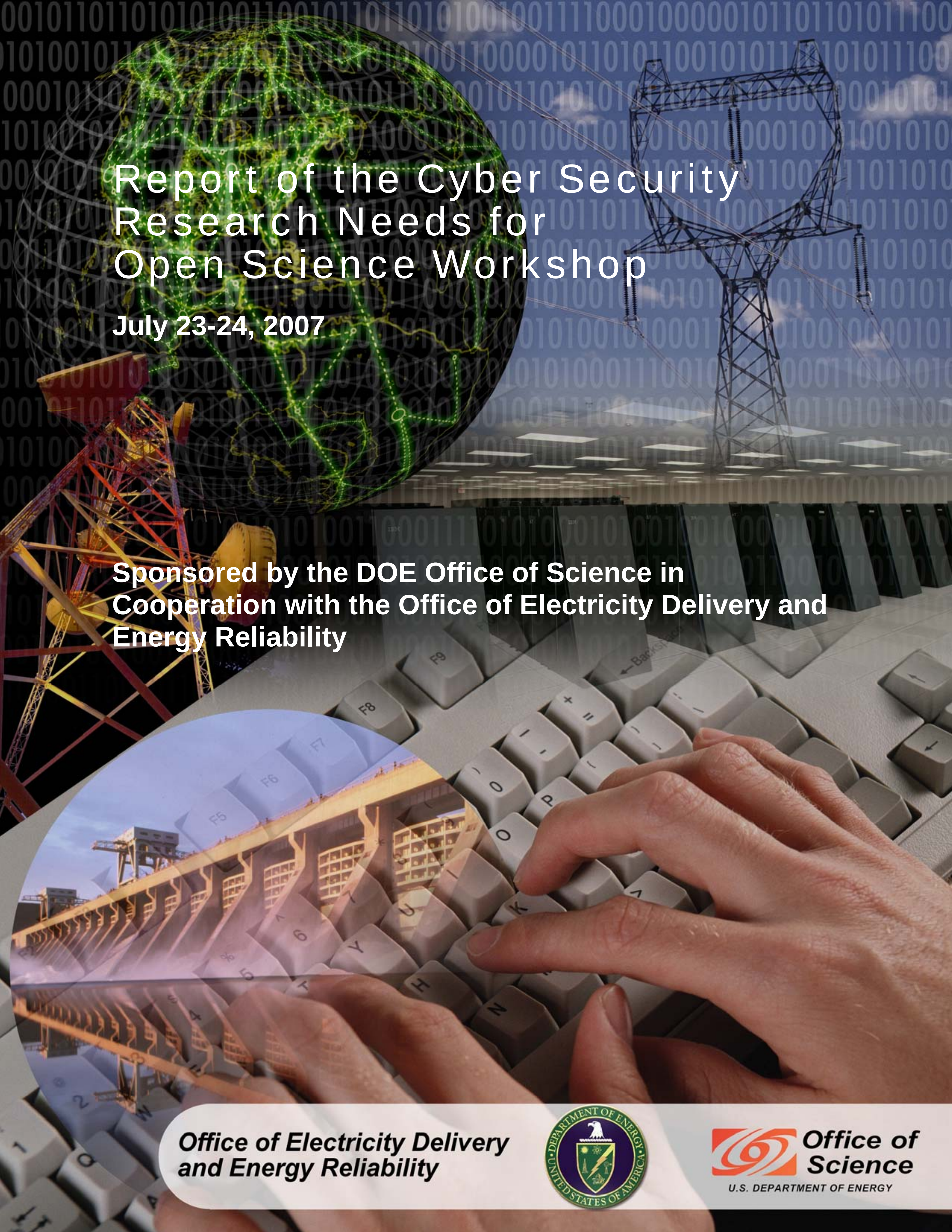


Report of the Cyber Security Research Needs for Open Science Workshop

July 23-24, 2007

Sponsored by the DOE Office of Science in Cooperation with the Office of Electricity Delivery and Energy Reliability

Office of Electricity Delivery and Energy Reliability

Office of Science
U.S. DEPARTMENT OF ENERGY

PNNL-16971

Report of the Cyber Security Research Needs for Open Science Workshop

July 23-24, 2007

Sponsored by the DOE Office of Science in Cooperation with the Office of Electricity Delivery and Energy Reliability

Acknowledgements

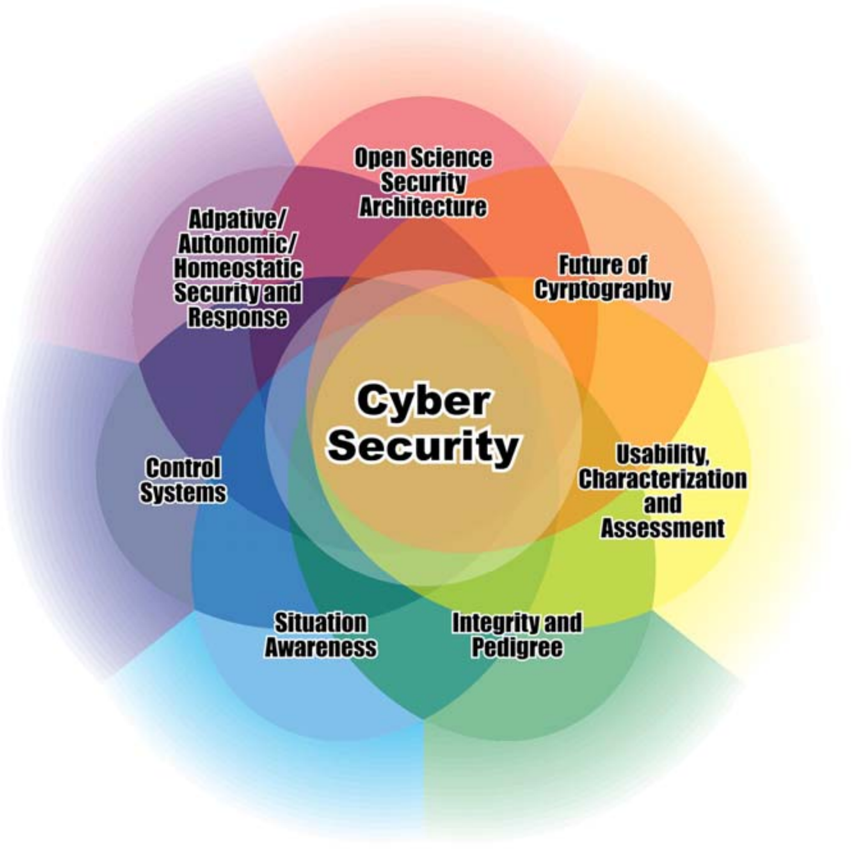
The workshop chairs wish to thank Joree O’Neal and Rachel Smith for all their help and support with organizing the logistics and registration activities for this workshop; Sue Chin, Ted Tanasse, Barbara Wilson, and Stacy Larsen for their expert help with the assembly, text editing, and graphics for this report; and Lance Baatz for his masterful job as webmaster for this project. Finally, we wish to thank the Pacific Northwest National Laboratory for supporting the development and assembly of the final report.

EXECUTIVE SUMMARY

Protecting systems and users, while maintaining ease of access, represents the "perfect storm" of challenges in the area of cyber security.

A fundamental tenet of every scientific investigation is the search for truth. Trust and integrity are immutable values that are fundamental to all research endeavors. By extension, the absolute integrity of information, encompassing systems, networks, and file systems is a critical requirement essential for promoting and facilitating discovery. These qualities are equally important for our nation's critical energy infrastructure. To determine the cyber security research needs for open science and energy control systems, the U.S. Department of Energy's (DOE) Office of Science (SC), in cooperation with the Office of Electricity Delivery and Energy Reliability (OE), organized a two-day workshop held July 23 and 24, 2007, at the Bethesda North Marriott Hotel in Bethesda, Maryland. This event brought together leading cyber security and infrastructure protection experts and researchers from academia, government laboratories, and industry.

This group of experts identified **seven long-term, priority research long-term thrust areas** that would provide a significant technological capability impact on the assured integrity of the nation's critical energy and science infrastructure. **The research thrust areas are: 1) Future Open Science Security Architecture; 2) Research for the development of Adaptive, Autonomic, and Homeostatic Security and Response systems; 3) Tools for Situational Awareness; 4) Integrity and Pedigree; 5) Analysis Tools for Systems Design for Usability, Characterization, and Assessment; 6) Control Systems Designs and Systems Assurance Capabilities; and finally (7) Future of Cryptography capabilities.** This reinforcing and synergistic research thrust approach will establish the scientific basis for transformation of our cyber security capabilities for the future.


Open Science Security Architecture
Adpative/ Autonomic/ Homeostatic Security and Response
Future of Cyrptography
Cyber Security
Control Systems
Usability, Characterization and Assessment
Situation Awareness
Integrity and Pedigree

To promote and facilitate the development of future systems that are intrinsically secure, the research defined in these thrust areas will mandate fundamental changes in computer and systems architectures; operating systems and programming models; large-scale data management, analytics, and visualization; real-time network and traffic analytics; mathematics research for complexity, discrete analysis, and multi-dimensional graph theory; applied statistics for anomaly detection and network-based behavioral analysis ; and finally, research to develop flexible and scaleable approaches for data fusion and analysis to enable real time co-operative network analytics. **Promotion of these research thrusts at a programmatic level will provide new capabilities for testing and implementations in both classified and open research environments.**

For this research to enable systemic and enduring change in the nation's critical systems for both energy and science, a coordinated, collaborative effort among academia, government laboratories, and industry is necessary. Appropriate off-the-shelf technology is unlikely to address the broad base and unique applications of the science research enterprise without such an effort. **At the core of these research thrusts is the recognition that now is the time to evolve beyond our current strategy that relies on forensics in a "catch-and-patch" approach to a strategy where cyber security is designed into critical systems and throughout all our cyber assets.**

Improving cyber security is a complex, especially daunting task in the Open Science environment, and there are obvious commonalities with the Energy Control Systems environment. But with a focus on scalability and flexibility in the research directions, the chairs and the attendees of the workshop believe that significant forward progress toward creating both usable and secure environments can be made in the next 5 to 10 years. The research directions suggested in this report will inspire new cyber security safeguards for both Open Science and Energy Control Systems.

TABLE OF CONTENTS

EXECUTIVE SUMMARY iv

INTRODUCTION 1

WORKSHOP LOGISTICS 1

CROSSCUTTING FINDINGS AND CHALLENGES 2

PANEL FINDINGS AND PROPOSED RESEARCH 4

CYBER SECURITY CHALLENGES 6

THRUST AREAS 12

CYBER SECURITY DEFENSE TAXONOMY 13

PRIORITY RESEARCH DIRECTIONS 14

Thrust Area 1: Future Open Science Security Architecture 14

Thrust Area 2: Adaptive/Autonomic/Homeostatic Security and Response 18

Thrust Area 3: Situational Awareness 20

Thrust Area 4: Integrity and Pedigree 23

Thrust Area 5: Usability, Characterization, and Assessment 26

Thrust Area 6: Control Systems 29

Thrust Area 7: Future of Cryptography 31

SPECIFIC RECOMMENDATIONS FOR A DOE CYBER SECURITY RESEARCH AGENDA 33

CONCLUSION 35

Appendix A – Workshop Organizers A.1

Appendix B – Workshop Participants B.1

Appendix C – Workshop Charge C.1

Appendix D – Panel Breakout Sessions and Reports D.1

FIGURES

1. Cyber Security Research Thrust Areas.. 5
2. Motivation for Enhanced Cyber Security. 6
3. Energy Control Systems and Our Critical Infrastructures 8
4. How Cyber Attacks Occur 9
5. Taxonomy of Cyber Security 13
6. Understanding Threat Behavior. 20
7. The “DOE Gold Seal of Approval” for Trusted Software. 23
8. Information Hierarchy 24
9. The Research Thrusts and Priority Research Directions Defining a Cyber Security Science Research Agenda. 36

INTRODUCTION

A fundamental tenet of every scientific investigation is the search for truth. Trust and integrity are immutable values that are fundamental to all research endeavors. By extension, the absolute integrity of information, encompassing systems, networks, and file systems, is a critical requirement for the effective conduct of science, essential to promoting and facilitating discovery. These qualities are equally important for our nation's critical energy infrastructure. To determine the cyber security research needs for open science and energy control systems, the U.S. Department of Energy's (DOE) Office of Science (SC), in cooperation with the Office of Electricity Delivery and Energy Reliability (OE), organized a two-day workshop held July 23 and 24, 2007 at the Bethesda North Marriott Hotel in Bethesda, Maryland. This event brought together leading cyber security and infrastructure protection experts and researchers from academia, government labs, and industry.

The charge to the workshop participants was to define Priority Research Directions (PRDs) relevant to DOE's Open Science and Electricity Delivery and Energy Reliability missions. The workshop participants focused on long-term research directions in the cyber sciences with more than a 5- to10-year time frame for advances. A parallel focus was on intersecting the research directions with control systems, also with a 3- to 10-year time frame for deployable delivery.

WORKSHOP LOGISTICS

Approximately 150 experts registered for the invitation-only workshop, with broad representation of experts selected for their knowledge of cyber security. The workshop participants came from 16 industries, 55 DOE laboratories and government agencies, 69 U.S. universities, 2 European universities, and 5 non-profit organizations.

The workshop included plenary presentations from DOE program leaders and distinguished speakers from the field of cyber security, notably Steve Crocker from Shinkuro, George Spix from Microsoft, and Jason Stamp from Sandia National Laboratories. Presentations from these speakers informed and motivated the participants of the breakout groups to focus their attention on long-term research issues.

The breakout sessions encompassed the following topics and leaders:

- **Securing Hardware, Software, and Data** –Frank Siebenlist and Len Napolitano
- **Monitoring and Detection** – Troy Thompson and John McHugh
- **Future Security Architectures and Information Assurance Technologies** – Tom Harper
- **Human Factors Analysis** - Anne Schur and Joe St. Sauver
- **Protecting Our Utility Infrastructure** – Jeff Dagle, Aaron Turner, and Bill Young.

CROSSCUTTING FINDINGS AND CHALLENGES

The charge to the breakout groups included the following directives:

- That the PRDs represent significant challenges, requiring 3 to 10 years to address in a sustained research program.
- That, although many of the PRDs are the subject of research and development (R&D) by other agencies, there are unique aspects to DOE's open science and energy control systems environments that merit new and different R&D by the Office of Advanced Scientific Computing Research (OASCR) and OE.
- That the PRDs encompass fundamental aspects of mathematics, algorithms, and computational science.
- That the PRDs characterize research needed and, by and large, would take at least 3 years to have an impact upon DOE's operational environments.
- That the results of the research should eventually be applied to DOE's open science and the energy control systems environments to render them both more secure and more accessible, leading to science that is conducted more efficiently and effectively, and greater security for, reliability of, and usability of the energy distribution environment.

The end goal was to adopt a proactive and forward-looking approach at defining a long-term research focus from a rigorous scientific, mathematical, and technical basis that would stimulate new open science research directions and have a lasting impact on cyber security. This philosophy is evident in the salient points that emerged from the workshop from Steve Crocker during his presentation and are included in the following frame titled, Cyber Security Messages.

Cyber Security Messages

Proactive is better than reactive

- Most of the money in cyber security goes to reactive systems
  - CERT, law enforcement, firewalls, intrusion detection systems, etc.

Building in good security pays off, but is less sexy

Security is many things

- Privacy/confidentiality
- Who's the threat?
  - Competitor? Thief? Colleague? Government? Carrier? (Net neutrality)
- Integrity
  - Buffer overflow and other untrusted sources
- Availability
  - DDoS attacks are the biggest threat today

IDEA: Explicit representation of trust issues

Security for the few or the many?

- Protection of high value systems for selected people

OR

- Protection of all systems for everybody
- it is not clear that either can be separated from the other

Usability is paramount

- Can the user comprehend and describe the protection?
- Good security enables and facilitates uses
  - Otherwise we put up barriers and do less
- Unix file settings and router access controls are not usable

IDEA: Evaluation of usability of security

IDEA: Subordinate security under usability(!)

Size matters

- Cyber security models are usually simple and built to handle lots of rules.
- Moore's law assures lots of processing power, lots of memory, etc.
- Human capacity is small. No scaling. No evolution.

Keep it simple and keep it small.

What's the adoption path?

- Sometimes it's possible to start over
- Most of the time it's not
- The end point design is just the beginning

Need a path for adoption

- Interoperate with existing systems
- "Encourage" adoption

Architecture is distributed & recursive

- Trust issues arise at every level
- Reference domain for cyber security is local and flat
- Build in explicit checks, protection at every level
  - Executables, of course
  - Remote inputs
  - Cookies, etc.
- Hardware only helps a little
  - Random numbers are important for crypto

All bindings should have finite lifetimes

- People move around
- Machines change
- Networks change
- Etc.

Force all bindings to be changed or refreshed

We are not alone

- Many other government programs
  - What's different about this one?
  - How does it fit in, leverage, cooperate with the others?
- Vendors make a lot of money
  - Short term focus
  - Also the path (or barrier) for adoption

These messages were part of the workshop keynote presentation by Steve Crocker. The pragmatic nature of the messages led the workshop attendees to focus on realistic research directions and catalyzed a forum for open discussion.

PANEL FINDINGS AND PROPOSED RESEARCH

Altogether, 28 PRDs were identified independently in the sessions. However, one was judged as outside the scope of the workshop charge and is not reported in this document. The remaining 27 PRDs are aggregated into thrust areas that have overlapping research requirements that may eventually result in the definition of a DOE cyber security research program. The seven thrust areas are illustrated graphically in Figure 1, and the thrust areas and PRDs under each are:

- Future Open Science Security Architecture
  - Open Science Security Architecture
  - Trusted Virtualization
  - Economics-Based Security Architecture
  - Cyber Security Information Framework for Open Science
  - Resilient Distributed Computing
  - Secure Software
  - Federated Cyber Security for Open Science
- Adaptive/Autonomic/Homeostatic Security and Response
  - Decentralized Monitoring, Detection, and Response (human and automated)
  - Autonomic Incident and Damage Containment
- Situational Awareness
  - Intrusion Prevention and Detection
  - Distributed Denial of Service (DDOS) Tolerance
  - Verification of Intended Use
  - Enabling Data and Code Sharing and Cooperative Analytics
  - Appropriate Distributed Defense
- Integrity and Pedigree
  - Long-Term Integrity and Authenticity of Large and Dynamic Datasets
  - End-to-end Data Security
  - Secure Information Management
- Usability, Characterization, and Assessment
  - Characterization of Human Threats for Open Science
  - Malware Research Lab
  - Security Policy Implementation Impacts on Usability
  - Usability of Security (secure) Systems
  - Improving Cyber Security Practice
- Control Systems
  - Survivable and Trustworthy Control Systems
  - Anomaly Detection in Control Systems
  - Understanding Risk and Survivability Assessment
- Future of Cryptography
  - Non-Cryptographic Security
  - Trusted Hardware and Crypto Acceleration

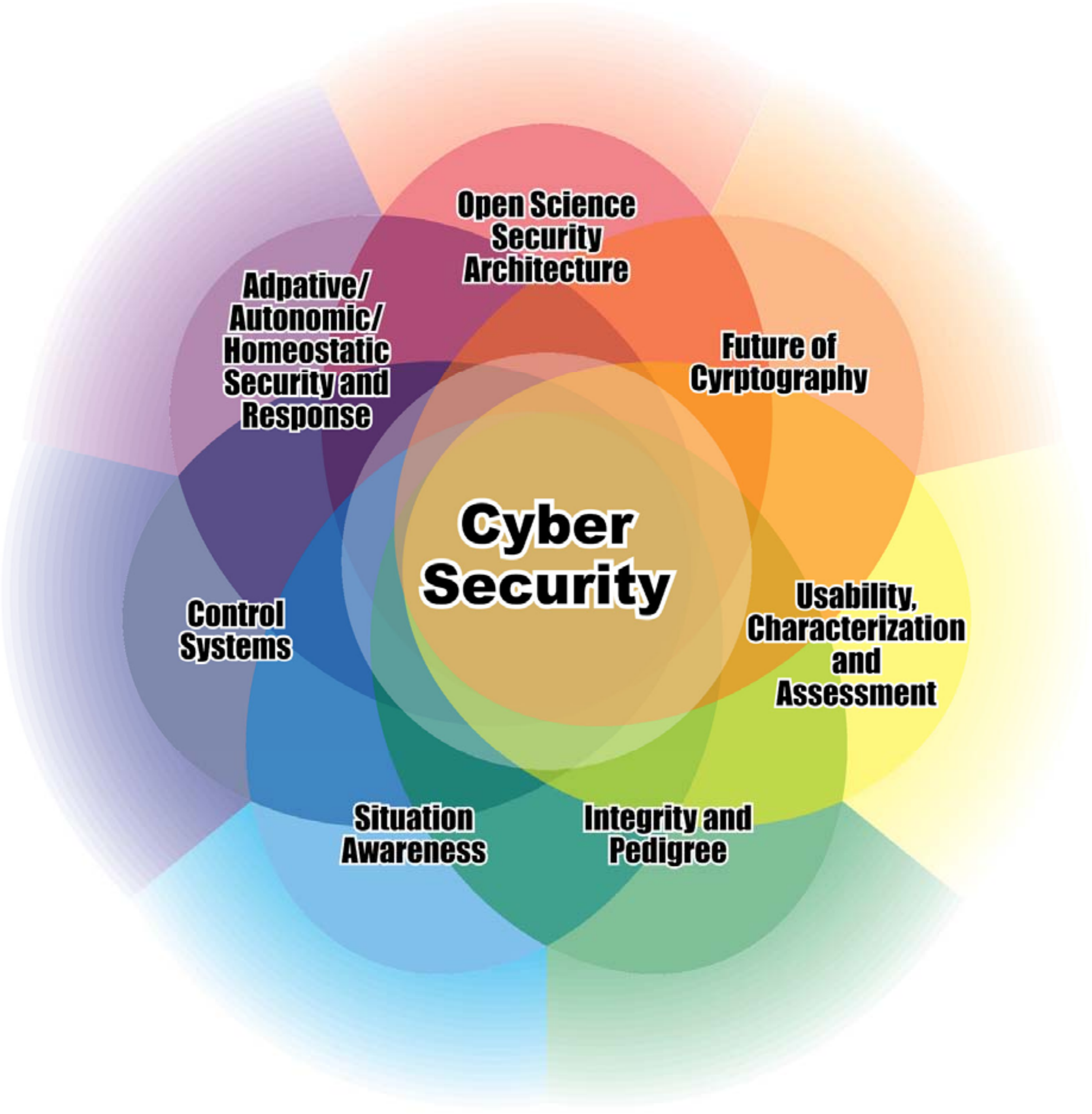

Open Science Security Architecture
Adpative/ Autonomic/ Homeostatic Security and Response
Future of Cyrptography
Cyber Security
Control Systems
Usability, Characterization and Assessment
Situation Awareness
Integrity and Pedigree

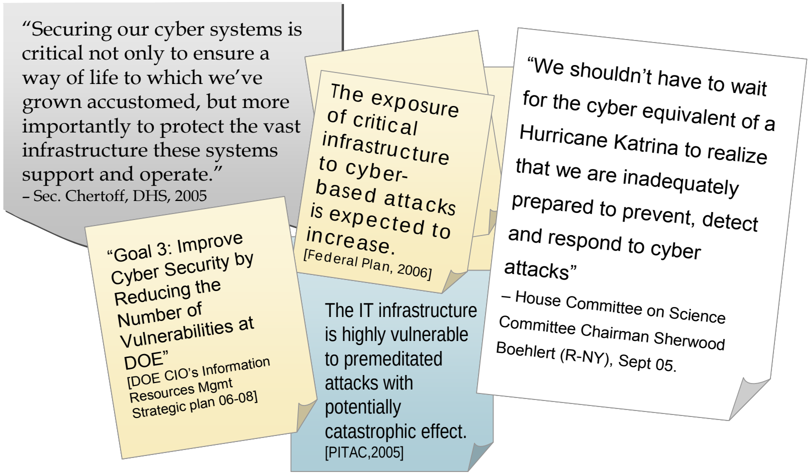
Figure 1. Cyber Security Research Thrust Areas. There are significant overlapping science research thrust areas where the intersection among areas will define the future capabilities to secure the Energy Control Systems and Open Science environments.

CYBER SECURITY CHALLENGES

A secure cyberspace is vitally important to the nation's welfare, but today's cyber environment is far from secure. The ability for individuals, organizations, and states to attack our nation's institutions and peoples' identities online in cyberspace has grown substantially to an alarming state.

Our heavy reliance on the Web and other emerging communication and collaboration technologies has exposed the nation to cyber attackers that now derive from anywhere in the world. We rely on Information Technology (IT) for the day-to-day operations of companies, organizations, and government. Peoples' personal lives also involve computing in areas ranging from communication with family and friends to online banking and other financial and household management activities. Companies large and small are ever more reliant on IT to support critical business processes, ranging from payroll and accounting to tracking of inventory, operations, sales, and support for research and development (R&D). Critical national infrastructures—such as those associated with energy, banking and finance, defense, law enforcement, transportation, water systems, communications, and government—and private emergency services—also depend on IT-based systems and networks; and the underlying telecommunications networks themselves are critical infrastructure for the nation.

Protecting systems and users, while maintaining ease of access, represents the "perfect storm" of challenges in the area of cyber security



"Securing our cyber systems is critical not only to ensure a way of life to which we've grown accustomed, but more importantly to protect the vast infrastructure these systems support and operate."
– Sec. Chertoff, DHS, 2005

The exposure of critical infrastructure to cyber-based attacks is expected to increase.
[Federal Plan, 2006]

"We shouldn't have to wait for the cyber equivalent of a Hurricane Katrina to realize that we are inadequately prepared to prevent, detect and respond to cyber attacks"
– House Committee on Science Committee Chairman Sherwood Boehlert (R-NY), Sept 05.

"Goal 3: Improve Cyber Security by Reducing the Number of Vulnerabilities at DOE"
[DOE CIO's Information Resources Mgmt Strategic plan 06-08]

The IT infrastructure is highly vulnerable to premeditated attacks with potentially catastrophic effect.
[PITAC,2005]

Figure 2. Motivation for Enhanced Cyber Security. Cyber security is a central issue that cuts across all agencies and activities. Identifying and maintaining innovative research for future technology developments and policy decisions that enable our national assets to stay ahead of today's cyber threats is critical to securing the nation's energy future and the infrastructure necessary for innovation and breakthrough discoveries in science attacks.

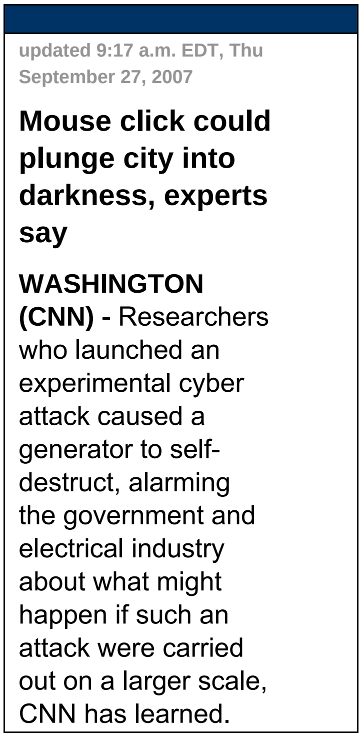
Our reliance on IT will only grow. The ability to realize the full benefits of IT depends on these systems being secure, yet the growing magnitude of the threat, whether associated with loss or damage, type of attack, or presence of vulnerability, indicates a worsening problem. Moreover, the actual scope of the threat is likely understated, because some successful attacks go unnoticed and others are noticed but not reported.

The potential consequences of inadequate security in cyberspace fall into three broad categories:

- Threat of catastrophe—a cyber attack, especially in conjunction with a physical attack—could result in thousands of deaths and billions of dollars of damage in a short time.
- Frictional drag—Frictional drag detracts from productivity and performance in important economic and security-related processes. Today, insecurities in cyberspace systems and networks allow adversaries to extract billions of dollars in fraud and extortion—and force businesses to expend additional resources to defend themselves against these threats. If cyberspace does not become more secure, the citizens, businesses, and governments of tomorrow will continue to face similar pressures, and on a greater scale.
- Lost opportunities—Concerns about inadequate cyber security may inhibit development and deployment of IT in the future, thereby minimizing the benefits that IT brings, benefits that will be needed to enhance the nation's global competitiveness as well as national and homeland security.

The charge to the workshop participants was to define Priority Research Directions (PRDs) relevant to the U.S. Department of Energy's (DOE's) Open Science and Electricity Delivery and Energy Reliability missions. The workshop participants focused on long-term research directions in the cyber sciences, typically with 5- to-10-year time frames for advances. A parallel focus was on intersecting the research directions of cyber security with control systems, on a 3- to 10-year time frame for initial deployment.

Improving cyber security is a complex, daunting task in DOE's Open Science environment. Several distinctive factors pertinent to DOE's open science mission are: 1) access is required to expensive, centralized resources, 2) emphasis is on "big science" that can be at unprecedented scales and exceedingly complex, and 3) users are numerous, highly decentralized and distributed, and exist in very diverse IT environments, most of which are not highly secured. The combination of these factors makes DOE unique in its cyber security needs, mandating fundamentally new approaches, illustrating the need for a sustained, coherent, and coordinated research program.



updated 9:17 a.m. EDT, Thu September 27, 2007

Mouse click could plunge city into darkness, experts say

WASHINGTON (CNN) - Researchers who launched an experimental cyber attack caused a generator to self-destruct, alarming the government and electrical industry about what might happen if such an attack were carried out on a larger scale, CNN has learned.

Then Secretary of Energy Spencer Abraham described DOE's unique Open Science environment in a 2004 DOE report: "DOE's state-of-the-art facilities are shared with the science community worldwide and contain technologies and instrumentation that are available nowhere else. Each year, these facilities are used by more than 18,000 researchers from universities, other government agencies, private industry, and foreign nations."

With expensive, one-of-a-kind facilities located around the world and with more than 18,000 researchers needing access to the systems, the data produced and the specialized computing resources that facilitate DOE's large worldwide collaborations, DOE is at a critical juncture. New approaches and technologies are needed to address cyber security in this environment. DOE must find flexible and scalable ways to provide a secure but usable scientific environment for its community of scientists and researchers.

Each of the Office of Science (SC) divisions has determined strategic hardware needs for the next 20 years as described in the two reports: the *Office of Science Strategic Plan* and the *Facilities of the Future.* No matter which SC division is examined, it is clear that their plans involve a significant need for strategic DOE computing and networking resources to enable their researchers to collaborate and to access strategic hardware. DOE researchers routinely use centralized, leadership-class computing resources, e.g., Argonne National Laboratory, National Energy Research Scientific Computing Center, Oak Ridge National Laboratory, and Pacific Northwest National Laboratory, decentralized processing capabilities, parallel computing facilities, massive data storage facilities, and movement and sharing of data for use in analysis. The SC's projection of new hardware and associated scientific endeavors for the next 5 to 10 years vastly expands the need for new cyber security research. An example is the near-term installation of the Large Hadron Collider where the expectation is that researchers will need, at a minimum, tens of gigabits per second of network capacity to conduct their research effectively.

The nation's energy delivery infrastructure shares many of the same cyber security issues that exist in the Open Science environment, mandating an imperative for a synergistic, joint approach. The OE report, *Roadmap to Secure Control Systems in the Energy Sector,* highlights the critical requirement for new cyber security research. Control systems—which include supervisory control and data acquisition (SCADA) systems and distributed control systems—perform vital functions across many of our nation's

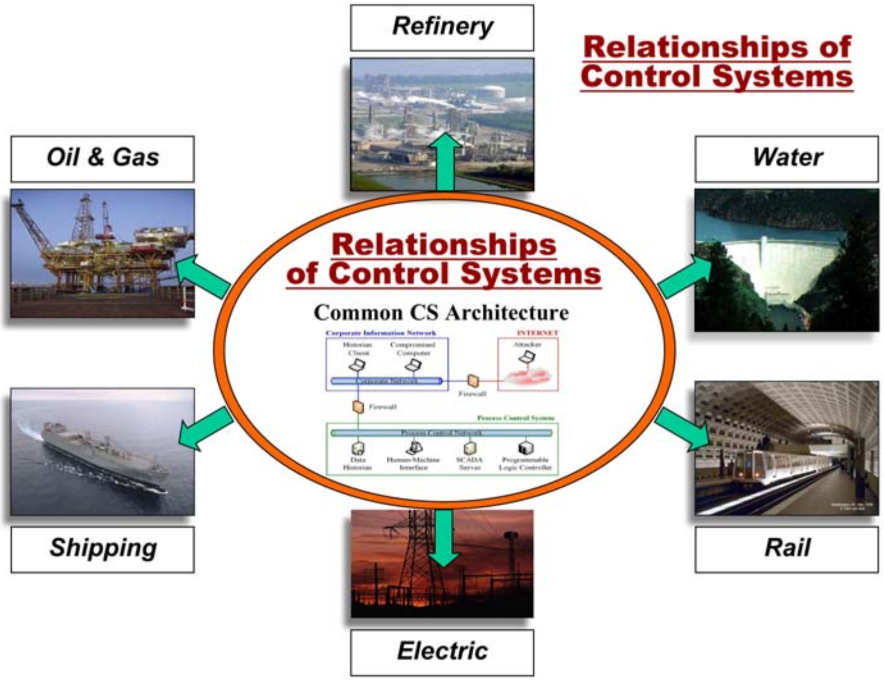


Figure 3. Energy Control Systems and Our Critical Infrastructures

critical infrastructures, including electric power generation, transmission, and distribution; oil and gas refining and pipelines; water treatment and distribution; chemical production and processing; railroads and mass transit; and manufacturing.

In addition to general cyber threats, which have been steadily increasing, several factors have contributed to the escalation of the risks of cyber attacks against control systems. These include the adoption of standardized technologies with known vulnerabilities and the increased connectivity of control systems to other systems. Control systems can be vulnerable to a variety of attacks. Successful attacks on control systems could have devastating consequences, such as endangering public health and safety. Securing control systems poses significant challenges, including limited specialized security technologies and lack of economic justification.

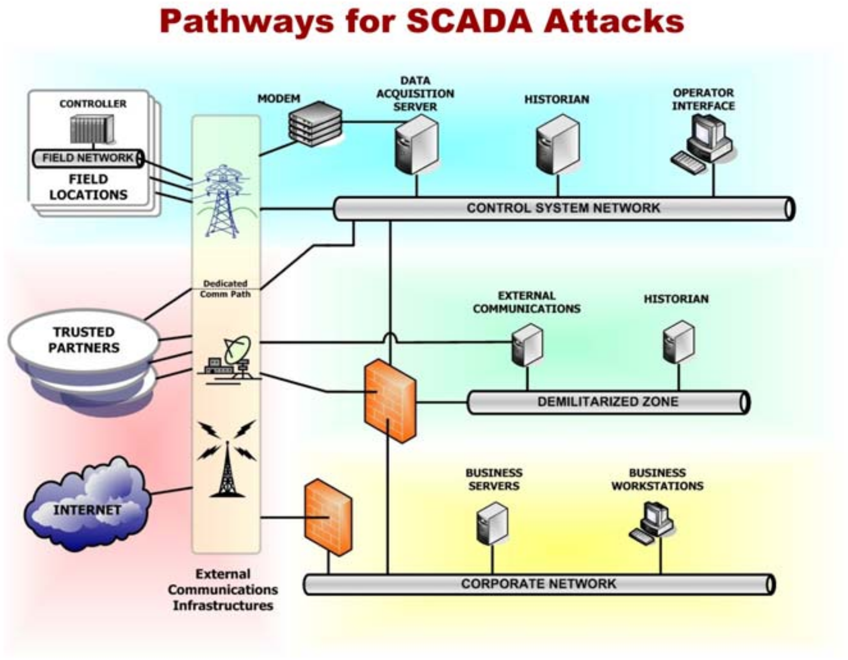


Figure 4. How Cyber Attacks Occur

Long-term efforts are needed to develop advanced methods and concepts for electricity delivery to and storage in the U.S. electric grid, ensuring that it remains among the most robust, reliable, secure, and technologically advanced in the world. Improving the security of energy control systems is a crucial requirement to protect our national energy delivery infrastructure.

