The ISC Int'l Journal of Information Security
ISeCure

July 2014, Volume 6, Number 2 (pp. 99–123)
http://www.isecure-journal.org

Invited Paper

A Risk Model for Cloud Processes

Ernesto Damiani [1,*], Stelvio Cimato [1], and Gabriele Gianini [1]
[1] *Department of Computer Science, Università degli Studi di Milano*

ARTICLE INFO.

Article history:
Received: 23 October 2014
Revised: 24 January 2015
Accepted: 30 January 2015
Published Online: 2 February 2015

Keywords:
Cloud Computing, Value of Information, Risk Assessment, Secure Computation.

ABSTRACT

Traditionally, risk assessment consists of evaluating the probability of "feared events", corresponding to known threats and attacks, as well as these events' *severity*, corresponding to their impact on one or more stakeholders. Assessing risks of cloud-based processes is particularly difficult due to lack of historical data on attacks, which has prevented frequency-based identification of "typical" threats and attack vectors. Also, the dynamic, multi-party nature of cloud-based processes makes severity assessment very dependent on the particular set of stakeholders involved in each process execution. In this paper, we tackle these problems by presenting a novel, *process-oriented* quantitative risk assessment methodology aimed at disclosure risks on cloud computing platforms. Key advantages of our methodology include (i) a fully quantitative and iterative approach, which enables stakeholders to compare alternative versions of cloud-based processes (e.g., with and without security controls) (ii) non-frequency-based probability estimates, which allow analyzing threats for which a detailed history is not available (iii) support for quick visual comparisons of risk profiles of alternative processes even when impact cannot be exactly quantified.

1 Introduction

As it often happens when new technologies are introduced, individuals and organizations wishing to adopt the cloud computing paradigm need to carefully consider all associated risks. Indeed, cloud-computing platforms have many risks in common with externally provided (outsourced) services, but also some specific features that require ad-hoc risk evaluation methods. From the economic standpoint, the *risk* of a "feared event" E for a given actor A is often represented as the product of the probability that E might happen, times the damage (expressed in currency units) to A if E really happened. In symbols:

$$R(A, E) = Pr(E) \times I_A(E) \qquad (1)$$

In the computer security context, one needs to identify feared events E as manifestations of security threats and estimate $I_A(E)$ and $Pr(E)$. As far as threat identification is concerned, it is often convenient to analyze a business process at a time. The risk analyst puts herself in the place of a specific actor (e.g. the *process owner*, i.e. the stakeholder in whose name or interest a business process P is executed) and asks

[*] Corresponding author.
Email addresses: ernesto.damiani@unimi.it (E. Damiani), stelvio.cimato@unimi.it (S. Cimato), gabriele.gianini@unimi.it (G. Gianini)
ISSN: 2008-2045 © 2014 ISC. All rights reserved.

the following questions [65]:

- *Threat Categorization*: Which unfortunate event can happen to the information assets involved in P?
- *Threat Impact*: How severe could that event be for the process owner?
- *(Frequentistic) Threat Probability*: How often might this event happen?

Accurately quantifying threat impact is often a challenge, as losses deriving, say, from decreased consumer trust after a security breach can only be estimated in the long run. As far as probability is concerned, there are cases where a frequency-based probability can be assigned to the feared events, and other cases where this is too difficult or misleading. Whatever the impact and probability assessment methodology, however, risk analysis has traditionally focused on composing (via a suitable aggregator, in the simplest case a summation) risks $R(A, E_i)$, $i = 1, \ldots, n$ for all known feared events that may affect a specific actor A.

An important category of feared events are disclosures of personal and private data. Business processes often involve storing or transmitting personal data that is subject to strict regulatory and compliance requirements. The choice of deploying such processes on a cloud hinges on the process owner being convinced that the cloud provider is fully compliant with regulations. Otherwise, the process owner risks liability for violating privacy, regulatory or other legal requirements. If a highly regulated business process (e.g. a e-health or e-government one) is to take place on a public cloud, then its deployment must fully meet all applicable regulations and laws regarding data confidentiality and leakage prevention.

Classic quantitative approaches to risk assessments proposed in the late Seventies are based on estimating threat probabilities as frequencies using available statistical information. As early as 1979, the US National Bureau of Standards (later absorbed into NIST) published its Federal Information Processing Standard (FIPS) 65, Guideline for Automatic Data Processing Risk Analysis [45], introducing the risk assessment standard for large data-processing centers and proposing a new metric for computer-related risks: the Annual Loss Expectancy (ALE), computed as follows

$$ALE = \sum_{i=1}^{n} I(O_i) \cdot F_i \tag{2}$$

where $\{O_1, \ldots, O_N\}$ is the Set of Harmful Outcomes, $I(O_i)$ represents the Impact of Outcome i in dollars, and F_i is the Frequency of Outcome i.

In more recent years, risk assessment methodologies have become a standard practice aimed to let organizations determine and demonstrate their privacy, security, and compliance with other policies. Most methodologies require some steps to be carried out, including system characterization, threat assessment, vulnerability analysis, impact analysis, and risk determination [15, 35].

There are three major categories of risk assessment methods. *Qualitative* methods are based on applying simple criteria to evaluate threats' severity, without separately assessing the value of the assets at stake and the threats' frequencies. In turn, *quantitative* methods are based on obtaining numerical estimates for the likelihood and impact of feared events. Finally, *semi-quantitative* (or hybrid) methods assess likelihood and severity of feared events separately, but prioritize risks according to threats' perceived consequences and users' belief in (rather than probability of) their taking place. In our methodology, we do not adopt a frequency-based probability estimate; rather, we rely on the knowledge of the business process model and its underlying micro-economics to attach probabilities to actors misbehavior/violation of confidentiality and to provide an evaluation of costs taking into account the value of disclosed information.

In this paper 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 would entail. These 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.

We present a novel, *process-oriented* quantitative risk assessment methodology aimed at assessing disclosure risks on cloud computing platforms. Key advantages of our methodology include (i) a fully quantitative and iterative approach, which enables stakeholders to compare alternative versions of cloud-based processes (e.g., with and without security controls) (ii) non-frequency-based probability estimates, which allow analyzing threats for which a detailed history is not available (iii) support for quick visual comparisons of risk profiles of alternative processes even when impact cannot be exactly quantified.

The paper is organized as follows: we start by surveying the state of the art (Section 2) and go on by introducing our notion of business process model (Section 3). Then we describe the basic concepts of our risk analysis methodology (Section 4) and detail the steps for its application to analyze risks of process-related data disclosures (Section 5). We specialize the methodology to multi-party processes taking place on clouds, adapting the model to describe cloud-based computations (Section 6), and showing our technique for assessing the impact and probability of threats in cloud-based processes (Section 7). Finally, we consider auctions as possible scenarios for the application of our methodology (Section 8), before drawing our conclusions (Section 9).

2 Related Work

Cloud computing is a computing paradigm where "massively scalable IT-enabled capabilities are delivered 'as a service' to external customers using Internet technologies" [29]. If on the one hand the adoption of such a model provides cost savings through economies of scale, on the other hand it introduces some peculiar risk challenges that increase the risks traditionally introduced by any externally provided IT service. Indeed, from the perspective of the security analyst, cloud-based services are outsourced in the least transparent way, since data are stored and processed on unspecified servers located in some unknown places, out of the control of the data owner. For these reasons, some researchers have started introducing cloud-oriented techniques to deal with specific cloud-related issues [21, 24, 51].

Various bodies such as the Cloud Security Alliance (CSA), the European Network and Information Security Agency (ENISA), and the US National Institute of Standards and Technology (NIST) have released documents assisting organizations and customers in the evaluation of the security issues related to cloud computing [12, 17, 46]. The Cloud Controls Matrix released by CSA provides an useful description of the security principles aiming to guide cloud vendors and help cloud clients in assessing overall security risks of a cloud service provider [17]. NIST Special Publication 800-144 provides an overview of the security and privacy challenges for public cloud computing and gives recommendations that organizations should consider when outsourcing data, applications, and infrastructure to a public cloud environment.

In an early study by ENISA [12], a cloud-specific, semi-qualitative risk assessment process was anecdotally described using three use-cases: the SMEs' perspective on cloud computing, the impact of cloud computing on service resilience, and a scenario on cloud-based e-Health applications. The ENISA study included a table showing the distribution of feared events' probabilities and impacts adopting a scale of 0 to 8, and classifying low risk values from 0 to 2, medium risk from 3 to 5 and high ones from 6 to 8. Feared events were classified into three categories: policy and organizational, technical, and legal. Within the first category, lock in, loss of governance, and compliance challenges were mentioned as having in some cases very high impact. Namely, *lock in* refers to the way tools, procedures and standard data formats or services interfaces are provided on the cloud; if portability is not guaranteed, the migration from one provider to another may be extremely difficulty for a customer. This could be the cause for a business failure should the cloud provider go bankrupt or be acquired by another company. *Loss of governance* over data and services can have a potentially severe impact on any organization's mission, leading to the impossibility of satisfying requirements about confidentiality, integrity and availability of data. *Compliance* problems may arise due to the migration of services to the cloud, since it is difficult for the cloud providers to provide evidence of meeting industry standards or regulatory requirements. Other feared events mentioned in the study include *loss of business reputation* due to co-tenant activities, and *unwanted disclosure of information* to co-tenants. The former event is linked to threats of malicious activities on the part of co-tenants that may affect the reputation of the other customers who are using the same cloud infrastructure. The latter event may be due to failure of mechanisms separating storage, memory and routing between different tenants of the shared infrastructure caused by different kind of threats, such as *guest-hopping attacks*, or SQL injection attacks exposing multiple customers' data stored in the same table.

A recent whitepaper [2] describes a qualitative risk assessment methodology specific for clouds. It starts by considering risk factors that change when an organization shifts from a traditional infrastructure to a cloud-based one. The analysis is based on the risk taxonomy presented by the Open Group [27]. The transition to a cloud infrastructure may change the probability of the occurrence of a harmful event, reducing the effort necessary to carry on an attack when a cloud specific vulnerability can be exploited. To denote a threat as cloud specific, four indicators are proposed. A first category collects all threats that are intrinsic to cloud computing, such as the possibility for an attacker to escape from the virtualized environment, the possibility to ride or hijack sessions in shared web applications, threats to the integrity and confidentiality of data caused by the insecure usage of cryptography or the selection of flawed implementation of cryptographic primitives. Other specific threats are the ones concerning problems with standard security controls, such as the difficulties to execute network security controls in virtualized environment, poor management or storage of the of encryption keys, the difficulty of establishing security metrics suitable to monitor the security status of cloud resources.

Another interesting case study showing qualitative risk assessment at work in a cloud computing scenario is described in [26], where the case of a software company developing business software and adopting a IAAS provided by another CSP is analyzed. The methodology is based on the *Risk IT* framework, which provides a detailed process model for the management of IT-related risk, as well as on the *COBIT 5* framework by the Information Systems Audit and Control Association (ISACA) [33], which assists enterprises in achieving their objectives for the governance and management of enterprise information technology (IT). RISK IT includes a list of generic high-level risk scenarios and a mapping between those scenarios and more general COBIT control objectives, so that a map of risks showing the impact/magnitude and likelihood/frequency of key risks can be created. Based on this map and on the prioritization of risks, a risk mitigation approach can be adopted, balancing the benefit from ensuring controls and the costs necessary for their implementation.

Some initial work toward a quantitative risk and impact assessment framework for cloud computing, called QUIRC, has been presented in [52]. The QUIRC framework classically defines risk as a combination of (a rough estimate of) the probability of a feared event and its severity, measured as its impact. QUIRC lists six key *Security Objectives* (SO) for cloud platforms, claiming that most of the typical attack vectors and feared events map to one of these six categories. QUIRC's strong point is its fully quantitative approach, which enables stakeholders to comparatively assess the robustness of cloud vendor offerings. However, lack of reliable data on the occurrences of cloud threats in many vertical domains can make QUIRC probability assignment (and the entire notion of "typical" attack vectors) somewhat arbitrary.

