Chapter 8
Preventing Social Engineering and Espionage in Collaborative Knowledge Management Systems (KMSs)

Oluwafemi S. Ogunseye
University of Agriculture, Nigeria

Olusegun Folorunso
University of Agriculture, Nigeria

Jeff Zhang
Ball State University, USA

ABSTRACT

Insider attack and espionage on computer-based information is a major problem for business organizations and governments. Knowledge Management Systems (KMSs) are not exempt from this threat. Prior research presented the Congenial Access Control Model (CAC), a relationship-based access control model, as a better access control method for KMS because it reduces the adverse effect of stringent security measures on the usability of KMSs. However, the CAC model, like other models, e.g., Role Based Access Control (RBAC), Time-Based Access Control (TBAC), and History Based Access Control (HBAC), does not provide adequate protection against privilege abuse by authorized users that can lead to industrial espionage. In this paper, the authors provide an Espionage Prevention Model (EP) that uses Semantic web-based annotations on knowledge assets to store relevant information and compares it to the Friend-Of-A-Friend (FOAF) data of the potential recipient of the resource. It can serve as an additional layer to previous access control models, preferably the Congenial Access Control (CAC) model.

DOI: 10.4018/978-1-4666-2053-7.ch008

INTRODUCTION

If business organizations and governments were cars, knowledge will be the fuel they require to achieve the purpose of their creation, which is movement. As on point as this analogy is, it seems to undermine the importance of knowledge to the different sectors of the world. While we will prevent harping on the issue, we live in a world of competition where there seems to be a conscious agreement (with few exceptions) that in order for knowledge to be valuable for competition, it must be rare, non-imitable and non-substitutable (Uren et al., 2005). Knowledge management concentrates on the processing and storage of documents and the business processes that build on them. These documents provide a rich resource describing what an organization knows (Uren et al., 2005; Sure et al., 2003). They are believed to account for 80-85% of the information stored by many companies. Uren et al. (2005) and Sure et al. (2003) cited contracts, consulting reports, and consumer surveys as examples of documents that can be stored as knowledge resources. Regular web pages can also be formats for knowledge assets.

For systems and organizations to remain relevant and competitive, these knowledge assets must be protected and made scarce to the outside world (Desouza & Vanapalli, 2005). Most research on security of knowledge assets has focused on security against threats from outside sources. These external threats, called intrusions, are handled by access control methods and other techniques. However, the Federal Bureau of Investigation in the US estimated that corporations lose $100 billion, annually, to industrial espionage (Winkler, 1996). This makes clear the fact that insider threats also pose a major problem to business and government systems. This issue of extrusion and insider abuse becomes more delicate when we consider the fact that there is now a continuous rise in alliances between organizations and arguably increasing interests in outsourcing (Desouza & Vanapalli, 2005). Employees, who have all requisite access rights, can send valuable knowledge resource(s) to remote locations or even to partnering (competing) organizations at the detriment of the source organization. In partnering organizations, if two companies A & B are partnering on a project, Company A's employees with access right to company B's Knowledge Systems can abuse that right; stealing valuable knowledge resources from B's organization. As KMSs become more and more semantic web compliant in nature and design, the advantages provided by the design and framework of semantic web can be put to good use in enhancing security for KMSs. Explored in this work are advantages and opportunities, such as this.

KMS FACILITIES OF THE SEMANTIC WEB

Tim Berners-Lee, one of the inventors of the World Wide Web, proposes a more machine-processable web as a development route for the current web. His work on the "semantic web" as an extension of the current web is under progressive research. For the semantic web to work, machines have to be able to not only read web-based information, but also understand it. The term "machines" as used in this statement refers to intelligent agents and software that work on the web. Therefore, these machines should be able to process web-based content including text documents, media, and graphics. This can only be possible through the concept of "intelligent" documents as imagined by the Delphi Group (1994). Intelligent documents are documents that have some degree "self-awareness", meaning that they know who created them, what they might contain, and other information that will enable a machine know what to do with them. This was traditionally accomplished through the use of metadata, but has been replaced with semantic annotations based on domain ontologies (Berners-Lee et al., 2001). The advantages of such annotations are quicker search and retrieval of documents, the automation of several web-based activities, etc. (Gardenfors, 2004; Frienland et

al., 2004; Dowman et al., 2005; Rinaldi et al., 2004; Plessers et al., 2005; Maynard et al., 2004; Hunter et al., 2004). Different methods have been employed to annotated knowledge assets; these are comprehensively tackled in Uren et al. (2005). However, they pointed out that semantic meta-data can be included in documents (knowledge assets) either manually, by humans, or automatically, by machines. Since humans are prone to error, many automated methods of semantic metadata inclusion were reviewed. Gardenfors (2004) agrees that any form of annotation deemed necessary by the developer can be put in a document. We will therefore focus on annotations that help us identify the source of a document and its creator's information.

