Zeszyty Naukowe AON nr 2 (103) 2016
ISSN 0867–2245

TOKENIZATION AS A DATA SECURITY TECHNIQUE

Maj. Florin OGIGAU-NEAMTIU
Regional Department of Defense Resources Management Studies
Brașov, Romania

Abstract

Nowadays, organisations, especially those which embrace the modern cloud computing technology, are facing the complex challenge of securing their data assets while providing modern IT services. Classical solutions for ensuring data security have consistent limitations in modern platforms due to factors like data sharing requirements, multi-tenancy, dynamic environment, high availability and reliability requirements, etc. One approach to address this problem is to implement encryption mechanisms which will provide the required security, but they depend on substantial investments in hardware and/or software and add supplementary complexity to those systems. This article analyses tokenization as an alternative strategy for ensuring data security in modern cloud computing systems by implementing mechanisms to hide sensitive data and restrict access to it. Tokenization systems are more easy to implement and can be used extensively for ensuring data security in IT systems.

Keywords: cloud computing, data security, encryption, obfuscation

Introduction

Since its appearance, cloud computing technology has lured modern organisations to implement it by promising multiple advantages (Salesforce 2016). However, the new technology raises many challenges starting from redesigning internal IT architecture and re-engineering internal business processes to accommodate the new technology and get maximum benefits from and identifying the required instruments to properly protect the companies' data assets. Specialists (Cloud Security Alliance 2016) conclude that data security is the highest hindrance of cloud acceptance and that losing data security is one of the main drawbacks which stops its spreading and user acceptance.

In order to address these issues, companies adopted different strategies depending on the business activity, nature of data and the cost of implementation. Each of these strategies has pros and cons and is best suited for certain environments.

Classical solutions for ensuring data protection based on a defence in depth approach and access control procedures have consistent limitations in public cloud computing systems due to the usage of this technology of untrusted systems for transmission, storage and data manipulation. The new strategy focus has to be on developing procedures and mechanisms for securing data in all of its forms: stored, processed or transferred, while considering each system to be unsecured.

Data obfuscation is a technique used to alter original data and to prevent unauthorised access to sensitive materials. It is one part of the organisation strategy which addresses the data security. There are multiple data obfuscation techniques but they can be grouped based on data alteration in 3 categories: encryption (all data is affected and the process is reversible), data masking (not all data is affected but the process is irreversible) and tokenization (not all data is affected and the process is reversible).

The main goal of this paper is to analyse and compare current data obfuscation techniques used in the industry, while evidencing the potential of the tokenization to be utilised in the wider area of data security. Nowadays, tokenization is used mainly in the payment card industry, but this paper considers the technology to be currently undervalued and, because of its advantages over other obfuscation techniques, it could be used extensively in other areas. The analysis goes beyond the pure technical capabilities comparison and presents the impact of these technologies upon an organisation. The analysis can be used by IT decision makers in their efforts to design and implement a viable data protection strategy for their organisation while trying to minimise IT costs in terms of resource consumption and disruption upon internal business processes.

Encryption as a data obfuscation technique

Data encryption is the process of conducting specific mathematical operations on an original data packet based on specific cryptographic algorithms in such a way that it cannot be deciphered without knowing the secret key. Modern encryption algorithms can be divided in two main categories (symmetric and asymmetric) and each of them are used in particular situations based on operational needs and constraints. The encryption/decryption process for symmetric and asymmetric encryption is presented in fig. 1, the difference between the two encryption processes being determined by the keys used. The symmetric encryption utilises the same key for encryption and decryption, while asymmetric encryption used different ones. In both cases, one of the most important aspects of encryption is that it provides a way of reconstructing the original message based on a key.