Another quantitative framework for assessing some security risks associated with cloud computing platforms has been proposed in [55]. The model relies on a fuzzy decision making technique, that allows the definition of the weights of the coefficients for the basic security properties (CIA - Confidentiality, Integrity, and Availability) and the corresponding values of assets relevant for the project, using the knowledge of experts. Then, vulnerability indices are defined for each asset separately and a final fuzzy model is created to compute the impact of each identified risk as product of asset values, vulnerability and threat effects. Even if the resulting prioritization of the risks is valuable, this approach only considers threats to CIA properties. Also, it relies on subjective assessments of likelihood and severity by experts that may be difficult to replicate in practice. Focusing on the same small set of security properties, Khan *et al.* [36] introduced a more systematic approach combining existing tools and techniques such as CORAS [19], and the IRAM (Information Risk Analysis Methodology) with the Threat and Vulnerability Assessment tool (T&VA) [1]. Their technique uses a list of threats provided by the Information Security Forum (ISF). Depending on the priority of the assets and on the perceived likelihood of the ISF threats, they construct an evaluation matrix and use it to rate the threats' impact on the business. Due to the anecdotal nature of the ISF threat list, whose entries often highlight new and emerging threats rather than frequent ones, this technique can be considered a semi-quantitative one.

2.1 Integration of Disclosure Risk Assessment with Privacy Risk Management Frameworks

Business processes involving personal data present specific risks due to the liability brought upon the process owner (often called *controller* in this context) and, possibly, upon other stakeholders by violations of the privacy of third parties (*data subjects*).

A special regulatory framework for personal data processing is currently in force at the European level, prescribing - among other things - that the purposes of the business process involving personal data are clearly defined, that personal data are relevant to such purposes, that personal data are erased at the end of a given time, and that all data subjects have the opportunity to exercise their rights (such as opposition, access, rectification and deletion of their personal data). In addition, the controller has an obligation to take all useful precautions in order to ensure the security of the personal data he processes.

Privacy authorities and regulators have been devoting a huge effort to develop *Privacy Risk Management* (PRM) frameworks [13]. As observed in [39], we still lack of a systematic approach to identify privacy threats and design privacy-supportive business processes. According to recent studies [66], privacy threat analysis should begin at the earliest possible stage of the lifecycle of any business process involving personal data, when there are more opportunities to influence the business process' implementation; also, it should continue along the business process lifecycle.

In principle, privacy risks may be targeted by all forms of risk analysis introduced in previous Sections. *Qualitative Analysis* uses ordinal scales expressed in words to quickly assess the relative severity of risks.

This technique is often used when numerical data is not available and/or the process targeted by privacy risk assessment is only partially known by the risk assessor, as is often the case at early design stage. [1] .

Semi-Quantitative Analysis adds a score expressed in points to the ordinal scale (e.g. “1” = Low; “5” = Very High) [2] . Finally, *Quantitative Analysis* computes real numeric values to assess impacts (expressed in monetary terms) and probabilities of privacy threats.

The quantitative approach is generally more complex to undertake, requiring full knowledge of the business processes to be analyzed and, in many cases, the development of organization-specific value models to assess the value of disclosed information as seen by different actors.

3 Process Model

Let us now formalize our notion of business process model. It is important to remark that the aim of our model is enabling risk assessment; so the process representation will focus on risk-related rather than design-related aspects. 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. [3] . In our view, security controls (when present) are an integral part of the business process definition. In order to be able to represent a complete set of security controls, message exchange in our process model is a general *timestamped choreography* [4] 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. [4]
- *Local computations* $(A_i, f(), INFO_{i,ts})$ i.e. functions computed by actors on (portions of) locally held information at a given time.

Figure 1 shows a sample process:

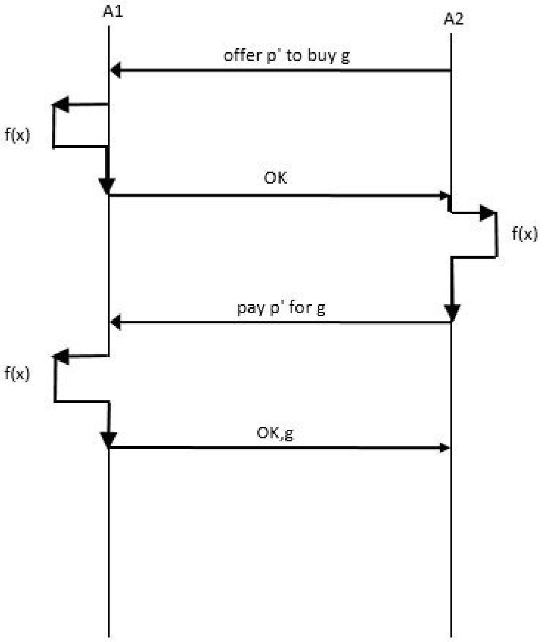

Figure 1. A sample process model

3.1 Process Model Assumptions

Our process model is completed by some additional assumptions. Here β denotes the probability of an event, however assessed (for our own probability assessment technique see Section 4.2):

- *Protocol efficacy*: Given a message delivery (A_s, A_d, m_{ts}), with $m_{ts} = INFO_s$ then $\beta_i(E_{sd}) = MAX$ for all actors A_i.
- *Information completeness*: Given a message delivery (A_s, A_d, m_{ts}), with $m_{ts} \leq INFO_s$, then $\beta_i(E_{sd}) = 0$
- *Strong local computation transparency*: Given a local computation $(A_i, f(), INFO_{i,ts})$, then $INFO_i = INFO_{i,ts} \cup f(INFO_{i,ts} \cup S_f)$ for $t \geq ts$, where S_f is the *specification* of f as an algorithm or a closed formula.
- *Belief propagation*: Given a message delivery (A_s, A_d, m_{ts}), then for $t \geq ts$, $\beta_i(E_{sk}) = \beta_i(C(A_d, A_k))$ for $k \neq d$, where $C(A_d, A_k)$ denotes the event of information sharing between A_d and A_k.

It is important to remark that the *local computation transparency* assumption implies that any actor computing a function $f()$ over its local data becomes aware of the results of that function as well as of its *specification* S_f, represented e.g. as a computer program. However, research has shown that this assumption may be weakened by *obfuscation* or *garbling* techniques [5].

[1] The Office of the Privacy Commissioner of Canada has provided support for early-stage qualitative analysis of privacy risks in its PIPEDA Self-Assessment Tool, `http://www.priv.gc.ca/information/pub/ar-vr/pipeda_sa_tool_200807_e.cfm`.

[2] This approach was first used in the US by the American Institute of Certified Public Accountants (AICPA) in their “Privacy Risk Assessment Tool”

[3] Of course, the impact of disclosing an empty item is always 0.

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

3.2 Garbling Outsourcing Scheme

Garbled circuits, a classical idea rooted in early work by Andrew Yao, are a well-known example of obfuscation techniques. Here, we follow the literature [5] to briefly describe a *garbling outsourcing scheme* corresponding to Yao's garbling technique. The purpose of our simplified description is to show how *obfuscation is represented within our process model.*

Let us assume Alice wants Bob to compute for her a function $f()$ on a set of inputs, some of which are held by herself and others by Bob, without sharing with Bob the function specification S_f . At a high level of abstraction, the scheme works as follows: Alice creates a "garbled circuit", i.e. the specification S'_f of a garbled function $f'()$ having the same input-output table as $f()$, and sends it to Bob. Bob uses S'_f to build $f'()$, compute it with his inputs B and returns the result to Alice. The result of $f'(B, x)$ evaluation with $x = A$ (where A is Alice's inputs) coincides with the function $f()$ that Alice wanted Bob to compute; but by computing $f'()$, Bob has learnt nothing about S_f. Note that in this scheme Alice does not send her inputs to Bob; rather, her inputs are encoded into the "garbled circuit" in such a way that Bob can not determine what they are. As an example, assume that Bob has $x = 2$ bits, (a, b), and Alice has $y = 2$ bits, (c, d). The function $f()$ is:

$$f(x, y) = (a + c) \vee (b + d) \qquad (3)$$

For the construction of the garbled circuit, one simply constructs a new truth table for each gate in the original circuit. A sample truth table for an AND gate is shown below (Table 1), with inputs p, q and output z. Alice picks two random keys for each wire and obtains the garbled truth table by encrypting the output-wire key with the corresponding pair of input-wire keys

Table 1. Garbled computation for an AND gate

input1	input2	output	garbled computation
k_q^0	k_p^0	k_z^0	$E_{k_q^0}(E_{k_p^0}(k_z^0))$
k_q^0	k_p^1	k_z^0	$E_{k_q^0}(E_{k_p^1}(k_z^0))$
k_q^1	k_p^0	k_z^0	$E_{k_q^1}(E_{k_p^0}(k_z^0))$
k_q^1	k_p^1	k_z^1	$E_{k_q^1}(E_{k_p^1}(k_z^1))$

After Bob has received the garbled specification $S_{f'}$ and the corresponding truth tables, he still needs Alice's inputs before he can evaluate the function. Bob can get these inputs using a 1-out-of-2 instantiation of Rabin's *oblivious transfer* protocol. [5]

Once Bob has received the input values from Alice via the oblivious transfer protocol, he can "decrypt" each of the gates, and using his own inputs he can evaluate the circuit. Today, efficient garbling schemes are available achieving privacy as well as obliviousness and authenticity, the latter properties being needed for private and verifiable outsourcing of computation. Highly efficient block-cipher-based instantiations of garbling schemes have been described in the literature.

For our purposes, it is sufficient to observe that when a garbling outsourcing scheme is in force within a process, a weaker assumption (*weak local computation transparency*) can be adopted for our business process model, where the party executing a local computation $f()$ learns the output of the function, but not its specification.

More formally, let us consider a process P including a local computation $(A_i, f(), INFO_{i,ts})$.

Let $G()$ be a functional acting on the $f()$ function specification S_f, so that

$$G(S_f) = S_{f'} \qquad (4)$$

We call $G(S_f) = S_{f'}$ a *garbled specification* of $f()$ if and only if $f'(x) = f(x)$ for all inputs x and:

$$t \geq ts : (A_i, f(), INFO_{i,ts}) \rightarrow$$
$$INFO_i = INFO_{i,ts} \cup f'(INFO_{i,ts}) \qquad (5)$$

It is important to remark that the computation of $G(S_f) = S'_f$ *can itself be a local computation of the process* P. This way, any actor can outsource a local computation to another actor, who will compute the garbled function under our weak transparency assumption.

4 A Methodology for Quantitative Assessment of Risk in Cloud Process Execution

In this section we describe some basic concepts behind our quantitative risk assessment approach, namely the identification of the feared events and threats, the estimation of threats' probabilities and of their impacts.

4.1 The Threat Space: Disclosure Events

Any risk model must clearly specify the event space where risk will be quantified. Indeed, a well-known problem of applying general-purpose risk assessment frameworks based on equation 1 is the effort required by threat assessment, since each practical scenario taken into consideration introduces new families of threats. In an effort to be comprehensive, risk model-

[5] In an oblivious transfer protocol, a sender transfers one of potentially many pieces of information to a receiver, but remains oblivious as to what piece (if any) has actually been transferred [49].

ers have been tempted to try to capture all threats, assets, vulnerabilities, and security concerns. Unfortunately, trying to be exhaustive can put risk analysis beyond the capabilities of both personnel and computing resources. Indeed, highly expensive assessment was one of the pivotal reasons why early risk models failed to achieve widespread acceptance [59].

In this paper 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.*

A major feared event happens when actors (including service and cloud providers) put together the partial information they hold to reconstruct knowledge that is not available to them when taken individually. We remark that this feared event is not caused uniquely by collusion among rogue participants. Indeed, different parties may put together their information for other reasons, including

- eDisclosure, i.e. the mandatory process of disclosing information to adversaries during litigation [6]
- An information request from a regulatory authority [7]
- Inadvertent or dysfunctional behavior of employees.

For the first factor, 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, e-mails containing bids for a auction held in one country may be stored on a server located in another jurisdiction, where a regulatory authority can ask the service provider - for reasons unrelated to the auction - 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 [16] polled more than 2000 employees and information technology professionals in 10 countries, including major EU markets. The study identified the feared event of unwanted information sharing, related to sloppy implementations of interchange protocols, or intentional communication with unauthorized parties. For instance, a plaintext 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. The methodology and models presented in the next sections are aimed at filling this gap.

