C. Fernandez-Gago and D. Nuez, "Metrics for Accountability in the Cloud", *Accountability and Security in the Cloud*, vol. 8937, pp. 129-153, 2015.
NICS Lab. Publications: https://www.nics.uma.es/publications

Metrics for Accountability in the Cloud

Carmen Fernández-Gago(✉) and David Nuñez

Network, Information and Computer Security Lab,
University of Malaga, 29071 Malaga, Spain
{mcgago,dnunez}@lcc.uma.es

Abstract. Accountability in the Cloud is a key concept that is determined by the *accountability attributes*. For assessing how accountable an organisation is we should be able to assess or provide techniques for measuring the attributes that influence on accountability. How much or to what extent they should be measured is a key issue. One of the goals of the A4Cloud project is, therefore, to develop a collection of metrics for performing meaningful measures on the attributes that influence accountability. This paper sets up the foundations towards the elicitation of metrics for accountability attributes. We describe here a metamodel for metrics for accountability attributes, which constitutes the basis for the process of elicitation of metrics for accountability. This metamodel is intended to serve as a language for describing accountability attributes and sub-attributes and for identifying the elements involved in their evaluation. One of the key components of the metamodel is the type of evidence the attribute use.

1 Introduction

One of the important aspects behind the accountability concept is the ability of an organization to demonstrate their conformity with required obligations [6]. The goal of an organisation is therefore the demonstration of accountability through the measurement of the degree of such conformity and the provision of meaningful evidence. Thus, measurement becomes an important tool for assessing the accountability of an organization by external authorities (and organizations themselves, in the case of self-assessment). It is then crucial to find suitable methodologies for eliciting metrics for accountability attributes. Metrics can be of different types (quantitative and qualitative), and they can be supported by different kinds of evidence. Thus, for the case of accountability we need to determine which are the most suitable ones for each case of the attributes. In fact, it is not the attributes themselves that we measure but the evidence related to them. It is then of paramount importance to identify the most suitable evidence for each aspect of the accountability attributes to be measured.

The methods and models that we introduce in this paper rely on different inputs for eliciting metrics such as the context and the nature of the attribute to be measured. In order to carry out the process of elicitation of metrics we start by a review of the definitions of the basic concepts and terminology regarding

M. Felici and C. Fernández-Gago (Eds.): A4Cloud 2014, LNCS 8937, pp. 129–153, 2015.
DOI: 10.1007/978-3-319-17199-9_6

metrics. The concepts related to Metrology range from what is to be measured (attributes) to what is a measure, scale or measurement method. Then, we perform an analysis of the accountability attributes from the metrics perspective. This analysis will allow us to identify the aspects or dimensions involved in the definitions of the attributes that are suitable to be measured. It is worth to note that it is not possible to measure the attributes themselves as they are very general but specific aspects and dimensions that are identified for them. For instance, it might not make much sense to measure how transparent an organization is but instead we could measure more specific aspects such as whether there exists a notification process. Once the attributes to be measured are analysed we have to define a methodology for performing meaningful measures. Thus, we define a metamodel for eleicitng metrics for accountability attributes that takes as inputs evidence and criteria that follows a top-down approach for the process of elicitation of metrics. We illustrate the use of the metamodel in the case of the transparency attribute.

The paper is organised as follows. Section 2 introduces the basic concepts and definitions on measurement and Metrology and Sect. 3 their application to security properties. An analysis of the definitions of accountability attributes from the metrics point of view is performed in Sect. 4. The methodology that we propose for measuring them is introduced in Sect. 5 and an example of its application is described in Sect. 6. Finally, Sect. 7 concludes the paper and outlines the open research areas in the field.

2 Background and Basic Definitions of Metrics

In this section we will summarise the main concepts concerning metrics that will be used for defining accountability metrics. We start with some notions of Metrology within the context of information security and privacy.

Metrology is defined as the scientific study of measurement [1]. As such, there already exists a broad selection of reference material regarding metrology concepts, including standards, books, research papers and guidelines. In this section we will provide a brief review of them from the most important sources. In particular, we will use the following material as the main reference on metrology and information security measurement:

- ISO/IEC 27004:2009 (E) Information Technology Security techniques Information Security Management Measurement [5]: This standard belongs to the ISO/IEC 27000 family on information security. In particular, the 27004 standard provides guidance on the development and use of measures with respect to Information Security Management Systems (ISMS). Most of the definitions regarding measurement proposed here are extracted or adapted from this standard.
- NIST SP 800-55 (revision 1) Performance Measurement Guide for Information Security [4].

- Complete Guide to Security and Privacy Metrics [11]. As its title states, this book provides extensive guidelines for developing security and privacy metrics, as it is based on a wide selection of metrology and information security standards and guidelines.
- Software Metrics and Software Metrology [7] is another useful source of metrology concepts, in this case with a focus on the software area. It provides basic concepts for designing measurement methods.

2.1 Definitions

Taking the above documents and [3] as main references we have adapted the following definitions for the scope of our work in the A4Cloud project [2].

- **Attribute:** property or characteristic of an object that can be distinguished quantitatively or qualitatively by human or automated means [3].
- **Metric or measurement result:** a set of indicators, together with an associated interpretation, that is designed to facilitate decision-making and improve performance and accountability through collection, analysis, and reporting of relevant data (adapted from [5,11]).
- **Measure:** variable whose value is assigned as a result of measurement [3].
- **Measurement method:** logical sequence of operations, described generically, used in quantifying an attribute with respect to a specified scale [3].
- **Indicator:** measure that provides an estimate or evaluation of specified attributes derived from an analytical model with respect to defined information needs [5].
- **Evidence:** data collected to support a metric, including the data inputs necessary to calculate and validate the metric (adapted from [4]).

Note that we prefer to use the term 'attribute' rather than 'property'. This decision is based on the fact that the ISO/IEC 27004:2009 standard uses the 'attribute' term for referring to a measurable concept. In addition, the term 'property' is often used to refer to functional properties of a system. Therefore, in our case, attribute is used as a synonym of 'non-functional property'. Also, the term 'attribute' is the one used in the Conceptual Framework for describing the main concepts that comprise accountability in the A4Cloud project.

2.2 Scales of Measurement

In the classical theory of measurement [17], the scales of measurement (or levels of measurement) are a set of categories for classifying measurement methods regarding their characteristics. Identifying the scale for each particular metric is essential for interpreting and analysing its results. Moreover, since each scale has a set of permitted operations, knowing its scale allows us to assess the validity of a metric, or at least, to discard senseless metrics.

