Adopting encryption to protect confidential data in public clouds: A review of solutions, implementation challenges and alternatives


All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
Adopting encryption to protect confidential data in public clouds: A review of solutions, implementation challenges and alternatives

Jyrki Kronqvist and Martti Lehto

Faculty of Information Technology, University of Jyväskylä, Finland

jyrki.kronqvist@jyu.fi

martti.lehto@jyu.fi

Abstract

A shift towards use of public cloud services is ongoing and more and more enterprises will start to use them in the near future. As public cloud services certainly promise to deliver many benefits, this new way of delivering services also introduces new types of risks. Due to the NSA's surveillance programs, non-US enterprises need to reassess the risks of public cloud services provided by US companies and look for available solutions to protect their confidential data transferred and stored in the cloud.

Encryption is seen as a solution to help enterprises full fill the requirements related to security and privacy, but is often challenging to implement. Encryption has its own security problems, like key management. Some cloud service providers have also announced that they are improving their security by encrypting all communications and other information flowing into their data centers.

This paper will explore the common encryption solutions available on the market for enterprises to protect their confidential data transferred and stored in the cloud. In this paper we will review possible challenges enterprises may face while implementing these solutions and if these challenges play a role in the decision to use a public service. We will review also alternatives for data encryption.

Keywords: cyber espionage, cloud services, data encryption, confidentiality, trustworthy

1. Introduction

Revelations of the NSA's electronic surveillance programs ensure that most non-US enterprises need to reassess the risks of public cloud services provided by US companies. The enterprises may cancel the cloud services provided by US high-tech companies or continue to use them while looking for solutions to protect their confidential data in the cloud. Both researchers and practitioners see encryption as a solution for protecting data in a public cloud, but there are practical limitations to encryption being used as a general solution for securing data (Stavinoha, 2013), (Sun et al., 2014). An encryption solution, no matter if it is an enterprises own implementation or provided by a public cloud service provider, will increase the complexity and costs related to the use of cloud services and may prevent enterprises from fully utilizing the flexibility and new capabilities offered by public cloud services.

Encrypted data makes data utilization a very challenging task, for instance keyword search functions on documents stored in the cloud need specific algorithms and tools. Without those usable data services, the cloud will become only remote storage which provides limited value to the users (European parliament, 2013), (Sun et al., 2014). Homomorphic encryption, which could enable some processing of data while it remains encrypted, is offered as a potential solution but it is not a realistic business solution and will require further research and testing (Stavinoha, 2013), (Lauter, et al., 2011).

This paper will review common encryption solutions available on the market for enterprises that protects their confidential data transferred and stored in the cloud. In this paper we will review possible challenges enterprises may face while implementing these solutions and if these challenges play a role in the decision to use a public service. We will also review alternatives for data encryption. This paper is based on online research and carried out as an extensive review of online news, vendor's web pages and articles related to the selected topics. The structure of this paper is as follows. Section two illustrates possible consequences of the NSA’s surveillance programs. Section three will explore the common encryption solutions available on the market for enterprises to protect their confidential data transferred and stored in the cloud and review possible challenges they may face while implementing these solutions. Section four reviews alternatives to encryption. Finally, in section five, conclusions are drawn.
2. Will the NSA’s revelations cause a significant slowness in the cloud adaptation?
A series of disclosures related to the NSA’s surveillance programs was started by Edward Snowden in June 2013. The Guardian (The Guardian, 2014) and The Washington Post (The Washington Post, 2014) revealed the first news about the existence of the NSA’s PRISM program, what it actually is and how it works. After the first news more and more new information about details and other programs have leaked out.

As described in the articles, PRISM is a system that the NSA used to get access to private communications of customers from nine popular public cloud services, including Microsoft, Yahoo, Google, Facebook, Skype, Apple and others. The program enables "collection of data directly from the servers" of the online companies including services like email, chat, stored data, file transfers, VoIP and video conferencing. Another project, called MUSCULAR, is the NSA’s tool to exploit the data links. The NSA and its British counterpart, the Government Communications Headquarters GCHQ are copying entire data flows across fibre-optic cables that carry information through the data centers of the internet companies (Gellman and Soltani, 2014). The XKeyscore system is another tool used by the NSA to collect data about what a user does on the internet. The system indexes e-mail addresses, file names, IP addresses and port numbers, cookies, webmail and chat usernames and buddy lists, phone numbers, and metadata from web browsing sessions. Documents provided by Edward Snowden also show a years-long effort by both the NSA and Britain’s GCHQ to weaken encryption systems so that they could tap emails and internet communications (BULLRUN). There is also suspicion that the NSA has undermined the strength of encryption protocols developed by NIST, the US National Institute for Standards and Technology (Ball, Borger and Greenwald, 2013). The legal base for the access is governed by Section 702 of the Foreign Intelligence Surveillance Act, enacted in 2008.

