International Journal of Advances in Computer Networks and Its Security– IJCNS
Volume 5: Issue 1 [ISSN : 2250-3757]
Publication Date : 30 April, 2015

The main aspects of data security in cloud computing

[Zuzana Priščáková, Ivana Rábová]

***Abstract*— Cloud computing is a modern trend in information technologies. Implementation of the cloud is associated with data security. Data security includes phishing, provider personnel with privileged access, data origin and lineage. Another security aspect is determination of data categorization. In this article we point to two ways for data categorization. The aim of this paper is to determine the safety aspects based on the special security points. We set user login, encryption and data security, risks and data integrity as the main aspects. These aspects we summarized in nondeterministic finite automaton. We determined a new methodology of security using automaton.**

***Keywords*—cloud computing, data security, data categorization, storage, data integrity**

I. Introduction

According to NIST Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. [1]

Cloud computing has "unique attributes that require risk assessment in areas such as data integrity, recovery, and privacy, and an evaluation of legal issues in areas such as e-discovery, regulatory compliance, and auditing," Gartner says. [2]

High popularity of cloud computing requires high demands on the data security. Solutions are being voiced through the adoption of this new security methodology for storing data. In this paper we attempt to clarify the issues from security aspects. The research methodology used to achieve this goal is based on software engineering and information systems design approaches. [3] The aim of the paper is to specify safety aspects through the nondeterministic finite automaton. This automaton is based on the information system and on the cloud model – software as a service, storage as a service.

Zuzana Priščáková
Mendel University in Brno
Czech Republic

Ivana Rábová
Mendel University in Brno
Czech Republic

II. Storage and data security

The main aspects of data security including processing of data (multitenancy, data lineage, data provenance, data remanence) [4]. According to Mather the primary risk is in not using a vetted encryption algorithm (data-in-transit). To information security is more important understand requirement when using a public cloud and a protocol provides confidentiality for data integrity. [5]

Merely encrypting data and using a non-secured protocol (e.g. “vanilla” or “straight” FTP or HTTP) can provide confidentiality, but does not ensure the integrity of the data (e.g. with the use of symmetric streaming ciphers). [4] For a variety of reasons, enterprises often rely on their cloud service providers to maintain ownership and management of the keys, believing that cloud data encryption can only be accomplished in this way. [6]

According to Gartner, organizations have a limit to the amount of time that staff can dedicate to becoming experts in a given solution. [7] Increasing the number of different vendor cryptographic solutions deployed within a given environment increases the level of overall complexity of the overall system due to higher demands on staffing, increased training and the greater risk of misunderstanding a particular deployment configuration dependency. [5], [8]

Although some cloud providers to examine their applications to third parties or to verify security applications third party tools, the data is not platform dedicated exclusively to one organization. Although an organization´s data-in-transit might be encrypted during transfer to and from a cloud provider, and its data-at-rest might be encrypted if using simple storage (i.e. i fit is not associated with a specification application), an organization´s data is definitely not encrypted if it is processed in the cloud (public or private). [3], [4], [5]

The data lineage is more important for an auditor´s assurance or management. Trying to provide reporting on data lineage for a public cloud service is really not possible. Even if data lineage can be established in a public cloud, for some customers there is an even more challenging requirement and problem. [4], [5]

The data integrity refers to data that has not been changed in an unauthorized manner or by an unauthorized person. Provenance means not only that the data has integrity, but also that it is computationally accurate. According to Ning, use of virtualized infrastructure as a springboard may introduce new attacks on the integrity of user data. [9]

The data integrity is defined in as accurate and consistent data stored in the absence of any modification of data between

two updates a file or record. The data integrity is a relevancy of data, consistency and availability. [4], [10]

When end user saved data to the cloud, he assumes that the data (or application, which works in a virtualized environment) are properly secured. Any security cloud model is based on the assumption that the user (customer) should believe in your cloud provider. This trust will file the documents SLA, which generally defines mutual expectations and obligations of the provider and the end user. [4], [10], [11], [12]

Data remanence is the residual representation of data that has been in some way nominally erased or removed. [3], [4], [12] This residue may be due to data being left intact by a nominal delete operation, or through physical properties of the storage medium. Data remanence may make inadvertent disclosure of sensitive information possible, should the storage media be released into an uncontrolled environment (e.g. thrown in the trash, or given to a third party). [4], [10], [12]

