First publ. in: 5. DFN-Forum Kommunikationstechnologien : Verteilte Systeme im Wissenschaftsbereich ; 21.05.-22.05.2012 in Regensburg ; [Beiträge der Fachtagung] / Paul Müller... (Hrsg.). - Bonn : Ges. für Informatik, 2012. - S. 63-72. - ISBN 978-3-88579-297-0

A Legal and Technical Perspective on Secure Cloud Storage

Sebastian Graf, Jörg Eisele and Marcel Waldvogel, University of Konstanz*
Marc Strittmatter, University of Applied Sciences, Konstanz+

*(firstname.lastname)@uni-konstanz.de
+(firstname.lastname)@htwg-konstanz.de

Abstract: Public cloud infrastructures represent alluring storage platforms supporting easy and flexible, location-independent access to the hosted information without any hassle for maintaining own infrastructures.

Already widely established and utilized by end-users as well as by institutions, the hosting of data on untrusted platforms, containing private and confidential information, generates concerns about the security. Technical measures establishing security rely thereby on the technical applicability. As a consequence, legal regulations must be applied to cover those measures even beyond this technical applicability.

This paper provides an evaluation of technical measures combined with legal aspects representing a guideline for secure cloud storage for end-users as well as for institutions. Based upon current approaches providing secure data storage on a technical level, german laws are applied and discussed to give an overview about correct treatment of even confidential data stored securely in the cloud.

As a result, a set of technical possibilities applied on fixed defined security requirements is presented and discussed. These technical measures are extended by legal aspects which must be provided from the site of the hosting Cloud Service Provider.

The presented combination of the technical and the legal perspective on secure cloud storage enables end-users as well as hosting institutions to store their data securely in the cloud in an accountable and transparent way.

1 Introduction

Internet Services such as Flickr, Dropbox, Wuala as well as Amazon S3 and Google Cloud Storage provide comfortable and ubiquitous storage and sharing for a wide class of data. These services relieve the user from hardware purchases, software bug fixes, and infrastructure maintenance, at the cost of the users implicitly granting the Cloud Service Provider and their administrators full access to all their sensitive data, including secret business data when used by a company or institution.

The world-wide accessibility of these public services not only enables external attackers to gain access to the data: It must be assumed that within these public clouds, the hosting companies like Amazon, Google and Yahoo use the data, representing an alluring mass of confidential information, for user-analysis and advertising. The different attack-models, ongoing with the geographical distribution of the data over different countries, make the identification of necessary security measures ongoing with corresponding legal aspects hard to accomplish.

This paper maps common security requirements to the peculiarities of cloud-based storage. Since security is only guaranteed by the satisfaction of all security requirements, a combination of different measures is discussed and extended by corresponding legal aspects.

The proposed set of techniques, guarding the data on a technical base extended by suitable regulations, represents a guideline for secure storage on public cloud infrastructures for end-users as well as for institutions.

2 Applying security measures to cloud-based storage

Public cloud infrastructures offer different Cloud Service levels of utilization defined as "Software as a Service" (SaaS), "Platform as a Service" (PaaS) and "Infrastructure as a Service " (Iaas) [MG09].

Applications are commonly deployed on one of the defined Cloud Service levels. Figure 1 maps these levels on the ability of technical control: Each service deployed in the cloud relies on an execution stack consisting of Services, Applications, Platforms, Operating Systems and Hardware. The ability to influence the application for a customer e.g. for establishing security bases on the cloud level utilized for deployment. As a consequence, the responsibility for ensuring trust is shared between the customer and the Cloud Service Provider. Within SaaS-infrastructures, the customer has only the ability to control the service itself without having influence on the lower levels, while in PaaS-infrastructures, the customer can control the service and the representing application. For the lower levels, only the Cloud Service Provider has technical control and is therefore responsible for providing security. This transition of responsibility based on the denoted application-stack is called *Threshold of Technical Control* within Fig. 1 and the rest of the paper. Consequently, technical security measures can only be applied by the customer on her site of the *Threshold of Technical Control* depending on the deployed level, while security on the site of the Cloud Service Provider must be covered by legal regulations to provide throughout security in the cloud.

