Copyright © 2019 Tech Science Press JCS, vol.1, no.1, pp.19-28, 2019

Research on Protecting Information Security Based on the Method of Hierarchical Classification in the Era of Big Data

Guangyong Yang[1, *], Mengke Yang[2, *], Shafaq Salam[3] and Jianqiu Zeng[4]

Abstract: Big data is becoming increasingly important because of the enormous information generation and storage in recent years. It has become a challenge to the data mining technique and management. Based on the characteristics of geometric explosion of information in the era of big data, this paper studies the possible approaches to balance the maximum value and privacy of information, and disposes the Nine-Cells information matrix, hierarchical classification. Furthermore, the paper uses the rough sets theory to proceed from the two dimensions of value and privacy, establishes information classification method, puts forward the countermeasures for information security. Taking spam messages for example, the massive spam messages can be classified, and then targeted hierarchical management strategy was put forward. This paper proposes personal Information index system, Information management platform and possible solutions to protect information security and utilize information value in the age of big data.

Keywords: Big data, hierarchical classification, rough sets, information index system.

1 Introduction

In the era of big data, the quantity of the information increases in geometric series, the risk and value attract much attention. Technically, the application of big data can be infinitely broad. Because of the lack of rational use of rules, the data used for commercial applications and public services will be far less than the theoretical data which can be collected and processed [Liu, Jia, Guo et al. (2014)]. In the long run, it will be harmful to the formation and development of big data industry. Especially in the construction of smart city, only revitalizing the stock of existing data and making full use of big data can we enhance the level of smarter cities and promote urban management from “experience management” to “scientific management”. With the gradual development of big data applications, how to make good use of big data to protect personal information security

[1] School of Economics & Management, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[2] School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[3] Beaconhouse School System, Peshawar, Pakistan.

[4] School of Economics & Management, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[*] Corresponding Authors: Yang Guangyong. Email: yangguangyong@bupt.edu.cn;
Yang Mengke. Email: yangmengke@139.com.

JCS. doi:10.32604/jcs.2019.05947 www.techscience.com/jcs

has become an important mission of the government and society. Therefore, the government and society should not only protect user privacy and personal information security, but also maximize the value of mining the information itself [Otani, Baba and Kashima (2016)].

In present, China still lacks management norms which contain reasonable opening and utilization of user’s data. The principles of user information protection and reasonable utilization are defined in Telecom and Internet users' personal information protection regulations which was promulgated in 2013 year. While there are no relative regulations about specific rules for the development and utilization of data. The problem of information security of data needs to be properly solved. Big data applications will inevitably bring the application and sharing of user data, and the multidimensional data interaction will mean more information disclosure risk. Once the protection of user information is ineffective or encounters information theft, it is bound to cause user panics, and pose a threat to social stability and national security.

The traditional privacy rules adopt “informing and permission” principle that people decide whether, how and who to deal with their information. This means that each individual citizen is responsible for the protection of personal privacy. While in the era of big data, "informing and permission" lacks feasibility in practice because of the secondary use of data. Therefore, scholars propose a strategy that change the traditional privacy protection system, transferring the responsibility of privacy protection from the individual citizen to the data user. This means that data users take responsibility for their actions, rather than whether they obtain the consent of the individual at the beginning of the data collection [Edwards and Rodriguez (2016)].

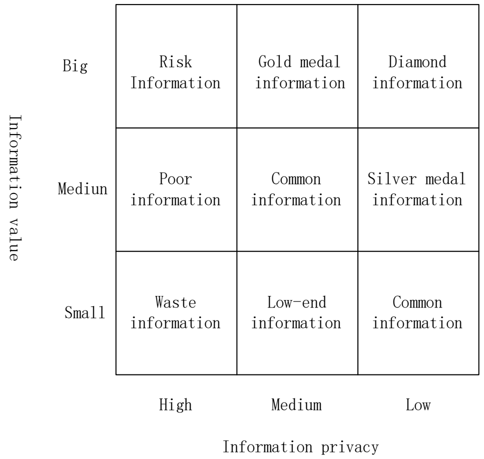
From the perspective of privacy security and protection cost, we should classify and rank data, and protect critical data according to the different needs. After entering the era of big data, personal information not only has more commercial value for Internet service providers, but also faces a greater security threat. How to protect the user's personal information in the era of big data has become a hot issue which has to be solved. It is concerning to network foundation development. Protecting the user's personal information must be based on the reality of the internet economics development. Using the hierarchical method to protect user information is an effective way to solve the problem of user personal information protection in the era of big data [Huang, Liu, Han et al. (2014)].