We can classify the scales of measurement as follows:

- **Nominal scales.** This type of scale is applicable for mapping entities to names or categories. It is also known as categorical scale. Values in a nominal scale are not related to each other. For this reason, only the equality operation (=) is permitted for nominal values. From a statistical viewpoint, only modes can be computed.
- **Ordinal scales.** This scale permits to assign an order relation to its values, which is used to put measured entities in order. For this reason, ordinal scales are said to have magnitude. However, there is no information for measuring the differences between values. A simple example of this scale is the set of values 'Low – Medium – High'. There is an order relation that permits to state that High is greater than Medium, which in turn is greater than Low, but it makes no sense to measure the difference between Low and Medium. Ordinal scales are also nominal. Ordinal scales therefore permit to use equality (=) and inequality ($\leq$) operations, as well as medians and percentiles. Certain non-parametric statistical tests that only require ordinal data, known as ranking tests [17], can also be performed.
- **Interval scales.** This type of scale permits to measure differences between values. Additionally, interval scales are also ordinal scales. Thus, their values can be compared and ordered. Interval scales permit additions and substractions of their values. Therefore, means and standard deviations can also be computed. However, multiplications and divisions, and hence any other operations that depend on those, such as ratios, cannot be performed.
- **Ratio scales.** This type of scales improve interval scales by adding a meaningful zero value. Ratio scales are also interval scales. All the operations that are valid for interval scales apply here too. In addition, multiplication and division are also meaningful.

Nominal and ordinal metrics are often grouped as qualitative metrics, whereas interval and ratio metrics are quantitative. This differentiation is very important when processing the results of metrics, which will happen when aggregating and composing metrics or when producing interpretation of the results of a metric. Qualitative metrics may need to be converted to quantitative, in order to make possible complex processing, such as aggregated metrics. Note that this process often consists on defining a transformation from a qualitative domain (which at most possess a partial ordering) to a numeric one, which implies making assumptions on the validity of such transformation. On the contrary, quantitative metrics may need to be converted to qualitative ones when facing the reporting of final assessments, in order to be easily interpreted by people. For example, a numeric metric could be transformed to a simple Green-Yellow-Red label.

2.3 Guidelines for the Development of Metrics

In this section, we will provide some guidelines for the development of metrics. We set up the following steps: design of the metric, application and exploitation of the result of the metric. In the following, we will explain these steps in detail:

Design of the Metric. This is the initial phase of the life cycle of the metric, which is composed of the following steps:

1. **Definition of the scope and measurement objectives.** This first step intends to provide a clear description of the purpose of the metric and aids to isolate its context. It is composed of the following sub-steps:
   (a) **Specification of the attribute and entity to be measured.** This sub-step is related to the characterization of the measured concept. In particular, it addresses questions such as 'What do we want to measure?', 'Which attribute of which entity?', etc.
   (b) **Objectives of the metric.** In this sub-step, we will clearly define the goals of this metric and its context. It addresses questions like 'What is the purpose of this metric?', 'Who will use it?', 'Whose viewpoint is used for defining this metric?', etc.
   (c) **Relation to the requirements.** This sub-step is important when the metric is related to any non-functional requirement. As we mentioned earlier, metrics are useful to verify the compliance of such requirements.
2. **Definition of the measurement method.** In this step, a mapping from the observations to a measure is established, as well as the associated details of the measurement method. This can be done in the following steps:
   (a) **Specification of the measurement scale and measurement unit.** This step is important for realizing the admissible operations on this metric, and therefore, correctly defining it.
   (b) **Specification of the mapping from observations to measures.** This step provides means for effectuating this mapping. Note that the measure could either be numerical or nominal. The mapping could be expressed in several ways, such as a mathematical expression, an algorithm or a generic procedure.
3. **Documentation of the metric.** Metrics have to be properly documented once they are specified.

Application of the Metric. This phase corresponds to the execution of the measurement method to the observations from the real world.

1. **Input data collection.** This step gathers the data that will be used for performing the measurement procedures.
2. **Application of the measurement method.** This step comprises the execution of the measurement method to the observed data gathered in the previous step.
3. **Verification of the measurement results.** The results from the application of the metric should be verified in order to guarantee its quality. Special attention should be given to delicate steps of the application, such as mathematical operations and input data gathering.

Exploitation of the Results of the Metric. Once the result of the metrics is obtained and verified, then it has to be exploited. The steps involved in this phase are the following:

1. **Reporting of the results.** This step is intended for the presentation of the results of the application of the metric. The output of this step should be documented with the main information from the application phase.
2. **Interpretation of the results.** Interested stakeholders could interpret the results of the application in relation to the objectives of the metrics and its associated requirements. For example, the management of an organization could make use of the results of the metrics in order to support management decisions, such as the initiation of corrective actions.

3 Metrics for Attributes Relevant to Accountability

In this section we will consider some attributes (or non-functional properties) that influence accountability to a certain extent, and will mention some metrics for them that have been considered in the literature. These attributes are especially relevant in the A4Cloud framework. We categorize these attributes in three main areas:

– Privacy attributes.
– Security attributes.
– Cloud-specific attributes.

3.1 Measuring Privacy Attributes

The relation between privacy and accountability is complex and materializes in several ways. For example, digital transactions are easily recorded by service providers and third parties, leading to increasing tracking and profiling of individuals. This fact clashes with the right to be informed (in some cases, even consent is needed) and with the right of the individuals to be forgotten (in the case of the EU, this right is proposed for recognition in the proposed regulation on data protection [10]). Privacy and accountability are in this case related concepts, as providers of IT systems (i.e. data-collecting parties) should be accountable for the protection and treatment of the personal information they gather.

This aspect is a particular case of a more general relation between privacy and accountability. Organizations should be accountable for the degree of conformity with their privacy-related obligations. These obligations could be either of regulatory, contractual or ethical nature, among others. Measuring the degree of conformity with these obligations is very important in order to assess the level of accountability of an organization.

A minor instance of the relation of privacy and accountability (within the A4Cloud context) can be seen also in the trade-off between anonymity and accountability of users. If users should be accountable of their use of IT resources,

some compromise must be established regarding their privacy. At one end of the spectrum, a fully anonymous system difficulties accountability, as it is not possible to trace the identity of users who misbehave. In this case, the accountability of the system will presumably be low. At the other end, a fully accountable system difficulties anonymity. Therefore, the level of users privacy in a system influences its accountability and viceversa, but it is not clear to what extent. However, this aspect is out of the A4Cloud scope.

