chapter.
Assuming that state-of-the-art-encryption is seen as sufficient, since the Encrypted HANA-provider would not be able to decrypt the data, the encrypted data would not be qualified as personal data (following the relative approach). From the perspective of the absolute approach, there would be at least one entity able to decipher the encrypted data - the Encrypted HANA client thus the data remains to be personal data (for everyone).

[20] Even if the single employee working with Encrypted HANA might not be able to decrypt all data, the legal entity he is working for, is considered to be the controller. The legal entity has in the end access to the data they work with in plaintext

[21] *Dammann*, in: Simitis, BDSG, par. 3, recital 34.

3.1.2.3 Compliance with Existing and Future Data Protection Law

3.1.2.3.1 Compliance with the DPD

As mentioned above, the Directive can be applicable to the storage and computation on encrypted data with Encrypted HANA depending on the approach that has been chosen in relation to the notion of personal data and the level of encryption. As neither the DPD nor the GDPR implement a clear definition of personal data, and moreover the ECJ still does not provide any clue to which approach the court favors, we have to do the analysis using a two-fold approach in order to take into account a "worst-case-scenario" (if, contrary to our legal opinion, the absolute approach will prevail in the future):
If in spite of encryption, we assume the applicability of the DPD, the principle of prohibition of data processing without explicit consent or legal permission would come into effect.
If the data subject was sufficiently informed about the computation carried out and had freely given their consent, the controller in effect complied with the requirements of the Directive. However, we have to note that the required information concerns every kind of processing data, particularly the purposes, etc.
If consent is not available or not given, the processing may only take place in the case of explicit legal permissions such as outweighing interests of the processor or fulfilling contractual obligations, etc. (in the relationship with the data subject not in relationship with the cloud provider and cloud user). However, this could hardly have been foreseen for all users of Encrypted HANA in every individual case.
Hence, even given the problems of providing sufficient information, it is highly recommended to obtain the data subject's consent for the storage of personal data on Encrypted HANA and the computation that runs with it. With regard to the sufficient information, it is advisable to use only physical machines in certain locations (within the EEA) as DBMS-servers, so the data subject can be told exactly where his data will be stored thus, the problem of sufficient information can be minimized.
Another option to comply with the Directive regarding the transfer of the data to the DBMS-server would be a contractual framework that binds the Encrypted HANA provider legally to the user so that the provider would process the personal data on behalf of the controller in accordance with Article 17 of the DPD "order processing". Thus, the controller would be treated as if they were running the Encrypted HANA-technology themselves. Therefore, there would be no 'transfer' to another entity when the data is stored on the DBMS-server and, no permission would be needed to do so. On the other hand, to run queries over the data, they would still need permission (see 2.4.3.1).
If the DBMS-server is located in a state outside the EU/EEA, then the transfer of data is only permitted either if this state has an adequate level of protection or if adequate safeguards had been adduced, as mentioned above (2.4.3.1). Even if the processing would be carried out on behalf of the controller, as described above ("order processing"), these requirements have to be met by the controller.
Reengineering the query-results in order to obtain personal information should be as difficult as possible. If not, providing a query-result would lead to transferring personal data to the entity receiving it thus bringing the DPD into play again.
The Encrypted HANA provider additionally has to comply with the Directive's requirements for technical and organizational measures to ensure data safety. Unauthorized access to the data has to be prevented.
By making computation of encrypted data possible, Encrypted HANA effectively minimizes the amount of personal data that has to be processed. [22]

[22] *Schaar*, Privacy by Design.

3.1.2.3.2 Compliance with the GDPR

Encrypted Data as not Being Qualified as Personal Data The application of the GDPR just like the DPD depends upon personal data being processed. As already mentioned, the forthcoming Regulation does not, unfortunately, solve the dispute concerning the approach for defining 'identifiable' data. However, recital 23 of the proposals for a Regulation states that all means reasonably *likely* to be used by the controller or by any other person should be taken into account. Whilst the words 'by any other person' might suggest an absolute approach, it is crucial to comprehend that only the means reasonably likely to be used have to be taken into account.
Concerning the specific case of Encrypted HANA, it is not reasonably likely that this state-of-the-art encryption could be overcome with reasonable efforts. Hence, the storage and computation of the encrypted data on the DBMS-server will not be affected by the GDPR, since no personal data will be processed by the Encrypted HANA-provider. It is important to understand that the relative approach (as it is followed here) does not treat the data processed by the DBMS-server provider as "encrypted data" for him as defined by Article 4 Par. 2b of the LIBE-proposal. The DBMS-server provider is not able to identify the affected persons using the encrypted data; therefore, for him the data is not encrypted personal data because it is not personal data, at all (from his perspective). However, from the perspective of those who are able to decrypt the data (the Encrypted HANA user), we have to qualify the data as personal data.
This distinction between different parties processing the data is the main disparity between the relative and the absolute approaches. *Technically*, whilst the DBMS-server provider stores encrypted data, they do not process "encrypted data" according to the relative approaches since there is no personal data anymore.
However, as described above (see 2.2.3) we cannot exclude that in the political process of adopting the GDPR an absolute approach may gain approval. For this worst case scenario the options to comply with the GDPR when using Encrypted HANA shall be described.

Absolute Approach: Encrypted Data as Personal Data The provider of the DBMS may be considered as a processor on behalf of the user of Encrypted HANA, whereas the user's client (client of the controller) would be the affected person, the data subject. In order to comply with the GDPR, the transfer to the provider and the processing done by them should be constituted as 'order processing', as described in 2.3.2. The controller (the Encrypted HANA user) can ensure that appropriate technical safeguards as described in 2.4.4.2 have been taken if the provider is offering an encrypted database, like Encrypted HANA fulfilling then their duties following Article 22 GDPR (see 2.3.2.1). To establish an order processing compliant with the GDPR, a contract between the provider and the user has to be concluded subduing the provider to the user's (controllers) instructions. It should enable the controller to document the processing as Article 28 GDPR demands and obliges the processor to take technical and organizational measures demanded by Article 30 GDPR (see 2.4.4.2). The controller has to report data breaches to the supervisory authority; hence, the contract with the processor should oblige him to inform the controller if such a breach occurs. If the controller has implemented appropriate technological and organizational protection measures such as a high level of encryption technic, according to Article 32 GDPR the controller does not need to inform the data subject about the breach (see 2.5.5).
A risk analysis of the potential impact of the data processing according to Article 32a GDPR (but only included in the LIBE-proposal) has to be carried out by the controller; moreover, according to Article 33 GDPR, under certain circumstances a data impact assessment should be carried out by either the controller or the processo. In this case, the contract regulating the order processing should clarify who will be responsible for this task.

In order to enable the user of Encrypted HANA (the controller) to comply with their duties, the provider the DBMS should apply for a certification (a privacy seal as described in 2.3.2.4) so that they can guarantee the processing in compliance with all technical and organizational measures required by the regulation. Such a certificate or privacy seal should ensure that all possible clients of the Encrypted HANA solution can comply with the Regulation's requirements to control the processor (the Encrypted HANA provider) processing on their behalf. If the provider has been certified the users of Encrypted HANA would not have to monitor and prove the compliance on their own, but rather be able to check the validity of the certificate (the data protection seal). For a DBMS-provider aiming at offering his service to multiple users, a certification is therefore highly recommended.
If the provider of the DBMS-server falls under the jurisdiction of a third country an order processing is not impossible (see 2.3.2.6); however, certain measures have to be taken in order to ensure adequate safeguards to comply with the GDPR. As described under 2.4.3.2, those measures may be an adequate acknowledgement of the EU Commission (Article 42 GDPR), 'binding corporate rules' (Article 42 Par. 2 a) GDPR), a European Data Protection seal (Article 42 Par. 2 aa) GDPR) (see 2.3.2.4), standard contract clauses (Article 42 Par. 2 c) GDPR) or contract clauses approved by a supervisory authority (Article 42 Par. 2 d) GDPR). If the controller is not able to ensure those measures, he has to obtain the data subject's consent or the processing has to take place on the basis of one of the permissions in Article 44 GDPR. The data protection seal can also be used in order to provide evidence for a lawful transfer to a third country.
Hence, the DBMS-server provider's chances of attaining a certification are even greater if they are-based in a third county. According to Article 43a of the LIBE-proposal, the controller and the processor have to notify the supervisory authority and obtain prior authorization before disclosing personal data to a third country authority because of a third country's judgment, a court tribunal or a decision of an administrative authority (see 2.5.1).
Moreover, the principle of privacy by design will be explicitly included in the data protection act. In addition, data controllers have to continuously check and improve their systems if new challenges from the perspective of privacy (new risks, etc.) come to the light. Hence, if a new system using Encrypted HANA is established, privacy issues have to be dealt with when developing the new technology as well as their usage, from the implementation of its software on the client's system and the application server to the encryption and storage of data on the DBMS and the computation over this encrypted data. Not only the accuracy, confidentiality and integrity, but also the physical security of the system has to be kept in mind. It should be ensured that the client is able to use the system while giving away the least amount of personal data possible. All those requirements are directed by the encrypted database system, Encrypted HANA.
According to Article 14 GDPR, before the controller collects their clients' data, the controller will have to provide the clients with information. Therefore, among other information, the controller will have to inform the clients who the processor (the DBMS-server provider) will be, where the server will be located and for which purposes the controller intends to process the data (see 2.4.2.4). If the client exercises their rights according to Article 17 GDPR (the so called 'right to erasure' or 'right to be forgotten'), the user of Encrypted HANA will have to ensure the complete deletion of the clients data from the DBMS. If the Article 35 GDPR's 'trigger' of processing data relating to more than 5000 data subjects in a consecutive 12-month period is pulled (only in the LIBE-proposal), the user of Encrypted HANA and the provider of the DBMS-server have to designate a data protection officer. In contrast to this, the proposal of the Council leaves the designation of a data protection officer directly to the discretion of the controller or the processor themselves except as required otherwise by Union or Member State law, see Article 35 Par. 1. Thus, the provision regarding the designation of a data protection officer in the Council's proposal on a voluntary basis is a huge relief for cloud users and providers, since 5000 data subjects can be reached very quickly when processing big data' in the

cloud.

3.2 Secret sharing

3.2.1 Sharemind

3.2.1.1 Functioning

Sharemind is a cloud-ready data analysis system for securely processing confidential information. By its design, it provides security without the risk of an insider attack. The input data never leaves the hand of the owner, only final results of the computation are shared with partners. It provides privacy because private information can be processed without compromising the data subject's rights and convenience of use, Sharemind can be run in a cloud and is compatible with existing tools. [23] Sharemind is designed to be deployed as a distributed secure computation service that can be used for outsourcing data storage and computations by splitting personal or secret information between a minimum of three servers, to ensure the security of the data. [24] Sharemind Server is an application and a database server that uses secret sharing technology to store and process information without leaking it. To achieve the best efficiency and privacy, three servers must be deployed by separate organizations that will likely not collude. The system is capable of performing computations on input data without compromising its privacy. Moreover, because the system is based on solid cryptographic foundations, it can process data whilst shielding data even from access of the server administrator.

3.2.1.1.1 Architecture of Sharemind

Sharemind uses Secure Multiparty Computation and secret sharing to protect the personal data of the user. [25]

3.2.1.1.2 Secure Multiparty Computation

Secure Multiparty Computation refers to a field of cryptography that deals with protocols involving two or more participants, who want to mutually compute a useful result. [26] Every party will provide an input value and learn only the result of their own individual value so that nobody is ableto access all the information. [27] In the case of data aggregation algorithms, it is generally not possible to learn the inputs of other parties from the result. [28] Figure 3.2 shows the data storage process with three servers. The data is collected from the users or exists already on other servers and is sent to the three servers. A data donor distributes the data into parts/shares using secret-sharing and sends one random share of each value to a single server.
If the Sharemind technology is used to compare information from two entities in such a way that no one knows the others values, then both entities function as data donors. It can be necessary that one donor specifies what kind of information the other donor has to provide from its database, for example the IDs of the persons whose data is about to be compared).

[23] *Cybernetica*, Sharemind Your secure service platform or data collection and analysis, p. 2.
[24] *Bogdanov*, Sharemind: programmable secure computations with practical applications, p. 30.
[25] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 5.
[26] *Bogdanov*, Sharemind: programmable secure computations with practical applications, p. 24.
[27] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 2.
[28] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 2.

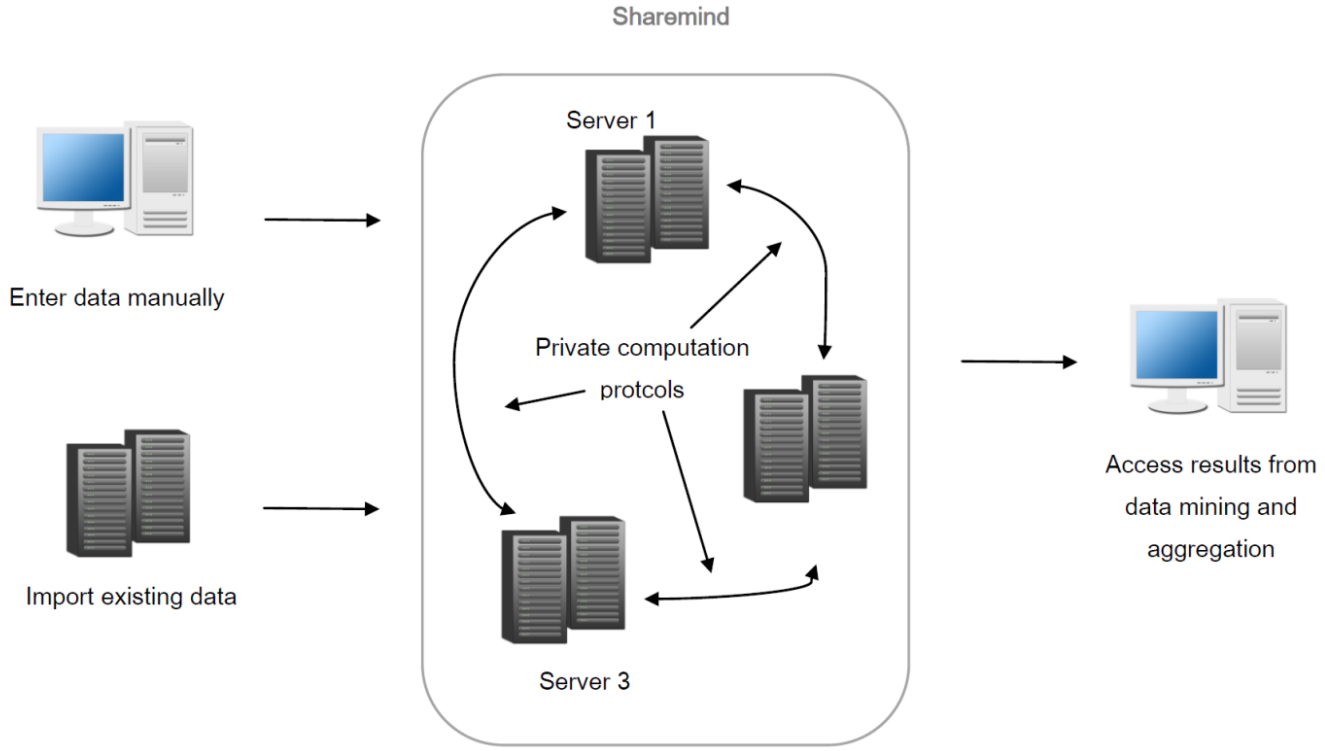

Figure 3.2: Function of Sharemind, three Sharemind Servers were deployed by three independent organizations, the information is divided between the three and every one of them receives a part of the information and works with it. At the end every organization sends his results back to the client.

The separation of servers between input donors and servers is useful as it does not force every party to run secure multiparty computation protocols. [29] After the data has been transmitted and stored, the server can perform computations on the shared data; however, the server does not share the information with other servers. This is done so that none of them can reconstruct the input values. [30] The number of three servers is used for efficiency and is needed to guarantee the security during the computation; otherwise, it would be too easy to reconstruct the data. Moreover, an increase of servers reduces any risk of collusions. Secure multiparty computation protocols specify which messages the server should exchange in order to compute new shares of the value that corresponds to the result. [31] After finishing the computation, the results of the servers are transmitted and published to the client of the computation (Figure 1: the user). The servers send the share of the results to the user who reconstructs the real result. [32]

3.2.1.1.3 Secret-Sharing

In Sharemind, each party will receive one share of every secret value. The original secret can only be reconstructed by collecting all the shares of a value and adding them up using the addition operation in the ring. [33]

[29] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 3.
[30] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 3.
[31] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 3.
[32] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 3.
[33] *Bogdanov*, Sharemind: programmable secure computations with practical applications, p. 34.

3.2.1.1.4 Use Case: Secure Floating-Point Arithmetic and Private Satellite Collision Analysis and Sharing of Medical Data

One prominent case regarding the use of Sharemind refers to the delicate issue of sharing information about locations of satellites, etc., in order to avoid collisions. In the orbit, nearly 7,000 spacecraft are flying around the Earth. [34] In the year 2009, two communication satellites belonging to the USA and Russia collided in orbit because the nations did not talk about the trajectory of their satellites. [35] The two orbital planes intersected at a nearly complete right angle, resulting in a collision velocity of more than 11 km/s.[36] The locations and orbits of the communication satellites are sensitive information; hence, governments or private enterprises want to protect this data. [37] Whilst a trusted third party gathering all the data and performing analysis could be a solution, this party would still need disclosure of information of all the parties involved, thus endangering privacy and security of data. [38] By its design, it ensures secrecy of information by using Multiparty Computation and Secret Sharing Sharemind, and can, therefore, be used for calculating the probability of a collision between two satellites. [39] Using secure multiparty computation instead of a trusted third party could solve the problem of disclosure, and would be more practical. The satellite operators would choose three independent parties as the data host/servers. [40] Subsequently, the operators would secret-share their data and upload the shares to the three servers. At this point, collision analysis is a collaborative effort between the three hosting parties and the satellite operator, who can query the results of the analysis. [41]
The same method could potentially be used for many types of sensitive information. For instance, for private health data, in order to ensure that no unauthorized person is able to obtain health information.

3.2.1.1.5 Difference between Sharemind and Encrypted HANA

The two presented solutions in this report, Sharemind and Encrypted HANA, use different techniques to avoid handling personal data. Sharemind breaks each value down to several random fragments, so that the information is anonymized. Encrypted HANA works with encrypted queries and encrypted data so that the administrator or an external attacker does not have access to the personal information.

3.2.1.2 Legal Evaluation and Risk Assessment

3.2.1.2.1 A Legal Classification of the Involved Parties and the Data Processing Activities

Sharemind requires three steps: the donors have to be informed whose data shall be provided; the data has to be divided; and then stored on the different servers. If it is necessary for one data donor to specify whose information the other donor has to provide, this has to be considered as processing of personal data in a legal sense. It would then be inevitable to identify the data subjects whose information is needed for the purposes of computation. The transfer of the ID information would need the data subject's consent or an explicit legal permission. Alternatively, all data can be loaded to Sharemind and securely joined to them using ciphertexts. This would reduce the amount of personal

[34] *NASA*, NSSDC Master Catalog, last accessed February 16, 2014 available at: http://nssdc.gsfc.nasa.gov/nmc/.
[35] *NASA*, Orbital Debris Quarterly News, 2009, vol. 13, Issue 2, 1 f.
[36] *NASA*, Orbital Debris Quarterly News, 2009, vol. 13, Issue 2, 1 f.; Kamm/Willemson, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 6.
[37] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 8.
[38] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 8.
[39] *NASA*, Orbital Debris Quarterly News 2009, Vol. 13, Issue 2, 1 f; Kamm/Willemson, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 1.
[40] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 8.
[41] *Kamm/Willemson*, Secure Floating-Point Arithmetic and Private Satellite Collision Analysis, p. 8.

data shared during the work done with Sharemind and, therefore, comply with the principle of Privacy by Design (if the new Regulation comes into force).
Before the data is stored on the different servers, it has to be divided. This process must be carried out through personal data in plaintext. The problem regarding the applicability of the Directive is not the qualification of this data as personal data, under the terms of Article 2 (a) of the directive; it is, rather, whether dividing of personal data still has to be seen as processing personal data. In Article 2 (b) the directive mentions the alteration of data as processing. However, as described above, this refers to the alteration of content, not of its appearance. [42] The secret-sharing of personal data by dividing it does not fall under the Directive's scope.
Once the data has been divided, it will be stored on the different servers. If the data chunks were to be considered personal data (according to the absolute approach which is in principal not followed here, see 3.2.1.2.2), this kind of processing also would be qualified as processing of personal data. Even so, we have to state again that we consider this approach to be extending the scope of the DPD in an tremendous way.
Concerning Sharemind, it is not as easy to determine who the controller is: The role of the controller is not fixed to one of the many participants (at least two data donors, three data server providers and the user). The entity (user or in other terms "client") who is in charge will likely be the one who has initiated the research carried out with Sharemind and decided to use Sharemind. According to the relevant criteria regarding the definition of controller" in particular, who can determine the purposes of data processing, etc. it shall be the user (the client) who initiates the process.
The provider of Sharemind can be one of the server providers who has the technical know-how to use the technology. However, this does not imply that they decide how the data processing is done with Sharemind. They should, rather, be seen as a usual provider of a cloud solution. In a way, Sharemind can thus be compared to Software as a Service (SaaS).
If more than one participant is determining the purpose of the processing, they each have to ensure compliance for all processing that is carried out.
Not easy to be answered (and still generally unresolved) is the issue of joint controllership if the "controllers" (the users, the clients) have initiated the data processing in such a manner that they have acted together but did it so without being aware of cooperating. Sharemind brings the "controllers" together but shields their identities from one another; hence, there is no common platform for them to decide and jointly determine the data processing. Thus, according to common legal thinking and notions of "joint partnership," etc., the crucial element for being jointly responsible for a data processing is clearly missing. However, we have to note that the legal discussion has only recently begun, concerning the interpretation and the necessary elements of "joint controllership"– in particular, if this notion has to be interpreted in the same way as traditional partnerships (like in corporate law). Currently, some sort of acting together is still required. Hence, in the case of Sharemind, the users will not be considered as "joint controllers," but rather as the controller for each section of data processing (which is difficult to handle regarding the obligations of data controllers).
However, we have to note that the ECJ ruling on the Google Spain case that was cited above had a very broad notion of joint controllership and failed to deliberate the elements more in-depth. Thus, we are currently confronted with legal uncertainty in a "worst-case scenario," we should take into account the fact that all users (clients) of Sharemind will, eventually, have to be considered as joint controllers, so that they are each responsible for actions to be taken, in light of the DPD.

[42] See also *Gola/Klug/Körffer*, in: *Gola/Schomerus*, Bundesdatenschutzgesetz, par. 3 recital 30.

3.2.1.2.2 Applicability of Data Protection Law

Multiparty computation is advantageous due to the fact that simply random fragments of personal data are used. The original data can only be restored (and thus turns into personal data) if all fragments are put together. Hence, it is crucial to determine whether the DPD is applicable to the computation over data fragments. The division of data cannot be treated as a traditional form of encryption. Thus, the controversy regarding the sufficient level of encryption is not relevant either.
The qualification of split data is still new to data protection law, yet, well-known in intellectual property law. By splitting works protected by copyright into 'chunks', as in peer-to-peer sharing, people try to circumvent the protection of the work provided by copyright law. Although there are differences between copyright law and data protection law, [43] one important parallel can be drawn. For example, if a copyright protected work is split in many parts, and those parts can be perceived, the copyright protection still affects the single parts. The single page of a book, for example, is protected just like the whole book. If the chunks of a copyright-protected work cannot be used for perception of the work (as is the case if archive files like .zip or .rar files are shared via peer-to-peer sharing for example) without having all other parts of the work, some authors argue that the copyright protection does not apply for a single part. [44]
This idea can be used to evaluate if one part of a secret-shared file is still personal data. Without the other two parts, this file cannot be read in any way. One fragment itself does not contain information regarding a person and should not be seen as personal data. For someone looking for information about a certain person, this fragment would be useless. Only if all fragments of the data were gathered and put together the directive would be applicable. Theoretically, all server providers may collude and reengineer the personal data. However, this is highly unlikely since the providers of the server, themselves, have a high interest in ensuring safety and confidentiality of Sharemind and should be legally bound by contract. Once again, from the stance of the relative approach, the unreasonable chance of collusion leads to ruling out the applicability of the Data Protection Directive. Concluding, we have to point out that the perspective from an absolute approach would differ in taking these chances into account, thus applying the Data Protection Directive.
So, we have to analyze the legal situation in the sense of a "worst-case scenario" if the absolute approach would prevail.

3.2.1.2.3 Compliance with Data Protection Law Now and in the Future

Compliance with the DPD As outlined already for HANA, it is highly recommended to obtain the data subject's (explicit) consent; if not, there has to be a specific legal permission for processing the data such as fulfillment of contractual obligations.
The consent given must be informed and given with free will on the basis of sufficient information; the same criteria as above apply. If the servers are in a third country, an adequate level of protection has to be guaranteed by the controller using safeguards, as described above. If the servers are within the jurisdiction of the Directive, a processing on behalf of the controller, in the sense of Article 17 of the Directive, could be ensured by a legal framework between the data donors, as controllers, and the Sharemind-provider, as the processor, ("order processing").
Especially for Sharemind, it is important that the output produced by the multiparty computation cannot be easily re-identified, since the user of Sharemind might be an entity completely different than the original data donors. For instance, data donors can be persons which are interviewed in a statistical

[43] Copyright law aims to protect the right holder against unlawful reproduction of his work whereas data protection law protects the data subject's right to decide what is done with its personal data.

[44] With further references regarding the copyright-law based discussion concerning illegal sharing of chunks see Heckmann/Nordmeyer, CR 2014, 41 (43).

query, such as students etc., an user can be a research (or commercial) organization that would use the data in order to combine with other data which stem from other parties. In such a scenario, the "user" would be qualified as a data controller. Moreover, further provider may also be involved, which raises the issue of joint controllership again; for instance, a provider of statistical methods and software who determines the data processing. Furthermore, the computing parties have to check that secret data is not published or made accessible by mistake, thus creating a joint controllership. The user (data controller) should be bound by contract to refrain from using Sharemind in a way that would reveal information about the persons which the data donors provided in the first place (for example, combining the donor's data with other data derived from Sharemind, thus creating new personal data). Considering the principle of Privacy by Design, this issue should be solved when setting up the contractual framework needed for Sharemind by forbidding the re-identification of persons using contractual penalties as an organizational measure as required by the law (see 2.4.4.1). In order to deter the user of Sharemind from re-identification, the use of auxiliary information to re-identify data subjects and the consequences of doing so have to be disproportionate (unreasonable) in comparison to the value of the personal data they would produce.

Compliance with the GDPR The GDPR will only be applicable to the data processing carried out by the participants of Sharemind if the secretly shared data would be considered 'personal data'. Due to the differences between traditional encryption and secret sharing, it is unlikely that a single part of a secretly-shared date enables the identification of a person, as more than a single key is needed to decrypt the date and they are distributed among different entities with strong interests in keeping the data confidential. A collusion of those parties is highly unlikely. Therefore, we do not think that the upcoming Regulation will be applicable to the computation over secretly shared data, even under the assumption that an absolute approach may prevail under the GDPR (see 2.2.3). To fully assess the possible legal risks the GDPR's main issues with regard to Sharemind shall be described in the following.

The role of the controller is not assigned to one of the participants; rather, it may change for every case Sharemind is used. Even two joint controllers are possible. To ensure a lawful processing under the GDPR all processors involved should process the data on behalf of the respective controller (the user, the client). Hence, before Sharemind is used the parties involved should enter into contractual relations ensuring the requirements described under 2.3.2 are met. It is the controllers' responsibility to bind all other participants legally (in the sense of a contract) and to ensure the necessary technical and organizational measures are implemented. Again, a certification of the processing parties as described in 2.3.2.4 is recommended. The proposal of the Council even states that certifications may be used as an element to demonstrate compliance with the obligations of the controller and of the processor.

If an 'order processing' takes place compliant to Article 22 ss. of the GDPR the controller has either to obtain the consent of affected persons or to benefit from an explicit legal permission. Therefore, like under the DPD (see 3.2.1.2.3) it is highly recommended to obtain the data subjects consent (see 2.4.2.4).

Akin to the DPD, the GDPR addresses, like the DPD, joint controllership. Both controllers are responsible for the use of Sharemind. They will be bound by Article 24 GDPR to enter into an arrangement that clarifies each controllers' duties, e.g. the information of the supervisory authority in case of a data breach (Art. 31 GDPR), or eventually (if necessary) the appointment of an data protection officer. The arrangement has to be made available to the data subjects, so they can know to whom they can turn if they want to exercise their rights according to Article 17 GDPR (see 2.5.3). Sharemind makes it possible for computation to be carried out over data without the computing parties learning it. If it is ensured that the system's output cannot be re-identified without disproportionate efforts, then the

goals of Privacy by Design in Article 23 GDPR can be met.

3.2.2 Secure Collaborative Statistics in Credit Rating

3.2.2.1 Functions

3.2.2.1.1 The Basic Concept

Secure Multiparty Computation (MPC) can be used to facilitate complementary decision support in a traditional credit rating. This business case involves small- to medium-sized Danish banks and an accounting firm. They merge their confidential data by using MPC to create a database. An implemented MPC-based LP-solver is used to compute relative performance analysis of the bank's customers directly on the secretly shared data set. Thus, traditional credit rating can be complemented by means of relative performance evaluations. It is difficult for banks to obtain traditional accountancy information on 'peer' farms, since most farmers are not required to publish and disclose such information (unlike many other businesses). However, their accountancy information which is processed by an accounting firm can be used by involving this accountancy firm in a way that shares the information with banks, without giving them direct access to the personal information of the farmers.
The relative performance analysis of the farms is computed using linear programming, which is one of the most basic and most useful optimization tools. It is widely used in operational research and applied micro economics.
Like in the other use cases mentioned here, none of the involved parties is required to disclose their data to others. Once again, a trusted third party may solve the problem, such as credit scoring agencies in Germany. However, such a solution may turn out too expensive or, on a smaller business scale, too complex to reach.
Another solution is again provided by Secure Multiparty Computation (MPC) which allows two or more parties to compute any function without leaking any additional information, other than the output of the function. In this scenario, an LP-solver using MPC primitives has been implemented. Instead of a third trusted party, the MPC coordinates the private information according to a comprehensive protocol; the MPC is used like a trusted third party. In contrast to a trusted third party, MPC does not require one entity to learn all inputs. As the parties involved in MPC are interested in keeping their data confidential, the risk of a privacy breach is lowered compared to a coordination by an uninvolved third party. Banks and the accountancy firm thus provide confidential data without the other party learning these data. An MPC-based LP-solver is used to compute over the datasets to produce a relative performance analysis of the bank's clients– in this case, 'peer' farms. This information can be helpful in evaluating the credit rating of a farm, as well as for evaluating the bank's portfolio of farms.
In the basic scenario, two parties (a bank and an accountancy firm) hold specific data that the other party shall not learn. As long as data is just stored, no encryption of the data is needed to prevent the other party from discovering it because solely the party holding the data can access it. Only when the secure computation of the benchmark takes place the data will be secretly shared between the two servers.
Every farm has a 'CVR Number' an identifier provided by the Danish Central Business register, i.e. the central register containing primary data on all businesses in Denmark. [45]The bank will be able to obtain a performance analysis of its client a specific farm that has either been a client before or asks the bank for a loan by providing the CVR Number to the software. Only the bank knows which farm is one of its clients.

[45] See https://cvr.dk/Site/Forms/CMS/DisplayPage.aspx?pageid=21.

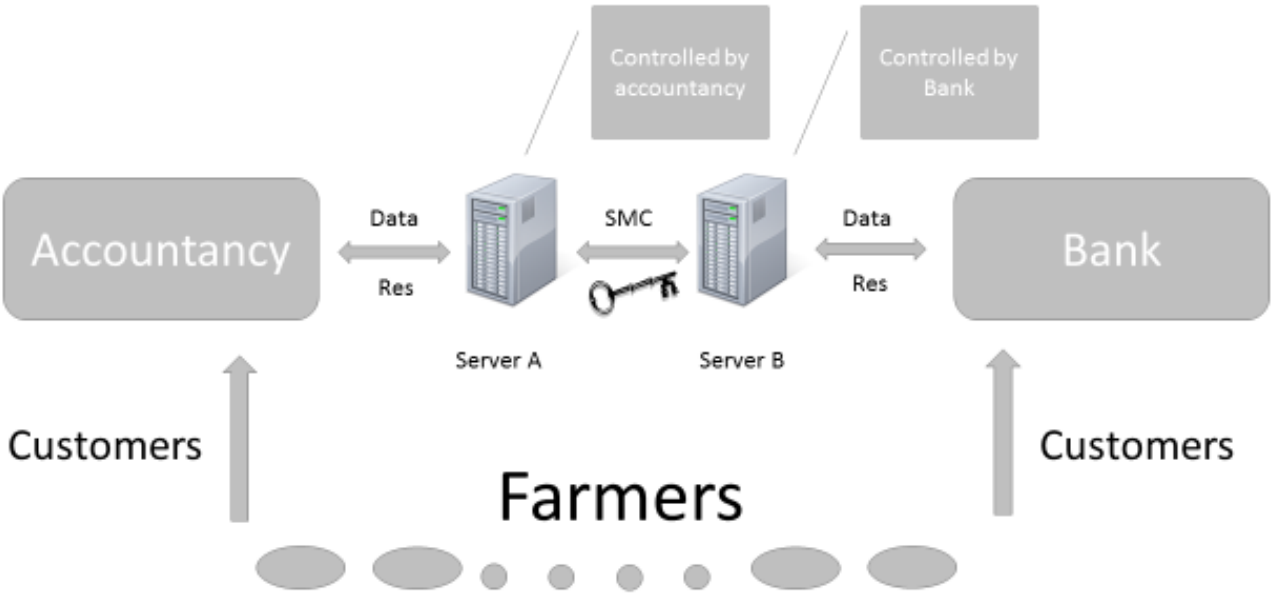

Figure 3.3: The basic functioning of secure collaborative statistics in credit rating using linear programming

3.2.2.1.2 The Systems Output: Secure Complementary Credit Ranking

The reason a bank may be interested in using this system is that reliable credit scoring and evaluation of a farm is needed in order to meet the banking regulations, concerning great amounts of lending. Since a bank only has information (typically limited information) about its own customers, smaller banks may lack sufficient data to conduct a proper credit rating analysis of its customers. The software provides a complementary analysis of the firm's relative economic performance, instead of trying to predict the risk of failure in fulfilling its financial commitments to the bank (ie. repaying a potential credit). The bank realizes that this is an efficient scoring method which measures the performance of a farm against its most important peers. The application uses benchmarks (ie. comparing the performance of one unit against that of best practice) with the Data Envelopment Analysis approach. [46] Data Envelopment Analysis can be formulated as an LP-problem. The performance analysis is presented to the bank as a single value without any information that would reveal further information about the farm.

3.2.2.1.3 The Possible Variations of the System

In a straightforward scenario, each involved party is using an Amazon EC2 as a server under the assumption that Amazon gives them full and exclusive control of their respective EC2 instance. However, the bank and the accountants may either use another cloud provider's service or their own IT-resources.
Moreover, the system can be extended to more parties. For instance, instead of having server A run by

[46] A frontier-evaluation technique that supports best practice comparisons in a multiple-inputs multiple-outputs framework, see *Charnes/Cooper/Rhodes*, European Journal of Operational Research 1978, 429 ff.; *Charnes/Cooper/Rhodes*, European Journal of Operational Research 1978, 339.

the accountants house and server B run by a bank, the Danish Bankers Association could run server B, so that every bank that is part of the association, could use server B provided by the association. Since the information is confidential, banks would not want the association to learn this information. Once again, this problem can be solved by secretly sharing the banks data between the two servers. None of the controlling parties of the servers would thus have knowledge of the bank's data. Trust is based on the assumption that the involved parties will not collude – the basic principle of secret sharing. The difference to the basic model described above (see 3.2.2.1.1) refers to the secret sharing of data even when the data is simply stored in contrast to secret sharing only when data is being computed. The computing would be done by MPC, not by the Banker's Association, or a trusted third party.

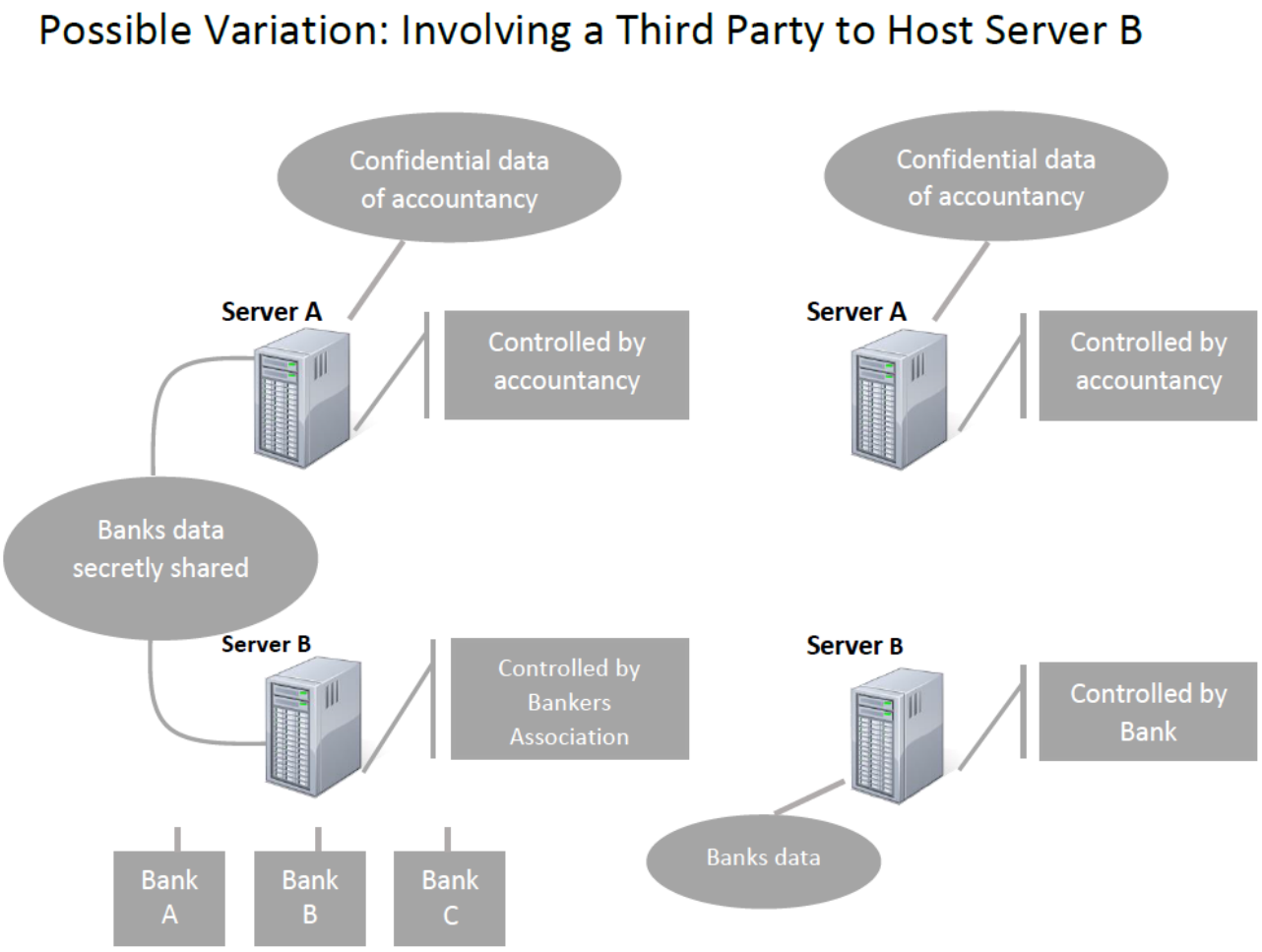

Figure 3.4: Possible Variation of the basic principle by using a third party to host server B

3.2.2.2 Legal Evaluation and Risk Assessment

3.2.2.2.1 A Legal Classification of the Involved Parties and the Data Processing Activities

The parties involved are the accountancy firm, the bank, the farms, and, eventually, the Danish Bankers Association and Amazon. The bank and the accountancy firm both determine the purposes and means of the processing, as they decide which data will be computed. Therefore, as joint controllers, they are mutually responsible for the processing of the data, Article 2 (d) DPD. If the servers are hosted using Amazon EC2 Instances, then data will be transferred to Amazon and the MPC will be run based on Amazon's Instances. Evidently, Amazon does not determine the purposes of the data processing and, therefore, is not qualified as a controller, but rather as a processor. The same would apply if the Danish Bankers Association would provide the hosting service. If the accountancy firm and/or the banks use their own servers they have to be considered to be controllers and processors at the same time.
In the straightforward scenario, there are four relevant data processing activities: the transfer of the data to the servers, the storage of the data on the two servers, the computation of the data stored on the two servers via MPC, and the production of the performance analysis. If Amazon EC2 Instances are used to run the two servers, a transfer to Amazon as a third party is needed. If the Amazon Instances are hosted on servers outside the EEA/EU, a transfer to a third country is implied. If the two servers are run on physical machines controlled by the bank and the accountancy firm, [47] the transfer of the data to the servers and the storage are not relevant since the entity storing the data will be the entity who is controlling the processing in the first place. This processing would be carried out internally and thus does not face legal challenges.
If the Danish Bankers Association is involved in a transfer of the bank's data to the server run by the association, then the same criteria and principles as for Amazon can be applied.
As explained earlier, the dividing of the data in plaintext cannot be considered as data processing (see 3.2.1.2.1).
Whereas the MPC runs over secret shared data that existed before, the computation of the system's outputs produces new data. If those values are to be considered personal data, the banks using the system would be collecting personal data. This, as well, is data processing that requires an explicit legal permission or the consent of the affected person, here the farmer.

