3-2005

Assessing the Usefulness of Visualization Tools to Investigate Hidden Patterns with Insider Attack Cases

Amy M. Rammel

ASSESSING THE USEFULNESS OF VISUALIZATION TOOLS TO INVESTIGATE HIDDEN PATTERNS WITHIN INSIDER ATTACK CASES

THESIS

Amy M. Rammel, Captain, USAF

AFIT/GIR/ENV/05M-14

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/GIR/ENV/05M-14

ASSESSING THE USEFULNESS OF VISUALIZATION TOOLS TO INVESTIGATE HIDDEN PATTERNS WITHIN INSIDER ATTACK CASES

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Information Resource Management

Amy M. Rammel, BS

Captain, USAF

March 2005

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GIR/ENV/05M-14

ASSESSING THE USEFULNESS OF VISUALIZATION TOOLS TO INVESTIGATE HIDDEN PATTERNS WITHIN INSIDER ATTACK CASES

Amy M. Rammel, BS
Captain, USAF

Approved:

//signed// 15 Mar 05
Alan R. Heminger, PhD (Advisor) date
Associate Professor, Information Resource Management
Department of Systems and Engineering Management

//signed// 15 Mar 05
Dennis D. Strouble, PhD (Reader) date
Assistant Professor, Information Resource Management
Department of Systems and Engineering Management

//signed// 15 Mar 05
Robert F. Mills, PhD (Reader) date
Assistant Professor, Electrical Engineering
Department of Electrical and Computer Engineering

AFIT/GIR/ENV/05M-14

Abstract

The insider threat is a major concern for organizations. Open markets, technological advances, and the evolving definition of employee have exacerbated the insider threat. Insider threat research efforts are focusing on both prevention and detection techniques. However, recent security violation trends highlight the damage insider attacks cause organizations and illuminate why organizations and researchers must develop new approaches to this challenge. Although fruitful research is being conducted and new technologies are being applied to the insider threat problem, companies remain susceptible to the costly damage generated by insider threat actions.

This research explored how visualization tools may be useful in highlighting patterns or relationships in insider attack case data and sought to determine if visualization software can assist in generating hypotheses for future insider threat research. The research analyzes cases of insider attack crimes committed during the period of 1998 to 2004 with an information visualization tool, IN-SPIRE. The results provide some evidence that visualization tools are useful in both finding patterns and generating hypotheses. By identifying new knowledge from insider threat cases, current insider threat models may be refined and other potential solutions may be discovered.

Acknowledgments

First, I would like to express my sincere appreciation to my faculty advisor, Dr. Alan Heminger, for having faith in me throughout this research effort. His guidance and support was invaluable in focusing me on the end goal. Next, I thank my committee members, Dr. Dennis Strouble and Dr. Robert Mills, for lending both their experience and insight to my research. A special thank you goes to Dr. Grimaila for sharing his data mining experience.

Finally, my deepest gratitude is to my family--my husband and three daughters--for their unwavering support and encouragement throughout the course of this endeavor.

Amy M. Rammel

Table of Contents

List of Figures

List of Figures (cont)

Figure Page

List of Tables

ASSESSING THE USEFULNESS OF VISUALIZATION TOOLS TO INVESTIGATE HIDDEN PATTERNS WITHIN INSIDER ATTACK CASES

I. Introduction

Overview

This chapter introduces the insider threat problem, including why open markets and technological advances have increased both the risk and ease of insider attacks. A brief review of recent security violation trends illuminates why organizations and researchers must develop new approaches to this challenge. Next, the case is made that data mining tools and exploratory analysis may provide further insight into the insider threat problem. Finally, the problem statement and research focus are presented.

Background

Outsourcing, contractors, consultants, partnership arrangements between organizations, and temporary employees have expanded the traditional definition of employee (Anderson, 1999; Schultz, 2002). In addition, globalization has frequently geographically separated the employees from their parent organization (Friedman, 2000). Due to these shifts, today’s organizations are finding it difficult to maintain a distinction between insiders and outsiders (Schultz, 2002).

An insider is a legitimate user of a computer system, such as a network administrator or financial officer. An outsider is someone without system authorization, such as a hacker or a virus writer. The focus of this research is on the insider threat--not the accidental misuse of an authorized user, but the deliberate misuse. As such, a malicious insider intentionally oversteps his or her positional or system authority and betrays the organization’s trust. Insiders can commit their crimes physically or electronically and can act alone or in tandem with groups outside the organization.

Advances in technology have expanded the resources of an organization, but this same progress has also advanced the techniques of our enemies and competitors (Kipp, 2001). The insider threat has also benefited from these new technologies. Technological advances such as networking capability, encryption, USB thumb drives, and CD burners have increased the ease with which malicious insiders can both conduct and conceal their attacks (NIPC, 2004). Instead of tediously photocopying hundreds of pages of documents that are awkward and bulky, inside attackers now have the capability to copy the data to a small medium that can be hidden in their pocket or briefcase, or even easier, the capability to encrypt and send the data right out of the organization with only a few clicks of a mouse.

Recent security surveys reveal only part of the scope of the insider threat. The 2002 survey sponsored by PricewaterhouseCoopers, the U.S. Chamber of Commerce, and the American Society of Industrial Security (ASIS) foundation reported that the greatest threat to proprietary information and intellectual property are former employees (Trends, 2002). In the 2004 E-Crime Watch Survey, 32% of the 500 respondents reported that insiders were the greatest cyber security threats to their organization (E-Crime, 2004).

Additionally, the Computer Security Institute (CSI) and Federal Bureau of Investigation's (FBI) ninth annual Computer Crime and Security Survey reported for the sixth straight year that insider abuse of network access was the second highest cited form of attack, second only to virus incidents (Gordon et al., 2004).

However, only 20% of the 481 organizations that experienced a computer intrusion in 2003 reported the crime to law enforcement (Gordon et al., 2004). The E-Crime survey results were worse; only 13% of the 140 organizations reported the intrusions to law enforcement (E-Crime, 2004). Since organizations seem reluctant to report any computer intrusions by either an insider threat or outsider attack, it is likely the insider threat is under reported. The full scope of the insider threat problem cannot be understood unless organizations report the crimes to law enforcement and researchers can analyze all of the available insider attack case data. Larry Johnson, a United States Secret Service agent in the Criminal Investigative Division stated:

> Many companies still seem unwilling to report e-crime for fear of damaging their reputation. However, as we see with this survey, ignoring the problem or dealing with it quietly is not working. The question is not why can't we stop these criminal acts from happening, but rather, why are we allowing them to take place? The technology and resources are there to effectively fight this. We just need to work smarter to do it (E-Crime, 2004:7).

One area of technology that may help is the use of data mining tools. Data mining tools may help researchers see important patterns of characteristics or behaviors that are present in information that we already have, but don't see because the patterns are not readily visible.

Newly developed data mining software and improved capabilities in older data mining software are being used in a variety of research areas. The U.S. government (U.S., 2004), hospitals (Cerrito, 2004; Lok, 2004), chemical and pharmaceutical companies (Robb, 2004; D'Amicom, 2002), retailers (Clark, 2002), and credit card companies use data mining technology successfully in non-security research efforts.

Data mining has evolved into two distinct areas: one for structured data and one for unstructured data (Mena, 2004). Structured data is organized data, such as in databases. Unstructured data is free form text, such as in documents, presentations, emails, and web pages. Unstructured data is the "wild, wild west" (Erramouspe, 2004:18) of information and accounts for the bulk of an organization's data stores (Meyers, 2002; Mena, 2004; Robb, 2004). Many organizations are 'data rich' yet 'knowledge poor' (Chen, 2001).

Fielden states that "considerable amounts of actionable information" is located in unstructured data (Fielden, 2000:88). Data mining makes it possible to automatically detect trends and patterns amongst the mass of unstructured text (Walter, 2003; Uramoto et al., 2004). This technology not only helps 'connect the dots,' but also helps decide which dots to connect (Sniffen, 2004).

A subset of data mining, visualization tools are designed specifically for unstructured data and operate in a form conducive to the strengths of the human brain (Hand et al., 2001; Fayyad et al., 2002; Mena, 2004). Visualization methods "harness the perceptual capabilities of humans to provide visual insight into the data" (Fayyad et al, 2002:4). Visualization methods rely on exploratory analysis techniques. Exploratory analysis looks for a hypothesis, unlike confirmatory analysis which starts with a hypothesis

(Chen, 2001). Exploratory analysis is the interactive collaboration between the software and the user (Cios et al., 1998). The user is “searching for structure or trends and is attempting to arrive at some hypothesis” (Grinstein and Ward, 2002:22). With exploratory analysis, there is “no indication of what the user expects and what type of discovery could be of interest” (Cios et al., 1998:3).

This research posits that exploratory analysis using data mining technology may help uncover new information regarding the insider threat which could be used to build new models or develop new technologies to help thwart future attacks. This study will focus on using information visualization technology to analyze reported insider attack cases to determine if these tools could be useful in understanding insider threat activities in new ways.

Problem Statement

Security surveys continue to report insiders as one of the major threats to today’s organization’s information systems (Trends, 2002; Gordon et al., 2004; E-Crime, 2004). Furthermore, organizations are not using all available technologies and resources to combat the insider threat (E-Crime, 2004). Advances in data mining tools have promising applications in a variety of research areas (U.S., 2004; Cerrito, 2004; Lok, 2004; Robb, 2004; D’Amicom, 2002; Clark, 2002). This research seeks to use a visualization tool to explore insider threat patterns or relationships from insider attack cases using a data mining information visualization tool to see how such a tool might be applied to enhance our knowledge of the insider threat.

Research Questions

A literature review will be conducted to learn what we know about the nature of the insider threat problem, including the technology tools and analysis approaches that are being used to investigate the insider threat. Insider threat models and frameworks will also be examined. A review will be conducted of data mining, unstructured data, and visualization tools to determine how these tools may be able to help advance insider threat research.

Armed with this “insider threat” knowledge, exploratory analysis using a data mining visualization tool, IN-SPIRE, will be performed on insider attack cases to seek to uncover relevant patterns or relationships that these perpetrators may have (or not have) in common. The data used in this study are insider attack cases obtained from the Department of Justice (DOJ). The specific questions that will be examined by this research are:

1. Using exploratory analysis, how can visualization tools be useful in highlighting patterns or relationships in insider attack case data?
2. Can visualization software assist in generating hypotheses for future insider threat research?

Summary

This chapter discussed the insider threat dilemma and introduced the proposed research to determine how data mining technology may be able to assist researchers in understanding and prevailing over the inside attacker. Chapter two will review literature on the insider threat, including previous insider threat research and data mining concepts.

Chapter three will discuss the methodology for analyzing the case data and conducting the research discussed in this chapter. Chapter four will detail the results obtained from the analysis research. Finally, Chapter five will present the conclusions and recommendations for the study and suggestions for further research.

II. Literature Review

Overview

This chapter summarizes the foundational literature this research will use to understand the insider threat phenomenon. This review includes insider threat indicators, insider threat motives and goals, and insider threat objectives previous researchers found would-be inside attackers may possess. Next, insider threat research efforts will be examined to look at insider threat models and frameworks and insider threat mitigation techniques. Following the review of research literature, data mining, unstructured data, visualization software, and exploratory analysis concepts are presented. Finally, rationale for applying visualization technology to the insider threat is discussed.

Introduction

The insider threat is not a new phenomenon. Schneier states that a glance at our past will show what to expect from the future (2001). Malicious insiders have stolen, sabotaged, destroyed, and misappropriated organizational assets centuries before computers were invented. In fact, the insider threat has not changed much since 1779 when Benedict Arnold conducted his traitor activity with the British (Robinson, 2001). These physical threats we saw in our past are mirrored in the digital world (Schneier, 2001). However, advances in technology have changed the methods in which insiders carry out their attacks. “They’re just repacking their old tricks for the new millennium” (Schneier, 2000:17). Also, new technologies have vastly increased the potential damage inside attackers may cause. As such, several security studies by both government and industry continue to report insiders as one of the major threats to today's organizations (*Trends*, 2002; Gordon et al., 2004; *E-Crime*, 2004; Jonas et al., 2001, Yager, 2003). In 2004, 32% of the 2004 E-Crime Watch Survey respondents reported that insiders were the greatest cyber security threat to their organization. In addition, "recent high profile fraud cases such as BCCI, Barings, and Enron show people in positions of trust habitually and often all too easily bypass internal control mechanisms" (Porter, 2003:12).

In the very recent past, few laws existed to prosecute these computer crimes. Law enforcement now takes these crimes seriously (Yager, 2003). International and national laws are in place to prosecute the crimes insiders commit. Killcrece and others include a list of these law resources in Appendix D: Cyber Crime Law Resources (Killcrece et al., 2003:163). However, with the exception of financial organizations who are required to report insider crimes (CSTB, 2000), other organizations are reluctant to report these crimes to law enforcement authorities (Gordon et al., 2004; *E-Crime*, 2004) due to loss of consumer confidence, damaged reputation, or loss of competitive advantage (Bateman et al., 2004).

In fact, only 20% of 481 organizations surveyed that experienced a computer intrusion in 2003 reported the crime to law enforcement (Gordon et al., 2004). This is comparable to the E-Crime survey results—of the 140 organizations that experienced an insider intrusion, only 13% reported to law enforcement (E-Crime, 2004). By not prosecuting these insider crimes, organizations are actually encouraging future insider attacks by perpetuating a low prosecution rate for these crimes. Organizations are also single-handedly absorbing the losses posed by insiders including 1) increased legal, research and development, and insurance costs, 2) loss of revenue, competitive advantage and market share, and 3) embarrassment (Trends, 2002). Non-reporting masks the true extent of the insider problem and denies researchers the ability to analyze the latest insider crimes to develop an effective response (Trends, 2002). In addition, globalization, technology, and the sheer volume of an organization's information have each contributed in their own way to the insider problem. Thus, the insider threat is probably a bigger problem than we can document.

Globalization.

Globalization has changed the organizational landscape. "Downsizing, outsourcing, transfer of jobs overseas, restructuring to adapt to the pressures of global economic competition, rapid technological change and increased hiring of part-time workers to avoid paying benefits are all eroding many employees' sense of job security and loyalty to employer" (Heuer, 2001:3). Low job security and employee loyalty increases the likelihood of an insider incident. Moreover, the "smooth functioning of our world" (Magklaras and Furnell, 2002:1) is highly dependent on computer systems and the connectivity the internet provides. The Internet Software Consortium identified nearly 250 million hosts on the internet in 2004. Figure 1 demonstrates the internet host count for the previous ten years.

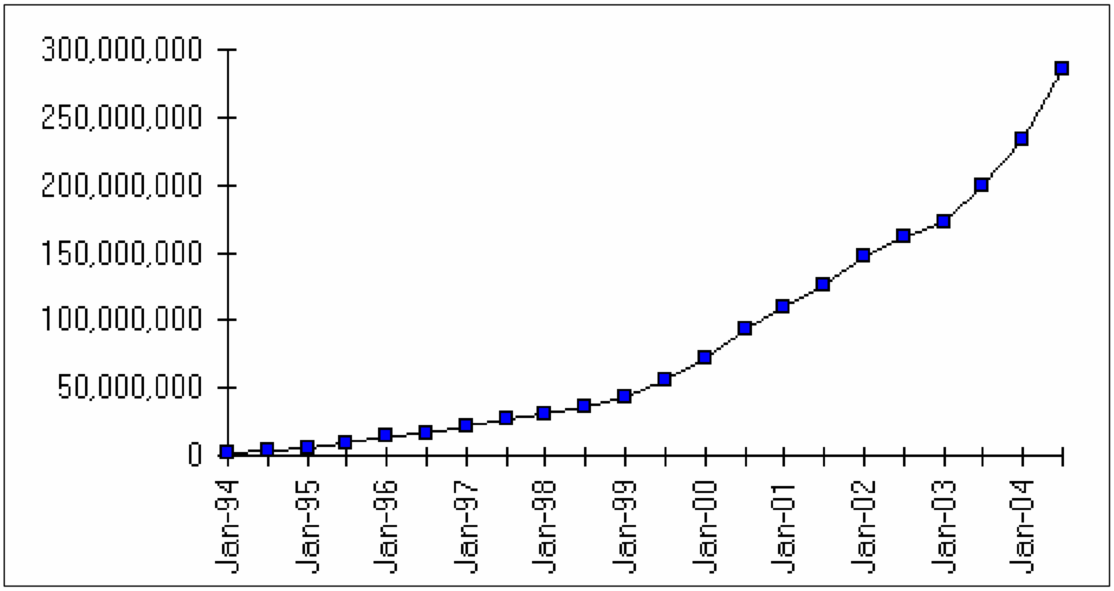


Figure 1. Internet Domain Host Count (ISC, 2004)

This interconnectivity is opening new doors in the business world, but at the same time, increasing the susceptibility of attack. The CIO at APL, a global shipping giant, states that "one of the largest threats facing us today is the interconnectivity between business associates" (Messmer, 2003:1). These business associates are foreign or domestic competitors, vendors or suppliers, strategic partners, intelligence services, and outsource manufacturers (Trends, 2002). The Computer Emergency Response Team (CERT) Coordination Center (CC) at Carnegie Mellon tracked the number of incidents that occurred against internet connected systems, though without any distinction between insider-initiated and outsider-initiated attacks. Figure 2 illustrates the perpetual growth of reported incidents since 1988. Incidents involve one, hundreds, or even thousands of sites. (CERT/CC, 2004).

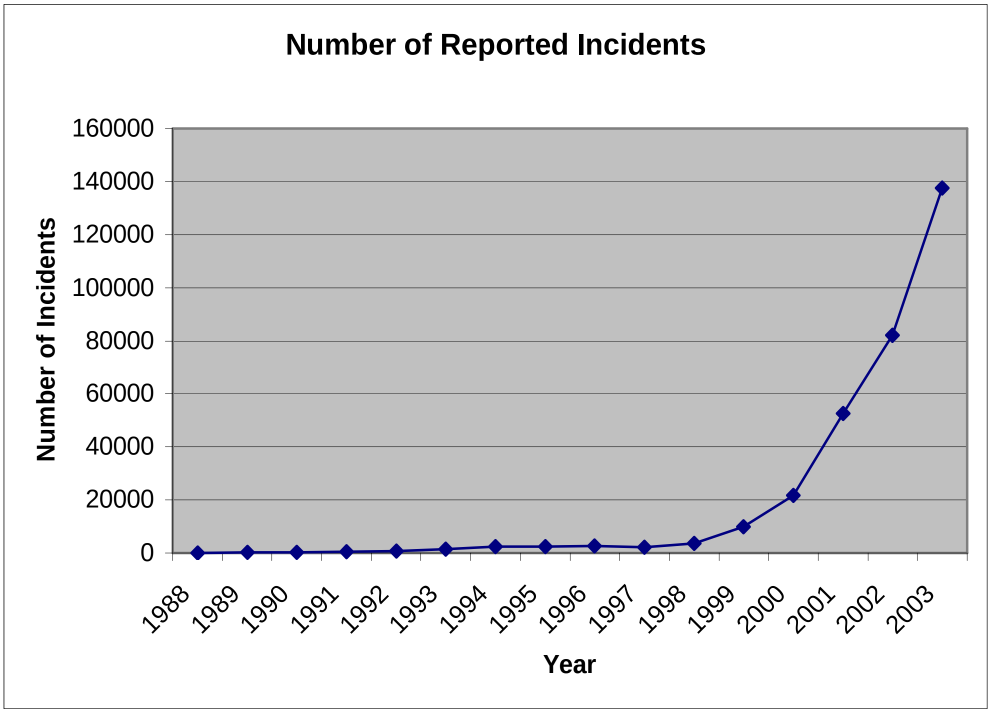


Figure 2. Incidents against Internet Connected Systems (Author)
Data from CERT/CC

In addition to interconnectivity, globalization has expanded an employee's job opportunities to the international level. Fewer American students are enrolled in science and engineering programs, thus, foreign born students are receiving doctoral degrees in these areas from U.S. universities in record numbers (NAS, 1995). Consequently, organizations are recruiting foreign students for their technical talent and to use them to break into new markets in their countries of origin. However, these employees' loyalty may be at odds between their country and their place of work. These conflicting loyalties increase their insider threat susceptibility.

Additionally, the increased use of Commercial-Off-the-Shelf (COTS) products and the development of standardized protocols have been a necessary requirement to expand the global marketplace. COTS has given “third parties access to hardware and software at many lifecycle points” (NSTISSC, 1999:4). COTS is increasingly developed by low cost, off-shore foreign nationals (Brackney and Anderson, 2004). This practice has increased “risk and removed certain checks and balances (Porter, 2003:12). Additionally, protocol standardization has given everyone the same baseline, giving others inside knowledge on how our systems and software operates.

Technology.

“Technology, too, has become a double-edged sword…its power, speed, pervasiveness, mobility, and anonymity offer attractive opportunities” (Porter, 2003:12). Like castle walls and moats that were used to stop invading armies (Kipp, 2001), organizations take many security measures, such as the use of firewalls, antivirus software, and intrusion detection systems to combat the malicious viruses that promulgate the Internet and to protect their data from hackers. Unfortunately, these technical security measures only protect outsiders from accessing the organization’s information system. Insiders are typically aware of these systems and often don’t need to bypass them to create havoc in an organization. The insiders are already ‘inside the castle walls,’ so many security measures do not deter them.

New tools and information technologies have made organizations more productive. Even so, just as we benefit from these advances, so can our enemies and competitors (Kipp, 2001). These tools and technologies have made it easier for an insider to conduct and conceal an attack. New technologies include encryption and networks. New tools include faxes, e-mail, CD burners, scanners, digital cameras, USB thumb drives, wireless technologies, anonymous remailers, and steganography (NIPC, 2004). Unfortunately, technology has advanced beyond our security measures (NSTISSC, 1999; Barnett, 2004). Security policies and technologies need to catch up with the newly created threats from these advanced technologies.

Volume of information.

For a number of years, companies have been collecting and storing huge amounts of data from a variety of sources (Bransten, 1999). Given today's high-technology environment and complex networked systems, our ability to gather and process information is unprecedented, as well as our ability to keep track of it all. Because as Charles Robertello stated "information is the only asset that can be in two places at the same time" (Schwartau, 1994:82), organizations may not even realize that they've been attacked.

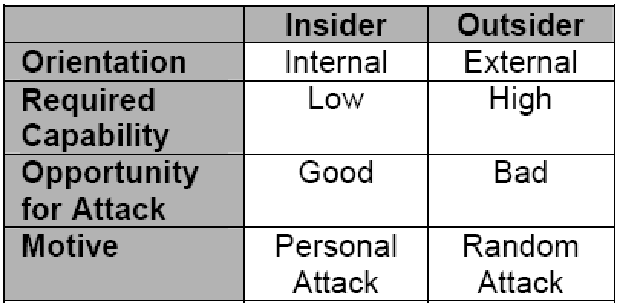
Several sources insist that the majority of insider losses are never discovered (NIST, 1994; Mitnick, 2002; Porter, 2003). "The National Computer Crimes Squad estimates that between 85% and 97% of computer intrusions are not even detected" (Icove et al., 2004:1). The role of chance plays a huge role in their discovery; often these crimes are detected accidentally. This suggests that those cases that are reported are just the tip of the iceberg.

Differences between Insider and Outsiders

Several security surveys have found that outsider attacks outnumber insider attacks (E-Crime, 2004; Gordon et al., 2004; Yager, 2003). Yet, security experts concede insider attacks are usually not only more successful, but also more costly (Shaw et al., 1998; E-

Crime, 2004; Schultz, 2002; Yager, 2003; Gordon et al., 2004; D'Arcy and Hovav 2004). It is valuable to examine the reasons why this may be so.

Important differences exist between insiders and outsiders. Gardiner noted the differences in the attacker's orientation, required capability, opportunity for attack, and motive (2003). Generally, outsiders are external to the organization, require time and skill to commit their attack, and usually choose who to commit their attacks against randomly. Insiders, however, are focused on their own organization, require little time or skill to circumvent the security controls, have regular system access to commit the attack, and the attack is more personal in nature (Gardiner, 2003). Gardiner's insider-outsider dichotomy is illustrated in Table 1.



	Insider	Outsider
Orientation	Internal	External
Required Capability	Low	High
Opportunity for Attack	Good	Bad
Motive	Personal Attack	Random Attack

Table 1. Comparison between Insider and Outsider Threat Agents (Gardiner, 2003:6)

Anderson also cited possible differences between insider and outsiders to include: 1.) Knowledge of the environment, 2.) Speed of attack, and 3.) Relative ease of accessibility (1999).