2 The hierarchical classification method in the era of big data

Hierarchical classification of information is beneficial to the rational development and utilization of information. And classified protection of information can avoid the imbalance brought by “one size fits all” [Zeng, Yang and Dong (2015); Bo, Wang, Liu et al. (2018)].

2.1 Nine-Cells information hierarchical classification

Nine-Cells matrix has two dimensions, one is the information value, and the other is information privacy. The value of information is the importance or the vital interests to the users, and information privacy refers to whether the information is related to personal privacy or public security information. Based on the value of information distinguishing as small, medium, large and the privacy of information distinguishing as high, medium, law, the matrix is divided into nine modules, corresponding waste information, law-end information, common information, poor information, silver medal information, risk information, gold medal information, diamond information respectively [Wang, Liu, Chao et al. (2018); Li, Yan, Han et al. (2014)].



Information value			
Big	Risk Information	Gold medal information	Diamond information
Mediun	Poor information	Common information	Silver medal information
Small	Waste information	Low-end information	Common information
	High	Medium	Low

Information privacy

Figure 1: Nine-Cells information matrix chart

Taking spam messages for example, there are huge amount of spam messages in the community. Some researchers propose that we can take corresponding strategies to manage the spam messages based on the social harmfulness and the spread degree. As shown in Fig. 2, spam messages hierarchical classification governance strategy is described.

Hierarchical classification protection mode needs to distinguish information content firstly, then gives different degrees of protection according to the value and security risks of all kinds of information. Meanwhile it puts forward different behavior requirements to service providers.

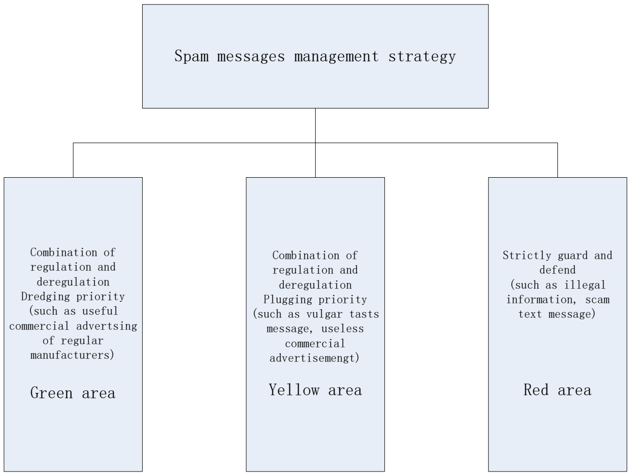

Spam messages management strategy
Combination of regulation and deregulation Dredging priority (such as useful commercial advertsing of regular manufacturers)
Green area
Combination of regulation and deregulation Plugging priority (such as vulgar tasts message, useless commercial advertisemengt)
Yellow area
Strictly guard and defend (such as illegal information, scam text message)
Red area

Figure 2: Spam messages hierarchical classification governance strategy chart

2.2 Information hierarchical classification based on Rough Set Theory

Assume that there are seven kinds of information that constitute set A , $A=\left\{ x1,x2,x3,x4,x5,x6,x7, \right\}$, Every kind has value attribute. In accordance with the different value, we can divide the information into $W1=\left\{ red,\ yellow,\ blue \right\}$ as three categories. For example,

Law value information constitutes a set $X1=\left\{ x1,x2,x5 \right\}$,

Medium value information constitutes a set $X2=\left\{ x3,\ x6 \right\}$,

High value information constitutes a set $X3=\left\{ x4,\ x7 \right\}$.