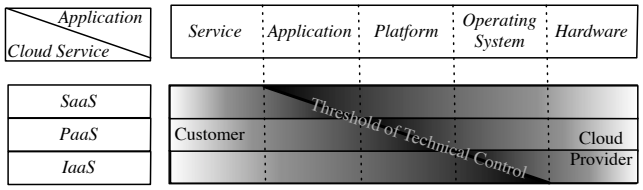

Figure 1: Point of interference

Cloud storage systems commonly fit the SaaS- and PaaS-levels: Security measures applied on native clients (like e.g. provided by Wuala [GMSW06] and Dropbox) as well as on common web services (like e.g. REST [Fie00] accessing Google Cloud Storage or the Amazon S3 system) therefore ensure security on the Service as well as on the Application level depending on the concrete system. For all lower levels down to the Hardware level, the Cloud Service Provider is responsible for guaranteeing secure data storage. This

responsibility increases since cloud services are often stacked resulting in cloud-service supply chains[1].

Based upon the *Threshold of Technical Control*, the concrete kinds of established security measures depends on the level of control namely the kind of service which is utilized for storage: Enabling storage on untrusted public infrastructures thereby fit two main kinds: User-centric cloud storage and application-centric cloud storage.

2.1 User-centric cloud storage

Figure 2 shows a schema of an user-centric cloud storage.

Clients, which are under user-control and therefore trusted denoted by the lockers, use the cloud to store the data directly. The cloud itself consists of abstraction layers mirroring the data world-wide. Since the storage is represented by a direct accessible service, it is utilized as SaaS. Consequently, technical measures to provide security must be applied on the client-site before the data is sent into the cloud. Each client accesses the storage directly, applying optional rules for sharing. Nevertheless, these access rights are only recognized by other clients while internal access is not technically restricted by default. Practical examples of this scenario include Dropbox and Wuala where native applications care about the synchronization between the clients and the cloud.

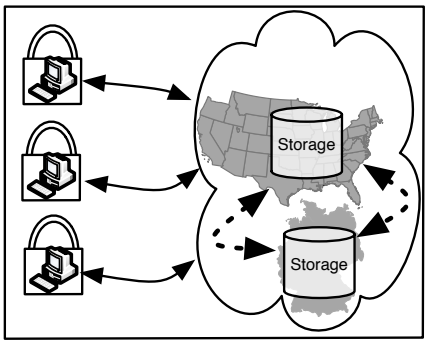

Figure 2: User-centric Cloud Storage

2.2 Application-centric cloud storage

Figure 3 shows a different schema of an application-centric cloud storage.

Institutions for example rely on trusted centralized components and therefore access the cloud in a centralized way. Again, the cloud represents an abstraction of services and storages opaque for the users: The concrete location of the data is unaware even though many Cloud Service Providers offer regional storage options in this scenario. The storage is often utilized as PaaS where either own defined applications care about the data handling in the cloud or the storage is accessed with the help of web services. This access is per-

[1] Dropbox for example utilizes Amazon S3 as storage backend.

formed by trusted centralized applications denoted as “Internal Service” in Fig. 3. Access rights as well as technical security measures are administrated over this service whereas the cloud has no deeper semantic knowledge about the data. Examples of this scenario are the Google App Engine, Microsoft Azure, Amazon S3 as well as the Google Cloud Storage.

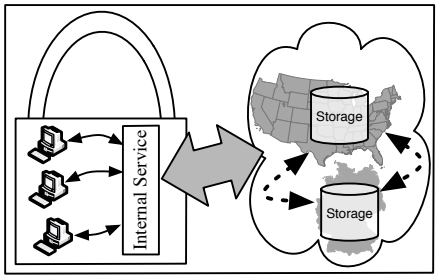

Figure 3: Application-centric Cloud Storage

In both scenarios, techniques to provide security must be established either on the clients or within the centralized “Internal Service”. Legal regulations must extend those techniques from the site of the Cloud Service Provider while the *Threshold of Technical Control* is represented by the transfer of the data into the cloud.

3 Defining a secure cloud storage

Before discussing technical measures as well as legal regulations, common security requirements [Sto01, Sch00, Lam01] must be mapped to the use case of secure cloud storage. Based on this mapping, technical approaches as well as concerned regulations are discussed.

- **Confidentiality in the cloud:**
  Confidentiality definitely represents a major security concern within the cloud. The question “Who can read my data?” is not only related to companies or institutions hosting third-party data. Internal malicious accessors as well as external attackers leverage from the world-wide accessibility and hosting of the data. Besides these attack scenarios, the inlying information is worth a mint for the Cloud Service Providers answering questions about the generosity of their often free offers.

- **Availability in the cloud:**
  “How can I ensure everlasting access to my data?” is one driving question behind putting data in the cloud extended by the wish for easy sharing. Since the physical control of the data is obscured by the *Threshold of Technical Control*, users have only the possibility to trust the Cloud Service Providers regarding their “number of nine’s”[2] and their promises not to harm any data.

