THE JOURNAL OF
DIGITAL FORENSICS,
SECURITY AND LAW

Journal of Digital Forensics,
Security and Law

Volume 10 | Number 1 Article 3

2015

Data Loss Prevention Management and Control: Inside Activity Incident Monitoring, Identification, and Tracking in Healthcare Enterprise Environments

Manghui Tu
Purdue University Calumet

Kimberly Spoa-Harty
Purdue University Calumet

Liangliang Xiao
Frostburg State University

Recommended Citation

Tu, Manghui; Spoa-Harty, Kimberly; and Xiao, Liangliang (2015) "Data Loss Prevention Management and Control: Inside Activity Incident Monitoring, Identification, and Tracking in Healthcare Enterprise Environments," *Journal of Digital Forensics, Security and Law*: Vol. 10 : No. 1 , Article 3.
DOI: https://doi.org/10.15394/jdfsl.2015.1196
Available at: https://commons.erau.edu/jdfsl/vol10/iss1/3

DATA LOSS PREVENTION AND CONTROL: INSIDE ACTIVITY INCIDENT MONITORING, IDENTIFICATION, AND TRACKING IN HEALTHCARE ENTERPRISE ENVIRONMENTS

Manghui Tu
Department of Computer Information Technology & Graphics
Purdue University Calumet

Kimberly Spoa-Harty
Department of Computer Information Technology & Graphics
Purdue University Calumet

Liangliang Xiao
Department of Computer Science and Information Technologies
Frostburg State University

ABSTRACT

As healthcare data are pushed online, **c**onsumers have raised big concerns on the breach of their personal information. Law and regulations have placed businesses and organizations under obligations to take actions to prevent data breach. Among various threats, insider threats have been identified as 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. After that, issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed. Finally, evidences of inside activities are modeled as evidence trees to provide guidance for inside activity identification, tracking, and reconstruction.

1. INTRODUCTION

As healthcare data are pushed online, **c**onsumers have raised big concerns on the breach of their personal information. Law and regulations have placed businesses and organizations under obligations to take actions to prevent data breach. Among various threats, insider threats have been identified as 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. After that, issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed. Finally, evidences of inside activities are modeled as evidence trees to provide guidance for inside activity identification, tracking, and reconstruction

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 Figure 1. The data module is a central and critical part of this architecture and is composed 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, including hospital services, lab test services, 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. The data management party is essentially the IT teams including database/web/network/computer system administers, and the security/compliance team. The client party is essentially the patients and their legal guardians.

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 and payment information.

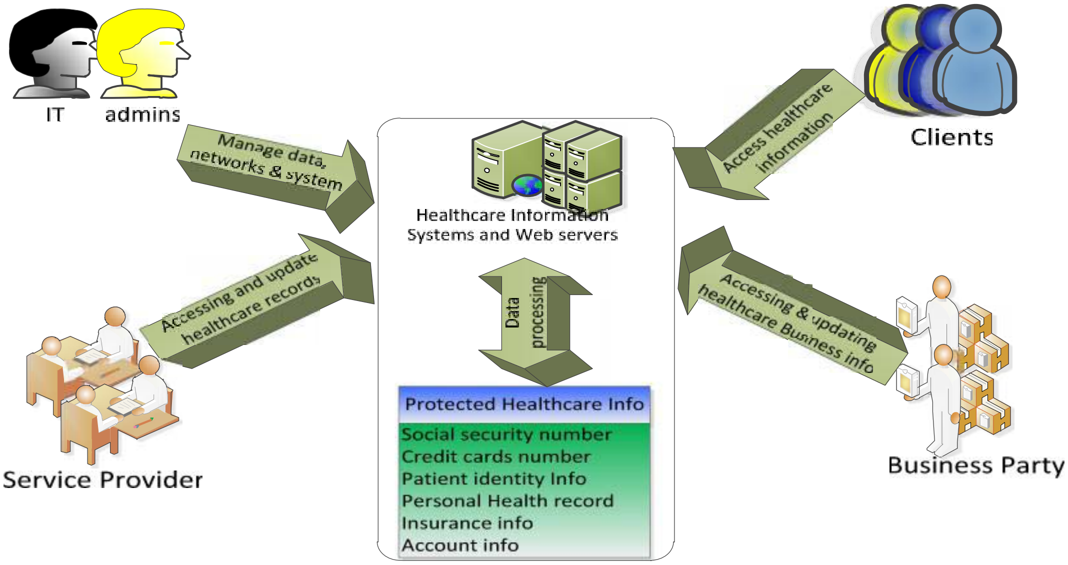

IT admins
Manage data, networks & system
Healthcare Information Systems and Web servers
Access healthcare information
Clients
Accessing and update healthcare records
Data processing
Accessing & updating healthcare Busness info
Service Provider
Business Party
Protected Healthcare Info
Social security number
Credit cards number
Patient identity Info
Personal Health record
Insurance info
Account info

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

3. RESEARCH METHODOLOGY

3.1 Inside Activity Identification Methods

In order to develop an effective method to identify inside activities from regular business activities, in this paper, two methods will be introduced to formally describe the relationship between inside activities and regular business activities. 1). The Work Role-Data Asset-User Operation-Access Preference (*WDOA*) model and, 2). The User-Operation-Data Asset-Access Path (*UODP*) model.

