12-4-2007

Importance of Verification and Validation of Data Sources in Attaining Information Superiority

Gautham Kasinath
Edith Cowan University

Leisa Armstrong
Edith Cowan University

DOI: 10.4225/75/57b54ce7b875e
5th Australian Information Security Management Conference, Edith Cowan University, Perth Western Australia, December 4th 2007

Importance of Verification and Validation of Data Sources in Attaining Information Superiority

Gautham Kasinath and Leisa Armstrong
School of Computer and Information Science
Edith Cowan University
Perth, Western Australia
gautham.kasinath@gmail.com
l.armstrong@ecu.edu.au

Abstract

Information superiority has been defined as a state that is achieved when a competitive advantage is derived from the ability to exploit a superior information position. To achieve such a superior information position enterprises and nations, alike, must not only collect and record correct, accurate, timely and useful information but also ensure that information recorded is not lost to competitors due to lack of comprehensive security and leaks. Further, enterprises that aim to attain information superiority must also ensure mechanisms of validating and verifying information to reduce the chances of mis-information. Although, research has been carried out into ways to increase the security of information and detect leaks, not enough focus has been given to the key elements of information, namely data and context. This paper outlines the importance of data in contributing to information superiority and highlights the lack of data centric approach in attaining information superiority. The paper also discusses the importance of verification and back tracking of information to ascertain the data, its source and context in validating information for its correctness, validity and accuracy. A brief list of consequences of information leaks is also provided in the document to emphasize the importance of information security in the context of data collected. Further, this paper examines the McCumber model, which outlines the various states and elements of information, to accommodate a data centric, quantitative approach. Outlining simple protocols for verification of data in the information superiority context, this paper also highlights a few steps that can be taken to verify the sources of data.

Keywords

Information superiority, information assurance, data centric information assurance.

INTRODUCTION

"Information is a develop once – use many times – simultaneously" commodity (Alberts and Garstka, 1999). Information once gained and checked for validity, is then used by different disparate entities, mostly at the same time, to make numerous decisions, in both the corporate and the defense arena (Alberts and Garstka, 1999). Hence, although good information will bring about good decisions, bad information has an equal, if not greater, impact on the decisions, in reverse (Alberts and Garstka, 1999).

In recent times, the world has seen a global effect of misinformation. One such was the U.S.A War on Terror and the subsequent invasion of Iraq during 2002. The decision makers in the U.S.A. had been misinformed of the presence of the weapons of mass destruction in Iraq. The aftermath of the war, saw that the information was indeed wrong, but had brought about irreversible impact on not only Iraq's but the global economy (Center for American Progress, 2004; Schakowsky, 2003; Whitney, 2005).

Information can be considered as collated data in context, and as such bad sources of information may cause more harm that good (Hutchinson and Warren 2001). With a proliferation of sources of information and the excessive amount of information coming into any decision making body, the key is to filter out the bad information and exploit the good ones (Hutchinson & Warren, 2001). Hence, importance of knowing the source of information, the original data, is vital to superior decision making. (Albert et al., 2001) define information superiority as a state that is achieved when a competitive advantage is derived from the ability to exploit a superior information position. The objective of those that seek information superiority is to use information superiority to create and maintain a competitive advantage (Alberts et al., 2001). This paper will outline the role that data/information source(s) plays in gaining information superiority. To attain information superiority, the quality and assurance of information is important. The McCumber model is one model which has been used to interpret the role that information plays to determine information assurance (McCumber, 1991; Maconachy et

al., 2001). The McCumber outlines the significance of various attributes of information, namely availability, integrity and confidentiality in determining the quality of information. However, the McCumber model takes an information centric approach and does not account for the role that data and context in the quality assurance and security of information. This paper aims to prove the importance of including a data centric component to the model. A proposed improvement on the original McCumber model which is data centric will be outlined.

