INFOSEC Research Council (IRC)

HARD PROBLEM LIST

November 2005

INFOSEC Research Council

Hard Problem List

The IRC Hard Problem List is Unclassified – Distribution is Unlimited

Acknowledgments

The INFOSEC Research Council extends its deepest thanks to the speakers, panelists, authors, contributors, reviewers, and technical editors whose time and work made essential contributions to the quality of this report. Their efforts are deeply appreciated. Special thanks are extended to the following individuals who contributed their time and talents to review and offered suggestions for improvements to this document.

Steve Kent
James R Gosler
Marc Donner
Steve Bellovin
Joan Feigenbaum
Fred Schneider
Peter G. Neumann

Although these reviewers were not asked to endorse the conclusions or recommendations of the report, nor did they see the final draft before its release, they provided many constructive comments and valuable suggestions.

Many organizations have representatives as regular members of the INFOSEC Research Council:

Advanced Research and Development Activity, Central Intelligence Agency, Department of Defense (including the Air Force, Army, Defense Advanced Research Projects Agency, National Reconnaissance Office, National Security Agency, Navy, and Office of the Secretary of Defense), Department of Energy, Department of Homeland Security, Federal Aviation Administration, National Aeronautics and Space Administration, National Institutes of Health, National Institute of Standards and Technology, National Science Foundation, and the Technical Support Working Group.

In addition, the IRC is regularly attended by partner organizations from Canada and the United Kingdom.

Table of Contents

Executive Summary

Members of the INFOSEC Research Council (IRC) are the major sponsors of information security research within the U.S. Government. The *Hard Problem List* defines desirable research topics by identifying a set of key problems from the Government perspective and in the context of IRC member missions. Solutions to these problems would remove major barriers to effective information security (INFOSEC). The Hard Problem List is intended to help guide the research program planning of the IRC member organizations. It is also hoped that nonmember organizations and industrial partners will consider these problems in the development of their research programs. Policy makers and planners may find this document useful in evaluating the contributions of ongoing and proposed INFOSEC research programs.

The original Hard Problem List, which was composed in 1997 and published in draft form in 1999, is included as Appendix A: *Retrospective on the Original Hard Problem List.* The original list has proven useful in guiding INFOSEC research. However, the significant evolution in technology and threats over the past several years requires an update to the list; therefore, the current Hard Problem List (HPL) was created. Additional updates to the HPL may be provided as warranted at the discretion of the IRC.

Looking forward over the next five to ten years, the INFOSEC technical hard problems included in the current Hard Problem List are:

1. **Global-Scale Identity Management**: Global-scale identification, authentication, access control, authorization, and management of identities and identity information

2. **Insider Threat**: Mitigation of insider threats in cyber space to an extent comparable to that of mitigation in physical space

3. **Availability of Time-Critical Systems**: Guaranteed availability of information and information services, even in resource-limited, geospatially distributed, on demand (ad hoc) environments

4. **Building Scalable Secure Systems**: Design, construction, verification, and validation of system components and systems ranging from crucial embedded devices to systems composing millions of lines of code

5. **Situational Understanding and Attack Attribution**: Reliable understanding of the status of information systems, including information concerning possible attacks, who or what is responsible for the attack, the extent of the attack, and recommended responses

6. **Information Provenance**: Ability to track the pedigree of information in very large systems that process petabytes of information

7. **Security with Privacy**: Technical means for improving information security without sacrificing privacy

8. **Enterprise-Level Security Metrics**: Ability to effectively measure the security of large systems with hundreds to millions of users

These eight problems were selected as the hardest and most critical challenges that must be addressed by the INFOSEC research community if trustworthy systems envisioned by the U.S. Government are to be built. INFOSEC problems may be characterized as "hard" for several reasons. Some problems are hard because of the fundamental technical challenges of building secure systems, others because of the complexity of information technology (IT) system applications. Contributing to these problems are conflicting regulatory and policy goals, poor understanding of operational needs and user interfaces, rapid changes in technology, large heterogeneous environments (including mixes of legacy systems), and the presence of significant, asymmetric threats.[1]

Many of the hard problems have seen remarkable progress in research labs, but have not transitioned to commercial availability or to deployment in operational systems. The transition of improved security into mainstream use is further hampered by issues that keep such changes from being attractive to developers or users. Complete solutions to these hard problems must be complemented by solutions to the issues that inhibit the transition to deployment such as limited incentives and resources for transitioning research; hardware, software, and systems engineering practices to produce inherently more secure or securable information technology; education, and training; the economic forces that drive the market; and the perceived need for speed and reduced cost at the expense of quality and security. Likewise, concerns over the current policy, legal, and regulatory contexts associated with information security will not be addressed by solving these technical hard problems. However, problem resolution will result in systems that are safer, more reliable and resilient in the face of attack, and perhaps more affordable. Although these sociological, economic, financial, and legal issues are not addressed as part of this report, they will have an impact on the effectiveness and completeness of the final solution.

[1] Asymmetric threats are unanticipated or nontraditional approaches to circumvent or undermine an adversary's strengths while exploiting vulnerabilities through unexpected technologies or innovative means.

Introduction

Many changes have occurred since 1997, the year in which the first INFOSEC Hard Problem List study was initiated. Dependence on Information Technology (IT) has continued to increase. IT is now rapidly finding its way into crucial roles in commercial, civil, and military applications of wired and wireless computing. IT and the Internet are now critical components of our national economy and the nation's infrastructure sectors. As IT continues to spread through society, the average level of user IT knowledge will continue to decline, creating fertile ground for hackers. As the sophistication of technology has increased, the sophistication of attack tools has also risen, and the time from vulnerability discovery to exploitation continues to shrink. Terrorism is now a real concern, and privacy maintenance is more prominent in the minds of citizens. In an effort to stem the tide of unwanted exploitation and cyber terrorism, leaders in the IT industry have begun to increase their already major investments in addressing information security; commercial purchases of security technology have grown into a multibillion-dollar business. Moreover, technological developments have precipitated changes in legislative and criminal environments. This revision to the Hard Problem List (HPL) is made in the context of these changes, and reflects five years of additional research and experience since creation of the original list.

Based on technology changes, research progress, and government needs that have become known to members of the IRC, the IRC convened a study panel to develop this report. The panel condensed information gathered from IRC members and from experts from academia, industry, and government, who described crucial hard problems from many divergent viewpoints. This process resulted in the current HPL, which is intended for people within the research community, organizational management, industrial partners, and members of congressional staffs responsible for reviewing and establishing research initiatives, as well as other interested parties.

This Hard Problem List is a complete revision of the original HPL. The purpose of the revision is to identify critical new problems that have emerged, identify hard problems that have remained critical despite dramatic technological and sociological changes, and apply several years of successes and struggles in research to refine our understanding of those problems. As before, the HPL identifies problems that are vital to building the information systems envisioned by departments and agencies of the U.S. Government and to trusting those systems to function as needed, even in the face of attacks by aggressors of all types. Given the level of investment in solving such INFOSEC problems via research, the HPL is limited to those unsolved problems that are unlikely to be solved in the next five to ten years without aggressive concerted research. For this reason, the HPL excludes problems expected to be solved by existing or emerging government systems, or by commercial systems already in production. New reports will be issued periodically. However, the INFOSEC Research Council will not commit to an update schedule or content of any revision, because of the administrative infrastructure required to produce such a document on a regular basis.

Identifying the hardest and most critical challenges facing INFOSEC research requires an understanding of the existing and proposed systems used by the U.S. Government and the assets that must be secured. Although specific needs vary between organizations within the U.S. Government, most agencies share a common vision of highly versatile and integrated seamless communications systems. These systems will stretch not only from coast to coast, but around the world (and in some cases even into space), and will provide interactive access to real-time data. Frequently, this access will require the strongest security possible. U.S. Government systems provide both wired and wireless access for warfighting and emergency response. In addition, access is used by humanitarian aid personnel at home and abroad, and with foreign partners in coalition and other inter-government affairs that require secure digital interactions. Networking systems include the Internet and the public switched telephone network, and all other domain specific networks, including sensor nets, on-demand networks, and military networks. Consistent with an emerging U.S. Government vision for networking research, this report uses the term "network" to include traditional networking layers and topological boundaries, in addition to end-to-end service, and application components of the network.

Systems must be secure, flexible, and rapidly adaptable in the face of changing topologies, dynamic adversaries, time-varying wireless channel conditions, mobility of communication nodes, and changing mission requirements. The definition of *system* takes into account the role of people in the system and the degree of harm they intentionally or accidentally can inflict on the system. Systems not only must meet the needs of people, but also must provide an environment that is conducive to the needs of sensors, robots, and autonomous platforms. Networks must not only support traditional data needs, but must be capable of swiftly moving high-quality still images (such as maps and satellite photos) or video with higher bandwidth needs and stronger quality of service requirements for military, medical, and homeland security personnel, and for most other government agencies. The systems required to meet these needs will be large in scale, involving millions and perhaps billions of users and devices, and interoperability will be crucial. These systems must be able to achieve the goals described above even while under attack, and when bandwidth, processing, memory, and energy are constrained. While many U.S. Government organizations have more modest needs, U.S. Government research programs must provide for common requirements and for the additional challenges of those organizations with the most critical needs.

Hard Problems in INFOSEC Research

Over the last several years, the original HPL has been influential in focusing research strategies within and among various organizations. Due consideration of the list's effectiveness and success requires a brief discussion of how the original Hard Problem List has changed over the last several years. Begun in 1997 and released in 1999, the original HPL was divided into functional problems and problems associated with design and development. Each of the 1999 problems is briefly described in Appendix A. More complete descriptions may be found at the IRC Web site http://www.infosec-research.org.

The current list identifies eight problems as the hardest and most critical challenges in INFOSEC research that must be addressed for the development and deployment of trustworthy systems for the U.S. Government. The following are included for each problem: problem definition, the impact of solving or failing to solve the problem, a threat description, a description of specific

challenges making the problem hard, sample approaches for handling the problem, and metrics for measuring progress against the problem. The threats span both intentional and accidental misuse, as well as environmental hazards. Each subsection on approaches includes one or more of the limitations of current approaches, promising approaches not yet fully explored and desirable approaches where relatively few ideas have been proposed. However, in no case should any list of possible approaches be taken as complete or exhaustive, or used to constrain the search for solutions. In many cases, the ideas offered here are only a few of the many possible starting points currently known, and far more innovative strategies are likely to be required to achieve success.

The remainder of this document contains a description of the Current Hard Problems in INFOSEC Research. Each section provides the definition of the problem, threats that make it a hard problem, motivations that caused selection of this item as a problem, challenges that may be faced in creating a solution, and approaches to possible solutions. In addition, an associated appendix for each problem provides more information, along with metrics to be used to determine success. The appendices are:

To avoid showing preference for any person's work, this report discusses problems and potential solutions in general terms without providing specific citations to prior or ongoing research.

Current Hard Problems in INFOSEC Research

The eight problems selected as the hardest and most critical challenges in INFOSEC research that must be addressed for the development and deployment of trustworthy systems for the U.S. Government are presented in the following sections.

1. Global-Scale Identity Management

Definition: Global-scale identity management is the problem of identifying and authenticating people, hardware devices, and software applications when accessing critical and sensitive Information Technology (IT) systems. The term *global scale* implies that identification and authentication (I&A) approaches should be capable of supporting all possible users of these systems around the world. In addition, the I&A approaches should be interoperable across U.S., state, and local governments' IT systems, and with systems of foreign governments and non-government institutions as needed. Eventually, global-scale identity management may require not only advances in technology, but also open standards and policies for the creation, use, and maintenance of identities and privilege information (e.g., rights or authorizations), particularly given complex issues of privacy and anonymity. The question of when identifying information must be provided is fundamentally a policy question and therefore beyond the scope of this study. Nevertheless, countless critical systems and services require authenticated authorization for access and use, and the technology needed to support these authorizations.

Threat: Identity-related threats are surprisingly ubiquitous. Although technological approaches to combating those threats are computer related, many threats are external to computer systems. Internal threats include outsiders masquerading as other users, bypassing authentication, subverting authorization under false identities, and setting up unauthorized intermediaries to enable further security problems (e.g., man-in-the-middle attacks). External threats typically involve social engineering, impersonation, or coercion, and may result in identity theft and other forms of privacy violations as well as misuse of resources.