3.1.1 The WDOA (Work Role-Data Asset-User Operation-Access Preference) Model

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 (Chen, Laih, Pouget, & Dacier, 2005; Ellard & Megquier, 2004; Johnson & Willey, 2011; Murphey, 2007). For such a healthcare information system, denoted as Ω., the inside activity model *WDOA* can be defined as below.

Definition 1.1 (user): A user $u_i \in U = \{u_1, u_2, ..., u_L\}$, where $L \geq 0$ and $0 \leq i \leq L$, is a specific subject to access and consume recourses in Ω to perform tasks defined by the work role, where L denote the number of users in Ω.

A user in Ω is a specific subject who can be a doctor, a nurse, a business user, an information technology staff, or a non-empty set of software processes in Ω.

Definition 1.2 (work role): A work role $w_i \in W = \{w_1, w_2, ..., w_M\}$, where $M \geq 0$ and $0 \leq i \leq M$, is a group of users who have the same type of tasks and the same set of privileges to access and consume recourses in Ω to perform tasks, where M denote the number of work roles in Ω.

A work role in Ω can be doctor, nurse, business users, information technology staff, or a non-empty set of software processes in Ω. 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.

Definition 1.3 (sensitive asset): An asset $d_i \in D = \{d_1, d_2, ..., d_N\}$, where $N \geq 0$ and $0 \leq \ i \leq N$, is a category of resource that is owned by owner of $\boldsymbol{\Omega}$ and can be consumed by user, where N denote the number of resource categories in $\boldsymbol{\Omega}$.

The healthcare business should identify its own set of assets D. For example, user accounts, computer and network resources, and protected healthcare information. Also, the healthcare data should be classified into different sensitive levels, and the set of sensitive levels is denoted as S, where $S = \{s_1, s_2, ..., s_m\}$. 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 A denote the set of preference level where $A = \{a_1, a_2, ..., a_n\}$, 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 preference.

Definition 1.4 (operation): An operation $o_i \in O = \{o_1, o_2, ..., o_Z\}$, where $Z \geq 0$ and $0 \leq \ i \leq Z$, is a user activity to access or consume resources defined in $\boldsymbol{\Omega}$, where Z denote the total number of operations that can be performed by work roles defined in $\boldsymbol{\Omega}$.

The specific operations in O include read, write, execute, delete, shutdown, print, copy, and any other operation that is defined in a specific business sector. Based on the preference levels, 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, and should not copy the data to personal devices; an application developer will need to query sensitive data objects a lot but should not modify sensitive data. Therefore, a user's accesses to sensitive data objects in the healthcare enterprise environment can be modeled into certain pattern based on the work role of different users.

Definition 1 (WDOA Model): The WDOA Model is a 4-tuple $\{W, D, O, A\}$ data access preference model, where the first field represents a work role in W, the second field represents an asset in D, the third field represents an operation in O, and the last field represents an access preference in A.

The WDOA Model can 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.

3.1.2 The UODP (User-Operation-Data Asset-Access Path) Model

An inside activity model *UODP* is defined for an arbitrary healthcare information system $\boldsymbol{\Omega}$. In such a model, to reach a sensitive asset d_i, an insider needs to have known or unknown access paths to asset d_i (Ellard & Megquier, 2004; Kowalski et al, 2008; Moore, Cappelli, & Trzeciak, 2008).

Definition 2.1 (access path): An access path $p_i \in P = \{p_1, p_2, ..., p_K\}$,where $K \geq 0$ and $0 \leq \ i \leq K$, denote the access channels or access media for users to access or consume assets in the healthcare enterprise environment, where K is number of access paths enabled in $\boldsymbol{\Omega}$.

A specific access path can be USB access, CD access, VM instance access, email access, or any other access mechanism that allows users to access the sensitive asset, in legitimately way or illegitimately way. For example, to steal an asset d_i for personal use, an insider may copy asset d_i from the data storage site and then send to a personal USB device that has been attached to a system within a healthcare enterprise environment. An insider may first create secret user account and setup a virtual machine (VM) instance, and then copy d_i to the VM instance to be accessed later. Let $Path(\{u_i\}, k)$ denote the set of access paths to asset d_k by a subset of users $\{u_i\}$, then $Path(\{u_i\}, k)$ can be defined by a set of 4-tuples (u_i, o_j, d_k, p_m).

