Combatting Insider Misuse, with Relevance to Integrity and Accountability in Elections and Other Applications

Peter G. Neumann
Principal Scientist
Computer Science Laboratory
SRI International EL-243
333 Ravenswood Avenue
Menlo Park CA 94025-3493
Tel 1-650/859-2375
Fax 1-650/859-2844
Neumann@CSL.sri.com
http://www.csl.sri.com/neumann/

Prepared for the Dagstuhl Workshop
on Insider Threats, 20-25 July 2008
July 25, 2008

[Note: I regret that I was unable to attend the workshop. However, Matt Bishop (an organizer of the workshop) graciously presented this paper for me at the workshop. He also prepared a set of slides for the talk, which are available at http://www.csl.sri.com/neumann/dagstuhl-bishop.pdf . My paper is at http://www.csl.sri.com/neumann/dagstuhl-neumann.pdf . PGN]

Abstract

Various risks of insider misuse arise at different layers of abstraction. This observation leads to a perspective on insiders that is both hierarchical and context-dependent. This position paper examines systemic approaches that might be most useful in overcoming the risks. It applies these approaches to the problems of developing and operating computer-related systems that would be suitable for use in applications requiring trustworthy systems and networking, such as critical infrastructures, privacy-preserving database systems, voting systems, and so on. It also examines the relevance of the Saltzer-Schroeder security principles to elections.

Ultimately, insider misuse cannot be sensibly addressed unless significant improvements are made in system and networking trustworthiness, architecturally, developmentally, and operationally.

Some of the distinctions presented here are intentionally not all clear-cut. There are nuances that must be considered, because blurrings exist among what some people might superficially think are dichotomies. These subtleties can be quite significant in assessing how we should approach insider misuse within the more general context of system and network trustworthiness.

This position paper draws on and extends an earlier one, The Challenges of Insider Misuse, that the author wrote for the 1999 RAND Workshop on Preventing, Detecting, and Responding to Malicious Insider Misuse, 16-18 August 1999, of which he was a co-organizer. Sadly, many of the conclusions of that earlier analysis are still relevant today.

1 Introduction

We consider a broad spectrum of problems and appropriate techniques for preventing, detecting, diagnosing, and understanding those problems. Although we focus primarily on insider misuse, we do not ignore the obvious problem that outsiders who are capable of penetrating system can often logically become insiders – for example, through a combination of exploited vulnerabilities and social engineering. Thus, the distinction between insiders and outsiders can be a slippery one in the absence of adequate overall system security aimed at reducing the likelihood of penetrations and external denials of service that do not even require system access. Similarly, the distinction between malicious acts and accidental events is often misleading, in that events occuring accidentally could often be triggered intentionally. Thus, a system typically needs to be protected against both kinds of events. As a consequence, sensible system architectures must address misuse by both both insiders and outsiders in a coordinated way.

1.1 A Hierarchical View of Insiders

For present purposes, an *insider* is simply a system user that can misuse certain privileges. For generality, we allow that the 'user' could be some sort of computer entity – process, agent, or system – that may or may not be acting on behalf of specific human users. There are many different definitions of insiders – some of which are in conflict with one another. For example, some definitions of 'insider' exclude outsiders who have usurped privileges of insiders, whereas other definitions those outsiders who have effectively gained the privileges of insiders. Some of these definitions fail to define either 'insider' or 'outsider', simply implying that one is not the other. A definition attributable to a National Research Council study report is "a person who is allowed inside the security perimeter of a system and consequently has some privileges not granted outsiders." That NRC CSTB definition suffers from its implicit distinction between insiders and outsiders, and its defining one as the opposite of the other – which is itself seemingly undefined. It also suffers from the reality that there is typically no single security perimeter; indeed, there may be different perimeters for different aspects of trustworthiness that satisfy different subsets of the set of given requirements, and compromises of one perimeter may also compromise others. Furthermore, as we see in the analysis here, the supposed perimeter may actually be the entire system and its operational environment, possibly including the entire Internet – particularly in badly designed systems. (See also the consideration of the insider threat in INFOSEC Research Council Hard Problem List [8].)

One man's outsider is another man's insider. The determination of who is an insider and who is an outsider is relative to what boundaries might be assumed to exist. For example, someone who can manipulate bits in memory or secondary storage using hardware diagnostic tools might be called a hardware insider. Someone who can manipulate operating system parameters because she has authorized use of some sort of root privileges would be considered an insider with respect to the operating system. Someone who can tamper with a browser because he is the maintainer of Web facilities would be considered an insider with respect to the webware. A similar analysis is given by Matt Bishop [2]. See also a subsequent paper [3].

A useful distinction exists among three alternative situations with respect to any particular layer of abstraction: *compromise from outside, compromise from within, and compromise from below.* The second alternative represents actions of an insider with respect to that layer of abstraction. The third one represents actions of an insider with respect to some lower layer of abstraction.

One historical explicit example of such a hierarchy was presented by the Multics ring structure [20]. Because of the inherent structure of the ring mechanism, each ring could be compromised only from within or below, and in principle could not be compromised from outside (that is, from a higher layer or externally from the same or higher logical layer, as in the case of network connections).

2 Risks of Insider Misuse

To understand the problems, we need to explore various kinds of insiders further, as well as classes of misuse, threats, vulnerabilities, risks, and the types of knowledge and experience that might be applied.

2.1 Classes of Insiders

Differences among users may involve physical presence and logical presence. For example, there may be logical insiders who are operationally physically outside, and physical insiders who are logically outside. For present purposes, we focus on both logical and physical insiders.

Clearly there are different degrees of logical insiders, relative to the nature of the systems and networks involved, the extent to which authentication is enforced, and the exact environment in which a user is operating at the moment. A user may be an insider at one moment and an outsider at another. A user may be also be an insider within one operational domain and an outsider with respect to another operational domain.

For example, if a system supports multilevel security (or multilevel integrity [1]), or even some form of multilevel availability or multilevel survivability [12]), then the existence of compartments suggests that a user can be an insider in one compartment but an outsider in another compartment, or an insider at Top Secret but an outsider with respect to all compartments. In that a user may operate at different levels and compartments at different times, the concept of insider is both temporal and spatial. In some sense, all users of a single-level Top-Secret system could be called insiders with respect to confidentiality, although they would appear to be outsiders relative to those others who were cleared into a particular Top Secret compartment. Similarly, a user could be an insider with respect to multilevel security and an outsider with repect to multilevel integrity. Thus, everything is relative to the frame of reference – what the user is trusted to be able to do, what privileges are required, and what data or programs are being referenced, and whether the user authentication is strong enough to ensure that user identities are not spoofed.

With respect to conventional operating systems, database management systems, and applications functioning as single-level systems (even if system high), at one extreme there are ordinary insiders who have passed the login authentication requirements; at the other extreme, there are users who are authorized to become superuser or an equivalent holder of ultraprivileges. However, consider a system in which the superuser privileges have been partitioned starkly (as in Trusted Xenix [7]), where no one user holds all of the privileges, and where the granted privileges are insufficient to gain possession of all other privileges. (The iterative closure of static privileges augmented by privilege-changing privileges must also be considered whenever we consider what privileges are actually attainable by a given user or group of collaborating users.) In that rather ideal case, we would have no complete insiders, but many different types of relative insiders. Unfortunately, in the absence of meaningfully secure systems and differential access controls that are properly defined, properly implemented, and properly administered, that ideal may still be mostly a fantasy.

Thus, we are confronted with a wide potential range of insiders, and conclude that the notion of "insider" is necessarily multidimensional. For the rest of this section, we consider insiders generically, and do not attempt to make fine nuances among different kinds of insiders. We assume that relative to a particular computational framework, insiders are users who have been authenticated to operate within that framework; where necessary, we qualify that to include reference to the authorized privileges that many be specifically associated with a particular instance of an authenticated user. For example, in several cases we make a distinction among insiders – for example, between ordinary users on one hand and system administrators or other users with extreme privileges on the other hand.

2.2 Classes of Insider Misuse

Along with the range of meanings of the term "insider" is associated a variety of classes of insider misuse.

