Four Grand Challenges in
TRUSTWORTHY COMPUTING

Second in a Series of Conferences on Grand Research Challenges in Computer Science and Engineering
November 16–19, 2003

Computing
Research
Association

Organizing Committee Members

Eugene H. Spafford, Purdue University (Committee Chair)
Richard A. DeMillo, Georgia Institute of Technology (Committee Co-Chair)

Andrew Bernat, Computing Research Association
Steve Crocker, Shinkuro, Inc.
David Farber, Carnegie Mellon University
Virgil Gligor, University of Maryland
Sy Goodman, Georgia Institute of Technology
Anita Jones, University of Virginia
Susan Landau, Sun Microsystems Laboratories
Peter Neumann, SRI
David Patterson, University of California, Berkeley
Fred Schneider, Cornell University
Douglas Tygar, University of California, Berkeley
William Wulf, National Academy of Engineering and University of Virginia

Acknowledgments

Funding for this conference was provided by National Science Foundation Grant No. CCR-033524.

For Additional Information

See: http://www.cra.org/Activities/grand.challenges/security/home.html

Contents

1. Introduction

In 2002, the Computing Research Association (CRA) sponsored its first "Grand Research Challenges in Computer Science and Engineering" conference. This was the first in a series of highly non-traditional conferences to define important questions rather than expose current research. Grand Challenge meetings seek "out-of-the-box" thinking to arrive at exciting, deep challenges yet to be met in computing research.

Because of the importance of information security and assurance, CRA's second Grand Challenges Conference was devoted to defining technical and social challenges in trustworthy computing.

Nearly fifty technology and policy experts in security, privacy and networking (see Appendix) met November 16-19, 2003, at Airlie House in Northern Virginia in a Gordon-style research conference under the sponsorship of CRA and the National Science Foundation (NSF). This report describes Four Grand Challenges in trustworthy computing identified by the conference participants, why these challenges were selected, why progress may be possible in each area, and the potential barriers in addressing them.

Trustworthy Computing

Information technologies and the computing base that enables them are pervasive. From the desktops of business and home users to the massive, distributed data centers that regulate commerce and control critical infrastructure, the world has come to depend on the availability of information and communications technology.

This infrastructure grows more complex as the underlying computational and communications capabilities double in speed and capacity every eighteen months, inexorably following a law articulated by Intel co-founder Gordon Moore.

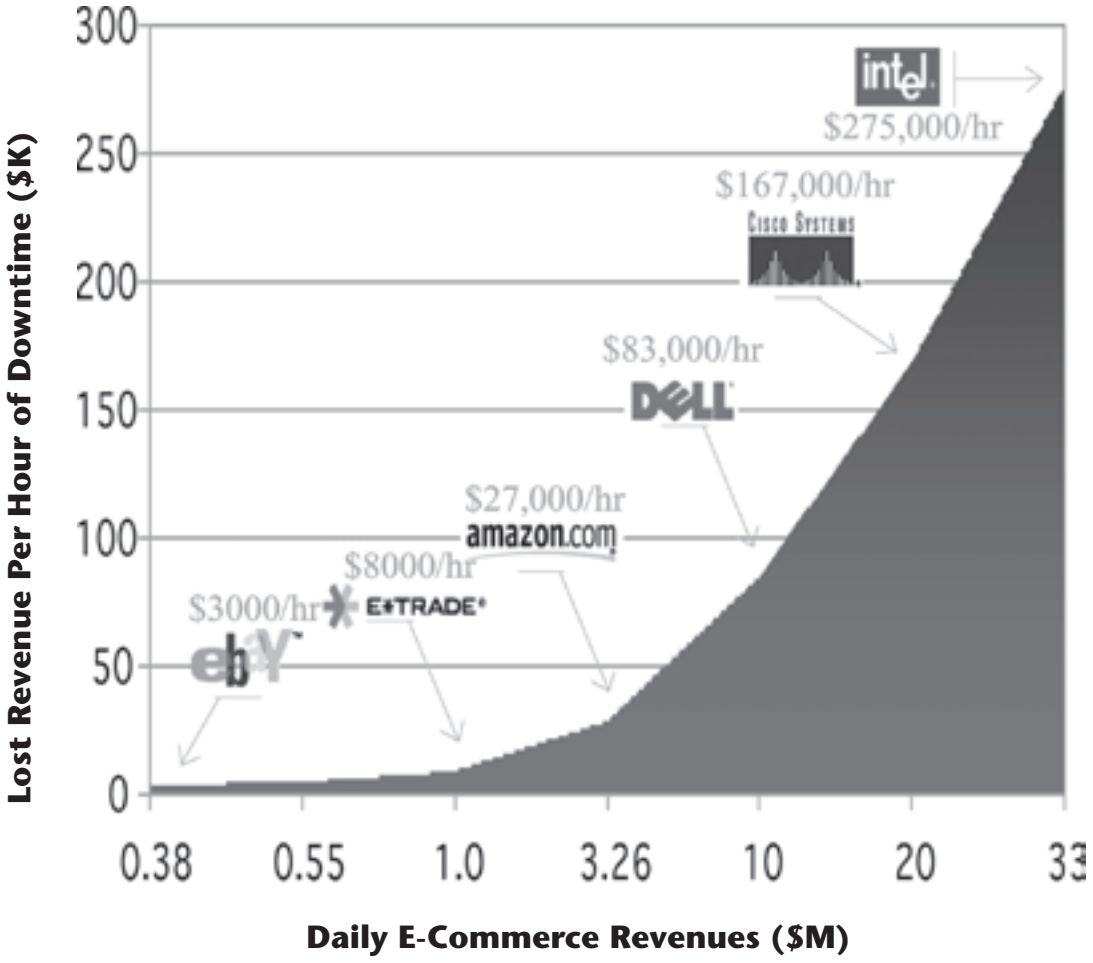
Every vision of the future includes predictions of ubiquitous computing and networking. Distributed, embedded and increasingly portable systems in every aspect of life will continue to transform the way we do business, govern and defend ourselves, maintain our health, communicate, keep records, control our environment, educate our children, create new knowledge and entertain ourselves.

