ANALYSIS OF CORE DOCUMENTS IN INFORMATION SECURITY BASED ON MAPPING KNOWLEDGE DOMAINS

Hong-zhou Shen, Qin-jian Yuan
Department of Information Management, Nanjing University, No.22 Hankou Road, Nanjing, China

Qian-jin Zong, Ling-yu Tong
Department of Information Management, Nanjing University, No.22 Hankou Road, Nanjing, China

Keywords: Information Security, Knowledge Mapping Domains, Core Documents, Evolution.

Abstract: It has been gradually realized that the research of Information Security should beyond the only perspective of technology, but from a perspective of inter-discipline. However, researchers belonging to a certain discipline often seem to have a poor awareness of the contributions made by researchers in other disciplines. In order to help researchers fully understand the contributions made by researchers in other disciplines, this paper, with the method of Mapping Knowledge Domains, identifies four clusters of core documents that get more attention and analyze the contents of core documents in each cluster. Finally, there is also an analysis on evolution of these core documents in this paper.

1 INTRODUCTION

Information Security means "protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, perusal, inspection, recording or destruction" (Wikipedia, 2011). Because of the rapid growth and widespread use of electronic data processing and electronic business conducted through the Internet, along with numerous occurrences of international terrorism, the Information Security has become more and more important. As a booming research area, Information Security is now the main subject in many journals and a large number of annual conferences and workshops, and it has also been gradually realized that the research of Information Security should beyond the only perspective of technology, but from a perspective of inter-discipline. Lots of contributions by researchers from different disciplines, such as computer science, mathematics, public policy, economics and management etc., have been done to the research of Information Security. However, these contributors usually have very different views on the key problems in Information Security and their respective solutions. Furthermore, we find that scholars belonging to a certain discipline often seem to have a poor awareness of the contributions made by researchers in other disciplines. In order to help the researchers easily find out the core documents and the relationships among these core documents so as to get a clearer understanding to the contributions made by researchers in other disciplines, this study, based on Mapping Knowledge Domains, is attempting to visually identify the clusters of related core documents and to give brief introduction of the related core documents in each cluster.

This paper is organized as follows. The second section presents the methodology and data we use in our work. In the third section, the clusters of related core documents were found and analyzed. Finally, the paper is concluded with a summary of our work.

2 METHODOLOGY AND DATA

2.1 Methodology

Based on the ideas from bibliometrics, social network analysis and information visualization, Mapping Knowledge Domains is aimed at easing information access, making evident the structure of knowledge, and allowing seekers of knowledge to succeed in their endeavours, through the process of

Shen H., Yuan Q., Zong Q. and Tong L..
ANALYSIS OF CORE DOCUMENTS IN INFORMATION SECURITY BASED ON MAPPING KNOWLEDGE DOMAINS.
DOI: 10.5220/0003576704210427
In *Proceedings of the 13th International Conference on Enterprise Information Systems* (EIT-2011), pages 421-427
ISBN: 978-989-8425-55-3

charting, mining, analyzing, sorting, enabling navigation of, and displaying knowledge (Shiffrin and Börner, 2004). Mapping Knowledge Domains can be used to identify the core documents in a certain area with the citation analysis and the corresponding cluster analysis. It can also be used to indicate the key topics drawing most attention in a research area by keyword analysis. Due to the dramatic increases in computational storage capacity and processing speed, there are many new techniques of analysis, retrieval and visualization that make it possible to deal with massive information. Various analysis tools are available for free, such as Bibexcel, Netdraw, Pajek, HistCite and so on. In this study, Citespace Ⅱ by Chaomei Chen is employed, and we will use the Co-citation Analysis function of the software.

Citation Analysis is "the examination of the frequency, patterns and graphs of citations in articles and books" (Wikipedia, 2011). As one of the most widely used methods of bibliometrics, it uses citations in scholarly works to establish links to other works or other researchers. Co-citation Analysis, an important aspect of Citation Analysis, represents how often two bibliographic items are cited together, for example, papers in document co-citation networks (Small, 1973). If two papers(A and B) are co-cited by one or more papers, A and B have a co-citation relationship. It is believed that the more the two papers are co-cited, the closer the co-citation relationship will be, which means the two papers may have the same research topics.

