On the Measurement of Data Protection Compliance of Cloud Services

Thomas Kunz, Annika Selzer, Ulrich Waldmann

Fraunhofer Institute for Secure Information Technology SIT
Rheinstr. 75, 64295 Darmstadt
{thomas.kunz|annika.selzer|ulrich.waldmann}@sit.fraunhofer.de

Abstract: Companies want to benefit from the numerous advantages of cloud services such as flexibility and cost efficiency. However, cloud services vary considerably with respect to the security and privacy mechanisms provided. Moreover, security-aware companies complain the lack of transparency concerning the security measures and processes the cloud provider has installed. As a solution for the latter one, auditors may evaluate cloud providers and issue certificates attesting whether the cloud provider meets the agreed requirements. However, due to the characteristics of cloud computing, on-site inspections in the data centers of a cloud provider do not seem to be realistic. In this paper we show how metrics can be derived from data protection requirements and how these metrics can be expressed in the form of formal policies, in order to be used for an automated evaluation of cloud services[1].

1 Introduction

Cloud services promise numerous advantages for companies with respect to flexibility and cost efficiency. They are hosted in data centers that are professionally managed, physically protected, and equipped with failsafe hardware. However, in business context, security and data protection aspects become very important [CPA12]. This is one of the reasons why many companies still have concerns regarding the use of cloud services: on the one hand, companies are still responsible for the protection of their data, even if they shift the storage and the processing of the data into the cloud. But on the other hand, they do not trust the cloud providers in terms of confidentiality and data protection even though they expect the cloud providers to act according to their contractual agreements.

In [KSW14] and [KNW13] we proposed an automatically generated data processing certificate for cloud services, in order to compensate this problem. This solution is based on the standardized audit suggested by the working group "Legislation" of the "Trusted Cloud"-technology program [Sel13][B+12][2]. We use secure log files as a trustworthy data base for the evaluation of a cloud service. For this reason, we introduced a secure logging

[1]This contribution has been created within the project CloudCycle (`http://www.cloudcycle.org`). CloudCycle is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) according to a decision of the German Federal Parliament within the funding priority "Trusted Cloud".

[2]`http://www.trusted-cloud.de`

method which is resistant against internal manipulation attempts through tamper-proof and confidential log data. In order to guarantee an automated control of the cloud service, we assumed that the underlying security and data protection requirements are described in a formal policy. Unfortunately, typical data protection requirements based on legal regulations like the Federal Data Protection Act (BDSG) are too generic for an automated evaluation. They do not define what has to be measured to control the requirements and which measured values are sufficient to fulfill the requirements.

In this paper we show how metrics may help to derive formal data protection policies from rather abstract data protection requirements. The metric definitions are essential for specific measurements that in turn are compared with parameters specified in the policies. This way an automated evaluation of cloud services seems possible.

The structure of this paper is as follows: Related work is discussed in Section 2. Section 3 shows how generic data protection requirements can be concretized in the form of metrics. Section 4 illustrates the translation of the metrics into formal data protection policies. The paper ends with an outlook in Section 5.

2 Related Work

To be able to measure the quality of data protection, quantitatively measurable indicators are needed – both for the technical components and for the organizational aspects. A promising project is the definition of security metrics to support management decisions in information security which is led by the Center for Internet Security (CIS). The current version from 2010 contains 28 definitions of metrics in the fields of incident management, vulnerability and patch management, application security, configuration management, change management, and financial indicators [CIS10]. However, the metrics focus rather on organizational and financial security indicators than on data protection indicators as required in BDSG.

[Jul09] also deals with typical security metrics like "number of security incidents". The paper introduces three properties that all security metrics have to fulfill (validity, accuracy, and precision), and gives a formal definition of the term "security metric". Furthermore, it classifies security metrics by means of the input data analyzed by the metrics. In [Böh10], Böhme covers security metrics from an economic perspective. In [Bez10], Bezzi provides an information theoretic approach for privacy metrics (e. g. k-anonymity) in order to express and compare disclosure risk metrics.

3 Derivation of Metrics from Data Protection Requirements

Data protection requirements derive primarily from legal regulations such as the BDSG. In Germany, the contract data processing according to Sec. 11 BDSG is crucial within the cloud computing context. One main requirement is taking technical and organizational

measures to protect the personal data processed by the processor. The annex of Sec. 9 BDSG includes eight requirements for technical and organizational measures like access control and disclosure control. In the remainder of this section, we will consider examples of measures and metrics for the following two requirements of the BDSG annex:

- Protection of data against accidental destruction or loss (availability control) and
- Separate processing of data collected for different purposes.

