Analysis and Detection of Malicious Insiders

Mark Maybury, Penny Chase, Brant Cheikes
Information Technology Division
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
{maybury, pc, bcheikes}@mitre.org

Dick Brackney
Advanced Research and Development Activity
in Information Technology
9800 Savage Road
Fort George G. Meade, MD
rcbrack@nsa.gov

**Sara Matzner
and Tom Hetherington**
Applied Research Laboratories
University of Texas
Austin, TX 78713
{matzner, tomh}@arlut.utexas.edu

**Brad Wood
Conner Sibley
and Jack Marin**
BBN Technologies
9861 Broken Land Parkway, Suite 400
Columbia MD 21046
{bwood,csibley, jamarin}@bbn.com

Tom Longstaff
CERT Research and Analysis Centers
Software Engineering Institute
Carnegie Mellon University
4500 Fifth Avenue
Pittsburgh, PA 15213-3890
tal@cert.org

**Lance Spitzner
and Jed Haile**
Honey Net Consortium
lance@honeynet.org
jed.haile@theloganggroup.biz

John Copeland
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0490
copeland@ece.gatech.edu

Scott Lewandowski
MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108
scl@ll.mit.edu

Keywords: insider threat, malicious insider, information assurance, cyber indications and warning, observables taxonomy, assets, data fusion, attack graphs, honeypots, StealthWatch, Robert Hanssen

Abstract

This paper summarizes a collaborative, six month ARDA NRRC[1] challenge workshop to characterize and create analysis methods to counter sophisticated malicious insiders in the United States Intelligence Community. Based upon a careful study of past and projected cases, we report a generic model of malicious insider behaviors, distinguishing motives, (cyber and physical) actions, and associated observables. The paper outlines several prototype techniques developed to provide early warning of insider activity, including novel algorithms for structured analysis and data fusion. We report the assessment of their performance in an operational network against three distinct classes of human insiders (an analyst, application administrator, and system administrator), measuring timeliness and accuracy of detection.

1. The Threat: Malicious Insiders

An *insider* as anyone in an organization with approved access, privilege, or knowledge of information systems, information services, and missions. A *malicious insider (MI)* is one motivated to adversely impact an organization's mission through a range of actions that compromise information confidentiality, integrity, and/or availability. This research explores three fundamental hypotheses motivated by our study of MIs.

1. While some MIs can be detected using a single cyber observable, other MIs could be detected only by using multiple and heterogeneous observables.

2. Fusing information from heterogeneous information sources (e.g., logs from printers, authentication, card readers, telephone calls) and various levels of the IP stack (e.g., application vs. network traffic) allows more accurate and timely indications and warning of malicious insiders.

[1] This effort was performed at The MITRE Corporation at the Northeast Regional Research Center (NRRC) which is sponsored by the Advanced Research and Development Activity in Information Technology (ARDA), a U.S. Government entity which sponsors and promotes research of import to the Intelligence Community which includes but is not limited to the CIA, DIA, NSA, NGA, and NRO.

3. Observables together with domain knowledge (e.g., user role, asset value to mission) can help detect inappropriate behavior (e.g., need to know violations).

To maximize progress in this challenge workshop, we created multiple working groups: one responsible for our experimentation data and network, one using StealthWatch sensors (which perform traffic and host profiling), another using honeynets, another using structured analysis models, and another using bottom up fusion across multiple sensors to detect insiders.

2. Historical MI Case Analysis

The first step in our approach was an analysis of prior malicious insiders. While we investigated information on dozens of insider cases (DSS 1999, Herbi and Wiskoff 2002), we performed detailed analysis on six cases. Maybury et al. (2004) summarizes some key features of three representative cases such as CIA's Aldrich "Rick" Ames, FBI's Robert Philip Hanssen (2003), and DIA's Ana Belen Montes (2001). In each of these cases we summarize their position, motive, foreign handlers, impact, sentence, computer skill, polygraph experience, cyber security violations, counter intelligence activities, physical and cyber access, cyber extraction and exfiltration, cyber communication, and the transfer of materials to foreign handlers.