Privacy metrics have been applied in anonymity networks, anonymity in databases, and unlinkability for individuals. One of the examples of metrics in anonymity networks is *anonymity set* [15]. This is a metric that is given by the numbers of members in a set, to whom the adversary is looking for. The adversary will take advantage of all he/she knows about the members of the set to exclude as many individuals as possible. If the adversary is able to reduce the number of members of the set to one such that there is only one individual the adversary would have been successful. Thus, the bigger the set is the better anonymity is preserved.

3.2 Measuring Security Attributes

Availability. In today's world where more and more of our every day lives rely upon automated processing, the inability to access or use personal data can have consequences that range from a minor inconvenience to life threatening consequences. For example, when a bank ATM network becomes unavailable, it will result in discontent from card owners, but if a hospital system is unable to access patient data it may have more serious effects.

It should be noted that data protection is not limited to protecting privacy, but also concerned to broader goals such as assuring that data is notably processed 'fairly and lawfully' and that this is processed in a secure manner [9]. Availability can then be as important as confidentiality and integrity of data. Additionally, data subjects generally have in the right to access, update or erase (depending on the cases) their data. For these rights to be exercised, the system supporting the data must be available. For all these reasons, the ability to measure availability of a system is relevant to accountability.

Availability can be usually defined as the **target percentage of total operational time or requests** for which a service should be considered available in a period of time [8]. It is necessary to define precisely what the service is and what constitutes an 'unavailability event'. Defining an unavailability event requires the definition of certain parameters [8]:

- Service request. This is related to the functions of the service that are included in the measurement.
- Failure of a service request. What is the criteria to determine that a service failed? Are there standards for that?
- Sample size. What is the time period where the availability criteria is applied? If the sample size is too low the measure might not be significant.

- The scope of the service. Does this apply to requests from a single customer, service-wide user requests, requests from a specific geographical region, etc.? Does the service cover end-to-end fulfilment of requests, or only as far as the nearest Internet connection point?
- The commitment period to measure availability has to be always specified. This period could be one year, one month, one week, etc.

We could measure availability based on the definition of a single request failure. Thus, the target percentage of total requests can be calculated as the ratio $\frac{T-F}{T}$, where T is the total number of requests and F the number of request failures.

Providers could define a **recovery time objective (RTO)** as the maximum acceptable delay for recovery from an availability incident. RTO can be measured against **mean recovery time (MRT)**, which is the necessary average time to recover from an unavailability event.

Incident Response. According to [8] an incident is any event which is not part of the normal operation of the service and which causes, or may cause an interruption or a reduction in the quality of the service as stated in the SLA. Usually, it is necessary to characterize incidents through the following parameters:

- A Severity level. A classification of the incident according to a severity scale.
- Time to respond. Time between the notification of the incident and the implementation of the remediation action.

Based on these parameters it is possible to define the following metrics:

- Percentage of incidents of a certain severity resolved in a period of time.
- Recovery process and expected time to recover.
- Time to report. This is the time since the occurrence of an incident and until it is reported to the user.
- Time since the last incident of a given severity level.
- Specific incident data (e.g. number of records breached, downtime, time to respond).

Data Lifecycle Management. Data lifecycle management is related to how well the practices to handle data are managed by the provider. The measurements that can be performed for it include measuring the efficiency and effectiveness of such practices: service's back up, data replication system, portability and data loss prevention systems. The following parameters can be measured:

- Back-up test frequency and results.
- Restoration speed.
- Success or failure of operational back-ups.
- Data recovery points.

- Export test results.
- Percentage of response to requests.
- Data loss protection.
- Data durability.
- Scheduled deletion failure.
- Legal disclosure of regulatory requests.

Data Confidentiality Level. We propose the following measure to indicate the level of encryption of data in a cloud-based system (see Table 1):

Table 1. Data confidentiality level

Level	Description
0	Data is not cryptographically protected by the cloud provider
1	Data is cryptographically protected in transit
2	Data is cryptographically protected at rest and in transit
3	Data is cryptographically protected even at execution time

A system with Level 0 of data confidentiality does not use any cryptographic protection. It may, however, use other types of security measures, such as access control policies. A system that achieves Level 1 protects data that is transmitted from and to the cloud provider. This kind of system achieves security against an eavesdropper, but data is in clear inside the cloud provider, and therefore, susceptible to insider attacks or security breaches. Level 2 implies that data is protected also at rest. Proper mechanisms for key management need to be used. However, data should be decrypted before processing and could be accessed by malicious software or insiders. In a system with Level 3, the cloud provider does not decrypt data prior processing because the cryptographic scheme enables the processing of encrypted data. This level could be achieved with the aid of Fully Homomorphic Encryption schemes, but current proposals are not viable in practice. However, for certain applications, such as secure auctions and e-voting, there are simpler homomorphic schemes that are efficient and usable, and could reach this level.

Confidentiality Objective. We propose a measure to indicate the level of confidentiality achieved by a system regarding client data independently of the means used to achieve this objective (see Table 2).

The last level represents the best possible protection for a cloud client, however it will limit the ability of cloud providers to process the data except for storage purposes.

Table 2. Level of confidentiality

Level	Description
0	Data confidentiality does not satisfy any of the above levels
1	Data may be accessible by the cloud provider personnel for regular operational purposes, under the control of an authentication, authorization and accounting (AAA) mechanism
2	Technical and organizational measures are in place so that data may only be accessible to privileged CSP personnel (administrators) for debugging or maintenance purposes, under the control of an AAA mechanism
3	Technical and organizational measures are in place so that data is only accessible to privileged CSP personnel to respond to law enforcement or extraordinary requests made by the client, under the control of an AAA mechanism
4	Data is encrypted by the client with cryptographic keys that cannot be ascertained by the provider

3.3 Cloud-Specific Metrics

Elasticity. Following a similar approach to [8], we propose to define the elasticity ratio, a quantitative measure of elasticity, as the ratio $\frac{T-F}{T}$, where F is the total number of failures of resource provisioning requests over a period P (the commitment period), and T is the total number of provisioning requests over period P.