4.2 The Probability Model

When performing a quantitative risk assessment, a key activity is the estimation of the uncertainty present in a variable. A major problem for the practical application of our risk assessment methodology is that the needed Probability Mass Functions (PMFs) and Probability Density Functions (PDFs) are not readily available, and have to be derived from available information and knowledge. Several methods have been proposed in the literature for the derivation of PMFs and PDFs in risk assessment.

The choice of an appropriate method depends on what information and knowledge is available. In this section, we will focus on choice of PMFs, i.e. on the discrete variable case, since it can be readily applied to events such as the realization of threats, in the case where an adverse event either occurs or does not occur at all. In this binary case the appropriate family of distributions among which the PMF has to be chosen is the Bernoulli family, characterized by a single parameter: the probability that the adverse event corresponding to the threat T actually takes place $Pr(T)$.

In the remainder of the Section we will discuss how to find the probability of the threat T corresponding to the event in which a subsets of actors in 2^A *colludes* (the actors put together the information they know). Our method does not rely on frequency-based probability estimates but on the elicitation of expert opinion.

Our technique is based on the notion that the dysfunctional behavior of actors taking part in a business process is often due the *unfairness of the redistribution of payoffs* in the process (like a benefit allocation structure that responds to organization efficiency more than to fairness). We will elicit the probability

[6] http://www.ediscloureinformation.co.uk

[7] Whether data is on premises or in the cloud, the obligation to comply with the demands of the court or regulatory authorities remains essentially the same.

$Pr(T)$ based on the opinion of experts to whom it has been provided as input parameter, the degree ϕ of perceived unfairness in the business process resource allocation computed on the basis of the economic tools taken from coalitional game theory.

Indeed, even a process configuration yielding the highest total surplus does not necessarily guarantee a fair distribution of this surplus. Efficiency says nothing about equity or fairness, i.e. *distributive justice*.

The problem of how profits of a coalition should be redistributed is a well-known one (it is an instance of the general problem of distributive justice). There are several solutions to the problem. Due to the subjectivity of satisfaction criteria for each agent an objectively optimal solution cannot in general be attained, however, a solution fulfilling some largely accepted requirement can be obtained by following the prescription dictated by the so-called *Shapley Value*[3]: 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 a hyperplane in an N-dimensional space. The Shapley value can be computed for the actors of an organization in order to find the fairness point and to compare it to the point representing the current allocation of the value in an organization. The distance between these two points can be related to an individual actor's probability of defection: the closer the two points, the more likely there will be no dysfunctional behavior on the part of that actor. Should the convenience become too low, the actor will be tempted to behave un-cooperatively. Such behavior may damage to the overall business process.

4.2.1 The Shapley Value

We now provide a formal definition of the Shapley value. Let us consider a general game with a set $\mathcal{N}$ participants. The Shapley Value is defined as an allocation of payoffs: a payoff u_i for each actor $i \in \mathcal{N}$. Any subset of players in N is a potential *coalition* C. A coalition can strike deals among its own members to exploit all the available knowledge for mutual advantage. Combinatorially, there are $(2^N - 1)$ possible coalitions altogether, including the so-called *grand coalition* consisting in $\mathcal{N}$ itself (and disregarding the empty set).
It is customary to call *security level* of a coalition C the quantity $s(C)$ expressing the total surplus that its members can achieve on their own even if the nonmembers took the action that was the worst from C's perspective. An allocation $(x_1, x_2, \ldots, x_N)$ is a list of amounts for the players (they are shares of a total value, and add up to the total added value), and it is said to be feasible if allowed by the rules of the game. A feasible allocation is blocked (i.e. not even considered) by a coalition C if $a(C) > \sum_{i \in C} x_i$, i.e. if the allocation values add up to an amount which is less the security level of the coalition.
We call *core* the set of allocations that cannot be blocked by any coalition: it contains all possible reasonable deals. The core can be a point, a range or a general set. For some games, it can even be empty. However, the core has some desirable properties. Since it cannot be reduced further by any groups searching for a better deal, including the grand coalition, it can be shown that it is *Pareto efficient*, i.e. that no allocation outside the core will improve everyone's payoff simultaneously. Note that even the core, being defined on the base of an inequality over a sum of the allocation array, does not give any guarantee over the distributive justice of an allocation: the elements of the core will all represent efficient allocations, but some will be fairer than others.
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 it makes to each of the possible coalitions underlying the process. In order to produce each coalition one has to run ideally over all the permutations of actors: each ideal ordering of the actors corresponds to a non-decreasing surplus value achieved by the members up to the considered index. When arriving at the actor i, whose Shapley Value is being computed, one has to take note of the added-value introduced him. The Shapley Value for an actor is then given by the average over all permutations of those added-values, which we denote by Δ_i:

$$u_i = \frac{1}{N!} \sum_{\pi} \Delta_i(\pi) \tag{6}$$

where the index π runs over all the permutations of N objects. Equation 6 can be rewritten in a more computable form by taking into account that, when scanning 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. Thus, the Shapley Value u_i can be computed as follows:

$$u_i = \sum_{C \subseteq \{\mathcal{N} \setminus i\}} \left[\frac{N!}{(|C|)!(N - |C| - 1)!}\right]^{-1} \times \Big(s\left(\{C \cup i\}\right) - s(C)\Big) \tag{7}$$

Here C represents the set of leading actors to which i brings his contribution as additional actor, $|C|$ is the cardinality of such a set, N the overall number of actors, while $\Delta_i(C) \equiv (s(\{C \cup i\}) - s(C))$ is the difference in security levels. The quantity in square brackets is a combinatorial factor that accounts for the fact that all the permutations have the same probability $1/N!$, and that for a given leading set C there are $(|C|)!$ equivalent orderings, while for the trailing set, consisting in $(N-|C|-1)$ elements, there are $(N-|C|-1)!$ equivalent orderings.

From Shapley value to feared event probability Given the Shapley value for each actor in a subset, we first define $\delta_i = u_i - x_i$ as the difference between the Shapley Value and the actual resource allocation for that actor, i.e. the benefit the actor expects from taking part to the process. If this difference is positive, it means that the actor is under-rewarded for his contribution and there may be a positive probability that this causes a defection; if, instead, it is negative, the actor is over-rewarded and this discrepancy will not contribute to its probability of defection; one needs also to relate the discrepancy, when positive, to the absolute value of u_i.
For the above considerations, the factor ϕ_i for an actor i can be defined as follows:

$$\phi_i \equiv \frac{\theta(u_i - x_i)}{u_i} \tag{8}$$

where $\theta(.)$, is a filter function defined for an argument $z \in R$ as:

$$\theta(z) \equiv \begin{cases} 0, & if \quad z < 0 \\ z, & otherwise \end{cases} \tag{9}$$

Elicitation of expert opinions. By itself, the above defined ϕ factor does not allow to compute the probability of each subset of actors potentially colluding to share the information they hold. However, based on *expert evaluation*, each value of ϕ can be mapped into a value of probability for an individual actor's defection (within a given context, denoted hereafter by the exponent c): for each value of ϕ, the probability distribution modeling the occurrence of the defection event will be a Bernoulli distribution characterized by the expert-provided value of the probability parameter p^c. In practice an expert can give its estimate of the dependence of this parameter from the factor ϕ – i.e. can provide its opinion about the function $p^c(\phi)$ – which takes into account not only the most plausible constraints, but also contextual conditions difficult to model mathematically. Among the plausible constraints are $p^c(\phi = 0) = 0$, the non-decreasing nature of the function $p^c(\phi)$ and the likely saturation behaviour ($p^c \to 1$ as $\phi \to 1$), which, altogether, give to the function a sigmoidal shape.

This technique is an instance of the well-known problem of eliciting experts' knowledge given the value of different context parameters. In order to elicit the shape of the function $p^c(\phi)$ our methodology uses the Bezier curves. Those models have often been used in computer graphics [28] to approximate a smooth (continuously differentiable) function on a bounded interval, by forcing the Bezier curve to pass in the vicinity of selected control points in two-dimensional Euclidean space. Indeed, using a Bezier curve, one can approximate a function to an arbitrary level of detail by taking a sufficiently high number of control points, with appropriate values for the coordinates. Sometimes those curves have also been used to represent expert opinion in terms of an univariate or bivariate non-parametric density (see for instance [14, 42, 62]), however, in our methodology, they are used to capture the (plausibly sigmoidal) functional dependence of the Bernoulli distribution parameter p^c upon the input factor ϕ.

Overall, the computation of the probability of adverse events involves the following steps:

- Expert opinion is elicited to determine the probability p^c as a function of the percentage Shapley value deviation ϕ. Bezier curves are suitable candidates for representing expert opinions in this case.
- Then, given a specific instance of the collaborative process definition, one computes the numerical value of $p^c_i(\phi_i)$ for every actor i.

Starting from these probabilities, the probability of collusions within subsets of actors can be computed by suitable aggregators. For instance, if the simplifying assumption of independence among actors's defections can be made, one can compute the collusion probability $Pr(S)$ of a subset $S \in 2^A$ as a simple product of the individual actors probabilities:

$$Pr(S) = \Pi_{i \in S}\, p^c(\phi_i) \tag{10}$$

4.3 Impact Assessment

The technique we use for estimating impact of information disclosure is loosely related to the one we just used for our probability estimate. Let us start with an example: if an information item sent via email via a cloud-based mail service contains, say, an attachment with the design information of a new product, what will be the impact of its disclosure? To answer this question, one can use two different approaches:

(1) Perform an accurate analysis to precisely quantify the impact of the disclosure, e.g. in terms of the financial loss the company would occur in when a new competitor enters the market (an event that would certainly happen once the information item containing the new product design has been disclosed to current competitors)
(2) Use an arbitrary discrete unit and quantify the perceived impact of loosing such a message to a conventionally high level, corresponding to a perceived "disaster".

Our process-oriented way to perform impact analysis relies on quantifying the *Value of Information* (VoI) for each knowledge set K_S potentially reconstructed by a subset $S \in 2^A$ of the process actors.

4.4 Value of information Analysis

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 [31]. From a purely rational perspective, it is clear that acquiring extra information is only useful for an actor A if knowing it has a significant probability of modifying its behavior.

Classic VoI analysis typically involves constructing a complex decision-analytic model to fully characterize all information items available to each process actor, the loss each actor would incur should these items become known to other actors, 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 prioritizing interventions [30]. As alternatives to full VoI, we identified three approaches to analyzing 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-alleviation countermeasure, and raises the question about how the use of VoI to assess disclosure risk in the framework of our methodology.

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 achieving 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). 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 The Overall Methodology

In our approach, managing risks related to the execution of a business process P in presence of threats constitutes itself a process (usually called *risk management process*, in symbols $M_{R(P)}$) where alternative techniques for dealing with threats are compared according to a procedure. The output of $M_{R(P)}$ is a *risk alleviation strategy*, which consists of modifications to P that have some effect on the risk of executing it, including the introduction or removal of security controls. In this Section, we put forward a methodology for comparing alternative risk alleviation strategies. Our methodology does not provide specific guidance on the choice of mechanisms that will actually counter the threats; rather, it allows comparing the residual risk of competing risk strategies. Although qualitative comparison is supported, the methodology aims to quantitative cost-benefit calculations, assessments of risk tolerance, and quantification of preferences involved in $M_{R(P)}$.

Before describing our methodology in detail, we remark that probability assessment strongly affects what can be practically done within $M_{R(P)}$. For instance, it is sometimes possible to ask the users for a rough estimate of perceived probabilities and impacts and then multiply them to get *risk coefficients*. Such coefficients can be used for a "quick-and-dirty" comparison of versions of P that include different security controls. On the other hand, a (more costly) best-effort computation of probabilities and impacts allows to use quantitative estimates of $R(A, E)$ to drive the organization's choice between alternative implementations of P, by comparing $R(A, E)$ values to the cost of adoption/deployment of proposed security patches or controls.