Motivation: Controlling access to an IT system and its resources is critical to having a secure and reliable system. Strong I&A mechanisms support this need by enabling authentication of individuals, networks, and system components and applications within and among various IT systems. Strong I&A mechanisms also mitigate insider threats and support the detection of system misuse. When strong I&A mechanisms are implemented, the management and distribution of electronic security policies can be securely automated, greatly improving the flexibility and responsiveness of systems to changing conditions. Also, with strong I&A mechanisms, automated network inventory and system configuration control management can reduce life-cycle costs and the total costs of system ownership. Perhaps more important, strong I&A mechanisms are part of the required controls to support the ability to securely share sensitive information among U.S., state, and local authorities, and with other governments.

As an example, consider a user and workstation logon scenario. When a workstation is powered on, the workstation could communicate with an edge network switch that would not allow the workstation to connect to the network until the user was authenticated. The user might use an

identity token and biometrics to authenticate to the workstation, and the use of the network and workstation could then be governed by a particular set of privileges assigned to the user.

Challenges: This is a hard problem for a number of reasons, including standardization, scale, churn, time criticality, and the prospect of quantum computing. The risks from insider threat should also be considered if people are required to administer critical or broad aspects of any identity-management system.

Achieving the goal of open, globally accepted standards for identifying individuals, system components, and processes is difficult and will take considerable coordination and cooperation between industry and governments. There is also the requirement to maintain the anonymity of personal information unless explicitly required. In addition, determining how system processes or threads should be identified and privileged is an even more complex and daunting undertaking. Finally, while sensor networks and radio frequency identification (RFID) have tremendous utility for government use, their current vulnerabilities and the desired scale of future deployment underscore the need to address the hard challenges of identity management on a global scale.

From an operational perspective, the life-cycle management of identity information and related tokens is extremely complex and costly, particularly given the scale of information systems today and tomorrow. In an enterprise as large as the U.S. Government, individuals constantly change positions, move to other organizations, leave, return as contractors, and retire. In addition, system components are modified, retired, and replaced by newer and faster components. This continual 'churn' makes it more difficult to manage identities and privileges associated with individuals and system components. Privileges and permissions also change, particularly when individuals and system components are compromised. Unfortunately, with changes in people, components, and permissions across geographically distributed and, in some cases, intermittently connected networks, I&A updates may take considerable time. Delays could cause mistakes, allowing unauthorized access to sensitive information or unauthorized control of sensitive systems. In a world of almost instantaneous computer attacks, such a position is not tenable.

These challenges are compounded by the existence of multiple authorities that assign identities and privileges. These authorities may represent different departments of the U.S. Government, different governments (state, local, or foreign), or even non-government groups (e.g., private industry). Thus, some sort of federated or cross-recognition of authorities must be supported for both identification and authentication.

In addition, identity management system(s) and any supporting infrastructure for managing cryptographic keys will be attractive targets for adversaries. Thus, it will be necessary to build these systems with a high degree of assurance that is currently difficult and costly to attain and maintain. Finally, although most systems emerging for large-scale I&A involve public-key cryptography, quantum computing may eventually render unsafe those systems based on factoring of very large prime products. (Elliptic-curve cryptography may prove to be advantageous in this and other respects in the long term.) While there is no immediate concern, it makes the longer-term solution to the problem harder to envision.

Approaches: Currently, there are several major initiatives involving large-scale identity management, including a government-wide e-Authentication initiative, the Defense Department's Common Access Card, and Public Key Infrastructure for the Global Information Grid. However, none of these scale to the levels required without substantial problems regarding federation of certification authorities and delays in handling revoked privileges. Moreover, they are all based on public-key cryptography and will therefore eventually be susceptible to attack by quantum computers. Research strategies to achieve a strong I&A architecture for the future include large-scale symmetric key infrastructures with key distribution *a priori*, federated systems of brokers to enable such a system to scale, strategies for scaling symmetric creation of one-time pads, schemes of cryptography not reliant upon a random oracle, and other schemes of cryptography not susceptible to attack by quantum computers, if possible. However, by no means should solutions be limited to these initial ideas.

Metrics: Ideally, any I&A system should be able to support millions of users with identity-based or role-based authentication. The system should be able to handle millions of privileges and a heavy churn rate of changes annually in users, devices, roles, and privileges. In addition, each user may have dozens of distinct identities across multiple organizations, with each identity having its own set of privileges. Revocation of privileges should be effective for near-real-time use. Last, it should be extremely difficult for any national-level adversary to spoof a mission-critical or critical infrastructure system into believing that anyone attempting access is anything other than the actual adversary or adversaries.

2. Insider Threat

Definition: According to the Computer Science and Telecommunications Board of the National Research Council, "A person who is allowed inside the security perimeter of a system and consequently has some privileges not granted outsiders is referred to as an *insider*. The insider threat is the threat that an insider maliciously, or even accidentally, abuses those privileges and compromises system security."[2] Insider threats originate from a variety of motivations (e.g., financial gain, personal grievances, revenge, recruitment, or coercion). Unlike unauthorized outsiders and insiders who must overcome security controls to access system resources, insiders have legitimate and (depending on their positions) minimally constrained access to computing resources. In addition, *trusted* insiders who design, maintain, or manage critical information systems are of particular concern because they possess the necessary skills and access to engage in serious abuse or harm. Typical trusted insiders are system administrators, system programmers, and security administrators, although ordinary users may sometimes acquire those privileges (sometimes as a result of design flaws and implementation bugs). Thus, there are different categories of insiders.

Threat: Insiders can quickly access, download, copy, corrupt, or remove large amounts of critical data, significantly damaging an organization's information assets. Moreover, they can do this in relative obscurity over potentially long periods of time. Perhaps worse, insiders can cause systems integral to operation of the power or transportation grids to fail dramatically by sabotaging the computers and software that control their operations. In combination with faulty identity management (Hard Problem 1), they may also be able to masquerade as other insiders. In combination with faulty system security (Hard Problems 4 and 5), they may be able to subvert the audit trails and effectively cover their tracks.

Motivation: A trusted insider who maliciously abuses his computer privileges is one of the greatest threats facing system owners because of the potentially catastrophic damage that can be inflicted. Some of the worst damage occurred from insider malice. As a senior agent in the U.S. Bureau of Investigation (FBI), Robert Hanssen had been clandestinely transferring highly classified documents to Russians for more than a decade before he was stopped. According to an FBI-sponsored survey in 2000, insiders were responsible for 71% of the unauthorized entries into U.S. corporate computer networks reported by the commercial sector. However, the extent of incidents related to insiders varies dramatically from one kind of system to another. In classified systems, essentially all risks arise from insiders; on the Internet, major risks are created by worldwide accessibility by outsiders. Some disgruntled insiders in system administrator positions have planted logic bombs, erased corporate databases, stolen proprietary information (e.g., the recent theft of an AOL subscriber list), and stolen large sums of money from banks.

[2] Some definitions for "insiders" include malicious hackers who have gained access to internal networks by obtaining legitimate credentials. This document does not include these individuals in the definition of insider. However, note that a Trojan horse can act as an insider, with all the privileges normally associated with the executing environment.

Challenges: Most tools and technologies intended to secure IT are designed to protect networks and systems from unauthorized access. In addition, many organizations employ various techniques to keep malicious insiders from compromising information systems. These techniques include background checks, requiring multiple individuals to perform security-relevant functions (i.e., two-party integrity compartmenting access), setting attractive lures to trap malicious insiders, and introducing traceable disinformation to determine the source of leaks. Unfortunately, no single description of an insider is applicable in all cases, and trusted insiders even have "inside knowledge" of these countermeasures and controls. Information systems programmers and operators control critical systems, operations, and confidential information. In addition, information systems provide a breadth of access across an organization and enable latent defects like unseen time bombs, and zero-day attacks. Some insiders may even be trained with outside help on how to counteract each of these countermeasures. Motivations, objectives, level of expertise, and degree of access vary from one incident to the next. The insider threat is a problem that can never be entirely removed. However, it might be possible to reduce the problem to an acceptable level of risk, for instance, by implementing countermeasures in cyber space that are comparable to the countermeasures in the physical space. Reducing the insider threat problem to this degree would greatly benefit national security.

Approaches: More advanced technologies, methods, and procedures are needed to better mitigate and counter malicious insider activities. Insider threat countermeasures must be resistant to compromise; effective and portable across a broad range of technologies; and capable of addressing changes to threat conditions, agency missions, and personnel access revisions. Finally, awareness and control of insider-threat countermeasures must be visible or available to only a limited number of individuals, in order to ensure their effectiveness.

A relatively straightforward approach would be to ensure the strong enforcement of fine-grained context-sensitive access controls, with all the attendant security administration necessary to manage the assignment and revocation of privileges. This is not quite as easy as it might sound, because it in turn relies on advances in system and network architecture (part of Hard Problem 4), situational awareness (Hard Problem 5), and identity management (Hard Problem 1).

Compartmentalization limits the extent of the damage that can be caused by a single insider. Unfortunately, compartments disrupt government business, and managing the compartmentalization can be extremely complex. Therefore, there is a need for compartmentalization schemes that do not adversely affect legitimate users in their legitimate roles and responsibilities. Better methods of policy specification and enforcement could result in compartment boundaries that provide more effective protection and are less likely to impede legitimate work. One way of improving policy specification and enforcement is through better digital rights management (DRM). DRM helps originators enforce control over information that others attempt to read, write, modify, print, copy, distribute, or destroy. A limitation of DRM is that the technology is not intended to protect against terrorists or hostile nation-states.

An alternative strategy involves monitoring all security-relevant actions. This will require technologies for pervasively auditing and tracking cyber transactions. In addition, the technologies must support real-time audit analyses and post-compromise forensic analyses without access by malicious insiders. Automated pattern matching and anomaly detection algorithms could supplement the audit and forensic analyses to detect and characterize insiders

whose observable behavior might be that of a malicious insider. Models could capture insiders' behaviors and other aspects such as motives, intentions, methods, skills, and risk tolerance.

Several other novel strategies might also be effective. For example, redundancy has been mentioned as a way to minimize dependency on a single individual. Implementing redundancy on a much larger scale will require a generalized scheme for a technological variant of the "two person rule." Autonomic and self-managing networks could dramatically reduce the need for system administrators with significant security-relevant responsibilities. In addition, new forms of cyber deception could go beyond current techniques such as lures, decoy networks, and honeypots, as a way to discover malicious insider activity.

Metrics: Measuring progress toward reducing the insider threat is hard, both in research and in practice. In research, several diverse reference data sets that represent "real world" cyber activity are needed to properly test prototype insider threat countermeasures against a variety of desired attributes. Some of the desired attributes of the countermeasures are assurance of low false positive and false negative rates and adequate coverage of the problem space. Countermeasures and reference data sets used to assess numerous approaches may result in progress against this exceptionally hard problem.

3. Availability of Time-Critical Systems

Definition: Frequently, availability of the information to individuals in a timely manner is more critical than secrecy of the information. For example, adversaries "reading the mail" of process control systems such as supervisory control and data acquisition (SCADA) systems and digital control systems are of less concern than adversaries who can disrupt the operation of those systems. The hard problem is assuring service availability of time-critical systems, particularly in adverse conditions such as when the systems are under attack.

Threat: The threats to time-critical systems are even more notable than those for conventional systems. System, network, and enterprise survivability critically depend not only on security, but also on reliability and fault tolerance, and the ability to recover sufficiently rapidly from outages and from losses or diminution of resources. The threats to time-critical systems thus include the entire gamut of security attacks from outsiders and insiders, and are pervasive. Recent, worms have disrupted 911 services in a few communities, financial networks that support Automated Teller Machines, and for several hours, the control system of a Nuclear Power plant. Worms can spread so quickly that human response is ineffective. Denial-of-Service attacks can often be mounted externally, with absolutely no user privileges required; worse yet, those attacks may be untraceable. Consideration of the threats to a particular time-critical system must encompass hardware failures, software faults, operational errors, and anything else on which availability may depend --- such as servers and networks, and people in the loop. Logical and physical attacks must both be considered. For those systems that support critical national infrastructures, the risks of not adequately addressing these threats can be enormous.

Motivation: Everyone is aware of our increasing dependence on the availability of information systems that for the most part have been designed prioritizing availability over security. The scope of public panic and outrage seen if major communication systems or if major parts of the national electrical power grid failed as a result of terrorist attacks would be unprecedented. There are many examples of power, telephone, and other communications outages resulting from natural disasters such as ice storms and hurricanes. The affects of these outages should be considered mild when compared to what a sophisticated adversary might accomplish by targeting vulnerable critical infrastructure information systems that support time-critical systems.

The nation needs an information infrastructure that can support the military, national emergency recovery operations, telemedicine, SCADA, and homeland security activities, and an infrastructure that remains available even while under attack. A disruption that causes a loss of service that results in loss of life while a system reboots or a distributed denial-of-service (DDOS) attack can be resolved is unacceptable.

