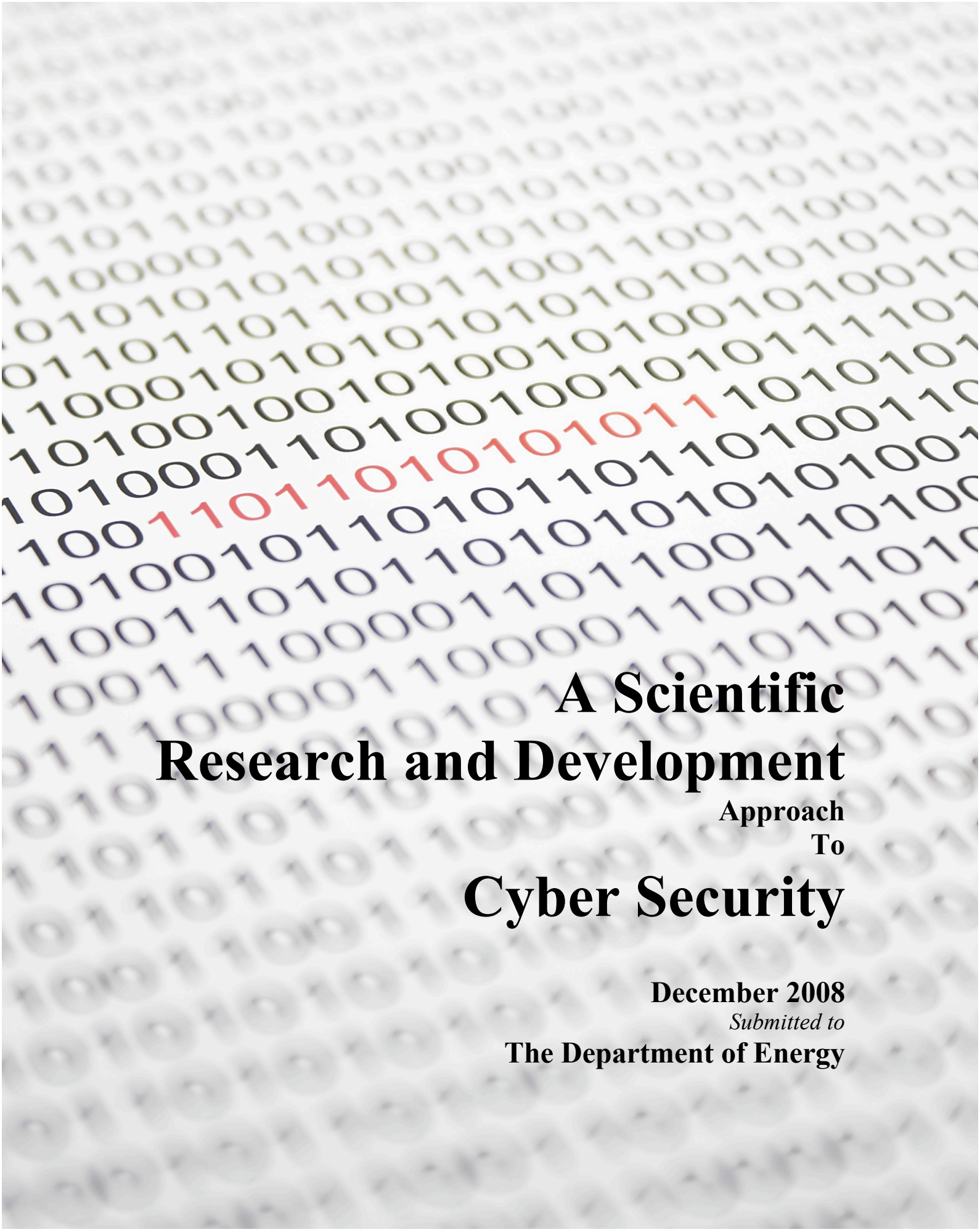

A Scientific Research and Development Approach To Cyber Security

December 2008

Submitted to

The Department of Energy

Cover Image:

Piedmont Photo

A Scientific Research and Development Approach to Cyber Security

December 2008

Submitted to the
Department of Energy

Abstract

The Department of Energy has the responsibility to address the energy, environmental, and nuclear security challenges that face our nation. In support of this mission, it operates national laboratories and scientific user facilities, performs basic and applied research and engineering, and works to assure reliable energy delivery and to maintain our nuclear deterrence capabilities.

Despite ubiquitous dependence on electronic information and on networked computing infrastructure, cyber security practice and policy is largely heuristic, reactive, and increasingly cumbersome, struggling to keep pace with rapidly evolving threats. Advancing beyond this reactive posture will require transformation in information system architecture and new capabilities that do not merely solve today's security challenges—they must render them obsolete.

The need is critical not only to the Department of Energy but also to other federal agencies and to the private sector. The Department of Energy is uniquely poised to undertake this work, complementing efforts at other agencies and industry.

Final Report

Submitted to the Department of Energy

On behalf of the Research and Development Community
(See *Acknowledgements*)

8 December 2008

Comments to:

Charlie Catlett, Argonne National Laboratory

catlett@anl.gov

Table of Contents

A Scientific Approach to Cyber Security **1**

Program Focus Areas **5**

Mathematics: Predictive Awareness for Secure Systems 6
Information: Self-Protective Data and Software 10
Platforms: Trustworthy Systems from Untrusted Components 14

A Science-Based Cyber Security Research Program **18**

Technical Organization and Management 18
Incentivizing Innovation 19
Relationship to Other Energy Programs 20
Relationship to Other Federal Agencies 20

Summary and Recommendations **22**

The Challenge: Obsolete Cyber Security Approaches 22
The Opportunity: Computational Science and Innovative Architecture 22
The Program: Incentivizing Innovation; Leveraging Science Programs 23

Acknowledgements **24**

References and Notes **25**

A Scientific Approach to Cyber Security

The Department of Energy has the responsibility to address the energy, environmental, and nuclear security challenges that face our nation. The Department relies on information and digitally based technology for every aspect of its mission, from operating national laboratories and scientific user facilities to performing basic and applied research and engineering, and from assuring reliable energy delivery to maintaining our nuclear deterrence capabilities. Much of the Department's enterprise involves distributed, collaborative teams; a significant fraction involves "open science," which depends on collaborations that must share significant amounts of information among institutions and over networks around the world.

The Department and its contractors produce millions of lines of new application-level and system-level software each year, while deploying substantial amounts of commercial and open source software and hardware systems. Energy infrastructures likewise involve complex interactions of the Supervisory Control and Data Acquisition (SCADA) system [1] and other digitally based devices. The operational and scientific work at the Department's national laboratories, plants, and scientific user facilities also create hundreds of thousands of data sets annually, ranging from fully open to highly classified, and varying in size from kilobytes to petabytes. The ability of the Department to execute its responsibilities depends critically on its ability to assure the integrity, confidentiality, and availability of the intellectual property embodied in its information systems and in the scientific, engineering, and operational software and data that support its mission.

Despite this ubiquitous dependence on information and networked computing infrastructure, the security of the Department's systems, software, and data is based on a largely heuristic, reactive and increasingly cumbersome methodology that struggles to keep pace with rapidly evolving threats. This situation puts at risk the Department's ability to ensure safe and secure operation of facilities and infrastructure, to protect and control vital information assets, and to engage in the open science research collaborations that are so essential for Department's success.

Innovation is needed in many areas—ranging from better authentication protocols to stronger encryption to better understanding of social and human factors. While some basic research is being done in these and other areas across the federal complex, the President's Information Technology Advisory Committee (PITAC) concluded that the paucity of investment in this area is "a crisis of prioritization." [2] The modest Federal investment in cyber security research and development is primarily focused on very long-range theoretical topics (such as NSF's roughly $100M program) or is classified (such as the Department of Defense programs) and thus not accessible for application to unclassified programs that comprise the majority of cyber security needs of the Department of Energy and indeed of our society as a whole.. Transformation toward sustainably secure infrastructure and operation within the Department of Energy and in our nation will require bridging research and

A Science-Based Approach

Significant, "game-changing" transformation requires a science-based approach that combines fundamental understanding with experimentation, theory, and modeling. The most successful scientific programs use peer review to maximize intellectual capital and prioritize research needs. The Department of Energy has applied this approach through programs such as SciDAC and ASCI, employing multidisciplinary teams, national-scale experimental facilities, and careful stewardship to create synergy between classified and unclassified sectors.

operation, classified and unclassified contexts, and public and private sector needs. The Department of Energy is unique in its need—and demonstrated ability—to effectively bring these together.

The Department is not alone in facing the cyber security issue and in the need to fundamentally reinvent current cyber security practice. As noted in August 2007 by the President's Council of Advisors on Science and Technology (PCAST), "Despite intensive efforts in government and the private sector in recent years to identify and patch vulnerabilities and to upgrade overall security, attackers continue to find new avenues for attack" [3]. PCAST concluded, "The ability to design and develop secure NIT (Networking and Information Technology) systems is a *national priority*" (see sidebar "President's Council of Advisors," p. 3). More recently, the National Science and Technology Council's September 2008 Federal Plan for Advanced Networking Research and Development [4] emphasizes the urgency of research and development in cyber security, stating that "special focus and prioritization are needed to respond to current national networking security concerns."

The protection and control of information within the context of the global, open, Internet are also essential for U.S. industry, where there is need to protect data across a spectrum from financial and strategic business data to proprietary engineering designs and processes. Recent cyber security breaches in the international financial sector, such as in the World Bank [5] and International Monetary Fund [6] systems further illustrate the widespread failure of today's cyber security strategies. As with open science, the entertainment and software industries have additional challenges with information assets that must be distributed—often internationally—in order to be valuable. Lacking effective methods for controlling the use of data products once they have been distributed, these industries rely on ineffective copyright laws and enforcement—at a loss of billions of dollars annually—or on overly restrictive and complex digital rights management schemes [7,8].

Cyber security has the objective of ensuring the *integrity*, *confidentiality*, and *availability* of information and information systems. These properties are balanced commensurate with the specific requirements of the information, information systems, and associated operations. The traditional approach focuses on a "layered defense," or "defense in depth," strategy in which the "crown jewels" are protected by walls and moats that form "air gaps" between the layers.

