# Data Loss Prevention Management in Healthcare Enterprise Environments

**Dr. Manghui Tu, Purdue University Calumet**

Manghui Tu, assistant professor of Computer Information Technology, Purdue University Calumet, USA. He received his Ph.D. degree of computer science from the University of Texas at Dallas in December 2006. His research interests include distributed computing, information security, and computer forensics.

**Ms. Kimberly Lynn Spoa-Harty, Purdue University Calumet Graduate Student**

System Engineer and Architect, work with security standards for desktop standardization and implementation. Experience is over 10 years in Information Technology. Currently working towards a Master of Science at the School of Technology.

©American Society for Engineering Education, 2014

# Data Loss Prevention Management in Healthcare Enterprise Environments


## Abstract

As healthcare data are pushed online, consumers have raised big concerns on the breach of their personal information. Law and regulations have placed businesses and public organizations under obligations to take actions to prevent data breach. Among various threats, insider threats have been identified to be a major threat on data loss. Thus, effective mechanisms to control insider threats on data loss are urgently needed. The objective of this research is to address data loss prevention challenges in healthcare enterprise environment. First, a novel approach is provided to model internal threat, specifically inside activities. With inside activities modeling, data loss paths and threat vectors are formally described and identified. Then, threat vectors and potential data loss paths have been investigated in a healthcare enterprise environment. Threat vectors have been enumerated and data loss statistics data for some threat vectors have been collected. Issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed.


## 1. Introduction

The advances in internet technologies, the proliferation of mobile devices, and the development of electronic healthcare records, have driven healthcare services online and ubiquitous to provide convenience and flexibility to users and patients[14, 31]. However, due to the untrustworthy internet environment and sophisticated healthcare service and business processes involved, healthcare sector faces severe challenges on securing protected healthcare information[9, 10, 12, 14, 31]. Over the past few years, millions of sensitive data records in healthcare and other private and public sectors were exposed[10, 12, 14, 20, 24] and has resulted in substantial financial and operational loss, which greatly hurts the confidence of customers, business partners and stakeholders[10, 13, 14, 25]. The average total cost per data breach has risen to $7.2 million or $214 per record lost [10] and the estimated total cost of data breach in healthcare industry is 6 billion dollars annually[14]. According to RSA's annual Consumer Online Fraud Survey, more consumers are concerned, more than ever before, about online private businesses and public organizations putting their data at risk[13]. Meanwhile, over the last two decades, government and industry bodies around the world have issued many laws and regulations such as *HIPAA* 1996, to ensure the security, integrity, and confidentiality of healthcare records, business data, and healthcare IT infrastructures. These mandates have placed businesses and organizations in healthcare industry under obligations to undertaken programs to ensure the compliance with these laws and regulations. Therefore, securing protected healthcare information and healthcare IT infrastructures to prevent data loss including data breach and industrial espionage is critical.

Many intrusion/fraud prevention, detection, and tolerance mechanisms, including security policy establishment, trust management, access control, authentication, and other data and communication security protection techniques, have been deployed by private businesses and public organizations to secure their IT infrastructures and stored sensitive data[11, 24]. However, data loss incidents have increased in the past few years due to various reasons that can be contributed to different parties in the security protection chain[5, 10, 11, 13, 14, 26]. Among them,

insider threats have been shown to be a major factor. Insider threats have the most significant impact on information and IT infrastructure security, and have recently been recognized by both private and public sectors[10, 14, 15, 16, 21]. Over the past few years, millions of sensitive data records were breached or stolen due to unintentional user mistakes or intentional fraudulent user activities[7, 15, 20, 24]. Hence, effective mechanisms to control insider threats to prevent data loss are necessary[10, 16, 19, 28]. However, there are a few challenges that need to be addressed. First, even though the general theory of threats and attacks may be well understood, the specific vectors and paths of data loss and the insight view of inside threats in healthcare industry remain unclear due to the high complexity of the healthcare information systems and business processes. Second, since the insiders have all the needed access privileges, a successful inside activity does not need to conduct malicious attack against the healthcare information system, thus, traditional intrusion detection techniques may not work to successfully detect or identify an inside data loss incident in a healthcare business.