Both Gardiner and Anderson acknowledge that insiders generally have a distinct advantage over outsiders. Insiders possess innate privileges, physical access, and indepth knowledge of the environment (Anderson, 2000; Gaudin, 2000; Shaw et al., 1998) including the organization's policy and procedures (Jonas et al., 2001) and knowledge of an organization's real or potential weaknesses and vulnerabilities (Schneier, 2000). For instance, insiders are aware of the "undocumented realities" (Crume, 2000:88) of how security policies are followed, the cultural norms such as the fact that shared passwords are never changed. Insiders "have intimate knowledge of where the valuable information resides, and where to hit the company to cause the most harm" (Mitnick and Simon, 2002:161). All of these factors can make insider attacks more damaging and costly to the organization.

In addition to the Gardiner and Anderson's insider/outsider distinction, Chuvakin divides insider crimes into roughly three categories: mistakes, crimes of opportunity, and malicious premeditated crimes (Chuvakin, 2003). Similarly, the Department of Defense's (DoD) Integrated Process Team (IPT) states that insider attacks stem from a variety of employee actions including maliciousness, disdain of security practices, carelessness, and ignorance (1999). The categories described by Chuvakin and the DoD IPT distinguish accidental and purposeful events. Mistakes, disdain, carelessness, and ignorance are largely accidental in nature and are categorized as nonmalicious. Crimes of opportunity, premeditated crimes, and maliciousness, however, focus on a deliberate attempt to cause damage or destruction to the IT system and are thus categorized as malicious.

Neumann further broke down insiders into *groups of classes* (physical vs. logical presence, temporal vs. spatial reference, multidimensional nature) and the various *classes of insider misuse* (intentional vs. accidental, overt operation vs. covert operation) (1999).

Diverse approaches are needed to handle the distinctively different threats nonmalicious and malicious users pose. Both types of users are serious threats to any information system. In spite of this, due to scoping restraints, this research will focus on the malicious insider.

Insider Threat Definition

Many definitions of the insider exist. This section will review some of the insider definitions from previous research studies. The Department of Defense (DoD) Insider Threat Mitigation Team Integrated Process Team (IPT) defined an *insider* as “anyone who is or has been authorized access to a DoD information system whether a military member, a DoD civilian employee, or employee of another Federal agency or the private sector” (1999). As illustrated in Table 2, the IPT definition includes employees, network connected users, and information technology (IT) providers.

Employee	**Network Connected User**	**IT Providers**
Civilian or Military	Other Federal (Executive, Legislative)	Vendors and Suppliers (e.g., software development, maintenance)
Contractors (e.g., outsourcing)	Contractors (e.g., acquisition systems)	
Full-time, part-time, and temporary	Colleges/universities	
	Foreign partners, State & local, Other (EC/EDI)	

Table 2. Insiders (DoD IPT, 1999:3)

It is important to note that the IPT’s definition included many groups (e.g. columns two and three) who are not the traditional DoD employee, college/universities and foreign countries for example. However, in their definition, vendors and suppliers were limited to IT providers. In Denning’s research, she included non-IT providing vendors in her insider definition (Denning, 1999), as well as Brackney and Anderson who specified maintenance and custodial personnel (2004).

What is missing from these definitions is the malicious intent; the recognition that these insiders somehow abuse the organization's trust. For instance, Schultz and Shumway define an *insider attack* as "the intentional misuse of computer systems by users who are authorized to access those systems and networks" (2001:189). Recognizing the ambiguity, forty participants at a RAND Corporation insider threat workshop acknowledged the difficulty in defining the term insider. They determined the insider term was "like a chameleon—its color can change depending upon both the insider and the insider's environment" (Anderson, 1999:7).

A couple of points need to be made to help clarify who the malicious insider is in this research effort. First, insiders are not just employees; they are any person who has (or has had) business-related interactions with the organization. Secondly, insiders intentionally choose to misuse the organization's resource(s). Finally, the insider concept refers to users who abuse IT—the data, the software or hardware, the system, or the network. In this research, the insider is defined as a current or former associate of an organization that intentionally attempts to steal, deny, damage, degrade, or destroy an organization's data, information system, or information technology resources. For the remainder of this report, the malicious insider will be referred to simply as the 'insider threat' and their crimes as 'insider attacks.'

In addition to the insider threat definition, it is important to recognize that the insider threat is encompassed within white collar or occupational crimes such as fraud, money laundering, espionage, and stealing resources such as intellectual property or inventory. In addition, inside attackers can commit their crimes physically or electronically and can act alone or in tandem with groups outside the organization. Researchers generally agree that insider attackers internally possess certain traits or characteristics and give subtle clues that magnify at risk employees.

Insider Threat Indicators

Political Psychology Associates, Ltd. contends the employment contexts and the personal and cultural vulnerabilities can aid researchers in understanding and recognizing the insider threat. The employment contexts include full- and part-time employees, contractors, partners, consultants, and temps, and former employees. They assert these types of employees are motivated differently and have different loyalties. Personal and cultural vulnerabilities, on the other hand, can identify employees whom are at risk for this illegal or destructive behavior. Personal vulnerabilities include introversion, social and personal frustrations, and computer dependency; cultural vulnerabilities include ethical flexibility, reduced loyalty, entitlement, and lack of empathy (Shaw et al, 1998).

Heuer, however, asserts that four conditions are present before an employee betrays an organization’s trust and commits insider threat crimes such as espionage, embezzlement, and sabotage. The four preconditions are “opportunity, motive, an ability to overcome natural inhibitions to criminal behavior, and a trigger” (Heuer, 2001:1).

Chuvakin takes a slightly different approach. He argues that because an employee may possess one or many of the insider threat characteristics that alone does not necessarily make him a likely attacker. He contends that a combination of the insider threat characteristics, emotional stress, and a lack of supervisor interaction are more indicative of a potential inside attacker (Chuvakin, 2003).

Motives/Goals of the Malicious Insider

In addition to studying insider threat indicators, numerous case studies have been performed by researchers to determine the motives (or goals) of inside attackers. Insider threat motives typically fall into either a financial, social, political, or personal cause. Revenge, retaliation, money, ideology, and sabotage are widely recognized insider threat motives (Denning, 1999). In addition, greed, a need for recognition, a desire to make him or her irreplaceable (Shaw et al., 1998; Chuvakin, 2003), provocation of change, and subversion (Wood, 2000) are other cited motives. Others commit insider attacks to cause mischief or to test their skills (Jarvis, 2001). Heuer introduced divided loyalties (2001), while Krause cites fear of falling (Krause, 2002).

Furthermore, Shaw, Post, and Ruby described eight insider threat categories to include: Explorers, Good Samaritans, Hackers, Machiavellians, Exceptions, Avengers, Career Thieves, and Moles (Shaw et al, 1999). These insider threat categories explain the typical motivations behind the insider attack and give some insight into the malicious insider's objectives.

Objectives of the Malicious Insider

The objective of the malicious insider is to violate the information security triad--confidentiality, integrity, and availability--of the system (Chuvakin, 2003). Confidentiality, integrity, and availability are the key components of Information

Assurance (IA). Brackney and Anderson contend that the greatest threat to IA may be the insider threat (2004).

Information Assurance is defined as the “measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. This includes providing for restoration of information systems by incorporating protection, detection, and reaction capabilities” (DoDI 8500.2, 2003:19). Confidentiality is “the property that information is not made available or disclosed to unauthorized individuals, entities, or processes” (DAF, 2004:68). Integrity is the “property that allows the preservation of known unaltered states between baseline certifications and allows information, access, and processing services to function according to specified expectations. It is composed of data and system integrity” (DAF, 2004:70). Availability is “ensuring that data transmission or computing processing systems are not denied to authorized users” (Joint, 1997:504).

In layman’s terms, confidentiality means that information is shared by only authorized users, integrity means that the information is authentic and complete, and availability means that the system is accessible by authorized users when needed.

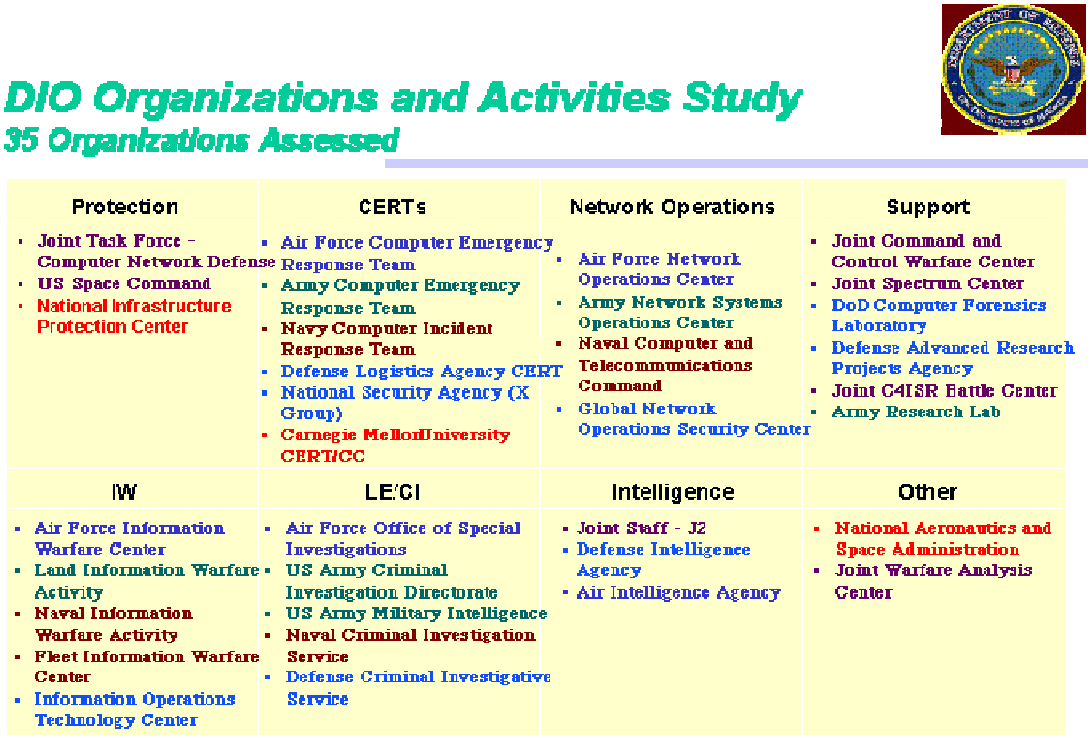
Charney provided a few examples of the types of crimes committed in these three areas. Confidentiality offenses include stolen customer lists and accessing someone else’s medical records or voice mail. Integrity offenses include defacing a web page or altering a credit report, criminal history, or investment advice. Availability offenses include denial-of-service attacks (Charney, 2004).

Insider Threat Models and Frameworks

To understand what research in the insider threat arena has been conducted, existing organizations that research the insider threat and insider threat models or frameworks were examined in both the government and private sector.

Government Sector.

The United States Government takes the insider threat seriously and has identified many organizations and activities that actively defend our country's information operations. Figure 3 lists 34 organizations that support Defensive Information Operations.



DIO Organizations and Activities Study

35 Organizations Assessed

Protection	CERTs	Network Operations	Support
- Joint Task Force - Computer Network Defense - US Space Command - National Infrastructure Protection Center	- Air Force Computer Emergency Response Team - Army Computer Emergency Response Team - Navy Computer Incident Response Team - Defense Logistics Agency CERT - National Security Agency (X Group) - Carnegie MellonUniversity CERT/CC	- Air Force Network Operations Center - Army Network Systems Operations Center - Naval Computer and Telecommunications Command - Global Network Operations Security Center	- Joint Command and Control Warfare Center - Joint Spectrum Center - DoD Computer Forensics Laboratory - Defense Advanced Research Projects Agency - Joint C4ISR Battle Center - Army Research Lab
IW	**LE/CI**	**Intelligence**	**Other**
- Air Force Information Warfare Center - Land Information Warfare Activity - Naval Information Warfare Activity - Fleet Information Warfare Center - Information Operations Technology Center	- Air Force Office of Special Investigations - US Army Criminal Investigation Directorate - US Army Military Intelligence - Naval Criminal Investigation Service - Defense Criminal Investigative Service	- Joint Staff - J2 - Defense Intelligence Agency - Air Intelligence Agency	- National Aeronautics and Space Administration - Joint Warfare Analysis Center

Figure 3. DIO Organizations and Activities (Anderson, 1999:27)

In addition, several government organizations are specifically in charge of conducting research (some collaboratively) in the insider threat area. (The following organizations are simply arranged in alphabetical order.)

- Advanced Research and Development Activity (ARDA) Advanced Intelligence Community (IC) Information Assurance – Focused on research in countering the insider threat. (www.ic-arda.org)

- Computer Emergency Response Team/Coordination Center (CERT/CC) – provides technical expertise in network systems survivability and security. (www.cert.org)

- Defense Personnel Security Research Center (PERSEC) – provides policy makers with research on personal security issues, such as espionage.

- Secret Service National Threat Assessment Center (NTAC) – develops and provides threat assessment training and conducts operational research relevant to public officials, workplace, stalking/domestic, and school-based violence. (www.secretservice.gov/ntac/)

The following paragraphs describe current areas these organizations are researching.

ARDA is currently designing Voltaire, a project aimed at protecting computer networks from insider threats. Designed for the intelligence community, the Voltaire system plans to integrate existing technology to detect suspicious activity and enforce access control (Jackson, 2004). ARDA is working to find or develop "technologies that better understand, prevent, detect, and react to malicious IC insider activities" (ARDA, 2004:2). This research seeks to determine if data mining technology helps to better understand the insider threat.

CERT/CC has been involved in many insider threat projects. The Secret Service's National Threat Assessment Center (NTAC) and the CERT/CC at Carnegie Mellon University conducted the Insider Threat Study to analyze the physical and online behavior of malicious insiders prior to and during network compromises (NTAC, 2004). The first critical infrastructure report studied malicious insider activity in the banking and finance sector from both a behavioral and technical perspective (Randazzo et al., 2004). Two interesting findings from this study were that many of the attacks did not require technical expertise to carry out the crime and that many of the crimes were conducted at the job during the normal workday (Randazzo et al., 2004).

Also, Adgar University College (Norway), CERT/CC, and TECNUN, University of Navarra (Spain) conducted a workshop and produced the Preliminary System Dynamics Maps of the Insider & Outsider Cyber-threat Problems. (CERT/CC, 2004). Presented at the System Dynamics Society conference, the maps modeled three areas regarding the insider threat: 1) Learning from experience, audits, and detection, 2) Growth of motive, and 3) Trust and deterrence (CERT/CC, 2004).

Personnel Security Research Center (PERSEC) created a database based entirely on open source information for espionage cases from 1947 to 2001. Researchers cataloged 150 cases on the personal and job characteristics of the espionage criminals and the characteristics of the acts of espionage they committed (Herbig et al., 2002). This database was statistically analyzed and discovered important criminal background findings. One interesting finding was that twice as many espionage criminals were not recruited by other countries or companies, but instead decided to commit the attack on their own accord.

In 1997, a DoD Inspector General (IG) report indicated that in an investigation, 87% of the intruders in DoD computer systems were employees or other malicious insiders (DoD IG, 1997). The senior civilian official at the Office of the Assistant Secretary of Defense (OASD) Command, Control, Communications, and Intelligence (C3I) chartered the Insider Threat IPT in 1998 to “foster the effective development of interdependent technical and procedural safeguards” to reduce malicious behavior by malicious insiders (OASD C3I, 1998:1).

The Insider Threat IPT conducted a risk management review on the insider threat. From this review, the IPT identified six security elements to create the framework illustrated in Figure 4.

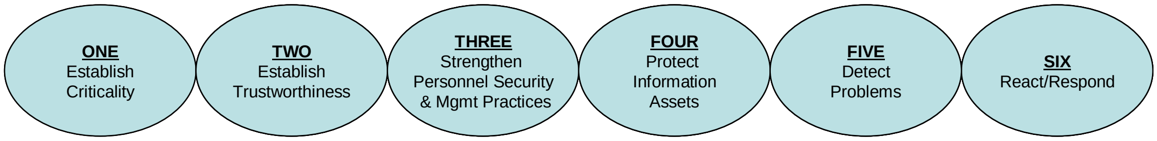


Figure 4. DoD Insider Threat IPT Strategy (Author)

In addition to creating this strategy, the IPT recommended that “the Department must also refine and update policies, procedures and practices to account for changes in operations attributable to changes in the military mission, the changing international security environment, and advances in technology” (DoD IPT, 1999:9). To address these issues, the DoD IPT identified 65 recommendations in the following seven areas: Policy and Strategic Initiatives, Personnel (Management and Security), Training and Awareness, Deterrence, Protection, Detection, and Reaction/Response. One of those recommendations was to conduct insider threat workshops on a recurring basis to examine technological approaches to mitigate the insider threat and to reduce information system vulnerabilities (DoD IPT, 1999).

The first workshop, sponsored by RAND Corporation, recommended specific technical research and development initiatives that would mitigate the insider threat. One of the recommendations was “to develop data correlation tools, including data reduction for forensics, and visualization tools focused on internal misuse” (Anderson, 1999:30). Because organizations must quickly gather and analyze data from a variety of sources when responding to a malicious insider incident, software tools are needed to visualize and analyze complex patterns before responding (Anderson, 1999). In addition to this recommendation, a Joint Task Force – Computer Network Defense (JTF-CND) chart characterizing an Information System Security Incident was modified. This new overview chart (Figure 5) made interesting distinctions among incidents, attacks, and events (Anderson, 1999).

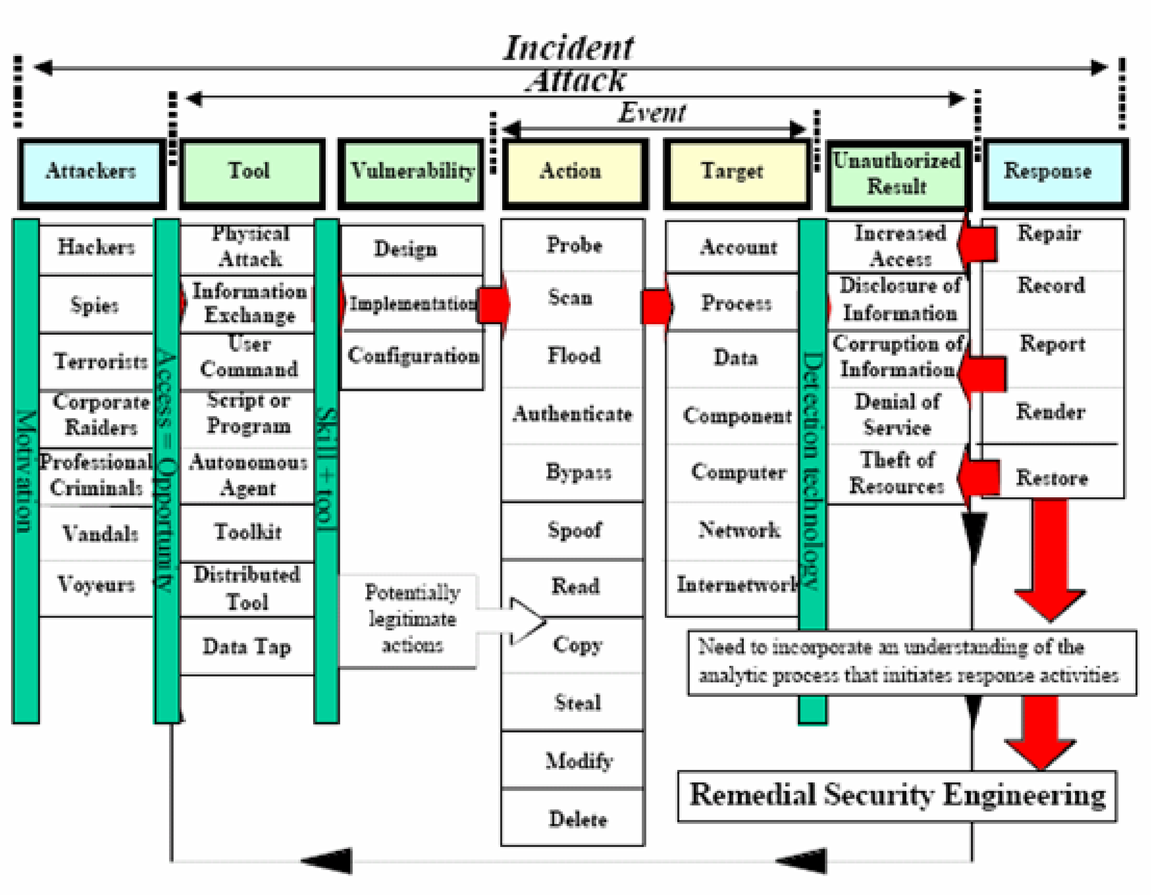


Figure 5. Information System Security Incident/Attack/Event (Anderson, 1999:13)

The second RAND workshop prioritized the DoD IPT’s recommendations. Ranked as one of the first priorities was the finding to “assess technologies currently available for dealing with the insider threat problem" (Anderson et al., 2000:21). The open research issues identified were:

- What existing, new, and projected technologies are being or could be used for dealing with the insider problem?
- What characteristics of technologies make them applicable to the insider problem (Anderson et al., 2000:21)?

Again, this research should be an extension of these issues.

Insider threat workshops were held at RAND Corporation in 2003 and 2004. A key component of the 2004 conference looked at collecting and analyzing the vulnerabilities and exploits of malicious insiders that have attacked intelligence systems in the past (Brackney and Anderson, 2004).

The U.S. government has done much to understand and mitigate the insider threat. The private sector has also contributed to insider threat research.

Private Sector.

The CMO Model is a very generic and basic model that asserts an individual needs the <u>capability</u> to commit an attack, the <u>motive</u> to do so, and the <u>opportunity</u> to commit the attack (Schultz, 2002). The CMO Model is very similar to Denning's information warfare means-motive-opportunity model (Denning, 1999). The mean (or capability) and opportunity is determined by the attacker's job position and technical skills. Thus, focusing on the attacker's motive seems to be a logical step in thwarting insider attacks.

Wood took a slightly different approach expanding the three components of the CMO model to create an insider threat model based on eight specific insider attributes. By focusing on these attributes—access, knowledge, privileges, skills, risk, tactics, motivation, and process—he contends organizations can realistically model the insider adversary (Wood, 2000). This model than can be used for insider threat simulation teams to test the security of IT systems.

Magklaras and Furnell developed an insider threat taxonomy and an insider threat prediction model (2001). The insider threat IT misuse taxonomy top level is *misusers*. Misusers are classified into three types: *system role*, *reasons of misuse*, and *system consequences*. Each of these groups is further categorized. *System role* is defined as what type of computer user the *misuser* is: *system masters*, *advanced users*, or *application users*. Figure 6 illustrates the Top Level and System Role views.

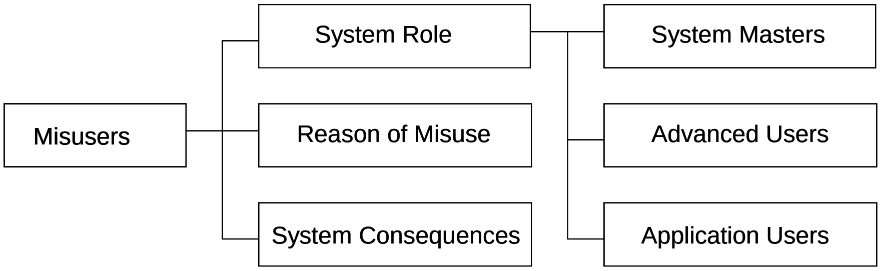


Figure 6. Top Level and System Role View (Magklaras and Furnell, 2001:64)

Reason of misuse is broken down into *intentional* and *accidental* incidents. *Intentional* incidents are *data theft*, *personal differences*, and *deliberate ignorance of rules*. Figure 7 illustrates the *reason of misuse* view.

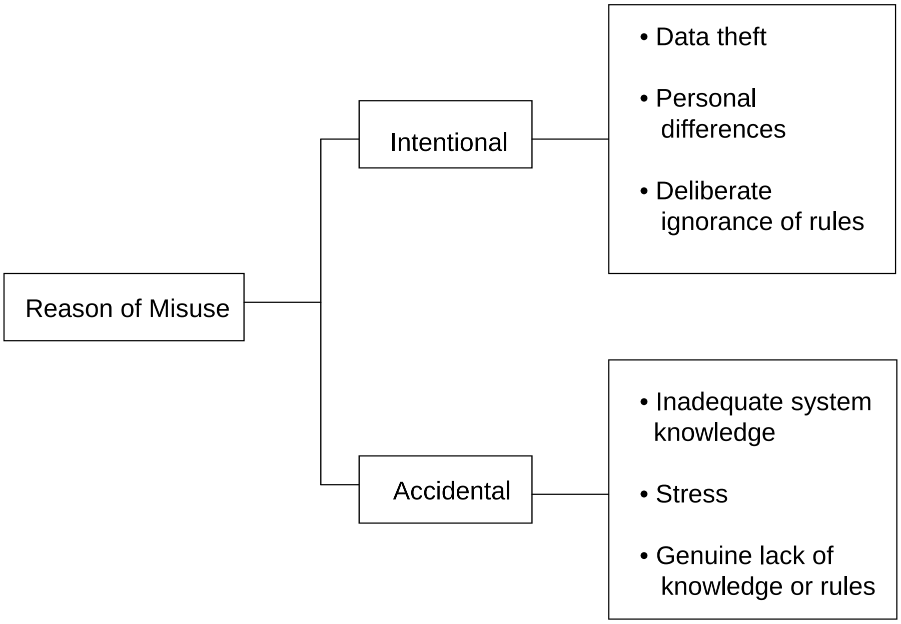

Reason of Misuse
Intentional
• Data theft
• Personal differences
• Deliberate ignorance of rules
Accidental
• Inadequate system knowledge
• Stress
• Genuine lack of knowledge or rules

Figure 7. Reason of Misuse View (Magklaras and Furnell, 2001:66)

System consequences describe what part of the information system the attacker damaged: *Operating System* (OS), *network*, or *hardware*. Figure 8 illustrates the *Systems consequences* view.