2.2 Data

In this study, we collected the data of the documents from the SCI-EXPANDED and SSCI databases in ISI Web of Science. We only retrieve the "Article" type of documents with the topic of "Information Security" and we only concerned about the documents published between "1981 and 2010". Finally, we got 740 records and the results were obtained on 27th Feb. 2011.

3 RESULTS AND ANALYSIS

3.1 Content Analysis of Core Documents

In the Citespace Ⅱ, Time Slicing is set as 3 years per slice so as to divide the 30 years (from 1981 to 2010) into 10 slices, which will be used to analyze the evolution of core documents in the next section. We choose "Cited Reference" as the node type in the Co-citation Networks, so the link between two nodes indicates how often the two references are cited together. Meanwhile, (c, cc, ccv) = (2, 2, 10; 2, 3, 20; 3, 3, 20) are set as the threshold value in the software.

Table 1: The top 10 cited references.

Freq	Author	Title	Year
35	Gordon L. A. and Loeb M. P.	The Economics of Information Security Investment	2002
32	Straub, D. W. and Welke, R. J.	Coping With Systems Risk: Security Planning Models for Management Decision Making	1998
24	Refregier, P. and Javidi, B.	Optical image encryption based on input plane and Fourier plane random encoding	1995
22	Rivest, R. L. et al.	A Method for Obtaining Digital Signatures and Public-Key Cryptosystems	1978
20	Straub, D. W.	Effective IS Security: An Empirical Study	1990
19	Diffie, W. and Hellman, M. E.	New Directions in Cryptography	1976
18	Campbell, K. et al.	The economic cost of publicly announced information security breaches: empirical evidence from the stock market	2003
16	Matoba, O. and Javidi, B.	Encrypted optical memory system using three-dimensional keys in the Fresnel domain	1999
14	Stanton, J. M. et al.	Analysis of end user security behaviours	2005
14	Javidi, B. and Nomura, T.	Securing information by use of digital holography	2000
14	Cavusoglu, H. et al.	A Model for Evaluating IT Security Investments	2004

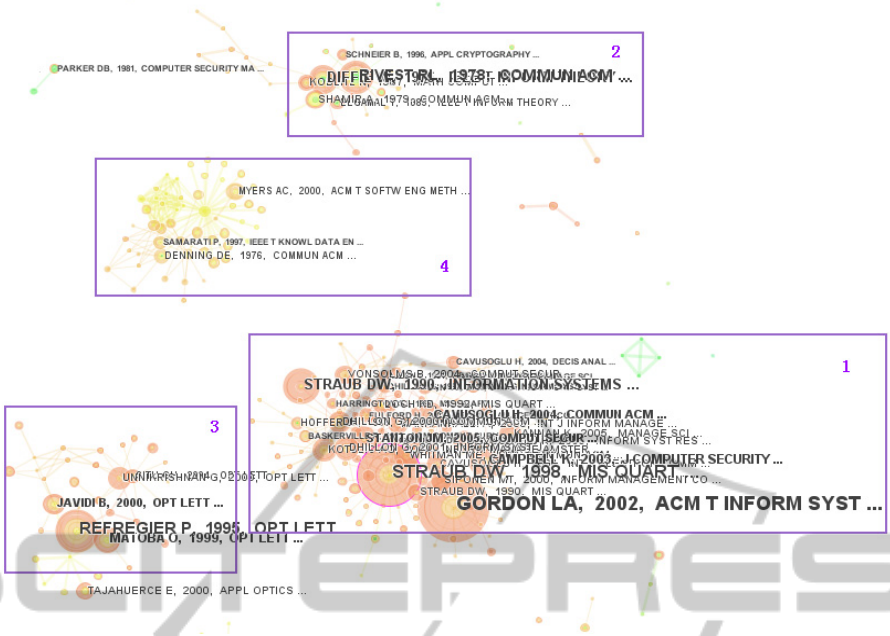

Figure 1: Clusters in Co-citation Networks.

Then, we make the tool to handle the data and documents co-citation networks are generated. We can get a list of cited references, and top 10 cited references are given in Table 1. The more frequent a document is cited, the more important the document will be, which means the document is a core document in Information Security. Finally, we use the "Find Optimal Clusters" function to find a partition so that documents within a cluster would be significantly more co-cited than documents from different clusters, which will help us to easily analyze the related documents and further to accurately understand the contents of these core documents.