This layered defense approach to protecting assets is a key element of most defense systems, whether physical or cyber. However, the complexity of information systems and platforms is such that these tactics often introduce vulnerabilities that are not easily anticipated or addressed. As such, today's cyber security methods, policies, and tools have proved to be increasingly inadequate when applied to the exponentially growing scope, scale, and complexity of information systems and the Internet. Even in highly isolated implementations, the approach has been shown to be vulnerable to compromise by widely available technologies such as USB drives, increasingly powerful (and small) mobile computing devices, and wireless networks. The Department of Energy's mission requirements involving work with industry and with the open science community drive unique new cyber security needs due to the fact that associated information and activities cannot be completely isolated without rendering them ineffective to support the mission.

Vulnerabilities and new exploitations of today's approach to cyber security are identified daily. Increasingly sophisticated adversaries with significant resources, including organized crime and nation states [9], rapidly develop exploits to take advantage of these vulnerabilities. Concurrently, automated attack tools [10,11] have expanded the volume of malicious activities by lowering the level of expertise required to launch an attack. Typically, as new vulnerabilities emerge, new products, policies, and initiatives are introduced to reactively counter these exploits. The result of this reactive approach has ultimately been an ineffective posture characterized by a cycle of patching vulnerabilities, more often than not discovered by exploits of those vulnerabilities. The inevitable outcome is that some vulnerabilities will exploited before they are patched.

Rather than continuing to approach cyber security problems in a reactive fashion, using variations of the same tools and approaches, the Department must fundamentally re-examine its approach to cyber security by moving to a proactive posture, anticipating and eliminating vulnerabilities while also being prepared to effectively and rapidly defend against attacks. During the past two years, a growing community of cyber security professionals and researchers from the laboratories, private industry, academia, and other government agencies has conducted a series of workshops to assess the state of cyber security in general and within the Department of Energy specifically. The conclusion reached is that the Department should develop a long-term strategy that applies science and mathematics to develop information system architectures and protective measures that go beyond stopping traditional threats to *rendering both traditional and new threats harmless*.

The complexity, interconnectedness, and scale of information systems suggest that important lessons can be learned from similarly complex systems that require integrity, confidentiality, and availability. For example, a Department-sponsored workshop in May 2008 brought together academics, industry experts, national laboratory scientists, and policy makers to explore metaphors such as biological immune systems, ecosystems, and markets and risk management [12]. A key conclusion was that any effective approach to cyber security must address complexity at scale, necessitating the use of scientific tools and techniques appropriate for such complex systems.

Over six decades, the Department of Energy and its predecessors have employed science, mathematics, and technology to solve challenges that require fundamental understanding of large-scale, complex systems—ranging from climate and genomics to the development and maintenance of our nuclear deterrent. To date, however, this system-level science [13] approach has not been applied to information infrastructure and systems, their behaviors, or their vulnerabilities. Such work requires advanced algorithms, high-performance computation, large-scale data analysis, and, most critically, the ability to organize and sustain multidisciplinary, multiyear research efforts that maintain the long-term

President's Council of Advisors on Science and Technology (PCAST)
"A National Priority"

The current portfolio of Federal investments in CSIA R&D is too heavily weighted toward shorter-term projects and the development of reactive rather than preventative technologies. CSIA R&D should focus on developing the scientific and technological foundations for future-generation NIT (Networking and Information Technology) systems that are inherently more secure than current technologies. The higher-priority investments for CSIA should include R&D in:

- Comprehensive analysis of potential system-level vulnerabilities to inform the design of inherently secure NIT systems
- Generation of the fundamental building blocks for the development of secure NIT systems
- Usability and related social sciences, because progress in improving the security of NIT systems also involves altering user behavior

Recommendation: The Federal NIT R&D agencies should give greater emphasis to fundamental, longer term Computer Security and Information Assurance R&D and the infrastructure for that R&D. The Federal NIT R&D agencies should accelerate development of an R&D infrastructure for creating, testing, and evaluating new generations of inherently more secure NIT systems.

perspective required to anticipate challenges that continue to evolve over decades. These are strengths unique to the Department's national laboratories and, in particular, to the Office of Science, which supports basic research programs that have demonstrated significant long-term impact on the Department's mission (see, e.g., [14]).

Through programs such as the Advanced Strategic Computing Initiative (ASCI [15]) and Scientific Discovery through Advanced Computing (SciDAC [16]), the Department has transformed the practice of science beyond traditional approaches, leveraging exponential improvements in capabilities and uniquely talented multidisciplinary teams to address problems of a complexity that was previously inaccessible. The success of these programs positions the Department to reinvent cyber security through a similar strategy of transformational research and development. Such reinvention is essential and timely for the achievement of the Department's mission to protect and assure the integrity of its scientific and nuclear deterrence information resources and the nation's energy system and scientific capabilities. It will also have a profound positive impact on the nation's economic and national security.

Within a broader federal program, the Department's unique mission and capabilities create opportunity to address the Department's needs while providing new cyber security capabilities to other agencies and to the private sector. The program outlined below addresses three focus areas: (1) developing realistic, at-scale models that can be used to make faithful predictions about the security properties of complex information and infrastructure systems, (2) ensuring the integrity, confidentiality, and accessibility of mission-critical data, and (3) devising information and command/control platforms and systems that enable operational integrity even given the presence of untrusted components in a hostile operating environment. Such a program will provide a firm scientific foundation for designing and operating critical information and digitally based command/control systems and infrastructure.

Transformational capabilities such as those outlined in this report have application well beyond computing platforms, software, and information infrastructure. They address increasingly critical needs in areas including command and control, supply chain management, and regional and national electrical distribution grids. As has been a hallmark of scientific programs in the Department, the program outlined here encompasses the needs and priorities of both open and classified aspects of the Department's mission, emphasizing the balance of information (and information systems) integrity, confidentiality, and access.

President's Information Technology Advisory Committee (PITAC)
Cyber Security Research Priorities

1. Authentication Technologies
2. Secure Fundamental Protocols
3. Secure Software Engineering and Software Assurance
4. Holistic System Security
5. Monitoring and Detection
6. Mitigation and Recovery Methodologies
7. Cyber Forensics: Catching Criminals and Deterring Criminal Activities
8. Modeling and Testbeds for New Technologies
9. Metrics, Benchmarks, and Best Practices
10. Non-Technology Issues That Can Compromise Cyber Security

Program Focus Areas

Today's national cyber security needs are broad, ranging from better authentication and authorization technologies and protocols to improved cryptographic techniques, from improved understanding of the human factors that underpin too-often-successful social engineering attacks to new protocols able to protect wireless networks. Many federal programs and initiatives as well as multiple reports on information technology and cyber security have outlined a broad range of research focus areas to address cyber security. For example, the President's Information Technology Advisory Committee (PITAC) laid out ten priorities, shown in the sidebar at left, in 2005 [17], and the President's Council of Advisors on Science and Technology (PCAST) identified three major areas, shown in the sidebar on p. 3, for high-priority investment.

A national focus on cyber security research and development has emerged with presidential directives as well as with the multiagency "National Cyber Leap-Ahead Year" [18]. These initiatives will support both operational and research programs across the federal government, providing a broad focus across important research areas such as those listed by PITAC.

This report outlines three focus areas that that form an integrated program within this broader context, leveraging specific, unique strengths, infrastructure, and programs within the Department. The aim of this program is to transform the security and operational integrity of national assets and capabilities essential to the Department's mission.

Mathematics: Predictive Awareness for Secure Systems. Provide capabilities to examine system or network behavior to anticipate failure or attack, including real-time detection of anomalous activity and adaptive "immune system" response. This work will require deeper understanding of complex applications and systems, through data-driven modeling, analysis, and simulation of architectures, techniques, and processes.

Information: Self-Protective Data and Software. Create "active" data systems and protocols to enable self-protective, self-advocating, and self-healing digital objects. This work will tackle the critical problem of data provenance and related research to provide information integrity; awareness of attributes such as source, modification, traceback, and actors; and mechanisms to enforce policy concerning data confidentiality and access.

Platforms: Trustworthy Systems from Untrusted Components. Develop techniques for specifying and maintaining overall trust properties for operating environments and platforms, quantifying and bounding security and protection, integrity, confidentiality, and access in the context of a "system" comprising individual components for which there are varying degrees of trust.

Mathematics: Predictive Awareness for Secure Systems

The inherent interdependence and complexity of modern cyber infrastructures suggests that understanding and predicting behavior and performance at scale requires the application of mathematical and computational tools and techniques. The Department has well-established strengths in using very large scale simulation and modeling approaches across a wide range of scientific disciplines. Leveraging these strengths will enable significant advances in understanding the trustworthiness of complex systems, assessing the effectiveness of cyber defenses, and understanding situational threat, vulnerability, and mission risk.

During the past several decades, the adoption of computational science—simulation and modeling—has revolutionized many scientific disciplines. Ironically, computational science and high-performance computation have played a much more modest role in the fields of computer science and engineering, and almost no role at all in the design and management of information and energy infrastructures. Where models exist at all in these areas, they are relatively simplistic [19]. Where cyber security is concerned, virtually all of today's policies, techniques, and protective systems have evolved from trial and error rather than being based on an underlying set of models regarding individual components, systems of components, or complex and dynamic information infrastructures. In short, cyber security today is more of a craft than a science. Consequently, cyber security solutions often resemble prescientific approaches, where systems are frequently inflexible, overengineered, and fraught with unanticipated failure modes, and where it is impossible to reasonably forecast the impact of a modification or series of events.

A scientific basis for the design of trustworthy systems, proactive protection, and methods for understanding behavior under a variety of conditions, including failure modes, is essential if we are to move beyond defensive, ad hoc, expensive, and ultimately vulnerable cyber security practices. Mathematics, modeling, simulation, and data analysis are the means by which we can design trustworthy systems as well as predict behavior and anticipate attacks before they occur. Simply put, these tools will elicit the detailed picture necessary to create game-changing solutions to the cyber security problem [20].