Location. In a cloud environment, providing the exact location of the data center that holds the clients data is neither strictly useful nor necessarily desirable (for physical security reasons). On the other hand, there is a strong case for providing a country or regional indicator of the location of the data, since it has strong regulatory implications. In some circumstances, the data may however reside in two or more datacenters during its life cycle. In practice, it is usually impossible to strictly 'prove' that data is only in a particular location (and not elsewhere) and we must rely on the trustworthiness of the cloud providers to provide that information. We propose to define a *location indicator* as a list of pairs *(location, certainty)*, where *location* refers to the ISO 3166-1 alpha-2 country or region code where the data resides, and *certainty* refers to the probability that the data will be located in this location at least once during its lifecycle (according to the CSP). Note that identical copies of data may be simultaneously in two different locations. Additionally, the proposed 'certainty parameter' could also be expressed as the average percentage of time that the data spends in a location during its lifecycle.

Data Isolation. This property is about ensuring the confidentiality, integrity and availability of data between different cloud clients [8]. When it comes to

data isolation, we can ask the following questions: can a cloud client read or modify a memory block, storage data or network packets produced by another client? Can a cloud client still read a memory block or storage data once another client has deleted it? Recent research shows the additional risk of side channel attacks, whereby a cloud client can discover information about another client, including in particular the value of secret cryptographic keys, by observing the temporal behaviour of the system [18]. To the best of our knowledge there are no metrics associated with data isolation in the cloud. As a first step, we propose to define the following indicator called *data isolation testing level*, which describes the level of testing that has been done by the cloud provider to assess how well data isolation is implemented (see Table 3).

Table 3. Data isolation testing level

Level	Description
0	No data isolation testing has been performed
1	Read/write isolation has been tested
2	Secure deletion has been tested, in addition to read/write isolation
3	Absence of known side channel attacks has been tested, in addition to read/write and secure deletion

It is important to note that in order to use such a metric, the resources in the scope of the measurement need to be well defined (storage, CPU, network, memory, database, etc.). Additionally, a standard set of tools or procedures need to be defined to establish the tests that should be conducted to assess each level.

4 A Review on Accountability Attributes from the Metrics Perspective

The accountability attributes have been defined in [16]. We remind here their definitions according to the Conceptual Framework for accountability introduced there and we will present an analysis of the accountability attributes in order to assess their usefulness with respect to metrics. Such an analysis will be carried out focusing on the following aspects:

– *Are the definitions of the accountability attributes valid from the point of view of metrics? Is there any ambiguity in the definition given by the Conceptual Framework? Is the attribute to be evaluated well identified from the definition?* These questions will help us to identify any inconsistencies, vagueness, and significant overlappings of the definitions of the attributes. Ideally, a correction should be proposed.

- *Can the attribute be decomposed in other sub-attributes?* The definitions of the attributes included in the conceptual framework are in some cases very abstract and high-level. Therefore, it might be useful to identify particular cases for each attribute depending on its nature and context that may be more concrete and useful from a metrics viewpoint.
- *Interdependencies with other attributes.*
- *What type of metrics could be defined for this attribute? Are there any requirements for a metric for this attribute?* We should identify the possible characteristics for a metric for such attribute depending on its nature and context and if possible, identify potential metrics for them.

For all of the attributes it might difficult to measure the attribute itself. For this reason, we identify dimensions of the attributes that are easier to be measured.

4.1 Transparency

Transparency is a property of an organization or a system about how well it implements and demonstrates the implementation of the following three transparency practices:

- Informing upstream stakeholders about data protection policies and their implementation practices.
- Notification in case of policy violation and other events that have been agreed upon in the policy, which includes explanation of the actions taken on such event.
- Responding to data subject access requests about data handling, e.g., data storing and processing.

A transparent organization will implement procedures for supporting these practices, and will provide means for demonstrating the existence and quality of such procedures.

From a high-level point of view, a transparency metric would measure how easy is for an external party to inspect the policies and procedures of an organisation regarding data protection.

There are several dimensions for assessing transparency that could be measured:

- Accessibility. This dimension is related to the level of easiness for obtaining the necessary information by the relevant stakeholders. The more transparent an organization is, the easier for stakeholders is to obtain the information they need.
- Effectiveness. Even if information is fully accessible, it may not be effective. It is necessary that the receptor is capable of processing, digesting and using the information [12]. This dimension is related to the usefulness of provided information. For example, the provision of excessive amounts of information, although accessible, renders it useless. The same aspect applies to the format and method of the provision of information.

– Timing. This dimension is related to assessing when the transparency actions are taken with regard to the event that triggered (this dimension has more sense with aspects such as notification). For example, it is possible to measure quantitatively the elapsed time between the event of the violation of a privacy policy and the corresponding notification.
– Other dimensions can be framed as combinations of accessibility, effectiveness, and timing. For instance, the provided information may be incomplete at the beginning (an accessibility problem) but may be completed after further user requests, which is also a timing problem.

4.2 Responsibility

Responsibility is a relationship between two entities regarding a specific Responsibility Target (policy, rules, states of affairs), such that the Responsibility Holder is responsible to the giver of the responsibility, the Responsibility Principal.

According to this definition, Responsibility should take into account any operation performed by the responsibility holder, then the policy should be used to evaluate if the performed action was according to the norm or not. As shown in Fig. 1, the important point for the responsibility attribute is that responsibilities cannot be looked at in an isolated way but must always be considered as a relationship between two agents. The Responsibility Target for which responsibilities are held may be at any level of granularity of the organization.

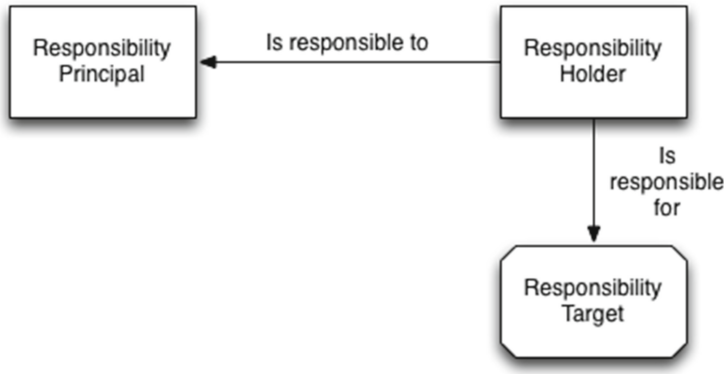

Fig. 1. Responsibility relationships

Some of the practices that we could identify for responsibility, derived from the definition given above, are:

– Responsibility granting. The process where the responsibility principal grants the actual responsibility to the responsibility holder. It should be noted that the granting of responsibility can actually involve a chain of Responsibility Holders, as shown in Fig. 2. For example, the primary Cloud Service Provider

(data processor) might be responsible towards the Cloud Customer (data controller), if one of the sub-processors carrying out processing operations on his behalf do not implement the appropriate security measures for the protection of personal data.
– Responsibility assessment/attribution. The process where conformance of the Responsibility Holders performed actions is evaluated with respect to the Responsibility Target. This practice can be subdivided into the following:
  • Non-repudiation. Comprehending the unambiguous authenticity and integrity of the Responsibility Holders identity.
  • Authentication. As required to assess the identity of the Responsibility Holder. As mentioned in the example above, in a chain of Responsibility it is possible that the person responsible for a malicious action, is not the legal responsible.
  • Integrity. Needed to assess that the Responsibility Holders identity and actions have not been tampered with.