DOE SC also has been spearheading the implementation of new, innovative leadership-class computing facilities and capacity. With a move to support exascale science and beyond, a new generation of computing technologies are emerging that will provide research opportunities for the next decade that could include new embedded, integrated cyber security features that result in more intrinsically secure, information-assured, open computing ecosystems.

Many of those areas have been captured in the PRDs from this workshop. The workshop has produced innovative research directions that have the following attributes:

- The PRDs represent significant challenges, requiring 3 to 10 years to address in a sustained research program.
- Although many of the PRDs are the subject of R&D by other agencies, there are unique aspects to DOE's open science and energy control systems environment that merit new and different R&D by the OASCR and OE.
- The PRDs encompass numerous fundamental aspects of mathematics, algorithms, and computer science.
- The PRDs characterize research needed that, by and large, would take at least 3 years to have an impact upon DOE's operational environments.
- Obvious synergies and overlaps exist between cyber security in the Open Science environment and the Energy Control Systems environment to render them more secure and more robust, leading to open science that is conducted more efficiently and effectively, and greater security, reliability, and usability in the energy distribution environment.

The PRDs that emerged from the workshop are consistent with the five principles defined in the 2007 National Research Council report, *Toward a Safer and More Secure Cyberspace,* and that will guide and focus the ongoing research agenda. These five principles are given in the frame below.

In addition, the National Research Council identified the protection of energy distribution services by improving security for supervisory control and data acquisition (SCADA) system as one of 14 most important technical initiatives in its 2002 report, *Making the Nation Safer: The Role of Science and Technology in Countering Terrorism*. Government, academia, and private industry have initiated efforts to strengthen the cyber security of control systems. The White House's 2003 report, *National Strategy to Secure Cyberspace,* states that securing SCADA is a national priority and establishes a role for the Department of Homeland Security to coordinate with these entities to improve the cyber security of control systems. DHS's coordination of these efforts could accelerate the development and implementation of more secure systems.

The five principles defined in the 2007 National Research Council report, *Toward a Safer and More Secure Cyberspace,* will guide and focus the ongoing research agenda.

Conduct cyber security research as though its application will be important. The scope of cyber security research must extend to understanding how cyber security technologies and practice can be applied in real-life contexts. Consequently, fundamental research in cyber security will embrace organizational, sociological, economic, legal, and psychological factors as well as technological ones.

Hedge against uncertainty in the nature and severity of the future cyber security threat. It seems prudent to take a balanced approach that hedges against the eventuality that a high-end cyber security threat emerges and becomes manifestly obvious to all. That hedge is an R&D agenda in cyber security that is both broader and deeper than might be required if only low-end threats were at issue. (Because of the long lead time for large-scale deployments of any measure, part of the research agenda must include research directed at reducing those long lead times.)

Ensure programmatic continuity. A sound research program should also support a substantial effort in research areas with a long time horizon for payoff. This is not to say that long-term research cannot have intermediate milestones, although such milestones should be treated as midcourse corrections rather than “go/no-go” decisions that demoralize and make researchers overly conservative. Long-term research should engage both academic and industrial actors, and it can involve collaboration early and often with technology-transition stakeholders, even in the basic science stages.

Respect the need for breadth in the research agenda. Cyber security risks will be on the rise for the foreseeable future, but few specifics about those risks can be known with high confidence. Thus, it is not realistic to imagine that one or even a few promising approaches will prevent or even substantially mitigate cyber security risks in the future, and cyber security research must be conducted across a broad front. In addition, because qualitatively new attacks can appear with little warning, a broad research agenda is likely to significantly decrease the time needed to develop countermeasures against these new attacks when they appear. Priorities are still important, but they should be determined by those in a position to respond most quickly to the changing environment—namely, the research constituencies that provide peer review and the program managers of the various research-supporting agencies. Notions of breadth and diversity in the cyber security research agenda should themselves be interpreted broadly as well, and might well be integrated into other research programs such as software and systems engineering, operating systems, programming languages, networks, Web applications, and so on.

Disseminate new knowledge and artifacts, e.g., software and hardware prototypes, to the research community. Dissemination of research results beyond one’s own laboratory is necessary if those results are to have a wide impact—a point that argues for cyber security research to be conducted on an unclassified basis as much as possible. Other information to be shared as widely as possible includes threat and 222 incident information that can help guide future research.

THRUST AREAS

The PRDs are organized into seven thrust areas that embody a DOE cyber security research program. The thrust areas are:

- **Future Open Science Security Architecture:** This thrust area involves understanding and developing a new, secure cyber architecture that can scale with emerging and planned DOE open science requirements, including extensive worldwide collaborations. The ultimate goal would be the definition of new, built-in capabilities and a set of baseline standards for cyber security that can be institutionalized by vendors.
- **Adaptive/Autonomic/Homeostatic Security and Response:** The enormous complexity of cyber security in DOE open science computing systems and in energy control systems is, at times, beyond the capabilities of humans by themselves to assimilate and manage in real time. Clearly, sophisticated, automated systems are required. This thrust area is focused on computerized self-management of cyber security, based on human-defined policies and rules.
- **Situational Awareness:** Effective cyber security requires effective command, control, and communications systems. Situational awareness is complex, dynamic, in DOE's high-risk Open Science and Electricity Delivery and Energy Reliability environments. This thrust area focuses attention on understanding the situational complexity needed for effective cyber security, including representations of objects, people, system states, interactions, environmental conditions, and other situation-specific factors.
- **Integrity and Pedigree:** Today, DOE's scientists use a myriad of commercial off-the-shelf software in a blind trust model, including "software of uncertain pedigree" (SOUP). This thrust area addresses the complex "back-ends" needed to define, measure, and make accessible to the open science community integrity and pedigree of software and data.
- **Usability, Characterization, and Assessment:** Understanding the effect of human behaviors on and in cyber environments is imperative. Indeed, human factors are often the weakest link in a cyber security environment. On the other hand, extreme cyber security measures can render systems inaccessible and/or unusable. To catalyze scientific discovery, a balance between usability and cyber security must exist, with an emphasis on usability. This thrust area provides for research in the primary understanding of human behavior both in launching attacks and in using cyber security in day-to-day work.
- **Control Systems:** Improving the cyber security of energy delivery and control systems is a crucial step for national infrastructure protection. This thrust area addresses research into the key cyber security attributes of control systems: survivability and trustworthiness.
- **Future of Cryptography:** Current cyber security practices depend extensively on cryptography. In DOE's open science environment, datasets are becoming too large for today's cryptographic methods. It also is obvious that computing systems will soon become sophisticated enough to easily crack current cryptographic schemes. This thrust area will define specific research in this open science domain.

CYBER SECURITY DEFENSE TAXONOMY

It is useful to categorize the elements that comprise end-to-end architectures and higher-level aspects that must be protected in a cyber security environment. Such a categorization will facilitate aggregating research directions into the aforementioned thrust areas to result in a coherent and consolidated cyber security defense research and development program. The illustration below shows a taxonomy that is useful for this purpose, beginning with hardware components, e.g., ASICS, PC boards, I/O elements, etc. that are assembled into hardware systems, to data and information. Note that cyber security can and should be implemented at each of the levels shown, and in many cases across the boundaries shown in the figure. Moreover, the areas of Middleware and Users cut across all elements and even extend “down” into hardware components. This extension of users and middleware across all cyber security areas merits discussion. For example, consider the Trusted Platform Module (TPM), a chip (a hardware component) that is currently integrated into motherboards on portable computers. The TPM chip serves multiple purposes, including authenticating the system and its configuration in a manner not subject to forgery, encrypting system and user files, and it even can be used in conjunction with a biometric reader, e.g., a fingerprint reader, for biometric user authentication. Thus, users interact directly even with specific, individual hardware components. In addition, Middleware, originally occupying the domain between the system and the user, now reaches all levels of the taxonomy, as authentication and encryption involve not only such hardware components as TPM chips, but users who must interact with individual hardware components of the system, the network, the software, and with other as they embark upon discovery in the Open Science environment.

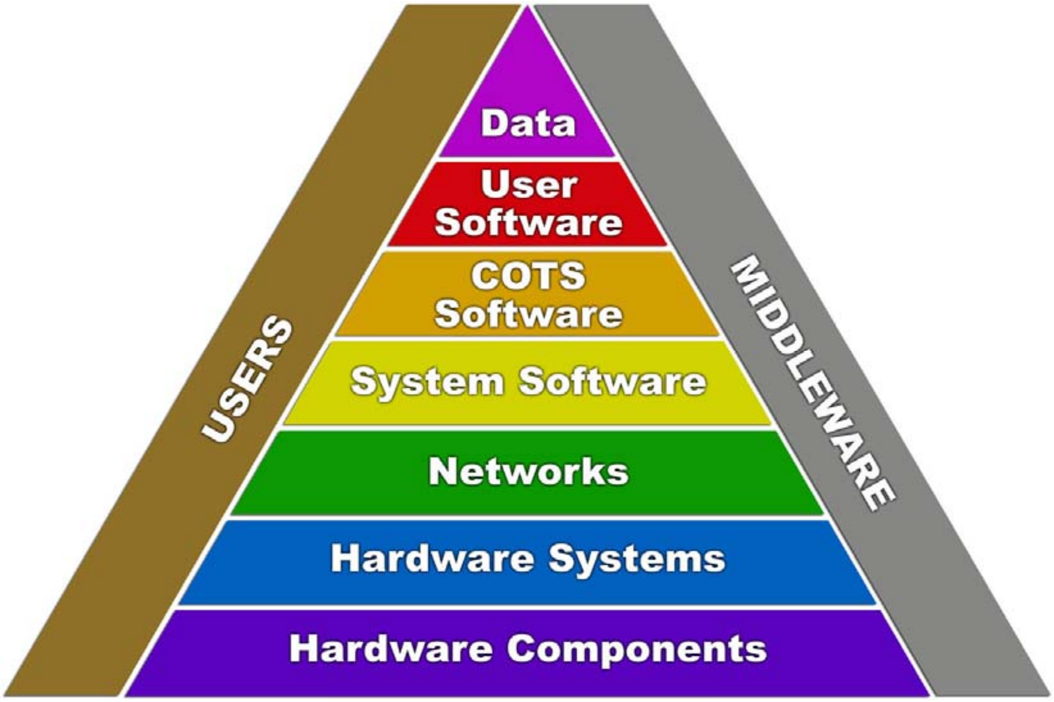


Figure 5. Taxonomy of Cyber Security

PRIORITY RESEARCH DIRECTIONS

Altogether, 28 PRDs were identified independently in the breakout sessions. However, one was determined to be outside the scope of the original charge for the workshop and was therefore omitted from this report leaving 27 PRDs in this section of this report. Because a number of these PRDs have sufficient commonality, they were aggregated into the 7 overall thrust areas noted above. In this section, the PRDs will be summarized individually within the aggregated thrust areas. The final PRDs that were submitted are included in the Appendix − no attempt was made to condense or consolidate PRDs in the Appendix for the sake of completeness. Finally, common research elements and areas of fit with the cyber security taxonomy are also presented.

> *"A common theme across all the PRDs was the need for common vocabularies, semantics, and ontologies of the security-related components and associated properties and attributes."*
> -from the Securing HW, SW, and Data session.

- Trusted Virtualization
- Economics-Based Security Architecture
- Cyber Security Information Framework for Open Science
- Resilient Distributed Computing
- Secure Software
- Federated Cyber Security for Open Science

This thrust area involves understanding and the development of a new, secure cyber architecture that can scale with forthcoming DOE open science requirements, including extensive worldwide collaborations. The ultimate goal would be new capabilities and a set of baseline standards for cyber security that can be institutionalized by vendors and embedded throughout new cyber systems.

> *"Security + Architecture = hard!"*
> - from the FSA session.

- **Trusted Virtualization** – This PRD proposes the development of a model for trusted virtualization of computational environments. Virtualization could provide a scalable computational ecosystem that would be based upon capabilities uncoupled from the hardware upon which software is deployed. The challenges involve extending trust from current hardware to emerging virtual environments, expressing and enforcing security goals at the level of a virtual machine, and maintaining cyber security when transitioning among virtual environments. New cyber security discoveries include the development of containment strategies, mechanisms for quantifiable verification of trust in virtual environments, and the addition of a cyber security layer to the virtual environment. The benefits of this research are greater trust/cyber security in virtualized environments, greater efficiency of usage due to the ability to load balance flexibly across and among virtual environments, and easier distribution and usability of software environments for the user. Additionally, this PRD will address fault tolerant and resilient computing systems beyond petascale systems. The research is expected to require 3 to 10 years to achieve fruition with advancements in computer science and mathematics. This PRD involves the lower-level areas of the cyber security taxonomy, from systems software down through hardware components and will require new approaches and algorithms from the field of computer science.

> *"With the explosion of the number of (virtualized) systems, there is a claimant requirement for globally, unique, and universal identifiers to which arbitrary naming authorities can bind attributes and properties that can be securely resolved by reliable parties."* -- abstracted from "Handle System" (http://handle.net).

- **Economics-Based Security Architecture** – This PRD proposes the development of a model to analyze interactions between cyber security and user policies, phrase problems as sets of games, derive Nash equilibria, prove scalability, identify inflection points, and define countermeasures and associated costs. The challenges include the uncertain and evolving black market economics associated with identity theft and system resources, inherent complexity, identification of security trade-offs, and determination of quantifiable model variables and parameters. New cyber security discoveries include the development of enforcement across multiple scales of heterogeneous systems, optimization of defense strategies, and a model with the capability to evaluate trade-offs in terms of hardware, network, software, policy, and human countermeasures. The primary benefit of this research is a consistent understanding of all elements of cyber security that can be used to deploy resources (both human and machine) optimally. The research is expected to require 3 to 10 years to achieve fruition. Ideally, this PRD should encompass all areas of the cyber security taxonomy, most especially including the economic factors associated with productivity. This PRD will benefit from new approaches and algorithms from the fields of computer science and mathematics.

- **Cyber Security Information Framework for Open Science** – This PRD proposes the development of a new framework for assessing security in open science environments, including unifying the semantics of security data. The challenges involve the numerous complexities inherent in the Open Science environment – especially the multitude of different components, systems, sites, and users. The framework must accommodate the complexity without impairing user accessibility and productivity, be scalable, and model appropriate trade-offs of cyber security against these areas. A self-consistent framework, in and of itself, would be a significant new cyber security discovery. The benefits of this research are more effective, trusted systems and greater integrity for using those systems in distributed, open environments. It was not determined how long this research might take to achieve fruition. This PRD involves all areas of the cyber security taxonomy and will require new approaches and algorithms from the field of computer science.
- **Resilient Distributed Computing** – This PRD proposes the exploration of a framework for fine-grained replication, replication on-demand, large-scale virtualization, and incremental migration as a means to assess, detect, minimize, prevent and/or mitigate the loss or corruption of computational resources including data. The challenges involve incorporating replication into systems that are already running at or near capacity on very large science problems that may execute for extended periods of time. This problem is known to be canonically hard. New cyber security discoveries include the development of procedures to evaluate, distribute, and (re)allocate computing resources in a highly distributed environment. R&D in this area will facilitate trust, increase availability, i.e., more cycles should become available, and enhance robustness in Open Science computing environments. The research is expected to require 5 to 10 years to achieve fruition. This PRD involves the lower-level areas of the cyber security taxonomy, from systems software down through hardware components, and will require new approaches and algorithms from the field of computer science.
- **Secure Software** – This PRD proposes taking a fresh look at securing software, including detection, diagnosis, moderation, and remediation of cyber security vulnerabilities. The challenges include the complexities of the diversity of software; distributed, heterogeneous systems upon which the software executes; the need for an end-to-end approach; and the long life cycles of some software. There also are human factors of how diagnostics, moderation, and remediation will be communicated to and interact with users, system administrators, and cyber security experts. New cyber security discoveries include parallel techniques for early detection of security vulnerabilities in distributed, heterogeneous software and systems. The primary benefit of this research is that it addresses cyber security at a fundamental level where cyber security must be inviolate, i.e., software. The research is expected to require less than 5 years to render cyber security more accessible to developers, 5 to 7 years to develop techniques applicable to commercial off-the-shelf (COTS) software, and more than 7 years to develop moderation and remediation techniques. This PRD involves the upper-level areas of the cyber security taxonomy, from systems software up through data, and will require new approaches and algorithms from the field of computer science.

- **Federated Cyber Security for Open Science** – This PRD addresses DOE's participation in federated identity management. The DOE open science community has unique needs to provide secure and easy access to DOE resources – systems will be more secure and more accessible to the open science community. The challenges are that the open science community exists in a highly decentralized environment, involving many sites, each with different environments and policies for cyber security. New cyber security discoveries include novel techniques for user privilege negotiation among systems, user authentication, user authorization, and possibly remote configuration of cyber resources in the remote environment. The benefits of this research are better cyber security, easier accessibility to DOE resources, and distribution of the effort required to implement cyber security. The research is expected to take 3 to 5 years for initial efforts (federated authentication and authorization), and 5 to 7 years for federated configuration. This PRD involves all areas of the cyber security taxonomy, especially users and middleware, and will require new approaches and algorithms from the field of computer science. However, new software and possibly hardware also are expected to result from this research.

The common elements in this thrust area are very broad, encompassing hardware, operating system software, architecture, algorithms, economic analysis and optimization, and middleware (trust models), with usability and users running throughout.

- Decentralized Monitoring, Detection, and Response
- Autonomic Incident and Damage Containment

The enormous complexity of cyber security in DOE's Open Science and in Energy Control Systems environments is, at times, beyond the capabilities of humans by themselves to assimilate and manage in real time. Clearly, sophisticated automated systems are required. This thrust area is focused on computerized self-management of cyber security, based on human-defined policies and rules.

> *"An ounce of prevention is worth a pound of cure."*
> - Benjamin Franklin

- **Decentralized Monitoring, Detection, and Response (human and automated)** – This PRD involves the development of research methods and approaches for: intelligent data reduction/analysis techniques, dynamic modeling, definitions of triggers for proactive response, proactive response mechanisms, and advanced visualization for analysts. The challenges are: 1) the need for a trusted link between data and human operators, 2) characterization of data and datasets to be analyzed, e.g., systems, components, networks, users, coordinating and correlating across multiple decentralized domains, mitigating the effects of "poisoned" data, and the massive amount of data. New cyber security discoveries include the development of self- and community-aware systems and next-generation systems that perform to DOE's requirements (ultra-high capacity and low latency). The benefits of this research include improved cyber security in DOE systems, better decision support, and graceful degradation rather than catastrophic failure when experiencing cyber attacks. The research is expected to require 5 to 10 years to achieve fruition with advancements in computer science, new hardwired architectures, large-scale data intensive analytics, and sensor developments. This PRD involves several areas of the cyber security taxonomy, notably hardware systems (possibly hardware components/ASICS), networks, software and data. This PRD will benefit from new approaches and algorithms from the fields of computer science, mathematics, statistics, and behavioral science.

> *"The emphasis should be on MyScience instead of MySpace."*
> – Discussion in the Workshop Session

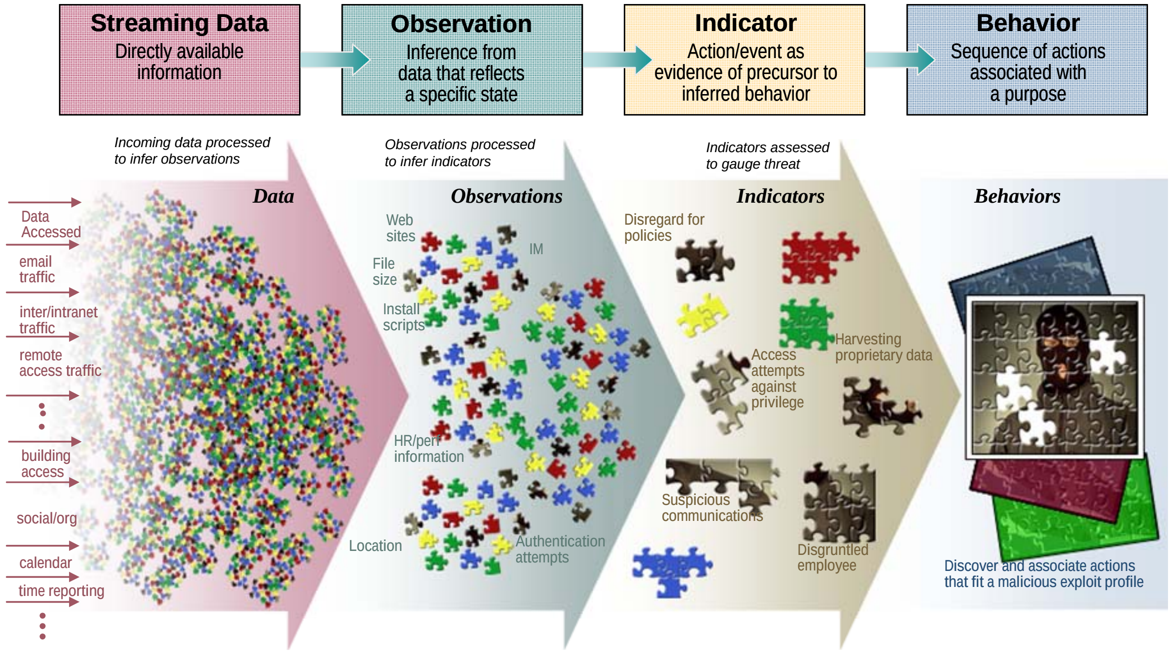
- **Autonomic Incident and Damage Containment** – This PRD proposes the development of secure approaches to autonomic cyber security incident and damage containment. The fields of control theory, group dynamics, machine learning, and software assurance are involved. The challenges of the research include the size, speed, broad scope, and complexities inherent in the DOE open science environment. The benefits of this research are shorter time to identify and react to cyber security incidents, accomplished by removing or distancing the human from the system. How long this research might take to achieve fruition remains an open question due to the rapid evolution and deployment of new technologies into communications and computing systems that continue to drive these systems to new performance levels and increase their complexity. This PRD involves several areas of the cyber security taxonomy, notably networks, software (and possibly hardware components or ASICS due to the large volume of information that must be processed), and data. This PRD will benefit from new approaches and algorithms from the fields of computer science, mathematics, and statistics.

Common elements in this thrust area encompass characterization, assimilation, and identification of threats, and will most likely require new hardware and systems to implement at the scales required in DOE environments.

- Intrusion Prevention and Detection
- Distributed Denial of Service (DDOS) Tolerance
- Verification of Intended Use
- Enabling Data and Code Sharing and Cooperative Analytics
- Appropriate Distributed Defense

Effective cyber security requires effective command, control, and communications systems. Situational awareness is complex, dynamic, and must be applied in DOE's high-risk open science and energy environments. This thrust area focuses attention on understanding the situational complexity needed for effective cyber security, including representations of objects, people, system states, interactions, environmental conditions, and other situation-specific factors.

Real-Time Analysis vs. Forensics
Model–Based Classification


Streaming Data Directly available information

Observation Inference from data that reflects a specific state

Indicator Action/event as evidence of precursor to inferred behavior

Behavior Sequence of actions associated with a purpose

Incoming data processed to infer observations

Observations processed to infer indicators

Indicators assessed to gauge threat

Figure 6. Understanding Threat Behavior. Pulling the pieces together is the problem of discovery analytics. Research enabling the development of understanding threat behavior using scalable analytics will require a new class of high performance computing tools that are based on new performance metrics to enable data fusion and analysis.

- **Intrusion Prevention and Detection** – This PRD proposes the development of a framework for analysis and characterization of the structure and nature of specific DOE open science data, control, and execution paths. Methods and technologies are required to capture and process, at extremely high speeds, the following elements: monitoring, packet filtering, anomaly detection, information fusion, integrated response, fewer false positives, failback mechanisms, containment, and forensics. The challenges of this PRD are scale (extremely high traffic and large number of users) and scope (vast geographical distribution and types of systems). R&D in this area will improve trust among users and systems for broad, diverse, open science. The research is expected to require 3 to 5 years to achieve fruition, and production deployment may be achieved in 4 to 10 years. This PRD involves primarily the network and software areas of the cyber security taxonomy and will benefit from new approaches and algorithms from the fields of electrical and computer engineering, computer science, mathematics, and statistics.
- **Distributed Denial of Service (DDOS) Tolerance** – This PRD proposes a focused examination of technology and techniques to defend large-scale, distributed computing and experimental systems against DDOS attacks while allowing the computation and/or experiment to proceed. The challenge includes the development of vastly more sophisticated techniques and algorithms than are currently available. R&D in this area will promote safe, resilient computing and experiments in open science. The research is expected to require 5 to 10 years to be put into production with advances in large-scale, data processing and analysis techniques. This PRD involves primarily the network and software areas of the cyber security taxonomy and will benefit from new approaches and algorithms from the fields of electrical and computer engineering, computer science, mathematics, and statistics.
- **Verification of Intended Use** – This PRD proposes the development of new frameworks and methodologies to verify that both users of DOE open science systems and the applications being run on them are as intended. New profiling application tools are needed to ensure validation and verification of software for systems of the future. Research may span the areas of biometric devices, user usage and behavior (human "signatures") semantics, watermarking binary and source code, etc. The challenges include complexities inherent with human factors, the large scale of traffic, large numbers of users, and complexity inherent in DOE open science distributed heterogeneous systems. New capabilities will be developed to detect human attacks, especially "insider" attacks, and enable more science by protecting systems from unintended use. The research is expected to require 5 to 10 years to achieve fruition, although intermediate aspects may become available in 7 to 15 years. This PRD involves primarily the network and software areas of the cyber security taxonomy and will benefit from new approaches and algorithms from the fields of computer science, mathematics, statistics, and behavioral science.
- **Enabling Data and Code Sharing and Cooperative Analytics** – This PRD proposes the development of secure approaches to encourage users to greater degrees of sharing data, algorithms, and applications. The challenges include non-deterministic social factors, all of the difficulty inherent in data of different time and length scales in differing formats, and

cyber security aspects. The benefits of this research include a shorter time-to-solution and reduced storage and transmission capacity needs in the Open Science environment. This research is expected to require 5 to 10 years to be put into production. This PRD involves primarily the software and data elements of the cyber security taxonomy, but also must be accessible to and usable by users. This PRD will benefit from new approaches and algorithms from the fields of computer science, mathematics, and behavioral science.

- **Appropriate Distributed Defense** – This PRD proposes research into techniques to query a wide variety of cyber security information sources that shield open science systems and users from known cyber security problems. Such sources might include databases and blacklists of known exploits, viruses, worms, malware, dangerous sites, etc. Also being proposed in this PRD is the development of an ecosystem-wide awareness capability, quantitative measures of the cyber security "health" of a system, and display of results visually and intuitively in a real-time presentation suitable for program managers and security analysts. The research challenges include the size, speed, broad scope, and complexities inherent in the DOE Open Science environment. The research benefits of this research include a shorter time to identify and react to cyber security incidents as well as more comprehensive information about the health of a system to inform human analysis and decision-making. The research is expected to take 3 years to begin to bear fruit, and 5 to 10 years to be put fully into production. This PRD involves primarily the network and software areas of the Cyber Security taxonomy but also must be accessible to and usable by network and system operators. Thus, new approaches and algorithms from the fields of computer science and statistics,

Common elements in this thrust area encompass capture, characterization, assimilation, and identification of threats from network and system traffic, and will most likely require new hardware and systems to implement at the scales required in DOE environment.

- Long-Term Integrity and Authenticity of Large and Dynamic Datasets
- End-To-End Data Security
- Secure Information Management

Today, DOE scientists use a myriad of commercial, off-the-shelf (COTS) software in a blind trust model, including “software of uncertain pedigree” (SOUP). This thrust area aims to develop the complex “back-ends” needed to define, measure, assess, and verify the integrity and pedigree of software before it is made accessible to the open science community. Moreover, some of the same techniques should be used to assess and verify the integrity and pedigree of data before it is allowed to be put on or accessed by DOE open science systems. Ideally, software and data that are verified would be granted a “stamp of approval” or a “gold seal” as illustrated in the graphic below. However, in reality, a virtual stamp of approval would be required before software or data is installed on or accessed by DOE open science systems. Different levels of software assurance would be appropriate for different levels of access to DOE open science systems.

Figure 7. The “DOE Gold Seal of Approval” for Trusted Software. Software and data that are verified would be granted a “stamp of approval” or a “gold seal.”

The PRDs in this area involve higher-level processing and intelligence. Viz., the information hierarchy below illustrates the process flow in this area: given a software instance, data must be collected and/or produced on that instance, information is then distilled from that data, and knowledge is gained from the exercise of these processes over many instances. Eventually, over time and after analysis of many cases, wisdom in knowing how to protect our systems and software should result.

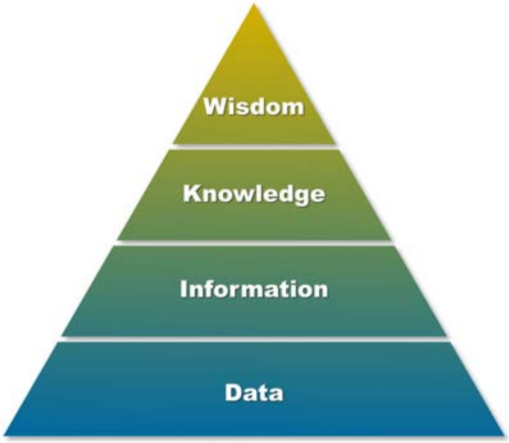


Figure 8. Information Hierarchy

The PRDs in this thrust area encompass both the information hierarchy and integrity and pedigree areas. Elements of all of these areas are needed for a fully developed, comprehensive program. Three PRDs have been defined in this area, as follows:

- **Long-Term Integrity and Authenticity of Large and Dynamic Datasets** – This PRD proposes the development of data integrity models and analyses that: survive reductions/abstractions, accommodate privacy constraints, maintain their usefulness over the long-term, are auditable, are efficient in terms of processing, are applicable over widely distributed, heterogeneous systems, and take uncertainty into account. The research challenges are: datasets can be exceedingly large; involve aggregation, reduction, and fusion; are often dynamic; and can persist over very long times. Techniques currently do not exist for measuring the integrity of datasets that are often the purview of distributed users and are not well understood by any individual. New cyber security discoveries in this area require fundamentally new approaches in analyzing, assessing, and evaluating data; in detecting and correcting errors; and in engaging users to adopt secure, effective behaviors. The research is expected to require 5 to 7 years to achieve fruition. This PRD involves primarily the software and data elements of the cyber security taxonomy, but also must be accessible to and usable by users. Thus, new approaches and algorithms are required from the fields of computer science, mathematics, and statistics.
- **End-to-End Data Security** – This PRD proposes the development of new frameworks and techniques for end-to-end data security on distributed, heterogeneous systems. Managing storage and provenance for dynamic international collaborations for thousands of scientists across many, diverse platforms and domains are the research challenges. New cyber security discoveries in this area include automatic metadata capture, validation, and transfer among systems and users. This research is expected to result in more effective, trusted collaborations and greater integrity for data in distributed environments and is expected to require 5 to 10 years to achieve fruition. This PRD involves primarily the software and data elements of the cyber security taxonomy, but also must be accessible to and usable by users. Thus, new approaches and algorithms are required from the fields of computer science, mathematics, and statistics.

- **Secure Information Management** – This PRD proposes the development of new frameworks to protect critical information distributed across millions of nodes. The research challenges involve complexity and scalability, i.e., datasets can be exceedingly complex, large, often aggregated, reduced, or fused. They also are often dynamic, can persist over very long times, and can exist on many, distributed, heterogeneous systems. Additionally, information involves much more than just data, including code (binary and source), metadata, and complex, sometimes unstructured relationships among data. New cyber security discoveries in this area require fundamentally new ontologies, models, processing approaches, and policies. This research is expected to result in more effective, trusted data and systems that will facilitate sharing data and resources among open science communities. The research is expected to require 5 to 7 years initially to produce results and 8 to10 years to achieve security and scalability. This PRD involves primarily the software and data elements of the cyber security taxonomy, but also must be accessible to and usable by users. Thus, new approaches and algorithms are required from the fields of computer science, mathematics, and statistics.

The common research elements in this thrust area include assessment and profiling tools to verify both software and data, operating system sentries to verify intended use, techniques for categorizing the degree to which software and data are trusted, techniques for analyzing extremely high volumes of network traffic, and categorizations of DOE open science resources as to which software and data must be verified and to what degree before being implemented on DOE systems.

The common approaches in this thrust area encompass the characterization, capture, assimilation, and identification of threats from network and system traffic, and will most likely require new hardware and systems to implement at the scales required in DOE environment. New, advanced algorithms will be required to implement and operate these elements.

- Characterization of Human Threats for Open Science
- Malware Research Lab
- Security Policy Implementation Impacts on Usability
- Usability of Security (secure) Systems
- Improving Cyber Security Practice

> *"In the Bentham calculus of protecting our systems, networks and data, the user is often forgotten, ignored, or even neglected, sometimes profoundly affecting productivity and impeding open science discoveries."*
> -from the Human Factors session

Understanding the effect of human behaviors in cyber security is imperative. To catalyze scientific discovery, a balance between usability and cyber security must exist, with an emphasis on usability. This thrust area involves the higher levels of the cyber security presented earlier. The emphasis is not on systems, but rather on the effectiveness of information capture and then the usability of that information in a production environment. Such human factors are often overlooked, as observed in the workshop's Human Factors session (see inset). Indeed, in a field as nascent and emerging as cyber security, human factors become apparent only after the technology matures sufficiently. First and foremost, the technology must work; then usability and vulnerability issues are addressed as the technology matures. This thrust area defines research to develop the predictive techno-social understanding of human behavior both in launching attacks and in using cyber security in day-to-day work.

