Department of Computer Sciences
University of Salzburg

Seminararbeit

SE Seminar aus Informatik
SS 2013

When the Clouds Disperse

Data Confidentiality and Privacy in Cloud Computing

14. Juli 2013

Author: Christof Kauba[1]

Stefan Mayer[2]

Supervisor: Univ.-Prof. Dipl.-Ing. Dr.techn. Wolfgang Pree[3]

Universität Salzburg
Fachbereich Computerwissenschaften
Jakob–Haringer–Straße 2
A–5020 Salzburg
Austria

[1] ckauba@cosy.sbg.ac.at - Matr.Nr.: 0825754
[2] smayer@cosy.sbg.ac.at - Matr.Nr.: 0930426
[3] office@cs.uni-salzburg.at

Abstract

Cloud computing has emerged as a main form of modern computing due to its many advantages. But there are also some disadvantages if users decide to store sensitive data in the cloud. This seminar paper deals with cloud computing security issues and threats, especially with data confidentiality and privacy issues. The interests of different groups on users private data, the legal aspects which enable governments and service providers itself to access their users' data and also the technical background how they can do so is covered. Several data confidentiality and privacy threats, attacks and case studies are presented. In addition also the attitudes and beliefs of cloud service users and their misconceptions about the rights they have are discussed. Finally some more or less feasible countermeasures are presented.

Contents

1 Introduction

Today almost all PC users have access to the internet. More and more users are using at least some cloud services, like e-mail, Facebook, Google Docs and so forth. But not only private users are switching to cloud services, also companies and governments are adopting them. Cloud computing offers many benefits for its users, e.g. cost savings, increased flexibility and ubiquitous access to the data just to mention a few. Cloud computing will become even more dominant in the future.

But as with every technological improvement there is also a dark side. Most users do not think of where their data is stored and how it is kept confidential. Hackers, cloud service providers and also governments have a growing interest in the data stored in the cloud and wish to covertly access users' data for different reasons. So cloud computing implies a higher risk of personal data disclosure on its users as if they would store their data locally. Governments may have a legitimate need to access private data, but with laws allowing them to do so without a court order and even without noticing the user the risk of abuse is rather high.

Especially recent debates about voluntary data disclosures of cloud service providers, NSA data accesses without a warrant[4], providers giving government agencies raw access to their whole network, fines for providers not complying with the agencies [5] and data and the US PRISM program[6] [7] show that cloud users should really be aware of that risk and think twice which data they store in the cloud.

This paper discusses the privacy and data confidentiality risks that users have to be aware of using cloud services, the legal aspects regarding privacy, users' attitudes and beliefs regarding cloud computing security and also some ways how existing technology can be used to tackle these issues.

It is organized as follows: In chapter 2 cloud computing and the different types of cloud services are defined. The benefits and possible issues of cloud computing are also presented. Then in chapter 3 the legal aspects regarding privacy and data confidentiality cloud computing are covered, including US laws and also the service providers' terms of service agreements. Afterwards in chapter 4 some possible threats and attacks mainly focusing on web applications are discussed. Chapter 5 deals with the opinions and attitudes of cloud users regarding their rights, privacy and the policies of the cloud service providers. Chapter 6 then presents some countermeasures to deal with the attacks presented in chapter 4. Finally chapter 7 concludes the paper.

2 Cloud Computing

The main concept of cloud computing is not new, but it took quite some time until its successful realization. Many authors state that cloud computing can be regarded as next step in the evolution of distributed computing. Cloud computing enables its users to access almost unlimited computing resources in a comfortable and scalable way.

2.1 Cloud Computing Definition

There is no single definition for cloud computing but according to recent literature cloud computing refers to:

> both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services.[2]

The US NIST (National Institute of Standards and Technology) defined the five main characteristics of cloud computing as following[8]:

1. On-demand self-service
2. Broad network access

[4] The top secret rules that allow NSA to use US data without a warrant `http://www.guardian.co.uk/world/2013/jun/20/fisa-court-nsa-without-warrant`

[5] Panel seeks to fine tech companies for noncompliance with wiretap orders `http://www.washingtonpost.com/world/national-security/proposal-seeks-to-fine-tech-companies-for-noncompliance-with-wiretap-orders/2013/04/28/29e7d9d8-a83c-11e2-b029-8fb7e977ef71_story.html`

[6] NSA slides explain the PRISM data-collection program `http://www.washingtonpost.com/wp-srv/special/politics/prism-collection-documents/`

[7] NSA Prism program taps in to user data of Apple, Google and others `http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data`

[8] See also: `http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf`

3. Resource pooling
4. Rapid elasticity
5. Measured service

Cloud computing is further distinguished between private and public clouds. Private clouds are mainly clouds which are operated by companies and are used inside their intranet only or via VPN access also from outside, but are not accessible for non-company members. Public clouds are open and accessible for everyone, not only business users but also private users can use the cloud services. Of course open means not free and many cloud service providers charge the users for providing their services.

Besides that there are also community and hybrid clouds. A community cloud can only be accessed by a specific community of consumers from different organizations that have shared concerns. A hybrid cloud is a composition of two or more distinct cloud infrastructures (e.g. a public and a private cloud).

2.1.1 Software as a Service (SaaS)

This is the level which consumers are most familiar with. Providers are offering access to an application running on their web servers, usually as web application. Common examples are file storage, email, social networking and other software applications like word processing.

2.1.2 Platform as a Service (PaaS)

Providers are offering a platform where the consumers can deploy and run their own applications on without having to manage the underlying hardware. Tools and libraries as well as the network and storage space are also provided. Examples are Windows Azure and Google App Engine.

2.1.3 Infrastructure as a Service (IaaS)

Raw computing power and storage space is provided. Consumers can fully control the underlying virtual machines, including operating system, network and storage space. Providers in this category are Amazon EC2 and Rackspace Cloud Services.

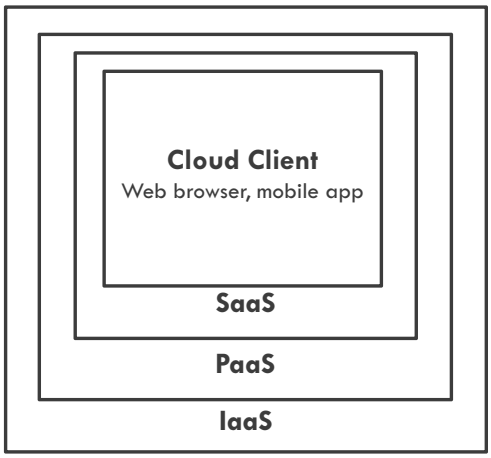

Figure 1: Cloud Service Hierarchy

2.2 Cloud Computing Advantages

In this subsection the benefits of cloud computing for consumers, service providers and governments are outlined.