One immediate categorization of insider misuse involves intent, as in intentional versus accidental misuse. Even among intentional misuse, there is a wide range of possible actions – from outright malice to relatively benign annoyance, with many degrees in between.

A second categorization involves the evidential nature of the misuse, that is, whether the misuse is intended to be detected or hidden. System and network denials of service may be overt, in that they are readily obvious once they are enabled. (Whether their cause is accidental or intentional is sometimes less clear.) On the other hand, insertion of stealthy Trojan horses that act as sniffers or that quietly leak information are typically intended to be covert, and their desired purpose may be to remain undetected for long periods.

Although the focus of this workshop is primarily on intentionally malicious misuse, it is generally unwise to ignore accidental misuse. For example, the apparent success of what might be considered accidental but tolerated misuse could easily inspire subsequent malicious misuse. Furthermore, it is generally unwise to ignore stealthy forms of misuse. To the extent that detecting accidental misuse can be dealt with by the same mechanisms that are used for intentional misuse, accidental misuse need not be treated separately. Similarly, to the extent that stealthy misuse can be dealt with by the same mechanisms that are used for more obvious misuse, stealthy misuse need not be treated separately. However, remember that seemingly accidental misuse may in fact be intentional misuse in disguise, and stealthy misuse may be extremely dangerous; thus, it is potentially risky to ignore any particular mode of insider misuse. Nevertheless, the responses may differ depending on whether the cause is deemed to be accidental or malicious.

3 Threats, Vulnerabilities, and Risks

There are clearly differences in the nature of the threats. However, it must be remembered throughout that once an apparent outsider succeeds in penetrating a system boundary, he/she effectively becomes an insider from a mechanistic point of view. Although an insider might conceivably have greater knowledge of the environment, and may thereby present greater threats, the differences between insider threats and outsider threats are often not stereotypically characterizable. Because of the flaky operating system and networking infrastructures today, outsiders have little difficulty in carrying out nasty denial of service attacks and destructive integrity attacks, whether or not they need to actually penetrate any systems. Virtual private networks may tend to complicate the situation, and demand special attention. Nevertheless, if a system complex has meaningful authentication, many of the outsider threats can be made much less riskful, whereas most of the insider threats clearly remain. Also, firewalls that are well-designed, well-implemented, and well-configured can help somewhat, but today are also largely vulnerable to many attacks (such as active pass-through attacks using http, JavaScript, Active-X, PostScript, other forms of executable content, cross-site scripting, SQL injection, and bogus URLs in phishing attacks). The availability of meaningful additional authentication for insiders could be useful in inhibiting masquerading. With extensive monitoring, robust authentication may also help discourage misuse – especially if the identity of the perpetrator could be established and traced reliably. This may be especially relevant to insider misuse, if the true identity of the apparent user can be unequivocally determined (subject to exploitations of operating-system vulnerabilities – including manipulations of audit trails).

It is of course useful to consider insider threats in their own right. In today's systems, insider vulnerabilities and outsider vulnerabilities are both out of control. Serious efforts are needed to improve security and reliability of system and networks, and indeed to improve the overall survivability in the face of a wide range of adversities. With good external security in critical systems, insider risks may be much more serious than outsider risks. However, meaningfully precise access control policies and meaningfully secure differential fine-grained access controls may alter the nature of the insider threats. This document assesses the situation as it exists today, how that situation might change, and what future research is essential to improve both detectability and response.

Table 1 itemizes some of the threats that appear to differ from outsiders to insiders. It ignores threats that

Table 1: Threats to Security

Attribute	Outsiders	Insiders
Access controls	Unprivileged exploitation of inadequate controls	Privileged manipulation of access controls
Confidentiality	Unencrypted password capture or compromise of encrypted passwords	National security leaks and other disclosures; access to crypto keys
Integrity	Creating Trojan horses in untrusted components, Word macro viruses, and untrustworthy Web code	Putting Trojan horses or trapdoors in trusted and untrusted components; ...-in-the-middle attacks
Denials of Service	External net attacks, flooding, physical harm to exposed equipment	Disabling of protected components, exhaustion of protected resources
Authentication	Penetrations, attacks on PKI/authentication infrastructures, war dialing	Misuse of intended authority by over-authorized users, usurpation of superuser, access to root keys
Accountability	Masquerading, attacks on accounting infrastructures	Hacking beneath the audit trails, altering audit logs, compromising misuse detection
Other misuses	Planting pirated software on the Web	Running a covert business, insider trading, resource theft

are common to both outsider and insider perpetrators, such as carrying out personal attacks on individuals or corporations through an anonymous e-mail remailer, sending spams, creating monster viruses from a toolkit, creating risky mobile code, tampering with existing mobile code, intentionally crashing a system or component (although there are potentially mechanistic differences in causing crashes between insiders and outsiders), and so on. (Although an outsider who has successfully penetrated a system may seem like an insider, the depth of personal knowledge may be different.) Nevertheless, to simplify the table, penetrating outsiders are logically considered as outsiders unless they are knowledgeable enough to appear indistinguishable in something like a Turing-test sense from the insiders as whom they are masquerading – as might be the case with disgruntled recent ex-employees. More realistically, the indistinguishability may be more like the ability of an outsider to masquerade as an insider if just a little social engineering is all that is required.

3.1 Knowledge Required and Knowledge Used

In systems with weak authentication, there may be naïve outsiders who, when they subsequently appear as insiders, are obviously distinguishable from experienced insiders. On the other hand, there may also be naïve insiders. Nevertheless, the knowledge used to perpetrate misuse may be of value to the analysis associated with detected misuses.

At least superficially, some differences typically exist in the knowledge available, the knowledge required for various types of misuse or already available without further study, experimentation, or effort, and the knowledge actually used in perpetrating insider misuse. Those outsiders who are in practice indistinguish-

Table 2: Knowledge Gained and Used

Outsiders	Ordinary Insiders	Privileged Insiders
Direct info and inferences from web info (such as penetration scripts), help files, social engineering; chats/ BBoards helpful	Experience gained from normal use and experiments; familiarity with sensitive files, project knowledge; collusion easy	Deep knowledge from experience; ability to change and abuse privileges; ability to create invisible accounts; collusion dicier?

able from insiders (as noted above) are thus considered as insiders.

For example, insiders might have greater knowledge of what to look for in terms of sensitive information and particularly vulnerable programs in which to plant Trojan horses. In system-high systems, legitimate insiders are already likely to be gratuitously granted information to which they do not necessarily need access. In compartmented multilevel-secure systems, users would have to be cleared, although that works both ways: a user not entitled to access a particularly compartment is effectively an outsider with respect to that compartment, and indeed may not even know of the existence of the compartment if the system is properly implemented and operational procedures are properly enforced. But users cleared into that compartment clearly have an enormous advantage for potential misuse over users who are not – assuming isolation is suitably enforced and operationally deployed.

In Table 2, we make a distinction among outsiders, general legitimate insiders, and specially privileged insiders such as system administrators (who tend to be totally trusted), recognizing that we are lumping together users with common logical characteristics.

4 Exploitations

There are also differences in how vulnerabilities can be exploited, and the risks that may ensue.

4.1 Exploitations of Vulnerabilities

There is a likelihood that an experienced insider can operate close to normal expected behavior (especially if engaged in a long-term effort at what in terms of a statistical misuse detection system would resemble statistical-profile retraining), which would be more difficult to detect. This increases the need for a variety of analysis techniques and correlation (see below).

Today, we have pervasive deficiencies in authentication, authorization, accountability, operating system security, network security, and intelligently deployed access controls. Given the absurdly poor state of the art in defensive security, the differences between outsider exploitations and insider exploitations may be less relevant than they would be in the presence of good security.

Insider exploitations would become conceptually different in the presence of better system security, but would still present a problem. Enormous benefits could result from intrinsically better operating system security, network security, authentication, and encryption. One of those benefits is that the detection and response problem could be much more precise, rather than encompassing all of security (or the lack of it), as is the case at present. This is absolutely vital, and must not be ignored.