- **Characterization of Human Threats for Open Science** – This PRD proposes the development of scalable techniques to understand, predict, and detect human behavior of users of DOE open science systems. Human signatures for evaluating and assessing use of systems also need to be developed as part of this PRD. The research challenges involve complexities inherent with human factors, the large scale of traffic, high numbers of users, and application on and in distributed, heterogeneous environments. New capabilities must be developed to predict, detect, and understand the intention of human attacks, especially "insider" attacks. This research is expected to require 5 to 10 years to achieve fruition, although intermediate benefit may become available in 3 to 5 years. This PRD involves primarily the software and data elements of the cyber security taxonomy, but also must be accessible to and usable by network, system, and security administrators. This PRD will benefit from new approaches and algorithms from the fields of computer science, mathematics, statistics, and behavioral science.
- **Malware Research Lab** – This PRD proposes the creation of an environment to implement and test cyber security malware in representative environments. The research challenges are the large, complex, and evolving instances of malware that must be "mapped onto" a large,

complex, and evolving set of hardware, software, and networks that is often distributed and interacting. New cyber security discoveries include novel techniques for copying and duplicating malware, duplicating or characterizing and representing systems and environments upon which to "map" the malware, and the ability to assess the effects of malware on these systems so that appropriate protective and countermeasures may be defined, tested, and implemented. This research area has many similarities to comparative genomics with the biological metaphor of DNA as the cellular operating system for complex networks of control networks has expressive power. New tools are needed that can rapidly assess and detect a dynamic threat in open computing environments. The research is expected to require 7 to 10 years to achieve fruition with advances in computer science, mathematics, computer architecture, large-scale data-intensive parallel computing and visualization. This PRD involves all areas of the cyber security taxonomy and will benefit from new approaches and algorithms from the fields of computer science, mathematics, statistics, and biological science.

- **Security Policy Implementation Impacts on Usability** – This PRD proposes the development of a flexible simulation testbed for modeling cyber security policies on open science systems so that they can be tested before being put into production. Metrics and measurements for evaluating and assessing the effectiveness of policies on cyber security and usability will be developed as part of this PRD. The research challenges involve the complexities inherent with cyber security policies in large-scale, highly distributed environments; the difficulty of translating cyber security policy into practice in complex, distributed, heterogeneous environments, and the difficulty in assessing implications of new cyber security policy. New cyber security discoveries will eventually include a framework to better understand the impacts of new cyber security policy upon systems and users; enable a more consistent application of cyber security policy across multiple, distributed systems; and provide a more useable, accessible, productive environment for open science. The research is expected to require 5 to 10 years to achieve fruition, although intermediate aspects may become available in 3 to 5 years. This PRD involves all areas of the cyber security taxonomy, but most of the effort and emphasis will be on human factors and the higher levels of the taxonomy. This PRD will benefit from new approaches and algorithms from the fields of computer science and biological science.
- **Usability of Security and/or Secured Systems** – This PRD proposes the development of new user interfaces to promote the ease and correctness of installation, configuration, operation, and maintenance of security systems. In addition, this PRD advocates the development of quantitative metrics and measurements for usability. The research challenges include the significant complexity inherent in cyber security systems and the difficulty communicating in this complex environment. New cyber security discoveries include the development of metrics and measures to quantify usability of security systems. Research in this area will provide increased assurance that systems and networks are better protected and will result in more efficient and effective cyber security operations and greater availability of secure systems, thereby facilitating open science. The research is expected to require 5 to 10

years to achieve fruition. This PRD involves all areas of the cyber security taxonomy, but most of the effort and emphasis will be on human factors and usability across all areas of the taxonomy. This PRD will benefit from new approaches and algorithms from the fields of mathematics, statistics, computer science, and biological science.

- **Improving Cyber Security Practice** – This PRD proposes the development of new trust frameworks and the tools to model, simulate, and analyze trust in open science environments. Risk-benefit analyses for cyber security practice must also be developed and exercised. The research must address the challenges of an open science community that exists in a highly decentralized environment, involving many sites, each with different policies and infrastructures for trust. New cyber security discoveries include novel techniques for trust negotiation among systems and users. It was not determined how long this research might take to achieve fruition. This PRD involves primarily the software area of the cyber security taxonomy and will benefit from new approaches and algorithms from the fields of mathematics and statistics.

Some of the common research elements in this thrust area are the development of a quantitative tool for assessing system usability, development of assessment and profiling tools to duplicate or simulate environments upon which security may be implemented and tested, and development of techniques for evaluating and fine tuning how security policy can best be implemented in and on complex systems and architectures.

The common approaches in this thrust area encompass new definitions, metrics, and frameworks for cyber security vulnerability and "health" of systems; new methods to capture, characterize, assimilate, and identify threats on the network and throughout (and even across) systems; and most likely will require new hardware and systems to implement at the scales required in DOE's Open Science environment. Also, a new malware research and operations laboratory is proposed to be defined and implemented. New, advanced algorithms and flexible systems (hardware and software) will be required in this research thrust area.

- Survivable and Trustworthy Control Systems
- Anomaly Detection in Control Systems
- Understanding Risk and Survivability Assessment

Improving the security of energy control systems is a crucial step for national infrastructure protection. This thrust area proposes research directly into key attributes of control systems: survivability and trustworthiness.

- **Survivable and Trustworthy Control Systems** – This PRD involves template architectures for control systems, including models of survivability, designs for graceful failure (controlled degradation), and improved support for human intervention. Protection is to be against malicious attacks and accidental failures, accommodate varying reliability requirements, and strike the appropriate balance between safety and performance. The research challenges are numerous, including: the distributed, heterogeneous nature of systems and system components; how to quantify, measure, and evaluate survivability and trustworthiness with respect to cyber and physical threats; how to identify and prioritize failure and degradation; and the requirement to maintain a high level of service during an incident. New cyber security discoveries include the development of comprehensive models of systems that will address holistic factors. The benefits of this PRD are strong and survivable systems − both DOE large-science systems and utility infrastructure. The research is expected to require 3 to 10 years to achieve fruition. This PRD involves primarily the lower levels of the cyber security taxonomy, from hardware components up through systems software. This PRD will benefit from new approaches and algorithms from the fields of mathematics and computer science.
- **Anomaly Detection in Control Systems** – This PRD involves the development of a model for system behavior, appropriate parameters for and sensitivity to control systems, and a generic template for anomaly detections for adaptive, self-healing control systems. The challenges facing research include the following: systems deployed today are more complex than our ability to understand them fully; failure modes are not completely predictable; a large diversity (age, high degree of geographical distribution, and technologies) of systems exist in production; and a wide variety of factors influence performance, including environmental, social, physical network, and the interface between humans and systems. New cyber security discoveries include novel techniques for early detection that will be widely applicable to power and control systems. The benefits include greater cyber security in control systems, better decision support, and graceful degradation rather than catastrophic failure. The research is expected to require 5 to 7 years to achieve fruition with advancements in computer science, computer architecture, statistics, and the mathematics of complexity. This PRD cuts across all levels of the cyber security taxonomy, but primarily involves the areas of software and human factors. This PRD will benefit from new approaches and algorithms from the fields of mathematics, statistics, and computer science.

- **Understanding Risk and Survivability Assessment** – This PRD proposes the development of a comprehensive, real-time, high-performance operational model of control systems, and the evaluation of its robustness during a security incident. A significant emphasis of this PRD will be on gathering, logging, distilling, anonymizing, and sharing threat data. A security investment model also is a component of this PRD. The research challenges are numerous, including: the distributed, heterogeneous nature of systems and system components; how to quantify, measure, and evaluate robustness with respect to cyber and physical threats; how to identify and prioritize incident response, including factors of cost; how to quantify and predict responses to operator interaction; how to enforce cyber security in the face of real-time requirements; and the requirement to maintain a high level of service during an incident. New cyber security discoveries include the development of comprehensive models of control systems that will address holistic factors. The benefit of this PRD is trusted and robust systems. The research is expected to require 3 to 10 years to achieve fruition. This PRD involves primarily the software level and the cross-cutting user area of the cyber security taxonomy. This PRD will benefit from new approaches and algorithms from the fields of mathematics, statistics, and computer science.

The common elements in this thrust area encompass the characterization, capture, assimilation, and identification of threats from the network and systems and will most likely require new approaches and algorithms to implement in DOE's energy delivery environment.

- Non-Cryptographic Security
- Trusted Hardware and Crypto Acceleration

Current cyber security practices depend extensively on cryptography. In DOE's Open Science environment, it is clear that datasets are becoming too large for today's cryptographic methods. It also is obvious that computing systems will soon become sophisticated enough to easily crack current cryptographic schemes. This thrust area will provide DOE with specific open science research into this topic.

> *"HW vendors are actively working to incorporate trusted-computing and crypto features in chips, CPUs, instruction sets, motherboards, and supporting software to harden and security-enhance compute platforms."*
> − From 1) "Intel® Trusted Execution Technology" (www.intel.com/technology/security/) and 2) "The Intel Safer Computing Initiative" (http://www.intel.com/intelpress/sum_secc.htm).

- **Non-Cryptographic Security** – This PRD involves taking a completely fresh look at cryptographic security. If malware that impairs or "breaks" mathematical algorithms for cryptography is emergent, virtually all networks and environments become open to packet sniffing. The challenge involves the vast scope of the problem – software encryption is embedded throughout our secure systems today and the entrenched and ubiquitous nature of our packet-switched networks. New cyber security discoveries include the development of new strategies and methodologies for protecting information, including possible circuit-based approaches or hardware-based cryptography. The research is expected to require 3 to 5 years to achieve fruition. On the surface, this PRD seems to involve areas that are not represented on the cyber security taxonomy, and may include new hardware, mathematics, and computer science algorithms. However, an implementation may cut across all areas of the taxonomy, as security must be embedded within and throughout cyber environments.

- **Trusted Hardware and Crypto Acceleration** – This PRD also involves taking another, completely fresh look at cryptographic technology, for which today's implementations represent a significant bottleneck to performance. As a result, encryption often is not selected for transport, thereby posing a cyber security vulnerability. Here, the focus is on developing next-generation cryptographic technologies that can keep up with the demands of the fastest networks. New cyber security discoveries include the development of new algorithms that perform much better, possibly hardware-based accelerators, and new strategies for secure key exchange. The research is expected to take 3 to 5 years to achieve fruition. This PRD involves primarily the hardware and software levels of the cyber security taxonomy. A fresh look at the mathematics, algorithms, and hardware/software implementations of cryptography are in order under this PRD.

The two elements in this thrust area may be disparate. One encompasses fundamentally new scientific approaches to cryptography. The other involves new algorithms, hardware and mathematics to accelerate traditional cryptography.

SPECIFIC RECOMMENDATIONS FOR A DOE CYBER SECURITY RESEARCH AGENDA

This section of the report contains research thrust areas that the workshop chairs feel are most pertinent to and perhaps even unique in DOE's Open Science and Energy Control Systems environments. Specifically, these are areas where DOE has both unique needs due to the scale and nature of its environments and unique capabilities evidenced in its mission. Indeed, these areas in particular may be where DOE can make the most significant and most enduring cyber security research contributions. Securing the design and operational integrity of the exascale computing enterprises of the future is by definition an exascale challenge. Cyber security science and technologies must keep pace with the computing architectures developed if there is a hope to keep scalable open computing resources available.

NEW ARCHITECTURES

To make significant progress in the myriad of cyber security, it is categorically apparent that cyber security must be built into systems from the ground up. This, in fact, may be the most important cyber security research direction that emerged from the workshop. Thus, new architectures containing new hardware (e.g. TPM+ chips), designed to include embedded cyber security monitoring and processing capabilities (e.g. on-board or peripheral cyber security processing, virtualized architectures), and even especially designed to accommodate new cyber security analytics (e.g. processors designed for ultra-fast data comparison and analysis encompassing searches, sorts, merges, joins and pattern recognition) and new encryption and decryption techniques are needed. Multi-core capability and FPGA processing offer promise in this regard. DOE has unique, very large-scale hardware platforms and associated expertise that provide the environment for processing at the rates required in DOE's large-scale environment and meet the need inherent in this environment. This research would apply to the lower levels of the cyber security taxonomy, involve principally computer scientists, and should include strong interactions with hardware and operating system vendors.

SITUATIONAL AWARENESS AND ADAPTIVE, AUTONOMIC AND HOMEOSTATIC ANALYSIS AND RESPONSE SYSTEMS

Cyber security is today reactive, and in far too many cases, accomplished only manually. Threats and vulnerabilities are defined and addressed only after they emerge, are then isolated, analyzed, and distilled into well-defined behaviors and even digital signatures. Clearly, new systems are needed to detect threats based upon more than just tabulated data, i.e., using sophisticated, predictive mathematical models, to "stay ahead of the curve." Again, DOE has both unique needs in this area, due to the massive rate at which data are generated, assimilated, and exchanged, and unique capabilities its vast computational resources, algorithm expertise, and analytics needed to process the data. This research would principally be conducted at the higher levels of the cyber security taxonomy. Although hardware and software would play an essential role, most of the discovery and development would occur at the higher levels involving computer science, mathematics, and statistics.

MIDDLEWARE AND FEDERATED IDENTITY MANAGEMENT

As mentioned previously, human factors are often the weakest link in any cyber security environment. DOE is unique in the large scale and scope of its Open Science environment. Users of DOE open science systems are vast in numbers and exist in highly distributed, unverified cyber environments. Better, “smarter” federated systems for authenticating users and authorizing access to varying classes of DOE assets will facilitate easier, more secure access to valuable DOE assets. Nowhere is the need greater for improved access control and cyber security than in DOE’s Open Science environment. This research and development would occur across all levels of the cyber security taxonomy, but mostly involve software and systems.

ENERGY CONTROL SYSTEMS

DOE is singularly responsible for the proper functioning of the nation’s energy delivery systems. This is an essential and incredibly important infrastructure that must be secured from cyber vulnerabilities. Much of the other DOE cyber security research is applicable in this sector and should be tested, productized, and implemented in this environment that has its own unique implementation nuances. Indeed, much of the cyber security research should be a joint program of both offices, so as to realize maximum benefit and quickest time to deployment in the energy sector.

INTEGRITY AND PEDIGREE

In addition to its unique hardware and system environments, DOE is unique in its vast scale and scope of its data and software environments. Fundamentally, software and data are the aspects of a cyber security environment that contain cyber vulnerabilities and allow systems to be exploited. In addition to its unique needs, DOE is uniquely positioned with expertise in its software environments. For decades, DOE has been a leader in compiler and assembler technologies, due to its unique hardware environment. Moreover, DOE has experience in providing both physical and virtual security. Thus, no agency is better positioned than DOE to conduct cyber security research in the integrity and pedigree of code, executables and data. Particular aspects that deserve attention are adding a cyber security layer (or “pass”) to compilers to provide a cyber security score for vulnerabilities that can be used in a federated system as a “gatekeeper” to specific DOE resources, and verifying the authorship and ownership of in the “web of trust” for shared software and data by watermarking, bonding, binding, etc. This research and development would occur across the higher levels of the cyber security taxonomy, but would mostly involve software and systems.

MALWARE R&D LABORATORY

A malware research and development laboratory, as mentioned previously, would provide an ideal development environment to implement, test, refine, and productize the results of DOE’s cyber security research. Indeed, a test and development environment is essential to transform the research into products and practices. Building a knowledge base of software configurations that are key to DOE assets will be critical to understanding the real state of the enterprise. With the proper security models in place, the results of this research could become a national cyber security technology asset. This research area should be well coordinated with other agencies that have similar issues and concerns, enabling the technology to be developed and rapidly deployed operationally in a timely fashion. Derivative technologies can then be transferred to the private sector and made more broadly deployable in open cyber security operational environments, thereby rendering the cyber security research relevant and enduring. This research and development would occur across the higher levels of the cyber security taxonomy, but would mostly involve software and systems.

CONCLUSION

IMPORTANCE OF CYBER SECURITY AND OPEN SCIENCE

DOE SC is responsible for the secure and efficient operation of some of the nation's most advanced R&D user facilities. These state-of-the-art, multi-billion-dollar facilities contain technologies and instrumentation that are not available anywhere else and are relied upon by the science community worldwide. These are only some of the national assets that DOE has stewardship for that must be protected. Through shared use of these facilities, the open science community produces discoveries and advances in numerous scientific areas of critical national and international importance that otherwise would be impossible.

These facilities and assets, however, are high-visibility targets for various forms of cyber attack that, if successful, can have extremely deleterious consequences. Compromise of a high-value user facility can deter scientific progress for weeks. Accidental or deliberate data or software corruption can invalidate literally years of work. Moreover, misuse of computational resources to launch attacks on other facilities can have devastating consequences.

The nation's energy delivery infrastructure shares many of the same cyber security issues that exist in the Open Science environment, mandating an imperative for a synergistic, joint approach.

Cyber security advances are critical to meeting Open Science and Control Systems needs. These advances will produce the missing components required to deploy effective open science cyber security systems and energy control systems. An advanced science and technology program, sustained, coherent, and coordinated, will bridge the ever-widening gap between modern academic and commercial cyber security research and the solutions necessary for open science.

A PATH FORWARD

Improving cyber security is a complex, daunting task in both the Open Science and Energy Control Systems environment. But with a focus on scalability and flexibility in the research directions, the workshop attendees believe that significant forward progress toward creating both usable and secure environments can be made in the next 10 years. The research directions suggested in this report will inspire new cyber security safeguards for both open science and energy control systems.

It is additionally recognized that while the goal is to enable a secure and open science infrastructure for research, it is often the case that the mechanisms to deliver that security involve an overlap with classified operations for implementation. **It is imperative that the DOE cyber infrastructure be effectively defended to ensure national energy security for the future.** Research in long-term science applicable to cyber security will have broad-ranging value to the DOE mission as well as for national and homeland security.

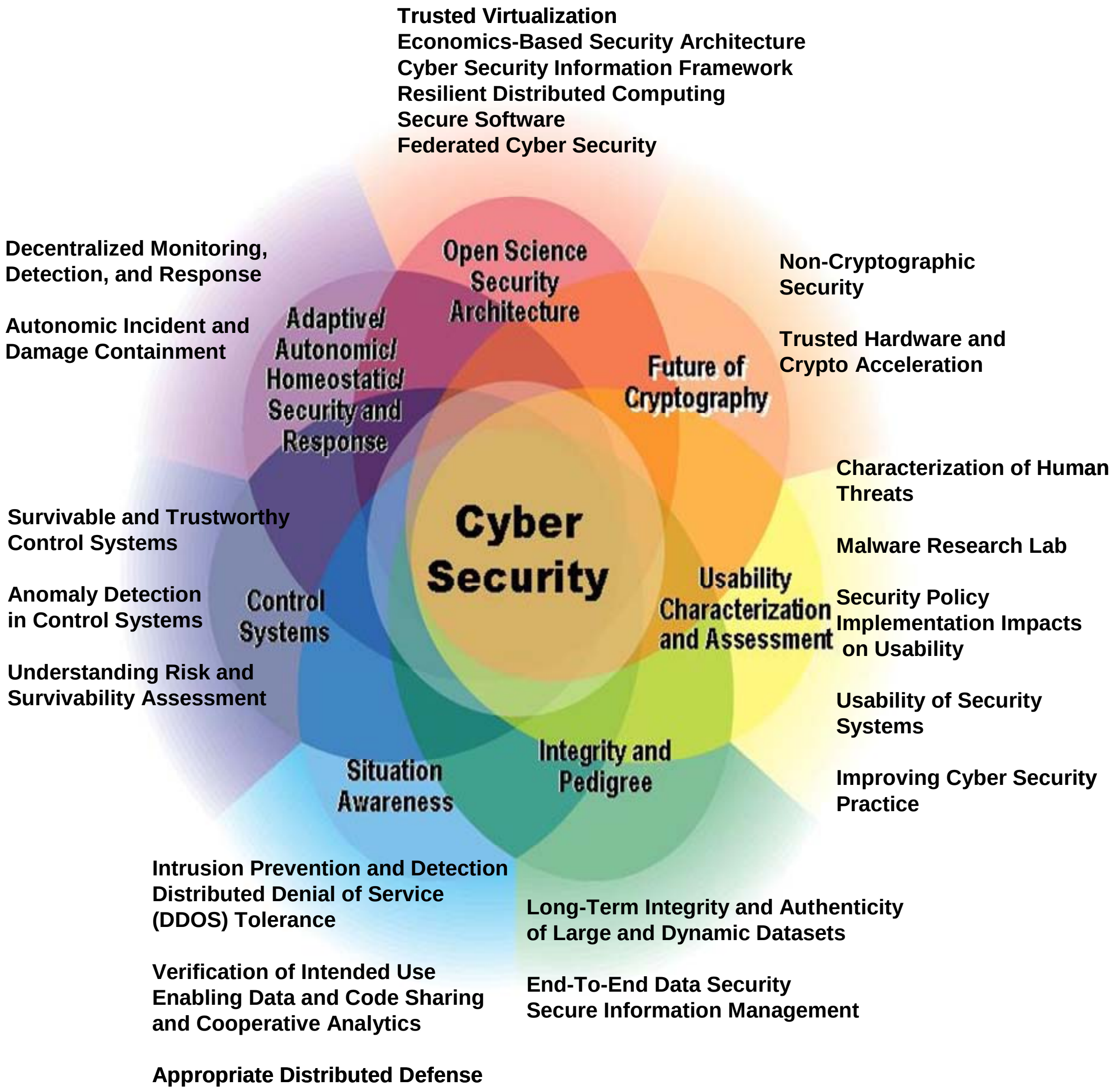

Trusted Virtualization
Economics-Based Security Architecture
Cyber Security Information Framework
Resilient Distributed Computing
Secure Software
Federated Cyber Security
Decentralized Monitoring, Detection, and Response
Autonomic Incident and Damage Containment
Open Science Security Architecture
Adaptive/ Autonomic/ Homeostatic/ Security and Response
Future of Cryptography
Non-Cryptographic Security
Trusted Hardware and Crypto Acceleration
Characterization of Human Threats
Malware Research Lab
Security Policy Implementation Impacts on Usability
Usability of Security Systems
Improving Cyber Security Practice
Cyber Security
Survivable and Trustworthy Control Systems
Anomaly Detection in Control Systems
Understanding Risk and Survivability Assessment
Control Systems
Usability Characterization and Assessment
Situation Awareness
Integrity and Pedigree
Intrusion Prevention and Detection
Distributed Denial of Service (DDOS) Tolerance
Verification of Intended Use
Enabling Data and Code Sharing and Cooperative Analytics
Appropriate Distributed Defense
Long-Term Integrity and Authenticity of Large and Dynamic Datasets
End-To-End Data Security
Secure Information Management

Figure 9. The Research Thrusts and Priority Research Directions Defining a Cyber Security Science Research Agenda.

Appendix A – Workshop Organizers

Appendix A – Workshop Organizers

A.1 Organizing Committee

- Patrick Burns, Colorado State University
- Susan Estrada, Aldea Communications
- George Michaels, Pacific Northwest National Laboratory
- Ron Bailey, Consultant
- Dave Zachman, Mesa Networks

A.2 Panel Leaders

Securing Hardware, Software and Data (SHSD)

- Frank Siebenlist, Argonne National Laboratory
- Len Napolitano, Sandia National Laboratories

Monitoring and Detection (MD)

- Troy Thompson, Pacific Northwest National Laboratory
- John McHugh, Dalhousie University

Future Security Architectures and Information Assurance Technologies (FSA)

- Tom Harper, Idaho National Laboratory

Human Factors Analysis (HF)

- Anne Schur, Pacific Northwest National Laboratory
- Joe St Sauver, Internet2

Protecting our Utility Infrastructure (UI)

- Jeff Dagle, Pacific Northwest National Laboratory
- Aaron Turner, Idaho National Laboratory
- Bill Young, Sandia National Laboratories

Appendix B – Workshop Participants

Appendix B – Workshop Participants

Industry

Dunagan, John; Microsoft Research; jdunagan@microsoft.com
Estrada, Susan; Aldea Communications; sestrada@aldea.com
Kounavis, Michael; Intel Corporation; michael.e.kounavis@intel.com
Kropp, Thomas; the Electric Power Research Institute; tkropp@epri.com
Lazarus, John; Symantec; john_lazarus@symantec.com
Lowry, John; BBN Technologies; jlowry@bbn.com
Mohan, Ram; Afilias; anna@afilias.info
Pato, Joe; HP Labs; joe.pato@hp.com
Peterson, Dale; Digital Bond, Inc.; peterson@digitalbond.com
Rakaczky, Ernest; Invensys Process Systems; ernest.rakaczky@ips.invensys.com
Sachs, Marcus; SRI International; marcus.sachs@sri.com
Schissel, David; General Atomics; schissel@fusion.gat.com
van Doorn, Leendert; AMD; leendert.vandoorn@amd.com
Wakid, Shukri; HP; Shukri.Wakid@hp.com
Wan, Tao; Nortel; twan@nortel.com
Zatko, Mudge; BBN Technologies; mudge@bbn.com

DOE Laboratories and Government Agencies

Agarwal, Deb; **Lawrence Berkeley National Laboratory; daagarwal@lbl.gov**
Allen, John; **Lawrence Livermore National Lab; allen24@llnl.gov**
Ashby, Steven; **Lawrence Livermore National Laboratory; sfashby@llnl.gov**
Bailey, F. Ronald; **AMTI/NASA Ames Research Center; fbailey@mail.arc.nasa.gov**
Baker, Ann; **Oak Ridge National Laboratory; bakerae@ornl.gov**
Bartoletti, Tony; **CIAC; bartoletti1@llnl.gov**
Bland, Arthur (Buddy); **Oak Ridge National Laboratory; blandas@ornl.gov**
Brown, David; **Lawrence Livermore National Laboratory/ASCR; brown@ascr.doe.gov**
Burnette, John; **Pacific Northwest National Laboratory; anne.ouderkirk@pnl.gov**
Casella, Rich; **Brookhaven National Lab; rac@bnl.gov**
Castro, Pedro; **Lawrence Livermore National Laboratory; pcastro@llnl.gov**
Corbett, Cherita; **Sandia National Labs; clcorbe@sandia.gov**
Dagle, Jeff; **Pacific Northwest National Laboratory; jeff.dagle@pnl.gov**
Diachin, Lori; **Lawrence Livermore National Laboratory; diachin2@llnl.gov**
Draelos, Timothy; **Sandia National Labs; tjdrael@sandia.gov**
Eidenbenz, Stephan; **Los Alamos National Laboratory; eidenben@lanl.gov**
Eliassi-Rad, Tina; **Lawrence Livermore National Laboratory; eliassirad1@llnl.gov**
Foster, Ian; **Argonne National Laboratory; foster@mcs.anl.gov**
Gaines, Irwin; **Fermilab; gaines@fnal.gov**
Goldfarb, Joshua; **US-CERT; joshua.goldfarb@us-cert.gov**
Goodwin, David; **Department of Energy; dave.goodwin@science.doe.gov**
Harper, Thomas; **Idaho National Laboratory; Thomas.Harper@inl.gov**
Hazlewood, Victor; **Oak Ridge National Laboratory; victor@ornl.gov**
Helland, Barbara; **Department of Energy; Barbara.Helland@science.doe.gov**

Hoisie, Adolfy ; **Los Alamos National Laboratory; hoisie@lanl.gov**
Hush, Don; **Los Alamos National Laboratory; dhush@lanl.gov**
Johnson, Gary; **Office of Science; Gary.Johnson@science.doe.gov**
Kemper, Chris; **Oak Ridge National Laboratory; kemperkb@ornl.gov**
Kent, Alex; **Los Alamos National Laboratory; alex@lanl.gov**
Kramer, William; **NERSC, Lawrence Berkeley National Laboratory; wtkramer@lbl.gov**
Landwehr, Carl; **DTO; clandwehr@casl.umd.edu**
Lindsay, Robert; **Department of Energy; lindsay@ascr.doe.gov**
May, Deborah; **Lawrence Livermore National Laboratory; may14@llnl.gov**
Michaels, George; **Pacific Northwest National Laboratory; george.michaels@pnl.gov**
Midkiff, Scott; **National Science Foundation; smidkiff@nsf.gov**
Minuzzo, Kim; **Lawrence Livermore National Laboratory; minuzzo@llnl.gov**
Napolitano, Leonard; **Sandia National Laboratories; napolitano@sandia.gov**
Oldfield, Ron; **Sandia National Laboratories; raoldfi@sandia.gov**
Pancerella, Carmen; **Sandia National Laboratories; carmen@sandia.gov**
Pederson, Perry; **DHS; perry.pederson@dhs.gov**
Petravick, Don; **Fermi National Accelerator Lab; petravick@fnal.gov**
Poole, Stephen; **Oak Ridge National Laboratory; spoole@ornl.gov**
Pundit, Neil; **Sandia National Labs; pundit@sandia.gov**
Quinlan, Dan; **Lawrence Livermore National Laboratory; dquinlan@llnl.gov**
Rao, Nagi; **Oak Ridge National Laboratory; raons@ornl.gov**
Riesen, Rolf; **Sandia National Laboratories; rolf@sandia.gov**
Rogers, Jim; **Oak Ridge National Lab; jrogers@ornl.gov**
Schmucker, Ron; **Lawrence Livermore National Lab; schmucker1@llnl.gov**
Schur, Anne; **Pacific Northwest National Laboratory; anne.schur@pnl.gov**
Sekine, Yukiko; **ASCR/Department of Energy; yukiko.sekine@science.doe.gov**
Siebenlist, Frank; **Argonne National Laboratory; franks@mcs.anl.gov**
Stamp, Jason; **Sandia National Laboratories; jestamp@sandia.gov**
Strip, David ; **Sandia National Laboratories; drstrip@sandia.gov**
Studham, Scott; **Oak Ridge National Laboratory; studham@ornl.gov**
Sumikawa, Denise; **Lawrence Livermore National Laboratory; dsumikawa@llnl.gov**
Tatar, John; **Argonne National Laboratory; tatar@anl.gov**
Thompson, Troy; **Pacific Northwest National Laboratory; troy.thompson@pnl.gov**
Torgerson, Mark; **Sandia National Laboratories; mdtorge@sandia.gov**
Turner, Aaron; **Idaho National Laboratory; aaron.turner@inl.gov**
Vasil, David; **Oak Ridge National Laboratory; dmvasil@ornl.gov**
Watson, Jean-Paul; **Sandia National Laboratories; jwatson@sandia.gov**
Weaver, Mike; **DOE/ASCR; weaver@ascr.doe.gov**
White, Greg; **Lawrence Livermore National Laboratory; white6@llnl.gov**
Young, Bill; **Sandia National Laboratory; wfyoung@sandia.gov**
Zacharia, Thomas; **UT-Battelle/Oak Ridge National Laboratory; zachariat@ornl.gov**

U.S. Universities

Ahn, Gail-Joon; **UNC Charlotte; gahn@uncc.edu**
Benzel, Terry; **USC-ISI; tbenzel@isi.edu**
Borisov, Nikita; **University of Illinois at Urbana-Champaign; nikita@uiuc.edu**
Bose, Anjan; **Washington State University; bose@wsu.edu**
Bradford, Wayne; **University of Utah; wayne.bradford@utah.edu**
Brooks, Richard; **Clemson University; rrb@acm.org**
Burns, Patrick; **Colorado State University; patrick.burns@colostate.edu**
Cowles, Robert; **SLAC; rdc@slac.stanford.edu**
Crandall, Jedidiah; **University of New Mexico; jedcrandall@hotmail.com**
Daniels, Thomas; **Iowa State University; daniels@iastate.edu**
Dasgupta, Dipankar; **The University of Memphis; dasgupta@memphis.edu**
Elhanany, Itamar; **University of Tennessee; itamar@ece.utk.edu**
Enbody, Richard; **Michigan State University; enbody@cse.msu.edu**
Evans, David; **University of Virginia; evans@cs.virginia.edu**
Franz, Michael; **University of California, Irvine; franz@uci.edu**
Fu, Kevin; **UMass Amherst; kevinfu@cs.umass.edu**
Giffin, Jon; **Georgia Institute of Technology; giffin@cc.gatech.edu**
Govindarasu, Manimaran; **Iowa State University; gmani@iastate.edu**
Grossman, Robert; **University of Illinois at Chicago; grossman@uic.edu**
Hauser, Carl; **Washington State Univ; hauser@eecs.wsu.edu**
Humphrey, Marty; **University of Virginia; humphrey@cs.virginia.edu**
Jaeger, Trent; **Pennsylvania State University; tjaeger@cse.psu.edu**
James, John; **EE&CS, USMA; john-james@usma.edu**
Jha, Somesh; **University of Wisconsin; jha@cs.wisc.edu**
Jiang, Xuxian; **George Mason University; xjiang@gmu.edu**
Johnson, Erin; **University of Wisconsin: Eau Claire; johnsone@uwec.edu**
Khurana, Himanshu; **University of Ilinois; hkhurana@ncsa.uiuc.edu**
Kohno, Tadayoshi; **University of Washington; yoshi@cs.washington.edu**
Kupsch, James; **University of Wisconsin; kupsch@cs.wisc.edu**
Lee, Ruby; **Princeton University; rblee@princeton.edu**
Lee, Wang-Chien; **Penn State University; wlee@cse.psu.edu**
Lee, Wenke; **Georgia Institute of Technology; wenke@cc.gatech.edu**
Li, Jun; **University of Oregon; lijun@cs.uoregon.edu**
Li, Ninghui; **Purdue University; ninghui@cs.purdue.edu**
Li, Xiangyang; **University of Michigan – Dearborn; xylum@umich.edu**
Liu, Alex; **Michigan State University; alexliu@cse.msu.edu**
Livny, Miron; **U of Wisconsin-Madison; miron@cs.wisc.edu**
Markowsky, George; **University of Maine – Computer Science Dept; markov@maine.edu**
Maxion, Roy; **Carnegie Mellon University; maxion@cs.cmu.edu**
McDaniel, Patrick; **Pennsylvania State University; mcdaniel@cse.psu.edu**
McKinley, Philip; **Michigan State University; mckinley@cse.msu.edu**
Mirkovic, Jelena; **University of Delaware; sunshine@cis.udel.edu**
Moh, Melody; **San Jose State University; moh@cs.sjsu.edu**
Myers, Andrew; **Cornell University; andru@cs.cornell.edu**

Neuman, Clifford; **University of Southern California; bcn@isi.edu**
Nicol, David; **University of Illinois; nicol@crhc.uiuc.edu**
Park, Jung-Min ; **Virginia Tech; jungmin@vt.edu**
Park, Kihong; **Purdue University; park@cs.purdue.edu**
Ramamurthy, Byrav; **University of Nebraska-Lincoln; byrav_ramamurthy@ieee.org**
Reeves, Douglas; **N.C. State University; reeves@eos.ncsu.edu**
Reiher, Peter; **UCLA; reiher@cs.ucla.edu**
Sanders, William; **University of Illinois; whs@uiuc.edu**
Shmatikov, Vitaly; **The University of Texas at Austin; shmat@cs.utexas.edu**
Sion, Radu; **Stony Brook Network Security and Applied Cryptography Lab; sion@cs.stonybrook.edu**
Smith, Sean; **Dartmouth College; sws@cs.dartmouth.edu**
St. Sauver, Joe; **Internet2 and the University of Oregon; joe@uoregon.edu**
Striegel, Aaron; **University of Notre Dame; striegel@nd.edu**
Sundaram, Ravi; **Northeastern University; koods@ccs.neu.edu**
Thuraisingham, Bhavani; **The University of Texas at Dallas; bhavani.thuraisingham@utdallas.edu**
Tsudik, Gene; **Computer Science Dept., UC Irvine; gts@ics.uci.edu**
Welch, Von; **NCSA; vwelch@ncsa.uiuc.edu**
Winslett, Marianne; **University of Illinois; winslett@cs.uiuc.edu**
Wu, Felix; **University of California, Davis; wu@cs.ucdavis.edu**
Xu, Dongyan; **Purdue University; dxu@cs.purdue.edu**
Xuan, Dong ; **Dept. of Computer Science and Engineering, The Ohio-State University; xuan@cse.ohio-state.edu**
Yasinsac, Alec; **SAIT Laboratory; yasinsac@fsu.edu**
Zachmann, Dave; **Colorado State University; dzach@mesanetworks.net**
Zhang, Zhi-Li; **University of Minnesota; zhzhang@cs.umn.edu**
Zou, Cliff; **University of Central Florida; czou@cs.ucf.edu**

European *Universities*

Dacier, Marc; Eurecom Institute; dacier@eurecom.fr
McHugh, John; Dalhousie University; mchugh@cs.dal.ca

Non-Profit Organizations

Aiken, Robert; Qwest Government Systems; aikenr@acm.org
Chatterjee, Lali; IOP Publishing; lali.chatterjee@iop.org
Ioannidis, John; Packet General Networks, Inc.; ji@tla.org
Reid, Brian; ISC**;** Brian_Reid@isc.org
Soda, Lisa; API**;** sodal@api.org

Appendix C – Workshop Charge

Appendix C – Workshop Charge

Approximately 150 experts participated in the Cyber Security Research Needs for Open Science Workshop. The attendees included broad representation from national laboratories, higher education, and industry.