2.2.1 Advantages for Users

Using cloud computing should first of all have advantages for the users. One of the most important advantages for users is cost saving. Many cloud services are provided for free and offer enough functionality for most of the users. Therefore, users can save much money by using cloud services. Also with paid services there is a big savings potential. Regular software has high one time costs for licenses and there is also the need for an IT technician to do the maintenance, updates etc. All these costs are not there if cloud services are used, only the regular costs per use.

Another not less important advantage is increased flexibility. Users only have to pay for services and capacity which they are really using. So if they need less they pay less and if they need more they can simply add more which of course leads to higher costs but it is much more flexible than adding another server to the company internal IT resources. The addition or removal of processing units or storage space does only take seconds to minutes and not days like it would in a company internal data centre.

Besides that data which is stored in the cloud can be accessed from anywhere in the world without the need for specific software, a simple web browser is sufficient for most of the applications. As all the calculations are done on the service provider's servers so also older and low power devices like mobile phones can be used to work with cloud services without any impact on the service quality.

2.2.2 Advantages for Service Providers

Cloud computing also has benefits for the service providers. First there are no more software piracy issues. As the cloud software runs on the server of the cloud service provider and only the user interface and the results are transferred to the user, there is no easy possibility for software pirates to just crack and copy the software, except if the server itself has some security vulnerabilities.

With the same argument also a protection from reverse engineering is provided. If the code never leaves the server of the service provider, also reverse engineering is nearly impossible. This protects patented algorithms from reimplementation not only by hackers but also by competitors.

Regarding customer retention (even if not voluntary) vendor lock-in is another benefit. There are no standardized interfaces for cloud applications and services, therefore if a user or a company decides to use one service, there is no easy way to switch to other service providers without modifying great parts of the application. So users are discouraged from changing the service provider, which has the benefit of keeping users, especially business users.

A cloud service provider may also gain higher revenues due to the subscription payment model. Most cloud services are paid per use (per hour, per gigabyte, etc.). Standard software is only paid once per license. If the cloud application or infrastructure is used frequently and for longer periods of time, in the long term the revenues for the service providers can be higher than with selling a single one-time license.

Cloud service providers do not only benefit from the subscription payment model, they can also do carefully targeted advertising. As most service providers have access to the private data, including emails, business letters, shopping lists, etc. of their users they can utilize this data to provide highly personalised ads, doing user profile marketing, data mining and so on which will lead to a quite reasonable profit for them.

2.2.3 Advantages for Governments

Governments profit much from cloud computing especially if it comes to surveillance. Whereas in former times it was necessary to physically observe a suspect, to manually install wiretaps, to do a "black bag job"[9] or to steam open mails to get useful information for criminal prosecution, nowadays surveillance can be achieved at near zero marginal cost with significantly reduced manpower and nearly total elimination of the physical risk for the agents. No raids have to be done and also the expensive extraction of the data from a suspect's PC after a successful raid does not have to be done. Surveillance causes only fixed costs, which are to pay at the time the necessary equipment is installed, e.g. digital switches with surveillance capabilities, but as soon as the equipment is in place no more costs are arising. Therefore, the costs are almost independent of how many suspects are to be observed, except a provider decides to charge for surveillance[10]. This not only induces higher costs[11] but also leaves a paper-trail so that at least the surveillance can be tracked (who, how long, which data, etc.). In addition if a provider charges for surveillance and as the governmental agencies only have a fixed budget they will be more careful on which suspects surveillance is performed instead of doing random "fishing-trips".

Providers have dedicated legal compliance departments, open 24h a day, which are only there to process enquiries from governmental agencies. So for the agencies just a phone call or a mail to the provider is sufficient to get the desired data. Some providers decide to give the agencies full access to their systems so that there is not even the need for them to contact the provider in advance, just a few clicks are necessary to get the desired data. This also means that there is nearly no possibility to track the surveillance and disclosure of personal data as the providers do not log the data access and the agencies are not willing to disclose any information about that. Another advantage for governments and their agencies is that many cloud providers do not delete data immediately, so they can also request previously deleted data[12]

2.3 Cloud Computing Issues

Cloud computing does not only have advantages. There are some disadvantages and issues arising with the use of cloud services as well. One obvious disadvantage is that an active internet connection is necessary for using a cloud service. Another concern of many users have is that their data is no longer stored locally and they may loose control over it as soon as it is stored in the cloud. But some users

[9] Agents break into a suspect's house and covertly install microphones with remote transmitters.
[10] Which he is allowed to do so by law
[11] Which are still much lower than the costs in previous days
[12] The storage periods range from a few weeks to over a year depending on the provider.

may not even have a choice to use cloud services or not or may not be aware that they are using a cloud service at all which can also be an issue.

The transmission of data over the internet also imposes threats to data confidentiality, on the one hand during transmission, especially in case an unsecure connection is used and on the other hand while the data is stored, if it is stored unencrypted because there may be possible data leaks on a provider's server.

Not only for private users but especially for companies several legal issues arise due to different laws. Many laws are only applicable in one country, or inside the European Union or the USA. As the exact position where the data is stored in the cloud is not known, it is also unclear which laws are applicable. In addition many local laws treat data which is stored locally different than data which is stored somewhere on the web. This makes it even harder to decide which law is applicable. Different countries also have different perspectives regarding privacy[13] so in some cases companies may not even be allowed to use cloud services due to an uncertain legal situation. To cope with this problem, e.g. Europe has introduced its safe harbour convention where a cloud service provider can decide to obey stricter privacy conventions than demanded by the law and is therefore treated like a European company.

2.4 Attackers and their Interest in Private Data

There are three different groups of attackers which may have an interest in getting access to private user data stored in the cloud: hackers, cloud service providers and governments.

2.4.1 Hackers

The main interest for hackers in a user's private data is for illegal activities. Therefore, the most interesting data is credit card information, bank account details, health records, bank login details and so on. Hackers may gain a reasonable profit by selling this data. Unlike the service providers and governments hackers have no legal possibilities to access the user's data; instead their activities are also subject to criminal prosecution.

2.4.2 Cloud Service Providers

Like hackers cloud service providers want to gain access to their users' data as well, mainly because of achieving profit. In contrast to hackers they have legal possibilities to access this data, mainly due to the terms of service agreement[14], a contract which every user has to agree. Many service providers scan user data on tags which are then used to show highly personalized ads. But also more complex data and statistics are recorded, bundled and analysed (data mining) to be able to do so called user profile marketing, making prediction on what items a user might buy in the near future, what is his next travel destination etc.

2.4.3 Governments

Governments have the most extensive legal possibilities to access private user data stored in the cloud. Their main interest in this data is to support fighting crime and terrorism and of course surveillance. Governments also have possibilities to force cloud service providers to disclose otherwise protected or encrypted user data to them.