According to the Nine-cells information in Fig. 1, Information has another privacy attribute as $W2=\left\{ strong,\ medium\ and\ weak \right\}$ corresponding three sets：

$\left\{ x1,x7 \right\}$,$\left\{ x3,x4 \right\}$,$\left\{ x2,x5,x6 \right\}$. Adding the $W1$ attribute to set A:

$A/W1=\left\{ X1,X2,X3 \right\}=\left\{ \left\{ x1,x2,x5 \right\},\left\{ x3,x6 \right\},\left\{ x4,x7 \right\} \right\}$ (value classification)

$A/W2=\left\{ Y1,Y2,Y3 \right\}=\left\{ \left\{ x1,x7 \right\},\left\{ x3,x4 \right\},\left\{ x2,x5,x6 \right\} \right\}$ (privacy classification)

Through the intersection operation:

Law value & weak privacy $\left\{ x1,x2,x5 \right\}$ Ç $\left\{ x2,x5,x6 \right\}=\left\{ x2,x5 \right\}$

Law value & medium privacy $\left\{ x1,x2,x5 \right\}$ Ç $\left\{ x3,x4 \right\}=0$

Law value & high privacy $\left\{ x1,x2,x5 \right\}$ Ç $\left\{ x1,x7 \right\}=\left\{ x1 \right\}$

Medium value & weak privacy $\left\{ x3,x6 \right\}$ Ç $\left\{ x2,x5,x6 \right\}=\left\{ x6 \right\}$

Medium value & medium privacy $\left\{ x3,x6 \right\}$ Ç $\left\{ x3,x4 \right\}=\left\{ x3 \right\}$

Medium value & high privacy $\left\{ x3,x6 \right\}$ Ç $\left\{ x1,x7 \right\}=0$

High value & weak privacy $\left\{ x4,x7 \right\}$ Ç $\left\{ x2,x5,x6 \right\}=0$

High value & medium privacy $\left\{ x4,x7 \right\}$ Ç $\left\{ x3,x4 \right\}=\left\{ x4 \right\}$

High value & high privacy $\left\{ x4,x7 \right\}$ Ç $\left\{ x1,x7 \right\}=\left\{ x7 \right\}$

All of these concepts can be calculated by intersection and union constitute a knowledge system $W=W1$ Ç $W2$ together with $\left(A/W1,A/W2 \right)$.

Next, we observe the attributes of information and decisions making of treatment information from two-dimensional Tab. 1.

Table 1: Information attributes and decision

Information	Value	Privacy	Usability
$X1$	Law	Strong	No
$X2$	Law	Weak	Free use
$X3$	medium	medium	Strive for use
$X4$	High	medium	Rational use
$X5$	Law	Weak	Free use
$X6$	medium	Weak	Active use
$X7$	High	Strong	Use after processing

$X1 \sim X7$ is all kinds of information. The value and privacy are the condition attributes and decision attribute is the availability attribute. From the above table, we use each kind of information in different strategy which has manifested the idea of hierarchical classification of information [Luo, Li, Yi et al. (2016); Hong, Liou and Wang (2009); Leung, Fischer, Wu et al. (2008); Huang, Li and Wei (2012); He, Wu, Chen et al. (2011)].

3 Construction of personal information index system and classification management platform

In this part, we will propose two ways of personal information classification. The first is based on content. The second is based on the level of protection. While proposes personal information index system and information management platform respectively.

3.1 User personal information classification based on content

Personal information can be divided into privacy information, identity information, log information and public information based on the content. Identity information refers to information that can be used alone or in combination to identify a particular user's identity, mainly including identity authentication information (such as password), address book information, user basic information and virtual identity information. Log information refers to the information generated by the user using the Internet service, mainly including user consumption information, service ordering relations, terminal information, access information (such as IP address), location information and network behavior records (such as web shopping records, search content) [Zhu, Liu and Chen (2010); Zhu (2006); Lee, Seo and Choi (2002)].