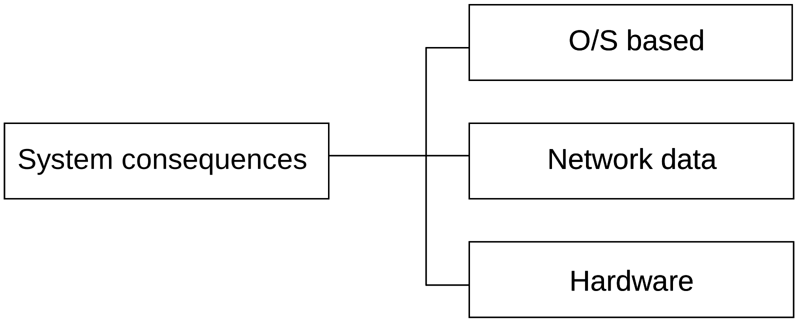


Figure 8. System Consequences View (Magklaras and Furnell, 2001:67)

Taking the taxonomy a step further, Magklaras and Furnell developed the Insider Threat Prediction Model (ITPM) as a means to quantify the taxonomy. Each component of the taxonomy was given a weighted rating in which to mathematically calculate the likelihood of an insider attack. Figure 9 illustrates the mathematical formula to calculate the threat.

(top level) **EPT** $\underset{\cdot}{=}$ **F**threat components
EPT = Fattrib **+ F**behavior **+ F**imsinfo
(second level) **EPT = C**role **+ C**tools **+ C**hardware **+ F**behavior **+ F**imsinfo
(third level) **EPT = C**role **+ C**data **+ C**hardware **+ F**knowledge **+ F**content **+ F**network **+ F**imsinfo

Figure 9. Three Layer ITPM Model (Magklaras and Furnell, 2001:72)

Similar to Wood, E. Eugene Schultz proposed that there was no single clue to predict or detect an insider attack, but multiple indicators with varying levels of contributions. He created a framework for predicting and detecting insider attacks. The insider threat indicators in Schultz's framework include personality traits, verbal behavior, correlated usage patterns, preparatory behavior, meaningful errors, and deliberate markers, as illustrated in Figure 10 (Schultz, 2002).

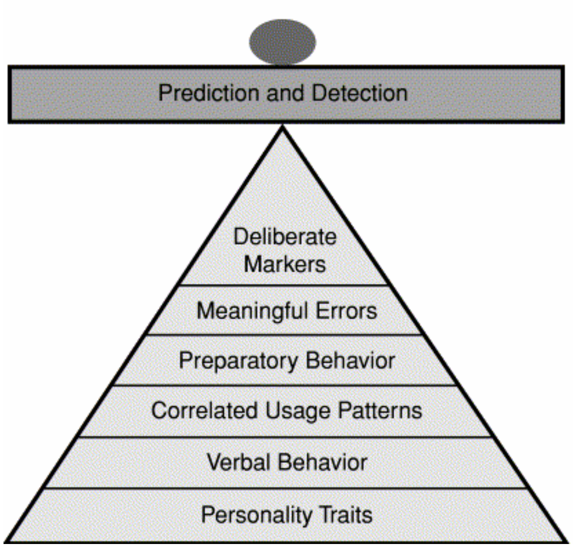


Figure 10. Framework for Potential Indicators of Insider Attacks (Schultz, 2002)

Taking a slightly different angle, Caruso developed an insider threat/outsourcing IT model using a grounded theory approach, Figure 11 (2003). She argues that outsourcing conditions, psychological conditions, socio-economic conditions, and systematic conditions each impact the likelihood of an employee to commit an insider attack.

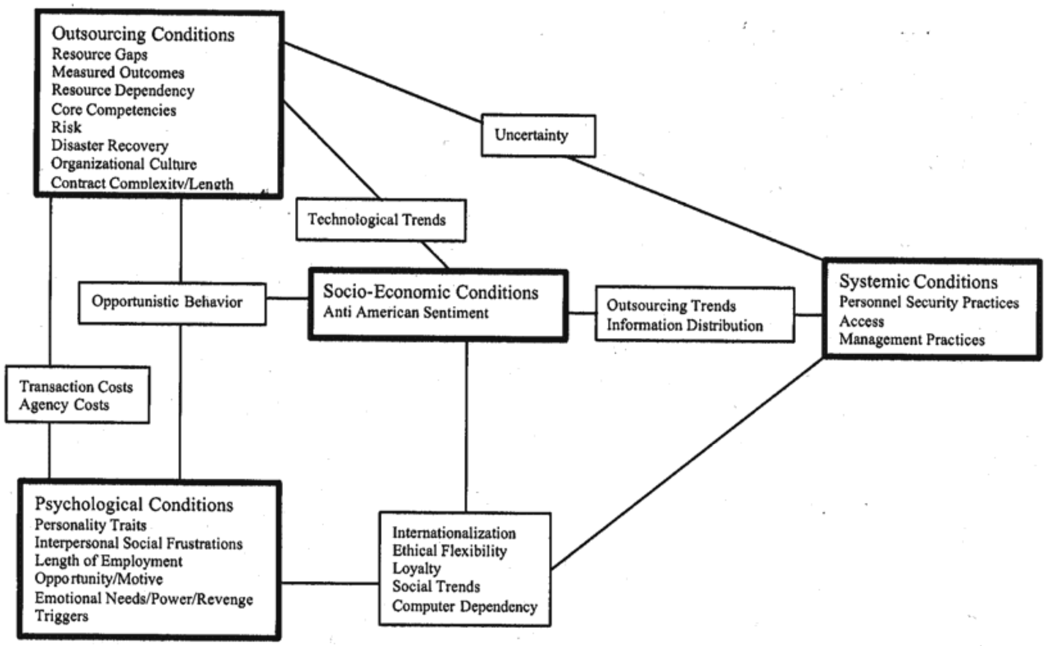


Figure 11. Insider Threat and Outsourcing IT Model (Caruso, 2003:83)

D'Arcy and Hovav proposed a conceptual model (see Figure 12) describing a relationship between security countermeasures and IS misuse intention with perceived certainty and perceived severity of sanctions as mediators and individual characteristics and employment context as moderators (D'Arcy and Hovav, 2004).

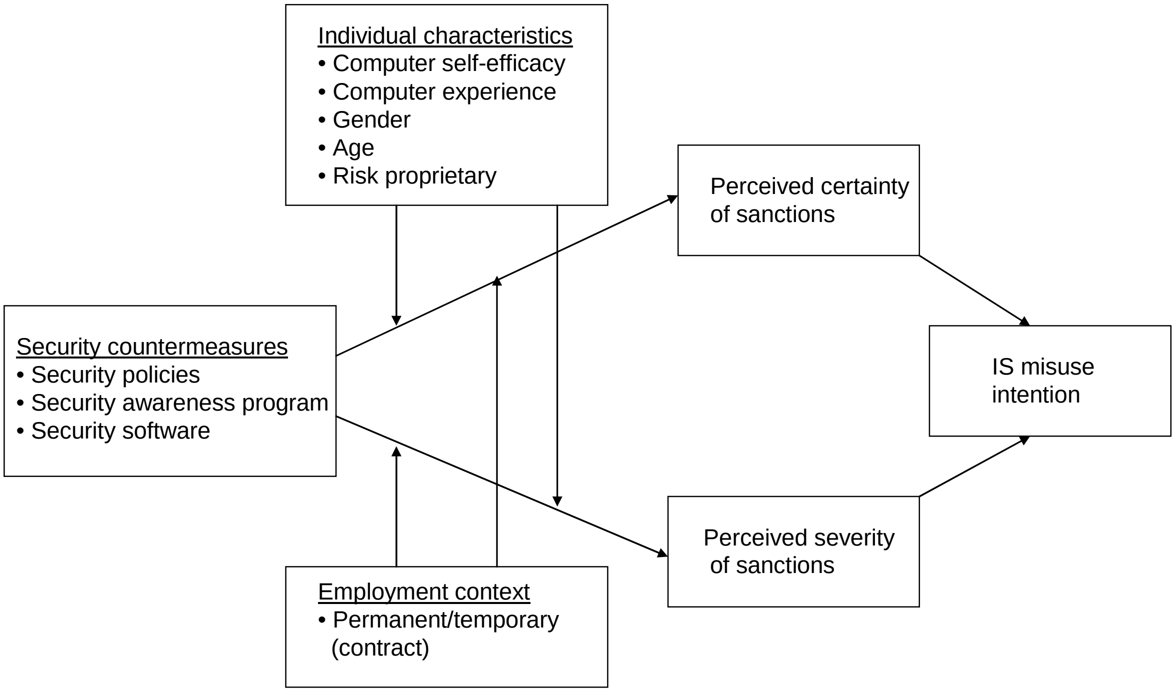
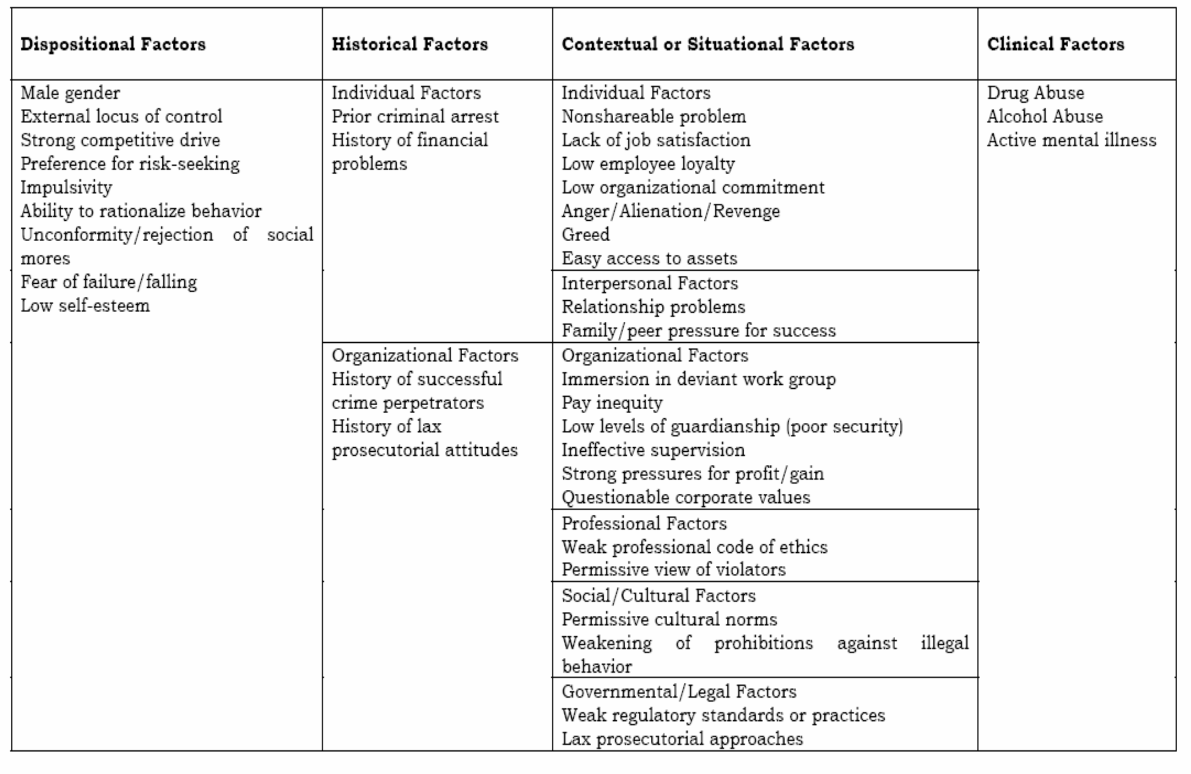


Figure 12. Model Linking Deterrent Security Countermeasures to IS Misuse Intention (D'Arcy and Hovav, 2004:4)

Krause focused on neutralization (e.g. rationalization) effects in white collar crime in her research of personnel security. Using Monahan's organizational scheme, Krause developed a list of dispositional factors, historical factors (individual and organizational), contextual or situational factors (individual, interpersonal, organizational, professional, social/cultural, government/legal and clinical) as it relates to occupational crime (2002). Table 3 illustrates the risk factors for occupational crime involvement. Several of these factors support previous insider threat research findings.



Dispositional Factors	**Historical Factors**	**Contextual or Situational Factors**	**Clinical Factors**
Male gender External locus of control Strong competitive drive Preference for risk-seeking Impulsivity Ability to rationalize behavior Unconformity/rejection of social mores Fear of failure/falling Low self-esteem	Individual Factors Prior criminal arrest History of financial problems	Individual Factors Nonshareable problem Lack of job satisfaction Low employee loyalty Low organizational commitment Anger/Alienation/Revenge Greed Easy access to assets	Drug Abuse Alcohol Abuse Active mental illness
		Interpersonal Factors Relationship problems Family/peer pressure for success	
	Organizational Factors History of successful crime perpetrators History of lax prosecutorial attitudes	Organizational Factors Immersion in deviant work group Pay inequity Low levels of guardianship (poor security) Ineffective supervision Strong pressures for profit/gain Questionable corporate values	
		Professional Factors Weak professional code of ethics Permissive view of violators	
		Social/Cultural Factors Permissive cultural norms Weakening of prohibitions against illegal behavior	
		Governmental/Legal Factors Weak regulatory standards or practices Lax prosecutorial approaches	

Table 3. Risk Factors for Occupational Crime Involvement (Krause, 2002:54)

Utilizing these factors with the other insider threat characteristics that have been researched, organizations and researchers can develop new policies, procedures, or tools to prevent or detect insider attacks.

In another fertile area of research, a non-profit group called the Honeynet project is researching tools, techniques, and activities of the intruder community (Killcrece et al., 2003). These honeynets, as well as honeypots and honeytokens, are a promising security resource. Killecrece and others defined a honeynet as

> essentially a network of systems deployed in a controlled environment that can be watched and monitored for attacks and intruder activity. By watching attacks and probes against the system or by monitoring how the system is compromised and used to attack others, the system owners can learn about the techniques and tools used by the intruder community. This information can then be used to improve the knowledge and understanding of other computer security professionals (2003:126).

Functioning like a one way mirror, honeynets may be the basis for effectively reducing the insider threat on IT systems.

Insider Threat Mitigation

Whatever mitigation controls are put into place should be in line with the identified risks. "Many companies spend many times more on security products than they are likely to lose from successful attacks" (Yager, 2003:44). To determine the appropriate dollar amount to spend on security efforts, organizational risks must be identified and managed. The risk management efforts regarding the insider threat have been categorized into four categories: technological, administrative, legal, and psychological methods. To mitigate the insider threat risk, a well-balanced prevention program should include all of these measures (Chuvakin, 2003).

Researchers are focusing on two methods to mitigate the insider threat: prevention and detection. "Prevention focuses on controls designed to reduce the opportunity for unauthorized use of corporate assets. Detection focuses on the controls designed to alert the appropriate personnel to the fact that a fraud has been perpetrated" (Porter, 2003:13). Prevention measures are used to thwart insider attacks before the crime is committed, whereas detection measures are used to minimize the damage caused by insider attacks.

Prevention should be the primary mitigation effort, but should prevention fail, systems should then focus on detection (Neumann, 1999).

Preventing the Insider Attack.

Prevention techniques include administrative and technical methods. Technological methods should include compliance activities such as auditing systems (Robinson, 2001) and honeypots to divert and detect possible attackers. Honeypots are a promising security resource that may be able to identify the insider threats and mitigate their damages while the 'attacks' are being conducted. Honeypots are a computer, a login/password, a document, a credit card number, or any item that attracts a person to a false entity. "Anything or anyone interacting with the honeypot is an anomaly, it should not be happening" (Spitzner, 2003: 2). "No single technical security solution can provide total system security; a proper balance of security mechanisms must be achieved" (Loscocco et al., 1998: 10). Compliance activities are also important, such as auditing the systems and the users (Robinson, 2001).

Administrative methods can be implemented by a variety of policies and procedures in both the management and security areas. Important human resource management practices include pre-employment screening and knowing how to terminate employees (Shaw et al., 1999; Scalet, 2002). Whether employees quit, are fired, or laid off, revocation of ID badges, changing key codes on doors, and disabling network and RAS accounts are important practices (Robinson, 2001). Supervisors must watch their employees for warning signs. Personnel changes, such as demotions, terminations, or reassignments, may be the event that triggers a malicious insider to attack (Shaw et al., 1999). Research has proven that these warning signs are not always recognized or acted upon.

Security polices would include education and awareness programs, a layering of security measures and concepts like least privilege. Least privilege, also called Just-enough Privilege (JeP) (Martzahn, 2003) and compartmentalizing (Schneier, 2000), is the security principle easiest to implement. Least privilege defines a unique demilitarized zone for each user of a system based upon the requirements of their job. Least privilege is defined as "every program and every user of the system should operate using the least set of privileges necessary to complete the job…to limit the damage that can result from an accident or error" (Saltzer and Schroeder, 1975: 1279). Least privilege is comparable to the military's 'need to know' rule (Saltzer and Schroeder, 1975; Langford, 2003). In addition to limiting privileges, Robinson recommends re-verification procedures for sensitive user accounts, group membership, and access control lists (2001).

Like Chuvakin, Dhillon and Moore believe there are safeguards organizations can put in place to minimize computer crime. The success of these controls is maintained by establishing the right balance between technical, formal, and informal interventions (Dhillon and Moore, 2001).

> Technical interventions essentially deal with restricting access, which may be to the buildings and rooms or to the systems and programs. Formal interventions deal with establishing rules and ensuring compliance to the laws and procedures. Informal interventions relate to the educational and awareness programs that could be put in place within organizations (Dhillon and Moore, 2001:720).

Informal controls are perhaps the most cost-effective type of control (Dhillon and Moore, 2001). However, the “amount spent (on these controls) should be in proportion to the criticality of the system, cost of the control, and probability of the occurrence of an event” (Dhillon and Moore, 2001:722).

Detecting the Insider Attack.

Insider detection efforts are put into place in case the prevention efforts fail. Like prevention efforts, detection efforts are both technical and administrative in nature. Anomaly and misuse detection software and systems, an extension of intrusion detection systems, have been developed to log and analyze user behavior (Neumann, 1999). By analyzing normal user behavior and system access history, these systems attempt to give organizations prior warning to prevent an insider attack. Should these systems miss the warning signs, however, they should then detect the attack. Interestingly enough, intrusion detection systems actually fall into both the prevention and detection domain depending on if the malicious insider was detected and stopped before or after a crime was committed. Keystroke monitors, voice recorders, and action logging are other detection technologies.

Since most malicious insider crimes are discovered accidentally (Porter, 2003; Icove et al., 2004), administrative methods are needed to assist discovery. Regularly scheduled reviews and audits can deter and uncover past crimes. Also, emphasizing employee awareness can reveal crimes due to employees sensing things that don’t seem quite right.

“Dealing with the insider threat inevitably involves organizational policies, practices, and processes as well as technological approaches” (CSTB, 2000:2).

Data Mining

The Institute for Management and Administration stated in its *2000 Report on Preventing Fraud* that “the analysis of company data is the single most effective way of preventing and detecting fraud, and computers and data analysis are generally underutilized” (Jonas et al., 2001:22).   In addition, many organizations are rich in data, yet poor in knowledge (Chen, 2001). A data analysis tool that has proven successful in identifying fraud, terrorists, new marketing strategies, health epidemics, and patent developments, is data mining (U.S., 2004; Cerrito, 2004; Lok, 2004; Robb, 2004; D’Amicom, 2002; Clark, 2002). Data mining is a software analysis that automatically detects trends and patterns among data (Walter, 2003; Uramoto et al., 2004). This technology not only helps ‘connect the dots,’ but also helps decide which dots to connect (Sniffen, 2004). Data mining comes from a variety of disciplines including: statistics, database technology, machine learning, pattern recognition, artificial analysis, and visualization (Cios et al., 1998; Hand et al., 2001; Chen, 2001; Mena, 2004; Fayyad et al., 2002).

Data mining is defined as “the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner” (Hand et al., 2001:1). Data mining tools provide a slightly different approach to data analysis than traditional statistical methods. For one, the data used in a data mining project is usually collected for some other reason than the data mining analysis and is often called ‘secondary’ data analysis (Hand et al., 2001). Furthermore, statistical methods relies largely on numerical data, whereas data mining can involve numerical or text data, or both.

Data mining has evolved into two distinct areas: one for structured data and one for unstructured data (Mena, 2004). Structured data is organized data, such as in databases. Unstructured data is free form text, such as in documents, presentations, emails, and web pages. Some researchers define data mining as the analysis of ‘structured’ data and text mining as the analysis of ‘unstructured’ data.

Depending on the software, data mining tools can be used to perform several different types of tasks. Hand and others describe five data mining tasks: 1) Exploratory data analysis (EDA), 2) Descriptive Modeling, 3) Predictive Modeling (classification and regression), 4) Discovering Patterns and Rules, and 5) Retrieval by content (2001). The output of these data mining tasks produces either a model or a pattern. Spiegler proposes that data mining technology can also be used to generate knowledge (2003).

Data mining has been used successfully in money laundering systems, identity theft services, name recognition software, and homeland security programs (Mena, 2004); however, data mining has challenges of its own. These challenges include: “synonymy, polysemy, uncertainty of language, scarcity, and human-like understanding” (Mena, 2004:251). Researchers agree that future advances in data mining technology will rely on the capability to process unstructured data (Walter, 2003; D’Amico, 2002; Mena, 2003).

Unstructured Data Challenges

Mining from unstructured data has proven to be challenging. Computers were designed to work with single letters, not words. Tim Fielden describes the problem eloquently.

> Computers only deal with words for human convenience. The only thing a computer understands about text is its American Standard Code for Information Exchange (ASCII) assignment. A word such as ‘hi’ has the same ASCII representation regardless of language, even though it does not have the same meaning. In fact, to a computer it has no meaning at all; it is simply the letter ‘H’ and the letter ‘I’ with no space in between. Therefore, making it searchable, or at least meaningfully searchable, is problematic to say the least (Fielden, 2000:88).

To further complicate the issue, the bulk of an organization’s information is in unstructured form (Mena, 2004; Meyers, 2002; Robb, 2004). “There is a distinct need for software capable of analyzing and categorizing unstructured data, a task to which computers are not innately suited” (Meyers, 2002:1). In fact, “recent studies indicate that information workers spend as much as a quarter of their time just finding and gathering job-related information. Nuanced information about trends and customer attitudes spend another quarter of their time” (Fielden, 2000:88). No doubt researchers suffer from this ‘time management’ problem as well. Today, several data mining products like visualization tools are being used to automatically “generate taxonomies and classify information” (Meyers, 2002:1) from this unstructured data.

While some organizations continue to throw technology at problems, insider threat included, others maintain a combination of humans and technology are a better approach. Mena contends that “it is in the marriage of humans and machines that the best chance of criminal detection lies” (2003:21). Due to the large volumes of data generated on a daily basis, researchers and analysts cannot physically look at every piece of data; instead they rely on the brute force of computers to assist (Uramoto, et al., 2004). “Computers enable us to view data in many different ways, both quickly and easily, and have led to the development of extremely powerful data visualization tools” (Hand et al., 2001:54).

Visualization Software

A subset of data mining, visualization tools are designed specifically for unstructured data and operate in a form conducive to the strengths of the human brain (Hand et al., 2001; Fayyad et al., 2002; Mena, 2004). The primary goal of data visualizations is "to find a view or projection of the data that reduces complexity while capturing important information" (Fayyad et al., 2002). As demonstrated by early statistical methods use of histograms and scatterplots (Fayyad et al., 2002), it is generally easier for humans to understand pictures than large amounts of text. Grinstein and Ward assert that "visualization is not a substitute for quantitative analysis" (2002:39). Rather visualization is another tool in a researcher's toolbox.