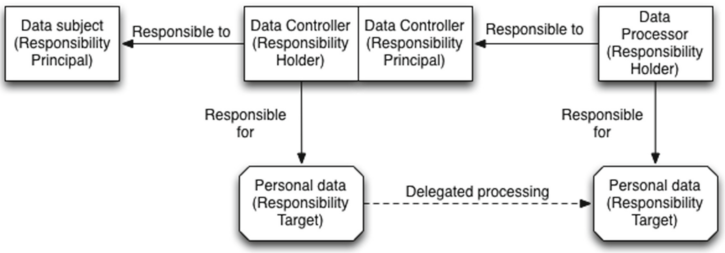

Fig. 2. Responsibility relationships

If we want to measure (either qualitatively or quantitatively) an entity's responsibility (i.e., its Responsibility Level) with respect to (i) some specific action and (ii) a set of policy/rules, then some high-level metrics to take into account are:

– Level of Authentication (LoA). Different organizations are likely to deploy different authentication mechanisms, therefore we cannot expect the same assurance in the responsible entity's (unambiguous) identification process.
– Delegation of Responsibility. This metric should assess the responsibility delegation process. It is clear that responsibility will attenuate in long delegation chains.
– Integrity. In analogy to the LoA metric(s), the inherent assurance of the adopted integrity mechanisms must be assessed to measure the organizations responsibility. For example, an organization using MD5 to protect the integrity of their log files cannot have the same Responsibility Level of other organization using SHA-512, due to the inferior integrity level offered by MD5 with respect to SHA-512.

– Duty/Role separation. The model used to split the responsibilities (e.g., n out of m), must be clearly stated in order to determine the responsible entity/entities. Notice that this metric is somehow related to the Delegation metric.

4.3 Remediability

Remediability is a property of an organization on the quality of its internal processes for taking corrective and compensatory actions in case of failing to comply with their commitments and policies. According to this definition, remediability is supported by the following main practices:

– Notification, which implies informing the relevant stakeholders (e.g., affected data subjects, regulators, services elsewhere in the service chain) about the failure, breach or disclosure.
– Reparation, which is related to taking corrective actions and technical remedies for restoring the system to the state prior the damage, if possible. This implies restoring data and supporting forensic recording.
– Redress, which implies legal remedies due to the damage suffered. These remedies may imply that the affected part claims compensatory, or even, punitive damages.

The Remediability concept is built upon the existence of a relation of responsibility between two entities, the responsibility holder and the responsibility principal (as described in Sect. 3.2) and the occurrence of a failure to comply with the responsibility target. Remediability also adds a fourth entity called remediation agent, such as a court or a dispute resolution entity, which may be used as a third-party by the responsibility principal and the responsibility holder in order to arbiter the remediation actions. Figure 3 shows these relationships.

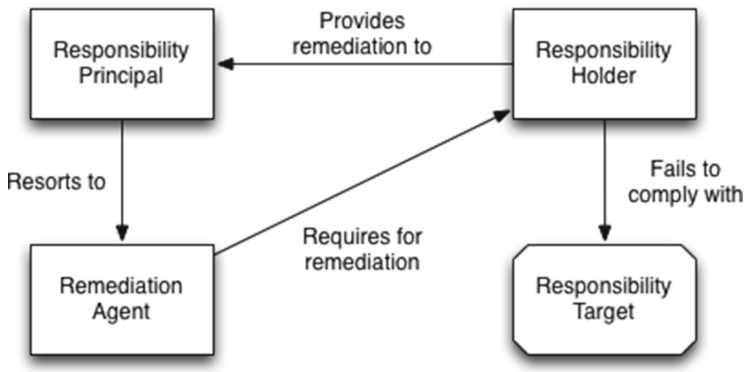

Fig. 3. Remediability relationships

A metric for remediability would measure the quality of the remediation practices held in place by an organization. There are several aspects that can be assessed with respect to the quality of remediation of an organization:

- As notification is part of the remediation process, one can evaluate the quality of the notification procedures. More aspects to be measured are:
  - Existence and quality of the notification processes. For example, simply to assess the existence of internal policies within an organization for addressing the notification of the affected parties after any damage has occurred.
  - Timing of notification. The relevance of notification is affected by the elapsed time between the occurrence of the damage and the effective time of notification.
  - Efectiveness of the notification. Even if notification is provided, it may not be useful for the relevant stakeholder. For example, indirect notification, such as publication of a notice in a web site, is not as useful as a direct notification by email. Also, the information included in the notification should be useful enough for the affected party, such as a proper explanation of the incident and the taken actions, and a description of the possible options for seeking for remediation.
- Reparation activities. Metrics could be defined to evaluate the quality of the technical remedies and corrective actions:
  - Preparedness level. Actions intended to prepare the organization in advance to the event of a failure and the necessity of restoring to a prior state. Some of these practices are data recovery and support forensics.
  - Repairability level. Assessment of the level of reparation of an organization to restore a failure, from the perspective of the affected party. For example, restoring damaged data from a back-up can be enough, while the disclosure of personal data that has already taken place cannot be entirely corrected.
- Redress. A metric for redress could measure aspects that impact the quality of the redress actions planned and taken by the organization such as proper definition of compensations, standard vs custom compensation, number of incidents that end up with compensatory/punitive damages, expenses due to compensatory damages (e.g. average/total redress per upheld complaint), number of complaints, time to resolve a complaint, etc.
- Proactivity towards remediation. That is, an organization can take either a proactive or a reactive attitude with respect to remediation actions. Hence, remediation actions can be taken in a proactive manner by the organization, or in a reactive way, after complaints of the customer.

4.4 Liability

Liability is related to the consequences that must be faced if an organization is found responsible for not fullling its obligations.