The pace of change is accelerating and so is our dependence on information technology (see Box 1). We currently face threats of massive disruption: outages in power, transportation and communications systems resulting from denial of service; loss of privacy; alteration of critical data; and new forms of theft and fraud on an unprecedented scale. Threats from criminals, anarchists, extremists, cyber terrorists and random attackers continue to grow even as we increase our reliance on computing infrastructure. Attacks on information technology infrastructure undermine security and, ultimately, trust. A trustworthy computing infrastructure should be immune from such attacks.

Box 1. Pace of Change and Dependence on IT Accelerating

By 2000, world spending on IT equipment and services had surged to over $2 trillion and represented nearly 10% of the US gross domestic product.[1]

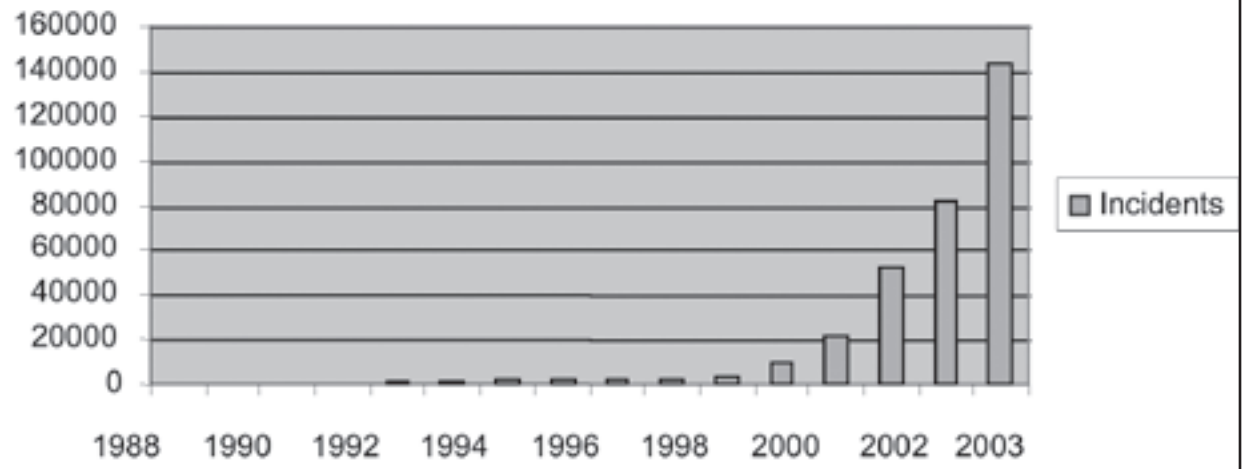
The growth of e-commerce had already resulted in a massive shift to reliance on the continuous availability of IT systems. In a 1999 study drawn from corporate 10Q reports, Forrester Research developed estimates of lost revenue due to downtime in Internet-based systems



[1] http://www.witsa.org/press/dp2000pr.pdf

The Computer Emergency Response Team Coordination Center (CERT/CC) at Carnegie Mellon University has been tracking these attacks since its inception in 1988, reporting them as "security incidents." Like Moore's Law, the CERT/CC statistics (see Figure 1 below) show exponential growth: the number of security incidents has been doubling every year for the past fifteen years. In 1988 there were six reported incidents; in 2003, more than 143,000. Even more significantly, the number of distinct vulnerabilities is also growing exponentially. In 2002, more than 4,000 distinct vulnerabilities were reported. It is important to note that the number of security incidents that occur is undoubtedly much higher than the number actually reported.