We are now ready to provide a step-by-step description of our risk analysis methodology :

- The first step is the *stakeholder identification*, where we identify the actor set A of our business process P and compute its power set 2^A. In our approach, process stakeholders include all participants to P. Namely, our actor set includes *all actors who, according to the risk assessor, may in any way get the capability of reading (or writing) information shared during P's execution.* As we shall see in the following Sections (Section 6.1), actors in A can be further refined by type according to their role in the computation.
- The second step consists in the *formalization of the business process model*, using the syntax introduced in Section 3, which represents two types of actions: (i) *message exchanges* and (ii) *local computations*. It is important to remark that while execution-oriented process models usually contain control structures like conditions and loops [53], our process model syntax expresses all possible execution paths *independently*, i.e. as separate models. The next step takes care of this.
- The third step consists of *process streamlining*, which includes *loop unrolling* and *re-encoding of conditions as parallel paths*. Here we do not enter into the details of business process streamlining, as process improvement techniques have been deeply studied since the Eighties and are discussed in detail in the technical literature (see for instance the rich bibliography of [53]). However, software toolkits supporting our methodology will have to provide guidance w.r.t. process streamlining.
- The fourth step, *identifying reconstructible knowledge*, consists in computing the *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.*
- The fifth step consists in *estimating the collusion probability* for each subset S in 2^A at each step of the process P. Once again it is important to remark that this estimate needs to be process-specific (as it will take into account the microeconomics and social relations underlying P) and take into account multiple causes of collusion, including dysfunctional behavior, intervention of regulatory authority and others (Section 4.1)
- The sixth step consists in *estimating the disclosure impact* of K_S for each subset $S \in 2^A$ at each step of the business process P
- The seventh and final step consists in *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.

It is important to remark that the choice of the probability assessment method will strongly depend on the threats whose probability we are trying to assess. When the threat is the behavior of one or more stakeholders of a business process, like putting in common the knowledge they hold at a given time, probability estimates will be computed according to the microeconomics model underlying the process (Section 3). However other techniques, such as the assessment of the perceived level of the event's likeliness in terms of the social network of relations between the individuals involved in the process [8] are also possible.

In turn, the *impact quantification method* used to assess disclosure impact can go from a simple ordinal prioritization of levels of information sensitiveness to complex analysis of potential loss that would be caused by the disclosure of specific data items.

6 A Cloud-based Process Model

In this section we specialize the process model presented in Section 3) to describe cloud-based computations. We rely on a variation of Bogdanov *et al.*'s representation of cloud actors[6].

6.1 The Cloud Process Model

In order to make our representation of our multi-party business process actors suitable for describing cloud-

[8] As an alternative, a *Bayesian Belief Network* (BBN) [60] can be used to model the process, by taking into account its different actors and their mutual influences. The BBN can be exploited in different ways, especially to support identification and evaluation of risk control options at the organizational level.

based computations, our actor set A becomes a (non-necessarily disjoint) 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 - looser versions of the ones in [6] - 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 2 shows a sample visual representation of a cloud-based process, where a buyer send messages to two sellers who respond with their offers:

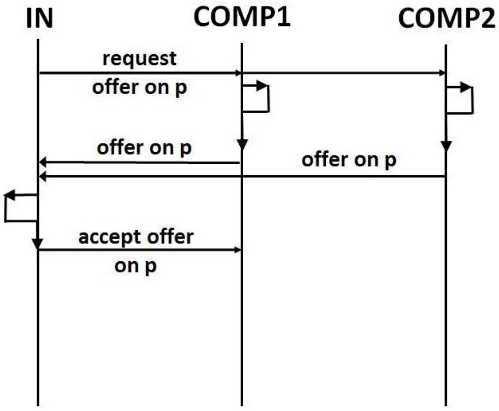

Figure 2. Visual representation of a sample cloud process model

7 Impact and Probability Assessment of Threats to Cloud-based Processes

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 $t_S \leq t$. 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:

$$I_{S,k,t} = \sum_{p=1}^{h} I_{j_p,k} \tag{11}$$

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. This risk can be written as follows:

$$R(A_k, E_S) = P_{S,t} I_{S,k,t} \tag{12}$$

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

$$R(A_k, 2^A, \infty) = \sum_{S \in 2^A} I_{S,k,\infty} P_S \tag{13}$$

However, it is clear that information sharing events - collusions - are in general not independent. We shall take care of dependency in the next version of our methodology.

7.1 Sample Assessments

Let us start with a very simple example: a business process where a client uses a cloud-based computation service to add two integer numbers and another one to publish the result. In this case, we have the actor set $A = (IN1, COMP1, RES)$ where actor $IN1$ holds the information item $INFO_1$ containing the two summands $INFO_1[1]$ and $INFO_1[2]$, actor $COMP1$ is the outsourced services that computes the addition, while actor RES publishes the result. The process is represented by the choreography shown in Fig. 3, where the input actor $IN1$ sends $INFO_1$ to $COMP1$, who computes the desired local function $f(INFO_1) = INFO_1[1] + INFO_1[2]$, i.e. adds the two summands and sends the result to the result node RES who outputs it.

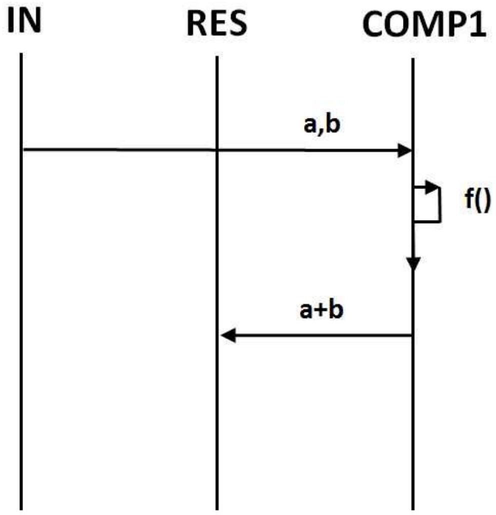

Figure 3. Our sample business process

According to the definitions given in the previous section, the (possibly empty) *common knowledge* of any subset of actors $S \in 2^A$ at time t, namely $K_S(t)$, is composed of the $INFO$ items that have been received in their entirety by all members of S at or before time t. The power set 2^A of the actor set is the simple Boolean lattice:

$$\{IN1\}, \{COMP1\}, \{RES\}$$

$$\{IN1, COMP1\} \qquad \{IN1, RES\} \qquad \{COMP1, RES\}$$

$$\{IN1, COMP1, RES\}$$

Figure 4. The Boolean lattice for the considered actor set

Running our sample business process P, we obtain the following knowledge sets $K_S(t)$ for $t = 1, 2, 3$ (we omit the formal step $K_\emptyset(0) = K_\emptyset(1) = K_\emptyset(2) = \emptyset$):

The Process Initialization $t = 0$

$$\begin{array}{c}\{IN1\}, \{COMP1\}, \{RES\}\\ [INFO], [\phi], [\phi]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{IN1, RES\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{COMP1, RES\}\\ [\phi]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1, RES\}\\ [INFO]\end{array}$$

Figure 5. The knowledge sets at time $t = 0$

First Step $t = 1$

$$\begin{array}{c}\{IN1\}, \{COMP1\}, \{RES\}\\ [INFO], [INFO], [\phi]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{IN1, RES\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{COMP1, RES\}\\ [INFO]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1, RES\}\\ [INFO]\end{array}$$

Figure 6. The knowledge sets at time $t = 1$

Second Step $t = 2$

$$\begin{array}{c}\{IN1\}, \{COMP1\}, \{RES\}\\ [INFO], [INFO], [INFO]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{IN1, RES\}\\ [INFO]\end{array} \qquad \begin{array}{c}\{COMP1, RES\}\\ [INFO]\end{array}$$

$$\begin{array}{c}\{IN1, COMP1, RES\}\\ [INFO]\end{array}$$

Figure 7. The knowledge sets at time $t = 2$

The disclosure risk estimated by actor $IN1$ at $t = 0$ is zero, as there are no subsets $X \in 2^A$ such that $K_X(0) \neq \emptyset$ but the singleton $\{IN1\}$, whose only member coincides with the risk evaluating actor $IN1$.
At $t = 1$, however, there is another singleton such that $K_X(1) \neq \emptyset$ (in particular, $K_X(1) = INFO_1 \cup S_f$, namely the subset $X = \{COMP1\}$. All the other subsets for which $K_X(1) \neq \emptyset$ can be obtained by computing the ideal generated by {IN1,COMP1} w.r.t. the Boolean lattice's join ($\cup$), so their contribution to the risk estimation is zero (all their members had the same knowledge separately than they have when taken together).
The estimate by actor $IN1$ of disclosure risk in the part of $COMP1$ of information $INFO_1 \cup S_f$ at $t = 1$ can therefore be written as follows:

$$R(A_k, E_S) = R(IN1, E_{\{COMP1\}}) =$$
$$= P_{\{COMP1\}}(1) I_{\{COMP1\}}(IN1, 1) \tag{14}$$

where $P_{\{COMP1\}}(1)$ is the probability (assessed by $IN1$) that $COMP1$ will disclose at $t = 1$ the information it now holds, i.e. the data $INFO_1$ and the specification S_f of the local function $f()$ it computes (i.e. the addition). $I_{\{COMP1\}}(IN1, 1)$ is the resulting total damage to $IN1$ of the service provider $COMP1$ disclosing what it knows, i.e. the summands $INFO_1$ and the specification S_f.

At $t = 2$, another singleton set such that $K_X(1) \neq \emptyset$ pops up, namely $X = \{RES\}$. Again, all the other subsets for which $K_X(1) \neq \emptyset$ can be obtained by computing the ideal generated by {IN1,COMP1,RES} w.r.t. the Boolean lattice's join ($\cup$) (in this case, the entire lattice) so their contribution to the risk estimation is zero (all their members had the same knowledge separately than they have together).

Under the assumption the two disclosure events to be independent, total risk estimate at $t = 2$ by $IN1$ is therefore, as expected:

$$R(A_k, E_S) = R(IN1, E_{\{COMP1\}} \cup E_{\{RES\}}) \tag{15}$$

that becomes:

$$R(A_k, E_S) = P_{\{COMP1\}}(1)I_{\{COMP1\}}(IN1, 1) + \\ +P_{\{RES\}}(2)I_{\{RES\}}(IN1, 2) \quad (16)$$

Of course, risk seen by other actors of P can also be evaluated by the same procedure: estimating the probability (or belief) that a disclosure event will occur as well as the damage they would incur should the disclosure event happen. For instance, risk estimated by RES at $t = 0$ is related to the singleton subset $\{IN1\}$ (the only one whose knowledge set is not empty; note that in this case, unlike before, it does not coincide with the risk assessor). We get:

$$R(A_k, E_S) = R(RES, E_{\{IN1\}}) = \\ = P_{\{IN1\}}(0)I_{\{IN1\}}(RES, 1) \quad (17)$$

In the same line, risk estimated by RES at $t = 1$ can be written as follows:

$$R(A_k, E_S) = R(RES, E_{\{IN1\}} \cup E_{\{COMP1\}}) \quad (18)$$

that becomes:

$$R(A_k, E_S) = P_{\{IN1\}}(0)I_{\{IN1\}}(RES, 1)+ \\ +P_{\{COMP1\}}(1)I_{\{COMP1\}}(RES, 1) \quad (19)$$

7.2 Alleviating Disclosure Risk

In order to mitigate disclosure risk, we apply our *risk management methodology* $M_{R(P)}$) to compare alternative strategies for dealing with risks connected to disclosure threats. The output of $M_{R(P)}$ is a *risk alleviation strategy*, which consists of modifications to P (including the deployment of security controls) that have the desired effect on the risk of executing it. While our methodology does not at present include specific guidance in the choice of such controls, we remark that the user can identify possible changes to P by searching pattern libraries offering alternative mechanisms for achieving and certifying the security properties of business process and services (see for instance [11]) [9].

We consider first a pattern of obfuscation of the local function $f()$. Instead of pushing the plaintext specification of S_f to the service provider $COMP1$, actor $IN1$ can use an obfuscation technique for computing the sum. While the obfuscation techniques themselves are outside the scope of this paper, we remark that a variety of obfuscation mechanisms have been proposed in the literature, including homomorphic encryption, evaluation of branching programs, and Garbled Circuits (GC). GC evaluation and homomorphic encryption are in principle both suitable for obfuscation of simple arithmetics operations like the one in our example. Let us assume that a GC technique is used for obfuscating addition (the size of the garbled adder circuit is small, linear in the size of the inputs), and its secure evaluation is efficient, as it is linear in the number of Oblivious Transfers (OT) and in the number of evaluations of a cryptographic hash function, for example SHA-256) [10].