These generic requirements are hardly precise enough for an automated evaluation of cloud services. Looking at the examples above, it is unclear how to decide whether a cloud provider sufficiently cares for the availability of the customers' data and whether the customers' data have been separated adequately or not. A concise specification of technical and organizational measures is required, in order to ensure the conformance with these requirements. So, the next step is to search for appropriate measures.

3.1 Selection of Measures

Since cloud customers often rely on the IT-Grundschutz methodology [BSI08] that includes threat analysis, requirements and measures regarding different IT security aspects, it may be a good source of measures for data protection. The methodology suggests that a list of possible requirements is divided into three protection categories: "normal", "high", and "very high". Then it may be easier to identifier the damages that could occur under loss of protection and to find appropriate countermeasures.

The availability control obviously comprises many different aspects as it regards all parts of a process, i. e. data, IT systems, facilities and personnel. The creation of backups and redundancy are standard measures, others are the usage of virus scanners or the defense against DoS attacks. The protection category may influence the strength of selected measures, as regards measures of backups e. g., the backup frequency and storage period. Measures to contribute to the separate data processing may be implemented, e. g. on application level (use of separate databases or software with multi-client capability), on virtual level (providing each customer with an own virtual machine) or, for even higher protection, on hardware level.

It should be noted that protection requirements are often interdependent and even opposing each other like availability and confidentiality to control disclosure: whereas perfect availability could be realized by publishing in the Internet, the availability of data should probably be restricted to authorized parties. On the other hand, one measure may contribute to more than one requirement, e. g. cryptographic encryption may help to prevent unauthorized access as well as disclosure. In general, a whole bundle of carefully balanced measures is needed to fulfill generic requirements and adhere to the data protection principles of data economy, data avoidance and transparency. A survey of measures in relation to several data protection requirements is given in [Pro12].

Additionally, cloud customers may define further individual protection requirements, such

as backup locations restricted to a certain Federal State in Germany. Both requirement types determine the measures a cloud provider has to implement. In the following, we derive two sample metrics from a backup measure and a measure for separate data processing.

3.2 Definition of Metrics

After deploying selected measures, it has to be determined how to decide if a measure has been adequately implemented. This can be done by a system of metrics that effectively relates to measurements of desired compliance features. Each metric should represent a part of the real compliance state and have the following characteristics. First, a good metric is quantifiable as it is based on quantitative measurements. Second, it is repeatable yielding the same result in a repeated assessment, and third, it is comparable as it is based on a linearly ordered set of normalized values. Each metric should provide useful information for a continuous business improvement towards a defined target value. Most metric definitions include measurement units combined with a time unit (e. g., per week or mean-time between two activities) or a proportion to a total (e. g., percentage) [ENI12], and a compliance target (setpoint) that will be considered in relation to regular measurements (actual values).

The Center of Internet Security (CIS) has developed a set of agreed security metrics including the underlying formula and units, recommended measurement frequencies, target values and possible measurement sources [CIS10]. The target describes the good and bad value ranges - including ideal targets, e. g. zero security incidents, or defines a precise threshold to decide whether the measured value can be considered compliant with the generic security requirement. The following two tables show data protection metrics defined with similar attributes. Table 1 shows the definition of a sample metric for the "backup" measure. It may contribute to the indicator for the degree of availability control.

Name	**Mean Time Interval between Backups**
Description	The average time between the last two backups of all storage media with the data of a specific customer.
Formula	The Mean Time Interval between Backups (MTIB) is calculated by determining the number of hours between the date of the last backup and the date of the current backup for each storage medium. These results are averaged across the number of the relevant storage media: $$MTIB = \frac{\sum(CurrentBackup\_Date - LastBackup\_Date)}{Count(Data\_Storage\_Media)} \quad (1)$$
Units	Hours per backup
Frequency	Daily, weekly, monthly, quarterly, annually
Targets	MTIB values should trend lower over time. Generally, the target time for MTIB depends on the desired availability level, e. g. normal: 24 hours; high: 6 hours; very high: 1 hour.

Table 1: Availability metric "Mean Time Interval between Backups"

In general, for a single measure more than one metric is required. For example, in the case of backups this may include further metrics such as "The mean distance between the storage location of the original data and their backup", and "The average storage period of backup data". As a result, we would end up with a whole catalog of metrics. Table 2 shows the definition of a sample metric for a "Customer-specific separate data processing". It may contribute to the indicator for the degree of separate processing of data collected for different purposes.