PROBLEMS WITH EXISTING ACCESS CONTROL MODELS

Okesola and Ogunseye (2010) and Cranor and Garfinkel (2005) submitted that security can have an inverse relationship with the usability of KMSs. Sodiya and Onashoga (2009) and Ogunseye and Okesola (2011) also showed gaping holes in many access control methods that can impede their ability to secure assets. In this section, for the sake of clarity, we review two groups of access control methods based on their effect on usability and their capabilities: the general access control techniques (common to most information systems) and the specialized access control techniques designed to meet the sensitive needs of knowledge management (KM).

General Access Control Techniques: An Example of RBAC

Role-Based Access Control (RBAC): it has been one of the most successful access control systems used in many information systems projects. There are many documented works on RBAC (Sandhu et al., 1996; Covington et al., 2001).

Okesola and Ogunseye (2010) established in their work that the RBAC is not a suitable access control method for knowledge management systems. Aside from being too rigid, researchers continue to work towards solving the problem of role engineering, which can be an exploitable loophole for corporate espionage as its complexities increase when companies merge or work together.

Figure 1. Congenial access control model

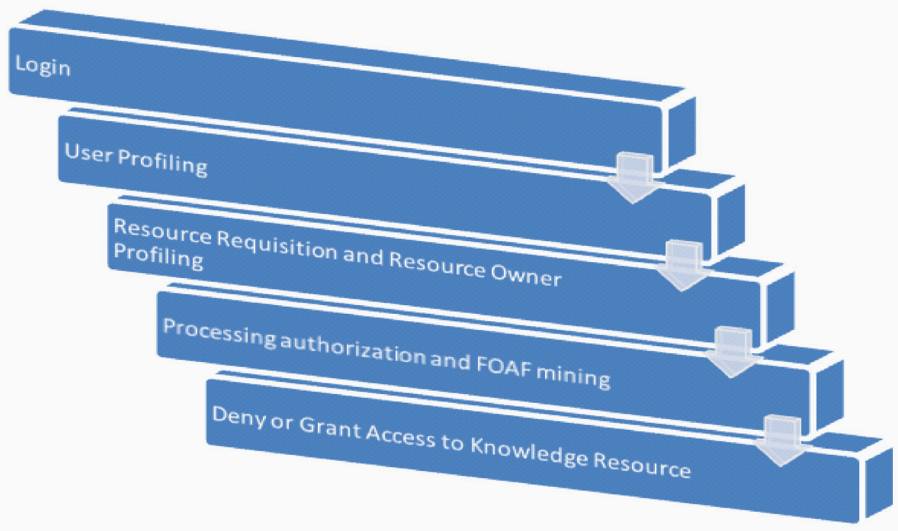
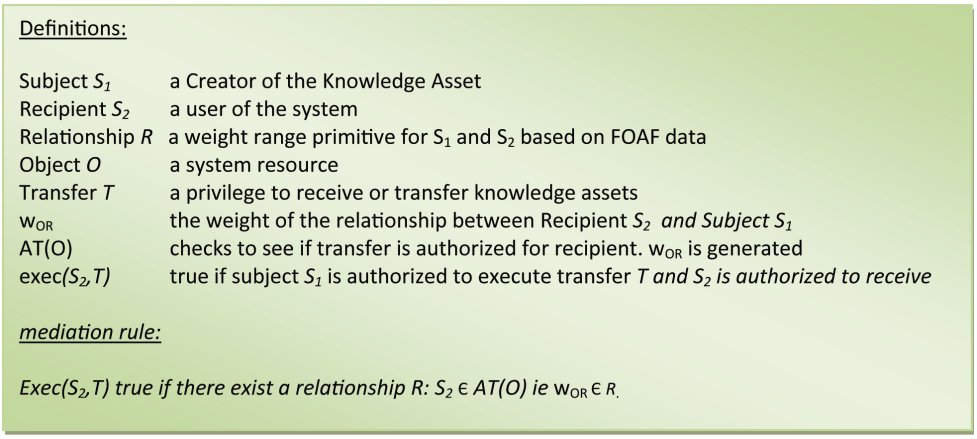