Table 3: Potential Severity of Risks Incurred

Outsiders	Ordinary Insiders	Extra-Privileged Insiders
Very serious in badly designed and poorly implemented systems, perhaps less serious with good user authentication and good auditing	Potentially very serious unless strong separation of roles, MLS, and differential access controls; beware of system-high systems	Extremely serious, even with strong separation of roles and separation of privilges, MLS levels and compartments; misuse of multipurpose root privileges is inherently risky

4.2 Potential Risks Resulting from Exploitations

The potential risks may vary significantly from outsiders to ordinary insiders to highly privileged system administrators. However, it is in itself risky to give too much credence to these differences, because of several factors:

- Outsiders who become insiders can in some respects potentially wreak havoc similar to that of malicious (or even benign) insiders, although in any particular case the damage could be greater, comparable to, or less.
- Some outsiders such as terrorists may have highly visible major havoc in mind. Other outsiders might try to mask the existence of clandestine Trojan horses, trapdoors, and other system aberrations. In general, exactly the same situation applies to insiders, although the stealthy route would generally be more likely in the presence of strong authentication that hinders insider masquerading and provides a fairly clear chain of evidence. Each type could create highly undetectable effects or massive disasters entailing major risks.
- When the security of systems, servers, firewalls, and networks is weak, outsiders can often do serious damage, whether or not they become insiders.
- Measures of risk are highly speculative, and strongly dependent on the application environment. (One man's feat is another man's poison.)

Thus, it is in general a huge mistake to conclude that outsiders cannot be as destructive as insiders, or vice versa. Nevertheless, there seem to be substantive differences in the potential risks.

5 Illustrative Examples of Insider Misuse

The ACM Risks Forum (http://www.risks.org) is laden with cases of insider misuse. For example, my Illustrative Risks annotated index [15] includes descriptors for each case. In particular, the descriptor "SHI" is relevant here, where S indicates a security problem, H indicates an intentionally deleterious action, and I indicates that that action was performed by one or more insiders; O indicates outsiders as well. Additionally, A indicates inadequate authentication, which can result in an outsider becoming an insider. P denotes a privacy violation. $ indicates financial losses were incurred. ! indicates a death resulted, and * denoted risks of death or injury.

Just a few cases are sufficient to suggest the wide range of applications and the effects of the misuse. (R i j) references refer to the ACM Risks Forum volume i number j; (S i j) references refer to the ACM SIGSOFT Software Engineering Notes volume i number j.

5.1 Higher-Tech Insider Misuse with Relatively Detailed System Knowledge

- $SHI Autotote ex-programmer with insider knowledge hacked the winning Breeders' Cup Pick Six horse-race off-track betting system, after a previous trial went undetected. The off-track system transmitted rsults to a central facility only after the fourth race. Dummy bets were placed for the first four races (which where then altered locally after the fourth race), with wild-cards spanning every possible outcome in the last two races. Drexel frat buddies were implicated (S 28 2:13; R 22 33,38-40). Programmer Chris Harn was sentenced for only a year and a day in jail, because he helped the authorities incriminate his buddies, who received two- and three-year sentences (R 22 65). [Note: the same vendor, Scientific Games, was the culprit in a quick-pick system that omitted the last horse in the field, and remained undetected for an undetermined long time – because of a lack of auditing. Whether this was intentional or accidental remains undetermined (R 25 16).

- $SHIO Harrah's $1.7 Million payoff scam – details never revealed, but Trojan horse chip or electronically triggered payoff suspected (S 8 5); see also *San Francisco Examiner/Chronicle,* 18 Sep 1983.

- $SHI Joseph Jett (Kidder Peabody) created $350M phantom profits, and got a bonus of $9M. His scheme was undetected by KP oversight (double meaning, not intended as a pun) (S 19 4:12).

- $SHI A frequent flier computer scam netted 1.7 million bonus miles (S 14 2).

- $SHI Travel agents in an American Airlines Frequent Flyer ticketing fraud received prison terms (S 16 2).

- $SHI Beijing Hotel managers embezzled $9K by rigging billing records (S 19 4:13).

- $SHI In a UK Clydesdale Bank cash machine fraud, insider fraud was suspected (S 16 2); a bank engineer recorded ATM PINs, fabricated cards, and profited (S 17 3).

- $SHI A Salomon Brothers scandal was aided by misuse of database confirmations (S 16 4).

- SHIe Remote mobile phone configuration changes were made via a Swisscom SMS service (R 21 89).

- $SHI An Australian security firm was forced to close after insider sabotage (R 22 62).

- $SHI Two Lucent scientists were charged with selling PathStar Access Server software to a Chinese firm (R 21 38).

- $SHI A Taco Bell register was reprogrammed to redirect funds (S 22 4:31, R 18 76).

- SHI An FAA programmer destroyed the only copy of source code for flight-control data transfer; authorities recovered encrypted copy at his home (R 20 64).

- SHI+O Hackers penetrated Russian Gazprom, and controlled pipeline flow. Russian police noted a twelve-fold increase in computer crime in 1999 over 1998 (R 20 87).

- SHIf Rogue code in Microsoft software, dvwssr.dll, included rogue password to access thousands of Web sites (R 20 87-89).

- $SHI Trojan-horsed chips in gas pumps enable an overcharging scam (R 20 03).

- $VSHI A disgruntled Reuters computer technie brought down a trading net (S 22 2:23).

- SHAI Former UBS PaineWebber programmer faces U.S. fraud charge in virus attack (R 22 44,46).

5.2 Lower-Tech Insider Misuse of Government Privileges

- !$SHI Spying activities of CIA agent Aldrich Ames, in charge of suses (S 18 4:7).
- $SHI NY police chief indicted for misusing confidential database (S 13 4).
- SHI 3 police officers sentenced for misusing Police Nat'l Computer (S 14 2).
- $SPHI 45 LA police cited for searching private computer records (S 18 1:21).
- $SPHI Theft of 8.5K criminal records; investigator, 2 police indicted (S 18 2:16).
- SHI$ Virginia DMV fraud again; illicit driver licenses cost up to $3,500 each (R 23 94); some years ago, the going rate was only $25 for one notorious Virginia DMV office that engaged in widespread issuing of unapproved driver licenses.
- SHI 24 California DMV clerks fired in fraudulent license scheme (S 23 1:14, R 19 27).
- SHI DMV security code disclosed at hospital in New Haven (R 18 28).
- $SHAI Massive NY City tax fraud wiped out $13M in taxes; many people were implicated (S 22 2:23,R 18 63).
- SHAI$ Former Drug Enforcement Agency employee sentenced to 27 months for selling protected data (R 22 44).
- SHPI FBI employee snooped through confidential police databases, sentenced to 12-month prison term (R 23 23).
- SHI IRS agent accused of giving defendant tax data on judges/jurors/... (S 16 3:9).
- $SHI Browsing by IRS employees: curiosity to fraud (S 19 4:13).
- $SHI Social (In)Security employees sold 11,000 SSNs to activate cards stolen in the mail (S 21 4),(R 18 02).
- $SHI Military pay fraud netted $169,000 using bogus account (S 23 1:14, R 19 26).
- $SHIO 40 arrested (9 postal workers) in massive D.C. credit-card fraud (S 19 2:7).
- $SHI Olympia WA Health Department check scam detected; four indicted (S 18 1:12).
- SHI Massachusetts welfare fraud investigators fired: tax-record misuse (S 22 1:20).
- SHI Database misuse by 11 prison guards in Brooklyn: leaking names of informants to prisoners, warning about searches, ... (R 19 20).
- $SHI Danish credit-card fraud: mailman intercepted postal mail (R 22 61).
- SHI Nova Scotia worker fired for deleting her speeding ticket (R 22 84).