3.2.2.2.2 Applicability of Data Protection Law

The Data Protection Directive only regulates the processing of personal data, which is data that relates to an identified or identifiable natural person. A natural person is a 'normal' human being, and not a company, a corporation or an association. Those are 'legal persons' by law. The processing of data concerning legal persons is not affected by the European Data Protection Directive:

> Recital 24 DPD: "(24) Whereas the legislation concerning the protection of legal persons with regard to the processing data which concerns them is not affected by this Directive"

The data processed in this scenario affects farms accounting and production data, their identities and valuations of their assets. If those farms are organized as companies in the sense of a legal entity, a legal person (i.e. not run just by a single farmer as a private individual or as a partnership), their data can hardly be considered to be personal data. Nevertheless, the business case deals with small farms and the system described might be useful for the financial performance evaluation of such farms (because there is little to no accountancy data of these farms publicly known). Most of them are privately owned and are not organized as a legal person/entity. However, Denmark opted to include

[47] Provided that the servers of banks and accountancy firms are based in the EU.

legal persons' data in the scope of its data protection law. Moreover, courts in Germany applied the DPD to legal entities run by a single natural person (one-man-company) in the case the data reflects information of the natural person. Hence, there are substantial reasons to assume that natural persons may be affected, specifically, in the business case of farmers and bank loans. Therefore, European data protection law should be taken into account when the system described is applied.
As in the Sharemind case, data will be secretly shared for secure collaborative statistics in credit rating. The secretly shared data fragments are no longer giving away information without the values stored on the other server. Due to the opposing interests of the involved parties (banks and the accountancy firm), it is highly unlikely that a party will learn the other party's information. As showed above, secretly shared information should not be considered to be personal data anymore (see 3.2.1.2.2) taking also into account the absolute approach.
In contrast, however, to the Sharemind case study not all data are secretly shared all the time. The data is stored on the two servers in plaintext so that all processing concerning this data have to comply with the DPD.
Another disparity with the Sharemind case study refers to the quality of the system's output: whilst with Sharemind, a researcher is not able to re-identify the affected persons and the results are supposed to be anonymized data (see 3.2.1.2.3.1) the case of farmers and banks refers to information being provided for a certain farm: The bank wants to obtain a financial performance scoring in order to use it in addition to the traditional credit rankings. At least for the bank who is aware which farm is involved these results are clearly personal data. Therefore, the data protection law is applicable by way of the collection of personal data. If an absolute approach would be applied to the definition of personal data, the performance analysis a single value without any identifiers would have to be considered personal data, even for persons who do not know to which farm this analysis refers.

3.2.2.2.3 Compliance with Existing and Future Data Protection Law

Compliance with the DPD Under the assumption that the data protection law is applicable to the business case both the absolute and the relative approach will lead to the same outcome for some processing activities. Thus, compliance with the DPD has to be ensured.
Using Amazon Instances to host the two servers raises the problem of data protection for the transfer of the data to the servers. The data will be stored in plaintext on those servers and only be secretly shared during the MPC. Hence, even by following the relative approach, this data has to be considered personal data. Since Amazon will be considered to be a third party, the bank and the accountancy firm would need the affected farms' consent or a legal permission for the transfer to Amazon. The consent must be given freely, incorporating all requirements mentioned before (see 2.4.2.3). [48]
In addition, as the system would be used by the banks to assess the credit worthiness of a farm, the transfer of the data can be justified on the grounds of Article 7 lit. (b) or (f) DPD: the processing (here in the form of transferring) can be necessary for the performance of a contract (the loan contract between the farm and the bank) or necessary for the legitimate interests pursued by the bank. As the controller -the bank solely has an interest in lending money to farms financially stable enough to loans it back on-time and with interest.
However, those legal permissions require a balance of interests; the interest of a farm in not having its personal data transferred to a third party may outweigh the interest of the bank in using the system run through Amazon.
The same results for the bank could be produced if the bank and the accountancy firm would run the servers on their own physical machines, without including Amazon. For the accountancy firm,

[48] Note that since the actual processing - the MPC - is done over secret shared data this data is not personal data anymore in our opinion, see 3.2.1.2.3.1.

which also would be transferring personal data to a server run through Amazon the interest would be even weaker in comparison to the affected farm's interest, since the accountancy firm would not gain a direct advantage (other than making money out of the service or that the bank recognizes the most efficient farms, which would come at the expense of the less efficient farms and therefore might possibly weaken the accountancy firm's interest even further). Thus, it is hard to justify a transfer to Amazon on the grounds of Article 7 lit. (b) or (f).
Moreover, in this version of the scenario, Amazon could hardly be considered a processor on behalf of the controller (see 2.3.1.4), since there would be no contract legally binding Amazon to process data. The processing done with MPC would only take place on Instances run on Amazon servers with Amazon doing nothing more than providing the cloud infrastructure for the system. The controllers would not benefit from the privileged status of a processing carried out on behalf of the controller.
In addition, the transfer to servers outside the EU requires a specific consent by the affected farm or an explicit legal permission concerning the transfer of data to a receiver under the jurisdiction of a third country (see 2.4.3.1). The Amazon Instances may be run on physical machines within the USA, which would allow a transfer on the grounds of the principles for data transfers to third countries (see 2.4.3.1) however, incorporating all problems already mentioned (2.5.1).
These legal problems can be avoided if the bank and the accountancy firm use physical machines of their own (located in the EU). In this case, there is no transfer to a third party. The only data processing other than the MPC would be the storage on the bank's and the accountancy firms' own servers. The affected farms' interest in not having its data transferred to a third party would not have to be considered in the balance of interests in this scenario. It is more likely that the banks interest in a valuable financial performance analysis before providing a loan to a farm will outweigh the farms interests in that case (note the still weaker interest of the accountancy firm, see above). From a legal perspective the basic scenario using Amazon EC2 Instances clearly involves risks that can be avoided. Still, encryption of the data before transferring it to Amazon may solve the problem if Amazon is to be involved as the DPD will not be applicable to the transfer (according to the relative approach, see 2.2.1).
If the Danish Bankers Association is to be involved (see 3.2.2.1.3), the data transfer to the association has to be evaluated. Since the Association is not allowed to learn the bank's information and would only be used as a means to simplify the system, if more than one bank would want to use it, then the bank's data will be secretly shared between the two servers (one controlled by the Association, one controlled by the accountancy firm). Hence, the two servers have to be treated like the data mining servers in the Sharemind use-case (for the banks data), see 3.2.1.2.3.1. Therefore, neither the storage nor the computation of the bank's data should, would require a legal permission or consent by the affected farm. It is highly unlikely that the two parties controlling the servers would collude so that one party cloud learn the other party's information in this scenario. If we assume a worst-case scenario where even storing and computation over secretly shared data would fall under the scope of European data protection law, the Association can be considered a processor on behalf of the controller, in this case the banks. The Association would run one of the used mining servers for the storage and for the MPC; however, the banks would still decide over the purposes and means of the processing. An 'order processing' by the Association would be possible if an appropriate contract would be drawn up. The difficulties arising from a cloud-provider functioning as a processor would not occur if the Association (as a 'normal' processor) would process the data. The legal requirements could be met (see 2.3.1.4.3) so that 'order processing' could be assumed. As an organizational measure (see 2.4.4.1) the Association should also agree to an enforceable non-collusion clause in the contract regulating the order processing.
Concerning the system's output the bank will be able to obtain a scoring value that provides information about the financial performance of a certain farm, compared to the performance of other farms.

The bank provides the farms CVR number for the system. The entity who knows to which farm the CVR number belongs to also knows which farm the systems' output value was computed for. Therefore, this value should be considered personal data (if the farms' data is regarded as personal, see 3.2.2.2.2). According to the absolute approach (see 2.2.1) the output value has to be considered to be personal data for everyone. Neither will the output-value be encrypted nor will it be secretly shared, thus differing from the data processing carried out via MPC to produce the output or the secret sharing of the data if the Association is involved. The system's purpose is to produce identifiable data as an output, so safeguards have to be taken in order to ensure compliance with the data protection law. Since the output-value is new data, its production ought to be considered as collection of personal data for the entity retrieving it. For the two entities in charge of producing the output (the accountancy firm and the bank jointly) the production and the provision of the output to the bank has to be considered a transfer of personal data (note that the bank is both one of the joint controllers and the entity receiving the data in the basic scenario). Both the accountancy firm and the bank are jointly responsible for this data processing to be compliant with data protection law; hence, they need, once again, the consent of the farmer or an explicit legal permission, as they do for storing the data on the servers (if it is stored in plaintext). The same result would be reached according to the relative approach.
However, consent may be more easily obtained from farmers asking for a loan without any cloud-specific problems; unlike in the Sharemind case, there is no greater number of affected individuals whose consent would be needed but rather one single farm. Even without consent, the collection of the value might be based on Article 7 lit (b), (f) DPD. Nevertheless, as accountancy firms do have weaker interests in regards to data protection compared to farms (see above), obtaining consent would be the favorable legal option.
A higher risk refers to the potential abuse of the system. Especially in the scenario involving several banks and the Bankers Association a bank may use the CVR number of a farm that is not asking for a loan and who is not one of the banks clients to produce a financial performance evaluation of the farm. The collection of data affecting a farm that is not a client of the 0 bank would be illegal, as there is no legal ground for the collection without any contractual performance etc. The eventual abuse raises a legal risk for the (joint) controller of the system that should be addressed by including technical and organizational safeguards (such as contractual cases for indemnization) before the system is put in use.

Compliance with the GDPR As explained under 3.2.1.2.3.2, the GDPR is not applicable to the processing of secretly shared data. However, the case of Danish banks and farmers implies the storage of data in plaintext, which thus may be used to produce personal data as an output to its user. Therefore, an assessment of the GDPRs impact on this financial-performance-analysis system is needed.
If Amazon EC2 Instances are used to host the two servers, a data transfer to the USA (if the physical machines those Instances are hosted on are seated there) can be legal according to the principles of data transfers to third countries, but with the legal uncertainties described in 2.4.3. If those agreements do not provide legal grounds for the data transfer, Amazon can apply for a certification – the proposed European privacy seal. This seal can enable the banks to provide evidence that they have ensured that the entity (Amazon) in a third country they are transferring data to provides for an adequate level of data protection. [49] Amazon will do no processing in the described scenario (see 3.2.2.2.3.1) and, therefore, no 'order processing' as regulated in Article 22 and the following GDPR will take place.
The situation changes if the Danish Bankers Association is running one of the two servers since the Association will process data via MPC and will secretly share data between the two servers together with the accountancy frim. According to the GDPR, the Association will process the data on behalf of the banks (the controllers). Therefore, as described under 3.2.1.2.3.2 the requirements of Article

[49] Note that this is only the second step, the transfer of data itself has to be lawful, too. See 2.3.4.

22 GDPR have to be respected. Again, a certification of the processor (the Association) in form of the data protection seal is recommended.
In any case, there will be joint controllers, either the accountancy firm and one bank or the accountancy firm and all participating banks, each using the Association as a processor to host their server. Therefore, a contractual framework has to be entered by each partner regulating responsibilities and duties of the involved parties (see 2.3.2.2). [50]
The system will provide an output that has to be considered personal data under the assumption that the affected farms are data subjects. This personal data has not existed before, but it will be newly created data. The financial-performance analysis, therefore, will be a collection of data, according to Article 4 lit (3) GDPR. Hence, according to Article 14 GDPR, the affected farm has to be informed, before the system is put in place. To fulfil Article 14 GDPR's requirements (especially Article 14 Par. 1 lit (f) of the Commission's and of the LIBE-proposal; Article 14 Par. 1a. (c) of the proposal of the Council), it is once again recommended not to use Amazon EC2 Instances to host the two servers.

[50] As described above: the information of the supervisory authority in case of a data breach (Art. 31 GDPR), if needed the appointment of a data protection officer (2.5.6), also the making available to the data subjects (the farms), in case they want to exercise their rights following Art. 17 GDPR (see 2.5.5).

List of Abbreviations

Abbreviation	German spelling	English spelling
AG	Amtsgericht	District Court
BB	Betriebs Berater	Operation advisor (journal)
BCR	-	Binding Corporate Rules
BDSG	Bundesdatenschutzgesetz	German Federal data protection act
BeckRS	Beck-Rechtsprechung	Beck-jurisdiction
CFR	-	Charter of Fundamental Rights of the European Union
CR	Computer und Recht	Computers and Law (journal)
DPA	-	Data Protection Authority
DPD	-	Data Protection Directive
DuD	Datenschutz und Datensicherheit	Data protection and data security (journal)
ECJ	-	European Court of Justice
EuZW	Europäische Zeitschrift für Wirtschaftsrecht	European journal of Business Law (journal)
GDPR	-	Proposal for a General Data Protection Regulation
GRUR	Gewerblicher Rechtschutz und Urheberrecht	Intellectual property and copyright (journal)
jurisPR-ITR	Juris Praxis Report - IT-Recht	Juris practice report - IT-law (online journal)
JZ	JuristenZeitung	Lawyers' Journal (journal)
K&R	Kommunikation und Recht	Communication and Law
KG	Kammergericht	See OLG
LG	Landgericht	Regional court
LMuR	Lebensmittel und Recht	Foodstuffs and law (journal)
LP	-	Linear Programming
MMR	MultiMedia und Recht	MultiMedia and law (journal)
NIST	-	National Institute of Standards and Technology
NJW	Neue Juristische Wochenschrift	New weekly report on legal issues (journal)
OLG	Oberlandesgericht	Higher regional court (or circuit court)
OVG	Oberverwaltungsgericht	Higher administrative Court (circuit court)
RDV	Recht der Datenverarbeitung	Law of data processing (journal)
SCA	-	Stored Communications Act
SMC	-	Secure Multiparty Computation
TMG	Telemediengesetz	Telemedia Act
TTP	-	Trusted Third Party
WP	-	Working Party
VG	Verwaltungsgericht	Administrative Court
ZD	Zeitschrift für Datenschutz	Journal of data protection
ZUM	Zeitschrift für Urheber- und Medienrecht	Journal of Copyright and Media Law

Bibliography

[1] Alich, Stefan; Nolte, Georg: Zur datenschutzrechtlichen Verantwortlichkeit (außereuropäischer) Hostprovider für Drittinhalte. In CR, 741 ff, 2011.

[2] Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Lnder sowie der Arbeitsgruppe Internationaler Datenverkehr des Dsseldorfer Kreises: Orientierungshilfe Cloud Computing, Version 2.0. Available at: https://www.datenschutz-bayern.de/technik/orient/oh_cloud.pdf.

[3] Art. 29-Working Party: Opinion 04/2012 on Cookie Consent Exemption, WP 194, 07/06/2012. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp194_en.pdf

[4] Art. 29-Working Party: Opinion 05/2012 on Cloud Computing, WP 196, 01/07/2012. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf

[5] Art. 29-Working Party: Opinion 15/2011 on the definition of consent, WP 187, 13/07/2011. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf

[6] Art. 29-Working Party: Opinion 03/2009 on the Draft Commission Decision on standard contractual clauses for the transfer of personal data to processors established in third countries, under Directive 95/46/EC, WP 161, 05/03/2009. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2009/wp161_en.pdf

[7] Art. 29-Working Party: Opinion 04/2007 on the concept of personal data, WP 136, 20/06/2007. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

[8] Art. 29-Working Part: Opinion 08/2010 on applicable law, WP 179, 16/12/2010. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp179_en.pdf

[9] Art. 29-Working Party: Opinion 01/2010 on the concepts of “controller” and “processor”, WP 169, 16/02/2010. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf

[10] Art. 29-Working Party: Working Document: Transfers of personal data to third countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers, WP 74, 03/06/2003. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2003/wp74_en.pdf

[11] Art. 29-Working Party: Working Document: Setting up a framework for the structure of Bindin Corporate Rules, WP 154, 24/06/2008. Available at http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp154_en.pdf.

[12] Bär, Wolfgang: Anmerkung zu: BGH: Speicherung von IP-Adressen durch die Bundesrepublik. In MMR, 134 ff., 2015.

[13] Berg, Kay Uwe: EU-Datenschutzgrundverordnung - Das Aus für Auskunfteien und Inkassounternehmen?. In PinG, 69 ff, 2013.

[14] Bergauer, Christian: Indirekt personenbezogene Daten - datenschutzrechtliche Kuriosa. In Jahrbuch Datenschutzrecht, 55 ff, 2011.

[15] Bergemann, Benjamin: EU-Datenschutzverordnung darf nicht Merkles NAS-Feigenblatt werden - Netzpolitik.org. 17/08/2013. Available at: https://netzpolitik.org/2013/eu-datenschutzverordnung-darf-nicht-merkels-nsa-feigenblatt-werden/

[16] Bergt, Matthias: Anmerkung zur Entscheidung des BGH (Beschluss vom 28.10.2014 VI ZR 135/13, ZD 2015, 80) zur Speicherung von IP-Adressen durch die Bundesrepublik. In ZD, 83 ff., 2015.

[17] Bergt, Matthias: Die Bestimmbarkeit als Grundproblem des Datenschutzrechts. In ZD, 365 ff., 2015.

[18] Bergt, Matthias: IP-Adressen: EU-Kommission gibt BGH Nachhilfe in Sachen Grundrechte. CRonline-Blog of 13/09/2015, available at: http://www.cr-online.de/blog/2015/09/13/ip-adressen-eu-kommission-gibt-bgh-nachhilfe-in-sachen-grundrechte/

[19] Bitkom: Leitfaden Cloud Computing, 2009. Available at: http://www.bitkom.org/files/documents/BITKOM-Leitfaden-CloudComputing_Web.pdf

[20] Bogdanov, Dan: Sharemind: programmable secure computations with practical applications, PhD thesis, University of Tartu, 2013. Available at: http://dspace.utlib.ee/dspace/bitstream/handle/10062/29041/bogdanov_dan_2.pdf?sequence=5

[21] Bogdanov, Dan; Kamm, Liiana; Laur, Sven; Pruulmann-Vengerfedt, Pille: Secure multi-party data analysis: end user validation and practical experiments, 2013. Available at: http://eprint.iacr.org/2013/826.pdf

[22] Brennscheid, Kristin: Cloud Computing und Datenschutz, Diss. Bochum, Baden-Baden, 2013.

[23] Brink, Stefan; Eckhardt, Jens: Wann ist ein Datum ein personenbezogenes Datum? Anwendungsbereich des Datenschutzrechts. In ZD, 205 ff., 2015.

[24] v. d. Bussche, Axel; Voigt, Paul: Konzerndatenschutz Rechtshandbuch. Munich, 2014.

[25] Charnes, A.; Cooper, W.W; Rhodes, E: Measuring the efficiency of decision making units. In European Journal of Operational Research, 429-444, 1978.

[26] Charnes, A.; Cooper, W.W; Rhodes, E.: Short communication: measuring the efficiency of decision making units. In European Journal of Operational Research, 339, 1978.

[27] Cybernetica: sharemind : Your secure service plattform for data collection and analysis. Available at: https://sharemindSharemind.cyber.ee/files/images/SharemindSharemind%20secure%20service%20platform%202012.pdf

[28] Däubler, Wolfgang; Klebe, Thomas; Wedde, Peter; Weichert, Thilo (ed.), Bundesdatenschutzgesetz Kompaktkommentar. 4th Edition, Frankfurt/Main, 2014.

[29] Dammann, Ulrich; Simitis, Spiros: EG-Datenschutzrichtlinie - Kommentar. 1st Edition, Baden-Baden, 1997.

[30] Decker, Florian: Die neue europäische Datenschutzgrundverordnung - welche änderungen sind für deutsche Unternehmen zu erwarten?, 2013. Available at: http://blog-it-recht.de/2013/12/02/die-neue-europaeische-datenschutzgrundverordnung-welche-aenderungen-sind-fuer-deutsche-unternehmen-zu-erwarten/

[31] Drews, Stefan; Moneal, Manfred: Grenzenlose Auftragsdatenverarbeitung. In PinG, 143 ff., 2014.

[32] Eckhardt, Jens: Kommentar zu: LG Berlin, Urteil vom 06.09.2007 - 23 S 3/07. In K&R, 601 ff., 2007.

[33] Eckhardt, Jens: IP-Adresse als personenbezogenes Datum - neues Öl ins Feuer. In CR, 339 ff., 2011.

[34] Eckhardt, Jens; Kramer, Rudi; Mester, Brita Alexandra: Auswirkungen der geplanten EU-DS-GVO auf den deutschen Datenschutz. In DUD, 623 ff., 2013.

[35] Eckhardt, Jens: Cloud Computing Orientierungshilfe 2.0 des Dsseldorfer Kreises. In DuD, 176 ff., 2015.

[36] Ehmann, Eugen; Helfrich, Marcus: EG-Datenschutzrichtlinie - Kurzkommentar. 1st. Edition, Cologne, 1999.

[37] Fazlioglu, Muge: Forget me not: the clash of the right to be forgotten and freedom of expression on the Internet. In International Data Privacy Law, p. 149 ff., 2013. Available at: http://idpl.oxfordjournals.org/content/3/3/149.full.pdf+html

[38] Frauenhofer Institut für Offene Kommunikationssysteme: ISPRAT-Studie, Cloud-Computing für die öffentliche Verwaltung, 11/2010. Available at: http://www.cloud.fraunhofer.de/content/dam/allianzcloud/de/documents/ISPRAT_cloud_studievorabversion20101129tcm421-76759.pdf

[39] Funke, Michael; Wittmann, Jörn: Cloud Computing - ein klassischer Fall der Auftragsdatenverarbeitung?. In ZD, 221 ff., 2013.

[40] Gerlach, Carsten: Personenbezug von IP-Adressen. In CR, 478 ff., 2013.

[41] German Federal Office for Information Security Technology: BSI-Standard 100-1

[42] German Federal Office for Information Security Technology: BSI-Standard 100-1 Information Security Management Systems (ISMS). Available at: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/BSIStandards/standard_100-1_e_pdf.pdf?__blob=publicationFile

[43] German Federal Office for Information Security Technology: Safety Recommendation for Cloud Computing Providers. Available at: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Mindestanforderungen/Eckpunktepapier-Sicherheitsempfehlungen-CloudComputing-Anbieter.pdf?__blob=publicationFile

[44] Giedke, Anna: Cloud Computing: Eine wirtschaftsrechtliche Analyse mit besonderer Berücksichtigung des Urheberrechts. Diss., Munich, 2013.

[45] Gola, Peter; Schomerus, Rudolf (ed.): BDSG Bundesdatenschutzgesetz - Kommentar. 11th Edition, Munich, 2012.

[46] Härting, Niko: Datenschutzreform in Europa: Einigung im EU-Parlament : Kritische Anmerkungen. In CR, 715 ff., 2013.

[47] Härting, Niko: Internetrecht. 5th Edition, Cologne, 2014.

[48] Härting, Niko: Starke Behörden, schwaches Recht - der neue EU-Datenschutzentwurf. In BB, 459 ff., 2012.

[49] Härting, Niko: Schutz von IP-Adressen. In ITRB, 35 ff., 2009.

[50] Heckmann, Dirk (ed.): Juris PraxisKommentar Internetrecht. 4th Edition, Saarbrcken, 2014.

[51] Heckmann, Jörn; Nordmeyer, Arne: Pars pro toto: Verletzung des Urheberrechtsgesetzes durch das öffentliche Zugänglichmachen von Dateifragmenten ("Chunks") in Peer-to-Peer-Tauschbörsen. In CR, 41-45, 2014.

[52] Heidrich, Joerg; Wegener, Christoph: Sichere Datenwolken Cloud Computing und Datenschutz. In MMR, 803, 2010.

[53] Heinemeyer, Dennis: Verfahrensstand-Anzeiger. In Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung). Available at: http://www.computerundrecht.de/26378.htm

[54] Hennrich, Thorsten: Compliance in Clouds. In CR, 546 ff, 2011.

[55] Hilber, Marc: Handbuch Cloud Computing. Cologne, 2014.

[56] Hon, W Kuan; Millard, Christopher; Walden, Ian: The Problem of 'Personal Data' in Cloud Computing - What Information is Regulated?, The Cloud of Unknowing, Part. 1. 10/03/2011. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577

[57] Hon, W Kuan; Hörnle, Julia; Millard, Christopher: Data Protection Jurisdiction and Cloud Computing - When are cloud Users and Providers Subject to EU Data Protection Law?, The Cloud of Unknowing, Part 3. 09/02/2012.

[58] Hon, W Kuan; Millard, Christopher; Walden, Ian: Who is Responsible of 'Personal Data" in Cloud Computing?, The Cloud of Unknowing, Part 2. 21/03/2011. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1794130

[59] Hon, W Kuan; Millard, Christopher: Data Export in Cloud Computing; How Can Personal Data Be Transferred Outside the EEA?, The Cloud of Unknowing, Part 4. 04/04/2012. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2034286

[60] Hornung, Gerrit; Sädtler, Stephan: Europas Wolken - Die Auswirkungen des Entwurfs für eine Datenschutz-Grundverordnung auf das Cloud Computing. In CR, 638 ff., 2012.

[61] Jandt, Silke; Roßnagel, Alexander: Datenschutz in Social Networks - Kollektive Verantwortlichkeit für die Datenverarbeitung. In ZD, 160 ff., 2011.

[62] Jaspers, Andreas: Die EU-Datenschutz-Grundverordnung : Auswirkungen der EU-Datenschutz-Grundverordnung auf die Datenschutzorganisation des Unternehmens. In DuD, 571 ff., 2012.

[63] Jotzo, Florian: Gilt deutsches Datenschutzrecht auch für Google, Facebook & Co. bei grenzüberschreitendem Datenverkehr?. In MMR, 232 ff., 2009.

[64] Jotzo, Florian: Der Schutz personenbezogener Daten in der Cloud. Diss., Kiel 2013.

[65] Kamm, Liina; Willemson, Jan: Secure Floating-Point Arithmetic and Private Satellite Collision Analysis. Available at: http://eprint.iacr.org/2013/850.pdf

[66] Karg, Moritz: Anonymität, Pseudonyme und Personenbezug revisited? In DuD, 520 ff., 2015.

[67] Kilian, Wolfgang; Heussen, Benno (ed.): Computerrechts-Handbuch: Computertechnologie in der Rechts- und Wirtschaftspraxis. Supplement 32, Munich, 2013.

[68] Klar, Manuel: Räumliche Anwendbarkeit des (europäischen) Datenschutzrechts. In ZD, 109 ff., 2013.

[69] Klinger, Markus: Vorschlag zur EU-Datenschutz-Grundverordnung i.d.F. des EU-Parlaments - Auswirkungen auf datenverarbeitende Unternehmen im Überblick. In jurisPR-ITR 6/2014 Anm. 2.

[70] Kokott, Juliane; Sobotta, Christoph: The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. In International Data Privacy Law, 222 ff., 2013. Available at: http://idpl.oxfordjournals.org/content/3/4/222.full.pdf+html

[71] Kos, Clemens; Englisch, Bastian: Auftragsdatenverarbeitung? - Gegenüberstellung der aktuellen Rechtslage und der DS-GVO in der Fassung des LIBE-Entwurfs. In ZD, 276 ff., 2014.

[72] Krempl, Stefan: EU-Datenschützer fordert Einbau von Datenschutz in die Technik. Available at: http://www.heise.de/newsticker/meldung/EU-Datenschuetzer-fordert-Einbau-von-Datenschutz-in-die-Technik-960735.html

[73] Kroschwald, Steffen: Verschlüsseltes Cloud Computing : Auswirkung der Kryptografie auf den Personenbezug in der Cloud. In ZD, 75 ff., 2014.

[74] Krügel, Tina: LIBE-Kompromissvorschlag zur DS-GVO. In ZD-Aktuell, 03870, 2014.

[75] Kühling, Jürgen: Auf dem Weg zum vollharmonisierten Datenschutz?!. In EuZW, 281 ff., 2012.

[76] Kühling, Jürgen; Klar, Manuel: Unsicherheitsfaktor Datenschutzrecht - Das Beispiel des Personenbezugs und der Anonymität. In NJW, 3611 ff., 2013.

[77] Kuner, Christopher: European Data Protection Law. 2nd Edition, New York, 2007.

[78] Lang, Markus: Reform des EU-Datenschutzrechts. In K&R, 145 ff., 2012.

[79] Leonard, Peter: Customer data analytics: privacy settings for 'Big Data' business. In International Data Privacy Law, Vol. 4, No. 1, 53 ff., 2014. Available at: http://idpl.oxfordjournals.org/content/4/1/53.full.pdf+html?sid=a11bd260-1434-4d5e-8e3b-c15b73574748

[80] Leucker, Franziska, Die zehn Mrchen der Datenschutzreform. In PinG 195 ff., 2015.

[81] Leupold, Andreas; Glossner, Silke: Mnchener Anwaltshandbuch IT-Recht, 3rd Edition, Munich, 2013.

[82] Leutheusser-Schnarrenberger, Sabine: Zur Reform des europäischen Datenschutzrechts. In MMR, 709 f., 2012.

[83] Maisch, Michael Marc: Nutzertracking im Internet. In ITRB, 13 ff., 2011.

[84] Marnau, Ninja; Schlehahn, Eva: Cloud Computing: Legal Analysis. In TClouds (D1.2.2). Available at: http://www.tclouds-project.eu/downloads/deliverables/TC-D1.2.2_Cloud_Computing-Legal_Analysis_M12.pdf

[85] Marnau, Ninja; Schlehahn, Eva: Cloud Computing und Safe Harbor. In DuD, 311 ff.; 2011.

[86] Marschall, Kevin: Datenpannen neue“ Meldepflicht nach der europischen DS-GVO? Rechtliche nderungen durch Art. 31 und Art. 32 DS-GVO. In DuD, 183 ff, 2015.

[87] Meyerdierks, Per: Sind IP-Adressen personenbezogene Daten?. In MMR, 8 ff., 2009.

[88] Millard, Christopher: Cloud Computing, Oxford, 2013.

[89] Nägele, Thomas; Jacobs, Sven: Rechtsfragen des Cloud Computing. In ZUM, 281 ff., 2010.

[90] NASA: Satellite Collision Leaves Significant Debris Clouds. In Orbital Debris Quarterly News, Volume 13, Issue 2, 1-2, 2009. Available at: http://orbitaldebris.jsc.nasa.gov/newsletter/pdfs/ODQNv13i2.pdf

[91] Natz, Alexander; Wolters, Chris: Geplante EU-Datenschutz-Verordnung: Auswirkungen für die Datenverarbeitung in Unternehmen. In LMuR, 3 ff., 2014.

[92] Nebel, Maxi, Richter, Philipp. Datenschutz bei Internetdiensten nach der DS-GVO - Vergleich der deutschen Rechtslage mit dem Kommissionsentwurf. In ZD, 407 ff., 2012.

[93] Niemann, Fabian; Paul, Jörg-Alexander: Bewölkt oder wolkenlos - rechtliche Herausforderungen des Cloud Computings. In K&R, 444 ff., 2009.

[94] Niemann, Fabian; Ammann, Jörg-Alexander: Praxishandbuch Rechtsfragen des Cloud Computing. Berlin, Bosten, 2014.

[95] Nink, Judith; Pohle, Jan: Die Bestimmbarkeit des Personenbezugs - Von der IP-Adresse zum Anwendungsbereich der Datenschutzgesetze. In MMR, 563 ff., 2015.

[96] Nord, Jantina; Manzel, Manzel: Datenschutzerklärungen- - misslungene Erlaubnisklauseln zur Datennutzung : -Happy-Digits- und die bedenklichen Folgen im E-Commerce. In NJW, 3756, 2010.

[97] Pahlen-Brandt, Ingrid: Datenschutz braucht scharfe Instrumente: Beitrag zur Diskussion um -personenbezogene Daten. In DuD, 34 ff., 2008.

[98] Peifer, Karl-Nikolaus: Verhaltensorientierte Nutzeransprache - Tod durch Datenschutz oder Moderation durch das Recht?. In K&R, 543 ff., 2011.

[99] Petri, Thomas, Auftragsdatenverarbeitung heute und morgen. Reformberlegungen zur Neuordnung des Europischen Datenschutzrechts. In ZD, 305 ff., 2015.

[100] Piltz, Carlo: Datenschutzreform: aktueller Stand der Verhandlungen im Rat. 20/01/2014. Available at: http://www.delegedata.de/2014/01/datenschutzreform-aktueller-stand-der-verhandlungen-im-rat/

[101] Plath, Kai-Uwe: Datenherausgabepflicht für Cloud-Anbieter nach US-Recht vs. EU-Datenschutzrecht. 13/05/2014. Available at: http://www.cr-online.de/blog/2014/05/13/datenherausgabepflicht-fuer-cloud-anbieter-nach-us-recht-vs-eu-datenschutzrecht/

[102] Pohle, Jan; Ammann, Thorsten: Software as a Service - auch rechtlich eine Evolution?. In K&R, 625 ff., 2009.

[103] Pollirer, Hans-Jürgen; Weiss, Ernst M.; Knyrim, Rainer: Datenschutzgesetz 2000 (DSG 2000) samt ausführlichen Erläuterungen. 2nd Edition, Vienna, 2014.

[104] Popa, Raluca Ada: Research Statement. Available at: http://www.mit.edu/~ralucap/researchstatement.pdf

[105] Popa, Raluca Ada; Zeldovich, Nickolai; Balakrishnan, Hari: CryptDB: A Practical Encrypted Relational DBMS. In Technical Report MIT-CSAIL-TR-2011-005, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, 01/2011. http://people.csail.mit.edu/nickolai/papers/raluca-cryptdb-tr.pdf

[106] Popa, Raluca Ada; Redfield, Cahterine M. S.; Zeldovich, Nickolai; Balakrishnan, Hari: CryptDB: Protecting Confidentiality with Encrypted Query Processing. In Proceedings of the 23rd ACM Symposium on Operating Systems Principles (SOSP), Cascais, Portugal, 10/2011. Available at: http://people.csail.mit.edu/nickolai/papers/raluca-cryptdb.pdf

[107] Rammos, Thanos: Datenschutzrechtliche Aspekte verschiedener Arten "verhaltensbezogener" Onlinewerbung. In K&R, 692 ff., 2011.

[108] Rath, Michael; Rothe, Britta: Cloud Computing: Ein datenschutzrechtliches Update. In K&R, 623 ff., 2013.

[109] Roßnagel, Alexander (ed.): Beck'scher Kommentar zum Recht der Telemediendienste. Munich, 2013.

[110] Roßnagel, Alexander (ed.): Handbuch Datenschutzrecht: Die neuen Grundlagen für Wirtschaft und Verwaltung. Munich, 2003.

[111] Roßnagel, Alexander; Scholz, Philip: Datenschutz durch Anonymität und Pseudonymität : Rechtsfolgen der Verwendung anonymer und pseudonymer Daten. In MMR, 721 ff., 2000.

[112] Roßnagel, Alexander; Richter,Philipp; Nebel, Maxi: Besserer Internetdatenschutz für Europa - Vorschläge zur Spezifizierung der DS-GVO. In ZD, 103 ff., 2013.

[113] Sartor, Giovanni: Providers' liabilities in the new EU Data Protection : Regulation: A threat to Internet freedoms?. In International Data Privacy Law, 3 ff., 2013. Available at: http://idpl.oxfordjournals.org/content/3/1/3.full.pdf+html

[114] Schaar, Peter: Privacy By Design. Available at: http://www.bfdi.bund.de/SharedDocs/Publikationen/EN/0610EUPrivacyByDesign.pdf?__blob=publicationFile

[115] Schneider, Jochen: Handbuch des EDV-Rechts. 4th Edition, Cologne, 2009.

[116] Schulz, Sebastian: Privacy by Design Datenschutz durch Technikgestaltung im nationalen und europäischen Kontext. In CR 204 ff., 2012.

[117] Simitis, Spiros (ed.): Bundesdatenschutzgesetz Kommentar. 7th Edition, Baden-Baden, 2011.

[118] Spies, Axel: Cloud Computing: Keine personenbezogenen Daten bei Verschlüsselung. In MMR-Aktuell, 313727, 2011.

[119] Spindler, Gerald: Persönlichkeitsschutz im Internet - Anforderungen und Grenzen einer Regulierung. In Verhandlungen des 69. Deutschen Juristentages, Band I Gutachten, Munich, 2012.

[120] Spindler, Gerald: Persönlichkeitsrecht und Datenschutz im Internet - Anforderungen und Grenzen einer Regulierung. In NJW-Beilage, 98 ff., 2012.

[121] Spindler, Gerald: Datenschutz- und Persönlichkeitsrechte im Internet - der Rahmen für Forschungsaufgaben und Reformbedarf. In GRUR, 996 ff., 2013.

[122] Spindler, Gerald: Datenschutz- und Persönlichkeitsrechte im Internet - Der Rahmen für Forschungsaufgaben und Reformbedarf. In GRUR-Beilage, 101 ff., 2014.

[123] Spindler, Gerald: Durchbruch für ein Recht auf Vergessen(werden)? - die Entscheidung des EuGH in Sachen Google Spain und ihre Auswirkungen auf das Datenschutz- und Zivilrecht. In JZ 2014.

[124] Spindler, Gerald; Schuster, Fabian (eds.): Recht der elektronischen Medien. 3rd Edition, Munich 2015.