Figure 2. Espionage prevention layer model



Following these lines of thought, we reiterate its unsuitability for protecting KMS against espionage and extrusion. Consider a scenario where a staff member of a business organization is duly logged in and abuses his privileges by transferring valuable knowledge resources to a competing organization that has access to company network through partnership.

In similar light, we see that many other traditional access control techniques are not suitable for preventing internal threats because it is possible for a legitimate users to abuse their privileges. Some popular access control methods and their flaws were pointed out in Ogunseye and Okesola (2011) and Sodiya and Onashoga (2009).

Specialized Access Control Techniques: An Example of CAC

The Congenial Access Control (CAC) model was proposed in 2010, by Okesola and Ogunseye, as a viable access control method for KMSs. The novel model uses a page rank algorithm-based reputation computation method to provide or deny users access to resources in a KMS. To decipher the relationship status between a resource requester and a resource owner, the model uses customized "friend of a friend" (FOAF) data which serves as a record set for each user's "user information" and "group information".

Potential Users' FOAF data and that of each member of their Community of Practice (CoP) are probed using web-mining techniques to see if there is a strong enough relationship between the potential resource user and the resource owner, or his group, using a Google "PageRank-like" algorithm to weigh the relationships. The steps involved in the CAC model are illustrated in Figure 1.

The activities of each step are explained as follows:

Step 1: Login

The user logs in to his computer without necessarily having to log into the KMS as a separate entity.

Step 2: User profiling

The user is immediately profiled and his FOAF file pulled from the database.

Step 3: Resource Requisition and Resource Owner Profiling

When the user requests a resource, the FOAF file of the owner is immediately pulled to an active memory.

Step 4: Processing Authorization and FOAF mining

The FOAF file is mined and the data it contains is extracted for use with the computation of the reputation score.

Step 4 involves four stages:

1. Get the prescribed threshold score for access authorization set by the KMS administrator.
2. Check if the user is directly connected to the resource owner. If yes, grant access. If no then go to Stage 3.
3. Check if the members of the group are connected to the owner or a member of the owner's group and compute their authorization score.
4. If authorization score ≥ threshold, then grant access. Otherwise, ask the user to request permission from owner.

The FOAF data would contain the person's name, the group(s) he/she belongs to, and their position, which is used in computing the level in the organogram scale.

The algorithm implies a member of a group can request resources belonging to someone or a group, which he is not directly connected to, but indirectly connected to (e.g., through a colleague). This would not be possible in a role-based architecture.

While the CAC model is considered a very strong and Pro-KM model, it does not secure against the document's creator or his CoP sending the knowledge asset out to their partners in collaborative KM systems even when those partners are ordinarily not authorized access to such a system/knowledge resource.

THE INSIDER ATTACK PREVENTION MODEL

Computer-Based Industrial Espionage and insider attacks, even by disgruntled employees, are a blind spot for many access control models. These models might work relatively well in preventing unauthorized access to knowledge assets in an organization, but the cogent question is how well do they fare when the attacker has all the required access rights? For instance, the creator of a knowledge asset, while working in an organization, might have used the opportunities presented by his environment to synthesize that knowledge. Therefore, that knowledge cannot be shared with

Figure 3. Transactions of the complete iCAC model (a)