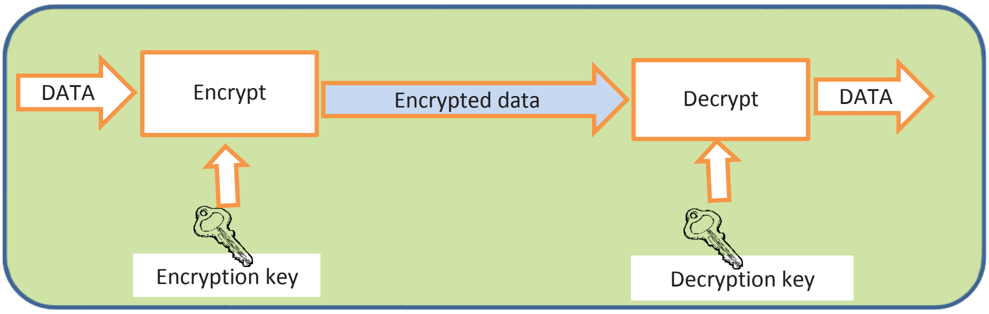

Fig. 1. Data encryption/decryption process

Most modern encryption systems are based on Kerckhoffs's principle, which states that a cryptosystem should be secure even if everything about the system, except the key is publicly known. So, in modern IT environments, keeping data secure actually means securing the keys used for data encryption and only sharing them with valid users. This process could be a very complex and difficult one, especially in untrusted shared working environments with high user mobility, as public cloud computing systems are.

Data masking as a data obfuscation technique

Data masking is a method of providing data security by creating a structurally similar but fake version of a section of the data packet that can be used for purposes such as software testing or user training (TechTarget 2016). Not all data is masked but the process parses the original data and applies masking techniques only when sensitive data is identified. The reason for conducting data masking processes is to provide de-identified, de-sensitised and anonymised data for application users, business intelligence, application testing, and outsourcing.

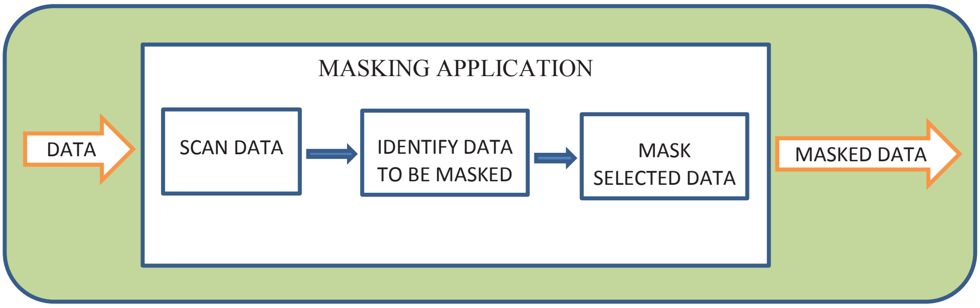

Fig. 2. The masking process

The process of data masking is presented in fig. 2. The process is simple and it starts with initial data scanning. Based on implemented policies, the system identifies sensitive data and performs alterations of it. Different masking techniques can be implemented, like substitution, shuffling, non-deterministic randomisation, blurring, nulling or deletion (TechTarget 2016) based on data processed and organisation needs. The data input can come from already stored data (static data masking) or for improved efficiency and increased security directly from application or processes (on-the-fly or dynamic data masking).

The most important characteristic of this technique is that data masking process is a one-way process. In other words, in masking methodology there is no reconstruction of original data from any intermediate data and there is no method for retrieving the original data from the masked one.

Data masking technology ensures the security of data by replacing sensitive information with a non-sensitive pattern and creates an output that can look and act as the original. This means that it can be used in business processes without imposing supplementary restrictions or procedures for applications or data storage capabilities.

Tokenization as a data obfuscation technique

Tokenization refers to a process by which a piece of sensitive data, such as a payment card number, is replaced by a surrogate value known as a token (Care & Litan 2016). The transformation of original data into tokens is done through one-way cryptographic function, which makes it impossible to reconstruct the original data without accessing the resources of the tokenization system.

The figure below is a generalised version of the tokenization process and is based on the PCI DSS Tokenization Guidelines (PCI Data Security Standards 2011).

The tokenization process:

- The application passes the data needed to be tokenised along with the authentication information to the tokenization system;
- The tokenization system checks the validity of authentication information. If authentication fails, then the process stops and the information is sent to the event collection system. This will allow administrators to identify issues and properly manage the system. If authentication is OK, then the system goes to the next step;
- The tokenization system generates, based on one-way cryptographic algorithms, the token for the data passed and both are stored in the highly secured data vault;
- The new token is passed to the application for further usage.

The critical point of this system and the most attractive target for hackers is the data vault where the actual sensitive data is stored. The vault needs to be protected with strong encryption capabilities and an enhanced key management system which will ensure that the sensitive data will be accessed only by authorised people and applications.