Extant infrastructure models focus on critical components or representative core subsystems but do not provide an overall view. The criticality of understanding the behavior and vulnerabilities, at scale, of SCADA systems, electrical power distribution networks, and the distributed, network-connected information infrastructure of the Department's complex provide clear focus areas for application. From a strictly cyber security point of view, these techniques will be equally useful for understanding networks, the Internet, and malware behavior. The benefits of better models range from improved strategies underlying policy and priority decisions to forming the basis for building predictive capabilities.

Analysis of existing networks and malware, as well as the currently intractable scale of network sensor data that must be analyzed, will be fundamental to providing inputs to support cyber security and modeling and simulation. Such analysis will define a new class of high-performance computing (HPC) application problems, which the Department is well suited to address. A diverse set of analytical capabilities will be required to extract information (large-scale data mining and knowledge discovery from high-speed networks) from cyber observables ranging from logins to network traffic to software behavior. A scientific approach to cyber security requires development and application of innovative approaches to quantify, process, display, and

communicate existing, future, and potential threats. The complex cyber world poses numerous analysis challenges that must be addressed to collect, manage, store, process, integrate, and understand massive, heterogeneous, distributed cyber data [21].

The Department has the opportunity, expertise, and infrastructure to apply the tools and techniques of computational science, including high-performance computing and the analysis of petascale data, to revolutionize cyber security. As with other disciplines, the first steps involve models for fundamental building blocks—individual programs, operating systems, and computing platforms—followed by composite models involving systems of such components. For example, the medical field is pursuing a range of models concurrently, including fundamental components (folding proteins to cells), critical systems (blood flow, nervous systems, immune systems), and the interaction of organisms (epidemiology [22, 23], sociology). Similarly, an adequate framework for trustworthy system design and predictive awareness will involve a range of cyber security–related efforts including fundamental components (programs, malware, operating systems, platforms), critical systems (networks, SCADA and electrical distribution systems), and interactions (epidemiology, sociology).

Several key capabilities are absent in today's cyber security approach:

- Provable methods for quantifying trustworthiness and risk within a component or system of components.
- Computational models that capture expected behavior in software, platforms, and networks of systems such that failure, compromise, or vulnerable conditions can be detected in real time or even predicted.
- Techniques for performing and analyzing ensembles of scenarios to develop effective responses to various events and vulnerabilities, leading to the ability to predict outcomes to potential remedies during an event.
- Techniques for understanding the necessary and sufficient conditions required to restore trust and yet maintain functioning and usable systems.

Such capabilities give rise to the potential for proactive cyber security scenarios such as:

- Component and infrastructure immune systems that detect failures or attacks and implement appropriate responses (isolation or destruction of pathogens, self-healing of systems) [24].
- Infrastructure models to predict and prevent modifications—whether to software, platforms, or an organization's infrastructure—that would introduce vulnerabilities or increase risk of failure [25].
- Defense responses designed to render infections or attacks ineffective, such as immediately instantiating a quarantined, virtual copy of an organization's infrastructure to isolate and examine the nature and intentions of the intrusion, or creating a "hall of mirrors" effect with thousands or tens of thousands of virtual targets, making it impractical for the attacker to locate assets of interest.

The research areas described in the following sections provide the basis for these and other fundamentally new approaches to cyber security. These areas also motivate the development of useful tools for risk assessments to guide cyber security investment priorities and policies. The proposed mathematics not only will aid in reducing today's vulnerabilities but also will provide guidance and modeling capabilities that are essential for the development of a more secure Internet in the future.

Modeling and Simulation Challenges

A strong mathematical foundation is essential for mathematical models that are to be used for computational simulation of critically important infrastructure, such as computer networks. Models must be well posed, meaning that small perturbations in input data do not result in unbounded changes in the model state. Computability is an important issue as well. Issues such as the required fidelity, scalability, and acceptability of various approximations must be considered in the context of the requirements of the application as well as the capability of the computer on which the model must execute.

For cyber security applications, large-scale modeling, simulation, and emulation approaches can be used to understand the inherent structure and evolution of networks (information, SCADA, or electrical distribution), software systems, or other complex infrastructure at various scales and time resolutions. Such a capability can potentially be used to predict network behavior that is consistent with observed data and to discover emergent behavior of such complex systems [26]. Capabilities anticipated with such models include the following:

- Provide methods to support ongoing assessment and experimentation associated with development of a new defensive posture, technologies, or opponent capabilities.
- Quantify the robustness and survivability of platforms, systems, and networks to attacks, comparing various architectures, policies, or changes.
- Real-time or retrospective discovery of large-scale attack kinetics.
- Evaluate the probable effectiveness and pitfalls of particular defenses, remedies and recovery strategies in advance of deployment.
- Understand the impact on cyber security of new or proposed technologies, security measures, and network topologies.
- Model the impact of human and social dynamics on the morphology and growth of critical infrastructure, e.g. by quantifying vulnerabilities to social engineering attacks.
- Provide real-time support for red-teaming activities when studying and evaluating new cyber security measures.

Realistic-scale simulation of critical infrastructure, from the electrical power distribution grids to distributed software systems to networks of computers, also requires precise understanding of subsystems and components. Often the significant actors lie in the particulars of protocol stacks, operating systems, or firmware of individual components. Work is needed to understand the propagation of these particulars from the subscale and their contribution to the observed overall system behavior. Mathematics and algorithms in the dynamics of large-scale graphs and renormalization schemes must preserve the essential dynamics and illuminate the mechanisms.

Research challenges and questions in this area include the following:

- Developing the multiscale mathematics techniques required to faithfully reproduce the observed emergent cyber security behavior of the network as a whole while preserving the essential characteristics of the fine scale (e.g., single attached node).
- Developing mathematical characterizations of normal network behavior so that anomalies can be identified (e.g., that indicate an attack or expose a vulnerability).
- Leveraging the Department's strengths in HPC to create a large-scale network emulation capability that reproduces observed Internet behavior and can inform the construction of mathematics, algorithms, and models for cyber security [27].
- Using modeling and simulation to evaluate means for turning complexity and/or scale against the attacker such as

- obfuscation of the instruction set or architecture forcing each attack to be a custom creation and
- architecture of "deceptive" networks with continually changing topologies and addresses or using virtual machines to populate every IP address on a network, confusing intruders as to the whereabouts of real assets.

Data Analysis and Underlying Mathematics Challenges

Observation and measurement complement modeling and simulation as tools for understanding the behavior of complex information systems. A particular opportunity for cyber security is to use and enhance analysis tools in order to provide an advanced cyber situational awareness capability. The purpose of such a capability is to provide immediate detection of anomalous and potentially dangerous activity on cyber networks and on computing platforms on the networks. In particular, statistical modeling, machine learning, graph theory, and network analysis techniques will play important roles in the development of such a capability.

For cyber security applications, analysis of cyber data applied research and tool development can be used to provide the following:

- Real-time ability to distinguish between harmless anomalies and malicious attacks.
- Capabilities for automated detection, warning, response, prevention, and preemption.
- Detection of hidden information and covert information flows.
- Statistical approaches for exploration, characterization and analysis of cyber activity.
- Forensic analysis, traceback, and attribution of cyber incidents.
- HPC-enabled "software wind tunnel" test harness capabilities to perform exhaustive software regression and usage tests for discovery of cyber security vulnerabilities.
- Evaluation of risk and quantification of trust through statistical traffic analysis.

The research challenges in this area include the following:

- Developing machine learning and data-mining techniques that can operate in real time on massive amounts of highly heteroscedastic, nonstationary data to distinguish between harmless anomalies and malicious attacks.
- Developing graph-theoretical methods to accurately characterize and measure the structure of the Internet.
- Leveraging HPC to understand emergent network behavior only observable at scale.
- Advancing the state of the art in graph theory, graph theoretic analysis, abstract network theory, and large-scale simulation to understand the spread of malware or the effects of an infrastructure attack.
- Developing knowledge discovery techniques on graphs that use patterns of data flow between nodes to characterize network behavior.
- Developing extensions to graph theory to provide meaningful theoretical statements about large, time-varying graphs and associated network information flows in order to understand the time-varying structure and behavior of the Internet.
- Developing advances in robust optimization and game theory in order to understand emergent behavior and develop methods to control the network.
- Understanding how to balance the risks of potential threats with the impact and costs of cyber responses through the development of statistical approaches for evaluating risk and quantifying trust.

Information: Self-Protective Data and Software

The orderly progress of both science and society depends on correct inferences and judgments drawn from data. In contexts ranging from high-energy physics to the corporate boardroom, the intelligence community, and each of the Department's mission areas, it is essential that these inferences be drawn from data whose provenance is assured and whose quality is understood. Although making the right decision based on available data is challenging an even more difficult task is assuring that the data itself has not been compromised as it is extracted from the original source, digitized, transformed, interpreted, filtered, and combined.

Concurrently, the protection of classified, private, and operational data and software from disclosure, unauthorized modification, or destruction is critical to the Department's mission. Data and software are protected by active barriers such as firewalls, authentication and authorization schemes, and physical isolation—the "crown jewels" metaphor that assumes information is passive and fundamentally subject to contamination, destruction, or theft. This approach cannot keep up with the rapidly evolving cyber threat space. A significant transformation is required that makes data self-protecting rather than dependent on external protections. Such an approach would render today's cyber security threats irrelevant.

Part of the Department's mission is to engage in large science projects and to provide infrastructure for international collaborations. For example, Open Science Grid [28] is revolutionizing scientific collaborations by enabling internationally distributed teams to operate as a single, coherent entity. A key feature of these collaborations is the cooperation of entities that need to protect proprietary material while sharing essential collaboration artifacts; a second key feature is a lower level of trust than would exist in a single institution; and a third key feature is the fact that data is typically produced by one or more organizations, transformed by others, and merged and filtered by yet others, before it is ultimately used to make a scientific judgment. Today, mechanisms for tracking the provenance of such data throughout the workflow exist only in rudimentary form and in a few large projects—no general-purpose system is available. Moreover, despite advanced understanding of the issues in some technical communities—for example, "hierarchy of evidence" in the medical community and "standards of evidence" in the legal community—these notions are not embodied in current approaches to digital infrastructure. Thus, the quantification of confidence and provenance in data and the workflows that manipulate data is left to individual scientists and projects.