ATTRIBUTES OF INFORMATION RICHNESS

Data are observations of the environment while information is the aspect that affects ongoing decisions (Anborg, 2000; Parker, 1995; Perry et al., 2004; Schou and Trimmer, 2004). As such, information is widely referred to as the interpretation of data. Though there are numerous tools and models, like the Information Security (INFOSEC), Information Assurance (IA), Operationally Critical, Threat, Asset and Vulnerability (OCTAVE), for measuring and controlling information availability, assurance and risk, there have been few research studies, that have measured the validity and usefulness of data and context alone collected to arrive at information. Few studies have reported on the ways to measure the usefulness of usefulness of information (Anachy, 2001; Albert et al., 2003). However, a study by (Alberts et al., 2001) has concluded that a number of attributes can be used to measure information richness including Completeness, Correctness, Currency, accuracy and Consistency. Alternatively, other research by (Parker, 1995) has supplemented attributes such as authenticity, utility and possession as a elements of information.

The completeness of information can be measured as a function of the completeness of data or the combination of data obtained from various sources (Alberts et al., 2001; Alberts and Garstka, 1999; Maconachy et al., 2001; McCumber, 1991). This attribute, can be expressed mathematically, in terms of data as $I(d) \alpha \Sigma (d)$, where I (d) is the Information derived out of data and Σ (d) is the sum of all data. The correctness of information is based on the outcome of the decision based on the information.

The correctness of information is by no means the accuracy of the information, its consistency nor its authenticity (Alberts et al., 2001; Alberts and Garstka, 1999; Maconachy et al., 2001; McCumber, 1991). The accuracy of information is directly proportional to the data, which leads to the information. This attribute can be expressed mathematically in terms of accuracy as $I(d) \alpha \Sigma (A(d))$, where I (d), is the Information derived out of data, A (d) is the measure of accuracy of the data (Alberts et al., 2001; Alberts and Garstka, 1999; Maconachy et al., 2001; McCumber, 1991; Perry et al., 2004). A classic scenario which could be used to demonstrate this concept is the sensors in the Hubble telescope. If the sensors are not accurate in collecting data, then the position of stars and planets will be malformed. This would lead to a different perspective of the spatial abundance around the world. A set of bad sensors/agents providing the same data, does not mean that the information is correct and reliable (Perry et al., 2004).

One further attribute which has been used a measure of information richness is consistency. The consistency of information is the measure of quality of data collected from numerous sources. Data can be said to be consistent, when over a large set of data sources, it has an allowable error. Although consistent data does not prove, with conviction, that the data is correct, the measure is perhaps an indicator on the accuracy and correctness of the data. This can also have the reverse effect. (Alberts et al., 2001; Alberts and Garstka, 1999; Maconachy et al., 2001; McCumber, 1991)

DOMAINS PERTAINING TO INFORMATION

Information and data may exist in three domains including physical, information and cognitive (Alberts et al., 2001; Perry et al., 2004, Figure 1). Since the 17th century, researchers have considered that the physical domain consists of the data collection agents, which are classified as direct and indirect sensors. The direct sensors like telescopes, field glasses etc provided useful data about enemy movements. In the corporate arena, the use of spies as direct sensors influenced the corporate policy and direction (Alberts, 2001; Denning, 1999; Perry et al., 2004). The impact of data in the physical domain has little significance because the data has not yet been interpreted. Hence, little understanding of the procured data can be derived and as such this may affect any decision made based on this data (Alberts et al., 2001).

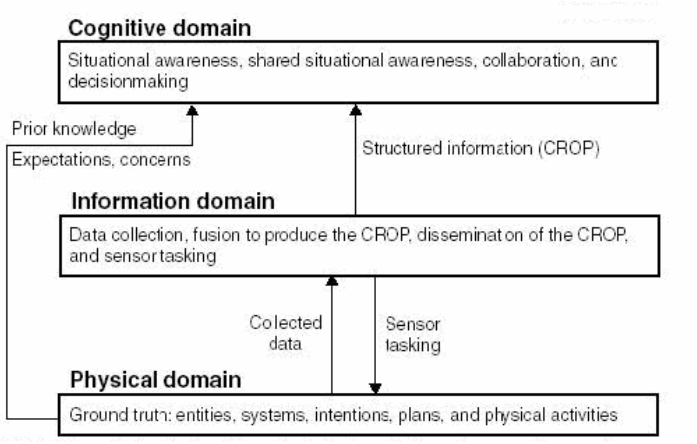

Figure 1: Information Superiority reference model (Perry et al., 2004).