Challenges: One of the biggest challenges associated with ensuring availability is addressing the full complement of components that need to be protected in a critical infrastructure information system. This includes end-to-end communication properties, processing power to meet computational deadlines, and timely access to mission-critical stored data.

Currently, critical infrastructure information systems carry and process a variety of information, for example, voice, video, transactional data, and broadcasts. These various types of information systems all have different functional requirements in terms of latency, jitter, and throughput, as do SCADA systems and process control systems. Historically, all this information was carried on different networks. Now, they utilize a single, converged network, which needs to guarantee adequate service for mission-critical needs and the application-specific needs of latency and throughput to fulfill those mission requirements. Availability requirements will not be met if only *a few* bits trickle through the system, yet today's technology finds it difficult to manage such qualities of service in wired networks, much less wireless networks. Moreover, demand for time-critical processing is growing rapidly in environments that use robotics, RFID, and sensor networks for real-time data collection.

The distributed nature of modern systems complicates efforts to ensure timely availability. Availability must be ensured on a host-by-host and resource-by-resource basis, as well as an end-to-end basis. Paradoxically, some attacks are enabled by the extra capacity built into the Internet itself. In DDOS attacks earlier this decade, a few people usurped control of thousands of machines and directed the traffic against a small number of victims, flooding them and denying access to millions of their users. Through government-funded research begun in the late 1990s, commercial operations now offer *limited* protection against these attacks for wired networks. However, general solutions for wired and ad hoc wireless networks are not yet available.

The source of the disruption of a wireless communications system is often difficult to identify. Disruption may be due to jamming, saturation, natural changes in paths, or changing characteristics of paths. Each situation may require a different response. Fixed communication lines are not an option for coordinating diagnostics and responses in networks that are entirely mobile, with no fixed base stations. Battlefield and emergency response networks have limited bandwidth and most likely no redundant communication paths, which can make them easier to disrupt.

Approaches: Survivable IT includes survivable computing and survivable communications. Limited-case solutions for both include, for example, fixed networks, overbuilding capacity, fixed provisioning of bandwidth, and diversification of platforms and protocols. Unfortunately, even information systems with these defenses frequently have single points of failure. Furthermore, these strategies are not effective in distributed, wireless, and mobile ad hoc environments that are critical to military and emergency response events. Even when spare resources are available in such environments, survivability techniques cannot yet guarantee time-critical delivery. Survivable computing must be available in these harsh environments to effectively protect mobile, ad hoc, and distributed wireless systems while they are under attack.

Ensuring guaranteed service for ad hoc networks represents a significant challenge requiring considerable research at all layers of the protocol stack. Some of the first steps toward robust caching schemes and adaptive bandwidth allocation have been taken to help mitigate DDOS attacks. There is significant work remaining to provide the end-to-end guarantee of availability for prioritized service. Providing circuit-level assurance of communications with the flexibility of packet-switched Internet technologies will require the research community to rethink some of the cornerstones of the Internet, such as how to achieve routing and establish transmission layer

connections. The current techniques do not enable mathematically or scientifically provable guarantees against denial-of-service attacks on Internet communications.

Metrics: Success should be measured by the range of environments over which the system is capable of delivering adequate service for top-priority tasks. These environments will vary by topology and spatial distribution, as well as number, type, and location of compromised machines, and a broad range of disruption strategies. Previous operations and exercises have assisted in identifying spatial distributions of interest; however, significant advances in availability attacks may require research in new areas.

4. Building Scalable Secure Systems

Definition: Over the last decade, significant computer security investments have attempted to create the highest assurance possible with predominantly commercial-off-the-shelf (COTS) components. Despite some progress, there are severe limits to this strategy. To ensure security, high assurance systems should be built from the ground up. They must have a fundamentally sound architecture, and be developed with well-established principles. Unfortunately, current methodologies used to create new systems with high assurance capabilities do not scale to the size of systems used today or envisioned for tomorrow.

Threat: Threats to the development process are both functional and procedural. Many failed developments have resulted from inadequate requirements, weak architectures, poor software development processes and inadequate support tools, plus copious human inadequacies. Ultimately, these systems have fallen apart without any external provocation. On the other hand, the development process must include extensive awareness of the threats from insider and outsider misuse, hardware malfunctions, and environmental disruptions. It must also concern itself with anticipated operational practice and human interface design, which if not properly considered represent further threats to the success of the development. The requirements for scalability must themselves be an important driving factor in the development. The acquisition process and trustworthy code distribution represent additional sources of threats.

Motivation: Imagine hackers or terrorists taking control of hundreds of passenger airplanes. Many systems require the most trustworthy components, including avionics, various emergency responses, linchpin financial services, and power production. Historically, many systems expected a secure computing base to provide a trustworthy foundation for such computing. However, the costs of full verification and validation (V&V) have prohibited any secure computing base from having the requisite assurance and functionality. This is particularly applicable given the scale and complexity often required to meet functionality needs. In addition, the inordinate length of the evaluation process has been incommensurate with the ongoing need for further patches and system upgrades. This has retarded the incorporation of new high assurance information technology. Furthermore, legacy constraints on COTS software, networking support, and serious interoperability constraints have also retarded progress. There are also marketplace issues limiting the extent of meaningful security in the mainstream.

Challenges: Designing secure systems from the ground up represents an exceptionally hard problem, particularly since systems of such size and complexity can obscure catastrophic flaws in their design and implementation. Catastrophic flaws in software may occur even in a relatively few lines of life-critical code, significantly less than the tens of millions of lines of code in today's systems. Given the minuscule size of catastrophic bugs, and the size of modern systems, scaling up more formal approaches to production and verification of bug-free systems seems more promising than many less formal approaches attempted to date. Better tools are needed for incorporating assurance in the development process and for automating formal V&V. These tools may provide the functionality to build a secure computing base to meet U.S. Government needs for assurance and functionality.

A huge challenge involves achieving a highly principled system development process that is based on detailed and farsighted requirements, sound architectures that can be predictably composed out of demonstrably trustworthy components and subsystems, and then subjected to a rigorous software engineering discipline for its implementation.

Approaches: Currently, searching for flaws in microprocessor design requires the use of formal verification tools to evaluate a chip's logic design. The tools help considerably, but still do not scale to full size for today's processors and applications. Scaling these formal verification tools will be critical to building systems of higher assurance than today's systems. Also, these tools should be available for pervasive uses in military systems, as well as commercial providers of process control systems and real-time operating systems. Scaling these tools will require the generation of new languages for formal verification, and progress in the capabilities of model checking and theorem proving. In particular, significant progress has been made in the past decade on static analysis of source code.

Second, verification of a poorly built system "after the fact" is inefficient and inadequate. Verification is expensive, and most COTS systems are built around functionality and optimized on cost to the detriment of security, often producing literally countless bugs. An alternative approach is to check the soundness of a system as it is being built. This approach will depend on new languages, environments that enable piecewise formal verification, and more scalable proof generation technology that requires less user input for proof carrying code. Also, these design environments must work effectively within the constraints of embedded devices. Success may require a computer automated secure software engineering environment (CASSEE) for the construction of secure systems.

Third, measuring confidentiality *holes* in trustworthy construction requires the ability to measure the covert channels through which information can leak out of a system. Covert channels have been well studied in the constrained, older, local sense of the term. However, covert channels are not just limited to the scope of microprocessor architectures, hardware/software interfaces, and operating systems, but may include more distributed forms of covert channels. In an increasingly connected world of cross-domain traffic, distributed covert channels become increasingly available. Although covert channels have been studied, few tools exist that capture that knowledge and expertise. Furthermore, for more distributed forms of covert channels, we lack the tools, mathematics, fundamental theory, and science for risk assessment.

A fourth critical element is the creation of comprehensible models of logic and behavior, with comprehensible interfaces so that developers can maintain an understanding of systems even as they increase in size and scale. Such models and interfaces should help developers avoid situations where catastrophic bugs lurk in the complexity of incomprehensible systems or in the complexity of the interactions between systems. Creation of a language for effectively specifying a policy between so many components represents a mathematically hard problem. Problems that emerge from interactions between components underscore the need for verifying behavior not only in the lab, but in the field as well.

Last, efficiently creating provably trustworthy systems will require creation of a secure but flexible foundation for computing. Without a secure computing base, developers will forever remain stuck in the intractable position of starting from scratch each time. This foundation must

include verified and validated hardware, software, compilers, and libraries with easily composable models and means of verifying compositions of those components.

Metrics: Properties that are important to the designers of systems should be measured in terms of the scale of systems that can be proven to be trustworthy. Success against covert channels (and distributed covert channels) should be measured for minimizing leakage. The effectiveness of CASSEE should be measured in the reduction of person-hours required to construct and verify systems of equal assurance and security. The reuse and size of components being reused should be measured, since the most commonly used components in mission-critical systems should be verified components.

5. Situational Understanding and Attack Attribution

Definition: The potential impact of attacks on critical infrastructure information systems becomes harder to determine as systems become more complex and cyber attacks become more sophisticated. Even after an attack has been detected, determining an appropriate response requires the ability to answer some hard questions. Who is attacking? Where is the source of the attack, and where are the attackers? Why are they attacking? How might the attack progress? How does this attack affect critical functionality? What are the best courses of action? Can what has been learned so far be used to prevent future in the future? IT Security decision makers need automated tools to help answer these questions in milliseconds, not months later, using traditional forensic reconstructive techniques. Without answers, the ability to respond effectively in defending critical systems is severely limited.

Threat: Irrespective of whether the primary misuse threats are from insiders, outsiders, hardware malfunctions, environmental considerations, or other factors, attacks are pervasive. Obtaining timely answers to the questions posed in the preceding paragraph has been an extremely difficult challenge, whether the attackers are individuals, nation states, or cyber-terrorists. The lack of adequate identity management (Hard Problem 1), the difficulties of maintaining system survivability (Hard Problem 3), and the complexities of developing trustworthy systems (Hard Problem 4) all make the problems of analysis and response more difficult.

Motivation: Some attacks might be a prelude to war by a nation-state, and others might be reckless teenagers engaged in criminal behavior. In any attack situation, the appropriate response depends on knowledge of who is attacking, understanding of the potential intent of the attack, and the possible consequences of both the attack and the responses to it. Many of the assumptions of the past may no longer be valid. For example, attackers may now be using tools that are even more sophisticated than those of the defenders. Physical location and barriers may now be more or less irrelevant if attacks may be coming from anywhere in the world. Simultaneous coordinated attacks are now easily conceived and perpetrated. Attacks may be multithreaded and polymorphic. Appropriate responses may be non-local, necessarily involving widely dispersed systems and system administrators.

Challenges: In cyber space, attackers may easily obscure their identity. Consequently, defenders have a more difficult task in discovering them. This hard problem involves the development of technologies that can rapidly and efficiently attribute an attack to the logical and physical source across multiple disparate network technologies and infrastructures. Traceback capabilities must work through multiple compromised hosts and across different jurisdictions, occasionally with hostile service providers. Attribution must work against the entire spectrum of attacks. Techniques must be effective against the variety of mechanisms that adversaries might use to avoid being tracked, including masquerading, multi-hop attacks through unwitting users, and covert communication of command signals and data. Attribution must provide enough confidence in the result to allow decisive actions. Technology limitations that add to the challenge of providing attribution arise because of the vulnerability of current protocols, inherent problems with anonymity and pseudo-anonymity, open enterprises, and public access points

(among others). All of these require much more sophisticated situational awareness and response capabilities.

Attribution must include identifying specific computers and their physical locations. There are three levels of attack attribution: (1) to the specific hosts involved in the attack, (2) to primary controlling host(s), and (3) to the actual human actor. Attribution to a higher organization behind an attack is typically a matter for traditional intelligence processes.

Effectively responding to attacks requires situational understanding that also reaches beyond attack attribution. Typically, the human defender of a network is deluged with data from network monitoring systems. The sensor data comes from a variety of systems, including firewalls, antivirus products, intrusion detection systems, intrusion prevention systems, and network management systems. Current technology has basic capabilities to reduce and correlate relevant data, but providing the operator with the best information possible requires improving the scalability of the audit reduction to support a fused picture from more pervasive auditing. Cyber visualization must provide the user with an accurate representation of the IT system's status. Also, cyber visualization must allow analysis to a fine level of granularity in a timely manner without losing any critical content.

Last, decision-makers need tools to assist in quickly deriving appropriate responses. Currently, the courses-of-action process is only partially automated and recommends coarse-grain responses (such as shutting off communication with suspected nefarious networks). More effective tools will assist in the selection of precise responses that take into account mission priorities, and allow the development of "what-if" scenarios.