The critical challenge to information or data integrity, accessibility, confidentiality, or trustworthiness is to move from today's paradigm of passive data (that must be protected by external means) to active data that can:

- Detect and prevent unauthorized access or use.
- Recover from damage or manipulation, retaining information regarding the nature of the event and initiator.
- Present verifiable credentials regarding its origins and subsequent transformations.
- Execute defensive protocols to identify attackers and attack methods.
- Develop immunities through learning and communicating with peers.

The concept of self-protection also involves active self-maintenance of provenance, integrity, and chain of evidence, or "self-advocacy." The applicability of such an approach extends beyond

large international collaborations and open science projects to contexts such as SCADA systems, the intelligence community, political governance, and military systems. Similarly, software and entertainment industries need to distribute information products while ensuring that use and further distribution continue to comply with copyright and licensing rules.

To achieve these capabilities will require changing our conventional notion of data from passive objects to active self-healing entities, moving from reactive to proactive approaches by deploying automated defense mechanisms. Not only will breakthroughs of this program transform the way we protect infrastructure, discover new sciences, and collaborate internationally, but they will provide essential building blocks for digital rights management in digital commerce.

The Department's pursuit of exascale computing capabilities, combined with the use of petabyte-scale data as will be produced by the Large Hadron Collider (LHC), introduces the dimension of scale to this challenge. The size and complexity of the data associated with such initiatives are well beyond the capacity of existing integrity approaches. Moreover, the data is distributed or stored at great geographical and temporal distances from the point of origin, making large-scale data integrity a critical cyber security research issue. In some instances, such as data collected from experiments or via sensor networks, data is difficult or even impossible to reproduce. As the Department pursues petascale computing, scientists must be able to establish and manage scientific integrity and provenance of exabytes of scientific data such as data generated and used by the INCITE [29] programs and associated with the scientific user facilities.

Self-protective information capabilities must address at least three fundamental challenges: knowledge or proofs about how data was originally constructed or gathered, and a measure of its trustworthiness and reliability when originally produced; the ability to determine whether changes have occurred since construction or capture of the data and whether these changes are acceptable; and the ability to express and enforce policies concerning how both original data and derived data products can be accessed and distributed (see sidebar "Self-Protection," p. 12). To achieve these capabilities will require movement from current ad hoc (or nonexistent) approaches to techniques such as referencing chain of successive custody, sources and operations, incorporating notions of pedigrees and dependencies, and tracking (including distribution and potentially source attribution). Simply put, decisions are based on information, and thus attributes such as chain of custody and intermediate transformations are essential.

Mathematical techniques such as encryption and digital signatures will be essential, but they are not solutions in themselves. Indeed, their application today often exemplifies the current, inadequate approach involving active systems manipulating passive data. Self-healing and self-protecting capabilities must enable the data itself to maintain key properties and provenance information over time and at scale.

The potential for self-protective, trustworthy data to transform and accelerate scientific discovery within the Department is in part related to empowering teams, as is clearly recognized in the medical domain, where loosely coupled "virtual biotechs" supported by e-commerce infrastructure are developing treatments for rare diseases whose impact is below the threshold of investment for large pharmaceutical companies [30]. Increased trust in data is a key enabler of novel workflows and virtual collaborations that have the potential to increase the pace at which virtual program teams operate in all scientific domains.

Self-protective data with the ability to maintain provenance and chain of evidence is also essential as a basis for information assurance more generally. Transactions in the physical world usually are more trusted than their cyberspace counterparts in large part because of accountability—when

people break the law they can be held accountable. Today's Internet-based networking technology, which arose in a trust-based academic context, completely lacks accountability. By providing mechanisms for data provenance and integrity, we will be taking a foundational step toward accountability in cyberspace.

Research Challenges and Objectives

Self-protecting data systems will have at least three critical capabilities: attributes, access, and protection. Key attributes include origin and history (chain of evidence, transformations, etc.), whereby the data set maintains information and history. Access, such as is necessary to protect private and classified data, must be actively governed by the data in contrast to externally governed data access in today's systems. Further, with the exponentially decreasing cost of data storage, it is feasible to consider data that is indestructible in that it can be reconstituted without loss of attributes or privacy.

These capabilities will require advances in mathematics, protocols, and data as well as software systems, including the following:

- Mathematical techniques that support deriving integrity and integrity checks at the exascale level, for potentially broadly distributed exascale datasets, and for high-performance data streaming.
- "Self-protecting" or "active data"—data with the ability to maintain provenance and chain of evidence and to recognize when such data has been compromised.
- Robust, trustworthy methods for proving that representations of data—claims made by self-protecting data—are consistent with the underlying data.
- Methods for measuring the trustworthiness or reliability of data when originally produced.
- Methods for specification of rules for changes or transformations applied to data, for representation of such modifications, and for validation and verification.
- Mechanisms to capture and retain information for traceability and accountability, including as necessary identity and context (location, tools used, time, etc.)
- Methods for combining data provenance from a variety of sources, that allow for uncertainty of provenance and propagation of uncertainties (analyze data provenance from disparate sources).

Self-Protection: From Biological Systems to Data Sets

One approach to developing self-protecting data is to imbue data sets with certain active, lifelike properties—contrasting sharply with today's inert datasets. For example, one can apply the concept of DNA fingerprinting to enable data sets to maintain information relating to identity, provenance, and integrity. When data sets were fused together, they would inherit the genetics of their parents, enabling users to determine "paternity" or "maternity" all the way back to the ultimate sources of their inferences. Certain data sets would be genetically incompatible, providing new ways to detect and avoid the "mosaic" problem in which unclassified sources may be combined to produce classified results. In the same way that biologists now use fluorescence to mark proteins, it would be possible to mark and trace data as it moves through a workflow or chain of custody. Given distributed storage, one should be able to reconstitute a data set from a single sample of its DNA. A "living dataset" is self-organizing, knows who has a "need to know" and where it needs to be. Living data can evolve to find a physical niche where it is protected from predators.

A number of recent scientific advances provide the basis for building such systems. Efficient, statistically based network monitoring techniques that detect the presence of adversaries in a network have recently been developed at Princeton [S1]; digital watermarking can be used in a stealthy fashion on certain kinds of media [S2]; new information-theoretic and cryptographic techniques for countering Byzantine pollution attacks are being developed in the context of network-coded systems that potentially combine multiple streams of data [S3]; peer-to-peer systems like Microsoft's Avalanche are beginning to recognize the value of network coding and swarming to ensure the ubiquity of data [S4]; and ample research arising from DARPA's "active networks" program provide s insight into ways for data to carry code about itself. The challenge is to bring these results together in a working system in the service of science. The Department has many opportunities in this regard, ranging from scientific user facilities at the laboratories to the LHC community and Open Science Grid.

- Data storage, organization, and replication techniques to support recovery and self-repair of large-scale data sets, including "cloning" with reconstitution of attributes.
- Schemes for enabling flexible integrity and provenance-sensitive policies for applications, thereby ensuring that only data that meets predetermined standards is incorporated in a computation or presented in a display.
- Algorithms and techniques for real-time detection of unauthorized access or modification ("tamper-proof data"), for self-repair, and for triggering defensive (or offensive) actions such as beaconing and/or self-destruction.
- Methods of protecting sensitive data provenance information during data transformations—recognizing and preventing the "mosaic problem" whereby multiple unclassified or sensitive items combine to reveal classified or sensitive information.

Related Research

Research in the area of data integrity and data provenance for scientific computing is still in its infancy. Biba's Integrity Model [31] is arguably the most influential paper in identifying the core research issues. Integrity deals with improper modification of data and is an important companion attribute to provenance, which considers how the data has been generated and handled.

The open science community has made useful strides in considering provenance architectures by addressing challenge problems through workshops [32, 33]. Researchers have identified key challenges of data provenance and have developed prototypes that show promise and could be leveraged by the Department. Examples include:

- The Proof Markup Language, part of a semantic web-based knowledge provenance/justification infrastructure, supports attributes such as provenance, interoperability, and trust as part of a document's markup [34].
- The Open Provenance Model [35] and the Pedigree Management and Assessment Framework [36] have been proposed for representing scientific data provenance.
- Progress has been made in identifying the security-relevant characteristics of provenance data, separately from that of the data associated [37]. Markings about provenance can often be far more sensitive than the data itself, particularly in communities where sources and methods must be protected.
- Though typically implemented in a "passive" data context, the Tripwire approach [38] addresses file system and configuration integrity through identifying changes in files. Integrity violations in the file system provide a way to detect the activities of some categories of intruders and malware. Researchers have also begun to consider proactive approaches to maintaining integrity, such as self-securing storage [39].
- Reputation systems [40] could be used to provide assurances about data integrity and assertions about the validity of a particular chain of custody.
- VXA (Virtual eXecutable Archives) is an architecture for active archives, a method whereby decoders for compression strategies can be incorporated directly within an archive. This is particularly useful in areas such as multimedia, since it allows for evolution of the compression schemes themselves without losing the ability to work with legacy archives [41].
- Rather than labels or graphs, other evidence could support provenance and integrity advances. For instance, proof-carrying authorization approaches have been successfully extended to provide an alternative way of thinking about credentials, so that they are proven as logical claims rather than simple identity bindings [42].

Platforms: Trustworthy Systems from Untrusted Components

Integral to effectively addressing the integrity, confidentiality, and availability of information and data is the notion of trust with respect to the platforms and systems that create, move, manipulate, and house this information and data. The ability to identify and manage the information integrity and provenance of a dataset is inextricably tied to understanding the trustworthiness of manipulations performed upon data (or performed by data, in the case of actively resilient data).