Data collected in the physical domain is fused with contextual information associated with the data to produce information as a Common Relevant Operating Picture (CROP). The value added to the data, in the process increases its impact in the physical domain, as depicted in figure 2. This is because, a value add to the data has been done, namely, context. The context is not reproducible and hence has a lesser time to live. This increases the value of the information, which can be expressed as a product of data and context (Alberts, et al., 2001). The expression, $I(d) \alpha \Sigma (d * c)$, is the authors mathematical attempt to define information. The expression is explained as I (d) being an interpretation of data being proportional, directly, to the sum of all data and the context associated with it.

After an analysis procured data obtained from physical domain, information is subsequently passed onto the cognitive domain where decisions are made. Decisions in a corporate arena may translate to expense or revenue. As such, data collected can bear greater importance, since it influences crucial decisions. Any increase in the accuracy and precision of data will improve decision making leading to increased revenue and vice-versa (Alberts et al., 2001). The decisions in the cognitive domain can either be derived out of data and be based on prior knowledge and experience as a value add to it or information derived out of data and the context in which it was collected. The significance of the elements of the physical domain increases in the cognitive domain. And likewise, the impact of the elements of the cognitive domain is more than that in the physical domain. (Alberts et al., 2001; Perry et al., 2004)

A number of research studies have concluded that the physical domain consists of sensors that collect data (Alberts and Garstka, 1999; Alberts et al. , 2001; Perry et al., 2004; Parker, 1995). It could be surmised that due to the lack fusion of data and its context, the usefulness of the data cannot be realized in the physical domain. This could imply that decisions cannot be made as yet with just the data collected in the physical domain and hence has a far lesser impact than that in the other two domains. As such, the information domain attempts to fuse the data with its context and derive a meaning out of it. It is in the information domain that the usefulness of data collected by sensors in the physical domain begins to emerge. Since mismatching the context and data can be possible in the physical domain, there is an increased possibility for false positives and/or false negatives. These however may cause wrong consequences in the cognitive domain and hence the impact it bears is larger than that in the physical domain. In the cognitive domain, however, crucial decisions are made based on the information from either the information domain or from data with prior experiences and knowledge. Wrong decisions in the cognitive domain not only leads to consequences with greater significance, but also questions the validity of the output of the other two domains, namely the physical and information domains. Hence, this domain bears greater impact than the other two domains. A summation of these conclusions is displayed in Figure 2.

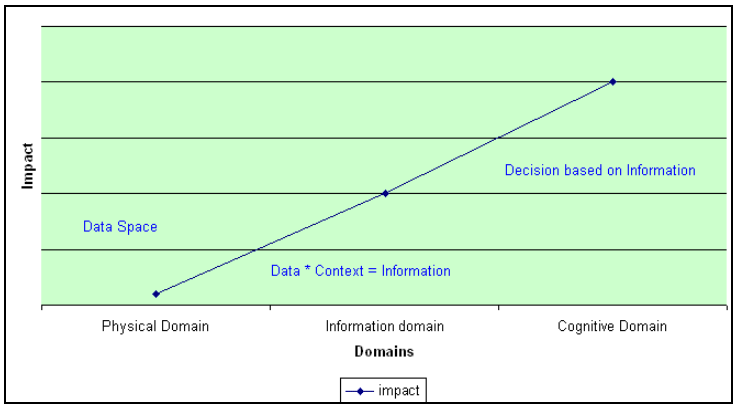

Figure 2: The impact of information in Physical, Information and Cognitive Domains.

EMPHASIS ON DATA AND DATA COLLECTION

"Risk to data is represented as the possibility of something adverse happening to data"

- (Berger, 2003)

Every organization works by the Sun Tzu principle of being informed for its survival. Organizations leverage on information assets, such as corporate spies, to be informed not only about the competitors, but also about themselves. These information assets vary from one another and are never equal in value. The same goes for the data that the assets pump into an organizations stream. The data collected, leads to information that is used in decision making. (Tzu (nd); SEI, 2005; Denning, 1999) Of the different approaches to the interpretation of information assets and security, the McCumber model, presented by John McCumber, in 1991 (McCumber, 1991), has been a popular and concise representation of information security (Maconachy et al., 2001). Figure 3 displays a visual interpretation of the McCumber model.