5.3 Lower-Tech Insider Misuse of Other Privileges

- SHI Hacker-nurse unauthorisedly changed prescriptions, treatments (S 19 2:5).
- $SHPI 48,000 Wachovia customers, 600,000 Bank of America customers, and others from Commerce Bank and PNC Bank of Pittsburgh notified that their financial records were potentially compromised by insider operation; laptop with information on 16,500 MCI employees stolen (R 23 88).
- SHI 6000 AIDS records stolen from Miami hospital PCs and diskettes (S 19 2:9); bad prank follows (S 20 5:10).
- SHI 4000-person AIDS database leaked to press, Pinellas County, FL (R 18 48,53); former Health Dept

employee and roommate charged (Reuters, 15 Feb 1997).

- $SHI Volkswagen Corp lost $260M to computer based foreign-exchange fraud (S 12 2); 5 people (4 insiders, 1 outsider) convicted, maximum sentence 6 years.

- $SHI Teller embezzled $15K, caught via computer audit-trail (S 19 3:10).

- $SHI Japanese bank workers stole 140 million yen by PC (S 20 2:12).

- $SHI Bank executive in Malaysia transferred $1.5M (S 15 5).

- $SHI $550,000 Tokyo bank fraud suspected in funds transfers (S 19 2:6).

- $SHI Visa victim of PC theft with info on 314,000 credit-card accounts (R 18 62).

- $SHI Tower Record credit-card info offloading; 2 convicted (S 21 4, R 18 02) .

- SHI Time Inc. employee peddled credit-card information (to detectives) (S 17 4).

- $SHI NJ car dealership in theft of 450 credit-card numbers, almost $4M (S 19 2:7).

- SHI During a Verizon strike, 2 NY employee saboteurs cut a power cable next to a telephone cable, suffering burns (R 21 03; S 26 1:28).

- SHI$ Former Cisco accountants sentenced to 34 months and restitution of $8M for unauthorized computer access, fraud (R 21 82).

- SHI ISP/bookseller intercepted Amazon messages, trying to get competitive advantage; fined $250,000 (R 20 66).

- *SHI Hacker-nurse unauthorisedly changed prescriptions, treatments (S 19 2:5; R 15 37,39).

- $SHI $1M internal computer fraud at Pinkerton (S 16 4).

- SHPI Indian call centre 'fraud' probe: info on 1000 customers sold (R 23 93).

- SHPI A Pakistani outsourcee of a University of California San Francisco health-care group threatened to release personal data files unless she were paid back wages (R 22 97). (The threat paid off.)

6 Countermeasures

6.1 Prevention of Insider Misuse

Where possible, prevention is vastly preferable to detection and attempted remediation. Clearly, there are cases in which detection is necessary – for example, whenever prevention is either not possible or uncertain. For example, the Multics experience beginning in 1965 (see [4, 5] and http://www.multicians.org/) should have taught everyone a lot about the importance and value of prevention – for example, defining what is meant by security, isolating privileged execution domains from less privileged executions (with 8 rings of protection), isolating one user from another while still permitting controlled sharing (via access control lists), access-checked dynamic linking and revocation, and using some sensible software-engineering concepts. In addition, use of some of the Saltzer-Schroeder [21] security principles would be directly relevant to minimizing insider misuse. The most obviously applicable principles here are separation of privileges, allocation of least privilege, and open design. In addition, ease of use (termed "psychological acceptability" by Saltzer and Schroeder) could provide incentives for insiders to avoid the excuse of security being too complicated, which otherwise often results in the creation of unnecessary vulnerabilities. These and other principles are discussed further in the context of election systems in Section 9.

If there is no meaningful security policy to begin with, then the task of detecting and identifying deviations from that policy is very difficult. If there is no fine-grained differential prevention in systems and networks, then even if there were a meaningful security policy, it would be difficult to implement it. With respect to insiders, any enterprise operating within a system-high approach is inherently implying that there is no such thing as “insider misuse” because everything is permitted to all insiders. Thus, to have any hope of detecting insider misuse, we first need to know what constitutes misuse. Ideally, it would then be much better to prevent it rather than to have to detect it after the fact.

Although relevant not specifically to insider misuse, but more generally to the development of trustworthy systems, several thrusts are of interest. A report on how to develop principled assuredly trustworthy composable architectures [13] and subsequent reflections on trustworthiness [14] are relevant here.

Fine-grained access controls are of particular interest in minimizing insider misuse. These controls date back to the Multics file system [6] in 1965, and have been the subject of refinement and alternative approaches ever since. Past work on strongly typed tagged capability-based systems (e.g., [16]) could also considerably reduce opportunities for insider misuse. Of course, various forms of multilevel security and multilevel integrity could also help to narrow down the possibilities for insider misuse, albeit with the associated administrative baggage.

Also relevant is paper by Paul Karger [9] that applies access controls to programs. That approach might be interesting in controlling the extent to which insider-introduced malware could do damage, assuming that the insider is not privileged to alter the access controls. However, it is otherwise not directly relevant to user access controls.

6.2 Specification of Sound Policies for Data Gathering and Monitoring

Commericial products for misuse and anomaly detection tend to assume a collection of known vulnerabilities whose outsider exploitations are associated with known policy violations. The existing systems tend to be aimed primarily at penetrators and intrusion detection, and are not easily applied to detecting insider misuse. Policies for insider misuse tend to be strongly application-domain specific, and should dictate what is to be monitored, at what layers of abstraction. Thus, it is essential to have a well-defined policy that either explicitly defines insider misuse, or else a policy that explicitly defines proper behavior and implicitly defines insider misuse by exclusion.

A much better understanding of the application domain is needed for monitoring users for potential insider misuse. Also, more detailed data may need to be collected. Furthermore, when someone is suspected of devious behavior, it may be desirable to go into a fine-grain monitoring mode (such as capturing keystrokes), although that has its own serious potential privacy problems.

Audit trails in existing operating systems and applications are typically not well suited to detecting insider misuse. Today, commercial systems for misuse detection generally rely on system audit trails, network packet collection, and occasionally physical sensors for their inputs. Other sources of input data are necessary for detecting insider misuse, including detailed database and application logs. In either case, the analysis systems need to obtain some knowledge of the perpetrator if they are to trace the detected misuses back to their initiators. In closed environments, there can be much better user authentication than in open environments, although masquerading is still possible in many operating systems and application environments. Whenever that is the case, the actual choices of data to be gathered for insider-misuse detection tend to differ from that of intrusion detection. However, the existence of logical insiders who are physically outside and logical outsiders who are physically inside may make such distinctions undesirable – suggesting that making the assumptions (such as there are no outsiders, or there is no insider misuse) is unwise. The necessary use of encryption for stored inforamtion in highly sensitive systems may also complicate the gathering of information on potential insider misuse, and necessitatge capture of unencrypted content – which raises serious some serious security and privacy concerns.

6.3 Detection, Analysis, and Identification of Misuse

In the absence of good prevention, it is of course desirable to detect known defined types of misuse (e.g., through rule-based detection) or otherwise unknown types of anomalous misuse (e.g., seemingly significant deviations from expected normal behavior). The latter type of detection could be particularly important in identifying early-warning signs of misuse.

Because there are potential differences in the data that may need to be collected, there may be some differences in the approach to detection of misuse among the different types of misuse, depending on the relative roles of insiders and insider misuse. If insiders can exist only within local confines (for example, as in the case of a multilevel security compartment in a system with no remote users and no Internet connectivity), it may be unnecessary to collect packets and other network data – which themselves constitute potential security and privacy risks. On the other hand, if privileged insiders are also able to access their systems remotely (for example, telnetting from outside) and are in some sense then indistinguishable from outsiders at least geographically or from their external Internet presence, then networking data may also be relevant. Clearly, the presence of strong authentication has an impact on carrying out insider misuse detection.

Similarly, there may be differences in data retention requirements for an anomaly and misuse system. If the intent is to gather sufficient information to prosecute insider misusers, then the situation is quite different from insider misuse detection whose aim is to merely detect the presence of misusers so that other extrinsic methods (such as wiretaps, cameras, and physical surveillance) can be invoked. (These differences may also apply to outsiders – although the relative priorities are likely to be different.) In general, long-term retention of raw audit logs and of digested (analyzed) data is recommended.

Detection techniques may be needed for both insider misuse and outsider misuse. Basic techniques for detection and analysis can be similar in both cases, although the parameters are likely to be different. However, in that existing analysis systems do not tend to adequately address insider misuse, some new techniques may be useful, particularly for interpretation and response.

