Selected Cloud Security Patterns to Improve End User Security and Privacy in Public Clouds

Thomas Länger[1], Henrich C. Pöhls[2], and Solange Ghernaouti[1]

[1] Swiss Cybersecurity Advisory and Research Group (SCARG), Université de Lausanne, Switzerland
[2] Institute of IT-Security and Security Law (ISL), Universität Passau, Germany

Abstract. Cloud computing has the potential to dramatically reduce the cost and complexity of provisioning information technology resources for end users. However, to make it secure and privacy-preserving for end users, additional technical safeguards must be added—the application of strong cryptography is such a safeguard. The Horizon 2020 project PRISMACLOUD surveys and advances several cryptographic protocols and primitives usable to cryptographically address common cloud security and privacy issues. The cryptographic functionality will entirely be encapsulated in five configurable tools, from which cloud services providing end-to-end security can be constructed. This approach relieves cloud service designers from dealing with the complex and error prone correct application of cryptographic functionality and shall spark the emergence of a multitude of privacy and security preserving cloud applications for the benefit of the end-users—who will no longer have to rely on contractual and legal instruments for ensuring, that privacy and security is enforced by cloud providers on their behalf. In order to support the privacy-by-design development of the tools, we developed several cloud security patterns for common critical situations in the cloud—in the three fields of data storage in the cloud, user privacy protection and data minimisation, and authentication of stored and processed data.

Keywords: cloud computing, privacy, security, user centric security, cloud security pattern, end-to-end security, cryptography, security-by-design

1 Introduction

1.1 Significance of Cloud Computing

Cloud computing[3] is the major growth area in information and communication technologies today, and with its huge processing capabilities and data storage architectures, and with all the data which is amassed, and even created through

[3] The authors' work is supported by the European Union Horizon 2020 research activity n° 644962 PRISMACLOUD: "Privacy and security maintaining services in the cloud" [17]; duration 2/2015-7/2018; 16 partners; `https://www.prismacloud.eu`

its use, it is closely related to another major growth area in Information and Communication Technologies (ICT), that of big data aggregation, processing and analysis. With an estimated size of about 150 billion US-Dollar an enormous rush to move into cloud computing is observed [28] [23]. The American business magazine Forbes has an overview of several forecasts and market estimates [13]. As a recent report by the Economist says: "Cloud technologies have gone mainstream" [27]. Today's biggest players to provide these capabilities are in fact companies which have enormous financial power at their disposal and are proficiently experienced in the field of ICT. They now aim at increasing revenue and domination in the developing information age, and invest huge efforts in the construction of new data centres and in new technologies for asserting their leading positions.

The biggest cloud provider today [24], Amazon.com Inc., started as an online book store in 1994 and has been generating enormous wealth as an e-commerce retailer. Since 2006, Amazon offers public cloud services (Platform as a service—PaaS, which it initially has developed to cater for its own retail infrastructures) on a commercial basis. The second and third biggest providers are Microsoft Corp. and Google Inc. (now the holding company Alphabet. Inc.) [24], who made their fortunes in Personal Computer operating systems and office software, and in search engines and internet advertising business, respectively. Besides the above mentioned three cloud providers, there are many other providers and players competing in this field over markets and governance of our future society.

1.2 Security Problems

In the history of ICT innovation several comparable situations are known, when companies have rushed into a newly developing market, while at the same time also shaping the market. In such a hurry, developments often do not respect the requirements and needs of the end users—but rather the needs of the companies, which want to grow quickly. The price in these situations is often paid by the end users: Systems and services are made available on a large scale before the data privacy and security concerns of the customers are fully addressed and resolved.

This situation, for valid reasons, keeps security aware customers currently away from the cloud—be it because they are forced by regulation to guarantee a certain degree of confidentiality for the data they are operating with (e.g. in the health sector, or in e-government), or that they are just companies, or individuals, who highly value the security of their data.

A comprehensive and authoritative Cloud Computing Security Risk Assessment is maintained by the European Union Agency for Network and Information Security (ENISA) [8, 9]. It references data protection risks, risks connected to governance and control, as well as technical risks related to cloud computing. Many of these risks can effectively be countered in the secure cloud services, that can be built from the PRISMACLOUD toolbox.

1.3 Proposed Solutions to Improve Cloud Privacy and Security

The European Commission, in its endeavour to strengthen European competitiveness and in its struggle to maintain European sovereignty over the data which is being moved to the cloud, has developed a proprietary European Cloud Computing Strategy [11], and supports the development of secure cloud systems in their Horizon 2020 strategic programme [10] of which the project PRISMACLOUD [17] is a part. The Commission recognises the enormous cost reduction potential of a move to the cloud for companies and entities of all sizes. Foremost, it recognises the strategic importance of a European share and participation in the development and commercialisation of cloud computing products and services, and what is more, the strategic importance of maintaining sovereignty by not losing "European data" to opaque conglomerates beyond European data protection legislation and control.

Whether European research and development will be able to economically contest with its American competitors on providing the basic cloud services on a large scale is questionable: Today, almost the entire cloud business is based in the United States of America, in the area of Seattle, Washington and in California in the San Francisco Bay Area. It is also there, and in huge data centres all across the United States, where the clouds are physically hosted, and the data is stored and processed.[4] European industries compete in the shadow of the American market giants, like in many other major fields of ICT. Yet, the European Commission sees an opportunity to focus on original European strengths of data security and privacy protection for the benefit of the end-users and customers.