Definition 2 (UODP Model): The UODP Model is a 4-tuple $\{U, O, D, P\}$ data access path model, Where the first field represents a user in U, the second field represents an operation in O, the third field represents an asset in D, and the last field represents an access path in P.

With the 4-tuples (U, O, D, P) model, it is possible to determine 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, d_k, un-monitored *virtual machine*) 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, d_k, monitored USB) is not an inside 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.

3.2 Inside Activity Modeling for Incident Tracking and Reconstruction

The attack tree approach that is first proposed by Schneier (1999) is used to systematically analyze security threats. Attacks are modeled and represented by a tree structure where the root node represents the final goal, other interior nodes represent subgoals, and leaf nodes are attacking approaches to achieve the final goal (Poolsapassit & Ray, 2007). Children of a node in the tree can be one of the two logical types: *AND* and *OR*. To reach the goal, all of its *AND* children, or at least one of its *OR* children, must be accomplished. Attack trees grow incrementally by time and they capture knowledge in a reusable form. First, possible attack goals must be identified. Each attack goal becomes the root of its own attack tree. Construction continues by considering all possible attacks against the given goal. These attacks form the *AND* and *OR* children of the goal. Next, each of these attacks becomes a goal and their children are generated. Figure 2 shows an example of an attack tree of the inside threat, "*achieving the root privilege*". In such an attack, the attacker is a regular user and has a lower access privilege to the target (which needs root privilege), and conducts a series of attacking operations to achieve the root privilege as the system user. Note that links that are connected with a line represents the "AND" relationship among the states or sub-goals, which are working together to achieve the same parent goal.

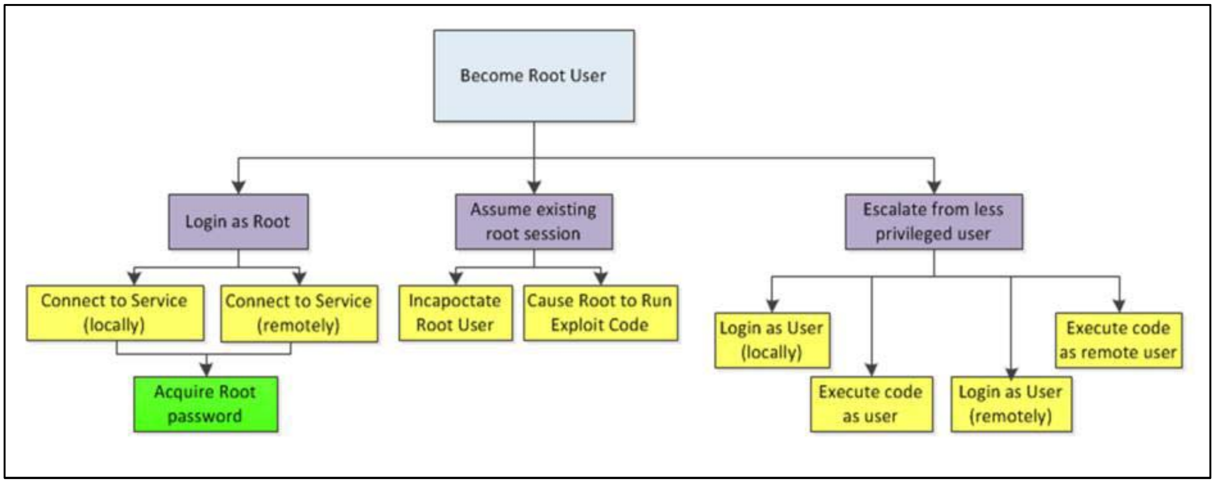

Become Root User
Login as Root
Assume existing root session
Escalate from less privileged user
Connect to Service (locally)
Connect to Service (remotely)
Incapoctate Root User
Cause Root to Run Exploit Code
Login as User (locally)
Execute code as remote user
Acquire Root password
Execute code as user
Login as User (remotely)

Figure 2. An attack tree of an internal threat "*achieving the root privilege*".

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, 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 data operations 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, and healthcare records. The results are shown in Table 1.

Table 1
The enumeration of data loss threat vectors in an enterprise healthcare environment

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 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

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 (Biggs & Vidalis, 2010; Bruening & Treacy, 2009; Brunette & Mogull, 2009), 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.

4.1 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.

Table 2

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

Threat Vector	**# Users/P**	**# Users/A**	**# Files/P**	**# Files/A**	**Data Size/P**	**Data Size/A**
USB	2765	413	4449429	374015	1123 G	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

As indicated in Table 2, the number of users 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 to charge devices without proper removable media protection, phones can be used to taking pictures and then 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 in Table 2, employees tend to abuse such 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.