Defining and differentiating liabililty and responsibility is pretty complex. On the one hand, responsibility is a requirement for liability to be established. On the other hand, although an entity might be responsible, it might not be considered at the end liable (for instance, due to an incident that happened, which the responsible entity could not predicted or prevented).

The first step towards eliciting relevant Liability metrics is to decide which are the actual consequences to consider. For example, if a re-definition of Liability

only considers economic consequences, then we can derive a set of economic-driven metrics (EDM). State of the art works on the EDM field (like Innerhofer [13]) have studied this topic in detail.

4.5 Observability

Observability can be defined as the ability of a system to expose (part of) its operations to authorized stakeholders.

An observable system will rely on processes and procedures for supporting these characteristics, and will provide means for demonstrating the existence and quality of such procedures. In the first case, an observable system will provide 'openings' for inspection, that is, means for independent inspection by third parties. In the second case, an observable organization must demonstrate and provide evidence of the low influence of unobservable actions in the state of the system. This aspect may be more difficult to fulfill. Hence, from a high-level viewpoint, an observability metric would measure the quality and effectiveness of such procedures.

Quality and effectiveness of observability can be assessed mainly from information based in certification from third parties. Organizations may be audited and/or certified by trusted third parties, who can then assert to what extent external inspections relate to internal system functioning.

4.6 Verifiability

Verifiability is a property of a process or system describing how well it implements and demonstrates the following practices:

- Compliance of process or system behaviour with rules is documentable.
- Continuous documentation.
- Scope of documentable compliance is a balance between benefits and costs.

A verifiable process or system will implement procedures for supporting these practices, and will provide means for demonstrating the existence and quality of such procedures. Accountability evidence relates to the documentation that should be collected in relation to compliance process. The scope of accountability evidence is based on the balance between benefits and costs.

4.7 Attributability

Attributability describes a property of an observation that discloses or can be assigned to actions of a particular actor (or system element). It implies the existence of two attributability processes:

- An evidence collection process that provides data regarding the effects of the actions of an actor in the system. For example, a logging component within an information system.

– An attribution process that maps evidence to actors. Log analysis is an example of this kind of process.

According to the attributability definition, attributability is independent of regulations, i.e., the attributability processes should function whether regulations exist or not. Accountability is what extends attributability by taking regulations into consideration.

A metric for attributability should measure the quality of the attributability processes of a system in order to ascribe actions to actors. Thus, when facing the assessment of attributability within an organization, the processes of attribution and evidence collection must be identified and described. These descriptions, which are considered the evidence of attributability, are what support a metric for attributability.

The implementation of an attributability metric has to be chosen depending on the use-case and the available evidence. For instance, in legal scenarios, it could be required that the observation is unambiguously and probably attributed to a set on entities (usually one). For example, the observation of the factual circumstances of processing might lead to the attribution of the role of data controllers to two or more entities (called joint data controllers). In this case, the set notation makes sense and the evidence must be good enough to reduce the set size of the set of entities that could have caused the observation to the minimum. In other scenarios, where strong indication for attribution is required, but not unambiguity, approaches based on information theory are more likely to yield the intended results. Aspects such as Data Stewardship, Data Lifecycle Management and Log Management also affect directly the quality of attributability of an organization. Thus, metrics for these subconcepts will be very useful for deriving metrics for attributability.

5 Measuring Accountability Attributes

In this section we propose a model-driven approach that includes the definition of a metamodel for describing metrics and accountability properties [14]. The goal of this metamodel is to serve as a language for describing: (i) accountability properties in terms of entities, evidence and actions, and (ii) metrics for measuring them.

One of the main features of this metamodel is that metrics are defined to take two main kinds of inputs: **Evidence** and **Criteria**. Any assessment or evaluation (i.e., a metric) can only be made using as input some tangible and empirical evidence, such as an observation, a system log, a certification asserted by a trusted party, a textual description of a procedure, etc. That is, a metric does not directly measure a property of a process, behaviour, or a system, but uses the evidence associated with them in order to derive a meaningful measure. Evidence is the fundamental support of any evaluation method and is what gives an objective dimension to assessments. Criteria are all the elements that convey contextual input that may constrain what should be measured, such as stakeholder's preferences, regulations and policies.

Next, we will describe the elements of the metamodel, which can be seen in Fig. 4.

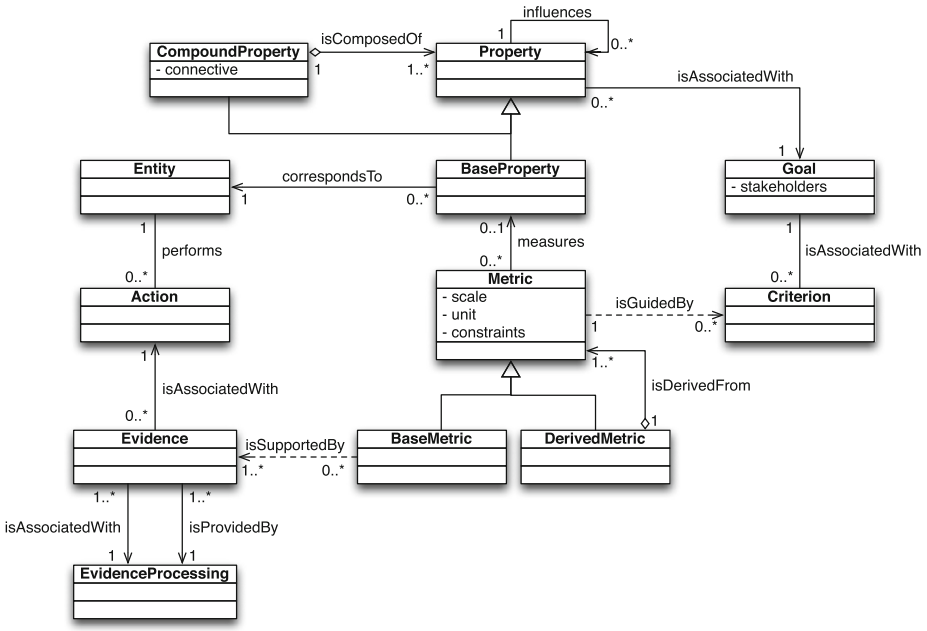

Fig. 4. Metamodel for metrics for accountability

- **Goal:** High-level description of the attribute (or family of attributes) that is modelled. These elements also contain a reference to the stakeholder (or stakeholders) for which the goal is oriented.
- **Attribute:** Attributes can be distinguished quantitatively or qualitatively by some evaluation method, however they may be defined as very-high level concepts. Thus, we consider that attributes can be further decomposed into more basic ones in some cases. In these cases, **BaseAttribute** elements can be defined in terms of entities and the actions between them, whereas **Compound Attribute** elements are defined in terms of other attributes, making possible a top-down decomposition of properties, from a high-level and abstract way to a tangible and more accessible one. In addition, attributes may also influence other attributes, not necessarily taking part of a composition relationship. The model then permits to express these influence relationships among attributes.
- **Entity:** This element is used to describe the entity that meets the modelled attribute. An entity is a physical or conceptual object that performs actions and that meets properties. For example, an organization, a process or a system can be considered as entities.
- **Action:** We define an action as a process that occurs over a period of time and it is performed by or has an effect on entities. Even though actions have