The risk posed by data remanence in cloud services is that an organization´s data can be inadvertently exposed to an unauthorized party – regardless of which cloud service you are using. Risk is almost unintentional when you using PaaS or SaaS. [12]

According to Winkler risks to cloud computing data security include:

A. *Phishing*

On indirect risk to data in motion in a cloud is phishing. Phishing is a simple process of sending an email to include links to a nefarious website or malware, and used by many criminals to provide a way in to a network with little comeback to the perpetrator. In addition, spear phishing, where the phishing attack is targeted at an individual or institution is on the increase. [8], [12]

The Chinese hacking group APT1 who are suspected to be state sponsored use spear phishing as a primary attack vector according to incident response consultant Mandiant whilst Websense in their 2013 Threat Report see only 1 in 5 emails as legitimate. [8], [12], [13]

B. *Provider personnel with privileged access*

This risk to data security has to do with a number of potential vectors for inappropriate access to customer sensitive data by cloud personnel. This risk is largely has to do with the potential for exposure with unencrypted data and with privileged cloud provider personnel access to that data. [12]

Evaluating this risk largely entails CSP practices and assurances that CSP personnel with privileged access will not access customer data. [12], [14]

C. *Data origin and lineage*

Proving the origin of information or data has importance in many areas, including patent or proving ownership of valuable data sets that are based on independent analysis of commonly available information sources. Reporting on data lineage may be very difficult to do so with a public cloud and largely due to the degree of abstraction that exists between actual physical resources and the virtualized resources that a public cloud user has access to. [12], [15]

For monitoring the data integrity is important to consider these protocols:

- Service Level Agreement [10], [12], [15], [16],
- Proof of Retrievability [15], [17], [18],
- Protocol based on inserting random guardian in the data file. [19]

To verify the data integrity:

- Protocol call – response [20], [21],
- Verify the correctness of the data stored in the cloud. [13], [22]

For specific information about how data security should be archived, providers should refer to the National Institute of Standards and Technology (NIST) Special Publicaton, 800-88. [12], [23]

III. Data categorization

Organizations across most industries are required by industry or governmental regulations to show how they handle and protect customer information in a secure way. Enterprises in the financial industry face numerous compliance requirements to ensure that they not only protect their customers' information, but that they are not compromising the integrity of the financial systems in which they participate. Organizations need to find ways to categorize this data in ways that make sense for their business and organizational requirements. [4], [12]

Data classification is a powerful tool, that can help determine what data is appropriate to store and/or process in different computing architectures, like the cloud or on premises. Without performing data classification, organizations might under-estimate or over-estimate the value of data sets, resulting in inaccurate risk assessments and potentially mismanaging the associated risk. [24], [25] Data classification is the process used to determine the value to their organizations of specific pieces of electronic information, be they e-mail messages, documents, or databases. [4], [12], [24], [25]

According to Noel, data classification is a process an organization uses to determine which information is more important or more sensitive. "Data across our organization has varying value. The value of the data serves as a guide (on how) to protect it. Data classification lets us determine the value of that data, whether an e-mail, a document, or a database." [24], [25]

Sensitive or otherwise valuable data should be categorized to support data security. By identifying data according to sensitivity, one can implement various strategies to better protect such data. [12] However, cloud data may require protection. Data that a user chooses to store in the cloud may not require protection if it is not sensitive or if it can easily be recovered. But generally, protecting data is a universal requirement regardless of its value, if for no other reason than failing to do so leads to all manner of complexity, consequence, and mischief. [12]

In addition to the prescribed classes of information that is sensitive or otherwise have value and labeling information according to its characteristics, you must protect this data (file permissions, encryption etc.). Also you need identity-based access controls to support organizational access policies. According to Winkler, procedures are also necessary for security across phases of the data life cycle, for instance, to limit exposure of such data when you create copies or backups. [12]

For data categorization you used information security technique – the hierarchical categories. In the hierarchical categories are five levels [12]:

- unclassified,
- confidential,
- secret,
- top secret,
- compartmented.

By Microsoft [24], [25] is data categorization following:

- confidential (restricted) – information that is classified as confidential or restricted includes data that can be catastrophic to one or more individuals and/or organizations if compromised or lost (e.g. personal data, financial records, business material, legal data, authentication data,
- for internal use only (sensitive) – information that is classified as being of medium sensitivity includes files and data that would not have a severe impact on an individual and/or organization if lost or destroyed (e.g. e-mail, documents and files that do not include confidential data),
- public (unrestricted) – Information that is classified as public includes data and files that are not critical to business needs or operations. This classification can also include data that has deliberately been released to the public for their use, such as marketing material or press announcements. In addition, this classification can include data such as spam email messages stored by an email service.