Name	**Separate Processing Coverage**
Description	This metric attempts to answer the question "Are the data of a specific customer processed separately from other customers' data?". The metric's unit is the percentage of total storage and processing media that are exclusively allocated for a specific customer or have data-protecting multi-client capability. Examples of media may include applications, virtual machines, storage systems like databases, server components, hardware components.
Formula	Separate Processing Coverage (SPC) is calculated by determining the number of storage and processing media with data-protecting multi-client capability and then averaging this across the total number of storage and processing media: $$SPC = \frac{\sum(MultiClient\_StorageProcessingMedia)}{Count(Media)} * 100 \quad (2)$$
Units	Percentage of storage and processing media
Frequency	Daily, weekly, monthly, quarterly, annually
Targets	The expected trend for this metric over time is to remain stable or increase towards 100%. The SPC scope depends on the desired separation level, e. g. normal: on application level by means of access control rules; high: on VM basis or separate database with access control rules; very high: on hardware level

Table 2: Metric "Separate Processing Coverage"

How can an effective system of metrics contribute to a continuous control of compliance features? The individual metrics should be aggregated, in order to calculate the significant compliance indicators, e. g. one indicator for each data protection requirement (each based on several metrics). The cloud provider may provide currently determined indicators that enable an evaluation of data protection-relevant qualities at every stage of a cloud service (service offer, instantiation, configuration, usage, termination). A broker service may use the indicators for advertising and matching the service qualities with individual customer requests. The cloud provider may empower the cloud customers to verify the compliance during operation by means of a reporting dashboard that visualize the relevant indicators.

However, there are obstacles to implement a customer-centric system of metrics. Most existing cloud services do not care about sector-specific requirements or even customers' individual requests. This is because "One size fits all" cloud solutions seem more effective and cost-efficient from the perspective of cloud providers. In the next section we show how machine-readable policies can be derived from metrics, in order to enable an automatic evaluation of cloud services. This may help to overcome the economic obstacles to customized services and metrics.

4 Data Protection Policies

Machine-readable policies are written in a formal language and may convey the properties of metrics into a technical control system for cloud services. The metrics precisely define what has to be measured to control the requirements and which measured values are sufficient to fulfill the requirements. The metrics and the corresponding thresholds serve as input for policies which in turn are the base to automatically evaluate the requirements.

Further use cases for such policies exist: applications may use them to automatically match cloud services to the customer's requirements. They may be used by cloud management environments to enable an automated enforcement of the security requirements. For these use cases, the policies should be standardized, in order to enable a widespread support through many cloud management environments.

A recently initiated standardization effort from OASIS, the "Topology and Orchestration Specification for Cloud Applications" (TOSCA) [LMPS13], provides an interesting approach. Using TOSCA, cloud service creators are able to describe the topology, the deployment, and the management of portable cloud services. TOSCA allows an automatic deployment and management of cloud services through the cloud platform. In addition, TOSCA allows the definition of policies that describe requirements of any kind. Current research deals with the definition of non-functional requirements in TOSCA as policies and how these policies can automatically be processed by a cloud environment [WWB+13].

In the following, we present two example policies that implement the data protection metrics in Section 3. The chosen syntax is close to the syntax of the TOSCA policies.

```
<PolicyTemplate type="BackupInterval">
  <Properties>
    <cc:Threshold>
      <cc:SecurityLevel value="normal">
        <cc:Frequency interval="24" timeUnit="hours"/>
      </cc:SecurityLevel>
    </cc:Threshold>
  </Properties>
</PolicyTemplate>
```

Listing 1: Policy for Backup Frequency

```
<PolicyTemplate type="DataSeparation">
  <Properties>
    <cc:Threshold>
      <cc:SecurityLevel value="high">
        <cc:Separation value="VM_Level"/>
      </cc:SecurityLevel>
    </cc:Threshold>
  </Properties>
</PolicyTemplate>
```

Listing 2: Policy for Data Separation

In order to enable an automated control of the cloud service, the data protection requirements are described in metrics and then transformed into formal policies. As part of a formal description of a cloud service the policies may contribute to customized services that are - due to the automation potential - not necessarily more complex or costly than the conventional pre-built solutions.

A software component capable of interpreting these policies is required for an automated evaluation of cloud services. Such a component has to implement the formulas presented in Section 3 to measure the thresholds defined in the policies. Furthermore, for performing the analysis, the component needs access to a trustworthy data basis such as secure log data [KSW14].