The PRISMACLOUD project will use a privacy-and-data-protection-by-design approach [6] [16] and provide the *advanced cryptographic tools* (in form of a software library which can be parametrized in various ways) for implementing privacy and security aware services on top of a potentially untrusted cloud. Thus, end users' effective governance and control over the storage and processing of their data shall be reinstated, following the spirit of the new European General Data Protection Regulation which has been adopted in June 2016. The feasibility of the PRISMACLOUD approach shall be validated in *eight sample cloud services* which will be provided as reference implementations: Data sharing service, secure archiving service, privacy enhancing identity management service, selective authentic exchange service, verifiable statistics service, infrastructure attestation service, anonymisation service, and encryption proxy service. The applicability of the services in real-world applications shall be verified in *three pilot applications* in the fields of Smart Cities, e-Health, and e-Government.

[4] It is now, that cloud providers have started to host their data centers in multiple locations world-wide, including Asia, South America, and countries of the European Union (see e.g. Amazon: `http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html`). Nevertheless, the headquarters and main installations of these businesses are certainly under U.S. American jurisdiction and it is at least possible that data, in whichever form and state of aggregation, might be consolidated with data residing in the U.S.A.

1.4 Contributions and Outline

This paper concentrates on the very tangible problem of how to practically tighten and increase for end users the security and privacy of data and computations in cloud settings, by applying suitable cryptographic tools. The PRISMACLOUD paradigm provides the tools encapsulating cryptographic protocols and primitives, thus enabling the required end-to-end security—much in the same way as encryption and digital signatures enable end-to-end security for communications over untrusted networks. In order to secure the aspired results, developers and application designers need to develop and use the suitable cryptographic tools right. To this goal, we developed nine cloud security design patterns, communicating and addressing the often conflicting requirements from different actors and explaining which existing cryptographic building blocks can be used to achieve the required functionalities.

In the Introduction (section 1) we framed the security context for end users in untrusted clouds. In section 2 we provide an introduction to the capabilities of design patterns in general by a historical approach on their evolution from architectural design patterns through software design patterns to cloud security patterns. In section 3 we present an overview of the *nine patterns* developed in the framework of the PRISMACLOUD project in the fields of *(i) data storage in the cloud*, *(ii) user privacy protection and data minimisation*, and *(iii) authentication of stored and processed data* and go into detail for one pattern of each of the three fields.[5] In Section 4 we introduce the *five configurable tools* which will be developed in the project, and list the cryptographic protocols and primitives they are composed of, as well as example services which can be built from them. The services' functionality and practicability will be evaluated by three pilot applications in the fields of Smart Cities, e-Health, and e-Government by project end. In Section 5 we present conclusions.

2 Design Patterns

2.1 Representation of Knowledge in Design Patterns

The Viennese Christopher Alexander, who has since 1963 been living and teaching in Berkeley, California, published his book "A Pattern Language: Towns, Buildings, Construction" [1][6] in 1977, where he and his co-authors introduced the concept of reusable design solutions for architectural problems. The idea behind the architectural patterns is to provide a collection of proven solutions for problems which occur over and over again. The 253 presented patterns contain the concentrated knowledge and experience of designers and are intended to be reused. Alexander defines a pattern language as a collection of patterns

[5] The other patterns can be studied in the public PRISMACLOUD deliverable D2.2 "Domain independent generic security models", available on the project web site `www.prismacloud.eu`.

[6] The entire book, 1218 pages, can be downloaded as pdf from `archive.org/details/APatternLanguage`.

from a specific domain. The proposed patterns were intended to be "alive and evolving". Alexander viewed them as "hypotheses", as "current best guess", to be improved and possibly replaced with more profound patterns, as a result of "new experience and observation". The idea of design patterns was taken up again in 1994 by computer scientists and especially software engineers who tried to tackle the reusability of software with a software design pattern approach. Reusability of software was then, after about 20 years of object oriented design, a big issue. The resulting book "Design Patterns: Elements of Reusable Object-Oriented Software" [14] has become a standard and has not lost its significance and relevance in software engineering today. The problem setting in software engineering is comparable to that in the field of architecture: Not to "solve every problem from first principles", but instead use a proven solution to a design problem.

The idea of design patterns was applied to other contexts as well. Security patterns, or security design patterns "codify basic security knowledge in a structured and understandable way" [25]. They represent a practical means to communicate end user needs and requirements. Security patterns are connected to one or more specific security goals. The Internet Privace Engineering Network (IPEN) of the European Data Protection Supervisor supports "(re)-usable building blocks, design patterns and other tools for selected Internet use cases where privacy is at stake".[7] IPEN's objective is "to integrate data protection and privacy into all phases of the development process (...) It supports networking between engineer groups and existing initiatives for engineering privacy into the Internet."[8] A comprehensive collection of security patterns which were discussed at the annual "Pattern Languages of Programs" (PLoP) conferences since 1997, is available on the homepage of the security researcher Munawar Hafiz (Auburn University, Alabama, USA).[9] It currently contains a catalogue of 97 security patterns. There is also on-going work on privacy patterns, which connect problems to solutions within the context of user privacy. The ability of design patterns to communicate and address the often conflicting requirements from different actors in different domains, is ideal for their application in designing information privacy into information systems: "Privacy Patterns that span across usability, engineering, security and other considerations can provide sharable descriptions of generative solutions to common design contentions. Since patterns focus on describing the resolutions of contradictory forces in a design context, the pros and cons of a specific solution can be easily debated. Unlike guidelines, regulations or best practices, patterns are descriptive, rather than normative, facilitating discussion and debate and providing education rather than insisting on particular solutions or practices" [7]. There are several websites online for joint develop-

[7] `https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/IPEN`

[8] ibid.