Approaches: A number of research areas can assist with both attack attribution and broader situational understanding. Detection of adversarial command and control systems could be accomplished by sensing and capturing control messages traveling from the attacker to other hosts involved in the attack. The detection system could identify the Internet hops used to relay the commands, or identify a probable point of origin without identifying each hop. Attack attribution could include advanced packet seeding and tracing techniques, automated and non-interactive network traffic correlation techniques, and advanced router design for packet capture and analysis.

Cyber network forensics is critical to determining attack attribution. This area deals with gathering digital evidence related to an attack or any event of interest. Effective cyber forensics must provide timely information on modern IT systems. This information is needed for mission-critical and critical-infrastructure information systems at a fine level of granularity and fidelity: for example, terabit-per-second (or faster) systems. Also, answers are needed in milliseconds, not months.

Metrics: Metrics are sorely lacking for the effectiveness of today's anomaly- and misuse-detection systems, which are themselves prone to false positives and false negatives. However, metrics for just false positives and false negatives are inadequate (and generally very inaccurate), and need to be augmented with additional metrics (for example) for how accurately the analysis is able to determine the origins of attacks, the actual perpetrators, and their intents.

Effectiveness of situational understanding can be measured by degradation of mission function during normal operations that may be tested by Red Team attacks. Red Teams can have objectives of either degrading missions or stealing mission-critical secrets, as defenders must often choose to degrade their service to stave off adversaries. Improvements in correlation, visualization, predictive technologies, and other decision support should minimize service degradation for all types of attacks. Defenders' capabilities should be measured on the reliability with which they correctly identify the adversary or class of adversary conducting the attack.

The use of large, real network-level datasets, which are now becoming available to the research community, will enable researchers to make repeatable measures of performance. With such metrics and technology, decision makers at all levels could have far better understanding of the situation evolving in their networks, and how best to defend critical infrastructure information systems.

6. Information Provenance

Definition: Information provenance is concerned with the original authorship and subsequent treatment of computational objects such as programs, data, and formatted display items, including changes and transformations from one medium to another. It is generally concerned with the integrity of the information rather than the actual content. Although the classification of information to some extent governs data handling, provenance relates to chain of custody at all levels of transformation and various granularities or layers of abstraction.

When people and machines make decisions based on information, they need to know that the information is "good." Identifying the source in the past was easier, because data came from fewer sources. Consequently, decision makers had more confidence in the data's reliability. Today, most information is aggregated from many sources and, even if the sources are sound, the transformation and aggregation processes are obvious targets for the adversary. Many new and emerging applications are able to store information in various formats for static, stream, and interactive multimedia use. Some products transform data from one format to another, thus making many products interoperable, but also compounding the security challenges. Separating releasable data from more sensitive data becomes harder with these transformation and aggregation processes. Furthermore, given the vulnerabilities in transformation processes, a completely new dimension of reliability must be considered: the set of transformations, as well as the sources that trace back to the data's origin.

Threat: Provenance-related threats arise in data migration, software distribution, and forensics efforts triggered by computer misuse. Specific problems arise in detecting and analyzing forgeries and resolving claims of authenticity --- including would-be repudiations of previous actions that may or may not have been genuine. Tampering with provenance itself is also a threat, including alterations to audit trails that are neither once-writable nor adequately protected, as well as disassociating provenance from the objects that it purports to represent. The integrity of provenance can also be complicated by the presence of man-in-the-middle attacks.

Motivation: Warfighters, emergency responders, and decision makers in general need reliable information from complex sources and analyses. Information must be carefully guarded from malicious individuals. Also, information accuracy and integrity must be maintained and tracked to ensure that users do not make serious mistakes based on misunderstandings of the pedigree of the data. Families of civilians killed by a military weapon will receive little solace when told that a mistake was made because of an inaccurate source or a compromised analysis process. As diverse data sets are combined, accurate information will be interspersed with less accurate information. Even the aggregation processes themselves might be suspect. It is of critical importance to identify and propagate the source and derivation (or aggregation) of the information with the information itself. Analysts will want to view the data provenance and the derivation and aggregation so that life-critical decisions (such as targeting a weapons system) can be made while considering the reliability and accuracy of the information. Tracking requires information provenance and pedigree at various scales, granularities, and levels of confidence.

Challenges: Several challenges make this problem exceptionally hard, including granularity, complexity, and volume.

Granularity has long been handled with labeled textual units in documents. Paper and electronic text can be tagged at the document, paragraph, sentence, word, or library level. These levels are discrete. Geospatial and multimedia data are not so easily made discrete. Video compounds this problem by adding the dimension of time and resolution. How should the pedigree be preserved as information is transformed between formats such as rasterized data, vectorized data, and symbolic, relational, or other formats? Each of these formats exists to support mission requirements. With increased demand for information, automated translators are finally emerging. While careful hand mapping of these translations may once have been practical, the sheer volume of today's computation makes such an approach uneconomical, time consuming, and inflexible.

Volume is where the largest challenge lies. Part of what made information provenance easier in the past was its small volume. Now, geospatial information-gathering systems are being planned that will have the capability of handling gigabytes of data per second, and the challenges of these data volumes will be exacerbated by collection via countless other sensor networks. Within 20 years, the Government will hold an exabyte of potentially sensitive data. No set of persons can manually track the pedigree of such information or determine its derivation and aggregation. The systems for handling and establishing provenance of such volumes of information must function autonomously.

Compounding the problems, many emerging applications such as sensor information and streaming data management must be current and of high quality, requiring processing in real time. If a deadline is missed, catastrophic events could result. Furthermore, the volume of data must be processed without sacrificing data quality and integrity. Provenance is crucial to decision making and also is important to declassification and release processes, which become more essential as coalitions and sensitive inter-governmental cooperation continue to increase.

The challenge may be conceptually bounded by the problem of maintaining the pedigree and integrity of all information that is digitally held in U.S. Government systems and that becomes traceable to the point of data collection where the information first entered the system.

Approaches: Few proposed solutions address these challenges. Rather, they address specific data types, without provision for different levels of security and information integrity on different scales and granularities. There is no widespread treatment of data pedigree based on geospatial and multimedia data scaling or granularity. Any approach will have to reconcile issues that appear to be in conflict. For intelligence data, it is common to intentionally hide the sources (and methods) used to acquire the data, so there is a conflict between the desire to provide provenance and the desire to distribute "product" more broadly than knowledge of the source would permit.

Metrics: Although it is possible to maintain information provenance today for small-scale systems, the real challenge is expanding the scope to large-scale systems and networks. One indicator of success will be the ability to track the pedigree of information in large systems that process and transform petabytes of data per day.

7. Security with Privacy

Definition: The goal of security with privacy is to create tools for users to better protect and control the confidentiality of their private information, even when they choose to (or are required to) share it with others in a limited and controlled fashion. This document does not attempt to address the question of what information should be protected or revealed under various circumstances, but it does highlight challenges and approaches to providing technological means for protecting both security and privacy.

Threat: Threats to privacy may be intrinsic or extrinsic to computer systems. Intrinsic computer security threats to privacy attributable to insider misuse include misuse of authorized privileges as well as insider exploitations of internal security flaws. Intrinsic threats attributable to outsiders (e.g., intruders) include a wide variety of intrusion techniques. Extrinsic threats arise once information has been viewed by users or made available to external media (via printers, e-mail, wireless emanations, and so on), and become completely outside the purview of authentication, computer access controls, audit trails and other monitoring on the originating system.

Motivation: Modern commerce and modern society rely on countless technologies and procedures for safety and convenience that often require individuals to provide information considered sensitive and private. Some societies depend on a pervasive and institutionalized implementation of privacy, and consider privacy to be an important civil liberty. One example of legislation that pertains to privacy is the Health Insurance Portability and Accountability Act (HIPAA) that makes organizations and individuals legally liable for failing to protect private health information from all unauthorized disclosures. HIPAA applies to healthcare, research, insurance, and a wide set of industries where private health information is handled. A growing concern is identity theft, a problem that, left unchecked, threatens to undermine consumer confidence in IT systems. Clearly, strategies must be developed whereby security and privacy coexist.

Data mining, sorting through data to identify patterns and establish relationships, may not only be used beneficially to prevent disasters, to save lives, to identify criminals, and to resolve forensic problems, but also be used to aggregate sensitive private information for nefarious purposes, of which identity theft is the most obvious example. Protecting the privacy of individuals and institutions in the face of widely accessible information is extremely difficult.

Challenges: Historically, governments have cited compelling national security needs for seeking to violate privacy. An example of such arguments is that security is most effective when based on actual and timely information that can be attained only through pervasive monitoring. In contrast, individuals and corporations cannot function in a modern society unless they have means to protect certain types of information. Protecting digital information relies on the use of encryption, steganography, sanitization, or anonymizers when sending electronic communications. Paradoxically, the use of these protection techniques can be viewed as a significant challenge to national security organizations responsible for identifying potential

threats and to preempting planned attacks. This apparent contradiction may not be amenable to a resolution. Even so, various possible approaches appear promising.

Many people believe it is absolutely necessary to maintain the confidentiality of medical records. However, in medical emergencies, such as allergic reaction to certain medication, information needs to be available to health care professionals responsible for making timely and accurate decisions. In some cases, the challenge is to ensure privacy but also to have the appropriate information at the right time during critical situations. In this regard, DRM techniques may be promising as a mechanism for protecting information in such diverse settings as healthcare records and corporate proprietary data; because the originator of the information thus (supposedly) retains varying degrees of access control even after the information has been given to third parties. A significant challenge to the DRM approach is the development of an indisputable definition of the originator. For example, the "originator" of medical information could be defined as the patient, doctors, nurses, hospitals, or insurance companies. In fact each of those listed may be the originator of different portions of the "medical information." Therefore, information provenance has an intersection with privacy to maintain a trail of who did what to the "medical information", and an intersection with both system and information integrity.

Approaches: Security with privacy appears to require establishment of fundamental trust structures to reflect demands of privacy, and means of reducing privacy breach risks through technologies such as data mining. (See Hard Problem 6, Information Provenance.) Data mining can be used to detect terrorist activities and fraudulent behavior, as well as to help correlate events and anomalies in medical research and disease control. Whereas these applications of data mining can benefit humans and save lives, data mining also has been viewed by many as a threat to the privacy of individuals. Ideas for reconciling data mining with privacy concerns include privacy-preserving data mining, distributed association-rule mining algorithms that preserve privacy of the individual sites, and a new formulation of privacy breaches that makes it possible to have limits on breaches without knowledge of original data distribution. These are only a few examples of many possible approaches for enabling data mining while preserving privacy.

Today, identity-theft related credit-card fraud has become a multibillion-dollar per year problem. Critical systems must be restructured so that the public has privacy with security, not security in lieu of privacy. In alternative schemes, a would-be consumer informs a credit provider of the desired intent, and the credit provider more directly assures the seller of compensation, without revealing the consumer's private information such as name and credit-card number. Although innovative strategies such as these can yield the functionality and security needed by all while protecting the identity and privacy of citizens, generalized solutions will be much harder to conceive and implement.

Although each of these approaches is promising and challenging, this is by no means an exhaustive list of ideas on how to achieve security with privacy.

Metrics: Although interesting approaches are beginning to emerge, time-proven metrics do not exist. A goal for addressing concerns regarding both data mining and identity theft is to provide users with the ability to retain control of sensitive information and its dissemination even after it

has left their hands. For data mining, quantitative measures of privacy have been proposed only recently, but are still fairly primitive. Refinement and validation of such metrics is certainly in order. For identity theft, a goal should be a simpler vision where citizens would not be required to reveal uniquely identifying information to anyone, with the exception of narrowly defined instances associated with law enforcement (e.g., cases where a warrant is obtained to require disclosure and where reuse of that information could be adequately controlled and monitored). Although certain sensitive systems and nationally critical capabilities will always require valid identification and authenticated authorization for use, there might other systems that serve national interests better by allowing citizens to retain more control of their private personal information.

8. Enterprise-Level Security Metrics

Definition: Along with the systems and component-level metrics that have been mentioned in the preceding "hard problems," and the technology-specific metrics that are continuing to emerge with new technologies year after year, it is essential to have a macro-level view of security within an organization. What happens when all the systems, processes, and tools are turned on? Today, government decision makers and corporate leaders do not have answers to important questions such as, "How secure is my organization? Has our security posture improved over the last year? To what degree has security improved in response to changing threats and technology? How do we compare with our peers? How secure is this product or software that we are purchasing? How does it fit into the existing systems and networks? What is the marginal change in our security, given the use of a new tool or practice?" Most organizations view the answers to these questions in the short term from a financial mind-set and make a cost-benefit trade analysis. The decisions resulting from this analysis will frequently be to the detriment of significant improvements in security in the long term, which may require costly new development.