an effect in the environment, we cannot deal directly with these consequences, but with the evidence associated to them.

- **Evidence:** We define evidence as a collection of information with tangible representation about the effect of actions. Evidence is used to support a metric. That is, evidence is not an abstract concept about the consequence of activities, but actual data that can even be processed by a machine. Note, however, that evidence may come from sources with different levels of certainty and validity, depending on the method of collection or generation of such evidence.
- **EvidenceProcessing:** In our model, we assume that evidence, although it is associated to the effect of actions, does not directly stem from them. Instead, evidence is originated or collected by means of an EvidenceProcessing element. In this way, we model the fact that there may not exist a perfect correlation between the effects or consequences of actions and the evidence associated with them. The EvidenceProcessing element makes this difference explicit. With the inclusion of this element in our metamodel, we emphasise that the method of collection and processing of evidence is as important as the evidence itself. For this reason, there should also be evidence associated to each EvidenceProcessing element, describing how it works. Such evidence may be used by a metric during the evaluation process.
- **Metric:** We define it as an evaluation method for assessing the level of satisfaction of a non-functional property (or attribute in our case) in a quantitative or qualitative way, on the basis of evidence and contextual criteria. Metrics can be of two types: **BaseMetric** for metrics that use evidence as inputs for their calculations, and **DerivedMetric** for aggregated metrics that are defined as a function of other metrics. Aggregated metrics may rely on auxiliary metrics that are not associated with any attribute and that are defined solely for facilitating the definition of the parent metric. In both cases, metrics may use Criterion elements for guiding the evaluation with respect to the context of the metric. This element has the following fields:
  - Scale: This field describes the type of measurement scale used in this metric. The scale can be either nominal, ordinal, interval or ratio.
  - Unit: This field represents the measurement unit adopted as standard for measuring the property. The definition of a measurement unit is only necessary in the case of quantitative metrics.
  - Constraints: This field conveys the contextual constraints that may affect the application and validity of the metric.
- **Criterion:** This element captures all the contextual input that may constrain what should be measured by the metric, such as regulation, best practices, organisational policies and contracts, and stakeholders' preferences. It could be the case that one could define different metrics for the same attribute. The assessment methodology for each metric will depend on the contextual input given for the metrics evaluation. The Criterion element will be the responsible of conveying such contextual information.

6 Measuring Transparency

In this section we show how the metamodel presented in Sect. 5 can be applied to any of the accountability attributes. We have chosen transparency for our example. According to the definition of transparency, a metric for it would measure how an organization implements and demonstrates some practices related to how well deals with policies and procedures regarding data protection, as well as the quality of the transparency processes held in place by the organization. Figure 5 shows how transparency is modelled by using our metamodel.

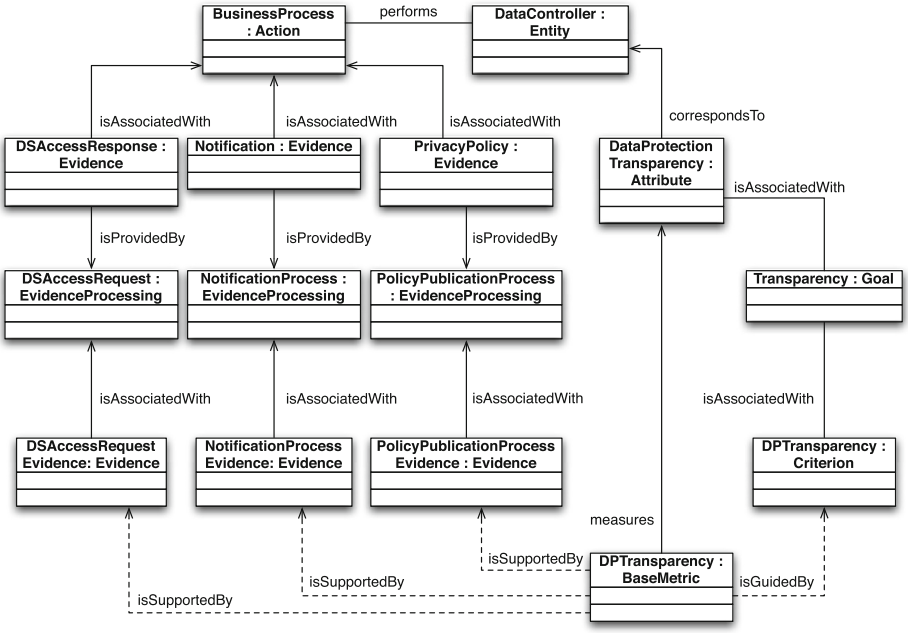

Fig. 5. Modeling transparency

The high-level goal in this example is represented by the **Transparency** element. This is a very generic goal that can have several properties associated to it. We are considering transparency with respect to data protection (**DataProtectionTransparency**). This property is defined upon an organization that acts as data controller (since it determines the purposes and means of the processing of personal data). In other words, a metric for this property would evaluate how transparent this organization (i.e., the **DataController** element) is with respect to data protection. In this example, the actions of the **DataController** are subsumed into one **Action** element and called **BusinessProcess**. One might want to be more specific and could model particular business processes, but in this case, it is not necessary.

The **DataController** must implement and demonstrate the transparency practices that we identified in Sect. 4 (informing stakeholders about data protection policies, notification of policy violations and responding to data subject access requests).