IV. Determination of safety aspects

We summarize these facts to the main points of safety:

A. User login

- Insert login through the form.
- The data are verified through the database. The database is located on a server.
- After successful verification, the access to the system and input data is granted. Input data depend to login.
- Input data are verified. If it is not permitted to enter the data, it is sent to the login request.
- When you finish working with the data, the end user enters requirements for the data storage.

B. Encryption and data security

- Data are classified into categories of sensitivity (public, sensitive, secret). Each category has different security.
- The security level has minimum privileges. Minimum privileges are permissions for the end user to work with data.
- If the minimum privileges are not sufficient, you need to determine new privileges. New privileges are assigned to the data set.
- The system requires encryption after determining authorizations.

C. Risks

- The risks are determined after the data encryption. Risks are procedures that cause data theft. The database stores the risks that have been identified so far. If you identify a new risk, it is inserted into the database. If the risk already is in the database, the procedure should be amended regarding to the information.
- The security policy is based on the identified risks. Security policy is a procedure to deal with a risk if there is any.
- The security policy is determined by boundary data protection. Boundary data protection determines appropriate and inappropriate processes for data. Boundary data protection can only be modified if the database is located.

D. Data integrity

- The data are identified by assigning a specific type of value in the case of external data storage medium.
- To ensure data, a hash value is assigned to data.
- We determine the type of security to data. Type of security is the value of data security. This value affects the boundary data protection, security policy, risk and minimal privileges.
- The data are stored. If the end user requests a continuance, he must provide input.

E. *Nondeterministic finite automaton (NFA)*

We chose nondeterministic finite automaton to illustrate the facts mentioned above. This type of formalism shows the sequence of steps which are necessary to ensure data security and data integrity by cloud solutions. We accent the crossing functions and requirements of data security. Figure 1 shows nondeterministic finite automaton.

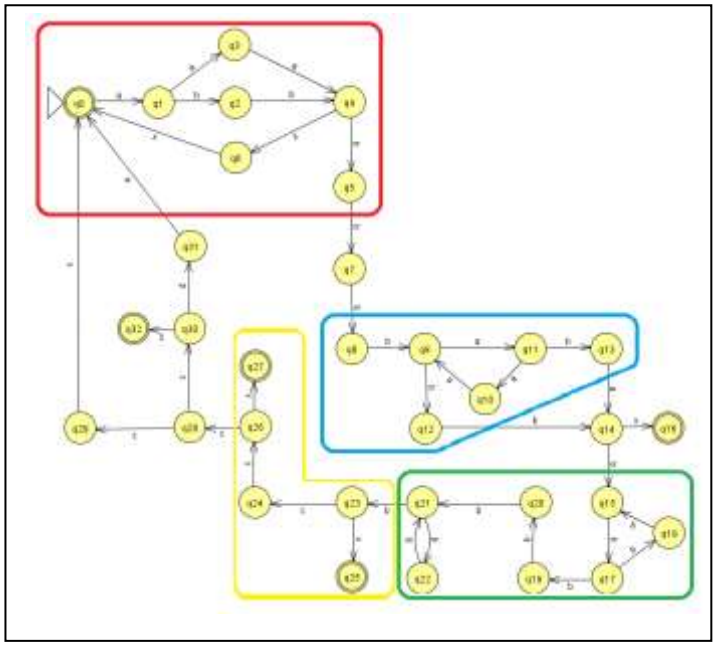

Figure 1. Nondeterministic finite automaton

Nondeterministic finite automaton is specified as:
A = (K, Σ, δ, q_0, F),
wherein,
K= {q0, q1, q2, q3, q4, q5, q6, q7, q8, q9, q10, q11, q12, q13, q14, q15, q16, q17, q18, q19, q20, q21, q22, q23, q24, q25, q26, q27, q28, q29, q30, q31, q32}

Σ={a, b, c, d, x, z}
wherein "a" is user login, "b" is unsecured data, "c" is data integrity, "d" is store data, "x" is wrong processing, "z" is work ending .

The transfer function is shown in the table 1.

Set of initial states is specified as:
q_0={q0}

Set of final states is specified as:
F={q0, q16, q25, q27, q32}
wherein
q0 – input user login,
q16 – unsuccessful encryption,
q25 – unlabeled data,
q27 – unsuccessful security,
q23 – complete the work.