[9] `www.munawarhafiz.com/securitypatterncatalog/index.php`. Munawar Hafiz is also author of several papers on security patterns, e.g. [15], which presents 4 design patterns that can aid the decision making process for the designers of privacy protecting systems.

ment of privacy design patterns, like `privacypatterns.org` by researchers of the University of California, Berkeley, School of Law (funded with grants from the U.S. Department of Homeland Security and from the NIST, among others), and the `privacypatterns.eu`—resulting from the European FP7 project PRIPARE (Preparing industry to privacy-by-design by supporting its application in research).[10]

2.2 Assumptions and Categories for the Pattern Descriptions

The cloud security patterns do not represent "hard requirements" on cloud applications and services, the patterns represent more a way of communicating a user need (and specifically a security need) to the system architects and developers of the services in an informal way. The system architects and developers themselves shall read from the pattern the information enabling them to develop the cryptographic building blocks in such a way, that the applications and systems using these building blocks, satisfy end users' security and privacy needs.

Different publications about security patterns (and about design patterns in general) define the patterns along different categories. We have taken into consideration the categories used in [1] [14] [25], as well the categories used on the security pattern websites `cloudcomputingpatterns.org` and `cloudpatterns.org` and have chosen a synthesis that seems suitable for us. We use the same main categories as in Alexander's et al. seminal pattern book [1] (problem, solution), as do all the other sources and complement them with other categories (intention, building block, consequences and countered threats).

3 PRISMACLOUD Cloud Security Patterns

3.1 Overview of Cloud Security Patterns

The nine cloud security patterns have been developed in the first year of the PRISMACLOUD project, in order to better understand the end user "situation" currently prevailing in cloud storage and computing. In the practical project context, the patterns will serve as additional input in the design phase of the PRISMACLOUD tools in another project work package. But the cloud security patterns will also provide input to an "impact analysis of cloud usage for end users", a main deliverable of the project, providing guidance for corporate, governmental, and individual end users in their confrontation with cloud services.

The nine cloud security patterns have been designed to varying level of detail and will, as design patterns are generally intended to be "alive and evolving" [1], be further developed while the PRISMACLOUD research activity continues. Because of space constraints, we will present here only one selected pattern from each of the categories *(i) data storage in the cloud*, *(ii) user privacy protection and data minimisation*, and *(iii) authentication of stored and processed data*. For the other patterns (which are not presented in detail), we give a summary

[10] `www.priparepproject.eu`

description after the short introductions to the single fields, in order to telegraph the basic "situation", and the idea behind the solution.[11]

3.2 Field 1: Data Storage in the Cloud

The security of data at rest represents one of the most fundamental problems regarding privacy. Too often data confidentiality is regarded as being easily fixable by "just employing client-side encryption". While this solution is viable, it requires the effort of a fully fledged infrastructure for managing cryptographic keys in order to still enjoy one of the true cloud benefits—the ability to share data with ease. There are two patterns in this specific field:

- *Pattern 1: Secure cloud storage by default* is applicable in any context where a user wants to securely store or share data objects in a cloud infrastructure.
- *Pattern 2: Moving a legacy application's database to the cloud* is applicable when an end user wants to deploy an existing database to a public cloud.

We describe only Pattern 1 as an example in the following.

Pattern 1: Secure Cloud Storage by Default

Summary. Describes the qualities of a cloud storage service, as most users would expect it when moving their digital assets to the cloud: The data in the cloud storage remains readily available when needed, and dependably and securely confidential against the cloud provider and other tenants in the vicinity of the cloud, as well as against other third parties which are not entitled by the user to access the data. The data may easily be shared with others, and easily be transferred to another cloud provider when the user wants to do so.

Intention. Provide a cloud storage service with strong confidentiality, integrity, and availability, from which the cloud user can anytime effectively pull away the stored data.

Problem. Currently, most cloud storage providers store the data either unencrypted, or apply encryption which remains completely under their control; some cloud users locally encrypt their data before they store it in the cloud in order to maintain the confidentiality of the data.

Whether the cloud provider encrypts or does not encrypt the data it stores, the cloud provider has in practice full access to the data—if it is not encrypted by the user in the first place. In many cases, especially in free-of-charge public cloud services from the big cloud providers, the end users have to consent to terms-of-reference granting the provider full rights to the data (including rights to store,

[11] For a more detailed description of all cloud security patterns we want to direct the attention to PRISMACLOUD deliverable D2.2 "Domain independent generic security models", available on the project web site `www.prismacloud.eu`

combine, or otherwise use the data in ways non-anticipated and not explicitly consented to by the user, in order to be able to sell or commercialise the data in any other imaginable way). Nevertheless, also in commercial cloud services, the cloud provider has to be trusted to maintain the confidentiality of the data—by not looking at the stored data itself, and by effectively protecting it against access by unauthorised third parties. This includes all copies and replications of the data which are created for availability purposes in all layers of a storage architecture.