3 Legal Aspects

The laws which are discussed in this section are mainly applicable in the US. But as the majority of cloud service providers is located inside the US or stores data on servers located inside the US, these laws are also affecting European users of these service providers. As the main focus of this paper is on the technical aspects and not the legal ones, the relevant laws are only outlined to get an impression of the legal aspects enabling the data disclosure.

[13] E.g. the USA has a much looser conception of privacy than the European Union, not to mention countries like India, China, and so on.

[14] The TOS is covered in detail in the following section.

3.1 Government's Legal Possibilities to Access Private Data

Governments have the legal authority to order firms to turn their own technology against their customers as long as there is no complete disruption of the service and to circumvent any privacy enhancing technology. Before we discuss the legal possibilities of governments doing so, we should define what private data is:

> "Any information relating to an identified or identifiable individual (data subject)." [15]

Some of the relevant laws which are either there to protect personal data or may help governments to force service providers to disclose the data should now be explained briefly.

The Fourth Amendment is there to protect US citizens against unreasonable search and seizure depending upon a person's reasonable expectation of privacy. But it does only protect data which is stored locally and is not applicable to data stored in the cloud.

The third-party doctrine enables the government to obtain files and other personal data with a mere subpoena, which means no court order is needed. The doctrine can be applied as soon as private information is shared with someone else, where courts consider storing data in the cloud is like sharing it with the service provider. The doctrine becomes useless as soon as the data is stored encrypted because it does not enable the government to obtain the encryption key.

Additional relevant laws include the USA Patriot Act, which enables the FBI to access any business record if a court order[16] is issued. The stored communications act determines which communication data is stored, how long this data is stored and distinguishes between electronic communication services (ECS) and remote computing services (RCS) and applies differing privacy protections to each. Most cloud services are regarded as RCS which has much lower privacy protections. The act distinguishes between two levels of privacy, one if the data disclosure is voluntary in case of an emergency and another if it is compelled by the government. However an RCS provider receives much lower protections against compelled data disclosure.

Among the other relevant laws there is the Wiretap Act (Title III) which regulates the collection of content passed over wire and electronic communication and is in combination with the All Writs Act the preferred legal tool for law enforcement agencies. In addition there is also FISA (Foreign Intelligence Surveillance Act) which permits governments to force service providers to install hidden backdoors.

Another law deals with the "emergency voluntary disclosure" where it states that a provider can decide to voluntary disclose user data if

> "... provider, in good faith, believes that an emergency involving danger of death or serious physical injury to any person requires disclosure without delay..."[17]

Whether a specific emergency gives reason to disclose the data relies on the judgement of the service provider. For further details the interested reader is referred to [6] and the corresponding legal texts.

3.2 Terms of Service Agreement

According to the contractarian paradigm the terms of service agreement (TOS) is the main lawful documents which constitutes the rights and liabilities of both, the cloud service provider and the user. So it would be really important for users to read it carefully before accepting it.

Most service providers do not guarantee that the data which a user stores will be kept confidential. They also preserve the right to manipulate, disclose and delete data even without noticing the user. They even preserve the right to disclose parts of the data to third parties[18]. Moreover, they disbar any liability in case of service interruption, data loss, error and inaccurate or untimely results. Another concern should be that the TOS may change without notification[19]. This may lead to the impression that the cloud service providers are capitalizing their users' lack of knowledge about privacy issues in cloud computing.

Following are some excerpts of actual TOS agreements.

[15] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Strasbourg, 28.I.1981, `http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm`

[16] But only a magistrate court order is needed.

[17] 18 U.S.C. § 2702 : US Code - Section 2702: Voluntary disclosure of customer communications or records `http://codes.lp.findlaw.com/uscode/18/I/121/2702`

[18] Mainly to be able to provide advertising companies with keywords.

[19] This is only allowed in several countries, not including the European Union.

3.2.1 Google's Free Cloud Services

Google's advertising system relies on user data as the following to excerpt from their TOS confirm:

> Google reserves the right . . . to prescreen, review, flag, filter, modify, refuse or remove any or all Content from any [Google] Service.[5] [20]
>
> ...the Gmail filtering system also scans for keywords in users' e-mails which are then used to match and serve ads. When a user opens an e-mail message, computers scan the text and then instantaneously display relevant information that is matched to the text of the message.[5] [21]

3.2.2 Yahoo's Privacy Policy

Same as Google, also Yahoo reserves the right to collect and use users' data for providing personalized ads.

> Yahoo! collects personal information when you register with Yahoo!, when you use Yahoo! products or services, when you visit Yahoo! pages or the pages of certain Yahoo! partners, and when you enter promotions or sweepstakes. Yahoo! may combine information about you that we have with information we obtain from business partners or other companies.
>
> We provide the information to trusted partners who work on behalf of or with Yahoo! under confidentiality agreements. These companies may use your personal information to help Yahoo! communicate with you about offers from Yahoo! and our marketing partners.
>
> Yahoo! displays targeted advertisements based on personal information.
>
> We respond to subpoenas, court orders, or legal process, or to establish or exercise our legal rights or defend against legal claims.[22]

3.2.3 Mozy's Privacy Policy

There are also some positive examples, as the one from Mozy, a cloud service provider who guarantees its customers the confidentiality of their stored content:

> We will not sell or market the email addresses or other collected personal information of registered Users to third parties. We will not view the files that you backup using the Service.
>
> Mozy does not disclose Personal Data to third parties, except to process credit card information for orders.
>
> Mozy does not disclose Personal Data, including the data you back up with the Service, unless disclosure is necessary to comply with an enforceable government request such as a warrant.[23]

4 Threats and Attacks

The threats and attacks discussed in this section will focus on web applications which are executed in a web browser, therefore on SaaS cloud services. This section is divided into the three main groups of attackers: hackers, cloud service providers and governments. For each of them several possible attacks are explained.

4.1 Hackers

4.1.1 Session Hijacking

One of the most well known attacks against cloud services is session hijacking. This attack is only possible if https is not used at all or only during the login process. Unfortunately most providers only offer https for login, not during the whole session. E.g. Google has now decided to use https by default on all it's services, Facebook has a hidden option to enable https and Yahoo does not provide https at all. The providers state that there is no customer demand for https. But many customers do not know what risks they are exposed to if not using https therefore the missing customer demand is due to a lack of

[20] Google Terms of Service, supra note 126, § 8.3 `http://www.google.com/intl/us/policies/terms/`
[21] Google Privacy Center: Advertising and Privacy, supra note 130 `http://www.google.com/intl/us/policies/privacy/`
[22] Yahoo! Privacy Policy, `http://info.yahoo.com/privacy/us/yahoo/details.html`
[23] Mozy Corporation Privacy Policy, `http://mozy.com/privacy`

information. The main reason why the providers do not offer https by default is an economic one. Https needs computing power, therefore more servers are necessary to maintain the service level constant if https is used instead of http. Adding more servers means higher running costs and a lower profit. So many providers tend to offer https as a hidden option, because only a very little percentage of all users will activate this option while the majority does not use https because the do not know about it. This minimizes the additional costs.