**The overall objective** of this research is to address data loss prevention challenges in healthcare enterprise environment. This paper will provide a novel approach to model internal threat, specifically inside activities, to provide a genuine method that can formally describe inside activities. It is expected that with inside activities modeling, data loss paths and threat vectors can be formally described and provide a guidance to identify potential inside activities threat vectors and data loss paths. We then investigate the threat vectors and potential data loss paths in a healthcare enterprise environment. Threat vectors will be enumerated and data loss statistics for some threat vectors have been collected. Finally, issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed.

## 2. System Model

An abstraction of classical healthcare enterprise environment is modeled as a multi-tier system that consists of multiple data access or management parties, including a data module, service providers, business users, data management team, and client party. The overview of the system is shown in Fig. 1. The data module is a central and critical part of this architecture and it consists of a data storage system, a data process module, and a data access module. The data storage module is essentially databases and files that contain the information to be protected. The data process module is composed of a healthcare information system which process the information stored in the data storage module for business clients, patients, government agencies, and healthcare service providers. The data access module is essentially web based interfaces for users. The service provider party is composed of different services that are provided by the healthcare business, for example, hospital services, lab test services, patient health prevention care services, disease diagnosis and treatment services, nursing, etc. These services involve many human users such as doctors, technicians, and nurses. The business party is essentially the interactions between the healthcare business entity and other entities such as other healthcare service providers, government agencies, insurance companies, healthcare equipment/pharmacy providers, etc. The data management party in this research is essentially the IT teams including database administrators, web administers, network administers, computer system administers, and the security/compliance team. The client party is essentially the patients and their legal guardians.

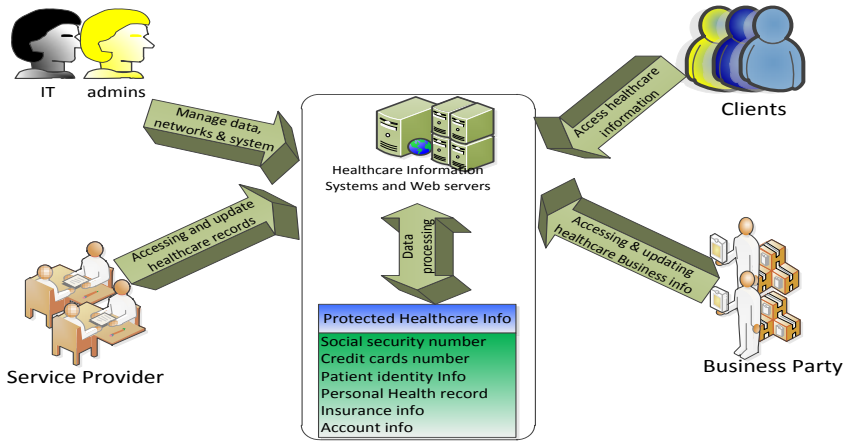


Fig.1 An architecture overview of the healthcare enterprise system.

In this research, the data items to be protected by using data loss prevention mechanisms include personal health information (PHI) such as social security numbers (SSNs), data of birth, payment data, insurance policy information in digital format, and personal electronic health records (EHRs), as well as business data such as business client information. In this architecture, the data module is interacted with other modules, for example, databases and file systems are managed by the database administers, protected by system and network administrators as well as the security/compliance team. The service module and the business module will not only read the information stored in the data module, it will also create and update the records in the data module. In most cases, the client module will only read the information stored in the data module such as patient's health record, payment information, etc.