The devastating impact of these three individuals included the violation of confidentiality, undermining of intelligence integrity, adverse influence of US policy, the revelation of sources and methods, and the death and compromise of field agents. Motives were diverse, ranging from financial to thrill to ideological. In each of these cases, handlers were professional foreign service agents. Two of the three passed polygraphs. While the computer skills of each of these insiders ranged significantly, all left trails of suspicious cyberactivity while performing cyber access, exfiltration, and/or communication. All engaged in counter intelligence to evade detection and/or destroy incriminating evidence. In each case we found opportunities to observe individual incidents and/or to detect anomalous behavior from correlated observables.

In addition to these historic cases, we also projected a future insider in the role of a systems or network administrator who would have significantly deeper computing skill and infrastructure access. This would enable, for example, more stealthy attacks (e.g., the MI might not have to perform network reconnaissance or could create private communication channels or open up backdoors) as well as new kinds of attacks such as on availability wherein the objective of the insider was to degrade, damage or destroy the network.

3 Simulated MIs: Pal, Jill, and Jack

Grounding our efforts in realistic insider behavior, we explored detecting three types of insiders in detail in this activity. The first was a historical insider modeled as a prototype of past need-to-know violators. We call this insider Pal. A second insider, named Jack, was a projected insider who would aim to disrupt, damage, or destroy the network or elements thereof. In the course of defining and simulating these insiders, the scenario team implemented a third category of insider, an application administrator, called News Admin or Jill. Only Pal's behavior model was disclosed to sensor builders prior to the experiment. For detail about these insiders including a log of specific actions taken by the insiders see Maybury et al. (2004). The three malicious insider cases were simulated on MITRE's Demilitarized Zone (DMZ) network. The DMZ consists of over 300 hosts with a range of missions utilizing services such as web (HTTP), news (NNTP), file transfer (FTP), messaging (SMTP), mail (POP, IMAP), database (SQL), and question answering. We instrumented 18 of 31 nodes on the NRRC (Northeast Regional Research Center) subnetwork which had 75 on-line, active users during the evaluation.

A semi-automated process captured, filtered, and anonymized the malicious insider collection to address security and privacy concerns. Figure 1 illustrates the heterogeneous nature of the collection consisting of over 11 million records which spans physical sensors (e.g., employee badge readers), network level sensors (e.g., Snort rules modified to detect inappropriate connections or behavior), host sensors (to detect user access and command sequences), and applications (e.g., mail server logs, web server logs, network news logs). A Common Data Repository (CDR) was established as a central database storing the over 11 million anonymized, time stamped audit-log records collected over three months.

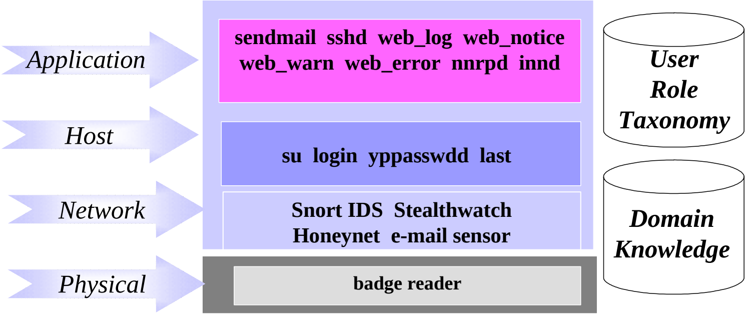

Application
Host
Network
Physical
sendmail sshd web_log web_notice web_warn web_error nnrpd innd
su login yppasswdd last
Snort IDS Stealthwatch Honeynet e-mail sensor
badge reader
User Role Taxonomy
Domain Knowledge

Figure 1. Heterogeneous and Multilevel Data Sources

4. Event and Observable Taxonomy

In order to access, exploit, or damage assets, a MI will necessarily need to perform (or have another person or process perform) a series of actions to gain privileges, access or manipulate assets. Derived from our analysis of MI cases, Figure 2 shows a taxonomy of cyber events which have associated observables that hold promise for the foundation of a detection system. The taxonomy distinguishes observables in the cyber domain from those in the physical domain. The taxonomy includes observables such as results of the polygraph, records of security violations, missing or misleading reports on finances, foreign travel or foreign contacts, physical facility access, personal finances, materials transfer, counter intelligence, social behavior, and communications. In this research we focused exclusively on cyber observables, including other observables that could be readily converted to a cyber signal (e.g., digitized facility access logs).