Visualization is a "mechanism to more tightly couple the user to the various applications and to harness the creative and exploratory capabilities of the human within the data analysis loop" (Fayyad et al., 2002:5). By examining the relationships of taxonomies and time lines, visualization can further aid understanding (Mena, 2004). To create the visualization, the software uses statistical clustering.

> In this method, the program uses algorithms to assess the relationship between documents. The algorithms break down a document and analyze various features statistically; this is known as feature or concept extraction. A simple example of such a feature is frequency of a particular word…The software expresses its analysis in numerical form and compares the computed values of the documents to determine their degree of similarity or difference. When graphed based on the numerical values produced by the analysis, similar documents will appear closer together (Meyer, 2002:1).

Visualization software uses several operations to support exploration including: data selection, data manipulation, representation, image orientation and viewing, and visualization interactions (Grinstein and Ward, 2002). A data selection operation allows the user to retrieve a subset of the dataset. Data manipulation operations permit the user to smooth, filter, or interpolate the data. Representation operations allow the user to modify how the data is mapped. Image orientation and viewing operations give the user the ability to manipulate the data by pan, zoom, and rotate. Visualization interactions, such as three-dimensional charts, paths, and links (Mena, 2003:126), permit the user to directly perform actions on the dataset via the graphical display.

> Visualization is a powerful technique for aiding users in understanding the data and suggesting relationships. It is weak at predictive and quantitative tasks. It does not build formal models of the data, but instead suggest models and aids the analyst in deciding what to model (Wills, 2002:708-709).

This interaction between visualization software and users utilizes exploratory analysis techniques. To assist the researchers in exploratory analysis, visualization tools, rather than the analyst, examine the dataset and cluster the data into groups based on content similarity. "Clustering is not the same as classification, where categories are usually defined by the investigator. Clustering attempts to extract categories from the data itself" (Rhodes:2002:28). Analysts then examine these automated clusters to "look at old patterns as well as new ones--both the classification of known patterns and the clustering analysis of anomalies and outliers" (Mena, 2003:276).

Exploratory Analysis

Mena contends data mining does not rely on a single methodology (Mena, 2003). In addition,

> traditional statistical techniques are more limited in their mining ability because their effectiveness depends on underlying assumptions such as data normality. Given this, the challenge lies not only in the design of new techniques, but also in developing criteria for using these techniques is specific problem domains (Rajagopalan and Krovi, 2002).

Exploratory analysis is an investigative technique that can be used to analyze the insider threat. Exploratory analysis looks for a hypothesis, unlike confirmatory analysis which starts with a hypothesis (Chen, 2001). Information visualization tools rely upon this form of analysis. Exploratory analysis is the interactive collaboration between the software and the user (Cios et al., 2001). Exploratory analysis gives the researcher the ability to guide the direction of his investigation based on each action (s)he selects. By having the ability to direct the course of research based on the users knowledge and experience, exploratory analysis tools can provide new insights that traditional statistical software packages cannot. “Discovery is an interactive process. The user dynamically both guides and is guided by the discovery process. The interaction between the two is what gives the system much of its power” (Feldman, 2002:632).

With the exception of data mining from intrusion detection logs (Wenke and Lee, 1998), most insider threat studies have either manually analyzed or used traditional statistical tools to conduct their research. Since data mining tools have been successfully used in other research areas, this research proposes to demonstrate the potential benefits of its use in insider threat research, a proof of concept if you will.

Friedman recognized that globalization requires people to use different lenses to view the world (2001). I propose that visualization software provides such a “lens” to analyze the insider threat.

Summary

This chapter examined the insider threat research that has been conducted including insider threat indicators, motives and goals, and objectives. Visualization software may provide insight into issues that may be beneficial to insider threat research. By identifying new knowledge from insider attack cases, current insider threat models may be refined and other potential solutions may be discovered. Various insider threat models and insider frameworks were examined as well as insider threat prevention and detection methods. After the literature review, data mining, unstructured data, and exploratory analysis was explained. The following chapter will discuss the methodology used to conduct this research. Chapter four will detail the results of the data analysis. Finally, Chapter five will discuss the research findings, research limitations, and recommendations for future research in this area.

III. Methodology

Overview

The previous chapters outlined the insider threat problem, reviewed previous insider threat research efforts, and described visualization capabilities that could be applied to insider threat research. This chapter outlines the methodology used to conduct the exploratory analysis on the dataset. It includes a description of the dataset, the data collection and cleaning methods, and a depiction of the visualization tool and its capabilities. Finally, the technique that will be used to analyze the dataset is explained.

Since data mining tools have been successfully used in other research areas, this research proposes to demonstrate the potential benefits of its use in insider threat research. The proposed theory is that visualization, a data mining technique, may provide potential benefits in insider threat research. Undiscovered insider threat patterns or insider threat relationships may be uncovered that have not been identified via the manual and statistical analysis methods that are widely used in insider threat research. By allowing the visualization software rather than a human being to categorize the data, the insider threat may be seen in a different light and reveal new knowledge regarding the insider threat. Since the data categorization is displayed graphically, a strength of the human brain, it may give the researcher useful insights regarding the insider threat that were previously "unseen." Additionally, instead of starting out with a hypothesis to find X out about the insider threat, by using a discovery tool such as visualization, the researcher instead seeks to discover hypotheses from within the insider threat data that may be tested in future research efforts. The next sections explain how this will be done.

Dataset

The data in this study was obtained from the United States Department of Justice (DOJ) Computer Crime and Intellectual Property Section (CCIPS) website. The data are DOJ written accounts of computer crime (US DOJ CI, 2004) and intellectual property (US DOJ IP, 2004) cases that have occurred in the United States from 1998 through 2004. The DOJ documents the cases, and then releases these electronically to the media and public via a press release. The releases include arrest, plea, indictment, and sentencing documents.

Data Characteristics.

Each of the documents is from one to two pages in length and is in .html format. For the research timeframe, there were 198 total cases--88 computer intrusion cases and 110 intellectual property case. These cases are not an exhaustive list of computer intrusion or intellectual property cases that have occurred during this time frame, but only a DOJ-provided sample of the cases that have occurred.

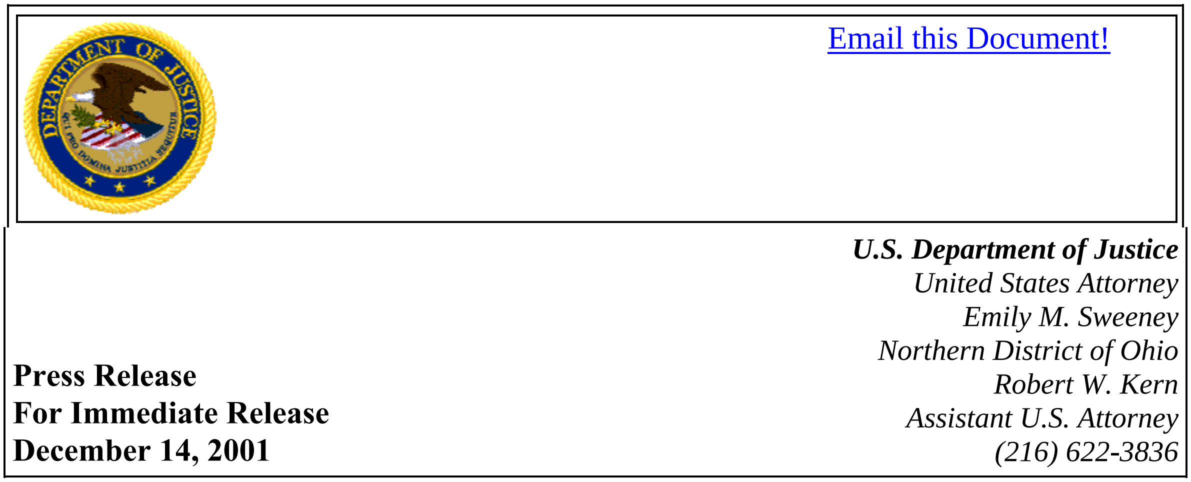
To ascertain an adequate number of insider attack cases existed in the dataset, the researcher manually reviewed the 198 DOJ cases to determine if the case represented an insider attack based on the definition of insider threat provided in Chapter two. Based on this definition, insider threat attacks were characterized as many events including: a current or former employee, an employee that passes the information to an outsider to commit a crime, a valid user of a computer network (although not necessarily an employee of that network's company—for example, a student at a university or a company shared computer network), a current or previous contractor to that company (person developed computer software or voice mail system for that company years ago), a current or former employee of the company’s network provider (person knows the company through the services he provides to them), a current or former authorized customer of a company, and a subsidiary of company. The remaining cases were considered outside attacks.

Several of the cases had more than one case within that document. These were considered a separate case and added to the insider threat or outsider case pile. In addition, one case was a duplicate so the sum of cases was reduced by one. For the computer intrusion cases, forty-one of the insider threat identified documents had forty-seven cases (six combined in another document). For the intellectual property cases, nine of the insider threat documents had eight cases (one duplicate). Overall, fifty-five insider threat cases and 143 outsider cases were identified. The researcher recognized that some insider threat cases may have been missed; however, since the study is solely based on insider threat cases, the one or two cases that may have been missed proved no harm. The 143 outsider cases were subsequently removed from the study.

To ensure validity in the researcher’s insider threat case selection, two fellow graduate students examined the fifty-five DOJ cases that the researcher determined to have been committed by an inside attacker. The first rater classified fifty-four of the fifty-five cases as insider attacks. The second rater classified the same fifty-four of the fifty-five cases as insider attacks. The one questionable case was removed from the dataset. The total dataset included fifty-four cases. Given that the minimum number of documents needed for the visualization tool is fifteen, the researcher considered fifty-four a sufficient number of cases to conduct the exploratory analysis.

Data Preparation.

Before loading the data into the software, several actions were taken to prepare the data for processing. First, the data was converted from fifty-four separate .html documents into a single .txt document. The conversion from .html was conducted to ensure IN-SPIRE^{TM} would cluster the documents based on the document text rather than the html tags (i.e. colspan, quot) that describe how to display the data within the document. The header and footer information on each case was deleted during the conversion since this information was irrelevant to the analysis. An example of the deleted header and footer information is shown in Figures 13 and 14. Next, a single row of dashes (-) was inserted as the first line of each case to ensure the software would identify the beginning of each of the fifty-four cases within this one dataset. Finally, the date and title fields of each case was labeled as *Date:* and *Title:*, respectively.

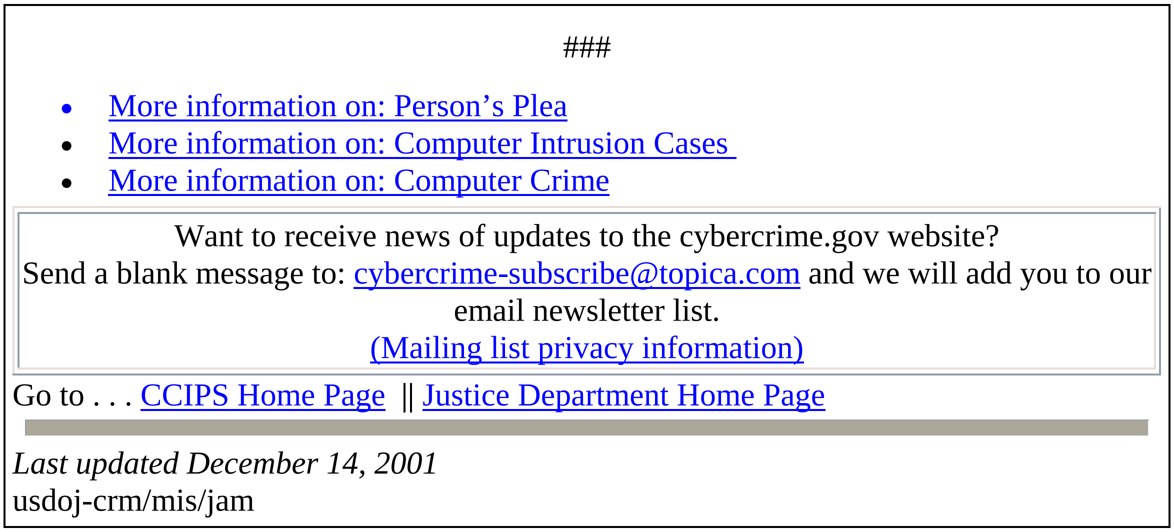


Email this Document!

U.S. Department of Justice
United States Attorney
Emily M. Sweeney
Northern District of Ohio
Robert W. Kern
Assistant U.S. Attorney
(216) 622-3836

Press Release
For Immediate Release
December 14, 2001

Figure 13. Deleted Header Information



###

- More information on: Person's Plea
- More information on: Computer Intrusion Cases
- More information on: Computer Crime

Want to receive news of updates to the cybercrime.gov website?
Send a blank message to: cybercrime-subscribe@topica.com and we will add you to our email newsletter list.
(Mailing list privacy information)

Go to . . . CCIPS Home Page || Justice Department Home Page

Last updated December 14, 2001
usdoj-crm/mis/jam

Figure 14. Deleted Footer Information

An explanation of the visualization software that will be used for the data analysis will now be described.

Visualization Tool

Originally designed for use on UNIX-based machines, Spatial Paradigm for Information Retrieval and Exploration (SPIRE) was developed by Pacific Northwest National Laboratory (PNNL) to assist the intelligence community in identifying trends, patterns, or unexpected occurrences of themes or topics within large document sets (Ginn, 2001). SPIRE uses advanced computer graphics technologies to allow the user to visually see and explore relationships among large collections of unstructured data. An information visualization software, SPIRE aids analysts in 1) identifying the fundamental nature of the dataset without having to read the entire collection of documents and 2) allowing the user to interactively guide the exploration of the dataset solely by what (s)he sees or does not see in the data. IN-SPIRE^{TM}, the SPIRE program designed for Windows platforms, is the discovery tool used to conduct this research effort.

IN-SPIRE^{TM} can process American Standard Code for Information Interchange (ASCII) or eXtensible Markup Language (XML) files. Once a dataset is loaded into the IN- SPIRE^{TM} program, the software creates a mathematical representation of the collection and organizes the documents into groups for visualization. IN- SPIRE^{TM} clusters the documents according to the most frequently occurring words and topics within the data. More specifically, IN- SPIRE^{TM} performs the following steps:

> 1. The text engine scans through the document collection and automatically determines the distinguishing words or topics within the collection, based upon statistical measurements of word distribution, frequency, and co-occurrence with other words. Distinguishing words are those that help describe each document in the dataset are different from any other document. (For example, the word "and" would not be considered a distinguishing word, because it is expected to occur frequently in every document. In a dataset where every document mentions "Iraq", "Iraq" would not be considered a distinguishing word.
>
> 2. The text engine uses these distinguishing words to create a mathematical signature for each document in the collection. Then it does a rough similarity comparison of all the signatures to create cluster groupings.
>
> 3. IN- SPIRE^{TM} compares the clusters against each other for similarity, and arranges them in high dimensional space (about 200 axes) so that similar clusters are located close together. The clusters can be thought of as a mass of bubbles, but in 200-dimensional space instead of just three.
>
> 4. That high-dimensional arrangement of clusters is then flattened down to a comprehensible two-dimensions—trying to preserve a picture where similar clusters are located close to each other, and dissimilar clusters are located far apart. Finally, the documents are added to the picture by arranging each within the invisible bubble of their respective cluster. All of this information is then mapped onto the Galaxy and ThemeView^{TM} visualizations that convey the document and topical relationships of the information (PNNL, 2004:3).

IN- SPIRE^{TM} has two visualization displays: Galaxy and ThemeView^{TM}. Galaxy visualization groups the documents as stars in the sky. The closer the stars (i.e. documents) are within the visualization, the more similar the documents' topical content will be as illustrated in Figure 15.

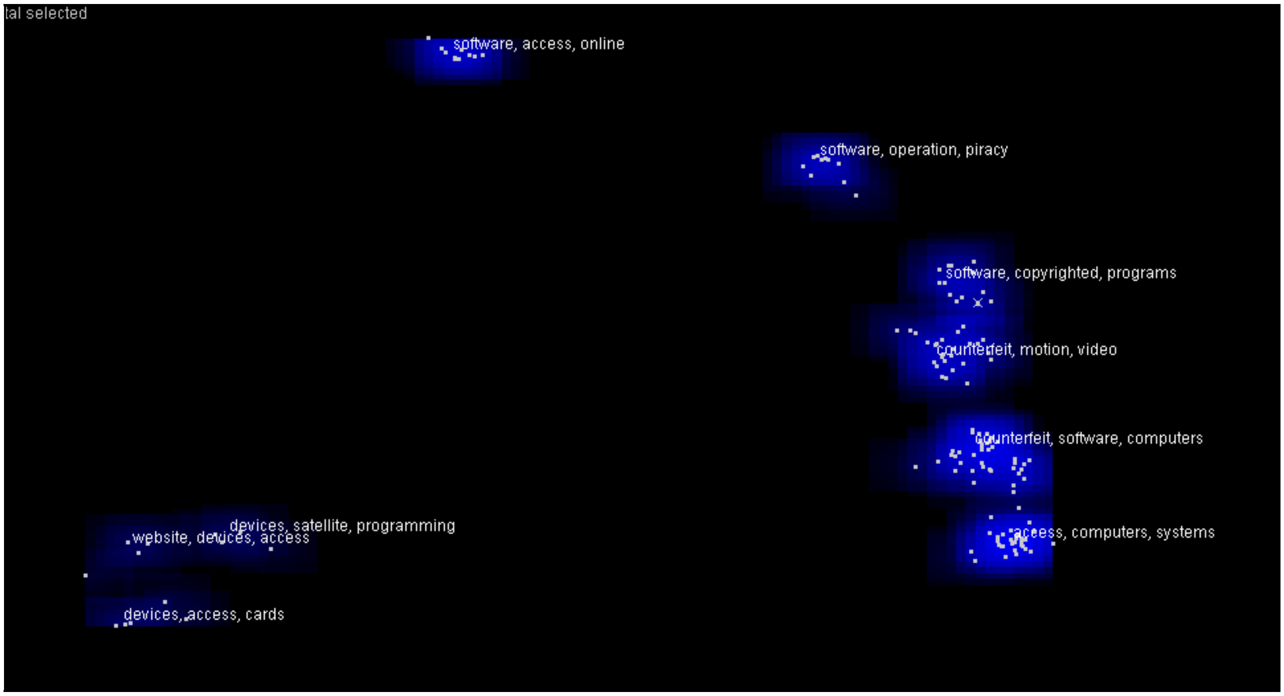

tal selected
software, access, online
software, operation, piracy
software, copyrighted, programs
counterfeit, motion, video
counterfeit, software, computers
devices, satellite, programming
website, devices, access
access, computers, systems
devices, access, cards

Figure 15. Sample Galaxy Visualization

The ThemeView^{TM} visualization displays the data on a three-dimensional terrain map. The highest peaks represent the most prevalent topics within the data. An example of ThemeView^{TM} is shown in Figure 16.

By grouping similar documents together, IN- SPIRE^{TM} reveals common themes and exposes hidden relationships within the collection that can lead to new knowledge and new insights in the area of interest. IN- SPIRE^{TM} gives analysts the ability to see something different in the data they have already collected. In this information age, analysts are overwhelmed with the amount of data that is available. Through these displays, analysts can learn which pieces of data are the most relevant and can focus their time appropriately. The documents are accessible individually, by cluster, or by the entire dataset in the document viewer.

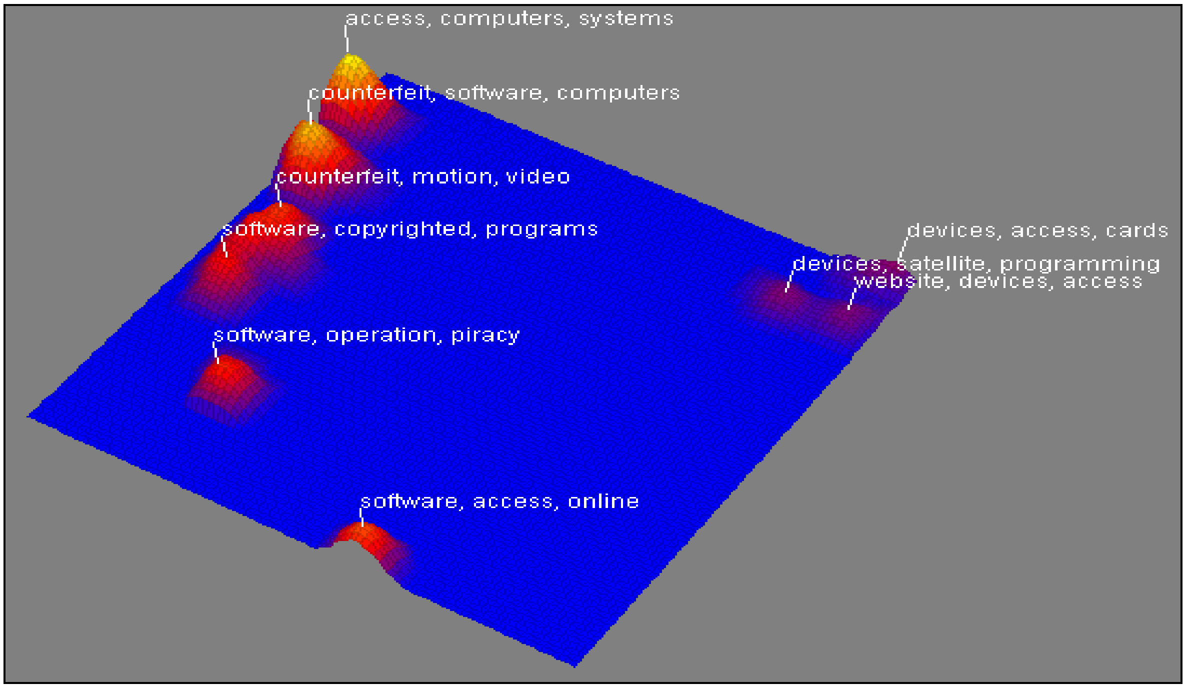

access, computers, systems
counterfeit, software, computers
counterfeit, motion, video
software, copyrighted, programs
software, operation, piracy
software, access, online
devices, access, cards
devices, satellite, programming
website, devices, access

Figure 16. Sample ThemeView^{TM} Visualization

In addition to the displays and document viewer, IN-SPIRE^{TM} provides numerous analysis tools to aid the user in the data exploration. These analysis tools give users the ability to drill down and examine other relationships within the dataset that may not be immediately apparent. The documents in the dataset can be grouped, gisted, probed, queried, and time sliced. The grouping tool allows users to assemble documents into user-defined collections. Gist provides the general idea, or essence, of a selection by displaying the most frequently used words and how many documents those words were found. The probe tool identifies the strongest topics and places them into a ranked list. Three types of queries are available—by word, by phrase, and by example. If a date field is defined, time slice permits users to view their datasets in year, month, week, or minute groupings. Now that the capabilities of the visualization software are understood, the analysis process will be explained.

Analysis Process

Initially, a pilot study was conducted with an experimental dataset. A default dataset of 425 Time magazine articles from 1963 is automatically installed when the IN-SPIRE^{TM} is loaded onto a computer. Without any a priori knowledge of the dataset contents, the researcher used the Time dataset to learn how the visualization tool worked. From this pilot study, the researcher, from trial and error, became skilled on how to analyze a dataset using the visualization software.

The first step in the visualization analysis process is to load the dataset. The dataset will be loaded as an ASCII Dataset. The document delimiter radio button will be selected and identified as a string of dashes. Two fields, date and title, will be formatted to be recognized by the software. Neither field will be used in the software computation. Stopwords, stopmajors, and punctuation rules will be set to the default options. Once these settings are entered, IN- SPIRE^{TM} will automatically process the dataset into a Galaxy and ThemeView^{TM} visualization.

Since the researcher is using a visualization tool, exploratory analysis will be used. Exploratory analysis, according to Grinstein and Ward, searches the data for structure or trends and attempts to arrive at a hypothesis (2002). Therefore, the specific analysis steps of this research cannot be determined in advance. It will be an interactive process between the software and the user. However, based on the insider threat models and frameworks that were reviewed in Chapter two, the exploratory analysis should be guided by some of the following issues:

- What types of employees committed the crimes (former, current, contractor, vendor, supplier)? (Shaw et al., 1998; DoD IPT, 1999; Denning, 1999; Brackney and Anderson, 2004; D'Arcy and Hovav; 2004)
- What type of job position was held by these criminals (network or system admin, computer or software programmer, accountant, bank teller)? (Maglakaras and Furnell, 2001)
- What types of crimes were committed (Denial of Service, hardware, software, operating system attacks)? (Anderson, 1999; Magklara and Furnell, 2001)
- What was the motive for the crime (revenge, anger, greed, money)? (Denning, 1999; Shaw et al., 1998; Chuvakin, 2003; Wood, 2000; Jarvis, 2001; Heuer, 2001; Krause, 2002)
- Did the attacker have any personal problems (drug, alcohol, mental, financial, prior arrests)? (Krause, 2002)
- Did outsourcing play a role in the crime? (Caruso, 2003)

The researcher will review both the Galaxy and ThemeView^{TM} visualizations to identify any initial "findings." The clusters and peaks will be examined to guide the usage of the analysis tools. The cluster titles will be examined to ensure the three provided terms are relevant words to describe that particular cluster. Outlier documents and terms will be examined and possibly removed. The gist, probe, and query tools will be used to analyze the clusters and to group the documents into like sets for further exploration. The analysis will be complete when significant findings can longer be discovered from the visualizations and the following research questions can be answered.