Below the Waterline

The importance of trusted information platforms is illustrated by looking at the Basic Input-Output System, or BIOS, software that is used to start a computer. It is "burned" into a memory chip and soldered onto the computer's motherboard. The original BIOSes were simple and fit in 8 kilobytes; newer BIOSes are extraordinarily complex and use up to 16 megabytes [S5]. BIOSes now include device drivers, file systems, and network protocol stacks. Increased complexity makes it ever harder to verify that the BIOS is doing only what it should be doing. Researchers have noted that this enormous increase in size opens up new exploitation opportunities. Because the software to replace the BIOS resides in the BIOS itself, a compromise can include code that makes changes impossible, or only appear to succeed. And because the BIOS is below the control of the operating system, no software reload or even disk replacement can eradicate a BIOS virus. Further, the BIOS is physically attached to the motherboard and not field-replaceable.

Since these BIOSes are available only in binary form, it is practically impossible to assure that they have not been compromised [S6] somewhere in the supply chain. In other words, they can arrive already compromised. Since most supply chains are now entirely outside the U.S., there are many places outside our control where a BIOS can be compromised in a manner we cannot detect. It is straightforward to embed a virus in the BIOS that is not detectable on the operating systems or application levels. Indeed this possibility was illustrated when code was embedded in a commercial BIOS from 1999 to 2001, allowing the originators to gather usage data and potentially to take control – transparently and undetectably – of several million PC desktops [S7].

Platforms comprise many components from many sources, ranging from hardware to embedded firmware to software, and they operate within an untrusted or hostile environment. They are subject to malicious attacks, manipulation, policy conflicts and gaps, unplanned-for circumstances, mis-configuration and accidental failures. Currently it is impossible to understand precisely (within an acceptable tolerance) the trustworthiness of a software or hardware platform. That is, we lack a quantitative understanding of the likelihood that the platform can provide confidentiality, integrity, and accessibility commensurate with the mission supported by the platform. Determining the level of trust, where a Boolean answer is insufficient and impractical, is a challenge that extends from computing and information platforms to any real-world system, from aircraft to supply chains, from SCADA control systems for electric power grids or instruments control and management systems.

Simply put, we no longer operate single computer systems with simple peripherals. Today's platforms are distributed systems in their own right, running several proprietary operating systems, on different types of CPUs, and with multiple interconnect subsystems. The challenge encompasses not only complexity and scale but also the reality that any such system includes components whose internals (whether hardware or software) are opaque because of practical or legal constraints, or both.

The traditional focus on securing the operating system is thus necessary but not sufficient because the operating system is only one of many sources of vulnerability in a complex trust chain. Even if the operating system is secured, significant vulnerabilities remain underneath.

Of particular concern today is the presence or insertion of malicious components (hardware or software) whose aim is not to fundamentally disable or alter the operation of the platform but to introduce modifications that are inherently difficult to detect. These include unauthorized resource usage, subtle modifications sufficient to undermine proper operation (e.g. to produce plausible, but incorrect, results), or the exfiltration of critical information [43].

Research in this thrust area is motivated by a number of challenging questions:

- Can we design a composite platform such that failure or compromise of one component is isolated, protecting the overall platform?
- Can we extend software inspection and development tools to identify and correct commonly known security programming errors?
- Can we develop data-processing algorithms for parallel platforms such that a limited number of compromised nodes will not affect the integrity of the computations?
- Can we build desktop and server platforms such that an adversary connected directly to one of our core networks cannot cause damage or have access to protected data?

Addressing these challenges and quantifying trust for individual platforms, much less networks of platforms, will require a number of breakthroughs:

- Frameworks and languages for specifying and enforcing expected (and thus preventing aberrant) behavior and interaction among components, quantifying trust levels, and precisely understanding the impact on trustworthiness of introducing new components or platform modifications [44].
- Architectures containing one or more trust points whereby the platform trustworthiness can be bounded by securing a subset of software and/or hardware components.
- Algorithms and techniques that enable quantification of trust in scientific or operational (e.g., control) results derived from ensembles of platforms where a subset of platforms is known to be untrusted. Essentially, what is needed is a computational analog to Redundant Array of Inexpensive Disks (RAID) or secret-splitting techniques that enable trust despite failure of individual components below particular thresholds.
- Mechanisms for isolating trust within platforms, such as protecting mission-critical applications and data from operating systems or protecting operating systems from device firmware [45].
- Approaches to enable platforms to detect threat behavior of subsystems or components, initiating protective platform response such as isolating or disabling the offending device or program.
- Approaches to rapid, precise, and effective incident recovery to re-establish positive control with minimal collateral damage.

Applied Research Opportunities

A number of emerging technologies and approaches provide the basis for pursuit of these objectives, enabling, for example:

- Developing effective strategies for use of emerging new hardware protections that partition memory and I/O access (and similar strategies to disable vulnerable components such as DMA [46] on machines without this support). Upcoming hardware from Intel, AMD, and IBM will all support this capability.

- Compartmentalizing the operating system by exploiting virtualization technologies to create multiple machine partitions, compartmentalizing functions such as I/O (e.g., device drivers), and explicitly enforcing interaction limitations to eliminate the current need for built-in trust within the operating system and among its subsystems.
- Providing secure, limited-functionality operating systems and applications for the virtual machine manager that enforces access policy to the physical resources and virtual systems under its management. Research in secure operating system technologies that are applied in a well-defined confined context should result in higher integrity guarantees for the virtual machine hosting environments.
- Limiting operating system functions in lieu of enabling user programs to interact directly with device drivers, placing the operating system in the focused role authenticating and authorizing user applications with respect to device usage.
- Exploring a Mandatory Access Control model [47] beyond device drivers and operating systems, encompassing user applications and interfaces as well.
- Investigating trustable and verifiable security use of trusted hardware modules within individual components. These trust anchors are to provide the basis for attestation of the integrity of any software module layered on the actual hardware. The Trusted Platform Module [48] specification is an example, and with current Multicore CPU technology a potential implementation might be to dedicate a core or subset of cores for use as trusted arbiters, monitors, or enforcement agents.
- Deploying virtual machine isolation properties and hypervisor physical access management for security purposes. So far, the commercial deployment of virtual machine technologies has focused mainly on the virtues of resource sharing. Many security-related advances are possible through the use of virtual machines, such as compartmentalization of applications, fine-grained policy enforcement, and monitoring of applications inside of virtual machine instances [49].
- Exploring embedded software design and implementation for components whose untrusted nature (e.g., closed proprietary and hence unverifiable) affects the overall trust in the composite platform. One example would be an open source, verifiable or attestable alternative that provides higher integrity guarantees of that platform.

Structural Vulnerabilities: The Operating System and I/O

Operating systems run within a protection domain with greater privilege than user programs, with access to the entire machine, including user program memory. However, the basic assumptions made by most operating systems today are based on trust paradigms that no longer hold for modern platforms.

In traditional designs, an important role of the operating system is to govern the interactions between the CPU and I/O (e.g., peripherals, networks), where these devices are assumed to be "dumb" hardware under control of the CPU. In this design the operating system kernel implicitly trusts the software (drivers) operating these I/O devices, traditionally assumed to be from the same source as the OS itself.

Today, however, I/O systems include dedicated processors and are autonomous subsystems rather than under control of the CPU. Indeed, today's I/O devices frequently run complex, dedicated operating systems themselves. These autonomous I/O devices have direct access to read or write platform memory, outside of the control of the CPU and thus the protection of the operating system.

This disconnect between operating system and platform architecture means that many contemporary platforms are inherently insecure and can be readily compromised. An untrusted device has the capability, for instance, to carry a "back door" within its operating system that, upon receiving a certain sequence of packets, could scan memory, look for sensitive data, and send the data anywhere, at low bandwidth, as email or even http requests.

Example Platform Challenges

Current approaches to platform design emphasize performance and economics, with fundamental cyber security requirements either assumed to be imposed externally (e.g., via policy or access control) or included based on principles that assume a level of platform homogeneity and simplicity that no longer exists. Evaluating new techniques for creating trustworthy platforms might involve challenges such as the following:

- Design and implement the minimal/smallest, proven-secure open operating system that can function in the role of a virtual machine appliance, hypervisor, or embedded software operating system (e.g., BIOS) [50].
- Given a computational problem and a known correct answer, reproduce the correct answer given a platform of N nodes where a certain percentage is untrustworthy [51].
- Demonstrate an alternative architecture to the conventional CPU that writes/burns code directly to the FPGA to produce a useful/provable platform [52].
- Create a machine code disassembler that can correctly map the functions, execution paths, and known security holes (race conditions, buffer overflows, etc.) for a provided executable [53].
- Create a "security analyzer" for a compiler that can find known security problems in a program (or operating system).

A Science-Based Cyber Security Research Program

Sustained collaborative effort between cyber experts and scientists from other disciplines must drive the cyber security research agenda. Because of the range of sensitivities and the direct impact on the Department's mission from the standpoint of operations in general and with respect to computing resources, both the articulation and the execution of this agenda require active involvement of both the DOE laboratories and academic research communities. Peer review processes must be used to identify the best research ideas. Opportunities for dissemination of research results—through workshops, conferences, traditional publications, or online journals—will be an important consideration in engaging the open science community. Involvement of postdoctoral researchers and students in this effort will help build the pipeline of trained cyber professionals. Additional partnerships with forward-looking, innovative commercial hardware and software vendors may be necessary to fully address the cyber threat.

Technical Organization and Management

Creating a cyber security R&D initiative as outlined above will require an integrated set of programs to both address the underlying scientific challenges and foster experimentation with new approaches to cyber security that are revolutionary rather than evolutionary in nature. The Department's mission encompasses both unclassified and classified work, and thus there is a significant culture with embedded processes to manage the interplay between these needs through the national laboratories. Because cyber security is critical to both classified and unclassified mission needs, this capability positions the Department to play a unique and important role in cyber security research. Similarly, the interdependency between scientific research and the operation of mission infrastructure—from user facilities to materials handling plants to national laboratories—enables the Department to guide research based on operational requirements while shepherding the deployment and adoption of new concepts and techniques.

A strong program must engage a wide range of talented researchers, from both universities and DOE laboratories, to consider transformative approaches to cyber security. It must also ensure that promising new approaches are developed at a scale that permits realistic evaluation. Two organizational structures developed within the Department's SciDAC program may prove useful here. *Institutes* bring together researchers from many institutions to discuss innovative approaches to complex science and engineering challenges, and may also undertake developing and provide targeted testbeds to enable controlled experiments. *Enabling Technology Centers*—comprising scientists, applied mathematicians, application scientists, and engineers—research, develop, and demonstrate new approaches to complex science and engineering challenges.