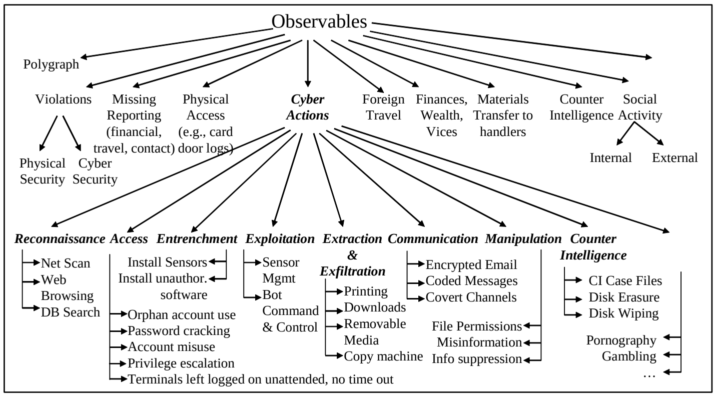


Figure 2. Cyber Event/Observable Taxonomy

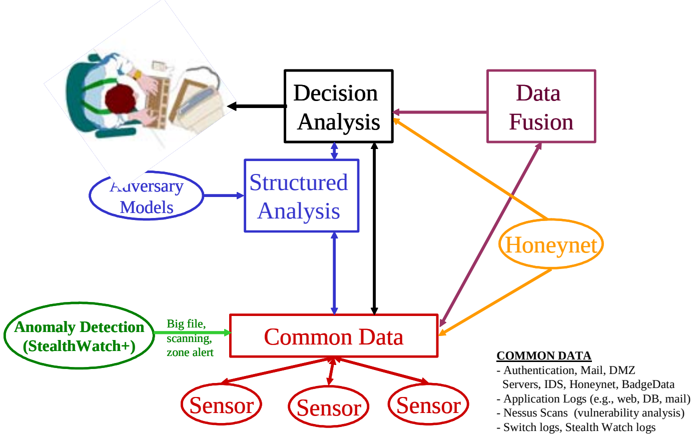
The core of the taxonomy incorporates a range of cyber observables encompassing a range of classes of cyber actions indicated in bold italics in Figure 2. These include activities of network, system, and information reconnaissance, access to assets (e.g., media, hosts, accounts), entrenchment (e.g., installing sensors or unauthorized software), exploitation (e.g., commanding and controlling entrenched assets such as software bots or zombie machines), extraction and exfiltration (e.g., of hardcopy, media, information), communication (e.g., encrypted messaging, encoded messages, covert channels), manipulation of cyber assets (e.g., changing file permissions, suppressing or altering information content), counter intelligence (e.g., wiping disks), and other cyber activities associated with unethical or addictive behavior (e.g., on line gambling). Some observables have been used in some historical cases as a tip-off of malicious activity; others serve as direct indicators of inappropriate behavior.