[125] Stadler, Thomas: Datenschutz: IP-Adressen als personenbezogene Daten. 27/06/2011. Available at: http://www.internet-law.de/2011/06/datenschutz-ip-adressen-als-personenbezogene-daten.html

[126] Stadler, Thomas: Der Datenschutz bietet keine Handhabe gegen die überwachungspraxis der Geheimdienste. 05/11/2013. Available at: http://www.internet-law.de/2013/11/der-datenschutz-bietet-keine-handhabe-gegen-die-ueberwachungspraxis-der-geheimdienste.html

[127] Sydow, Gernot; Kring, Markus: Die Datenschutzgrundverordnung zwischen Technikneutralität und Technikbezug - Konkurrierende Leitbilder für den europäischen Rechtsrahmen. In ZD, 271 ff., 2014.

[128] Taeger, Jürgen; Gabel, Detlev (ed.): Kommentar zum BDSG und zu den Datenschutzvorschriften des TKG und TMG. 2nd Edition, Frankfurt/Main, 2013.

[129] Tene, Omer: Privacy: The new generations. In International Data Privacy Law, 15 ff., 2011. Available at: http://idpl.oxfordjournals.org/content/1/1/15.full.pdf+html?sid=a11bd260-1434-4d5e-8e3b-c15b73574748

[130] Voigt, Paul: Datenschutz bei Google. In MMR, 377 ff., 2009.

[131] Weichert, Thilo: Cloud Computing und Datenschutz. In DuD, 679 ff., 2010.

[132] Wieczorek, Mirko: Der räumliche Anwendungsbereich der EU-Datenschutz-Grundverordnung. In DuD, 644 ff., 2013.

[133] Wisskirchen, Gerlind: Grenzberschreitender Transfer von Arbeitnehmerdaten. In CR, 862 ff., 2004

[134] Wolff, Amadeus; Brink, Stefan: Datenschutz in Bund und Ländern - Kommentar. Munich; 2013.

Part II

Part II - Risk Assessment

Chapter 4

The methodology

4.1 Introduction

Risk Management is defined by standards as ”the systematic application of management policies, procedures and practices to the tasks of establishing the context, identifying, analyzing, assessing, treating, monitoring and communicating risks” (AS/NZS 4360:2004), while risk is defined as ”the chance of something happening that will have a negative impact on objectives”. Managing risks concerning the execution of a business process is itself a process, the *risk management process*, where alternative techniques for dealing with threats are compared, and criteria for the selection of *risk alleviation strategies* are defined. This iterative process can, with each cycle, provide management with a greater insight into risks and their impact.
The basic steps of a classic risk management procedure include:

- Identify the Threats, i.e. the negative events that could jeopardise achieving the business process' objectives.
- Estimate the Likelihood. For each threat, provide a quantitative or qualitative estimate of its likelihood.
- Estimate the Impact. For each threat, quantify the damage that could follow missing the corresponding objectives.
- Identify the Controls. Identify all the devices (Controls) that can be put in place taimed at reducing the likelihood of your risks from happening in the first place and, if they do happen, to reduce their impact (Consequence).
- Make a Decision: decide which controls to deploy, based on their cost and on the risk alleviation they can provide.

In this deliverable (D31.2) we revise and extend the risk assessment methodology we introduced in D31.1. The ultimate goal of the *quantitative* methodology we propose is to provide an easy way to compute the risk of competing risk reduction controls, providing quantitative cost-benefit calculations, assessments of risk tolerance, and quantification of preferences. We focus on process-related *leakage threats*, i.e. the disclosure of one or more information items to be exchanged in a multi-party protocol to participating parties who are not the originally intended recipients.
The basic idea underlying our approach is using the *micro-economics* underlying the process to compute a quantitative estimate of the process' data leakage probability and impact. We start by using the

micro-economics process to compute the *perceived unfairness* of the business process resource allocation. Then, we use perceived unfairness to derive a continuous probability distribution associated to a formal random variable. Starting from the probability values of this distribution, we compute discrete probabilities of disclosure attacks associated to each subset of actors, on the basis of the information items they could access.

4.2 A Methodology for Risk aware deployment of Secure Computation

Our approach is to insert in the risk management process an iterative sub-process, where the risk assessor can estimate during the risk profiles associated to subsets of participants (henceforth called *actors*) who could share among themselves or disclose to external parties the information they get to know by taking part to the process. This sub-process consists of a simulation where, step by step, the information exchanged among the actors is traced, and the collusion probability of each subset of actors is quantified, to compute the overall likelihood of information disclosure. If the risk assessor decides that the risk level is not acceptable, she can explore alternative solutions based on the deployment of Controls like secure computation techniques, run a different simulation and obtain a new risk analysis. At the end, cost-benefit analysis will allow the risk assessor to decide if the cost of additional security controls is worth the reduction of the risk she has obtained in the new business configuration.
Figure 4.1 shows the overall process.

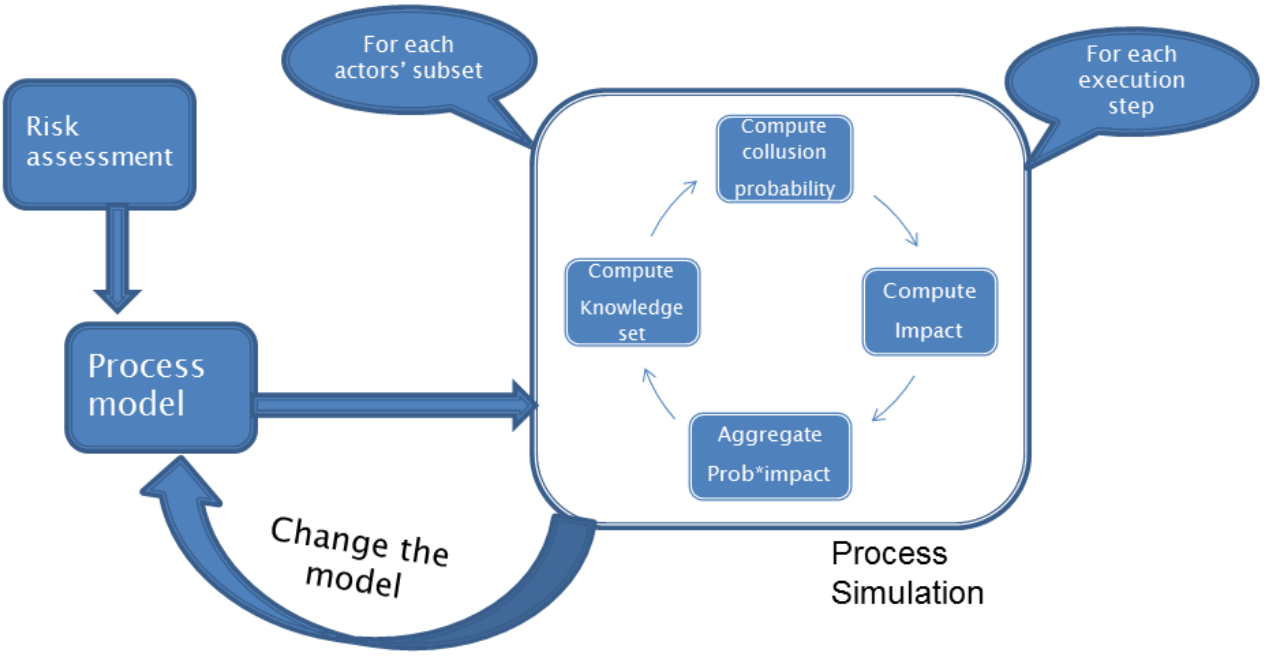

Figure 4.1: Our iterative process.

Let us now examine in detail our risk assessment methodology.

- We represent the business process P whose risk we want to assess via our *formalization of the business process model*, whose syntax is detailed in section 4.4. Two types of actions are represented: (i) *message exchanges* and (ii) *local computations*. It is important to remark that

our process model syntax expresses all possible execution paths *independently*, i.e. as separate models, providing a *process streamlining*, which includes *loop unrolling* and *re-encoding of conditions as parallel paths* [86].

- We compute (see again section 4.4), the probability of malicious behavior on the part of each actor on the basis of a number of features, such as the unfairness of the redistribution of payoffs in the business process, (e.g. a benefit allocation structure that responds to organization efficiency more than to fairness), the actor's greed, plus other context-related factors. At the same time, expert knowledge is used to estimate the impact of information disclosure on the business process, quantifying the Value of Information (VoI) associated to each information item.

Once the process model has been designed, and the probabilities and impact values estimated, an iterative process starts by considering for each execution step of P and for each possible subset of actors:

- *identifying reconstructible knowledge*, i.e *knowledge set* K_S for each subset $S \in 2^A$. The knowledge set includes all the knowledge that members of A can achieve by putting together the information they hold.
- *identifying reconstructible knowledge*, i.e *knowledge set* K_S for each subset $S \in 2^A$. The knowledge set includes all the knowledge that members of A can achieve by putting together the information they hold.
- *estimating the collusion probability* for each subset S in 2^A. Once again it is important to remark that this estimate needs to be process-specific (as it will take into account the micro-economics and social relations underlying P) and take into account multiple causes of collusion, including dysfunctional behavior, intervention of regulatory authority and others (section 5.1).
- *estimating the disclosure impact* of K_S for each subset S in 2^A.
- *aggregating the products* between (i) the collusion probability of each subset S in 2^A and (ii) the disclosure impact of K_S at each step of the process P, obtaining the *total risk* related to the process.

4.3 Modeling Security Controls

Our methodology aims to provide the risk assessor with the capability to evaluate different alternative control deployment to secure the business process computation. For this purpose on one side it is possible to model the information exchange according to the selected security control adopted, and on the other side to automatically trace how the risk profiles associated to the information that actors can access change along the execution of the business process. The techniques involve also different communication facilities and services that affect also the overall cost of the provided solution. On the basis of a comparative analysis among the different configurations, the manager then could finally take a decision and trade-off money for security. Here we focus on some specific techniques used within PRACTICE to perform secure computation among a number of parties. See also D31.1 for a more complete landscape of the available security controls. They all belong to the family of *Secure Multiparty Computation* (SMC) techniques, which includes a wide range of different protocols, all sharing the feature that the involved multiple parties collaborate to securely compute some agreed function on their private inputs. There exist also special-purpose multi-party protocols supporting the evaluation of functions tailored to solve a specific problem such as *private set intersection* or *electronic voting*.

4.3.1 Modeling Secret Sharing-Based SMC

Secret sharing is a technique based on the splitting of secret data into *shares* that can then be distributed among the parties taking part to the execution of the protocol [55]. Since none of the individual share does not reveal any information on the secret data, the secret information is protected unless one party gets a sufficient amount of the shares enabling to *reconstruct* the original secret. Usually, in secret sharing-based SMC protocols, each party starts by sharing her secret input data in order to let the function, represented as a circuit, to be evaluated among all parties. Then to evaluate a gate of the functions circuit, the parties use the secret shared input to that gate to run a gate-specific obtaining a secret sharing of the result of the given gate. In this way, the parties can evaluate the entire circuit gate by gate. After evaluating the last gate of the circuit, the parties will hold secret shares of all intermediate values on the wires of the circuit. To reveal the result of the function, the parties exchange their shares of the output wires in order to reconstruct the value of the output.

4.3.2 Homomorphic Threshold Encryption-Based SMC

Homomorphic encryption schemes rely on a particular form of encryption enabling the computation on values while they are encrypted, i.e. without the real knowledge of those values. In this way, a party can perform operations on the encryptions to obtain a new encryption of the result of the given operation on the encrypted values. The parties proceed to evaluate the circuit gate by gate, using the homomorphic property of the encryption scheme in such a way that they obtain an encryption of the output of each gate. Once all gates have been evaluated, the parties can collaborate to decrypt the result. Introduced by Rivest et al. [85], homomorphic encryption has recently been reconsidered after Gentry et al. in [43] showed the possibility of constructing fully homomorphic encryption schemes, i.e. schemes supporting homomorphism for multiple operations.

4.3.3 Garbled Circuits

The technique of garbled circuits has been introduced by Yao's in [97]. A garbled circuit (GC) allows the evaluation of any function in a secure way involving two parties (in the semi-honest model). The function to be evaluated is provided in advance to both parties and is represented as a boolean circuit. A GC can be considered the encrypted representation of the original boolean circuit and allows the computation of the output value without revealing the original inputs nor the intermediate values carried on the wires of the circuit. One party is considered to be the *creator* C of the garbled circuit and the other the *evaluator* E: On a very general level, the scheme starts with C constructing a *garbled circuit* from the original circuit which represents the function they want to compute. C sends the circuit to E who evaluates it and obtains the result, which is shared with C.
In more detail, we model GC computation following [9]. A garbling algorithm GC can be modeled as a randomized algorithm that transforms a function $f : 0,1^n \to 0,1^m$ into a triple of functions $(F,e,d) \leftarrow Gb(f)$, where $f = d \cdot F \cdot e$. The encoding function e transforms the initial and private input $x \in 0,1^n$ into a garbled input $X = e(x)$ that can be evaluated by the garbled function F. When F is applied on the garbled input X it returns the garbled output $Y = F(X)$. Finally the decoding function d turns the garbled output Y into the final output $y = d(Y)$, which must coincide with $f(x)$.
To model the required privacy, we say that a party acquiring (F,X,d) shouldnt learn anything impermissible beyond that which is revealed by knowing just the final output y. To formalize that which it is permissible to reveal, a side-information function ϕ parametrizes the definition; an adversary should be able to ascertain from (F,X,d) nothing beyond $\phi(f)$ and y. By varying one can encompass the customary setting for SFE, for example requiring that the knowledge on the evaluation circuit is leaked and then let $\phi(f) = f$.

4.4 The Process Model

Let us now formalize our syntax for expressing business process models for risk assessment purposes. We start by representing the business process' set of actors as a set $A = \{A_1, \ldots, A_n\}$. Each actor A_j holds a (possibly empty) information item $INFO_j$ whose content is used to generate messages to be exchanged during the business process' execution. Also, we denote by $\{I_{j,k}\}$ the impact of the disclosure of $INFO_j$ to A_k (as assessed by A_j). In principle, this impact can be positive or negative, and can depend on a number of factors, including the content of $INFO_j$ or of other information items. In our view, security controls (when present) are an integral part of the business process definition. In order to be able to represent all security controls of PRACTICE, message exchange in our process model is a general *timestamped choreography* [6] consisting of:

- *Messages*, i.e. triples (A_i, A_j, m_{ts}), where m_{ts} is (a part of) an $INFO$ item and ts is an integer representing a discrete time value[1]
- *Local computations* $(A_i, f(), INFO_{i,ts})$ i.e. functions computed by actors on (portions of) locally held information at a given time.

Next, we enrich our representation of business process actors and make it suitable for representing cloud-based computations. To this end, our actor set A becomes (non-necessarily disjoint) a triple $\{IN, COMP, RES\}$ where IN denotes actors holding non-empty information items (a.k.a. input nodes), while $COMP$ and RES are auxiliary sets of actors (a.k.a. *compute* and *result* actors) whose information items are initially empty. Such actors respectively perform local computations ($COMP$) and publish results (RES).
The following constraints are in place for our cloud model:

- *Separation of duties*: Sender actors belong to IN and $COMP$ only.
- *Local information integrity*: Any actor can send part of an $INFO$ item it holds entirely, or relay parts it has previously received from other actors.

Figure 4.2 shows a sample visual representation of a cloud-based process, where a buyer sends messages to two sellers who respond with their offers:

The Knowledge Set For each subset $S \in 2^A$, we can now compute the risk of disclosure for information shared within S, at each time t. We proceed as follows: we consider all messages in the process incoming to actors belonging to S with timing $ts \leq t$. The (possibly empty) *common knowledge* of S, $K_S(t)$ is then composed of the $INFO$ items whose shares have been all received by members of S before time t, say $K_S(t) = \{INFO_{j_1}, \ldots, INFO_{j_h}\}$. The impact of the disclosure of this common knowledge on any actor $A_k \in A$ can be expressed in symbols as follows:

$$I_{S,k,t} = \sum_{p=1}^{h} I_{j_p,k} \tag{4.1}$$

and, in words, as the damage that members of S can do to A_k by getting to know all information items they can jointly reconstruct from the shares they hold at time t. Computing the risk posed by S to A_k also requires estimating the probability of members of S having colluded at time t (section 4.4). This risk can be written as follows:

$$R(A_k, E_S) = P_{S,t} I_{S,k,t} \tag{4.2}$$

[1] For the sake of simplicity, in our model we assume synchronous clocks and instant message delivery.

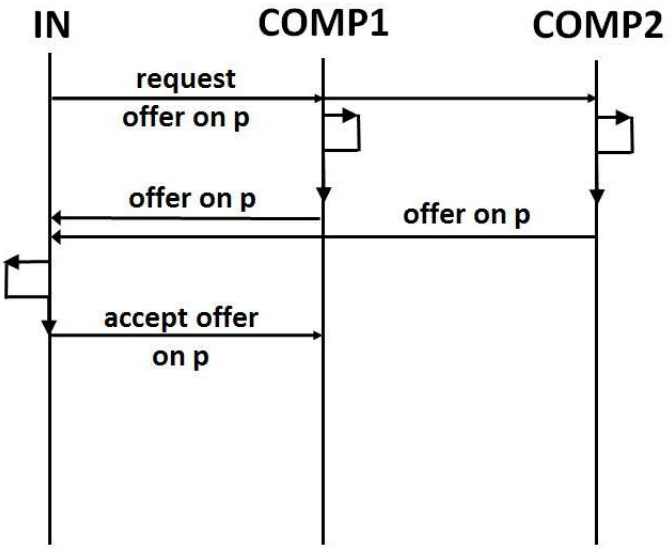

Figure 4.2: Visual representation of a sample cloud process model.

Assuming that collusions happen independently, we can also write the total risk for A_k taking part to the process, as follows:

$$R(A_k, 2^A, \infty) = \sum_{S \in 2^A} I_{S,k,\infty} P_S \tag{4.3}$$

4.4.1 Reasoning about Shared Knowledge

A large amount of work has been dedicated to understanding the process of reasoning about knowledge in a group and using this understanding for the analysis of complex systems [88, 38]. As regards the verification of cryptographic protocols modeled as sequences of encrypted messages exchanged to achieve some goals, *logic of knowledge* and *belief calculus* have been used extensively [14, 63]. Inference construction approaches try to use inference in some specialized logics to establish required beliefs at the protocol participants. The *logic of authentication* introduced in [14], and commonly referred to as "BAN", was one of the first successful attempts at representing and reasoning about the security properties of protocols, and has been extended and improved in successive works, as well as used as basis for the development of automatic verification tools [89, 78, 23]. In many cases, logics of knowledge are used together with a formal model of a multi-agent system and a precise definition of a message trace [37, 69].
Here, we adopt a semi-formal approach, where we explicitly represent the knowledge sets, but do not equip our notation with the burden of formal semantics. Our goal is to trace the evolution of the knowledge set associated to each subset of actors, and compute the overall risk by considering both the probability of a malicious behaviour on the part of the actors within the set and the value of the information they hold.

Securing processes First of all, our methodology helps the risk assessor to construct a representation of an unsecured process, and then to add controls to secure it. We use a simple syntax for expressing the operation of security controls over process messages, starting from a plain-text exchange.

- Plaintext exchange $A \rightarrow A$
- Data Encryption $A \rightarrow E_{KA}[A]$
- Homomorphic Encryption $A \rightarrow HE_{KA}[A]$
- Secret sharing $A \rightarrow Sh_{1,n}[A] \oplus Sh_{2,n}[A] \oplus \cdots \oplus Sh_{n,n}[A]]$

4.4.2 The Role of the Cloud Provider

Under EU and US data protection laws, organizations remain responsible for the personal data of their customers and employees and must guarantee its security even when a third-party like a cloud provider processes the data on their behalf. For this reason, data access issues have become a key consideration in any cloud provisioning arrangement. Data owners increasingly require outsourcing contracts to specify all details of cloud providers' security procedures, including prompt notification of any security breach and, above all, which techniques will be available to support cloud-based encryption/decryption services. Still, many cloud users have trouble in understanding and comparing available solutions.
When analysing the configurations of actors taking part in a business process that is outsourced to the cloud, it is also important to examine the relations among the actors and a special participant, the *cloud provider*, since the messages exchanged among the actors and the information items in the actors local memories could be potentially exposed to the cloud provider and its employees. Of course, from the process model point of view the cloud provider may or may not explicitly take part in the process, i.e. be the source/destination of information exchanges or simply provide the environment where such exchanges take place. In general, we want to model this situation giving to the cloud provider the role of a potential *inside attacker*, who can access and/or manipulate the information items that are exchanged by the actors hosted on his platform. To this end, we enriched the D31.1 original notation for denoting process actors, by introducing the notion of actor attached to (running/using) services of a cloud provider. Each actor is labeled taking into consideration the hosting cloud. For example if C and D are two cloud providers, and A_1 and A_2 are two actors, the notation $A_1.C$ and $A_2.D$ means that the two actors run on different cloud providers. If also A_2 was running on the cloud provided by C, then the knowledge set of C would include all the information items that A_1 and A_2 observe. This situation is particularly relevant in the case of a secret sharing based control, where if secret information is shared between A_1 and A_2, the cloud provider can directly reconstruct the secret. In notation: if at some time t the share $Sh_{1,2}[A]$ is held by A_1 and $Sh_{2,2}[A]$ is held by A_2 then C knows the entire information item A. This scenario can be modelled by what we call the *Cloud Transparency Equation* (CTE):

$$\cup_i KS(A_i.C) = KS(C) \tag{4.4}$$

4.4.2.1 Formalizing Cloud Transparency

In principle, data owners wishing to use the cloud for running their processes could handle data protection by hosting controls (e.g., the ones performing encryption and decryption) on their own (or on a trusted third party's) premises. Unfortunately, a single untrusted cloud provider who is not allowed to decrypt customer data will be unable to deliver processing and display services. Using a different cloud for executing security controls may look a viable option; indeed PRACTICE itself deals with on scenarios where multiple untrusted clouds and combinations of FHE, secret sharing and other Secure Multi-Party Computation (SMC) techniques are used to process encrypted data without decryption.
Generally speaking, however, our Cloud Transparency Equation (CTE) can be adapted to represent

Cloud transparency, i.e. the effect of techniques for controlling access on the part of the cloud provider's *internal roles* to information exchanged by actors taking part to outsourced processes.
Most of these techniques (nested virtualisation, domain disaggregation and secure execution environment) [15, 66, 87, 10, 99], require architectural modifications,such as instrumenting one or more levels of the cloud stack with tamper-resistant hardware/software components aimed at protecting users' messages from unauthorised inspection and misuse on the part of internal roles - for instance, in Infrastructure-as-a-Service (IaaS) cloud offerings, the employee of the cloud provider who administers the domain where the business process takes place. To represent the effect of such architectural modification techniques, one can write in general that in a modified cloud stack the knowledge of the community of the actors will be larger that the one of the domain administrator role in charge of the process:

$$\cup_i KS(A_i.C) = K(C) > KS(Dom_C) \tag{4.5}$$

It is important to remark that our equational formalism can be used to specify which provider role has access to which information item. For instance, equation $KS(A.C) \in KS(Dom_C)$ specifies that the domain administrator has full monitoring of the information held by actor A, and similar equations can be written to denote that the domain administrator can observe a part of the messages and local memories involved in the process. The development of the equational annotations will be completed in the next version of this deliverable D31.3.

4.4.3 The Knowledge Transformation Rules

Our model includes a set of rules defining how the Knowledge Set evolves, and what are the information items that the actors can reconstruct during their participation to the business process. To this purpose, our notation provides a set of *knowledge transformation rules*, that can be used to automatically compute the information items shared among a given set of actors, as resulting from the previously exchanged messages and a-priori knowledge. The rule explicitly represent basic facts, such as the possibility to access an encrypted message if both the key and the message are held by the involved actors, or that a set of actors can reconstruct a shared secret, if they hold all the shares. Such rules are necessary to automatically reason on the shared knowledge and are the basis for the tool described in detail in section 7.

The Encryption rule This rule expresses the fact that holding the decryption key together with holding a data item encrypted with it is equivalent to holding the plaintext.

$$K_A \oplus E_{K_A}[X] \rightarrow X \tag{4.6}$$

Sharing reconstruction This rule expresses the fact that holding all the shares of a data item is equivalent to holding the plaintext.

$$Sh_{1,n}[A] \oplus Sh_{2,n}[A] \oplus \cdots \oplus Sh_{n,n}[A] \rightarrow A \tag{4.7}$$

The Homomorphic Encryption rule This rule expresses the fact that holding operands homomorphically encrypted w.r.t. an arithmetical operation is equivalent to holding the encrypted result of the operation.

$$HE_{K_A}[X] \oplus HE_{K_A}[Y] \rightarrow HE_{K_A}[X \otimes Y] \tag{4.8}$$

As the CTE, these equational annotations can be developed to express the properties of a number of different security controls applicable to the business processes. The development of the equational annotations will be completed in the next version of this deliverable D31.3.

Chapter 5

The Likelihood Assessment Module: A Possibilistic Approach

In this chapter we focus on a specific but important category of data disclosure events, the ones that bring one or more parties taking part to a cloud-based business process to know more information than the process execution would entail. We briefly recall the nature of these threats, which have been described in detail in D31.1; then we describe our new possibilistic approach for likelihood estimation, discussing when and where it should be used instead of the probabilistic one reported in 31.1.

5.1 The Threat Space

Any risk model must clearly specify the *event space* where risk will be quantified. Here we focus on a single, albeit large, family of threats, namely *data process-related leakage threats*, i.e. the disclosure of one or more information items to be exchanged in a multi-party protocol to participating parties who are not the originally intended recipients. These unwanted disclosures may be due to intentional publishing of supposedly protected information items, or to carelessness in the communication protocol implementation and deployment, e.g. when one party is using the same mobile terminal previously used by another and can reconstruct the information items held. We call these events *process-related data disclosures*, in order to distinguish them from disclosures due to conventional eavesdropping attacks.
Often *process-related data disclosure* happens when actors (*including service and cloud providers*, as outlined in Chapter 4) put together the partial information they hold to reconstruct knowledge that is not available to them when taken individually. We remark that unexpected sharing of partial information is not always caused by collusion among rogue participants. Indeed, actors taking part to a business process may put together their information for other reasons, including:

- eDisclosure, i.e. the mandatory process of disclosing information to adversaries during litigation.
- Information requests from regulatory authorities.[1]
- Inadvertent or dysfunctional behavior of employees.

[1] Whether data is stored on premises or in the cloud, the data owner's obligation to comply with the demands of the courts or regulatory authorities remains essentially the same. See for instance http://www.edisclosureinformation.co.uk.

As far as the first factor is concerned, we remark that data sharing imposed by courts of law may generate leaks that are difficult to identify a priori even for experienced security auditors. The second factor - the intervention of a regulatory authority - is also difficult to predict. For instance, a customer's emails containing bids for an auction held in one country may be stored on a cloud server located in another jurisdiction, where a regulatory authority can ask the cloud provider - even for reasons unrelated to the auction - to get full access to the storage of the mail server, without informing the auctioneer. This way, a third party would get to know in advance the outcome of the auction.
As one would expect, the third factor has the strongest documentary evidence. A global security study on data leakage, commissioned by Cisco and conducted by a U.S.-based market research firm [21] polled more than 2000 employees and information technology professionals in 10 countries, including major EU markets. The study identified the adverse event of unwanted information sharing, related to sloppy implementations of interchange protocols, or intentional communication with un-authorized parties. For instance, a plain text email containing a business offer sent in good faith through a "secure" cloud-based mail service poses a danger if disclosed by the cloud provider to a competitor of the original sender. Today, it is very challenging even for experienced process owners to fully identify, analyze and handle data leakage risks, due to the complexity and diversity of business processes and of the underlying IT systems; the trend toward outsourcing and the cloud is further blurring the scenario. Many organizations have little visibility into where their confidential data is stored on the cloud or control over where that data is transferred during the execution of a process. Even when insight is available, organizations often lack a clear methodology to assess whether the process involves an acceptable level of risk.
Hereafter we will focus on the information disclosures taking place during the provision of the cloud service and distinguish between two main categories of attacks/threats:

- Information disclosure by actors taking part in the cloud based collaboration process (either individual actors or colluding subsets).
- Information disclosure by the cloud provider orbits personnel.

5.2 The Approach

We start from the notion that a limited set of major motivations, which can drive a business process actor or the cloud provider towards an attack: we group them under the categories of *perceived unfairness*, *greed* and *generic contextual factors*. The first two factors derive from the model of individuals and corporations as rational profit maximisers borrowed from neo-classical economics, where shareholders are modelled as having no interests other than increasing the value of the stock they hold or their bonuses, while directors and managers are represented as entirely self-interested [51]. Our approach consists in identifying the role played by the different actors, and in modelling quantitatively their different drivers in relation to each actor and each relevant coalition, either by probabilistic or possibilistic techniques; we then propagate the uncertainty associated to the elements of the model and give a synthesis estimate of the risk associated to each player. The same kind of computation is carried out in relation to scenarios in which the different available countermeasures are deployed (e.g. use of secret sharing techniques over business process actors data shares) and the cost/benefit balance in terms of countermeasures cost and risk reduction is drawn.
At the core of our methodology there is also the Value of Information (VoI) technique to the estimate of the economic worth of the information assets involved. Value of Information (VoI) is a concept from decision analysis: how much holding a piece of information allows a decision-maker to improve its decision. From a decision making standpoint, extra information is only useful if it has a non-zero

probability of changing the current strategy. The penalty of extra information is usually valued as its cost, and sometimes also the delay incurred in waiting for it.
The usual starting point of a VoI analysis is to consider the *Value of Perfect Information* (VoPI), i.e. answering the question ”What would be the benefit for an actor (usually quantified as money, but it could be in terms of material assets like customer satisfaction), to be able to know some information item(s) perfectly?”.
If perfect knowledge would not change the actor’s behaviour, then the extra information is worthless. If it it does change it, then the value is the difference between the expected net benefit of the new behaviour compared to that of the original behaviour. VoPI is a useful limiting tool, because it provides the maximum value that any data may have for a given actor. If learning the information involves a cost more than that maximum value, a fully rational actor will not pursue it any further.

In this chapter we discuss quantitative models of the above mentioned drivers of *perceived unfairness*, *greed* and *generic contextual factors*. The driver represented by greed can be modelled within decision theory by representing the attacker as a rational agent with a specific greed related utility function: the option of whether to deliver or not an attack corresponds to the decision of whether to enter or not in a bet, structured around possible benefits, possible drawbacks and their likelihoods. Another core element which will be modeled is the representation of the driver of *perceived unfairness* by means of an index based on the game theoretic concept known as *Shapley Value*.
Since the uncertainty associated with the relevant knowledge does not have necessarily a frequentistic form which could be used in Probabilistic models, we start by outlining the basic elements of the possibilistic representation of uncertainty, suitable to our methodology.
The material that follows extends and develops the discussion in D31.1. Namely, in the next section (Section 5.3) we discuss the general elements entering the assessment of risk; in the following section (Section 5.5) we recall the basics of Possibility Theory, then (Section 5.6) we illustrate our methodology.

5.3 Risk modeling: Probabilistic vs. Possibilistic approach

5.3.1 The Components of Risk

Risk definition From the economic standpoint, the *risk* $R(A,E)$ of an "adverse event" (a.k.a. "feared event") E for a given actor A is often represented as the product of the damage $I(E)$ (expressed in currency units) in which A incurs when E really happens, times the likelihood that E might happen, traditionally represented in terms of probabilities. In symbols: $R(A,E) = I(E)\,Pr(E)$.
In the computer security context, one needs to identify all the adverse events as manifestations of security threats and for each estimate $I(E)$ and $Pr(E)$; then the overall risk is computed by a suitable aggregator. Typically [93], the risk assessor puts herself in the place of a specific actor (e.g. the *process owner*, i.e. the stakeholder in whose name, a business process P is executed) and asks the following questions:

- Which adverse event can happen to the information assets involved in P? (*threat categorization*)
- How severe could that event be for the process owner? (*threat impact assessment*)
- How much is this event plausible? (*threat likelihood assessment*)

Likelihood As far as the estimate of the likelihood is concerned, many risk assessment methods use predictive models involving a certain number of parameters, and each parameter is usually affected by some degree of uncertainty. There are cases where the probability of adverse events can eventually be obtained by propagating this uncertainty (i.e. either by analytical or Monte Carlo methods). However, in many cases, assessing the probability of the parameter values is too difficult due to incompleteness of information. As put forward in [83] "*Probability is perfect, but we cannot elicit it perfectly*". In fact, the uncertainty involved in risk assessment has typically two distinct origins: it can be due either to *variability* or to *imprecision* (or to both factors). *Variability* (also referred to as "objective uncertainty") arises often (but not exclusively) from the *random* character of natural processes. *Imprecision* (also referred to as incomplete information, partial ignorance or "subjective uncertainty") arises from the *partial* character of *individual knowledge* about the state of the world. Traditionally, in risk assessment no distinction was made between these two types of uncertainty, both being represented by means of a single probability distribution. However, in case of partial ignorance, the use of probability measures introduces information that is in fact not available: this may seriously bias the outcome of a risk analysis [7]. While random variability can be suitably represented by probability measures and propagated by the methods of *Probability Theory*, incomplete information is better accounted for and propagated by the methods of *Possibility Theory*.
Possibility Theory is similar to Probability Theory in that it is based on set functions, but differs from it by the use of *a pair* of dual set-functions (the *possibility measure* and the *necessity measure*) instead of only one (the probability measure) and by the fact that is not additive, but sub-additive, and can be cast either in an ordinal or in a numerical setting (the two formulations differ in terms of conditioning, and independence notions). The latter form will be used in this chapter. The development of Possibility Theory is due to a large number of authors, especially to Lotfi Zadeh [81], Didier Dubois and Henri Prade [27, 31] (a comprehensive review can be found in [34, 35, 22]). As we have seen in the previous chapter, our methodology relies on the knowledge of the business process model and its underlying micro-economics. Here we use them to attach *possibilities* to actors' misbehavior or violation of confidentiality. Namely, we are going to describe in section 5.6.1 how knowledge about the likelihood of an adverse event can be elicited and suitably represented in possibilistic form.

Impact As far as the impact (a.k.a. *severity*) is concerned, its precise quantification is often a challenge. In our methodology, we provide an evaluation of costs taking into account the value of the disclosed information by means of a set of techniques known as VoI analysis. As we shall see, uncertain knowledge about the impact of adverse events on processes can be suitably represented in possibilistic form (section 5.6.2). By composing possibilistic information about the likelihood of an event and the possibilistic information about the impact of that event, one can obtain a possibilistic representation of the risk of an adverse event and use this as the basis for taking decisions.

5.4 Possibility Theory in Risk Assessment

The drawbacks of computing the risk as a product of perfectly known factors, such as probability and impact, have been pointed out by several authors. A major criticism regards the way experts assign precise numerical values to risk parameters (see for instance [96, 42, 41]). Possibility Theory has found applications in several domains including project investment decision [80], project network analysis [72], contract decision making [94] and safety performance [90]. Within risk assessment, Possibility Theory has been used in various hybrid forms and in a limited number of use cases (see for instance [64]). From the methodological point of view, several works have focused on the problem of merging data and expert opinion [32, 33]. In the context of general decision making with possible applications to risk management, most works focus on the joint propagation of variability and imprecision in the risk assessment process [7, 28, 84].

5.5 Elements of Possibility Theory

5.5.1 Possibility Distributions

The basic entity of Possibility Theory is the *possibility distribution*, denoted by π, which represents the subjective knowledge of an agent about the actual state of affairs x of a quantity. It consists in a mapping π from a set of states of affairs S to a totally ordered scale of plausibility, such as the unit interval $[0,1]$.

$$\pi_x : \quad s \in S \rightarrow \pi(s) \in [0,1]$$

it distinguishes what is plausible from what is less plausible for the ill-known quantity x using the following conventions:

- $\pi_x(s) = 0$ means that the state s is rejected as *impossible* for x;
- $\pi_x(s) = 1$ means that the state s is *totally possible* (= unsurprising) for x.

If the state space is complete, at least one of its elements should be the actual state (closed word hypothesis), so that at least one state is totally possible: this is referred to as the *normalisation condition*:

$$\max_{s \in S} \pi_x(s) = 1$$

Notice that distinct values within S may simultaneously have a degree of possibility equal to 1. Possibility theory is driven by the principle of minimal specificity, which states that any hypothesis not known to be impossible cannot be ruled out.
This framework can represent extreme forms of partial knowledge:

- *complete knowledge*: for some state s_0, $\pi(s_0) = 1$ and for all the others $\pi(s) = 0$, i.e. only s_0 is possible;

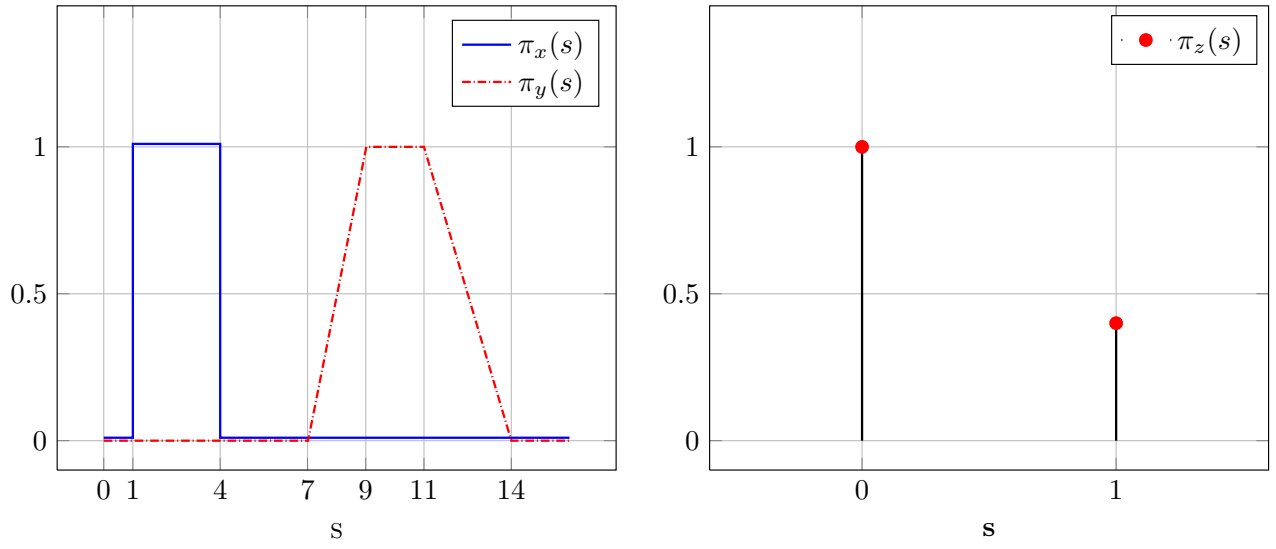

Figure 5.1: Left: (solid – blue – line) a crisp set membership function $\pi_x(s)$ representing the knowledge about the real quantity x: "x cannot lie outside the set $E = [1,4] \subset S = \mathbb{R}$"; (dashed-dotted – red – line) a fuzzy membership function π_y representing the knowledge about the quantity y . Right: a fuzzy membership function π_z representing the knowledge about the quantity y that can take only values in $\{0,1\}$: the value $z = 0$ is totally possible, or unsurprising, i.e. $\pi_z(0) = 1$ while $\pi_z(1) = 0.4$. Notice that, unlike probability, the sum of possibility values does not necessarily amount to 1.

- *complete ignorance*: $\pi(s) = 1 \quad \forall s \in S$, i.e. all the states are totally possible.