Also with respect to availability of data and of cloud services, the user is dependent on the provider. There are cases known, where bankruptcy of a cloud provider led to sudden loss of access to customer data. Deletion of data in clouds is also a big issue and it is not sufficiently solved how an effective deletion of data in all replications and backups can be achieved and substantiated.

When cloud users use end-to-end encryption to mitigate some of the mentioned problems and threats they are required to implement and maintain a cryptographic key management system and an access control mechanism, with all its known complexities and implications.

Solution. Cloud users do not want to give up their property rights and privacy rights on the data. Cloud users want to maintain full control over their cloud storage by default. They want strong confidentiality guarantees by default, while being able to share data with other cloud users or with the cloud provider at their own discretion. The data needs to be protected against loss by some kind of redundancy in a way that the confidentiality remains upheld. The cloud user wants to be able to withdraw the data from one cloud provider and give it to another provider for hosting at any time without having to rely on any form of cooperation with the cloud provider. The cloud user wants to be sure that the data can be completely withdrawn, with no copies of the plain information remaining at the provider.

Building Block. PRISMACLOUD proposes the *cryptographic storage solution tool* with increased practical usability for the secure, distributed storage of data. This tool uses information dispersal, based on a *secret sharing primitive* [26, 2, 4].

Consequences and countered threats. The pattern *secure cloud storage by default* counters almost all identified risks related to confidentiality, integrity, and availability of stored data in the cloud and therefore constitutes a disruptive technology of highest potential. A cryptographically secure storage solution can potentially entirely transform cloud provisioning world-wide. One new assumption which is introduced by this tool is the non-collusion assumption, i.e. that sufficiently many of the cloud providers do not maliciously cooperate to discover the secret. This means, that the number of shares necessary to reconstruct the secret in the threshold scheme of the information dispersal algorithm is a crucial design parameter. The non-collusion assumption can only be substantiated by other assumptions on the trustworthiness of the single involved cloud

providers. However, that risk can be deliberately reduced by continuous renewal and replacement of the shares. This reduces the attack window for procuring a sufficient number of shares for reconstructing the information. On the other hand, the data owner does not have to rely on computational assumptions for the confidentiality of the data. The pattern covers the following threats:

- *Loss of governance* with respect to losing the authority to effectively decide about access to the data, about moving the data and deleting the data.
- *Lock-in* is effectively countered by the ability to exclude shares from the data set and to generate new shares to be stored at a different providers.
- Many other technical risks are covered by the implicit encryption, e.g. *isolation failure, management interface compromise, data protection failure, insecure or incomplete data deletion, malicious insider.*
- *Availability* improves as even in the case of one storage provider being off-line, the secret still may be reconstructed with the shares from other providers. On the other hand, if many providers would be off-line simultaneously, the reconstruction may (temporarily) not be possible.

The leakage of metadata, which occurs during storage and retrieval of the single shares, and by synchronisation activity between the single storage providers during share renewal, may still present a privacy problem.

3.3 Field 2: User Privacy Protection and Data Minimisation

Privacy protection requires to minimise the access to information following a need-to-know principle, which means, that the cloud provider shall only have access to what is needed to fulfil the delegated task. This is a known principle with respect to data, but it also applies for the meta-data created through the interaction of the user with the cloud. The most common interaction of a user with a cloud is to prove that he or she is authorised to use a service, but doing so shall not reveal more information than necessary, and shall not allow user tracking by the cloud.

- *Pattern 3: Non-identifiable and untrackable use of a cloud service* has anonymity as its goal, and linkable data is to be completely excluded, while in pattern 4 some information is revealed.
- *Pattern 4: Minimise exposure of private data during authentication in the cloud* assumes that some information is revealed in order to get authorised, but which information exactly is revealed, remains under the control of the user.
- *Pattern 5: Big data anonymisation* is applicable when user privacy is at stake in big data analysis.

The patterns 3 and 4 are closely related to each other—both are concerned with effectively reducing the amount of data which is exposed during interaction with cloud services and applications, and both can be realised with the cryptographic building block of anonymous credentials. We describe pattern 4 as an example next.

Pattern 4: Minimise Exposure of Private Data During Authentication

Summary. Only expose the minimum amount of data necessary when authenticating for a cloud service. During the process of authentication, a user wants to present some attributes, without revealing other attributes he or she may additionally have. The user may also only want to prove the possession of an attribute, or some quality of an attribute (e.g. a statement on a range it is in) without revealing the exact value of the attribute. Moreover, the user may want to show or prove attributes to different sites in a manner, that the single showings cannot be linked to the same user.

Intention. The pattern wants to reduce the data which is unnecessarily exposed during authentication situations.

Problem. Disclosing more data than necessary for performing or delegating a specific task represents a severe privacy threat for the user. Such data is prone to being accumulated and data-mined by the cloud provider and by other parties eventually getting in possession of the data. For example, authentication for a service in the cloud is often performed by the use of an identity certificate. The user shows the certificate to the verifier who verifies the digital signature on the certificate with the public key of a certifier. The verifier thus learns all the data contained in the identity certificate, although for a proper authentication it might be sufficient to access only a small subset of the data . Identity certificates also make interactions attributable to the bearer of the identity certificate, i.e. interactions can be tracked across services. All these side effects are problematic from a privacy point of view and the principle of data minimisation actually calls for avoiding such unnecessary disclosure of data in information infrastructure transactions.