Table 3

The encrypted and unencrypted data loss path accesses and operations after the deployment of Safend

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

5. INSIDER ACTIVITY IDENTIFICATION & TRACKING

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 (Tu et al, 2012). Current research on inside threat detection and identification (Eberle & Holder, 2009; Moore, Cappelli, & Trzeciak, 2008; Phua, Lee, Smith, & Gayler, 2007) and event reconstruction mechanisms (Case et al, 2008; Tang, & Daniels, 2005; Tu et al, 2012) 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.

5.1 Data Loss Identification

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 if the read access is permitted, or 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.

(3) A combination of uncontrolled accesses 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 (Wippich, 2007).

5.2 Inside Activity Operation and Evidence Modeling

Inside activity identification and track require that inside activities to be thoroughly studied. In such study, inside activities will be modeled as attack tree and then conducted in a simulated environment and a forensic investigation will be followed for each attack successfully committed. Fingerprints will be located and identified for each operation of the inside activity. The metadata of the fingerprints of each attack operation, such as log name, format, location, timestamps, and security features. are composed into nodes, which will then become child nodes of the leaf nodes in the augmented attack tree. This entire process will finally result in an evidence tree for each inside activity studied. Fingerprints of sensitive operations of the evidence trees will be identified as incident identifiers and the evidence tree can provide the contextual information to reconstruct security incidents automatically.

In this research, two inside activities have been studied, both of which utilize removable media (USB drive and CD-ROM) as the access paths leading to data loss in the healthcare enterprise environment.

5.2.1 Inside Activity Operation Modeling

Inside Activity A, as shown in Figure 2, is a typical industrial espionage inside activity**.** In such incident, the insider has all the needed privileges to access data and the USB ports which are required to perform the user's duty. However, those sensitive data should not be copied to personal USB devices since this may result in potential information leakage. To perform such an attack, the user is logged into system with all needed privileges, navigate to sensitive data, copy and paste the sensitive data into the USB device. The USB device is then removed and the user is logged out of the system later.

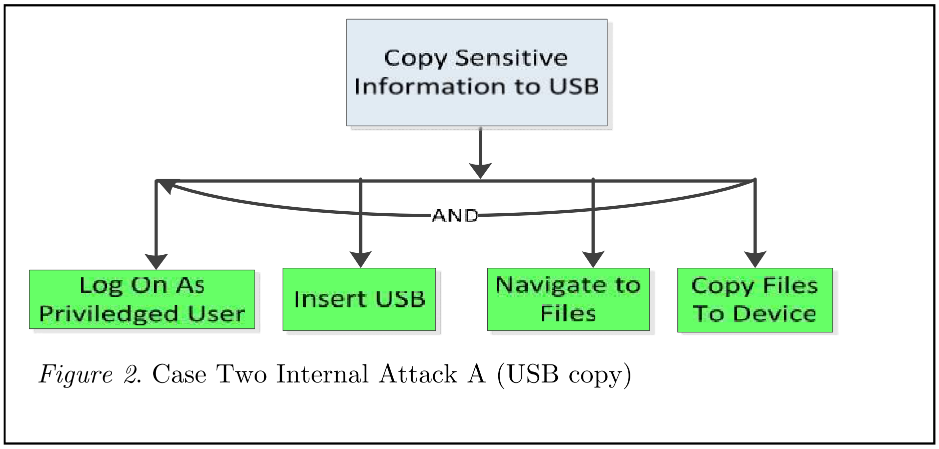


Figure 2. Case Two Internal Attack A (USB copy)

Inside Activity B, as shown in Figure 3, is also a typical industrial espionage inside attack**.** In such an attack, the attacker has all the needed privileges to access sensitive data and to access the CD-ROM Drive, which is needed to perform the user's duty. However, those sensitive data should not be copied to CD-ROM since this may result in potential information leakage. To perform such an attack, the user can log into system with all needed privileges and navigate to sensitive data, and then burns the sensitive data onto a CD-ROM. The CD-ROM is then removed and the user is logged out of the system.