The simplest non-extreme form of a *possibility distribution* on the set S is the characteristic function of a subset E of S, i.e. $\pi(x \in E) = 1$ and $\pi(x \notin E) = 0$. It models the situation when all that is known about x is that it cannot lie outside E.
This type of possibility distribution is naturally obtained from *experts* stating that a quantity x lies between values a and b: in which case E is the interval $[a,b]$. However, this binary representation is not entirely satisfactory: sometimes, even the widest set of possible values does not rule out some residual possibility that the value of x lies outside it: so it is natural to use a graded notion of possibility.
In this case, *formally, a possibility distribution π coincides with the membership function μ_F* of a *fuzzy subset* $F \subseteq S$, such that $\mu_F(s) = \pi_x(s)$ (see Figure 5.1 for illustrative examples). The core of a possibility distribution $\pi_x(s)$ is the set of s such that $\pi_x(s) = 1$; the support is the set such that $\pi_x(s) > 0$.
On the real line $S = \mathbb{R}$ one can interpret the possibility distribution $\pi_x(s)$ as a family of nested confidence intervals [29].

5.5.2 Recall of Fuzzy Set Theory

Fuzzy sets Fuzzy sets are sets whose elements have degrees of membership, expressed by a membership function. The latter is a generalisation of the indicator function of classical sets (often called "crisp sets" in this context). An element s is said to be fully included in F if $\mu_F(s) = 1$ and not included if $\mu_F(s) = 0$. A fuzzy set whose membership function is convex, normalized, at least segmentally continuous and has the functional value 1 at precisely one element is called *fuzzy number* (typical examples are triangular membership functions); if it has the functional value 1 at exactly one mid interval point it is called *fuzzy interval* (typical examples are trapezoidal membership functions); a triangular fuzzy number is clearly a special case of a trapezoidal fuzzy interval. Conversely, sometimes the latter is called trapezoidal fuzzy number. The traditional distinction between fuzzy numbers and fuzzy intervals is not always taken in account: often, both are called fuzzy numbers. Since a fuzzy set is a collection of objects with various degrees of membership, it is often useful to consider those elements that have *at least* some specified level of membership α: this leads to the concept of α*-cut*. Given a value of $\alpha \in [0,1]$ an α-cut of the fuzzy set F is the crisp set

$$F^\alpha = \{s : \mu_F(s) \geq \alpha\} \quad \text{for} \quad 0 < \alpha \leq 1$$

A fuzzy set F is completely characterised by its α-cuts: $F = \cup_{\alpha\in[0,1]} \alpha F^\alpha$.
If μ is a fuzzy interval, its α-cuts are crisp intervals included in its support. They have the form $[\underline{a}^\alpha, \overline{a}^\alpha]$, where $\underline{a}^\alpha = \min\{s|\mu_F(s) \geq \alpha\}$ and $\overline{a}^\alpha = \max\{s|\mu_F(s) \geq \alpha\}$. The α-cuts are often denoted in the shorthand interval form $F_\alpha = [\mu_F]^\alpha$. The α-cut representation allows a straightforward extension of classical operations from crisp numbers to fuzzy numbers, discussed below.

Fuzzy aggregators Classical set operations such as complement, intersection and union can be given a fuzzy analog: there are several alternative ways for defining those classical notions to the fuzzy domain still preserving desirable axioms (such as that the operator behaves as the classical operator when applied to crisp sets, or that it has the same associativity properties and commutativity properties of its classical counterpart). The functions which provide an extension to the *intersection operator* and fulfill the axioms of commutativity, associativity, neutral element-1 and monotonicity are called T-norms. The term originates from the studies of probabilistic metric spaces, in which T-norms and T-conorms (a.k.a. *S-norms*) were used to extend the Triangular Inequality (the seminal paper on T-norms in fuzzy logic is [59], for a review see [53] for a complete discussion of a use case see [26]). It is worth noticing that the concept of T-norms overlaps with the concept of copulas [1] and conveys the semantics of the assumed relationship between the fuzzy sets involved. The so-called standard T-norm (for the intersection) and T-conorm (for the union) which were originally used in Lotfi Zadeh seminal paper [98] correspond to the following definitions for the three basic operators:

- standard *complement*: $\mu_{\overline{F}}(s) = 1 - \mu_F(s)$
- standard *intersection*: $\mu_{F\cap G}(s) = \min(\mu_F(s), \mu_G(s))$
- standard *union*: $\mu_{F\cup G}(s) = \max(\mu_F(s), \mu_G(s))$.

These aggregators are known also as *min-max norms*, or Gödel's norms. Other frequently used norms (see [53]) are the so called *probabilistic norms*, where the T-norm is based on the product ($\mu_{F\cap G}(s) = \mu_F(s)\mu_G(s)$) and the T-conorms is based on sums ($\mu_{F\cup G}(s) = \mu_F(s) + \mu_G(s) - \mu_F(s)\mu_G(s)$).
In the context of Possibility Theory, the membership function is used in the meaning of *possibility distribution* of ill-known quantities: therefore the value $\pi_{(x_1,x_2)}(s_1,s_2)$ – from now on for short $\pi(s_1,s_2)$ – takes the meaning of joint possibility for two values s_1 and s_2 of two ill-known quantities x_1 and x_2.

It expresses the plausibility of the event $(x_1, x_2) = (s_1, s_2)$ defined by the simultaneous occurrence of the two values. In Possibility Theory the aggregators T-norm and T-conorm specify the way in which one can obtain the joint possibility, when given the individual possibility distribution and allow to express the assumptions about the general relationship between two possibilities (notably, in probability, the choice of adopting the product copula reflects assumption of independence between variables). Hereafter we adopt the standard T-norm

$$\pi(s_1, s_2) = \min(\ \pi_{x_1}(s_1)\,,\, \pi_{x_2}(s_2)\ )\,.$$

In the possibilistic framework, our choice for the joint distribution reflects the already mentioned *principle of minimal specificity*: the above expression presupposes nothing on the possible dependence between quantities x and y. The corresponding T-conorm, defined as the max of the two possibility distributions reflects in some sense (in the risk assessment context) a "pessimistic" assumption about the fact that the joint possibility is the possibility of the most possible adverse event. When the definition above is adopted, the variables are called *separable*.

The Extension Principle It is possible to extend n-ary functions from ordinary crisp numbers and intervals to fuzzy numbers and intervals, using the *extension principle* by Zadeh [98]. In the case of a two place function f it works as follows.
If a joint possibility function relating two quantities x_1 and x_2 is *separable*,
i.e. if $\pi(s_1, s_2) = \min(\pi_{x_1}(s_1), \pi_{x_2}(s_2))$, the possibility distribution for the value of the (ill-known) function $z = f(s_1, s_2)$ is defined (with the convention of not indicating the regions where the possibility is zero) by the following convolution

$$\pi_f(z) = \bigcup_{(s_1,s_2)\in f^{-1}(z)} \Big(\pi_{x_1}(s_1) \bigcap \pi_{x_1}(s_1)\Big) = \sup_{(s_1,s_2)\in f^{-1}(z)} \{\ \min(\pi_{x_1}(s_1), \pi_{x_2}(s_2))\ \}$$

In the case of the sum operation $\oplus$ and of the product operation $\otimes$ on fuzzy numbers and intervals, the above principle reduces respectively to

$$\pi_{x_1\oplus x_2}(z) = \bigcup_{s_1+s_2=z} \Big(\pi_{x_1}(s_1) \bigcap \pi_{x_1}(s_1)\Big) = \max_{s_1+s_2=z} \min(\pi_{x_1}(s_1), \pi_{x_2}(s_2))$$

$$\pi_{x_1\otimes x_2}(z) = \bigcup_{s_1\times s_2=z} \Big(\pi_{x_1}(s_1) \bigcap \pi_{x_1}(s_1)\Big) = \max_{s_1\times s_2=z} \min(\pi_{x_1}(s_1), \pi_{x_2}(s_2))$$

A straightforward way to implement these operations consists in developing each input possibility distribution, seen as a fuzzy interval, in terms of the crisp intervals defined by their α-cuts, and then apply the operations to (the extremes of) those intervals and finally reconstructing the output fuzzy interval from those α-cuts. For instance, if one wants to extend the sum operator $+$ from crisp numbers to fuzzy numbers so as to define a new operator $\oplus$, one can observe that the sum of two intervals $[\underline{c}, \overline{c}]$ and $[\underline{d}, \overline{d}]$ can be defined as another interval $[\underline{c}+\underline{d}, \overline{c}+\overline{d}]$; if we need to add two fuzzy numbers we can do it by adding the corresponding α-cuts for every level α. This construction based on α-cuts, due to Nguyen [82], can be used for defining the full arithmetic of fuzzy numbers [65]. Thanks to the α-cut based representation of fuzzy numbers and intervals and to interval arithmetics, the following results can be obtained for possibility distributions. Given two (fuzzy) possibility distributions $a \equiv \pi_{x_1}(s_1)$, $b \equiv \pi_{x_2}(s_2)$ and a scalar $\lambda \in \mathbb{R}$ and $\lambda \geq 0$, let us use the notation $[b]^\alpha = [\underline{b}^\alpha, \overline{b}^\alpha]$, we have

$$[a \oplus b]^\alpha = \left[\underline{(a \oplus b)^\alpha}\ ,\ \overline{(a \oplus b)^\alpha}\right] = \left[\underline{a}^\alpha + \underline{b}^\alpha\ ,\ \overline{a}^\alpha + \overline{b}^\alpha\right]$$

$$[\lambda \odot b]^{\alpha} = \left[\underline{(\lambda \cdot b)^{\alpha}}\ ,\ \overline{(\lambda \cdot b)^{\alpha}}\right] = \left[\lambda \cdot \underline{b}^{\alpha}\ ,\ \lambda \cdot \overline{b}^{\alpha}\right]$$

while

$$[a \otimes b]^{\alpha} = \left[\underline{(a \otimes b)^{\alpha}}\ ,\ \overline{(a \otimes b)^{\alpha}}\right]$$

with

$$\underline{(a \otimes b)^{\alpha}} = \min\{\underline{a^{\alpha}}\,\underline{b^{\alpha}}, \underline{a^{\alpha}}\,\overline{b^{\alpha}}, \overline{a^{\alpha}}\,\underline{b^{\alpha}}, \overline{a^{\alpha}}\,\overline{b^{\alpha}}\}$$

$$\overline{(a \otimes b)^{\alpha}} = \max\{\underline{a^{\alpha}}\,\underline{b^{\alpha}}, \underline{a^{\alpha}}\,\overline{b^{\alpha}}, \overline{a^{\alpha}}\,\underline{b^{\alpha}}, \overline{a^{\alpha}}\,\overline{b^{\alpha}}\}$$

Thanks to the definition of suitable logical aggregators and to the extension of the necessary operations from crisp to fuzzy values, it is thus possible to propagate the uncertainty through system/process models from the input variables to the output variables.

5.5.3 Possibility and Necessity

Possibility and Necessity degree (measures) of an event Based on the knowledge captured by the *possibility distribution*, Possibility Theory provides two dual evaluations of the likelihood of an event (for instance, that the actual value x of an ill-know quantity, for instance a memory cell, should lie within a certain interval): the *possibility* Π and the *necessity* N of the event. Possibility measures refer to the idea of *plausibility*, while the dual necessity measures refer to the idea of *certainty*. Given a subset of states A and the event "*the actual value x of the unknown quantity lies in A*" the normalized measure of possibility Π and necessity N are defined from the possibility distribution $\pi : S \to [0,1]$ such that $\sup_{s \in S} \pi_x(s) = 1$ as follows:

$$\Pi(A) = \sup_{s \in A} \pi_x(s) \tag{5.1}$$

$$N(A) = 1 - \Pi(\overline{A}) = \inf_{s \notin A}(1 - \pi_x(s)) \tag{5.2}$$

Π tells to what extent *at least one element of A* is consistent with the knowledge π_x (i.e. is possible), while $N(A)$ tells to what extent *no element outside A* is possible (i.e. to what extent A is implied by the knowledge π_x). The possibility-necessity duality is expressed by $N(A) = 1 - \Pi(A^c)$, where A^c is the complement of A. Generally $\Pi(S) = N(S) = 1$ while $\Pi(\emptyset) = N(\emptyset) = 0$.
See Figure 5.2 for a visual illustration of the Possibility-Necessity space.
The two measures can be seen as measures of *consistency* and *implication* respectively. They are dually related in that the certainty of an event reflects a lack of plausibility of its opposite. However, they cannot be obtained one from another. This represents a remarkable difference with Probability Theory, where probability is a self-dual measure (i.e., in probability theory $Pr(A) = 1 - Pr(\overline{A})$). However, degrees of necessity can be equated to lower probability bounds and degrees of possibility to upper probability bounds.

Maxivity and minivity axioms The possibility measure Π verifies the *maxivity* axiom

$$\forall A, B \subseteq S \quad \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

The necessity measure verifies a dual axiom, the *minitivity* axiom

$$\forall A, B \subseteq S \quad N(A \cap B) = \min(N(A), N(B))$$

In our context of Risk Assessment, the possibility measure of a harmful event can be thought of as the criterium that would be adopted by a pessimistic decision maker, whereas the necessity measure can

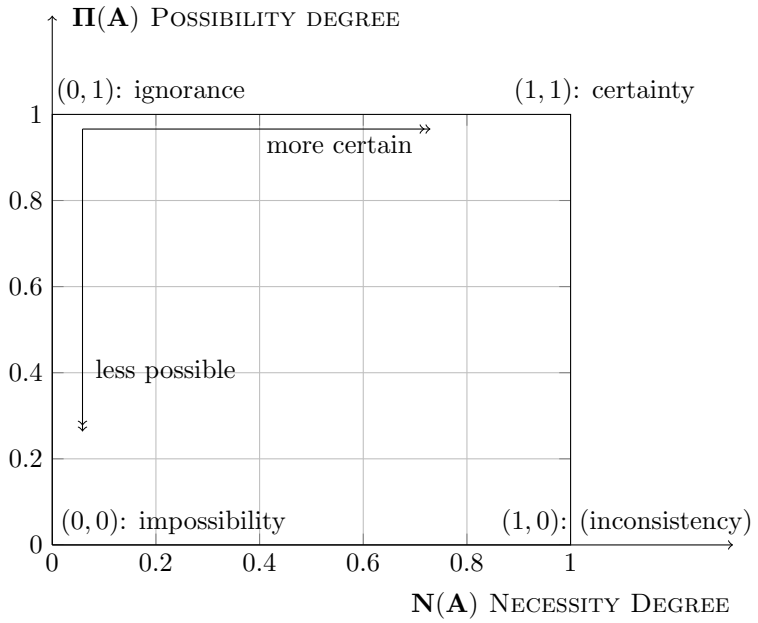

Figure 5.2: The Possibility-Necessity space

be thought of as the criterium which would be adopted by an optimistic decision maker. Sometimes, risk analysts use a synthesis indicator of the Possibility and Necessity of an event, the so called Credibility degree [95]. The Credibility measure of an event A is defined as its arithmetic average of its Possibility and its Necessity:

$$Cr(A) = \frac{1}{2}(\Pi(A) + N(A)) \tag{5.3}$$

There are several approaches to the definition within Possibility Theory of indicators formally analogous to the Expected Value, Variance and Covariance of Probability Theory [47, 70, 71]: the definition of possibilistic expected value adopted in several risk assessment contexts [44, 48, 49, 45] is the following, based on α-cuts. Given a fuzzy number A with alpha cuts $A_\alpha = [\underline{a}, \overline{a}]$, and given a weighting function $f(\alpha) : [0,1] \to \mathbb{R}$ which is non-negative, monotone increasing and such that $\int_0^1 f(\alpha) d\alpha = 1$, the *$f$-weighted possibilistic expected value* of A is

$$E_f(A) = \frac{1}{2} \int_0^1 d\alpha \, (\, \underline{a}(\alpha) + \overline{a}(\alpha) \,) \, f(\alpha) \tag{5.4}$$

For $f(\alpha) = 2\alpha$ the quantity $E_f(A)$ defines a possibilistic mean value. Following [46] the weighted possibilistic variance can be defined as

$$E_f(A) = \frac{1}{12} \int_0^1 d\alpha \, (\, \underline{a}(\alpha) - \overline{a}(\alpha) \,)^2 \, f(\alpha) \tag{5.5}$$

5.5.4 Possibility Propagation in Risk Assessment

As we have seen in the previous chapter, most Risk Analysis methodologies prescribe to proceed through a bottom-up compositional approach towards the risk estimate. Typically, one should start by breaking the system/process under analysis down to the component level. For each component, one should list all the possible threats or failure modes; obtain the threat likelihood and the impact of each mode and each component; propagate that information from component level to the next level of hierarchy up to system level and finally obtain the risk value for each failure mode. At this point, depending on the type of decision that has to be supported by the analysis, one can take different steps, for instance prioritise risks by failure mode or by component and take consequent actions or state the degree of conformance to the objectives.
In risk assessment the following issues are considered fundamental:

1. Representing all available information faithfully;
2. Accounting for dependencies, correlations between the parameters;
3. Choosing a suitable propagation technique.

We will now discuss these issues with reference to our methodology.

5.5.4.1 Input Variables and their Relationships

Input information representation Most of the information relevant to the risk assessment of the process of interest come from elicitation of expert opinion (for a review of elicitation methods see [11]): for this kind of information the possibilistic representation is better than the probabilistic one because it is difficult for experts to assign crisp numerical values to risk parameters. If different opinions about a parameter are provided by different experts or sources, they can be merged by the use of suitable operators or knowledge fusion procedures [32]: in the context of crisp variables, it is common to use weighted averages (the weight being the level of expertise of the source); an analogous of the weighted average of fuzzy sets can be devised for fuzzy numbers by applying the extension principle. In order to explain in detail this aspect of our methodology, we assume, for sake of simplicity, that a single expert has been consulted, and that, provided with the available knowledge, she has expressed its opinion/knowledge in terms of possibility distributions over the space of the possible events.

Relationship among variables Also the assumptions on dependencies can be elicited from expert opinion. The most conservative attitude in absence of contrary expert statements would be to make the least specific assumption and that the variables are separable. For the sake of simplicity, here we adopt this simplifying assumption.

5.5.4.2 Reliability Propagation

Knowledge propagation from the system/process components level description to the system/process as a whole can be carried on through the application of logical aggregators and functional aggregators defined on the basis of the extension principle. For instance, following [62, 58] (other possibilistic approaches to reliability are reviewed in [57]) one can perform a fault-tree based propagation of the *possibility of failure*: from the possibility distributions of a variable describing the failure of individual system components it is possible to work out the possibility distribution of variables describing the failure of the whole system.

Let us consider a reliability model where event behavior is fully characterized in terms of possibility measures and at any instant of time a system component is in one of two crisp states: perfectly functioning $s = 1$ or completely failed $s = 0$, i.e. $s \in S = \{0,1\} = \{fail, work\}$. This setup is called Posbist (from "Possibilistic, binary states"). We assume that expert opinion has assigned to the ill-known state x_i of the i-th component the possibilities $\pi_{x_i}(1) = \pi(x_i = 1) = r_i$ and $\pi_{x_i}(0) = \pi(x_i 0) = u_i$. Under the hypothesis that the state of the system is determined completely by the states of its components, the structure function of an n components system can be denoted by

$$G = G(\pi_{x_1}, ..., \pi_{x_i}, ...\pi_{x_n}).$$

Also for the ill-known system state G we adopt the convention $G = 1$ if it is functioning and $G = 0$ if it is failed: the system Possibilistic Reliability is defined as $R_G = \pi(G = 1) = \pi_G(1)$ and the system Possibilistic Unreliability is defined as $U_G = \pi(G = 0) = \pi_G(0)$.

Series systems In the case the system structure function is structured with all AND gates (i.e. as a *series* of components), the system works if all the components work correctly, and fails if at least one fails. The possibilistic Reliability is

$$R_G^{series} = \pi(G = 1) = \pi\left(\bigcap_{i=1}^{n}(x_i = 1)\right) = \min(r_1, ...r_n),$$

the possibilistic Unreliability

$$U_G^{series} = \pi(G = 0) = \pi\left(\bigcup_{i=1}^{n}(x_i = 0)\right) = \max(u_1, ...u_n).$$

Notice that neither in R nor in U there is an explicit dependence on the number of components.

Parallel systems In the case the system structure function is structured with all OR gates (as a *parallel* of components), the system works if at least one works and fails if all fail. The possibilistic Reliability is

$$R_G^{parallel} = \pi(G = 1) = \pi\left(\bigcup_{i=1}^{n}(x_i = 1)\right) = \max(r_1, ..., r_n)$$

the possibilistic Unreliability

$$U_G^{parallel} = \pi(G = 0) = \pi\left(\bigcap_{i=1}^{n}(x_i = 0)\right) = \min(u_1, ...u_n)$$

Neither in R nor in U, again, there is an explicit dependence on the number of components.

k-out-of-n systems A system with n unrelated components is said to have a k-out-of-n G structure if it operates when at least k components operate. Fully parallel systems are 1-out-of-n G systems. Also for general k-out-of-n systems, other than the parallel system, it is possible to define a Possibilistic Reliability $R^{k/n}$ and a Possibilistic Unreliability $U^{k/n}$ following the same procedure as above.
For the sake of illustration let us consider an $n = 3$ component system and indicate by A, B and C the events $(x_1 = 1)$, $(x_2 = 1)$ and $(x_3 = 1)$ corresponding to the correct operation of component one, two and three respectively; let, as usual, $\pi_{x_1}(s)$, $\pi_{x_2}(s)$ and $\pi_{x_3}(s)$ be the possibility distributions of the ill

known states x_1, x_2 and x_3 of the components. Let us assume that $\pi_{x_1}(1) > \pi_{x_2}(1) > \pi_{x_3}(1)$ and that $\pi_{x_1}(0) < \pi_{x_2}(0) < \pi_{x_3}(0)$ – in shorthand notation $r_1 > r_2 > r_3$ and $u_1 < u_2 < u_3$.
Suppose this is a 2-out-of-3 system. Its Possibilistic Reliability is derived considering that the system works if either exactly 3 components work or if exactly 2 work.

$$\begin{aligned} R_G^{2/3} = \pi(G=1) &= \pi\left((A\cap B\cap C)\bigcup(A\cap B)\bigcup(A\cap C)\bigcup(B\cap C)\right) \\ &= \max\left(\min(r_1,r_2,r_3),\min(r_1,r_2),\min(r_2,r_3),\min(r_1,r_3)\right) = r_2 \end{aligned}$$

We remark that the outcome can be derived also by the following general considerations, holding for the k-out-of-n case. The computation is structured in three steps:

1. Get the power-set of the actors and select the sets with at least k elements.
2. Take the minimum of r_i out of each set (if the values are strictly ordered all those minima are different form one another).
3. Choose the maximum among them (in a strict ordering, the latter operation also identifies the set from which the element comes, so, in fact, the procedure identifies the set that contributes the highest minimum).

Now, no set with cardinality greater than k can contribute the highest minimum, because those sets would have to be eventually compared with the set of the top k elements, whose minimum would prevail. Hence one can focus only on the k-elements sets: however also among them the one which has the highest minimum is the top-k-elements set. Hence the chosen value must be the k-th value of the ordered list. In synthesis: the Possibilistic Reliability $R_G^{k/n}$ of a k-out-of-n system based on a set A of components whose possibilistic reliabilities form a strictly descending ordered list $(r_1, r_2, \ldots, r_n)$ is

$$R_G^{k/n} = \max_{E\in\mathscr{P}(A):|E|=k}\left(\min_{i\in E}(r_i)\right) = r_k \qquad r_i > r_{(i-1)} \ \ \forall i \in \{2,n\}$$

By analogy one can derive that the possibilistic Unreliability $U_G^{k/n}$ of a k-out-of-n system based on a set A of components whose possibilistic unreliabilities form a strictly descending ordered list is

$$U_G^{k/n} = \max_{E\in\mathscr{P}(A):|E|=(n-k+1)}\left(\min_{i\in E}(u_i)\right) = u_{(n-k+1)} \qquad r_i > r_{(i-1)} \ \ \forall i \in \{2,n\}$$

which implies that for a strictly ascending ordered list $(u_1, u_2, \ldots, u_n)$ is

$$U_G^{k/n} = \max_{E\in\mathscr{P}(A):|E|=(n-k+1)}\left(\min_{i\in E}(u_i)\right) = u_k \qquad u_{(i-1)} < u_i \ \ \forall i \in \{2,n\}$$

So if a specific list of components represents a strictly descending order of the individual components' possibilistic reliabilities and at the same time a strictly ascending order of the possibilistic unreliabilities, then the resulting k-out-of-n system has possibilistic reliability and unreliability equal to those of the k-th component. Following the above approach, arbitrarily complex fault trees consisting in OR and AND gates – whose details depends on the system/process structure – can be used to propagate the uncertainty through any specified system/process, so as to obtain the possibility of system/process failure events.

5.5.4.3 Risk Assessment

The propagation of possibilistic reliability from the components to the system allows us to obtain the possibility $\pi_G(G=0)$ of the adverse event $(G=0)$ (failure) at system/process level. However, to compute risk one needs also to assess impact.

Impact Assessment The impact of the adverse events can be estimated by different approaches: we will illustrate the approach, based on Value of Information analysis, used by our methodology in section 5.6.
It is important to remark that also the VoI about the impact is suitable to a possibilistic representation. Typically the output of impact assessment is a fuzzy possibility distribution π_I over the economical value of the incurred damage rather than a single crisp value. The shape of the distribution is obtained by propagation of uncertainty from the input parameters of the impact to the overall impact through fuzzy algebraic operations.
In general, the propagation problem can be stated as follows. Let $I : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function of m variables x_i $(\mathbf{x} = (x_1, ..., x_m))$, possibly represented as an array of fuzzy sets $(\pi_{x_1}(s_1), \ldots, \pi_{x_m}(s_m)$ carrying the semantics of possibility distributions. The problem consists in carrying over to $I(\mathbf{x})$ the uncertainty attached to the variables so as to obtain I's possibility distribution associated to a specific adverse event $(G = 0)$, the distribution $\pi_I(I = y) = \pi_I(y)$, with the least possible loss of initial information. The function $I((x_1, ..., x_m))$ reflects the structure of the impact computation and is specific to the system/process instance.

Risk computation Once the knowledge about the occurrence of the adverse event A is represented in terms of the possibility distribution π_G and the knowledge about its impact is represented in terms of the possibility distribution π_I a possibility distribution π_R of the risk R can be obtained by the product-convolution of the two distributions, corresponding to the product of the two variables.

$$\pi_R = \pi_G \otimes \pi_I$$

The output of this operation is a possibility distribution $\pi_R(R = r)$ over the admissible economic values of risk (typically $(0 \leq r < +\infty)$). Based on this distribution, one can make further computations to support the decision process. Given $\pi_R(r)$, one can choose different approaches to support the decision. For instance if the aim of the risk analysis is to assess whether the risk exceeds some predefined threshold t, one can define the event F as $r > t$ and compute possibility $\Pi(F)$ and necessity $N(F)$ of the event F according to the equations (5.1) and (5.2). A decision can then be based, for instance, on the synthesis value given by their average, the credibility $Cr(F)$ defined in equation (5.3). Another approach could base the decisions on the computation of a different synthesis value, defined by the expected value of the possibility distribution, e.g. as defined in equation (5.4). A technique for obtaining synthesis values suitable to risk assessment consists in taking account the subjective risk aversion of the decision maker by adopting the so called Hurwicz criterion, generalized to possibility distributions as proposed by Gyuonnet at al. [54]: if the ill-known utility value of a decision lies in the interval $[a, b]$, let w measure the degree of pessimism of the decision maker, then one can consider the actual utility value to be $wa + (1 - w)b$.
Another important remark about the methodology described in this chapter is that it allows comparison among the residual risks of competing risk mitigation strategies. In this case the decision can be taken either based on the ordering of synthesis indicators mentioned above or based on ranking criteria for fuzzy intervals [13] and particularly for possibility distributions as proposed by Dubois and Prade (called the *four grades of dominance* method) [30].

5.6 Methodology for Disclosure-risk Assessment in Cloud Processes

In this section we describe the key elements of our quantitative risk assessment approach, namely, the identification of the adverse events and threats, the estimation of threats' possibility distributions and

of their impacts.

5.6.1 The Model for Information Disclosure Attacks

In this section we will discuss how to estimate the probability/possibility that process players assume behaviors classifiable as disclosure attacks.
We make use of the notion that the dysfunctional behaviour of actors in a business process is most often due to the *unfairness of the redistribution of payoffs* in the process, (e.g. a benefit allocation structure that responds to organization efficiency more than to fairness), to the actor's greed, plus other environmental factors, which can be assessed by eliciting expert opinion. We translate resulting values for each actor computed on the basis of three points in probability or possibility values.

1. The actor's perception of the business process's fairness, modelled as a distance between its fairness value and the actual profit it obtains from the process,
2. The actor's perception of the potential outcome of an attack, modelled as the expected payoff of an attack,
3. Some context factors, such as the actor's role and position within the chain.

Qualitatively these factors express different drivers for an attack: the first one corresponds to the degree of unfairness an actor may be subject to, the second is linked to an actor's greed, the third encompasses the remainder, possibly less important, factors: among those we singled out, as a representative, the degree of replaceability of an actor.
Thus in the upcoming three subsections, we consider the following main drivers/factors for an attack in the following order

1. perceived unfairness by an actor,
2. greed,
3. and contextual factors.

Furthermore, we distinguish between *individual player* attacks and *coalition based* attacks.
We distinguish between attacks by business process actors and attacks by the Service Provider(s).

- *Individual player attacks* refer to the attacks either by an individual process actor or by individual service provider
- *Coalition based attacks* refer to the attacks delivered by groups of colluding players, who put together the information they know. Typically the collusion can take place among process actors. In fact the likelihood that an attack is delivered by a coalition of two or more Cloud Service Providers can be considered negligible.

With regard to the above mentioned drivers/factors, not all factors are relevant to all the kinds of players and types of attacks.

- The perceived unfairness factor applies to process actors who, for their contribution, receive a compensation from the organisation for which they work. Typically for business process actors this driver is relevant. On the contrary, typically, such driver is not relevant for the cloud provider.
- The greed factor applies to all players.

- Contextual factors apply to all players.

Hereafter the drivers will be modelled along the following lines.

- The perceived unfairness driver will be modelled quantitatively by an index which has the role of "unsatisfaction index" related to a player or to a coalition of players, and is defined on the basis of the Shapley Value.
- The driver represented by greed can be modelled within decision theory by representing the attacker as a rational agent with a specific greed related utility function the option of whether to deliver or not an attack corresponds to the decision of whether to enter or not in a bet, structured around possible benefits, possible drawbacks and their likelihoods.
- The combined effect of the two above factors can be modelled by a suitable aggregator function.
- Contextual factors are modelled as fixed parameters of the aggregator functions.

5.6.1.1 First Driver: Process Unfairness Assessment based on Shapley Value

As stated above, in our methodology we derive the probability or the possibility of dysfunctional behaviour by an actor or by a group of actors taking part to a business process on the basis of the perceived unfairness at the individual level and at coalition level. This can be done by computing an index ϕ of unfairness in the process resource allocation *for each relevant player* and *for each relevant coalition*, then one associates to each actor and subset of actors the probability or the possibility of the adverse event (an individual player leaking her information or the members of a subset putting together the information they hold, so as to obtain together knowledge that they were not supposed to access).
The problem of how profits of a coalition should be redistributed is a well-known one. A principled solution can be obtained by attributing to each actor an amount corresponding to the actor's *Shapley Value* [4]. Indeed, given a coalition, the contributions of the actors to the process, and the value of the surplus value produced by the process, the Shapley value yields a unique ideal allocation of that value fulfilling some largely accepted requirements. With N actors, this solution can be visualized as a point on the hyperplane of the feasible allocations in an N-dimensional space. This computation can be applied to the actors of an organisation to find the *fairness point* and compare it to the point representing the current allocation of the value in the organisation: the distance between these two points provides for each actor an estimate of its individual dissatisfaction. The farther the two points, the more likely there will be dysfunctional behaviour on the part of that actor. Elicitation of expert opinion can transform this information, into a possibility distribution for the defection of the individual actor (for the use of a similar method within a probabilistic approach see [40, 2, 67, 24, 3, 16, 25]).

Hereafter, we first provide a formal definition of the Shapley Value and of the computation of the dissatisfaction parameter, then illustrate the use of this information in the determination of the actors' unreliability. The Shapley Value is an allocation respecting some generally accepted criterion of fairness. The idea behind it is that each party taking part to a process should be given a payoff equal to the average of the contribution that she could make considering each of the possible coalitions underlying the process.
In coalitional Game Theory it is customary to call *security level* of a coalition C the quantity $\mu(C)$ expressing the total surplus that its members can achieve on their own even if the non-members took the action that was the worst from C's perspective. This function is, formally, a set measure: let us consider a general game with a set $\mathscr{N}$ of N participants; any subset of players in $\mathscr{N}$ is a potential

coalition C (combinatorially, there are $(2^N - 1)$ possible non-empty coalitions altogether); the concept of security level allows to define a function from the power set of $\mathcal{N}$ onto the scalar field of real numbers, $\mu : 2^{\mathcal{N}} \to \mathbb{R}$.
This, in turn, allows to define the added value of an actor to a coalition as

$$\Delta_i(C) \equiv \Delta(C, \{i\}) \equiv (\mu(\{C \cup i\}) - \mu(C)) \tag{5.6}$$

i.e. as the *marginal contribution of the actor* computed as the difference of two security levels.
A fair distribution of payoffs, according to the Shapley concept, is one that attributes to each actor a share proportional to the average of its marginal contributions. Formally: the Shapley Value is defined as an allocation of payoffs, a payoff v_i for each actor $i \in \mathcal{N}$; the Shapley Value $v_i^{Shapley}$ for the player i in the game $(\mathcal{N}, \mu)$ is obtained averaging the $\Delta_i(C)$ over all the coalitions. For computation purposes, the Shapley Value may be equivalently interpreted as follows. Suppose that all the actors are arranged in some order, all orderings being equally likely: each ideal ordering of the actors corresponds to a non-decreasing surplus value achieved by the members. For each ordering one can take note of the added value introduced by the actor i, whose Shapley Value is being computed; the Shapley Value for the actor is then computed as the average over all permutations of its added values.

$$v_i^{Shapley} = \frac{1}{N!} \sum_{\sigma} \Delta_i(\sigma) \tag{5.7}$$

where the index σ runs over all the $N!$ permutations of N objects.
Equation (5.7) can be rewritten in a more computable form by taking into account that, when going all over a permutation, the actors following i (the *trailing* actors) are irrelevant to the computation of that actor's added value, and that, to the same computation, it is irrelevant the order, in which the actors preceding i (the leading actors) are ordered, provided that the composition of the set of those actors is the same. Denoting by $|C|$ the cardinality of the coalition C, the above defined v_i can be computed as follows:

$$v_i^{Shapley} = \sum_{C \subseteq \{\mathcal{N} \setminus i\}} \left[\frac{N!}{(|C|)!(N - |C| - 1)!} \right]^{-1} \cdot \Delta_i(C)$$

To highlight its possible interpretation as a weighted average of the marginal contributions $\Delta_i(C)$ of the element i alone in all combinations one can rewrite the index as follows

$$v_i^{Shapley} = \frac{1}{N!} \sum_{c=0}^{N-1} \frac{1}{\binom{N-1}{c}} \sum_{C \subseteq \{\mathcal{N} \setminus i\}:\ |C|=c} \Delta_i(C) = \frac{1}{N!} \sum_{C \subseteq \{\mathcal{N} \setminus i\}} \left[\frac{(N-1)!}{(|C|)!(N - |C| - 1)!} \right]^{-1} \cdot \Delta_i(C) \tag{5.8}$$

The coefficient $1/\binom{N-1}{c}$ in the last expression can be interpreted as the probability of a coalition of size c, i.e. for any coalition C is the probability that it corresponds precisely to the set of players preceding player i within a given ordering.

Non-additive measures (or capacities) and Choquet integral To the benefit of the following discussion, we recall that the above defined Shapley values can be interpreted as averages of the *Choquet integral* of *non-additive* measures (also called *capacities*). To clarify our use of this technique, we recall that any measures are additive (e.g. length, areas, volumes or probabilities), that is, the measure of the union of two disjoint sets is the sum of these two measures: $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A \cap B = \emptyset$. The standard way of aggregating *additive measures* is the weighted sum (linear superposition), which can be viewed as a *discrete integral*: those measures and integral reflect models whose

input components do not "interact" in the determination of the output. A more accurate representation of many situations may need the use of *non-additive measures*. For example the values emerging from a cooperative game can be seen as non-additive measures on sets. The *Choquet integral*, whose formal definition is given below, is the integral of a real function with respect to a non-additive measure, by analogy with the Lebesgue integral (which is defined with respect to an ordinary, i.e., additive, measure). As the integral of a function, in a sense, represents its average value, the discrete Choquet integral can be viewed as a mean or an averaging aggregation operator. The Choquet integral generalises the weighted arithmetic mean by substituting the weights with suitable set functions. The Shapley definition is an example of such an integral. Choquet capacities contribute to the foundations of several fields: in cooperative game theory, and multi-criteria decision making they are used to model the importance of coalitions in the Dempster-Shafer theory of evidence, and in possibility theory they are generally used to model uncertainty (a discrete Choquet capacity can be interpreted as a generalization of a discrete probability distribution). The use of Choquet integral based aggregation has found application in several fields. Among them is information fusion. For instance they have been used in the aggregation of classifiers when the output of different classifiers can be dependent. In that case, a fuzzy measure for classifiers is used to define a weight on each combination, thus making it possible to model the interaction among them. In service-oriented computing, the Choquet integral has been often used for integrating non-additive measures, e.g. of service replaceability. This wide range of application is due to the fact that in any situation with more than minimal complexity, the effect of the ensemble of parts is different from the sum of the effects of the parts taken individually. It is easy to see that the additivity requirement of classical measures severely limits their applicability, however it is not obvious how to generalize classical measures: typically one needs to lift additivity and, as put forward by Klir "...replace it with an appropriate weaker requirement. It is generally recognized that the highest generalization of classical measures that is meaningful for formalizing uncertainty functions is the one that replaces the additivity requirement with a weaker requirement of monotonicity with respect to the subset-hood ordering. Generalized measures of this kind are called *monotone measures*."

By the term *non-additive measures* one refers by convention to those non-additive measures which replace additivity by the weaker constraint of monotonicity: $\mu(A) \leq \mu(B) \iff A \subseteq B$. This expresses the fact that the measure of a set must not be smaller than the one of any of its subsets. All additive measures are monotonic, thus non-additive monotonic measures generalize the additive ones.
The following inequalities hold for every monotone measure μ. If $A, B, A \cup B \in \mathscr{F}$, where $\mathscr{F}$ is a non-empty family of subsets of $\mathscr{N}$, then $\mu(A \cap B) \leq \min\{\mu(A), \mu(B)\}$ and $\mu(A \cup B) \geq \max\{\mu(A), \mu(B)\}$ (they come respectively from $A \cap B \subseteq A$ and $A \cup B \supseteq A, B$). If, in addition, *for all* $A, B, A \cup B \in \mathscr{F}$ such that $A \cap B = \emptyset$ either the inequality $\mu(A \cap B) \leq \mu(A) + \mu(B)$ or the inequality $\mu(A \cup B) \geq \mu(A) + \mu(B)$ hold, then the monotone measure is called *sub-additive* or *super-additive*.
Non-additive measures allow us to represent interaction between the elements. For example, we might have $\mu(A \cup B) < \mu(A) + \mu(B)$ (negative interaction between A and B), and $\mu(A \cup B) > \mu(A) + \mu(B)$ (positive interaction between A and B). We return on those concepts below.
We note that *non-additive measures* are also known for historical reasons by the term *capacities*, and by the term *fuzzy measures*. The former term, widely used, derives from the fact that this kind of quantity was first defined in the context of potential theory and used to model capacitance [19]; the latter term, in the words by Klir [68] "is somewhat confusing, since no fuzzy sets are involved in the definition of monotone measures. To avoid this confusion, the term *fuzzy measures* should be reserved to measures (additive or nonadditive) that are defined on families of fuzzy sets." This is also the choice we have made in the present discussion.