Threat: One of the most insidious threats to security metrics lies in the metrics themselves. The mere existence of a metric may encourage its purveyors to over endow the significance of the metric. A common risk is that analyses may be based on spurious assumptions, inadequate models, and flawed tools, and that the metrics themselves are inherently incomplete --- often a one-dimensional projection of a multidimensional situation. Furthermore, a combination of metrics in the small (e.g., regarding specific attributes of specific components) typically do not compose into metrics in the large (e.g., regarding the enterprise as a whole).

Motivation: Without answers to these important questions, management is mired in a quandary without meaningful direction. The dearth of metrics and decision-making tools places the determination of information security risk to the enterprise on the judgment of IT security practitioners. The gathering and sharing of information about threats, vulnerabilities, and attacks is critical to establishment of a scientific approach to managing these risks.

Metrics and a risk management framework must guide decision makers. First, recent events (like 9/11 and its economic impacts), along with intelligence reporting, have shown the existence of considerable threats to the critical infrastructures of the United States. Second, financial restrictions require explicit understanding of how funds invested in security will affect an organization. Last, regulations such as the U.S. Information Security Management Act (FISMA) and the Public Company Accounting and Investor Protection Act require the government and private sector firms to become accountable in the area of IT security. These factors support the need for decision makers to have sound metrics and a decision-making framework that embraces risk management principles.

As technology continues to advance into every facet of society, societal dependence on technology grows. This dependence has increased unabated. Technologies are at risk not only from highly publicized hackers, but also from more deceptive and dangerous nation-states and terrorists. In addition, systems that are poorly designed, implemented, and maintained tend to

fall apart on their own, without any attacks. Organizations need a metric-based approach built on qualitative and quantitative risk management principles for the effective allocation of IT security resources, in addition to empirical methods.

Challenges: Many challenges still exist in this area. First, in a world where technology, threats, and users change so quickly, tomorrow's risks may be quite different from yesterday's risks, and historical data is not a sufficiently reliable predictor of the future. Second, organizations are reluctant to share information, thus making data on emerging threats difficult to collect. Even when network owners are aware of threats, the constant barrage and high volume of low-level threats (e.g., phishing attacks and spam) distract many organizations from defending against potentially devastating attacks representing more serious threats. Third, risk management is complicated by a dearth of adequate information on capabilities and intentions of threat agents, such as terrorists and hostile nations. To estimate the potential costs of downtime, loss, or impairment of tangible and intangible assets across an entire organization for previously unseen events is almost impossible. Finally, *complete* security is unattainable at any price, and security is not simply a matter of technology.

Many factors complicate the statistical foundations of any approach to predict the likelihood of attacks for a range of impacts. Better protection for some resources often merely increases the likelihood of other resources being attacked. Attackers will shift their focus from more protected resources to less well protected resources. Furthermore, IT security technology is often bought through a principle of adverse selection: Groups that are the most lucrative targets will buy the most defensive technology, and although those defenses may decrease attacks, those organizations may still be attacked more than their peers that are less lucrative targets. This creates a misperception that defenses draw attacks. Amplifying this perception, the best defended groups often have the best sensors, catching and reporting more successful attacks than other groups, leading to the imprecise conclusion that funds spent on defenses have allowed the number of successful attacks to rise when in reality the number of successful attacks may have fallen although the fraction being detected may have risen. Also, even as the fraction of attacks detected rises, that fraction is never known, because "you never know what you don't know." IT security also experiences self-falsification through a set of moral hazards similar to the claim that "seatbelts cause accidents" --- in that such protection can lower users' risk aversion, causing them to operate systems less cautiously. These factors make formal metrics for IT security difficult.

Approaches: Many disciplines operate in environments of decision making under uncertainty, but most have proven methods to determine risk, for example, financial metrics and risk management practices; balanced scorecard, six-sigma, insurance models; complexity theory; and data mining. The field of finance, for example, has various metrics that help decision makers understand what is transpiring in their organizations. These metrics provide insight into liquidity, asset management, debt management, profitability, and market value of a firm. Capital budgeting tools such as net present value and internal rate of return allow insight in the return that can be expected from an investment in different projects. In addition, there are decision-making frameworks such as the Capital Asset Pricing Model and Options Pricing Model that link risk and return to provide a perspective of the entire portfolio. These frameworks have demonstrated some usefulness and can be applied across industries to support decision making.

A possible analog for IT security would be sound systems development frameworks that support an enterprise view of an organization's security.

Metrics: The IRC supports the Computing Research Association's finding that an excellent goal or Grand Challenge for this area would be that, within 10 years, quantitative information-systems risk management should be at least as good as quantitative financial risk management.

However, a caveat is needed. This goal has serious pitfalls based on some inherent differences between the more or less continuous mathematics of multidimensional econometric and financial models on one hand and the more or less discrete nature of computers on the other hand. For example, a one-bit change in a program or piece of data may be all that is required to transform something that is extremely secure to something that is completely insecure. Metrics for the validity of metrics for security need to be taken with a grain of salt. Indeed, metrics about metrics always seem to be speculative.

Conclusions

This report identifies eight hard problems of INFOSEC science and technology that are important to the security of IT systems crucial in the U.S. Government's responsibilities to the nation. These problems were selected because of their importance to Government missions and the inherent difficulties in obtaining solutions. However, by no means should these problems be mistaken as the only challenges in the field of IT security. Even solutions to all eight of these problems would not ensure the security of any given system, because these eight problems by no means span the technological and non-technological spectrum, and because sound information security is an ongoing process that involves more than technology. Although not within the scope of this study, several non-technical issues impact the protection profile of information and systems. These non-technical issues include policy issues, legal issues, technology transition challenges, cost of leveraging good research, economics and market forces that drive those costs, and academic education and training driven at least in part by those market forces. Just as humans are an essential part of information systems when viewed in a broad sense, so is the human element a critical piece of what is needed for effective IT security.

The IRC members recognize the extent and complexity of providing sound information security. Our members also recognize the role that technology plays in providing a fundamentally sound base on which to build a secure IT infrastructure. Without that sound basis, few information systems can fully succeed. We also place a great deal of confidence in the aid that can be achieved through the use of scientific and formal methodologies in assuring that advanced technologies are implemented with comprehensive security.

Each of these eight problems represents a major challenge that currently cannot be solved, but which must be appropriately addressed to provide adequate security and functionality for current and future government systems and nationally important non-government systems. As such, the IRC puts forward these eight hard problems as worthy goals for research and technology development.

Appendix A: Retrospective on the Original Hard Problem List

Over the last several years, the original Hard Problem List has been influential in helping shape and coordinate research strategies within and between various organizations. Due consideration of the list's effectiveness and success requires at least some brief discussion of how our understanding of each of the original hard problems has changed over the last five years. Begun in 1997 and released in 1999, the original Hard Problem List was divided into functional problems and problems associated with design and development. The functional problems in 1997 were

1. Intrusion and Misuse Detection
2. Intrusion and Misuse Response
3. Security of Foreign and Mobile Code
4. Controlled Sharing of Sensitive Information
5. Application Security
6. Denial of Service
7. Communications Security
8. Security Management Infrastructure
9. Information Security for Mobile Warfare

In 1997, the problems associated with design and developments were

A. Secure System Composition
B. High Assurance Development
C. Metrics for Security

Additional details about the above problems may be found at www.infosec-research.org.

Five years of failures and successes against many of these problems have refined the common understanding of them. Similarly, five years of change in U.S. systems and available technology have made some of these problems more relevant and others less so, particularly as threats to U.S. Government systems have increased in number, sophistication, and variety. The remainder of this section describes each of the previous hard problems in light of these changes. It should be mentioned that the absence of a hard problem from the new list does not mean that it has been solved or that the problem has gone away. In some cases, the focus has been narrowed to the harder problem buried within. In other cases, the hardest research challenges have been solved, but the larger expenses of development toward solutions in those areas have only just now begun. In each case, this retrospective on the old list of INFOSEC research hard problems should help put each challenge in the context of these changes and progress.

The Functional Hard Problems of 1997

1. **Intrusion and Misuse Detection:** Although progress in intrusion and misuse detection has been hard earned, the problems of detecting new and novel attacks are still far from solved. Much progress has been made in these areas, with many but not all research goals met. In fact, systems under research can now detect many attacks missed by previous generations of intrusion detection systems, but commercial systems are still riddled with false positives and false negatives, especially in high-volume situations such as networking. Although very few of these research systems have been fielded, this represents tremendous progress in the lab. However, five years of experience has shown that the general problem of intrusion detection leaves adversaries too much room to maneuver, and that the general approaches to intrusion detection are completely blind to certain classes of attack, such as life-cycle attacks. With this progress and these limitations in mind, the more appropriate and more narrowly defined hard problems now seem to be detecting and mitigating the insider threat (new Hard Problem 2), and detecting and measuring covert channels (new Hard Problem 4). In addition, while substantial progress has been made on the general problem of detection, research into fully automated correlation has only recently begun in earnest, with more research yet to be done as described in Situational Understanding and Attack Attribution (new Hard Problem 5). However, the most fundamental problem is the inadequate security of the computer systems themselves (addressed in part in new Hard Problem 4). If that problem were more adequately addressed, many of the intrusion opportunities could disappear (although denial-of-service attacks and insider threats would still need to be considered). The critical dependence on systems that are inherently flawed is a serious obstacle to all of the hard problems --- old and new.

2. **Intrusion and Misuse Response:** Although difficult research persists, intrusion prevention technologies in the market now can respond to many kinds of attacks, and other technologies can mitigate distributed denial of service (DDOS) via quick response. However, more progress remains to be made in reliability and safety of these technologies in life-critical applications, particularly given the presence of false alarms. Given this progress, and the degree to which response depends upon detection, the emphasis of this area has been refocused on insider threat detection (new Hard Problem 2) and detection of covert channels (new Hard Problem 4), and Situational Understanding and Attack Attribution (new Hard Problem 5) as described in the previous paragraph.

3. **Security of Foreign and Mobile Code:** Although difficult research remains, proof-carrying code and sandboxing represent important advances in limiting the potential negative effects of foreign and mobile code. However, even domestic production of software is being outsourced to firms offshore. Moreover, even at reputable software companies, insiders can be bought to plant malicious code into key products used by the U.S. Government. This is true whether software is installed from disk or downloaded from the Internet. Given these realities, the focus of security attention for foreign and mobile code seems best shifted to the challenge of developing trustworthy software in the first place, and in conducting extensive static analysis of all critical software --- especially foreign and mobile code. Some significant progress in the use of formal methods has been made and deserves further

emphasis, although such techniques must be applied judiciously where they can have the greatest effect. (See new Hard Problem 4.) As noted above in how to simplify the problems of intrusion and misuse detection, the most fundamental problem is the lack of computer operating system protection that can protect the operating system against itself and against applications (including mobile code). Such systems would inherently make opportunities for sandboxing of potentially malicious mobile code much more effective. Procedural and policy activities have also changed the way this hard problem has been addressed.

4. **Controlled Sharing of Sensitive Information:** Progress has been made in terms of languages, labeling, and cryptographic schemes for controlled sharing of sensitive information. Progress in digital rights management may ease remaining policy specification challenges by empowering end users to set policies as required on information as it is being created. However, without a foundation of trustworthy enforcement mechanisms for enforcing separation, the value of this progress will be substantially diminished. For these reasons, high emphasis is focused on the challenges of building trustworthy foundations in building scalable secure systems (new Hard Problem 4). Such a foundation is also necessary (but not sufficient) for any potentially useful approaches to digital rights management, which also require some additional hardware support. In addition, progress in tracking the provenance of information (new Hard Problem 6) will also aid controlled sharing of sensitive information by easing automation of ability to release decisions. Furthermore, better means of analyzing covert channels (new Hard Problem 4) will also help controlled sharing.

5. **Application Security:** Application security has seen important progress toward intrusion tolerant applications that are able to function despite flawed components and are less reliant on an underlying Trustworthy Computing Base (TCB) than traditional applications. However, research remains toward getting these techniques to work in distributed, asynchronous, time-critical environments (new Hard Problem 3). In addition, one of the most painful lessons has been that there will always be situations where a TCB is needed. This helps motivate emphasis toward a truly trustworthy TCB in building scalable secure systems (new Hard Problem 4).

6. **Denial of Service:** Although critical hard research remains, progress has been made toward assuring the availability of information systems against denial-of-service attacks. Technology now exists for mitigating distributed denial-of-service attacks. Moreover, progress from traditional fault tolerance can now help mitigate other denial-of-service attacks. However, many of these defenses require tight coupling of systems in relatively high-bandwidth environments. Guaranteeing availability of time-critical systems in lossy, loosely-coupled, wireless environments (new Hard Problem 3) remains a hard and increasingly important problem as the U.S. depends more and more on time-critical, life-critical information systems.