Attack Procedure The user logs in to the cloud service via https. After the login data is verified, the service provider sends some session information (e.g. a cookie or a session ID) via plain http. At this point an attacker which is able to intercept the network traffic, e.g. if he is in the same public Wi-Fi network or uses the same network hub, captures the session information. But he does not have to capture it at this point, as the session information is transmitted with every action the user performs, so the attacker has just to be able to intercept the network traffic at any time. As soon as he got the session information he can do everything also the user can without anyone, neither the user nor the service provider, noticing. He can impersonate the real user as long as the real user stays logged in and can read all the information which is transmitted between the cloud service provider and the user. Therefore, it is also very important for the user to log out every time he leaves the website or the cloud service and not just to close the browser windows. Else the session cookie or ID remains valid and the attacker can just continue using the website like he is the real user as long as he wants.

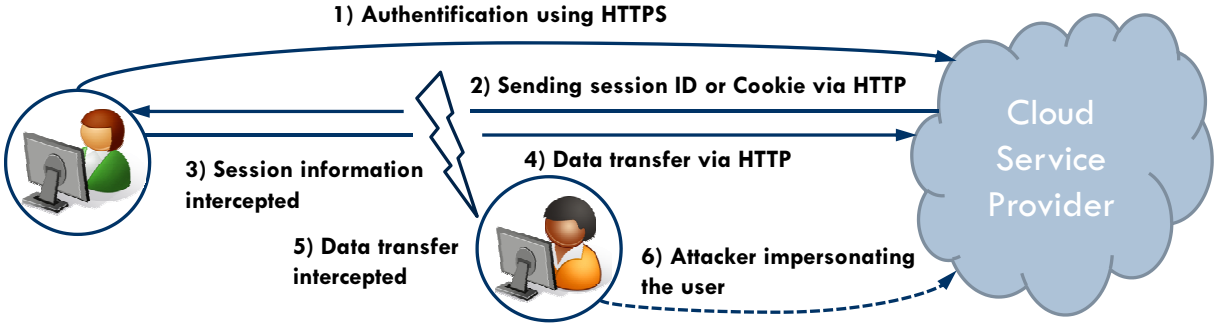

Figure 2: Session Hijacking Attack Diagram

There are also some easy to use, automated tools which exploit this weakness, e.g. DroidSheed, FaceSniff, etc. The main problem with this attack is that especially if a cloud email service is affected, the attacker might get to know several other passwords and personal information. As such a data theft is not really obvious[24], like a stolen car or stolen personal documents are, it can take a rather long time until the theft is discovered. If it is discovered, neither the user nor the service provider knows what lead to that theft. It could also be a leaked password, an eavesdropper getting to know login information and so on.

The only solution to this problem is to force the service providers to turn on https by default. Providers which do not want to do so should be forced by law to show a warning message to users which are using http only, like:

> "You are using an insecure connection. Every data which is transferred could be intercepted. For your own safety it is highly recommended to turn on https. Do you want to proceed?"

4.1.2 Man-in-the-Middle Attack

There are several different types of the man in the middle attack and one of them is explained here. This one only works if the login page is a plain http one and only the actual login information is transferred via https.

An attacker tries to modify the DNS information requested by the user, e.g. by forging DNS packets, DNS cache poisoning or ARP spoofing. If he is able to do so, he can redirect the user to his own "login page" which may look exactly the same as the original login page. The user does not see any difference and since it is plain http also the browser does not warn the user that he is on a different site now (as according to the DNS information it is indeed the correct one). Now the user enters his login information

[24] Neither the user nor the service provider notices anything about the attack

on the fake login page, the attacker gets this information in plaintext, encrypts it using https and sends it to the cloud service provider. For the provider it is exactly the same as if the user logged in directly. Now the attacker can relay all the traffic between the user and the service provider. Of course he can read this traffic, as it is unencrypted. In addition the attacker also knows the login information of the user.

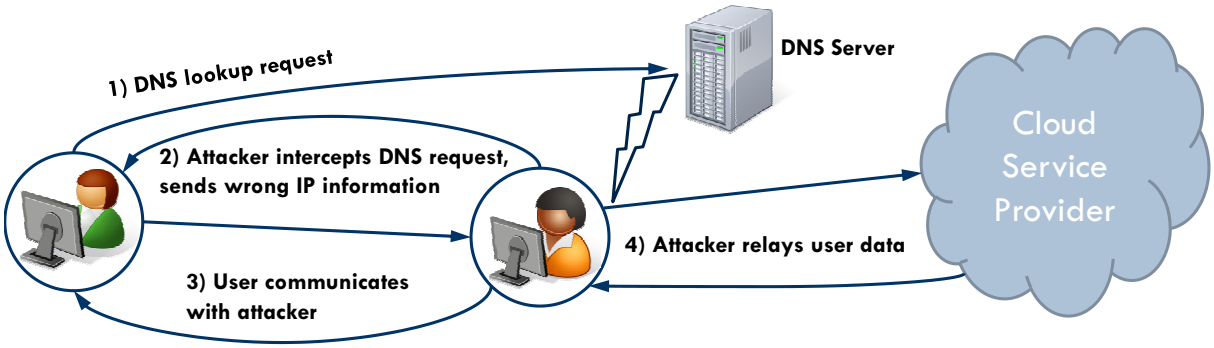

Figure 3: Man-in-the-Middle Attack Diagram

The majority of cloud service providers are using an https login page, where the traffic redirection is not longer possible this way[25] due to the SSL certificates which are used with https. The browser would warn the user if the traffic is redirected as the certificate does not match the hostname or a plain http site is received instead of an https one. Therefore, this type of man-in-the-middle attack is not a big concern any more.

4.1.3 User Interface Attacks

Most if not all cloud applications are accessed using a web browser. Modern web browsers tend to store all kinds of information, including browsing history, login information, form data etc. All attacks under this category mainly exploit bugs and other security issues present in web browser software. If some specific web browser has known vulnerabilities these can be used to trick the browser to show a fake URL instead of the correct one and therefore showing a fake login page to the user. It may also be possible to gain access to the stored browser cache, including stored login information and passwords. Another possibility is to trick the browser to insert stored auto-complete data (e.g. user name and password) on a different website than the one the browser stored it for. The user may also be fooled to think a fake website is the real one by homographic attacks which are using international characters in the URL which look like national ones and there is no difference at the first glimpse. Moreover, an attacker can also try to fake the https lock icon by exploiting browser vulnerabilities.