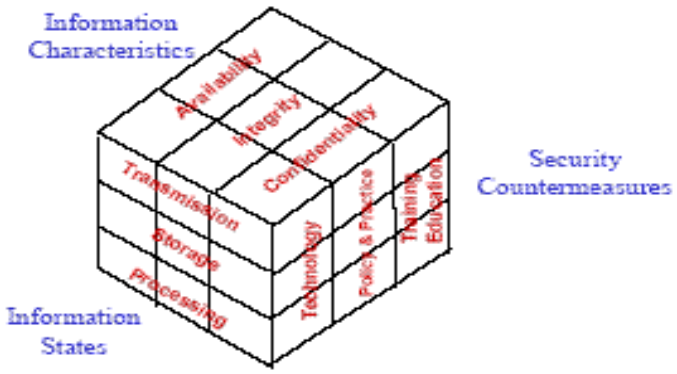

Figure 3: Original McCumber Model (McCumber, 1991).

According to the McCumber model, information state is where information is created, manipulated and shared. The information that exists in information domain may or may not reflect ground truth (McCumber, 1991). Although the McCumber model is quite concise and accepted widely, it has been a basis for further research and improvisations. The McCumber model splits information security into three domains, the Information characteristics, Security counter measures and the information states. The McCumber model revolves around information and not on the data or the data collection agents/sensors. Other models based on the McCumber, also focus on the information and not on the data which is the atomic level of the information. The model by Victor Maconachy et al., 2001 (as displayed in Figure 4); does shed more light on the Security Services domain of INFOSEC, however it does not reflect a finer level of information, data. (Maconachy et al., 2001)

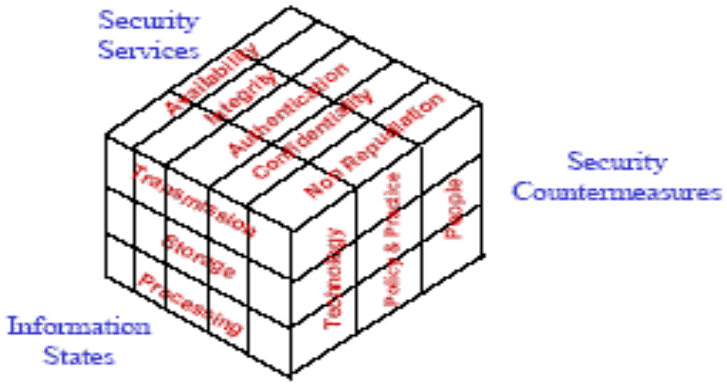

Figure 4: The Modified McCumber Model (Maconachy et al., 2001).

THE PROPOSED MODEL

This paper proposes that an improvement can be made on the McCumber model and its previously described adaptations, by including a data centric aspect. Figure 5 displays a visual interpretation of the new proposed model. This model will include the characteristics of information as applicable to data. The limitation of the McCumber model (McCumber, 1991) and that of its variant by Maconachy et al., (2001) is that no effort is taken to ensure that information gathered can be back tracked and verified (Maconachy et al., 2001; McCumber, 1991). The following paragraphs outline how the improvements can be made on the McCumber model and its previously described adaptations by justifying the need to include a process of verification and back tracking of collated information.

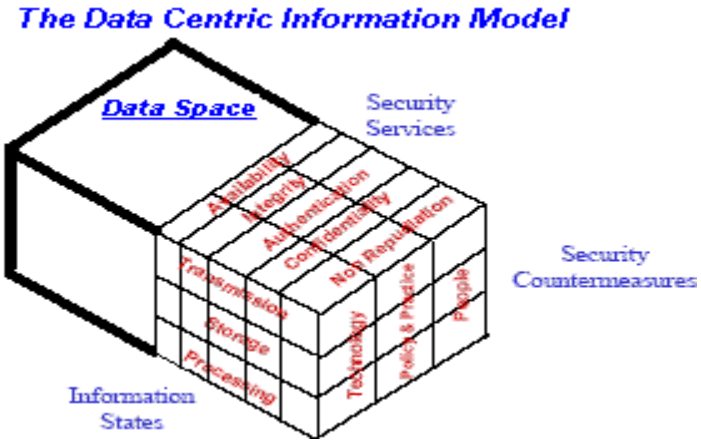

Figure 5: The proposed Model based on modification to McCumber Model.