- **Integrity in the cloud:**
  Consistence is a major issue often only applied on the transfer of the data. The

[2]The “number of nines” represent the exact percentage of availability of the services (e.g. 99.99 % related to Amazon in 2007 [Gar07]).

question “Is my data still intact in the cloud?” demands the integrity. Guarding integrity represents a major security challenge since proofs of integrity as well as restoration of data are less reliable when performed directly in the cloud.

- **Accountability in the cloud:**
  The traceability of actions on the data is covered by the question “What actions occurred on my data?”. Accountability defines the ability to trace any kind of access as fine-granular as possible. Tracing read access is thereby hardly realizable in the cloud due to the physical abstraction of the storage and the various possibilities of access. Accountability applied to secure cloud storage thereby focusses on modifications on the data including procedural approaches.

- **Assurance on the cloud:**
  Assurance is the overall trust even beyond the applied security measures. Applied to cloud storage, it is formulated as the question “How secure is my data in the cloud?”. The answer to this question is a combination of all applied security techniques combined with regulations and policies. Assurance thereby includes legal aspects as well, since a fixed definition of reliances is mandatory to provide security even beyond the *Threshold of Technical Control*.

To store data “securely” in the cloud, a combination of measures to satisfy all denoted security requirements becomes necessary. The *Threshold of Technical Control* defines the possible field of the appliance of technical measures whereas the aspects must not only adhere the characteristics of cloud-based storage, namely its high availability, the mistrustfulness of the hosting infrastructure, the distant location from the data as well as the loss of physical possession of the data. Furthermore, technical approaches should neither hamper collaboration and sharing nor complicate synchronization of data between different locations. All security measures must, based on the *Threshold of Technical Control*, be applicable on the trusted components only, even though the denoted benefits of the cloud should be utilized.

3.1 Technical approaches

All technical measures, applied on the trusted site of cloud-storage architectures, must adhere the defined security requirements.

3.1.1 Confidential data handling

Confidentiality represents the most obvious security requirement to be satisfied in the cloud. Straight-forward encryption ensures confidentiality of the data, yet synchronization mechanisms necessary for pushing data efficiently in the cloud should respect the modifications in an encrypted way. Based on diff-algorithms, transferring only deltas between two versions enables performant synchronization. Ignoring the deltas within establishing confidentiality increases the synchronization effort in an unscalable way. Besides the

awareness of the deltas, a suitable key management must be established to provide secure data sharing. Since sharing and collaboration represent main use cases for cloud storage, this functionality should no be hampered by establishing confidentiality. Confidentiality-awareness in the cloud is as a consequence less a question of encryption but more a field of encryption-aware synchronization and key management enabling efficient access on encrypted data for disjunct clients.

3.1.2 Keeping the data available

Besides the necessity for confidentiality without reducing cloud-based functionalities, the availability of cloud-based data must be guaranteed as well. The high availability of cloud services leads to the perception that data stored in the cloud will remain accessible forever. As a consequence, users often do not backup their data when pushed into the cloud. Nevertheless, errors in cloud infrastructures occur[3]. Besides disturbances within the cloud infrastructure itself, the access to the cloud represents not only a bottleneck related to data transfer rates but also a vulnerability related to the access.

Current approaches provide redundancy by storing data on multiple clouds, namely in a cloud-of-clouds like DepSky [BCQ+11]. Besides the utilization of multiple clouds, local caching of the data buffers possible network disturbances.

3.1.3 Consistency checks of the data

Erasure codes like provided within DepSky guard furthermore the integrity of the data. Due to the physical loss of control, such checks become necessary to be aware about the status of the data. Based upon at least local partial caching, out-of-the-box integrity checks like provided by Amazon can be utilized even though the main purpose of those checks is the awareness of transmission errors.

Specialized approaches like HAIL [BJO09] try to fill this gap based upon replication and checking of blocks utilizing signatures and checksums. Since the data at rest in the cloud stays behind the *Threshold of Technical Control*, integrity-checks in the cloud rely on the trust against the hosting provider resulting in the necessity for suitable regulations established between the customer and the Cloud Service Provider.

3.1.4 Tracing actions on the data

Accountability describes the ability to trace actions on single entities within the data. A straight forward approach represents logging and auditing as well as establishing policies and regulations for the access. Since we rely on data storage only, sophisticated versioning of the data represents a straight-forward mechanism to ensure accountability. Such a versioning must be robust against the damage of single versions to offer easy reconstruction of any version. Due to the distance to the storage, the deltas between consecutive

[3]The Amazon EC2 cloud crashed at the beginning of 2011 generating some data lost without any possibility of reconstruction.

modifications must be balanced and encryption-aware to ensure efficient transfer of the data.