Nearly every software has some bugs and security vulnerabilities but most modern web browsers only have minor ones and if there are some more severe, they are usually fixed in a few days, so this attacks are only a minor issue if it comes to data confidentiality and privacy in cloud computing.

4.1.4 TLS Null-Prefix Attack

This attack exploits the different implementation of the TLS common name (CN) used by the certificate authorities (CAs) and web browsers. TLS/SSL and also https relies on a chain-of-trust using certificates for each website. Such a certificate is issued by a so called certificate authority and only valid for a given host name. This host name is specified in the common name (CN) field. According to the TLS standard, the CN field is a Pascal string, where the length of the string is stored before the actual string. All CAs treat the CN field as a Pascal string. In contrast, most web browsers are using the C string representation, where a string is always terminated by a \0, therefore the length does not have to be stored explicitly.

An attacker who wants to attack the website www.website.com may ask a CA to issue a certificate for www.website.com\0.attacker.org, where he really owns the domain attacker.org. As subdomains are allowed, and the \0 is ignored inside a Pascal string, the CA will check if he really is the owner of attacker.org and after successful validate issue a valid certificate to the attacker.

Now the attacker just has to modify the DNS information sent to the user to redirect the traffic to his website www.website.com\0.attacker.org instead of www.website.com. The users web browser uses

[25] But is possible as described at the TLS Null-prefix and the SSL interception attack

the C string representation for checking if the certificate is valid and also the current URL is correct, so it will only compare the string until it reaches \0, which is until www.website.com. As this is equal to the website the user entered inside the browser, the web browser shows that it is a valid certificate for the site of the attacker and the user does not notice that he is on a fake website, as also the https lock icon is shown.

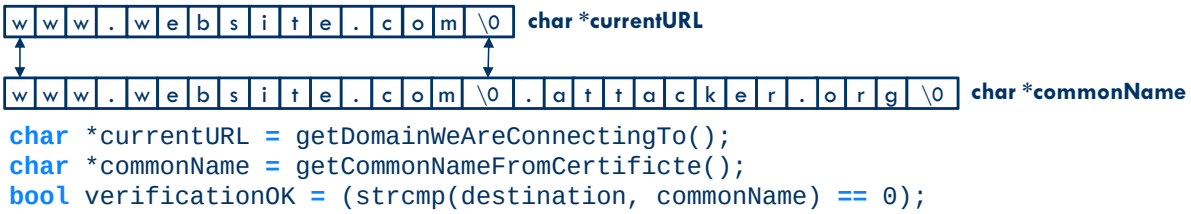

Figure 4: TLS Null-Prefix String Comparison

Of course soon after this attack got public most TLS implementations were corrected to treat the CN field as Pascal string, therefore this attack should not be a security problem any more.

4.2 Governments

4.2.1 Bypassing Storage Encryption

Storage encryption deals with the encryption of the data as soon as it is stored on the servers of the cloud service provider, not to confuse with transport encryption like https and SSL/TLS does. There are two different types of storage encryption according to where the encryption key is stored:

- Only the user has the encryption key
- Only the provider or both the user and the provider have the key

The second option may be better for the provider but is definitely a big threat for data confidentiality. Because the provider knows the key he can decrypt the user's data without the user's knowledge. Moreover it can also be forced by law to hand over the key to governmental institutions. In addition also a data leak at the service provider may lead to a disclosure of the encrypted data. This option may provide a certain protection against insider attacks, at least if the key is only known to a few employees of the service provider but it definitely provides no protection against governmental surveillance.

Thus the first option is preferable. If only the user has the key, the provider cannot be forced to hand it over to governments. Furthermore also data leaks at the service provider do not cause problems regarding data confidentiality as only the encrypted data gets disclosed which cannot be decrypted without the key. But there are also some possibilities how the provider might get to know the key. If for example the key is entered on a web interface, the provider may store the key only in RAM and use it as long as the user stays logged in for encryption and decryption but it may also log it to HDD without the user noticing. The same is true if software provided by the cloud service provider is used. Even if he states that the whole encryption and decryption is done on the users PC, the software might be manipulated so that the key is covertly transferred to the cloud service provider without the knowledge of the user.

Most service providers do not offer any kind of encryption and justify that with a lack of consumer demand. But the main reason for them not to provide storage encryption is that if the first option is used, they are not able to decrypt the data and therefore not able to read the contents. As many service providers are gaining profit by scanning user data and showing highly personalized ads they have to be able to read the contents of users emails, texts and other personal data or else their business model will fail.

Case Study - Hushmail Hushmail is an US email service which provides an encrypted mail service using PGP. They are offering a server-side and a client-side encryption[26]. In the case described here, the US law enforcement wanted to have a look at the emails of a drug dealer who was using the email service from Hushmail with the server-side encryption. Therefore, the court ordered Hushmail to store the user's encryption key to HDD and not only in RAM. Hushmail changed the software, without anyone noticing,

[26] Corresponding to option 2 and option 1 from the explanation.

according to the court order. The next time the dealer logged in, they recorded the key and were now able to decrypt all his emails. In this case also the client-side encryption would not have helped him. As explained above Hushmail could just have changed the java applet, including a backdoor in the way that the key is silently transferred to their servers.

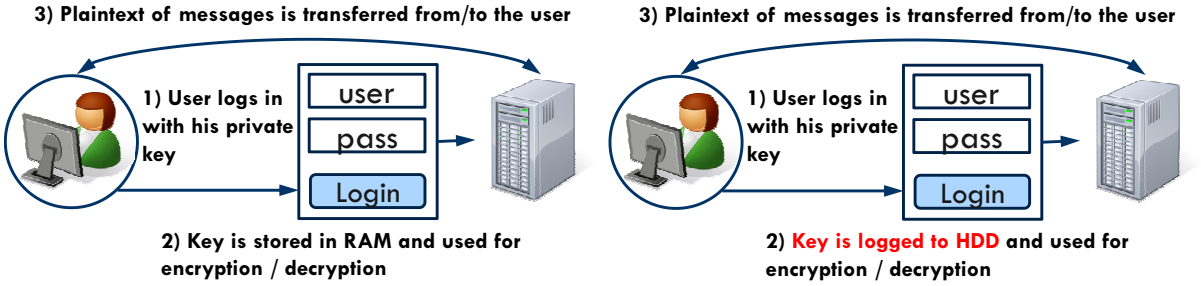

Figure 5: Hushmail - Insertion of a Hidden Backdoor

4.2.2 SSL Interception Attack

As previously explained, SSL relies on a public-key infrastructure involving a chain-of-trust, built up of several certificate authorities (CAs) to validate a website's certificate. The top level CAs in this chain are called root CAs. Browser vendors trust the root CAs. Over a chain starting from the CA which issued the server's certificate it is always possible to go up until reaching any of the root CAs and validate the certificate. CAs below the root CAs are called intermediate CAs. There are also intermediate certificates which can be issued by any CA and can be used to sign other certificates to be valid certificates like the ones signed by the original CA.