The charge to the workshop participants was to define Priority Research Directions (PRDs) relevant to DOE's open science and electricity delivery and energy reliability missions. The workshop participants focused on long-term research directions in the cyber sciences with a 5 to 10-year time frame for advances. A parallel focus was on intersecting the research directions with energy control systems with a 3 to 10-year time frame for deployable delivery.

The breakout sessions encompassed the following topics and leaders:

- **Securing Hardware, Software and Data (SHSD)** –Frank Siebenlist and Len Napolitano
- **Monitoring and Detection (MD)** – Troy Thompson and John McHugh
- **Future Security Architectures and Information Assurance Technologies (FSA)** – Tom Harper
- **Human Factors Analysis (HF)** – Anne Schur and Joe St. Sauver
- **Protecting Our Utility Infrastructure (UI)** – Jeff Dagle, Aaron Turner, and Bill Young.

Altogether, 27 final PRDs were identified in the breakout sessions. However, some were concatenated, and two were not submitted. Appendix D contains the concatenated, submitted PRDs.

Appendix D – Panel Breakout Sessions and Reports

Appendix D – Panel Breakout Sessions and Reports

Securing Hardware, Software and Data (SHSD)

- Frank Siebenlist, Argonne National Laboratory
- Len Napolitano, Sandia National Laboratories

Monitoring and Detection (MD)

- Troy Thompson, Pacific Northwest National Laboratory
- John McHugh, Dalhousie University

Future Security Architectures and Information Assurance Technologies (FSA)

- Tom Harper, Idaho National Laboratory

Human Factors Analysis (HF)

- Anne Schur, Pacific Northwest National Laboratory
- Joe St Sauver, Internet2

Protecting our Utility Infrastructure (UI)

- Jeff Dagle, Pacific Northwest National Laboratory
- Aaron Turner, Idaho National Laboratory
- Bill Young, Sandia National Laboratories

BREAKOUT SESSION
SECURING HARDWARE, SOFTWARE AND DATA (SHSD)

Breakout Leads: *Frank Siebenlist, Argonne National Laboratory*
Len Napolitano, Sandia National Laboratories

Securing hardware, software and data to provide data and communication integrity is a fundamental Mathematics and Computer Science R&D problem. While today many believe that the network is to “blame” for security problems, it is a far more basic problem that needs to be looked at from an end-to-end systems perspective. Each component of a system in an open science environment, including the hardware, software, or the data itself, may allow a possible security breach point. Each component must be protected to enable a fully secure and trusted system in an open science environment.

The charge to this panel is to identify Priority Research Directions in this area related to:

a. TPM chip and embedded hardware identity technologies
b. Theft reporting and discovery technologies
c. Tripwire technologies
d. Secure operating systems and applications
e. Virtualized technologies
f. Domain name service integration, registering special services and machines
g. Encryption and decryption technologies
h. Data integrity technologies
i. Special challenges in mobile devices

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-1: EVALUATING CYBER SECURITY PRACTICES IMPACT (MANAGING TRUST)

ABSTRACT

Many common cyber security practices are not appropriate for the global open science environment, which is characterized by enormous scale and scope, demand for high-performance, geographic diversity, competing multi-national interests, heterogeneity of computing systems, and non-hierarchical management. In addition, no good method is available to develop and manage trust relationships between different parties−a key to the operation of open science collaborations. This Priority Research Direction will address these issues by supporting research that will develop tools to manage the full lifecycle of a trust relationship and to model and analyze cyber security environments so as to evaluate the effectiveness and risk/benefits of particular practices. This evaluation is especially important where differing practices are mandated by different international and organizational bodies.

EXECUTIVE SUMMARY

The problems of inappropriate and inconsistent cyber practices and inability to manage trust will be attacked by a research program that includes:

- Development of a comprehensive set of tools to manage the full lifetime of a trust relationship between parties, including how the relationship is defined (naming the entities to the relationship, the actions to be taken, and the conditions under which the actions are performed); categorized (determining the necessary level of assurance); monitored (so all parties are assured that the trusted behavior takes place); published (so all parties can see at all times what relationships exist); terminated; and restored when broken. These trust relationships are dynamic and evolving and can be both one-to-one and many-to-many.
- Development of modeling, simulation, and analysis tools that will allow a cyber security environment instantiating a particular set of practices to be described and evaluated for its impact on various science goals (especially computing system performance). This includes the adoption of a cyber security ontology to allow commonly understood descriptions of security environments and practices.
- Development of mechanisms and methodologies to resolve conflicts between policies imposed by differing geographic or organizational bodies, to assess risk/benefit tradeoffs of particular practices, and to evaluate practices based on effectiveness in the real world.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Current challenges in the cyber security practices domain fall into three major areas:

- evaluating security practice impacts, especially in cases of conflicting policies from multiple policy domains
- managing trust relationships between collaborating bodies
- bringing practices in line with open science needs and incorporating real-world, effectiveness-driven policy validation.

In more detail:

- Evaluating policy: any imposition of security policies will restrict or limit computing activities. The benefits from the added security provided due to these policies should outweigh the costs (both direct costs and indirect costs due to foregone computing activities) incurred due to these policies. The makers of policy have developed good feelings for balancing these costs and benefits in ordinary, single-site computing environments.

  However, these standard cost/benefit tradeoffs fail in the open science environment due to the fundamental collaborative-distributed nature of open science. Policies are imposed on multiple sites with differing computing environments, making it difficult to correctly evaluate the true costs and benefits of particular policies. Moreover, since the open science environment spans multiple policy domains, there are often conflicting policies imposed by different policymaking authorities, making the cost/benefit evaluations even more difficult to perform. These differing policy domains will frequently have differing goals, missions, and concerns, and be subject to differing legal or regulatory constraints.

  Thus, new modeling and evaluation tools are needed to allow the cost/benefit tradeoffs to be performed in distributed open science environments that are subject to multiple policy domains.

- Trust relationships: the operation of open science computing environments, which are fundamentally collaborative rather than hierarchical in management structure, requires a new form of management tools. In particular, such enterprises rely on trust relationships, defined as an instance where partners in a collaborative enterprise can rely on other partners to perform specific actions under well-defined circumstances for a defined period of time, without needing to check each instance where the action is performed.

  These trust relationships can be characterized by several lifetime stages. They must be defined (naming the entities party to the relationship, the actions to be taken, and the conditions under which the actions are performed); categorized (determining the necessary level of assurance); monitored (all parties are assured that the trusted behavior takes place); published (all parties can see at all times what relationships exist); terminated; and restored when broken. Until now such relationships are managed in an ad hoc manner. Fully performant open science computing environments require more precise definition and management of these relationships.

- Policy formulation and validation: even with the use of modeling and simulation tools to allow selection of appropriate policies in multiple domain environments, open science progress will still be held back unless policy validation moves away from a compliance-oriented checklist approach to a real-world, data-driven effectiveness approach. Auditing tools must be developed that measure the results of policy implementations by their effect on actual improvements in security as assessed by historical performance rather than by assessing compliance with an arbitrary set of prescriptive standards.

  Finally, contact must be made with standards and policy-formulating bodies on an ongoing basis to ensure they are aware of the impact of their recommendations on open science communities.

SUMMARY OF RESEARCH DIRECTION

- Evaluation of cyber security practice impact: this research will first develop a set of procedures and tools that can characterize a security policy environment, including adoption of a standard security ontology, and evaluate the impact of these practices on science objectives. This will consist of:
  - a list of areas that requires security policies that spans the full space of policy requirements
  - a categorization of policy needs of different environments
  - a set of modeling and simulation tools that allows any given set of policies to be compared against the performance needs of the open science community
  - a process for monitoring and comparing policies and their results on security and science productivity
  - processes for evaluating and comparing quantitative or qualitative benefits of risk mitigation (by security policy) vs. costs (both direct and opportunity costs) of implementing said policies.

At this stage, the project plays only an observational role, using descriptive and analytical techniques.

Finally, at a later stage of the research, processes will be developed to suggest syntheses of differing policies imposed on open science collaborations by different policy bodies. Making use of modeling tools developed in earlier stages of the research, sets of model/template policies will be developed that meet needs of multiple policy domains, and general purpose processes will be described that can coordinate policies without stifling the needs of the open science community.

- Trust relationships: this research will analyze the stages of a trust relationship, produce thorough descriptions of each stage in the lifecycle of a trust relationship (including those stages described above together with others that arise in the course of the research), and develop and support a set of tools to carry out and enforce the various lifecycle stages.
- Real-world policy formulation and validation. Research in this area will be focused on three main topics:
  - Multi-environment use of modeling and simulation tools to predict effectiveness of policies in real-world situations. The complex interdependencies inherent in open science require a sophisticated use of the policy analysis tools developed above. Particularly when multiple sets of policies are imposed on heterogeneous global collaborations, the impact of each policy may not be obvious. Standard processes will be developed to avoid careless imposition of inappropriate policies in such domains.
  - Data-driven policy validation based on real-world measures of policy effectiveness. All security policies should be designed so as to provide a concrete improvement in cyber security. The effectiveness of these policies should thus be assessed by looking at concrete instances of improvements in security metrics in real-life situations. This research will develop a comprehensive set of such metrics together with tools for computing them, and incorporate these into procedures to perform policy assessments. Particular attention should be paid to policies that do not provide any actual improvement in real-world security, with an eye towards removing such policies as requirements.
  - Processes for expressing scientific needs and incorporating these requirements into the deliberations of standards and policy-formulating bodies.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Many communities other than open science collaborations will benefit from the use of organized mechanisms to manage trust. The field of cyber security at large also will benefit from techniques that ensure formulation of tailored and appropriate policies that provide real-world effective security rather than mere compliance and defense against last year's threats. The research program will involve collaboration between security experts, computational scientists, and domain science experts as well as other less traditional disciplines (sociologists, engineering psychologists, etc.) to better match policies to security and science objectives. Beneficial results of the research will spread as the results are fed back to policy formulation and standards bodies such as the National Institute of Standards and Technology in the United States.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Open science collaborations will be more effective and secure because they avoid policy collisions between multiple policy domains and are governed by security policies and practices that are appropriate to their scientific goals and risk environment. Policies will be more dynamic and flexible, which will allow open science to take advantage of technological advances more rapidly. Since much of this work is done at the leading edge of computing technology, it is vital not to hold back the exploitation of new technology with outdated or inappropriate security policies. Global collaboration will be facilitated by trust management tools that encourage worldwide open science projects.

TIME FRAME

- Development of trust relationship tools should be pursued in the short term (3 to 5 years) because these tools will have an immediate impact as soon as they are available.
- Modeling and other security practice evaluation tools will be developed over a longer term (5 to 10 years) incrementally so that early versions will have an immediate impact as fuller versions dealing with increases in complexity and scope of open science collaborations are under development.

Authors: Irwin Gaines, FermiLab; Don Petravick, FermiLab; Von Welch, NCSA; Bob Cowles, SLAC/Stanford; Bhavani Thuraisingham, University of Texas at Dallas; Carmen M. Pancerella, Sandia National Laboratories.

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-2: TRUST HARDWARE AND CRYPTO ACCELERATION

ABSTRACT

Research vectors related to building trusted hardware and crypto acceleration mechanisms for open science are described in this report. The potential impact of this work to the global computation science community and to open science is significant. First, mechanisms for fast crypto will be available, allowing users to access data at link speeds much greater than what is supported today. Second, there will be a set of scalable mechanisms for trusting hardware, software, users, and data. Third, secure and scalable user authentication will be made possible. Finally, fast integrity mechanisms will be able to ensure the trusted execution of shared software, guaranteeing that no viruses or adversarial interferences disrupt the proper execution of code.

EXECUTIVE SUMMARY

Four main challenges associated with enhancing cyber security for open science are described:

- the need for trust anchors between entities in a virtualized world
- the need for accelerating the performance of cryptographic algorithms
- a discussion that we do not yet know how to build scalable and secure hardware for user authentication
- the challenge of having mechanisms to inform us whether programs and data have been modified by viruses, worms, or malicious users.

A research vector on building hardware and software systems that can potentially execute cryptographic algorithms much faster than the state of the art is described, and research for building distributed mechanisms for scalable key management and user authentication is discussed. Another research vector is related to the need for development of hardware that allows virtual machines to trust physical resources. Finally, research concerning the development of security mechanisms for using multi-core processors and the development of algorithms for runtime checking of software vulnerabilities is outlined.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

One of the main challenges associated with building trusted hardware for open science is creating trust anchors between entities in a virtualized world. Current and future processors are likely to support an increasing number of features for hardware, operating systems, and application virtualization. While virtualization allows for more efficient resource management and better usability of a computer system, it creates a number of security challenges because users need to trust the entities providing physical resources. For example, it would be unacceptable if the physical resource supporting the execution of a virtual machine is located in a compromised server or under the control of an adversary. The current state of the research on trusted hardware is reflected in the following selection of references: Lee et al. 2005; Kwan and Durfee 2007; Dwoskin and Lee 2007; Arora et al. 2005; Piromsopa and Enbody 2006; Lee et al.

2003.Another challenge has to do with the performance of the cryptographic algorithms used for open science. The open science project will need to support fast, secure access to large amounts of data by many different users simultaneously. Therefore, the performance of cryptographic algorithms supporting secure communication becomes critical. Developing cryptographic hardware and software for accelerating symmetric key cryptography, e.g., efficient advanced encryption standard (AES) acceleration hardware, as well as public key operations, e.g., a modular exponentiation engine for accelerating RSA, and cryptographic hash functions is an open research issue, especially for the speeds and inter-operability needed for open science. The following references provide background for the current state of research in crypto acceleration: Koc et al. 1996; Satoh et al. 2001; Schroeppel et al. 2005; Fiskiran and Lee 2005; Scheibelhofer 2007; Kounavis 2007; Hilewitz and Lee 2007.

A third challenge is that we do not yet know how to build scalable and secure hardware for user authentication and secure key management of a very large number of keys. As the number of users scales up—as expected for open science—issuing, managing, and verifying certificates and keys for all these users becomes difficult. Open issues include whether there should be a centralized authority for issuing certificates or whether trust should be supported in a distributed manner and whether keys should be managed by a separate trusted platform module or by software with trust anchored in a secure processor. One aspect of this broad challenge is to investigate hardware-based trust anchors that can significantly enhance distributed user authentication, as well as data and program confidentiality and integrity.

A last challenge is related to the need for trusted execution of shared software. It is important for open science to have mechanisms to inform us whether programs and data have been modified either by viruses, worms, or malicious users or other sources of error such as noise.

SUMMARY OF RESEARCH DIRECTIONS

The first research vector proposed for supporting cyber security in open science is crypto acceleration. Crypto acceleration is the development of hardware and software systems tailored to open science applications that execute well-known, as well as new, cryptographic algorithms faster than the current state-of-the-art. Since most operations involved in cryptographic processing are complex mathematical computations (large-number multiplications, inversions and exponentiations, and elliptic-curve operations) or sophisticated bit manipulations as in some block or stream ciphers and hash functions, the plan is to investigate new algorithms for such computations and potentially discover much faster ways to implement crypto in hardware and software (Schroeppel et al. 2002). This includes both specialized hardware accelerators and more generalized microprocessor enhancements, as well as faster software-hardware implementations.

A second research vector is related to the development of distributed mechanisms for scalable key management and user authentication. One approach to user authentication is to have a centralized authority issuing certificates and signing public keys, as mentioned earlier. While such an approach is easy to build, works well for small or medium-size networks, and also can leverage the hierarchical public-key infrastructure already established in many areas, there are still associated scalability and trust problems when applied to a large, international community such as open science. Furthermore, each user or entity also has lots of different keys for encryption, decryption, and keyed hashes, in addition to public/private keys used primarily for authentication. A need exists to investigate scalable, distributed trust models for secure user authentication that use hardware-rooted trust for enabling more secure storing and managing of keys by software systems.

A third research vector is related to the development of hardware for allowing users of virtual machines to trust physical resources. We would like to investigate mechanisms for building a hardware-based trust anchor between entities in a virtualized world such as virtual machines and virtual machine monitors. This research vector is related to the key management challenge, because each virtual machine and virtual machine monitor may need to support its own public/private key pair and credentials, or some other means for authentication and attestation. It is also related to the crypto acceleration challenge because entity authentication and attestation will need to be done as quickly as possible.

Other research vectors proposed include the development of security mechanisms for using multi-core processors, including core authentication and core-to-core encryption protection against side channel attacks, and the development of algorithms for runtime checking of software vulnerabilities. Such research may include the investigation of significantly faster cryptographic hash functions for dynamic code integrity checking, and the interaction between mechanisms for integrity and confidentiality.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The potential impact of this work to the global computational science community is significant. First, we will have mechanisms for fast crypto, allowing users to access data and communicate at link speeds much greater than what is supported today. For example, today software implementations of AES can support 1 Gbps communication. With appropriate hardware support this processing may be accelerated by factors of 10 or 100 or more, allowing links of 10-100 Gbps or even greater speeds to be fully utilized when carrying encrypted traffic. In addition, multiple encryption, hashing, and authentication algorithms may be supported by generalized rather than specialized crypto acceleration features in programmable processor cores to enhance flexibility and inter-operability in international settings.

Second, a set of scalable mechanisms for hardware trust anchoring (beyond the trusted platform module) will be available for trusting hardware, software, users, and data. This will enable rapid use of processor virtualization and efficient resource management. With trust anchoring extended to multi-core processor technology, it will increase the level of usability of multi-core processor architectures that are likely to be the basis for the computing nodes in future infrastructures including the global network for open science.

Third, secure and scalable user authentication and key management will be made possible. Millions of users will be able to authenticate themselves without incurring the overheads of centralized key management while new inventions will allow the dynamic introduction or deletion of users at very high speeds. Last, fast integrity mechanisms will be able to ensure the trusted execution of shared software, guaranteeing that no viruses or other forms of malware disrupt the proper execution of code.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Specifically to open science, the research will allow securing data and metadata in an open collaborative environment as well as trusted sharing of code, data, and tools.

TIME FRAME

Crypto Acceleration

- Algorithm design, implementation, testing, and certification: 3 to 5 years
- Hardware design, implementation, and recommendations: 3 to 5 years

Key Management/User Authentication

- Trust model design: 2 to 3 years
- Hardware implementation and recommendations: 3 to 5 years

Trust Anchoring

- Process/algorithm design: 2 to 3 years
- Hardware implementation and experimentation: 3 to 5 years

Safe Code Execution

- Design for trusted execution of shared software: 3 to 5 years
- Cryptanalysis, hardware implementation, experimentation: 5 years+

REFERENCES

Koc, C., T. Acar, and B. Kaliski, 1996. “Analyzing and Comparing Montgomery Multiplication Algorithms,” IEEE Micro, 16, 3, 26-33.

Satoh, A., S. Moriokah, J. Takano, and S. Munetoh. 2001. “A Compact Rijndael Hardware Architecture with S-box Optimization,” in Proceedings, Workshop on Cryptographic Hardware and Embedded Systems (CHES 2001), 175-188.

Schroeppel, R., C. Beaver, R. Gonzales, R. Miller, and T. Draelos. 2002. “A Low-Power Design for an Elliptic Curve Digital Signature Chip,” in Proceedings, Workshop on Cryptographic Hardware and Embedded Systems (CHES 2002), 366-380.

Fiskiran, A. M., and R. B. Lee. 2005. "On-Chip Lookup Tables for Fast Symmetric-Key Encryption," in Proceedings of the IEEE 16th International Conference on Application-Specific Systems, Architectures and Processors (ASAP), 356-363.

Scheibelhofer, K. 2007. "A Bit-Slice Implementation of the Whirlpool Hash Function," in Proceedings, RSA Conference 2007, Cryptographer's Track (CT-RSA), 385-401.

Kounavis, M. 2007. "A New Method for Fast Integer Multiplication and its Application to Cryptography," in Proceedings, Symposium on the Performance Evaluation of Computer and Telecommunication Systems (SPECTS), San Diego, California.

Hilewitz, Y., and R. B. Lee. 2007. "Performing Advanced Bit Manipulations Efficiently in General-Purpose Processors," in Proceedings of 18th IEEE Symposium on Computer Arithmetic (ARITH-18), 251-260.

Lee, R. B., P. C. S. Kwan, J. P. McGregor, J. Dwoskin, and Z. Wang. 2005. "Architecture for Protecting Critical Secrets in Microprocessors," in *Proceedings of the 32nd International Symposium on Computer Architecture (ISCA 2005)*, 2-13.

Kwan, P. C. S., and G. Durfee. 2007. "Practical Uses of Virtual Machines for Protection of Sensitive User Data," in *Proceedings of the 3rd Information Security Practice and Experience Conference (ISPEC 2007)*, LNCS 4464, 145-161.

Dwoskin, J., and R. Lee. 2007. "Hardware-rooted Trust for Secure Key Management and Transient Trust," *Proceedings of the 14th ACM Conference on Computer and Communications Security.*

Arora, D., A. Raghunathan, S. S. Ravi, and N. K. Jha. 2005. "Enhancing Security through Hardware-assisted Run-time Validation of Program Data Properties," in *Proceedings of ACM/IEEE International Conference on Hardware Software Co-design and System Synthesis (CODES+ISSS)*, 190-195.

Piromsopa, K., and R. Enbody. 2006. "Secure Bit: Transparent, Hardware Buffer-Overflow Protection." *IEEE Transactions on Dependable and Secure Computing*, **3**, 4, 365-376.

Lee, R. B., D. K. Karig, J. P. McGregor, and Z. Shi. 2003. "Enlisting Hardware Architecture to Thwart Malicious Code Injection," in *Proceedings of the International Conference on Security in Pervasive Computing (SPC-2003)*, D. Hutter, Ed., pp. 237-252, Springer Verlag, Berlin.

Authors: Mike Kounavis, Intel; Ruby Lee, Princeton University; Tim Draelos, Sandia National Laboratories; Richard Enbody, Michigan State University

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-3: SOFTWARE SECURITY

ABSTRACT

Systems that are used by the U.S. Department of Energy (DOE), such as the Open Science Grid (OSG), are composed of several diverse software components. As a consequence, vulnerability in one software component can impact a large system and have dire consequences. Hence, there is a critical need to develop analysis techniques and methodologies to analyze large systems for security properties. There have been significant advances in developing analysis techniques and tools for security analysis of software. However, these techniques do not have adequate diagnostics and the more sophisticated analyses are not scalable enough to handle large DOE software systems.

Systems in DOE are frequently distributed and have a large number of software components (perhaps from diverse sources). Techniques need to be developed to analyze end-to-end security properties of large distributed systems. Some of the existing analysis tools are very cumbersome to use and thus are not widely deployed. Programmers need to be trained to use these tools and a methodology designed to develop secure software. Moreover, usability of these analysis tools must be improved. New research is required to address the scale and unique features that effect security of open science within DOE.

EXECUTIVE SUMMARY

Research is required to develop analysis techniques and methodologies that will enable analysis of large-scale DOE systems (perhaps composed of many components). Such work will provide a measurable degree of security assurance for DOE open science. Analysis techniques and tools will have to provide better diagnostics and support for fixing the identified vulnerabilities. Research also is important to improve usability and scalability of these tools.

New analysis techniques need to be designed, and such analyses must provide better diagnostics to developers when vulnerabilities are discovered. This will allow developers to understand and design fixes for discovered vulnerabilities. Currently, a large portion of DOE software, such as the OSG, executes in the privileged mode, which causes the consequences of an exploit to be more severe. New techniques are required to minimize the trusted computing base (the portion of software executing in privileged mode) of large DOE applications.

It is crucial to determine the end-to-end security of the software stacks that power our open science infrastructure by performing independent vulnerability assessment that proactively identifies and repairs security problems at all layers and components. Just as program correctness is best achieved by an independent quality assurance team, so software security can only be achieved through an independent effort. Such an in-depth effort requires a multi-modal security analysis framework that is a based on a balanced combination of skilled practitioners and automated tools. As such, a comprehensive research program is needed to support the building of knowledge, methodologies, active training and dissemination of skills, alongside an investment towards significant advances in analytical tool technology.

We have identified various tasks to be accomplished to achieve these goals. If accomplished, the results will be usable and scalable analysis tools. These tools can be used to analyze components used in DOE applications. Because DOE applications are composed of several components, we also have identified tasks related to establishing end-to-end security.

SUMMARY OF RESEARCH DIRECTION

Improve diagnostics of analysis tools. Significant advances in tools for analyzing software components have occurred. However, these analysis techniques need to be designed for providing explanation of errors/violations. These analysis techniques also need to be enhanced to provide feedback to help suggest fixes. Current analysis techniques are geared towards just reporting violations (this is due to bias of these traditional applications to compiler optimization). We need additional research to address false positive rates, prioritization of flaws, ties to security flaw classifications (CWE, SAMATE, etc.), and accommodating distributed computing (grid computing).

Language restrictions/extensions to support secure software development. Concepts of language restriction can be applied to existing languages currently used for DOE applications to make the languages more secure. Language extensions can be defined through libraries to provide security-specific abstractions whose proper use can be enforced at compile time. The development of applications within such secure language subsets with security abstractions provides for more powerful security analysis techniques to be applied, such as theorem proving, bounded model checking, and related techniques from formal methods. Research in this area would be particularly practical since it does not require development of new languages and can be applied to legacy code.

Automated correction of selected security flaws. The development of patches to automatically correct restricted sets of security flaws would significantly improve the productivity of software developers and increase awareness of security flaws within commonly developed applications. This research would especially impact security of legacy DOE applications. A basis for this work would be research defining source-to-source code generation mechanisms that match the original application code.

Reducing the trusted computing base (TCB). Currently, software components that require sensitive operations are all executed in trusted mode. We need analysis techniques for partitioning code into trusted and untrusted components. This will enable software that requires trusted operations to be small, i.e., it reduces the trusted computing base. There are techniques for partitioning (Brumley and Song 2004), but we need advances to make these techniques scalable and address distributed applications.

Enable software components to comply with site policy. To support the development of more secure software, libraries (abstractions) are needed that implement a customizable site policy so that software components follow the site policy as well as policies to interact with these libraries and ways to define such policy. We also need analysis techniques to certify that software components comply with the policy. This will enable software components to comply with site policy.

Need to analyze end-to-end security. We need to enhance existing analysis tools and methodologies to enforce end-to-end software security. Analysis techniques need to deal with large interconnection of components from different vendors. Techniques also need to be developed to aid in the analysis of vulnerabilities inherent in the architecture of the system as a whole.

Improve the usability of analysis tools. Frequently, these analysis tools are very cumbersome to use. We need a training methodology for programmers to enable them to use these tools. Analysis techniques need to be made more usable. One aspect of usability is presentation of identified errors/violations to the programmer, e.g., how do you present a vulnerability that involves multiple lines of code?

Analyzing Commercial Off-the-Shelf (COTS) software binaries. Frequently, software components are used in DOE systems, such as the Open Science Grid (OSG) [Open Science Grid], for which source code is not available (such as COTS). We need techniques for analyzing COTS binaries (Miller, et al. 1995; Reps, et al. 2005) to increase trust in this software. In this context, trust is defined as the software which correctly performs its intended purpose, and does not do anything additional which might compromise the integrity or security of the system or the data stored on the system. This is especially important with security significant software, such as anti-virus, encryption, virtual machine monitors, and operating systems. This will allow COTS to be used in a more secure manner.

Framework to accomplish multi-modal analysis. Developing a multi-modal frame requires research in techniques and methodologies to aid in all phases of the vulnerability assessment, especially those unique to the domain of distributed open science middleware and applications. Recent experience has shown that only half of currently identified vulnerabilities can be found using existing tools. This gap comes from extreme false-positive rates (from finding both non-vulnerabilities and vulnerabilities that are not effectively exploitable) and missed vulnerabilities due to complex interactions. Therefore, any recent agenda on automating and improving the ability to find vulnerabilities must be driven by a top-down, architecture based analysis of the code. This synergistic approach requires two parallel research directions. First, we must develop new analysis techniques and tools for efficiently extracting architectural structure, privilege levels, and key resources used in complex distributed codes. Second, we must evaluate the current gap (the 50% gap) in automated vulnerability assessment tools to find new joint techniques that can use the analysis techniques to drive a more effective search for exploitable vulnerabilities.

Virtual-Machine reference monitor rules. The use of virtual machines to execute software permits new opportunities to define restricted subsets of the general machine to support the application. Applications that use functionality outside of a tightly enforced subset of the operating system and machines can be regarded as violating the virtual-machine rules. Defining the rules to specify such restricted virtual machines is a software analysis issue. Software source code analysis and/or binary analysis (static and/or dynamic analysis) will form a critical part of defining the rules by which to instantiate virtual machines. Binary analysis has been used to generate policies for host-based intrusion detection systems (Feng et al. 2004). However, new techniques have to be developed for the virtual machine context. Such techniques can be expected to provide a significant additional level of security for DOE open science software.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The large scale and distributed nature of open science software will be a significant challenge that is specific to DOE open science. Verification of trusted inputs and techniques to do analysis of impact of non-trusted inputs on software of this scale within the DOE OSG will be of critical importance to attaining high degrees of software assurance and credibility of results.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Software security is critical to the protection of DOE resources controlled by software developed for open research science (such resources include computers, experimental equipment, etc.). Software security is similarly important for critical software-oriented control systems that are used within national infrastructure regulated by DOE. Open science is conducted on a worldwide scale using software written by diverse sources, often without regard or knowledge of cyber security issues or secure software development practices. DOE software applications are complex, large scale and can control particularly critical national infrastructure. Automated forms of software analysis and a methodology for using it will form a critical part of improving cyber security of DOE open science in the future.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

The requirements of open science within DOE are such that new motivations will drive the development of new technologies and broaden existing technologies. The required application of security analysis to large scale DOE applications and distributed component infrastructure along with the availability of significant DOE parallel computing resources will pioneer the development of parallel algorithms for security analysis. The requirements of security will drive a different style of analysis. Specifically, forms of analysis that can save meta-data required to provide developers with information and contribute insight to understand identified security flaws. In contrast, more conventional program analysis has targeted performance optimization where there are no such requirements to provide diagnostic information to developers about identified security flaws. The requirement to report security flaws in large scale applications will in turn also stimulate research to prioritize security flaws, such work could even outline the consequences of security flaws.

At present, no technology exists to automatically heal large scale software applications of automatically identified security flaws. Such work could significantly simplify making large scale DOE applications secure and address the security vulnerabilities of DOE substantial legacy code base. Such work would advance security for open science while pushing the state-of-the-art for cyber security.

TIME FRAME

- A number of different level efforts from delivery of existing security analysis technologies to the research required to address larger-scale software assurance within DOE will need to be proposed.
- The release of existing technologies into tools would occur on a 3-5-year schedule while the development of capabilities requiring research would occur on a 4-8- year timetable.

REFERENCES

Brumley, D. and D. Song. 2004. “Privtrans: Automatically Partitioning Programs for Privilege Separation.” In *Proceedings of the 13th Usenix Security Symposium.*

Feng, H. H., J. T. Giffin, Y. Huang, S. Jha, W. Lee, and B. P. Miller. 2004. Formalizing Sensitivity in Static Analysis for Intrusion Detection,” *IEEE Symposium on Security and Privacy*, 194-208.

Miller, B. P., M. D. Callaghan, J. M. Cargille, J. K. Hollingsworth, R. B. Irvin, K. L. Karavanic, K. Kunchithapadam, and T. Newhall. 1995. "The Paradyn Parallel Performance Measurement Tool," *IEEE Computer* ***28****, 11*, 37-46. (Special issue on performance evaluation tools for parallel and distributed computer systems.)

Reps, T., G. Balakrishnan, J. Lim, and T. Teitelbaum. 2005. “A Next-generation Platform for Analyzing Executables.” In *Proceedings of the 3rd Asian Symposium on Programming Languages and Systems*, Tsukuba, Japan, Springer-Verlag, New York.

Open Science Grid, http://www.openscience.org.

Authors: D. Quinlan, Lawrence Livermore National Laboratory; J. Kupsch, University of Wisconsin; S. Jha, University of Wisconsin; G. White, Lawrence Livermore National Laboratory; M. Livny, University of Wisconsin; B. Miller, University of Wisconsin

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-4: END-TO-END DATA SECURITY

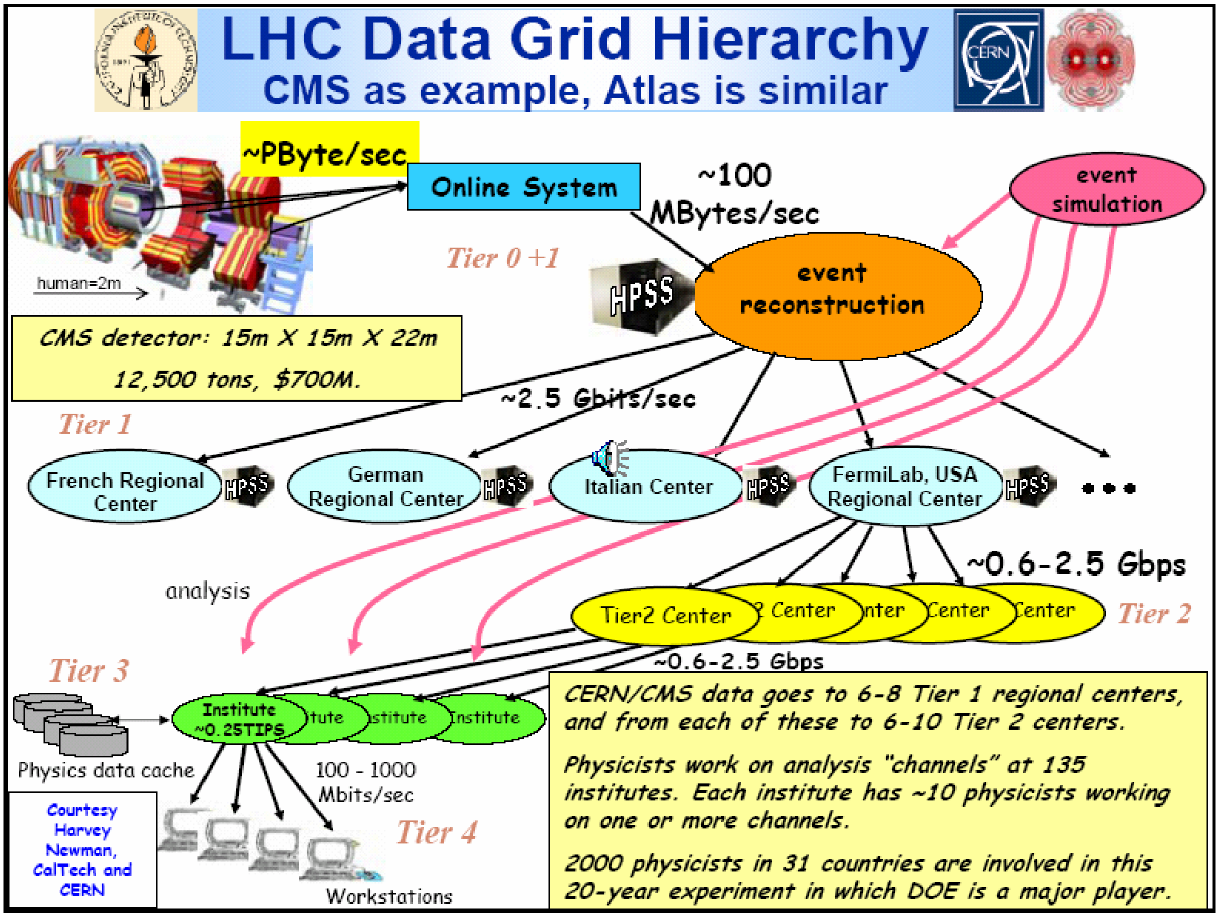
ABSTRACT

The ultimate goal of research on securing open science data is for users to have complete confidence in the ability of the system to protect their data against inappropriate modification and use. The large scale, shifting composition, and diverse nature of the U.S. Department of Energy (DOE) open science community present unique challenges in meeting this goal. To guarantee end-to-end data security, **new approaches are needed for managing information sharing in distributed collaborations with heterogeneous environments and potentially conflicting policies.**

EXECUTIVE SUMMARY

DOE open science projects are unique in having enormous geographically distributed datasets that are read and written by thousands of collaborators all over the globe. For example, the Large Hadron Collider (LHC) project involves 2,000 physicists at over 100 institutes in 31 countries, all of whom store and transfer portions of LHC data and actively participate in its analysis, at data rates of 2.5 Gb/s (see Figure 1). These unique characteristics lead to special challenges in securing open science data, both at rest and on the wire–especially given the limited security expertise of the participating scientists.