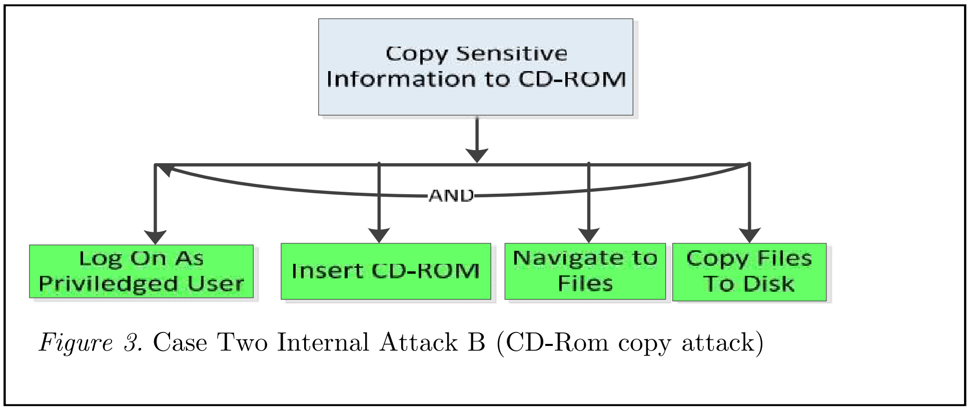
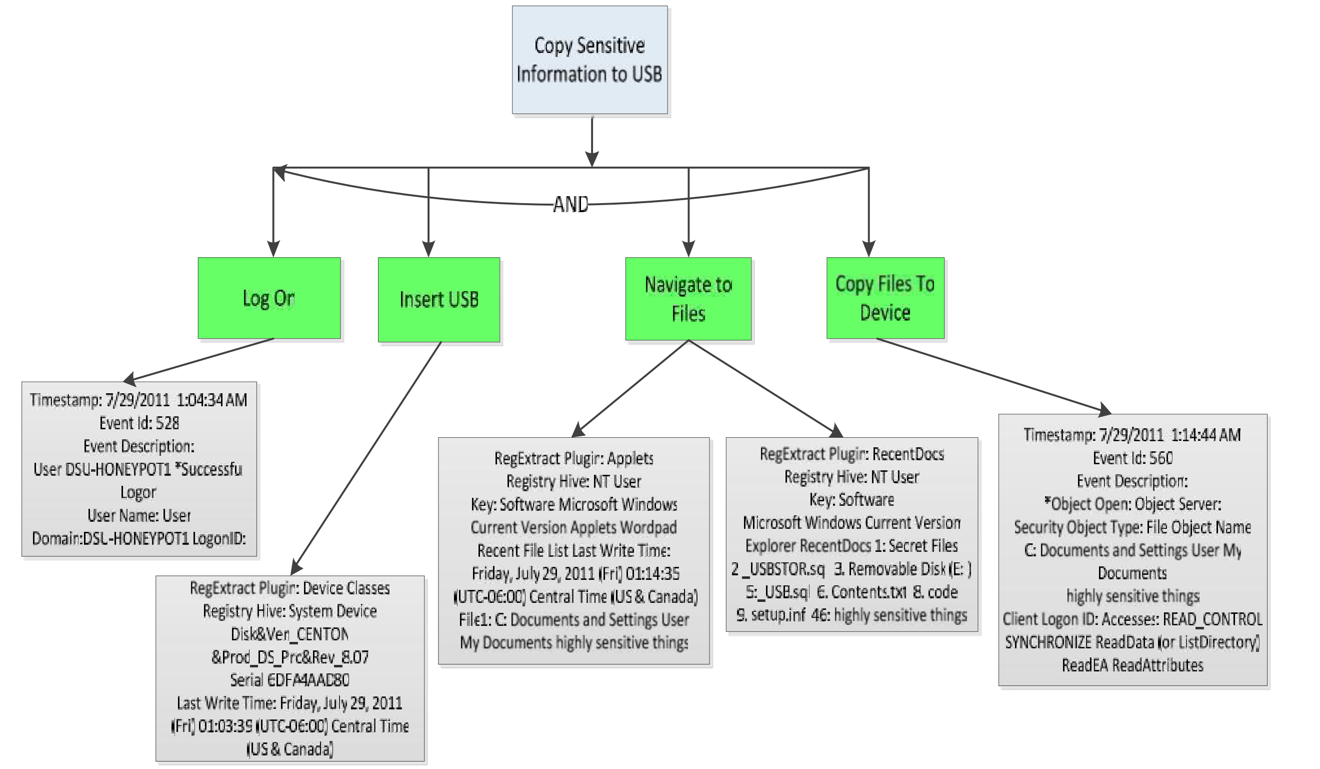

Copy Sensitive Information to CD-ROM
AND
Log On As Priviledged User
Insert CD-ROM
Navigate to Files
Copy Files To Disk

Figure 3. Case Two Internal Attack B (CD-Rom copy attack)

5.2.2 Inside Activity Evidence Modeling

The results of Attack A and Attack B are shown in Figures 4 and 5. Each figure contains an augmented threat tree that represents the vulnerability exploited, the steps needed to exploit it, the attacker's operations, and the fingerprint generated by those operations. The final goal of both attacks is to steal sensitive information from a business information system with desired system permissions. Operations conducted on a Windows machine may leave some forensic traces in the registry, some are persistent for a long time and some are volatile. If a piece of registry fingerprint is coupled with information from the event logs and file systems, the insider attack may be tracked and reconstructed. Based on our observation, relevant fingerprints can be located in machine's *System* hive, *Software* hive, the user's *NTuser.dat* hive, the *setupapi.log* that keeps a history of all devices installed via plug and play, and the Security event log.