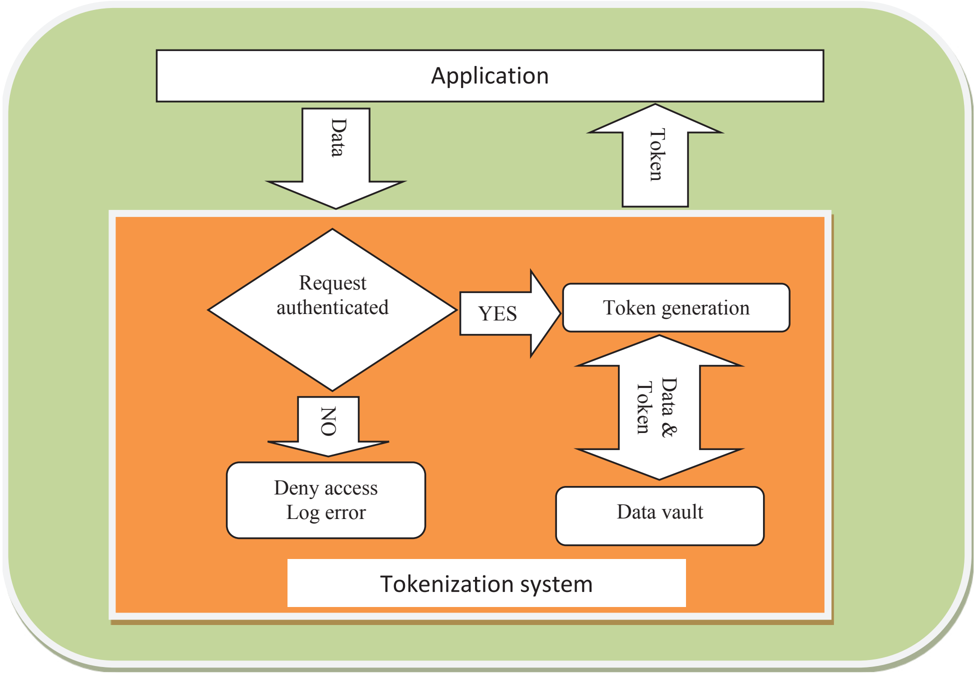

Fig. 3. The tokenization process

The development of a new vaultless generation of tokenization systems (Jonathan Care 2015) attempts to eliminate these limitations which will not be based on a central storage database and will eliminate the corresponding limitations.

Comparing obfuscation techniques

In order to provide a better understanding of the three obfuscation techniques and reveal their advantages and weaknesses, a comparative analysis had been conducted from three key perspectives: hardware and software requirements, impact upon organisation business processes and data security capabilities.

Hardware and software requirement

Different technologies have different prerequisites to support their implementation. Adopting encryption systems requires high investments in dedicated hardware and software solutions, as the encryption process is, in many cases, a consistent consumer of computing power. The systems need to be properly calibrated in order to support maximum workload capabilities and minimise the risk of failure in peak demand times.

The keys are critical to the security of data and, if compromised or lost, the protected data has no value. This poses a great administrative burden on designing, implementing and managing a key management system, especially in dynamic environments with frequent key operations (provision, deletion, renewal, archive, etc.) as most modern cloud IT organisations are. Minimising the key management risk can be achieved by outsourcing services and transferring the risk to third party entities (security as a service), but this increases the risk of losing data confidentiality.

Specialists have proposed and analysed different techniques to overcome these limitations which take into consideration the usage of third party key management solutions (Ogîgău-Neamțiu 2015), classic (Chunming Tang 2012) or quantum (Cloud Security Alliance 2015) key distribution and renewal processes, proxy re-encryption schemes (Poonkodi, Kavitha, Suresh 2013) or even cryptographic systems based on homomorphic encryption (Maha Tebaa 2013) which provide the ability of computers to directly process encrypted data.

In respect of the infrastructure requirements and maintenance costs, adopting data masking and tokenization technologies does not need complex systems or algorithms for conducting the corresponding process. The masking process totally eliminates the complex PKI infrastructure and Key Management Systems and requires limited computing resources and minimal administrative burden.

Tokenization requires limited IT resources; however, dedicated investments have to be committed in the area of token management and vault administration. The Key Management System is limited to the vault system and has minimal influence upon system complexity. One thing to mention here is that tokenization requirements grow exponentially with linear data growth, as increasing the vault databases will increase the amplitude and complexity of maintenance operations and will limit the system availability.

Impact upon business processes

In modern IT organisations, technology does not only have a support function but is also considered a business enabler capability. Using technology has a great impact upon organisation business processes and, based on the implementation model, can lower or improve organisation success rate. In most cases, implementing data security policies in organisations imposes limitations which have a bad impact upon the business processes and limit their efficiency.