## 3. Insider Activity Modeling

An insider with the desired privilege does not need to conduct any malicious activity (or attack) to obtain the privilege to access sensitive assets. In a well-managed healthcare enterprise environment, appropriate security policy and acceptable user policy should be in place and enforced by policy based access control[6, 8, 14, 17]. Suppose that there are $L$ users $U = \{u_1, u_2, \ldots, u_L\}$ and $M$ work roles $W = \{w_1, w_2, \ldots, w_M\}$ in an enterprise environment. Each user $u_i$ should have a well-defined work role $w_i$ and been assigned with data access privilege based on the *need-to-know* principle. The healthcare business should identify its own set of sensitive assets $D = \{d_1, d_2, \ldots, d_N\}$, for example, user accounts, computer and network resources, protected healthcare information, etc.. Also, the healthcare data can be classified into different sensitive levels, and the set of sensitive levels is denoted as $S$, where $S = \{s_1, s_2, \ldots, s_m\}$. Each data object in $D$, $d_i$, is assigned a sensitive label $s_j$. A data item that is labeled as $s_i$ has higher sensitive level than a data item that is labeled as $s_j$ if $i > j$. To fulfill tasks defined by the work role, a user accesses sensitive assets with certain preference. Let $O = \{o_1, o_2, \ldots, o_Z\}$ denote the set of $Z$ access operations, for example, read, write, execute, delete, shutdown, print, copy, etc. Let $A$ denote the set of preference level where $A = \{a_1, a_2, \ldots, a_n\}$, then the sensitive access preference can be defined by a set of 2-tuples, $(d_i, a_j)$. A data item with access preference $a_i$ is accessed with higher frequency than a data item with access preference $a_j$ if $i > j$, and $a_1$ is defined as the lowest access preference, e.g., zero access. Some operations will be performed regularly, and some should rarely be performed or may never be performed. For example, an IT system

administrator may have to copy and move sensitive data objects around but should not delete or modify sensitive data objects; an application developer will need to modify sensitive data objects a lot but should not copy the data objects to a personal USB device. Therefore, a user's accesses to sensitive data objects in the healthcare enterprise environment can be modeled into certain pattern that can be defined by the 4-tuple {*W, D, O, A*} based on the work role of different users. The 4-tuple {*W, D, O, A*} data access preference model, denoted as WDOA, may give a hint on whether a data access activity is normal or not. However, the WDOA model cannot precisely determine whether an access activity is an inside activity or not. For example, application developer should have access preference $a_1$ to user account information, and any access to such data would be suspicious. An IT administrator has low access preference $a_i$ ($i > 1$) to personal healthcare records for data management purpose, and a copy access to those data items cannot determine whether such access is suspicious or not.

To reach a sensitive asset $d_i$, an insider needs to have known or unknown access paths to $d_i$[8, 22, 23] through a set of access paths. For example, to steal $d_i$ for personal use, an insider may copy sensitive data $d_i$ from the data storage site and then send to a personal USB device that has been attached to a system inside the healthcare enterprise environment. An insider may first create secret user account and setup a virtual machine (VM) instance, and then copy sensitive data $d_i$ to the VM instance to be accessed later. Let $P = \{p_1, p_2, \ldots, p_K\}$ denote the set of $K$ access paths in the healthcare enterprise environment, where $p_i$ is one specific access path such as USB access, CD access, VM instance access, email access, etc. Let $Path(\{u_i\}, k)$ denote the set of access paths to data $d_k$ by a subset of users, $\{u_i\}$, in $U$, then $Path(\{u_i\}, k)$ can be defined by a set of 4-tuples ($u_i$, $o_j$, $d_k$, $p_m$,). With the 4-tuples ($U$, $O$, $D$, $P$) data access path modeling, denoted as UODP, it is possible to identify whether an access activity is an inside activity or not. An application developer $u_i$ copy healthcare records $d_k$ to a un-monitored VM instance, the 4-tuples (application developer, copy, un-monitored *virtual machine*, $d_k$) can be definitely considered as a suspicious inside activity since an un-monitored VM instance is beyond the control of the healthcare enterprise and can lead to data loss of $d_k$. While an IT system administrator $u_i$ copy healthcare records $d_k$ to a monitored USB device, the 4-tuples (IT system administrator, copy, monitored USB, $d_k$) should not be considered as a suspicious activity since the monitored USB device is still under the control of the healthcare enterprise and will not lead to data loss of $d_k$ at the current stage. With sufficient resources, the elements in $U$, $O$, $D$ can be well classified and identified based on current technologies. However, due to the complexity of the healthcare information system, data storage techniques, user access controls, and usage obligations, the identification and classification of access paths is still challenging to for healthcare enterprises.