Timestamp: 7/29/2011 1:04:34 AM
Event Id: 528
Event Description:
User DSU-HONEYPOT1 *Successfu
Logor
User Name: User
Domain:DSL-HONEYPOT1 LogonID:

RegExtract Plugin: Device Classes
Registry Hive: System Device
Disk&Ven_CENTON
&Prod_DS_Prc&Rev_8.07
Serial 6DFA4AAD80
Last Write Time: Friday, July 29, 2011
(Fri) 01:03:39 (UTC-06:00) Central Time
(US & Canada)

RegExtract Plugin: Applets
Registry Hive: NT User
Key: Software Microsoft Windows
Current Version Applets Wordpad
Recent File List Last Write Time:
Friday, July 29, 2011 (Fri) 01:14:35
(UTC-06:00) Central Time (US & Canada)
File1: C: Documents and Settings User
My Documents highly sensitive things

RegExtract Plugin: RecentDocs
Registry Hive: NT User
Key: Software
Microsoft Windows Current Version
Explorer RecentDocs 1: Secret Files
2 _USBSTOR.sq 3. Removable Disk (E:)
5:_USB.sql 6. Contents.txt 8. code
9. setup.inf 46: highly sensitive things

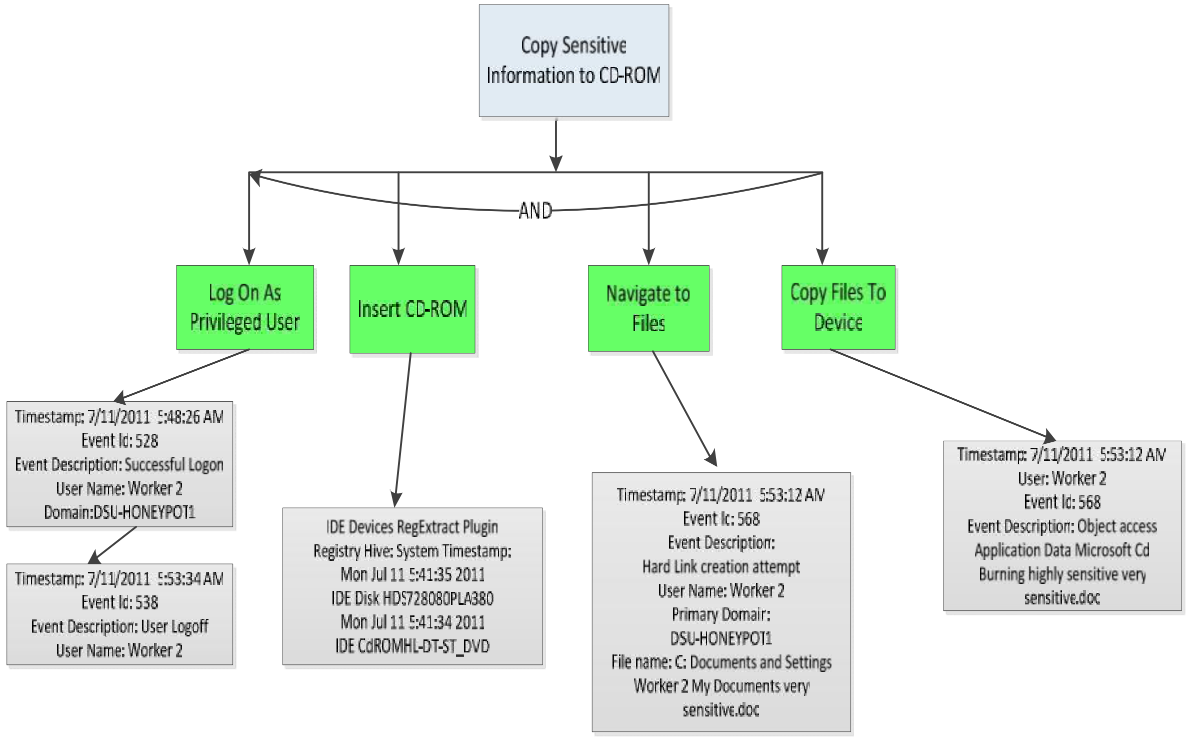
Timestamp: 7/29/2011 1:14:44 AM
Event Id: 560
Event Description:
*Object Open: Object Server:
Security Object Type: File Object Name
C: Documents and Settings User My
Documents
highly sensitive things
Client Logon ID: Accesses: READ_CONTROL
SYNCHRONIZE ReadData (or ListDirectory)
ReadEA ReadAttributes

Figure 4. A part of the evidence tree for Inside Activity A.

The inside attack A is conducted on 7/29/2011. Based on information in the registry, at 1:03:39 AM, a *Centon* USB device with a serial number of *6AFA4AAD80* was attached to the machine. At 1:04:34 AM, the attack was logged into the system and left fingerprints in the security event log. Based on additional fingerprints in the registry, the USB device with serial number *6AFA4AAD80* can be linked with the disk with driver letter *E*. Examining the *RecentDocs* registry key with the tool *RegExtract* shows that _USBSTOR.sql, Removable Disk (E:), _USB.sql, and a file named “highly sensitive things” which is flagged in the honeypot as a sensitive file, were recently accessed. At 1:14:44 AM, User synchronized the document titled with “highly sensitive things”, with the Removable

Disk (E:). The evidence tree of attack A is shown in Figure 4.