Figure 1 is the final Co-citation Networks, from which we can identify 4 biggest clusters of co-cited documents intuitively. We will analyze the contents of the core documents and summarize the main research topics in each cluster.

3.1.1 Cluster1: Economics and Management of Information Security

From the contents of the papers in cluster1, it can be clearly noticed that economics and management of Information Security are the main research topics.

As the Information Security needs high and sustained investments, the benefits of these investments are certainly the greatest concern of managers. Gordon and Loeb (2002) presented an economic model that determined the optimal amount to invest to protect a given set of information, and it was shown that for a given potential loss, a firm should not necessarily focus its investments on information sets with the highest vulnerability, but be better off concentrating its efforts on information sets with midrange vulnerabilities. As tools such as risk analysis and cost effectiveness analysis have to work with very high-level aggregate data, so these tools are of limited value in evaluating of IT security investment. Cavusoglu, Mishra and Raghunathan (2004) proposed a comprehensive model to analyze IT security investment problems and they used the Game Theory to overcome some of the limitations. Losses caused by information security breaches lie not only in the breach itself, such as information disclosure, but also in the stock market reaction, especially when the breaches of traded corporations are publicly announced. Campbell, Gordon, Loeb and Zhou (2003) found that a highly significant negative market reaction for information security breach involving unauthorized access to confidential data, but no significant reaction when the breach does not involve confidential information.

In order to protect the information sets effectively, managers should pay more attention to kinds of management countermeasures in an organizational level. Straub and Welke (1998) proposed 3 management actions for managers to deal with the system risk, and the viability of the approach was validated through qualitative studies in two Fortune 500 information services firms.

Specifically, establishing clear information security policies and improving user's awareness of information security are the most important two countermeasures. Straub (1990) indicated that organizations that articulate their policy on computer abuse and actively enforce this policy should benefit from these activities, and security measures such as computer security awareness training sessions also reduced losses from abuse. Fulford and Doherty (2003) studied the uptake, content, dissemination and impact of information security policies within UK-based organizations, and they found that, while policies were then fairly common, at least amongst the sample, there was still a high degree of variety in terms of their content and dissemination. Actually, the effectiveness of the policy does not so much rely on the right content, but rather the way in which the content is addressed in the document and ultimately communicated to the users (Höne and Eloff, 2002). Sometimes, the failures of information security are caused by the unconscious behavior of users, so, training users and improving their security awareness are also the key steps. Stanton, Stam, Mastrangelo and Jolton (2005) developed taxonomy of end user security-related behaviors, tested the consistency of that taxonomy, and used behaviors from that taxonomy to conduct a U.S. survey of an important set of end user behaviors. Their U.S. survey of non-malicious, low technical knowledge behaviors related to password creation and sharing showed that password ''hygiene'' was generally poor but varied substantially across different organization type. Further, they documented evidence that good password hygiene was related to training, awareness, monitoring, and motivation. Obviously, in addition to confidentiality, integrity, and availability, the responsibility, integrity, trust and ethicality principles hold the key for successfully managing information security in the new millennium (Dhillon and Backhouse, 2000).

3.1.2 Cluster2: Cryptography

The cluster2 is mainly about the discussion of Cryptography. Several classic papers in Cryptography can be found in the cluster2. The paper from Diffie and Hellman (1976) was a key paper in this research area. It proposed two approaches to transmitting keying information over public (i.e., insecure) channels without compromising the security of the system. It also discussed the problem of providing a true, digital, message dependent signature. Finally, the paper considered the interrelation of various cryptographic problems and introduced the even more difficult problem of trap doors. Most importantly, Diffie and Hellman (1976) invented the concept of "public key cryptosystem" for the first time in this paper. Rivest, Shamir and Adleman (1978) then was motivated by the concept and presented an implementation of the "public key cryptosystem", which was named RSA, the most famous algorithm for public key cryptosystem. About 10 years later, because the analog of the discrete logarithm problem on elliptic curves is likely to be harder than the classical discrete logarithm problem, Koblitz (1987) introduced a more secure public key cryptosystem named Elliptic Curve Cryptosystems. In another paper that can be found from the cluster2, Shamir (1979) gave a technique that enabled the construction of robust key management schemes for cryptographic systems that could function securely and reliably even when misfortunes destroy half the pieces and security breaches expose all but one of the remaining pieces.