Open science data faces confidentiality and integrity threats all along the path from the devices where it is generated, on the network, intermediate caches in the network, and the ultimate storage devices. Further, there can be no single uniform policy about how to protect all data from an experiment; for example, some data owners may require strong encryption for privacy, while most will not. The primary threat that we aim to guard against is inappropriate creation, modification, or loss of data, whether inadvertent, e.g., packet loss, disk failure, a user giving the wrong name to a new file, or deliberate. For users to be able to trust data from open science, we need **user-friendly tools for managing information sharing** across boundaries of organizations with potentially conflicting policies and little security expertise. We also need **flexible security models** that adapt the employed policy based on a combination of properties including data classification, user attributes, and environment. Finally, user trust in data integrity requires an understanding of how the data were generated and processed; for this we need automated tools for **provenance tracking** and validation, and automatic transfer of such metadata when data are moved. Cost-effective approaches also are critical for guaranteeing the **long-term integrity of data** that are too big for conventional backup approaches.



LHC Data Grid Hierarchy

CMS as example, Atlas is similar

CERN/CMS data goes to 6-8 Tier 1 regional centers, and from each of these to 6-10 Tier 2 centers.

Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

2000 physicists in 31 countries are involved in this 20-year experiment in which DOE is a major player.

Figure 1: Illustration of the LHC Data Grid.

Summary of Research Directions

The protection of data in an open science setting introduces unique security challenges related to storage. An open science project can be thought of as a large virtual organization (VO) (Foster et al. 2001), in the form of an international collaboration involving thousands of scientists. A complete solution for open science data security has to account for a variety of users, computing environments, and data classifications; it has to provide verifiable proof of enforcement; and the security model has to be applied at all points between the data source and the storage device. To address these areas, this paper recommends research be directed in three particular areas:

1. user-friendly tools for management of a coherent security policy across administrative domains
2. flexible security models that adapt the security policy based on document classification, user permissions, and environmental considerations
3. identifying critical provenance information and developing methods to automatically extract and manage provenance for security considerations.

Scientific and Computational Challenges

VO and grid security is an active topic of research today (Foster et al. 1998, 2001; Welch et al. 2003), but the state-of-the-art grid security does not offer tools to help set up and manage large collaborations in a distributed manner, analyze sets of policies to determine what end-to-end guarantees are provided, or to explain authorization decisions to frustrated users. This section discusses the challenges specific to open science for each of the targeted research directions.

- Tools to Manage Distributed Security Policy

  VO members typically have little interest or expertise in security, yet are very concerned with ensuring data integrity; they need tools to help them manage authorizations in the VO environment. Because VO membership can change daily, e.g., as graduate students come and go, it must be easy to update VO membership lists and security-related attributes. No single approach to managing authorization will be appropriate for all open science projects, or for all sites participating in a particular VO. For example, different data products and tools will intrinsically have different levels of sensitivity, e.g., export controls. The policies regarding access confidentiality and privacy may be different in different countries; policies can even directly conflict with one another. Sites participating in the VO will have heterogeneous hardware and software and operate with substantial or complete autonomy.

- Flexible Security Models

  A heterogeneous user community introduces a number of interesting research questions about appropriate security models. The traditional UNIX model for data protection provides an access-control scheme that allows three types of access (read, write, and execute) to three different user lists (owner, group, and other); however, this scheme is insufficient to capture the variety of users and document classifications that may be shared in an open science project. For example, a document may have particular export controls that prevent it from being readable by group members from certain countries. Similarly, a dataset may have proprietary constraints that prevent collaborators that are not part of a non-disclosure agreement from a particular type of access. Rather than exclude these documents from an open science collection, we should have more expressive models for access control that allow a particular type of access (more than just read, write, and execute) based on an extensible set of attributes for users and data.

  Another interesting issue deals with selecting the appropriate security model based on environmental concerns. For example, in a tightly controlled environment such as a DOE facility for supercomputing, the security policies to enforce on a dataset may be more relaxed. Imposing unnecessary security, e.g., encryption, in a secure network adds overheads that severely hinder performance of a tightly coupled scientific simulation. However, as the data leaves that environment to perhaps move to a shared archive, we need to ensure that it receives the appropriate protections for privacy and integrity. This issue is also relevant to *distributed workflow* security (Gudes et al. 1999). A workflow consists of a number of distributed "tasks," each tied to a potentially different security policy. While it might seem desirable to have a consistent security policy applied to a distributed dataset across all environments, it is more logical (and practical) to define a level of protection to apply to the dataset as a whole. The decision about how to apply that protection could depend on a number of properties, including the environment.

Finally, complex security models introduce new challenges related to providing guarantees that a required level of protection can and will be applied. How can one site trust that another site will provide a sufficient level of protection to allow the remote site to host a dataset with unusual protection requirements, e.g., proprietary data? To provide these guarantees, we need to explore security metrics (Jaquith 2007) that have a particular relevance to the open science community.

- Provenance Tracking

  Provenance tracking tools are an open area of research (Clifford et al. 2007; Frew et al. 2007) that is particularly important for the open science community. Because the interpretation of and trust in open science data depends entirely on how it was produced, the VO needs automated tools for capturing provenance information, validating it, and ensuring that the relevant provenance information is available whenever and wherever the data is accessed in the future. Further, data will move through the system, passing from one member of the VO to another. Thus although all VO members will agree on the importance of data integrity, no single approach can be adopted to ensure the integrity of the data and its associated metadata, such as provenance information.

  And, while significant research efforts (Braun et al. 2006; Buneman et al. 2000, 2006) have been focused on the collection, semantic analysis, and dissemination, very little has been done in securing provenance data, a vital step in achieving trust and ultimately usability of provenance as a concept. Yet, unless provenance information is secured–and under the incidence of appropriate access control policies for confidentiality and privacy–it simply cannot be trusted.

Potential Open Science Impact

- greater trust in the data and conclusions produced by open science
- more effective scientific collaborations, through improved ability to share data in a controlled manner
- improved ability to reuse data in subsequent projects, through improved provenance information.

Potential Impact on Cyber Security for Open Science

- quantifiable guarantees of confidentiality and integrity for distributed open science data produced by large international collaborations
- user-friendly tools for managing information-sharing in large collaborations of scientists, with potentially conflicting security policies at different sites

- tools for capturing, tracking, and validating provenance information for open science data.

Time Frame

- **5-year goal:** user-friendly tools for managing information sharing in large collaborations; provenance tracking tools
- **10-year goal:** Quantifiable guarantees of confidentiality and integrity for long-lived distributed open science data

References

Braun, U., S. L. Garfinkel, D. A. Holland, K.-K. Muniswamy-Reddy, and M. I. Seltzer. 2006. "Issues in Automatic Provenance Collection," in *IPAW'06: International Provenance and Annotation Workshop,* 171-183.

Buneman, P., S. Khanna, and W. C. Tan. 2000. "Data Provenance: Some Basic Issues," in *FST TCS 2000: Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science,* pp. 87-93. London, United Kingdom, Springer-Verlag.

Buneman, P., A. Chapman, and J. Cheney. 2006. "Provenance Management in Curated Databases," in *SIGMOD'06: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data,* 539-550. ACM Press, New York.

Clifford, B., I. Foster, J.-S. Voeckler, M. Wilde, and Y. Zhao. 2007. "Tracking Provenance in a Virtual Data Grid." *Concurrency and Computation: Practice and Experience*, Doi: 10.1002/cpe.1256.

Foster, I., C. Kesselman, G. Tsudik, and S. Tuecke. 1998. "A Security Architecture for Computational Grids," in *Proceedings of the ACM Conference on Computers and Security, 83-89*.

Foster, I., C. Kesselman, and S. Tuecke. 2001. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations." *The International Journal of High Performance Computing Applications*, **15**, 3, 200-222.

Frew, J., D. Metzger, and P. Slaughter. 2007. "Automatic Capture and Reconstruction of Computational Provenance," in *Concurrency and Computation: Practice and Experience*, Doi:10.1002/cpe.1247.

Gudes, E., M. S. Olivier, and R. P. van de Riet. 1999. "Modelling, Specifying and Implementing Workflow Security in Cyberspace." *Journal of Computer Security*, **7,** 4, 287–315.

Jaquith, A. 2007. *Security Metrics: Replacing Fear, Uncertainty, and Doubt*, Addison-Wesley, Upper Saddle River, New Jersey.

Welch, V., F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, and S. Tuecke. 2003. "Security for Grid Services," *In: Proceedings of the Twelfth IEEE International Symposium on High Performance Distributed Computing, 48-57.*

Authors: Marianne Winslett, University of Illinois ; Ron Oldfield, Sandia National Laboratories; Patrick McDaniel, Penn State

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-5: TRUSTED VIRTUALIZATION

ABSTRACT

This Priority Research Direction (PRD) recommends a cyber security research agenda that focuses on the security aspects of virtual machine (VM) technologies likely to appear pervasively in open science infrastructure designs and deployments over the next 3 to 10 years. A number of research areas are highlighted either because no satisfactory solutions exist for identified issues or because potential benefits remain unrealized. The results from our research agenda are expected to substantially improve the secure deployment of this highly virtualized information technology (IT) infrastructure through enhancements in compromise isolation, detection, and recovery, and through improved assurances about the trust fabric that ties the virtual resources to the physical hardware.

EXECUTIVE SUMMARY

A much more manageable resource infrastructure as currently articulated through fashionable terms like autonomic-, self-healing-, utility-based-, on-demand-computing, or organic IT, will materialize in 3 to 10 years. All these paradigms rely on some form of virtualization of the resources to provide transparent, dynamic, and real-time properties to features like resource migration, pooling, replacement, repairing, sharing, and load balancing. These features provide enhanced support for the U.S. Department of Energy's (DOE's) deployment of physical resources on the open science infrastructure as well as on the power grid, and will therefore provide a more holistic view of general resource management within DOE. Note that deployment of VMs on the DOE's resources of the Center for Enabling Distributed Petascale Science project is already in proof-of-concept stage and planned for production (Keahey et al. 2007).

However, a number of security issues arise through the additional abstraction of resource virtualization, such as losing the direct connection to the physical hardware associated with the resource, i.e., losing the assurance that the resource actually resides on known, trusted physical hardware (Garfinkel and Rosenblum 2005; Ormandy 2007).

Virtualization also provides us with new opportunities to add enhanced security features to our resource deployment. Real-time replacement of compromised resources, much improved isolation properties that substantially limit the consequences of compromise, real-time and transparent monitoring, and policy enforcement of the use of physical resources such as CPUs, disks, and the network are facilitated by the use of virtualization (Sailer et al. 2005; Kuhlmann et al. 2006). This PRD recommends a 3- to10-year research agenda that will focus on ensuring trusted use of virtual resources while unlocking the advanced potential security features that the virtualization technologies can deliver to a safer and more robust DOE open science infrastructure.

SUMMARY OF RESEARCH DIRECTION

For the "Trusted Virtualization" PRD, the following research areas have been identified:

- **Assurance of VMs hosting environment**: The virtualization of resources introduces an additional abstraction that complicates the policy enforcement for a VM user who requires assurances about the location, type, or kind of hardware that hosts the hypervisor. The use of secure hardware components, such as an integrated TPM, could help to attest the trust chain from the application service running on a VM running on a hypervisor running on a specific machine that has an embedded TPM (Marchesini et al. 2003; McCune et al. 2006; St. Clair et al. 2007). We believe this assurance will become critical in highly virtualized environment where resources from many different sites are discovered, brokered, and matched, and the user's policy requires Service Level Agreements (SLAs) that stipulate certain acceptable HW properties of the resource. Note that this particular research area overlaps with our "Trusted Hardware and Crypto Acceleration" PRD.
- **Correctness of Hypervisor Security Execution**: The overall protection of the VMs from the outside world as well as from the other hosted VMs relies on the integrity of the hosting system, i.e., the integrity of the hypervisor software and correctness of the policy enforced by its reference monitor. In order to limit the number of bugs in the hypervisor code, the code base must remain as small as possible and must be formally proven secure where possible. The correct and unambiguous enforcement of the policy by the reference monitor as it is derived from the SLAs and higher-level site policies is another concern. All areas will require continued focus from the research community.
- **VM-instance identity and lifecycle**: The execution of VMs differs from conventional computing environments in that applications can be stopped, frozen, serialized, replicated, migrated, and restarted/resumed on other hosting environments transparently. These features allow the higher-level ability to migrate, load-balance, and mirror resources based on demand and on deployment considerations. Unsuspecting applications, however, may yield unintended results if application contexts are replayed. In particular, the data-sets and memory-snapshots associated with such VM-images include long- and short-lived secrets that are used for authentication of the resource and the integrity of the communications which can be compromised if execution expectations are invalidated. How to deal with such issues correctly and properly is an open question and requires investigation.
- **Trusted security service VMs**: Because of the excellent isolation properties of the hypervisor, the access to a VM can be restricted to only a single other VM managed by the same hypervisor and further restricted to a single communication mechanism and protocol. Such a setup, for example, could off-load the secrets and crypto processing from a network attached VM to a non-network-accessible VM. This is the equivalent of using a VM as a smartcard or secure hardware device. Such applications have the potential to limit the consequences of compromise but their feasibility requires further research.
- **Secure proxy service VMs**: The inter-process communication between VMs is subject to the reference monitor's policy enforcement and is safe from snooping by other VMs or the

outside world. This property can be used to transparently provide security to insecure versions of protocols, like dns, snmp, smtp, by hosting a proxy service in a dedicated VM that uses the insecure protocols for the inter-VM communication while communicating securely with the outside world through the secure versions of the protocols. The advantages are that only the outward-facing proxy-services have to be pre-configured with the correct trust-root information. The development and deployment of such set-ups require further investigations.

- **Compromise detection**: The ability of the hypervisor to observe the detailed use of the physical resources by a VM in real time, can be used to detect abnormal actions, like access to unknown outside IP addresses, modification of critical disk files, calls to new libraries, and unexpected CPU-usage spikes. The issue becomes how to define “normal or expected behavior,” and we can see three research areas that have potential: 1) let the VM user identify expected behavior as part of the SLA with the hosting party, like the use of ports, external services, local library calls, etc., 2) the hypervisor can observe a known non-compromised VM over time and deduce “normal” patterns of resource usage, and 3) scan the source/binary code of the VM for resource access calls, like open(). The latter research could be combined with similar areas that the “Secure Software” PRD is proposing.
- **Isolate compromise**: VMs hosted by a hypervisor have the nice property that they are isolated from each other such that a compromised VM will not be able to compromise another VM or the hypervisor directly, such as via a rootkit equivalent. A compromised VM could still attack other VMs through any of the communication mechanisms that the hypervisor allows it to use. By using well-defined access control policies over VM resources and integrity-protecting interfaces for communication, we could further isolate the VM and limit its ability to compromise others.
- **Investigation of compromises**: As intruders and compromises become more sophisticated, more advanced forensic analysis options are needed. Hypervisors can freeze a complete VM-image that includes OS, application, memory and disk-data, which constitutes a substantial amount of forensic information. In addition, when a compromise is expected, the hypervisor with its reference monitor could change the running application’s environment into a honey-pot configuration for real-time tracking of the intruder’s actions. Lastly, the hypervisor could record a VMs detailed actions such that one could literally rewind and playback through the VMs life, which could facilitate investigations.
- **Compromise recovery**: After detecting and studying a compromise, the affected environment has to be cleaned-up and restarted in a known safe state. The hypervisor’s ability to freeze a VMs state can be used to “snapshot” VMs during their lifecycle. These snapshots provide safe recoverable images, which could potentially save substantially on the time and nuisance associated with recovery from security violations. We have to investigate how to use and optimize these VM features and learn what the possible pitfalls are.
- **Overlay-VM-VPNs**: Distributed applications used by collaborating groups may choose to host all the application and infrastructure services on VMs distributed over different hosting

hypervisors based on the required SLAs associated with the different components. To enhance the secrecy and integrity of such distributed computations, the collaboration may choose to deploy an overlay VPN across all the different components hosted on the VMs. Setting up such a VPN could be achieved independently from the applications/services themselves through the VM configurations by the hypervisors and should be driven by the SLA negotiation between the VM users and the hypervisor. Further, secure hardware components may be used to justify the integrity of the VPN infrastructure at each system. Future work is needed to examine the process of establishing and maintaining trustworthy VPNs.

All described areas and scenarios require substantially more research and development of tools.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The research areas enumerated in the previous section have a number of scientific and computational challenges associated with them:

- **Assess what is happening inside VM**: Compromise detection and forensic analysis require the ability for a hypervisor to comprehend the runtime behavior of all software executing within a VM. This software is complex and includes the scientific computation as well as the operating system layer between the computational software and the VM. Previous research has demonstrated an ability to extract profiles automatically characterizing execution of binary computational applications (Giffin et al. 2002), but the incorporation of operating system code tremendously increases the difficulty of such analysis. New source code and binary code analyses able to cope with complete system software will be needed to satisfactorily address this PRD.

- **Designing, expressing, and enforcing security goals at VM abstraction**: High-level policies embodying security goals able to be written and understood by humans are at a different level of abstraction than the policy enforcement software contained in a hypervisor. Automated policy compilers or translation mechanisms must be developed to transform the high-level statements into correct and enforceable statements at the virtual hardware interface. Understanding this transformation is non-trivial and may require techniques providing knowledge of how information flows through the software inside a VM.

- **Quantifying level of assurance, satisfaction of higher-level collaboration policies**: By virtue of the coarser granularity of VM resources, it may now be possible to quantify the security afforded by higher-level collaboration policies. Policy analysis for mandatory access control systems, such as SELinux, demonstrates where secrecy and integrity problems may exist in systems, but using VM policies may both reduce the number of such problems, thus enabling comprehensive management, and provide more options for resolving such problems. Tools that enable VM policy analysis and support design for security in VM systems are needed.

- **Dynamically overlaying VPN with VM components**: Using VPNs enables secure network communication as well as control of VM communication between machines. The result is that all VM communications can be mediated by the system’s access control policies and all communication between systems can be protected. Management tools for IPsec will be necessary to support the dynamic configuration of VM computations and to enable effective access control and network security.

- **Maintaining security of data when resuming/rolling back/migrating computation and data**: VM migration and compromise response that restarts a VM from a prior safe snapshot offer compelling usefulness to open science infrastructure. However, these mechanisms may compromise data security in unexpected and poorly understood ways. For example, rewinding application execution may lead to reuse of cryptographic key material, and this reuse may enable an attacker to break the crypto system. Trusted computing relies upon a hardware root of trust. Similarly, migration changes the underlying hardware without notification to the software, and the effect of this change upon trusted computing is unpredictable. We must first study such issues and second, develop secure rollback and migration algorithms.
- **Leveraging secure hardware/TPM for VM security**: Much of the security impact of VM technology depends on the ability to establish the integrity of the trusted computing base for such systems. Secure hardware/TPM provides mechanisms for establishing, measuring, and maintaining the integrity of software. However, approaches to leverage such mechanisms are still immature. Future work will explore new approaches, their deployment, and support for the above security guarantees.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The impact of our recommended research direction on computational science is:

- **Reduction of computation to appliance**: The appliance model has the advantage of easier reuse, support, replaceability, etc., which would facilitate the discovery of matching compute services and the creation of the computational workflows.
- **More flexible consumption of utility computational resources**: Secure and trusted virtualization-based open science infrastructures allow users to choose resources based on a containment policy rather than on compatibility of the execution environment. Restrictions on node usability occur in current infrastructure designs when the installation includes operating systems on individual nodes. Virtualization allows any operating system to run on any hardware that presents the same virtual device interface. When OS heterogeneity is desired, secure VM-based architectures allow more jobs to execute on more nodes in the system.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

The impact of our recommended research direction on cyber security for open science is:

- **Containment of compromise**: Virtual machines provide stiff boundaries between the VMs and the hypervisor and among all VMs executing on a single machine. As a result, any compromise of applications or the operating systems will have limited consequences because the hypervisor enforces containment properties. An attacker may escalate their access from a compromised application up to a compromised OS, but they cannot easily escalate to full machine access because hypervisors have a severely restricted attack surface. This provides a layer of security not present in current open science infrastructure designs.
- **Quantitative policy analysis**: The coarse granularity of VM resources enables comprehensive, quantitative assessment of access policy, permitting verification as to whether the policy satisfies security goals (secrecy and integrity) and identification of where such goals are not met to guide resolution.
- **Least-privilege operations**: The ability to enforce fine-grained policy about the use of physical resources with the ability to describe in detail what physical resources will be used by applications and VMs, allow for an operational least-privilege mode that enhances overall security.
- **Simplified policy definition/easier management**: The focus on deriving low-level hypervisor-enforceable security policy from the high-level SLAs will bring the abstraction level up, resulting in easier and more precise expression of policy.
- **Reduced TCB**: When the physical hardware can be tied to the virtual resources that are hosted, the resource user's policy is able to be more specific about the resources that can be trusted, which results in a reduced trusted computing base.

TIME FRAME

- This cyber security research agenda−focusing on security aspects of VM technologies−is likely to appear pervasively in open science infrastructure designs and deployments over the next 3 to 10 years.

REFERENCES

Garfinkel, T., and M. Rosenblum. 2005. "When Virtual is Harder than Real: Security Challenges in Virtual Machine Based Computing Environments," in 10th Workshop on *Hot Topics in Operating Systems (HotOS X)* held in Santa Fe, New Mexico, available at http://www.usenix.org/events/hotos05/final_papers/full_papers/garfinkel/garfinkel.pdf

Giffin, J. T., S. Jha, and B. P. Miller. 2002. "Detecting Manipulated Remote Call Streams," in 11th *USENIX Security Symposium,* held in San Francisco, California, available at http://www.usenix.org/publications/library/proceedings/sec02/full_papers/giffin/giffin.pdf.

Keahey, K., T. Freeman, J. Lauret, and D. Olson. 2007. “Virtual Workspaces for Scientific Applications,” *SciDAC 2007 Conference,* Boston, Massachusetts, available at http://workspace.globus.org/papers/SciDAC_STAR_POC.pdf.

Kuhlmann, D., R. Landfermann, H. Ramasamy, M. Schunter, G. Ramunno, and D. Vemizzi. 2006. “An Open Trusted Computing Architecture—Secure Virtual Machines Enabling User-Defined Policy Enforcement,” *Research Report RZ 3655 (#99675), IBM Research*, available at http://domino.research.ibm.com/library/cyberdig.nsf/papers/7024C307EA0DFAEE852571D0003B10F3/$File/rz3655.pdf.

Marchesini, J., S. W, Smith, R. MacDonald, and O. Wild. 2003. “Experimenting with TCPA/TCG Hardware, Or: How I Learned to Stop Worrying and Love the Bear,” Dartmouth Computer Science Technical Report TR2003-476, Hanover, New Hampshire.

McCune, J., T. Jaeger, S. Berger, R. Caceres, and R. Sailer. 2006. “Shamon: A System for Distributed Mandatory Access Control,” in *Proceedings of the 22nd Annual Computer Security Applications Conference,* IEEE Computer Society, Los Alamitos, California, ed. D. Thomsen, 23-32, available at http://ieeexplore.ieee.org/iel5/4041138/4041139/04041151.pdf?tp=&arnumber=4041151&isnumber=4041139

Ormandy, T. 2007. “An Empirical Study into the Security Exposure to Hosts of Hostile Virtualized Environments,” *CanSecWest,* Vancouver, British Columbia, available at http://taviso.decsystem.org/virtsec.pdf

Sailer, R., T. Jaeger, E. Valdez, R. Caceres, R. Perez, S. Berger, J.L. Griffin, and L. van Doorn. 2005. “Building a MAC-based Security Architecture for the Xen Open-Source Hypervisor,” in *Proceedings of the 21st Annual Computer Security Applications Conference,* Tucson, Arizona, IEEE Computer Society, Los Alamitos, California, 276-285, Doi: 10.1109/CSAC.2005.13.

St. Clair, L., J. Schiffman, T. Jaeger, and P. McDaniel. 2007. “Establishing and Sustaining System Integrity via Root of Trust Installation,” to appear in *2007 Annual Computer Security Applications Conference,* to be held in December.

Authors: J. Giffin, Georgia Tech; T. Jaeger, Penn State; J. Pato, HP Labs; F. Siebenlist, Argonne National Laboratory

SECURING HARDWARE, SOFTWARE, AND DATA
PRD-6: SECURE INFORMATION MANAGEMENT

ABSTRACT

This Priority Research Direction (PRD) recommends a cyber security research agenda that focuses on the security aspects of information management. The need to protect open science information is increasing because of social and economic impact. Information may include the inferred/discovered data, results of experiments, computations, equations, metadata, code, binaries, security policy statements, and relationships between the data. There is a critical need to maintain confidentiality, privacy, availability, and integrity of open science information distributed across millions of nodes. Due to these requirements, the open science research community needs a secure information management system.

EXECUTIVE SUMMARY

Recent developments in information systems technologies have resulted in computerizing many applications in various business areas. Data has become a critical resource in many organizations, and therefore, efficient access to data, sharing the data, extracting information from the data, and making use of the information has become an urgent need. As a result, there have been many efforts on not only integrating the various data sources scattered across several sites, but also on extracting information from these databases in the form of patterns and trends. These data sources may be databases managed by database management systems, or they could be data warehoused in a repository from multiple data sources.

The advent of the World Wide Web (WWW) in the mid 1990s has resulted in even greater demand for managing data, information and knowledge effectively. Today a second generation of web-based communities and hosted services, such as social networking sites and wikis, are emerging that facilitate collaboration and sharing between users. The term Web 2.0 has been coined to embrace all the new collaborative applications and also to indicate a new "social" approach to generating and distributing web content, characterized by open communication, share and re-use. There is now so much data on the web that managing it with conventional tools is becoming almost impossible. New tools and techniques are needed to effectively manage this data and conduct experiments and collaboration. Therefore, to provide interoperability as well as warehousing between the multiple data sources and systems, and to extract information from the databases and warehouses on the web, as well as to share the information and conduct collaboration, we need to invent and develop efficient tools, many of which are currently not available.

As the demand for data and information management increases, there is also a critical need for maintaining the security of the databases, applications, and information systems. Data and information have to be protected from unauthorized access as well as from malicious corruption. With the advent of the web and openness of the environment, it is even more important to protect the data and information as far greater numbers of individuals now have access to this data and information. Therefore, we need to develop and enhance effective mechanisms for securing data and applications.

The objective of this PRD is to determine the directions for secure information management for the open science community.

SUMMARY OF RESEARCH DIRECTION

A number of research areas have been identified:

Policy Management: What are the appropriate languages to specify policies such as confidentiality, privacy, and trust policies? How can policies be discovered and designed? How can ontologies be used for policy management? What sorts of tools are needed for policy integration, policy interoperability, policy consistency chancing and policy reasoning? Which tools and mechanisms are needed for distributed, decentralized, and collaborative policy enforcement?

Discretionary Security: While discretionary security for relational database is a mature technology. There is a lot to do on discretionary security for XML (eXtensible Markup Language) and RDF (Resource Description Framework) information bases. For example, how can we specify policies in XML? How can XML be secured? What sorts of temporal authorization models are appropriate for the emerging database systems? These are all interesting challenges.

Mandatory Security: We have focused on multilevel security for various types of databases. While research in this area is not as active as it used to be, we have learnt a lot in conducting research in multilevel information management. Furthermore, such systems are still needed for certain DOE applications. The challenges here include developing new kinds of models and architectures for multilevel information management as well as building high assurance systems for open science environment.

Secure Grid Computing and Infrastructures: How can security be incorporated into service-oriented architectures and web services so that secure infrastructures can be developed to host the information management applications? How can the grid service-meta-data be securely published, discovered, and shared? How can different services be composed securely? How can we incorporate security into grid information management? How can security functions be organized as services (Security as a Service – SaaS) so that they can be shared by multiple applications and parties?

Secure Information Management Models and Functions: What are the appropriate models for secure information management? How does security impact functions such as query processing transactions management and storage management? What are the challenges in secure collaboration?

Accountability: How do we develop fine-grained and efficient accountability mechanisms, driven by policies, for large-scale distributed environments? Which environments and tools are needed to support accountability queries and analysis?

Digital Identity Management: How do we manage identity information concerning users and other entities across large-scale, distributed systems? How do we ensure that identity information is correct and at the same time maintained confidential? How to specify and enforce differentiated authentication policies, depending on context and situations?

Inference Problem: While this is a very difficult problem, it continues to fascinate researchers. We need to build constraint processors that are more efficient and manage prior knowledge. The complexity of the problem also needs to be examined. There is a lot of interesting theoretical work to do in this area. Furthermore, in an open science environment, there is a possibility for researchers to assemble collections of data and infer information that is highly classified or private. Tools are needed to protect sensitive information in such situations.

Secure Distributed and Heterogeneous Information Repositories: While some progress has been made, an extensive investigation of security for distributed, heterogeneous, and federated databases and information repositories is needed. What sorts of access controls and models are appropriate for such systems? How can we share data and still have security and autonomy? How can security policies be integrated across organizations? How can distributed transactions be executed securely?

Secure Object Information Management and Applications: There has been work on both discretionary and mandatory security for object databases. How can we apply the principles for object-relational systems since such systems are dominating the marketplace? Are the security mechanisms for distributed object management systems sufficient? How can we provide fine-grained access control? How can UML be used to design secure applications?

Secure Data Warehousing, Mining, Security and Privacy: There challenges are many. How can we build a secure warehouse from the data sources? How can we develop an integrated security policy? What is the security impact on the functions of a warehouse? What are the data mining techniques appropriate for national security and cyber security? How can we solve the privacy problem? How can we build effective privacy controllers? What is the complexity of the privacy problem?

Secure Web Data, Information and Knowledge Management: There is a lot of work to be done on secure web data and information management. For example, how can we build secure web database systems? What are the security issues for digital libraries? How do we secure the semantic web? How can we maintain trust on the semantic web? How can we secure emerging applications such as knowledge management, multimedia, collaboration, e-commerce and peer-to-peer data management? How can we use ontologies for policy specification and management?

Data Quality and Provenance: How can we maintain data quality? How can we determine data provenance so as to prevent/detect misused? What the appropriate models for data quality representation? How can we reason about the quality of the data?

Emerging Security Technologies: Little work has been reported on secure dependable data management. For example, how can we build systems with flexible policies that can handle security, real-time processing, fault tolerance and integrity? How can we secure sensor database systems? What are the security issues for wireless information management? Finally how can we further the developments in digital identity management, digital forensics, and biometrics?

Societal Impact: How can we ensure that societal concerns such as data privacy and data confidentiality are handled appropriately? How can we ensure that the data is not mishandled or misused? How can we create an environment that will foster collaboration between natural scientists, computer scientists, and social scientists?

Risk, Trust and Economics: What are the risks involved to security? What are the costs involved in incorporating security and trust? What are the tradeoffs between risk and cost? What are the appropriate models for risk and cost analysis?

Scalability: Last but not least, we need to ensure the scalability of the techniques developed. The issues are: What sorts of tools do we need to ensure scalability? What sorts of experiments do we need to carry out to determine that a technique will scale to millions of nodes?

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The research areas enumerated in the previous section have a number of scientific and computational challenges associated with them:

- The need to protect critical open science information distributed across millions of nodes is increasing due to social and economic impact.
- Information may include certain inferred/discovered data, results of experiments, computations, equations, metadata, code, binaries, unstructured data (images, text) and relationships between data.
- Secure distributed information management system is required to ensure confidentiality, privacy and integrity of security and other information.

POTENTIAL OPEN SCIENCE IMPACT

Open science researchers will be able to securely share information of assured authenticity and integrity in addition to maintaining their privacy.

TIME FRAME

- While security solutions could be provided within a 5 to 7 year time frame, it will take up to 8 to 10 years to ensure scalability and security.

Authors: B. Thuraisingham, University of Texas at Dallas; E. Bertino, Purdue University

BREAKOUT SESSION
MONITORING AND DETECTION (MD)

Breakout Leads: *Troy Thompson, Pacific Northwest National Laboratory;*
John McHugh, Dalhousie University

MONITORING AND DETECTION
PRD-1: VERIFICATION OF INTENDED USE

ABSTRACT

Both open science security and control systems security operate in environments in which individuals, organizations, and governments can compromise systems, software, and data. The key objective of this Priority Research Direction (PRD) is to develop a rigorous theoretical and deep operational understanding of how a complex and decentralized system is used and what kinds of users and organizations it has.

EXECUTIVE SUMMARY

Verification of intended use is a crucial aspect of any cyber security infrastructure. First, accurate verification of intended use provides us with a deeper understanding of secure and vulnerable environments. This leads to better design of infrastructure. Second, real-time accurate verification of intended use facilitates early detection of threats (whether they be insider or outsider; deliberate or accidental). This naturally leads to early mitigation.

Verification of intended use involves assimilation of computational and statistical models that represent 1) usage patterns of the network, software, and data; 2) behavior of applications, users, and organizations; and 3) interdependencies between 1 and 2. The novelty of this PRD to open science and control systems is with respect to the scale and diversity of the verification problem. Furthermore, the state-of-the-art lacks any systematic formal study of algorithm robustness or benchmarking with respect to noisy, incomplete, and uncertain observations of the environment.

To achieve real-time accurate verification of intended use, we need multidisciplinary teams with expertise in computer networks, sensor networks, dynamic social networks, machine learning, data mining, statistics, and cognitive science, to name a few.