The inside attack B is conducted on 7/29/2011. Based on fingerprints in the *security event log*, user *Worker 2* logged into the system at 5:46: 26, and attempted to create a hard link with "highly sensitive very sensitive" at 5:53:12. Analysis of the IDE Device Class registry shows that a CD ROM was documented at 5:47:34, a minute after *Worker 2* logged on to the system. Finally, the user *Worker 2* is found to burn the file "highly sensitive things" to the CD ROM at 5:53:12. The evidence tree of attack B is shown in Figure 5.


Copy Sensitive Information to CD-ROM
AND
Log On As Privileged User
Insert CD-ROM
Navigate to Files
Copy Files To Device

Timestamp: 7/11/2011 5:48:26 AM
Event Id: 528
Event Description: Successful Logon
User Name: Worker 2
Domain:DSU-HONEYPOT1

Timestamp: 7/11/2011 5:53:34 AM
Event Id: 538
Event Description: User Logoff
User Name: Worker 2

IDE Devices RegExtract Plugin
Registry Hive: System Timestamp:
Mon Jul 11 5:41:35 2011
IDE Disk HDS728080PLA380
Mon Jul 11 5:41:34 2011
IDE CdROMHL-DT-ST_DVD

Timestamp: 7/11/2011 5:53:12 AM
Event Ic: 568
Event Description:
Hard Link creation attempt
User Name: Worker 2
Primary Domair:
DSU-HONEYPOT1
File name: C: Documents and Settings
Worker 2 My Documents very
sensitive.doc

Timestamp: 7/11/2011 5:53:12 AM
User: Worker 2
Event Id: 568
Event Description: Object access
Application Data Microsoft Cd
Burning highly sensitive very
sensitive.doc

Figure 5. A part of the evidence tree for **Inside Activity B**

Now we discuss how to determine the incident identifiers for the two inside activities. 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 ∨ *email access*) 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 (Sheyner, 2002; Todtmann, Riebach, & Rathgeb, 2007; Tu et al, 2012). 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 (Tu et al, 2012).

6. RELATED WORK

Forensics readiness has recently been a big research concern in digital forensic investigation and information assurance (Carrier & Spafford, 2003; Carrier & Spafford, 2004; Popovsky, Frincke, and Taylor, 2007; Rowlinson, 2004; Tan, 2001; Tang & Daniels, 2005; Wilson & Wolfe, 2003; Yasinsac & Manzano, 2001). These existing research efforts focus on organization-level framework design such as policy or management. None of them has addressed the details of the technology part of forensics readiness, e.g., mechanisms of the application and system event logging, fingerprint storage and archiving, and evidence-handling procedures, which are essential to enable forensics readiness for computer information systems. Our research presented in this paper attempts to provide a practical mechanism to automatically identify, track, and reconstruct attacks or inside activities, through the identification and tracking of the evidences of the attacks or inside activities.

An insider usually has the desired privilege and does not need to conduct any malicious activity (or attack) to obtain the privilege to access sensitive assets. Current malicious activity monitoring and detection techniques have limitations to effectively detect inside activities (Moore, Cappelli, & Trzeciak, 2008; Tu et al, 2012). Some research works have attempted to address inside threat modeling and detection issues (Bradford, Brown & Perdue, 2004; Burford, Lewis, & Jakobson, 2008; Chivers, 2009; Eberle & Holder, 2009). Bradford, Brown, & Perdue (2004) proposed principles for proactive computer-system forensics investigation on security incidents include internal threats, but no technical implementation of the proposed principles has been given and their focus is not threat detection. Burford, Lewis, & Jakobson (2008) proposed a comprehensive framework defining a large set of internal threat 'observables', and a graph theory based method to model individuals' behavior

(Chivers, 2009). Eberle & Holder (2009) proposed an inside activity detection method in which behavioral events are modeled as graphs and abnormal behaviors such as inside activities can be identified by searching abnormal subgraphs. The above approaches offer the advantage of modeling potential attacker and providing interesting insights into observable behavior (Chivers, 2009). However, their applications are limited by the availability of social knowledge of the insiders. Our research, however, will simply require the locating and identification the fingerprints left in the systems by operations of attacks or inside activities, with the guidance of evidence models, attack identifiers, and access preference models.