3.1.3 Cluster3: Optical Encryption and Decryption

Optical encryption and decryption is the main topic in cluster3, and several influential articles can be found in this cluster. Refregier and Javidi (1995) proposed a new optical encoding method of images for security applications. The encoded image was obtained by random-phase encoding in both the input and the Fourier planes. They also analyzed the statistical properties of this technique and showed that the encoding converted the input signal to stationary white noise and that the reconstruction method was robust. An encrypted optical memory system using double random phase codes in the Fresnel domain was proposed by Matoba and Javidi (1999). In the system, two random phase codes and their positions formed three-dimensional keys for encryption of images and were used as keys to recover the original data. In another paper, Javidi and Nomura (2000) proposed a security system that combined double-random phase encryption with a digital holographic technique. The proposed system enables us to store, transmit, and decrypt the encrypted data digitally. Unnikrishnan, Joseph and Singh (2000) proposed an optical architecture that encoded a primary image to stationary white noise by using two statistically independent random phase codes. The encoding was done in the fractional Fourier domain. In the paper of Situ and Zhang (2004), a lensless optical security system based on double random-phase encoding in the Fresnel

domain was proposed. This technique can encrypt a primary image to random noise by use of two statistically independent random-phase masks in the input and transform planes, respectively. In this system the positions of the significant planes and the operation wavelength, as well as the phase codes, are used as keys to encrypt and recover the primary image. Therefore higher security is achieved.

3.1.4 Cluster4: Information Flow Control

The cluster4 is mainly about the information flow control. Denning (1976) investigated mechanisms that guarantee secure information flow in a computer system. These mechanisms were examined within a mathematical framework suitable for formulating the requirements of secure information flow among security classes. The results showed that suitable constraints did indeed exist, and moreover within the context of a richly structured model. Role-based access control (RBAC) is the most famous approach to restricting information access to authorized users. It was thought that although the recognized usefulness of the RBAC concept, there was little agreement on what RBAC means. As a result, RBAC was an amorphous concept interpreted in different ways by various researchers and system developers, ranging from simple to elaborate and sophisticated. Sandhu, Coyne, Feinstein and Youman (1996) described a novel framework of four reference models developed to provide a systematic approach to understanding RBAC, and to categorizing its implementation in different systems. Their framework also separated the administration of RBAC from its use for controlling access to data and other resources. Actually, although RBAC models had received broad support as a generalized approach to access control, and were well recognized for their many advantages in performing large-scale authorization management, no single authoritative definition of RBAC existed before the appearance of the paper of Ferraiolo, Sandhu, Gavrila, Kuhn and Chandramouli (2001). This lack of a widely accepted model resulted in uncertainty and confusion about RBAC's utility and meaning. The standard proposed in this paper sought to resolve the situation by unifying ideas from a base of frequently referenced RBAC models, commercial products, and research prototypes. It was intended to serve as a foundation for product development, evaluation, and procurement specification. Except for the RBAC models, there are some other ideas proposed to control the information flow. Samarati, Bertino, Ciampichetti and Jajodia (1997) described a high assurance discretionary access control model for object-oriented systems. The model not only ensured protection against Trojan horses leaking information, but provided the flexibility of discretionary access control at the same time. The basic idea of their approach was to check all information flows among objects in the system in order to block possible illegal flows. Purpose-oriented access rules indicate what operation in each object can invoke operations of other objects. Information flow among the objects occurs if the requests and responses of the operations carry data. Only the purpose-oriented access rules which imply legal information flow are allowed. In the paper from Yasuda, Tachikawa and Takizawa (1998), they discussed how to specify the access rules so that the information flow occurring in the nested invocation of the operations was legal. The article of Myers and Liskov (2000) described the decentralized label model, a new label model for control of information flow in systems with mutual distrust and decentralized authority. The model improved on existing multilevel security models by allowing users to declassify information in a decentralized way, and by improving support for fine-grained data sharing. It supported static program analysis of information flow, so that programs could be certified to permit only acceptable information flows, while largely avoiding the overhead of run-time checking.