5. Insider Detection

While the live network instrumentation describe in Section 3 provided an unprecedented and essential set of MI experimental data, the thrust of our activity was developing novel algorithms to detect MIs. Figure 3 illustrates the high level architecture of a proof of concept system that was designed, implemented, and tested to detect MIs. Distributed, heterogeneous sensors provide input to a Common Data Repository (CDR) from which a range of analyses are performed including data fusion and structural analysis to identify potential suspects on a watch list or issue an alert of an insider threat. As illustrated in Figure 3, our technical approach is novel in the following respects:

- A Common Data Repository (CDR) captures and anonymizes heterogeneous sensor input.
- Multilevel monitoring occurs at the packet level, system level, and application level.
- StealthWatch sensors detect abnormal insider behavior on the network such as scanning, file transfer, or internal network connections.
- Distributed honeynets acquire attacker properties, pre-attack intensions, and potential attack strategies.
- A real-time, top-down structural analysis drawing upon functional models of MIs maps pre-attack indicators to models of potential MIs.
- Traditional and non-traditional indicators (e.g., logs of network activity, physical access, PBX, help desks), including non-digital sources, are fused bottom-up.

Sensor inputs are then exploited by a decision analysis component to determine watch list membership and insider detection. We next consider each of the primary detection strategies.



COMMON DATA
- Authentication, Mail, DMZ Servers, IDS, Honeynet, BadgeData
- Application Logs (e.g., web, DB, mail)
- Nessus Scans (vulnerability analysis)
- Switch logs, Stealth Watch logs

Figure 3. Integrated Architecture for Insider Detection System

5.1 HoneyTokens

Honeypots are realistic but dummy systems that reflect true production systems and are designed to attract malicious users to inappropriately access resources. Combined with subtlely advertised enticements to potential insider threats, honeypots provide a mechanism to determine what motivates the inside attacker and what capabilities the attacker possesses.

A novel idea developed during the workshop and applied in the insider detection process is the notion of a *honeytoken*. A honeytoken is a semi-valuable piece of information whose use can be readily tracked. This could be a credit card number, an Excel spread sheet, a database entry, or a login and password. A honeytoken is an entity that has no authorized use. Honeytokens can be used for the initial detection of insider threats, then those threats can be redirected to honeynets to confirm if a violation has occurred, potentially learning more about the threat.

In the Pal scenario described in Section 3, the honeytoken takes the form of a web page which lists (fictitious) operatives in the geographic region of interest to the MI. The data fusion group, detailed in a subsequent section, exploits the detection of honeytoken access as one of a range of indicators of malicious behavior. In other examples of honeytokens (e.g., login password) it is possible to use the false information to track activities (e.g., in a controlled account) to more readily discover MI actions, capabilities, and intentions.

5.2 Stealthwatch

One of the sensor types deployed in support of the insider threat workshop was StealthWatch. StealthWatch (http://www.lancope.com) provides traffic profiling (e.g., data flow analysis) and host profiling such as zone profiling (e.g., monitoring which hosts can talk to which). StealthWatch creates a concern index (CI) based on flow analysis, reconnaissance activity, and anomaly detection. Any host reaching an accumulated CI level above 20,000 generates an alarm. Stealthwatch was deployed to monitor scanning, downloads, and inside connections. StealthWatch enabled an analyst to identify Jack within a day of his malicious activity, which initiated on February 11th. However, StealthWatch did not detect Pal and Jill who exihibited no scanning, download or insider connection anomalies.

5.3 Structured Analysis

Using the cyber observable framework introduced at the beginning of this article in Figure 2, the structured analysis approach considered a range of cyber observables associated with adversaries to perform top-down, real-time, model-based detection of MIs. Knowledge of network mission and network configuration together with likely adversarial actions were used to filter and highlight adversary behavior. MI actions modeled included reconnaissance (e.g., via web browsing or net scans), access (e.g., privilege escalation), entrenchment (e.g., via deployment of sensors), extraction and exfiltration (e.g., unauthorized downloads), and communications (e.g., coded messages or covert channels). A range of sensors and data logs are relevant to detecting many of the cyber actions. For example email communication patterns would provide insight into the social network of a malicious insider. Conversely, large downloads might signal data exfiltration.

The structural analysis group (SAG) modeled two insiders, Pal and Jack, considering temporal characteristics of protocols such as event proximity (e.g., immediate vs. days vs. years) and observable ordering. The Pal detector exhibited 3% false positives and no false negatives and the Jack detector had 1% false positives and 50% false negatives.