From this perspective, encryption is the most invasive and limits process efficiency by adding high complexities to the systems and increasing the risk of system failure. It also requires redesigning internal business processes to accommodate the new workflow and retraining human resources in order to educate people on how to use the new systems and how to service them.

That is why, in many cases, organisations only implement encryption in production environments and leave the internal testing environments without proper protection. In order to provide a testing environment which mimics realistic conditions as many times as possible, companies import real data into the testing environments (Oracle 2013) without altering them and increasing the risks of compromising data.

The influence of masking technology upon business process is limited and is conducted by adding a data masking layer. The technology requires limited business processes changes and is greatly leveraged by companies for establishing a non-sensitive data information environment. The masking layer can be user transparent, and once established requires minimal administrative burden and minimal user training and it converts sensitive data into non-sensitive data in a very small amount of time.

Tokenization has a limited impact upon organisation business processes. The system needs to be integrated with other systems in the organisation and increases the complexity, especially in the area of token and vault management. Implementing tokenization introduces minimal time delays, requires application integration and the redesigning of some processes, imposes operating procedure updates and user training. Overall, adopting tokenization has a greater impact upon organisation processes than the adoption of the masking technology, but is considerably less when compared to the encryption one.

Data security capabilities

When analysing the data security capabilities, these technologies offer the CIA triad (confidentiality, integrity and availability; Panmore Institute 2016) complemented with authentication and nonrepudiation.

Encryption based systems ensure confidentiality by obscuring data and requiring a key for getting access to it. The management of the encryption keys is a very sensitive process and requires additional secure communication systems in symmetric encryption or complex and costly public key infrastructure (PKI) systems in an asymmetric one. Implementing and managing key management systems in increased shared environments with high user mobility and dynamics, as modern cloud IT environments are, is a very challenging task, introduces great system complexity and hinders data availability. Providing data integrity in encryption based systems is carried out by implementing cryptographic hash functions which are used to rapidly determine if the original message has been altered.

Encryption is used in modern systems to provide authentication capabilities by using secret passwords and digital signatures, while nonrepudiation is ensured by implementing complex PKI systems. Encryption is suitable to be used for all kinds of data in an organiation, while in transit or at rest and modern research (Maha Tebaa 2013) has found that encryption can even be used to secure data while processed.

These capabilities in some encryption based systems tend to grow unjustified because users do not classify data properly and use encryption even if it is not necessary.

Data confidentiality and integrity in masked based systems is achieved by substituting sensitive data with non-sensitive in a one-way process and by limiting the access to the sensitive data. Once converted, original data cannot be reassembled from masked data and information disclosure is avoided.

The integrity of masked data is not a huge problem but some requirements could be developed if integrated with other applications. The masking systems can be designed to generate masks based on wildcards that have instruments for integrity checking, like sum control. Masking based systems do not provide availability, authentication or nonrepudiation themselves and additional systems have to be added and integrated.

Data masking can be used for obscuring all kinds of data in an organization, but it needs prior identification and development of procedures on how it will be masked. Sensitive data needs to be at rest, as when considering data in transit or in process the masking process eliminates it, so there is no sensitive data transferred or processed. Data masking systems do not provide instruments for authentication of non-repudiation.

In tokenization based systems, confidentiality of data is ensured by replacing original sensitive data with non-sensitive data and restricting the access to the token vault. Access to that information is achieved by implementing authentication systems based on additional technologies, as tokenization does not provide such capabilities.

Storing all sensitive data in one location and limiting data access ensures the integrity of data. Data availability is easy to ensure in small tokenization systems, but it becomes a huge problem in large environments; however, the risk can be minimised by implementing consistent data classification procedures within the environment. Distributing token databases between different systems can increase availability, but introduces other problems like complex database back-up operations and continuous database synchronisation, which make this techniques efficient until a certain transaction threshold is reached.

Tokenization systems can be used to ensure data security while the data is at rest, but does not provide capabilities for secure data in transit or in process and this needs to rely on other systems. Tokenization does not provide instruments for authentication or non-repudiation.