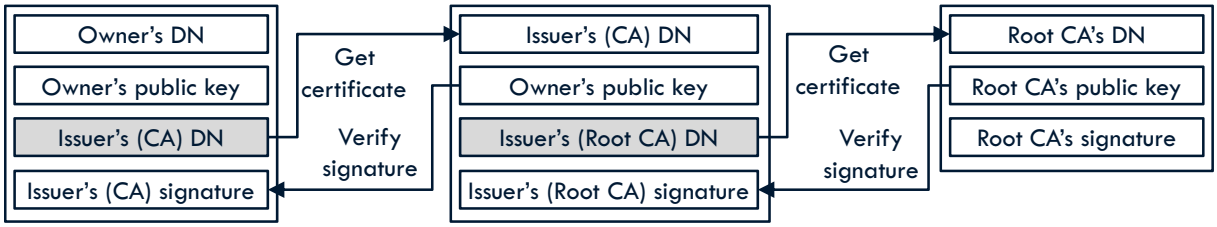

Figure 6: SSL Chain-of-Trust with one intermediate CA (DN...Distinguished name, similar to CN field)

For governments it is possible to do a so called compelled certificate creation attack. This attack is possible because using SSL it is technically possible for any CA to issue a certificate for a website, which has already obtained a valid certificate from any or even the same CA. So there can be more than one certificate per website[27]. Therefore, it is possible for a government to compel a CA[28] by law to issue a false certificate for any website. The government may also compel the CA to issue an intermediate certificate which can then be used to sign further false certificate without involving the CA. As soon as the government has the false certificate, the only thing left to do is to redirect the user's traffic to a website also owned by the government, e.g. using DNS spoofing or simply by manipulating the DNS entries if the DNS server is owned by the same government anyhow.

If the user now wants to access, e.g. the website of his email provider, the DNS lookup request is redirected to the government's website, showing the user the correct URL. As the government has a certificate with the same hostname as the one of the users email provider, the browser will validate this certificate and indicate the user that everything is fine. Now the user enters his login data, which is transferred to the government's server over https, the government is able to decrypt the data[29], read, store or process it. Then the government's server opens a https connection with the email provider using the real certificate of the provider, sends the user's data, which is decrypt it and then encrypt it again

[27] Which means per host name or common name (CN) field in the certificate.

[28] Governments also own CAs but they normally force other CAs to issue this false certificates because the issuer of the certificate is recorded in the certificate and then the attack might be discovered more easily and in addition can be traced back to the government.

[29] As the user created an https connection with the government's server and not as he believes with the email provider.

to send it via https, to the email provider. The only thing which is different for the provider is that the government's server has a different IP than the user, but as he cannot know the user's IP[30], he does not notice anything. From now on, all the https traffic is relayed over the government's server, enabling them silently and covertly intercept, read, store and process the entire user's traffic.

For some people this attack might seem rather theoretical, but there are already specific devices which can be used to automatically perform this attack. One example is Packet Forensics 5-Series, a device not much bigger than a standard internet switch. If provided with an intermediate certificate, this device can be plugged in and will do all the necessary steps automatically, including the creation of the fake certificates. Moreover, there is some evidence that this attack and also this device are already used in practice.

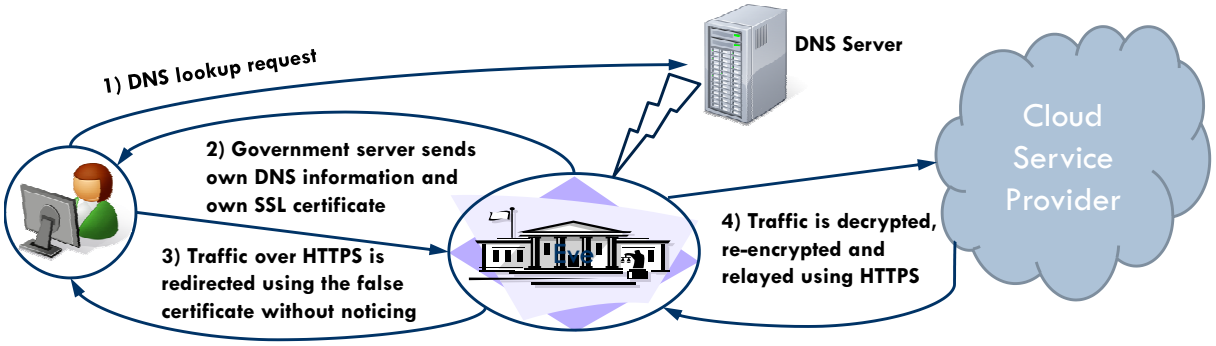

Figure 7: SSL Interception Attack Diagram

As the compelled certificate creation attack is a severe threat for data confidentiality and privacy, a possible countermeasure, the Firefox plugin CertLock is presented in Section 9. For an in-depth explanation of the attack and also the CertLock plugin please refer to [8].

4.2.3 Hidden Backdoors and Unnoticeable Software Changes

Regular software products may have rather slow adoption rates because they require updates to be installed manually. For software companies this is a problem, because the time, which passes between a security vulnerability becomes known and all users have the patched version of the software, might be long. The user has the advantage to be able to decide if he wants to update the software, or at least in case it is automatically updated he can check if something has changed[31].

Cloud software can be updated fast and has high adoption rates. The user does not have to do anything, only the cloud service provider has to update the software on his server and the next time the user logs in he already uses the new version. The disadvantage for the user is that he does not know if anything has changed, if it is only a minor change. Most cloud products do not have a version number and also standard checksums cannot be applied. So the service provider can do arbitrary modifications to the software without the users noticing. This makes the installation of hidden backdoors much easier as nobody notices.

Another advantage for a cloud service provider who decides or is forced to install a hidden backdoor is that a specific user can be supplied with a specific (backdoored) version of the software, while all other users are still using the unmodified one. The decision which software is loaded is done at login time. If only a single user has the modified software version, the change that the modification will be discovered is much lower than if all users would use this version, as there are always users which are analysing patches and try to work out what has changed (with regular software).

Another problem with these unnoticeable software changes is that there is no technology which could prevent a company from executing a lawful order compelling to insert a backdoor in one of its products.

Case Study - JAP JAP means Java Anonymous Proxy and is open-source software written as a project of a German university which uses different proxy servers enabling users to visit websites without revealing their real IP address. The German government forced the developers to implement a hidden backdoor which logs accesses to certain illegal websites and sends the logs to a government server afterwards. As

[30] The user may also use a proxy server which also shows a different IP and not the user's IP to the provider.
[31] Using version numbers and file checksums.