3.1.5 Combining measure to assure data security

Violating one security requirement results in a vulnerable cloud storage. As a consequence, confidentiality, availability, integrity as well as accountability must be applied synchronously to gain assurance.

Fig. 4 recapitulates the proposed measures. Some of these techniques thereby satisfy more than one security requirement. Erasure codes ensuring availability within DepSky for example guard the integrity utilizing the computation they are based on. Checksums and signatures, guarding mainly the consistence of the data, play an important role for providing accountability as well: Combined with versioning of the data, higher level security goals like non-repudiation can be established.

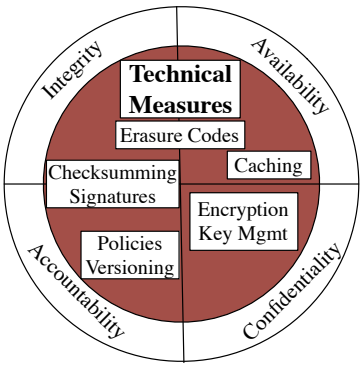

Figure 4: Technical measures

Since the measures can only be applied on the customer site of the *Threshold of Technical Control*, legal implications ensuring secure cloud storage on the Cloud Service Provider site become necessary to provide throughout assurance.

3.2 Legal implications

Legal concepts applied on security do not depart from the classification of security requirements like technical measures because of the various legal disciplines which are touched e.g. data protection law, torts law, contracts and criminal law. Applying acts to cloud storage is in practice thereby not as easy as it seems: First, the geographical distribution of services make the application of mandatory law, relying on the border of countries, often hard to determine. Data stored in a cloud while uploaded from e.g. Germany might be illegal in other countries. Within this paper, we mainly focus on the application of German statutory laws based on Sec. 9 I StGB which relies on the place of action (namely the initial push of the data into the cloud). Sec. 7 I StGB furthermore defines the application of repressive measures if the action is unlawful in other countries as well.

Technical undefined terms in statues like “notable” and “necessary” make an interpretation of law even more complicated. We will therefore evaluate the implications of security to cloud storage based upon different scenarios:

3.2.1 Unauthorized access

Even though from a technical point of view there is an immense difference how to establish confidentiality by encrypting or by just blocking the access, from a legal point of view Sec. 202a StGB and Sec. 202b StGB prevent any unauthorized access. Only accessing restricted content in an unauthorized manner is sufficient to harm those statutes regardless if the accessor attacks from outside or is represented by an internal person. It is important to know that even the preparation of unauthorized data access is indictable by Sec. 202 c StGB.

3.2.2 Harming data

Unauthorized modifications or deletions of data are covered by Sec. 303a StGB. Any unexpected status of the data is not only harming the integrity but also the availability. If a copy of the unauthorized removed/modified data exists, this act might not be impinged. The preparation to make data inaccessible in an unauthorized way is covered by Sec. 202 c StGB as well. Similar to possible unauthorized access to the data, it is unimportant if the attack harming the data occurs from outside or inside the cloud.

3.2.3 Data privacy

Data privacy is a huge field within cloud infrastructure utilization. The storage of information in untrusted infrastructures not only harms confidentiality, it touches, from a legal point of view, all security requirements. German law about data privacy is rather strict when personal information is stored. From an EU perspective, any stored personal information must be handled in a way that the user keeps control over the data, directly or indirectly by installing a contractual data controller-data processor relationship while restricting data processing to countries with acceptable levels of data security. Harming the related German statute Sec. 43 II BDSG thereby can be based upon unauthorized modifications (mapping the confidentiality and integrity) as well as the accountability since the user has the ability to order a reconstruction of all actions taking place on the data. Data privacy is handled differently within different countries which complicates related user-requests. European harmonization has installed a minimum level of protection. Current 2011 ECJ (European Court of Justice) decisions have triggered legal discussions of the need for a maximum protection level by EU law overruling more protective country laws (such as German BDSG).

3.2.4 Author's rights

The ease of collaboration brings concerns about authors right into the focus of security. Unauthorized access thereby not only covers the field of confidentiality, it furthermore harms the accountability. Unauthorized copies of data are harming authors rights especially when the attackers intent is to make unlawful profit. Related statues harmed in such scenarios are Sec. 106 and 108 UrG.