## 4. Data Loss Results and Analysis

Based on the UODP access model, if a legitimate user accesses and operates sensitive data through an uncontrolled access path $p_i$, it can lead to potential data loss. The combination of such uncontrolled access paths and access operations are threats to data loss, and are defined as the data loss threat vectors. Therefore, to prevent and control data loss in healthcare enterprise environment, the first critical task is to identify the set of data loss threat vectors, more specifically, the set of uncontrolled access paths. In this research, we will explore the potential threat vectors in the healthcare enterprise environment. Safend Data Protection Suite, an end point security product from Wave, has been used in this research to regulate data loss prevention in an enterprise healthcare environment. Data loss results are collected before and after the

placement and enforcement of end point security protection. To identify data loss threat vectors, examine potential data loss threats, and to analyze potential data loss controls, the following studies will be conducted. First, potential threat vectors will be enumerated and feasible operation controls will be listed for each threat vector. The status of the enforcement of such controls is also indicated. Second, statistical data loss prevention results are provided.

### 4.1 Data Loss Threat Vectors Identification

Data loss threat vectors can be categorized as external storage media and transmission media. External hard disks, USB flash drives, PDA's, CD/DVD, floppy disks, and tapes are traditional storage media, while cell phones, SD card readers, IPAD, FTP, web sites, and printing can be categorized as transmission media. The only exception is cloud storage which is a new technology combining transmission and storage. To control operations on sensitive data, port controls such as block, allow, force encryption, set to read only are enforced. Data filtering technologies based on expressions are deployed to filter sensitive data such as credit card numbers, social security numbers, healthcare records, etc.

| **Threat Vectors** | **Port Control Options** | **Enforcement Status** |
|---|---|---|
| External Hard Drives | Block/Allow/Force Encryption/Set To Read Only | Enforced |
| USB Flash Drives | Block/Allow/Force Encryption/Set To Read Only | Enforced |
| Cell Phones | Block/Allow | Not Implemented |
| PDA's | Block/Allow | Enforced as external storage media |
| SD Card Readers | Block/Allow/Set To Read Only | Not Implemented |
| iPad | Block/Allow | Not Implemented |
| CD/DVD | Block/Allow/Force Encryption/Set To Read Only | Enforced |
| Floppy Drives | Block/Allow | No data to report - technology in environment does not allow for floppy drives |
| Tape Drives | Block/Allow | No data to report - technology in environment does not allow for tape drives |
| Websites | none | Due to product, high administration efforts to identify and analyze risks. |
| FTP | none | Blocked by perimeter within the domain, by both static IP, site IP allowed to use, and user security - 3 factor authentications. |
| Cloud Storage | none | Not Implemented |
| Email | none | Email filtering, algorithms to look for sensitive data, will force encryption |
| Printing | can block physical printers from connecting, but not network printers | Not Implemented |

Table 1: the enumeration of data loss threat vectors in an enterprise healthcare environment.

As indicated in table 1, traditional storage media are usually well controlled by enforcing port controls, since they have been well documented and the monitoring and control technologies have been well designed. Cloud storage is a new technology and not well documented[1, 2, 3], thus, mature control technologies are not ready yet. Some transmission media such as FTP can easily be controlled since FTP can easily be replaced with an alternative secure technology. It means that these technologies are not required to accomplish healthcare activities and thus can be blocked. Some other transmission media such as printing are not easily controlled since printing

is required for routine business activities. Also, due to the nature of printing (graphical presentation of information), sophisticated identification and examine technologies are needed to filter sensitive data. Currently, efficient deployment of such technologies has not been ready yet.