The structured analysis approach detected all three simulated MIs, Pal, Jill, and Jack, at various times. Pal was put on the watch on December 11th, 2003, two days after Pal initiated behavior. Six days later an alert was issued for Pal. The News Administrator Jill was put on the watch list on January 24th, 2004, even before her suspicious behavior started on February 11th. Four days later, on February 15th, SAG issued an alert for Jill. Finally, SAG issued an alert on 17 February for Jack, six days after he initiated malicious administrator behavior.

5.4 Data Fusion

In contrast to the preceding model-based, top-down analysis of observables, a data fusion approach explored bottom-up combination of observables from multiple sensors including the combination, correlation and analysis of both traditional and novel indicators stored in the Common Data Repository. This low-level data came from available resources on the MITRE DMZ network along with inputs from other teams (StealthWatch and honeynets). Thus data fusion occurred over input from the physical, network, host and application levels. Fusion can occur along a number of dimensions such as the type of sensor (e.g., card reader, authentication, printer, telephone calls) and the level of IP stack (e.g., from network to application).

A proof-of-concept data fusion engine was developed that:

- Fused multiple indicators that were related by IP, user name, etc. into a single indicator
- Allowed customer selectable weights for each element in the observable taxonomy
- Allowed site specific criteria for generating an alert and for submission of names to a watch list based on:
  - Accumulated indicator weight
  - Accumulated indicator breadth, i.e., the number of unique sources of evidence

In addition, a novel sensor was developed that performed email consistency checking. The email sensor analyzed PGP signatures and PGP-encrypted attachments and generated error messages to report if the PGP data did not decode properly. A simple, flexible data fusion engine was implemented in the JESS rule system.

Figure 4 illustrates a series of logged messages from the operation of the data fusion engine that report various inferences being made on the basis of sensor input for the scenarios on the DMZ network introduced in Section 3.

The data fusion engine was able to find indications of malicious activity in two out of the three scenarios, with little tuning. One in-the-wild incident was discovered, in particular a user who was excessively downloading news. Perhaps equally significant is the substantial data reduction that resulted. The data fusion engine needed to examine 7.4 millon records to analyze the activity of the 75 users on the system. 259 indicators were generated for 24 of those users. The data fusion approach exhibited very low false positive rates (0 or 3%) with moderate to high false negative rates (33 to 66%) using a threshold of the maximum change between any two detected users.

```
Cyber-Access, user324, weight 1, at 2003-12-10 11:14:38, from
news.mitre.org
   su to user9676 failed for non-admin user user324 on /dev/pts/0
--------------------------------------------------------------------------------
Physical-Access, user295, weight 5, at 2003-12-15 19:19:37,
   After hours badge access for user295
--------------------------------------------------------------------------------
Cyber-Extraction-Exfiltration, user2649, weight 5, at 2004-01-06
15:37:28, from nrrc-springfield.mitre.org,
   Data was uploaded to an external server via FTP protocol
--------------------------------------------------------------------------------
Cyber-Reconnaissance, user295, weight 10, at 2004-01-09 20:57:18, from
nrrc-springfield.mitre.org,
   User user295 searching in non-need-to-know country korea
--------------------------------------------------------------------------------
Cyber-Communication, user9, weight 15, at 2004-02-10 22:14:48, from
cvw.mitre.org,
   User user9 received email with masqueraded content from
user11649@yahoo.com
--------------------------------------------------------------------------------
Cyber-Reconnaissance, user1, weight 5, at 2004-02-10 13:54:15, from
nrrc-plymouth.mitre.org,
   Ongoing CI violation -- 066.170.227.074 currently has 49613 alerts
of this type…
--------------------------------------------------------------------------------
Cyber-Extraction-Exfiltration, user295, weight 8, at 2004-02-12 23:54:58,
from dmzsrv1.mitre.org,
  User user295 sent encrypted email to user9983@comcast.net
--------------------------------------------------------------------------------
Cyber-Extraction-Exfiltration, user1, weight 15, at 2004-02-20 12:25:03,
from nrrc-erie.mitre.org,
  user1 sent email with masqueraded content to user1@mitre.org
```