Any verification of information must apply and evaluate the context in which the data concerned has been interpreted. This proposed variation of the McCumber model aims to explore other possible contexts that can influence a certain interpretation, by including a data component. This data component can re-enforce the validity of the information and also factor out other interpretations of the information. This may add clarity to the information and to the decision maker (Grabner-Krauter, 2002; Perry et al., 2004). Further, decision makers can base their decisions on information which can be dug deeper into and analyzed. Since sometimes, accurate and precise information can also lead to wrong decisions, in such circumstances, the interpretation and the context applied, can be re-evaluated as a learning process, so that in similar situations, the judgment can be better and hence productivity increased (Grabner-Krauter, 2002).

Research by (Perry et al., 2004) has indicated that the tasks mentioned above can be easily performed with a well defined protocol, which will lay down every aspect of data collection and interpretation. For instance, if the data collectors are non-human contact sensors, then a protocol can define the sensitivity of the sensors. Further, the sensors can also be made to transmit the data, which will be logged into a central storage. Parameters, like pressure, temperature and other physical parameters, which may influence the sensor or its operation, must also

be logged into the system. During the interpretation of the data, the interpretation systems must log the various contexts possible for interpretation, including the reason for accepting or rejecting a set of contexts. All decisions must be logged with the information and the added contexts which influence a decision. This knowledge must be available to the decision makers or decision evaluators. The availability of all the parameters influencing a decision makes it possible for an effective learning process, which may be used to judge a decision or enhance it.

Such reverse engineering is not uncommon especially when decisions made bring wrong results. In spite of this, today, reverse engineering is possible only down to the level of information and not to that of the data or the data collecting sensors. Further, human errors in interpretations most often pass by without being questioned, due to insufficient or inaccurate protocol or its enforcement. (Anderson, 2001; Arnborg et al., 2000; Perry et al., 2004; Schou and Trimmer, 2004)

The Security state in the McCumber model is responsible for ensuring that information is available at all times, with the loss of integrity. In addition, the security state ensures the authenticity of the information and access to the same by authorized persons or systems (McCumber, 1991). Work by (Parker, 1995) has demonstrated the significance of utility and possession of information by indicating that more elements of information security may be involved in addition to those outlined in the McCumber model. The comprehensive list of threats to information security includes "threats to availability and usefulness; threats to integrity and authenticity; and threats to confidentiality and possession;"

A closer examination of the description of these threats, outlined by (Parker, 1995), reveals that the same can be applied not only to information, but also to data. However, managing data and context separately eliminates some of the threats. For instance, if data and context are separate and made available, the usefulness of either one of them separately is not as much as them together. The integrity and authenticity of data can be easily ascertained than that of information, making it easier to eliminate the threat to the same. (Parker, 1995) indicates that the threat to integrity and authenticity can be of the form:

Enter, use or produce false data: When focussing on securing data, rather than information, this threat however has a high impact. Suitable security measures like encryption based on strong cryptographic algorithms are required to secure data (Perry et al., 2004; Ryutov et al., 2003; Alberts et al., 2001; Alberts and Garstka 1999; Anderson, 2001; Parker, 1995).

Modify, replace or reorder: Data and context are as much susceptible to modification, replacement or re-ordering. Such intrusions must be expected by the information security analyst when securing an information system and authorization mechanisms with string cryptographic algorithms must be installed to not only prevent intrusion, but also to detect the same. When intrusion(s) are detected, suitable corrective measures must be taken to ensure data corruption is not elaborate and propagated to the decision making systems/processes (Perry et al., 2004; Ryutov et al., 2003; Alberts et al., 2001; Alberts and Garstka 1999; Anderson, 2001; Parker, 1995).

Repudiate: The repudiation of information, as alluded to by Parker (1995), is a challenge for information security experts. However, when information is separated from its context as data, the repudiation of data can only be authenticated by the data collectors or collection systems. If the data collection processes is protected by methods mentioned above, and are tamper proof, then the repudiation of data is impossible. To further protect the same, string cryptographic algorithms may be used, together with digital signatures, to advocate the correctness of data (Perry et al., 2004; Ryutov et al., 2003; Alberts et al, 2001; Alberts and Garstka, 1999; Anderson, 2001; Parker, 1995).