### 4.2 Data Loss Analysis

A 90-day time period data collection is conducted prior to the deployment of any end point security protection technology (denoted as /P). After the 90-day time period, Safend Security Protection Suite was deployed in the enterprise healthcare environment to control data loss. Then, a 90-day time period data collection is conducted with the deployment of the end point security protection technology (denoted as /A). Due to the limitation of the technology and the feasibility of the policy enforcement in the enterprise environment, only part of the threat vectors, USB, CD/DVD, external hard disk, and phone, are controlled.

| Threat Vector | # Users/P | # Users/A | # Files/P | # Files/A | Data Size/P | Data Size/A |
|---|---|---|---|---|---|---|
| USB | 2765 | 413 | 4449429 | 374015 | 1123G | 432.4G |
| CD/DVD | 157 | 44 | 212067 | 8530 | 291.7 G | 76.5G |
| External Hard Disk | 161 | 21 | 443805 | 9356 | 804.39 G | 2.4G |
| Phone | 426 | 5805 | 0 | 0 | 0 | 0 |

Table 2: The potential data loss path accesses and operations before and after the deployment of Safend.

As indicated in Table 2, the number of users that access potential data loss threat vectors, such as USB, CD/DVD, and external hard disks have been significantly reduced (USB users from 2765 to 413, CD/DVD users from 157 to 44, external hard disk users from 161 to 21). The only exception is the use of phone and the usage has been significantly increased (from 426 to 5805). One reason could be the block or reluctance of the use of email and other controlled communication paths. However, users may plug in for charging of devices without proper removable media protection, thus, phones can be used to take pictures which can then be transmitted out without control. Therefore, such an abrupt increase needs to be carefully analyzed and better to conduct a thorough investigation. A countermeasure to such data loss threat through phone can be achieved to enforce non-personal phone policy in sensitive working environment. As indicated by the number of files and the size of files moved around in Table 2, employees tend to abuse the data loss threat vector accesses and operations without data loss control, since such significant reductions (for examples, the number of USB accessed files from 4449429 to 374015, the number of CD/DVD accessed files from 212067 to 8530, the number of external disk accessed files from 443805 to 9536) does not affect business activities in the enterprise healthcare environment. Please note that no data is transferred to phones due to the reason that most phones when connected are seen as removable storage or external hard drives, and thus will adhere to the policies already enforced for that media type. In Table 3, it indicates that a large part of accesses are encrypted, which can significantly reduce the potential of unintentional data loss due to theft, mis-sent, and misconfiguration. However, there are some accesses and operations are unencrypted but all of them are approved. As stated in the security usage policy and recorded in the usage logs, such approved usages are required to follow predesigned processes such that all files accessed and operations on files are all logged in audit reports.

| Threat Vector | # Users/A | # Files/A | Data Size/A | # Users/A | # Files/A | Data Size/A |
|---|---|---|---|---|---|---|
| | Encrypted Use | | | Unencrypted Use | | |
| USB | 187 | 247680 | 334 | 226 | 126335 | 98.4 |
| CD/DVD | 32 | N/A | 64.6 | 12 | N/A | 11.9 |
| Ext. Hard Disk | N/A | N/A | N/A | N/A | N/A | N/A |
| Phone | N/A | N/A | N/A | N/A | N/A | N/A |

Table 3: The encrypted and unencrypted data loss path accesses and operations after the deployment of Safend.

# 5 Discussion

## 5.1 Data Loss Prevention Issues

With deployment of end point security protection product such as Safend Security Protection Suite, it is possible to control data loss through traditional external storage media. For example, with appropriate access control, any data accessed can be logged and can be blocked to be moved to USB storage media or other external storage media. However, potential uncontrolled data loss access paths could still exist.