Solution. Authentication allows a claimant in a protocol to convince the verifier that the required set of attributes is correctly held by the claimant. A solution must enable this functionality without revealing any additional attributes and potentially also without being able to link several interactions of the same user.

Building Block. PRISMACLOUD proposes the *flexible authentication with selective disclosure tool* to achieve the desired solution. This tool could implement the technology of "anonymous credentials" following [5].

Consequences and countered threats. The pattern allows an effective reduction of the amount of data which is revealed during authentication and other transactions requiring the presentation of user data. The pattern enables, that statements about the encoded attributes can be proven to a verifier without revealing the values of the attributes. The pattern enables, that different credential shows are unlinkable or can be implemented to be unlinkable. If events need to be linkable, it is possible to anonymously prove the possession of a

pseudonym. The most important technical risks can be excluded because of the cryptographic security of the primitives which are used for its implementation. The current pattern is also effective for countering data protection risks. It allows a fine grained control of which data is exposed to whom. It thus reduces risks connected to the processing of personal data collected by the service provider without effective necessity. It reduces the lack of transparency, and all the risks involved by chain processing involving multiple processors and by moving data between jurisdictions, especially also out of the control of a local data protection regulation.

3.4 Field 3: Authentication of Stored and Processed Data

The patterns in this field are concerned with integrity of data and with a verifiable authentication of origin of data. This is particularly of interest when data is entrusted to cloud systems outside the immediate control of the data owner. But the rigidity of previously used data authentication schemes, e.g. digital signature schemes, did not allow to authorise certain subsequent modifications. Thus, using them did increase the security of a cloud service, but still represented a severe privacy threat as the authenticity proof required to show the entire authenticity protected data to the third party. Moreover, verifiable authenticity is also required for the results of processing and for properties of the involved infrastructures, which cannot be achieved with previously used schemes.

- *Pattern 6: Protect the authenticity of a data set and possible subsets* is applicable whenever data originates at a credible source and its trustworthiness depends on (a) the source staying verifiably authentic and (b) the data being subjected only to authorised subsequent modifications. It is applicable as a substitute for integrity protection by standard signatures.
- *Pattern 7: Authorise controlled subsequent modifications of signed data* is closely related to a pattern known as "delegation". It applies whenever a third party shall be authorised to do subsequent changes, for which the verifier is able to cryptographically verify the authorisation by the original signer. It maintains the confidentiality of processing steps and the original data, as the verifier does not know the changes done nor the original data.
- *Pattern 8: Controlling the correctness of delegated computations* is relevant whenever cloud providers are performing computations on data but cannot be considered fully trustworthy or immune to attacks on data integrity.
- *Pattern 9: Controlling your virtual infrastructures* applies to situations where a customer or end user rents a virtual infrastructure from a cloud service provider. Using recently developed methods for representing the topology of virtualised infrastructure as a graph and issuing a signature on that graph, one can extend current audit procedures with a means for proving the correct configuration of virtualised infrastructures.

In the following we will display Pattern 6 as a selected example.

Pattern 6: Protect the Authenticity of a Data Set and Possible Subsets

Summary. It shall be possible to subsequently cloak and/or remove information from an authentic data set, e.g. a signed data structure, while attaining two additional properties: (1) to protect the confidentiality of the information that was removed, (2) to retain (or have only minimal impact on) the authenticity guarantee of the remaining data.

Intention. The pattern allows for future subsequent removal of data from a data set for which integrity and authenticity protection mechanisms such as digital signatures are usually applied to protect the data set (1) against unauthorised subsequent changes and (2) to authenticate the source of the data.

Problem. Currently well accepted and widely used standard digital signatures do not support any subsequent editing of the data. Whether authorised or not, it will be detected and as a result the integrity and authenticity can no longer be established for the remaining unchanged data. Obvious and naïve solutions to the integrity problem exist, but offer no privacy with sufficient cryptographic strength.[12] Assume, a number of tests is carried out an a blood sample and a report is being created, containing e.g., (1) blood sugar, (2) total cholesterol, (3) haemoglobin, (4) vitamin D, (5) tuberculosis (TB). If only blood sugar, cholesterol and vitamin D (tests 1, 2, and 4) are given to the patient's ecotrophologist, the problem, generally also known as the document sanitization problem [19], is how the remaining data is protected against malicious tampering and the credible source remains verifiable. Moreover, the removal must eliminate all traces such that the ecotrophologist as a potential attacker is prohibited from reconstructing removed data. This must go as far as to even remove any trace that there ever has been done a tuberculosis test (test 5), as it is only conducted for patients in high risk groups or already treated for TB, which reveals private information.

Solution. Employ a different set of cryptographic functionalities, or conventional digital signature schemes in a different way, such that malleability is enabled while authenticity for the remaining data and confidentiality of the removed data is preserved. The allowed modifications must be formally described and the special digital signature for the data set is created. Subsequently the authenticity of the modified data set can be verified, thus giving the cryptographic assurance about the origin of the modified data and that only allowed modifications were made.

