CiteSpace Comparison of Citations in the fields of Information Security and Information Assurance

Joe Raczkowski
Drexel University
Joeraz5@verizon.net

Bander Alsulami
Drexel University
bander@alsulamy.com

ABSTRACT

The fields of Information Security and Information Assurance appear to be very similar with some overlap. In order to determine the overlap between these two fields, we have constructed CiteSpace visualizations for each of them, and then analyzed them, and tried to find their overlap. Upon closer analysis, there appears to be a well-defined, and very clear, difference between these two fields and there is actually very little overlap.

Keywords: Information Security, Information Assurance, visualization.

Index Terms: Comparison, CiteSpace, Security, Assurance, visualization.

1 INTRODUCTION

In today's world, Information Security is an increasingly important field due to our increasing dependence on, and numerous threats to the confidentiality and safety of information, both internal and external. However, related to the subject of Information Security is Information Assurance. I have taken a class on each of these subjects, and could not always identify the difference. There was definitely a sign of strong overlap. Plus, it can be difficult to distinguish between the two from the definitions alone.

In hoping to better understand these topics, and their differences, better, we have decided to analyze the co-citation networks in the subjects of Information Security and Information Assurance by creating visualizations using CiteSpace and comparing them. Through this, we hope to better understand the similarities and differences between these two topics.

2 PROCEDURE

For each of the two topics, we followed a similar procedure. The first step was to download the article information from Web of Science. In order to produce a comprehensive web for each of these two topics, we downloaded 2,000 entries in the web of science for information security topic and 1,000 entries for information assurance topic. These entries were saved into text files and used in CiteSpace.

Once we had collected all of the results for both topics, they were inputted into text files, which were then used to create project files for CiteSpace. For each topic, we opened CiteSpace, creating a project containing the appropriate data that was downloaded from Web of Science. After configuring CiteSpace, we pressed the "Go" button to process the data, and then "visualize" to display the results.

To attempt to get the best results possible, we tested several different sets of results for each visualization, and created visualizations for each, to determine which results produced the most interesting and informative results.

Once we had done our initial analysis of the result, we configured CiteSpace to identify the clusters within the visualization and to number and name them, making it easier for us to identify which clusters are which.

The visualizations in CiteSpace indicate the individual articles and patterns among co-citations of each of the articles. By analyzing the web that was produced, we can find the important papers and some interesting trends in the research on each of these two topics.

3 INFORMATION SECURITY

Information Security has been one of the most increasing concerns worldwide. Different systems serve multiple clients. Therefore, they are greatly exposed to more threats and attacks. Analyzing information security research with CiteSpace will give us a great vision and a clear picture about the historical pattern and evolving trends. It also helps us to identify the future trends of research areas.

3.1 Clusters

As shown in Figure 1, there are more than 50 clusters that have been generated with Citespace. Most of the clusters are concentrated in the center . They are also linked together through more than one paper. However, other small clusters are scattered in the visualization. They aren't linked to any other clusters in anyway. As shown in Table 1, the largest clusters in information security. The largest cluster has 35 members. The second large cluster has 29. It's worth point out that the average year for most of the largest clusters is around 1996. We will cover the largest five clusters in detail.

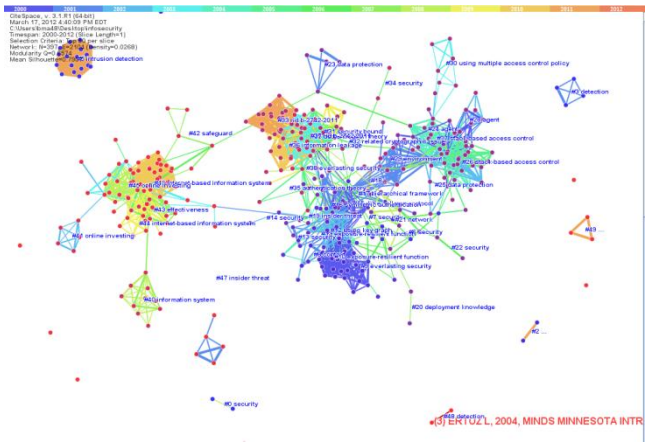

Figure 1: Information security article clusters

The first one is cluster number 46. It's the largest cluster that has 35 members in the network. It's labeled as "internet-based information system" by mutual information. Most of the articles in this cluster focus on secure any an information system that uses Internet technologies to deliver information and services. They focus on forcing and implementing security policy. They also cover management and investigation of policy violations. Many clusters have emerged from this cluster such as online investing and effectiveness.