TABLE I. TRANSITION FUNCTION NONDETERMINISTIC FINITE AUTOMATON

δ(q0, a) = q1	δ(q1, a) = q3 δ(q1, b) = q2	δ(q2, b) = q4	δ(q3, a) = q4	δ(q4, b) = q5 δ(q4, x) = q6	δ(q5, b) = q7
δ(q6, x) = q0	δ(q7, b) = q8	δ(q8, b) = q9	δ(q9, a) = q11 δ(q9, b) = q12	δ(q10, a) = q9	δ(q11, a) = q10 δ(q11, b) = q13
δ(q12, b) = q14	δ(q13, a) = q14	δ(q14, b) = q15	δ(q15, b) = q17	δ(q17, b) = q19,q18	δ(q18, b) = q15
δ(q19, b) = q20	δ(q20, b) = q21	δ(q21, b) = q22 δ(q21, b) = q23	δ(q22, b) = q21	δ(q23, c) = q24 δ(q23, x) = q25	δ(q24, c) = q26
δ(q26, x) = q27 δ(q26, c) = q28	δ(q28, c) = q29 δ(q28, c) = q30	δ(q29, x) = q0	δ(q30, d) = q31 δ(q30, z) = q32	δ(q31, a) = q0	

The example of a correct input word is: abbbbbbbbbbbbbbccccda.

For this automaton, these requirements result:

- A correct input word must contain letters "a","b","c","d".

- The input word should be the string “bcccc” to ensure data integrity.
- The input word for accepted automaton satisfying the safety conditions must begin with the letter "a" and end with the letter "a". This string saves data and continues working.
- If the input word has the string cx" at the end of the word, then data integrity is fulfilled and data is not stored.
- If the input word has the letter “x” at the end of the word, then the input word is wrong. NFA is terminated before it meets the security conditions.
- If the input word has the letter "z" at the end of the word, it means that the input word is correct. Data are secured. The end user finished his work.
- If the input word has the string “xx” at the end of the word, it means that data is not available for input user login.

Above requirements are aggregated and requirements are the main aspects of data security. Data security influences the end user and his privileges for working with data, type of data based on the sensitivity, risk procedures and boundary data protection, security type, saving on external media. Unsecured data is data, which did not meet data integrity. This data type is not stored.

v. Discussion and conclusion

As companies turn to burgeoning cloud computing technology to streamline and save money, security is a fundamental concern. Loss of certain control and lack of trust make this transition difficult unless you know how to handle it. [12] According to Mather, the main aspects of data security are multitenancy, data lineage, data provenance and data remanence. [4]

This article describes new aspects of data security according to cloud´s implementations as SaaS and StaaS model. The main security requirements result from the NFA. The research indicates that aspects of data security are influenced by the environment.

Another aspect may be the data categorization, but wrong data categorization may result in low data security. Data categorization should reflect the environment and type of storage. Before implementing cloud, you have to know the requirements of environment.

References

[1] P. Mell and T. Grance, “The NIST Definition of Cloud Computing.,” National Institute of Standards and Technology Gaithersburghil, Special Publication 800-145, September 2011.

[2] GARTNER, IT Glossary - defining in IT industry, 2012. [online]. Available: http://www.gartner.com/it-glossary/cloud-computing/

[3] D. Zissis and D. Lekkas, Addressing cloud computing security issues, Future Generation Computer Systems, 2012, pp.583–592.

[4] T. Mather, S. Kumaraswamy and S. Latif, “Cloud Security and Privacy,” USA: O´Reilly, 2009, p. 338.

[5] M. Lynch, “The cloud wars: $100+ billion at stake,” Merrill Lynch, 2008.

[6] A. Nagarajan and V. Varadharajan, “Dynamic trust enhanced security model for trusted platform based,” Future Generation Computer Systems, 2010.

[7] G. Grealish, “Why protecting encryption keys is critical to keeping cloud data private,” March 2014. [online]. Available: http://www.cloudcomputing-news.net/news/2014/mar/20/when-deploying-cloud-data-encryption-protecting-encryption-keys-is-critical-to-keeping-cloud-data-private/

[8] K. Sheetal and S. Sandeep, “ECC-based anti-phishing protocol for cloud computing services,” International Journal Security and Networks, vol. 8,no.3, 2013.