- For example, consider signature-based detection tools, such as expert systems. The rules for outsider misuse are generally established based on trying to detect exploitations of known vulnerabilities. The rules for insider misuse must be based on knowledge of what kind of misuse is considered dangerous. These rules must take into account discrepancies between the actual access controls and what kind of access is considered appropriate. Once again, the extent to which explicit differential access controls can be enforced has a direct influence on what kinds of insider misuse need to be detected. However, the same detection tools can be used in both insider and outsider misuse.

For identification of hitherto unknown modes of insider misuse, some effort is required to apply statistical techniques.

As noted above, the marketplace for intrusion detection is aimed primarily at detecting known attacks by outsiders. The idea of rapidly deploying an analysis system is meaningful for a given firewall, or for a given operating system, or for a given application for which a set of rules have already been written. Insider attacks tend to be much more domain specific, and thus the deployment of a system for insider analysis requires some detailed analysis of the threats and risks, some skilled implementation of rules, judicious setting of statistical parameters, and some further work on analysis of the results. This is not a straightforward off-the-shelf installation process. The applicable term here might better be anomaly and misuse detection, particularly if it has to encompass both insiders and outsiders.

In a multilevel compartmented system/network environment, in which there are presumably no outsiders and in which the insider threat predominates, monitoring and analysis take on multilevel security implications, with many opportunities for covert channels. Monitoring can be done compartmentally, but aggregation, higher-level and cross-compartment correlation on an enterprise-wide basis present serious potential multilevel security problems.

More emphasis is needed on not-well-known forms of insider misuse, on interpretation of detected anomalies, and hierarchical and distributed correlation. Much more emphasis is needed on tools to aid in the deployment and configuration of analysis tools for domain-specific applications. Serious effort might also be devoted to multilevel-secure analysis (and response) in contexts in which MLS systems might be important. Although, procedural and psychological approaches are likely to predominate, much greater awareness of the threats and risks of insider misuse is likely to drive new approaches.

An enormous risk exists relating to false accusations of supposed culprits despite the inability to carry out any definitive traceback to host systems and individual logins. This problem is exacerbated by the long-time retention and undeletable mirroring of erroneous data throughout the Internet. A political analog is what is known as Swift Boating (after the attacks on John Kerry's integrity in the 2004 U.S. Presidential election campaign), where considerable damage can be done.

6.4 Desired Responses to Detected Anomalies and Misuses

In some cases of outsider attacks (particularly denials of service), it is more important to stave off the attacks than to let them continue. In other cases, it may be appropriate to let the attacks continue but to somehow confine their effects (as in the case of honeypots). A similar range of responses exists for insiders. In some cases of insider misuse (particularly where the perpetrator has been identified and prosecution is anticipated), it may be particularly important to detect the misuse, to allow it to continue (perhaps under special system constraints and extended data gathering such as key-stroke capture), and monitor it carefully – without giving away the fact that detailed surveillance is being done.

Thus, there are clearly differences in the desired responses that may be considered once misuses have been detected. However, the full range of possible responses may also be applicable to both insiders and outsiders – although possibly in different degrees in the two cases. In any case in which continued misuse is allowed, serious risks exist that undetected contamination and other integrity problems may occur and remain subsequently. This must be factored into any dynamic strategies for real-time response to detected misuse.

7 Decomposition of Insider Misuse Problems

This section looks at insider misuse in the context of the bigger picture of security. It considers the development process (Section 7.1), operational aspects (Section 7.2, security implications (Section 7.3), the effects of those issues on anomaly and misuse detection (Section 7.4), and response (Section 7.6). It also specifically addresses the importance of user profiling and the desirability of extending it to include psychological factors (Section 7.5).

7.1 Development Stages

- System development methodologies: See [13, 14]

- Requirements: Insider threats are often ignored, even in highly sensitive closed systems that tend to run at system high. Security, reliability, survivability requirements are often short-sighted or left unattended in the subsequent development.

- System architecture and design: Many commercial systems are short-sighted, hindered by their needs for backward compatibility with earlier nonsecure systems and networking, and handicapped by a serious lack of commitment to robust systems. These systems are primarily aimed at low-hanging fruit, or else must have their rule bases updated frequently to keep up with the malware du jour. On the other hand, the research community has progressed significantly in recent years. For example, [17] discusses some of the research directions as well as the desirable characteristics of future-seeking systems for anomaly and misuse detection that could be applicable to insider misuse as well as intruders.

- Implementation: Many serious security-related implementation flaws persist. As just one example, buffer overflows continue to appear despite years of knowledge of their origins and the ensuing risks. Serious attention to software engineering discipline is sorely lacking.

7.2 Operational Aspects

Operational practice and immediate palliatives are important, particularly as they apply to the anomaly and misuse detection platforms themselves. They must be included in the characterization of the problem – because they are part of the solution, without which detection and response are rendered ineffective, and because their inadequacies directly feed into the needs for insider-misuse detection.

- System support. Although vendors may not always have the customer's best interests at heart, many customers seem to be overly naive. In an insider misuse workshop on 12 April 1999, Ed Amoroso mentioned AT&T's experience with Net Ranger (later provided by Cisco). When he and his colleagues finally tried to install it, months after receiving the CD, they discovering that certain files were missing from the installation CD. When they complained to Cisco, the Cisco folks indicated they had never before heard about this problem; apparently no one had ever successfully installed it!

- System administration. System administrators are by and large not particularly experienced at coping with security flaws, security patches, and administering intrusion detection. Much greater help is needed, such as self-configuring detection and analysis tools that can be easily tuned to the threats of greatest significance.

- High-level enterprise management. Existing intrusion-detection systems do little for enterprise-wide monitoring and correlation across multiple network and system platforms.

7.3 Security Implications

- Authentication can seriously impede outsiders, but not if the systems rely on fixed passwords (especially if those passwords are transmitted unencrypted or are replayable).

- Authorization is typically not fine-grained enough, and limits the effectiveness of access controls and misuse detection.

- Accountability: Flexible misuse and anomaly detection must be tuned to insider misuse.

- Boundary controllers are typically vulnerable to denial-of-service attacks.

7.4 Anomaly and Misuse Detection in Context

- Monitoring of appropriate audit trails (operating systems, DBMSs, applications) and network data (packets, network management information) needs to be heterogeneous, diversified, and able to extract just what information is needed. Existing data sources tend to have little abstraction and contain huge quantities of relatively useless information.
- Detection of known misuses is handled fairly simplistically by existing products.
- Detection of anomalies and hitherto unrecognized forms of misuse deserves much greater attention, using a variety of approaches (signature based, profile based). Insider misuse is almost totally ignored.
- Misuses and anomalies that have been detected require some abstraction in their reporting and diversity of sites to which reporting occurs. (For example, EMERALD allows for a wide variety of destinations, including passing the results to higher-layer instances of EMERALD and directly to system administrators.) Correlation is needed across wider scopes across multiple target systems and networks, and across multiple analysis platforms. Much more sophisticated and understandable interpretation of analysis results is essential at varying layers of abstraction.

7.5 Extended Profiling Including Psychological and Other Factors

It is clear from the above discussion that detecting insider misuse must rely heavily on user profiling of expected normal behavior (although some use can be made of application-specific rules). Efforts to date have concentrated on relatively straightforward statistical measures, thresholds, weightings, and statistical aging of the profiles, independent of particular users. Considering that much is already known about insiders, it would seem highly desirable to include additional information in the profiles.

- Physical access to buildings, rooms, and computers based on real-time access records – using badges, logins to local machines, biometric authentications, interactive pager probes, cameras, and other sensor data
- Planning data as to expected activities – travel schedules, special arrangements regarding working hours, facility restrictions.

The combination of physical whereabouts and expected whereabouts could also be used to detect stolen badges or stolen authentication information of people who are highly trusted.