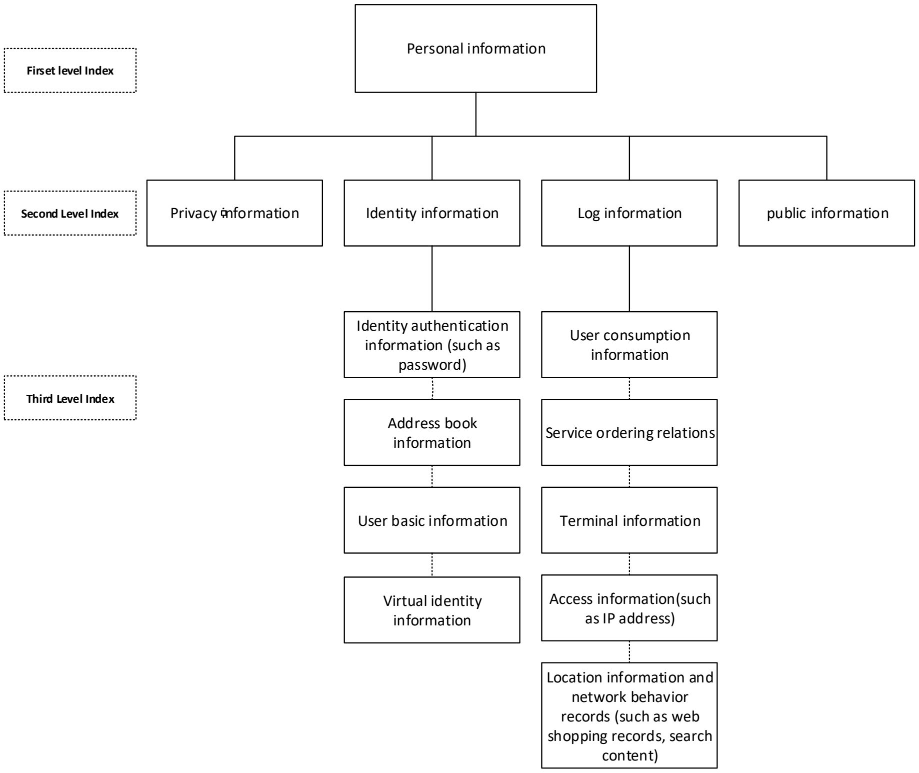

Firset level Index
Second Level Index
Third Level Index
Personal information
Privacy information
Identity information
Log information
public information
Identity authentication information (such as password)
Address book information
User basic information
Virtual identity information
User consumption information
Service ordering relations
Terminal information
Access information(such as IP address)
Location information and network behavior records (such as web shopping records, search content)

Figure 3: Personal information index system

3.2 User personal information classification based on the level of protection

On the basis of classification, personal information for all types of users can be classified based on protection level to match different management requirements. The level of protection is mainly divided into the following four factors: the first is whether a specific user can be identified directly based on information; the second is the affinity of offline life and information; the third is the ability to obtain other related information through the information; the forth is information security risk. The above four factors can classify the level of information protection, and the level of identity information protection is higher than that of log information. Among all kinds of identity information, the protection level of the identity authentication information is the highest, the address book secondly, the user basic information thirdly, and the virtual identity information is the lowest [Zeng, Cheng and Yang (2016)].

The protection degree is reflected in the behavior of the enterprise in the information flaw. Personal information flaws from the producer to the end-user and complete a life cycle, which contains five links, collection, storage, utilize, transfer, delete. The utilize refers to information collector use the information themselves; the transfer part refers to the information flaws from the collector to another, including the public or the specific object of public, information sharing between cooperative partner, commissioned by the processing situation, etc.

Based on above content, we need propose an Information management model and platform. It includes three elements in this platform. The first one is policy, law and standard, the second one is technology methods, the third one is security management system [Yang, Zhou, Zeng et al. (2016); Li, Yang, Chen et al. (2015)].