Figure 1. Security Incidents



Several factors fuel this explosion. First, information technology is reactive when it comes to information security. We have constructed a quilt of patched fabric in response to newly identified vulnerabilities and threats. Moreover, no patch carries with it a guarantee of future security. It is an indication of how pervasive the problem has become that the word "patch" has acquired a technical meaning in the industry. Software developers issue modifications to their products in the form of small program "patches" that customers can download and install on their computers. Sometimes patches are used to add features to or correct defects in software products but, with increasing frequency, patches are used to correct security flaws and vulnerabilities that can lead to system attack and compromise. Microsoft, the world's largest software vendor, recently changed the frequency of security-related patch releases to its ubiquitous Windows™ operating systems from weekly to monthly because it found that system administrators could not keep up with the sheer volume of releases.[2]

[2] "Ballmer Addresses Security," Michael Cherry, Directions on Microsoft, posted Oct. 27, 2003. http://www.directionsonmicrosoft.com/sample/DOMIS/update/2003/12dec/1203bas.htm

Current methods of patching our way to security strain credibility. We simply cannot depend on hundreds of millions of users to individually and continually upgrade their hardware and software. Some attackers even wait for a patch to be announced to discover flaws in systems. They then launch attacks using those flaws because they know that many computers will not be patched and will thus be vulnerable for months to come.[3]

Second, the world has become much more dangerous. The IT industry has not kept pace with worms, viruses, Trojan horses, spam, and denial-of-service attacks that have increasingly threatened e-commerce and Internet communications. Even more troubling, however, is the rapid emergence of civilian and military groups—at home and abroad—with the resources and resolve to mount serious attacks on critical infrastructure.

Third, computing technology has been turned against us. Attackers have been able to exploit Moore's Law by automating attacks and flooding networks to deny critical information and services to legitimate users. Even modestly well-equipped attackers can easily harness massive computing resources, and inexpensive network access makes it easy for worldwide networks of enemies to mount coordinated attacks.

Lastly, asymmetric warfare has arrived in cyberspace. In the Internet age, open interfaces connect users, data, business processes, commodity hardware and software in dynamic webs of trust and dependence. Organizational boundaries are essentially meaningless, so the idea of a defendable perimeter for computer security has become meaningless as well. The enemy to defend against may well be a trusted employee acting alone—a trusted "insider"—and not an identifiable external force mounting an attack.

In this environment there are clear challenges to be met. In the near term, the IT industry will certainly continue to arm itself against the growing threat. This strategy will not be effective for very long. The attackers are gaining ground daily. Rather, we look toward the research community to innovate along four dimensions, the grand challenges for trustworthy computing:

1. Develop new approaches for eradicating widespread, epidemic attacks in cyberspace.
2. Ensure that new, critical systems currently on the drawing board are immune from destructive attack.
3. Provide tools to decision-makers in government and industry to guide future investment in information security.
4. Design new computing systems so that the security and privacy aspects of those systems are understandable and controllable by the average user.

[3] http://asia.cnet.com/newstech/security/0,39001150,39156953,00.htm

Why Are Grand Challenges Needed?

Many prior academic, government, and industry studies have pointed with alarm to these and similar threat trends (see Box 2). It has been articulated clearly in each of these studies that the IT industry needs to apply the best and most mature trust-enhancing technologies to stave off impending disaster.

We endorse these recommendations, but such an agenda focuses money, energy and attention on incremental improvements and updates to existing systems, rather than seeking fundamental advances. CERT/CC statistics document an exponentially expanding threat, and there are no explanations for how an incremental approach to trust can succeed in such an environment.

The goal of the CRA Grand Research Challenges conferences is to encourage thinking beyond incremental improvements. Some important problems simply cannot be solved by narrow investigation aimed at short-term payoffs. Multiple approaches, carried out over a long period of time, will be required. The community is, in effect, looking for big advances that require vision and cannot be achieved by small evolutionary steps. The February 2005 report by the President's Information Technology Advisory Committee (PITAC) supported a long-term view of research by agencies such as DARPA and NSA, arguing that the trends "favoring short-term research over long-term research. . . should concern policymakers because they threaten to constrict the pipeline of fundamental cyber security research that. . .is vital to securing the Nation's IT infrastructure."[4]

Breaking the CERT/CC Curves

The overriding challenge for information security research is to fundamentally change the CERT/CC incident and vulnerability curves. In effect, we challenge the community to neutralize the threat by seeking fundamental changes in the underlying technology. With fundamental advances, it is possible that computing infrastructure will become *more* trustworthy over time—in effect, the number of incidents and vulnerabilities as a function of time could be fixed or even decreasing.

Such advances would change the ground rules for attackers. At the moment, threats are riding the same technology curve as the rest of us. There is an immense economic and technological engine that fuels Internet growth and Moore's Law, and neutralizing the threats would effectively decouple attackers from that engine.

[4] "Cyber Security: A Crisis of Prioritization," President's Information Technology Advisory Committee, 2005, p. 23. http://www.nitrd.gov/pitac/reports/20050301_cybersecurity/cybersecurity.pdf

Box 2. Major Studies and Reports in the Past Twenty Years

President's Information Technology Advisory Committee. 2005. "Cyber Security: A Crisis of Prioritization." http://www.nitrd.gov/pitac/reports/20050301_cybersecurity/cybersecurity.pdf

Keeney, Michelle. 2005. "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors." National Threat Assessment Center, United States Secret Service. http://www.cert.org/archive/pdf/insidercross051105.pdf

Schneider, Fred B. 1999. "Trust in Cyberspace," National Research Council.