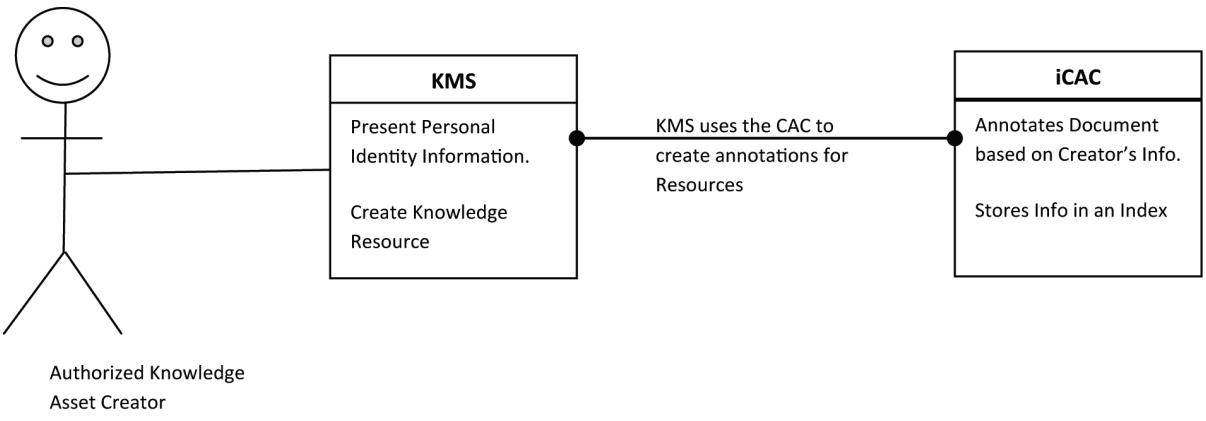
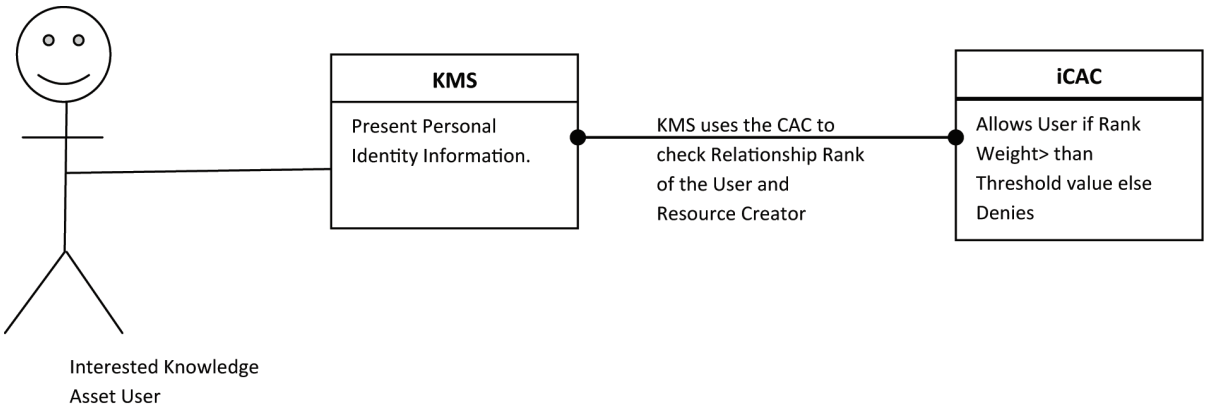


Figure 4. Transactions of the complete iCAC model (b)



(b)

the company's competitors without the company's permission, even when there is collaboration between them. Current access control methods makes sure that only those with authorized access can use the knowledge resource, but they do not cater to privilege abuse by these authorized persons. Every other information security measure comes to naught when the user is authorized. Matasano (2007) emphasizes that many extrusion prevention systems try to protect information when they are already in the wrong hands. For instance a KMS that encrypts its content will, of course, decrypt it for authorized users. As the saying goes, the only system that will be exempted from an insider attack will be the one that is isolated, not on a network, not turned on, and without a memory or a processor. Consider another scenario where the Chief Information Security Officer is the one that has been compromised. He has overall authority on roles in RBAC systems and can set and unset privileges as he pleases. The audit trails are also within his reach. In such a situation, the organization is never safe.