1. Using exploratory analysis, how can visualization tools be useful in highlighting patterns or relationships in insider attack case data?
2. Can visualization software assist in generating hypotheses for future insider threat research?

Summary

This chapter described the methodology used in conducting the insider threat analysis research. The dataset was illustrated, the visualization software was explained, and the analysis process was defined. In the following chapter, the results of the visualization data analysis are summarized. Chapter five presents the conclusions and recommendations for the overall study and suggestions for further research.

IV. Data Analysis

Overview

The previous chapters outlined the current problem statement, reviewed literature pertaining to insider threat research, and presented the research questions examined in this study. In addition, Chapter three described the data and outlined the methodology for analyzing the insider threat case data. This chapter examines the results of the exploratory analysis conducted on the Department of Justice (DOJ) data.

Results of Exploratory Analysis

Based on the insider threat models and frameworks that were reviewed in Chapter two, the researcher identified the following insider threat issues that the exploratory analysis should at the very least investigate: types of employees who committed the crimes, the job position of the employee, the type of crime committed, the motive for the crime, any personal problems of the attacker, and did outsourcing play a role in the crime. As these themes were examined, other areas of interest that were identified during the analysis were examined as well.

Document Clusters.

The fifty-four insider threat cases were loaded into a single insider threat dataset. The initial visualization of the dataset displayed the following Galaxy and ThemeView^{TM} visualizations as shown in Figures 17 and 18.

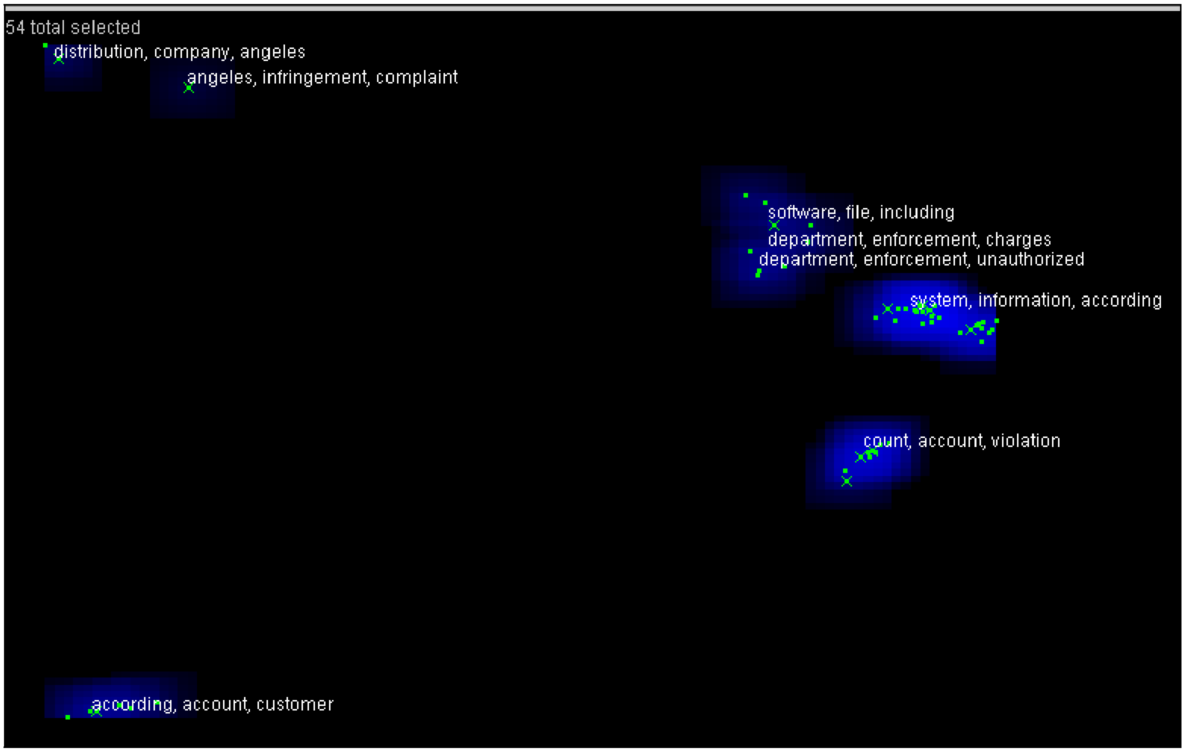

54 total selected
distribution, company, angeles
angeles, infringement, complaint
software, file, including
department, enforcement, charges
department, enforcement, unauthorized
system, information, according
count, account, violation
according, account, customer

Figure 17. Initial Galaxy view of Insider Dataset

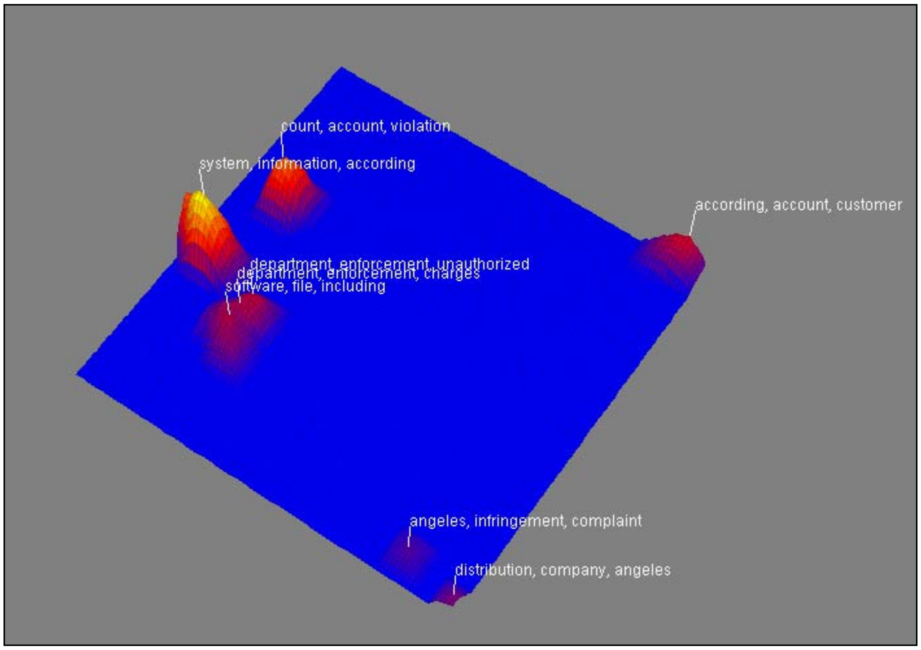
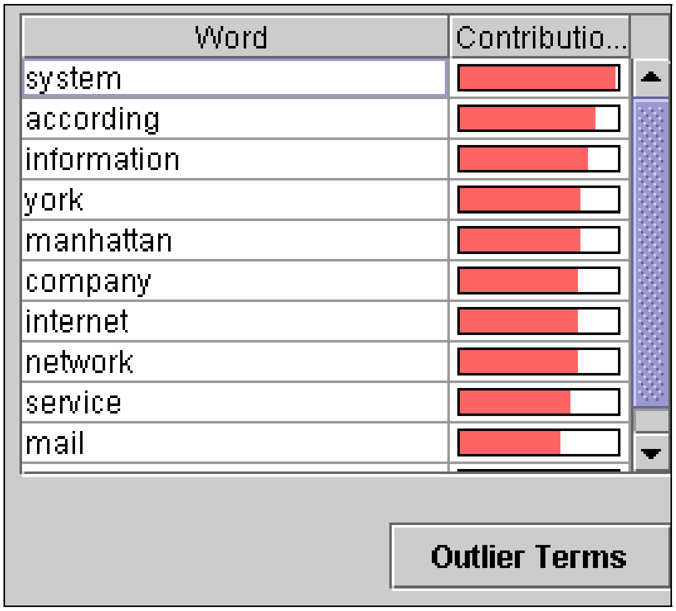

count, account, violation
system, information, according
according, account, customer
department, enforcement, unauthorized
department, enforcement, charges
software, file, including
angeles, infringement, complaint
distribution, company, angeles

Figure 18. Initial Themeview^{TM} of Insider Dataset

Several of the clusters had irrelevant cluster terms. In order to understand the contents of the insider threat cases, the probe tool was used to show the high frequency words and their relative weighting in each of the clusters. For example, the probe for the “system, information, according” cluster title on the middle right of the Galaxy display (in Figure 17) is displayed in Figure 19.



Word	Contributio...
system	🟥
according	🟥
information	🟥
york	🟥
manhattan	🟥
company	🟥
internet	🟥
network	🟥
service	🟥
mail	🟥

Figure 19. Probe Analysis Tool

To remove some of the irrelevant cluster title terms that impeded the understanding of those particular clusters (such as according and information), terms were moved to the outlier panel. This is done by selecting a term from the ‘word’ column in Figure 19 and clicking the ‘Outlier Terms’ button. However, once the visualization was recalculated with those terms removed, more irrelevant terms appeared. The researcher continued to move irrelevant terms to the outlier panel from each of the clusters and recalculated the visualization until the cluster titles were more telling. Moving these terms to the outlier panel did not remove the terms from the dataset; the terms are still available for queries, gists, probes, and other analysis tools. The terms are simply not included as words in the cluster titles. All told, fifty-nine terms were moved to the outlier panel. These terms are included in Table 4.

Accessed	Charges	Defendants	Five	Northern	Set
According	Chip	Department	Formerly	Pleaded	Seven
Admitted	Company	Distribution	Including	Received	Statement
Agent	Complaint	Employees	Indictment	San	System
Alleges	Conduct	Enforcement	Information	Secret	Term
Angeles	Conspiracy	Evidence	Infringement	Section	Unauthorized
Bank	Copied	False	Internet	Seized	Unit
Business	Copies	Fbi	Investigations	Sentenced	Violation
California	Count	File	Manhattan	Sentencing	York
Charged	Counts	Files	Months	Service	

Table 4. Cluster Terms Removed to Outlier Term Panel

Once the terms were removed, the new Galaxy and ThemeView^{TM} visualizations were recalculated and are illustrated in Figures 20 and 21, respectively.

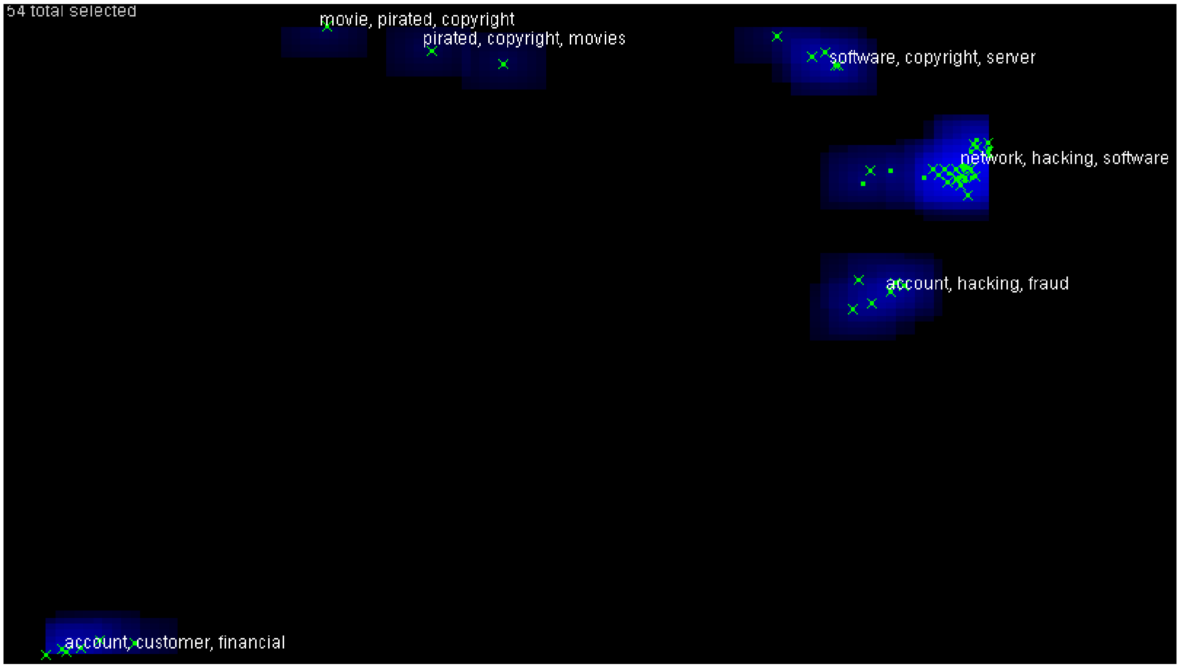

54 total selected
movie, pirated, copyright
pirated, copyright, movies
software, copyright, server
network, hacking, software
account, hacking, fraud
account, customer, financial

Figure 20. Second Galaxy Visualization

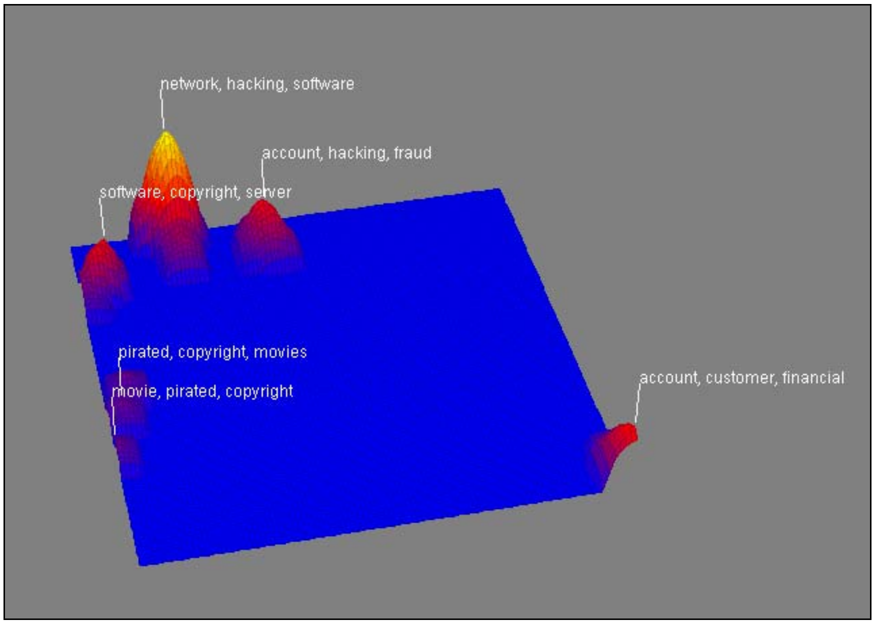

network, hacking, software
account, hacking, fraud
software, copyright, server
pirated, copyright, movies
movie, pirated, copyright
account, customer, financial

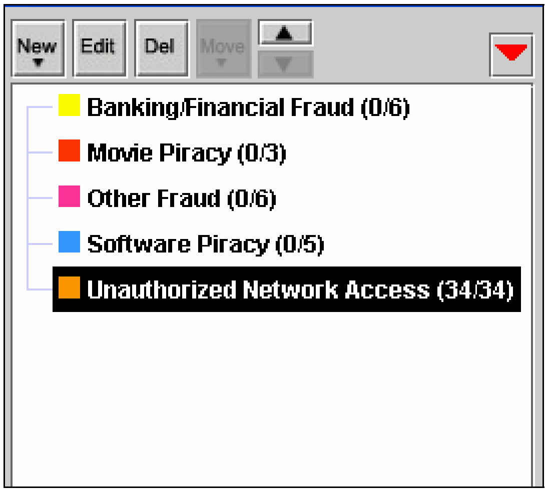
Figure 21. Second ThemeView^{TM} Visualization

Cluster Groups.

From these new visualizations, the researcher was able to identify five unique clusters of cases: movie crimes, software crimes, network hacking crimes, fraud crimes, and financial crimes. The researcher classified these five clusters into the following groups (number of cases within that group):

- Movie piracy (3)
- Software piracy (5)
- Banking/Financial Fraud (6)
- Other Fraud (6)
- Unauthorized network access (34)

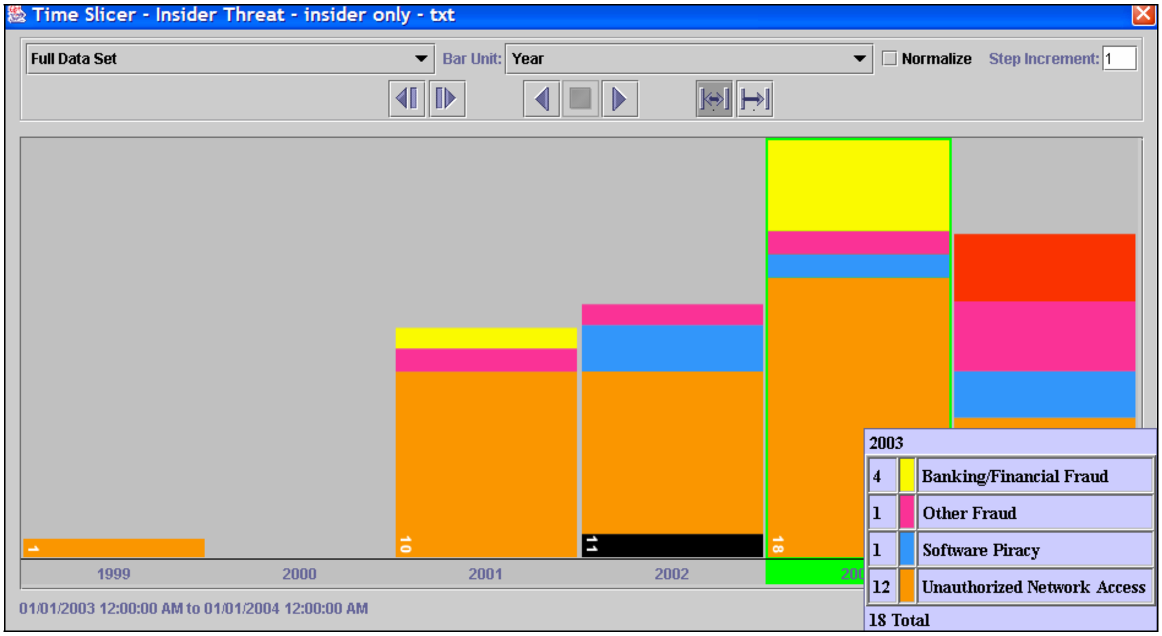
Figure 22 illustrates how each group is assigned a separate color to identify the documents in its respective group as well as the number of cases contained in that group. If this group is selected, the documents within that group displays in that color in the Galaxy visualizations.



🟨 Banking/Financial Fraud (0/6)
🟥 Movie Piracy (0/3)
🟪 Other Fraud (0/6)
🟦 Software Piracy (0/5)
🟧 Unauthorized Network Access (34/34)

Figure 22. Defined Groups

Next, the data was viewed by the case dates to see how these cases occurred during the 1998 to 2004 timeframe. In the Time Slicer analysis tool, the color bands correlate to the assigned group colors; the wider the band, the more cases exist for that group. Furthermore, each year can be displayed with the number of cases that fall into each group for that year. Figure 23 illustrates the Time Slicer view for the groups' cases arranged by date. (Note none of the 198 cases were classified as an insider threat case from 1998.)



Time Slicer - Insider Threat - insider only - txt

Full Data Set

Bar Unit: Year

□ Normalize

Step Increment: 1

01/01/2003 12:00:00 AM to 01/01/2004 12:00:00 AM

2003		
4	🟨	Banking/Financial Fraud
1	🟪	Other Fraud
1	🟦	Software Piracy
12	🟧	Unauthorized Network Access
18 Total		

Figure 23. Time Slicer Group View by Year

To ensure the clusters were an accurate depiction of the dataset, the evidence panel was used to review the cases in each group to see if each case actually fit the "profile" of the defined group. Upon review of the case text, the following changes were made:

- An additional group, Malicious Code, was added after noticing a large amount of Unauthorized Network Access cases were malicious code or logic bombs attacks.

- The terms *malicious*, *code*, *logic*, and *bomb* were added to the highlighting panel as the remainder of the Unauthorized Network Access cases was reviewed.

- One case (Fox cable network- 5/14/2004) was ID'd in the Music Piracy group only. Upon review, it was co-grouped into the Software Piracy group.

- A second case (Alta vista source code- 7/02/2004) was ID'd in the Other Fraud group. Upon review, it was moved into the Unauthorized Network Access group.

- A third case (Hulk movie- 6/25/2003) was ID'd in the Unauthorized Network Access group. Upon review, it was moved to the Movie Piracy group.

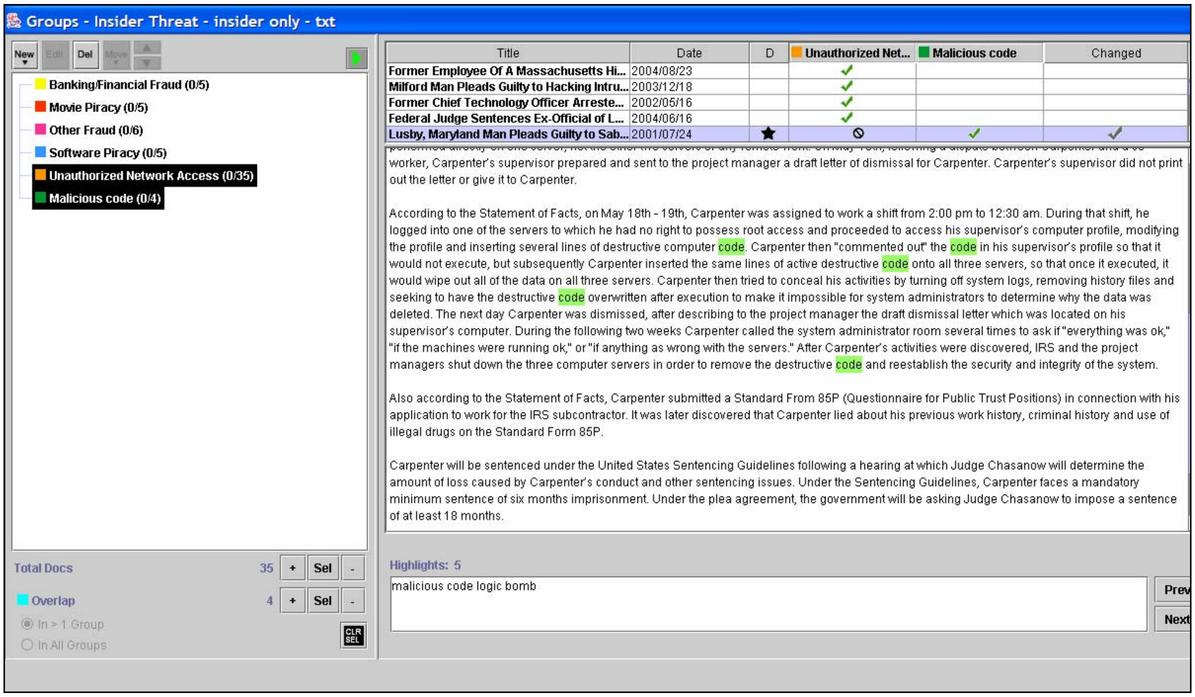
- A fourth case (IRS- 7/24/2001) was ID'd in the Unauthorized Network Access group. Upon reviewing the highlights, it was moved to the Malicious Code group.

- A fifth case (Paine Weber- 2/17/2002) was ID'd in the Unauthorized Network Access group. Upon review, it was moved to the Malicious Code group.

- A six case (Omega- 2/26/2002) was ID'd in the Unauthorized Network Access group. Upon review, it was moved to the Malicious Code group.

- A seventh case (Lance- 4/13/2001) was ID'd in the Unauthorized Network Access group. Upon review, it was moved to the Malicious Code group.

Based on this review, seven cases of fifty-four were recoded. Figure 24 illustrates the evidence panel with both the document viewer window and the highlight panel.



Title	Date	D	🟧 Unauthorized Net...	🟩 Malicious code	Changed
Former Employee Of A Massachusetts Hi...	2004/08/23		✓		
Milford Man Pleads Guilty to Hacking Intru...	2003/12/18		✓		
Former Chief Technology Officer Arreste...	2002/05/16		✓		
Federal Judge Sentences Ex-Official of L...	2004/06/16		✓		
Lusby, Maryland Man Pleads Guilty to Sab...	2001/07/24	★	⃠	✓	✓

worker, Carpenter's supervisor prepared and sent to the project manager a draft letter of dismissal for Carpenter. Carpenter's supervisor did not print out the letter or give it to Carpenter.