1) A combination of multiple access technologies in the extended healthcare enterprise environment. For the purpose of business trip, an employee $u_i$ with work role $w_j$ (e.g., sales representative) may need to move data (e.g., $d_n$) outside of the enterprise network by applying access operation (i.e., *copy*), and such access has a high access preference for $w_j$. Then, based on the WDOA model, the 4-tuple (sales representative, $d_n$, *copy*, high access preference), will be defined as a legitimate access without an alert. By applying UODP, it can help to detect potential data loss due to access violations. End point security product can monitor regular access to the data and any violation (e.g., *copy* $d_n$ to an unauthorized personnel USB device $p_m$) may result in an alert since the union of *copy* and $p_m$ ($copy \cup p_m$) has been pre-defined as data loss threat vector. In this way, the combination of UODP and end security protection product together can create an extended enterprise environment outside the physical enterprise network boundary. However, there will be other techniques available to bypass the control. For example, employee can photograph the data when the data is open for read, or the external storage media can be bit-by-bit imaged without leaving any evidence on the media if it is connected through write block devices.

2). Forgotten paths due to unsuccessful change management. For example, misconfiguration could be unnoticed during system update, security product update, or human resource change. With such misconfiguration, media block and media access log may not be enabled, etc.

3). A combination of uncontrolled access technologies within the physical network of healthcare enterprise environment. In Table 1, some access technologies, such as web site, phone, and cloud technology, have not been controlled. As analyzed in the above section, phone technology has the potential to result in data loss. With appropriate surveillance technology deployed, such data loss access path can be monitored and detected. Cloud technology, web site, and local virtualization technology can provide a perfect uncontrolled data loss access path. Local data can be moved between physical storage within the network and a local VM instance, which can then connect to remote private cloud storage website. With this secret path, data

encrypted in the VM instance can be moved outside the physical network boundary of the healthcare business. After the local VM instance deleted, little evidence will be left within the boundary of the healthcare enterprise network. The only feasible control is to block any encrypted traffic[32].

### 5.2 Incident Identification, Tracking, and Reconstructions

Even with data loss threat vectors identification, control, and monitoring, inside activities cannot be detected or identified with current access control techniques since the access operation and access path are both legitimate user privileges. Therefore, forensics investigation on inside activities in healthcare enterprise environment, including incident detection and reconstruction is critically needed[31]. Current research on inside threat detection and identification[7, 16, 18] and event reconstruction mechanisms[4, 29, 31] are limited in real world since they require a comprehensive set of information including social information and explicit dependence knowledge, which are not available in an enterprise environment. Hence, a novel mechanisms are critical to identify potential inside activity and reconstruct the inside activity for tracking.

To perform the job allowed for a user's work role, the operation and access path of an inside activity are allowed and cannot be prevented, thus, any individual operation on data and access path cannot identify an inside activity. One approach is to apply the UODP model and utilize the contextual information of the incident such as a joint of the operations of data access and path access (e.g., *copy* $d_i \wedge$ *access USB*) to identify a potential inside activity. Operations on sensitive assets can be labeled as safe or highly risky for each work role $w_i$, and can be defined by a tuple $\{\{W, O, D, P\}, R\}$, where $R$ defines the risk levels. Hence, a risk table containing entries of $\{\{W, O, D, P\}, R\}$ can be developed for each service. Once a risky operation (e.g., *copy* $d_i$) has been performed by user $u_i$ (with work role of $w_j$), the $u_i$'s operations will be tracked to look for an operation $p_i$ (an access path) associate with $d_i$ (*e.g., USB access* $\vee$ *CD access* $\vee$ *email access*, etc.) performed by $u_i$. If the two operations (data access $O$ or access path $P$) are discovered, then a potential inside activity is identified. Since such combinations cannot be totally prohibited and the knowledge of such combinations cannot be obtained from access control, the detection has to rely on the logged information in the system.