A non-additive measure is said to be an order 2 Choquet capacity (or 2-monotone Choquet capacity) if for all pairs of subsets $A,B \in \mathscr{N}$ the relation $\mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$ holds (or if the corresponding relation with $\leq$ holds). A non-additive measure is said to be an order k Choquet capacity (or k-monotone Choquet capacity) if for all collections of subsets $C_1, C_2, \ldots C_k \in \mathscr{N}$ the relation $\mu(\cup_{j=1}^{k} A_j) \geq \sum_{K \subseteq \mathscr{N}: K \neq \emptyset} (-1)^{|K|+1} \mu(\cap_{j \in K} C_j)$ holds (or if the corresponding relation with $\leq$ holds). For convenience, monotone measures that are not required to satisfy the above equations are viewed as Choquet capacities of order 1.
There are two main types of integral-like constructs based on this nonadditive setting monotonic: the Sugeno integral (introduced by Sugeno in 1972), which is ordinal, and the Choquet integral. We focus on the second.
The Choquet integral is defined as follows. Let μ be a non-additive monotonic measure, i.e. a capacity, on $\mathscr{N}$.
We now introduce some conventional notation. Denote by σ a permutation on N, denote by $\sigma(i)$ the index in which the permutation maps the i-th element. Each permutation determines two sequences of nested coalitions: we use the trailing coalitions family consisting in the full coalition $\{\sigma(1), \sigma(2), \ldots, \sigma(N)\}$, the same without the first leading elements $\{\sigma(2), \ldots, \sigma(N)\}$ and so on; given a reference index i, denote by $C^{Tr}_{\sigma(i)}$ the trailing coalition $C^{Tr}_{\sigma(i)} = \{\sigma(i), \sigma(i+1), \ldots, \sigma(N)\}$ and set $C^{Tr}_{N+1} = \emptyset$. Using this notation, the added value of an element to a (trailing) coalition can be rewritten, indeed

$$\mu(i \cup C^{Tr}_{\sigma(i+1)}) - \mu(C^{Tr}_{\sigma(i+1)}) = \mu(C^{Tr}_{\sigma(i)}) - \mu(C^{Tr}_{\sigma(i+1)})$$

Consider now a scalar function of the indexes, $\mathbf{f} = \{f_1, f_2, \ldots, f_N\} \in [0,1]^N$, which have the meaning of the scores for the elements (players, criteria etc...): this represents the integrand function, it will be integrated using the non-linear measure μ.
Let σ be a permutation of the elements, determined by the ordering of the scores and let σ be chosen so that $f_{\sigma(1)} \leq f_{\sigma(2)} \leq \cdots \leq f_{\sigma(N)}$, (there can be more than one such permutations).
The Choquet integral of the function $\mathbf{f}$ with respect to the capacity μ is the set function $\mathscr{C}_\mu : [0,1]^n \to [0,1]$

$$\mathscr{C}_\mu(\mathbf{f}) \equiv \sum_{i=1}^{N} f_{\sigma(i)} \left[\mu(C^{Tr}_{\sigma(i)}) \; - \; \mu(C^{Tr}_{\sigma(i+1)}) \right] \tag{5.9}$$

An equivalent form is

$$\mathscr{C}_\mu(\mathbf{f}) \equiv \sum_{i=1}^{N} \left[f_{\sigma(i)} - f_{\sigma(i-1)} \right] \; \mu(C^{Tr}_{\sigma(i)}) \tag{5.10}$$

with $C^{Tr}_{\sigma(N+1)} = \emptyset$. In the latter expression one can recognize the form of the Lebesgue integral: the only difference is that the used measure is non-additive.
It can be seen by comparing the expressions that the Shapley Value corresponds to the Choquet weight of the element $i \in N$, however not determined from a single (integrand ordering) permutation, but averaged over all the permutations (monotonicity "cones" of the measure).
The interrelated devices of Choquet integral and Shapley Value respond to different goals. In both, the measure is given as input and is problem-specific. In the Choquet integral also the function values/scores are given as input and are problem specific: one can say that there the interest is in computing the effect of the non-additive measure over the function, or the other way round. In the case of the Shapley value computation, function values/scores are not problem specific, but come directly from the definition of the quantity, they consist in combinatorial coefficients: here one can say that the interest is in computing for at least one player, the average of the non-additive measure weighted by the combinatorial coefficients, the output quantity is interpreted as a power index associated to the player.

Higher Order Shapley Values Returning to the Shapley Value defined by the equations above, one can notice that it is based on the difference $(\mu(\{C\cup i\})-\mu(C))$ (the added value, or marginal contribution, by an actor), which is a *first-order* quantity in the number of the elements deviation from C. This quantity has the formal properties of a derivative and it is called *derivative of the capacity* μ *with respect to the element* i at the "*point*" C. Higher order quantities in the number of elements can be built as well, i.e. one can define derivatives of a capacity μ w.r.t. a subset $\{i,j\}$ at "point" C: if $C\subseteq\mathcal{N}\setminus\{i,j\}$, one can define

$$\Delta_{ij}(C)\equiv\Delta(C,\{i,j\})\equiv\mu\left(\ \{C\cup\{i,j\}\}\ \right)-\mu\left(\ \{C\cup\{i\}\}\ \right)-\mu\left(\ \{C\cup\{j\}\}\ \right)+\mu(C) \quad (5.11)$$

This quantity takes into account the correlation between the contributions of different elements. Indeed, intuitively, a strictly positive correlation between i and j corresponds to the following inequality $\mu(\{i,j\})<\mu(i)+\mu(j)$: if i and j are positively correlated, then the marginal contribution of j to any coalition containing i is strictly less than the marginal contribution of j when i is excluded

$$\mu(\ C\cup\{i,j\}\ )-\mu(\ C\cup\{i\}\ )\ <\ \mu(\ C\cup\{j\}\ )-\mu(C)$$

so a positive correlation between contributions results in negative marginal contribution of the pair as defined by equation (5.11). In short, a negative marginal contribution of the pair arises from the *substitutability* or *redundance* between i and j.
On the other hand, a negative correlation between i and j corresponds to the inequality $\mu(\{i,j\})>\mu(i)+\mu(j)$: if i and j are negatively correlated, then the marginal contribution of j to any coalition containing i is strictly greater than the marginal contribution of j when i is excluded

$$\mu(\ C\cup\{i,j\}\ )-\mu(\ C\cup\{i\}\ )\ >\ \mu(\ C\cup\{j\}\ )-\mu(C)$$

so, a negative correlation between contributions results into a positive marginal contribution of the pair as defined by expression (5.11). In short, a positive marginal contribution of the pair arises from the *complementarity* between i and j.
When i and j are neither positively nor negatively correlated one speaks of independence. Due to the desirable properties of the second-order marginal contribution one can define a higher order Shapley value: suitably weighted averages of second-order differences (also called interaction values in [76]) result in the *second-order Shapley value* (also called *Shapley interaction index*):

$$v_{ij}^{Shapley}=\sum_{C\subseteq\{\mathcal{N}\setminus\{i,j\}\}}\left[\frac{(N-1)!}{(|C|)!(N-|C|-2)!}\right]^{-1}\cdot\Delta_{ij}(C)$$

Following [76] one can say that i and j show a negative interaction if $v_{ij}^{Shapley}<0$, a positive interaction if $v_{ij}^{Shapley}>0$ and that they do not interact if $v_{ij}^{Shapley}=0$.
One can generalize also to a set S containing more than two elements i and j: the *derivative of a capacity* μ *w.r.t. a subset* B *at point* C becomes

$$\Delta_S(C)\equiv\sum_{B\subseteq S}(-1)^{|S\setminus B|}\mu(C\cup B)\qquad\forall B\subseteq N,\ \forall C\subseteq N\setminus B,$$

and the *Shapley interaction index* of S (with $|S|\geq 2$) related to the measure μ is

$$v_S^{Shapley}=\sum_{C\subseteq\{\mathcal{N}\setminus S\}}\frac{(N-|C|-|S|)!}{(N-|S|+1)!}\Delta_S(C). \quad (5.12)$$

All the above expressions can be simplified by using the Möbious transform $m_\mu(\cdot)$ of a set measure $\mu(\cdot)$, defined implicitly by

$$\mu(A) = \sum_{B \subseteq A} m(B)$$

and explicitly by

$$m_\mu(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B) \qquad with \quad S \subseteq N.$$

In terms of $m_\mu(\cdot)$ the general equation (5.12) for interactions can be rewritten [76]:

$$v_S^{Shapley} = \sum_{T \supseteq S} \frac{m_\mu(T)}{t-s+1} = \sum_{C \subseteq \{N \setminus S\}} \frac{m_\mu(C \cup S)}{c+1},$$

with $t = |T|$ and $s = |S|$ and $c = |C| = t - s$ so that the (single-element) Shapley value corresponds to

$$v_i^{Shapley} = \sum_{C \subseteq \{N \setminus i\}} \frac{m_\mu(C \cup i)}{c+1} \qquad i \in N \tag{5.13}$$

while the (two-elements) interaction Shapley value is

$$v_{ij}^{Shapley} = \sum_{C \subseteq \{N \setminus \{i,j\}\}} \frac{m_\mu(C \cup \{i,j\})}{c+1} \qquad i,j \in N. \tag{5.14}$$

A drawback of the Choquet integral we already mentioned for the Shapley value is its exponential computational complexity (indeed if we consider a finite space of cardinality n, only $(n-1)$ values are needed in order to completely determine a probability, while $2^n - 2$ coefficients are needed to define a non-additive measure on the same referential). Several particular sub-families of capacities have been introduced to cope with this issue: in these subfamilies some extra constraints are added in order to decrease the number of coefficients but, at the same time, keep the modelling capabilities of the measures in the subfamily as rich as possible.
Among such subfamilies are the k-additive capacities by Grabisch [50], the p-symmetric capacities by Miranda et al. [79], the k-tolerant and k-intolerant capacities by Marichal [77]. It is worth mentioning the peculiar case of non-additive measures directly inspired by probability theory and represented by the λ-fuzzy measures, introduced by Sugeno. They are based on measures for disjoint sets: $\forall A, B \in 2^{\mathcal{N}}$ such that $A \cap B = \emptyset$ λ-fuzzy measures satisfy $\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B)$. There, the parameter λ corresponds to the intensity of the interaction: $\lambda = 0$ represents to the case where A and B are independent (i.e. to the additive case) if $\lambda < 0$ there exist a substitutive effect, if $\lambda > 0$ a multiplicative effect.
Of the cited subfamilies, frequently used in applications are the k-additive capacities. A capacity μ is said to be k-additive if its Möbius transform vanishes for subsets of more than k elements, i.e. $\forall A \subseteq N$ with $|A| > k, m_\mu(A) = 0$, and there is at least one subset A such that $|A| = k$ and $m(A) \neq 0$. Clearly 1-additive capacities coincide with additive capacities: a capacity is additive if and only if its Möbius transform is non zero only for singletons. The simplest non-linear capacity model is represented by 2-additive capacities: a capacity μ on the set N is said to be 2-additive if its Möbius transform satisfies $m_\mu(T) = 0$ for all $T \subseteq N$ with $t > 2$ and there exist at least one coalition $T \subseteq N$ with $t = 2$ such that $m_\mu(T) \neq 0$. In the 2-additive case the expression for the capacity can be expanded in

$$\mu(T) = \sum_{i \in T} m_\mu(i) + \sum_{\{i,j\} \subseteq T} m_\mu(ij) \qquad T \subseteq N$$

and the Shapley value (5.13) takes the following simple form

$$v_i^{Shapley} = m_\mu(i) + \frac{1}{2}\sum_{j\in\{N\setminus i\}} m_\mu(ij) \qquad ij \in N.$$

A completely different approach, with respect to the computational complexity, is based on *numerical* or *Monte Carlo* approaches. Since the publication of the work by Shapley it was remarked that the complexity of the direct computation of the Shapley value *for all* the elements of a set of actors is factorial in the number of actors: among the simplest approximate methods is the one proposed by Mann and Shapley [75] and based on randomly sampling the space of permutations. A review of the various methods used to compute exactly or approximate the Shapley Value can be found in [39, 5, 74].

From the Shapley value to the insatisfaction index Now let $v_i^{factual}$ be the actual resource allocation for an actor. If the difference between this quantity and the actor's Shapley Value is positive, the actor is under-rewarded for her contribution: this situation may feed its propensity toward a defection; if, instead, this difference is negative, the actor is over-rewarded and the discrepancy will not contribute to its propensity towards defection. One needs also to relate the discrepancy, to the absolute value of $v_i^{Shapley}$. For all the above considerations, the dissatisfaction parameter ϕ_i for an actor i can be defined as follows:

$$\phi_i \equiv \theta\left(v_i^{Shapley} - v_i^{factual}\right)/v_i^{Shapley}, \tag{5.15}$$

where $\theta(z)$ is such that $\theta(z) \equiv 0$ if $z < 0$ and as $\theta(z) \equiv z$ otherwise.
Similarly, based on two-elements interaction Shapley value, one can define the dissatisfaction parameter for a coalition of two actors

$$\phi_{ij} \equiv \theta\left(v_{ij}^{Shapley} - (v_i^{factual} + v_j^{factual})\right)/v_{ij}^{Shapley} \tag{5.16}$$

and so on for a coalition S of actors

$$\phi_S \equiv \theta\left(v_S^{Shapley} - \sum_{i\in S} v_i^{factual}\right)/v_S^{Shapley} \tag{5.17}$$

5.6.1.2 Second Driver: Expected Percentage Gain from an Attack

The expected gain from an attack is defined as the difference between the payoff obtained from a unilateral attack (for instance consisting in selling in un illegal market the information acquired) and the payoff to which the actor is entitled based on a business contract. The degree to which this quantity influences the probability of attack quantifies the greed of an actor.
In greed motivated attacks the assumption is that the probability of an attack is proportional to the expected value of the economical gain from the attack (in which the possibility of being identified as the source of the attack and undergoing the corresponding negative consequences is also taken into account).
The computation is based on expert knowledge and follows the decision theory paradigm of the problem consisting in choosing whether to enter or not in a bet:

- Leaking information = entering the bet,
- Not leaking information = not entering the bet.

The adoption of the bet scheme follows from the classical approach used in the Economics of Crime (see for instance [36]) which assumes the at least approximate rationality of the potential offender. Becker [8] (as put forward by [36]) "*argues that criminals are like anyone else, and assumes that an individual behaves as if he is a rational utility maximizer. As the total outcome of a criminal act is uncertain, Becker employs the usual assumption that people act as if they were maximizing expected utility, and also that utility is a positive function of income.*" The individual's expected utility, or gain g, for committing an offense is

$$g = p^{(-)}\,u(v^{illegal} - d) + (1 - p^{(-)})u(v^{illegal}),$$

where u is the individual von Neumann-Morgenstern utility function, $p^{(-)}$ is the subjective probability of being caught and convicted, Y is the monetary plus psychic income (i.e. the monetary equivalent) from an offense, and d is the monetary equivalent of the damage deriving from the inflicted punishment. The individual will commit the offense if the expected utility is positive, and he will not if it is negative. In this view the variance of the bet is considered negligible and the agents risk-neutral.
If the utility function is approximately linear, one can write

$$g \;\simeq\; u(v^{illegal}) \;-\; p^{(-)}u(d).$$

In the present scenario $v^{illegal}$ is the monetary value of the leaked information (shadow) market, while d represents the monetary value of the punishment, typically in terms of business loss and fines to pay if caught; $p^{(-)}$ is the probability of being caught; the utility function can be elicited based on expert opinion.

Point of view of the defendant Notice that the potential attacker has the possibility to probe the (underground) market of industrial information and to know with precision what is the worth of the information she owns. From the point of view of the defendant the reasoning is the same, however the information available about the parameters is typically poorer. The point of view of the defendant is the most relevant one for the methodology, since it is the one taken by the decision makers, deciding at design time the choreography of the process and the kind of measures to be taken.
The defendant does not know the exact worth $v^{illegal}$ of the information on the shadow market, nor the exact worth of the damage d which could be suffered by the attacker: in place of a rather precise estimate of the above expected value, the defender can obtain from experts distributions of such parameters and eventually the probability distribution over possible values of the attacker expected gain g. So while for the attacker g is a number, for the defendant it is an ill-known variable, and typically is treated as a random variable.
Let us indicate the event of an attack from the attacker by A. Since a potential attacker attacks when her valuation of the expected gain is positive, from the point of view of the defendant, the probability $P(A)$ of an attack equals $Prob(g > 0)$. This probability enters the process of decision by the defendant and we are going to return shortly on its estimate from the stand point of the defendant.
The discussion on this driver for an attack is developed in the next chapter, contextually with a numerical example.

5.6.1.3 Context Factors: the Example of Role Cardinality

Several context factors (e.g., culture-related) can influence or reinforce individual propensity towards an attack. A complete analysis of these factors is beyond the scope of the present document; however, some problems and preliminary solutions related to aggregating multiple context factors to obtain a single environmental factor are presented below.

For the sake of clarity, let us underline that here we focus on the merely topological aspect of *multi-sourcing*. Namely, we argue that the higher the number of actors in a role of the business process which can provide a service, the less likely the actors of that layer will attack. We capture this aspect by defining a metric χ as the reciprocal of the number of the actors in a role of the supply chain. The value χ_i for an actor i, by definition, belongs to the interval, where the value 1 corresponds to the case where the actor is the only one who is able to provide a given service.
This represents a rough measure of actor centrality, and it complements the ϕ_i based on the Shapley Value (which is defined as a percentage). Notice furthermore that sometimes, due to the unavailability of some parameter of the process, ϕ_i might not be computable for all the individual actors: in those cases, however the cardinality of a role is likely to be still known.

5.6.1.4 Putting All Factors Together

Let us now focus on applying the panoplia of techniques described so far to the evaluation of the likelihood of a disclosure attack. We derive its a priori probability based on expert evaluation of the three above identified risk factors. This is an instance of the well-known problem of eliciting experts' knowledge in order to obtain probability distributions for an event, given the value of different context parameters. In probabilistic settings, this type of analysis can result in a multivariate possibility distribution over a number of parameters, or, alternatively in an approximation which factors a multivariate density into the product of uni-variate densities: this happens because in some setting it is important to take into account and to model the correlation between variables, while in other settings either the correlations are irrelevant or their effect can be approximated by using suitable components in the factorisation. The setting in which our risk assessment problem is set corresponds better to the second type of situation, as the following considerations explain.
Our problem consists in finding the dependence of the likelihood distribution from the three joint variables, i.e. a function

$$\pi(\phi, g, m) \tag{5.18}$$

mapping each triple of values in a value of an a-priori likelihood/possibility of an attack, where ϕ is the percentage deviation from the Shapley value, g the expected gain from an unilateral attack, and m the reciprocal of the number of direct competitors. Even without the simplifying assumption of the irrelevance of the correlations, the actual dimensionality of the problem is lower than the above expression suggests. Indeed, the topology of a business is by definition known and for some time it remains unchanged: therefore one can consider as fixed the number of competitors $1/m$ and can mine experts knowledge in order to evaluate the bivariate density

$$\pi^{(m)}(\phi, g) \tag{5.19}$$

Furthermore, in our case we can make the assumption that the variables ϕ and g should be independent, i.e. contribute independently to the determination of the possibility of an attack, i.e.

$$\pi^{(m)}(\phi, g) \simeq \pi_{\phi}^{(m)}(\phi) \otimes \pi_{g}^{(m)}(g). \tag{5.20}$$

The rationale behind this choice is the fact that the two quantities quantify two "psychologically" distinct behavioural driving forces: s reflects the feeling of being treated unfairly and the probability of seeking a remedy if attacked, whereas g reflects the availability of possible increased pay-offs, i.e., the consequent temptation and the probability of acting on one's greed and hence to behave in a malevolent manner.
In short, the elicitation of the expert opinion over the likelihood as a function of the three above identified parameters, can be reduced to the elicitation of expert opinion over univariate possibility distributions.

5.6.1.5 Elicitation of Expert Opinions for the ϕ Possibility Distribution.

In this section, we discuss the elicitation of the expert opinion for the possibility distribution of ϕ. This procedure contains as a special case the elicitation of the probability of defection as a function of ϕ discussed above. Following the conventions used so far, we adopt the following notation: x_i indicates the ill known quantity representing the state of the individual actor, the actor can be *defecting*, i.e. contribute to data disclosures ($x_i = 0$) or *non-defecting*, i.e. behaving correctly with respect to data disclosures ($x_i = 1$); the possibility distribution for the individual actor, given its ϕ (in the context tagged by m), is denoted

$$\pi_{x_i}^{(m)}(\cdot|\phi_i) : s \in \{0,1\} \rightarrow \pi_{x_i}^{(m)}(s|\phi_i) \in [0,1].$$

Based on this, one can define, in correspondence to the two values of s, two distinct functions of ϕ: the *possibilistic reliability function* as a function of ϕ, i.e. $r_i(\phi) \equiv \pi_{x_i}(s = 1|\phi)$, and the *possibilistic unreliability function*, as a function of ϕ, i.e. $u_i(\phi) \equiv \pi_{x_i}(s = 0|\phi)$.
The possibilistic reliability $r(\phi)$ and unreliability $u(\phi)$ can be obtained by the elicitation of expert opinion and the application of obvious constraints – namely $r(\phi = 0) = 1$ and $u(\phi = 0) = 0$, the non-increasing character of $r(\phi)$ and non-decreasing character of $u(\phi)$.
We note, in passing, that due to the monotonicity of r and u, already the value of the parameter ϕ alone, without further elicitation of knowledge, can be used to rank the individual actors' reliabilities and unreliabilities. However expressing, in addition, expert knowledge in terms of r and u enables the quantitative assessment of the disclosure risk.
The elicitation of experts' knowledge in order to obtain a function of a continuous variable is a problem which has attracted a lot of attention from the research community. In our methodology we approach it using a technique based on the *Bézier curves*. These curves have often been used in the elicitation of expert opinion (see for instance [17, 92]) also in the form of membership functions [73]: they can be used to approximate a smooth (continuously differentiable) function on a bounded interval up to an arbitrary level of detail by forcing the curve to pass in the vicinity of an arbitrarily high number of control points (in two-dimensional Euclidean space) selected by an expert.
In synthesis, overall, the computation of the possibility of defection for the individual actor involves the following steps:

- Expert opinion is elicited to determine the possibility distributions possibilistic reliability $r(\phi)$ and unreliability $u(\phi)$ as a function of the dissatisfaction parameter ϕ and the possibilistic reliability $r(g)$ and unreliability $u(g)$ as a function of the greed parameter g: Bézier curve based methods are suitable candidates for eliciting expert opinion.

- Given a specific instance of the collaborative process definition, and the actual allocation to the actors of the resources deriving from the process surplus, the numerical value of ϕ_i and of g_i is computed for each actor.

- Based on $r(\phi)$ and $u(\phi)$ and based on $r(g)$ and $u(g)$ and using standard fuzzy aggregators, one computes possibilistic reliability r_i and unreliability u_i for each actor.

Note that the individual level unreliabilities already hint at the weak points of the process: when units of valuable information are fully known by individual process actors, the individual unreliabilities can be used to quantitatively assess the risk of its disclosure. Furthermore, this knowledge can be used to support the many-actor level risk assessment. As it has been shown in section 5.5.4.2, from individual level (un)reliabilities one can obtain many-actors (un)reliabilities based on the structure function of the failure model. The main results can be translated into the possibility of disclosure. For instance given the reliabilities r_i and unreliabilities u_i of n actors:

- (*Series* analogy) if the failure model is such that it is necessary that all the actors do not defect in order for the information to be disclosed, i.e. if it is sufficient that $k = 1$ actor defects for the disclosure to take place, then the possibilistic reliability is $r = \min_i(r_i) = r_n$ and the possibilistic unreliability is $\max_i(u_i) = u_1$.

- (*Parallel* analogy) if the failure model is such that it is *sufficient* that $k = 1$ actor does not defect for the disclosure not to take place, i.e. it is *necessary that all* the actors defect in order for the information to be disclosed, then the possibilistic reliability is $r = \max_i(r_i) = r_1$ and the possibilistic unreliability is $\min_i(u_i) = u_n$.

- (k-out-of-n analogy): in this case the failure model is such that it is sufficient that k (out of n) actors do not defect for the disclosure not to take place, i.e. it is necessary that $(n-k+1)$ actors defect in order for the information to be disclosed; as a consequence, the possibilistic reliability is equal to the k-th largest reliability r_k and the possibilistic unreliability is equal to the k-th smallest unreliability u_k.

5.6.1.6 Decision Theoretical Approach and Elicitation of Expert Opinion for g

The likelihood (under the form of a probability or of a possibility distribution) of an attack due to the greediness factor, as quantified by g, can be estimated using Decision Theory. One can characterize a player as a rational agent confronted with the following choice between two options, i.e. entering or not entering a bet:

- entering the bet corresponds in delivering the attack,

- not entering the bet corresponds to not delivering the attack.

A rational agent takes one choice or the other based on the expected value of the bet, which in turn is determined by the agent utility function and by the likelihoods of the favourable and adverse outcome from the bet. All the mentioned elements are normally not known with precision and their forms and values have to be obtained by eliciting expert opinions. In particular the functional form of the utility function must be obtained from experts, on the one hand, according to our methodology the argument of the function in a specific situation is the value of the greed factor g obtained in the previous sections.

Let us illustrate the concept in the probabilistic setting, where the bet has two possible outcomes, from the point of view of the attacker:

- The adverse outcome brings a net damage d with the complementary probability $q = 1 - p$. Such an event could consist of being identified as the responsible fo the leak and having to pay a fine or being thrown out of business or suffering a severe business loss.

- The favourable outcome is the opposite, it brings a net benefit $b(g)$ (a function of the greed factor g) with likelihood p (in this case a probability).

The balance E, the expected value of the bet is defined by $E \equiv p\,b(g) - q\,d$. The choice of a risk-neutral player will be the following: if $E > 0$ she will enter the bet, otherwise she will not (this simple model can be corrected for non-risk-neutral players by acting on the utility function, however this element is secondary).
The basic utility function $b(g)$ can be obtained from experts by the use of the Bézier curve methodology described above and considering the following constraints: the curve must be monotonically non-decreasing, furthermore $b(0) = 0$. Also the probability of the adverse outcome can be obtained by experts, when it is not zero or negligible.

For the likelihood of the adverse and favourable events. in place of probabilities, which could be not known with precision, one can use expert provided possibilities distributions.

5.6.2 Impact Assessment by Value of Information Analysis

As we have outlined above, the technique we use for estimating impact of information disclosure relies on quantifying the *Value of Information* (VoI) for each knowledge item (or set of items) potentially reconstructed by a subset of colluding process actors. Also, the estimated value of a knowledge item has an intrinsic possibilistic character since it will be known to lie in a range but its precise economical value will depend on several incompletely known factors.

5.6.2.1 Value of Information Analysis

As we have anticipated in the previous section, VoI has been defined as the *analytic framework used to establish the value of acquiring additional information to solve a decision problem*. In the risk management domain, VoI has been successfully used since the sixties in several areas of research including engineering and environmental risk analysis [61]. From a purely rational perspective, it is clear that acquiring extra information is only useful for an actor A if knowing it can significantly modify its behaviour.
Classic VoI analysis typically involves constructing a complex decision-analytic model to fully characterise all information items available to each process actor, the loss each actor would incur, should these items become known to other actors and the costs of interventions that could be executed to prevent them. This comprehensive approach to VoI often turns out to be prohibitively expensive for use in prioritising interventions [60]. As alternatives to full VoI, we identified three approaches to analysing the value of information that are less burdensome:

1. The *conceptual* approach to VoI, where context information is used to provide informative bounds on the value of information without formally quantifying it through modeling. For instance, the VoI of the design information about a device that is already available on the market cannot be higher than the cost of reverse-engineering the device itself;

2. The *minimal* approach to VoI, which is possible when evidence of the net benefit of holding a piece of information, are readily available from existing research. For example, the VoI of the design information about a device that is currently available on the market cannot be higher than the net profit coming from its sales to its current supplier.

3. The *maximal* modeling approach to VoI, where the value of an information item is estimated from previous VoI studies concerning similar information in different contexts. For instance, the VoI of the design information about a solid-state storage device is quantified according to previous VoI studies on disks.

These three low-cost VoI methods can be readily applied in priority-setting of risk-mitigation countermeasure, and raises the question about how the use of VoI to assess disclosure risk in the framework of our methodology.

Value of Total Information and Value of Partial Information Here, we take a process-oriented view of VoI, in order to assess the impact of information disclosure. Let us consider once again a set of actors $A = \{A_1, \ldots, A_n\}$ who take part to a business process P, and the expected benefit for each actor A_k, Ben_{A_k} resulting from the execution of P. The starting point of our VoI analysis of P is to consider the *Value of Total Information* (VoTI), i.e. answering the question "What would be the change to

Ben_{A_i} should A_i know all information (local memory plus messages) held by the other actors of P?". If there is no such change, then obtaining extra information is worthless. If such a change exists, then the impact on A_k of A_i's ($i \neq k$) complete knowledge can be estimated as the corresponding change in the value of Ben_{A_k}.
For the security-aware process designer, our simple VoTI provides a useful upper bound, because it tells the maximum value that any information held by other actors may have for each participant to P. If that value is negligible, or achieving that information would cost more than that, a rational actor will not pursue disclosure any further (i.e., it would not enter agreements for information sharing with other actors).
A different type of check involves looking at the *Value of Partial Information* (VoPI) compared with the *Value of Total Information* (VoTI). For any process participant A_i, getting to know some information beyond the one that is strictly necessary to carry out its part in the process (e.g., the messages exchanged among other actors, or the content of another actor's local memory) may or may not bring a benefit, i.e. a change in Ben_{A_i}. For each subset K of knowledge items used in the process, VoPI focuses on (i) checking whether the benefit of knowing K would match the cost of collecting it and (ii) quantifying the impact of each actor A_i getting to know K on the benefits Ben_{A_k} of the other participants (for $i \neq k$).

5.6.2.2 Possibilistic Value of Information

The fact that the output of the Value of Information analysis is in nature possibilistic and fuzzy has been broadly recognised and accepted in several domains [91, 52, 18, 20, 56]. Indeed, the knowledge of an expert who has performed VoI analysis can be best conveyed by means of a possibility distribution (in this case it will not be a distribution over discrete values but over continuous values). In line with the traditional representation of the risk as the product of the likelihood of the adverse event by the impact of the event, one can represent the possibilistic risk by the (fuzzy) product of the possibility distribution of the event and the possibility distribution of the values that the impact can take.

Chapter 6

Case study: Spare Part Management

In this chapter, we apply our methodology to the one of the processes described in PRACTICE Deliverable D24.1, Chapter 3, namely the Collaborative Planning System for Aircraft Engine Maintenance. Specifically, we focus on the overhaul management process and on the spare part management module presented in Section 3.1.3.1 of that deliverable.

6.1 The Spare Part Management Process

The spare part management module deals with determining the optimal demand of service parts for the Maintenance Repair Overhaul (MRO) to decrease inventory costs and achieve higher service level. Figure 6.1 shows a view over the spare part management module, whose input data is provided by the "decision tree technical based on historic data" module. The decision tree is used in D24.1 to predict the probability that a specific component needs to be replaced. More specifically, in order to predict spare part demand f a probabilistic binary decision tree is learned on an historic database D. The MRO estimate of spare part demand F_{spares} for a certain component is then computed by summing the demand of all airlines $i \in \{1,\ldots,A\}$ over all the entries in their real time databases S_i, denoted as $F_{spares} = \sum_{i=1}^{A} \sum_{x \in S_i} f_D(x)$.

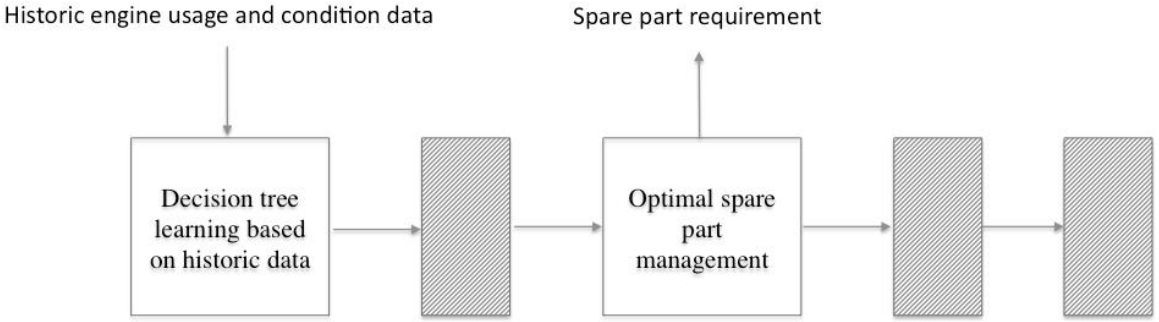

Figure 6.1: View over the overhaul management process showing the optimal spare part management.

The conceptual schema of the spare management sub-process is shown in Figure 6.2.

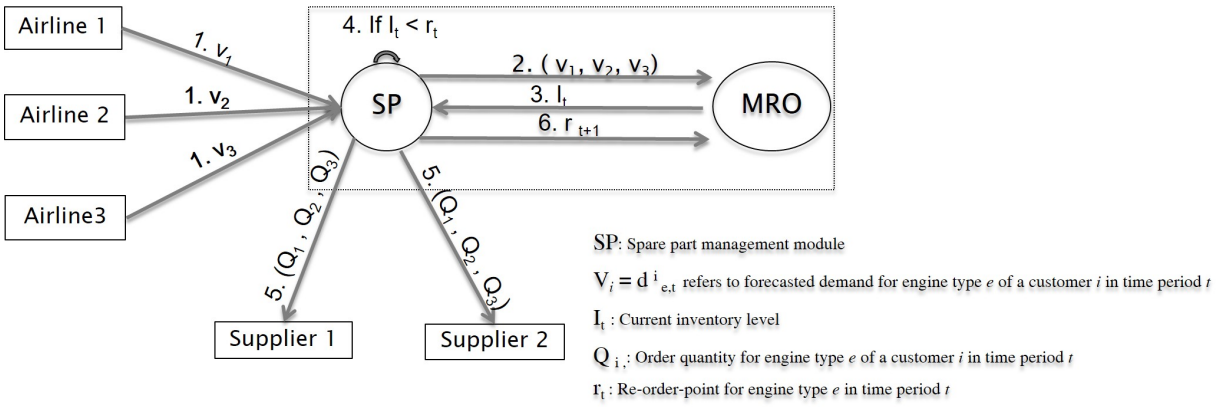

Figure 6.2: Spare management process.

Figure 6.3 shows a simplified event-based representation with 3 airlines and 2 suppliers of the spare part management process. It is a non-secure incarnation of the process where each airline sends its spare part demand in plaintext to the cloud-based collaborative system. In our example, the actor set is $A = (IN_1, IN_2, IN_3, IN_4, IN_5, COMP, RES)$ where actors IN_1, IN_2 and IN_3 are the airlines and actors IN_4 and IN_5 are the suppliers. Actor $COMP$ is the cloud-based service consisting in the spare management module and the rest of the MRO. The spare management module SP computes the optimal re-order point for each component, and actor RES securely sends the estimated overall spare part demand for the component to the MRO.

Each airline i sends to SP the expected demand for each part that needs to be replaced. As we said before, demand is computed using a probabilistic binary decision tree on historical data S_i as $V_i = \sum_{x \in S_i} f_D(x)$. The SP module sends the requested demand for the component to the MRO, which answers with the current inventory level l_t for that component per the period t. Then, the SP compares the received inventory level and re-order point r_t for the component. If the inventory level has fallen below r_t then SP releases the order quantity Q_i for each airline to the suppliers. The suppliers carry out the orders and send the confirmation to the RES node that notifies the results to the SP. At the end, the SP calculates the optimal re-order point r_{t+1} for the subsequent period of the component and sends it to the MRO.

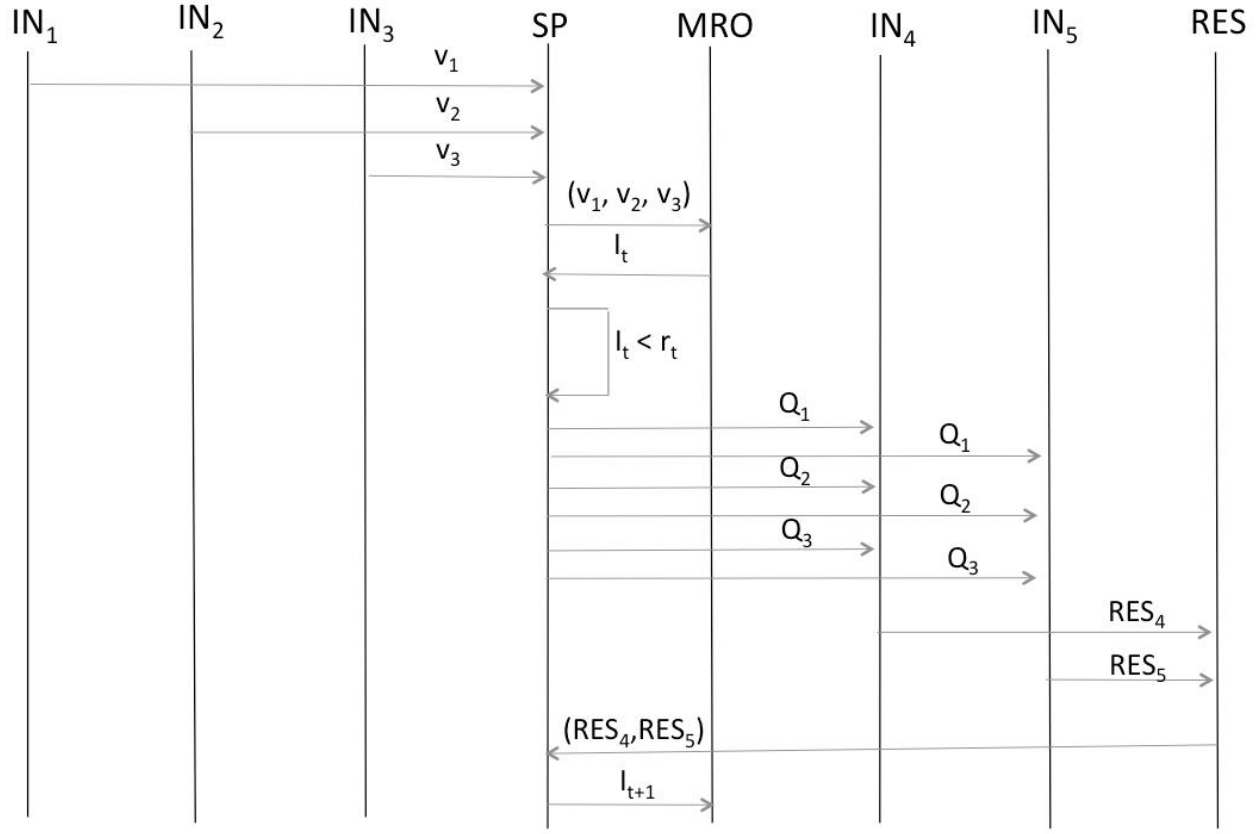

Figure 6.3: Event-based representation of Optimal Spare Part Management with 3 airlines and 2 suppliers.