White, Steve R. 1998. "Open Problems in Computer Virus Research." IBM Thomas J. Watson Research Center. Yorktown Heights, NY USA. htttp://www.research.ibm.com/antivirus/SciPapers/White/Problems/Problems.html

National Institute of Standards and Technology. 1998. "Guide for Developing Security Plans for Information Technology Systems." NIST SPEC PUB 800-18.

National Computer Security Association. 1997. "NCSA 1997 Computer Virus Prevalence Survey." http://www.delta-logic.fr/docs/ncsa97.pdf

Government Accountability Office. 1996. "Information Security: Computer Attacks at Department of Defense Pose Increasing Risks." GAO/AIMD-96-84. http://www.epic.org/security/GAO_DOD_security.html

National Institute of Standards and Technology. 1996. "Generally Accepted Principles and Practices for Securing Information Technology Systems." NIST SPEC PUB 800-14.

National Institute of Standards and Technology. 1995. "An Introduction to Computer Security: The NIST Handbook." NIST SPEC PUB 800-12.

National Institute of Standards and Technology. 1991. "Public-Key Cryptography." NIST SPEC PUB 800-2.

National Research Council, National Academy of Sciences. 1991. *Computers at Risk.*

2. Computing in the Future

Every vision of the future of information technology includes predictions of increased dependence on computing infrastructure in daily life. While Moore's Law describes the inescapable commoditization of computing power, similar technology growth curves exist for both wired and wireless bandwidth and for Internet utilization.

One consequence of such trends is that smaller, cheaper computers will be embedded everywhere. It will be so inexpensive to power and to interconnect devices that computing infrastructure will be pervasive in daily life. Mobility is another revolution whose influence on computing is accelerating. No longer tied to desktops or confined within enterprise boundaries, users connect spontaneously using intelligent devices, interact with services available through the Internet and then disconnect, effectively disappearing from the networked infrastructure. Wireless communication in a ubiquitous computing environment enables mobility in ways that were not possible a generation ago, so that between successive Internet sessions, a user may have traveled halfway around the world.

Table 1. Examples of Pervasive Computing

• Wearable Computers	• Smart Classrooms
• Wearable Keyboards	• Enhanced Learning Environments
• Smart Homes	• Telematics
• Active Badges	• GPS-equipped Automobiles
• Active RFID tags	

Computing in the future will therefore have global reach and global participation as new services based on these capabilities become abundant. Many of these services will result in the creation and storage of data at unprecedented rates. In 2002, by some current estimates,[5] mankind crossed the "exabyte" boundary by creating more than a million terabytes of information, the equivalent of a half million libraries the size of the Library of Congress. In 2003, the worldwide production of information surpassed the amount of information previously produced by the human species in all of recorded history.

The implication is that massively powerful computing and information storage capabilities will be so inexpensive to manufacture and network that new services will be abundant. We can see the roots of many of these services today in the growing number of user-centric services:

- Internet commerce
- E-government
- On-demand, online services
- Telecommuting
- Individualized entertainment
- Telematics
- Telemedicine
- Defense and warfare

A principle attributed to computer networking pioneer Robert Metcalfe asserts that the value of networked infrastructure is proportional to the number of possible interactions in the network. It is not

Table 2. Two Alternate Futures

Future Given Current Trends	Trustworthy Future
Spam	Hassle-free systems
Identity theft	User-controlled privacy
Network outages	Self-aware networks
Malware	Self-adjusting networks
Frequent manual intervention	Self-healing networks
Unchecked abuses of laws and rights	Balanced regulation and law enforcement

[5] Lyman, Peter and Hal R. Varian, "How Much Information 2003?" Retrieved from http://www.sims.berkeley.edu/how-much-info-2003 on November 28, 2005.

hard to imagine that a significant fraction of the wealth created by the human species and its agents will be embedded in the computing infrastructure that is being created today.

Metcalfe's Law provides a powerful incentive for criminals, terrorists, and anarchists. By focusing on networked systems, attackers can impact value and wealth networks—the impact of system compromise extends far beyond the individual machine that is attacked.

The future course of the underlying technology is predictable. How the technology will affect the quality of daily life is less certain. Simply trying to adapt current information technology to globally networked and ubiquitous computing infrastructure suggests a world in which none of us would want to live.

It would be a world in which we were overwhelmed by unsolicited junk. Easy access to massive databases and online services would give rise to rampant identity theft as adversaries acquired phony credentials. Mission-critical services would be hostage to frequent (natural or malicious) network outages; indeed, the economics of automation would probably collapse because of the need for frequent manual intervention. In all aspects of life, user-citizens would be exposed to the largely unchecked abuses of laws and rights.

The challenge is to create an alternative future in which spam, viruses and worms, the plagues of modern information technology, have been eliminated.[6] In this vision of the future individuals would control their own privacy and could count on the infrastructure to deliver uninterrupted services. Because even expert users could not be expected to understand and act on the sheer volume of information flowing through the global networks, it would have to be a world of "hassle-free" computing in which self-aware, self-adjusting, and self-healing systems served as human-scale tools. In such a world, policy and technology fit together in a rational way, balancing human needs with regulation and law enforcement.