Another challenge in digital forensics investigation is the lack of efficient digital forensics investigation mechanisms. Huge amount of artifacts of events and operations are logged in the system, which may introduce inefficiency to internal incident tracking and reconstruction. Many of the security breaches are not investigated due to the unaffordable effort required to perform a forensics investigation[27, 30, 31]. Therefore, to improve the responsiveness and to free businesses and public organizations' burden on the incident report and investigation process, an incident reconstruction mechanism should be in place to track inside activity incident automatically. To automate the reconstruction of an inside activity incident, external contextual information is needed to correlate individual operations of such incident, which can only be learned from logged information from the networks and information systems within the healthcare enterprise environment. Therefore, mechanisms such as automatic tracking and reconstruction of a crime scene should be designed[31].

## 6 CONCLUSION

This paper addressed the data loss prevention management problem in healthcare enterprise environment. First, a novel approach is provided to model inside activities and a UODP inside activity modeling mechanism is proposed. With inside activities modeling, data loss paths and threat vectors are formally described and identified. Second, threat vectors and potential data loss paths have been investigated in a healthcare enterprise environment. Threat vectors have been enumerated and data loss statistics results for some threat vectors have been collected and analyzed. Finally, issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed.

## References


[1] Biggs, S. and Vidalis, S. (2010). Cloud Computing Storms: *IJICR* **1**(1), pp. 61-68.

[2] Bruening, P. J. and Treacy, B. C. (2009). Cloud computing: privacy, security challenges. *Privacy & Security Law Report by The Bureau of National Affairs, Inc*. [online]. Available: http://www.bna.com.

[3] Brunette, G. and Mogull, R. (2009). *Security Guidance for critical areas of focus in Cloud Computing V2. 1*. CSA (Cloud Security Alliance), USA. [online]. Available: http://www.cloudsecurityalliance.org/guidance/csaguide.

[4] Case, A. Cristina, A., Marziale, L., Richard G., & Roussev, V. (2008). FACE: automated digital evidence discovery and correlation. *Digital Investigation*. 5, pp.s65-s75 .

[5] CENZIC. (2008) Q1 Cenzic application security trends report. [online]. Available: http://www.cenzic.com/downloads/Cenzic_AppSecTrends_Q3_Q4-2008.pdf.

[6] Chen, P., Laih, C., Pouget, E. and Dacier, M. (2005). Comarative survey of local honeypot sensor to assist network forensics. In *Proceedings of the 1st International Workshop on Systematic Approach to Digital Forensics Engineering*, pp. 120-132.

[7] Eberle, W. and Holder, L. (2009). Insider threat detection using graph-based approaches. In *Proceedings of IEEE Cybersecurity Applications & Technology Conference for Homeland Security* (*CATCH*), pp. 237-241.

[8] Ellard, D. and Megquier, J. (2004). DISP: practical, efficient, secure and fault-tolerant distributed data storage. *ACM Transactions on Storage*. **1**(1).pp. 71-94.

[9] El Emam K. et al., (2010). The inadvertent disclosure of personal health information through peer-to-peer file sharing programs. *J. American Medical Informatics Assoc.*, **17**(2), pp. 148–158.

[10] Ernst & Young. (2011). Data loss prevention: keeping your sensitive data out of the public domain. White Paper. [online]. Available: https://www.watchguard.com/tips-resources/grc/wp-data-loss-prevention.asp.

[11] Fratto, M. (2008). Security survey: we're spending more, but data's no safer than last year. [online]. Available: http://www.informationweek.com/news/security/management/showArticle.jhtml?articleID=208800942.

[12] Halbesleben, J.R.B, Wakefield, D.S. and Wakefield, B.J. Work-arounds in healthcare settings: literature review and research agenda. *Health Care Management Rev.*, **33**(1), 2008, pp. 2–12.

[13] Hoffman, P. (2007). RSA security reports low level of trust in online banking security. eWeek News. [online]. Available:http://www.eweek.com/c/a/Security/RSA-Survey-Reports-Low-Level-of-Trust-in-Online-Banking-Security/.