In previous systems aimed at insider misuse as well as intruders, statistical profiling (e.g., in NIDES and EMERALD) provided the capability of monitoring individualized computer activities, such as which editors the user prefers, which programming languages, which mail environment, which variants of commands, and so on. This approach seems to be less relevant today, but still has potential where intrusion is not a primary concern.

Personal on-line behavior can also be profiled statistically by extending the analysis information that is recorded, such as with whom an individual tends to exchange e-mail, which Web sites are visited regularly, and even what level of sophistication the user appears to exhibit.

There are also biological factors that might be monitored, such as how often a user gets up to walk around, go to the washroom, or go out of the building for a smoke (activities which themselves could be monitored by physical access controls!).

In environments in which monitoring key strokes is not considered intrusive, some effort has been made to monitor key-stroke dynamics. This approach tends to be much less reliable in general, particularly with confronted with network and satellite delays. Also, if you are typing with one hand because you are drinking a cup of hot coffee with the other hand, your typing dynamics go all to hell.

In addition to providing a real-time database relating to physical whereabouts, and extending statistical profiling to accommodate subtle computer usage variants, it would also be appropriate to represent certain external information regarding personal behavior, such as intellectual and psychological attributes.

As an example of an intellectual attribute, consider writing styles. There are already a few tools for analyzing natural-language writing styles. Profiles of individual-specific "msipelings", the frequency of obscenities and the choice of explicit expletives, the relative use of obscure words, and measures of obfuscational proclivities and Joycean meanderings might also be quite useful. (Recall Tom Lehrer's warning: Don't write naughty words on walls if you can't spell.)

Psychological factors do not seem to have been explored much in the past, especially in the context of insider misuse. Psychologists routinely observe certain standard behavioral characteristics and analyze deviations therefrom. Some of those characteristics that are particularly relevant to potential insider misuse might be modeled in extended user profiles. As one specific example, we might be able to develop measures of relative paranoia, based on how often a particular user invoked certain commands to observe who else might be observing what that user was doing in real time, or the use of aliases in posting to newsgroups. A measure of aggressive behavior could be interesting, but would probably require some human reporting of perceived relative hostility levels in e-mail messages received from a given individual. Measures of anger and stress levels in general computer usage could also be conceived. However, considerably more effort is needed to characterize which psychological attributes might be effectively utilized. However, in my opinion, this is not likely to have much success, because there are not well established characteristics, and human variabilities are likely to confound them anyway.

If this approach is considered possibly fruitful, we should approach some psychologists who are familiar with computer users and ask them to speculate on psychological factors that might be both computer detectable and behaviorally discriminative with respect to insider misuse. On the other hand, users tend to be not particularly inherently risk-aware. However, see a *CACM* Inside Risks column by Dr. Leonard Zegans [22], which observes that, with respect to computer technology, users tend to take risks unconsciously and in many cases unwillingly.

7.6 Responses to Detected Insider Misuse

Responses must be tailored to the detected and interpreted misuses, including recommendations for further real-time analysis, human investigation, immediate reactions such as reconfiguration, and intelligent responses based on additional derived knowledge. The basic response framework in the desired misuse-detection systems should be directly applicable to responding to insider misuse, although different types of responses may be desirable – such as triggering a more detailed monitoring mode or invoking human observation and possible intervention.

7.7 Important Observations

A basic gap exists in computer systems and networks between what kinds of system uses are intended to be permitted and what uses are actually thought to be implemented as permissible. In addition to that gap, there is a further gap between what is thought to be permissible and what is actually possible. For example, flaws in the implementation of system security tend to contribute to the latter gap. Desired prevention of insider misuse suggests that better system security through differential access controls is necessary as one part of the solution. However, that may impose administrative problems.

Differential access controls might include fine-grained access control lists, fine-grained roles and fine-grained role-based access controls, compartmentalized protection, attribute-based encryption, and so on. In addition, better user authentication could not only prevent intruders from gaining insider access, but could also provide positive identification of insiders that might diminish their ability to masquerade as other insiders and to otherwise hide their identities.

Ignoring the importance of authentication and constructive access controls is putting the cart before the horse. After all, what does unauthorized use mean when everything is authorized? Recall the Internet Worm of 1988. Robert Morris was accused of exceeding authority; yet, no authority was required to use the `sendmail debug` option, the `finger` daemon buffer overflow, the `.rhost` mechanism, and the copying of encrypted but unprotected password files.

In the absence of an explicit security policy on what access is supposed to be permitted, it is difficult to ascertain what constitutes misuse. A far-reaching example of the impact of the difficulties thus presented is given by the PC virus detection problem, which would be a nonproblem if the PC software had any meaningful inherent security.

Despite the obvious truths that arise from the gross inadequacy of many existing systems, and precisely because of the risks that are thereby created, we focus here on detection and response to insider misuse, assuming that someone else has defined what is meant by insider misuse.

8 Requirements for High-Integrity Elections

The general problem of dramatically increasing the integrity of election processes throughout the world is in some sense a paradigmatic hard problem that encompasses a nice diversity of requirements addressing system integrity, data integrity, data confidentiality, system survivability, accessibility, and other issues. This problem also nicely illustrates the notion that the definition of an "insider" is hierarchical, distributed, and highly context dependent.

These requirements appear in various guises across the entire spectrum of operations – which includes all of the following stages. They apply to some extent irrespective of whether computer technology is used to prepare ballots, to tabulate ballots, or just to compile results. The early stages may be computer aided, but are also heavily based on human processes.

- Registration. Prospective voters should be fairly vetted for eligibility. Databases of eligible registrants must be correct, nontamperable, and carefully audited for misuse. Discriminatory procedures should be illegal and offenses should be penalized.

- Authentication. Verification of voter registration should be nonpartisan. Disputed challenges should result in provisional ballots being cast and should be fairly considered. Overzealous disenfranchisement (as in the Choicepoint list of supposed convicted felons in 2000 in Florida) should not be permitted.

- Authorization. Voters should be given correct ballots for the correct precinct, especially in multiprecinct polling places.

- Voter information. Voters should be officially notified in a manner that is clearly differentiated from bogus information such as last-minute phone calls informing voters that their polling places have been changed.

- Polling place availability. Adequate voting machines, paper ballots, provisional ballots, and so on must be available equanmiously for all polling places. Early voting is advisable. Absentee voting should be on the basis of voter convenience, not based on draconian rules.

- Polling place accessibility. Disabled and disadvantaged voters should be allowed to vote conveniently and without time pressures.
- Vote casting. Erroneous and poorly designed human interfaces should be avoided. The voting process should be voter-friendly. Votes should be correctly recording, and that process should be demonstrably verifiable.
- Vote counting. The tabulation of votes should be demonstably correct, reproducible through independent cross-checking, and demonstrably not subject to accidental errors, manipulation, and tampering.
- Monitoring. Audit trails should be sufficiently comprehensive to allow for detection of errors, manipulation, and tampering.
- Remediation following detected irregularities. Meaningful definitive recounts should be possible, with mandated do-overs whenever sufficient evidence of malfeasance warrants.

An obvious conclusion from this itemization is that the insider risks are ubiquitous and pervasive. Every step in the election process is a potential weak link that can be easily exploited or accidentally exercised – especially in the absence of thorough monitoring and audit trails.

9 Relevance of the Countermeasures to Elections

Picking up on the paradigmatic nature of the election problems, it is interesting to consider how the countermeasures suggested above might apply to elections.

Immediately obvious in that context are the following principles, attributable to Saltzer and Schroeder [21], or to extensions thereof:

- Economy of mechanism. Having to trust an enormous operating system in its entirety is clearly a bad idea. In contrast, Ka-Ping Yee's PhD thesis shows how the amount of software that must be trusted can be reduced to something like 500 lines of Python.
- Fail-safe defaults. Access controls should generally permit access explictly with a default of no access, rather than a default of total access unless explicitly denied. This simple principle can considerably reduce the likelihood of insider misuse, especially when combined with the other principles.
- Complete mediation. Whatever security controls are in place, they should not be bypassable. This principle also simplifies auditing.
- Open design. Reliance on the secrecy of a design (and the nondisclosure of proprietary source code) is generally an inherently bad idea. And yet it pervades the general marketplace for electronic election systems. The commercial vendors in the U.S. manage to keep their software proprietary, the internal election data proprietary, and the evalutions proprietary (and paid for by the vendors), and steadfastly resist attempts for scrutiny.
- Separation of privileges. In general, vendors should not have to share privileges with election officials, and vice versa.