7. **Communications Security:** The foundation of secure communications is the infrastructure for managing cryptographic keys. Furthermore, communications systems are not truly secure without verifying and authenticating identities of people as well as computers and communications media. With government and commercial systems for secure communications on the shelf, but predicated upon such infrastructures for authentication and key management, it seems to make sense to narrow this hard problem to Global-scale identity management (new Hard Problem 1) and all the challenges therein. This narrowing also

deemphasizes broader cryptography as part of the hard problem, as the need for sustained investment in cryptography research is driven less by unmet challenges, and driven more by a need to keep ahead of an adversary's cryptographic measures and countermeasures. Moreover, given this success at building systems to protect secrecy of communications, and the remaining challenges of assuring availability of communications, much of this problem also has been narrowed toward assuring availability of time-critical systems (new Hard Problem 3).

8. **Security Management Infrastructure:** Although critical research remains, industry has already begun acquiring emerging security response management technologies, and investing in a next generation of such technology. However, aside from the infrastructure for managing security response, infrastructure is also required for managing cryptographic foundations of security. As described under Communications Security (old Hard Problem 7), this subset of hard problems seem to rest on the challenges of establishing key-management and authentication infrastructures (new Hard Problem 1), as described in the previous paragraph. Moreover, additional research into security response management requires better situational awareness and attack attribution (new Hard Problem 5).

9. **Information Security for Mobile Warfare:** Both homeland defenders and military now depend upon mobile and secure networked computing, particularly given risks of attack and the need for fire, police, rescue, and recovery personnel to be able to securely coordinate crisis response via information systems, services, and networks. Given the importance of these challenges, it seems to make more sense to divide the problem into its constituent sub-problems, and to address more directly the challenges of availability in time-critical systems (new Hard Problem 3) and building scalable secure systems (new Hard Problem 4) within the constraints of heat, power, timing, bandwidth, and size faced in mobile environments.

Problems Associated with Design and Development

A. Secure System Composition: Predictable composition of subsystems into systems with desired trustworthiness properties remains a crucial hard problem. Recent experience demonstrates many of the difficulties. Furthermore, simplistic approaches such as just adding firewalls and intrusion detection systems and increasing the length of easily compromised fixed passwords have been recognized for the fantasies that they actually are. The discussion of building scalable secure systems (new Hard Problem 4) describes various approaches for developing predictably secure systems, including formal and other promising techniques for composing secure systems out of trustworthy components.

B. High-Assurance Development: Although some high-assurance development tools do now exist, they do not scale. For this reason, the Computer Automated Secure Software Engineering Environment (CASSEE) is proposed as a focus of building scalable secure systems (new Hard Problem 4).

C. Metrics for Security: This problem remains hard, and is described at length under Enterprise-Level Security Metrics (new Hard Problem 8), with some descriptions of the areas of INFOSEC where progress has been made, and discussion of areas that remain

exceptionally challenging and remain exceptionally important. Unfortunately, there remains a fundamental difference between the security of a system product (with no applications, no users, and no network connectivity) and the security of a system in its actual use, relative to all of the threats that may be encountered in its actual operational environment. Even more difficult is the much more broadly scoped problem of the trustworthiness (including security, reliability, survivability, timeliness, and so on) of all of the interrelated computers that support an entire enterprise --- that is, security in the large.

Appendix B: Global-Scale Identity Management

Concepts relating to identity management have been well-known for many years. These concepts include establishment of distinguishable identities for users (where *users* is a term that encompasses people, systems, network nodes, and other computational entities in hardware and software); varying granularities that might range from single bits to entire systems and sub-networks; monitoring and misuse detection with respect to identities, access controls, and resource use; and revocation of access privileges for particular identities.

All of these concepts depend in varying degrees on the whether the perceived user identities are in fact correct, which is becoming increasingly difficult to ensure as information infrastructures transition into a mode of ubiquitous global access to highly distributed networked computer-communication resources across which identities may be interoperably recognized and validated and in which vast collections of remote resources may effectively be considered as virtual extensions of local computing. In particular, the scalability of these concepts is in question. Identities may have to be globally unique, or at least globally resolvable. Authentication may in some cases need to be trustworthy, even though it may have taken place on untrustworthy platforms. Authorization must be sensitive to the potential untrustworthiness of identities and authentication, rather than simply assuming the trustworthiness of external evaluations. Revocation may in some cases have to be instantaneous, and it might also have to be global rather than local, irrespective of the presence of unknown mirrored copies of software and data. As a consequence, each concept relating to global-scale identity management needs to be reexamined in light of the emerging global scope.

Further concerns are introduced by incompatibilities among different conventions for identification, different access-control regimens, different protocols, and different languages. Even if open interface standards exist across radically different systems and networks, incompatibilities may still exist across different vendors. The problems are further complicated by anonymous and pseudonymous identities, anonymizers that mask true identities, aliases that allow multiple identities, pseudo-anonymous servers with mechanisms by which identities can be obtained by law enforcement but that are also susceptible to misuse by untrustworthy trusted insiders, and weak security that allows identities to be spoofed. Privacy implications must also be respected, although many of those privacy issues are extrinsic. (See Hard Problem 7 for further discussion of privacy and forms of anonymity.) Some additional problems arising in global-scale identity management are enumerated as follows.

- **Authentication.** Identities are unlikely to be meaningful uniformly everywhere in the world. In some cases, authenticated identities are much less important (for example, in net browsing public and widely mirrored Web sites) than in other environments (at the other extreme, classified computing). However, an identity that is authenticated by one host or server is unlikely to be accepted outside of the perimeter of (un)trustworthiness in which it was first established. For example, the notion of a single sign-on is extremely dangerous in an open context, and is only marginally sensible within closed contexts. Different user roles may still have to be identified and authenticated within the scope of any particular user. Also, even the most elaborate authentication based on a combination of biometrics, user knowledge, and cryptography may be meaningless outside its original context. As a consequence, determination of areas of substantially comparable

trustworthiness is likely to limit the extent of authentication across institutional and other boundaries, and new strategies for transitive authentication are needed. In addition, authentication may need to be established with respect to the intended user roles, with different and more stringent requirements whenever greater permissions are sought.

- **Access control**. Authorization needs to exist within the context of intended user roles, with different permissions granted for different roles. Trusted insiders (see Hard Problem 2) must be treated differently from untrusted outsiders. Furthermore, what may work adequately in the small may not work at all in the large. For example, super-user privileges are inherently unsound across boundaries in a global environment. Similarly, fine-grained differential access controls are not likely to be meaningful outside of local contexts. As a consequence, repeated and perhaps more stringent authentication and access controls that are based on the trustworthiness of the authentication process and the trustworthiness of the authenticated identity may be needed for sensitive remote actions.

- **Monitoring and misuse detection**. Accountability issues will tend to take on a broader character in global environments, with the need to interpret the soundness of authentication and access controls and the trustworthiness of users, as well as extending the concepts of misuse. In addition, correlating and interpreting the significance of detected anomalies across multiple systems and networks becomes both more important and more difficult. (See Hard Problem 5.)

- **Revocation.** Disabling authentication methods when they have been compromised and revoking conferred privileges that are being misused both become much more difficult problems on a global scale. For example, a distributed variant of David Redell's 1974 scheme for indirecting privileges through a primary source and centralizing certain aspects of identity management to a few controllable locations might be used advantageously when revocation is an important global consideration.

- **Long-term integrity of approaches to identity management**. A concern is sometimes raised about a future in which factoring of large primes and other mathematical attacks on public-key crypto compromise the use of cryptographic approaches to authentication. Given that techniques are likely to change as better methods become available; this is perhaps less of a concern for authentication than it is for confidentiality --- for which today's long-term storage (including backups) may eventually be compromised. For one thing, commonly used reusable fixed-password mechanisms need to be eliminated and authentication techniques need to improve dramatically. Furthermore, the persistence of serious security flaws in systems on which authentication techniques may be implemented suggests that the integrity of the authentication of today's systems may be compromised without compromising the identity management. Incidentally, the same statement holds for confidentiality.

Identity theft is an increasingly critical problem, requiring a combination of technological, social, and legal approaches. Some of the above technological approaches can certainly help, but are inherently incomplete.

Appendix C: Insider Threat

This appendix examines the implications of insider threats and considers a few areas that at present are not adequately pursued with respect to preventing, detecting, diagnosing, and understanding insider misuse of computers and networks.

In general, it is unwise to consider defenses against insider threats in isolation of the overall problem of misuse. Considerable benefits can be gained from a coordinated approach that encompasses defenses against both insiders and outsiders --- especially in applications in which outsiders can potentially become functionally indistinguishable from insiders after they have penetrated a system. This is especially true if outsiders have succeeded in acquiring privileges of insiders. There are certainly distinctions between the two classes, but there are also considerable overlaps.

Furthermore, potential distinctions may be blurred by usage patterns, and are thus not always black and white. For example, a physical outsider (e.g., entering from a remote system) can actually be a logical insider (e.g., privileged within the system under attack). Furthermore, a user who is a logical insider in one context (e.g., a multilevel-security compartment) may be a logical outsider in other contexts (unless of course the environment is run rather unimaginatively at system-high with everyone having equal privileges --- which is of course a common phenomenon that seriously impedes logical defenses against insider misuse).

Nevertheless, this appendix is concerned specifically with insider misuse, within the context of architectures for scalable secure systems (Hard Problem 4). At present, it can in some cases be difficult to determine unambiguously and with certainty the identity as well as the physical and logical whereabouts of insider misusers. It is of course even more difficult to determine unambiguously the identity and source of outside attackers. In addition, it is usually also very difficult to determine unambiguously the resources and capabilities that are available to misusers, as well as to determine the resources and capabilities that are actually being used at any particular time during any particular misuse --- whether the misuse originates from insiders or outsiders.

Significant improvements in cybersecurity measures against insider misuse must be accompanied by better prevention of outsider misuse, including authorization technology and other perimeter defenses, and more easily used access controls. Each of these areas represents a set of weak links. However, significant improvements in defenses against outsider misuse (including denials of service) can also make insider misuse easier to detect and distinguish.

Particularly important are the following considerations relating to insider threats.

- **Differential access controls**. With respect to the confidentiality and data integrity aspects, fine-grained access controls can have an important role in narrowing down opportunities for insider misuse, in single-level and multilevel secure systems and networks. Coarser access controls such as mandatory multilevel security with narrowly scoped compartments and carefully enforced restrictions as to which users have access can also be important. (See Hard Problem 1.)

- **Quarantine and isolation**. With respect to system and data integrity, compartmentalization, domain confinement, and process isolation can help narrow down opportunities for insider misuse. Multilevel integrity has considerable appeal in principle, although it has practical limitations that must be overcome.

- **Misuse detection**. Anomaly and misuse detection, and determining the intent of each detected misuser are vital components of dealing with insider misuse. (Needs for early detection, prompt notification, and intelligent autonomic response are relevant to Hard Problem 5.)

- **Dependable traceability of users and resource use**. Dealing adequately with insider misuse relies on being able to uniquely identify the actual misuser(s) and their access path, to obtain as definitively as possible a map of their operating configurations and of every action they may have taken that could compromise security, and to determine the provenance (pedigree) (Hard Problem 6) of any system and network objects (programs and data) that they may introduce or disseminate. Thus, audit trails and general accountability are also important. (See also Hard Problem 5.)

- **Tracking and forensic analysis**, including discovery, preservation of evidence, the need for data integrity, and recovery. With respect to traceback, two basic situations need to be considered: (1) traceback within a given system, local-network enclave, or extended web of trust, within which there is some likely knowledge or expectation of the trustworthiness of the environment and the identity of its participants, and therefore some reasonable expectation of the veracity of the traceback; and (2) traceback beyond any reasonable expectations of trustworthiness, potentially reaching out into totally unknown territory as a result of multiple indirections and network weaving, and therefore a very limited realistic expectation of the veracity of the traceback. The second case is considered in Appendix F, although it is also somewhat relevant here.

    - The first case requires (or, perhaps naively, implicitly assumes) the presence of meaningfully strong authentication, authorization, and accountability. Examples include a single isolated compartmented enclave or a multilevel secure (MLS) local network. The second case is obviously more difficult. In that case, the ability to achieve any reasonable traceback might benefit somewhat from better authentication, authorization, and accountability within any areas of presumed trustworthiness. However, these features could be mere palliatives whenever untrustworthy components are involved; after the cat is out of the bag, or the Trojan horses are inside the supposedly protected barn, it may already be too late. Furthermore, the traceback approach may be of use only within a restricted scope in which it can be trusted, and may be of limited use in the case of network weaving through hostile territory.