[14] Johnson, M. E, and Willey, N. (2011). Usability failures and healthcare data hemorrhages. *IEEE Security and Privacy*. Issue March/April 2011, pp. 18-25.

[15] Kowalski, E., Conway, T., Keverline, S., Williams, M., Cappelli, D. and Moore, A. (2008). Insider threat study: illicit cyber activity in the government sector. [online]. Available: *http://www.cert.org/insider_threat/*.

[16] Moore, A., Cappelli, D. & Trzeciak, R. (2008). The "big picture" of insider IT sabotage across U.S. critical infrastructures. *Advances in Information Security*. **39**, pp. 17-52 .

[17] Murphey, R. (2007). Automated windows event logs forensics. *Journal of Digital Investigations*. **4S**. pp. S92-S100.

[18] Phua, C., Lee, V., Smith, K. and Gayler, R. (2007). A comprehensive survey of data mining-based fraud detection research. [online]. Available: http://www.bsys.monash.edu.au/people/cphua/.
[19] Popovsky B. and Frincke, D. (2004). Adding the fourth "R". *In Proceeding of the 2004 IEEE Workshop on Information Assurance.* pp.442-443.
[20] Ramzan, Z. (2008). Security trends of 2008 and predictions for 2009. *Net Security News*, [online]. Available: http://www.net-security.org/article.php?id=1194. Dec. 24.
[21] Randazzo, M. Keeney, M., Kowalski, E., Cappelli, D. and Moore, A. (2004). Insider threat study: illicit cyber activity in the banking and finance sector," [online]. Available: *http://www.cert.org/insider_threat/*, 2004.
[22] Rozinat, A. van der Aalst, W., Dustdar, S., Fiadeiro, J. and Sheth, A. (2006). Decision mining in ProM. In: *Lecture Notes in Computer Science*. **4102**. Springer, Berli
[23] Rozinat, A., Mans, R., Song, M. and van der Aalst, W. (2008). Discovering colored petri nets from event logs. *International Journal on Software Tools for Technology Transfer*, **10**(1).
[24] RSA Security. (2008). CSI computer crime & security survey. [online]. Available: http://i.zdnet.com/blogs/csisurvey2008.pdf .
[25] Seltxer, L. (2006). Is online banking too dangerous? *eWeek News*. [online]. Available: http://www.eweek.com/c/a/Security/Is-Online-Banking-Too-Dangerous/.
[26] Shah, A. (2009). More employees neglecting data security, survey says. [online]. Available: http://www.networkworld.com/news/2009/061009-more-employees-neglecting-data-security.html. *IDG News Service*.
[27] Sheyner, O., Haines, J., Jha, S., Lippmann, R. and Wing, J. (2002). Automated generation and analysis of attack graphs. *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 273-284.
[28] Siponen, M. and Oinas-Kukkonen, H. (2007). A review of information security issues and respective research contributions. *Database for Advances in Information Systems*. **38**(1), pp.60-80.
[29] Tang, Y. and Daniels, T. (2005). A simple framework for distributed forensics. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops*, pp.163-169.
[30] Todtmann, B., Riebach, S. and Rathgeb, E. (2007). The honeynet quarantine: reducing collateral damage caused by early intrusion response. In *proceedings of the 6th international Conferenc*e on Networking, pp.464-465.
[31] Tu, M., Xu, D., Butler, E., and Schwartz, A. (2012). Locating and identifying forensic evidence for attacks against online business information systems by using honeynet. *Journal of Digital Forensics, Security, and Law*. **7**(4), pp 73- 97.
[32] Wippich, B. (2007). Detecting and preventing unauthorized outbound traffic. White Paper, *SANs Institute Reading Room*. [online]. Available: https://www.sans.org/reading-room/whitepapers/detection/detecting-preventing-unauthorized-outbound-traffic-1951.