The second cluster is number 26. It's labeled with "stack-based access control" by mutual information. It's about securing information flow in programming languages, operating system and applications. The idea of the control system was born in BELL DE (1973) article. It interests many multiple fields in information security such as Network Security and Secure Mobile Ad. This cluster has been highly active since it emerged from authentication theory cluster, see Figure 1.

ClusterID	Size	mean(Year)	Label (MI)
46	35	1995	internet-based information system
26	29	1990	stack-based access control
33	22	1996	rid b-2742-2011
47	18	1996	insider threat
45	18	1996	online investing
5	17	2002	intrusion detection
10	15	1991	exposure-resilient function
31	15	1997	security bound
9	14	1991	everlasting security
27	14	1995	stack-based access control
29	13	1995	environment
39	11	1996	authentication theory
12	10	1993	using key graph
36	10	2000	information leakage
40	10	2004	information system
43	9	1995	effectiveness
16	8	1990	symmetric authentication
28	7	1996	agent
35	7	1992	authentication theory

Table 1: Largest clusters in information security

The third cluster is number 33. It's labeled with securing fingerprint minutia with TFIDF. It's about providing a way to guarantee the integrity of secure message. It's been highly cited in many other clusters that focus on the protection of data between end points and the quality of cryptographic. The cluster has 22 members. The members share a common topic about securing the communication with different fingerprints such as providing a hash function. Although the cluster emerged after "Blachman Nm" paper in (1965), the researches have been continued since then.

The fourth largest cluster is number 47. It has 18 members. It's labeled as "Insider threat" by mutual information and camouflage with TFIDF. Most of the cluster's members focus on security of central part of systems from any inside threat.

The fifth cluster is number 45. It's labeled with "online investing" by mutual information. It's about implementing a secure infrastructure for online exchanging information between nodes. The paper Schneier B (1996) is very essential in this cluster. It has a high centrality ratio. It has also been cited from different clusters. This cluster was emerged from internet-based information system, see Figure 2.

Figure 2: Online investing cluster

3.2 Papers

There are many important papers in the field of information security. However, we've chosen the most important and highly cited papers in that field. The first chosen paper is "Communication Theory of Secrecy Systems"[1]. This paper, for many reasons, is the most important paper in this area. It's the fundamental paper for recent papers. Therefore, it's the most cited paper among all papers in the network. The paper deals with the basic mathematical structure of secrecy systems. It provides solid cryptographic theory. It is the foundational paper of modern cryptography. It's highly cited for many other foundational papers in cryptography area. As it is shown in Figure 3,

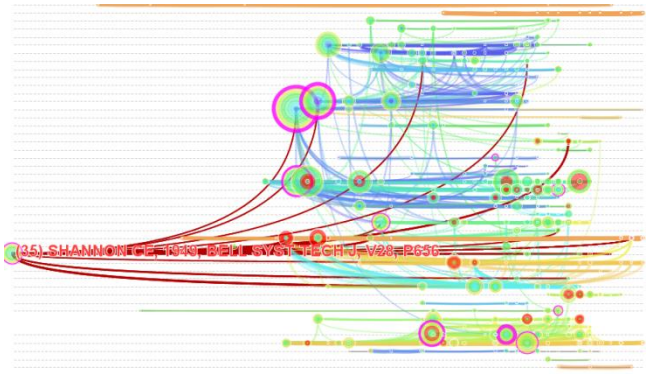

Figure 3: E. Shannon's paper influence on other papers

The second paper is "New directions in cryptography". This paper has the highest frequency. The paper provides a secure mechanism to exchange keys between two entities with no prior knowledge. This is known as "public key cryptosystem" The paper also explores one way authentication in details. This paper was a revolution in cryptography. It is also the foundation for cluster information security. The high frequency of this paper in that cluster shows how much it is fundamental to this field. Besides its high frequency in information security cluster, it has the highest centrality too with 0.44. See Figure 4.