The analysis above revealed that each obfuscation technique has its advantages and can be leveraged by companies in certain environments/applications to satisfy organisation needs:

Encryption:

- Can be applied to all types of data in an organisation;
- Can be used to ensure sensitive data transfer and even data processing in untrusted environments;
- Provides advanced data security capabilities like authentication and non-repudiation;

Masking:

- Creates a sanitised testing environment which resembles the real production one as closely as possible for the research and development teams to perform realistic testing, increase product capabilities and minimise failures;

- Provides realistic data to profile researchers in order to determine the effectiveness of the products, identify customer needs and satisfaction level;
- Shares sanitised customer monitoring data with other profile entities in order to provide customers quick access to desired resources or to advertise the proper products to them;
- Provides instruments to enhance collaborative work, while ensuring sensitive data is protected;
- Has a minimal influence upon the organisation business processes;
- Requires limited hardware and software resources and minimal administrative burden.

Tokenization:

- Requires limited hardware, software and maintenance costs;
- Has a minimal impact upon organisation business processes;
- Leverages outsourcing capabilities by limiting the risk of compromising data (even if the cloud provider storage environment has been breached, only tokenised data is accessible and not the sensitive data);
- Provides compliance with governmental regulations dealing with personal data manipulation (PCI DSS, HIPAA-HITECH, GLBA, ITAR, and the EU Data Protection Regulation);
- Provides compliance with regulations dealing with the location of personal data. In cloud computing systems, data is processed many times by globally distributed applications or saved for backup resiliency in multiple locations geographically scattered around the world. However, this is not always in accordance with local legislation that forbids processing/saving personal data outside country boundaries;
- Provides sanitised data for testing and developing environments;
- Improves collaboration capabilities by providing sanitised data to partners for further processing and analysis.

Tokenization potential

Today's most used domain for tokenization is the establishment of secure payment card data systems. The PCI Data Security Standard refers only to the tokenization of Primary Account Number (PAN) and imposes a set of requirements regarding the system and token format which is specific to the financial sector. The technology, however, is not limited to this data and it can be used for any kind of data: social security number, financial identification number, account numbers, names, email, phone number, assets and other sensitive data. Because of that, tokenization does not have to be limited to the financial sector only and other industry sectors can benefit from its advantages.

For security reasons, tokens are generated by using random number functions, but they can take certain formats, like the masking wildcards, to ensure proper application integration with other systems. In this way, tokenization can substitute masking and organisations will be able to provide an integrated secure environment with sanitised data for testing and development purposes. The system will have the capability to recover the original data through detokenization if required.

The tokenization systems has benefits over encryption by minimising investments, reducing complexity, minimising administrative burden, reducing influence upon business processes and enhancing collaboration.

Tokenization is not the perfect solution for providing data security in modern cloud computing systems and it comes with drawbacks such as:

- Tokenization is not suitable for all data in organisations and organisation data needs to be scanned and filtered;
- The vault is a database that maps sensitive information to tokens. If used on a great scale, the databases expand humongously determining a time increase in the search process, limiting system performance and increasing back-up procedures;
- The vault maintenance work increase exponentially with new data being added;
- Ensuring consistency across databases requires continuous synchronisation of token databases;
- Secure communication links have to be established between sensitive data and vault, so that data would not be compromised while being transferred to or from the storage;

Reports (Dimensional Research 2015) estimate that the most sensitive data in an organisation is credit cards and health records, passwords or other authentication credentials and personal employee information. Organisations should analyse their data, classify it based on its importance and then shrink their data security problem. Instead of implementing an overall complex security system based on encryption and masking, they should apply different policies to mitigate the data security problem with minimal cost expenditure and minimal influence over the business processes.