The US high-tech cloud service providers are the ones repeatedly referred to in the NSA’s surveillance programs. As described in the article (Maxwell, and Wolf, 2014), most developed countries have mutual legal assistance treaties (MLATs) which allow them to access data from third parties whether or not the data is stored domestically. The authors state that “the governmental access to data stored in the cloud exists in every jurisdiction, not just limited to the United States (PATRIOT Act), but including also European countries with strict privacy laws also have anti-terrorism laws that allow expedited government access to cloud data”. Are the NSA’s surveillance programs something on a completely different level compared to other countries? Based on the news articles the NSA surveillance state is robust politically, legally and technically and thus may not be easily compared to what other countries have achieved even if some countries like China continue to use the Internet as a giant surveillance platform (Schneier, 2014). A similar answer was given by John Kerry, the US secretary of state, as he conceded that some of the NSA’s surveillance activities had gone too far and certain practices had occurred without the knowledge of senior officials in the Obama administration (Roberts and Ackerman, 2014). European parliament has requested all EU member states and in particular, the United Kingdom, France, Germany, Sweden, the Netherlands and Poland to ensure that their current or future legislative frameworks and oversight mechanisms governing the activities of intelligence agencies are in line with the standards of the European Convention on Human Rights and European Union data protection legislation and to clarify the allegations of mass surveillance activities, including mass surveillance of cross border telecommunications, untargeted surveillance on cable-bound communications, potential agreements between intelligence services and telecommunication companies (European Parliament, 2014).

A relevant question that may be asked is if the NSA’s revelations caused a significant slowness in the cloud adaptation. Enterprises have adopted cloud primarily because of economic reasons, the cloud is less expensive, more efficient, and offers powerful new capabilities that most organizations may not implement easily in-house. Non-US enterprises may not massively cancel their cloud services provided by US high-tech companies because there are few other competitive options (Castro, 2014 and Milian, 2014). Because of the NSA’s revelations, enterprises are looking after security solutions to protect their data stored in the cloud, implementing in-house private clouds, using non-US cloud vendors or keep the existent legacy on premise systems (Cloud Security Alliance Survey, 2014). Implementation of additional safeguards means extra costs and it slows down the overall public cloud adaption. Due to these changes, enterprises may not receive the significant benefits from cloud services that they expected.

3. Encryption - a cure-all solution to secure data stored in the public cloud
The disclosure of documents on the NSA’s surveillance programs created great concerns within large enterprises about how to protect their data transferred and stored in the cloud. The main concern is the loss of confidential data, like business confidentiality and privacy. This concern has caused large enterprises to look after encryption solutions software such as data encryption applications, to securely use public cloud services.
There is a strong argument that business confidential and privacy data needs to be encrypted before transferring and storing it to the public cloud (Sun et al., 2014). A common objection to the use of encryption for cloud data is that it is complicated for enterprises to deploy in their organizations. The proper encryption implementation requires (Cloud Security Alliance, 2011):

- **Enforcement of appropriate policies and instructions.** Enterprise’s policies and instructions illustrating good data classification and management practices whether all or just some of the data need to be encrypted or protected by alternative method or not protected at all. The importance to properly identify an enterprise’s information assets, who should have access to information and who should not and what kind of information it is going to transfer to the cloud and how to protect it may not be underestimated.

- **Encryption software and tools.** A solution based on strong, known algorithms to encrypt the data. Large enterprises may need to have several encryption solutions for different purposes, like an encryption solution for email exchange and transferring data between systems. Taking a new or a complimentary encryption solution needs to be evaluated that it is fulfils the requirements and is aligned with the existing enterprise ICT architecture and services. New encryption software and tools requires extra efforts from users and thus the user experience needs to be taken into the account.

- **Training of employees.** Enterprise’s higher management, business managers, administrators, users and support organizations need to have a proper training related to the policies and also encryption tools in use. Different users groups and encryption solutions need their own specific training programs.

- **Management of encryption keys.** Key management is one of most difficult processes in public cloud computing. This includes the generation, exchange, storage, use, and replacement of keys. It is highly recommended that the keys that encrypt and decipher information have to be under the control of the user organization.