In short, it would be a world in which information technology could be trusted. It would be a world of trustworthy computing.

Overarching Vision

Our overarching vision for trustworthy computing is that it should be:

- Intuitive
- Controllable
- Reliable
- Predictable

Trustworthy infrastructures should support a range of reasonable policies but, because change is inevitable, they should be adaptable so that when the environment changes the technology is still useful. Because computing must deliver value to its users, it should enable, rather than constrain, them.

[6] "The Next Plague," Vincent Kiernan, *Chronicle of Higher Education*, January 29, 2005. http://chronicle.com/free/v51/i21/21a03601.htm

A key to achieving this vision is identity. As in the real world, cybersecurity demands a trust relationship between individuals. The reason that spam spreads so easily in the current Internet is the difficulty of determining the identity of an email sender. Virus authors have become expert at "scanning"—that is, determining the identity and capabilities of millions of Internet-attached computers. Owners of digital property legitimately want to know to whom their property has been licensed. Identity must be shared to be useful, but individuals should make individual choices about their personal privacy and the technology should support those choices.

This vision is only achievable if security and trust are designed into systems as integral properties, rather than as afterthoughts. It is, in fact, one of the brewing tragedies of the digital world that existing infrastructure was not designed with trust as a primary consideration. We are on the verge of creating a new wave of digital technology; if we are to avoid repeating the mistakes of the past decade, it is essential that these new systems be designed to operate securely "out of the box." That is to say, security should be the default condition, not an option.

Role of Security

The very concept of information security has undergone a massive refinement over the last decade. Once confined to methods for keeping potentially harmful users out, security is currently much more focused on enabling users to extract value from computing infrastructure—that is, security is concerned with letting the right people access the right information and services in a trusted environment. Security features in IT systems are, in a sense, like brakes on automobiles. Although brakes are used to slow or stop vehicles, their real purpose is to enable drivers to go faster by enabling them to avoid accidents caused by external threats (such as mechanical failure in other vehicles, rude or reckless drivers, road hazards, stop signals and heavy traffic). Better security is an enabler for greater freedom and confidence in the cyber world.

Why is it Difficult?

As noted above, adversaries have conspired among themselves and with the technology itself to create a very threatening world. Potential attackers come with a variety of motives and backgrounds. In recent months we have seen successful attacks mounted by thrill-seeking teenage attackers, by anarchists bent on bold political statements, and by thieves. Insiders are among the most disconcerting threats because they have the most access and knowledge to construct devastating attacks.

All potential attackers benefit from the decreasing cost of computing power and communications connectivity. Even more threatening are transnationals, sometimes funded by rogue states or wealthy individuals who can sustain huge infrastructure, research, and development investments to mount terrorist attacks on the cyber infrastructure of the civilized world.

Another difficulty is the disappearance of a useful notion of a defensive "perimeter" around which various technologies and policies can be deployed to thwart attacks. The threat has become "asymmetric." We are literally surrounded by the threat, and our recent experience has shown that any attempt to defeat attacks by excluding attackers from our midst is doomed to failure.

Traditional regulation and law enforcement is ineffective because geo-political boundaries are increasingly irrelevant to cyberspace.

Lastly, the ground rules keep changing. Computing infrastructure is dynamic and threats evolve in unpredictable ways. Security measures are met with counter-measures designed to defeat them. Forensic analysis is frustrated by technology aimed at hiding an attacker's identity. Rogue programs modify themselves and mutate in an imitation of biology to survive and reproduce. It seems that the attackers understand the nature of the arms race in which they have engaged us.

Need Focus on Long-Term Research

The immediacy of the threat has led to a focus on near-term needs. Because near-term needs mainly address methods for securing existing systems, this has led to investment in patching existing infrastructure rather than technological innovation of the sort that will be needed to devise the next-generation trustworthy computing base. Policy tends to lag innovation, so too much focus on near-term problems has also hindered the development of effective policy at all levels.

Innovation requires focus on long-term research, a kind of investment in which progress is measured by the extent and level of investment. In trustworthy computing, this focus has been episodic and so progress has not been sustained. Furthermore, the main source of long-term research funding for information security has been the defense agencies, and the problems of cybersecurity clearly go beyond the needs of any single federal agency.

A natural question is whether industrial investment alone will provide the stimulus needed. In recent years, the appetite of industrial research organizations for basic, long-range research has diminished dramatically. Indeed, many of the central research labs that would have launched significant projects in information security have simply disappeared, or have dramatically altered their mission to focus on near-term, product-oriented practical engineering.

Lastly, the talent pool at all levels is inadequate. Trained engineers and research scientists are needed to combat the growing and changing threats. Therefore, universities must be engaged to make real progress. One recommendation of the PITAC report was to double the size of the civilian cybersecurity fundamental research community by the end of the decade.[7]

[7] "Cyber Security: A Crisis of Prioritization," op. cit., p. 30.