    - Preservation of evidence requires the imposition of trustworthy (i.e., non-subvertible and non-repudiatable) integrity checks such as cryptographic seals to detect alterations, and encryption to provide confidentiality of evidence as needed.

- **Recovery of information**. Cryptologic techniques may be needed for cracking encrypted files. Remote access and special analysis tools may be needed if information has been

distributed and fragmented among multiple databases on multiple sites. Recovery can also be complicated by intentional obfuscation, including steganography. However, mechanisms that enable emergency trapdoors for recovering encrypted data are always suspect, being prone to misuse.

- **Recovery of secure system status**. Following detected destructive and possibly security-compromising misuse, Heisenberg uncertainty issues can arise in attempting to do nondestructive analysis. Furthermore, restoring a compromised system to an assuredly secure status can be very difficult.

- **Behavioral modeling.** Given an advanced system for anomaly and misuse detection and analysis, it should be possible to detect a very wide range of attacks, to identify the nature of the attack, and to respond accordingly. However, fundamental to the ability to make an astute response is the ability to have some relative certainty as to the nature of the attacker's intent. Serious research and development is needed here, to characterize intent and to recommend appropriate responses.

Overall, even in the presence of improved authentication, authorization, access controls, and accountability, much work still remains to be done to determine the identity, source, and resources of misusers --- especially if required in near-real time.

Appendix D: Availability of Time-Critical Systems

To provide end-to-end service availability with real-time constraints in the presence of attacks, many challenges must be overcome. From a systems perspective, requirements for availability and survivability in the presence of diverse adversities must be explicitly defined, and interdependencies among different architectural components must be identified and addressed. The components may be geographically distributed, may be inherently unreliable, and may be managed by entities that have different policies or even have conflicts of interest. For example, availability of a network service may depend on the correct functioning of certain network infrastructure components such as the router infrastructure and the domain name system. However, existing network architecture and protocols provide little support for quality-of-service guarantees. Another factor to consider involves highly distributed operating environments. For example, with the rapid growth of the Internet and widespread deployment of high-speed network connections, an adversary may leverage on dispersed computing and network resources to perform a powerful coordinated denial-of-service attack.

Some of the important hard problems are as follows:

- **Wireless mobile ad hoc networks**. Because of their deployment flexibility and adaptive nature, wireless mobile ad hoc networks have been gaining momentum in both commercial and military environments. However, these networks also present additional challenges from the availability perspectives. For instance, in these networks, communication links may be more bandwidth limited, and end-to-end connectivity may be significantly less stable than the wired counterparts. Thus, it may be easier to launch an attack to degrade or even deny services in such an environment. Moreover, the wireless medium is more accessible and thus more vulnerable to certain denial-of-service attacks such as signal jamming. These networks usually do not have a fixed network infrastructure. Instead, untrusted third-party intermediate nodes may be used to relay packets from the sender to the destination. In some cases, such as in sensor networks, nodes may have severe resource constraints in that they may have limited memory, energy reserve, and processing power. As a result, solutions for protecting conventional wired networks may not be applicable for these wireless ad hoc networks.

- **Mission-aware quality of service**. It is necessary to relate system-level configuration parameters to mission-level availability requirements. In addition, the plan of a mission may change as a result of events such as natural disasters or attacks. For example, when there is an earthquake, it is desirable to ensure that emergency response teams can coordinate among themselves and with each other, and can obtain relevant sensor data in a timely fashion --- even at the expense of preempting less critical communication in the area. This not only calls for a scheme that provides quality-of-service guarantees and conflict resolution strategies involving activities of multiple priorities, but also may require interoperability among heterogeneous devices and potentially incompatible systems. A framework is needed for specifying and reasoning about mission requirements and environment conditions, as are techniques for constructing networks and systems that satisfy the mission requirements even in the presence of attacks.

- **Evaluation methodology**. To date, there has been little work on developing a methodology for evaluating the effectiveness of various techniques for protecting the availability of time-critical systems. From a testing viewpoint, an obstacle is created by a lack of testing environments and datasets pertaining to actual denial-of-service attacks. There are recent research efforts to address this issue, namely DETER/EMIST and PREDICT, which are being funded by NSF and DHS. Developing a testing methodology for availability of time-critical systems that gives provable guarantees for coverage and soundness is a hard problem. To that end, more efforts might be put in modeling the threats, mission requirements, and systems, and in developing analysis techniques to enable formal reasoning.

Appendix E: Building Scalable Secure Systems

Designing and developing secure systems (including networked systems) is already very difficult, especially those with stringent requirements for trustworthiness. Developing such systems and networks that can operate effectively when scaled up to widespread use including global applications is significantly more difficult. Several challenges are considered here that are fundamental to the ability to develop scalable trustworthy systems.

- **Requirements**. Requirements should encompass not only security, but also other essentials such as reliability, guaranteed performance, interoperability, and survivability in the face of realistic adversities. The requirements must be clearly specified and able to be evaluated for feasibility and implementability.

- **Architectures**. System and network architectures are needed that are suitably well defined and inherently capable of satisfying the given requirements and that are demonstrably able to avoid a wide range of characteristic design flaws that continually reemerge. The architecture should explicitly address the desire to build the system out of predictably composable components that can avoid incompatibilities and unforeseen interactions. The architecture and design specifications should also pay particular attention to accessible interfaces and to operational needs.

- **System development**. The development effort should use sensible programming languages whose use contributes rather than hinders to the avoidance of implementation flaws, and should embody good software engineering practice.

- **Principles**. Attention to well-known principles for security and trustworthiness should be considered throughout, as relevant. However, these principles must be used wisely, and never considered to be the end goals themselves.

With respect to the future of trustworthy systems and networks, perhaps the most important recommendations involve the urgent establishment and use of soundly based, highly disciplined, and principle-driven architectures, as well as development practices that systematically encompass trustworthiness and assurance as integral parts of what must become coherent development processes and sound subsequent operational practices. Only then can we have any realistic assurances that our computer-communication infrastructures --- and indeed our critical national infrastructures --- will be able to behave as needed, in times of crisis as well as in normal operation. These challenges do not have easy turn-the-crank solutions. Addressing them requires considerable skills, understanding, experience, education, and enlightened management. Success can be greatly increased in many ways, including the availability of reliable hardware components, robust and resilient network architectures and systems, consistent use of good software engineering practices, careful attention to human-oriented interface design, well-conceived and sensibly used programming languages, compilers that are capable of enhancing the trustworthiness of source code, techniques for increasing interoperability among heterogeneous distributed systems and subsystems, methods and tools for analysis and assurance, design and development of systems that are inherently easier to administer and that provide

better support for operational personnel, and many other factors. The absence or relative inadequacy with respect to each of these factors today represents a collection of potential weak links in a process that is currently riddled with vastly too many weak links. On the other hand, much greater emphasis on these factors can result in substantially greater trustworthiness, with predictable results. Many lessons relating to the development of scalable trustworthy systems and networks can be drawn from past research and prototype development, and also from gaps in essential R&D that remain unfilled. However, many of those lessons have been largely ignored by developers of both commercial and source-available software. Thus, a major culture change is needed.

- **Construction methods**. Scalable architectures and use of appropriate software engineering and programming languages are very important. The principle of minimizing what has to be trusted is also important. Architectures that provide high-assurance trustworthiness precisely where it is most needed (for example, in critical servers and basic utilities) are preferable to unstructured systems in which almost everything must be trusted. However, the record for developing large and complex computer-based systems is generally abysmal. For example, The Global Information Grid and other defense applications desperately need various forms of multilevel security in systems and networks, and at the same extensive interoperability; many difficulties must be overcome.

- **Composability**. Ideally, we should be able to compose a system out of components, with predictable behavior of the resulting system. Ideally, systems should be small and analyzable. However, it is generally not advisable to attempt to eviscerate a large system in order to make it more trustworthy. Instead, it would be much more effective to be able to predictably compose a minimal system out of trustworthy components selected specifically for the intended maximum range of applications, and with the resulting assurance that it would be trustworthy (consistent with its requirements). Design tools for determining the composability of a collection of components would be extremely valuable.

- **Verification and validation**. Formal and semiformal methods have long been touted as having great potential for developing trustworthy systems and for increasing assurance. Benefits are now being reaped in the hardware industry by using formal analyses of chips and circuits, a lesson that was learned in response to the Intel Floating Divide flaw. Formal methods also have increasing applicability in subsystems and systems that have extremely critical requirements for security and reliability, with particular relevance to detecting and eliminating flaws and covert channels in multilevel security systems. Static analysis tools (some of which include model checking, some of which are more ad hoc) can also be very effective at detecting security flaws in algorithms and protocols. However, there is a need for carefully worked and carefully documented examples of the application of these techniques to the development of real systems, including multilevel secure systems and covert channel analysis.

- **Criteria and metrics**. Evaluation of system behavior depends on having suitable criteria and metrics against which systems can be evaluated. Metrics for reliable systems are fairly standard. Metrics for trustworthy systems that must be secure, survivable, and provide guaranteed service are speculative at best. (See Hard Problem 8.)

- **Evaluation and assessment**. Despite the historical progression from ad hoc development to evaluation criteria (from the DoD Trusted Computer System Evaluation Criteria to the Common Criteria), much is lacking: relatively complete protection profiles against which evaluation can be based, the discipline of developing systems to be evaluated throughout the development cycle rather than just when "completed" (especially relevant because systems are rarely if ever completed), the ability to overcome the inertia of bad development practice, the need for rapid reevaluation when systems undergo changes, the need for evaluation of applications rather than just operating systems, the need for evaluation of compositions of subsystems, the need for evaluating networks of systems, and overcoming many other difficulties that have hindered the evaluation process in the past. Perhaps most important is the need for continuing evaluations and assessments throughout the extended life cycle of critical systems.

The above enumeration barely touches on the tip of an enormous iceberg. For extensive background on these and related issues, see a recent report for DARPA by Peter G. Neumann, Principled Assuredly Trustworthy Composable Architectures.

Appendix F: Situational Understanding and Attack Attribution

Situational awareness has many aspects, of which some of the most important are noted here.

- **Real-time analysis**. When attacks and other forms of misuse are in progress, one of the main challenges is to be able to quickly derive an understanding of what is happening --- for example, which state information and which configurations have changed, whether multiple attacks exist and if so whether they are coordinated, whether current attacks might be correlated with earlier attacks, who or what entities are causing the attacks (including insiders and outsiders), what system resources and configurations are being used, and so on. The problems are further compounded by malicious users who can alter system audit trails, spoof IP addresses, and anonymize their identities or masquerade as other users, and by hostile or merely non-cooperative administrators in regions of untrustworthy servers and network components. These issues suggest the need for integrated techniques and tools (including analysis of system logs, application audit trails and network packets) that enable going much further into interactive diagnosis. Although research and development has been under way for over 20 years in this area, the existing commercial systems are still quite limited in their abilities. (Of course, documentation and training must accompany those tools.)

- **Cyber forensics**. Although it is often slighted in favor of real-time analysis, post-mortem analysis is also important. At least with respect to the corresponding tools, considerable commonality should exist between real-time analysis and post-mortem analysis, although various operational differences will of course exist. (A recent book on this subject is *Forensic Discovery* by Dan Farmer and Wietse Venema, Addison-Wesley, 2005.)

- **Configuration and network mapping**. Mapping techniques and tools are needed that produce accurate results in the presence of untrustworthy computer systems, untrustworthy networking components (routers, firewalls, etc.), and untrustworthy communications (e.g., media that are subjected to denial-of-service attacks, or that transmit data unencrypted), under varying assumptions. These tools should attempt to make explicit some of the primary tradeoffs among security, survivability, trustworthiness, traceability, and other requirements.

- **Traceback across domains**. One of the most difficult problems today is the inability to have any real assurance as to the identity and location of many remote users and hosts. (See Hard Problem 1 and Appendix B.) (Various graphical maps demonstrate the extraordinary complexity of Internet connectivity.) It may be desirable to architecturally partition networks into sub-networks of relatively comparable trustworthiness with respect to their ability to perform traceback accurately, and provide only carefully controlled connectivity among different sub-networks. Wireless connectivity can add further complexities. (See Tracking and forensic analysis in Appendix C for traceback within domains.)