A set of such centers tied with DOE programs such as the Small Business Innovation Research / Small Business Technology Transfer (SBIR/STTR, [54]) and Entrepreneur in Residence (EIR, [55]) programs for commercialization of results would address the need to move research and development results into practice in order to enact transformative change.

Analysis and provability of inherently secure architectures as well as predictive awareness require complex, large-scale modeling such as is characteristic of SciDAC's *Scientific Challenge Teams* and of ASCI and INCITE projects. These teams research, develop, and deploy advanced

computational modeling and simulation codes and new mathematical models and computational methods that can take advantage of petascale computers.

As with ASCI, INCITE, and SciDAC challenges, access to high-performance computational and associated data analysis resources—and expertise—will be critical. Indeed, many cyber security R&D projects will be ideal INCITE project candidates. However, as with other disciplines that currently lack an established critical mass of computational science work, the community will require assistance in moving from present methods (using small-scale simulations on PCs or modest clusters) to new methods capable of exploiting petascale systems.

Incentivizing Innovation

In the 21st century, innovation is no longer the exclusive domain of large organizations. Small, ad hoc teams can now self-assemble using Internet communication and coordination capabilities that simply did not exist a decade ago [56]. As a result, "game-changing" technologies and techniques are increasingly emerging from a worldwide pool of expertise [57].

Each of the research sections of this document includes examples of capabilities underscoring the fact that fundamental research must be done, that no obvious solution or approach exists today, and that a solution is conceivable nonetheless. The Department has the opportunity to leverage the innovation of today's academic, research, and commercial talent by carefully defining a set of target, disruptive capabilities to serve as challenges whereby individuals and teams from academia, industry, and national laboratories develop proofs-of-concept in competition for funding to pursue developing the capability.

The X Prize Foundation [58], modeled after the Ortiez Prize [59] won by Charles Lindbergh in 1927, is an example of such a program, with four such challenges undertaken over the past two years involving reusable manned spacecraft [60], gene sequencing devices [61], environmentally friendly vehicles [62], and lunar vehicles [63]. The Department has annually supported a similar project called *Challenge X* [64], in which university engineering teams compete to design environmentally friendly vehicles.

A key characteristic of these various challenge competitions is the specification of clear goals that can be objectively judged. This document outlines potential challenge areas that would lend themselves to similar competitions, ideally combined with the strengths of the Department's current approaches outlined below.

For example, with the goal of engaging the broadest research community to explore the greatest solution space, a multilevel challenge competition focused on cyber security might involve:

- Open competition for written descriptions of approaches.
- Selection of the top 25% of entries and award of planning grants to develop detailed designs and research objectives.
- Selection of the top 10% designs for prototype development.
- Award of the top 1-2 prototypes for creation of working system.

Relationship to Other Energy Programs

As noted throughout this document, cyber security research and development are increasingly critical to all aspects of the Department's mission. The Office of Science, NNSA, and EERE for example rely on secure operation of national laboratories. The scientific mission of the Department involves national- and international-scale collaboration on problems ranging from studying the human genome to climate and particle physics, and including access to and provision of international user facilities. The Department is also focused on energy and the environment, where a secure and reliable electric power grid represents a significant cyber security challenge.

The Department's investment in and reliance on high-performance and high-capacity computing and information infrastructure is longstanding, including the NNSA's Advanced Simulation and Computing (ASC) as well as the Quantification of Margins and Uncertainties (QMU) programs, and more recently the DOE Leadership Computing centers. The cyber security requirements of these initiatives encompass the secure operation of the facilities, the integrity and security of the information and software, and the ability of scientists to readily access these assets.

Relationship to Other Federal Agencies

Nearly every federal agency is involved in cyber security initiatives; and without exception these agencies require many—in some cases, most—of the advances laid out in this document. The Department's collaboration on key multiagency coordination teams to date provides opportunity to work to ensure complementary efforts and to avoid unplanned duplication. The diversity of agencies suggests that, within the research thrusts outlined here, there exist opportunities for collaboration. For example, the National Science Foundation conducts long-term research, primarily with the university community where, relative to the national laboratories, there are fewer natural opportunities to engage operational security experts on mission-critical infrastructure. Conversely, a number of agencies including the Department of Defense have national laboratories and mission-critical infrastructure, but the preponderance of classified requirements makes it more difficult to work with the open academic research community.

The Department of Energy has a unique posture that provides opportunity for complementary efforts with these and other agencies. The national laboratories are consistently strong in creating and supporting sustained research collaborations with universities, while also operating mission-critical infrastructure and supporting classified and unclassified programs. The assets also offer potential for bridging operational environments with unclassified research programs such as at NSF and DNS as well as classified programs at other agencies.

Each of the proposed research thrusts begins with fundamental effort that better defines the long-term objectives and success criteria, before moving toward increasingly well-defined and applied tasks working toward a long-term vision. Notably, none of the thrusts is intended to result in classified projects or classified byproducts. The projects are intended to support the direct DOE missions, but they also are intended to have broader societal applications. Each project will require close collaborations among academia, industry, and government. And, because the proposed program is long running, the role of the national laboratories becomes increasingly important as the repository of organizational knowledge

These characteristics suggest a framework for analyzing the relationship between this work and other agencies, as suggested in the table below.

Table 1: Comparison of Federal agency approaches relevant to cyber security research and development as described in this report.

	DARPA	NSF	DOD Labs	NIH	NSA, IARPA	DHS	DOE
Programmatic Orientation	Project	Project	Vision	Vision and Project	Project	Project	**Vision**
"Customer"	DOD	Society	DOD	Society & medical community	Intelligence community	National infrastructure	**Energy & Society**
National laboratory assets	-	-	Yes	Yes	Yes	-	**Yes**
Research horizon	Mid-term	Long-term	Long-term	Near, mid, and long-term	Near, mid, and long-term	Near, mid, and long-term	**Near, mid, and long-term**
Typical performers	Academia, Industry, Government	Academia	Academia, DOD Labs	Academia, Industry	Academia, Industry, Government	Academia, Industry	**Academia, Industry, Government**
Cyber security expertise	Yes	Some	Some	-	Yes	Yes	**Yes**
Classified approach to cyber security	Mostly	-	Some	N/A	Mostly	Some	**Flexible**

Summary and Recommendations

Every facet of the Department's mission relies on networked information technology, from the command and control of infrastructure such as the electrical power grid to petascale supercomputers that enable mathematical modeling at unprecedented fidelity. The challenges of ensuring the security of infrastructure and the confidentiality, integrity, and accessibility of information are illustrated by the fact that every federal agency has made cyber security a critical mission requirement and that multiple presidential advisory committees [2, 3, 4] have declared cyber security to be a *national priority* (see sidebars, p. 3 and 5).

The Challenge: Obsolete Cyber Security Approaches

Although the complexity of networks, software, and platforms has grown by many orders of magnitude in the past several decades, today's cyber security practice and policy remain essentially heuristic and reactive. We have few models with which to verify the effectiveness of security policies, nor adequate methods to extract knowledge and awareness from situational data. Current approaches to protecting and controlling digital information effectively disable its digital nature in order to reduce the problem to one of physical access, rather than exploiting that digital nature to create self-protection mechanisms. Platform architectures and operating systems rely on the principles developed for stand-alone mainframes three decades ago. Today, precisely 20 years after the Morris Worm [65], our network security architecture has not fundamentally changed. Hardware and firmware are implicitly trusted irrespective of source, and we continue to erect walls and insert gaps to protect passive data, with decreasing effectiveness and increasing cost.

The Opportunity: Computational Science and Innovative Architecture

This report outlines a set of opportunities for altering the very nature of cyber security. Advances in mathematics and computational science offer the possibility to create model-based tools that introduce anticipation and evasion capabilities to platforms and networks, data systems that actively contribute to their control and protection, and platform architectures that operate with integrity despite the presence of untrusted components. These motivate three research and development thrusts with the potential to provide new, game-changing capabilities to the Department, capabilities that are also directly applicable to other agencies, industry, and society.

Mathematics: Predictive Awareness for Secure Systems.

Goal: Provide capabilities to examine system or network behavior to anticipate failure or attack, including real-time detection of anomalous activity and adaptive immune-system response.

Research: Develop mathematical modeling techniques for complex information applications and systems, enabling data-driven modeling, analysis, and simulation of architectures, techniques, and optimal response to threats, failures, and attacks.

Information: Self-Protective Data and Software.

Goal: Create active data systems and protocols to enable self-protective, self-advocating, and self-healing digital objects.

Research: Develop techniques and protocols to provide data provenance; information integrity; awareness of attributes such as source, modification, trace back, and actors; and mechanisms to enforce policy concerning data confidentiality and access.

Platforms: Trustworthy Systems from Untrusted Components.

Goal: Develop techniques for specifying and maintaining overall trust properties for operating environments and platforms.

Research: Develop approaches for quantifying and bounding security and protection, integrity, confidentiality, and access in the context of a system comprising individual components for which there are varying degrees of trust.

The Program: Incentivizing Innovation; Leveraging Science Programs

To achieve the objectives outlined above will require sustained investment in a broad range of topics, guided by specific "challenge" capability objectives, on the scale of the SciDAC program in terms of funding and diverse participation from laboratories, universities, and industry. Such an effort must reach beyond traditional research and development programs, in which proven approaches are scaled up or problems are constrained to fit established methods. Innovative thinking and new architectures—involving risk and even failure—will be needed in order to affect the necessary transformation of cyber security.

It will be essential to ensure that research and development results are deployed in operational systems as they are proven in test and laboratory settings. Joint funding of research in highly advanced capabilities would enable both the Department and industry to explore areas with very high potential reward, sharing the associated risk. While joint funding works well with large companies, particularly those with extensive research organizations, the Department also has a number of successful programs that address commercialization (and thus operational availability and deployment) such as the SBIR/SBTT [57] and EIR [58] programs referenced earlier.