3. Four Grand Challenges

The CRA Grand Challenges Conference participants considered the long-term needs of future computing environments in light of the need to escape the exponentially growing threat represented by the CERT/CC curves. Given the alternative futures and the difficulty of making progress with incremental approaches to trustworthy computing, many important suggestions were made that would lead to significant advances. Some of these suggestions are currently being investigated in university and industrial research labs around the world. Many others have already been identified by agencies such as the National Science Foundation. Still others were stimulated by the unique environment of the Grand Challenges Conference, and the conference organizers expect that they will be explored more fully as a direct result of the meeting.

In the end, conference participants expressed a need for a coherent set of challenges aimed at immediate threats, emerging technologies, and the needs of the future computing environment over a much longer term. The several discussion groups coalesced into four working groups aimed at defining the research challenges in each of these time frames. The Grand Research Challenges identified in this way were the following:

1. Within the decade eliminate the threat of all epidemic-style attacks such as viruses and worms, spam, and denial-of-service attacks.
2. As many new systems with great societal impact are currently planned or under development, develop tools and design principles that will allow these systems to be highly trustworthy.
3. Develop and validate quantitative models of risk and reward and deploy them to decision-makers so that progress can be made.
4. Lastly, setting its sights on the dynamic, pervasive computing environments of the future, provide understandable security and privacy to tens of millions of new users.

Challenge 1: Eliminate Epidemic Attacks by 2014

Epidemic attacks are intended to provoke catastrophic losses in the worldwide computing and communications infrastructure. Characterized by their extremely rapid spread, the costs (to legitimate users) of successful epidemics have risen in recent years to billions of dollars per attack. Examples of epidemic attacks include:

- Computer viruses and worms: programs that are launched from one or more sites, infect and sometimes damage individual computers, and spread geometrically to adjacent or connected machines.
- Spam: massive, unsolicited and unauthorized floods of email that exploit weaknesses in current infrastructure to clog enterprise and individual mailboxes with junk that effectively denies users access to legitimate information.
- Distributed Denial of Service attacks: coordinated direction of overwhelming unauthorized and malicious network traffic at one or more critical suppliers of IT or communications services to disable them and deny legitimate users access to their services.

These attacks spread very quickly, following a model of infection and transmission that parallels biological epidemics. A major difference between these epidemics and biological epidemics is the speed at which they propagate. Biological infections must incubate for days, months, or (as is the case with AIDS, BSE and other emerging health threats) years before they can be diagnosed. Similarly, propagation of biological infection follows pathways that require the movement of hosts, carriers or victims over macroscopic distances. Smallpox spreads dangerously, but the virus must travel meters to infect a new host, often requiring minutes or hours. Plague requires the proximity of carriers and humans to propagate.

Moore's Law favors cyber-epidemics. High bandwidth connections essentially make the instantaneous transmittal of infections possible. Sheer computational power is all that is needed to increase the number of new hosts that can be infected per unit of time. Thus, whereas global biological epidemics require at least months to establish themselves in a population, epidemic-style cyber attacks can propagate on a global scale within minutes. The Slammer worms, for example, infected 90 percent of the vulnerable hosts in less than 30 minutes. A sophisticated e-mail worm called MyDoom wreaked more than $20 billion in damages (see Box 3).

Box 3. MyDoom Worm—Final Toll and Disposition

Processing between 50,000 and 60,000 new copies per hour, "W32/Mydoom.A has exceeded the infamous SoBig.F virus in terms of copies intercepted, and the number continues to rise."

Message Labs collected over 1.2 Million copies of W32/Mydoom.A-mm.

At its peak infection rate, about 1 in 12 emails on the Internet were MyDoom Viruses.

Infected six times more computers than Bugbear.B.

300K computers infected worldwide.

Sources:
Message Labs, January 17, 2004. http://www.messagelabs.com
Panda Software, January 16, 2004. *Computer World* article.
http://www.pandasoftware.com/about/press/viewNews.aspx?noticia=4658

The price of launching an epidemic can be very low. Global Internet connectivity is ubiquitous and essentially free. Computing power that a decade ago would have been available only to the most advanced scientific research labs is now available for less than $1,000 from common retail outlets. There are no particular knowledge barriers to be overcome in launching such an attack—the Internet itself provides hundreds of how-to libraries for would-be attackers.

The sheer unpredictability of new attacks presents special problems in devising adequate defenses. Existing technology has enabled high-quality, pattern-matching antivirus tools that quickly look for suspicious signatures and isolate potential attacks. Virus-writers have learned how to construct "polymorphic" worms and viruses that change themselves to avoid easy detection. These mutations then propagate as new epidemics. It is also difficult to pinpoint the source of an attack. Attackers hide their identities using natural anonymizing features of existing network protocols or by using any of the hundreds of "anonymizers" available at attackers' websites.