3.2 Evolution of Core Documents

We can rearrange the nodes of the Co-citation Networks in Timezone view (see Figure2), from which we can analyze the evolution of the core documents over time which we mentioned above.

It could be clearly noticed that the study of Cryptography was the main research topic in the core documents around year of 1981. About 10 years later, some core documents appeared to discuss the research of optical encryption and decryption, information flow control, and economics and management of Information Security. However, their paces of development are not consistent. The quantity of the documents on research of optical encryption and decryption and information flow control reached a peak around the year of 2000. However, with the participation of researchers from Economics and Management, the research of economics and management of Information Security underwent a sharp increase from 2000 and there were many classic documents appeared around 2005.

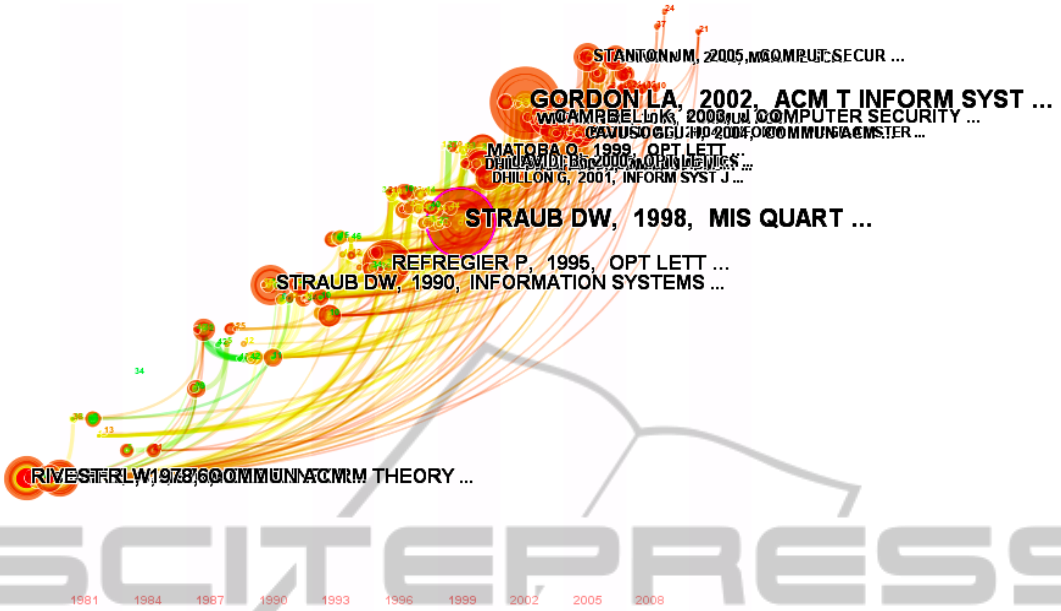

Figure 2: The Timezone view of Co-citation Networks.

Actually, in 2001, this trend was already identified by Dhillon and Backhouse (2001), who thought the research of information system security would move away from a narrow technical viewpoint to a socio-organizational perspective. This can also explain why the classic book "Security in Computing, 4th Edition" goes beyond technology, covering crucial management issues faced in protecting infrastructure and information, and contains an all-new chapter on the economics of cybersecurity, explaining ways to make a business case for security investments.

4 CONCLUSIONS

This paper presents a study of the core documents of Information Security, with the method of Mapping Knowledge Domains. We visually identify four clusters of core documents that get more attention and analyze the contents of core documents in each cluster, from which we summarize the main research topics in each cluster. They are economics and management of Information Security, Cryptography, optical encryption and decryption, and information flow control. Besides, there is also an analysis on the evolution of core documents, which illustrates that the core documents focus on the Cryptography initially, and finally pay more attention to the economics and management of Information Security.

ACKNOWLEDGEMENTS

This research is included in the project "The Study of Network Communication Model of Scientific Papers based on Academic Blog" supported by Ministry of Education. And the number of the project is 20100091110106.

REFERENCES

Campbell, K., Gordon, L. A., Loeb, M. P. and Zhou, L. (2003). The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security*, 11, 431-448.