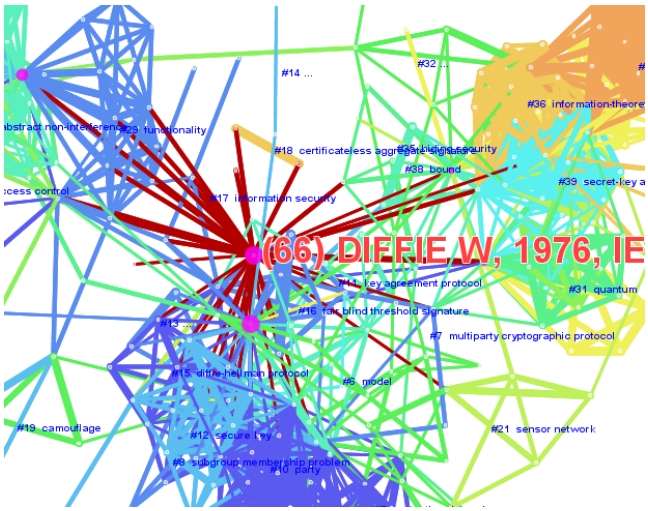

Figure 4. The importance of paper W. Diffie and M.E. Hellman.

The third chosen paper is "A method for obtaining digital signatures and public-key cryptosystems". A cluster 16 is labeled with "Symmetric authentication". Even though, the paper presented a solid mathematical mechanism to for , the symmetric authentication has improved since then. That explains the emerging of the trend symmetric authentication. The importance of this paper is that its mechanism is still being used until now. It has the highest frequency, citation burst, centrality in that cluster. This paper is know now as RSA paper.

3.3 Historical Patterns

Researches have been conducted in Information security field for a long time. As it was mentioned earlier in this paper, the average year of the largest clusters in this field is 1996. That means members range between 1970 and 2011. The field was focused on authentication theories. It started in Shannon's paper (1949). Then the field started to evolve in many different areas such as cryptography, control policy and secure exchange. However, The real revolution was after Rivest and Shamir paper. That paper presented an algorithm that enable two parties to securely exchange information. That algorithm has been used since then in many applications and different systems.

3.4 Trends

As Shown in Figure 1, there are many visible trends in information security field. The oldest trend was authentication theory. Then two trends (stack-based access control and internet-based information system) emerged in almost close date. The field reached a significant achievement with Diffie and Hellman's paper and RSA paper. Diffie-hellman protocol and symmetric authentication trends had emerged from those papers. They lasted for a while and they were frequently appeared in many researches back then. The field continued to evolve in different ways simultaneously. Intrusion detection along with online investing were have been trending for more than 10 years. Data protection

4 Information Assurance

After creating the Information Security web, we created the visualization for Information Assurance by gathering the top articles from Web of Science in a similar way, and running them through CiteSpace to create the visualization of co-citations among Information Assurance articles. The visualization looks dissimilar to the visualization for Information Security, as the big group of clusters is not quite as large and there are fewer smaller clusters on the outside.

4.1 Clusters

The visualization has numerous clusters, all spread apart, ranging from clusters of two or three papers to a few larger clusters that contain many papers. In the center of the visualization lies a group of interconnected clusters, numbered 28-34. This section contains the papers that are at the heart of the study of information assurance.

These clusters are held together by several papers, and hiding one or two of the major papers in this group of clusters will not separate them. An exception to this is the cluster 28 "Examining Internet Banking", which is connected to the others only by the paper "DEVARAJ S, 2002, INFORM SYST RES, V13, P316". Hiding this paper will separate this one cluster from the rest. Two other smaller clusters (number 30, "trust" and number 33, "delphi study") can be separated in a similar way, by hiding "NUNNALLY JC, 1978, PSYCHOMETRIC THEORY, V, P" and "MCKNIGHT DH, 1998, ACAD MANAGE REV, V223, P473" respectively.

Clusters closer to the inside of this group, such as cluster numbers 29 ("privacy") and 31 ("trust") are so tightly together that it is difficult to determine which papers are part of which. Number 34, "lithuanian company" also appears to be in the center of this group, but is difficult to determine where as the label is separate from the rest of the cluster.

The largest cluster is number 34, with 55 articles in it and the second-largest cluster is 31, with 39 articles in it.