The nature of the Internet hinders an organized active defense. The Internet operates under highly distributed and decentralized control. Its operations are the result of millions of poorly coordinated decisions. This organization has enabled the phenomenal growth of the Internet in the past ten years and makes it one of the most robust systems ever constructed. But there is poor visibility into its global operations and there are few methods for providing emergency global control to thwart an attack. Techniques that have thus far been tolerably successful involve voluntary cooperation and collaboration. These methods cannot react with the speed that would be required to stop a truly virulent viral attack.

Urgency of Halting Cyber Epidemics

As global business becomes increasingly dependent on the Internet and electronic commerce, minutes or hours of disruption become less easily tolerated. In 1999, the Forrester Group estimated the hourly downtime costs for emerging e-commerce ventures to be hundreds of thousands of dollars. Today, these costs have increased by at least an order of magnitude.[8]

Why is Progress Possible?

All stakeholders—researchers, regulators, policymakers, software and hardware vendors, telecommunications suppliers, educators, and "average" users worldwide—now recognize epidemic attacks as a critical problem, so investment in solutions is much more evident today than a decade ago. Researchers are already investigating new technologies that are promising. There is every indication that this investment will translate into the multiple streams of research that will be necessary to identify and cull the most promising approaches.

[8] One newspaper article reported that in October 2003 the economic losses due to malware amounted to $10.4 billion worldwide. "Spam Harmed Economy More than Hackers, Viruses," Tim Lemke, *The Washington Times*, November 10, 2003. According to Trend Micro, in 2001 viruses, worm and spyware cost businesses $13 billion; in 2002 the cost rose to $20-$30 billion; and in 2003 viruses, worm and spyware cost a record $55 billion in damages (http://www.securitystats.com/virusstats.html).

It is possible to imagine approaches that might be effective at deterring epidemic attacks if they were further developed and deployed:

- Immune System for Networks – capabilities built into a network to recognize, respond to and disable viruses and worms by dynamically managed connectivity.
- Composability – rules guaranteeing that two systems operating together will not introduce vulnerabilities that neither have individually.
- Knowledge Confinement – partitioning information in such way that an attacker never has enough knowledge to successfully propagate through the network.
- Malice Tolerance – much as many current systems can tolerate failures by continuing to operate even when many components do fail, tolerate malice by continuing to operate in spite of arbitrarily destructive behavior of a minority of system components.
- Trusted Hardware – tie software and service guarantees to the physical security of hardware devices.

Barriers to Success

Similar to the way in which the World Health Organization's Smallpox Eradication Programme demonstrated success with the elimination of naturally occurring smallpox outbreaks, the success of cyber-epidemic eradication will be demonstrated by the absence of naturally occurring attacks by Internet worms and viruses. Internet-wide service disruptions will no longer occur. Businesses will verify that massive spam attacks have subsided. Success will also be demonstrated by deployed technology for protecting the Internet. All new computers, servers, routers and appliances will supply standard protection features, and a visible mitigation strategy will help protect existing infrastructure.

There are significant barriers to be overcome in meeting such a challenge. It is not clear at the outset whose problem this is, which leads to finger-pointing among developers, network operators, system administrators, and users. Data are needed to drive experimental and theoretical work, but there is no current capability for gathering global network traffic data (see Box 4).

Box 4. No Internet-Scale Data Collection Effort

No situational awareness.

No early warning for spread of malware.

Decreased ability to track spread of malware.

No traffic trend analysis.

Unable to track distributed attacks.

Unable to assess consequences of adding new protocols and technology into network.

Source: NSF 98-120 Project Description. "Correlating Heterogeneous Measurement Data to Achieve System-Level Analysis of Internet Traffic Trends." http://www.caida.org/projects/trends/proposal/desc.xml

The distributed, decentralized nature of the Internet means that there is no simple formula that guarantees success. Experimentation will be needed and the only acceptable laboratory environments are those testbeds with the scale and the complexity of the Internet itself. If such testbeds cannot be constructed, extremely high fidelity simulations will be required. Currently, the technology to devise such simulations is not available. The February 2005 PITAC report placed the improvement in system modeling and the creation of testbeds on its list of priorities.[9]

Much productivity is lost in the current environment, and eliminating epidemics will enable the redirection of significant human, financial and technical capital to other value-producing activities.

Challenge 2: Enable Trusted Systems for Important Societal Applications

The first challenge addresses a critical problem for currently deployed IT systems. There are many new systems planned or currently under design that have significant societal impact, and there is a high probability that we will come to rely on these systems immediately upon their deployment. Among these systems are electronic voting systems, healthcare record databases, and information systems to enable effective law enforcement. A grand research challenge is to ensure that these systems are highly trustworthy despite being attractive targets for attackers.

Critical systems such as the ones mentioned above are being designed today, and it is a challenge to the research community to ensure that the mistakes of the past are not repeated. Despite many advances in computer and communications hardware and software, existing technology has not enabled us to build systems that resist failures and repel attacks. Decision-makers are today mandating the widespread deployment of electronic and Internet-based systems for uses that—should widespread attacks succeed—would undermine public institutions and structures to a catastrophic degree.