Encryption helps enterprises preserve some of the benefits of maintaining data on the premises, but is often challenging to implement. Encrypted data makes data utilization a very challenging task, for instance keyword search functions on the documents stored in the cloud need specific algorithms and tools. Without those usable data services, the cloud will become only remote storage, which provides limited value to the users (European parliament, 2013), (Sun et al., 2014).

The recent progress in this area looks promising. As they become available new turnkey encryption solutions will significantly help enterprises, who are concerned about the NSA’s surveillance programs to implement a new type of encryption solution to protect their data in the cloud. The encryption solutions, offered by CipherCloud (CipherCloud, 2014) are specifically designed to work with popular public cloud services such as Salesforce.com, Google Apps, and Microsoft Office 365 (Gould, 2014).

CipherCloud (CipherCloud, 2014), (Jasim et al., 2013) provides a solution to enable enterprises to securely adopt public cloud service and appropriately manage the risks related to data privacy, security, or regulatory compliance. Their offering provides a platform, which is located at the enterprise’s premises including security controls like encryption, tokenization, cloud data loss prevention, cloud malware detection, and activity monitoring. The solution encrypts sensitive information in real time, before it is sent to the cloud, preserving application usability and functionality and keeps the keys that encrypt and decipher information under the control of the enterprise (Figure 1.).
While reviewing the service it was noted that large enterprises operating in several countries may find that implementing the CipherCloud gateway solution requires changes to their existing ICT infrastructure. The main changes are related to access to the cloud services (via the CipherCloud gateway), performance issues and ensuring business continuity. Implementation of the CipherCloud gateway solution may mean implementation of two (or more) geographically separated internet access points having a CipherCloud gateways instance. Redundant internet access and CipherCloud gateways instances are required to allow smooth user access to the public cloud services, ensure business continuity and performance. Internet gateways also require other security equipment like firewalls, anti-virus protection, proxies and routers. Required changes to existing infrastructure will increase the costs related to implementation (see the Figure 2 below).
Implementing the CipherCloud solution also means changes to the users. Those enterprise users working remotely or users using their mobile devices need to first connect to the enterprise's internal network before they may access the data and services available in a public cloud. This may not be ideal from the end user experience point of view as it slows down and complicates the use of the cloud services, like the email service. There are also limitations regarding the services available via CipherCloud gateways, e.g. Google email (Gmail) is available by default, but not Google Calendar or Drive. This prevents users from fully utilizing the flexibility and new capabilities offered by cloud services.

As CipherCloud describes, their encryption gateway provides a Searchable Strong Encryption (SSE) solution, which enables the original application to perform operations such as search and sorting on the encrypted data without changing that application. The solution appears to break up sensitive data (specified by application-specific rules) into tokens (such as words in a string), and encrypt each of these tokens using an order-preserving encryption scheme, which allows token-level searching and sorting (Popa et al., 2011). This type of approach may limit a set of operations available and thus prevent full utilization and functionality of the cloud service.

Another example of the recent changes in the ICT industry is the increased use of open source solutions, like OpenPGP (Pretty Good Privacy, PGP), a data encryption program used most often for email encryption. Based on the report published by sks-keyservers.net (sks-keyservers.net, 2014) the daily creation of unique keys nearly tripled since June 2013, just after the NSA’s surveillance programs first became public (Rushe, 2014). The statistics related to over 80 key servers around the world (Figure 3). The data demonstrated the growth of new PGP key generation in July and August 2013, revealing a trend that has gone from 500 to 2,200 new keys added every day. Today, between 800 and 1,200 new keys are being added every day. The Figure 6 below indicates an increase in adopting OpenPGP encryption across the enterprises, showing that enterprises are becoming aware of security issues and implementing new countermeasures to protect their business critical data.
Google has an initiative to make OpenPGP easier for Gmail users. Google End-to-End is an extension for Chrome-compatible web browsers that implements OpenPGP in the browser so it can be more easily used by webmail applications like Gmail. The extension helps a user encrypt, decrypt, digital sign, and verify signed emails within the browser.

End-to-end encryption means data leaving the sender’s browser will be encrypted until the message’s intended recipient decrypts it, and that similarly encrypted messages sent will remain that way until the receiver decrypts them in a browser. While end-to-end encryption tools like PGP and GnuPG have been around for a long time, they require a great deal of technical know-how and manual effort to use. To help make this kind of encryption a bit easier, Google have released the code for a new Chrome extension that uses OpenPGP, an open standard supported by many existing encryption tools (Google Online Security Blog, 2014). As stated in the blog, this kind of encryption solution will mainly be used for very sensitive messages or by those enterprises that need additional protection. The solution preserves the Gmail functionality and the users may use the solution from any device and anywhere. As a desktop Chrome extension End-To-End is not supported on mobile devices. We may expect that Google’s new extension will significantly boost the adoption rate of the open source based PGP.