Figure 4. Data Fusion Log

6. Evaluation: Accuracy and Speed

We designed evaluation methods to assess both the accuracy and timeliness of insider threat detection. Accuracy was measured in terms of false positives and false negatives ad defined:

- False positives = # users incorrectly detected as malicious insiders
- False negatives = total # actual malicious insiders - # correctly detected

For each sensor, the maximum change between any two detected users is computed, and then false positive and false negative measures are computed at several points along the graph. Detection results were measured at three weighted thresholds: maximum change, 50% of the highest factor, and 10% of the highest factor.

In contrast to accuracy, timeliness were measured as the difference between the time when an insider's malicious activity begins, the time at which they are put on a watch list, and the time at which an insider threat alarm occurs. We consider these measures stricter than would be found in an operational environment which would include a human analyst in the loop to make alert decisions (Stealthwatch results did include a human analyst, whereas the other methods were autonomous).

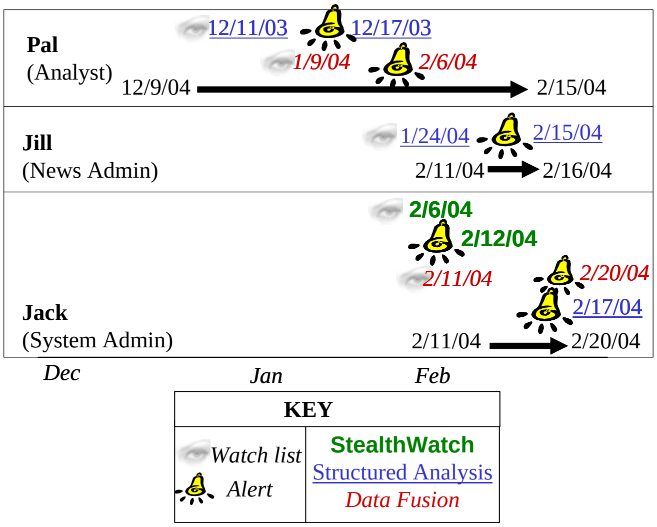


Figure 5. Timeliness of Detection Results

Figure 5 summarizes the timeliness performance of the StealthWatch, structured analysis, and data fusion detectors for the three insiders: Pal, Jill and Jack. One objective was to reduce the time from defection to the time of detection from years to months to weeks to days if not minutes. In Figure 5, the eye icon indicates the day when the sensor put the insider on the watch list. The bell indicates the first day when an alert is issued. The black vectors indicate the start (left of the vector) and stop (right of the arrow head) times of the malicious behavior by the three human MIs. The bold font dates are associated with the StealthWatch sensor (which only detected Jack), the underlined dates are for the structured analysis method, and the italicized ones are for the data fusion performance. In summary, when reviewed across all sensors and methods, of our three MIs, two were detected within one week of their initiation of suspicious activity and the third was detected within two weeks. Even removing some more obvious indicators such as the scanning behavior of Jack, because a multiplicity of sensors provide evidence for inferences Jack would still be detected.

7. Summary

Malicious insiders pose perhaps the most serious threat to organizational cyber assets. Malicious insider behavior is distinct from that of classical external intruders and cannot be detected using traditional intrusion detection methods. In this article, we report results from a challenge workshop that demonstrated how an integration of multiple approaches promises early and effective warning and detection for a range of insider threats. The primary contributions of this work include:

- A taxonomy of cyber assets and cyber actions associated with known malicious insider behavior
- An attribute-based model of known insiders correlated with cyber indicators - classification of classic insider classes (e.g., need to know violators motivated by moral objectives like Montes as opposed to vengeful system administrators) and measures of detection difficulty
- A live network test using simulated malicious insiders modeled on known and projected cases.
- Creation of an eleven million record data set of heterogeneous cyber events including physical access (e.g., badge logs), host access/administration (e.g., password, su, login), user/application level (e.g., web, mail, network news), and network security (e.g., StealthWatch, snort).
- Real-time detection of insider Pal (analyst representing a historical need to know violator), Jill (application administrator) and Jack (system administrator and projected network attacker) exploiting data fusion using a carefully selected set of heterogeneous sensors