[9] C. Ning, W. Cong, W. Ming, L. Kui and R. Wenjing, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” IEEE Trans, pp. 222-233, 2014.

[10] J. Rhoton, J. Clercg and D. Graves, “Cloud Computing Protected,” USA: Recursive Press, p. 412, 2013.

[11] S. Pearson and G. Yee, “Privacy and Security for Cloud Computing,” London: Springer, p. 300, 2013.

[12] V. Winkler, “Securing the Cloud,” USA:Syngress, p. 290, 2011.

[13] K. Yang, “Security for Cloud Storage Systems,” London:Springer, p. 83, 2014.

[14] P. Mell and T. Grance, “Effectively and Securely Using the Cloud Computing Paradigm,” 2013.

[15] A. Eswaran and S. Abburu, “Identifying Data Integrity in the Cloud Storage,” International Journal of Computer Science Issues, 2012. [online]. Available: http://ijcsi.org/papers/IJCSI-9-2-1-403-408.pdf

[16] Ch. Marsh, “Data Integrity In The Cloud,” Computer Technology, 2011. [online]. Available: http://www.wwpi.com/index.php?option=com_content&view=article&catid=99:cover-story&id=12800:data-integrity-in-the-cloud&Itemid=2701018

[17] A. Juels and B.S. Kaliski, “Pors: proofs of retrievability for large files,” USA:ACM, pp. 584-597, 2007.

[18] T. Neha and P.S. Murthy, “A novel approach to data integrity proofs in cloud storage,” Journal of Advanced Research in Computer Science and Software Engineering, 2012. [online]. Available: http://www.ijarcsse.com/docs/papers/10_October2012/Volume_2_issue_10_October2012/V2I10-0114.pdf

[19] K. R. Sravan and A. Saxema, “Data integrity proofs in cloud storage,” Communication system and networks, pp. 1-4, 2011.

[20] B. Thuraisingham, “Developing and Securing the Cloud,” Auerbach Publications, p. 730, 2013.

[21] W. Cong, S.M.Ch. Sherman, W. Qian, R. Kui and L. Wenjing, “Privacy-Preserving Public Auditing for Secure Cloud Storage,” IEEE Trans.Computers, pp. 362-375, 2013.

[22] L. Ming, Y. Shucheng, R. Kui, L. Wenjing and H. Thomas, “Toward privacy-assured and searchable cloud data storage services,” IEEE Network, 2013.

[23] NIST, “Guidelines for Media Sanitization,” National Institute of Standards and Technology Gaithersburghil, Special Publication 800-88, September 2006. [online]. Available: http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_with-errata.pdf

[24] Microsoft, “Data classification for cloud readiness,” pp. 1-22, 2014.

[25] Microsoft, “Trustworthy Computing,” pp. 1-5, January 2014.

About Authors:

ZUZANA PRIŠČÁKOVÁ

She graduated at Faculty of Natural Sciences UKF Nitra in 2012. From graduating she began studying doctoral studies at Faculty of Business and Economics MENDELU Brno in 2012. In 2014 she graduated her doctoral study in applied informatics and she obtained RNDr.
In 2008-2012 she worked as an associate scientific personnel in University UKF in Nitra. In 2010-2011 she worked as an assistant of IT project manager. Since 2013 she has worked as an academic staff (assistant) on the Department of informatics MENDELU in Brno.
In 2014 she taken part in internships at Faculty of Cybernetics Taras Shevchenko National University of Kyiv.
Her research is focus on data integrity, data security in cloud computing and IT project management.
In 2013-2014 she was the solver of project Deployment of open-source virtualization technology at MENDELU.
She is currently working on several projects with universities and firms as a data security solver and as a project manager.

IVANA RÁBOVÁ

She graduated at Electronic faculty VUT Brno in 1981. From graduating she always worked in areas of software applications and development of information systems and information technology. In 1990-1996 she worked as a leader of the information systems department in a large company in Brno.
Since 1997 she has worked as a university teacher on the Department of informatics PEF MZLU in Brno. She is the guarantee of followed subjects: Information technologies, Information systems, Management information systems, Information system projecting and modeling, Process management.
In 2002 she graduated her doctor study and she obtained Ph.D. The theme of dissertation thesis is The Business Metamodel in Information System Development. She is the member of research team.
In 2006 she was habilitated in Management informatics subject the name of her thesis is Enterprise Architecture; Analyse, Modeling and Value in Business Management.
She was the solver and co-solver of several projects of innovation education within the scope.