Some cloud providers have announced that they will improve their security by encryption, e.g. Google and Yahoo are expanding their efforts to protect their customers’ online activities by encrypting all the communications and other information flowing into companies’ data centers around the world (Rushe, 2014), (Gellman and Soltani, 2014). These improvements focus mainly on how to protect data “in transit”. There are other challenges of encrypting data “at rest” which are typically separate actions requiring different sets of encryption keys, additional key management, and separate processing (Stavinoha, 2013). The cloud providers may offer in the near future encryption solutions which encrypt also the data “at rest”, but they may resist cloud data encryption because encryption with customer controlled keys in a multi-tenant environment is challenging to implement, may be inconsistent with their business model and might require them to modify their existing software systems (Falkenrath and Rosenzweig, 2014). From the user organization point of view the encryption solution provided by the cloud service provider is the easiest to use. It is also fully transparent from the users point view. The downside is that encryption of data “at rest” may not be available in all cloud services and the encryption keys that encrypt and decipher information are not under the control of the user organization.

There is a proposal for a business model based on the concept of using a separate encryption and decryption service (Hwang et al., 2011). In the model, data storage and decryption of user data are provided separately by two distinct providers. Those working with the data storage system will have no access to decrypted user data, and those working with user data encryption and decryption will delete all encrypted and decrypted user data after transferring the encrypted data to the system of the data storage service provider. This model mitigates the risk that the encryption keys that encrypt and decipher information are
under the control of the storage cloud provider, but as a new business model proposal it is not supported by many cloud providers.

The enterprises are adopting the cloud primarily because of economic reasons, it is less expensive, more efficient, and offers powerful new capabilities. Implementing an enterprise wide encryption solution to protect data stored in the cloud is not an easy thing to do despite the recent progress in this area. An encryption solution, no matter if it is an enterprises own implementation or provided by a public cloud service provider, will increase the costs and complexity related to the use of cloud services and it prevents enterprises from fully utilizing the flexibility and new capabilities offered by public cloud services. Furthermore because of encryption they need to change their policies and instruct their users how to manage confidential data and change their behaviour while being online. The comparison of the main encryption solutions is presented in the Table 1 below.

<table>
<thead>
<tr>
<th>Encryption solutions</th>
<th>Pros (+)</th>
<th>Cons (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encryption gateway</td>
<td>a) The encryption keys are under the control of the user organization.</td>
<td>a) Required changes to existing infrastructure will increase the costs related.</td>
</tr>
<tr>
<td></td>
<td>b) Preserves the email service functionality</td>
<td>b) Service may not be accessed from anywhere and any device,</td>
</tr>
<tr>
<td></td>
<td>c) Support mobile devices.</td>
<td>c) Requires users to connect to the enterprise’s internal network before accessing the cloud service.</td>
</tr>
<tr>
<td></td>
<td>d) Support (limited) search of encrypted data</td>
<td></td>
</tr>
<tr>
<td>Browser extension based encryption</td>
<td>a) A solution for very sensitive messages, as it provides end to end encryption.</td>
<td>a) Browser extension based encryption solution may not be supported on mobile devices.</td>
</tr>
<tr>
<td></td>
<td>b) New browser based implementation makes the use rather effortless to users.</td>
<td>b) Preserves the email service functionality, except email content based search</td>
</tr>
<tr>
<td></td>
<td>c) The encryption keys are under the control of the user organization.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d) Service may be accessed from anywhere.</td>
<td></td>
</tr>
<tr>
<td>Encryption provided by the cloud vendor</td>
<td>a) Easy to take into the use.</td>
<td>a) The encryption keys are not under the control of the user organization.</td>
</tr>
<tr>
<td></td>
<td>b) Fully transparent to the users.</td>
<td>b) Encryption of data “at rest” may not be available in all cloud services and by all vendors</td>
</tr>
<tr>
<td></td>
<td>c) Service may be accessed from anywhere and any device.</td>
<td>c) May require changes in the vendor’s own infrastructure.</td>
</tr>
<tr>
<td></td>
<td>d) Preserves the email service functionality.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>e) Support mobile devices.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Comparison of common encryption solutions