JAP is open-source, one of the users had a detailed look on the recent changes and discovered the hidden backdoor. This is one more piece of evidence that open-source software cannot be easily modified without anyone noticing.

4.3 Cloud Service Providers

4.3.1 Exploiting User Data

The business model of many cloud service providers offering free services is built on highly personalized advertisements. Therefore, they scan the user data on tags and keywords and look for other companies which are offering products matching these tags. Then they charge these companies a small fee for showing the user an advertisement on the web interface. But they are not only scanning for simply keywords, also more sophisticated methods are used to obtain all kinds of statistics from a user's emails, documents etc. This is described under the term data mining, which means extraction, analysis and usage of data in a way that it was not originally stored for. The providers can link several different kinds of user data together to get highly accurate user profiles which can then be used to do behaviour prediction. This can go as far as the cloud provider knowing when the user has to buy a new pack of toilet paper and then showing an advertisement for toilet paper just in the right moment, even if the user might not have thought about buying it just yet.

It is obvious that storage encryption is against their business model as long as the provider is not able to decrypt the data. The only thing he could do with encrypted user data is to show random advertisements which is much less efficient and therefore much less profitable for the provider. Moreover, service providers claim that users would not even pay for storage encryption or enhanced privacy protection, so there is no need for them to offer it.

Some providers state that they do not scan the data if a paid service is used, but the problem is that nobody is able to check if they do so or not. One can only say that they have the technology to scan and analyse the data and it would be easy for them to use it also with paying users.

4.3.2 Data Leaks through Employees

If the data is stored unencrypted, a displeased employee of the cloud service provider may access the data and disclose it to a third party. As mentioned above storage encryption is an effective countermeasure against this type of data leak, especially if only the user is in possession of the key.

5 User Attitudes and Beliefs

In [3], Ion et al. performed several in-depth interviews with average cloud users from India and Switzerland. They studied users' privacy attitudes and beliefs regarding cloud storage systems. They only interviewed users which do not have an IT background. Their results show that almost all users had severe misconceptions about the rights and the level of privacy which the cloud service providers are offering. Many users still consider local storage safer than the cloud due to the fact that they consider the internet as highly insecure in general. In addition, they fear that they might loose the control over their data once stored in the cloud. Moreover, they consider physical protection, e.g. storing an USB-stick inside a safe more secure than encryption. Therefore, most users do not use the cloud as their main storage and only store less sensitive data, but the perception of what sensitive data is differs among them.

5.1 Perceived Privacy

Users' understanding of the cloud infrastructure is rather limited. They do not know how their data is stored nor exactly where it might be stored. As mentioned above the users consider the internet itself as highly insecure and therefore they assume that it would be easy for hackers to access their private data. They are also aware that governments and cloud service providers might gain access to their data. But on the other hand, they assume that their private data is not interesting for anyone, so the practical risk of a data disclosure is only minimal. This means that they are unaware of the actual privacy risks which they are exposed to by using cloud services.

5.1.1 Terms and Conditions

First of all users usually do not read the terms of service agreements and privacy policies of the cloud service providers. In fact, some users do not even know that such an agreement exists. As a consequence they assume that the provider guarantees higher availability, integrity, ownership rights and also privacy protection. Most users are unaware that the providers reserve the right of unauthorized modification, delayed deletion, arbitrary account disabling and they are not liable for any data loss. If users are presented with the actual TOS agreements they would agree to pay more for better privacy, which is in contrast to the statements of service providers claiming that users would not even pay for better security and privacy protection.

5.1.2 National Differences

The interviewers asked citizens of India and Switzerland and discovered some interesting differences between the two nationalities.

Switzerland: In Switzerland privacy is guaranteed by the constitution. The Swiss are considering surveillance and governmental monitoring as a severe privacy infringement. So it is not surprising that the Swiss are more privacy aware than Indians. In their opinion, everyone should have the right of privacy, including terrorists and criminals.

India: In India privacy is not explicitly mentioned in any law and is not explicitly recognized. Indians do know the concept of privacy but they are not insisting in protection of their privacy. They regard surveillance and governmental monitoring as necessary in fighting crime and terrorism. Indians are convinced that tap-proof technology should not exist because someone, especially bad people, would misuse it for sure.

6 Countermeasures

Note that some of the countermeasures can be taken solely by the cloud service providers or the infrastructure providers, e.g. using https, DNSSEC and the use of homomorphic encryption. Other countermeasures can be taken by the user so it relies on him if he exposes himself to the attacks or applies one of the countermeasures if possible.

6.1 CertLock against SSL Interception Attack

The authors of [8] created a Firefox plugin called CertLock which enables a user to detect the SSL interception attack. CertLock caches the certificates of websites when the user visits it the first time. At each visit the current certificate of the website is compared with the cached one on country-level changes. Because the compelled certificate creation attack is a low probability one, which means it will not occur very often, it is extremely important that CertLock has a very low false positive rate. Otherwise it would unnecessarily scare the user. As certificates of websites can change due to expiration, revocation and other reasons, it would not be feasible to just compare the whole certificate as the false positive rate would be rather high. The authors found out that if an SSL interception attack is going on it is mainly performed by foreign governments. E.g. an US citizen is travelling through Europe and the French government wants to have a look at his private mail. His mail provider is located in the US therefore it also uses a SSL certificate issued from a US CA. If now the French government intercepts the transmission, they will use a false certificate issued from a French CA, so the country information in the certificate has changed and CertLock will detect the attack. It is important to note that the attack is discovered only if the country information of the certificate changes. In the described scenario the US government could invoke a SSL interception attack without CertLock warning the user as the country information in the false and the original certificate are equal.

As the US CAs totally dominate the certificate market this imposes a big security problem. Many European service providers also using certificates issued by US CAs. If now the US government decides to invoke an SSL interception attack, they will just ask a CA for a false certificate or will create one themselves and CertLock will not warn the user of the ongoing attack. Therefore, each service provider should use a certificate issued by a CA located in its country if he does not want to expose users' data to the compelled disclosure by an additional government.

6.2 HTTPS/TLS

As mentioned before, providers should be forced by law to turn on https/TLS by default or if they do not want to do so, they should be forced to display a warning to the user. In online banking https is used by default since many years now, not least because the banks would be reliable for damage resulting from using an insecure connection.

Most https implementations are very efficient today so the overhead imposed by using https is less than 1%. This means that the cost argument of the service providers is obsolete. The usage of https/TLS is relying on the provider's support which has to enable it. The user cannot decide for himself to use it if the provider does not support it.