In other words, in this representation of our business process P, our functional $G(f)$ denotes a Garbling Outsourcing Scheme mechanism that locally (i.e., within its own trusted environment) computes a garbled function $G(f) = f'()$ corresponding to the sum and pushes the garbled specification of $f'()$, namely $S_{f'}$ to $COMP1$. The corresponding sample choreography is depicted in Figure 8.

The subset analysis carried out in Section 7.1 is now repeated after applying our modifications to the process P, but we can now apply the weaker version of our computation transparency assumption (Section 3). This way, the information disclosed to $COMP1$ does not include the local function specification any more. We get:

$$R(A_k, E_S) = R(IN1, E_{\{COMP1\}}) = \\ = P_{\{COMP1\}}(1)I'_{\{COMP1\}}(IN1) \quad (20)$$

where, since the knowledge reconstructible by $COMP1$ is now smaller than before, $I'_{\{COMP1\}} \leq I_{\{COMP1\}}$. The modification to P has therefore decreased risk; however, the amount of such decrease needs to be compared with the combined costs of (i) the local computation of garbling $G(f)$ on the part of $IN1$ and (ii) the additional complexity of computing the garbled function $f'()$ - instead of the original addition $f()$ - on the part of $COMP1$.

Another version of P that can be envisioned in order to decrease disclosure risk features *multiple service provisioning*, where the confidentiality of $INFO_1$ is increased by outsourcing the computation of $f()$ to multiple services, each getting to know only a portion (a share) of $INFO_1$. In this case of course we will need to extend our actor set to become $A = (IN1, COMP1, COMP2, RES)$ where, once more, at $t = 0$ actor $IN1$ holds the entire information item

[9] Also, links between security properties and the corresponding threat spaces have been defined in the framework of several certification schemes [18].

[10] Garbled integer arithmetics has attracted much attention in the past few years [41] [37] both following Yao's original formulation and the alternative Goldreich-Micali-Wigderson (GMW) protocol. Also, [54] summarizes several depth-optimized circuit constructions for various standard arithmetic tasks.

$INFO_1$ containing the two summands $INFO_1[1]$ and $INFO_1[2]$.

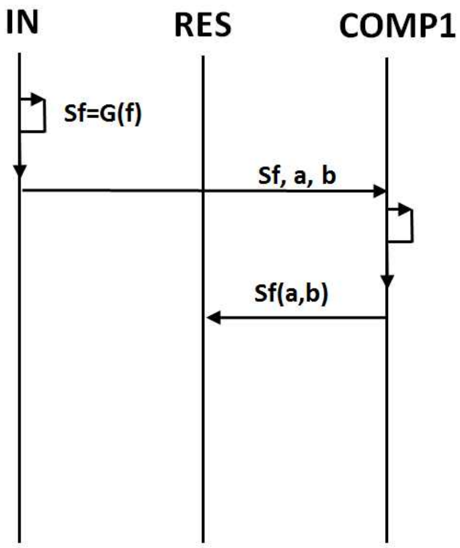

Figure 8. The choreography for the garbled sum computation

The power set 2^A of the actors is:

In words, the alternative version of our process P can be described as follows:

(1) The input actor $IN1$ computes a local function to divide each summand into two shares, obtaining $INFO_1[1,1], INFO_1[1,2], INFO_1[2,1], INFO_1[2,2]$.
(2) $IN1$ sends $INFO_1[1,1]$ and $INFO_1[2,2]$ to $COMP1$, and $INFO_1[1,2]$ and $INFO_1[2,1]$ to $COMP2$
(3) The two computation nodes compute a local function each on the shares they received, namely $f_{COMP1} = INFO_1[1,1] + INFO_1[2,2]$ and $f_{COMP2} = INFO_1[1,2] + INFO_1[2,1]$,
(4) The two computation nodes send the results to the result node RES
(5) RES computes $f_{RES} = f_{COMP1} + f_{COMP2}$ and outpu the result

For the sake of simplicity, let us assume for the moment that $IN1$ will generate two shares of $INFO_1$ using a naive technique, i.e. by taking respectively the Most Significant and the Least Significant Part (MSP-LSP) from the original value $INFO_1$.

For instance, if $INFO_1[1] = 25$ and $INFO_1[2] = 31$, then $COMP_1$ receives $INFO_1[1,1] = 20$ and $INFO_1[2,2] = 01$ and computes 21, while $COMP_2$ receives $INFO_1[1,2] = 05$ and $INFO_1[2,1] = 30$ and computes 35. Finally, RES receives 21 and 35 and computes 56.

Of course, this simplified share generation would not really prevent $COMP$ nodes from guessing the original values, so our assumption of *Information completeness*: "Given a message delivery (A_s, A_d, m_{t_S}), with $m_{t_S} \leq INFO_s$, then $\beta_i(E_{sd}) = 0$" (see Section 3) should now be revised here to, say, $\beta_i(E_{sd}) = \frac{1}{10}$, assuming they both $COMP$ nodes know that the summands are two-figure integers. However, we will not deal with probability of autonomous guessing in this example, as the threat space we are considering involves only collusions among multiple parties.

After defining this revised version of process P and the underlying assumptions, we can estimate the knowledge sets corresponding to this new, secured version of the business process. Figures 10–13 show the evolution of the knowledge sets starting from the initialization time $t = 0$ to time $t = 3$.

Our risk estimate at $t = 1$ by $IN1$ is therefore:

$$R(A_k, E_S) = R(IN1, E_{\{COMP1\}} \cup E_{\{COMP1\}}) = \\ = P_{\{COMP1\}}(1) I_{\{COMP1\}}(IN1, 1) \tag{21}$$

where, again, $P_{\{COMP1\}}(1)$ is $IN1$'s estimated probability that $COMP1$ will disclose at $t = 1$ the information it now holds, and $I_{\{COMP1\}}(IN1, 1)$ is the resulting damage to $IN1$. We recall the assumption of *Information completeness* (Section 3): given a message delivery (A_s, A_d, m_{t_S}), with $m_{t_S} \leq INFO_s$, then $\beta_i(E_{sd}) = 0$. Once again this property expresses a zero estimated probability that the sharing generation scheme can be broken. Therefore $IN1$ will attribute no risk to this stage, where no subset of actors not including itself has knowledge of both $INFO_1$ shares. Once again, we remark that a weaker version of the *information completeness* assumption can be adopted here to reflect the weakness of the naive share generation scheme, which could be easily broken by $COMP1$ via an educated guess. However, the threat space under consideration does not include autonomous guesses, and the original assumption is kept.

At $t = 2$, unlike the previous example, no other singleton exists such that $K_X(1) = INFO_1$. However, this time other subsets for which $K_X(1) = INFO_1$ can be obtained, namely $\{COMP1, COMP2\}$. Under the assumption the two disclosure events to be independent, risk estimate at $t = 2$ by $IN1$ is therefore:

$$R(A_k, E_S) = R(IN1, E_{\{COMP1,COMP2\}}) \tag{22}$$

that becomes:

$$R(A_k, E_S) = P_{\{COMP1\}}(1) I_{\{COMP1\}}(IN1, 1) + \\ + P_{\{COMP1,COMP2\}}(2) I_{\{COMP1,COMP2\}}(IN1, 2) \tag{23}$$

$\{IN1\}, \{COMP1\}, \{COMP2\}, \{RES\}$

$\{IN1, COMP1\}$ $\{IN1, COMP2\}$ $\{IN1, RES\}$ $\{COMP1, COMP2\}$ $\{COMP1, RES\}$ $\{COMP2, RES\}$

$\{IN1, COMP1, COMP2\}$ $\{IN1, COMP1, RES\}$ $\{IN1, COMP2, RES\}$ $\{COMP1, COMP2, RES\}$

$\{IN1, COMP1, COMP2, RES\}$

Figure 9. The Boolean lattice for the new actor set.

$\{IN1\}$ $\{COMP1\}$ $\{COMP2\}$ $\{RES\}$
$INFO$ $[\phi]$ $[\phi]$ $[\phi]$

$\{IN1, COMP1\}$ $\{IN1, COMP2\}$ $\{IN1, RES\}$ $\{COMP1, COMP2\}$ $\{COMP1, RES\}$ $\{COMP2, RES\}$
$[INFO]$ $[INFO]$ $[INFO]$ $[\phi]$ $[\phi]$ $[\phi]$

$\{IN1, COMP1, COMP2\}$ $\{IN1, COMP1, RES\}$ $\{IN1, COMP2, RES\}$ $\{COMP1, COMP2, RES\}$
$[INFO]$ $[INFO]$ $[INFO]$ $[\phi]$

$\{IN1, COMP1, COMP2, RES\}$
$[INFO]$

Figure 10. The knowledge sets at time $t = 0$

$\{IN1\}$ $\{COMP1\}$ $\{COMP2\}$ $\{RES\}$
$INFO$ $[INFO(1,1)]$ $[INFO(1,2)]$ $[\phi]$

$\{IN1, COMP1\}$ $\{IN1, COMP2\}$ $\{IN1, RES\}$ $\{COMP1, COMP2\}$ $\{COMP1, RES\}$ $\{COMP2, RES\}$
$[INFO(1,1)]$ $[INFO(1,2)]$ $[INFO]$ $[INFO(1,*)]$ $[INFO(1,1)]$ $[INFO(1,2)]$

$\{IN1, COMP1, COMP2\}$ $\{IN1, COMP1, RES\}$ $\{IN1, COMP2, RES\}$ $\{COMP1, COMP2, RES\}$
$[INFO]$ $[INFO]$ $[INFO]$ $[INFO(1,*)]$

$\{IN1, COMP1, COMP2, RES\}$
$[INFO]$

Figure 11. The knowledge sets at time $t = 1$

$P_{\{COMP1,COMP2\}}$ is the probability that $\{COMP1, COMP2\}$ will actually share the information they have to reconstruct $INFO[1]$ times the damage $IN1$ would incur in, should the disclosure event actually happen. It is important to remark that, if this probability is considered null by default (for example, the assessor is completely sure that $COMP1$ and $COMP2$ do not know of each other, operate on different clouds and are not under the jurisdiction of the same regulatory authority) risk at $t = 2$ is also 0.

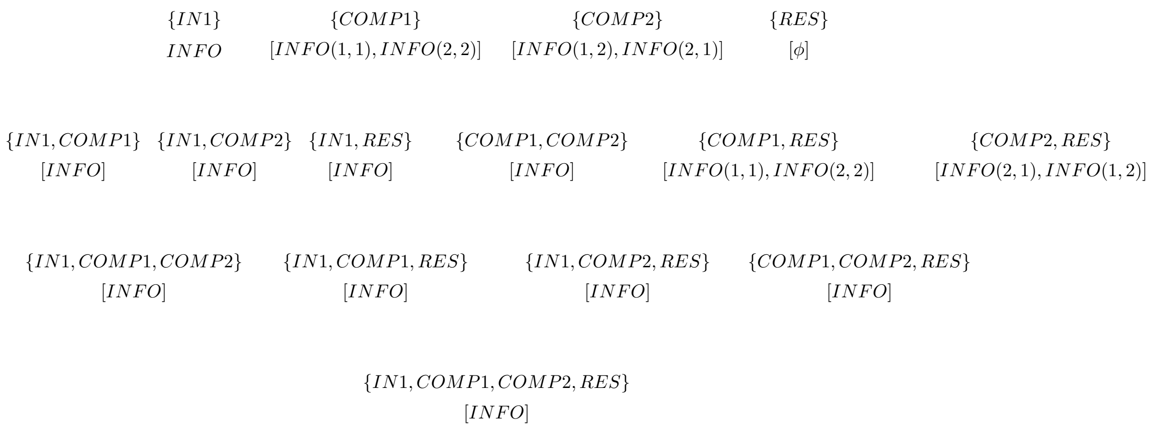
$\{IN1\}$ $INFO$
$\{COMP1\}$ $[INFO(1,1), INFO(2,2)]$
$\{COMP2\}$ $[INFO(1,2), INFO(2,1)]$
$\{RES\}$ $[\phi]$
$\{IN1, COMP1\}$ $[INFO]$
$\{IN1, COMP2\}$ $[INFO]$
$\{IN1, RES\}$ $[INFO]$
$\{COMP1, COMP2\}$ $[INFO]$
$\{COMP1, RES\}$ $[INFO(1,1), INFO(2,2)]$
$\{COMP2, RES\}$ $[INFO(2,1), INFO(1,2)]$
$\{IN1, COMP1, COMP2\}$ $[INFO]$
$\{IN1, COMP1, RES\}$ $[INFO]$
$\{IN1, COMP2, RES\}$ $[INFO]$
$\{COMP1, COMP2, RES\}$ $[INFO]$
$\{IN1, COMP1, COMP2, RES\}$ $[INFO]$