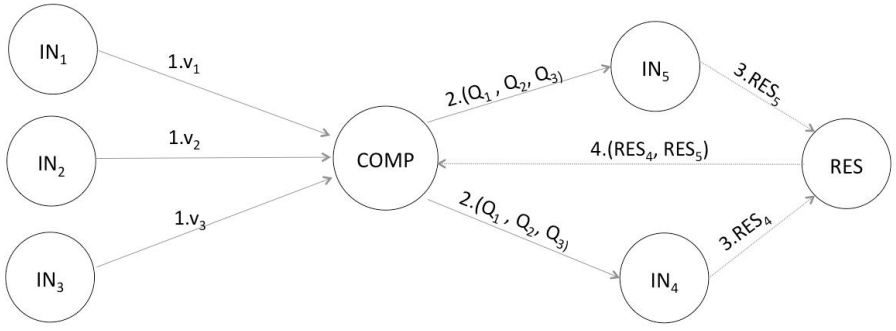

Figure 6.4: Process model representation of optimal spare part management with 3 airlines and 2 suppliers.

Figure 6.4 indicates, using our notation, the process model of the view over spare part management for 3 airlines and 2 suppliers.

6.1.1 The role of the Cloud Provider

The Spare Part Management process can be executed on-premises, but if it is outsourced to a cloud provider, an additional threat must be considered: some of the provider organisational roles may gain access to the local memory of the actors and to the exchanged messages. Our equational formalism (Chapter 5) allows to annotate the process model with equations expressing the degree of transparency

of the process exchanges for the cloud provider roles. In case the cloud is fully transparent, the basic CET equation described in Chapter 5 holds:

$$\cup_i KS(A_i.C) = KS(C) \tag{6.1}$$

Let us now assume that the companies taking part to the *MRO* process (the airlines and the maintenance centers) are willing to assume that their cloud provider, seen as a corporation, is trustworthy (for instance, because it is a highly reputed public company), which corresponds to a defection probability (and possibility) of zero in terms of the computation carried out in this chapter.
Let us also assume that, since misbehaviour on the part of individuals in the cloud provider staff is anyway conceivable, the actors have requested to the trusted cloud provider to enact *domain partitioning* (see Chapter 5), i.e. to assign each sub-process (including Spare Part Management) or even each block of related activities to a different cloud domain (handled by a different administrator with different credentials).
Domain partitioning will obfuscate the cloud transparency of the MRO process for specific organisational roles. In other words, we will be in a situation of *non-uniform transparency*. To represent this situation using our formalism, a custom transparency equation needs to be written for each domain administrator. For instance, assuming that the communication between *SP* and the airlines takes place only on the cloud's *DomA* we can write:

$$\cup_i KS(A_i.C_{DomA}) = KS(C_{DomA}) \tag{6.2}$$

The above equation looks very similar to the basic CET discussed in Chapter 5, but applying it to the *SP* process model gives very different results. Namely, a new "silent" actor corresponding to DomA admin will be added to the actor set[1]. At each exchange of messages between the airlines and the spare part demand planner, *DomA* administrator's knowledge set gets augmented with the content of the messages and of the local memories of the actors who are in its scope (the airlines and the planner).
We remark that, as it happens with all the provider roles (see Chapter 6), the risk of *DomA* admin misbehaving will be quantified using a different probability model and a different VoI assessment than the one of participants. The full provider-related risk analysis will be presented in the next release of this deliverable D31.3.

6.1.2 Representing and Comparing Security Controls

Of course, no airline wants to disclose the portion of its V_i data that is sensitive to its competitors and the suppliers, especially to the ones working on the same routes, which could use the competitors' information to improve their own decision-making. For this reason, a security control can be deployed to computed the estimated overall spare part demand F_{spares} for a certain component using an additive homomorphic encryption scheme, e.g. (Pailier) with encryption algorithm $Enc_{pk}^{Hom}(.)$ using public key pk and decryption algorithm $Dec_{sk}^{Hom}(.)$ using secret key sk.
As shown in Figure 6.5, this deployment of an encryption-based security control brings about a secure incarnation of the spare part management process, where each airline i sends the encrypted form of demand for a each component that needs to be replaced as $V_i' = Enc_{pk}^{Hom}(\sum_{x \in S_i} f_D(x))$ to the SP without access to the secret key. The SP aggregates all ciphers to $V' = \prod_{i=1}^{3}(V_i')$ and sends this aggregation to the MRO who can decrypt it getting $Dec_{sk}^{Hom}(\prod_{i=1}^{3} Enc_{pk}^{Hom}(\sum_{x \in S_i} f_D(x)) = \sum_{i=1}^{3}(\sum_{x \in S_i} f_D(x)) = F_{spares}$. The MRO sends the current inventory level l_t of the component per period t to the SP. Then,

[1]Note that this addition is not necessary for the cloud provider, as the probability of misbehaviour on its part has been forced *a priori* to be zero

the SP compares the received inventory level and re-order point r_t for the component, if the inventory level is fallen below r_t, the SP releases the total order quantity of all the airlines Q for the component to the suppliers. The suppliers provide the orders and send the result to the RES node that notifies the results to the SP. At the end, the SP calculates the optimal re-order point r_{t+1} for the subsequent period of the component and sends it to the MRO. The diagram shown in Figure 6.6 shows the process model according to our notation of secure representation of spare part management for 3 airlines and 2 suppliers.

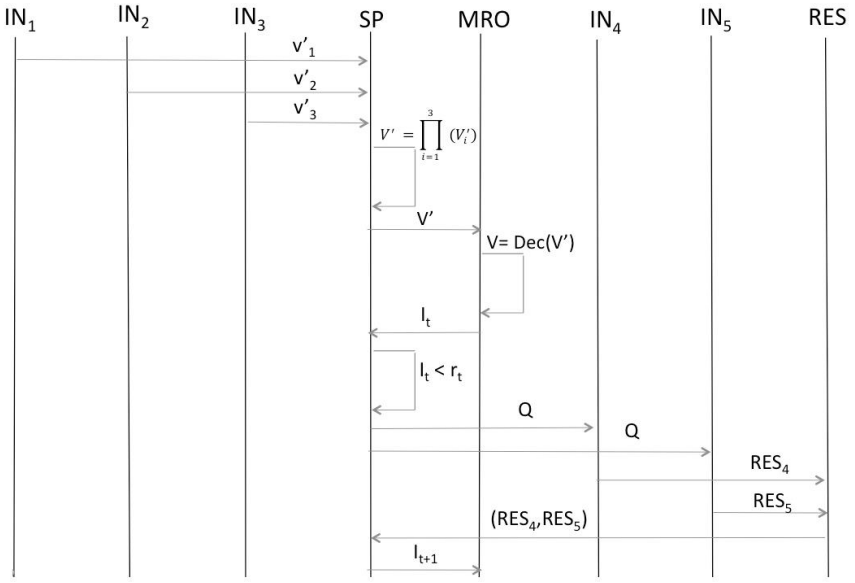

Figure 6.5: Event-based secured representation of optimal spare part management with 3 airlines and 2 suppliers.

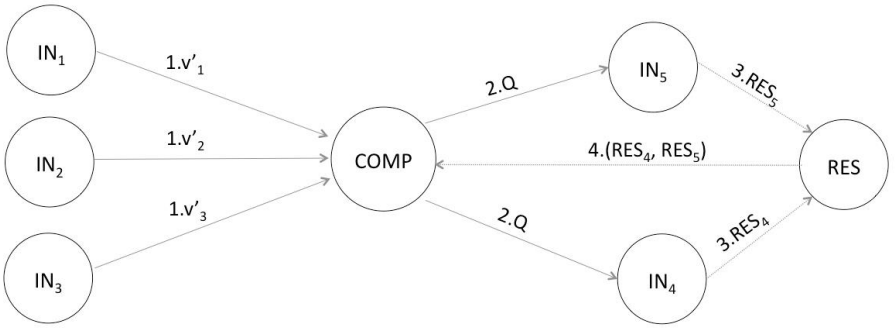

Figure 6.6: Process model secure representation of optimal spare part management with 3 airlines and 2 suppliers using homomorphic encryption communication.

Let us now see what happens when we introduce a different security control, namely a secret sharing one. Figure 6.7 shows the process model of an example (again with 3 airlines and 2 suppliers) where the cloud-based service is performed by two nodes, possibly located on different clouds.
In this way, the incarnation of our business process P is modified introducing a new node $(COMP2)$ and the function that permits to determine the optimal demand of service parts for the MRO.
Each airline i uses the secret sharing control to create two shares of the expected demand for a each component $sh_1(Vi)$ and $sh_2(Vi)$. Due to this control, each computational node knows only partial data and cannot disclose the entire data provided by the airlines.

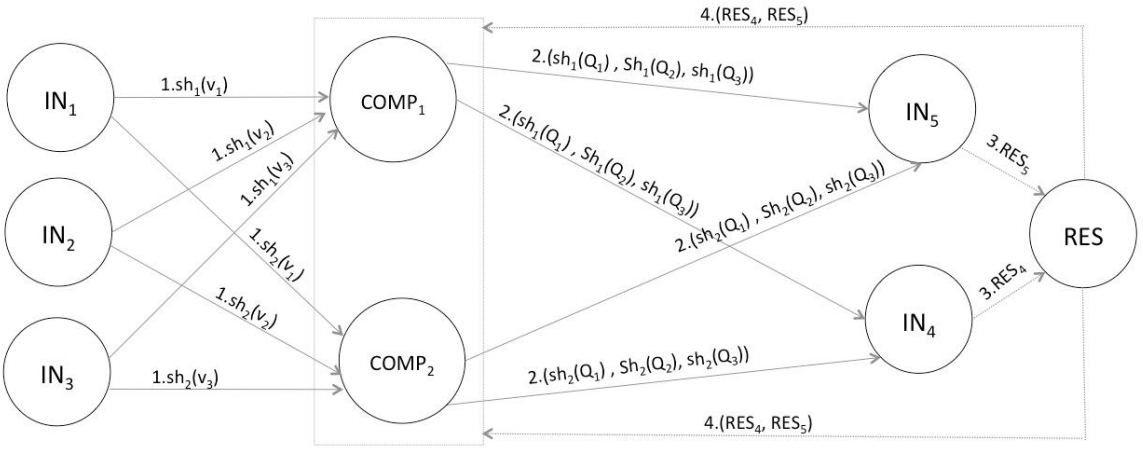

Figure 6.7: Process model secure representation of optimal spare part management with 3 airlines and 2 suppliers using secret sharing algorithm.

6.2 Risk Assessment and Decision Model Example

Here we illustrate the risk computation using stylized examples referred to the above described scenarios. We consider the possibility for the computational nodes and the supplier nodes to deliver an attack if sufficiently motivated by perceived unfairness or by greed. The attack contemplated here consists in the leakage of an airline specific information to competitor airlines. We will consider a scenario where the attack is possible and one where it is prevented by the security countermeasures. The economic value of the risk reduction and the cost of the countermeasures can help a decision maker in the choice of whether or not deploy countermeasures.
In order to illustrate the risk computation we develop some elements of the previous example scenarios using a slightly different notation for the sake of clarity: the differences are the following

- the airline companies IN_1, IN_2 and IN_3 are denoted by C_1, C_2 and C_3 respectively, so as to better distinguish them from the supplier companies
- the supplier companies IN_4, IN_5 are denoted by S_1, S_2
- the three different types of parts are indicated by a, b, c
- We do not mention the order messages v_1, v_2, v_3 from C_1,C_2,C_3, but mention the details of their break-down in terms of the number of parts of kind a, b and c ordered by each airline

We consider two following scenarios, corresponding to the first and third scenarios described above

- The first scenario, the white board scenario, which we call scenario A, corresponds to Figures 6.1 and 6.2:
  here COMP,S_1,S_2 can see the break down of the order among C_1,C_2,C_3 and the break down of the allocation of the spare part provision among S_1 and S_2.
- The second scenario, the secret sharing based, which we call scenario B, corresponds to Figures 6.5 and 6.6:
  here COMP,S_1,S_2 do not see neither the break down of the order, nor the break down of the allocation.

Hereafter we are going to show how to

- Compute the probability of attack due to reaction of unfair distribution of payoffs (Shapley value based method)
  - It applies to S_1 and S_2 and is used in scenario A
  - In scenarios B the information is hidden from S_1 and S_2 and they cannot use it for an attack
- Compute the probability of attack due to "greed"
  - It applies to COMP, S_1 and S_2 in scenario A
  - In scenario B the information is hidden from COMP, S_1 and S_2 and they cannot use it for an attack

Notice that the computation of the probability of attack due to greed applies also to the Cloud Service Provider.

6.2.1 Scenario A: White Board Condition

In this scenario, corresponding to the above Figures 6.6 and 6.7, COMP,S_1, S_2 can see the break down of the order among C_1,C_2,C_3 and the break down of the allocation of the spare part provision among S_1 and S_2.

6.2.1.1 Modeling the Likelihood of the Perceived Unfairness Attack

	a	b	c
Quantity ordered by C_1	20	20	20
Quantity ordered by C_1	30	30	30
Quantity ordered by C_1	50	50	50
Total	100	100	100

Table 6.1: This table shows the break-down of the order: *a*, *b* and *c* are the type of parts ordered by C1, C2 and C3.

Assume now that the production capacity of the suppliers is the one indicated in the following table.

	a	b	c
Unitary profit	2	3	4
Production capacity of S_1	80	80	50
Production capacity of S_2	60	60	70

Table 6.2: This table shows the production capacities of S_1 and S_2 with respect to the individual part types *a*, *b* and *c*.

Based on some optimization criteria (typically considering the most competitive price for each type: this info is not needed by the present discussion) COMP chooses for S_1 and S_2 the following allocation of the work: the table indicates also the factual gain obtained by the two suppliers in the hypothesis that the revenue from *a*, *b* and *c* are respectively 2, 3 and 4 units of cost (the unit of cost, for instance 1000 Euros, needs not to be mentioned here, since all the computations are eventually translated into a relative value).
We assume that COMP has no reason to be considered as a potential attacker motivated by perceived unfairness: its task is performing a critical computation and its earnings are not determined at run

	a	b	c	Total
Quantity from S_1	80	80	50	
Quantity from S_2	20	20	50	
Unitary profit	2	3	4	
Factual gain for S_1	160	240	200	600
Factual gain for S_2	20	20	50	300
Gran Total				900

Table 6.3: This table shows the allocation of the work over of S_1 and S_2 with respect to the individual part types a, b and c and the corresponding revenue (see text).

time, but by a contract established before the start of the process. So we perform the Shapley value based risk computation only for actors S_1 and S_2 (still COMP can be considered as a potential attacker motivated by greed, as we see later).
The Shapley value computation for S_1 and S_2, can be carried on using the table in Figure 6.2.1.1 and using the expression from the previous chapter, which we recall hereafter,

$$v_i^{Shapley} \propto \frac{1}{N!}\sum_s \Delta_i(\sigma^{(s)})$$

with the index s running over the permutations; in each permutation the element i is preceded by a set of elements C w.r.t. one computed the marginal contribution/added value $\Delta_i(C)$ defined by

$$\Delta_i(C) \equiv \Delta(C,\{i\}) \equiv (\mu(\{C \cup i\}) - \mu(C))$$

and C represents a coalition and σ is a permutation.

Permutation		a	b	c		a	b	c	
S_1,S_2		80	80	50		20	20	50	
S_1,S_2		40	40	30		60	60	70	
Total		120	120	80		80	80	120	
Shapley Value		60	60	40		40	40	60	

Table 6.4: Table for the computation of the Shapley value the first three columns refer to the number of items served/contributed by S_1, the last three columns to the number of items served/contributed by S_2. For the sake of simplicity measures are not normalized so the Shapley value is a multiple of the actual Shapley value (see text).

In this simple case there are only two permutations of the elements, namely $\sigma^{(1)} = (S_1,S_2) \equiv (1,2)$ and $\sigma^{(2)} = (S_2,S_1) \equiv (2,1)$. For the sake of simplicity we use non-normalized measures: For illustrating the meaning of the table elements, let us consider the items of type a.

- the total quantity of items of type a ordered by the airline companies is 100 units (see Figure 6.2.1.1)
- the production capacity of S_1 and S_2 w.r.t. a is 80 and 60 units (see Figure 6.2.1.1)

- in the permutation $(1,2)$ player S_1 arrives first and is assigned a number of units able to saturate her capacity, i.e. is assigned 80 units

$$\Delta_1(\sigma^{(1)}) = \Delta_1((1,2)) = \mu(\emptyset \cup 1) - \mu(\emptyset) = 80$$

  and player S_2 the remainder, i.e. only 20 units

$$\Delta_2(\sigma^{(1)}) = \Delta_2((1,2)) = \mu(1 \cup 2) - \mu(1) = 20$$

  i.e. the added value of the first is 80, the one of the second is 20.

- in the permutation $(2,2)$ player S_2 arrives first and is assigned a number of units able to saturate her capacity, i.e. is assigned 60 units

$$\Delta_2(\sigma^{(2)}) = \Delta_2((2,1)) = \mu(\emptyset \cup 2) - \mu(\emptyset) = 60$$

  and player S_1 the remainder, i.e. only 40 units

$$\Delta_1(\sigma^{(2)}) = \Delta_1((2,1)) = \mu(2 \cup 1) - \mu(2) = 40$$

  i.e. the added value of the first is 40, the one of the second is 60 units.

- thus (in terms of type a service provision) the average added value of S_1 is 60 and the average added value of S_2 is 40
- the respective Shapley values are (proportional to) $v_{S_1}^{Shapley} = 60$ $v_{S_1}^{Shapley} = 60$
- the ideal allocation, proportional to the average added value is shown in the following table along with the corresponding revenue for the two players.

	a	b	c	Total
Fair allocation to S_1	60	60	40	
Fair allocation to S_2	40	40	60	
Unitary profit	2	3	4	
Fair gain for S_1	120	180	160	460
Fair gain for S_2	80	120	240	440
Gran Total				900

Table 6.5: Table representing the hypothetical fair allocation dictated by the Shapley criterium.

The relative deviation (or gap) and the perceived unfairness index ϕ can now be computed using the expression

$$\phi_i \equiv \theta\left(v_i^{Shapley} - v_i^{factual}\right)/v_i^{Shapley}$$

where $\theta(z)$ is such that $\theta(z) \equiv 0$ if $z < 0$ and as $\theta(z) \equiv z$ otherwise.
They are shown in the following table.
The ϕ values are $\phi_{S_1} = 0$ and $\phi_{S_2} = 0.32$.

Player		Factual gain	Ideal Gain	Gap		ϕ	Probability
S_1		600	460	-140		0.00	0.00
S_2		300	440	140		0.32	0.14

Table 6.6: Table representing the index ϕ and the attack probability.

We can now translate this value in a probability, based on the expert opinion. Let us finally assume that the experts have given us the following sigmoid function, mapping the index ϕ into a probability (the probability parameter of a Bernoulli model):

$$p = (1 + \sqrt[3]{2\phi - 1})/2$$

notice that $p = 0$ for $\phi = 0$, $p = 1/2$ for $\phi = 1/2$ and $p = 1$ for $\phi = 1$.
In our case the function issues the probabilities of attack $p_{S_1} = 0$ and $p_{S_2} = 0.14$.
Finally obtaining from, experts or other means, the impact I_1 of the attack by S_1 it is possible to compute the risk associated to the actor S_1 originating from S_1 perceived unfairness alone. It will be $R_{perceived\_unfairness} = I_1 p_1$.

6.2.1.2 Modeling the Greed Motivated Attacks

In the plain text scenario A, the actors COMP, S_1 and S_2 can be motivated by greed to deliver an attack. In greed motivated attacks the tenet is that the probability of an attack is proportional to the expected value of the economical gain from the attack (in which the possibility of being identified as the source of the attack and undergoing the corresponding negative consequences is also taken into account).
The computation is based on expert knowledge and follows the decision theory paradigm of the problem consisting in choosing wether to enter or not into a bet:

- Leaking information = entering the bet
- Not leaking information = not entering the bet

We recall from the previous chapter that the individual's expected utility for committing an offense is

$$g = p^{(-)} u(v^{illegal} - d) + (1 - p^{(-)}) u(v^{illegal})$$

where u is the individual's utility function, $p^{(-)}$ is the subjective probability of being caught and convicted, Y is the monetary equivalent from an offense, and d is the monetary equivalent of the damage deriving from the inflicted punishment. The individual will commit the offense if the expected utility is positive, and he will not if it is negative. If the utility function is approximately linear, one can write

$$g \simeq u(v^{illegal}) - p^{(-)} u(d)$$

In the present scenario $v^{illegal}$ is the monetary value of the leaked information (shadow) market, while d represents the monetary value of the punishment, typically in terms of business loss and fines to pay if caught; $p^{(-)}$ is the probability of being caught; the utility function can be elicited based on expert opinion.
Thus, the elements involved in the computation are the following

- Experts of the airline domain provide

  - An evaluation of the economic value of the airline information for each of the available buyers: the maximum (or another aggregator) can be used to determine the reference value v^{info}; the fraction r which can be asked by the leaker: the illegal gain for the leaker is $v^{illegal} = r\,v^{info}$.
  - An evaluation of the probability $p^{(-)}$ for the leaker that, should she leaks information, she would be actually identified as the source of the leakage.

- Behavioral experts provide utility functions $u(v^{illegal})$ and $u(d)$. In the example hereafter we identify the value of the utility function with the economical value so that $u(v^{illegal}) = v^{illegal}$ and $u(d) = d$

Based on this the choice by the potential attacker on whether to enter the bet (to attack by leaking information), or not enter the bet, can be modeled *from the point of view of the potential attacker*. We use the COMP provider point of view as an example, in this case the meaning and illustrative of the bet elements are the following

- Negative part: being identified as the attacker and to suffer negative consequences
  - The economical damage from the negative consequences d.
    For a COMP service provider can be determined for instance by the content of the contract stipulating a penalty for leakage, and based on the damage of loosing the segment of high-information-value customers.
    Such a value can be provided by an expert.
    We assume that an IT domain expert estimates that the loss in customer fees corresponding the temporary loss of the market segment will cost to the provider $d = 400$ M (million euros).
  - The probability $p^{(-)}$ that the attack is detected and the attacker is identified as such. This depends on the details of the scenario: here if COMP, S_1 and S_2 have all access to the information, however it is likely that if COMP attacks a thorough investigation will determine that she is the responsible.
    Let us assume the experts assign to this event probability $p^{(-)} \simeq 0.70$.

- Positive side: benefit from selling the information, consisting in
  - The economic value $v^{illegal}$ that the attacker can obtain on the market, this is set by the market and Corresponds to a fraction of the gain in competitiveness of the buying party (the competitor airline company).
    Also this value of can be reliably assessed by a domain expert.
    Suppose airline domain expert estimates that this value amounts to $v^{illegal} = 350$ M.

Using the mentioned figures, the bet, from the point of view of the attacker turns out to be favourable, indeed:

$$g = 350 - 400 * 0.70 = 70$$

Thus COMP can enter the bet: if adopting a rational behavior, she will attack.

Point of view of the defendant Notice that the potential attacker has the possibility to probe the (underground) market of industrial information and to know with precision what is the worth of the information she owns. From the point of view of the defendant the reasoning is the same, however the information available about the parameters is typically poorer. The point of view of the defendant is

the most relevant one for the methodology, since it is the one taken by the decision makers, deciding at design time the choreography of the process and the kind of measures to be taken.
The defendant does not know the exact worth $v^{illegal}$ of the information on the shadow market, nor the exact worth of the damage d which could be suffered by the attacker: in place of a rather precise estimate of the above expected value, the defender can obtain from experts distributions of such parameters and eventually the probability distribution over possible values of the attacker expected gain g. So while for the attacker g is a number, for the defendant it is an ill-known variable, and typically is treated as a random variable.
Let us indicate the event of an attack from the attacker by A. Since a potential attacker attacks when her valuation of the expected gain is positive, from the point of view of the defendant, the probability $P(A)$ of an attack equals $P(g > 0)$. This probability enters the process of decision by the defendant and we are going to return shortly on its estimate from the stand point of the defendant.
In the example of COMP, we assume the distribution of the variable g as known by the defender has turned out to be a Gaussian $G(g|\mu = -50, \sigma = 50)$: then

$$P(A_{COMP}) = Prob(g > 0) = G(g > 0|\mu = -50, \sigma = 50) = N(z > 1) \simeq 0.17$$

where $G(\cdot)$ indicates the gaussian density, $N(\cdot)$ indicates the standard normal density and z is the standardized variable $z = (g - \mu)/\sigma$.
The bottom line for this example is that the estimated probability of attack by COMP is $P(A_{COMP}) = 0.17$

6.2.1.3 Probability of attack due to both drivers

An attack can, in our model, be considered as determined by two drivers: perceived unfairness and greed.
So far we have considered the two drivers separately, using the example of S_1 and S_2 for the first driver and the example of COMP for the second driver. However, they can happen to be both relevant for an actor.
For instance in the present scenario they are both relevant to S_1 and both relevant to S_2.
To deal with both drivers at the same time we enrich slightly the notation: $P(A_i^D)$ will indicate the Probability of an attack from actor i due to the driver D alone; we describe hereafter how to compute $P(A_i^{D_1 \cup D_2})$, i.e. the probability of an attack by an actor due to either one or both the drivers. Since we assumed in the previous chapter that the two drivers can be considered approximately independent, the computation of $P(A_i^{D_1 \cup D_2})$ can be carried on starting from $P(A_i^{D_1})$ and $P(A_i^{D_1})$ by means of standard probability manipulations; indeed, using the symbol $\neg$ for negation, no attack means no attack due to the first driver and no attack due to the second driver:

$$P\left(\neg A_i^{D_1 \cup D_2}\right) = P\left(\neg A_i^{D_1}\right) P\left(\neg A_i^{D_2}\right)$$

so

$$1 - P\left(A_i^{D_1 \cup D_2}\right) = \left(1 - P\left(\neg A_i^{D_1}\right)\right)\left(1 - P\left(\neg A_i^{D_2}\right)\right)$$

or

$$P\left(A_i^{D_1 \cup D_2}\right) = 1 - \left(1 - P\left(\neg A_i^{D_1}\right)\right)\left(1 - P\left(\neg A_i^{D_2}\right)\right)$$

Let us now consider the example of S_2 (let D_1 the driver of perceived unfairness, and D_2 the driver of greed)

- The modeling of the first driver for this actor has yielded in the previous subsection an estimated probability of attack of $p_2 = P(A_{S_2}^{D_1}) = 0.14$

- Let us assume that the modeling of the second driver (analogous to the treatment used above for COMP) has yielded $P(A_{S_2}^{D_1}) = 0.23$

Then

$$P\left(A_{S_2}^{D_1 \cup D_2}\right) = 1 - (1 - 0.14)(1 - 0.23) = 1 - 0.66 = 0.34$$

Finally assuming symmetry in greed between S_1 and S_2 (i.e. $P\left(A_{S_1}^{D_2}\right) = P\left(A_{S_2}^{D_2}\right)$) and observing that neither for COMP, nor for S_1 the unfairness represents an actual driver (see above), we have the following probabilities, to be used in the risk profiles of all the actors S_1, S_2 and COMP.

$$P\left(A_{S_1}^{D_1 \cup D_2}\right) = P\left(A_{S_1}^{D_2}\right) = 0.23 \tag{6.3}$$

$$P\left(A_{S_2}^{D_1 \cup D_2}\right) = 0.34 \tag{6.4}$$

$$P\left(A_{COMP}^{D_1 \cup D_2}\right) = P\left(A_{COMP}^{D_2}\right) = 0.34 \tag{6.5}$$

6.2.2 Scenario B: Two Cloud Providers and Use of Secret Sharing

Consider the scenario in which countermeasures are taken: the computation now uses secret sharing (the secret can now be reconstructed only using all the shares, in this illustrative case just two). In this case,

- the attack from S_1 and S_2 motivated by perceived unfairness is impossible, since they do not know the details of the spare part provision allocation
- the attack motivated by greed by S_1 and S_2 is impossible for the same reason,
- the attack motivated by greed from the COMP providers is impossible unless the two providers collude: it is very unlikely that two market competitors collude to attack their customers:

one can safely assume that in this way the attacks are disabled.

6.2.3 Modeling the Point of View of the Decision Maker (the Defendant)

Now we consider point of view of the decision maker, who in this context takes the role of the defender. He has to choose between two strategies $\{Defend, Not\ Defend\}$, based on the following elements of the computation

- Likelihood $P(A)$ of undergoing an attack: its computation has been illustrated in the previous subsection
- Damage from the attack $Damage(A)$,
- Cost of the defense $Cost(Defense)$,
- The expected value of the strategy Defend is $-Cost(Defend)$,
- The expected value of the strategy Do Not Defend is $-Damage(A) * P(A)$,

The defendant will choose the strategy with the highest expected value. The options are illustrated in Figure 6.2.3. So, the choice depends on the following balance

$$Damage(A) * P(A) - Cost(Defend)$$

(if balance > 0, then choose Defend).
In the expression

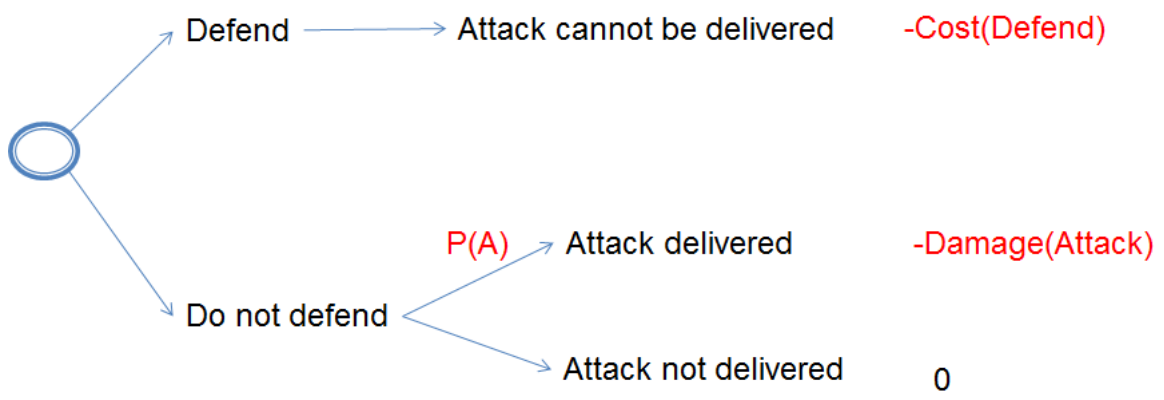

Figure 6.8: The tree of the possibility from the point of view of the Defendant.

- $Cost(Defend)$ is the cost of buying, deploying and maintaining the defense (e.g. using secret sharing and an extra Cloud Provider)
  This information is known with precision by the defender
- $Damage(A)$ is the decrease in business revenues due to the attack: e.g. a loss in competitiveness.
  This information can be assessed with reasonable precision by an domain expert
- $P(A)$ is the likelihood that the potential attacker(s) will attack
  This information is the one obtained from the analysis developed above, namely the one based on the Shapley Value modeling the perceived unfairness driver and the one based on the greed driver.

An example calculation with some example values is the following.
Suppose the airline domain expert estimates with reasonable precision the damage from ”OR”-ing of all the possible attacks as

$$Damage(A) = 300M$$

Then any value of $Cost(Defend)$ such that

$$Cost(Defend) < Damage(A) * L(A) = 300/4 = 75M$$

It is acceptable for the defender and the defender should take action by deploying the defenses.

Chapter 7

Tool Description

In this chapter we describe our open source web tool that supports our methodology for risk assessment, supporting both the modeling and the simulation of business processes executed on the cloud. In detail we will present the requirements, the technology solutions and the interfaces of the software, while in the second part we show some sample applications.

7.1 Simulator Requirements

The editor we developed has the primary aim to support the process of representing and simulating cloud-based business processes for risk analysis purposes, enforcing at the same time some compliance and consistency checks. The procedure to implement the protocol starts with the creation of a graph, where the nodes represent the actors participating in the process, modeled according to the Bogdanov model [12], while the edges are the exchanged messages.

Once the model has been defined, a step-by-step simulation can be started in order to analyse the information flow among the actors. Our editor allows to:

- Create, import and export processes.
- Check the consistency of the process against the constraints imposed by the simulated encryption techniques. For example, in the model only actors *IN* and *COMP* are allowed to send data and the management of the correct number of shares in a system.
- Compute the risk of collusion-to-misbehave for each subset of actors considering external input.
- Display charts and graphs to help measure, step by step, the risk of disclosure during the execution of the protocol.

The tool's modular architecture facilitates the future implementation of additional features.
Each process incarnation is stored in the standard JSON format[1], so as to facilitate the sharing of the process and to enable interoperability between our software and future working groups.

[1] JSON (JavaScript Object Notation) is a lightweight data-interchange format, easy for humans to read and write and for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999.

7.2 Simulation Editor

Our editor represents processes as graphs. At the implementation level, the editor takes as input a graph $G = (V, E)$ containing a set of edges E, representing messages exchanged, such that each edge is assigned a time instant from 0 to t, and a set of nodes V, representing three categories of actors:

- *IN* - Actors who collect and send data to the process.
- *COMP* - Actors that perform a computation in response to the input data of the *IN* actors.
- *RES* - Actors who receive and display the result of the computation performed by the *COMP* actors.

Once the input graph has been stored, the editor verifies the following constraints:

1. The *IN* actors cannot have input edges (i.e., if $v \in IN$ then $uv \notin E$ for any $v \in V$).
2. The *RES* actors (nodes) belonging to V cannot have output edges (i.e., if $u \in RES$ then $uv \notin E$ for any $v \in V$).

This way, only *IN* and *COMP* actors (nodes) have fan-out output edges.

7.3 Design and Implementation

The design of our toolkit has been divided into five phases:

1. Verification of the requirements imposed by security controls that can be modeled.
2. Identification of the technologies needed to meet the requirements.
3. Identification of the needed resources.
4. Definition of the MVC components (transitions, ordered transitions, passed parameters).
5. Identification of standards and data structures to be used.

7.3.1 Technology Solutions

Our editor has been developed as a Web-based application. All content that the application generates and manages is defined with the standard JSON so as to enable seamless interoperability between the different parts of the project. In the initial design, visualisation technologies and simulation were separated. The visualisation functionality has been implemented using client-side technology through the use of the JavaScript language. The simulation has been implemented server-side using Java. Given the continuous technological progress of client-side JavaScript interpreters in modern browsers, the work done on the server side was subsequently brought to the client side, allowing the usage of the computational power of the user only (and avoiding a server dedicated service). The technologies we used are the following:

1. *JavaScript*. The Web application is fully implemented through the use of JavaScript. Some precautions and optimizations have been used to obtain a satisfactory product, overcoming the limitations of the selected language.

2. *AngularJS*. It is a framework that allows the creation of web applications using MVC client-side. The choice has been motivated by the possibility to manage dynamic and organization projects in medium / large as the following ones. This framework also helped to break down the part of the inherent logic of the simulation with the layout of the process through the use of the properties of two-way data binding where the data model and controller are linked by an internal loop of AngularJS. By doing so, a modification to the model is instantly passed to the controller and vice-versa.

3. *SVG*. Scalable Vector Graphics is a technology that can display objects of vector graphics through the XML specification. This technology has allowed, along with the library d3.js, to bring the whole process graphically in the graph used for simulation, allowing smooth and easy visual interpretation.

4. *Canvas*. Canvas is an extension of HTML standard that allows dynamic rendering of bitmap images manageable through JavaScript. This technology has allowed us to bring graphically, through the CanvasJS library, the probability of collusion between all possible subsets of actors involved in the process.

5. *Bootstrap*. Bootstrap allowed to give a standard layout to the project, also allowing a responsive design which provides for the adaptation of the style for multiple resolutions and devices.

7.4 Software Architecture

The modular architecture of our toolkit is shown in Figure 7.1. The Main page is the main resource which contains the command interface. Each view (View) can access the template through the use of $ scope, i.e. an object containing instances of variables managed by the controller. The link between model, view and controller is mediated from the $ scope and dynamically updated through two-way data binding with the inner loop of *AngularJS*. As you can see in Figure 7.1, each controller is assigned to a view and each controller can in turn call a provider that provides a particular service. This technique called dependency injection, has enabled a dynamic structure of the project that will be easily manageable and maintainable in the future.

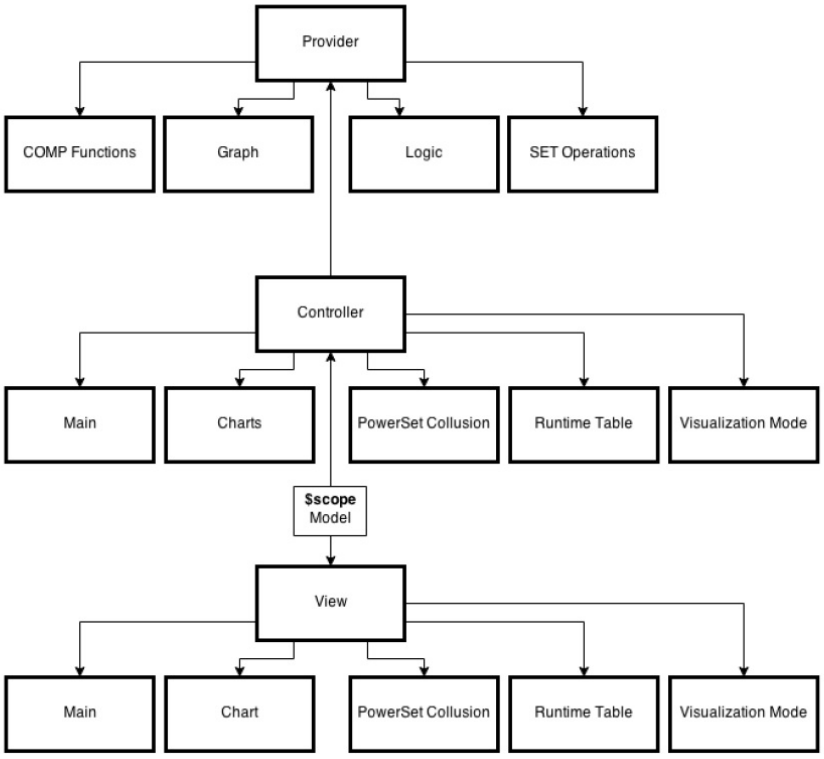

Figure 7.1: Software Architecture

7.4.1 Services

Our Providers allow handling of almost all of the logic of the project by providing a specific service used by one or more controllers. Below are described the functions carried out by each provider:

- *COMP* Functions. It provides the definition and the application of all functions $f()$ available to the actors performing computations. Some simple functions are pre-implemented such as sum (SUM), minimum (MIN) and maximum (MAX). Additional implementations can be easily made available thanks to dependency injection.

- Graph. The following service provides visualisation of the graph within the editor. Any change is delegated to it through the use of the library $d3.js$, modelling the graph according to the SVG specification. Each node is stored in a data structure in dictionary such that, every addition, change, or removal of elements within the process has a constant complexity $O(1)$.

- Logic. This service is the most important one, since it handles all the simulation logic of the application. The service allows the reorganisation of the data structure from the service Graph to make easy the simulation of the protocol both from the visual point of view and from the logic level. The reorganisation of the data structure has linear complexity $O(n)$ where n is the number of actors involved in the process.

- SET Operations. The service provides a set of operations that can be performed on sets. Particular importance was given to the function for generating power-set. The complexity of the latter is equal to $O(2^n)$.

7.5 Interface

Each controller has a set of associated functions:

MAIN. This controller refers to the main page. It is used to call the remaining controllers based on events generated by user actions.