The data space in the Security state would separate data and the context that is used to derive information. Hence, any adversary laying hands on either data or the context cannot really benefit. This would result in a requirement for both data and the context of the correct information. Hence protecting data and the context rather than information itself will increase the security by two fold. Further, since data can be made available at all times, those requiring only data for use within own contexts may benefit.

MODEL GAINS OVER CORPORATE ESPIONAGE

A number of examples exist of how compromise or breaches of information security has affected corporate strategies. For example, Denning 1999, reports on a article in the Forbes magazine, in April 1998, telling how an IBM Chief executive officer, shortly after taking over a top spot at Big Blue, sets up a team as a part of an extensive human intelligence network to gather information about the competitor. The information procured is placed in a central database accessible to 450 top executives of the firm. The information is perhaps used in critical decision making by the firm and thus the correctness and reliability of the information obtained is crucial. Wrong information in the database accessible to the executives will impact the decisions made and thus

will impact the firm as a whole (Denning, 1999). In another example, General Motors, in 1996, sued Volkswagen and won $100 million in damages for purchasing information from defecting employees of General Motors. (Denning, 1999) In 1994, an employee of a Colorado based software firm, sold $1 million worth software to a competing firm (Denning, 1999). Other breaches are regularly reported by news agencies. In a CNET news article, (Lemos, 2005), reported one the negative consequences of recent data leak from PayPal. Although PayPal acknowledges the leak and has implemented corrective measures since, the consequence of the same, is not "minimal" as stated by PayPal spokeswoman (Lemos, 2005). Further breaches have been reported by (McCullagh, 2005) of the unintended consequences of information leaks from the Internet portals, such as Yahoo. Such portals allow users to customize the way in which information is presented. Further reports by Fantuzzi, (2005) reports of information leaks from the Pentagon have been suggested to have resulted in international tensions between United States of America and Italian governments. Such leaks in governments have recognized the need to curb such information leaks in an attempt to improve national security and diplomatic relations (Hoekstra, 2005).

The incidents detailed above provide an insight into what a company goes through in the information and security states of the McCumber model. Information gained and stored in a company can be easily accessed, since it is "available", which opens a security hole. Willing and un-suspecting employees sometimes, give away the information, sometimes underestimating the damaging outcome of the same (Denning, 1999; McCumber, 1991). With the model proposed in this paper, employees may leak either data or context, but never both. Any access to data and context will be logged, through a protocol. This ensures that "information" perse does not reach the wrong hands, and in the event that the information does leak, a quick recovery can be done, since all accesses are logged and hence detected. This improves the security, but does not tighten it up to the extent that leaks can be avoided, which perhaps can never be done in real time.

CONCLUSION

A bad source of data can cause more harm than good since information is collated data in context. The key to, information superiority is not only filtering out bad information and exploit the good ones, but also applying the principles of information to data and context that lead to information. Data has a larger role in the information domain. It increases the value of the information, which can be expressed as a product of data and context. Data collected bears greater importance, since it influences crucial decisions. Data is also at times used in decision making with support from prior experiences and commonly known contexts, as perceived by (Perry et al., 2004). Complete security can be attained only as security to protect data is approached. The data space in the model proposed participates not only in the Information state, but also the security state, enhancing data protection and derivation of valuable information. Since data is the atomic representation of information, finer grained security measures need to be taken to protect data.

REFERENCES

Alberts, C., Dorofee, A., Stevens, J., & Woody, C. 2003. Introduction to the OCTAVE Approach. Carnegie Mellon University. Pittsburg, USA.

Alberts, D. S., & Garstka, J. 1999. *Information Superiority and Network Centric Warfare*. Retrieved April, 2005, from http://www.iwar.org.uk/iwar/resources/info-superiority1999/index.htm

Alberts, D. S., Garstka, J. J., Hayes, R. E., & Signori, D. A. 2001. *Understanding Information Age Warfare*: CCRP Publications, Washington, DC

Anderson, R. 2001. Why Information Security is Hard - An Economic Perspective. Paper presented at the *Proceedings 17th Annual Computer Security Applications Conference*. IEEE Computer Society, Los Alamitos, California.

Arnborg, S., Brynielsson, J., Artman, H., & Wallenius, K. 2000. Information Awareness in Command and Control: Precision, Quality, Utility. Paper presented at the *Third International Conference on Information Fusion (FUSION 2000.* Vol. 2, pp THB1/25- THB1/32. IEEE Computer Society. Paris, France.