Figure 12. The knowledge sets at time $t = 2$

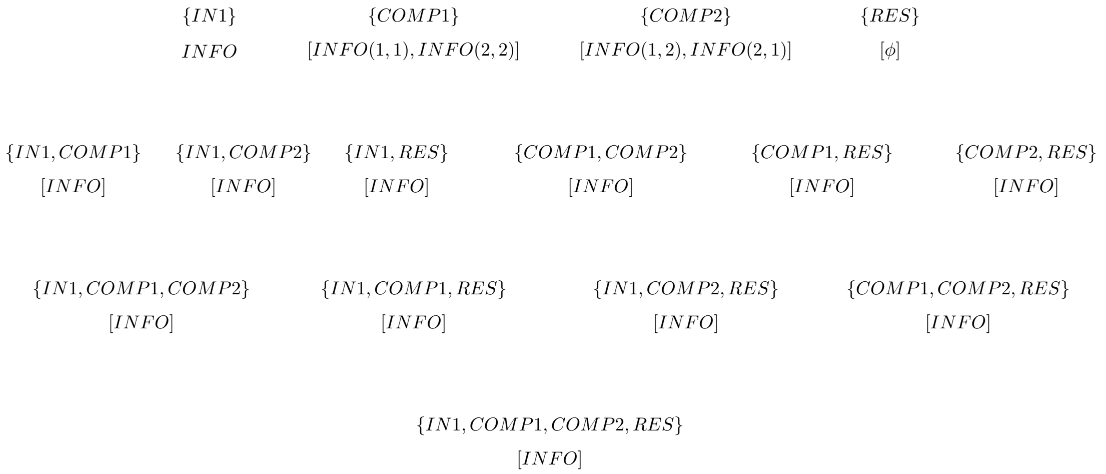
$\{IN1\}$ $INFO$
$\{COMP1\}$ $[INFO(1,1), INFO(2,2)]$
$\{COMP2\}$ $[INFO(1,2), INFO(2,1)]$
$\{RES\}$ $[\phi]$
$\{IN1, COMP1\}$ $[INFO]$
$\{IN1, COMP2\}$ $[INFO]$
$\{IN1, RES\}$ $[INFO]$
$\{COMP1, COMP2\}$ $[INFO]$
$\{COMP1, RES\}$ $[INFO]$
$\{COMP2, RES\}$ $[INFO]$
$\{IN1, COMP1, COMP2\}$ $[INFO]$
$\{IN1, COMP1, RES\}$ $[INFO]$
$\{IN1, COMP2, RES\}$ $[INFO]$
$\{COMP1, COMP2, RES\}$ $[INFO]$
$\{IN1, COMP1, COMP2, RES\}$ $[INFO]$

Figure 13. The knowledge sets at time $t = 3$

7.3 Comparing Alternative Processes via Risk Profiles

As we have seen, our two alternative versions of process P have different *risk profiles*, which the risk analyst can compare. In this section we show how a quick version of this comparison can be carried out over visual representations of the two profiles, *even when we are not able to exactly quantify probabilities ad impacts*. In this case, we simply assign a conventional value to the impact of the disclosure of each knowledge set and represent the different risk profiles associated to different versions of P on the same plane, where the horizontal axis shows the subsets $S \in 2^A$ and the vertical one shows the impacts of the disclosure of K_S in the two versions of the process.

Of course, different versions of the same business process P will in general have different actor sets A. Figure 14 shows the Boolean lattices corresponding to our sample comparison; in this particular case, it is easy to see that there is a total embedding of the lattice corresponding to the original version of P into the lattice corresponding to the modified.

In general, however, there will be no complete embedding of one lattice into the other; rather, a *partial mapping* μ will have to be defined. In the visual representation of our profiles, subsets of the two lattices that are connected by μ will correspond the same entry in the horizontal axis. Figure 15 shows the risk profiles associated to the two competing versions of our sample business process P for t=1 to 4. The risk assessor can thus get a first visual representation of the difference between the profiles, to be later tuned with more accurate impact and probability assessments.

8 The Auction Scenario

Auctions provide an ideal playground for the ideas discussed in the previous sections, since disclosure threats to auction fairness have been experimentally analyzed in the literature [22].

Also, much research has been devoted to the quantification of the impact of collusion. In this section, we briefly describe the main auction processes, point to some known security problems of auctions and outline

$$\{\mathbf{IN}\}, \{\mathbf{COMP1}\}, \{COMP2\}, \{\mathbf{RES}\}$$

$$\{\mathbf{IN,\ COMP1}\}\{IN, COMP2\}\ \{\mathbf{IN,\ RES}\}\quad \{COMP1, COMP2\}\ \{\mathbf{COMP1,\ RES}\}\ \{COMP2, RES\}$$

$$\{IN, COMP1, COMP2\}\ \{\mathbf{IN,\ COMP1,\ RES}\}\quad \{IN, COMP2, RES\}\ \{COMP1, COMP2, RES\}$$

$$\{IN, COMP1, COMP2, RES\}$$

Figure 14. The lattice embedding

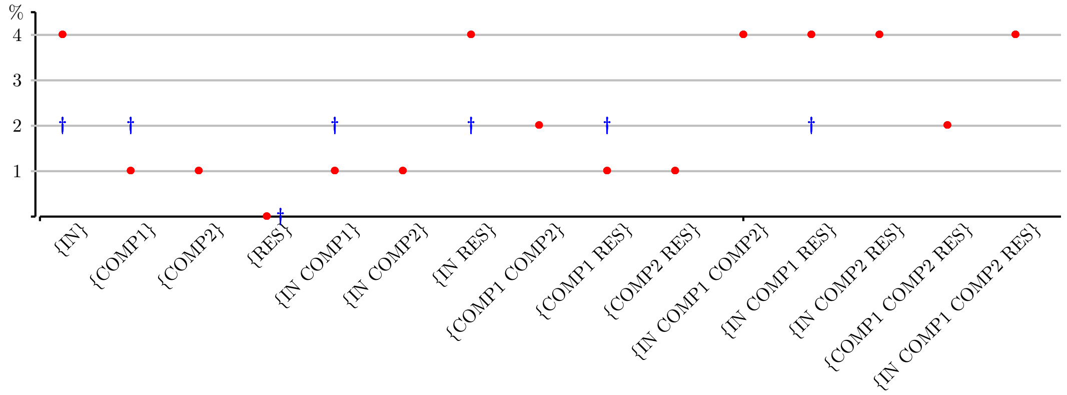

Figure 15. The risk profile of the secured P process compared with the original version

two auction scenarios exemplifying the computation of impacts and the application of our risk assessment methodology.

8.1 The Auction Process

Auctions provide efficient, distributed ways of solving goods and resource allocation problems [20, 32, 40, 50, 57, 63]. An auction consists of a set of potential *bidders* and one or more *auctioneers*, or bid takers. In some settings the auctioneer is the representative of a seller who wants to sell an item and get the highest possible payment for it while each bidder is a buyer (or the representative of one) who wants to acquire the item at the lowest possible price. In other settings the auctioneer represents a buyer who wants to acquire a service at the lowest possible price, while each bidder represents a seller who wants to offer the service at the highest possible payment. From the point of view of auction theory the two forms are totally analogous.

Although auctions can be safely run among cooperative agents, a key problem is how to design auction processes when auctioneers and bidders are self-interested agents. By looking for strategies that self-interested agents will follow, auction theory pursues a main goal: designing interaction protocols (auction processes) that achieve desirable social outcomes even though agents act based on self-interest. Auctions have acquired great relevance over the Internet thanks to commercial services which allow to auction tangible items, as well as in cloud environments, where auctions are used for the provision of services. Online auctions are often called *e-auctions*.

The auction process An e-auction process involves several main activities (or tasks):

- *Initialization*: the auctioneer sets up the auction and advertises it (i.e., type of good or service, starting time, etc.).
- *Registration*: in order to participate in the auction, bidders must first register with the auctioneer (or a registration manager); this ensures that only valid bids are made and that bidders can be identified for payment purposes; registered bidders should be able to participate in any number

of auctions rather than re-registering for each new auction.

- *Bidding*: a registered bidder computes his/her bid and submits it to the auctioneer (in some cases in a sealed envelope, in some other by an open declaration); the auctioneer checks the received bid to ensure that it conforms to the auction rules.
- *Winner determination*: the auctioneer determines the winner applying the auction rules (e.g. in an ordered set of bids, computing the first-best or the second-best). In sealed-bid auctions this phase is also called *Opening*.
- *Winner communication*: the auctioneer announces the winner to the parties.
- *Contract issuing*: the auctioneer issues to buyer and seller the contracts for the exchange of goods or services and payments. The enforcement of the contract is a key point in auctions: in case of weakly enforced contracts the auction can become object of attacks.

Among the problems already present in traditional auctions that are still present in e-auctions we mention the following:

- a buyer can cheat by colluding with other bidders to affect the settlement price; in the next subsection we will discuss this case in detail;
- a buyer can repudiate bids or fail to pay;
- the seller of the item could collude with some of the buyers
- the seller might fail to deliver the goods or services;
- buyers/sellers could also forge a bid in an attempt to introduce fake bids in order to influence the auction operation;
- a corrupt auctioneer could award the auction to someone other than the legitimate winner;
- a bidder's personal information could be sold to external parties, or used for malicious purposes.

E-auctions add security concerns relating to bid privacy, bidder anonymity, correct evaluation and declaration of winner, etc. For instance, bidders may try to eavesdrop the offers of other bidders, based on weaknesses of the communication infrastructure; the auctioneer may try to manipulate results; an incorrect outcome may result from introduction of false bids or modification of submitted bids, undue extension or shortening of bidding period and introduction of new bids based on information about submitted bids. For bidders, bid values may be sensitive information and loss of bid-privacy may reveal important information such as financial status etc. against their wishes.

Security of various types of e-auctions has received considerable attention from researchers during the past two decades [7–10, 25, 34, 43, 44, 47, 48, 58, 64]. The main goals for a secure and anonymous e-auction scheme are the following: *unforgeable bids* (if bids are forgeable a bidder can, for instance, be impersonated); *non-repudiation* (once a bidder has submitted a bid they must not be able to repudiate having made it: for example, if a bidder wins and does not want to pay, they might deny that they submitted the bid); *public verifiability* (there must be some publicly available information by which all parties can be verified as having correctly followed the auction protocol: this should include evidence of registration, bidding, and proof of winner/loser); *robustness* (for instance, the auction process must not be affected by invalid bids or by participants not correctly following the auction protocol).

Vickrey Auctions The Vickrey auction consists in a second-price sealed-bid auctions and is close to eBay's system of proxy bidding. A slightly generalized version of it – named generalized second-price auction – is used in Google's and Yahoo!'s online advertisement programs [23, 61].

During Vickrey auctions' bidding stage, bidders seal their bid (e.g., place it in an envelope) and submit it to the auctioneer. During the opening stage, the auctioneer opens all of the bids and determines the winner. The winner is the bidder with the highest bid however he is required to pay an amount equal to the second highest bid (i.e., the highest losing bid).

In private value Vickrey auctions, it is a weakly dominating strategy for both buyers to seal their true valuations of the auctioned good in to the envelopes they submit to the auctioneer. Indeed, whatever bids the other Buyers may have sealed in their envelopes, one can never benefit bidding below his true valuation because this can only lessen one's own probability of winning the auction without altering the amount paid in case of victory. Equally, one can never benefit from bidding above one's own true valuation, because this higher valuation will be useful to the victory only if another player has submitted a bid that is at least equal to one's own true valuation: in which case one would have to pay, in case of victory, at least the true valuation with the possibility to pay more.