6.3 DNSSEC

DNSSEC introduces a public-key infrastructure like the one used with https also for DNS. All DNS entries along the path while resolving a DNS name to an IP address are signed with a private key and can be validated using the public key of the next higher authority. This makes it much harder if not impossible for attackers to simply change the DNS information sent to the user by DNS or ARP spoofing. But it does not prevent governments from changing DNS information because it is just as easy for them to compel the DNS authorities as it is with the SSL certificate authorities. Also with DNSSEC the user cannot decide on its own to use DNSSEC. Instead the whole DNS infrastructure has to be changed to support DNSSEC which can only be done by ISPs and service providers.

6.4 Homomorphic Encryption

Using homomorphic encryption it is possible to process encrypted data like plaintext one without revealing the actual plaintext. It is mainly used for doing calculations on the data and it is not sure if it can be used to do tag or keyword searches as this would at least reveal parts of the plaintext. In addition homomorphic encryption is very complex, needs high computational power and it is not feasible for higher amounts of data. Unless there are some breakthroughs in the research community, homomorphic encryption is not suitable to solve the data confidentiality issues in cloud computing.

6.5 Single Site Browser

A single site browser is a specific kind of web browser. Unlike a standard web browser it is used for just a single cloud application or site as the name suggests. It has no address bar, no forward and backward buttons and may look like regular desktop software at a first glance. It is intended that the user has several single site browser instances, one for each cloud application he uses. The advantages of single site browsers are that they only store the history of the specific cloud application, they have no access to the cache of the other instances or the default web browser of the user and as they are limited in functionality they are less prone to security vulnerabilities. This provides a good protection against the user interface attacks described above. The problem is that for the user it may look like he is using desktop software, so inexperienced users may be unaware that they are actually using cloud software.

6.6 Web Application Fingerprinting

This approach tries to create a hash or fingerprint like one which can be created using checksums for files but for web applications. This is done by an external analysis of the web application, comparing link patterns, forms and keywords and generating a unique application and version pair. Of course this method has its limitations as it cannot detect "hidden" changes which do not affect the actual user interface and instead only affecting the algorithms running in the background. As the implementation of a hidden backdoor will not affect the user interface, it can not be detected with today's web application fingerprinting methods. Therefore, these methods have to become more advanced and able to detect background changes to be useful in the future.

6.7 Open-Source Software

The nature of open-source software is that the source code is public and accessible for anyone who is interested. If there are changes made to the software it is easy for an experienced user to get the source code of the new and the old version and compare them. Therefore, if one wants to implement a hidden

backdoor in open-source software, sooner or later one user would discover the specific change and figure out that there is a hidden backdoor now. As soon as this becomes public, which may be rather soon after the change, the hidden backdoor becomes useless because no informed user will use the software any longer.

6.8 Provider Independent Encryption

If a user wants to be sure that neither the key nor the data gets compromised by the provider, the only possible option is to use a separate encryption software and encrypt all the files to be stored in the cloud on the own PC before transmitting them. Advantageously it should be open-source software. Then it can be guaranteed that only the user[32] knows the key which is needed for decryption. Of course this solution is only usable if the cloud is used as file storage, whereas while using cloud applications or an email service no separate encryption on the user's pc can be done. For email services PGP can be used as with storage encryption separately on the user's PC. The plaintext is entered in the PGP software, the ciphertext is then just copied and pasted into the webmail form and the only thing which has to be known therefore is the public key of the recipient.

A problem with this separate storage encryption is that some providers prohibit the storage of encrypted files, at least if a free service is used, and may close[33] the user's account if he stores encrypted files.

7 Conclusion

We have seen the basic concept of cloud computing as well as its advantages. Of course the usage of cloud computing will definitely increase in the next few years. There are many challenges with have yet to be solved, especially with data confidentiality and privacy. Especially privacy is a relevant aspect for users of cloud services. Due to its nature, data stored in the cloud has a higher risk of being accessed by unauthorized individuals, during storage but also during transmission, than data which is stored locally. The discussed attacks are evidence for these threats. But many users are unaware of the actual security risks they are exposed to while using cloud services. Unfortunately cloud service providers have no incentives to provide better security and privacy for economic reasons.

The second actor, if it comes to privacy, are governments. Of course governments have a legitimate need to access data, e.g. for crime prosecution, fighting terrorism, etc. They have established several legal ways in which they can force cloud service providers to disclose a user's personal data. But as it is true for every technology and law, it can also be misused. Especially if no court order is needed and the user does not have to be noticed about the data disclose, the danger of misusing these legal possibilities is rather high. In this context compelled backdoors are a serious problem for cloud services as no one, except the governmental agencies and the service provider, is able to track these backdoors.

Also the use of a separate storage encryption is no magic bullet, as on the one hand, if there is enough time, every encryption can eventually be broken. On the other hand many countries have laws forcing users to reveal the encryption key. Regarding encryption we want to end this paper with a citation of Adi Shamir:

> Cryptography is typically bypassed, not penetrated.[34]

[32] Assuming that the encryption software itself does not contain any backdoor and there are no trojan horses or other malware installed on the user's PC

[33] They reserve the right to do so in the terms of service agreement

[34] Adi Shamir, one of the developers of the RSA cryptosystem.

References

[1] Moritz Borgmann, Tobias Hahn, Michael Herfert, Thomas Kunz, Marcel Richter, Ursula Viebeg, and Sven Vowé. On the security of cloud storage services. *Fraunhofer Institute for Secure Information Technology-SIT, Tech. Rep. SIT-TR-001*, 2012.

[2] Armando Fox, Rean Griffith, A Joseph, R Katz, A Konwinski, G Lee, D Patterson, A Rabkin, and I Stoica. Above the clouds: A berkeley view of cloud computing. *Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS*, 28, 2009.

[3] Iulia Ion, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. Home is safer than the cloud!: privacy concerns for consumer cloud storage. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 13. ACM, 2011.

[4] Harry Katzan Jr et al. On the privacy of cloud computing. *International Journal of Management & Information Systems (IJMIS)*, 14(2), 2011.

[5] William Jeremy Robison. Free at what cost: Cloud computing privacy under the stored communications act. *Geo. LJ*, 98:1195, 2009.

[6] Christopher Soghoian. Caught in the cloud: Privacy, encryption, and government back doors in the web 2.0 era. *J. on Telecomm. & High Tech. L.*, 8:359, 2010.

[7] Christopher Soghoian. *THE SPIES WE TRUST: THIRD PARTY SERVICE PROVIDERS AND LAW ENFORCEMENT SURVEILLANCE*. PhD thesis, Indiana University, 2012.

[8] Christopher Soghoian and Sid Stamm. Certified lies: Detecting and defeating government interception attacks against ssl (short paper). In *Financial Cryptography and Data Security*, pages 250–259. Springer, 2012.