- Least privilege. Given the separation of privileges, only the necessary privileges should be allocated. For example, vendors should not be permitted to alter certified code (unless there is a complete audit record of what they have done and why, authorized only under special circumstances). Election officials should not be permitted to make any alterations to system software or election data. Before and after elections, they may have to perform some carefully controlled and audited reconfigurations – such as initializing an election, or closing it down. Neither vendors nor election officials should be able to cast or alter any votes when serving in the roles of vendors or election officials. Test programs and test results should have absolutely no effects on the live election results. And so on.

- Least common mechanism. Building a voting system on top of a widely used operating system whose vulnerabilities are widely known does not seem particularly wise. This greatly increases opportunities for insider misuse.

- Ease of use. Given systems that are not easily used, not easily maintained systems, and difficult to configure correctly with respect to security controls (or the lack thereof), several consequences are common, such as outsourcing to system vendors certain responsibilities such as ballot-face preparation, voter registration pruning (such as the removal of supposedly convicted felons based merely on a name match), and oversight of the integrity and ultimately the authority of elections.

- Pervasive auditing. Stringent auditing within trustworthy system could enable meaningful recounts in cases of disputes (a facility that is almost nonexistent at present) and further provide forensic-quality evidence in cases of tampering or accidental changes. Such auditing should achieve high integrity, completeness, nonalterability, and nonsubvertibility. (This is a generalization and strengthening of what is termed "compromise recording" by Saltzer and Schroeder.)

- Other principles relevant to limiting insider misuse are also considered in [13]. For example, the principle of separation of policy and mechanism suggests that policy that may need to be altered should not be embedded in mechanism. A rather astounding example of the violation of that principle is found in one vendor's election systems that are reprogrammed *after the system is certified* in order to support many different ballot faces. In the case of Diebold, in every one of 17 counties in a California election in November 2003, the systems that were in use were versions other than the certified systems. Such violations of what is a common principle in software engineering represent one of the most dangerous examples of opportunities for insider misuse in existing voting machines.

Overall, election officials today have only superficial control over the entire lifecycle including all operations. The complexity of some of the all-electronic systems is such that the major vendors tend to provide their own personnel to help with setting up ballot faces and addressing technical problems that occur before, during, and even after elections. As a consequence, in the existing commercial systems the system developers and vendors have considerable latitude in making surreptitious system changes that could alter the results of elections. For example, in many cases in the past, the software that was actually used in elections was not identical to the software that had been certified. The absence of meaningful audit trails further complicates the process, because of the lack of demonstrable provenance and an almost complete lack of records on the alteration of code and election data.

10 Desired Research and Development Directions

What needs to be done that is not already being done? We consider general research and development directions (Section 10.1), as well as specific short-term (Section 10.2) and long-term (Section 10.3) directions. Because of its fundamental importance in the future, correlation is treated by itself (Section 10.4).

10.1 General Research and Development Directions

Overall, there is still much research and development work to be done in detecting and responding to insider misuse.

- We must recognize the commonalities between insider misuse and outsider misuse, with respect to threats, methods, exploitations, detection techniques, and response approaches, and take advantage of those commonalities where possible, resorting to different but compatible approaches where commonality is not immediately evident – all within the context of developing trustworthy systems that can address insider misuse as well as other critical requirements.
- Significant effort must be devoted to defining characteristic types of insider misuse. (For example, see [10, 11].)
- Finer-grain access policies and differential access controls are needed to help define what constitutes proper usage, thus facilitating the role of insider-misuse detection.
- Much greater effort needs to be devoted to detecting unknown modes of misuse, rather than just focusing so heavily on detecting known attacks. The existing statistical paradigms must be pursued and refined. However, new paradigms must also be considered.
- The community at large needs to address hierarchical and distributed correlation of results aggregated across different sensors, different application platforms, and different analytic tools. The correlation must seek to identify common patterns and intent, such as those resulting from coordinated distributed attacks.
- Anomaly and misuse detection must be integrated with network management in a trustworthy bidirectional manner.
- Much better software engineering is needed to make the analysis systems interoperable, robust, evolvable, and extensible in their application domains to monitoring other attributes such as reliability, fault-tolerance thresholds, survivability, performance, etc. (See [17].)
- Anomaly and misuse detection platforms must themselves be
  tamperproofed to hinder integrity attacks on the platforms, alterations of evidence (either by malfeasors to cover their tracks, or by law enforcement in attempting to fake evidence)
  spoofproofed to hinder bogus denial-of-service attacks on the platforms
  robustified to provide stability of the analysis platforms.
- Investigation of extrinsic individual characteristics such as psychological behavior that might be included in profiling user activities.

10.2 Short-term Research and Development Directions

Each of the above general topics has short-term implications.

Immediate payoff can be gained from identifying and carefully defining certain high-risk types of insider misuse, and extending existing signature-based systems to detect misuse of those types.

Some short-term progress can also result in detecting anomalous insider misuse, although some human analysis will be necessary in many cases to determine the significance of those anomalies. That effort should be transitioned into long-term research in which automated analysis is effective, as noted below.

Early experiments should be conducted to demonstrate the feasibility of wider-scope correlation across different detection capabilities and different instances of misuse and anomaly detection systems.

Some short-term efforts should be conducted in monitoring the status of network management tools, and experiment with detecting network anomalies – perhaps before they become serious problems. Initial efforts could also provide some simple automated analyses of recommended actions, such as system reconfiguration.

Some effort should be devoted to the short-term robustification of existing analysis platforms, at least to apply insider misuse detection to those platforms themselves, and to a few efforts that might make those platforms more resistant to tampering. Clearly, the longer-term efforts are important (see below), but some short-term activities could be very beneficial.

The Dagstuhl insider misuse workshop could be an important contributor to exploring the idiosyncrasies of insider misuse and to elaborating on these and other research directions.

10.3 Long-term Research and Development Directions

Each of the above general topics has long-term implications, and indeed many of the above short-term topics can be transitioned into longer-term efforts. Indeed, it is important that some of the short-term R&D efforts be conceived as efforts that can evolve gracefully into longer-term efforts. Although the U.S. Government funding agencies are not normally good at such longer-term planning, such evolvability can to some extent be made a byproduct of the way in which the R&D is carried out – for example, if good software engineering and evolvable architectures are requirements of the short-term efforts. This should be made a really important evaluation criterion for proposals for new work.

Applying anomaly detection to insider misuse has long-term implications, because it requires significant improvements in inference and hierarchical correlation.

Hierarchical and distributed correlation also has long-term needs for research, including new techniques for analysis.

In the long term, close integration of misuse analysis systems with network management monitoring platforms is highly desirable. Robust, stable, secure, reliable control of network reconfiguration as a result of detected anomalies and the resulting analysis is an important long-term goal. Although it might seem somewhat tangential to insider misuse, networking issues could otherwise be easily misused by insiders.

Developing robust platforms for anomaly and misuse detection for insider threats is a problem that is exacerbated by the presence of insider misuse in the first place. Insiders are inherently more likely to be able to compromise the analysis platforms than outsiders. Thus, robustification becomes all the more essential in the long run.

10.4 Hierarchical and Distributed Correlation and Other Reasoning

In his oral presentation of [17], Phil Porras outlined a collection of research directions related to correlation that are relevant here:

- Equivalence recognition, including various approaches to detecting significant security-related events
- Recognition of interrelated vulnerabilities and coordinated attacks
- Recognition of commonality among different distributed or local alerts
- Recognition of sequential trends

- Recognition of the intensity of misuses
- Coordinated aggregate analyses and aggregate scoring

It is important to have abstract representations of the results so that aggregation can be done effectively across heterogeneous target environments and heterogeneous analysis platforms (e.g., different Unix variants, different Microsoft systems, etc.).