Acknowledgements

This report captures the work of many individuals from the Department of Energy national laboratories, other agencies, universities, and private industry. Contributors include scientists, operational technical experts, and executives who collaborated via open workshops held at national laboratories. Preliminary information that formed the basis for this report was also presented and discussed in a classified, inter-agency workshop with cyber security experts and policy-makers from 27 federal organizations.

Authors
Charlie Catlett, Argonne National Laboratory, *Editor*
Mine Altunay, Fermi National Accelerator Laboratory
Robert Armstrong, Sandia National Laboratories (CA)
Kirk Bailey, University of Washington
David Brown, Lawrence Livermore National Laboratory
Robert R. Burleson, Oak Ridge National Laboratory
Matt Crawford, Fermi National Accelerator Laboratory
John Daly, Los Alamos National Laboratory
Don Dixon, Texas A&M University
Barbara Endicott-Popovsky, University of Washington
Ian Foster, Argonne National Laboratory
Deborah Frincke, Pacific Northwest National Laboratory
Irwin Gaines, Fermi National Accelerator Laboratory
Josh Goldfarb, BBN Technologies
Christopher Griffin, Oak Ridge National Laboratory
Yu Jiao, Oak Ridge National Laboratory
Tammy Kolda, Sandia National Laboratories
Ron Minnich, Sandia National Laboratories (CA)
Carmen Pancerella, Sandia National Laboratory
Don Petravick, Fermi National Accelerator Laboratory
J. Christopher Ramming, DARPA
Chad Scherrer, Pacific Northwest National Laboratory
Anne Schur, Pacific Northwest National Laboratory
Frank Siebenlist, Argonne National Laboratory
Dane Skow, Argonne National Laboratory
Adam Stone, Lawrence Berkeley National Laboratory
Chris Strasburg, Ames Laboratory
Richard Strelitz, Los Alamos National Laboratory
Denise Sumikawa, Lawrence Berkeley National Laboratory
Craig Swietlik, Argonne National Laboratory
Edward Talbot, Sandia National Laboratories (CA)
Troy Thompson, Pacific Northwest National Laboratory
Keith Vanderveen, Sandia National Laboratories (CA)
Von Welch, NCSA, University of Illinois at Urbana-Champaign
Joanne R. Wendelberger, Los Alamos National Laboratory
Paul Whitney, Pacific Northwest Laboratory
Louis Wilder, Oak Ridge National Laboratory
Brian Worley, Oak Ridge National Laboratory

Workshop Participants

John P Abbott, Sandia Nat'l Lab
Deb Agarwal, Lawrence Berkeley Nat'l Lab
Mine Altunay, Fermilab
Aaron S Alva, Pacific Northwest Lab
Robert Armstrong, Sandia Nat'l Lab
Ron Bailey, NASA Ames Research Center
Tony Bartoletti, Lawrence Livermore Nat'l Lab
Jonathan Berry, Sandia Nat'l Lab
Brett Bode, Ames Lab
David L. Brown, Lawrence Livermore Nat'l Lab
Bob Burleson, Oak Ridge Nat'l Lab
Pat Burns, Colorado State Univ.
Charlie Catlett, Argonne Nat'l Lab
David Chavarria, Pacific Northwest Lab
Oliver Chevassut, Lawrence Berkeley Nat'l Lab
Benjamin K Cook, Sandia Nat'l Lab
Matt Crawford, Fermilab
Ron Cudzewicz, Fermilab
Jeffery E Dagle, Pacific Northwest Lab
Kimberly A Deitchler, Pacific Northwest Lab
Don Dickson, Texas A&M Univ.
Brent Draney, Lawrence Berkeley Nat'l Lab
Barbara Endicott-Popovsky, Univ. of Washington
Doug Engert, Argonne Nat'l Lab
Susan Estrada, Aldea
Ian Foster, Argonne Nat'l Lab
Deborah A Frincke, Pacific Northwest Lab
Irwin Gaines, Fermilab
Ann Gentile, Sandia Nat'l Lab
Joshua Goldfarb, BBN Technologies
Bonnie Green, Sandia Nat'l Lab
Christopher Griffin, Oak Ridge Nat'l Lab
John Grosh, Lawrence Livermore Nat'l Lab
Tom Harper, Idaho Nat'l Lab
Yu Jiao, Oak Ridge Nat'l Lab
Mark Kaletka, Fermilab
Chris Kemper, Oak Ridge Nat'l Lab
Himanshu Khurana, Univ. of Illinois at U-C
William Kramer, Lawrence Berkeley Nat'l Lab
Matt Kwiatkowski, Argonne Nat'l Lab
Craig Lant, Lawrence Berkeley Nat'l Lab
Mark Leininger, Fermilab
John Lowry, BBN Technologies
Tami Martin, Argonne Nat'l Lab
Celeste Matarazzo, Lawrence Livermore Nat'l Lab
John Matthews, Galois, Inc.
Deborah May, Lawrence Livermore Nat'l Lab
Jackson Mayo, Sandia Nat'l Lab
John McHugh, Dalhousie Univ.
Miles McQueen, Idaho Nat'l Lab
Juan Meza, Lawrence Berkeley Nat'l Lab
Scott Midkiff, Nat'l Science Foundation
Ronald G Minnich, Sandia Nat'l Lab
Richard Mount, Stanford Linear Accelerator Lab
Len Napolitano, Sandia Nat'l Lab
Michael North, Argonne Nat'l Lab
Christopher Oehmen, Pacific Northwest Lab
Carmen M Pancerella, Sandia Nat'l Lab
Joe Pato, Hewlett-Packard
Chris Poetzel, Argonne Nat'l Lab
Alex Protopopesva, Oak Ridge Nat'l Lab
Neil D Pundit, Sandia Nat'l Lab
Daniel Quinlan, Lawrence Livermore Nat'l Lab
Gene Rackow, Argonne Nat'l Lab
J. Christopher Ramming, DARPA
Anne Schur, Pacific Northwest Lab
Patty Schwindt, Colorado State Univ.
Frederick Sheldon, Oak Ridge Nat'l Lab
Frank Siebenlist, Argonne Nat'l Lab
Dane Skow, Argonne Nat'l Lab
Mike Skwerak, Argonne Nat'l Lab
Robin Sommer, Lawrence Berkeley Nat'l Lab
Joe St. Sauver, Univ. of Oregon
John Stewart, Cisco Systems
Adam Stone, Lawrence Berkeley Nat'l Lab
Chris Strasburg, Ames Lab
Richard Strelitz, Los Alamos Nat'l Lab
Forest E Jr Strycker, Pacific Northwest Lab
Denise Sumikawa, Lawrence Berkeley Nat'l Lab
Craig Swietlik, Argonne Nat'l Lab
Ed Talbot, Sandia Nat'l Lab
Ricky Tam, Sandia Nat'l Lab
Troy K Thompson, Pacific Northwest Lab
Aaron Turner, Idaho Nat'l Lab
Scott A. Vander Wiel, Los Alamos Nat'l Lab
Keith B Vanderveen, Sandia Nat'l Lab
John Volmer, Argonne Nat'l Lab
Scott VonderWiel, Los Alamos Nat'l Lab
Von Welch, NCSA, Univ. of Illinois at U-C
Joanne R. Wendelberger, Los Alamos Nat'l Lab
Greg White, Lawrence Livermore Nat'l Lab
Paul Whitney, Pacific Northwest Lab
Louis Wilder, Oak Ridge Nat'l Lab
Brian Worley, Oak Ridge Nat'l Lab
Dave Zachman, Mesa Networks

Classified Multi-Agency Workshop Participants

Classified (TS/SCI) workshop were held in August and September 2008 to discuss and review the thrust areas in this report and their applicability to classified requirements. The workshop brought together 48 participants from the following 29 organizations: AFCYBER, ANL, CIA, CND, DHS, DISA, DOD, DOE-IN, DOE-OS, DOE-JIACTF, FDIC, G2, IARPA, INEL, JIACTF, KCP, LANL, LLNL, NIARL, NIST, NSA, ODNI, ORNL, OSD/DoD, OSTP, PNL, SNL, State, and Treasury.

References and Notes

Sidebar References

[S1] S. Goldberg, D. Xiao, E. Tromer, B. Barak, and J. Rexford, "Path Quality Monitoring in the Presence of Adversaries," SIGMETRICS, 46, no. 1 (2008) 193-204.

[S2] D. Kirovski and H. Malvar, "Stealthy Audio Watermarking," US patent number: 7266697, 2007.

[S3] M. Kim, M. Medard, and J. Barros, "Counteracting Byzantine Adversaries With Network Coding: An Overhead Analysis," http://arxiv.org/abs/0806.4451, 2008

[S4] C. Gkantsidis, J. Mimller, and P. Rodriquez, "Comprehensive View of a Live Network Coding P2P System," in Proceedings of SIGCOMM 2006, pp. 177-188.

[S5] An example of how much can be stored in contemporary BIOS space is a complete X11 server: http://www.hermann-uwe.de/blog/linuxbios-with-x11-server-completely-in-flash-rom

[S6] Coreboot (http://www.coreboot.org/Welcome_to_coreboot) is an example of an open source BIOS effort.

[S7] Phoenix Technologies operated the "eBetween" subsidiary, announced in 1999, which used code in the PC BIOS to enable companies to "better reach, register and retain online users and subscribers" by offering options during PC startup, before loading the operating system. http://findarticles.com/p/articles/mi_m0EIN/is_1999_June_21/ai_54937342

Endnotes

[1] Supervisory Control And Data Acquisition (SCADA) systems are central to the operation of a wide range of infrastructure including utilities and, of particular relevance to the Department, electrical power grids.

[2] "Cyber Security: A Crisis of Prioritization," President's Information Technology Advisory Committee, February 2005.

[3] "Leadership Under Challenge: Information Technology R&D in a Competitive World; An Assessment of the Federal Networking and Information Technology R&D Program," President's Council of Advisors on Science and Technology, August 2007.

[4] Federal Plan for Advanced Networking Research and Development, National Science and Technology Council, Interagency Task Force on Advanced Networking. September 2008. http://www.nitrd.gov/

[5] "Cyber Security Questions Persist at World Bank," Fox News, 2-Nov-2008.