According to the Statement of Facts, on May 18th - 19th, Carpenter was assigned to work a shift from 2:00 pm to 12:30 am. During that shift, he logged into one of the servers to which he had no right to possess root access and proceeded to access his supervisor's computer profile, modifying the profile and inserting several lines of destructive computer code. Carpenter then "commented out" the code in his supervisor's profile so that it would not execute, but subsequently Carpenter inserted the same lines of active destructive code onto all three servers, so that once it executed, it would wipe out all of the data on all three servers. Carpenter then tried to conceal his activities by turning off system logs, removing history files and seeking to have the destructive code overwritten after execution to make it impossible for system administrators to determine why the data was deleted. The next day Carpenter was dismissed, after describing to the project manager the draft dismissal letter which was located on his supervisor's computer. During the following two weeks Carpenter called the system administrator room several times to ask if "everything was ok," "if the machines were running ok," or "if anything as wrong with the servers." After Carpenter's activities were discovered, IRS and the project managers shut down the three computer servers in order to remove the destructive code and reestablish the security and integrity of the system.

Also according to the Statement of Facts, Carpenter submitted a Standard From 85P (Questionnaire for Public Trust Positions) in connection with his application to work for the IRS subcontractor. It was later discovered that Carpenter lied about his previous work history, criminal history and use of illegal drugs on the Standard Form 85P.

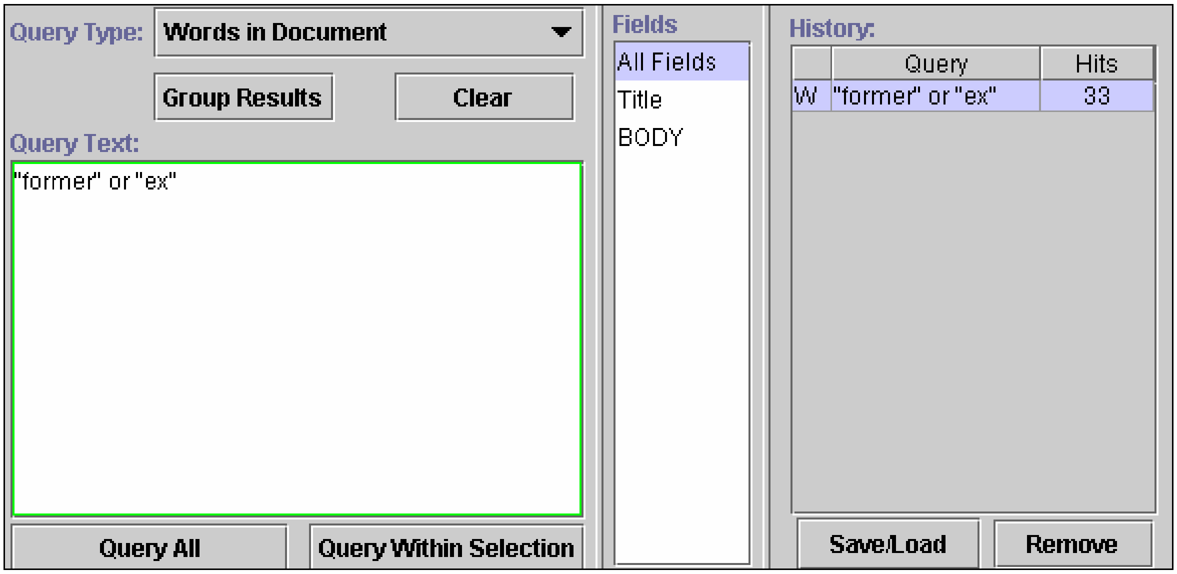
Carpenter will be sentenced under the United States Sentencing Guidelines following a hearing at which Judge Chasanow will determine the amount of loss caused by Carpenter's conduct and other sentencing issues. Under the Sentencing Guidelines, Carpenter faces a mandatory minimum sentence of six months imprisonment. Under the plea agreement, the government will be asking Judge Chasanow to impose a sentence of at least 18 months.

malicious code logic bomb

Figure 24. Evidence Panel and Highlights

Queries.

Next, the researcher began to query the dataset for unique terms that may prove supportive to the existing insider threat models and frameworks. The queries first focused on type of employee. A query was conducted on the terms *"former" or "ex"* to see how many of the crimes were committed after the insiders employment ended. Thirty-three of the case documents were identified (Figure 25). These thirty-three documents were located in the following groups (Figure 26). Interestingly, in at least half of each of the cases within all the groups, a former or ex employee committed the crime. It is important to keep in mind throughout the following queries that the visualization tool simply found the word within 'X' number of documents. The context of the terms former or ex may not be used with employee, for example. The results are merely indication, not confirmed or substantiated in any way.


Query Type: Words in Document

Query Text:

"former" or "ex"

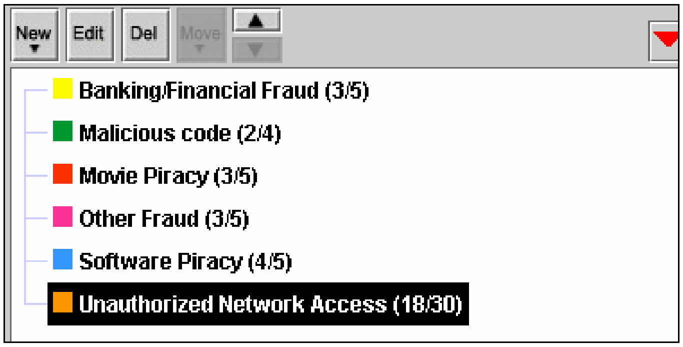
Fields

- All Fields
- Title
- BODY

History:

	Query	Hits
W	"former" or "ex"	33

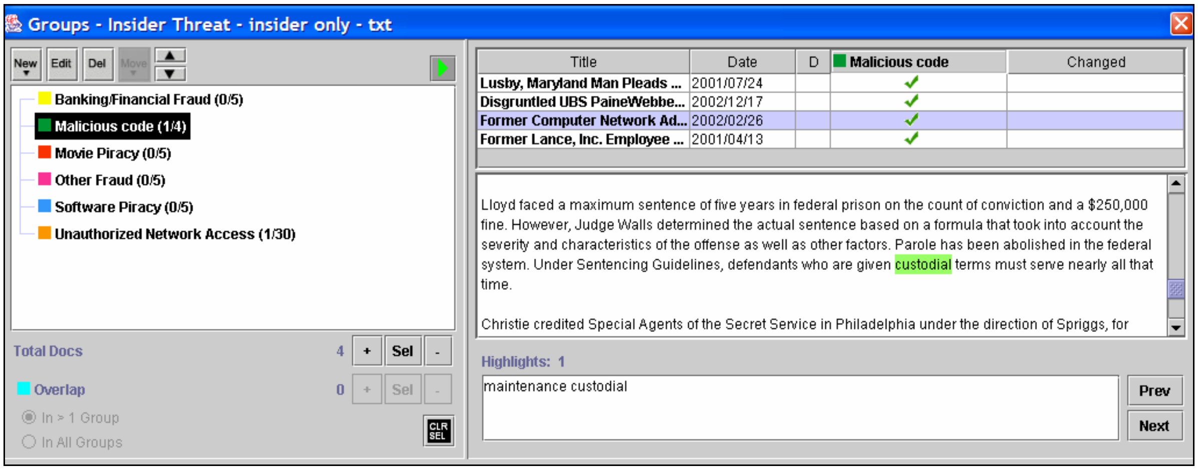
Figure 25. Former/Ex Query View



- 🟨 Banking/Financial Fraud (3/5)
- 🟩 Malicious code (2/4)
- 🟥 Movie Piracy (3/5)
- 🟪 Other Fraud (3/5)
- 🟦 Software Piracy (4/5)
- 🟧 Unauthorized Network Access (18/30)

Figure 26. Former/Ex Group View

Based on this previous 'finding', a query was performed on the terms *"current" or "present"* to see how many of the crimes may have been committed during the insider's employment with the organization. Only two of the fifty-four cases were identified as current or present employees. One case occurred in the Banking/Financial Fraud group, the other in the Unauthorized Network Access group. Another query was conducted on the term *"subcontractor"* to see how many cases had a subcontractor commit the crime. Only one case was identified. Similarly, a query was performed on the term *"student"* to see how many cases had a student commit the crime. Only one case was identified. Finally, a query was conducted on the term *"maintenance" or "custodial"* to see how many cases had a maintenance or custodial person commit the crime. Only two cases were identified. Skeptical that either term could be used in another context, the researcher highlighted and stepped through the two cases. The term maintenance and custodial were used in another context so neither of the identified cases were performed by maintenance or custodial personnel. Figure 27 demonstrates this review.



Title	Date	D	🟩 Malicious code	Changed
Lusby, Maryland Man Pleads ...	2001/07/24		✓	
Disgruntled UBS PaineWebbe...	2002/12/17		✓	
Former Computer Network Ad...	2002/02/26		✓	
Former Lance, Inc. Employee ...	2001/04/13		✓	

Lloyd faced a maximum sentence of five years in federal prison on the count of conviction and a $250,000 fine. However, Judge Walls determined the actual sentence based on a formula that took into account the severity and characteristics of the offense as well as other factors. Parole has been abolished in the federal system. Under Sentencing Guidelines, defendants who are given custodial terms must serve nearly all that time.

Christie credited Special Agents of the Secret Service in Philadelphia under the direction of Spriggs, for

Highlights: 1

maintenance custodial

Figure 27. Out of Context Term Review

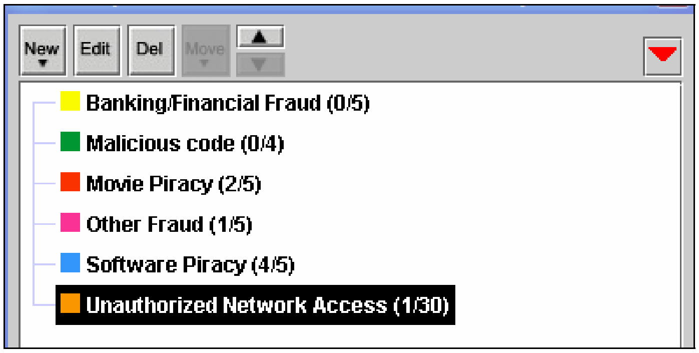
Next, the researcher focused the queries on <u>job position</u>. A query was performed on the term *"network administrator" or "system administrator" or "network admin" or "system admin" or "sys admin" or "administrator"* to see how many cases had a Network or System Administrators abuse their position to commit the crime. Fourteen cases were identified with nine of them located within the Unauthorized Network Access group. A query was also conducted on the term *"computer programmer" or "software programmer" or "programmer"* to see how many cases had a Computer or Software Programmer abuse their position to commit the crime. Only one case was identified, also in the Unauthorized Network Access group.

Yet another group of queries focused on the <u>type of crime committed</u>. A query was conducted on the terms *"Denial of Service" or "DOS" or "DoS"* to see how many cases were involved in DoS type crimes. three cases were identified as a DoS attack. Another query was performed on the term "hardware" to see how many cases had involved a computer hardware crime. Five cases were identified. Finally, a query was conducted on the term *"software"* to see how many cases had involved a computer software crime. Nineteen cases were identified in the following groups: three in Malicious Code, one in Movie Piracy, one in Other Fraud, five in Software Piracy, and ten in Unauthorized Network Access.

The queries next examined <u>motive of crime</u>. A query was performed on the terms *"motive" or "motives"* to see how many cases had a specified motive. Only one case was identified. Additionally, a query was conducted on the terms *"revenge" or "retaliation"* to see how many cases were conducted for this type of motive. Two cases were identified.

Two areas worthy of examination that did not produce any findings were <u>personal problems of the attacker</u> and <u>outsourcing role</u>.

Next, a query was performed on the terms *"group" or "ring"* to see how many crimes may be committed by people belonging to <u>groups or crime rings</u>. Seven cases were identified with the cases distributed in four groups as demonstrated in Figure 28.



- 🟨 Banking/Financial Fraud (0/5)
- 🟩 Malicious code (0/4)
- 🟥 Movie Piracy (2/5)
- 🟪 Other Fraud (1/5)
- 🟦 Software Piracy (4/5)
- 🟧 Unauthorized Network Access (1/30)

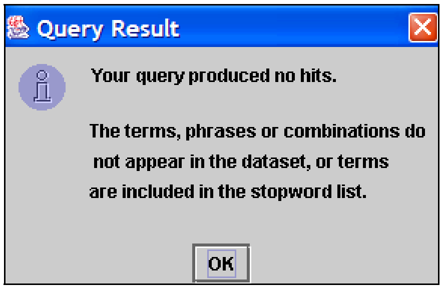
Figure 28. Group or Ring Query Distribution by Group

Two additional queries were performed outside of these predefined categories. A query was conducted on the term *"United State Code"* to see how many cases identified the law that was broken in the crime. The researcher was curious to determine if certain laws were used for prosecution more than others in the dataset. Only seven cases were identified. Also, a query was performed on the term *"password"* to see how many crimes may have been committed by a password-type vulnerability. Eleven insider attacks identified password. Nine of the cases occurred within the Unauthorized Network Access group; two of the cases occurred in the Software Piracy group. The terms confidentiality, integrity, and availability were also queried. Theses queries produced zero, one, and zero hits, respectively.

Finally, queries were performed on a handful of terms listed in Table 5; however no cases were identified by these queries as illustrated by the query result in Figure 29.

vendor	supplier
“network provider”	behavior
contractor	anger
drug	greed
alcohol	outsourcing
“mental illness”	Help desk
“prior arrest”	“operating system” or “OS”
“financial problem”	

Table 5. Queries Performed with No Results Returned



Your query produced no hits.

The terms, phrases or combinations do not appear in the dataset, or terms are included in the stopword list.

Figure 29. Negative Query Result Display

Visualization Outliers.

The next step in the exploratory analysis is to manipulate the visualization by removing outliers from the display. Like the outlier terms discussed earlier, outliers remain in the dataset and can be queried, gisted, and probed. However, the clusters within the visualization display are mathematically recalculated giving a new view of the dataset without these less significant documents. From the original Galaxy visualization in Figure 17, six cases in the ‘*account, customer, financial*’ cluster (Banking/Finance group) and three cases in the ‘*pirated, copyright, movies*’ cluster (Movie Piracy group) were moved to the outlier panel. The recalculated visualizations are displayed in Figure 30 and 31.

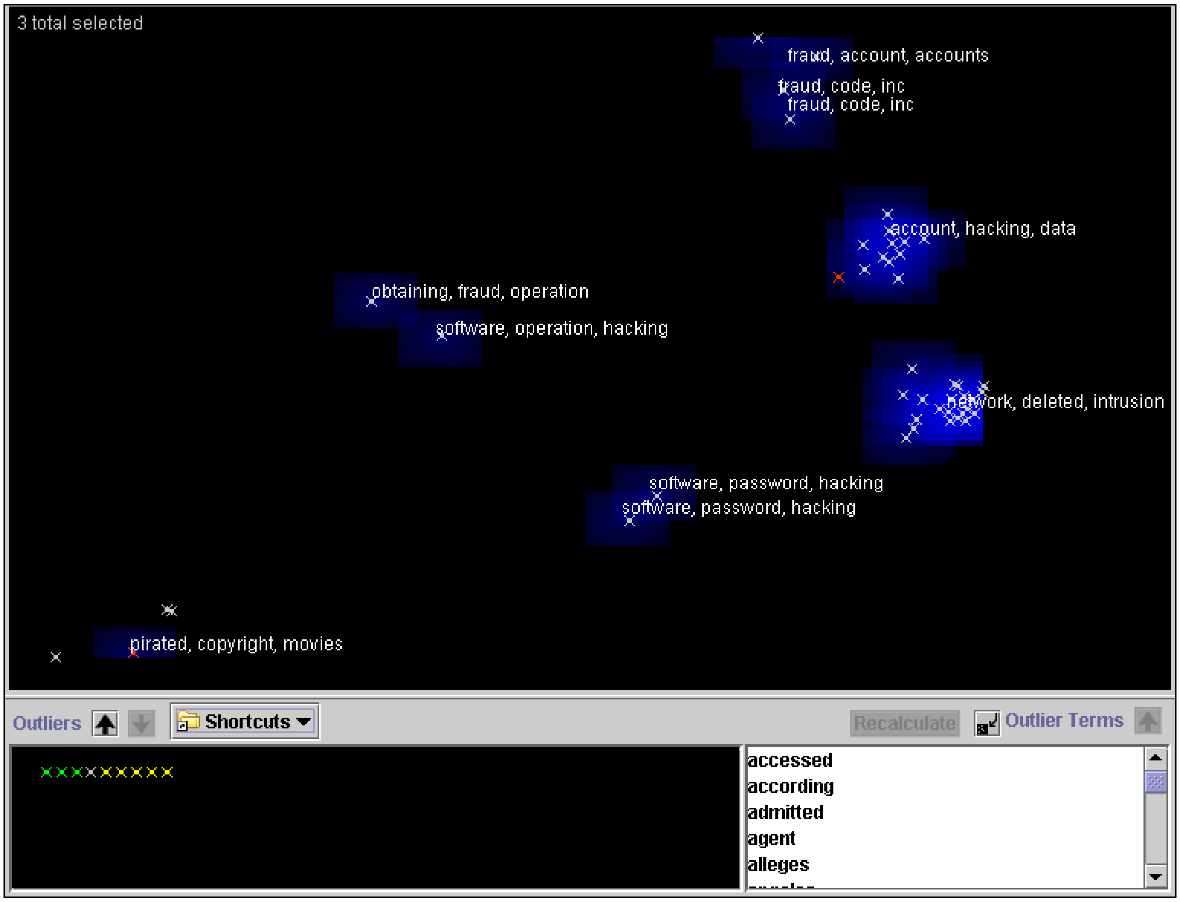


Figure 30. Recalculated Galaxy View #1 (same as Figure 17)

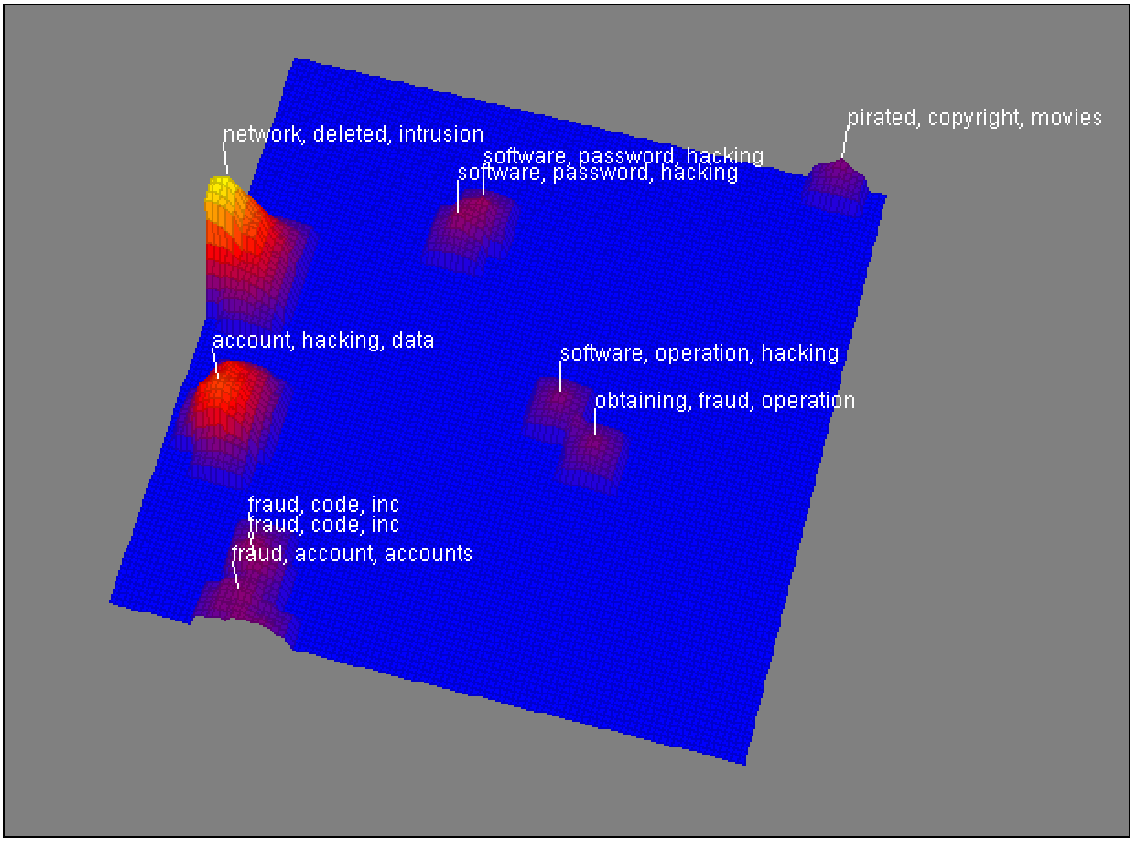
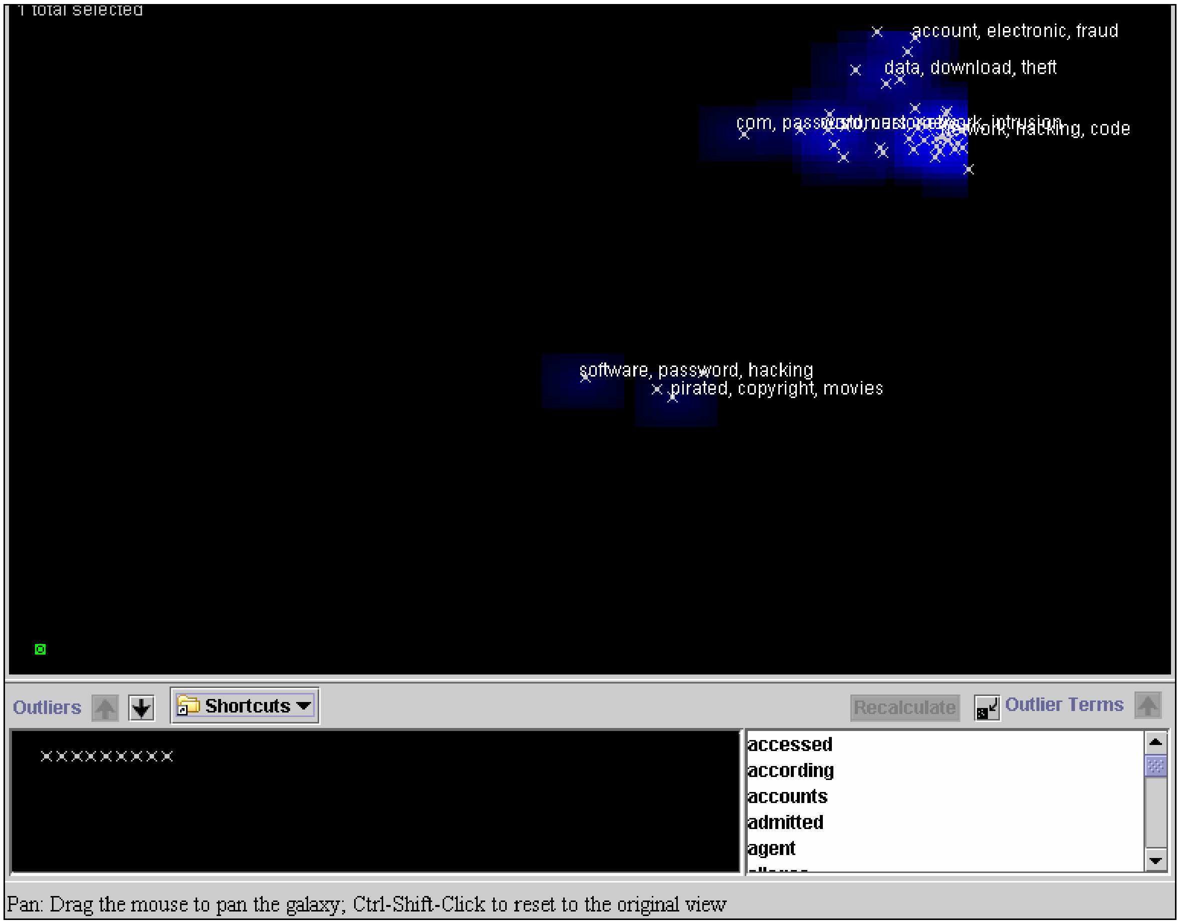

network, deleted, intrusion
software, password, hacking
software, password, hacking
pirated, copyright, movies
account, hacking, data
software, operation, hacking
obtaining, fraud, operation
fraud, code, inc
fraud, code, inc
fraud, account, accounts

Figure 31. Recalculated ThemeView^{TM} #1

Next, the cluster titles terms were examined for relevancy. The words “*inc*” and “*accounts*” were moved to the outlier terms panel and the visualization recalculated. The new Galaxy visualization is located in Figure 32. The researcher found the Galaxy view more informative and interactive than the ThemeView^{TM} visualization so further visualization figures will only include the Galaxy display.