Building Block. PRISMACLOUD proposes the *flexible authentication with selective disclosure tool* which enables transparent redactable signature functional-

[12] Whenever the signature mathematically still depends on some removed data, like in hash trees, they cryptographically do not offer a sophisticated level of privacy [3]

ity [22], as e.g. to authorise a subsequent removal while keeping the authenticity of the remaining data protected and to hide the fact that something was removed. The tool can be tailored, e.g. [3], to offer a similar legal assurance [29, 21].

Consequences and countered threats. The pattern combines the strength of cryptographic end-to-end integrity protection with the ability to remove data for data minimisation purposes. The pattern counters at least the following threats:

- *Loss of data integrity*: The remaining data is still integrity protected, any unauthorised change will be detected.
- *Loss of accountability*: The origin of the remaining data can still be authenticated using the public key that is used for digital signature verification. Further accountability depends on the tool and can be tailored.
- *Data leakage*: Unneeded data, if marked as removable, can be removed without reducing the remaining data's verification of origin and integrity.
- *Insecure or incomplete data deletion*: Data requested to be removed is marked as removable in an integrity protected data structure and can be removed with no negative effects on the integrity of the other data. This removes a potential hinderance to delete data at all occurrences.

4 PRISMACLOUD Tools

4.1 Introduction

The PRISMACLOUD project proposes a set of five configurable tools, encapsulating several cryptographic protocols and primitives. Without exception, the cryptographic protocols and primitives are either extensions or adaptations of existing cryptographic protocols or primitives of Technology Readiness Level [12] (TRL) 3 or higher. The novelty and added value of the project is, that the single primitives are advanced to TRL 7 (" system prototype demonstration in operational environment").

The encapsulation of complex cryptographic functionality shall leave the complex and error-prone correct implementation and application to cryptographers and specialised software engineers and prevent likely mistakes by service developers. The tools will be provided as a software library. The single tools can be parametrised in various different ways and thus be customised for use in a specific service. The services provide interfaces in form of (restful) application programming interfaces (APIs) and are suitable to be deployed in the cloud.[18].

Table 1 presents which tools can be applied as solution to which patterns.

4.2 Prismacloud Tools and Employed Cryptographic Primitives

In the following, we provide a summary of the functionalities of the single tools used in the single patterns, as well as the cryptographic protocols and primitives

Field 1: Data storage in the cloud
Pattern 1: Secure cloud storage by default Tool 1: Secure object storage tool
Pattern 2: Moving a legacy application's database to the cloud Tool 5: Data privacy tool
Field 2: User privacy protection and data minimisation
Pattern 3: Non-identifiable and untrackable use of a cloud service Tool 2: Flexible authentication with selective disclosure tool
Pattern 4: Minimise exposure of private data during authentication Tool 2: Flexible authentication with selective disclosure tool
Pattern 5: Big data anonymisation Tool 5: Data privacy tool
Field 3: Authentication of stored and processed data
Pattern 6: Protect the authenticity of a data set and possible subsets Tool 2: Flexible authentication with selective disclosure tool
Pattern 7: Authorise controlled modifications of signed data Tool 2: Flexible authentication with selective disclosure tool
Pattern 8: Controlling the correctness of delegated computations Tool 3: Verifiable data processing tool
Pattern 9: Controlling your virtual infrastructures Tool 4: Topology certification tool

Table 1. Cloud security patterns and related cryptographic building blocks

they are based on. A detailed descriptions of the tools and primitives, including references can again be found in [18].

Tool 1: Secure object storage tool. PRISMACLOUD proposes to split the data to be stored into a number of shares which are distributed to several cloud storage providers in a way, that no single provider can access the plain data, which can only be reconstructed from a fixed number of shares. Under the assumption that a certain number of providers do not maliciously cooperate, the secret sharing algorithm itself is considerably stronger than commonly used cryptographic systems and is capable of long-term security [20]. Therefore, it can be applied also in scenarios with highest confidentiality requirements, like in e-Health or e-Government. It requires an explicit access control system to the split shares, but then provides a kind of key-less encryption with provable security. The tool allows checking the integrity of remotely stored shares without having to retrieve the shares first. It also solves the availability problem at the user level,

without the need of explicit backups. Single shares can also be taken out of the system and be replaced by newly generated ones. This prevents vendor lock-in and, when shares are continuously renewed, enables long-term data security as it minimises the chance of an attacker to get a sufficient number of shares for reconstructing the information by attacking one cloud provider after the other. The used cryptographic protocols and primitives are:

- *Secret sharing schemes*: A secret sharing protocol is used to split the information into several parts, of which any subset of a given number of shares is necessary to access the information.
- *Remote data checking*: Allows for efficient checking of the availability and correctness of remote shares
- *Private information retrieval*: Allows clients to retrieve data items from a storage provider without revealing to the provider which items were retrieved

Tool 2: Flexible authentication with selective disclosure tool. PRISMACLOUD supports the authentication of arbitrary messages (or documents). This tools encapsulates cryptographic primitives to offer three abstract functionalities: *authentication*, *selective disclosure*, and *verification*. The data originator authenticates by signing a message, together with a disclosure policy describing which parts of the message can be selectively disclosed. Selective disclosure allows to disclose parts of the information from such a signed message to other receiving parties. The verification functionality checks if only authorised modifications, i.e. modifications conforming to the disclosure policy, were done. The selective disclosure is achieved by the concept of malleable signature schemes—although the direct application of a selective disclosure primitive would also be possible. The desired granularity of verification can be controlled by the signature primitive used. The cryptographic protocols and primitives are:

- *Malleable signatures schemes*: Allows to authorise subsequent modifications of certain parts of the signed data without the signature losing its validity; integrity against unauthorised modifications and authentication of origin are as protected as by classical digital signatures.
- *Attribute-based credentials*: Provides anonymous authentication; a multi-show credential system allows an arbitrary number of unlinkable showings.
- *Functional signatures schemes*: Allow to certify computations and processes; allow to delegate signature generation to other parties for a class of messages meeting certain conditions.
- *Zero-knowledge proofs*: Allow one party to convince another party of the validity of a statement without revealing any more information than the validity of the statement.
- *Group signature schemes*: Allow the signer to stay anonymously towards the verifier as the verifier only sees a signature that is valid for a group of signers.