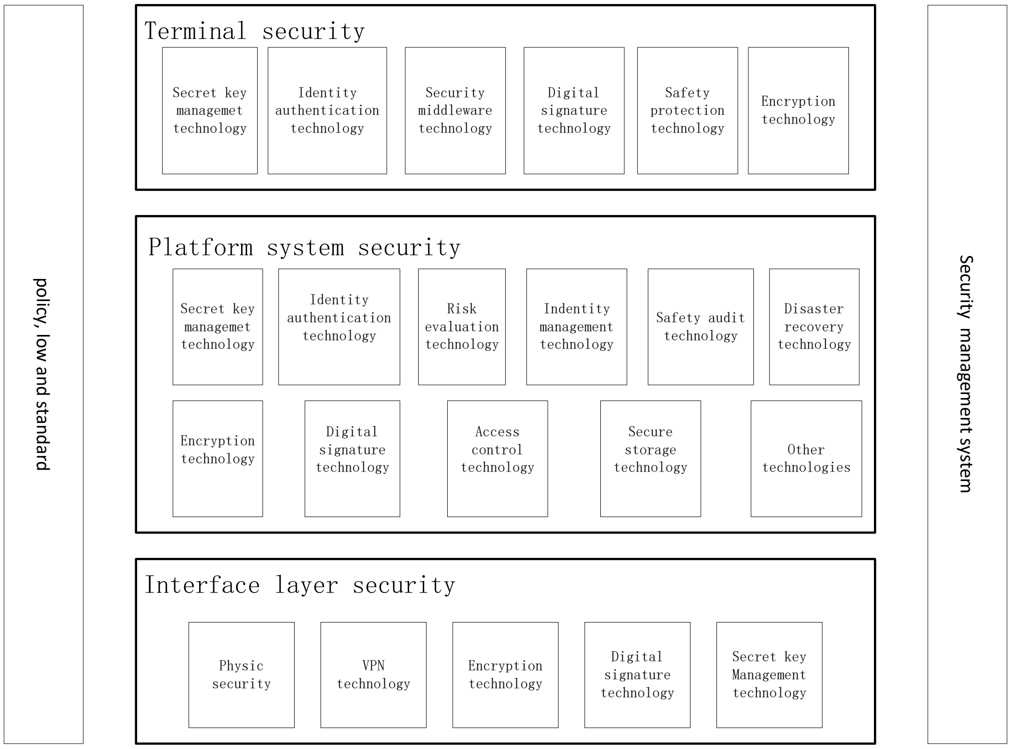

policy, low and standard
Terminal security
Secret key managemet technology
Identity authentication technology
Security middleware technology
Digital signature technology
Safety protection technology
Encryption technology
Platform system security
Secret key managemet technology
Identity authentication technology
Risk evaluation technology
Indentity management technology
Safety audit technology
Disaster recovery technology
Encryption technology
Digital signature technology
Access control technology
Secure storage technology
Other technologies
Interface layer security
Physic security
VPN technology
Encryption technology
Digital signature technology
Secret key Management technology
Security management system

Figure 4: Information management platform

4 Solutions to protect information security and utilize information value in the age of big data

The solutions to protect information security and utilize information value measures should be integrated and flexible. From the perspective of main bodies, they should include user, enterprise and government. From the perspective of method of view, they should include technologies, management, policy, law et al.

4.1 Make joint efforts to protect and utilize information among user, enterprise and government

In the above chapter we know personal information can be divided into privacy information, identity information, log information and public information. Considering the type and nature of the related data, comical secrets and personal privacy should be strictly protected. Except for the permission of the data owner, enterprise, organization or individual is prohibited to use, trade or disclose. We keep zero tolerant against disclosure, theft and illegal use of trade secrets and personal privacy information. And such an illegal act should be cracked down. However, business data and personal information, if cannot be processed to identify specific individuals and recovered or deleted identity, it can be used, exchanged and shared under certain conditions. For example, Guiyang Big Data Exchange has been developing rapidly because of the performance of protecting business secrets and personal privacy.

It does not mean that all the business data and personal information should be "one size fits all" type. Accordingly permit or ban all data absolutely. Contrarily, we should distinguish business data information which can be used and traded from business data which cannot be used and traded such as business secret, and personal privacy or sensitive information. We should classify based on the attributes of the relevant data information, the influence of the right of the data information, domain and category, then give the appropriate level of protection according to specific categories.

4.2 Taking measures including technologies, management, policy and law to protect and utilize information.

In the above chapter we propose information management platform. The platform includes safe management system, policy, law and standard and technologies guarantee. Generally speaking, technologies are the basic means, management is necessary support, the law is important guarantee. Security technologies contain three aspects of interface layer security, platform system security and terminal security. Security management needs users or organizations take scientific method like hierarchical classification strategy. The hierarchical classification strategy depends on administrators advanced concept, quality and technologies like data mining processing technology and sound legal system. Information protection law system will be built in the future. Firstly, it will improve the information protection system from the entities and procedures, and effectively protect the legitimate rights and interests of the main body; Secondly, it will promote the construction of the legal system of network management, strengthen the early warning of the risk of data information, reduce the risk of personal information leakage; Thirdly, it will establish an information security certification system, build the third party rating agencies, and according to the level of data information protection and evaluation, enterprises can take the corresponding technical and institutional measures to strengthen the protection of enterprise data information; Finally, we should encourage the development and research of information security enterprise, and suggest the government departments to encourage enterprises to strengthen the development of information protection technology and the research of the information protection system by means of tax and financial subsidies.

5 Conclusions

The paper proposes nine-cells information hierarchical classification chart based on recognition and importance of information. It uses rough set theory to establish information classification method. Taking spam messages for example, it classifies mass of spam messages and proposes targeted hierarchical classification governance strategies. Meanwhile the paper proposes the information fuzzy characteristic processing to strengthen the secondary uses of data.