Cavusoglu, H., Mishra, B. and Raghunathan, S. (2004). A Model for Evaluating IT Security Investments. *Communications of the ACM*, 47(7), 87-92.

Denning D. E. (1976). A Lattice Model of Secure Information Flow. *Communications of the ACM*, 19(5), 236-243.

Dhillon, G. and Backhouse, J. (2001). Current directions in IS security research: towards socio-organizational perspectives. *Information systems journal*, 11(2), 127-154.

Dhillon, G. and Backhouse, J. (2000). Information System Security Management in the New Millennium. *Communications of the ACM*, 43(7), 125-128.

Diffie, W. and Hellman, M. E. (1976). New Directions in Cryptography. *IEEE Transactions on Information Theory*, 12(6), 644-654.

Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R. and Chandramouli, R. (2001). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, 4(3), 224-274.

Fulford, H. and Doherty, N. F. (2003). The application of information security policies in large UK-based organizations: an exploratory investigation. *Information Management & Computer Security*, 11(3), 106-114.

Gordon L. A. and Loeb M. P. (2002). The Economics of Information Security Investment. *ACM Transactions on Information and System Security*, 5(4), 438–457.

Höne, K. and Eloff J. H. P. (2002). What Makes an Effective Information Security Policy. *Network Security*, 6, 14-16.

Javidi, B. and Nomura, T. (2000). Securing information by use of digital holography. *Optics Letters*, 25(1), 28-30.

Koblitz, N. (1987). Elliptic Curve Cryptosystems. *Mathematics of Computation*, 48(177), 203-209.

Matoba, O. and Javidi, B. (1999). Encrypted optical memory system using three-dimensional keys in the Fresnel domain. *Optics Letters*, 24(11), 762-764.

Myers, A. C. and Liskov, B. (2000). Protecting privacy using the decentralized label model. *ACM Transactions on Software Engineering and Methodology*, 9(4), 410–442.

Refregier, P. and Javidi, B. (1995). Optical image encryption based on input plane and Fourier plane random encoding. *Optics Letters*, 20(7), 767-769.

Rivest, R. L., Shamir, A. and Adleman, L. (1978). A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21(2), 120-126.

Samarati, P., Bertino, E., Ciampichetti, A. and Jajodia, S. (1997). Information Flow Control in Object-Oriented Systems. *IEEE Transactions on Knowledge and Data Engineering*, 9(4), 524-538.

Sandhu, R. S., Coynek, E. J., Feinsteink, H. L. and Youmank, C. E. (1996). Role-Based Access Control Models. *IEEE Computer*, 29(2), 38-47.

Shamir, A. (1979). How to Share a Secret. *Communications of the ACM*, 22(11), 612-613.

Shiffrin R. M., Borner K. (2004). Mapping knowledge domains. *Proceedings of the National Academy of Sciences of the USA*, 101(suppl.1), 5183-5185.

Situ,G. and Zhang, J. (2004). Double random-phase encoding in the Fresnel domain. *Optics Letters*, 29(14), 1584-1586.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265-269.

Stanton, J. M., Stam, K. R., Mastrangelo, P. and Jolton, J. (2005). Analysis of end user security behaviours. *Computers & Security*, 24, 124-133.

Straub, D. W. (1990). Effective IS Security: An Empirical Study. *Information System Research*, 1(3), 255-276.

Straub, D. W. and Welke, R. J. (1998). Coping With Systems Risk: Security Planning Models for Management Decision Making. *MIS Quarterly*, 22(4), 441-469.

Unnikrishnan, G., Joseph, J. and Singh, K. (2000). Optical encryption by double-random phase encoding in the fractional Fourier domain. *Optics Letters*, 25(12), 887-889.

Wikipedia. *Information Security*. Retrieved March 24, 2011, from http://en.wikipedia.org/wiki/Information_Security

Wikipedia. *Citation Analysis*. Retrieved March 24, 2011, from http://en.wikipedia.org/wiki/Citation_analysis

Yasuda, M., Tachikawa, T. and Takizawa, M. (1998). A purpose-oriented access control model. *Proceedings of the 13th International Conference on Information Networking*, 168-173.