4.2 Papers

The papers in the large section of clusters in the center of the visualization are obviously important for the topic of Information Assurance. In the center of this cluster appears to be a paper called "GEFEN D, 2003, MIS QUART, V27, P51". This paper connects to several of the clusters, and has the most connections within this cluster, at 15. Another paper in this cluster "MAYER RC, 1995, ACAD MANAGE REV, V20, P709" comes in a close second with connections to fourteen other papers throughout these clusters. These are also the two most frequently cited papers that appear on the visualization.

Figure 5. The complete web produced from the Information Assurance results from Web of Science

The third and fourth most frequently cited papers, both with a frequency of 11, are also in this large cluster. It is also interesting to note that these papers are close to the centre of this cluster, and almost all of the connections are from 2010-2011.

4.3 Historical Patterns

Interesting to note is that in the large group of clusters, most of the papers are only in the 2009-2011 area. Some of the smaller clusters near the outside of the connected cluster group contain papers cited in 2000 or 2006. This seems to indicate a transition. Information Assurance is still a relatively young field, so the papers on the outer clusters indicate what was important in the earlier days, and as you go close to the center, you find papers that are more relevant today, as the nature of the subject changes.

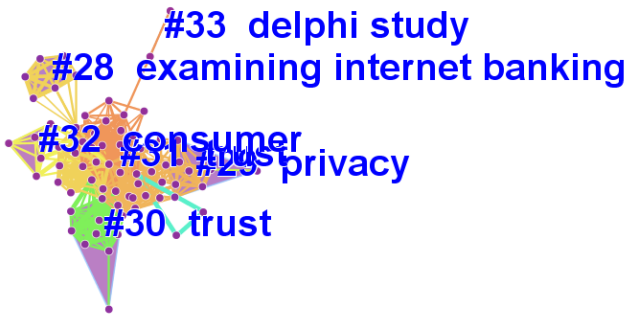

Figure 6. A close-up of the group of clusters 28-34

Also interesting to note is that there are no papers linked with lines with the colors for a year between 2000 and 2006. Perhaps this suggests that there was little significant change in the field between those years.

4.4 Trends

There are no noticeable trends in clusters other than the group of clusters in the center. In this cluster, from the visualization, you can notice several trends in the research in the field, as over time, citations transition from clusters toward the outside of that group (like number 30, "trust") towards the inner clusters (such as number 31, also called "trust"). As you get closer to the center, the clusters are joined together and less well-defined, perhaps indicating the increasing complexity of the research as the study of Information Assurance gets older.

5 Overlap Points

After analyzing both subjects, we looked for overlapping points. Although information security and information assurance are using interchangeably to refer to protection of data, there was no overlapping between those two subjects at all. The information security trends were cryptography, online investing, security, data protection, information leakage, etc. On the other hand,

information assurance trends were quality assurance, data collection, system integrity, quality paradigm. Information security has a lot of connected clusters whereas clusters in information assurance are scattered. Even the classic papers in information security are not present in information assurance visualization.

Based on the generated visualization, information security focuses on methods and tools to protect and control data. It's interested in defining ways to preserve the confidentiality, integrity, and availability of data. Information assurance, on the other hand, focuses preserving the quality and integrity of a specific organization in general. It doesn't involve information security methods or technique.

6 SUMMARY

In this paper, we presented two different visualizations for information security and information assurance topics. We chose those topics in order to show the similarity and overlap between them. We started by collecting 1,000 records or more for each topic. We ran CiteSpace on those records to create clusters. By using the timeline, cluster layout, we used those clusters to analysis and identify the historical patterns, trends and the development of the fields.We identified the important papers in each field along with their trends. Then we tried to predict the near future trends in both fields. We compared those results to identify the overlapping points.

7 CONCLUSION

This paper presented that information security and information assurance have no distinct overlap. Both fields have different clusters, historical patterns and trends. There are no similar networks in the research development. The classic and fundamental papers in information security are not present in information assurance graph and vice versa.

Despite our initial perception, Information Security and Information Assurance are two distinct, and very different fields, and there is little overlap. Perhaps the overlap we perceived was the result of the two fields being applied together.

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," BellSyst. Tech. J., vol. 28, pp. 656-715, Oct. 1949.

[2] W. Diffie and M.E. Hellman, "New Directions in Cryptography," IEEETrans. Information Theory, vol. 22, pp. 644-654, 1976..

[3] R. L. Rivest , A. Shamir , L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Communications of the ACM, v.21 n.2, p.120-126, Feb. 1978.