SUMMARY OF RESEARCH DIRECTION

Research direction areas include: 1) develop mathematical theory for modeling usage and behavior in a decentralized setting, e.g., a framework for decentralized anomaly detection; 2) develop semantics for representing the relationships between a system’s use and its users’ behavior; 3) develop algorithm benchmarks for operational understanding of large-scale, diverse, and complex open science environments; and 4) develop scalable algorithms for signature discovery, such as biometric user tracking and watermarking binaries.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Underlying scientific and computational challenges include: 1) linking human actions with cyber actions and effects, 2) modeling large variations of dynamic intended use behaviors, and 3) balancing security with flexible and scalable capabilities in the face of large number of users in various organization and geographical locations (Interagency Working Group 2006).

These challenges are not unique to DOE's open science efforts. The most recent and highly relevant BAA on cyber security is from HSARPA (2007) – announced in May 2007. Specifically, the HSAPRA BAA has a technical topic area on *Insider Threat Detection and Mitigation*. While their focus is on securing environments that interface government communities e.g., classified (Jones 2000), unclassified, local, state, and foreign and private industry, they too face non-trivial issues relating to scalability and diversity. This PRD's challenges are more general since verification of intended use encompasses insider threat detection and many other aspects relating to the security of decentralized systems e.g., accidental misuse and outsider threat.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The potential impact of this PRD on computational science is two-fold. The first impact is the development of an analytical framework that provides for accurate and efficient computational procedures for automatic generation and evaluation of models that represent both usage of complex decentralized systems and human/agent behavior on those systems. The second impact is the development of computable semantics i.e., automatically generated ontologies, that describe usage, behavior, and effect of users and applications in complex decentralized systems.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

With sound theoretical and operational understanding, we can design better environments for open science, patch vulnerabilities, and mitigate threats sooner. All of these will significantly enhance the productivity of open science.

TIME FRAME

The research is envisioned to take 7 to 10 years to mature and have a tangible impact.

REFERENCES

HSARPA. 2007. "Cyber Security Research and Development." U.S. Department of Homeland Security Broad Agency Announcement BAA07-09, available at http://www.hsarpabaa.com/Solicitations/BAA07-09_CyberSecurityRD_Posted_05162007.pdf

Interagency Working Group. 2006. *Interagency Working Group on Cyber Security and Information Assurance*, “Federal Plan for Cyber Security and Information Assurance Research and Development,” sponsored by the National Science and Technology Council. Available at: http://www.nitrd.gov/pubs/csia/ csia_federal_plan.pdf.

Jones, A. K. 2000. “Summary of Discussions at a Planning Meeting on Cyber Security and the Insider Threat to Classified Information,” sponsored by the National Research Council (NRC). Available at: http://www7.nationalacademies.org/CSTB/wp_insiderthreat.pdf.

Authors: T. Eliassi-Rad, Lawrence Livermore National Laboratory; ***eliassirad1@llnl.gov****; D. Xuan, Ohio State University; C. Corbett, Sandia National Laboratories*

MONITORING AND DETECTION
PRD-2: ENABLING DATA SHARING AND COOPERATIVE ANALYTICS

ABSTRACT

This Priority Research Direction (PRD) includes research to enable the sharing of data among entities across multiple administrative domains and to support a cooperative analysis of data to detect new and emerging threats that span multiple organizations. The key challenges in this PRD are to manage privacy and confidentiality of sensitive data, to ensure integrity of shared data and how it affects monitoring subsystems, and to maintain sufficient information in the shared data to enable new forms of analytics.

EXECUTIVE SUMMARY

Current cyber security threats are growing more distributed in nature and require a global monitoring infrastructure in order to detect them, analyze them, and effect a response, and the trend points to even more global reach of attacks in the future. Global monitoring raises concerns of privacy as data is shared across entities and international boundaries; data can both violate personal privacy and contain sensitive information that can itself be used to identify weak points to attack. Data integrity is also a major concern, as false data injected into a monitoring infrastructure could cause it to fail to respond to attacks or even shut down essential services.

Current research into privacy-preserving data sharing does not sufficiently address issues of inferences that are possible or integrity of the computed results. Furthermore, the bulk of the research focuses on anonymizing logs for research, i.e. scrubbing them of identifying information. This is a fundamentally different problem than supporting cooperative analytics, both because the information patterns that need to be preserved are likely to be different than the packet header and other such statistics preserved by the current methods, and because cooperative analytics can employ interactive privacy-preserving algorithms in order to support online queries and analysis while minimizing information leaks.

The goal of this PRD is to support research to develop techniques that can dynamically adapt to an evolving set of constraints, both on the side of privacy and on the information that needs to persist for analytics, while being robust to potential misinformation from corrupt or compromised entities. This research will support cooperative analytics in the open science environment and thus better protect the infrastructure from emerging threats. It will also include basic research that will help data sharing in all fields of open science. The expected time to deployment of such techniques is 7 to 8 years.

SUMMARY OF RESEARCH DIRECTION

Attacks on cyber security have been growing in scale and complexity. While it used to be possible to detect and respond to attacks on a system-by-system basis, the attacks of today and the future require a more global perspective. Worms, botnets, stepping stones, and privilege escalation attacks exploit vulnerabilities across multiple computers, networks, and organizations to escape detection or cause greater damage. Thus, to deal with the evolving threats, a monitoring infrastructure must combine and correlate information from multiple sources.

However, the sharing and aggregation of monitor data across organizational boundaries presents privacy challenges. Monitor data can be used to learn sensitive information and compromise both personal privacy and organizational secrets, so wider distribution of such data must be safeguarded by techniques to minimize the amount of sensitive information that is revealed. Monitor data also can be used for attacks, as it can help identify weak spots in defense systems, or even the monitoring infrastructure itself. At the same time, external monitor data is used to support decision processes, perhaps automated, that implement responses to attacks. As such, the integrity of such data is important, to prevent attacks that exploit the monitoring infrastructure to shut down essential services. A key research challenge is, thus, to satisfy both privacy and integrity requirements for a shared monitoring infrastructure.

This problem is particularly important for the DOE open science environment, as it is composed of multiple semi-autonomous entities that span the globe. Crossing international boundaries presents a special challenge; for example, both CERT and the PREDICT projects have their data producers and consumers restricted to within the United States. However, the open science environment also presents an opportunity because there is both a lower expectation of privacy and a shared value of open collaboration. In many organizations, laws and customs suggest that nearly any information about internal functions must be kept private, be it sensitive product information, data about personal habits, or even the structure of the network. In an open science environment, however, it is possible to draw a line between information that is (or should be) available to the public and information that is genuinely privacy sensitive, enabling a different space of solutions than is feasible for other applications. Therefore, mechanisms to support the sharing of monitor data in the DOE open science environment should be a Priority Research Direction.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Prior work on support for data sharing has focused on anonymizing, or scrubbing, of shared log data (Fan et al. 2004; Li et al. 2004; Lincoln et al. 2004; Pang and Paxson 2003). However, such techniques can remove at the same time too little and too much information. Too little, because, as recent results show (Brekne et al. 2005; Coull et al. 2007a, 2007b), inference attacks that combine anonymized data with prior knowledge or probed observations can defeat many anonymization schemes. And too much, because the information that is left is tailored towards today's manual data analysis tools such as NetFlows, and much information that would be useful to other types of analysis, such as would be useful to support a shared monitoring and response infrastructure, is removed.

A monitoring infrastructure for open science will need to employ novel detection algorithms to address emerging threats. Therefore, the informational requirements will be very different than what is provided by simply anonymizing logs. Most importantly, per-record transformations of today will need to be replaced by transformations that preserve multi-dimensional patterns that are needed for global detection and also remove other cross-record patterns that can lead to inferences that violate privacy and confidentiality. A data transformation infrastructure must adapt to, on one side, a set of privacy constraints that will evolve as our understanding of both privacy needs and potential inference attacks grows, and as privacy sensibilities themselves evolve, and on the other side, constraints on what data and relationships are to be preserved that will be specified by novel monitoring applications that have not yet been designed.

The context of a monitoring infrastructure also allows a different class of approaches than data sharing for research; in particular, interactive analysis of remote data with privacy-preserving techniques is possible. Recent years have seen the development of many techniques that allow queries over remote data or correlations of several datasets to be performed while minimizing information disclosure and preserving privacy of the participants (Brickell and Shmatikov 2005; Frikken and Golle 2006; Kissner and Song 2005; Saint-Jean et al. 2007). These techniques rely on either a trusted or semi-trusted third party, or use cryptographic tools to emulate such a party.

Outstanding research issues for these techniques are to make them dynamically adapt to the aforementioned constraints and to address issues of integrity. Current techniques allow a party to specify its own data and therefore potentially influence other parties into undesirable decisions. The ability to deal with malicious collaboration parties is important both as a way to maintain greater autonomy and to deal with potential compromise of another entity or the communication path. Without such safeguards, a compromised entity may use the monitoring data sharing mechanism itself to attack other systems and cause them to shut down or weaken their defense posture. A combination of probes and correlation with other observations will likely be necessary to ensure the integrity of shared data.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The ability to obtain datasets is essential to all scientific research and sharing data is a key ingredient in reproducibility, which is a hallmark of true science. Issues of confidentiality and integrity are frequent barriers to obtaining data for research or for reproducing results. Work addressing this PRD will involve basic research into data sharing and will explore new perspectives on how collaborative use of data can be carried out while preserving important constraints. Advances in this area will enable new experiments and accelerate the progress of computational science.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Cyber security, both in the context of the DOE open science environment and in the larger Internet, is increasingly dependent on maintaining a global perspective and the ability to track attacks as they traverse different computer systems and organizations. Advances in the support for data sharing will enable collaborative cyber security monitoring and response across entities in the DOE environment who can nevertheless preserve some measure of autonomy, thus mitigating exposure and preventing large-scale catastrophes that can result from failures in a more centralized or hierarchical approach.

TIME FRAME

- The basic research into adaptive data transformation and sharing is expected to take approximately 3 years.
- An additional 2 more years will be devoted to integrating the techniques with a decentralized, automated, and adaptive monitoring and response infrastructure for the DOE environment, with a special focus on analyzing robustness to incorrect data.
- Deployment can be expected in 7 to 8 years.

REFERENCES

Brekne, T, A. Årnes, and A. Øslebø. 2005. "Anonymization of IP Traffic Monitoring Data—Attacks on Two Prefix-Preserving Anonymization Schemes and Some Proposed Remedies," *Proceedings of 5th International Workshop on Privacy Enhancing Technologies,* ed. G. Danezis and D. Martin.

Brickell, J., and V. Shmatikov. 2005. "Privacy-Preserving Graph Algorithms in the Semi-Honest Model," in Advances in Cryptology: ASIACRYPT 2005, 11th International Conference on the Theory and Application of Cryptology and Information Security, ed. B. Roy, 236-252, Springer, Berlin. Held at Chennai, India.

Coull, S. E., M. P. Collins, C. V. Wright, F. Monrose, and M. K. Reiter. 2007a. "On Web Browsing Privacy in Anonymized NetFlows," *Proceedings of 16th USENIX Security Symposium,* ed. N. Provos.

Coull, S. E., M. P. Collins, C. V. Wright, F. Monrose, and M. K. Reiter. 2007b. "Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces," *Proceedings of the 14th Network and Distributed Systems Security Symposium*, eds. W. Arbaugh and C. Cowan.

Fan, J., J. Xu, M. Ammar, and S. Moon. 2004. "Prefix-preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme," *Computer Networks* ***46***, 2, 253-272.

Frikken, K., and P. Golle.   2006. "Private Social Network Analysis," in *Proceedings of ACM Workshop on Privacy in Electronic Society*, eds. R. Dingledine and T. Yu.

Kissner, L., and D. Song. 2005. "Privacy-Preserving Set Operations," in *Proceedings of Advances of Cryptology – CRYPTO*, ed. V. Shoup.

Li, Y., A. Slagell, K. Luo, and W. Yurcik. 2005. “CANINE: A Combined Conversion and Anonymization Tool for Processing NetFlows for Security,” *Proceedings of the 10th International Conference on Telecommunication Systems,* eds. B. Gavish and A. Bordetsky.

Lincoln, P., P. Porras, and V. Shmatikov. 2004. “Privacy-Preserving Sharing and Correlation of Security Alerts,” *Proceedings of 13th USENIX Security Symposium,* ed. M. Blaze.

Pang., R., and V. Paxson. 2003. “A High-Level Environment for Packet Trace Anonymization and Transformation,” *Proceedings of the ACM Special Interest Group in Communications (SIGCOMM) Conference,* eds. J. Crowcroft and D. Wetherall.

Saint-Jean, F., A. Johnson, D. Boneh, and J. Feigenbaum. 2007. “Private Web Search,” in *Proceedings of ACM Workshop on Privacy in Electronic Society*, ed. T. Yu.

BREAKOUT SESSION
FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES (FSA)

Breakout Lead: *Tom Harper—Idaho National Laboratory*

FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES

PRD-2: INTRUSION PREVENTION, DETECTION, AND RESPONSE

ABSTRACT

Intrusion prevention, detection, and response within the open science environments of the U.S. Department of Energy (DOE) have to deal with the challenges of two large-scale types: 1) large datasets produced at supercomputers and experimental facilities as well as fine controls needed for monitoring and steering of computations and experiments, which require networks with large capacity and dynamically stable connections, and 2) open science collaborations by large international research teams, which require highly dispersed, distributed, and heterogeneous network environments. Together, these two areas require cyber security approaches that are much more sophisticated than the current single-firewall approaches used in intellectual property (IP) networks. The first set of challenges arises due to the very high bandwidths (10-100Gbps) of dedicated or special-purpose network connections that connect the computing and experimental facilities. The challenge in the second area is due to the network connections that cross multiple open domains and span several countries with hundreds of users connected over different middleware and grid environments. The research areas identified here will deliver the capability to deploy and operate extremely large-scale cyber environments for open science with fast detection and precise and informed responses to internal and external cyber intrusions, thereby reducing the vulnerabilities to external and insider attacks.

EXECUTIVE SUMMARY

The next-generation networks needed for DOE large computational and experimental facilities, and large, international collaborations must securely operate at unprecedented scales involving supercomputers that operate at exaflops processing speeds, data and file transfers at bandwidths in excess of terabps at thousands of miles, data and storage systems with exabyte or higher capacities, and collaborations involving between 100 and 1000 users distributed worldwide. Cyber security at this scale is beyond the evolutionary path of industrial products and is tangential to the Internet-centric research efforts of other agencies. We outline the research component areas for an integrated intrusion prevention, detection and response framework by exploiting the structure and nature of specific DOE open science data, control, and execution paths. These efforts will enable security and trust in geographically dispersed, high-performance critical open science infrastructure by developing the underlying tools and methods. These technologies capture and process the network and host data at extreme speeds to support monitoring, packet filtering, anomaly detection, reduced false positives, intrusion forensics, information fusion for complex attacks, and integrated response involving fallback mechanisms and containment.

SUMMARY OF RESEARCH DIRECTION

DOE's open science environments offer unprecedented capabilities to advance scientific discovery. They offer supercomputers with unparalleled speeds to the open science community; currently the National Leadership Computing Facility operates at 100 teraflops, and these facilities are expected to reach petaflops to exaflops rates in near future. DOE also operates or participates in large scale experimental facilities, such as Spallation Neutron Source, which provide unique capabilities to scientists. Both the computing and experimental facilities generate petabyte to exabytes of data and also require remote monitoring, steering, and control of experiments and computations. In another direction, DOE supports open science collaborations of international research teams consisting of diverse domain experts who may be distributed over heterogeneous network environments spanning multiple countries. These teams require access to large storage sites that house exabytes of experimental and computed datasets, and collaboration tools for jointly steering computations on supercomputers and jointly visualizing datasets at remote powerful visualization facilities. To support these tasks, high-performance networks such as ESnet, Science Data Network, LHCnet as well as testbeds such as UltraScienceNet and CHEETAH are being developed by DOE and other agencies. These networks, together, provide unprecedented bandwidths as well as capabilities for the users and applications to co-schedule the network connections with their allocations on computing and experimental facilities. It is very important that cyber security measures be built into these networks and end systems to protect these valuable resources against cyber attacks, in particular intrusions of various types, including external and insider attacks.

Several intrusion detection, prevention, and response methods developed for Internet environments by industries and other federal agencies, in particular by NSF for eScience applications, will contribute to the cyber solutions for these DOE open science environments. However, there are several challenges that are unique to these environments, which are outside the projected trajectories of solutions from industry and other federal agencies. The challenges of large-scale are due to dedicated or special-purpose network connections with large (10-1000Gbps) and stable bandwidths that connect the computing and experimental facilities. The challenge of widespread collaboration is due to the network connections that span several countries with hundreds of users connected over middleware and grid environments. The challenges in this area include the scale of operations both in terms of data volumes and sizes of collaborative teams. In both cases, graded responses are needed for containing the effects of attacks while still safely operating these facilities, albeit at lower performance levels. Together, these two areas require a cyber security approach for DOE open science that is much more sophisticated than the current single-firewall approaches used in IP networks.

Research is required to analyze and characterize the structure and nature of specific DOE open science data, control and execution paths. Methods and technologies are needed to capture and process network data at extreme speeds for monitoring, packet filtering, anomaly detection, forensics, information fusion, integrated response, reduced false positives, fallback mechanisms, containment, and forensics. The networks that connect large computational and experimental facilities carry flows with known characteristics such as data transfer using custom protocols, but operate at bandwidths that are beyond the traditional firewalls. These special features must be exploited to design the intrusion prevention, detection and response methods that are optimized for these environments.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Challenges of Scale: Integrated intrusion prevention, detection and response methods are needed to handle network flows at terabps data rates and/or 1000 or more users. The following challenges must be addressed within the context this scale:

a. Profiling of threats must be carried out by taking into account the specific footprints and data rates, and transport dynamics of these environments.

b. Analysis of vulnerabilities of these open science environments must be carried out by taking into account the security measures at these facilities and networks at high data rates of these environments.

c. Monitoring methods and technologies must be developed to capture in real-time the network and host data at extreme speeds and volumes.

d. Processing and filtering methods and technologies must be developed to handle the network flows at extreme speeds.

e. Signature and packet filters must be developed to characterize both legitimate and attack network flows with special attention paid to data rates.

f. Anomaly detection methods must be developed by characterizing the special flows to and from these large-scale facilities and unacceptable deviations from them.

g. Methods must be developed for the reduction of false positives and unnecessary alert floods by combining signature- and anomaly-based methods, and by combining domain-specific and expert inputs.

h. Information fusion and correlation methods must be developed to detect complex and facility or DOE-wide coordinated attacks.

i. Forensic analysis methods for compromised systems must be developed to handle the massive amounts of network and host data collected during security incidents.

Graded Response Methods: Completely shutting down open science facilities or ongoing experiments due to intrusions is too expensive. Therefore, it is important to develop integrated responses that mitigate the effects of intrusions using attack containment methods and initiate fallback measures by selectively operating them at lower operational levels.

Non-Traditional Transport Methods: To effectively support massive storage and file systems that are located in DOE facilities separated by thousands of miles, newer methods such as Infiniband over SONET and FiberChannel over Ethernet are being developed. Intrusion prevention, detection, and response methods must be developed for these non-traditional transport methods, which are significantly different from traditional intrusion prevention methods.

Insider Attacks: Detection and prevention of attacks from insiders, and outsiders who masquerade with valid insider credentials is particularly critical for these expensive infrastructures. The structured nature of the user activities and the limited set of user codes and applications must be exploited to quickly detect suspicious user/applications activities that originate inside these infrastructures.

Specialized Network Attacks: The special networks that allow users and applications to co-schedule network connections with their facility allocations expose the network control-plane, either directly or through a proxy. It is important to protect these networks from user/application-based attacks that might compromise the network infrastructures.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

These research areas will enable the security and trust in geographically dispersed, high-performance and highly utilized critical open science infrastructure. Open science researchers located in geographically separated areas will be able to gain access to valuable computational and experimental facilities with unprecedented capabilities. They also can form world-wide collaborative teams to address complex open science problems in highly secure cyber infrastructures.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

This research will deliver the capability to field extremely large-scale cyber environments with fast detection and precise response to cyber intrusions from inside and outside the infrastructures. The research will result in reduced vulnerability to insider and external attacks and dramatically reduce the potential for catastrophic damages. More generally, research in these areas will contribute to the underlying tools and methods for “large-scale cyber security” issues such as monitoring, analysis and forensics, and integrated response including intrusion containment and robust fallback operations.

TIME FRAME

The research outlined here will be carried out in two phases:

- In the first phase, various technologies will be developed and tested over research testbed environments over the time period of the next 3 to 5 years.
- In the second phase, these solutions will be matured and field-hardened, and will be commissioned into production-level operations within the next 4 to 10 years.

REFERENCES

Agarwal, D., AS Bland, J. Bunn, C. Catlett, C. W. Cork, D. Dixon, T. N. Earnest, I. Foster, D. Gannon, M. J. Greenwald, J. Hodges, W. Johnson, W. Kramer, J. Leighton, G. McDermott, S. Merola, T. Ndousse-Fetter, H. Newman, L. Rahn, D. Schissel, M. A. Scott, G. Strand, R. Stevens, J. R. Taylor, B. Tierney, J. B. White III, M. Wilde, and L. Winkler. 2002. *High-Performance Networks for High-Impact Science*. R. Bair, ed. *Report of the High-Performance Network Planning Workshop*, available at http://www.doecollaboratory.org/meetings/hpnpw/finalreport. DOE Workshop. 2003. *DOE Workshop on Ultra High-Speed Transport Protocols and Network Provisioning for Large-Scale Science Applications*. 2003. Argonne National Laboratory, Argonne, Illinois, available at http://www.csm.ornl.gov/ghpn/wk2003.

DOE Science. 2003. *DOE Science Networking Challenge: Roadmap to 2008 Workshop*, June 3-5, 2003, Jefferson Laboratory, available at http://www.es.net/hypertext/welcome/pr/Roadmap/index.html.

Rao, N. S. V., S. M. Carter, Q.Wu, W. R. Wing, M. Zhu, A. Mezzacappa, M. Veeraraghavan, and J. M. Blondin. 2005. "Networking for Large-scale Science: Infrastructure, Provisioning, Transport and Application Mapping," *Journal of Physics: Conference Series*, **16**, 1, 541-545.

Rao, N. S. V., W. R. Wing, S. M. Carter, and Q. Wu. 2005. "UltraScience Net: Network Testbed for Large-scale Science Applications," *IEEE Communications Magazine*, **43**, 11, s12-s17.

Zheng, X., M. Veeraraghavan, N. S. V. Rao, Q. Wu, and M. Zhu. 2005. CHEETAH: Circuit-switched Highspeed End-to-End Transport Architecture Testbed," *IEEE Communications Magazine*, **43**, 8, s11-s17.

Enlightened Computing, available at http://www.enlightenedcomputing.org/.

Dynamic Resource Allocation via GMPLS Optical Networks, available at http://dragon.maxgigapop.net.

JGN II: Advanced Testbed Network for R&D, available at http://www.jgn.nict.go.jp/english/index.html.

Geant2, available at http://www.geant2.net.

ESNet On-demand Secure Circuits and Advance Reservation System(Oscars), available at http://www.es.net/oscars.

The Hybrid Optical and Packet Infrastructure, available at http://networks.internet2.edu/hopi.

Rao, N. S. V., Q. Wu, S. Carter, and W. Wing. 2006. "High-speed Dedicated Channels and Experimental Results with Hurricane Protocol," *Annales des Telecommunications*, **61**, 1-2, 21-45.

Falk A., T. Faber, J. Bannister, A. Chien, R. Grossman, and J. Leigh. 2002. "Transport Protocols for High Performance," *Communications of the ACM*, **46**, 11, 43-49.

Carter, S. M., M. Minich, and N. S. V. Rao. 2007. "Experimental Evaluation of Infiniband Transport over Local and Wide-area Networks," *Proceedings of the High Performance Computing Conference*, available at http://www.obsidianresearch.com/press/hpc2007.pdf.

Rao, N. S. V., W. R. Wing, Q. Wu, N. Ghani, T. Lehman, and E. Dart. 2007. "Measurements on Hybrid Dedicated Bandwidth Connections," *Proceedings of the INFOCOM2007 Workshop on High Speed Networks, 41-45, Anchorage, Alaska.*

Authors: N. Rao, Oak Ridge National Laboratory; M. Gupta, Indiana University; A. Striegel, University of Notre Dame; R. Brooks, Clemson University; P. Reiher, University of Southern California at Los Angeles

FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES
PRD-3 AND 4: RESILIENT COMPUTING IN FACE OF ATTACKS AND ACCIDENTAL FAILURES

ABSTRACT

Many U.S. Department of Energy (DOE) projects require long-lasting, distributed computation, making them an attractive target because a single weak link can disrupt and invalidate a computation that ran for days or weeks. Scientific computations are costly to repeat, so the infrastructure should be secured to be resilient to malicious and accidental service disruptions. The DOE's future security architecture must therefore:

- be resilient to intentional and accidental node and communication link failures
- detect and be resilient to data and computation corruption at a reasonable cost
- provide sophisticated mechanisms to prevent or detect and respond to attacks that target grid nodes and communication patterns.

EXECUTIVE SUMMARY

The DOE's future security architecture must be resilient to malicious and accidental failures of nodes and links. This resilience cannot be achieved by simply cloning resources and computations on multiple redundant nodes because most scientific computations operate at the edge of the available resources.

The following research directions will help provide resiliency at an acceptable cost:

- detection of attacks, data corruption, and accidental failures during computation
- extension and application of existing DOS defense mechanisms to DOE infrastructure, as well as development of DOE-specific DOS defenses
- attack response by fine-grained, on-demand replication of nodes or computation
- fine-grained error-recovery mechanisms to reconstruct lost data and results
- large-scale virtualization and compartmentalization
- disruption-resilient and/or randomized communication protocols.

Unlike the Internet environment, where the emphasis is either on surviving attacks through overprovisioning or responding to attacks via dynamic filtering and attribution, the focus of the proposed research must be on keeping the computation correct and efficient, in face of attacks and at an acceptable defense cost. Resilience techniques are thus the primary approach. Researchers should consider sophisticated attacks, such as corruption of data during computation, disturbance of communication between nodes, and “whack-a-mole” attacks that iteratively target responses to the attack.

SUMMARY OF RESEARCH DIRECTION

The DOE’s infrastructure is used for scientific computations that frequently require a large number of nodes and last for long periods of time (days or weeks). Malicious attacks targeting any node or communication link involved in distributed computation have a potential to halt the entire computation and invalidate results produced up to the point of the attack. Sophisticated attacks also are possible that corrupt data on one node– the corruption propagates when the results are merged at the end of the computation and may invalidate all results obtained in that run. Accidental failures resemble failures due to malicious activities but occur independently, randomly, and are usually limited to a small number of nodes. Any solutions that make DOE infrastructure resilient to malicious attacks also will handle accidental failures.

Resilient computation is usually achieved by naïve replication of resources and duplication of computation on replicated nodes. Such solutions, while effectively providing resiliency in face of failures, are not appropriate for DOE requirements. Scientific computations at the DOE require enormous amount of resources already and operate at the limits of the available infrastructure. Redundancy must be introduced selectively, at a fine resolution, and only when absolutely necessary. In addition to this, approaches are needed to detect and respond to sophisticated attacks launched by attackers familiar with the DOE’s infrastructure and computation/communication patterns. Informed, directed attacks are costly to handle, and resiliency is expensive under such attacks. For example, an attack that brings one node down should be handled by resiliency mechanisms developed under this program. But an attack that brings one node down, then moves to target another node involved in the same computation, and repeats this behavior indefinitely (a “whack-a-mole” attack) cannot be effectively handled through careful replication, because incremental costs of recovering from multiple node failures are higher than the attackers’ cost to target multiple nodes in sequence. Techniques are thus needed to detect and respond to sophisticated attacks. Such techniques should complement replication mechanisms to provide compact and resilient future security architecture.

Research is needed in the following directions:

- Develop techniques to detect sophisticated attacks that involve communication disruption, node failure, or overload and data corruption, and that may change a target dynamically. Another need is to develop techniques to respond to sophisticated attacks.
- Develop fine-grained, on-demand replication, computation, and communication redundancy mechanisms. Research should also address low-cost recovery mechanisms that facilitate reconstruction of lost data and computation results without full restart.
- Understand the analytic characteristics of particular types of attacks and defense techniques, providing insight into the practicalities and economics from both the attacker’s and defender’s perspective.

- Investigate large-scale virtualization and compartmentalization techniques to minimize the damage of attacks, and that acts as a preventive layer.

Existing DOS and intrusion defenses and fault tolerant mechanisms are likely to become parts of the final solution, but their combination is not a complete solution. Existing mechanisms need to be adapted to the unique DOE environment, where massive-scale computations need to be supported at minimal resource cost. We expect that significant novel approaches will be developed. These techniques are likely to be specific to the special situations that open science computations face, and thus are much less likely to be developed and deployed if the DOE does not guide and fund research efforts in this direction.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

- Resilient computing at an affordable cost has been an open problem for the last 50 years. While some advances have been made, the problem is still challenging.
- Detection of and recovery from data corruption during computation requires sophisticated techniques to model the computation flow and detect anomalies, or requires redundant computation. The first approach is difficult because new models for computation representation and intermediate result checking need to be developed. The second approach is difficult because its naïve application is too expensive for the DOE environment.
- Sophisticated DOS attacks are very difficult to detect and to defend against. This challenge is even more prominent in DOE environment because the defense costs must be kept low. On the other side, participants in the DOE infrastructure can be required to deploy a given security solution and can be authenticated reliably, thus the DOS problem is more constrained in certain aspects than the problem of protecting communication in the open Internet.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

This proposed research has many potential advantages for science. We expect this research to produce:

- cost-effective replication methods for resilient computing and communication
- sophisticated attack detection and response mechanisms
- error-recovery mechanisms and models for checking the validity of computation during the run
- virtualization and compartmentalization techniques to protect distributed computation from intrusions, corruption and error propagation.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

This proposed research has many potential advantages for the DOE. We expect this research to produce:

- prevention of the loss of vast amounts of computer and network resources due to attacks targeted at open science computations
- higher assurance that open science computational results have not been corrupted by malicious attackers
- resilient computing and communication mechanisms that guarantee on-time, correct execution for open science applications
- sophisticated DOS detection and defense mechanisms will increase safe computing capacity available to open science researchers.

TIME FRAME

- The development of isolated replication and error-recovery mechanisms will take up to 5 years.
- The development of isolated sophisticated attack detection and response mechanisms will take up to 5 years.
- The integration of promising techniques into a compact security architecture is expected to take up to 10 years.

Authors: J. Mirkovic, B. Reid, P. Reiher, S. Wakid

FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES
PRD-5: ANOMALY DETECTION IN CONTROL SYSTEMS

ABSTRACT

Engineers are much better at building reliable and useful devices than at understanding why the devices work. For open science, large-scale problems, this characteristic exposes us to the certain knowledge that we will have not yet discovered all system failure modes, even for systems that have been accredited for widespread use. For future complex systems, such failures can result from malicious or inadvertent human actions at the man-in-the-loop level or malicious or inadvertent control law automaton actions at the automatic-control-loop level. Science does not currently exist to support construction of predictive models of large-scale system evolution for open science projects. Furthermore, without significant effort, science will not exist to predict the future state of the self-healing, adaptive, complex systems being contemplated. Thus, how will future decision makers decide whether system failures or degradations occur due to unforeseen malicious attacks or unanticipated operator error or a previously unobserved combination of component failures? In order to detect anomalous behaviors in future control systems, we must first achieve a clearer understanding of the meaning of observed behaviors for future complex systems. This Priority Research Direction (PRD) seeks to resolve this shortfall through creation of technologies for detection of anomalous open science control system behaviors.

EXECUTIVE SUMMARY

Success in certification and accreditation of open science computing and communication resources for widespread use will only occur if we are successful in resolving the current inability to adequately understand complex control system behaviors well enough to build predictive models of those behaviors.

We propose a two-pronged approach to achieving the needed assurance that the complex open science control systems being contemplated will be adequately modeled:

- a lower-risk approach of constructing templates of expected behaviors under expected operational constraints and parameter variability. Such templates could be used to predict expected behaviors for a wide range of system operating conditions and inputs
- a higher-risk approach of building predictive models of components and composing the components to achieve predictive models of the resulting system.

This PRD will enable open science system operators to make more informed decisions during anomalous system operation to restore the system to a normal state. Without being able to understand the origin of an anomalous condition, autonomous correction will be unattainable and manual corrections initiated by operators will be based upon experience and heuristics.

SUMMARY OF RESEARCH DIRECTION

This PRD will result in technologies for developing models of open science system behaviors, especially under appropriate ranges of parameters and complexities of control systems. The primary results will be:

- development of a generic template for anomaly detections in the context of adaptive, self-healing, open science control systems
- development of predictive models of open system components and approaches for composing the components to achieve predictive models of the resulting system.

The modeling and simulation community has been consistently improving our ability to validate that the system under analysis meets user requirements and to verify that the system as built meets the requirements as stated. However, current science does not support achieving this level of support for the adaptive, self-healing system needed to achieve the goals of open science projects. Extensions to current science are needed to achieve the ability to detect anomalous behaviors of complex control systems for a wide variety of conditions.

The template is a coarse-grained, pattern-recognition approach to anomaly detection. For anomalous events in which structural changes to the system architecture are not encountered, the template approach should be effective in achieving reliable estimates of a future system state from a current system state and intermediate system inputs that differ parametrically from expected inputs. These template results can then be applied to identify the character of an anomalous condition that may arise.

The predictive model is a fine-grained approach to anomaly detection and analysis. This approach also should be effective for structural changes to system architectures as well as for parametric variability for a given architectural structure. While this achievement is preferred, such a result also will be harder to achieve than building templates for expected behaviors for a range of functional relationships and parametric values.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The complexity of the open science cyber security challenge is more than the scale of computing, storage, and communication systems being contemplated (exaflops, exabytes, and terabits per second). That is, while the magnitude of the system interactions alone exceeds our modeling capabilities, the problem is made more difficult by the fact that the nature of the individual interactions are more complex than our ability to fully understand. Science does not exist to build predictive models of the complex interactions of social networks engaged in distributed decision-making (BAA 07-56) and distributed large-scale dynamical systems (Final Report 2003) under control. The control system community has been actively involved for more than a decade in creating the science and technology for understanding interactions of event-based systems and continuous systems (Lee and Varaiya 2003) but the capabilities do not extend to understanding (predicting) the variety of intrusion mechanisms for high-speed communications networks (Bro 2007).

Anomaly detection in future power generation and distribution control systems also will be complicated by the fact that new systems will be interacting with the legacy systems.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Successful completion of this PRD will substantially improve the ability of computational science to support complex system design, implementation, and operation. More importantly, the ability to perform anomaly detection and analysis also will provide the basis for achieving proactive management instead of reactive control of cyber events. Results from this effort will be widely applicable in electric power systems. The results also will be valuable in the context of the future mix of analog and digital and man-in-the-loop control systems.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

We intend to move from reactive response to detected control anomalies to proactive prevention of open system control anomalies. These anomalies may be due to malicious-or-inadvertent, human-or-automaton activity. This Priority Research Direction will lay the groundwork necessary to move from reactive response to anomalous control behaviors due to cyber events to proactive prevention of anomalous control behaviors due to cyber events.

In addition to enabling improved management of cyber events, the research results will be widely applicable to power systems and process control systems in general.