Pan: Drag the mouse to pan the galaxy; Ctrl-Shift-Click to reset to the original view

Figure 32. Recalculated Galaxy View #2

The researcher noted that the outlier in the bottom left of the display may be causing the other document clusters to be compressed. This outlier, a software piracy document, was moved to the outlier panel and the visualization recalculated as illustrated in Figure 33. Also, to gain further understanding as to which type of group documents were appearing in which clusters, the groups were highlighted to display in their respective colors (Figure 34).

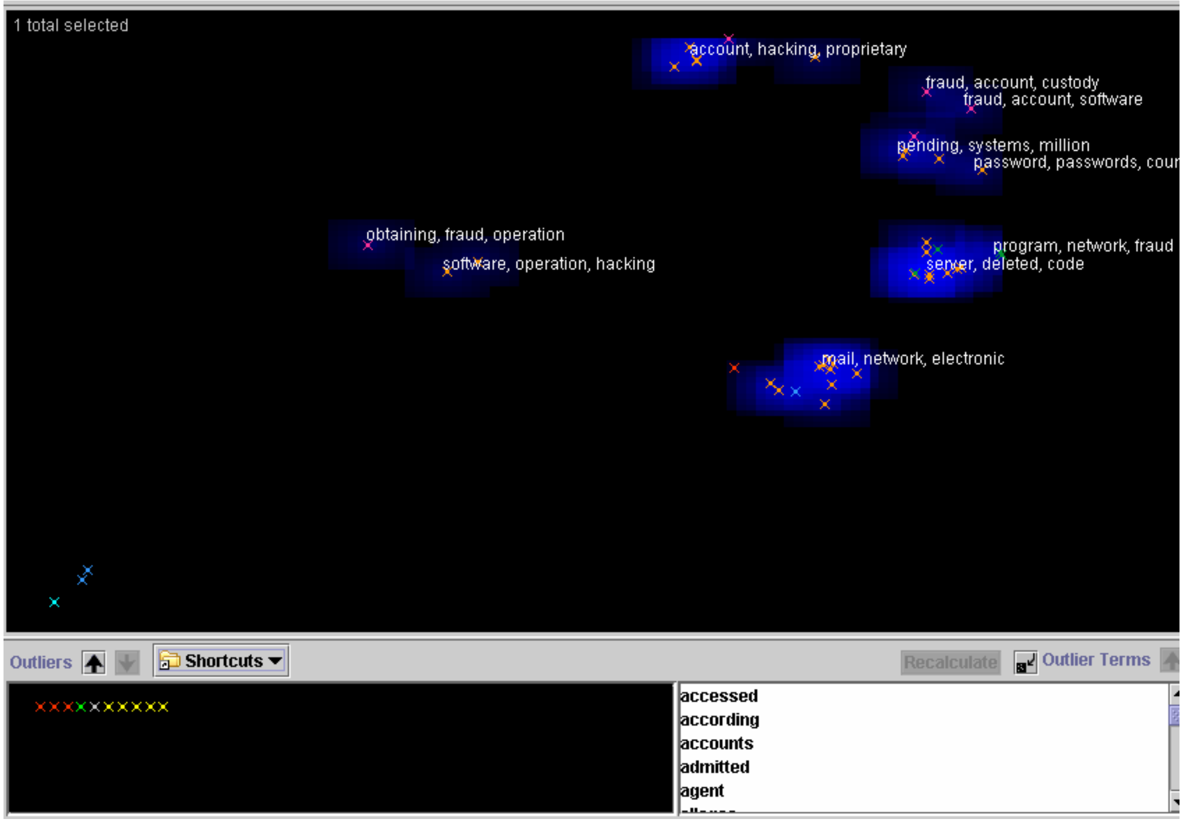
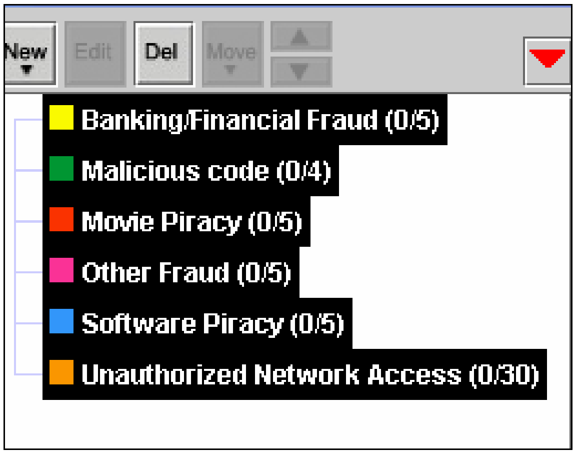


Figure 33. Recalculated Galaxy View #3



🟨 Banking/Financial Fraud (0/5)
🟩 Malicious code (0/4)
🟥 Movie Piracy (0/5)
🟪 Other Fraud (0/5)
🟦 Software Piracy (0/5)
🟧 Unauthorized Network Access (0/30)

Figure 34. Group Color Identification

The three outliers in the bottom left corner of the Galaxy display were software, music, and movie piracy cases. These three documents were moved to the outlier panel.

The remaining cluster titles were examined for irrelevant terms; the words "*obtaining*", "*april*", "*passwords*", and "*comey*" were moved to the outlier term panel and the visualization was recalculated (Figure 35).

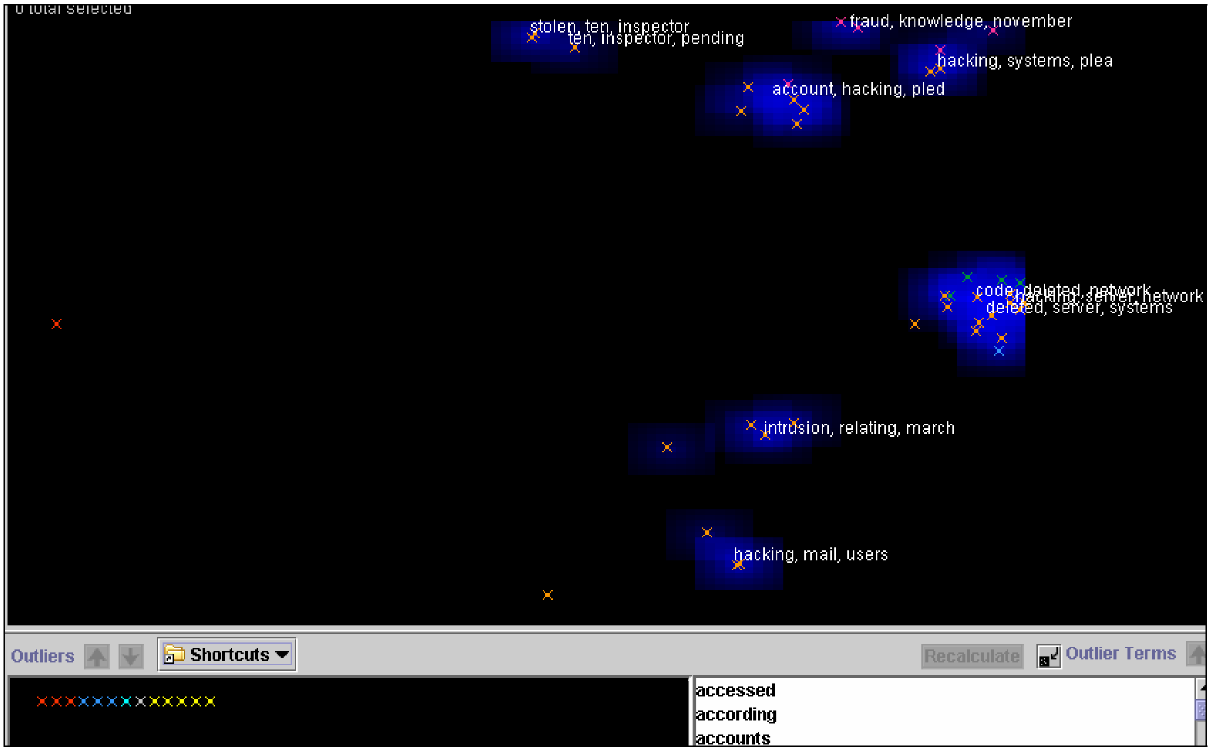


Figure 35. Recalculated Galaxy View #4

It appeared that the majority of the documents remaining in the visualization were narrowed down to the Unauthorized Network Access group documents. Thirteen terms were in the outlier panel and all but eleven of the remaining forty-one cases in the visualization were Unauthorized Network Access group documents. Next, the document on the middle left of the visualization and the four documents in the upper right, cluster title "*fraud*, *knowledge*, *November*," were moved to the outliers panel. The recalculated visualization appears in Figure 36.

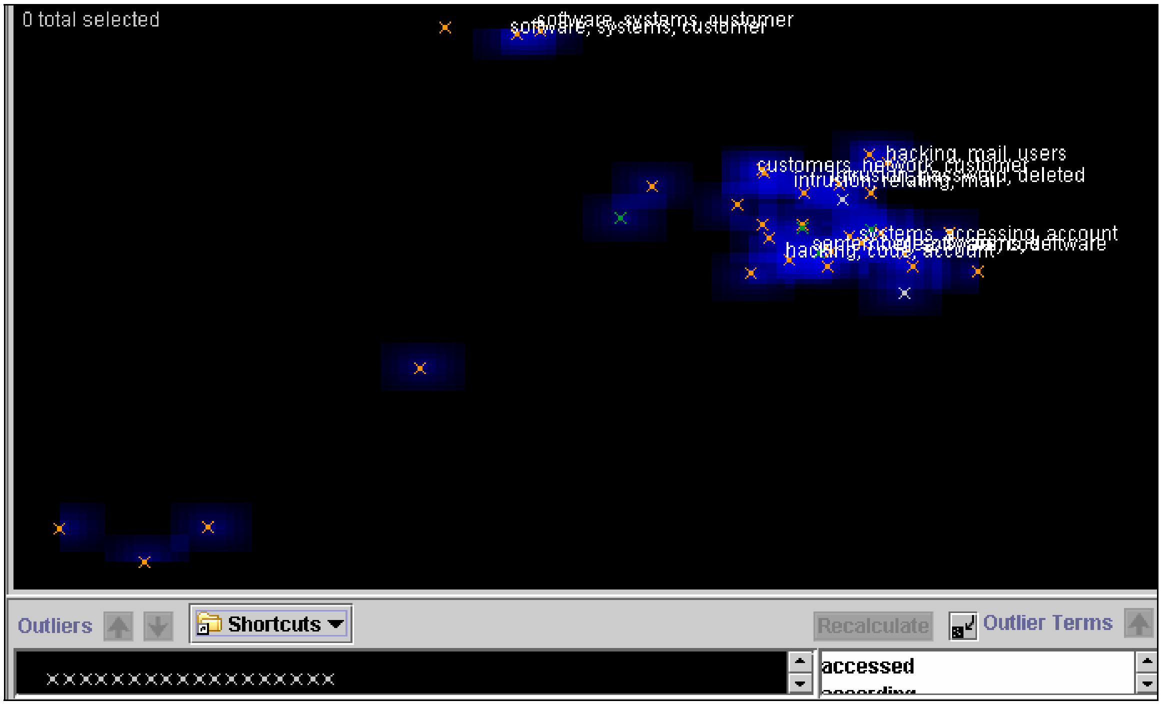

0 total selected
software, systems, customer
software, systems, customer
hacking, mail, users
customers, network, customer
intrusion, relating, mail
deleted
systems, accessing, account
hacking, code, account
software
Outliers
Shortcuts
Recalculate
Outlier Terms
accessed
according

Figure 36. Recalculated Galaxy View #5

The probe tool indicates the remaining thirty-six documents in the Galaxy, of which thirty-one are located in the Unauthorized Network Access group, seem to be related to hacking of some sort. A query was conducted on the terms *"hacking" or "hack" or "hacker"* to examine this insight. Twenty-one of these documents were identified as a hacking type case. Also, a query was conducted on the term “*cracker*”; zero cases were identified. The three documents in the bottom left were moved to the outlier panel and the visualization was recalculated. Again, irrelevant cluster title terms, “*michael*” and “*million*” were also removed to the outlier terms panel and recalculated (Figure 37).

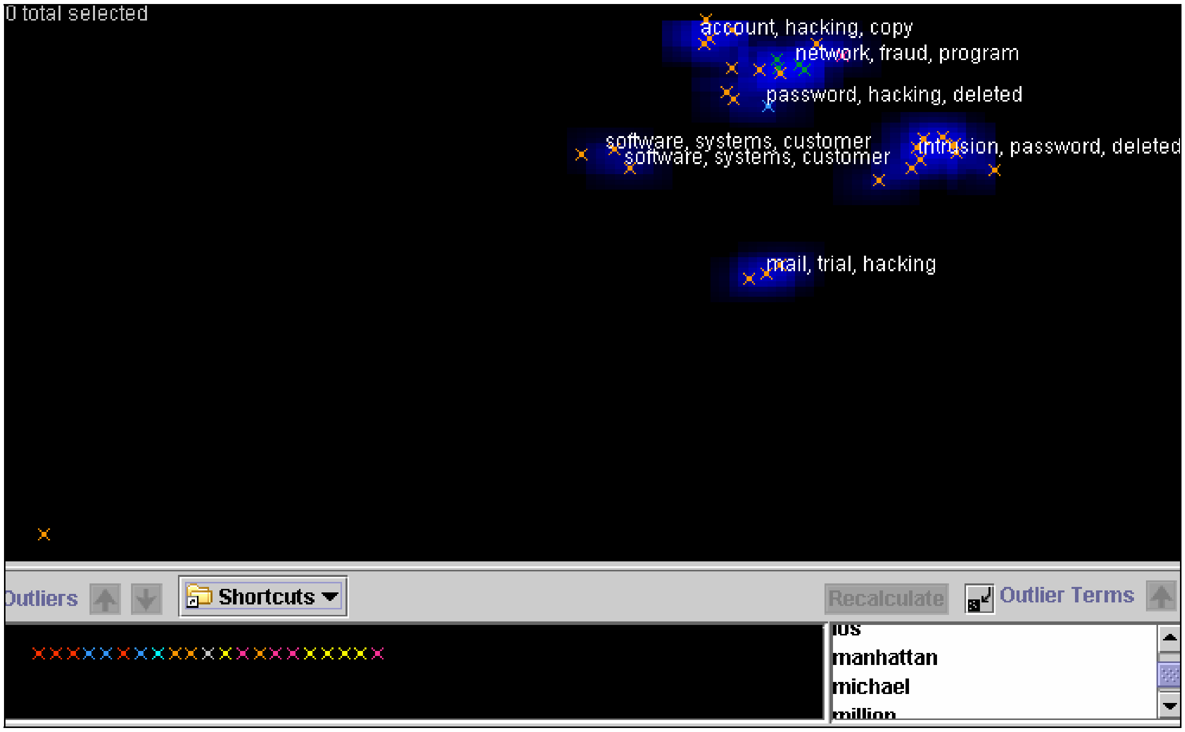


Figure 37. Recalculated Galaxy View #6

The nucleus of the remaining visualization documents centers on these Unauthorized Network Access group documents. With the exception of the outlier in the bottom left hand corner, the documents have been categorized into specific types of unauthorized access crimes. Interestingly, the outlier document is a company crime (internet service provider) that betrayed its customer's trust, rather than an individual crime that betrayed an organization's trust. This was the only organizational crime case in the dataset.

The researcher removed this outlier and conducted some further analysis; however, future visualizations did not produce any additional insight into the dataset or the insider threat problem so the exploratory analysis was ended.

Supplementary Analysis

Although not part of the initial research focus, the researcher limitedly tested to see if the exploratory analysis could be replicated. Interestingly enough, when the same dataset was loaded under a pseudonym, IN-SPIRE^{TM} created the same Galaxy and ThemeView^{TM} visualization with identical cluster title terms. The researcher then removed irrelevant cluster title terms to the outlier term panel as was done previously, however, not necessarily in the same cluster order. If the outlier terms were not added in the same order as the original analysis, the intermediate analysis step results were slightly different, but in the end, the analysis ultimately showed the same results. The same five groups were also able to be created.

Summary

This chapter presented the results obtained from the exploratory analysis of the insider threat case data. Visualization tools provide an uncomplicated, time saving, and applicable analysis approach to explore insider threat cases. Several patterns, both significant and insignificant, were found. The following chapter will provide conclusions and recommendations based on the results presented in this chapter.

V. Conclusions and Recommendations

Overview

The purpose of this study was to investigate if visualization tools could be useful in analyzing the insider threat. Fifty-four computer and intellectual property crimes were analyzed using visualization software to determine if new insight could be gleaned on the insider threat problem, specifically new patterns or relationships. Exploratory analysis was used to conduct this research. The study also examined whether visualization tools could be helpful in generating hypotheses for future insider threat research. This chapter presents conclusions, implication for researchers, limitations of the study, and recommendations for future research based on the exploratory analysis of the case data.

Discussion

Fifty-four insider threat cases from the Department of Justice that occurred during the period from 1998 to 2004 were examined via exploratory analysis to answer the following research questions:

1. Using exploratory analysis, how can visualization tools be useful in highlighting patterns or relationships in insider attack case data?
2. Can visualization software assist in generating hypotheses for future insider threat research?

It appears that the insider threat models and frameworks discussed in Chapter two provided numerous constructs in which to analyze during the exploration. Using these models and frameworks, the dataset analysis was able to perceive several findings.

For one, the majority of the insider attack cases, thirty-three of fifty-four cases, were conducted by former or ex employees. It would seem that these crimes may have been either set up prior to the employee's departure or were conducted as an externally initiated attack after their employment with the organization was terminated. Although the visualization tool query did not indicate this (two of fifty-four cases), this may suggest that the motive for these crimes is some type of revenge or retaliation. This may indicate that organizations are not implementing or enforcing the proper security policies and practices following an employee's termination, such as disabling network or remote access accounts, or the organization may need to be more tactful in the laying off or firing of its employees.

Additionally, eleven of the fifty-four cases involved the unauthorized use of a password. Forced password changes for individual and shared accounts and protection of these passwords (not sharing them or writing them down) may reduce some of the insider attacks from occurring. Both of these are easy and inexpensive solutions that may mitigate the insider threat. Mitnick may be correct in his statement that people are the weakest link: "Security is not a technology problem; it's a people and management problem" (Mitnick, 2002:4).

This insider attack dataset also indicated that for movie piracy and especially for software piracy crimes, the criminals tended to work in groups rather than alone. This suggests that when a person is involved in a movie or software piracy crime, law enforcement should focus part of its investigative research on determining if others may be involved in the transgression.

Before the research findings are discussed, it is important to note that exploratory analysis is not a science, but an ambiguous methodology driven by the whims of the analyst. Due to different educations, backgrounds, and experiences, various researchers will ‘see’ different things within a dataset. These differences affect the researcher’s exploratory path. Because there is no defined beginning or end to the analysis process, it may be difficult to determine when the analysis is complete. Also, the fact that “data miners typically have no control over the data gathering process…the data may be ideally suited to the purposes for which it was collected, but not adequate for its data mining uses” (Hand et al., 2001:213) must be remembered.

Research Question #1

Research question one, “Using exploratory analysis, how can visualization tools be useful in highlighting patterns or relationships in insider attack case data?”, was answered during the exploratory analysis. Based on this analysis, both patterns and relationships were discovered. The researcher was able to show the types of employees who committed crimes, how insider crimes were committed, and unique aspects of insider crimes.

The IN-SPIRE analysis tools that are particularly effective in analyzing the insider threat dataset included the grouping tool and its use of colors, the number of viewing options with this color distinction, the time slice tool, and the highlighting function. The grouping tool shows the researcher how many of each type of crime occurred and provides the ability to isolate this data group for further analysis of its own. The ability to view the group colors in the various analysis tools (galaxy visualization, document viewer, outlier panel, and group viewer/evidence panel) provides a cross pollination of these data views that are especially helpful to the analyst in identifying hidden patterns. Also, the group tools' evidence panel provided a means to somewhat verify these findings. The time slicer is an excellent trend detector, identifying when crimes occurred and the types of crime groups that occurred within the time unit specified. This researcher suspects the time slicer would provide even greater insight with larger datasets. Finally, the highlighting function provides the capability to identify analyst-supplied words that occur in the dataset documents by color, removing the requirement to manually read through each document to find the desired word.

Research Question #2

The second research question, "Can visualization software assist in generating hypotheses for future insider threat research?" was also answered during the course of the exploratory analysis. During the analysis, it was found that new knowledge is discovered using visualization tools. In addition, this research supported that visualization tools can assist with hypotheses generation for insider threat research. One hypothesis generated from the visualization tool concerns the analysis for former and ex employees.

H1: Former or ex employees conduct a majority of insider attacks

Since former or ex employees did not attack the organization when s(he) was an employee of the organization when the crime may have been easier to conduct, it suggests that the employee may have believed s(he) was ill-treated before or during his or her employment termination. Therefore, the crimes by former or ex employees may have been motivated by this ill will as a form of revenge or retaliation against the organization for this perceived wrong. A second hypothesis that would logically extend from this finding is:

H2: Revenge or retaliation is the motive for former or ex employee insider attacks.

Supplementary Findings

In addition to the two research questions, several supplementary finding were discovered. The literature supports that visualization tools are an efficient method for researchers to analyze large, unstructured datasets with minimum effort. However, due to the volumes of information in today's environment, the majority of analysts have limited time to format their datasets for processing. As such, IN-SPIRE^{TM} accepts unstructured data in a variety of contexts. To keep from having terms cluster on common words (such as the, a, and and), IN-SPIRE^{TM} uses a default stopword list to avoid this problem. A stopword is a non-information bearing word identified so the software will not cluster documents by this word. However, when webpages are used as the dataset source, the visualizations are clustered by some of their html tags. Although IN-SPIRE^{TM} has a default web stopword list, it is not comprehensive. The analyst is forced to recalculate many visualizations in order to remove dozens of terms just to reveal relevant cluster titles on which they can then focus their efforts on. Until an inclusive default list is developed, the time spent manipulating the stopword list indicates only a limited potential for unstructured web data.

Also, because insider threat researchers tend to agree that insider attacks are more costly than outsider attacks (Shaw et al., 1998; E-Crime, 2004; Schultz, 2002; Yager, 2003; Gordon et al., 2004; D'Arcy and Hovav 2004), it would be interesting to determine an average dollar damage caused by the insider attacks within this dataset. However, due to the impossibility for each of the insider attacks to have the same dollar amount of damage caused by the attack, the only way to determine a sum or average of the insider threat damages is to query the entire set for a dollar sign. Then each of the cases must be examined for this highlighted ‘$’, and then the dollar amounts summed and averaged. This approach is not only burdensome, but time consuming as well.

Finally, since IN-SPIRE’s^{TM} clustering algorithm works on word frequency and does not understand the different variations of the same word, a word has to be examined by query or removed as an outlier term several ways, such as copy, copied, and copies. Other visualization software is able to distinguish this similarity and remove all of the terms when only one of them is specified.

Implications for Researchers

Results from this study contributed to the existing body of knowledge on insider threat research as well as introduced new insider threat hypotheses and data for further exploration. The primary weapon against crime will not be bullets; it will be information (Mena, 2003). Ultimately, visualization tools may provide researchers in all disciplines with large amounts of unstructured data an additional tool to use in their analysis.

By using a visualization tool, the study also highlighted the successful combination of human and IS/IT capabilities. “The most important aspect of information systems development is to adjust the IS to meet human characteristics and behavior. This means humanization of IS” (Koskinen et al., 2005:1). Data mining and visualization software developers should maximize this liaison in future software release.

Limitations

This research focuses on identifying characteristics, patterns, or relationships of an inside attacker that may not be visible upon first glance. Yet as with all research, limitations exist.

The methodology of exploratory analysis is a limitation in its own right. Exploratory analysis is not a science; it is an ambiguous methodology. The results found during the exploratory analysis are not statistically validated results. Thus, care must be taken when interpreting the results. Also, because exploratory analysis, for the most part, is 'secondary analysis', the dataset may not be suitable for the data mining purpose.