Berger, B. 2003. Data-Centric Quantitative Computer Risk Assessment. Retrieved April, 2005, from http://www.sans.org/rr/whitepapers/auditing/1209.php

Center for American Progress. 2004. *Bush Legacy on Iraq: Misinformation and False Pretense*. Retrieved 10/6/2005, 2005, from http://www.americanprogress.org/site/pp.aspx?c=biJRJ8OVF&b=19351 & printmode=1

Denning, D. E. 1999. *Industrial Espionage*. Retrieved March, 2005, from http://infosecuritymag. techtarget.com/articles/1999/aprilcover.shtml

Fantuzzi, J. 2005. *Document security? Tell me another joke*. Retrieved 10/10/2005, 2005, from http://news.com.com/Document+security+Tell+me+another+joke/2010-1071_3-5783062.html?tag=st.ref.goo

Grabner-Krauter, S. 2002. The role of consumers' trust in online-shopping. *Journal of Business Ethics*. Vol. 39, pp 43-50.

Hoekstra, P. 2005. *Secrets and Leaks: The Costs and Consequences for National Security*. Retrieved 10/10/2005 from http://www.heritage.org/Research/NationalSecurity/hl897.cfm

Hutchinson, W., & Warren, M. 2001. Principles of Information Warfare. *Journal of Information Warfare*, Vol. 1, No. 1 pp 1-6.

Lemos, R. 2005. *Data leak puts PayPal users at phishing risk*. Retrieved 10/10/2005, 2005, from http://news.com.com/Data+leak+puts+PayPal+users+at+phishing+risk/2100-1029_3-5550046.html

Maconachy, V. W., Schou and Trimmer, C. D., Ragsdale, D., & Welch, D. 2001. A Model for Information Assurance: An Integrated Approach. Paper presented at the *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security.* United States Military Academy. IEEE Computer Society, West Point, NY.

McCullagh, D. 2005. *Navigating the law of unintended consequences*. Retrieved 10/10/2005, 2005, from http://news.com.com/Navigating+the+law+of+unintended+ consequences/2010-7348_3-5611746.html

McCumber, J. 1991. Information Systems Security: A Comprehensible Model. Paper presented at the *Proceedings of 14th National Computer Security Conference*. National Institute of Standards and Technology, Baltimore, MD.

McCumber, J. 1991. Information Systems Security: A Comprehensible Model. Paper presented at the *Proceedings of 14th National Computer Security Conference.* National Institute of Standards and Technology, Baltimore, MD.

Parker, D. B. 1995. Using Threats to demonstrate the elements of Information Superiority. Paper presented at the *European Convention on Security and Detection.* Brighton, UK

Perry, W., Signori, D., & Boon, J. 2004. *Exploring Information Superiority: A Methodology for Measuring the Quality of Information and Its Impact on Shared Awareness*. Retrieved 6/10/2005, 2005, from http://www.rand.org/publications/MR/MR1467/MR1467.pdf

Ryutov, T., Neuman, C., & Kim, D. 2003. Dynamic Authorization and Intrusion Response in Distributed Systems. Paper presented at the *DARPA Information Survivability Conference and Exposition (DISCEX'03).* Vol. 1, p. 50, South Carolina, USA.

Schakowsky, J. 2003. *Schakowsky: Bush Administration's misstatement of the day – Iraqi weapons on mass destruction.* Retrieved 6/10/2005, 2005, from http://www.house.gov/schakowsky/press2003/pr10_01_2003miswmd.html

Schou, C. D., & Trimmer, K. J. 2004. Information Assurance and Security. *Journal of Organizational and End User Computing*, Vol. 16, No . 3, pp 1.

SEI, 2005. *Principles of Survivability and Information Assurance*. Retrieved April, 2005, from http://www.cert.org/archive/pdf/SIAPrinciples.pdf

Tzu, S. *(nd) Art of War.* 167 pages. Published by Cloud Hands Press, U.S.A.

Whitney, C. R. 2005. *The WMD Mirage: Iraq's Decade of Deception and America's False Premise for War* PublicAffairs Publishing.

COPYRIGHT