3.2.5 Contracts

It should be noted that the contractual definition of "confidentiality" and "security" is typically subject to the parties appraisal. Depending on the applicable law (typically freely eligible by merchant parties to contracts, with some restrictions also by parties of contracts where one party is an end-consumer) the definition of what the parties accept as "secure" or define as "confidential" has a large gamut of variances. At times, the Cloud Service Provider even tie the minimum level of security to the one of its contractual partner[4]. Jurisdictions which rely on statutory, codified law (esp. Continental) do have less leniency in the interpretation of legal concepts than the common law ones (Anglo-Saxon) due to the limitations codified law. In essence, there is a considerable need to trigger a discussion around standardized legal concepts which are intended to be used for multi-jurisdictional relationships.

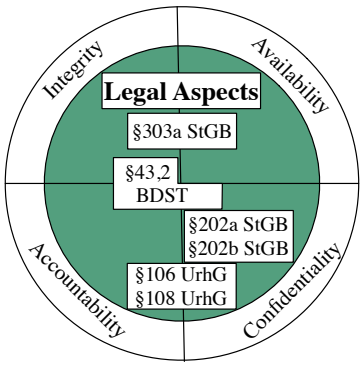

Figure 5: Legal measures

Fig. 5 summarizes the denoted regulations mapping the different security requirements. Similar to the technical measures, many regulations match multiple security requirements: Privacy law for example guards all security requirements since the data must handled in a way like a physical possession is present. Criminal law focus mainly on the availability and the integrity of the data as well as on the access. Further regulations are possible depending on concrete cases they could be applied on. Contract-based policies are excluded in Fig. 5 since they represent such a special case applicable only between the participating parties.

4 Conclusion and outlook

Secure cloud storage can neither be guaranteed by satisfying single security requirements like confidentiality or integrity only, nor by taking technical measures without suitable legal interpretations into account. Technical measures satisfying even multiple security requirements, must be established within trusted components up to the *Threshold of Technical Control*. Beyond this threshold, legal regulations must be established to guarantee throughout security. The corresponding legal applications cover thereby multiple disjunct areas of law science and heavily depend on the locality of the applied law. Nevertheless, it is mandatory that institutions and end-users are aware of the security requirements and the resulting mapping of at least the local regulations since they guard the stored data even

[4] e.g. by regulations like "You will take all reasonable measures to avoid disclosure, dissemination or unauthorized use of XY Confidential Information, including, at a minimum, those measures you take to protect your own confidential information of a similar nature."

beyond technical possibilities.

Our next steps on a technical site include the ongoing development of a client called Treetank [Gra11, GKW11]. Treetank combines the proposed technical measures to satisfy security upon the *Threshold of Technical Control* communicating with SaaS applications as well as with specialized servers deployed on PaaS instances. Additionally to the technical development of this infrastructure, we will continue our evaluation on corresponding legal aspects and apply our findings so far to international statues as well, satisfying the global-aware nature of cloud infrastructures.

References

[BCQ+11] Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando André, and Paulo Sousa. DepSky: dependable and secure storage in a cloud-of-clouds. In *Proceedings of the sixth conference on Computer systems*, EuroSys '11, 2011.

[BJO09] Kevin D. Bowers, Ari Juels, and Alina Oprea. HAIL: a high-availability and integrity layer for cloud storage. In *Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, 2009.

[Fie00] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, Irvine, 2000. Chair-Taylor, Richard N.

[Gar07] Simson L. Garfinkel. An Evaluation of Amazons Grid Computing Services: EC2, S3, and SQS. Technical report, Center for, 2007.

[GKW11] Sebastian Graf, Marc Kramis, and Marcel Waldvogel. Treetank: Designing a Versioned XML Storage. In *XMLPrague'11*, 2011.

[GMSW06] Dominik Grolimund, Luzius Meisser, Stefan Schmid, and Roger Wattenhofer. Cryptree: A Folder Tree Structure for Cryptographic File Systems. In *25th IEEE Symposium on Reliable Distributed Systems (SRDS), Leeds, United Kingdom*, October 2006.

[Gra11] Sebastian Graf. A secure cloud gateway based upon XML and web services. In *PhD Symposium, ECOWS'11*, 2011.

[Lam01] Pradip Lamsal. Understanding Trust and Security, 2001.

[MG09] Peter Mell and Tim Grance. The NIST definition of cloud computing. *National Institute of Standards and Technology*, 53(6), 2009.

[Sch00] Bruce Schneier. *Secrets and lies: digital security in a networked world*. John Wiley, 2000.

[Sto01] Gary Stoneburner. Underlaying Technical Models for Information Technology Security. *National Institute of Standards and Technology*, 2001.