The WDOA (Work Role-Data Asset-User Operation-Access Preference) Model and the UODP (User-Operation-Data Asset-Access Path) model reply on the classification of work role and data asset. Similarly, access control models such as role based access models and Lattice based access models [Harris, 2012] all rely on such on the classification or labeling of subjects and objects. The role based access control models classify users to a set of work roles and each role is assigned with a set of access privileges. A subject (or a user) can exercise a permission only if the permission is authorized for the subject's (or user's) active role. The Lattice based access models are mandatory access control models. The Bell-LaPadula Model focuses on the protection of confidentiality of information such that an object (data) can only be read by a subject (or a user) with higher (or equal) security clearance and an object (data) can only be write by a subject (or a user) with lower (or equal) security clearance. The Biba Model focuses on the protection of system integrity such that an object (data) can only be write by a subject (or a user) with higher (or equal) security clearance and an object (data) can only be read by a subject (or a user) with lower (or equal) security clearance. These access control mechanisms can protect confidentiality and integrity upon the authorization of access requests from users, however, they have no control on data after accesses are granted. Also, insiders usually have the desired access privileges to access data objects, thus, access control mechanisms will have limited effectiveness on inside activity identification and tracking.

7. 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. After that, issues on data loss prevention and inside activity incident identification, tracking, and reconstruction are discussed. Finally, inside activities are conducted in a simulated healthcare environment, evidences of inside activities are collected, analyzed and then modeled. Evidence trees have been developed for inside activities, which are expected to provide guidance for internal activity incident identification and reconstruction.

REFERENCES

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

Bradford, P., Brown, M., Perdue, J. (2004). Towards proactive computer-system forensics. IEEE International Conference on Information Technology: Coding and Computing (ITCC 2004).

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.

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.

Burford, J., Lewis, L., and Jakobson, G. (2008). Insider threat detection using situation-aware MAS. In IEEE 11th International Conference on Information Fusion, 1–8, Germany.

Carrier, B. & Spafford, E. (2003). Getting physical with the digital investigation process. *International Journal of Digital Evidence*, *2*(2).

Carrier, B. & Spafford, E. (2004, July). An event-based digital forensic investigation framework. In Proceedings of Digital Forensic Research Workshop.

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

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

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

Chivers, H., Nobles, P., Shaikh, S., Clark, J., Chen, H. (2009). Accumulating Evidence of Insider Attacks. 1st International Workshop on Managing Inside Security Threats (MIST09).

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

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

El Emam, K., Neri, E., Jonker, E., Sokolova, M., Peyton, L., Neisa, A., Scassa, T. (2010). The inadvertent disclosure of personal health information through peer-to-peer file sharing programs. J. American Medical Informatics Assoc., 17(2), 148–158.

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

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.

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

Harris, S. (2012). CISSP All-In-One Exam Guide. 6th edition, ISBN: 978-0071781749.

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/.

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

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/.

Mauw, S. & Oostdijk, M. (2005). Foundations of attack trees. In Won, D., Kim, S., eds.: International Conference on Information Security and Cryptology – ICISC 2005.Volume 3935 of LNCS, Springer 186–198.

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

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

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/.

Poolsapassit, N. & Ray, I. (2007). Investigating computer attacks using attack trees. IFIP International Federation for Information Processing, Vol. 242. Advanced Digital Forensics III.

Popovsky, B. E. & Frincke, D. (2004). Adding the fourth "R". In Proceeding of the 2004 IEEE Workshop on Information Assurance.

Popovsky, B. E., Frincke, D., and Taylor, C. (2007). A theoretical framework for organizational network forensic readiness. Journal of Computers. Vol. 2, No. 3.

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.

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/.

Rowlinson, R. (2004). Ten steps to forensic readiness. *International Journal of Digital Evidence*, *2*(3).

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

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).

RSA Security. (2008). CSI computer crime & security survey. [online]. Available: http://i.zdnet.com/blogs/csisurvey2008.pdf.

Saini, V., Duan, Q., Paruchuri, V. (2008). Threat modeling using attack trees. J. Comput.Small Coll. *23*(4).

Schneier, B. (1999). Attack trees: modeling security threats. Dr. Dobb's Journal.

Seltxer, L. (2006). Is online banking too dangerous? eWeek News. [online]. Available: http://www.eweek.com/c/a/Security/Is-Online-Banking-Too-Dangerous/.

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.

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, 273-284.

Singleton, T., Singleton, A., Bologna, G., and Lindquist, R. (2006). Fraud Auditing and Forensic Accounting, 3rd edition. ISBN: 9780471785910. Wiley.

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), 60-80.

Tan, J. (2001). Forensics readiness. Electronic version available at HTUhttp://www.arcert.gov.ar/webs/textos/forensic_readiness.pdf.

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*, 163-169.

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 Conference on Networking, 464-465.

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), 73- 97.

Wilson, W. & Wolfe, H. (2003). *Management strategies for implementing forensic security measures. Information Security Technical Report*, *8*(2).

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.

Yasinsac, A. and Manzano, Y. (2001). Policies to enhance computer and network forensics. Proceedings of the 2001 IEEE Workshop on Information Assurance and Security.