[6] "Cyber-Hackers Break into IMF Computer System," Fox News, 14-Nov-2008.

[7] "Sony BMG's Copy Protection Incites Global Controversy," Dow Jones Factiva, 19-Nov-2005.

[8] "Wal-Mart Latest Store to Shut DRM Key Servers," Ars Technica, 28 Sept. 2008.

[9] SANS Institute. "Top Ten Cyber Security Menaces for 2008," January 2008. http://www.sans.org/info/22218

[10] Markoff, John. "Thieves Winning Online War, Maybe Even in Your Computer," New York Times, 6 December 2008. http://www.nytimes.com/2008/12/06/technology/internet/06security.html?_r=1&emc=eta1

[11] Epstein, Keith. "U.S. Is Losing Global Cyberwar, Commission Says," Business Week, 7 December 2008. http://www.businessweek.com/bwdaily/dnflash/content/dec2008/db2008127_817606.htm

[12] Karas, T. H., J. H. Moore, and L. K. Parrott. "Metaphors for Cyber Security," Sandia Report SAND2008-5381, August 2008.

[13] Foster, I., and C. Kesselman, *Scaling System-level Science: Scientific Exploration and IT Implications.* IEEE Computer, 2006(November): p. 32-39.

[14] Brown, David L., et al., "Applied Mathematics at the U.S. Department of Energy: Past, Present and a View to the Future," http://www.sc.doe.gov/ascr/ProgramDocuments/Docs/Brown_Report_May_08.pdf/, 2008.

[15] Advanced Strategic Computing Initiative (ASCI). http://www.nitrd.gov/pubs/bluebooks/2000/asci.html

[16] Scientific Discovery through Advanced Computing (SciDAC), http://www.scidac.gov/

[17] "Cyber Security: A Crisis of Prioritization," President's Information Technology Advisory Committee, February 2005.

[18] National Cyber Leap Year, http://www.nitrd.gov/leapyear/, 2008.

[19] As noted by a March 2003 workshop, "Scalable Cyber-Security Challenges in Large-Scale Networks: Deployment Obstacles," organized by the Interagency Working Group for IT R&D, "…meteorology provides proof that complex, evolving, large-scale systems are amenable to mathematical analysis and that the network-security community need not necessarily restrict itself to the (probably oversimplified) models now in the literature."

[20] Ryan, Peter Y. A., Mathematical Models of Computer Security, Lecture Notes in Computer Science, Springer, Berlin, ISSN 0302-9743 (Print) 1611-3349, (Online) 2001, http://www.springerlink.com/content/31868dj6te26etev/

[21] Moitra, Soumyo D., and Suresh L. Konda, "A Simulation Model for Managing Survivability of Networked Information Systems," Technical Report CMU/SEI-2000-TR-020, Software Engineering Institute, December 2000, http://www.sei.cmu.edu/pub/documents/00.reports/pdf/00tr020.pdf

[22] Hofmeyr, S., and S. Forrest, "Architecture for an Artificial Immune System," Evolutionary Computation 7(1), Morgan-Kaufmann, San Francisco, 2000, pp. 1289-1296.

[23] Dasgupta, D., and S. Forrest, "Novelty Detection in Time Series Data Using Ideas from Immunology," in Proceedings of the International Conference on Intelligent Systems, 1999.

[24] Adaptive and Resilient Computing Security Workshop, Imperial College, London, Department of Electrical and Electronic Engineering, 27 September 2007, http://www.santafe.edu/events/workshops/images/5/5a/Arcs2007-cfp.pdf

[25] You-lei, C., and S. Chang-xiang, "A Security Kernel Architecture Based Trusted Computing Platform," Wuhan University Journal of Natural Sciences, 10, no. 1 (January 2005), 1-4, http://www.springerlink.com/content/0t82282377t07u0m/

[26] Sole, Ricard V., and Sergi Valverde, "Information Transfer and Phase Transitions in a Model of Internet Traffic," Santa Fe Institute, New Mexico, and Complex Systems Research Group, Department of Physics, FEN.

[27] Park, Kihong, and Walter Willinger, "Network Dynamics, The Internet as a Complex System: Scaling, Complexity, and Control," http://www.santafe.edu/research/publications/bookinforev/ics.php

[28] Pordes, R., et al., "The Open Science Grid," in Proceedings of the Scientific Discovery through Advanced Computing (SciDAC) Conference, Boston, 2007.

[29] The Department of Energy *Innovative and Novel Computational Impact on Theory and Experiment* (INCITE) program is managed by the Office of Science Advanced Scientific Computing Research (ASCR) program. INCITE was established in 2003 and provides access to high-performance computing resources on a peer-reviewed basis. http://www.er.doe.gov/ascr/incite/

[30] CollabRx,www.collabrx.com

[31] Biba, K. J., Integrity Considerations for Secure Systems, Technical report, 1977, http://handle.dtic.mil/100.2/ADA039324

[32] Moreau, L., J. Freire, J. Futrelle, R. McGrath, J. Myers, and P. Paulson, The Open Provenance Model," Technical Report, ECS, University of Southampton, 2007, http://eprints.ecs.soton.ac.uk/14979/

[33] Bose, R., Ian Foster, and Luc Moreau, "Report on the International Provenance and Annotation Workshop (IPAW06)," Sigmod Records, 35(3):51–53, September 2006.

[34] Pinheiro da Silva, Paulo, Deborah L. McGuinness, and Richard Fikes, "A Proof Markup Language for Semantic Web Services," Information Systems, 31, no. 4-5 (June-July 2006) 381-395, (http://www.inference-web.org/).

[35] Munroe, S., V. Tan, P. Groth, Sheng Jiang, S. Miles, and L. Moreau, "The Provenance Standardization Vision," 2006. See also http://provenance.org and http://twiki.gridprovenance.org/bin/view/Provenance/OpenSpecification

[36] Architecture Technology Corporation, http://atcorp.com/Research/pmaf.html

[37] Braun, Uri, Avraham Shinnar, and Margo Seltzer, "Securing Provenanc," in Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec), San Jose, CA, July 2008.

[38] Kim, G. H., and Spafford, E. H., "The Design and Implementation of Tripwire: A File System Integrity Checker," in Proceedings of the 2nd ACM Conference on Computer and Communications Security, ACM, New York, 1994, pp. 18-29.

[39] Strunk, J., Goodson, G., Scheinholtz, M., Soules, C., and Granger, G., "Self-Securing Storage: Protecting Data in Compromised Systems," in Proceedings of the 4th Symposium on Operating Systems Design and Implementation. San Diego, October 2000.

[40] Resnick, P., K. Kuwabara, R. Zeckhauser, and E. Friedman, " Reputation Systems," Commun. ACM 43, 12 (Dec. 2000) 45-48.

[41] Ford, B., "VXA: A Virtual Architecture for Durable Compressed Archives," USENIX File and Storage Technologies, December 2005.

[42] Lesniewski-Laas, C., B. Ford, J. Strauss, R. Morris, and M. F. Kaashoek, "Alpaca: Extensible Authorization for Distributed Services," in Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS-2007), 2007.

[43] See SANS analysis at http://preview.tinyurl.com/2sf38b

[44] He, R., Jianwei Niu, and Guangwei Zhang, "CBTM: A Trust Model with Uncertainty Quantification and Reasoning for Pervasive Computing," by http://www.springerlink.com/content/f3874512680q6926/

[45] Claude, Benoit, Noel De Palma, Renaud Lachaize, and Daniel Hagimont, "Self-Protection for Distributed Component-Based Applications," http://www.springerlink.com/content/n415wv2245166115/.

[46] Direct Memory Access, or DMA, is a common I/O technique that, by virtue of implicit trust by the operating system, creates inherent vulnerabilities in the event that the I/O device driver is untrusted or vulnerable.

[47] http://en.wikipedia.org/wiki/Mandatory_access_control

[48] http://en.wikipedia.org/wiki/Trusted_Platform_Module, https://www.trustedcomputinggroup.org/groups/tpm/

[49] O'Rourke, Patrick, "Extending VMs," http://blogs.technet.com/virtualization/archive/2008/04/07/isolation-of-vms.aspx

[50] For example, Embedded Real-Time Operating Systems (ERTOS) http://ertos.nicta.com.au/ .

[51] Gilbert, S., Rachid Guerraoui, and Calvin Newport, "Of Malicious Motes and Suspicious Sensors: On the Efficiency of Malicious Interference in Wireless Networks," in Principles of Distributed Systems, Springer, 2006.

[52] For example, MicroBlaze. http://en.wikipedia.org/wiki/MicroBlaze

[53] For example, the ROSE project: http://www.rosecompiler.org/

[54] SBIR/SBTT: http://www.science.doe.gov/SBIR/

[55] DOE's "Entrepreneur in Residence" program, initiated in October 2007, involves DOE Laboratories hosting individuals from start-up companies to identify technologies for commercialization in the private sector. http://www.energy.gov/news/5661.htm

[56] Shirky, Clay, Here Comes Everybody: The Power of Organizing Without Organizations, Penguin Press HC, 2008.

[57] Friedman, Thomas, The World Is Flat 3.0: A Brief History of the Twenty-first Century, Picador, 2007.

[58] X Prize Foundation, http://www.xprize.org/

[59] The Ortieg Prize was a $25,000 reward offered for the first non-stop trans-Atlantic flight between New York and Paris, http://en.wikipedia.org/wiki/Orteig_Prize

[60] Ansari X Prize, http://en.wikipedia.org/wiki/Ansari_X_PRIZE

[61] Archon X Prize, http://en.wikipedia.org/wiki/Archon_X_PRIZE

[62] Automotive X Prize, http://en.wikipedia.org/wiki/Automotive_X_PRIZE

[63] Google Lunar X Prize, http://en.wikipedia.org/wiki/Google_Lunar_X_PRIZE

[64] ChallengeX, http://www.challengex.org/

[65] Spafford, Eugene, "The Internet Worm Program: An Analysis," Purdue Technical Report CSD-TR-823, 1988. http://homes.cerias.purdue.edu/~spaf/tech-reps/823.pdf