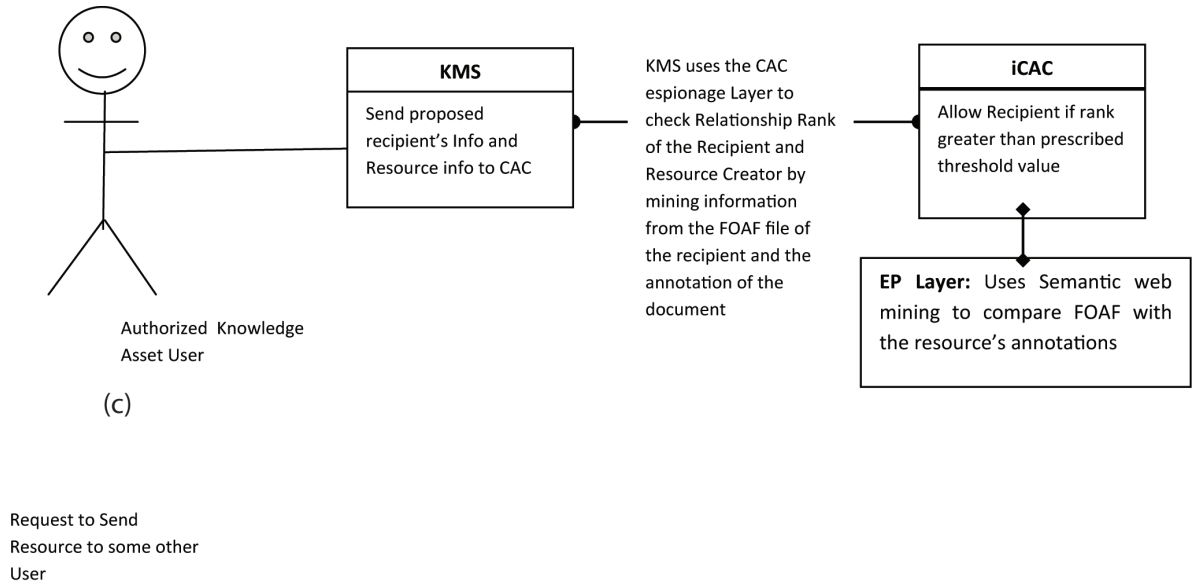
The CAC model will perform better in this scenario because it reviews the relationship between the resource requester and the owner of the resource before granting or denying access. You are not granted access because you are the system administrator (although some users can be granted superior access rights), but when you are the owner of the resource or it came from your CoP, then you can abuse that privilege. The "Insider attack prevention" model is therefore designed to prevent exactly such cases.

The Model Make-Up

In this work we assume that the KMS is on a network and organizations can partner together towards business ends and projects. We also assume that all forms of secondary storage devices are not allowed, but the system is on a network in an expanding and dynamic organization. We use customized annotations that identify the document's source and owner. Uren et al. (2005) gave an example of how this might be useful in providing access to information to only a particular set of people in their paper. But we take it a step further. We view the transaction between an authorized user and the person he is sending the knowledge asset to as "message sending".

In the model representation in Figure 2, we see that the object of focus is not the "transferor", who we assume must be a user with access permissions

Figure 5. Transactions of the complete iCAC model (c)



(enforced by the original CAC), but the potential recipient. The simplistic view and explanation to the specifications given in Figure 2 is presented.

1. The authorized user accesses the system.
2. The user accesses a knowledge resource.
3. The user decides to transfer the resource to a another user.
4. The document's annotations are pulled and parsed by an annotation parser.
5. The document's owner information and sensitivity ranking of the document is checked.
6. The potential recipient's FOAF file is checked for links with the document's owner.
7. If the recipient fits the profile of allowed users, then transfer access is granted. Otherwise, transfer access is denied and the transaction is logged into different location(s) for future records.

The operations of the improved CAC (iCAC) model are shown using use case diagrams for the actors as seen in Figures 3 through 5.

The use case scenarios above depict the major transactions of the Semantic KMS using the CAC. The EP layer performs functions as previously described in Figure 2.

Stumme et al. (2006) describe resourceful ways in which annotations can be mined.

Ensuring a multi-level authorization requirement for threshold value modification can further strengthen this model. This implies that the system administrator does not have the sole right to modify threshold values. As few as three people can be required to approve his request for threshold value modification before it is effective. This is to ensure that no one man can steal company knowledge resources because he has the power or is granted such privilege to.

CONCLUSION

The EP layer model above improves on an existing model of the CAC adding facilities for insider attack prevention and detection. The design aims at reducing the possibility of privilege abuse and social engineering to the barest minimum for semantic web based KMS. The CAC model of access control is easier on the KM processes

and semantic web-based annotations continue to reveal opportunities for improvement, which cannot be ignored. We believe this work will help stem the age-long problem of insider attacks, espionage, and some forms of extrusion, which the other access control models could not, all while maintaining the desired user friendliness requirement of access controls that work on knowledge management systems.