The paper builds the personal information index system and classification management platform, puts forward some solutions to protect information security and utilizes information value in the era of big data. Basic prerequisites have been well provided for mining the potential value of the user's personal information with the rapid development of technology. It is possible to gather a large scale of information because of acquisition and storage cost reduction. Data mining and data analysis technology provide opportunities and conditions for the secondary development of user's personal information. The potential value of information is released.

Acknowledgement: This work was supported by Research on the Guidance Effect of Graduate Tutor Based on Big Data Resource Pool, Key Project of Graduate Education Reform and Research of Beijing University of Posts and Telecommunications (No. 2018Y0403).

References

Edwards, J. S.; Rodriguez, E. (2016): Using knowledge management to give context to analytics and big data and reduce strategic risk. *Procedia Computer Science*, no. 99, pp. 36-49.

He, Q.; Wu, C.; Chen, D.; Zhao, S. (2011): Fuzzy rough set based attribute reduction for information systems with fuzzy decisions. *Knowledge-Based Systems*, vol. 24, no. 5, pp. 689-696.

Hong, T. P.; Liou, Y. L.; Wang, S. L. (2009): Fuzzy rough sets with hierarchical quantitative attributes. *Expert Systems with Applications*, vol. 36, no. 3, pp. 6790-6799.

Huang, B.; Li, H.; Wei, D. (2012): Dominace-based rough set model in intuitionistic fuzzy information systems. *Knowledge-Based Systems*, vol. 28, pp. 115-123.

Huang, X.; Liu, J.; Han, Z.; Yang, J. (2014): A new anonymity model for privacy-preserving data publishing. *China Communications*, vol. 11, no. 9, pp. 47-59.

Lee, C. H.; Seo, S. H.; Choi, S. C. (2002): Rule discovery using hierarchical classification structure with rough sets. *IFSA World Congress & NAFIPS International Conference.*

Leung, Y.; Fischer, M. M.; Wu, W. Z.; Mi, J. S. (2008): A rough set approach for the discovery of classification rules in interval-valued information systems. *International Journal of Approximate Reasoning*, vol. 47, no. 2, pp. 233-246.

Li, H. E.; Yan, J.; Han, W.; Ding, Z. (2014): Mining user interest in microblogs with a user-topic model. *China Communications,* vol. 11, no. 8, 131-144.

Li, Q.; Yang, S.; Chen, J.; Liu, Z. (2015): Design and implementation of multi-network integration information security gateway based on emergency communication. *Mobile Communications*, no. 8, pp. 65-68.

Liu, Y.; Jia, H. E.; Guo, M.; Yang, Q.; Zhang, X. (2014): An overview of big data industry in china. *China Communications*, vol. 11, no. 12, pp. 1-10.

Luo, C.; Li, T.; Yi, Z.; Fujita, H. (2016): Matrix approach to decision-theoretic rough sets for evolving data. *Knowledge-Based Systems*, vol. 99, pp. 123-134.

Otani, N.; Baba, Y.; Kashima, H. (2016): *Quality Control of Crowdsourced Classification Using Hierarchical Class Structures*. Pergamon Press, Inc.

Wang, B.; Liu, P.; Chao, Z.; Wang, J.; Chen, W. et al. (2018): Research on hybrid model of garlic short-term price forecasting based on big data. *Computers, Materials & Continua*, vol. 57 no. 2, pp. 283-296.

Yang, M.; Zhou, X.; Zeng, J.; Xu, J. (2016): Challenges and solutions of information security issues in the age of big data. *China Communications*, vol. 13, no. 3, pp. 193-202.

Zeng, J.; Cheng, G.; Yang, M. (2016): Research on information security risk management system of telecom operators. *Science & Technology Management Research*.

Zeng, J.; Yang, G.; Dong, H. (2015): Spam SMS classification governance strategies. *Journal of Beijing University of Posts and Telecommunications in China (Social Sciences Edition)*, no. 12, pp. 39-44.

Zhu J. (2006): Research on evaluation method and index system of information security management. *Science & Technology Economy Market*, no. 10, pp. 139-152.

Zhu, Y.; Liu H.; Chen C. (2010): Research of measuring information security management effectiveness. *Journal of Intelligence*, vol. 41, no. 1, pp. 73-76.