8.2 Impact of Collusion in a Simple Auction

We now consider issues that arise when a subset of, or possibly all, the bidders act collusively and engage in bid rigging with the purpose of obtaining lower prices. The resulting arrangement is called a *bidding ring* or *cartel*. While bidding rings are illegal, in real world auctions they appear to be widely prevalent. Investigations of collusion in real world auctions constitute a significant component of antitrust activity [38]. In e-auctions the same issue is present if the bidders can know each other in real word, but can be present also when bidders who do not know each other get in contact and exploit the information leaked from the system. Theoretical models of collusion among bidders involve a mix of cooperative and non-cooperative game theory: the former is needed to allocate to ring members the gain possibly obtained thanks to the the collusion: the problem can be faced for instance with an approach based on the Shapley Values of the ring members. Here, we focus on the noncooperative part and take an example scenario from the on-line equivalent of a sealed-bid second-price auctions, which could be synthesized as follows.

Let us refer to the valuation of the bidders Alice, Bob and Carol by X_A, X_B, X_C respectively: in an honest second-price auction all the bidders bid their real valuation of the item. Suppose Alice and Carol have already submitted their valuations, while Bob still has not submitted his; Bob – thanks to a leakage – gets to know Alice's valuation and other contact information and discovers that $X_A > X_B$. Bob contacts Alice to establish a bidding ring: Bob will not bid X_B, but 0, so that in case of victory, Alice will not risk paying Bob's true valuation, but will pay Carol's true valuation: in case Carol's is lower than Bob's, this represents a gain $X_B - X_C$ for the ring, and a corresponding loss for the seller. We now extend this schema to N independent bidders, following the lines of [38].

Specifically, we assume that each bidder's value is a random variable X_i, distributed according to the cumulative distribution function F_i over some common interval $[0, 1]$. Let $\mathcal{I} \subseteq \mathcal{N}$ be the set of bidders in the bidding ring; we indicate by $\mathcal{I} = \{1, 2, \dots, I\}$ the set if their indexes and by $\mathcal{N} \setminus \mathcal{I} = \{I+1, I+2, \dots, N\}$ the set of bidders outside the ring. For any set of bidders let the random variable $Y_1^{\mathcal{S}}$ denote the highest of the values of the bidders in S.

The presence of a ring in a second-price auction does not affect the behavior of bidders who are not members of the ring. It is still a weakly dominant strategy for a bidder $j \in \mathcal{I}$ to bid his or her value X_j. It is also weakly dominant for the ring to submit a bid equal to the highest value among its members, that is, $Y_1^{\mathcal{I}}$.

A bidding ring generates profits for its members by suppressing competition. Specifically, instead of N effective bids, only $N - I + 1$ effective bids are submitted, since only the member of the cartel with the highest value in the ring submits bids his or her true valuation: the reminders submit non-serious bids by bidding at or below the reserve price. The ring's profits come from the fact that, in certain circumstances, the price paid by a winning bidder from the ring is lower than it would be if there were no ring at all.

Specifically, suppose that one of the ring members $i \in \mathcal{I}$ has a value X_i and that this is the highest of all bidders in the ring or outside the ring, i.e. $X_i = Y_1^{\mathcal{N}}$. Assuming that $X_i > r$, in the absence of a ring, this bidder would pay an amount equal to $P_i = max\{Y_1^{\mathcal{N}\setminus i}, r\}$ for the object. But if he were part of a functioning ring, his fellow members in $\mathcal{I}$ would bid at most r, so he would pay only

$$\widehat{P}_{\mathcal{I}} = max\{Y_1^{\mathcal{N}\setminus\mathcal{I}}, r\}$$

Thus, the expected payments of ring members are lower than they would be if the ring did not exist.

For a fixed reserve price r, let $m_i(X_i)$ denote the expected payment of bidder i with value x_i when there is no ring operating and all bidders behave non-cooperatively. Likewise, let $\widehat{m}_i(X_i)$ denote it's expected payment when there is a ring, then

$$t_i(x_i) \equiv m_i(X_i) - \widehat{m}_i(X_i)$$

represents the contribution of bidder i to the ring's expected profits when his value is x_i. The total ex ante expected profits of the ring amount to

$$t_{\mathcal{I}} \equiv \sum_{i\in\mathcal{I}} E[t_i(X_i)]$$

Notice that the probability that a bidder outside the ring will win the object is the same whether or not the ring is functioning; in both cases it is just the probability that she has the highest value among all bidders. Furthermore, the price that a bidder $j \in \mathcal{I}$ would pay in the event that she wins is:

$$max\{Y_1^{\mathcal{I}}, Y_1^{\mathcal{N}\setminus\mathcal{I}\setminus j}, r\} = max\{Y_1^{\mathcal{N}\setminus j}, r\}$$

the same as the price she would pay if there were no ring. Since for all bidders who are not part of the ring, neither the probability of winning nor the price upon winning is affected, the expected payments in the two situations are the same: the profits of these bidders are also unaffected. Since the profits of bidders outside the cartel are unaffected by its presence, the gains accruing to the cartel as a whole are equal to the loss suffered by the seller. This reasoning also leads to the conclusion that the gains from collusion increase as the size of the ring increases.

The reasoning done so far can be easily given a quantitative exemplification in the case of uniform prior introduced previously. Let us assume, as above, that the common distribution of individual valuations is a uniform density $f(v)$ on $[0, 1]$ and the reserve price is $r = 0$. The expected payment of a player with valuation v in a non-rigged, second-price auction is given by the expected value of the the distribution of the maximum of $(N - 1)$ players (see above)

$$m_i(v) = \frac{N-1}{N} v^N$$

in a rigged second-price auction with a ring of I participants, instead, is given by the expected value of the the distribution of the maximum of $(N - I)$ players

$$\widehat{m}_{\mathcal{I}}(v) = \frac{N - I}{N - I + 1} v^{N-I+1}$$

which is clearly lower. Thus, the expected advantage for the ring members translates in an expected damage for the seller.

This closes our example of computation of the impact of collusion due to information leakage.

8.3 The Vickrey Auction Process Model

We are now ready to model the Vickrey sealed-bid auction process using the process model of Section 3. Bidders submit written bids without knowing the bid of the other people in the auction. The highest bidder wins, but the price paid is the second-highest bid. Figure 16 shows our process model of a Vickrey auction process VAP with two bidders. Node IN models the auctioneer, while a trusted $COMP$ node is used to compute via a suitable function $f()$ the second best bid. The RES node publishes both the amount to be paid (the second highest offer) and the winning bidder (the one who submitted the highest offer).

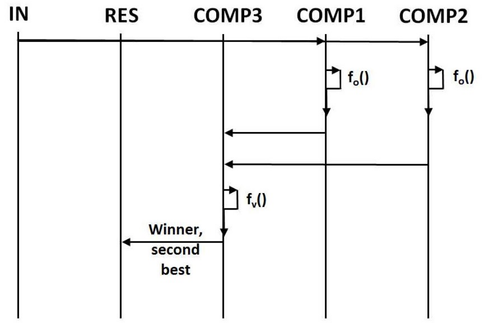

Figure 16. The Vickrey bid protocol

According to our methodology, the risk management process of VAP (M_{VAP}) first identifies the obfuscation mechanism for function $f()$. In principle, this risk alleviation could be relevant in this case even if the specification G_f of the Vickrey auction choice function is public from the start and accepted by all actors as part of the auction model. In principle, the obfuscation of the function, e.g. via GSC, can be tailored to hide from $COMP3$ whose offer is the second best, preventing it from learning more than the identity of the highest bidder and the amount of the second highest bid. However, the M_{VAP} comparison in this case largely coincides with the example in the previous section, and is therefore omitted [11]. The description below is the representation in our model of the approach described in [56]. After auctioneer IN has solicited their offers, the participating bidders $COMP1$ and $COMP2$ store their information items $INFO_i$ (their bids) in binary format , and the number of bits in each value is kept equal. in order to streamline the process representation avoiding (unrolling) loop, we have to choose a fixed value of this bit length (say 3) [12] Both bidders $COMP1$ and $COMP2$ send the most significant bit of their information items ($INFO_1$ and $INFO_2$) to $COMP3$. The latter actor computes a local function $f()$ (a logical OR) of bits received. The result of $f()$ is sent from COMP3 to RES who publishes it. If the result of $f()$ is zero, $COMP3$ does nothing and waits for the next bit. If the result of $f()$ is a 1, all those parties who sent a 0 bit stop sending further bits to $COMP3$. It means those actors who sent a 1 bit continue sending the bits. When all the bits of the last party are sent, $COMP3$ publishes to RES (i) the winner (ii) the set of the results of the OR operations of the second-greatest value [13] .

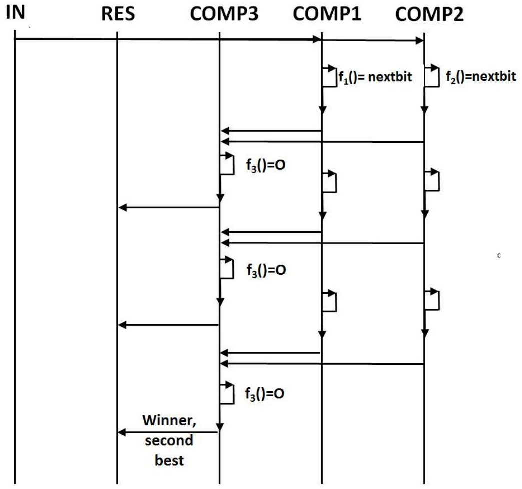

Figure 17. The alternative version of the Vickrey process

For the sake of conciseness, we do not show the entire Boolean lattice here. A quick visual comparison between the risk profiles of original and the modified version of the protocol VAP shows however that in the

[11] An alternative corresponds to a version of VAP where a Secure Multiparty Computation (SMC) protocol is used to run the auction.

[12] According to the methodology, the analysis should in principle be repeated for all 64 possible pairs of bid values. It is however easy to see that in this process the risk profile is independent from any specific set of bids.

[13] In the case of this small example, the identity of the winner becomes known anyway. When multiple bidders are present and the privacy of the winner is to be preserved, RES can communicate the identity of the winner and the amount only to IN

modified version the knowledge sets $K_{COMP1,COMP3}$ and $K_{COMP2,COMP3}$ do not increase the knowledge held by $COMP1$ and $COMP2$ until the tie-break moment, when the result is published anyway. This remains true even if we adopt a slightly weaker version of our *Information completeness assumption*, where knowing the MSB part of two information items allows to learn their relative order with respect to a domain total order relation.

Therefore, as intuition suggests, the role of $COMP3$ can be safely assigned to an untrusted party (in terms of availability to "collude", i.e. to share with any of the bidders the bits of other bidders it receives) with no additional risk. It is important to remark that this result only holds for this specific version of VAP process with two bidders. Moving to a three bidders version with bidders $COMP1$, $COMP2$, $COMP3$ and compute node $COMP4$, the latter's availability at a given time $t = k$ to collude with, say, $COMP1$ by sharing the bit flows received from $COMP2$ and $COMP3$, together with the weaker version of our *Information completeness assumption* would allow $COMP1$ to change its $k + 1$-th bit in order to keep up with competitors.

9 Conclusions

The risk analysis methodology presented in this paper provides a fresh look at fully quantitative risk management on the cloud, enabling the comparison of cloud-based process models including different security mechanisms from the point of view of the changes in risk they imply. While the methodology is still under evolution and refinement, especially as far as the scalability of the process models is concerned, we claim that our approach is extendable to cover most "cost versus risk" assessment activities. Also, the process model used in our methodology gracefully extends existing machine-readable specification of processes like the W3C candidate recommendation for choreographies WS-CDL (`http://www.w3.org/2002/ws/chor/`) and lends itself to be supported by an innovative software toolkit integrating existing choreography editors.

Acknowledgements

The work described in this paper is based on an invited talk presented by Ernesto Damiani at the ISCISC 2014 Conference in Tehran, Iran. The authors wish to thanks colleagues taking part to the conference for their useful comments and valuable insight. This work was partly supported by the EU-funded project PRACTICE (contract n. FP7-609611) and by the Italian MIUR project SecurityHorizons (c.n. 2010XSEMLC) and by the CMIRA2014/AcceuilPro program of the Region Rhone-Alpes.

References

[1] Information risk analysis methodology IRAM. `https://www.securityforum.org/iram#iramtva`.

[2] ATOS. Risk analysis framework for a cloud specific environment, 2008.

[3] 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.