REFERENCES

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1),99–120.doi:10.1177/014920639101700108

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 34–43. doi:10.1038/scientificamerican0501-34

Covington, M., Moyer, M., & Ahmad, M. (2008). *Generalized role based access control for securing future applications* (Tech. Rep. GIT-CC-00-02). Atlanta, GA: Georgia Institute of Technology.

Cranor, L., & Garfinkel, S. (2005). *Security and usability: Designing secure systems that people can use*. Sebastopol, CA: O'Reilly Media.

Delphi Group. (1994). *The document process.* Retrieved from http://www.delphigroup.com/research/whitepapers/DocIsProcess.pdf

Desouza, K., & Vanapalli, G. (2005). Securing knowledge assets and processes: Lessons from the defense and intelligence sectors. In *Proceedings of the 38th Hawaii International Conference on System Sciences.*

Dowman, M., Tablan, V., Cunningham, H., & Popov, B. (2005, May 10-14). Web assisted annotation, semantic indexing and searching of television and radio news. In *Proceedings of the 14th International World Wide Web Conference*, Chiba, Japan (pp. 225-234).

Friedland, S., Allen, G., Mathews, G., Witbrock, M., Baxter, D., & Curtis, J. (2004). Project Halo: Towards a digital Aristotle. *AI Magazine*, 29–48.

Gardenfors, P. (2004). How to make the semantic web more semantic. In *Proceedings of the 3rd International Conference on Formal Ontology in Information Systems.*

Hunter, J., Schroeter, R., Koopman, B., & Henderson, M. (2004, June 10). Using the semantic grid to build bridges between museums and indigenous communities. In *Proceedings of the GGF11- Semantic Grid Applications Workshop*, Honolulu, HI.

Jiang, X., & Tan, A. (2005). Mining ontological knowledge from domain specific text documents. In *Proceedings of the ICDM Conference* (pp. 665-668).

Matasano. (2007). *Defeating extrusion detection.* Retrieved from http://www.blackhat.com/presentations/bh-usa-07/Monti_and_Moniz/Presentation/bh-07-monti_and_moniz.pdf

Maynard, D., Yankova, M., Aswani, N., & Cunningham, H. (2004). Automatic creation and monitoring of semantic metadata in a dynamic knowledge portal. In *Artificial Intelligence: Methodology, Systems, Applications* (LNAI 3192, pp. 65-74).

Ogunseye, O., & Okesola, J. (2011). Meta-heuristics based multi-layer access control technique (MBMAC). *ANALE Seria Informatica*, 145-154.

Okesola, J., & Ogunseye, O. (2010). A congenial access control technique for knowledge management systems. *Global Journal of Computer Science and Technology*, *10*(14), 2–6.

Plesser, P., Casteleln, S., Yesilada, Y., De Troyer, O., Stevens, R., Harper, S., & Goble, C. (2005, May 10-14). Accessibility: A web engineering approach. In *Proceedings of the 14th International World Wide Web Conference*, Chiba, Japan (pp. 353-362).

Rinaldi, F., Schneider, G., Kaljurand, K., Dowdall, J., Persidis, A., & Konstanti, O. (2004, September 24). Mining relations in the GENIA corpus. In *Proceedings of the 2nd European Workshop on Data Mining and Text Mining for Bioinformatics*, Pisa, Italy.

Sandhu, R., Coyne, E., Feinstein, H., & Youman, C. (1996). Role based access control models. *IEEE Computer*, *29*(2), 38–47. doi:10.1109/2.485845

Sodiya, A., & Onashoga, A. (2009). Components-based access control architecture. *Issues in Informing Science and Information Technology, 6*.

Stumme, G., Hotho, A., & Berendt, B. (2006). Semantic web mining state of the art and future directions. *Journal of Web Semantics*, *4*(2). doi:10.1016/j.websem.2006.02.001

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., & Ciravegna, F. (2005). Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, *4*(1).

Winkler, I. (1996). *Case study of industrial Espionage through social engineering*. Carlisle, PA: National Computer Security Association.

This work was previously published in the International Journal of E-Adoption, Volume 3, Issue 4, edited by Sushil K. Sharma, pp. 44-51, copyright 2011 by IGI Publishing (an imprint of IGI Global).