4. Alternatives to encryption
Enterprises adopting public cloud services but finding encryption too challenging to implement may look at alternate approaches to protect data in the public cloud. For enterprises using public cloud services having issues with sending sensitive data outside their organization there are alternatives (Cloud Security Alliance, 2011), (Stavinoha, 2013):

- **Tokenization.** A solution a public cloud service is paired with a private cloud that stores sensitive data. The data sent to the public cloud is altered and contains a reference to the data residing in the private cloud.
- **Data Anonymization.** A technology that converts clear text data into a nonhuman readable and irreversible form.
- **Rely on contracts.** A model where an enterprise relies on the contract with the cloud provider to protect the data.
- **Access controls.** A solution where the access controls implemented, e.g. into the data base, provides adequate level of segregation.

In a survey (Stavinoha, 2013), consisting of IT professionals, they were asked to choose a course of action if encryption was not available as an option to secure data in the cloud. The survey indicated that the majority of respondents (56.4%) would not use a cloud service if encryption was not an option and the next highest percentage of respondents (26.4%) replied that access controls would be relied upon in cases where encryption was not available. Only 9.2% of respondents felt that anonymization of the data was a suitable choice and the lowest number of respondents (8%) felt that relying on the contract with the cloud provider to protect the data was an acceptable option. The results were similar when the same question was asked from management.

The enterprises looking for alternatives to encryption may find tokenization as a valid solution to protect sensitive data. Nowadays tokenization may not be supported by many cloud services, but some cloud security vendors, like CipherCloud, provide gateways that transform data into tokens as it traverses the gateway, storing the token mapping in its local cache (CipherCloud, 2014). This tokenization gateway has similar issues as the encryption gateway, such as, required changes to the existing infrastructure and there are limitations to access the cloud services. Tokenization may not be a solution for a cloud service like email, but it may be a valid solution for enterprises planning to transfer some legacy services into the cloud and looking solutions other than encryption to protect data used in cloud services.

A new type of cloud offering, tokenization-as-a-service (TaaS) is available from providers like Akamai (Akamai, 2015). These offerings based on tokenization can help enterprises reduce compliance scope for regulations like the Payment Card Industry Data Security Standard, having a third-party manage this could offer cost savings and ease implementation concerns.

Tokenization may be seen as an alternative to encryption, but it may not be supported by many cloud services. It solves issues related to encryption, like managing encryption keys and infrastructure, as well as avoiding sharing them with cloud service providers. Tokenization is faster than encryption and requires less management overhead. One useful scenario in the future may be a hybrid solution, which employs both tokenization and encryption. Tokenization is used for critical, real-time services that support it and encryption is used for storing data and cloud services that support robust key management. (Shackleford, 2014).

5. **Summary**

The revelations of the NSA's electronic surveillance programs ensure that most non-US enterprises need to reassess the risks of public cloud services provided by US companies and how to protect their data transferred and stored in the cloud. Encryption is seen as a solution to help enterprises to full fill the requirements related to security and privacy and protect their data stored in the cloud, but is often challenging to implement. Encryption has its own security problems, like the key management (Sun et al., 2014). An encryption solution, no matter if it is an enterprises own implementation or provided by a public cloud service provider, will increase the costs and complexity related to the use of cloud services and prevents enterprises from fully utilizing the flexibility and new capabilities offered by cloud public services.

Tokenization might be seen as an alternative to encryption, but it may not be supported by many cloud services. Tokenization may not be a solution for a cloud service like email, but it may be a valid solution for enterprise’s legacy applications handling massive amount of data in the cloud. Tokenization may be seen as a new tool, which enhances the solutions available for enterprises to protect their confidential data in the cloud. One useful scenario in the future may be a hybrid solution which employs both tokenization and encryption. Tokenization may be used for critical, real-time services that support it and encryption is used for storing data and cloud services that support robust key management. (Shackleford, 2014).

Non-US enterprises may not massively cancel their cloud services provided by US high-tech companies because there are few other competitive options (Castro, 2014 and Millian, 2014). Enterprises are searching for security solutions to protect their data stored in the cloud, implementing in-house private clouds, using non-US cloud vendors or keep the existent legacy on premise systems (Cloud Security Alliance Survey, 2014). In some cases enterprises may do an extreme decision not use a cloud service for their business confidential data if encryption is not available (Stavinoha, 2013). Implementation of additional safeguards means extra costs and it slows down the overall public cloud adaption as enterprises are expecting to receive significant benefits from the cloud services and these benefits may not materialize for them in a way as they expect.
References