Tool 3: Verifiable data processing tool. This tool allows the verification of results of computations on signed data, delegated to a computing cloud. When a client gets back the result of the computation, he or she can efficiently decide

whether the requested function was correctly applied to the data. The used cryptographic protocols and primitives are:

- *Secret sharing schemes*: see tool 1.
- *Malleable signatures schemes*: see tool 2.
- *Functional signature schemes*: see tool 2.
- *Zero knowledge proofs*: see tool 2.

Tool 4: Topology certification tool. Current cloud audit procedures can be extended with a means for proving security properties of virtualised infrastructures. An auditor (a human or a software agent) verifies an actual infrastructure, represents it as a graph, and issues a digital certificate on the graph. A prover component issues a zero-knowledge proof on the certificate, capable of convincing a cloud customer of the requested security properties, without revealing to the customer actual details of the topology. The tool encompasses the following:

- *Graph signature schemes*: Allows digitally signing a set of vertices and edges.
- *Zero-knowledge proofs*: see tool 2.

Tool 5: Data privacy tool. This tool provides the functionalities of the following two cryptographic primitives:

- *Format- and order-preserving encryption*: Adds a layer of cryptography directly into the data fields of a database applications: Format preserving encryption applies encryption in a manner such that the ciphertext has the same format as the plaintext (e.g. a social security number is mapped to a cryptogram with the format of a social security number).
- *k-anonymity*: K-anonymisation of data anonymises data in a way, that for each entry, there are at least $(k-1)$ other entries, from which it cannot be distinguished. While k-anonymity is a NP hard problem, new, more efficient approaches to anonymising big sets of data have improved in efficiency and are now capable of anonymising very large data sets.

5 Conclusions

In the current article we pointed out how cloud security patterns can be used to support the privacy-by-design process of a large scale development effort for reusable software tools, enabling the construction of privacy and security aware cloud services. In this context, the patterns act as medium between two groups: towards developers of cryptographic protocols and primitives, and to software engineers they communicate the problems which need to be cryptographically solved—and towards cloud service developers they convey which functionalities of existing software libraries ("the tools") can be re-used for the creation of cloud services. In addition to a commonly employed requirements approach, the cloud security patterns are used in the on-going H2020 PRISMACLOUD project to communicate the security requirements of involved stakeholders in a descriptive and informal way, thus enabling an on-going discussion, resulting in a generative approach towards resolving design contentions.

References

1. Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press (1977)
2. Backes, M., Datta, A., Kate, A.: Asynchronous computational vss with reduced communication complexity. In: Dawson, E. (ed.) Topics in Cryptology – CT-RSA 2013: The Cryptographers' Track at the RSA Conference 2013, San Francisco,CA, USA, February 25-March 1, 2013. Proceedings. pp. 259–276. Springer Berlin Heidelberg, Berlin, Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-36095-4_17
3. Brzuska, C., Pöhls, H.C., Samelin, K.: Non-interactive public accountability for sanitizable signatures. In: De Capitani di Vimercati, S., Mitchell, C. (eds.) Public Key Infrastructures, Services and Applications: 9th European Workshop, EuroPKI 2012, Pisa, Italy, September 13-14, 2012, Revised Selected Papers. pp. 178–193. Springer Berlin Heidelberg, Berlin, Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-40012-4_12
4. Buchmann, J., Demirel, D., Happe, A., Krenn, S., Lorünser, T., Traverso, G.: PRISMACLOUD D4.1: Secret Sharing Protocols for Various Adversary Models (2015), www.prismacloud.eu, H2020 project PRISMACLOUD deliverable
5. Camenisch, J., Herreweghen, E.V.: Design and implementation of the *idemix* anonymous credential system. In: ACM CCS. pp. 21–30. ACM (2002), http://doi.acm.org/10.1145/586110.586114
6. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.H., Le Mtayer, D., Tirtea, R., Schiffner, S.: Privacy and Data Protection by Design. Tech. rep., European Union Agency for Network and Information Security (ENISA) (2015)
7. Doty, N., Gupta, M.: Privacy Design Patterns and Anti-Patterns. In: Workshop "A Turn for the Worse: Trustbusters for User Interfaces Workshop" at SOUPS 2013 Newcastle, UK (2013)
8. ENISA European Union Agency for Network and Information Security: Cloud computing repository, http://www.enisa.europa.eu/activities/Resilience-and-CIIP/cloud-computing, (online 31.3.2015)
9. ENISA European Union Agency for Network and Information Security: Cloud computing; Benefits, risks and recommendations for information security; Rev. B. December 2012 (2012), https://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport, (online 1.3.2016)
10. European Commission: Establishing Horizon 2020 - The Framework Programme for Research and Innovation (2012), http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011PC0809:EN:NOT, (online 1.6.2016)
11. European Commission: European Cloud Computing Strategy "Unleashing the Potential of Cloud Computing in Europe" (2012), http://ec.europa.eu/digital-agenda/en/european-cloud-computing-strategy, (online 31.3.2015)
12. European Commission: Technology readiness levels (TRL) (2014), http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf, (online 1.6.2016)
13. Forbes magazine: Roundup Of Cloud Computing Forecasts And Market Estimates Q3 Update, 2015 (2015), http://www.forbes.com/sites/louiscolumbus/2015/09/27/roundup-of-cloud-computing-forecasts-and-market-estimates-q3-update-2015/#35e2a3576c7a, (online 1.3.2016)

14. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley. ISBN 0-201-63361-2 (1994)
15. Hafiz, M.: A collection of privacy design patterns. In: Proceedings of the 2006 Conference on Pattern Languages of Programs. pp. 7:1–7:13. PLoP '06, ACM, New York, NY, USA (2006), http://doi.acm.org/10.1145/1415472.1415481
16. Lorünser, T., Länger, T., Slamanig, D.: Cloud Security and Privacy by Design. In: Katsikas, K.S., Sideridis, B.A. (eds.) E-Democracy – Citizen Rights in the World of the New Computing Paradigms: 6th International Conference, E-Democracy 2015, Athens, Greece, December 10-11, 2015, Proceedings. pp. 202–206. Springer International Publishing, Cham (2015), http://dx.doi.org/10.1007/978-3-319-27164-4_16
17. Lorünser, T., Rodriguez, C.B., Demirel, D., Fischer-Hübner, S., Groß, T., Länger, T., des Noes, M., Pöhls, H.C., Rozenberg, B., Slamanig, D.: Towards a New Paradigm for Privacy and Security in Cloud Services. In: CSP Forum 2015. LNCS, vol. 8874, pp. 1–12. Springer (2015)
18. Lorünser, T., Slamanig, D., Länger, T., Pöhls, H.C.: PRISMACLOUD Tools: A Cryptographic Toolbox for Increasing Security in Cloud Services. In: Proceedings of the International Conference on Availability, Reliability and Security (ARES 2016). IEEE (2016), (to be published Sept. 2016)
19. Miyazaki, K., Hanaoka, G., Imai, H.: Digitally signed document sanitizing scheme based on bilinear maps. In: Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security. pp. 343–354. ASIACCS '06, ACM, New York, NY, USA (2006), http://doi.acm.org/10.1145/1128817.1128868
20. Müller-Quade, J., Unruh, D.: Long-term security and universal composability. In: Vadhan, S.P. (ed.) Theory of Cryptography: 4th Theory of Cryptography Conference, TCC 2007, Amsterdam, The Netherlands, February 21-24, 2007. Proceedings. pp. 41–60. Springer Berlin Heidelberg, Berlin, Heidelberg (2007), http://dx.doi.org/10.1007/978-3-540-70936-7_3
21. Pöhls, H.C., Höhne, F.: The Role of Data Integrity in EU Digital Signature Legislation - Achieving Statutory Trust for Sanitizable Signature Schemes. In: International Workshop on Security and Trust Management (STM). Springer LNCS (2011)
22. Pöhls, H.C., Samelin, K.: On Updatable Redactable Signatures0. In: Proceedings of the 12th International Conference on Applied Cryptography and Network Security (ACNS 2014). Lecture Notes in Computer Science (LNCS), Springer (2014)
23. PRWeb: A Cloud Computing Forecast Summary for 2013-2017 from IDC, Gartner and KPMG, citing a study by Accenture (2013), http://www.prweb.com/releases/2013/11/prweb11341594.htm, (online 31.3.2015)
24. RightScale Inc.: State of the Cloud Report (2015), http://assets.rightscale.com/uploads/pdfs/RightScale-2015-State-of-the-Cloud-Report.pdf, (online 31.3.2015)
25. Schumacher, M., Fernandez-Buglioni, E., Hybertson, D., Buschmann, F., Sommerlad, P.: Security Patterns - Integrating Security and Systems Engineering. John Wiley & Sons, Ltd. West Sussex, England (2006)
26. Shamir, A.: How to share a secret. Commun. ACM 22(11), 612–613 (Nov 1979), http://doi.acm.org/10.1145/359168.359176
27. The Economist Intelligence Unit: Mapping the cloud maturity curve (May 2015), http://www.economistinsights.com/analysis/mapping-cloud-maturity-curve, (online 31.3.2015)

28. Transparency Market Research: Cloud Computing Services Market – Global Industry Size, Share, Trends, Analysis And Forecasts 2012-2018 (2012), http://www.transparencymarketresearch.com/cloud-computing-services-market.html, (online 31.3.2015)
29. Van Geelkerken, F., Pöhls, H.C., Fischer-Hübner, S.: The legal status of malleable- and functional signatures in light of Regulation (EU) No 910/2014. In: Proceedings of 3rd International Academic Conference of Young Scientists on Law & Psychology 2015 (LPS 2015). pp. 404–410. L'viv Polytechnic Publishing House (Nov 2015), https://drive.google.com/file/d/0B-Yu3Ni9z3PXM2lBajhCXzhoWk0/view