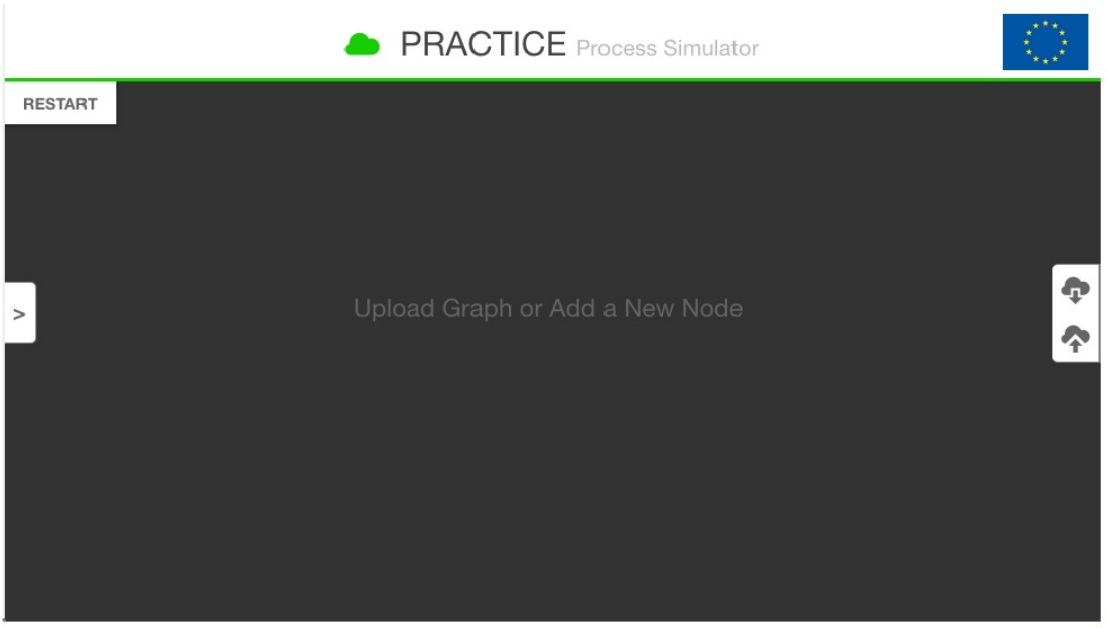

Figure 7.2: Main page, controlled by the MAIN controller

In Figure 7.2 the main interface is shown. Through the buttons on the right side, the user can upload a JSON representation of a previously created project and download it through the second button. Figure 7.3 shows how it is possible to create a new process through the addition of nodes and arcs which represent, respectively, actors and communications. The creation of an actor requires the following information:

- **Name**. The identification of the actor.
- **Descr**. The description of the actor.
- **Mem**. the input data of the process at the time of its creation. It is not available in the case of nodes that are part of the subsets COMP or RES, as specified.
- **PMAL**. It defines the probability of a node to be malicious. The value is constrained to be between 0 and 1.0.
- **TYPE**. Identifies the type of actor (IN, COMP, RES). By choosing the generation of an actor COMP you will also be asked to select the type of function F that the node should perform.
- **Provider**. Designates which cloud providers the actors resides. It allows to study the risk based on a prior evaluation of the provider.

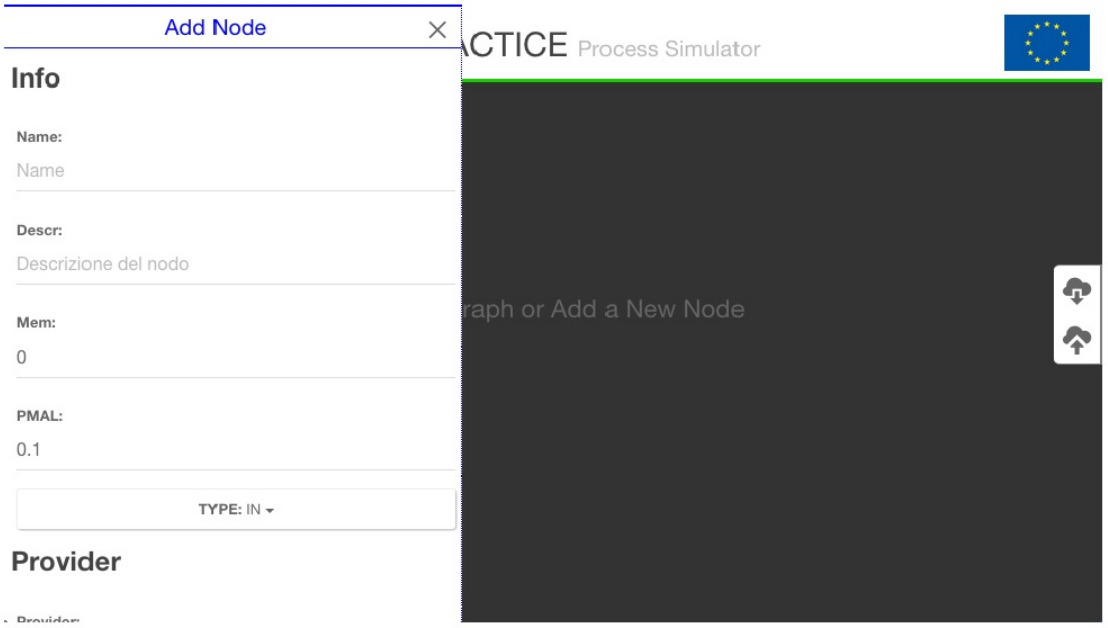

Figure 7.3: Main Page with overlay used to add nodes

The creation of a communication between actors requires the following information:

- **From**. The identification of the actor from which a particular piece of information is produced.
- **To**. The identification of the actor to which information flows.
- **Time**. The instant of time when the communication occurs.
- **TYPE**. The type of channel to use. The editors implements PLAIN TEXT, ENCRYPTED, SECRET SHARED and GARBLED CIRCUITS.

Figure 7.4 shows a process consisting of three actors: an input node ($IN1$), a computational node ($COMP1$) and a node setting out the results ($RES1$). Every actor type is identified by a different color: blue for IN, red for $COMP$ and yellow for RES. Similarly, the messages adopt a different color depending on the security control that is deployed: yellow for secret sharing, blue for encryption, and red for garbled circuits; plaintext it will be gray. By default, the editor uses the following notation regarding the transfer of data through security controls:

- **value** for encryption,
- **[value]** for secret sharing,
- ***value** for garbled circuit.

Once the process model has been created, the buttons in the interface allow the execution of a consistency check (CHECK GRAPH COHERENCE) according to the constraints, the running of the simulation (RUN VISUAL SIMULATION), the computation of risk by means of the external component implementing the algorithms presented in this deliverable (ESTIMATE RISK) and, finally, the computation of the risk in the event of a collusion between the different subsets of actors (Coalitions RISK SETS). This controller uses the service Graph to generate the graph of the controlled process

and COMP Functions to derive some of the functions available during the creation of the actors. A detailed description of the components of an actor in a simulation is started in figure 7.4. It is of course possible to modify the process by clicking on the actor or on the communication so that an input mask in an overlay level is displayed reporting the parameters of the object that can be corrected.

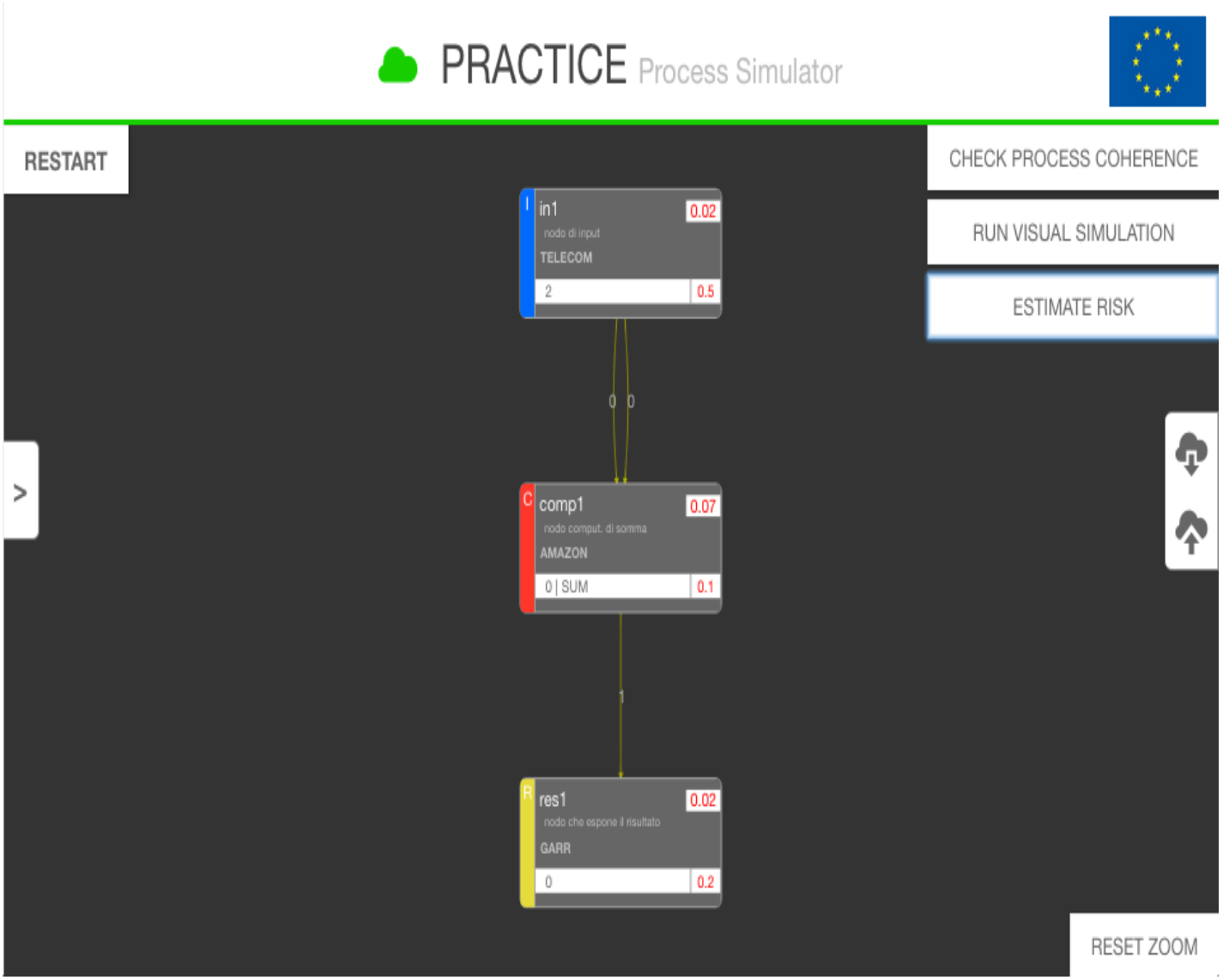

Figure 7.4: Main Page with the visualization of the process.

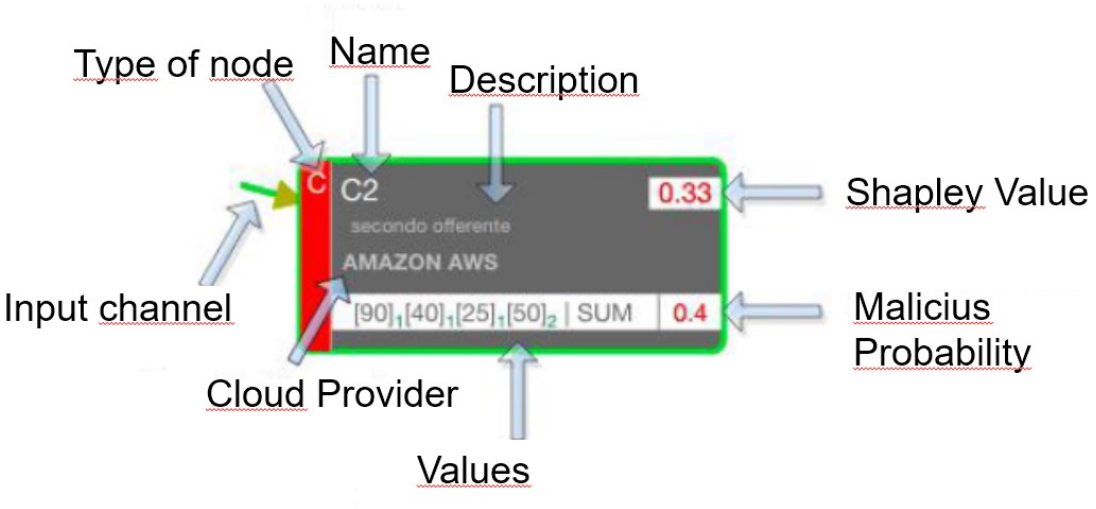

Figure 7.5: Node description.

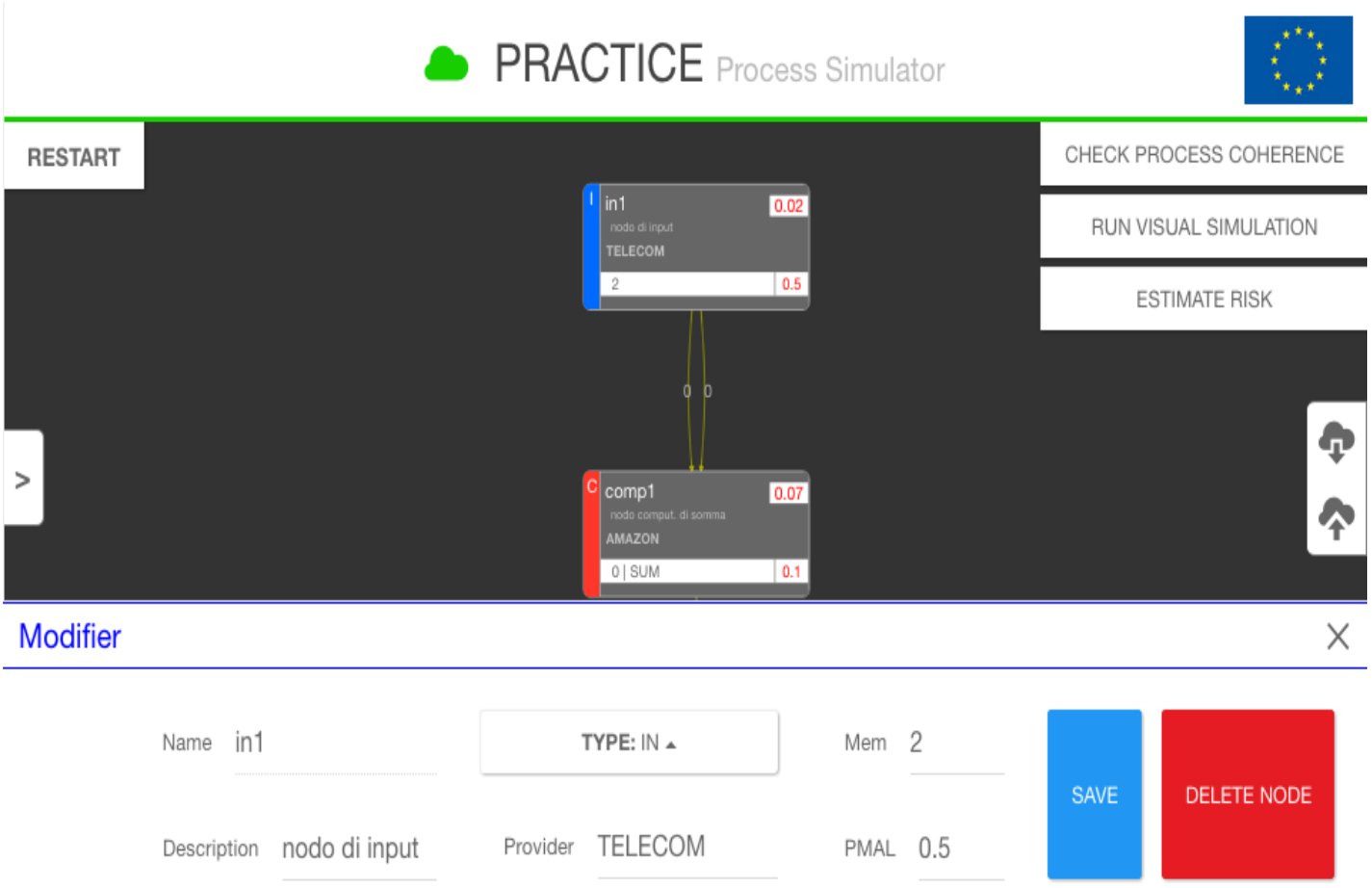

Figure 7.6: Edit interface of a single node.

Charts. The task of this controller is the management of graphics produced by the simulation. The generation is done by the ”Coalitions RISK SETS” that generates a graph as in Figure 7.7. Currently it is possible to display the probability of collusion between multiple actors through a representation of the Cartesian plane where the x-axis contains the number of actors that come together and ordered on the real chances of any coalition they can join. The latter is calculated by the product of the probability $P(A)$ of each actor A to be misbehave (parameter $PMAL$).

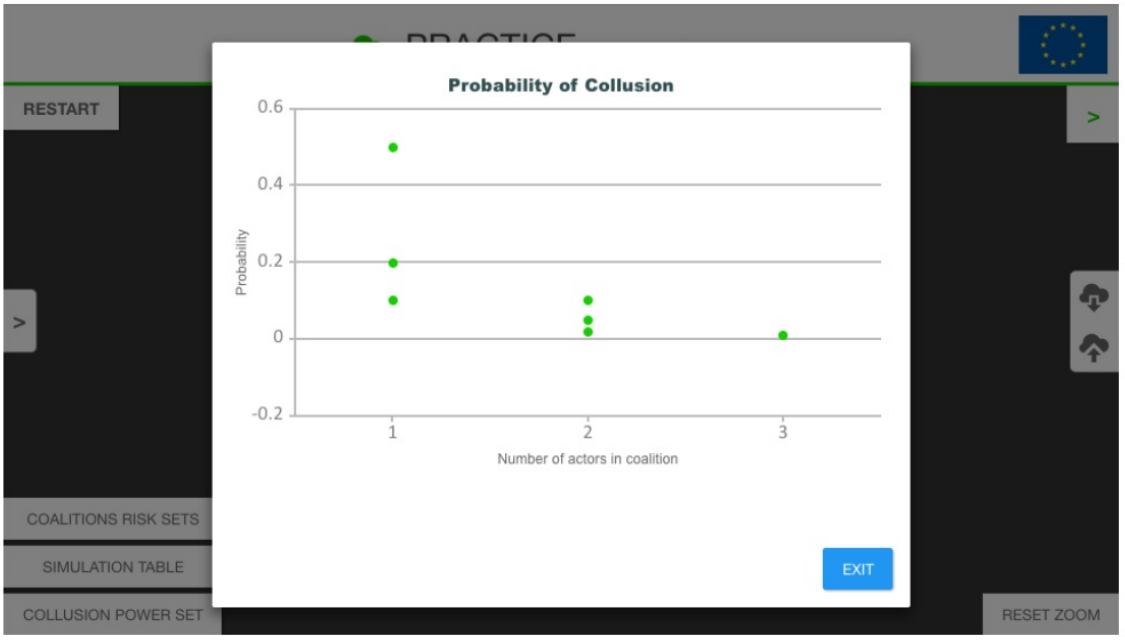

Figure 7.7: Charts that represent the probability of collusion.

Visualization Mode. The task of the following controller is to manage the way in which the simulation is executed. As can be seen in Figure 7.8, a simple control interface allows the selection of a (”STEP-BY-STEP”) simulation, or the execution of the protocol through automatic update (”AUTO INTERVAL ”) of the simulation at the end of an input time period (expressed in milliseconds).

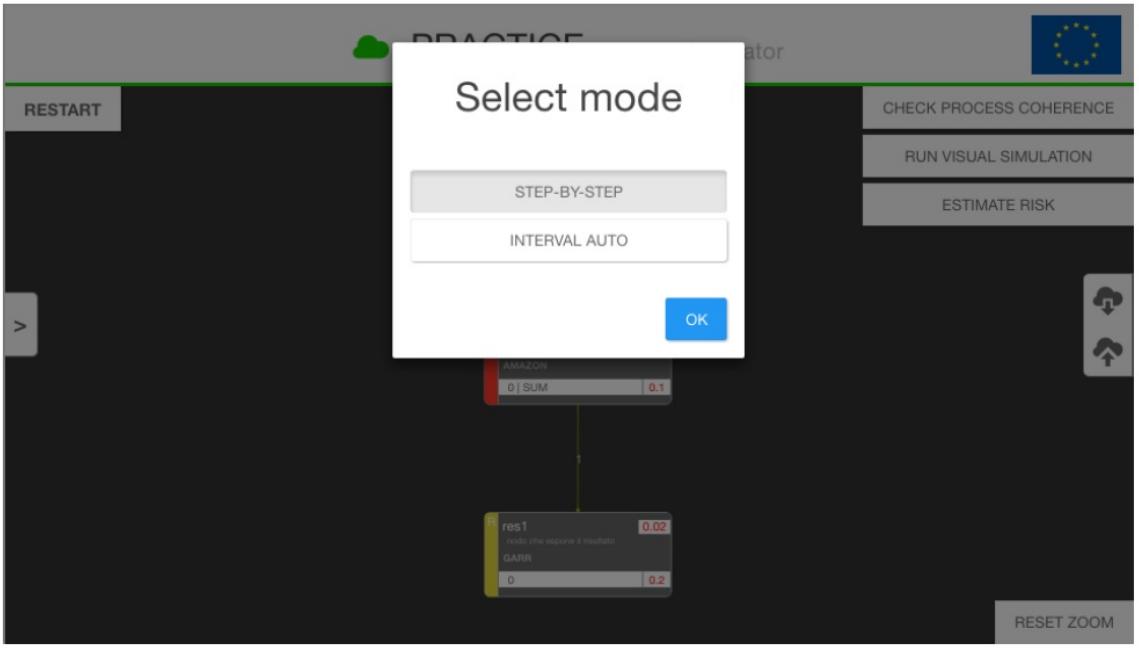

Figure 7.8: Selection of the type of simulation to run.

Power Set Collusion. This module deals with the computation of subsets of actors which and the Knowledge Set they jointly can access. For example, in the case of secret sharing, it highlights the possible reconstruction of the shared secret if all the shares are held by the subset of actors. As shown in Figure 7.8, each line represents one possible subset $S \in 2^A$ where 2^A is the Power Set of the actors of the process. Each column shows, for each instant of time from 0 to t, possible leaks of confidential information that, while encrypted, can be reconstructed in case of collusion among the actors.

Runtime Table. The following controller allows the analysis, once started the simulation, for each time instant. For each actor, in a particular time, it is possible to display the information contained in its local memory (field $INFO$). As can be seen in Fig. 7.10, table rows correspond to moments in time and columns represent the actors of the business process being analysed.

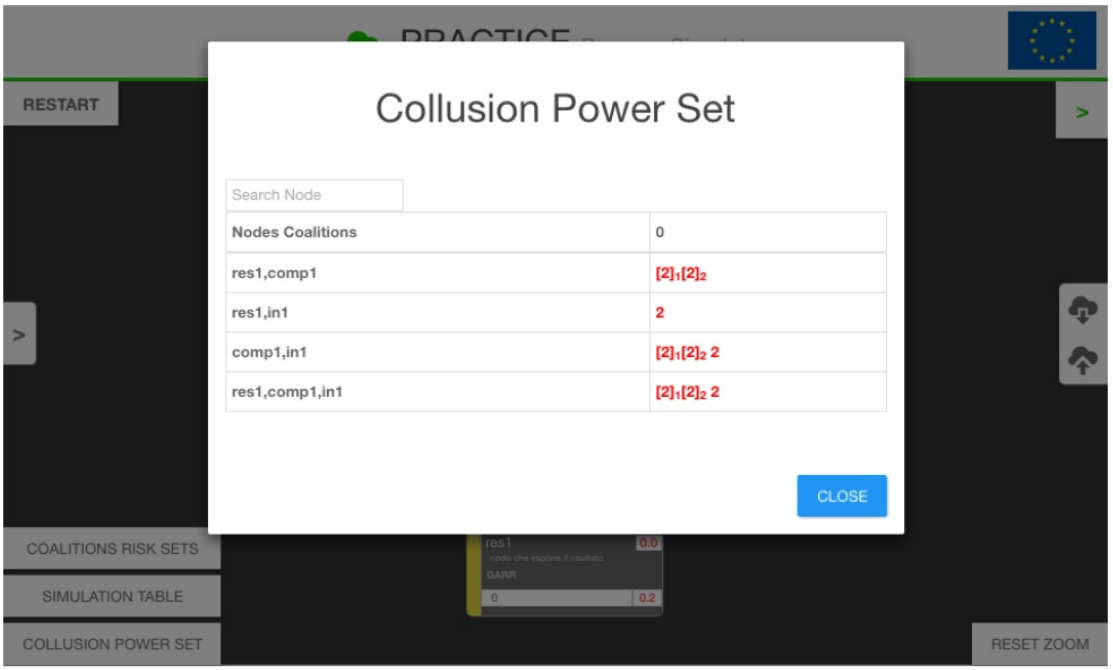

Figure 7.9: Table that shows aggregated information in the case of collusion.

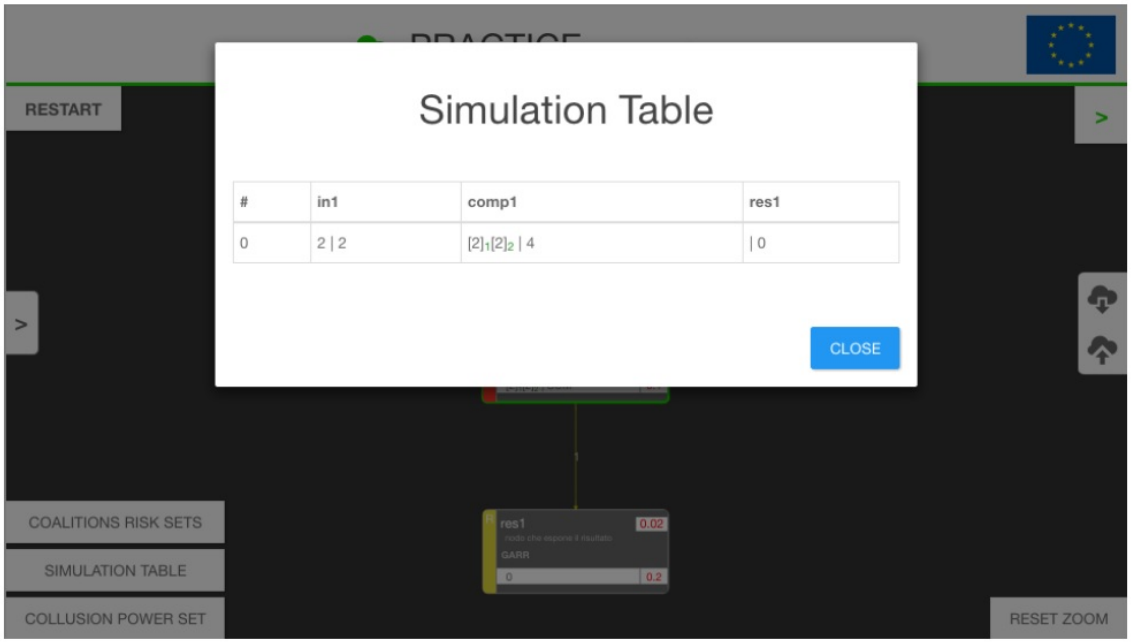

Figure 7.10: Table that shows information about every actor.

7.6 Simulation Process

The simulation of a process including Secure Multiparty Computation (SMC) controls can be done in the editor in two modalities (Fig. 7.11). In the first mode, the data exchanged among the actors is displayed automatically. A time interval (in milliseconds) is defined by the user; when the time expires, the simulation time is incremented (until it reaches a threshold value representing the end

of the simulation). In this type of simulation, only the graphical evolution of the protocol can be displayed. In the second modality, the simulation is interactive, so that it is possibile to follow the execution of the protocol step-by-step.

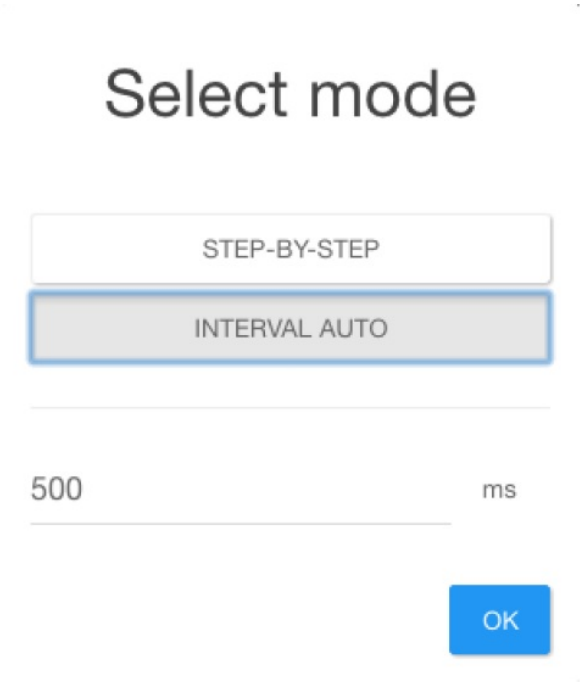

Figure 7.11: The interactive modality.

It is also possible to inspect, at any point in time t, the field $INFO$ of each involved actor and check whether a certain collusion can result in disclosure of sensitive information. The simulation process starts the Logic service, which encodes the data structures in the input graph in order to improve simulation performance. Specifically, at simulation time $t = 0$ all nodes and edges in the model representing actors are selected and a dictionary data structure called *Nodes* is created, whose key is the unique ID of each actor. In this way, it is possible to have a direct access to additional information such as each actor's communications with the other nodes, the type, the probability of misbehaving, the dynamic memory updated moment by moment, the shares generating and other information useful for the simulation. Another data structure is *Instants* which, by means of a dictionary where the key identifies the temporal instants from 0 to t, points to an array of objects which includes all actors who communicate with others, and the type of communication. The field *Edge* is structured to identify the edge on the page with an id formatted as follows: *Outgoing Node - Incoming Node - Temporal Instants - Number of Current Share*. This way of managing communication enables our analysis of memory nodes and collusion checking at runtime.

7.7 Sample Application

In this section, we use our toolkit to model an auction process for a tender, where one or more bidders can collude to obtain mutual benefits. The participants are: bidders, a contracting authority who will choose the lowest offer and an agent who will check who, among the participants, has won the contract. The actors listed above are classified in the editor type space as *IN* for suppliers, *COMP* for both bidders and client, *RES* for the agent in charge of publishing the winning offer. An example is shown in Figure 7.12.

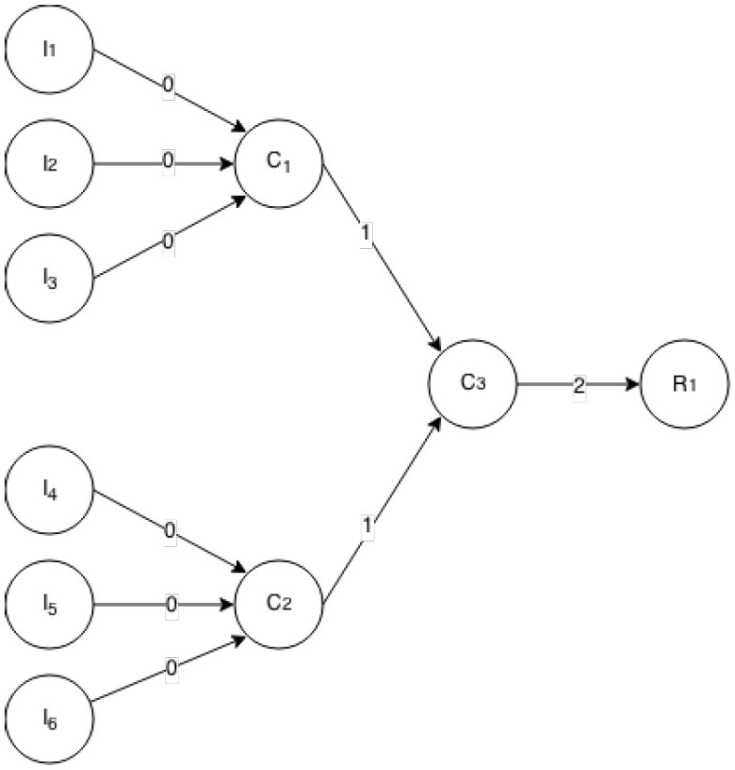

Figure 7.12: Example scenario.

The subsets are as follows:

- $IN = I_1, I_2, I_3, I_4, I_5, I_6$
- $COMP = C_1, C_2, C_3$
- $RES = R1$

C_1, C_2 represent the bidders, who receive as collect the costs of their suppliers and, in turn, submit the sum to C_3 that represents the client. The latter will select the minimum bid. Our editor will compute, from $t = 0$ to $t = 2$, the power set of all involved actors 2^A in the current case turns out to include $2^{10} = 1024$ possible coalitions and, at the same time, scan the memory contents to find instances where there was collusion. In Figure 7.13 one can see the implementation of the previous example, and the display of the editor.

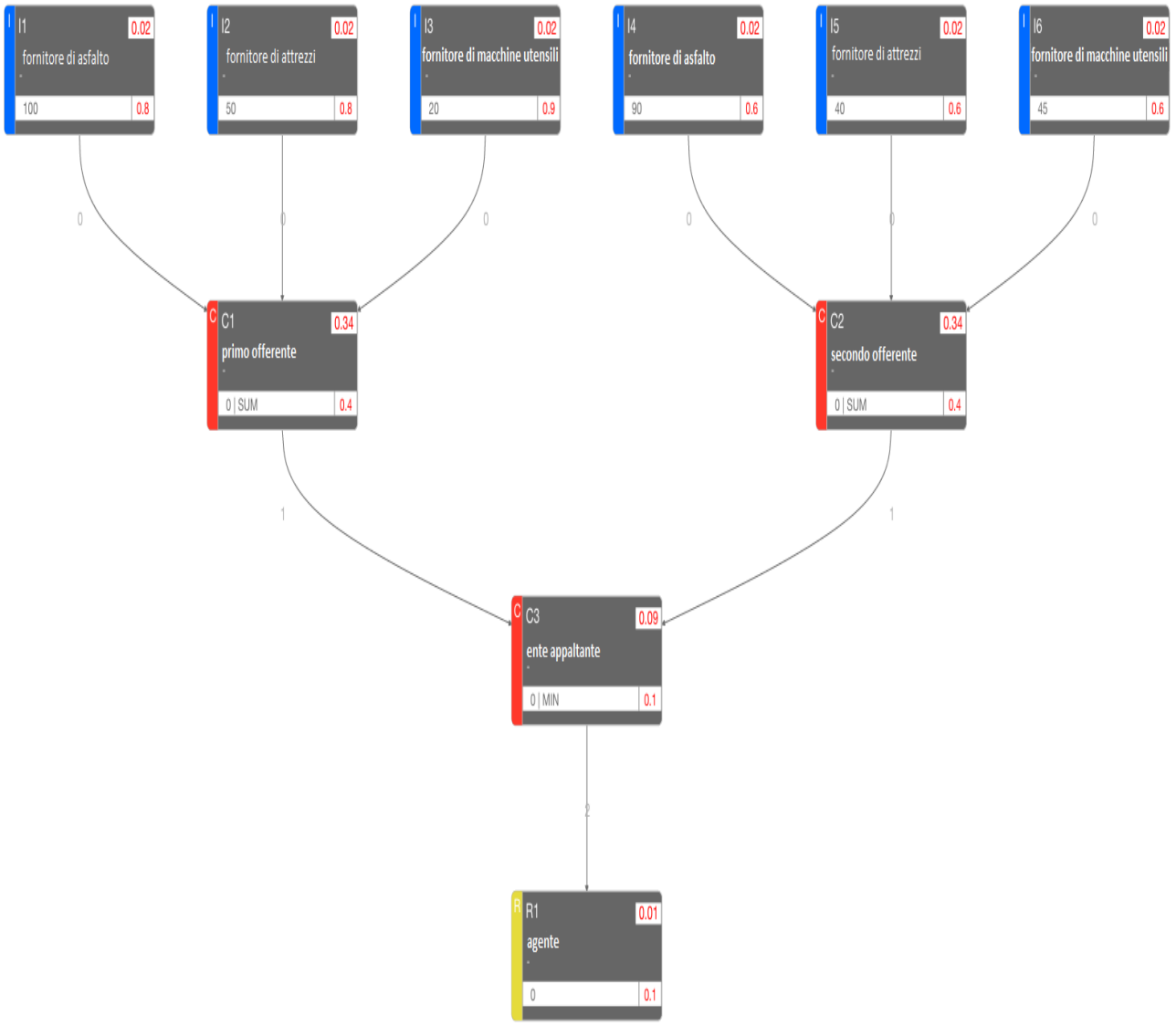

Figure 7.13: Example scenario represented in the editor.

The first step is to analyze a possible collusion sets. Using the graph shown in Figure 7.14, we can see that the highest probability of collusion is reached by sets $I_3 - I_1$ and $I_3 - I_2$. However, looking at the graph, we can observe that the nodes I_1, I_2 and I_3 provide different services to the same bidder. For these reasons, we focus on the actors C_1 and C_2. In fact, in case of collusion of C_1 and C_2, the offers submitted to C_3 are fraudulent.

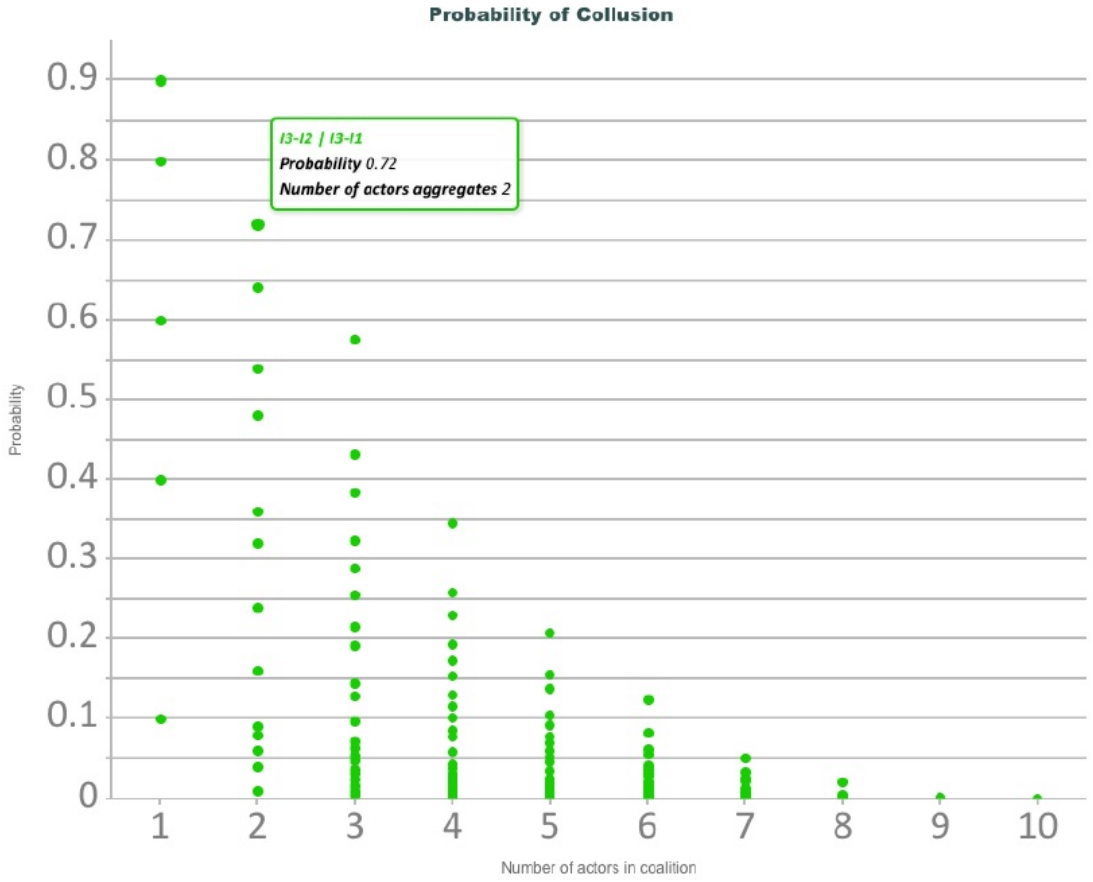

Figure 7.14: Probability of collusion.