The dataset had several limitations of its own. Although a minimum of fifteen documents are needed for the IN-SPIRE^{TM} visualization software, larger datasets may reveal more significant information. The researcher believed that the small sample size of fifty-four cases did not fully examine the capabilities of the visualization tool. The word weightings may have been skewed during the analysis showing the wrong 'picture.' In addition, because this research is limited to cases that have already occurred and so few organizations report insider crimes to law enforcement (Gordon et al., 2004), the results may not be a representative sample of all insider attackers. The cases provided by the Department of Justice were a convenience sample, as is typical of data mining datasets. However, these cases were identified as not being an "exhaustive" list of computer intrusion and intellectual property cases, but only a sample. Also, since financial organizations are legally forced to report all crimes, other sectors are probably underrepresented. In addition, the cases were written by many different individuals. Although some standardization existed in what was included in the case release, certain data was not included in each and every case. For instance, when the dataset was queried for motive or revenge/retaliation, only three cases had this data reported in the release. Overall, the word choice or content (or lack thereof) of the cases may have affected the findings of the data analysis. As such, treat the findings as suggestive, but not conclusive.

Finally, the assumptions and bias on the part of the researcher during the exploratory analysis was a limitation. The researcher tried to eliminate some of the assumptions by reviewing the previous insider threat research to provide some focus to the analysis. To minimize the biases, a pilot study was conducted on an unrelated topic to learn the visualization tool. However, since exploratory analysis is dependent on the active collaboration of the software and the user, not all of the bias can (or should) be removed.

Future Research

There are several opportunities for research in this area. In identifying the insider attack cases from the outsider attacks to establish the insider threat dataset, the researcher manually examined the 198 cases. In retrospect, it would have been interesting to see if the visualization tool could have correctly categorized the dataset into these two groups. Also, another researcher could replicate this research effort using the same visualization tool and dataset to see if they discover the same (or different) patterns or relationships. In addition, different visualization software could be used with the same dataset to see if the findings are similar to this research effort, maybe one with summarization ability. Also, a different dataset with detailed information regarding the perpetrators (such as medical, employment history, criminal, and educational records) could be analyzed within IN-

SPIRE^{TM} to get a more comprehensive view of the criminal as well as the crime they have committed. Additionally, in searching for a dataset to use in this study, the researcher could not find a location that specifically collects and reports details of insider threat crimes. A centralized location, possibly in the form of a database, should capture the details of the insider threat crimes that have been committed to date for future researchers to analyze. Furthermore, additional unstructured data tools using a variety of algorithms should be developed with more robust features and tools to assist organizations with the mountains of data they are buried beneath. Also, researchers should determine a method to test the results of this analysis in the real world. The hypotheses generated from this exploratory analysis, H1: Former or ex employees conduct a majority of insider attacks and H2: Revenge or retaliation is the motive for former or ex employee insider attacks, should be tested and validated.

Summary

Results of this study suggest that visualization tools may be useful for the analysis of unstructured data such as the data found in the insider attack cases. The visualization tool provided an effective categorization of the insider threat dataset once the data was converted to a .txt dataset. Both interesting and mundane information was culled from the insider attack dataset. It appears that visualization tools can be used to generate possible hypotheses for future insider threat research. Further research in unstructured data is needed to determine the most effective algorithm and visualization display. Be that as it may, IN-SPIRE^{TM} did find several fruitful areas of insider threat research to explore.

Bibliography

Anderson, Robert H. *Research and Developmental Initiatives Focused on Preventing, Detecting, and Responding to Insider Misuse of Critical Defense Information Systems*. CF-151-OSD. Santa Monica CA: RAND Corporation, August 1999.

Anderson, Robert H., Richard Brackney, and Thomas Bozek. *Advanced Network Defense Research.* CF-159-NSA. Santa Monica CA: RAND Corporation, August 2000.

Advanced Research and Development Activity (ARDA). Advanced IC Information Assurance. Retrieved on 21 April 2004 from http://www.ic-arda.org/Advanced_IC/index.html.

Barnett, Thomas. *The Pentagon's New Map*. New York: G.P. Putnam's Sons, 2004.

Bateman, Chris, Dawn Cappelli, Casey Dunlevy, Andrew Moore, Dave Mundie, Stephanie Rogers, Tim Shimeall, Jose Sarriegui, Jeff Stanton, and Jose Gonzalez. "System Dynamics Modeling for Information Security: A Group Modeling Workshop." Hosted at CERT/CC, Sofware Engineering Institute, Carnegie Mellon University, Pittsburgh. 16-20 February 2004.

Brackney, Richard and Robert Anderson. *Understanding the Insider Threat*, CF-176-ARDA. Santa Monica CA: RAND Corporation, March 2004.

Bransten, Lisa. "Technology (A Special Report) – Power Tools – Looking for Patterns: Data Mining enable companies to better manage the reams of statistics they collect; The goal: spot the unexpected," *Wall Street Journal* (Eastern Edition), 21 June 1999.

Carney, Ralph. *Trends in Background Issues of Applicants for Access to Classified Information*, Monterey CA: Defense Personnel Security Research Center, 30 June 1999.

Caruso, Valerie L. *Outsourcing Information Technology and the Insider Threat*. Air Force Institute of Technology (AU): Wright-Patterson AFB OH, March 2003. (ADA415113).

Cerrito, Patricia. "Inside Text Mining," *Health Management Technology*, 25: 28 (March 2004).

Charney, Scott. *Cybercrime*. PricewaterhouseCoopers. Retrieved on 8 September 2004 from http://www.pwc.com/extweb/newcolth.nsf/docid/1A47C7356C3D57AC 85256AD1005834D1.

Chen, Zhengxin. *Data Mining and Uncertain Reasoning: An Integrated Approach*. New York: John Wiley & Sons, 2001.

Cios, Krzysztof, Witold Pedrycz, and Roman Swiniarski. *Data Mining: Methods for Knowledge Discovery*. Boston: Kluwer Academic Publishers, 1998.

Clark, Ken. "From Data to Decisions," *Chain Store Age*, 78: 62 (November 2002).

Computer Emergency Response Team/Coordination Center (CERT/CC). "Preliminary System Dynamics Maps of the Insider & Outsider Cyber-threat Problems." *Proceedings of the 22nd International Conference of the System Dynamics Society*, July 2004.

-----Statistics, Retrieved from 20 April 2004 from http://www.cert.org/stats/cert_stats.html.

Computer Science and Telecommunications Board (CSTB). *Cyber-Security and the Insider Threat to Classified Information*. National Research Council. 2 November 2000. Retrieved on 30 September 2004 from http://www7.nationalacademies.org/cstb/wp_insiderthreat.pdf.

Corbitt, Terry. "Business Intelligence and Data Mining," *Management Services*, 47: 18 (November 2003).

Crume, Jeff. *Inside Internet Security: What Hackers Don't Want You to Know*. New York: Addison-Wesley, 2000.

D'Amico, Esther. "Sorting out the Facts," *Chemical Week*, 164: 22 (16 October 2002).

D'Arcy, John and Anat Hovav. "The Role of Individual Characteristics on the Effectiveness of IS Security Countermeasures." *Proceedings of the Tenth Americas Conference on Information Systems*. 1-8. New York, August 2004.

Denning, Dorothy E. *Information Warfare and Security*. New York: Addison-Wesley Publishing Company, Inc, 1999.

Department of Defense (DoD). *Information Assurance (IA) Implementation*. DoD Instruction 8500.2. Washington DC: GPO, 6 February 2003.

-----Insider Threat Integrated Process Team (IPT). *DoD Insider Threat Mitigation: Final Report of the Insider Threat Integrated Process Team*, 24 Apr 2000. Retrieved on 15 Apr 2004 from https://acc.dau.mil/simplify/ev.php?ID=64069_201&ID2=DO_TOPIC.

-----Office of the Inspector General (IG). *DoD Management of Information Assurance Efforts to Protect Automated Information Systems*, Report No. PO 97-049. Washington DC: GPO. 25 September 1997.

Department of the Air Force (DAF). *Network and Computer Security*. AFI 33-202. Washington DC: HQ USAF, 17 June 2004.

Dhillon, Gurpreet and Steve Moores. "Computer Crimes: Theorizing about the Enemy Within," *Computers & Security*, 20: 715-723 (December 2001).

E-Crime Watch Survey: Summary of Findings, 2004. CSO Magazine, 1-39, 13 September 2004. Retrieved on 30 September 2004 from http://www.csoonline.com/read/090104/2004ecrimewatchsummary.pdf.

Erramouspe, Jeff. "Unstructured Data," *Computer Technology Review*, 24: 17 (August 2004).

Fayyad, Usama, Georges Grinstein, and Andreas Wierse. *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco: Morgan Kauffman Publishers, 2002.

Fielden, Tim. "Text mining promises to cull answers from random text – understanding the text in emails, marketing materials, and documents could prove invaluable," *InfoWorld*. 22: 88 (16 October 2000).

Feldman, Ronen. From *Handbook of Data Mining and Knowledge Discovery*. Eds. W. Klosgen, and J. Zytkow. Oxford NY: Oxford University Press, 2002.

Friedman, Thomas. *The Lexus and the Olive Tree* (Updated Version). New York: Anchor Books, 2001

Gardiner, Brian. "E-Business Security in RAG order," Dublin Institute of Technology, Dublin, Ireland, 2003. http://www.comp.dit.ie/rfitzpatrick/MSc_Publications/2003_Bryan_Gardiner.pdf.

Gaudin, Sharon. "Case Study of Insider Sabotage: The Tim Lloyd/Omega Case," *Computer Security Journal*, 16: 1-8 (2000).

Gill, Lisa. "IT Nightmare: The Enemy Within," *NewsFactor Network* (29 July 2002). Retrieved on 20 September 2004 from http://www.newsfactor.com/perl/story/18778.html.

Ginn, Patrick. *Correlation Analysis of Fleet Information Warfare Center Network Incidents*. Naval Postgraduate School, Monterey CA, June 2001. (ADA396275).

Gordon, Lewis, Martin Loeb, William Lucyshyn, and Robert Richardson. *2004 CSI/FBI Computer Crime and Security Survey*. Computer Security Institute. Retrieved 4 June 2004 from http://www.gocsi.com/pdfs.fbi/FBI2004.pdf.

Grinstein, Georges and Matthew Ward. "Introduction to Data Visualization," in Chapter 1 of *Information Visualization in Data Mining and Knowledge Discovery*. Eds. U. Fayyad, G. Grinstein, and A. Wierse. San Francisco: Morgan Kauffman Publishers, 2002.

Hand, D., H. Mannila, and P. Smyth. *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.

Herbig, Katherine and Martin Wiskoff. *Espionage Against the United States by American Citizens 1947 – 2001*, Defense Personal Security Research Center, Technical Report 02-05. Monterey CA: PERSEC, July 2002.

Heuer, Richard. *The Insider Espionage Threat*. Defense Personnel Security Research Center. Monterey CA: PERSEC, 2001. Retrieved on 15 Apr 2004 from http://www.dss.mil/search-dir/training/csg.security/treason/insider.htm.

Icove, David, Seger, Karl, and VonStorch, William. *Fighting Computer Crime*. Computer Crime Research Center (CCRC). Retrieved on 13 December 2004 from http://www.crime-research.org/library/crime1.htm.

ISC Internet Domain Survey. Retrieved on 8 April 2004 from http://www.isc.org/ops/ds/.

Jackson, William. "Intelligence community seeks protection from inside threats," *Government Computer News*. January 12, 2004. Retrieved on 20 April 2004 from http://www.gcn.com/vol1_no1/daily-updates/24622-1.html.

Joint Staff, The. *Information Assurance (IA): Legal, Regulatory, Policy and Organizational Considerations (3rd Edition)*. Washington DC: GPO. 17 September 1997.

Jonas, Jeff, Paul Byron Pattak, and James P. Litchko. "Using Advanced Information Technology to Combat Insider Threats," *Journal of Organization Excellence*, (Autumn 2001).

Kabay, M. "Insider Attacks are a Thorny Problem," *Network World Security Newsletter*, 12 August 2003. Retrieved on 30 September 2004 from http://www.nwfusion.com/newsletters/sec/2003/0811sec1.html.

Kilcrece, Georgia, Klaus-Peter Kossakowski, Robin Ruefle, and Mark Zajicek. *State of the Practice of Computer Security Incident Response Teams (CSIRTs)*, Technical Report CMU/SEI-2003-TR-001. Pittsburgh PA: Software Engineering Institute, October 2003.

Kipp, Steven P. *Espionage and the Insider*. Information Security Reading Room, SANS Institute, 2001. Retrieved on 20 April 2004 from http://www.sans.org/rr/papers/48/426.pdf.

Klosgen, Willi and Jan Zytkow. *Handbook of Data Mining and Knowledge Discovery*. Oxford NY: Oxford University Press, 2002.

Koskinen, Minna, Katja Liimatainen, Eleni Berki, and Mikko Jakala. "The Human Context of Information Systems," *Proceedings of the 38th Hawaii International Conference on System Sciences*. 1-10. Hawaii, 2005.

Krause, M.S. *Contemporary White Collar Crime Research: A Survey of Findings Relevant to Personnel Security Research and Practice*. The Personnel Security Managers' Research Program. August 2002. Retrieved on 16 May 2004 from http://www.navysecurity.navy.mil/White%20Collar%20Crime.pdf.

Kumagai, Jean. "Mission Impossible?," *IEEE Spectrum*, 40: 26 – 31 (April 2003).

Lanford, Jeff. *Implementing Least Privilege at Your Enterprise*. Information Security Reading Room, SANS Institute, July 2003. Retrieved on 20 April 2004 from http://www.sans.org/rr/whitpapers/bestprac/1188.php.

Lee, Wenke Salvatore Stolfo, and Kui Mok. "A Data Mining Framework for Building Intrusion Detection Models," *Proceedings of the 1999 IEEE Symposium on Security and Privacy*. 1-12. Oakland CA, 1999.

Levin, John, Richard Labella, Henry Owen, Didier Contis, and Brin Culver, "The Use of Honeynets to Detect Exploited Systems Across Large Enterprise Networks", IEEE Proceedings, http://www.tracking-hackers.com/papers/gatech-honeynet.pdf. June 2003.

Lok, Corie. "Fighting Infections with Data," Technology Review, 107: 24. October 2004.

Loscocco, Peter, Stephen Smalley, Patrick Muckelbauer, Ruth Taylor, S. Jeff Turner, and John Farrell. "The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments," *Proceedings of the 21st National Information Systems Security Conference*. 303-314. National Security Agency, 1998.

Magklaras, G.B. and S.M. Furnell. "Insider Threat Prediction Tool: Evaluating the Probability of IT Misuse," *Computers & Security*, 21: 62-73 (January 2001).

Melara, Carlos, Jose Sarriegui, Jose Gonzalez, Agata Sawicka, and David Cooke. "A System Dynamics Model of an Insider Attack on an Information System." *Proceedings of the 21st International Conference of the System Dynamics Society*, New York, 20-24 July 2003.

Mena, Jesus. *Investigative Data Mining for Security and Criminal Detection*. Boston: Butterworth Heinemann, 2003.

-----*Homeland Security: Techniques and Technologies*. Hingham MA: Charles RiverMedia, Inc, 2004.

Messmer, Ellen. "IT Execs Share Security Concerns," *Network World*, 20: 12 (9 June 2003).

----- "Security Experts: Insider Threat Looms Largest," *Network World*, 20: 12-13 (8 Dec 2003).

Meyers, Jason. "Automatic Categorization,Taxonomies, and the World of Information: Can't Live With Them, Can't Live Without Them," *E-Doc*, 16: 20 (November/December 2002).

Mitnick, Kevin and William Simon. *The Art of Deception: Controlling the Human Element of Security*. Indianapolis: Wiley Publishing, Inc, 2002.

National Security Telecommunications and Information Systems Security Committee (NSTISSC). *The Insider Threat to U.S. Government Information Systems* (Report No. NTISSAM INFOSEC/1-99). Fort Meade MD: NTISSC Secretariat, July 1999.

Neumann, Peter G. "The Challenges of Insider Misuse," SRI Computer Science Lab. *Testimony at House Science Committee on technology,* 15 August 1999.

National Academy of Sciences (NAS). *Reshaping the Graduate Education of Scientists and Engineers*. Committee on Science, Engineering, and Public Policy (COSEPUP). Washington DC: National Academy Press, 1995.

National Infrastructure Protection Center (NIPC). *Special Technologies and Applications Unit (STAU): Insiders and Information Technology*. Retrieved on 20 April 2004 from http://www.hpcc-usa.org/pics/02-pres/wright.ppt.

National Institute of Standards and Technology (NIST). *Threats to Computer Systems. Computer Systems Laboratory Bulletin*. March 1994. Retrieved on 15 April 2004 from http:/csrc.nist.gov/publications/nistbul/csl94-03.txt.

National Threat Assessment Center (NTAC). *NTAC webpage*, United States Secret Service. Retrieved on 4 June 2004 from http://www.secretservice.gov/ntac.html.

Office of the Assistant Secretary of Defense (OASD) Command, Control, Communications, and Intelligence (C3I), "Insider Threat Integrated Process Team (IPT) Memo," December 22, 1998.

Pacific Northwest National Laboratory (PNNL). "IN-SPIRE Frequently Asked Questions." Retrieved on 12 November 2004 from http://in-spire.pnl.gov/faq.html.

Porter, David. "Insider Fraud: Spotting the Wolf in Sheep's Clothing," *Computer Fraud & Security*, 2003: 12-15 (April 2003).

Rajagopalan, Balaji and Ravi Krovi. "Benchmarking Data Mining Algorithms," *Journal of Database Management*, 13: 25-35 (January – March 2002).

Randazzo, Marisa, Michelle Keeney, Eileen Kowalski, Dawn Capelli, and Andrew Moore. *Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector*, United States Secret Service and CERT Coordination Center/SEI, August 2004. Retrieved on 30 Sep from http://www.secretservice.gov/ntac_its.shtml.

Reamy, Thomas. "Cyborg Categorization: the Salvation of Search?," *Intranet Professional*, 6: 20-21 (January/February 2002).

Rhodes, Philip. "Discovering New Relationships: A Brief Overview of Data Mining and Knowledge Discovery," in Chapter 21 of *Information Visualization in Data Mining and Knowledge Discovery*. Eds. U. Fayyad, G. Grinstein, and A. Wierse. San Francisco: Morgan Kauffman Publishers, 2002.

Robb, Drew. "Taming Text," *Computerworld*. 38: 40-41 (21 June 2004).

Robinson, Jarvis. *Internal Threat – Risk and Countermeasures*. Information Security Reading Room, SANS Institute, 15 November 2001. Retrieved on 20 April 2004 from http://www.sans.org/rr/papers/60/475.pdf.

Saltzer, J.H. and M.D. Schroeder. The Protection of Information in Computer Systems, Proceeding of IEEE, 63: 1278-1308 (April 1975).

Scalet, Sarah. Dr. Crime's Terminal of Doom and Other Tales of Betrayal, Sabotage, & Skullduggery. CIO Magazine. 15: 1-7 (1 June 2002). Retrieved on 20 September 2004 from http://www.cio.com/archive/060102/doom.html.

Schneier, Bruce. *Secrets and Lies: Digital Security in a Networked World*. New York: John Wiley & Sons, Inc, 2000.

Schwartau, Winn. *Information Warfare: Chaos on the Electronic Superhighway*. New York: Thunder's Mouth Press, 1994

Shaw, Eric D, Jerrold M. Post, and Keven G. Ruby. "Inside the Mind of the Insider," *Security Management*, 43: 34-41 (December 1999).

Shaw, Eric D, Keven G. Ruby, and Jerrold M. Post. "The Insider Threat to Information Systems," *Security Awareness Bulletin*, No 2-98, Department of Defense Security Institute (September 1998).

Shultz, E. E. "A Framework for Understanding and Predicting Insider Attacks," *Computers and Security*, 21: 526-531 (October 2002).

Shultz, E. E. and Russell Shumway. *Incident Response: A Strategic Guide to Handling System and Network Security Breeches*. Indianapolis: NewRiders, 2002.

Sniffen, Michael. "Controversial Data Mining Project Lives On…," *Information Week*, 23 February 2004.

Spiegler, Israel. "Technology and Knowledge: Bridging a "Generating" Gap," *Information and Management Journal*, 40: 533-539 (July 2003).

Spitzner, Lance. *Honeypots: Catching the Insider Threat.* 2003. Retrieved on 20 April 2004 from http://www.acsac.org/2003/papers/spitzner.pdf.

Swartz, Nikki. "Data-Mining Initiatives," *Information Management Journal*, 37: 17 (March/April 2003).

Trends in Proprietary Information Loss: Survey Report, September 2002. Sponsored by PricewaterhouseCoopers, U.S. Chamber of Commerce, and American Society of Industrial Security (ASIS) Foundation. Retrieved on 30 September 2004 from http://www.asisonline.org/newsroom/surveys/spi2.pdf.

United States Department of Justice, Computer Crime and Intellectual Property Section, Computer Intrusion Cases. Retrieved on 5 January 2005 from http://www.usdoj.gov/criminal/cybercrime/cccases.html.

-----, Computer Crime and Intellectual Property Section, Intellectual Property Cases. Retrieved on 5 January 2005 from http://www.cybercrime.gov/ipcases.htm.

Uramotoa, N, H. Matsuzawa, T. Nagano, A. Murakami, H. Takeuchi, and K. Takeda. "A Text-Mining System for Knowledge Discovery from Biomedical Documents," *IBM Systems Journal*, 43: 516-533 (September 2004).

"U.S. Government Still Mining Data," *The Information Management Journal*, 38: 7 (July/August 2004).

Walter, Jim. "The Ins and Outs of Data Mining: Data Mining Software Allows Researchers to Access, Analyze, Model, and Deploy Results Quickly," *R & D*, 45: 33 (April 2003).

Wills, Graham. "Visualization," in Chapter 33 of *Handbook of Data Mining and Knowledge Discovery*. Eds. W. Klosgen, and J. Zytkow. Oxford NY: Oxford University Press, 2002.

Wood, Bradley. An Insider Threat Model for Adversary Simulation. Cyber Defense Research Center. SRI International, 2000. Retrieved on 30 September 2004 from http://www.rand.org/publications/CF/CF163/CF163.appb.pdf.

Yager, Tom. "Security Lockdown," *InfoWorld* 25: 42-51 (14 July 2003).

Vita

Captain Amy M. Rammel was born in Dayton, Ohio. She graduated from Milton-Union High School, West Milton, Ohio in 1987. She enlisted in the Air Force in 1988 and served nearly ten years as both a Computer Operator and a Computer Programmer. While on active duty, she completed two Associate of Applied Science degrees from the Community College of the Air Force--Information Systems Technology in 1995 and Computer Science Technology in 1998. She completed her Bachelor of Science degree from Park College in Management Information Systems in 1998. Upon graduation, she was accepted into Officer Training School at Maxwell Air Force Base, Alabama.

After her commissioning in November 1998, Captain Rammel was assigned to the 325 Communications Squadron, Tyndall Air Force Base, Florida, where she served as the base's Year 2000 Action Officer and as Deputy Flight Commander. In 2000 she was reassigned to the Electronic Systems Center, Detachment 5, Peterson Air Force Base, Colorado, as the communications expert on a multi-million dollar source selection team. Upon contract award, she was reassigned as Support Flight Commander.

In August 2003, Captain Rammel entered the Graduate School of Engineering and Management, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, to pursue a Masters of Science in Information Resource Management. Upon graduation, she will be assigned to the National Air and Space Intelligence Center, Wright-Patterson Air Force Base, Ohio.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 21-03-2004	**2. REPORT TYPE** **Master's Thesis**	**3. DATES COVERED** *(From – To)* Aug 2003 – Mar 2005

4. TITLE AND SUBTITLE Assessing the Usefulness of Visualization Tools to Investigate Hidden Patterns within Insider Attack Cases	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Rammel, Amy M., Captain, USAF	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 641 WPAFB OH 45433-7765	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/GIR/ENV/05M-14
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The insider threat is a major concern for organizations. Open markets, technological advances, and the evolving definition of employee have exacerbated the insider threat. Insider threat research efforts are focusing on both prevention and detection techniques. However, recent security violation trends highlight the damage insider attacks cause organizations and illuminate why organizations and researchers must develop new approaches to this challenge. Although fruitful research is being conducted and new technologies are being applied to the insider threat problem, companies remain susceptible to the costly damage generated by insider threat actions.

This research explored how visualization tools may be useful in highlighting patterns or relationships in insider attack case data and sought to determine if visualization software can assist in generating hypotheses for future insider threat research. The research analyzes cases of insider attack crimes committed during the period of 1998 to 2004 with an information visualization tool, IN-SPIRE. The results provide some evidence that visualization tools are useful in both finding patterns and generating hypotheses. By identifying new knowledge from insider threat cases, current insider threat models may be refined and other potential solutions may be discovered.

15. SUBJECT TERMS

insider, insider threat, data mining, unstructured data, exploratory analysis, Exploratory Data Analysis (EDA)

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Alan R. Heminger, PhD, AFIT/ENV
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4797; e-mail: alan.heminger@afit.edu
U	U	U	UU	109	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18