The workshop insider cases and dataset are being reused by researchers and have inspired new sensor development. However, while this research makes initial contributions to the malicious insider, it equally raises many new research directions. These include the need for more refined malicious insider models, more elaborate cyber actions/observables taxonomies, more comprehensive test corpora, and more sophisticated detection algorithms.

Acknowledgments

Special appreciation to Jeff Sebring and Rich Pietravalle who provided information security expertise and to Penny Chase, Laurie Damianos, and Tony Ricciatti who acted as malicious insiders.

References

1. Anderson, Robert H.; Bozek, Thomas; Longstaff, Tom; Meitzler, Wayne; Skroch, Michael; and Van Wyk, Ken. August, 2000. Research on Mitigating the Insider Threat to Information Systems - #2. Workshop Proceedings. http://www.rand.org/publications/CF/CF163.

2. [DSS 1999] Recent Espionage Cases 1975-1999 (Defense Security Service). Security Research Center. Defense Security Service. Monterey, California September 1999. http://www.dss.mil/training/espionage

3. [Hanssen 2003] A Review of the FBI's Performance in Deterring, Detecting, and Investigating the Espionage Activities of Robert Philip Hanssen August 14, 2003 Office of the Inspector General. http://www.usdoj.gov/oig/special/03-08/index.htm

4. Hayden, Lt Gen Michael V. July 1999. The Insider Threat to U.S. government Information Systems. National Security Telecommunications and Information Systems Security Committee (NSTISSAM) INFOSEC 1-99. http://www.nstissc.gov/Assets/pdf/NSTISSAM_INFOSEC1-99.pdf

5. Herbig, Katherine L. and Wiskoff, Martin F. July 2002. Espionage Against the United States by American Citizens 1947-2001. Defense Personnel Security Research Center PERSEREC-TR 02-5. http://www.ncix.gov/news/2002/oct/Espionage.pdf

6. Jones, Anita K. (chair). November 1-2, 2001. White Paper: Cyber-Security and the Insider Threat to Classified Information. Computer Science and Telecommunications Board, National Research Council. http://www7.nationalacademies.org/CSTB/whitepaper_insiderthreat.html

7. Matzner, Sara Nov. 2004. Approaches to Insider Threat Mitigation. *ISSA Journal*, Feature article pp.6-8.

8. Matzner, Sara and Tom Hetherington. Summer 2004. Detecting Early Indications of a Malicious Insider", *IA Newsletter*, 7(2): 42-45.

9. Maybury, Mark, Sebring, Jeff, Chase, Penny, Chiekes, Brant, Pietravalle, Richard, Costa, Mick, Zarrella, Guido; Gaimari, Bob; Brackney, Dick, Lehtola, Penny, Matzner, Sara; Hetherington, Tom; Marin, Jack; Wood, Brad; Sibley, Conor; Longstaff, Tom; Spitzner, Lance; Haile, Jed; Copeland, John; and Lewandowski, Scott. 2004. Insider Threat Challenge Workshop: Final Report. MITRE Technical Report 04B-14.

10. Montes, Anna. September, 2001. Affidavit. http://news.findlaw.com/hdocs/docs/montes/usmontesaff901.pdf

11. Shaw, Eric D.; Post, Jerrold M.; and Ruby, Kevin G. Inside the Mind of the Insider. http://www.securitymanagement.com/library/000762.html

12. Spitzner, Lance. “Honeypots: Catching the Insider Threat” ACSAC, Las Vegas, Dec 2003

13. Webster, William H. (chair). March 2002. A Review of FBI Security Programs Commission for Review of FBI Security Programs. U.S. Department of Justice. http://www.fas.org/irp/agency/doj/fbi/websterreport.html