Starting the simulation, our tool permits to monitor the data flow during process execution. At each instant, the tool shows the Knowledge Set of each node. Figure 7.15 shows the knowledge of all nodes at time $t = 0$ (e.g., node C_1 has in input the set $100, 50, 20$ and in output 170).

#	I1	I2	I3	I4	I5	I6	C1	C2	C3	R1
0	100 | 100	50 | 50	20 | 20	90 | 90	40 | 40	45 | 45	100 50 20 | 170	90 40 45 | 175	| 0	| 0

Figure 7.15: Representation of knowledge in the new process at time $t = 0$.

Using the *Collusion table* at each time t the user can monitor the Knowledge Sets held by each possible collusion. For example, Figure 7.16 shows the Knowledge Set held by the set $C_1 - C_2$

Nodes Coalitions	0
C2,C1	90 40 45 100 50 20

Figure 7.16: Knowledge Set of C_1 and C_2.

Figures 7.17 and 7.18 show the evolution of process in the next steps, while Figure 7.19 shows the Knowledge Sets at the end of the simulation process.

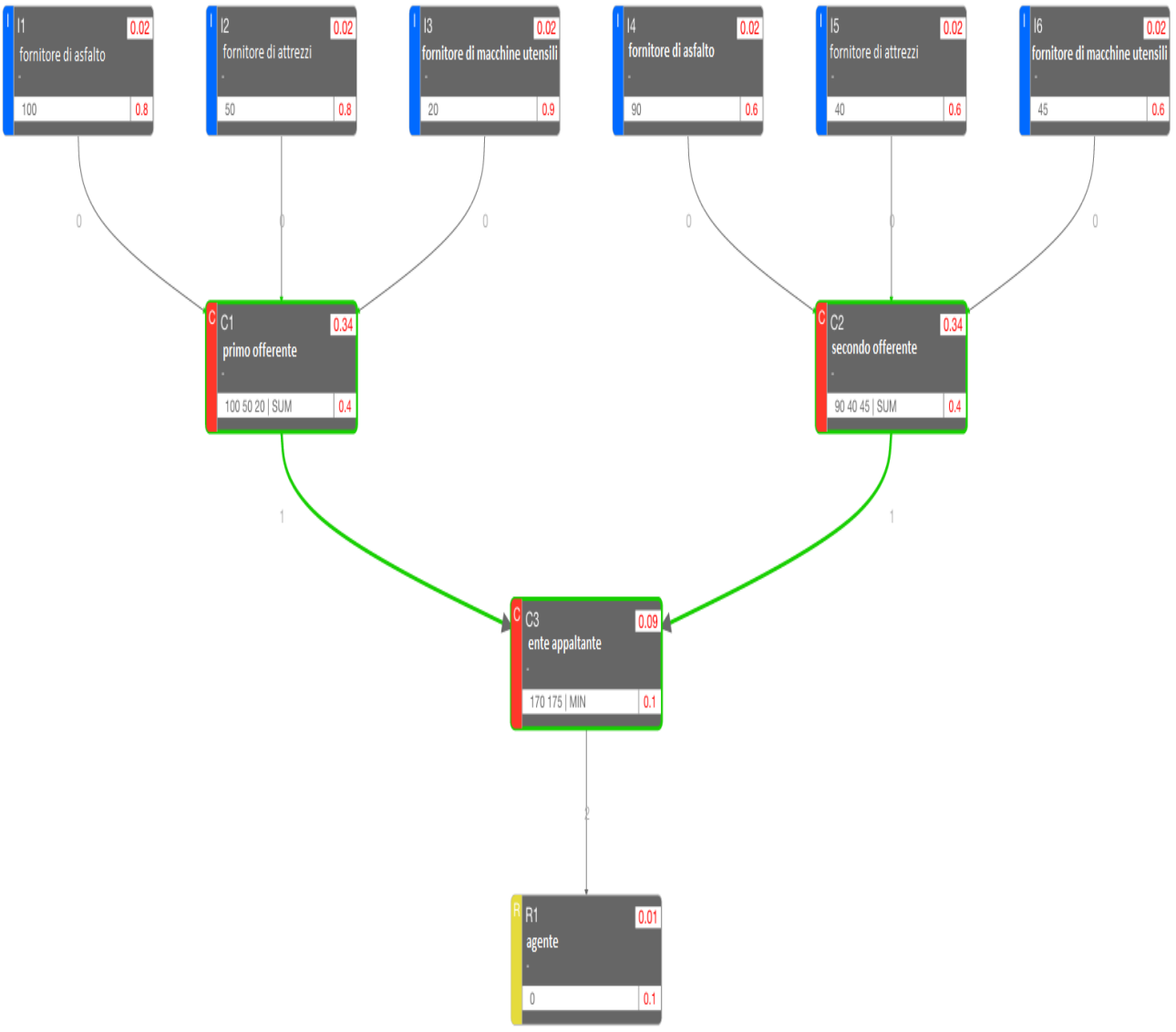

Figure 7.17: Process status at time $t = 1$.

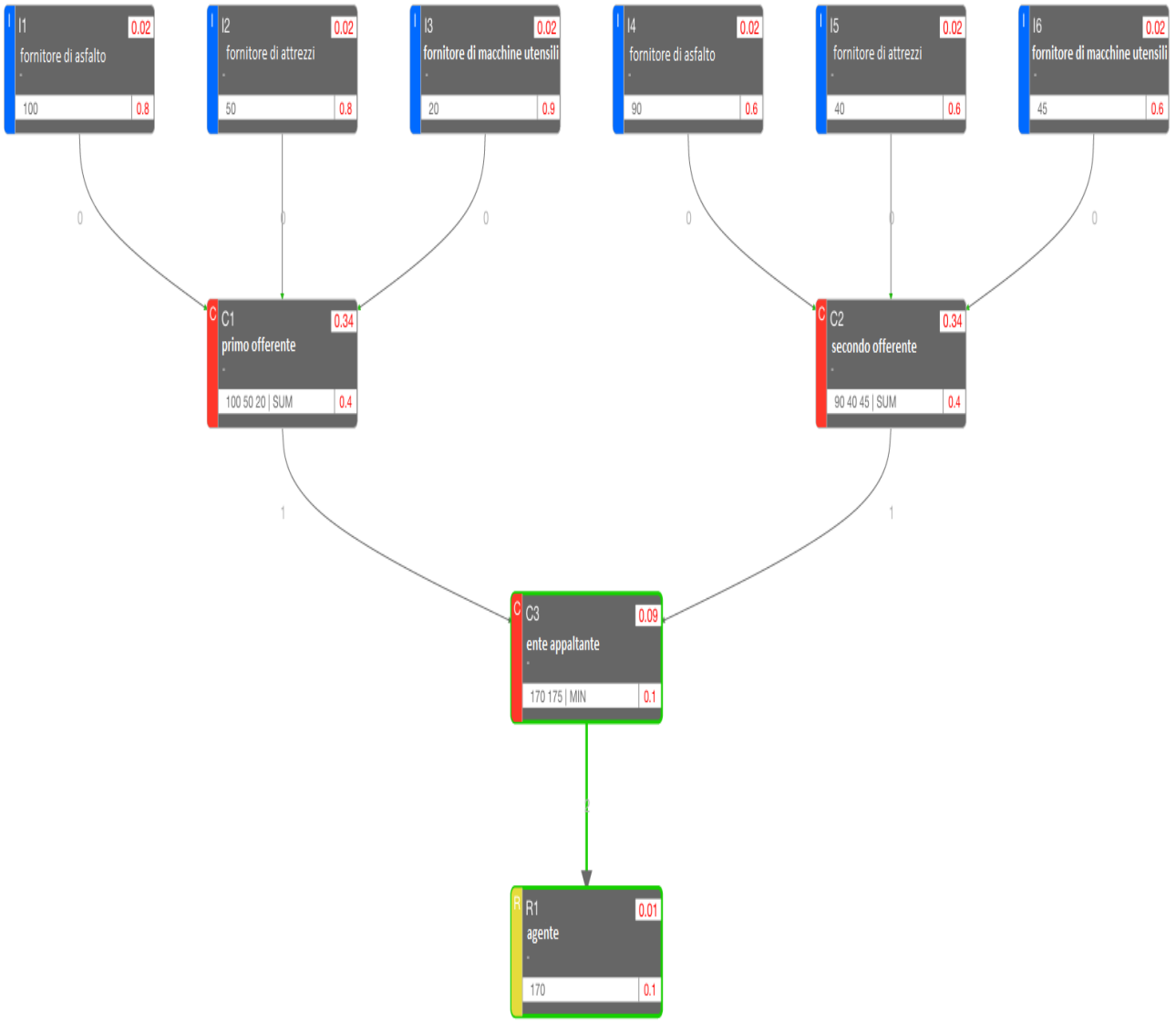

Figure 7.18: Process status at time $t = 2$

#	I1	I2	I3	I4	I5	I6	C1	C2	C3	R1
0	100 | 100	50 | 50	20 | 20	90 | 90	40 | 40	45 | 45	100 50 20 | 170	90 40 45 | 175	| 0	| 0
1	100 | 100	50 | 50	20 | 20	90 | 90	40 | 40	45 | 45	100 50 20 | 170	90 40 45 | 175	170 175 | 170	| 0
2	100 | 100	50 | 50	20 | 20	90 | 90	40 | 40	45 | 45	100 50 20 | 170	90 40 45 | 175	170 175 | 170	170 | 0

Figure 7.19: Knowledge Sets at the end of the process

Let us now try to increase the level of confidentiality. We redeploy a modified business process introducing some security controls implementing simulating a secret sharing technique. Figure 7.20 shows the new architecture. We remark that the number of the required computational nodes increases from 3 to 7.

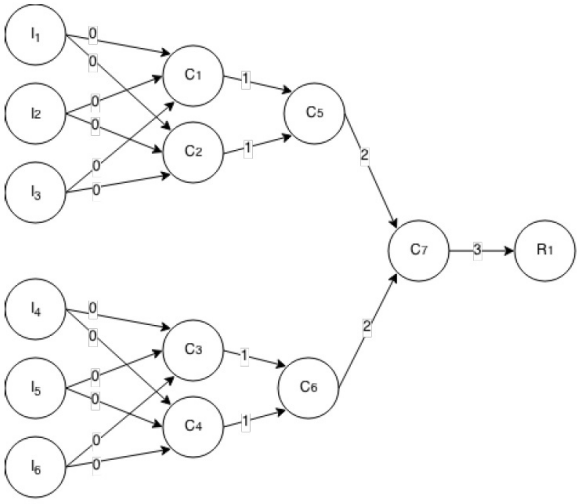

Figure 7.20: New business process architecture with secret share algorithm

The new subsets are:

- $IN = I_1, I_2, I_3, I_4, I_5, I_6$
- $COMP = C_1, C_2, C_3, C_4, C_5, C_6, C7$
- $RES = R1$

Figure 7.21 show the new process in the editor's interface at time $t = 0$. Differently from the original process, the collusion set composed of C_1 and C_3 is now unable to access the knowledge provided by the IN nodes, as data are split into two secret shares and sent to two different nodes.

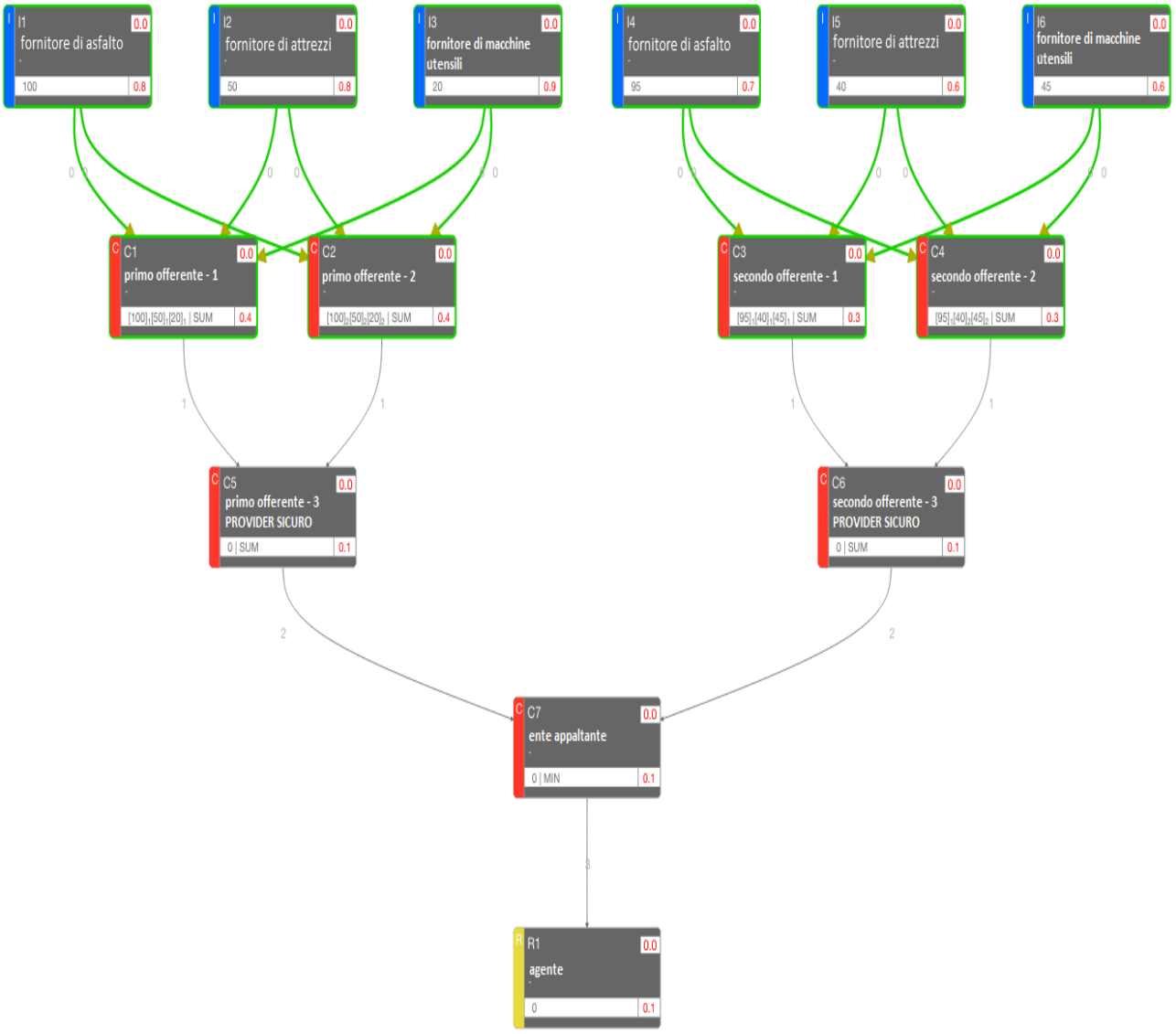

Figure 7.21: New process status at time $t = 0$

Figures 7.22 and 7.23 show respectively the Knowledge Sets at $timet = 0$ by the collusion set composed by C_1 and C_3.

#	I1	I2	I3	I4	I5	I6	C1	C2	C3	C4	C5	C6	C7	R1
0	100 | 100	50 | 50	20 | 20	95 | 95	40 | 40	45 | 45	$[100]_1[50]_1[20]_1$ | 170	$[100]_2[50]_2[20]_2$ | 170	$[95]_1[40]_1[45]_1$ | 180	$[95]_1[40]_2[45]_2$ | 180	| 0	| 0	| 0	| 0

Figure 7.22: Data knowledge at time $t = 0$ in the new process

C3,C1	$[95]_1[40]_1[45]_1[100]_1[50]_1[20]_1$

Figure 7.23: Knowledge Set of C_1 and C_3

At the end we show in the figure 7.24 the Data Knowledge at the end of the process.

#	I1	I2	I3	I4	I5	I6	C1	C2	C3	C4	C5	C6	C7	R1
0	100 | 100	50 | 50	20 | 20	95 | 95	40 | 40	45 | 45	$[100]_1[50]_1[20]_1$ | 170	$[100]_2[50]_2[20]_2$ | 170	$[95]_1[40]_1[45]_1$ | 180	$[95]_1[40]_2[45]_2$ | 180	| 0	| 0	| 0	| 0
1	100 | 100	50 | 50	20 | 20	95 | 95	40 | 40	45 | 45	$[100]_1[50]_1[20]_1$ | 170	$[100]_2[50]_2[20]_2$ | 170	$[95]_1[40]_1[45]_1$ | 180	$[95]_1[40]_2[45]_2$ | 180	170 170 | 340	180 180 | 360	| 0	| 0
2	100 | 100	50 | 50	20 | 20	95 | 95	40 | 40	45 | 45	$[100]_1[50]_1[20]_1$ | 170	$[100]_2[50]_2[20]_2$ | 170	$[95]_1[40]_1[45]_1$ | 180	$[95]_1[40]_2[45]_2$ | 180	170 170 | 340	180 180 | 360	340 360 | 340	| 0
3	100 | 100	50 | 50	20 | 20	95 | 95	40 | 40	45 | 45	$[100]_1[50]_1[20]_1$ | 170	$[100]_2[50]_2[20]_2$ | 170	$[95]_1[40]_1[45]_1$ | 180	$[95]_1[40]_2[45]_2$ | 180	170 170 | 340	180 180 | 360	340 360 | 340	340 | 0

Figure 7.24: Data Knowledge at the end of the new process

Chapter 8

Conclusion

To sum up, Part 1 of the deliverable examines the legal challenges of European data protection law regarding cloud computing and encryption, as well as the technologies developed by PRACTICE. Especially the proposals for a GDPR raise complex new issues regarding cloud computing. The question whether the DPD is applicable when personal data is encrypted is not solved yet, the EJC may answer this dispute by favouring either an absolute or a relative approach. The proposals for a GDPR don't provide a uniform definition of personal data. With the absolute-relative approaches of the proposals of the Commission and of the Council, supplementary knowledge of third persons has to be considered, but encryption of personal data could be a way to anonymize personal data. With the LIBE-Proposal's absolute approach, the GDPR would always be applicable, even if the data is encrypted. A lawful way to process personal data in the cloud both in the DPD as in the GDPR is using order processing on behalf of the controller. Data transfer to third countries is possible by using instruments such as binding corporate rules. Nevertheless, after the Safe-Harbor-Decision of the ECJ, the transfer of personal data to the USA is currently difficult to fulfil. Furthermore, the technologies developed by PRACTICE comply ideally with the principle of Privacy by Design.
If "Encrypted HANA" and "Sharemind" are used, according to the relative approach the encrypted data would not be qualified as personal data. Using "Sharemind", the user who initiates the process has to be considered as controller, whereas further users will not be considered as joint controllers in most of the cases. The final version of the GDPR is expected to be released at the end of 2015. It will have an enormous impact on the European data protection. Analysing the Regulation and its impact on cloud computing and encryption technologies will be the main task for D31.3.
In the Part II of the deliverable a methodology for the risk-aware deployment of secure computation is presented aiming to provide the analysis and the quantitative evaluation of risks related to privacy and confidentiality breaches during the execution of a multi-party business process on the cloud. The main contributions can be subdivided in four areas, regarding: (i) the methodology: We extend our previous approach including an iterative risk management process, where the risk assessor can estimate the risk profiles associated to subsets of potentially colluding actors; (ii) the possibilistic approach: We introduce in our methodology a novel approach for likelihood estimation, and discuss its use in alternative to the traditional probabilistic approach; (iii) an application to a case study: We apply our methodology in a real case exposed in the D24.2., modeling and analyzing the collaborative planning system for aircraft engine maintenance; (iv) a tool: We present an open source web based editor, enabling the representation and the simulation of the cloud based business process, along with the computation of the risk profiles associated to the actors.
In next year we will extend our methodology considering incentive and penalty schemes targeting specific class of actors. Those techniques will enable redressing the balance between confidentiality and efficiency to adapt it to the available computational resources over the cloud.

Bibliography

[1] Claudi Alsina, Berthold Schweizer, and Maurice J Frank. *Associative functions: triangular norms and copulas*. World Scientific, 2006.

[2] Marco Anisetti, Valerio Bellandi, Ernesto Damiani, Fulvio Frati, Gabriele Gianini, Gwanggil Jeon, and Jechang Jeong. Supply chain risk analysis: Open source simulator. In *Proceedings of the 2009 Fifth International Conference on Signal Image Technology and Internet Based Systems*, SITIS '09, pages 443–450, Washington, DC, USA, 2009. IEEE Computer Society.

[3] Marco Anisetti, Ernesto Damiani, Fulvio Frati, Stelvio Cimato, and Gabriele Gianini. Using incentive schemes to alleviate supply chain risks. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, MEDES '10, pages 221–228, New York, NY, USA, 2010. ACM.

[4] RobertJ. Aumann and RogerB. Myerson. Endogenous formation of links between players and of coalitions: An application of the shapley value. In Bhaskar Dutta and MatthewO. Jackson, editors, *Networks and Groups*, Studies in Economic Design, pages 207–220. Springer Berlin Heidelberg, 2003.

[5] Yoram Bachrach, Evangelos Markakis, Ezra Resnick, Ariel D Procaccia, Jeffrey S Rosenschein, and Amin Saberi. Approximating power indices: theoretical and empirical analysis. *Autonomous Agents and Multi-Agent Systems*, 20(2):105–122, 2010.

[6] Samik Basu and Tevfik Bultan. Choreography conformance via synchronizability. In Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors, *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, pages 795–804. ACM, 2011.

[7] Cédric Baudrit, Inés Couso, and Didier Dubois. Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *Int. J. Approx. Reasoning*, 45(1):82–105, May 2007.

[8] Gary S Becker. Crime and punishment: An economic approach. In *Essays in the Economics of Crime and Punishment*, pages 1–54. NBER, 1974.

[9] Mihir Bellare, Viet Tung Hoang, and Phillip Rogaway. Foundations of garbled circuits. In Ting Yu, George Danezis, and Virgil D. Gligor, editors, *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012*, pages 784–796. ACM, 2012.

[10] Muli Ben-Yehuda, Michael D. Day, Zvi Dubitzky, Michael Factor, Nadav Har'El, Abel Gordon, Anthony Liguori, Orit Wasserman, and Ben-Ami Yassour. The turtles project: Design and

implementation of nested virtualization. In *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*, OSDI'10, pages 1–6, Berkeley, CA, USA, 2010. USENIX Association.

[11] Taner Bilgiç and I Burhan Türkşen. Measurement of membership functions: theoretical and empirical work. In *Fundamentals of fuzzy sets*, pages 195–227. Springer, 2000.

[12] Dan Bogdanov, Liina Kamm, Sven Laur, and Pille Pruulmann-Vengerfeldt. Secure multi-party data analysis: end user validation and practical experiments. Cryptology ePrint Archive, Report 2013/826, 2013.

[13] G. Bortolan and R. Degani. A review of some methods for ranking fuzzy subsets. *Fuzzy Sets and Systems*, 15(1):1 – 19, 1985.

[14] Michael Burrows, Martin Abadi, and Roger Needham. A logic of authentication. *ACM Trans. Comput. Syst.*, 8(1):18–36, February 1990.

[15] Shakeel Butt, H. Andrés Lagar-Cavilla, Abhinav Srivastava, and Vinod Ganapathy. Self-service cloud computing. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 253–264, New York, NY, USA, 2012. ACM.

[16] P. Ceravolo, E. Damiani, D. Fasoli, and G. Gianini. Representing immaterial value in business model. In *Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2010 14th IEEE International*, pages 323–329, Oct 2010.

[17] SY Chan. An alternative approach to the modeling of probability distributions. *Risk Analysis*, 13(1):97–102, 1993.

[18] Pau Chen Cheng, Pankaj Rohatgi, Claudia Keser, Paul A Karger, Grant M Wagner, and Angela Schuett Reninger. Fuzzy multi-level security: An experiment on quantified risk-adaptive access control. In *Security and Privacy, 2007. SP'07. IEEE Symposium on*, pages 222–230. IEEE, 2007.

[19] Gustave Choquet. Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295. Institut Fourier, 1954.

[20] Shakhawat Chowdhury, Pascale Champagne, and P. James McLellan. Uncertainty characterization approaches for risk assessment of {DBPs} in drinking water: A review. *Journal of Environmental Management*, 90(5):1680 – 1691, 2009.

[21] CISCO. Data leakage worldwide white paper: The high cost of insider threats, 2011.

[22] Ins Couso, Didier Dubois, and Luciano Sanchez. *Random Sets and Random Fuzzy Sets As Ill-Perceived Random Variables: An Introduction for Ph.D. Students and Practitioners*. Springer Publishing Company, Incorporated, 2014.

[23] Cas J. Cremers. The scyther tool: Verification, falsification, and analysis of security protocols. In *Proceedings of the 20th International Conference on Computer Aided Verification*, CAV '08, pages 414–418, Berlin, Heidelberg, 2008. Springer-Verlag.

[24] Ernesto Damiani. Risk-aware collaborative processes. In *International Conference on Enterprise Information Systems (ICEIS)*. Institute for systems and technologies of information, control and communication, 2009.

[25] Ernesto Damiani, Fulvio Frati, and Romaric Tchokpon. The role of information sharing in supply chain management: The securescm approach. *International Journal of Innovation and Technology Management*, 08(03):455–467, 2011.

[26] Ernesto Damiani, Letizia Tanca, and Francesca Arcelli Fontana. Fuzzy xml queries via context-based choice of aggregations. *Kybernetika*, 36(6):635–655, 2000.

[27] Didier Dubois. *Fuzzy sets and systems: theory and applications*, volume 144. Academic press, 1980.

[28] Didier Dubois. Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. *Risk Analysis*, 30(3):361–368, 2010.

[29] Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.

[30] Didier Dubois and Henri Prade. Ranking fuzzy numbers in the setting of possibility theory. *Information Sciences*, 30(3):183 – 224, 1983.

[31] Didier Dubois and Henri Prade. Default reasoning and possibility theory. *Artif. Intell.*, 35(2):243–257, 1988.

[32] Didier Dubois and Henri Prade. Possibility theory in information fusion. In Giacomo Della Riccia, Hans-Joachim Lenz, and Rudolf Kruse, editors, *Data Fusion and Perception*, volume 431 of *International Centre for Mechanical Sciences*, pages 53–76. Springer Vienna, 2001.

[33] Didier Dubois and Henri Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142(1):143–161, 2004.

[34] Didier Dubois and Henri Prade. *Possibility theory and its applications: a retrospective and prospective view*. Springer, 2006.

[35] Didier Dubois and Henri Prade. Possibility theory. *Scholarpedia*, 2(10):2074, 2007.

[36] Erling Eide, Paul H Rubin, and Joanna Mehlop Shepherd. *Economics of crime*. Now Publishers Inc, 2006.

[37] F. Javier Thayer Fábrega. Strand spaces: Proving security protocols correct. *J. Comput. Secur.*, 7(2-3):191–230, March 1999.

[38] Ronald Fagin, Yoram Moses, Joseph Y Halpern, and Moshe Y Vardi. *Reasoning about knowledge*. MIT press, 2003.

[39] Shaheen S Fatima, Michael Wooldridge, and Nicholas R Jennings. A linear approximation method for the shapley value. *Artificial Intelligence*, 172(14):1673–1699, 2008.

[40] F Frati, E Damiani, P Ceravolo, S Cimato, C Fugazza, G Gianini, S Marrara, and O Scotti. Hazards in full-disclosure supply chains. In *Conference on Advanced Information Technologies for Management (AITM)*. Publishing house of the Wrocław University of economics, 2008.

[41] PAA Garcia, R Schirru, et al. A fuzzy data envelopment analysis approach for fmea. *Progress in Nuclear Energy*, 46(3):359–373, 2005.

[42] Heeralal Gargama and Sanjay Kumar Chaturvedi. Criticality assessment models for failure mode effects and criticality analysis using fuzzy logic. *Reliability, IEEE Transactions on*, 60(1):102–110, 2011.

[43] Craig Gentry, Shai Halevi, and Nigel P. Smart. Homomorphic evaluation of the AES circuit. In *Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings*, pages 850–867, 2012.

[44] Irina Georgescu. A possibilistic approach to risk aversion. *Soft Computing*, 15(4):795–801, 2010.

[45] Irina Georgescu. Computing the risk indicators in fuzzy systems. *Journal of Information Technology Research (JITR)*, 5(4):63–84, 2012.

[46] Irina Georgescu. Expected utility operators and possibilistic risk aversion. *Soft Computing*, 16(10):1671–1680, 2012.

[47] Irina Georgescu. *Possibility theory and the risk*, volume 274. Springer, 2012.

[48] Irina Georgescu and Jani Kinnunen. Multidimensional possibilistic risk aversion. *Mathematical and Computer Modelling*, 54(1):689–696, 2011.

[49] Irina Georgescu and Jani Kinnunen. Multidimensional risk aversion with mixed parameters. In *Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on*, pages 63–68. IEEE, 2011.

[50] Michel Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167 – 189, 1997. Fuzzy Measures and Integrals.

[51] Daniel JH Greenwood. Discussing corporate misbehavior. *Brooklyn Law Review*, 70:1213–1237, 2005.

[52] Anshuman Gupta and Costas D. Maranas. Managing demand uncertainty in supply chain planning. *Computers and Chemical Engineering*, 27(89):1219 – 1227, 2003. 2nd Pan American Workshop in Process Systems Engineering.

[53] MM Gupta and J Qi. Theory of t-norms and fuzzy inference methods. *Fuzzy sets and systems*, 40(3):431–450, 1991.

[54] Dominique Guyonnet, Gal Bellenfant, and Olivier Bouc. Soft methods for treating uncertainties: Applications in the field of environmental risks. In Didier Dubois, M.Asuncin Lubiano, Henri Prade, Marangeles Gil, Przemysaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, volume 48 of *Advances in Soft Computing*, pages 16–26. Springer Berlin Heidelberg, 2008.

[55] Mohammad Ali Hadavi, Ernesto Damiani, Rasool Jalili, Stelvio Cimato, and Zeinab Ganjei. AS5: A secure searchable secret sharing scheme for privacy preserving database outsourcing. In Roberto Di Pietro, Javier Herranz, Ernesto Damiani, and Radu State, editors, *Data Privacy Management and Autonomous Spontaneous Security, 7th International Workshop, DPM 2012, and 5th International Workshop, SETOP 2012, Pisa, Italy, September 13-14, 2012. Revised Selected Papers*, volume 7731 of *Lecture Notes in Computer Science*, pages 201–216. Springer, 2012.

[56] Timothy Hanratty, II Hammell, RobertJ., and Eric Heilman. A fuzzy-based approach to the value of information in complex military environments. In Salem Benferhat and John Grant, editors, *Scalable Uncertainty Management*, volume 6929 of *Lecture Notes in Computer Science*, pages 539–546. Springer Berlin Heidelberg, 2011.

[57] Li-Ping He, Hong-Zhong Huang, Li Du, Xu-Dong Zhang, and Qiang Miao. A review of possibilistic approaches to reliability analysis and optimization in engineering design. In JulieA. Jacko, editor, *Human-Computer Interaction. HCI Applications and Services*, volume 4553 of *Lecture Notes in Computer Science*, pages 1075–1084. Springer Berlin Heidelberg, 2007.

[58] Liping He, Jian Xiao, Hong-Zhong Huang, and Zhiqiang Luo. System reliability modeling and analysis in the possibility context. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on*, pages 361–367. IEEE, 2012.

[59] Ulrich Höhle. Probabilistic uniformization of fuzzy topologies. *Fuzzy Sets and Systems*, 1(4):311–332, 1978.

[60] T. Hoomans, J. Seidenfeld, A. Basu, and D. Meltzer. Systematizing the use of value of information analysis in prioritizing systematic reviews. Technical Report 12-EHC109-EF, Agency for Healthcare Research and Quality, 2012.

[61] Ronald A. Howard. Information value theory. *IEEE Trans. Systems Science and Cybernetics*, 2(1):22–26, 1966.

[62] Hong-Zhong Huang, Xin Tong, and Ming J Zuo. Posbist fault tree analysis of coherent systems. *Reliability Engineering and System Safety*, 84(2):141–148, 2004.

[63] Aaron Hunter and James P. Delgrande. Belief change and cryptographic protocol verification. In *Proceedings of the 22Nd National Conference on Artificial Intelligence - Volume 1*, AAAI'07, pages 427–433. AAAI Press, 2007.

[64] Iman Karimi and Eyke Hllermeier. Risk assessment system of natural hazards: A new approach based on fuzzy probability. *Fuzzy Sets and Systems*, 158(9):987 – 999, 2007. Selected papers from {IFSA} 2005 11th World Congress of International Fuzzy Systems Association.

[65] Arnold Kaufmann, Madan M Gupta, and A Kaufmann. *Introduction to fuzzy arithmetic: theory and applications*. Van Nostrand Reinhold Company New York, 1985.

[66] Eric Keller, Jakub Szefer, Jennifer Rexford, and Ruby B. Lee. Nohype: Virtualized cloud infrastructure without the virtualization. *SIGARCH Comput. Archit. News*, 38(3):350–361, June 2010.

[67] Florian Kerschbaum, Richard Pibernik, Ernesto Damiani, and Gabriele Gianini. Toward value-based control of knowledge sharing in networked services design. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (85):51–65, 2009.

[68] George J Klir. *Uncertainty and information: foundations of generalized information theory*. John Wiley & Sons, 2005.

[69] Apurva Kumar. A belief logic for analyzing security of web protocols. In *Proceedings of the 5th International Conference on Trust and Trustworthy Computing*, TRUST'12, pages 239–254, Berlin, Heidelberg, 2012. Springer-Verlag.

[70] Baoding Liu. Uncertain risk analysis and uncertain reliability analysis. *Journal of Uncertain Systems*, 4(3):163–170, 2010.

[71] YH Liu and Minghu Ha. Expected value of function of uncertain variables. *Journal of Uncertain Systems*, 4(3):181–186, 2010.

[72] Pasit Lorterapong and Osama Moselhi. Project-network analysis using fuzzy sets theory. *Journal of Construction Engineering and Management*, 122(4):308–318, 1996.

[73] Stephen G MacDonell, Andrew R Gray, and James M Calvert. Fulsome: A fuzzy logic modeling tool for software metricians. In *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American*, pages 263–267. IEEE, 1999.

[74] Sasan Maleki, Long Tran-Thanh, Greg Hines, Talal Rahwan, and Alex Rogers. Bounding the estimation error of sampling-based shapley value approximation. *arXiv preprint arXiv:1306.4265*, 2013.

[75] Irwin Mann and Lloyd S Shapley. Values of large games - iv. 1960.

[76] J.-L. Marichal. An axiomatic approach of the discrete choquet integral as a tool to aggregate interacting criteria. *Fuzzy Systems, IEEE Transactions on*, 8(6):800–807, Dec 2000.

[77] Jean-Luc Marichal. Tolerant or intolerant character of interacting criteria in aggregation by the choquet integral. *European Journal of Operational Research*, 155(3):771 – 791, 2004. Traffic and Transportation Systems Analysis.

[78] Fabio Massacci. Automated reasoning and the verification of security protocols. In *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, TABLEAUX '99, pages 32–33, London, UK, UK, 1999. Springer-Verlag.

[79] PEDRO MIRANDA, MICHEL GRABISCH, and PEDRO GIL. p-symmetric fuzzy measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(supp01):105–123, 2002.

[80] Sherif Mohamed and Alison K McCowan. Modelling project investment decisions under uncertainty using possibility theory. *International Journal of Project Management*, 19(4):231–241, 2001.

[81] C Negoita, L Zadeh, and H Zimmermann. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.

[82] Hung T Nguyen. A note on the extension principle for fuzzy sets. *Journal of Mathematical Analysis and Applications*, 64(2):369–380, 1978.

[83] Anthony O'Hagan and Jeremy E. Oakley. Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering and System Safety*, 85(13):239 – 248, 2004. Alternative Representations of Epistemic Uncertainty.

[84] Nicola Pedroni and Enrico Zio. Empirical comparison of methods for the hierarchical propagation of hybrid uncertainty in risk assessment, in presence of dependences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20(04):509–557, 2012.

[85] Ronald L Rivest, Len Adleman, and Michael L Dertouzos. On data banks and privacy homomorphisms. *Foundations of secure computation*, 4(11):169–180, 1978.

[86] August-Wilhelm Scheer and Markus Nüttgens. ARIS architecture and reference models for business process management. In Wil M. P. van der Aalst, Jörg Desel, and Andreas Oberweis, editors, *Business Process Management, Models, Techniques, and Empirical Studies*, volume 1806 of *Lecture Notes in Computer Science*, pages 376–389. Springer, 2000.

[87] Udo Steinberg and Bernhard Kauer. Nova: A microhypervisor-based secure virtualization architecture. In *Proceedings of the 5th European Conference on Computer Systems*, EuroSys '10, pages 209–222, New York, NY, USA, 2010. ACM.

[88] Paul F. Syverson. Knowledge, belief, and semantics in the analysis of cryptographic protocols. *J. Comput. Secur.*, 1(3-4):317–334, May 1992.

[89] Paul F. Syverson. Adding time to a logic of authentication. In *Proceedings of the 1st ACM Conference on Computer and Communications Security*, CCS '93, pages 97–101, New York, NY, USA, 1993. ACM.

[90] CM Tam and I Fung. Assessing safety performance by fuzzy reasoning. *Asia Pacific Building and Construction Management Journal*, 2(1):6–13, 1996.

[91] H Tanaka, H Ichihashi, and K Asai. A value of information in flp problems via sensitivity analysis. *Fuzzy Sets and Systems*, 18(2):119–129, 1986.

[92] Johan René van Dorp, Salvador Cruz Rambaud, José García Pérez, and Rafael Herrerías Pleguezuelo. An elicitation procedure for the generalized trapezoidal distribution with a uniform central stage. *Decision Analysis*, 4(3):156–166, 2007.

[93] Vic Winkler. Cloud computing: Risk assessment for the cloud. *Technet Magazine*, January 2012.

[94] KC Wong and Albert TP So. A fuzzy expert system for contract decision making. *Construction Management and Economics*, 13(2):95–103, 1995.

[95] Li Yang, Fushuan Wen, FF Wu, Yixin Ni, and Jiaju Qiu. Development of bidding strategies in electricity markets using possibility theory. In *Power System Technology, 2002. Proceedings. PowerCon 2002. International Conference on*, volume 1, pages 182–187. IEEE, 2002.

[96] Zaili Yang, Stephen Bonsall, and Jin Wang. Fuzzy rule-based bayesian reasoning approach for prioritization of failures in fmea. *Reliability, IEEE Transactions on*, 57(3):517–528, 2008.

[97] Andrew Chi-Chih Yao. Protocols for secure computations. In *FOCS*, volume 82, pages 160–164, 1982.

[98] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[99] Fengzhe Zhang, Jin Chen, Haibo Chen, and Binyu Zang. Cloudvisor: Retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, SOSP '11, pages 203–216, New York, NY, USA, 2011. ACM.