TIME FRAME

- The templates will be available in 5 to 7 years.
- The composable models for predictive control will be available in 10 to 15 years.

REFERENCES

(BAA 07-56.) BAA 07-56, Deep Green Broad Agency Announcement (BAA) for Information Processing Technology Office (IPTO) Defense Advanced Research Projects Agency (DARPA), available at http://fs2.fbo.gov/EPSData/ODA/Synopses/4965/BAA07-56/BAA07-56DeepGreen.pdf.

Bro. 2007. *Bro Intrusion Detection System*, available at http://www.bro-ids.org/.

Final Report. 2003. Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, U.S.-Canada Power System Outage Task Force, available at http://www.nerc.com/~filez/blackout.html.

Lee, E. A., and P. Varaiya. 2003. "Structure and Interpretation of Signals and Systems," Addison-Wesley, Reading, Massachusetts.

Authors: N. Pundit, J. James, W. Bradford

FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES
PRD-7: ECONOMICS-BASED SECURITY ARCHITECTURE

ABSTRACT

To promote scientific research and maintain the U.S. energy infrastructure, the U.S. Department of Energy (DOE) must guarantee the security of large heterogeneous networks maintained by coalitions of entities not under its direct control. The DOE's future security architecture must therefore:

- be self-enforcing
- reflect security needs and motivations of all participants
- encourage the efficient use of resources
- reflect the motivations of malicious entities
- allow analysis of both existing and possible future attack vectors.

Metaphorically, we view the architecture as a security marketplace. Only through careful analysis of the motivations of all participants and their associated costs will it be possible to develop a security framework that provides durable and adaptive security. This framework will use the tools of game theory, the branch of mathematics devoted to adversarial relationships to express, among other things:

- the black market motivations of attackers
- the costs to users of attack countermeasures
- system design, implementation, and maintenance costs.

EXECUTIVE SUMMARY

The DOE's future security architecture needs to scale well, enforce itself, support interaction among many independent actors, and be resilient to attack. One major problem with current security approaches, as identified by Ross Anderson at Cambridge, is that economics is currently on the side of the attacker, i.e., it is much cheaper to mount attacks than to create error-free systems. The goal of this research thrust is to use the tools of economics, including game theory, to better understand the computer and network security problem domain and reverse the situation.

Researchers need to create a framework including a hierarchy of game models that express:

- the motivations of all parties
- actions available to malicious parties
- possible security measures.

The framework needs to be open-ended to allow for analysis of emerging threat vectors. One goal of the research is to find Nash equilibria for the system when they exist (Stackelberg equilibria when they do not) so that optimal self-enforcing security strategies can be developed.

SUMMARY OF RESEARCH DIRECTION

Deregulation of the electrical grid has produced a large-scale free market system for energy distribution. Enron Corporation successfully manipulated the electrical grid to defraud the state of California. Recent surveys have shown that computer criminals are now motivated mainly by commercial profit; intellectual curiosity and idealism are no longer the driving forces in the hacker community. Ross Anderson of Cambridge has shown that the cost of breaking into computer systems is more than an order of magnitude less than the cost of securing the same system, and as the size of the system scales this difference only becomes greater. These examples are only the most obvious DOE-relevant interactions between economics and security.

The future security architecture for DOE systems needs to function both for the national critical infrastructure and for DOE's computational infrastructure for science. On an abstract level, the energy and computation grids have much in common. They are massive systems built from small components embedded in complex networks. The networks are continually evolving and their dynamics are poorly understood. Both grids are accessed internationally and administered locally by independent entities. Many aspects of local administration are not directly under DOE's control. In spite of this, DOE's mission requires it to find a way to maintain the security of the grid.

To address these issues, researchers need to construct models that express the realities of infrastructure security:

- Attackers are driven by market forces, searching for ways to subvert the system in their favor.
- Security countermeasures have real costs; attackers may even fake attacks in order to provoke security responses that hurt system response time.
- Attackers and security personnel play a zero-sum game in an abstract environment.

In phase I:

1. A model of the system infrastructure needs to be developed.
2. Payoffs to all parties need to be expressed in a common format.
3. A complete set of strategies (attacks and countermeasures) for all parties needs to be developed.
4. The set of strategies needs to be open to allow new classes of threats to be analyzed.
5. Nash equilibria of the system need to be derived when possible.
6. Solutions to games of kind need to be found, so that system inflection points are known.

Phase II research should transition these insights into practical applications. The goal is to have a self-enforcing security framework. All non-malicious participants will provide near-optimal security for their part of the infrastructure, because it is in their self-interest and it is the least costly course of action.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The proposed work has multiple challenges, but carries with it a very good chance of success. While it is clearly a challenge to adequately model the security infrastructure as a set of games, computers and their networks are more tractable than the economic motivations typically studied by economists. Countermeasure costs in terms of real dollars and infrastructure availability can be clearly defined. The monetary motivations of attackers also are easily captured.

Considering the purely technical aspects of the analysis:

- If computer networks are naively modeled as an abstract board game, the number of potential moves is astronomical. Note that the chess board consists of an 8-by-8 grid, and Go is played on a 16-by-16 grid. The number of possible positions for an attacker on even a small campus network is enormous in comparison. While this seems daunting, it should be noted that work by Conway and Berlekamp on combinatorial game theory has developed a number of techniques where near-optimal solutions to intractably large problems can be found tractably by pruning the search space. Unfortunately, Conway and Berlekamp's tools currently assume perfect knowledge by all players. This is unlikely to be the case.
- While the interactions between attackers and defenders can be viewed as zero-sum, interactions between defenders working for different autonomous systems (ASs) will be modeled as cooperative games. When Nash equilibria exist, they all have the same values for zero-sum games. Any combination of zero-sum equilibria forms an equally good Nash equilibrium. Mixing cooperative Nash equilibria can be problematic.
- The system can be expressed using a pay-off matrix if all players have a finite number of moves, Markov decision problems, or differential equations. In the first case, linear programming can be used to find the optimal solutions, in the sense of Nash, directly. When Markov decision models are used, dynamic programming is typically used to find optimal solutions. This tends not to scale well. Differential games may not have Nash equilibria and it can be challenging to prove whether or not solutions exist. In general, this problem space is P-Space completer (worse than NP-complete).
- If Nash solutions do not exist, Stackelberg (leader-follower) solutions always exist. It is not unreasonable in this situation to allow defenders to be the leader and define the game to be played.

But the problem also requires knowledge of the problem domain:

- It is challenging to develop a good model of the computer criminal black market.

- Many aspects of the computer and electrical grids are continually in flux; the infrastructure model must include stochastic and dynamic factors.
- It is difficult to quantify system interactions.
- The model must be open-ended to support exploratory analysis of the problem space and allow new threats to be analyzed.
- The system is likely to contain second- and third-order effects that are difficult to capture.
- The set of observations available to the defenders is likely to be quite small, making it difficult to distinguish between enemy moves.
- Daily system operations are so full of errors that it is often impossible to detect attacks while they are under way.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

This proposed research has many potential advantages for science. We expect this research to produce:

- a mathematical basis for computer and network security
- techniques for distributed enforcement adapting across scales
- tools for analyzing system security as a part of the design process
- tools for analyzing the quality of the software development and testing process as a function of possible security vulnerabilities.

The game theory tools available to the researcher include number theoretic tools, differential equation models, and optimization tools (mathematical programming). This work can integrate these tools into the network and software development process.

This work can integrate the effect of false positive effects into system countermeasure strategies.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

This proposed research has many potential advantages for DOE. We expect this research to produce:

- self-enforcing security policies
- distributed enforcement adapting across scales
- security for large heterogeneous systems without centralized control
- determine whether or not optimal (sense of Nash) security strategies exist
- determine optimal strategies (where possible) for attack and defense
- integrate the effect of false positive effects into system countermeasure strategies.

TIME FRAME

- The defense of electric grid applications will be available in 3 to 5 years.
- DOE frameworks for scientific collaborations will be available in 3 years.
- Attack-resilient network infrastructure will be available in 10 years.

Authors: R. Brooks, M. Sachs

FUTURE SECURITY ARCHITECTURES AND INFORMATION ASSURANCE TECHNOLOGIES
PRD-8: MALWARE RESEARCH FOR OPEN SCIENCE

ABSTRACT

The high value of U.S. Department of Energy (DOE) open science targets will engender an increased and unique malware threat. Risk is the product of vulnerability and threat, and large-scale open science systems will be just as vulnerable as conventional systems while also facing a distinctive, pronounced threat from sophisticated attackers dedicated to compromising specific DOE open science systems. The unique nature of this threat places it outside the scope of what conventional anti-virus vendors will offer solutions for, *and specific characteristics of DOE systems require research particular to those systems.*

EXECUTIVE SUMMARY

Attackers, such as agencies of foreign governments, who target DOE open science systems, will have a variety of custom malware attacks at their disposal. Commercial anti-virus products are designed to protect a general customer base against the most common malware attacks, so that no malware products will be available on the market to protect DOE open science systems.

We propose the development of a controlled cyber environment for the creation and experimentation of advanced malware. To build such an environment will push the limits of our ability to build large-scale experimental environments and advance our still-developing understanding of malware and defending against it.

The impact of this research direction on computational science will come from a combination of system building for the controlled, realistic environment necessary and novel algorithms for malware analysis and detection. This research will impact cyber security for open science by enabling researchers to conceive, investigate, and develop the tools to mitigate an increased, unique malware threat that will challenge open science systems.

The time frame for this work to impact the defense of DOE open science systems against malware is 5-7 years.

SUMMARY OF RESEARCH DIRECTION

Open science systems have unique characteristics, such as high-speed network connections and large computational and storage resources that can affect malware behavior. More importantly, traditional high-value government targets can draw a clearer line between the systems being protected and the outside world than is possible in open science. The insider threat is a problem for other government agencies, but the insiders are typically employed by that agency. For open science systems, there is a unique threat of Trojans, backdoors, and other targeted malware installed by foreign agencies that have been granted access to the system. For example, Trojans that steal information through system inference channels (Percival 2005; Zalewski 2005; Wang and Lee 2007) can be mitigated in other scenarios by protecting a perimeter around the system or through process isolation. For open science systems, the sharing of system resources, through which inference channels leak information, is a requirement.

Furthermore, there are DOE systems such as the power grid on which conventional malware can have indirect effects that are unique to DOE missions and DOE systems. An example is the effect that the Blaster worm had on the August 14, 2003, blackout (Final Report 2003). These indirect effects can only be measured only with testbeds that mirror a DOE-specific environment.

Research is needed to develop a controlled cyber environment for the creation and experimentation of advanced malware. When sophisticated attackers, such as agencies of foreign governments, seek to compromise DOE and open science systems they will have an exceptional arsenal of custom malware at their disposal–including not just viruses and worms, but a variety of Trojans and other components that are parts of larger coordinated attacks. *The open sharing of DOE system resources presents a unique threat of Trojans that is not faced by other high-target systems. While other government systems do face an insider threat, it is distinct from the outsider threat and can be handled differently. With open science systems, no distinction exists between insider and outsider.*

We must be able to anticipate advanced malware techniques, such as cryptovirology (Young and Yung 2004)–where advanced cryptographic techniques hide what the malware is doing, worms that are diversions to hide a more directed attack (Kumar et al. 2005), timebomb attacks (Crandall et al. 2006), advanced rootkit techniques (King et al. 2006; Krugel et al., 2004), and polymorphic/metamorphic malware (Newsome et al., 2005; Crandall et al. 2005; Christodorescu et al. 2005). The business model of anti-virus vendors is centered around analysis and response resources devoted to common threats, or those that pertain to the majority of the vendors' customers. To protect against threats that are specific to DOE and Open Science Initiatives will require analysis and response resources that are dedicated to DOE and open science needs, and techniques that are specific to unique threats.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

To build a controlled cyber environment for the creation and experimentation of advanced malware will push the limits of our ability to build large-scale experimental environments and advance our still-developing understanding of malware and malware defense.

The challenge of building an experimental environment for malware research is that the environment must be realistic while maintaining other important properties, including containment and reproducibility of results (DETER 2007). A current trend in malware research is toward behavior-based analysis and detection (Kirda et al. 2006; Crandall et al. 2006; Christodorescu et al. 2005). The power of behavior-based techniques lies in the way that they incorporate the complexity of the environment into the results of the analysis or detection.

Malware is defined by its malicious behavior and this behavior is as much, and sometimes more so, a reflection of the environment as it is a property of the malware itself.

This means that malware must be analyzed in an environment that matches the complexity of its natural environment as closely and in as much detail as possible. It also means that other testbeds built for malware analysis, such as DETER/EMIST (DETER 2007), will not be appropriate for testing malware that is developed for specific DOE and open science environments, which have characteristics that are very different from conventional environments. *For example, in order to test the indirect effects of a worm on the power grid, the testbed must incorporate aspects of the power grid network that are not modeled by generic Internet testbeds. In another example, it may be possible to compromise scientific data of a batch process running on a supercomputer with many nodes simply through inference channels via another process on the system. This is a threat not faced in typical supercomputer applications; evaluating information-theft Trojans for this particular threat requires a testbed for the specific environment.*

Building an environment for malware creation and experimentation also will help to advance our understanding of malware and malware defense. Academic research on malware is still developing and the tools we need to address advanced malware threats can only be developed on a solid theoretical foundation of malware that is firmly planted in practice. The proposed testbed will help to build that foundation.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The impact of this research direction on computational science will come from a combination of system building for the controlled, realistic environment and novel algorithms for malware analysis and detection.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

This research will impact cyber security for open science by enabling researchers to conceive, investigate, and develop the tools to mitigate an increased, unique malware threat that will challenge open science systems.

TIME FRAME

- Due to the need for a theoretical foundation for malware research that is firmly planted in practice and the unique nature of the malware threat DOE open science faces, it is expected that this research direction will inform actual malware defense initiatives in 5 to 7 years.

REFERENCES

Christodorescu, M., S. Jha, S. A. Seshia, D. Song, and R. E. Bryant. 2005. “Semantics-Aware Malware Detection.” In *2005 IEEE Symposium on Security and Privacy,* held in Oakland, California, 32-46. IEEE Computer Society, Los Alamitos, California.

Crandall, J. R., Z. Su, S. F.Wu, and F. T. Chong. 2005. “On Deriving Unknown Vulnerabilities from Zero-Day Polymorphic and Metamorphic Worm Exploits,” in *Proceedings of the 12th ACM Conference on Computer and Communications Security (CCS),* held in Alexandria, Virginia, eds. C. Meadows and P. Syverson, 235-238, ACM Press, New York.

Crandall, J. R., G. Wassermann, D.A.S. Oliveira, Z. Su, S.F. Wu, and F.T. Chong. 2006. “Temporal Search: Detecting Hidden Malware Timebombs with Virtual Machines,” in 12th *International Conference on Architectural Support for Programming Languages and Operating Systems* (ASPLOS XII), held in San Jose, California, 25-36. ACM Press, New York.

DETER. 2007. *DETER A Laboratory for Security Research: Community Workshop on Cyber Security and Test 2007*, Boston, Massachusetts, available at http://www.isi.edu/deter/.

Final Report. 2003. Final Report on the August 14th Blackout in the United States and Canada: Causes and Recommendations, U.S.-Canada Power System Outage Task Force, available at http://www.nerc.com/~filez/blackout.html.

King, S. T., P.M. Chen, Y.-M. Wang, C. Verbowski, H.J. Wang, and J.R. Lorch. 2006. “SubVirt: Implementing Malware with Virtual Machines,” in 2006 *IEEE Symposium on Security and Privacy,* IEEE Computer Society, Los Alamitos, California, held at Berkeley, California.

Kirda, E., C. Kruegel, G. Banks, G. Vigna, and R. Kemmerer. 2006. “Behavior-based Spyware Detection.” In *Proceedings of the 15th Usenix Security Symposium*, held in Vancouver, British Columbia, 273-288, Usenix Association, Berkeley, California

Kruegel, C., W. Robertson, and G. Vigna. 2004. “Detecting Kernel-Level Rootkits Through Binary Analysis,” in *Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC'04),* held in Tucson, Arizona, 91-100, IEEE Computer Society, Los Alamitos, California.

Kumar, A., V. Paxson, and N. Weaver. 2005. “Exploiting Underlying Structure for Detailed Reconstruction of an Internet-scale Event,” in *IMC ’05, 2005 Internet Measurement Conference,* held in Berkeley, California, 351-364, Usenix Association, Berkeley, California.

Newsome, J., B. Karp, and D. Song. 2005. “Polygraph: Automatically Generating Signatures for Polymorphic Worms,” in *2005 IEEE Symposium on Security and Privacy,* held in Oakland, California, 226-241, IEEE Computer Society, Los Alamitos, California.

Percival, C. 2005. “Cache Missing for Fun and Profit,” available at http://www.daemonology.net/papers/htt.pdf.

Wang, Z., and R. B. Lee. 2007. “New Cache Designs for Thwarting Software Cache-based Side Channel Attacks,” in *Proceedings of the 34th International Symposium on Computer Architecture (ISCA’07*), held in San Diego, California. Also published in ACM SIGRCH Computer Architecture News, **35,** 2, 494-505.

Young, A., and M. Yung. 2004. “Malicious Cryptography: Exposing Cryptovirology,” Wiley, Hoboken, New Jersey.

Zalewski, M. 2005. *Silence on the Wire*, No Starch Press, Inc., San Francisco, California.

Authors: N. Pundit, W. Bradford, J. Crandall

BREAKOUT SESSION
HUMAN FACTORS (HF) ANALYSIS

Breakout Leads: Anne Schur, Pacific Northwest National Laboratory;
Joe St. Sauver, Internet2

CHALLENGES OF HF IN CYBER SECURITY IN OPEN SCIENCE

While researchers are accustomed to protecting their research data, they have not been focused, historically, on broader cyber security considerations. Moreover, little information is available that addresses the specific needs of the open science research community from an HF perspective.

This workshop was focused on identifying specific HF cyber security research areas that needed to be addressed to enable researchers to perform their science without noticing intrusive cyber security provisions.

To improve the security of the open science community, we need to investigate and understand how the community does its work. Security measures must be unobtrusively integrated with the science research work activity processes.

A unique challenge that the open science community must address is the increasingly distributed nature of eScience. No longer do researchers run their jobs on a single local system−now they may use networked resources distributed around the country and/or around the world. In this operational environment, even "simple" things, such as authentication, can be a challenge. As the number of systems grows, the number of accounts and passwords associated with those systems also increases and scaleably supporting authentication becomes harder. We need to look for new authentication paradigms that will usably **scale** to increasingly complex networked environments. One example, which was highlighted as part of our group's work, was the promising approach embodied by federated authentication.

Because of the complexity of the networked environment, scientists also no longer are able to self-monitor their infrastructure and respond to adverse events. Techniques are needed to assist them to be situationally aware of the status of their infrastructure, make decisions, and determine actionable coordinated responses that mitigate the adverse event. A research agenda must also address cyber security methods and success metrics, perhaps via a modeling and simulation testbed that may give insights into the impact of policies and new technologies on the practice of science. We also address access. For example, do researchers need fundamentally new ways to interact with the open science environment?

All of the areas we recommend look at research needed to address long-term needs−things that are at or beyond the 5-to-10-year horizon. The understandings gained from this research will help characterize potential threat vectors and define new architectures and tools of the future, which can better eliminate vulnerabilities and malicious behavior and, most importantly, ease the process of doing scientific research. When identifying these areas, consideration was given but not limited to:

- profiles of a young unsophisticated hacker as well as experienced, sophisticated attacker
- usability issues: “deployability,” supportability, accessibility, complexity, etc., vs. security
- cyber warfare and cyber terrorism by individuals and groups
- legal issues, privacy, Communications Assistance for Law Enforcement Act (CALEA), beyond CALEA
- infragard, Research and Technology Protection, Secure Internet Gateway
- malware issues including: anti-spam, anti-virus, anti-spyware; history and projection− how will we be protected from these 5 to 10 years hence?
- real-time systems behavioral analytics
- intellectual property issues (ownership and rights of developers)
- training information technology security specialists to detect, isolate, and deal with cyber security threats.

HUMAN FACTORS (HF) ANALYSIS
PRD-1: USABILITY OF SECURITY SYSTEMS

SUMMARY OF RESEARCH DIRECTION

The scale and diversity of the U.S. Department of Energy (DOE) open science environment is unprecedented; it offers access to leading-edge supercomputer and large-scale experimental facilities while supporting open science collaborations of international researchers. Due to its increasing complexity of services, interconnection, and scale, providing the DOE open science environment with security that is easy to use and maintain is a great challenge. It is essential that the human interface to this complex, heterogeneous environment be designed for ease of use so that users can routinely and automatically apply protection mechanisms correctly. Today, many approaches to security fail this requirement. Security systems often are subverted or left open to exploitation due to misconfiguration brought about by poor human design and interfaces. For example, users often subvert secure authentication by keeping unprotected copies of complex passwords. Grid middleware that provides secure access to many different resources is often too complex and difficult to properly configure and maintain. Due to poor user interfaces, users often ignore or make inappropriate responses to security alerts.

Research is required into human performance in the use of security systems, design of secure systems for ease of use, and human-computer interfaces. The goal of this research is to ensure that users will routinely and correctly apply security mechanisms, that those mechanisms are easy to use and maintain, and human-computer interaction is as effective as possible. The areas of particular interest include: user interfaces to authentication mechanisms, user response to security alerts, protection mechanism design for ease of use, implementation and maintenance training, and human-computer interaction in detection and response to attacks.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

HF research includes human performance, technology design, and human-computer interaction. All three research areas are aimed at improving the usability and effectiveness of security systems for DOE open science.

Human performance research is conducted to improve our understanding of how humans interact, reason, adapt, and respond to different situations, technology, and environments. Through observation, monitoring, and analysis, we are better able to predict and enable human performance. The challenge is to apply human performance research to specific security mechanisms, such as authentication, to ensure users will routinely and automatically use them, as well as to secure design to ensure correctness and ease of training and maintenance. To minimize user mistakes, human performance research is needed to investigate the extent to which the user’s mental image of his protection goals matches the mechanisms he must use. Furthermore, research is needed into the HF related to the implementation of security policy.

Human-computer interaction research is designed to bring together humans and technology in order to more rapidly and effectively deal with attacks. Intrusion and detection systems generate massive amounts of data; the challenge is to quickly find the threats, characterize them, and take appropriate action.

Achieving knowledge discovery involves processes, dialogues, and actions that a user employs to interact with a computer, such as visual analytics. Research is needed into developing scalable tools that can more quickly and accurately discover knowledge from ever-expanding intrusion detection databases.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

A more secure and usable environment for open science will result when security protection mechanisms are made easier to use routinely and correctly, protection mechanisms are easier to configure, and the ability to discover attacks are enhanced. Making security easier to use and implement will increase accessibility to larger communities of researchers and provide more assurance of sharing science. Finally, knowledge discovery will increase availability and protection through improved response to attacks.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

HF research will reduce vulnerability of diverse, large-scale computing environments by enabling users to routinely use security system protection mechanisms correctly, for systems administrators to configure systems correctly, software analysts to better maintain systems, and for operators to effect security measures.

TIME FRAME

The HF research described above involved both applied and fundamental research:

- Applied research, such as knowledge discovery and user interfaces, can be accomplished in 3 to 5 years.
- More fundamental research into human performance and design criteria for usability and assurance of routine and correct use may take up to 10 years to see production implementation.

HUMAN FACTORS (HF) ANALYSIS
PRD-2: SECURITY POLICY IMPLEMENTATION IMPACTS ON USABILITY

ABSTRACT

Security policies often are the cornerstone of the actual security that we experience. Usually created by diverse independent agencies, security policies often are heterogeneous and conflicting. Even when there is consensus among the diverse stakeholders, security policies are difficult to translate into practice such that they result in configurations that comply with policy, that are usable, and that can be understood and the implications assessed in a timely fashion. The consequences of not considering the special HF security needs of open science research are great and can potentially put open science at risk. Researchers conducting their work activities purposefully or inadvertently circumvent cyber security protections if those protections are not usable. The proposed research addresses the development of i) a flexible modeling and simulation testbed capability to understand the impacts of and gain insights into the usability of policies on systems for open science and open science practices, and ii) methods and metrics to implement and audit policy implementation correctness and usefulness. Providing this capability to the research community: 1) affords security and other policymakers the ability to understand the implications of the policy prior to deployment, 2) allows more reliable, effective, and timely transfer and implementation of policy into practice, for example, by enhanced strategic and tactical planning, and 3) enables a better investment of resources.

EXECUTIVE SUMMARY

Policies often are the cornerstones of the actual security that we experience. In the security arena, policies are created by diverse independent agencies, resulting in heterogeneous and often conflicting policies. Even when there is consensus among the diverse stakeholders, it is difficult to translate security policy into practices that result in configurations that comply with policy and are usable. Finally, in today’s open science environment, it is difficult to understand and assess the implications of new security policies in a timely fashion.

This Priority Research Direction (PRD) specifically addresses these needs by developing a flexible modeling and simulation testbed (M&STB) capability to understand and gain insights about the impacts that policies have on the usability of systems utilized for open science and open science practices. A critical aspect of this capability will be the development of methods and metrics to carry into effect and audit policy implementation correctness and usefulness.

For scientific research to be efficient and to continue to be fruitful, the infrastructure of the open science environment must be easy to use by the community that it serves. Ease-of-use is about how readily and intuitively people can employ a particular tool to achieve a particular goal. This usability also can include methods and metrics to assess ease-of-use and the study of the principles behind entities perceived efficient and or esthetically pleasing.

Providing usability is challenging when a critical requirement of the infrastructure for conducting open science is to maintain its security and integrity for use anytime and anywhere by a broad spectrum of users involved. The consequences of not considering the HF security needs for open science research are great and can potentially put open science at risk. Cyber security protections that are in place and hard to use can be purposefully or inadvertently circumvented by researchers performing their work activities, thus compromising security. In a science community that is electronically interconnected, this can result, at worst, in widespread consequences that are an order of magnitude equivalent to the August 14, 2003, power blackout. The scientific enterprise could come to a halt, impacting a wide range of users from the researcher and network administrators to those maintaining experimental instruments and the public who benefits from the science. Additionally, much data and instruments can be irrevocably damaged.

A policy modeling and simulation testbed capability provides the opportunity for timely discovery of flaws in policies and their implementation before they are put in place in multiple contexts. Often, it is not until after the fact and over much time that problems of use and vulnerabilities become apparent. In an age where scientific research is conducted globally via interconnected systems, security and the impact of scale are important. An M&STB capability could address this need.

While there are simulation and modeling testbeds that address policy, many are from the perspective of exploring whether the cyber security that *exists*, for example, a tool or procedure, or if the new proposed cyber solution *supports* the *existing* policy (Rue et al. 2007; Firmino 2005). This capability is important but is not proactive and results in "patching" what exists to comply with the mandate. To address this proactive gap, the capability to address *how proposed policies impact security usability prior* to these mandates being implemented in real-world operational contexts is needed.

Historically, cyber security and information security (CIS) are approached from a technology-centric viewpoint. Solutions for CIS vulnerabilities and protection from breaches tend to focus on technical mechanisms, e.g., stronger firewalls and better encryption. In the last several years technical solutions have included HF. These efforts are seen in multiple complex domains−from cyber attack by terrorists and national infrastructure protections, to simulations for training of cyber awareness and many others, but none for open science research. There is need to extend this to our open science enterprise and address security with ease-of-use.

While there are simulation and modeling testbeds that address policy, many are from the perspective of exploring whether the cyber security that *exists*, for example, a tool or procedure, or if the new proposed cyber solution *supports* the *existing* policy (Rue et al. 2007; Firmino 2005). This capability is important but is not proactive and results in "patching" what exists to comply with the mandate. To address this proactive gap, the capability to address *how proposed policies impact security usability prior* to these mandates being implemented in real-world operational contexts is needed.

We propose a research agenda that specifically addresses the provision of a proactive capability to policymakers to enable them to understand the implications of the policy from users' viewpoints integrated with the technologies that are serving them. The impact of providing this capability is better allocation of resources, more reliable, effective, and timely transfer and implementation of policies, and more effective science.

SUMMARY OF RESEARCH DIRECTION

A dual path effort is envisioned where each path addresses the key outcomes of this research: 1) the development of a flexible M&STB capability to understand and gain insights about the impacts that policies have on the usability of systems used for open science and open science practices, and 2) the development of methods and metrics to carry into effect and audit policy implementation correctness and usefulness.

User-centered System Software Engineering (Norman and Draper 1986): Fundamentally, a security policy is a high-level definition of secure behavior for a technical system or an organization. From an organizational viewpoint, constraints on the human behavior of non-adversaries and its members as well as constraints imposed on adversaries are addressed. For systems, constraints are addressed on functions and flow among them, on access by external systems, including software programs, and access to data by people. To succeed in this new research, it is critical that the human and technical aspects of this program be tightly coupled. A user-centered software engineering approach will be used. A core multidisciplinary team will be established that, at a minimum, is composed of expertise in the following areas: parallel system architectures, networking, policy making, decision- making, modeling and simulation algorithm development, HF, information systems, statistical analysis, and SCIENTISTS.

Establishing Requirements: As a starting point, this program will focus on identifying the needs of the open science community from the perspectives of the users' role and technology in an eScience enterprise context. In addition to understanding what people do and what they need in order to perform their variety of roles, "how" questions will be asked. "How-type" questions offer insights into how science is done and the relationships that exist between those with different roles and the technologies they need. How questions also can indicate the cognitive ("thinking") aspects of their work that also must be supported. The resulting specifications of the M&STB will integrate technical requirements and the HF requirements.

Defining Models of the Scientific Processes and *Support to the Open Science Enterprise*: One critical aspect of this research is the characterization of *how* the work processes of the many stakeholders are manifested in addition to the kinds of cognitive work activities that need to be supported (Klein 1999). For example, individuals collaborating on problem solving a specific science issue via the infrastructure also will need to coordinate an activity with each other. These two work processes are cognitively different and may require different capabilities to provide security. Model and simulation capabilities will be developed to provide, at a minimum: 1) intelligent monitoring–for example, the recognition of policy conflicts and temporal conflicts including sequential (“when” conflicts) and spatial (“where” conflicts), that, if not addressed, are potential vulnerabilities, 2) predictive investigations–insights into dependency relationships duplicity, and 3) interdiction recommendations–recognition of actionable situations (behaviors) that could be vulnerable and candidate options to mitigate.

Defining and Developing M&STB Architecture: A challenge will be the development of an architecture that can accommodate the many forms of knowledge (Hayes 2005) that will need to be represented for dynamic and interactive gaming by policymakers to perform usability analysis under different policy conditions and implementation strategies. Research into how best to represent security policies and their interplay with policies that are typically considered outside the realm of security and are known in practice to influence security policy will be included. Another challenge will be providing capabilities that enable scenario-based assessments for the investigation of:

- What is possible from what is not–are there barriers to implementing the policy?
- “What-if” scenarios−what could happen if a policy is implemented in a particular manner in context of other influencing factors and/or policies? How a policy is implemented can have very different outcomes. This environment could provide insight into these outcomes; for example, when different components of a policy are changed and how they might interact with other in-practice human mechanisms that are in place.
- Testing assessment methods and metrics of security policies applied to open science.

Identifying and Development of Methods and Metrics: To claim an entity is secure it is important to understand what secure means in the context of operations and institute a policy that is appropriate to that context. A challenge to this research area will be developing methods and metrics that enables quantitative assessments in an enterprise that appears to need multi-parameter metrics to inform 1) policy development and provide insights about 2) how current science practices and technologies can be changed to accommodate needed security policies of the future. Three critical areas are envisioned. The first addresses usability and its basic tenants (Smith 2007) such as:

- Utility – the resulting product can be used to complete the desired task.
- Goal/Task support – the product is designed to efficiently achieve the end goal of the activity.
- Accommodation – the end product is designed to accommodate diverse user populations and their perspective roles of use.
- Adoption – the end product will be adopted by the expected user audiences.
- Extensibility/adaptability – the end product.

Another area is the representation of parameters that address “policy acceptability.” Tenants for this area will need to be identified and frameworks developed to enable their use by policymakers. A third research area is to enable the measure of “worth value” (Schur and

Hohimer 2006). The concept is that any aspect of system has worth. Determining its value and what and how its value can be changed will be critical to policy assessments.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Three critical challenges must be addressed:

- heterogeneous and often conflicting policies caused by security policies being set by diverse independent agencies
- difficulty in translating security policy into practice that results in configurations that comply with the policy and are consistent within the eScience enterprise
- difficulty in understanding and assessing the implications of new security policies, especially predicting the future outcomes and assessing how well one predicted the future.

As a subset of these key areas consideration must be given to:

- understanding the impact of implementing the policy from a human behavior perspective and exploiting weaknesses as a combination of human and technology vulnerabilities
- gaining insights into what will be needed to coordinate the implementation of policies between the numerous stakeholders in context of different regions−doing science does not take place in a homogeneous environment
- predicting and interacting (what-ifs) with new vulnerabilities that would otherwise go unnoticed and exploring their outcomes
- human-computer interactions: innovative approaches to communicate the simulation results to enable ease-of-use capabilities, such as exploration of the impacts of manipulating different aspects of policies and comparison of outcomes in different contexts
- privacy issues beyond data access and accuracy to address questions as to how data are used and whether citizens are informed about the collection and use of their personal data, as well as about their ability to correct inaccurate data
- developing models that are dynamic and scalable to investigate the anticipated complex environments in real time.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The findings from the assessment of the impacts of new policies will inform the providers of computational capabilities and the consumer about new areas that need innovation to successfully continue the pursuit of science. For example:

- new architecture that accommodates the human element as an inherent part of the eScience infrastructure functionality

- fundamental changes in the way current science practices are performed, which in turn may involve the need to change attitudes about the vulnerabilities of the computational infrastructures that are in daily use
- new computational capabilities that can be transferred into other knowledge work areas
- an eScience enterprise which, over time, becomes as ubiquitous as our telecommunication and power infrastructures are to us today.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

- give security and other policymakers the ability to understand the implications of the policy
- more reliable, effective, and timely transfer and implementation of policy into practice
- better investment of resources.

TIME FRAME

- Deployment of a baseline system can be achieved in 5 years. Within this 5-year time frame, an appropriately structured program should enable a proof-of-principle system to be fielded to a small number of targeted communities.
- Years 3 to 7 are envisioned to be a parallel effort; the development of a robust baseline system that could be deployed in the field to a wide audience of user communities and continued research and development that further the development of the simulation and modeling capabilities.
- Because the needs of science change as well as the methods of how potential adversaries adapt to our protection practices, it is anticipated that this capability will become a dynamic enterprise that is part of the infrastructure to support science.

REFERENCES

Firmino, R. J. 2005. "Planning the Unplannable: How Local Authorities Integrate Urban and ICT Policy Making." *Journal of Urban Technology* 12, 2*,* 49-69.

Hayes, B. 2005. Infrastructure: A Field Guide to the Industrial Landscape. W.W. Norton & Company, New York.

Klein, G. 1999. *Sources of Power: How People Make Decisions*. MIT Press, Boston, Massachusetts.

Norman, D. A., and S. W. Draper. 1986. User Centered System Design: New Perspective on Human-Computer Interaction. CRC Press, Boca Raton, Florida.

Rue, R., S. L. Pleeger, and D. Ortiz. 2007. "A Framework for Classifying and Comparing Models of Cyber Security Investments to Support Policy and Decision Making," in *2007*

Workshop on the Economics of Information Security, Rand Corporation. Also available at http://weis2007.econinfosec.org/papers/76.pdf.