These practices are mapped to the following **EvidenceProcessing** elements:

- **PolicyPublicationProcess:** This element describes the internal procedures of the DataController with respect to the publication and communication of data protection policies to the relevant stakeholders.
  - **PrivacyPolicy:** This **Evidence** is produced by the **PolicyPublicationProcess**. The result of this process is a description of the data protection policy accessible by the relevant stakeholders. This element by itself is not relevant for measuring the property that we are considering as individual policies are not assessed by a metric for transparency, as such metrics focus on making the policies known. Thus, what it could be measured is the existence of these elements.
  - **PolicyPublicationProcessEvidence:** This is associated to the transparency process that published policies and its features. This element could answer questions like 'Are all the policies published?'
- **NotificationProcess:** This element is related to the practices of the **DataController** with respect to the notification of any violation of daa protection policies to relevant stakeholders. The **Evidence** elements associated to this element are:
  - **Notification:** This element represents the Evidence generated by the **NotificationProcess** in case of a policy violation.
  - **NotificationProcessEvidence:** This element describes the nature of the process of notification. This element answers questions such as 'Does a notification process exist?'
- **DataSubjectAccesssProcess:** This element represents the internal procedures of the **DataController** for permitting data subjects to request access to their data and for properly responding to such requests. The two **Evidence** elements associated to it are:
  - **DataSubjectAccessResponse:** This element is the evidence representing the response generated by the **DataSubjectAccessRequestProcess** in case of an access request from a data subject.
  - **DataSubjectAccessRequestEvidence:** This element represents a description of the characteristics of the process for permitting data subject access requests. This element answers questions such as 'Does a process for data subject access requests exist?'

It is the **Evidence** elements associated to the **EvidenceProcessing** elements, and not the evidence produced by them, the ones that are evaluated by the **DataProtectionTransparency** metric.

Based on the existence of the transparency processes identified by the definition of transparency, we could define a metric for **DataProtectionTransparency** (see Table 4).

Table 4. Naive example of a metric for transparency

Level	Description of the level
None	No transparency processes are implemented by the Data Controller
Low	One transparency process is implemented by the Data Controller
Medium	Two transparency processes are implemented by the Data Controller
High	All the transparency processes are implemented by the Data Controller

The stakeholders' criteria for this particular metric is conveyed by the **DataProtectionCriterion**. In this case the measurement is about the existence of transparency processes but the metric does not evaluate their quality.

A more complex metric could be one that counts the existence of a transparency process if it has been audited by a trusted third party. We could define an ordered scale as described in Table 5.

Table 5. Another example of a metric for transparency

Level	Description of the level
0	No transparency processes are implemented by the Data Controller
1	Only a process for data subject access requests is implemented by the Data Controller
2	Only a process for notification is implemented by the Data Controller
3	Either the process for publication of policies or the processes for notification and data subject access requests are implemented by the Data Controller
4	The processes for publication of policies and data subject access requests are implemented by the Data Controller
5	The processes for publication of policies and notification are implemented by the Data Controller
6	All the transparency processes are implemented by the Data Controller

Note that a different definition of transparency could lead to a different model; that is the reason why we consider that a first requirement towards creating metrics is agreeing on a clear, concise and stable definition of the property to be measured, so that an appropriate model can be defined.

7 Conclusion and Future Work

This paper lays the foundations for the development of metrics in the context of the A4Cloud project. This includes: (i) the definition of basic concepts for developing metrics, (ii) an analysis of the accountability attributes from the

metrics perspective, and (iii) a metamodel for describing such attributes and their metrics.

The analysis of accountability attributes has helped us to refine the concepts involved in the definitions and to identify plausible sources of evidence in order to support the evaluation of such attributes. The metamodel for accountability metrics constitutes the first step in the metrics elicitation process. It serves as a language for describing the accountability attributes in terms of entities and activities among them. Moreover, it also allows us to describe the sources of evidence involved in those activities and to identify the evidence elements that can be used to support metrics. Thus, this metamodel is a valuable tool for guiding the process of defining metrics.

In the future, we intend to apply the metamodel to all the accountability attributes and not only transparency. We are also going to explore a bottom-up approach for the elicitation of metrics, as well as the top-down approach that we have introduced in this paper. In this new approach we are going to use as input for the elicitation of the metrics, control frameworks and how they influence on accountability.

Acknowledgements. This work has been partially funded by the European Commission through the FP7/2007–2013 project A4Cloud under grant agreement number 317550. The first author is funded by a FPI fellowship from the Junta de Andalucía through the project PISCIS (P10-TIC-06334).

References

1. New Oxford American Dictionary
2. The Cloud Accountability Project. http://www.a4cloud.eu/
3. ISO/IEC 15939:2007 – Systems and software engineering – Measurement process (2007)
4. NIST SP 800-55 – Performance measurement guide for information security. National Institute of Standards and Technology (2008)
5. ISO/IEC 27004:2009 – Information Technology – Security techniques – Information Security Management – Measurement (2009)
6. Implementing accountability in the marketplace – a discussion document. accountability phase iii – the madrid project. Centre for Information Policy Leadership (CIPL), November 2011
7. Abran, A.: Software Metrics and Software Metrology. Wiley, New York (2010)
8. ENISA. Procure secure – a guide to monitoring of security service levels in cloud contracts (2012)
9. EU Parliament and EU Council. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
10. EU Parliament and EU Council. Proposal for a regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation) (2012)

11. Herrmann, D.S.: Complete Guide to Security and Privacy Metrics: Measuring Regulatory Compliance, Operational Resilience, and ROI. CRC Press, Boca Raton (2007)
12. Hood, C., Heald, D. (eds.): Transparency: The Key to Better Governance?, vol. 135. Oxford University Press, Oxford (2006)
13. Innerhofer-Oberperfler, F., Breu, R.: An empirically derived loss taxonomy based on publicly known security incidents. In: International Conference on Availability, Reliability and Security, ARES 2009, pp. 66–73. IEEE (2009)
14. Nuñez, D., Fernandez-Gago, C., Pearson, S., Felici, M.: A metamodel for measuring accountability attributes in the cloud. In: 2013 IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), vol. 1, pp. 355–362. IEEE (2013)
15. Pfitzmann, A., Hansen, M.: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management–A Consolidated Proposal for Terminology. Version v0.31 (2008)
16. A4Cloud project. MS:C-2.2 - Initial framework description report, February 2013
17. Stevens, S.S.: On the theory of scales of measurement. Science **103**(2684), 677–680 (1946)
18. Zhang, Y., Juels, A., Reiter, M.K., Ristenpart, T.: Cross-vm side channels and their use to extract private keys. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security, pp. 305–316. ACM (2012)