In concept, the EMERALD resolver represented another type of analysis engine with the capability of mediating the results of rule-based, profile-based, Baysian, and other components, as well as analyses at different layers of abstraction.

11 Conclusions

Detecting insider misuse is typically more context dependent than trying to detect intrusions. It generally involves specifying what analysis information might need to be gathered, what rules might be given to an expert system, what parameters might be used to tune statistical analyses, what priorities might be associated with different modes of misuse, and what urgency might be accorded to various responses. Some new inference tools might be useful, but they could also be developed generally enough to be applicable to outsider misuse as well.

One difference that is perhaps substantive involves the use of individual user profiles. The SRI/Trusted Information System Safeguard project demonstrated significant advantages of profiling systems and subsystems rather than profiling individual users. However, in a critical environment in which there is only a modest number of insiders and *no* outsiders, profiling individual users makes much greater sense. (This was the basis for two early SRI exploratory projects for detecting insider misuse: one for the CIA involving IBM SMF records, which began in 1983, and one for the FBI involving the classified Field Office Information Management System Adabas mainframe database system, which took place in the 1990s.)

The reality that intrusion-detection tools are not oriented toward *insider attacks, unknown forms of misuse, intelligent results interpretation, generality of target systems, and long-term evolvability* presents a significant reason to question the prevailing DoD assumption that commercial off-the-shelf (COTS) products are adequate. The reality is that some of the COTS operating systems and networking software is badly flawed and easily subvertible.

The fact that multilevel security is not adequately implemented in either target systems or in analysis and misuse platforms need not be a handicap in the detection of insider misuse. We know how to build multilevel-secure systems out of nonmultilevel-secure user systems and trustworthy servers (e.g., see John Rushby and Brian Randell [19] and, Norm Proctor and Peter Neumann [18], as well as ongoing current work on Multiple Independent Levels of Security (MILS). Unfortunately, commercial realities seem to preclude the widespread availability of such systems.

Carefully directed research is essential, such as is outlined in the previous section. Better system security is essential – particularly if it can be achieved in open-source environments.

Unfortunately, the U.S. DoD has strongly endorsed the use of proprietary commercial off-the-shelf (COTS) software, most particularly including some unsecure unreliable nonsurvivable operating systems and applications. But as I state here, COTS "intrusion-detection" systems are not really useful for detecting hitherto unrecognized insider misuse. Furthermore, I believe that proprietary monocultures are extremely dangerous in the long run. (This is clearly exhibited in existing voting machines, where states and counties may become totally dependent on a single vendor to provide not only the election systems, but also all of the technical support and supplies.

I believe that efforts to robustify open-source software could result in enormous payoffs for everone, especially with some strong support and guidance from NSA, DoD, DHS, and NIST; if nothing else, those efforts could serve to inspire (or jawbone?) COTS developers into producing better systems. If governments roll over and play dead the way parts of DoD have done thus far (with the Yorktown dead in the water, Cloverdales, Melissas, Website takeovers, and frequent hostile denial of service attacks such as BotNets), DoD may run itself into the ground. The existing COTS strategy is doomed as long as the most commonly used products are so weak from the perspective of security, reliability, and survivability – although I recognize that at the same time there are other advantages. Alternatives to monocultures must be sought, and can lead to massive improvements. Cyberdiversity is essential to the future.

Acknowledgement

This position paper was prepared with funding from National Science Foundation Grant Number 0524111, under the SRI project for ACCURATE: A Center for Correct, Usable, Reliable, Auditable and Transparent Elections.

References

[1] K.J. Biba. Integrity considerations for secure computer systems. Technical Report MTR 3153, The Mitre Corporation, Bedford, Massachusetts, June 1975. Also available from USAF Electronic Systems Division, Bedford, Massachusetts, as ESD-TR-76-372, April 1977.

[2] M. Bishop. Position: 'insider' is relative. In *Proceedings of the 2005 New Security Paradigms Workshop*, pages 77–78, Lake Arrowhead, California, October 2005.

[3] M. Bishop, S. Engle, C. Gates, S. Peisert, and S. Whalen. We have met the enemy and he is us. In *Proceedings of the 2008 New Security Paradigms Workshop*, Olympic Valley, California, 2008.

[4] F.J. Corbató. On building systems that will fail (1990 Turing Award Lecture, with a following interview by Karen Frenkel). *Communications of the ACM*, 34(9):72–90, September 1991.

[5] F.J. Corbató, J. Saltzer, and C.T. Clingen. Multics: The first seven years. In *Proceedings of the Spring Joint Computer Conference*, volume 40, Montvale, New Jersey, 1972. AFIPS Press.

[6] R.C. Daley and P.G. Neumann. A general-purpose file system for secondary storage. In *AFIPS Conference Proceedings, Fall Joint Computer Conference*, pages 213–229. Spartan Books, November 1965.

[7] V.D. Gligor et al. Design and implementation of Secure Xenix[TM]. In *Proceedings of the 2004 Symposium on Security and Privacy*, Oakland, California, April 1986. IEEE Computer Society. also in *IEEE Transactions on Software Engineering,* vol. SE-13, 2, February 1987, 208–221.

[8] IRC. Hard problem list. Technical report, INFOSEC Research Council, November 2005.

[9] P.A. Karger. Limiting the damage potential of discretionary Trojan horses. In *Proceedings of the 1987 Symposium on Security and Privacy*, pages 32–37, Oakland, California, April 1987. IEEE Computer Society.

[10] C.E. Landwehr, A.R. Bull, J.P. McDermott, and W.S. Choi. A taxonomy of computer program security flaws, with examples. Technical report, Center for Secure Information Technology, Information Technology Division, Naval Research Laboratory, Washington, D.C., November 1993.

[11] P.G. Neumann. *Computer-Related Risks.* ACM Press, New York, and Addison-Wesley, Reading, Massachusetts, 1995.

[12] P.G. Neumann. Practical architectures for survivable systems and networks. Technical report, Final Report, Phase Two, Project 1688, SRI International, Menlo Park, California, June 2000.

[13] P.G. Neumann. Principled assuredly trustworthy composable architectures. Technical report, Computer Science Laboratory, SRI International, Menlo Park, California, December 2004. http://www.csl.sri.com/neumann/chats4.html, .pdf, and .ps.

[14] P.G. Neumann. Reflections on system trustworthiness. In Marvin Zelkowitz, editor, *Advances in Computers, volume 70*, pages 269–310. Elsevier Inc., 2007.

[15] P.G. Neumann. Illustrative risks to the public in the use of computer systems and related technology, index to RISKS cases. Technical report, Computer Science Laboratory, SRI International, Menlo Park, California, 2008. Updated regularly at http://www.csl.sri.com/neumann/illustrative.html; also in .ps and .pdf form for printing in a denser format.

[16] P.G. Neumann, R.S. Boyer, R.J. Feiertag, K.N. Levitt, and L. Robinson. A Provably Secure Operating System: The system, its applications, and proofs. Technical report, Computer Science Laboratory, SRI International, Menlo Park, California, May 1980. 2nd edition, Report CSL-116.

[17] P.G. Neumann and P.A. Porras. Experience with EMERALD to date. In *Proceedings of the First USENIX Workshop on Intrusion Detection and Network Monitoring*, pages 73–80, Santa Clara, California, April 1999. USENIX. Best paper.

[18] N.E. Proctor and P.G. Neumann. Architectural implications of covert channels. In *Proceedings of the Fifteenth National Computer Security Conference*, pages 28–43, Baltimore, Maryland, 13–16 October 1992.

[19] J.M. Rushby and B. Randell. A distributed secure system (extended abstract). In *Proceedings of the 1983 IEEE Symposium on Security and Privacy*, pages 127–135, Oakland, California, April 1983. IEEE Computer Society.

[20] J.H. Saltzer. Protection and the control of information sharing in Multics. *Communications of the ACM*, 17(7):388–402, July 1974.

[21] J.H. Saltzer and M.D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, September 1975.

[22] L.S. Zegans. The psychology of risks. *Communications of the ACM*, 51(1):152, January 2008. *Inside Risks* column.