Smith, T. “FErgs,” available at http://en.wikipedia.org/wiki/Usability.

HUMAN FACTORS (HF) ANALYSIS
PRD-3: CHARACTERIZATION OF HUMAN THREATS

ABSTRACT

A better understanding of human attacker skills, behaviors, and motivations can ultimately lead to a more effective ability to detect, predict, and mitigate cyber threats to open science computing environments. The two critical components to investigate in this domain are 1) development of a quantitative understanding of the threat from both external attackers and insiders, and 2) development of techniques, interfaces, and systems to support signature detection and ultimately action against the threat. Both components should focus strongly on the need to reduce data overload and automate data analysis to the greatest extent possible to turn out high-value, actionable information that could ultimately support an operational cyber security function.

EXECUTIVE SUMMARY

This report describes the critical aspects of the *Priority Research Direction in Human Factors Analysis: Characterization of Human Threats*. Research in signature characterization and techniques and systems to enable the detection, prediction, and mitigation of those signatures are vital to the cyber security of our existing and next-generation open science computing environments. This report is devoted to describing this research direction for the next 5 to 10 years. We summarize the essential goals of this research direction as well as outline many of the scientific challenges that lie ahead and the core computational disciplines that require significant progress and investment. Due to the multidisciplinary and multiple component nature of this research direction, we also strongly recommend particular attention to the integration of and "system solution" perspective to research conducted in this area. We conclude with a discussion of the benefits of this work for open science and elucidate the primary area of risk—inappropriate use of personal and research data—and recommend a foundational principle to help mitigate this risk.

SUMMARY OF RESEARCH DIRECTION

This Priority Research Direction (PRD) is composed of two thrust areas: 1) the development of signatures of potentially malicious activities and actors, and 2) the development of techniques, systems, and interfaces that enable the practitioner to identify these signatures in the wild.

Signature development: The ultimate goal of this research component is to understand the cyber threat to open science computing environments so that threatening activities can be monitored, detected, predicted, prevented, and mitigated. The critical first step is to understand the threat and, in particular, the person making the threat. The cyber world provides us with a tremendous amount of data—data that can be monitored and analyzed. The flip side is that the data volumes are so large that we must find smarter ways of improving the information to noise ratio so we can find what we are looking for faster and take action.

Developing threat signatures is a promising scientific pursuit to improve the efficiency and effectiveness of cyber data analysis. There are myriad ways to develop signatures, including: modeling techniques such as pathway analysis, adversarial modeling, and other agent-based modeling techniques; statistical approaches such as classification and outlier detection techniques, unsupervised clustering, supervised learning, and others; and red/blue-teaming exercises, just to name a few. Novel and combination approaches to signature development and validation also are needed, especially dynamic and temporally aware methods and methods that incorporate a variety of data sources. The key to future work in this arena is to reduce the data that must be analyzed, both by human and machine, to enable real-time detection and response.

Technique and system development: Even with signatures in hand, and a thorough understanding of threatening activities, it is challenging to identify them in the wild, and even more difficult to predict future signatures. Improved computational techniques to parse, interpret, reduce, and present high-value information in real-time operational environments are needed to make signatures useful to practitioners. The second step is to develop predictive techniques, especially those that incorporate expert knowledge of attack forensics and attack evolution patterns. Development of such techniques, which could be instantiated in either hardware or software, comprises the second component of this PRD.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The challenges in these two thrust areas are many and draw on numerous computer and computational science disciplines. These challenges are inherently multidisciplinary and require a “system approach” to ultimately result in solutions that can be deployed to support security operations for open science. Although the items are listed below as separate bullets, an overarching scientific challenge will be to effectively integrate the many components of the solution.

Scientific challenges to signature development include:

- Selecting or developing appropriate modeling techniques and methodologies. Because human behavior is notoriously difficult to quantify and predict, models that work with noisy, low-confidence, and incomplete data are required.
- Evaluating and validating models and methodologies is another challenge, particularly since historical data may not be available, is likely outdated, and so sparse that it is anecdotal at best.
- Developing models that can evolve over time—as attack methods and attacker modus operandi change.
- Developing signatures that are not just theoretically sound, but that can be observed in the wild.
- Determining and enforcing an acceptable level of false positives for the open science community.
- Developing models that can be trained or tuned for a large variety of open science compute environments.

- Developing not just signatures of “known” attacks and attackers, but predictive methods to support identification of previously unseen activities.
- Developing novel approaches to characterizing network traffic and computing system activities that are privacy- and security-sensitive to the open science community, their intellectual property, and business requirements.
- Developing novel approaches to identifying multiple-attacker and attacker-in- collusion signatures.

Scientific challenges to system development include:

- Research methods of rapidly processing and transforming network and system data streams. This includes algorithms that could be instantiated in either hardware or software.
- Developing methods and systems that can scalably aggregate and correlate network data and other data sources so signatures can be identified even when they occur over time and within different portions of the computing environment. This is critical to identifying the signature of a savvy adversary.
- Developing network sensor systems with reduced data storage requirements.
- Novel systems and interfaces that enable human analysis of processed network and system data.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Research in the above arenas could advance computational science along many fronts. These include:

- computational statistical modeling
- automated clustering and feature extraction
- automated pattern recognition and analysis
- machine learning and adaptive algorithms
- predictive methods
- scalable data processing
- scalable agent-based systems
- information privacy and security models, including provable security and audit capabilities
- network and system security forensic methods
- storage systems
- human-computer interfaces.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Expanding research in human threats to the cyber component of open science offers everyone in the open science community an improved environment to pursue collaborative science, conduct experiments, and contribute to the larger scientific community and their customers. Improving security operations—particularly operations enabled by scientific solutions—offers more scientific and collaborative freedom that can be less restricted by policy and bureaucracy. Additionally, scientists can pursue their work with greater assurance that their data, systems, and personal information are safe from malicious persons and that these systems and data will have integrity and be highly available,

The greatest risk to incorporating scientifically advanced human factors analysis into operational security for open science is the potential for inappropriate use of personal information, communications, or research data. This is particularly challenging in a physically dispersed and heterogeneous environment. Any scientific pursuits in signature and system development should incorporate appropriate privacy, data security, and data sensitivity protection mechanisms from the beginning. This concept—“intrinsic security” —should be a requirement for scientific research activities in this arena.

TIME FRAME

- Full demonstration of automated attacker signature detection, prediction, and mitigation can be realized within 5 to 10 years.
- Pilot demonstrations and other research components may be available in 3 to 5 years.

Authors: D. W. May and J. Neuss

HUMAN FACTORS (HF) ANALYSIS
PRD-4: FEDERATED IT SECURITY FOR DOE OPEN SCIENCE

ABSTRACT

This Priority Research Direction (PRD) encompasses research and development of federated middleware infrastructure and applications applicable to cyber security in the U.S. Department of Energy (DOE) open science environment. Two substantial benefits will be derived from this PRD: 1) access to DOE open science resources will be simplified for both DOE open science sites and users, and 2) research and development in this area will increase cyber security for both DOE open science sites and users. All told, this should render DOE users more productive, use of DOE resources more effective and efficient, and thereby indirectly promote advances in science and scientific discovery.

EXECUTIVE SUMMARY

This PRD entails research and development of new, usable federated frameworks and mechanisms for implementing, ensuring, and maintaining cyber security in DOE's open science environment. An early framework for federated access exists under the auspices of Internet2's Middleware Initiative (see http://middleware.internet2.edu/), but much more research and development is required to render that framework more usable and robust, and to implement it in DOE's open science environment. Specifically, new and extended frameworks are necessary for federated authentication, authorization, and configuration. Research and development in this area of federated access and cyber security will fundamentally enhance cyber security and increase accessibility and usability of DOE's open science environment. Finally, it is eminently possible to extend this framework to control systems in the utility infrastructure area.

SUMMARY OF RESEARCH DIRECTION

A longstanding ideal in the domain of DOE open science has been easy, secure access to DOE open science resources. Currently, access typically entails different logins, passwords, policies, and interfaces on many different systems. For DOE open science researchers, differences can exist within the environments at their sites, among the environments at DOE open science sites, and at other sites, e.g., federal research proposal submissions. Thus, accessibility is an issue.

Furthermore, in the decentralized open science community, users often exist at locations where cyber security policies are not enforced to the degree they are at DOE open science sites. This lower level of cyber security may exist at U.S. and foreign universities, where accessibility issues are often paramount. This introduces a significant vulnerability into the cyber security environment where cyber security must be enforced end to end. Indeed, consider that an open science user at his or her home may be using a computer shared by his or her family and accessing the Internet over a shared, "sniffable" Internet connection. It is extremely atypical for cyber security policies to be enforced and implemented on computers and networks in users' homes. Currently, the only means of controlling access into DOE open science systems is via login and password, and this falls far short of ensuring end-to-end cyber security. This is a well-known, longstanding issue.

The complexity inherent in this multiplicity of environments presents several problems:

1. Users are forced to deal with different passwords and policies at different sites, impairing easy access to resources.
2. In fact, stories abound about login names and passwords for various user accounts being written on “yellow stickies” and affixed to a user’s computer display. This represents a decidedly cyber security vulnerability.
3. DOE open science sites must provide user support to overcome these issues, with the requisite staffing and resource requirements.

In fact, a framework for accommodating the issues elucidated above is being developed under the auspices of the Internet2 Middleware Initiative. Specifically, this PRD suggests that DOE should become engaged in the ongoing federated middleware research project of Internet2 (http://middleware.internet2.edu/). This would enhance cyber security and increase accessibility and usability in DOE’s open science environment. Currently, the Internet2 Middleware Initiative is focused on federated *authentication*. However, there are plans to extend the effort into the realm of federated *authorization*. Additional research is required to extend that initiative one second step further into a third aspect of *configuration*. These three aspects are further explored below:

- Authentication – Authentication entails associating an individual with credentials that are presented to a system and is typically accomplished through entry of a username and a password that are supposedly private and uniquely associated with an individual. Typically, login name and password are entered at the DOE open science site when a user accesses system resources. However, authentication in the federated mode is accomplished by a new federated application running both at the user’s site and at the DOE open science site. Upon a user’s request to access resources, the application running at the DOE open science site communicates with the user’s home domain, e.g., on the user’s campus or other home site, where authentication is performed. Then, after the user authenticates in the home domain, the federated application at the user’s home site securely transmits only the minimum credentials required to the remote site (here, the DOE open science site). This has several advantages: 1) usability – the user need remember only one login and password, 2) privacy – only the minimum amount of personal information is transmitted outside the user’s home domain, 3) cyber security – only the minimum amount of sensitive information, e.g., no login and password information is transmitted outside the user’s home domain, and the transmission is over a secure link. Typically, the credentials transmitted to the remote site from the home site contain a quantitative Level of Assurance (LoA) that is indicative of the level of confidence that the user’s home site has that the user is actually the individual in question. LoA’s are determined based upon the cyber security policies, practices, and systems in place at the user’s home site.

- Authorization – Authorization permits access to resources. Typically, authorization is binary–"yes" or "no"–based upon the nature of the resources being accessed and the LoA returned in the user's credentials from the user's home domain. While efforts are well under way for authentication, there are significant research and development needs in the area of authorization, including the taxonomy of LoAs mapping onto DOE open science resources, and the infrastructure and application needs in the DOE open science environment necessary to support this model. For example, a flexible LoA framework for authorization may involve discovery of the nature and location of the system and network the user is on. Access across Internet2 and DOE's Energy Science network from a well-run university environment may result in a higher LoA than access from a home computer via the public internet. This also should be an active area of research.
- Configuration – Configuration entails a negotiation between the DOE open science site and a user's environment, e.g., the user's computer and/or network, first to discover and then to change system and/or network cyber security parameters from the current settings, upward to the minimum settings required to access DOE open science resources. Although possible under the federated structure, this is a rich area for research and development.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Many DOE open science users exist in highly decentralized cyber environments that are often not secured to the degree of some DOE open science environments. Login and password combinations, even using actual two-factor authentication, do not address end-to-end cyber security. Implementing a framework in such a disparate, decentralized environment is an extremely daunting proposition. Among the numerous challenges, the following may be the most significant:

- A significant number of sites must implement the cyber security federated framework for it to become a de facto standard that DOE adopts and requires for access. That the Internet2 Middleware effort is about 5 years old and only a small number of sites are still experimenting with deploying this framework illustrates the complexity and daunting nature of this proposition. DOE has significant resources that scientists worldwide must access to conduct their work, making it an ideal candidate to participate in this research and development activity.
- Each participating site must implement a technological framework that performs the negotiation, including authentication and authorization. Consult the Internet2 Shibboleth project for additional details of these requirements (http://shibboleth.internet2.edu/). Many sites have attempted to implement this framework, and the experience has been that it is not for the "faint of heart." Additional development is needed to render the system easier to deploy.
- In addition to a technological framework, a logical/policy cyber security framework must be implemented at all sites. Consult the InCommon project for additional details of these requirements (http://www.incommonfederation.org/). This requires each institution to define

and document its policies and procedures (practice) for cyber security and culminates in an external review of an institution's policies and procedures for cyber security. Although many institutions have cyber security policies documented, not nearly as many have cyber security practices documented. Such documentation is reviewed by an external auditing agency, and LoAs are then established for all conditions of access. A relatively small number of institutions have undergone this process, and there are yet many lessons to be learned. Again, the level of attractiveness of DOE open science resources and needs to secure them makes DOE an ideal participant in this effort.

- This framework also must be implemented at DOE open science sites, on DOE open Science systems, in DOE open science environments. How this framework maps technologically onto DOE open science environments is yet to be determined.
- The issues associated with mobile and home uses have not yet been fully explored. Extending policies into these environments is possible, but extending information technology support for cyber security into the home environment is generally not practicable at present.
- Human factors associated with the balance between usability and accessibility are yet to be determined in this federated environment. This framework must gain widespread acceptance among sites and among users for it to be successful. DOE could add tremendous impetus to this activity.
- A framework for including the cyber security aspects of the system(s) and network(s) the user employs to access DOE open science systems does not even exist. Indeed, one of the recommendations that emerged from various sessions in the workshop is the development of a quantitative metric understandable by an unsophisticated user for the cyber security health of a system (that might appear just like a CPU or I/O active icon under Windows Vista). This metric could be "stirred into" a dynamic determination of an LoA for a specific user employing a specific infrastructure for accessing DOE open science systems. Both of these aspects are open areas ripe for research and development.
- In addition to determining an appropriate LoA, systems might be configured to respond to requests from or requirements by DOE open science environments for access. For example, if the current LoA is insufficient to access the DOE open science resource, the cyber security application that configures the user's remote system may choose to negotiate that encryption to the DOE open science environment and be implemented as a remedial strategy. A new LoA could then be computed, and if above the threshold required for access to the DOE open science system, then access could be granted. This, too, is an area ripe for research and development.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The potential impact upon computational science is primarily in the following thematic areas:

- Because users will log into DOE open science systems using their "home" authentication and authorization infrastructure, DOE open science systems will be more accessible and usable.

- Because users will have a simplified manner of accessing systems, access will no longer be a barrier for users. Therefore, users should access more systems and be more productive.
- An ancillary benefit of this research is that the remote sites will do most of the "heavy lifting" for implementing an enhanced level of cyber security, removing this burden from or at least reducing this burden on DOE open science sites. Thus, DOE open science sites should be able to redirect the resources that otherwise would have been expended on this activity into other, more productive areas.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

As a result of this research, there should be significant, direct benefits upon DOE's open science cyber security environment. In addition, users' cyber security environments should be more secure. Specific benefits that should accrue from this research are:

- DOE's open science environment should become more secure as a result of this research. Users will have to remember only a single username and password set to access systems. This should significantly reduce the number of "yellow stickies" affixed to surfaces in users' offices.
- The entire cyber security environment, end-to-end, from the DOE open Science to the end user, should become more secure.
- The general cyber security effort internationally will benefit from the development of the cyber security meter.

TIME FRAME

There are several time frames for this research that are pertinent to the various aspects discussed above:

- It is anticipated that it will require about 3 years to develop and test the federated cyber security authentication framework, including the development to begin to implement it in DOE's open science environment.
- It is anticipated that it will require about two additional years, or 5 years total, to implement and test federated authorization in DOE's open science environment.
- Finally, it is anticipated to require up to 5 additional years, or 10 years total, to develop federated configuration, as this will require additional elements to be developed, tested, and implemented, including the cyber security meters for various systems.

REFERENCES

All materials referenced for this PRD are online:

The federated IT environment, including logical and policy elements with Levels of Assurance (LoAs), http://www.incommonfederation.org/

Internet2's Middleware Effort, http://middleware.internet2.edu/.

Internet2's Shibboleth project, providing the technical framework for federated authentication and authorization, http://shibboleth.internet2.edu/

HUMAN FACTORS ANALYSIS
PRD-6: NON-CRYPTOGRAPHIC SECURITY

INTRODUCTION

Many technical cyber security measures are built upon cryptographic underpinnings. Well-known examples of this include:

- password-based login mechanisms
- checksum-based file integrity (using MD5, etc.)
- protection of data transmissions from eavesdropping (such as SSH, SSL, etc.)
- message signing and encryption (using S/MIME, PGP/GNU Privacy Guard, etc.)
- email anti-spoofing techniques (such as DK/DKIM)
- digital rights management for controlling access to intellectual property,
- DNSSEC,
- s*BGP, etc.

The pervasiveness of cryptology as a foundation security technology means that if cryptographic approaches are successfully attacked, the community will find itself in a very difficult position.

Cryptographic approaches are under ongoing and overlapping attacks from a number of directions, including:

- malware-based attacks on endpoint security via hostile programs capable of capturing passwords entered by users or stored on workstations (see, for example, password stealing Trojans mentioned at http://www.viruslist.com/en/virusesdescribed?chapter=153317860)
- ever-increasing raw computational horsepower, horsepower which can be used to brute force cryptographic systems (including decentralized collaborative approaches to tackling cryptographic problems, such as distributed.net; low-cost, field-programmable gate arrays, as mentioned at http://www.cl.cam.ac.uk/~rnc1/descrack/DEScracker.html, etc.)
- approaches to cryptography, which trade storage for computational power, such as Rainbow Table approaches to password cracking (see http://rainbowtables.shmoo.com/)
- various ongoing analyses of potential algorithmic weaknesses, or their implementation.

Given those threats, investigating non-cryptographic alternatives to cyber security seems prudent, both to provide defense-in-depth (in conjunction with existing cryptographic approaches), and as a standalone alternative, in the event cryptographic approaches become unusable or untrustworthy.

What Is An Example of Potential Non-Cryptographic Approaches?

A number of breakout group participants wanted to avoid unduly influencing potential investigators through providing a laundry list of examples, but there was a perceived need to provide at least one example of what's meant by "non-cryptographic" approaches to security.

An example of a non-cryptographic approach to cyber security would be replacement of shared packet-switched network connections and shared multi-user systems with dedicated circuit-switched networks and dedicated physically isolated system endpoints.

That alternative is meant merely to serve as an illustration of a potential approach, but we expect that numerous other alternatives also will be identified over the course of a multiyear research program, and the identification and evaluation of those alternatives would be an integral part of that work.

IMPACT ON OPEN SCIENCE

It is also worth noting that non-cryptographic methods may have important benefits that extend beyond just security. For example, many eScience investigators need to routinely move large datasets across the country or overseas. Normally cryptography guarantees that transmitted data will arrive unaltered and without being eavesdropped upon, but it can be hard to support encrypted transmission of multi-gigabit-per-second flows without specialized hardware encryptors normally unavailable to members of the open science community.

Users thus can face a difficult choice: transfer their data slowly (but securely, and with assurance that it arrives intact and is not eavesdropped upon), or transfer their data rapidly, but with little or no protection.

Identification of non-cryptographic approaches to ensuring system and network security will serve to eliminate that cryptographic bottleneck, although that is not the primary motivation underlying this program of work.

TIME FRAME

- Because this topic is truly asking for out-of-the-box thought, and a genuine paradigm shift, this topic will likely require 3 to 5 years for the identification and development of alternative approaches.
- Additional years will be required for production deployment across the Internet community.

Author: J. E. St. Sauver, Ph.D.

HUMAN FACTORS (HF) ANALYSIS
PRD-7: APPROPRIATE DISTRIBUTED DEFENSE

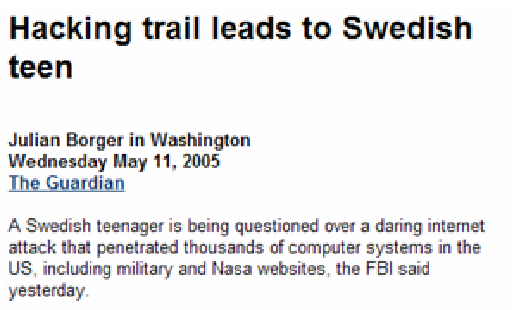
ABSTRACT

Many people view anomaly detection and automated defense as standalone activities to be triggered by information collected just from a targeted machine. A more effective way to determine the nature of a cyber attack as well as to better characterize the cyber attacker (individual hacker, cyber criminal, nation state, etc.) would be to use information available on the Internet and information available from allied networks, in addition to information collected from a targeted machine. Understanding the nature of a cyber attack and being able to better characterize the cyber-attacker could help take immediate, automated protective action as well as lay the groundwork for a more complete and systematic analysis of the attack that can then be used to minimize the impact of such attacks in the future.

EXECUTIVE SUMMARY

This Priority Research Direction (PRD) seeks to create a unified system that draws on information from a targeted machine, allied networks, and Internet archives and databases to automatically analyze the nature of a cyber attack and to help classify the cyber attacker into several categories (individual, criminal organization, terrorist organization, nation state, etc.). This research would need to determine methods for usefully querying disparate databases and archives on the Internet automatically, determine how to use allied networks in a manner consistent with their privacy and security policies, and integrate all of this external information usefully with information that can be collected from the targeted machine.

Security is often thought of as a standalone enterprise, but we believe that it can be significantly enhanced by drawing on information from sources external to the machine or network being attacked. These sources include databases and archives on the Internet such as blacklists, lists of spam sites, lists of phishing sites, archives of exploits, as well as information from allied networks. For example, many institutions have multiple networks; often cyber attacks will target more than one of these networks. Pooling information across allied networks can enable spotting of suspicious activity and alerting members of the network to take appropriate action. The Stakkato Intrusions (analyzed by Nixon (2006)) show that cyber attackers exploit networks to compromise individual machines, so it seems reasonable to use information gathered from networks to help those networks themselves. The following picture shows part of an article describing the attack.



Hacking trail leads to Swedish teen

Julian Borger in Washington
Wednesday May 11, 2005
The Guardian

A Swedish teenager is being questioned over a daring internet attack that penetrated thousands of computer systems in the US, including military and Nasa websites, the FBI said yesterday.

Cyber attacks proceed at computer speeds, so it is difficult for humans to respond appropriately. Quickly identifying an attack can permit quick automated action, such as protectively disconnecting a computer from the Internet for a short period of time, and alerting system operators of the suspicious activity so they can take appropriate action. The system also can provide operators with a characterization of the cyber attacker, which will help the system operators to better understand the nature of the threat. Cyber attackers range from curious individuals to nation states, and having an indication of who is attacking would be of great help in understanding the significance of the attack and taking appropriate action.

SUMMARY OF RESEARCH DIRECTION

A broad-based research effort will be needed to include research in human factors, database theory, pattern analysis, operating systems, network systems, and algorithms.

One thread of the research would require characterizing Internet resources and developing techniques for querying them automatically in a useful manner. The amount of information available about viruses, worms, malware, exploits, bad sites, etc., is staggering and figuring out how to use this information effectively would be of great help. Many important issues that need to be addressed to maximize the usefulness of available information. Some information is easy to use because it is commonly reported, such as lists of suspicious intellectual property addresses. On the other hand, other information is presented in a way that makes it difficult to use in an automated manner. For example, how can one use descriptions of exploits to determine automatically whether a suspicious incident is making use of a particular exploit?

Aside from technical issues, organizational and human factors issues also must be considered in using Internet resources. Methods for determining levels of trust of external resources need to be developed to be able to understand the trustworthiness of sites from the point of view of accuracy and whether they can be trusted to not leak information about our querying of them.

There are technical and policy issues that need to be resolved in using allied networks to help determine the nature of an attack and to characterize the attacker. Clearly, questions abound about what data to collect and how to collect it without seriously compromising the performance of the networks involved. We need to determine the packet patterns that characterize common activity across networks. Such knowledge might be helpful not only for dealing with cyber attacks, but might also help to reduce spam and other nuisance network traffic. Another issue is to develop operating procedures that do not violate confidentiality and privacy policies in force on the various networks. To better gauge the scale of an attack, statistical sampling techniques may need to be developed for use in case of a large number of allied networks participating.

Techniques also will have to be developed to measure the health of an attacked system. We would need to understand what constitutes normal and abnormal activities on a system. Of special interest for the Open Science Initiative is to understand the security needs and characteristics of cluster computers. The papers by Florez et al. (2004) and Lee et al. (2005) describe interesting work in this area that could be applied to this problem. To minimize performance penalties, we need to better understand how to use operating systems more effectively. There are also techniques such as writing loadable kernel modules that can permit us to introduce monitoring into operating systems without seriously impacting the performance of the system.

An effort also must be identified to attack patterns that characterize different types of attackers. In particular, it is critical that we distinguish among individual hackers, criminal organizations, terrorist organizations, industrial spies, and nation states. Understanding the differences can provide insight into which assets need to be protected and what might be the most effective way to protect them.

A hierarchy of possible defenses and appropriate triggers also needs to be developed. For many attacks, automatically taking a system offline for a short period of time can prevent the attack from being successful without seriously interfering with the tasks being worked on. More work needs to be done on understanding this and other approaches, and knowing when to use different defenses. Appropriate notification needs to be given to users and system operators so that they can take further, appropriate actions.

Of course, users in the Open Science Initiative are extremely interested in maximizing the number of cycles they can use. For this reason, we will need to develop algorithms that can defend the system while using as few cycles as possible.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

There are many interesting scientific and computational challenges. They are as follows:

1. We must be able to bring together, in real time, a myriad of disparate and possibly contradictory sources to make an informed decision about events happening on a system.
2. We must integrate this information with information coming from allied networks in a way that does not compromise security or privacy.
3. We must better understand how to gauge the health of systems and how to correlate this information with information gathered from external sources.
4. We must understand how to determine the nature of the attacker (individual, criminal group, terrorist group, nation state, etc.) from the information that we are able to gather.

5. We must determine the sorts of protective actions that can be taken automatically, and which ones require human intervention.
6. We must do all of this so efficiently that it does not detract from the computations that members of the Open Science Initiative are carrying out.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

We believe that the creation of this system will be of great interest to many people concerned with cyber security. While Open Science Initiative systems are tempting targets for cyber attackers, many common elements exist between the cyber threats faced by the open science community and other computer user communities.

Insight into the important human factors issues of getting different groups to work together to provide a common defense also will be of great interest to other workers in the cyber security area.

The ability to do this analysis in real time also would be of great interest to cyber security researchers worldwide.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

The majority of researchers in the Open Science Initiative will have cyber security as their primary interest. By and large they have important work to carry out and want to do so as easily as possible with the fewest distractions. Primarily, they will need to use the Internet and various networks to accomplish their tasks. Without a good security system that can provide protection to the Open Science Initiative assets and to the researchers' systems, the initiative might be much less effective than it could be. We believe that the proposed system will enable the users of the Open Science Initiative to get the most from its assets and their efforts.

TIME FRAME

- Deployment of this system can begin immediately and a rudimentary capability can be achieved in about 3 years.
- Implementing a very robust and comprehensive system can be done in the 5-10-year time frame.
- Because attackers never stop developing new attacks, we expect that the system will continue developing for the indefinite future.

REFERENCES

Florez, G., Z. Liu, S. Bridges, R. Vaughn, and A. Skjellum. 2004. "Detecting Anomalies in High-Performance Parallel Programs," in *ITCC 2004*: *International Conference on Information Technology*: *Coding and Computing*, ed. P. K. Srimani, IEEE Computer Society, Los alamitos, California. Held April 5-7 in Las Vegas, Nevada. **2**, 30-34. Available at http://ieeexplore.ieee.org/iel5/9035/28683/01286585.pdf?isnumber=28683&prod=CNF&arnumber=1286585&arSt=+30&ared=+34+Vol.2&arAuthor=Florez%2C+G.%3B+Liu%2C+Z.%3B+Bridges%2C+S.%3B+Vaughn%2C+R.%3B+Skjellum%2C+A.

Lee, A. J., G. A. Koenig, X. Meng, and W. Yurcik. 2005. "Searching for Open Windows and Unlocked Doors: Port Scanning in Large-Scale Commodity Clusters." *2005 IEEE International Symposium on Cluster Computing and the Grid, CCGrid 2005*, 146-151.

Nixon, L. 2006. "The Stakkato Intrusions: What Happened and What Have We Learned?" in 6th *IEEE International Symposium on Cluster Computing and the Grid*, *CCGRID 06*, May 16-19 2006, Singapore. Also available at http://www.nsc.liu.se/~nixon/stakkato.pdf.

BREAKOUT SESSION
PROTECTING OUR UTILITY INFRASTRUCTURE

Breakout Leads: *Jeff Dagle, Pacific Northwest National Laboratory;*
Aaron Turner, Idaho National Laboratory;
Bill Young, Sandia National Laboratory

PROTECTING OUR UTILITY INFRASTRUCTURE
PRD-1: SURVIVABLE AND TRUSTWORTHY CONTROL SYSTEMS

ABSTRACT

The integration of computing systems into the utility infrastructure of the United States provides measurable benefits to infrastructure owners but also introduces significant vulnerabilities. Especially in cases where commercially developed computing systems are deployed to serve in critical roles, those vulnerabilities can be exploited to the detriment of our country. To solve security problems in existing infrastructure control systems, new control system architectures and components need to be developed that would be survivable and trustworthy.

EXECUTIVE SUMMARY

The development of survivable and trustworthy systems will require a concerted effort among system vendors and end-users to collaborate and revolutionize the way technology is used to deliver the services that our society relies on for enjoyment of our present quality of life.
A collaborative process will need to be developed that addresses system vulnerabilities on a lifecycle level, building survivability into the architectures and components−not attempting to add on security after the product is developed. To achieve this revolutionary approach to using technology to manage utility infrastructures, research needs to be conducted to demonstrate revolutionary uses of technology in the delivery of critical services.

SUMMARY OF RESEARCH DIRECTION

The main research directions necessary for survivable and trustworthy control systems fall into the following main areas:

- template architectures specific to the control system domain
    a. models to facilitate reasoning about survivability
    b. improved system support for human intervention
- designs for failure proportionality and graceful degradation (failsafe)
    a. resistant to both malicious attacks and accidental failures
    b. enable tradeoffs between safety and performance
    c. accommodate different reliability requirements
- real-time anomaly detection/prevention/response to increase survivability
- architectures that are sufficiently understandable that operators place trust in them.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

The research effort must address the following challenges. While some work in these areas is already under way, unique aspects of future control system architectures and technologies demands research that anticipates, to the extent possible, the future of process control systems.

- monitoring, detection, and response with the timeliness and scalability needed for the control system domain
- large scale, geographically dispersed, authentication and authorization spanning all the people and devices in the larger control system
- survivable and trustworthy control systems designed to preserve these attributes under composition
- extending fault-tolerance to tolerate malicious and accidental faults in the control system domain.

While some work in these areas is under way, unique aspects of future control system architectures and technologies demands research that anticipates, to the extent possible, the future of process control systems.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

Infrastructure control system cyber security poses significant risks–control system compromises can cause significant loss of life and resources. Increasing survivability is a cross-cutting approach to addressing this risk. Research into intentionally robust architectures, novel approaches to anomaly detection, and designs that gracefully degrade are critical to increasing the survivability of control systems and reducing this security risk. Improved survivability is required for the evolving business demands facing process control owners and to promote the effective application of scientific rigor to the process control system environment.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Increasing survivability of control systems is a significant need for open science because of the desire to share capital-intensive resources (ranging from particle accelerators to high-powered microscopes). Strong system survivability is a prerequisite for widespread sharing of such resources.

TIME FRAME

- Applying research to solve security problems in existing infrastructure control systems by developing new survivable and trustworthy control system architectures and components will take 5-10 years.

PROTECTING OUR UTILITY INFRASTRUCTURE
PRD-2: UNDERSTANDING RISK AND SURVIVABILITY ASSESSMENT

ABSTRACT

Despite the dependency that infrastructure owners have on technology, few organizations truly understand their risk profile as it relates to system vulnerabilities that could adversely impact their operations and customers. Using established risk models has resulted in poor mitigation decisions that have not addressed the root cause of the system vulnerabilities. For this reason, it is important that we take a new look at how to categorize, measure, track, and respond to risks that affect the utility infrastructure.

EXECUTIVE SUMMARY

As technology allows greater efficiencies for infrastructure owners, it also enables greater efficiencies for those who wish to do harm to the system or system dependents. New approaches need to be developed that allow for more efficient response to control system risk, accurately measure the benefit of the response, and then track the long-term effectiveness of the response as it relates to protecting utility infrastructure.

SUMMARY OF RESEARCH DIRECTION

This research is essential to establish meaningful risk analysis and survivability assessments within in the control system environment. Multiple factors, both physical and cyber, must be accounted for in such analysis. The six items listed below represent the critical research areas that need exploration for success:

- Investigate how to effectively gather, log, collect, and share anonymized threat data.
- Develop a comprehensive operational model of control systems and evaluate robustness against security events.
- Develop a high-level, usable, real-time assessment and consequence analysis of security threats.
- Develop a security investment analysis capability appropriate for process control systems.
- Design high-performance algorithmic techniques for simulation and analysis.
- Quantify tradeoffs among survivability metrics.

SCIENTIFIC AND COMPUTATIONAL CHALLENGES

Success in the research directions listed above requires meeting the following challenges.

- how to measure and analyze the robustness of control systems with respect to cyber threats and physical threats to cyber assets

- how to identify and prioritize cost-effective means to improve security of control systems
- how to quantify and predict the response to operator actions and defensive mechanisms
- how to enforce security in the face of real-time requirements and heterogeneity of systems, devices, technologies, and emerging threats, while maintaining business productivity.

POTENTIAL COMPUTATIONAL SCIENCE IMPACT

The methods, techniques, and algorithms could apply to the risk and survivability analysis of computational environment in open science, which may rely on similar information system architectures. From the control-system-specific research, the major specific technical value would be a methodology to:

- Evaluate alternatives and configuration choices leading to cost-effective deployment of security solutions.
- Quantify assurance of control system survivability through awareness and operational efficiencies.
- Establish the baseline for control system security compliance.

Adaptation to the computational environment would be required, but at least some of the research will likely benefit both environments.

POTENTIAL IMPACT ON CYBER SECURITY FOR OPEN SCIENCE

Understanding risk and survivability of control systems is a significant need for open science because of the desire to share capital-intensive resources (ranging from particle accelerators to high-powered microscopes). Strong system survivability awareness is a foundation for widespread sharing of such resources.

TIME FRAME

Meeting the objectives of this PRD, below, will take from 5-10 years:

- Develop new approaches that allow for more efficient response to control system risk.
- Accurately measure the benefit of the response.
- Track the long-term effectiveness of the response as it relates to protecting utility infrastructure.