[4] Samik Basu and Tevfik Bultan. Choreography conformance via synchronizability. In *Proc.s International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, pages 795–804, 2011.

[5] Mihir Bellare, Viet Tung Hoang, and Phillip Rogaway. Foundations of garbled circuits. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012*, pages 784–796, 2012.

[6] 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.

[7] Peter Bogetoft, Ivan Damgård, Thomas Jakobsen, Kurt Nielsen, Jakob Pagter, and Tomas Toft. A practical implementation of secure auctions based on multiparty integer computation. In *Financial Cryptography and Data Security*, pages 142–147. Springer, 2006.

[8] Colin Boyd and Wenbo Mao. *Security issues for electronic auctions*. Hewlett-Packard Laboratories, 2000.

[9] Phillip G Bradford, Sunju Park, Michael H Rothkopf, and Heejin Park. Protocol completion incentive problems in cryptographic vickrey auctions. *Electronic Commerce Research*, 8(1-2):57–77, 2008.

[10] Felix Brandt. Fully private auctions in a constant number of rounds. In *Financial Cryptography*, pages 223–238. Springer, 2003.

[11] Ingrid Buckley, Eduardo B. Fernández, Marco Anisetti, Claudio Agostino Ardagna, Seyed Masoud Sadjadi, and Ernesto Damiani. Towards pattern-based reliability certification of services. In *On the Move to Meaningful Internet Systems: OTM 2011 - Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2011,*

Hersonissos, Crete, Greece, October 17-21, 2011, Proceedings, Part II, pages 560–576, 2011.

[12] Daniele Catteddu and Giles Hogben. Cloud computing: Benefits, risks and recommendations for information security. Technical report, ENISA, 2009.

[13] Ann Cavoukian. Privacy risk management: Building privacy protection into a risk management framework to ensure that privacy risks are managed by default. Technical report, Information and Privacy Commissioner - Ontario - Canada, 2010.

[14] SY Chan. An alternative approach to the modeling of probability distributions. *Risk Analysis*, 13(1):97–102, 1993.

[15] T. Chen. *Information and Risk Management*. 2009.

[16] CISCO. Data leakage worldwide white paper: The high cost of insider threats, 2011.

[17] Cloud Security Alliance. Security guidance for critical areas of focus in cloud computing v2.1, 2009.

[18] Ernesto Damiani, Claudio Agostino Ardagna, and Nabil El Ioini. *Open Source Systems Security Certification*. Springer, 2009.

[19] Folker den Braber, Gyrd Brndeland, Heidi E. I. Dahl, Iselin Engan, Ida Hogganvik, Mass S. Lundand Bjrnar Solhaug, Ketil Stlen, and Fredrik Vraalsen. The coras model-based method for security risk analysis. Technical report, SINTEF, 2006.

[20] K Eric Drexler and Mark S Miller. Incentive engineering for computational resource management. *The ecology of Computation*, 2:231–266, 1988.

[21] S Drissi, H Houmani, and H Medromi. Survey: Risk assessment for cloud computing. *International Journal of Advanced Computer Science and Applications*, 4:143–148, 2013.

[22] Martin Dufwenberg and Uri Gneezy. Information disclosure in auctions: an experiment. *Journal of Economic Behavior & Organization*, 48(4):431–444, August 2002.

[23] Benjamin Edelman and Michael Schwarz. Internet advertising and optimal auction design. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, page 1, 2008.

[24] Josep Oriol Fitó and Jordi Guitart. Introducing risk management into cloud computing. Technical Report UPC-DAC-RR-2010-33, Technical University of Catalonia, 2010.

[25] Matthew K Franklin and Michael K Reiter. The design and implementation of a secure auction service. *Software Engineering, IEEE Transactions on*, 22(5):302–312, 1996.

[26] Sailesh Gadia. Cloud computing risk assessment: A case study. *ISACA Journal*, (1):1–6, 2012.

[27] The Open Group. Risk taxonomy, 2008.

[28] Koichi Harada and Eihachiro Nakamae. Application of the bzier curve to data interpolation. *Computer-Aided Design*, 14(1):55 – 59, 1982.

[29] Jay Heiser and Mark Nicolett. Assessing the security risks of cloud computing, 2008.

[30] 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.

[31] Ronald A. Howard. Information value theory. *IEEE Trans. Systems Science and Cybernetics*, 2(1):22–26, 1966.

[32] Bernardo A Huberman and Scott H Clearwater. A multi-agent system for controlling building environments. In *ICMAS*, pages 171–176, 1995.

[33] Information Systems Audit and Control Association. Cobit 5. `http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Risk-Scenarios-Using-COBIT-5-for-Risk.aspx`, 2013.

[34] Ari Juels and Michael Szydlo. A two-server, sealed-bid auction protocol. In *Financial Cryptography*, pages 72–86. Springer, 2003.

[35] Burton S. Kaliski, Jr. and Wayne Pauley. Toward risk assessment as a service in cloud environments. In *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing*, HotCloud'10, pages 13–13, Berkeley, CA, USA, 2010. USENIX Association.

[36] A.U. Khan, M. Oriol, M. Kiran, Ming Jiang, and K. Djemame. Security risks and their management in cloud computing. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, pages 121–128, Dec 2012.

[37] Vladimir Kolesnikov. Gate evaluation secret sharing and secure one-round two-party computation. In *Advances in Cryptology - ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4-8, 2005, Proceedings*, pages 136–155, 2005.

[38] Vijay Krishna. *Auction theory*. Academic press, 2009.

[39] Antonio Kung, Alberto Crespo Garcia, Nicols Notario McDonnell, Inga Kroener, Daniel Le Mtayer, Carmela Troncoso, Jos Mara del lamo, and Yod Samuel Martns. Pripare: A new vision on engineering privacy and security by design. Technical report, PRIPARE, 2014.

[40] Jeffrey K MacKie-Mason and Hal R Varian. Pric-

ing the internet. Technical report, EconWPA, 1994.

[41] Dahlia Malkhi, Noam Nisan, Benny Pinkas, and Yaron Sella. Fairplay - secure two-party computation system. In *Proceedings of the 13th USENIX Security Symposium, August 9-13, 2004, San Diego, CA, USA*, pages 287–302, 2004.

[42] Thomas A Mazzuchi and Johan René van Dorp. A bayesian expert judgement model to determine lifetime distributions for maintenance optimisation. *Structure and Infrastructure Engineering*, 8(4):307–315, 2012.

[43] Moni Naor, Benny Pinkas, and Reuban Sumner. Privacy preserving auctions and mechanism design. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 129–139. ACM, 1999.

[44] Khanh Quoc Nguyen and Jacques Traoré. An online public auction protocol protecting bidder privacy. In *Information Security and Privacy*, pages 427–442. Springer, 2000.

[45] NIST. Federal information processing standard (fips) 65, guideline for automatic data processing risk analysis, 1979.

[46] NIST. Recommended security controls for federal information systems and organizations, 2009.

[47] Kazumasa Omote and Atsuko Miyaji. A practical english auction with one-time registration. In *Information Security and Privacy*, pages 221–234. Springer, 2001.

[48] David C Parkes, Michael O Rabin, and Christopher Thorpe. Cryptographic combinatorial clock-proxy auctions. In *Financial Cryptography and Data Security*, pages 305–324. Springer, 2009.

[49] Michael O. Rabin. How to exchange secrets with oblivious transfer. *IACR Cryptology ePrint Archive*, 2005:187, 2005.

[50] Tuomas Sandholm. An implementation of the contract net protocol based on marginal cost calculations. In *AAAI*, volume 93, pages 256–262, 1993.

[51] Amit Sangroya, Saurabh Kumar, Jaideep Dhok, and Vasudeva Varma. Towards analyzing data security risks in cloud computing environments. In *Information Systems, Technology and Management - 4th International Conference, ICISTM 2010, Bangkok, Thailand, March 11-13, 2010. Proceedings*, pages 255–265, 2010.

[52] P. Saripalli and B. Walters. Quirc: A quantitative impact and risk assessment framework for cloud security. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, pages 280–288, July 2010.

[53] August-Wilhelm Scheer and Markus Nüttgens. ARIS architecture and reference models for business process management. In *Business Process Management, Models, Techniques, and Empirical Studies*, pages 376–389, 2000.

[54] Thomas Schneider and Michael Zohner. GMW vs. yao? efficient secure two-party computation with low depth circuits. In *Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers*, pages 275–292, 2013.

[55] A.S. Sendi and M. Cheriet. Cloud computing: A risk assessment model. In *Cloud Engineering (IC2E), 2014 IEEE International Conference on*, pages 147–152, March 2014.

[56] R. Sheikh and D.K. Mishra. Protocols for getting maximum value for multi-party computations. In *Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on*, pages 597–600, May 2010.

[57] R Smith. Communication and control in problem solver. *IEEE Transactions on computers*, 29:12, 1980.

[58] Koutarou Suzuki and Makoto Yokoo. Secure generalized vickrey auction using homomorphic encryption. In *Financial Cryptography*, pages 239–249. Springer, 2003.

[59] The Economist Intelligence Unit. Managing business risks in the information age, 1998.

[60] Paolo Trucco, Enrico Cagno, Fabrizio Ruggeri, and Ottavio Grande. A bayesian belief network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Rel. Eng. & Sys. Safety*, 93(6):845–856, 2008.

[61] Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163 – 1178, 2007.

[62] Mary Ann Flanigan Wagner and James R Wilson. Using univariate be´ zier distributions to model simulation input processes. In *Proceedings of the 25th conference on Winter simulation*, pages 365–373. ACM, 1993.

[63] Carl A Waldspurger, Tad Hogg, Bernardo A. Huberman, Jeffrey O. Kephart, and W. Scott Stornetta. Spawn: A distributed computational economy. *Software Engineering, IEEE Transactions on*, 18(2):103–117, 1992.

[64] Changjie Wang and Ho-fung Leung. Anonymity and security in continuous double auctions for internet retails market. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.

[65] Vic Winkler. Cloud computing: Risk assessment for the cloud. *Technet Magazine*, January 2012.

[66] David Wright. Should privacy impact assessments be mandatory? *Commun. ACM*, 54(8):121–131, 2011.

Ernesto Damiani is a full professor at Università degli Studi di Milano, the director of the University's PhD program in computer science and the coordinator of SESAR Research Lab (`http://sesar.di.unimi.it`). He has held visiting positions at a number of international institutions, including George Mason University in Virginia, US, LaTrobe University in Melbourne, Australia, Tokyo Denki University, Japan and INSA-Lyon, France. Ernesto Damiani's research interests include cloud assurance, Web services and business process security and Big Data processing. He has served as Chair of many conferences, including the IEEE International Conference on Web Services (ICWS), the IEEE Conference on Digital Ecosystems series (IEEEDEST) and the IFIP Working Conference on Open Source Systems (OSS). Ernesto Damiani has published more that 300 papers, books and international patents. He is a senior member of the IEEE and a Distinguished Scientist of the ACM. He is the author of "Open Source Security Certification", Springer, 2009.

Stelvio Cimato is an Assistant Professor with the Dipartimento di Informatica of the Università degli Studi di Milano. He got the Ph.D in Computer Science at University of Bologna, Italy in 1999. His main research interests are in the area of cryptography, network security, and web applications. He has published several papers in the field and is active in the community, serving as member of the program committee of several international conferences in the area of cryptography and data security. He is also involved in the organization of conferences and workshops, and has participated to the activities included in several national and European research projects (SecureSCM (STREP), CUMULUS (STREP), PRACTICE (IP)).

Gabriele Gianini is an Assistant Professor at the Department of Information Technology of the University of Milan. He holds a visiting lectureship at the Computer Science Faculty of the Free University of Bolzano-Bozen. His research interests focus on software engineering and include the application of inferential statistics and of data mining methods applied to the study of open source software development data. Between 1990 and 2000 he has been working within the High Energy Physics community at the Fermilab Tevatron in Chicago (E299, E687 and FOCUS collaborations) and at the CERN in Geneva (CMS collaboration). Within the SESAR Lab, he is conducting research activities within the research project MAPS (Agile Methodologies for Software Production), and in the research project KIWI, both funded by the Italian Minister of University and Research, investigating the subject of information extraction from automatically collected software process logs and of semantic-aware knowledge discovery over open source development data.