5 Outlook

In this paper we have shown how metrics can be derived from abstract data protection requirements and how in turn these metrics can be expressed as formal policies. Such policies may contribute to an automatical control of the technical and organizational measures implemented by the cloud provider. These controls could be performed at certain time intervals, e. g. every year, and even at any time during regular operation as well. The results of the controls may serve as input for data protection certificates as proposed by the working group "Legislation" of the technology program "Trusted Cloud" [B+12].

The metrics and corresponding indicators can be used for an independent certification of individual cloud services. Most existing cloud certificates do not take account of sector-specific requirements. In general, customers are not able to evaluate the quality of certificates or to compare the offers of different certification bodies. Furthermore, most available cloud certificates refer to general qualities of the cloud provider rather than to specific cloud services [SLS13].

Future research should focus on developing new tools that enable companies and authorities to describe and measure automatically standardized data protection requirements. In order to make the privacy characteristics of different cloud providers comparable, a catalog of data protection requirements and measures including measurement characteristics has to be defined. Such a catalog could serve as a guidance for cloud providers to implement sector-specific requirements (such as for health care or financial services) as well as for auditors to define corresponding certification practice statements and evaluate the level of technical and organizational measures taken by the cloud provider.

Further efforts should address the technical, economic and legal challenges of implementing a system of data protection metrics, its effectiveness (attainment of objectives) and efficiency (minimized investment). For example, a cloud provider should not be able to selectively optimize the specified measurements without really improving the data protection level (strategy resilience). These challenges are within the scope of the BMBF-sponsored follow-up project VeriMetrix[3].

[3] http://www.verimetrix.de

References

[B+12] G. Borges et al. Datenschutzrechtliche Lösungen für Cloud Computing. Kompetenzzentrum Trusted Cloud, Oct 2012.

[Bez10] M. Bezzi. An information theoretic approach for privacy metrics. *Transactions on Data Privacy*, 3(3):199–215, 2010.

[Böh10] R. Böhme. Security Metrics and Security Investment Models. In I. Echizen, N. Kunihiro, and R. Sasaki, editors, *Advances in Information and Computer Security*, volume 6434 of *Lecture Notes in Computer Science*, pages 10–24. Springer Berlin / Heidelberg, 2010.

[BSI08] BSI Standard 100-2: IT-Grundschutz-Vorgehensweise. Technical report, Bundesamt für Sicherheit in der Informationstechnik, 2008.

[CIS10] The CIS Security Metrics v1.1.0. Technical report, The Center for Internet Security, 2010.

[CPA12] R. Contu, L. Pingree, and E. Ahlm. Predicts 2013: Security Solutions. Technical report, Gartner, Nov 2012.

[ENI12] Critical Cloud Computing - A CIIP perspective on cloud computing services, Version 1.0. Technical report, European Network and Information Security Agency (ENISA), Dec 2012.

[Jul09] K. Julisch. A Unifying Theory of Security Metrics with Applications. Technical report, IBM Research – Zurich, 2009.

[KNW13] T. Kunz, P. Niehues, and U. Waldmann. Technische Unterstützung von Audits bei Cloud-Betreibern. *Datenschutz und Datensicherheit – DuD*, 37(8):521–525, 2013.

[KSW14] T. Kunz, A. Selzer, and U. Waldmann. Automatic Data Protection Certicates for Cloud-Services based on Secure Logging. In *WTC 2013 : Wissenschaftliche Ergebnisse des Trusted Cloud Technologieprogramms*, 2014.

[LMPS13] P. Lipton, S. Moser, D. Palma, and T. Spatzier. Topology and Orchestration Specification for Cloud Applications (TOSCA),Version 1.0, Mar 2013.

[Pro12] T. Probst. Generische Schutzmanahmen für Datenschutz-Schutzziele. *Datenschutz und Datensicherheit - DuD*, 36(6):439–444, 2012.

[Sel13] A. Selzer. Die Kontrollpflicht nach §11 Abs. 2 Satz 4 BDSG im Zeitalter des Cloud Computing – Alternativen zur Vor-Ort-Kontrolle des Auftragnehmers durch den Auftraggeber. *Datenschutz und Datensicherheit – DuD*, 4:215–219, 2013.

[SLS13] S. Schneider, J. Lansing, and A. Sunyaev. Empfehlungen zur Gestaltung von Cloud-Service-Zertifizierungen. *Industrie Management – Zeitschrift für industrielle Geschäftsprozesse*, 4:13–17, 2013.

[WWB+13] T. Waizenegger, M. Wieland, T. Binz, U. Breitenbücher, and F. Leymann. Towards a Policy-Framework for the Deployment and Management of Cloud Services. In *SECURWARE 2013 : The Seventh International Conference on Emerging Security Information, Systems and Technologies*, 2013.