- **Data fusion, aggregation, and correlation**. Assessing configuration information is relatively straightforward with respect to misuse by physical insiders, especially in the presence of meaningful authentication, dedicated access lines for physical insiders, and sensible system management. The situation is more complex for misuse by logical insiders who are physical outsiders, and depends heavily on the soundness of the user authentication, the cryptographic protection of the external communications. (See Hard Problem 2 and Appendix C.) The situation is potentially very difficult for misuse by outsiders who are both logical *and* physical outsiders, and whose identity and access routes may not be adequately known. (See Hard Problem 1.) However, the most difficult challenges in this area today seem to arise in the enormous amount of system log data and network packet data, picking through and abstracting out the relevant suspicious events, aggregating the resulting knowledge across attacks, users, and platforms, correlating the results, and determining appropriate responses. If multiple attacks are made on different systems and sub-networks, the correlation and cooperative management across different platforms is also a serious challenge, as is achieving a common operational understanding as seen from different vantage points. Automated tools for analysis must also provide incisive recommendations for reacting to crises. The hard problems include not just accurate situational analysis, but also taking appropriate countermeasures without overreacting and making the situations worse.

- **Visualization**. Understanding of the situational information noted above and its implications can be greatly aided by visualization tools that graphically depict the parameters of attacks as they change over time, and that allow administrators to drill down as far as needed to obtain the relevant details. This is another area in which research and tool developments are urgently needed.

Appendix G: Information Provenance

Information provenance is concerned with the original authorship of and subsequent changes to programs, data, and other computational objects. It is generally also indirectly concerned with content as well as ownership, because whenever ownership allows or delegates the ability for content to be created, altered, and deleted, provenance must also include which entities (for example, users or automated programs) have performed those operations, and when. Information provenance is relevant to the confidentiality and integrity of information; for example, it may be useful in analyzing information flow in the former case, and dependence on potentially untrustworthy programs and data in the latter case.

- **Granularity**. Granularity issues vary considerably depending on the environment. For example, classified documents require separate markings defining the classification level of each paragraph, figure, table, or any arbitrary unit (e.g., word or character), even if the document is stored as a single object. The same is true of the collection of data from sensors with different sensitivities and integrity. Furthermore, aggregations of information derived from multiple sources need to have the provenance of their constituent parts separately recorded, so that the accuracy and integrity of the aggregations can be meaningfully addressed. In conventional systems, modular subsystems generally need to be considered independently, even if they may be accessible only as complete systems.

- **Tracking**. In essence, information provenance provides a historical trace that must live at least as long as the objects that it encompasses --- and sometimes much longer. If an object is derived in part from others, those relationships should also be recorded, particularly in the case of data mining, where the integrity and accuracy of the alleged data sources present fundamental questions. Information provenance could be particularly useful in pinpointing the supposed origin of Trojan horses and contaminated data, although it is only as trustworthy as the chain of provenance itself. However, the lack of authenticated trustworthy provenance (or gaps in the chain of provenance) may itself be an indication of untrustworthiness. Information provenance is therefore also related to the consistent use of audit trails. In addition, privacy considerations (Hard Problem 7) may arise if the provenance details are themselves sensitive, as in the case of intelligence information.

- **Attribution and trustworthy maintenance of provenance**. Within a trustworthy system, the maintenance of provenance must also be trustworthy, assuming trustworthy code distribution, bilateral trusted paths between users and computational entities, and so on. Proper operational practice must also be maintained. However, provenance may not be trustworthy whenever information content traverses untrustworthy systems. Trustworthy determination of information provenance requires reliable and secure associations of provenance with information content, and encompasses a non-spoofable and non-alterable binding of content and entities, along with appropriate time stamps. For example, program source code and object code should contain a demonstrably high-integrity record of all changes once that code has been subjected to configuration control. Cryptographic integrity checks and digital certificates may be helpful in making the

content and the binding resistant to subversion. Trustworthy maintenance of provenance addresses a wide variety of threats, including forensic evaluations of violations of data confidentiality, data integrity, system integrity, proprietary content, and rights management.

- **Assurance**. Numerous assurance issues arise relating to information provenance, including all of those noted in the previous bulleted item and more. The primary question is this: How trustworthy is the provenance itself (including its binding to the corresponding objects)? How trustworthy the content is to which the provenance corresponds may seem to be a secondary question; however, if the provenance itself cannot be trusted, then neither can the content. For example, given the untrustworthiness of many systems and many users on the Internet, the trustworthiness of the identity of an entity represented in the provenance of an object may also be in question. (See Hard Problem 1.)

- **Long-term key management**. It is highly likely that some keys that must protect information secrecy for many years into the future will eventually be compromised. As a consequence, reliance on long-term keys must be considered as suspect. For example, cryptographic representations of provenance may themselves be compromised over time as well as the encryptions of the content to which they relate.

Appendix H: Security with Privacy

Threats to privacy can be rather insidious and stealthy, or painfully evident. Computer security (in particular, authentication, access controls, and accountability in terms of misuse detection and detailed auditing) as enforced today is generally too weak to prevent privacy violations within computer systems. Even in the presence of fine-grained discretionary access controls and possibly multilevel security levels and compartments, misuse by insiders and beneficiaries of outsourcing and offshoring remains a problem. Furthermore, many of the privacy violations transcend the computer systems, and are extrinsic; that is, they involve misuses outside of the purview of information security --- e.g., once information has been viewed or offloaded. Nevertheless, security technology offers many opportunities for enhancing privacy.

- **Mechanisms for anonymity and pseudo-anonymity**. True anonymity is a double-edged sword. There are cases in which it can be life critical (for example, for whistle-blowers and protected witnesses), but also cases in which it can be seriously misused (character assassination, harassment, physical attacks). Pseudo-anonymity provides a middle ground in which a person's identity is in essence escrowed under conditions that might require a warrant or other controlled process to determine. (Note that such escrow mechanisms themselves may be compromised, particularly in the presence of weak security controls and weakly administered security.) Cryptography, anonymizers, temporary aliases that can allow responses and other techniques are approaches relevant to privacy, although they must also satisfy the needs of identity management (Hard Problem 1).

- **Multiple identities**. Mechanisms for multiple identities such as aliases (also related to Hard Problem 1) must be coupled with authentication, security access controls, and accountability concerns. Closely related are the needs for different roles corresponding to different duties that an individual might perform, with corresponding access rights. These further complicate security, but enforcement and monitoring of the use of these roles is desirable in any event to help augment privacy enforcement.

- **Privacy-preserving data mining**. Techniques are known for analyzing sets of would-be queries with respect to inference and aggregation, although they have limitations in the absence of secure system environments. Furthermore, recent research is developing techniques by which certain queries can be made on encrypted databases without requiring decryption of the actual data and without revealing more content than necessary to fulfill the query.

- **Regulation and legal implications**. Although privacy policies are perhaps considered outside of the realm of technology, privacy is a meaningless concept in the absence of well-defined policies that explicitly state what is expected of individuals and organizations. Technological approaches to privacy enforcement must be established with respect to the relevant policies, and the limitations of those approaches must be sufficiently well understood. It is unwise to rely on legislation and regulation, but it is also unwise to believe that technology is by itself an adequate solution. A combination of approaches is essential.

- **Mechanisms for determining and enforcing data ownership**. Some combination of techniques such as watermarking, steganography, cryptographic integrity seals, and embedded provenance indicators (see Hard Problem 6) might provide a basis for dealing with data ownership.

- **Accountability for information use and privacy violations.** Some kind of auditing and tracking abilities may be required to be able to keep track of use of sensitive information and to identify serious privacy violations. However, such mechanisms themselves can create serious privacy problems, which must also be considered. (Accountability is of course also closely related to Hard Problem 6, information provenance.)

Other issues that also need to be considered include rectification of incorrect personal data, revocation of access privileges when necessary, and many issues associated with surveillance --- including spy gear, spyware, and other approaches to interception. Outsourcing of sensitive information is also a potential problem. Perhaps the biggest problem from the perspective of individuals is the enormous increase in identity theft and the indirect problems that it causes. For example, 73% of all contacts with the California Office of Privacy Protection during 2004 related to ID theft. (See Hard Problem 1 for needs for global identity management.)

Appendix I: Enterprise-Level Security Metrics

Metrics have proven extremely valuable for assessing concepts that are relatively easily quantified. Examples include hardware reliability as well as many enterprise-wide properties such as overall system safety and returns on investment. However, metrics for security have been much less satisfactory. Most of the existing metrics are of questionable utility, even with respect to individual software systems. Furthermore, meaningful security-related metrics for networked systems and for entire enterprises are significantly more difficult to define, evaluate, interpret, and use intelligently. Security properties of systems in the large and of enterprises as a whole are emergent properties that cannot be analyzed in the small (i.e., locally). Nevertheless, effective subsystem and system security metrics would be extremely desirable if they can be made more relevant to reality, and if they can be composable in the same way that subsystems should be predictably composable into systems. That is, the metrics should compose correspondingly.

Some of the problems that must be addressed before security metrics can have greater utility include the following.

- **Defining meaningful metrics**. Existing metrics are generally inadequate for purposes of evaluation of management decisions. One fundamental problem with such metrics is that the assumptions on which they are based are poorly founded or changing with time. Also, system developers tend to optimize systems and situations to satisfy the metrics, and evaluators tend to concentrate only on those metrics --- ignoring fundamental security flaws that are not encompassed by the metrics. The problems are compounded by measures that suggest, for example, that a system is *n* percent secure. This is clearly meaningless, because a slight change in assumptions can reduce perceived 99% security to 0% in an instant. For example, a published exploit that can be invoked worldwide renders a hitherto supposedly secure environment vulnerable to massive attacks. Furthermore, metrics that make assumptions about how many vulnerabilities remain undetected are generally suspect. On the other hand, biometric devices (for example) lend themselves to detailed measures of their accuracy, the frequency of false positives, and false negatives, and so on --- at least in the small. In the large, those measures may not be meaningful if the implementations of the biometric techniques are poorly embedded in systems that can easily be compromised, or if they simply do not work in enterprise-wide contexts (as in the case of would-be global security solutions that can be subverted by playback attacks or during communication outages). Simple enterprise-wide metrics can be informative to system administrators, such as those that track vulnerabilities and their disposition, as well as exploits and their prevention. Other metrics might also be helpful to management at a much higher abstraction layer, such as the ratio of detected misuse to criminal prosecutions, and the ratio of prosecutions to convictions. However, there are typically serious risks in over endowing the significance of particular metrics.

- **Composing metrics across enterprises**. The desire for enterprise-level security metrics suggests the need for being able to derive the evaluation of those metrics from the evaluations of finer-grained (e.g., single-system) metrics --- that is, through hierarchical composition of metrics from low levels to entire enterprises. Considering that

composition compatibility and interoperability are already unpredictable among requirements, specifications, software modules, subsystems, and networked systems, it is not surprising that composability of metrics is a difficult problem --- perhaps as a consequence of different and incompatible assumptions among different metrics.

- **Sharing of metric information**. Enterprise-wide metrics require timely communications of inputs for evaluation across different organizational boundaries. This can be a problem within a large enterprise, but is exceedingly difficult across different enterprises --- especially those that are competitors. For example, aggregation of information on enterprise threat models, system and enterprise security vulnerabilities, and reports of actual incidents of security exploitations is complicated by unwillingness to provide the requisite information, both within and across enterprises. (Note that provenance of shared metric information is also a relevant consideration, along with corresponding estimates as to the accuracy and integrity of the information.)

- **Real-time measurement**. Assuming metric information is available in a timely fashion, the importance of real-time measurements may depend on the enterprise. However, in reacting to detected security misuse, any delays can be costly. This is particularly true of firewalls and other perimeter defenses.

- **What should be monitored and what should be measured**? Anomaly and misuse detection systems over the past two decades have monitored system attributes, network packets, and other low-level measurables. Some of those analysis systems have also been applied to applications such as entire database systems and transaction processing systems (such as credit-card misuse detectors). Ideally, metrics should be dealing with concepts at appropriate layers of abstraction rather than trying to infer behavior from low-level data.

- **Audit and forensic tools**. Tools exist for both online and offline measurement and analysis. (The interpretation of the results of such tools is considered in Hard Problem 5 and Appendix F.)

- **Metrics for evaluating return on investment of security measures**. One of the biggest obstacles to greater security is the perception that commitment to security does not contribute directly to marketing's bottom line. Metrics, evaluation techniques, and supporting analysis tools would be particularly valuable if they can clearly demonstrate the long-term wisdom of security measures in enterprise contexts, and, contrarily, the short-term lack of wisdom from a security perspective that usually results from narrowly scoped optimization (e.g., based primarily on marketing pressures). Such metrics could also be used to demonstrate the return on investment of establishing such metrics in the first place with respect to achieving significantly greater enterprise-wide security.