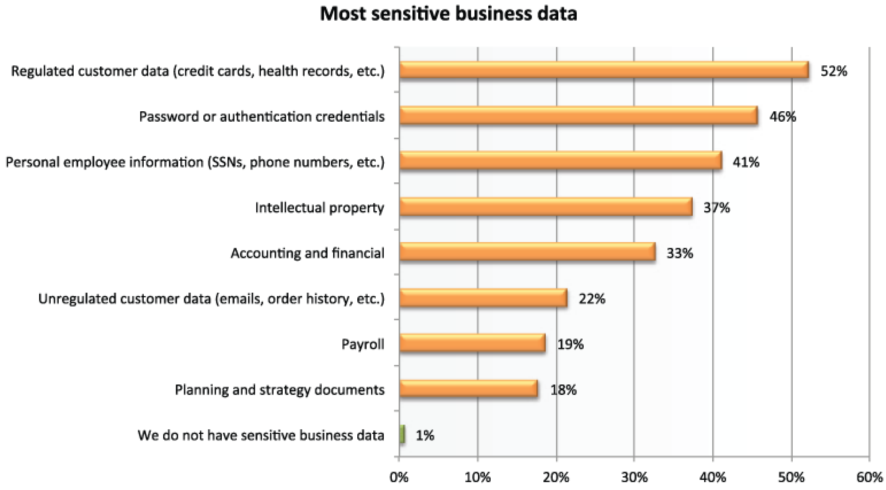

Fig. 4. Sensitive business data

Conclusion

Because of the great influence it has upon organisation's internal business processes, IT strategy and data security, a strategy needs to be developed upon a thorough analysis of the organisation requirements and estimated costs. Senior leaders have to decide how to approach this problem and address the risks of data compromising while maintaining operational status.

The most used data obfuscation techniques today are encryption and data masking, but they come with big disadvantages. The utilisation of high complexity procedures and systems for securing low level or non-sensitive data mainly generates unreasonable growth of system intricacy, interaction difficulties, hardships in data management (copy, move, processing, saving, backup, restoration, deletion), increases the system management effort and generates unjustified costs of human, financial and material resources.

The costs required for implementing data security strategies in terms of financial investments and limitations and constraints imposed on the organisation system have to be sustainable, correlated with the importance of the information being protected and in accordance with its mission.

The analysis conducted in this article reveals that tokenization has a huge potential and can be used in modern organisational environments to overcome encryption and masking limitations. Tokenization requires reduced IT resources, minimises the impact upon organisation business processes, limits performance impact and increases organisation collaborative capabilities. Implementing a tokenization framework in an organisation has to be done based on a consistent strategy plan. The decision makers have to identify challenges, analyse their impact upon their organisation and establish measures to address those challenges while ensuring business success.

References

Care, J., Litan, A. (2016), *Hype Cycle for Application Security*, Gartner.

Chunming Tang, X.H. (2012), *An efficient key distribution scheme in cloud computing*, „Cloud Computing Technology and Science, CloudCom.

Cloud Security Alliance (2015), *Quantum-Safe Security Working Group*, Cloud Security Alliance.

Cloud Security Alliance (2016), *The Treacherous 12 Cloud Computing Top Threats in 2016*, Cloud Security Alliance, https://downloads.cloudsecurityalliance.org/assets/research/top-threats/Treacherous-12_Cloud-Computing_Top-Threats.pdf.

Dimensional Research (2015), *The state of data privacy in 2015. A survey of IT professionals*.

Jonathan Care, R.K. (2015), *Market Guide for Merchant/Acquirer Tokenization of Payment Card Data*, Gartner, http://www.gartner.com/document/3177018?ref=solrAll&refval=173965545&qid=1adf8c9d58697bdfcaeba59872b301df.

Maha Tebaa, S.E. (2013), *Secure Cloud Computing through Homomorphic Encryption*, „International Journal of Advancements in Computing Technology", p. 29–38.

Ogîgău-Neamțiu, F. (2015), *Cryptographic Key Management In Cloud Computing*, „Defense Resources Management In The 21st Century", Brașov : National Defense University „Carol I" Publishing House, p. 225–229.

Oracle (2013), *Data Masking Best Practice*, Oracle.

Panmore Institute (2016), *The CIA triad*, http://panmore.com/: http://panmore.com/the-cia-triad-confidentiality-integrity-availability [access: 9.09.2016].

PCI Data Security Standards (2011), *PCI Data Security Standard (PCI DSS), v2.0.*, Tokenization Taskforce PCI Security Standards Council, https://www.pcisecuritystandards.org/documents/Tokenization_Guidelines_Info_Supplement.pdf.

Poonkodi, S., Kavitha, V., Suresh, K. (2013), *Providing a secure data forwarding in cloud storage system using threshold proxy re-encryption scheme*, „International Journal of Emerging Technology and Advanced Engineering", p. 468–472.

Salesforce (2016), *Why Move To The Cloud? 10 Benefits Of Cloud Computing*, https://www.salesforce.com: https://www.salesforce.com/uk/blog/2015/11/why-move-to-the-cloud-10-benefits-of-cloud-computing.html [access: 9.09.2016].

TechTarget (2016a), *Data masking best practices for protecting sensitive information*, http://searchfinancialsecurity.techtarget.com/tip/Data-masking-best-practices-for-protecting-sensitive-information [access: 10.09.2016].

TechTarget (2016b), *What is data masking*, http://searchsecurity.techtarget.com/definition/data-masking [access: 10.09.2016].