In many of these applications, computer systems are not used in isolation. Automated expressway tollbooths consist of computerized instrumentation that senses the approach of a vehicle with a certain identifying tag. This system communicates with a central database of registered vehicles to determine whether the owner of the vehicle has deposited sufficient funds to pay the relevant toll. This system communicates with a back-end billing system that records deposits from credit cards, bank accounts or other sources and issues invoices to drivers. For these systems to be useful, banking and credit card processing systems must have interfaces that allow the billing systems to operate properly. All of these systems rely on networking hardware and software that sends and receives packetized information, authorizes communication and transaction processing, and guards against intrusion and theft of personal information. Even perfect systems can be composed imperfectly, so the trustworthiness of the entire automated tollway system depends furthermore on the ability to compose systems into networks of trustworthy systems. It is beyond the ability of current technology to do that in a systematic, predictable way.

Future critical systems will be many more times complex than automated tollbooths, and the consequences of failure will be many times more severe (see Box 5).

[9] "Cyber Security: A Crisis of Prioritization", op. cit., p. 44.

Box 5. Example of Potentially Catastrophic Failure Resulting from Immature IT System

US railroad uses Wi-Fi to run 'driverless' trains (R 23 05; S 29 2:8); Union Pacific worker killed by locomotive he was operating remotely (R 23 07; S 29 2:8); Caltrain railroad accident results from deactivated crossing gate (R 23 08; S 29 2:8).

Spirit Rover failure on Mars: software upload to delete files failed, file space exceeded, caused reboot with insufficient file space, causing reboot loop (R 23 14, 15, see final summary in R 23 24).

Neumann, Peter G. 2005-06-29. "Illustrative Risks to the Public in the Use of Computer Systems and Related Technology," http://www.csl.sri.com/users/neumann/illustrative.html

Critical Applications Must Be Trustworthy

A number of critical technologies now in use or under development suffer from real vulnerabilities. Electronic voting is susceptible to both random and coordinated errors. Medical data and the corresponding personal records are stored, extracted and transmitted in electronic form under the weakest of trust guarantees. Federally mandated protections open the population as a whole to widespread abuses and risks. From privacy risks to threats of terrorist attacks by corrupting medical or pharmaceutical supply chains, emerging medical databases are a prime target for cyber attacks. Law enforcement agencies are increasingly interconnected using information technology and, as we have learned from the many lapses leading up to the events of September 11, 2001, the composition of systems in a web of trust that is adequate for law enforcement purposes, but which still protects individual privacy, is not yet feasible. War-fighting, intelligence–gathering, banking and finance, public utilities and transportation are also attractive targets as they become increasingly dependent on information technology (see Table 3).

There is very little reason to believe that such systems, if developed under current technology, will be trustworthy.

Table 3. Examples of Critical Systems and Infrastructure

Information and Communications Technologies	Transportation
Defense	Food Supply
Electrical Power	Gas and Oil
Banking and Finance	Water Supply Systems
Government Services	Emergency Services

Why is Progress Possible?

The fact that there is an early recognition that traditional methods will not suffice is an indication that progress is possible in this area. There has been a paradigm shift from perimeter defenses to asymmetric methods that emphasize intrusion detection, failure tolerance, denial of service protection and survivability. We expect that advances in these areas will provide tools that are useful in the design of new systems.

Similarly, the roles of social engineering and insider attacks are being re-examined to search for behavioral, usability and other ethnographic clues that can be applied to the design of such systems.

Cryptography provides extremely fertile ground for new approaches to system design that can be directly applied to problems such as voting. Indeed, there are systems in which mathematically strong guarantees can be given to participants about both the reconstructability and the secrecy of a given vote. If such techniques can be made practical and can be deployed in electronic voting platforms, progress can be assured.

Lastly, Moore's Law is an ally in the search for new technologies. Increased computing, storage and communications capabilities mean that we can deploy many purely defensive approaches that would not have been feasible a generation ago. Examples include intrusion detection systems, intrusion prevention systems, hybrid firewall/IDS/IPS, and heuristic-based anti-virus software.

Barriers to Success

Success will ultimately be demonstrated by the new systems that are developed. We will know, for example, that trusted healthcare databases have been deployed if we can create online systems that survive severe disasters and attacks without human intervention. Such systems will allow storage of medical information that provides:

1. Confidentiality, so there are no unauthorized disclosures of records.
2. Integrity, so it is impossible for unauthorized persons to alter records.
3. Auditability, so it is possible to reliably determine who has accessed records.
4. Availability, so no individual is denied access to medical services because records are not available.
5. Global accessibility, to allow projection of the most advanced medical delivery technology into countries and regions with little opportunity to directly access doctors, hospitals, and instruments.

The barriers to be overcome in developing effective tools are as much social and political as technological. Any successful attack on the problems posed above will have to reconcile various legal regimes with new technology. Social and cultural concerns may delay the widespread deployment of privacy-enhancing measures, for example.

The likely approaches to solving these problems are also certain to increase the cost and complexity of the technology. The new trust capabilities may have to achieve unprecedented levels of protection in order to be politically acceptable.

Lastly, all of the systems we have looked at involve integrating new technologies with legacy applications that have little or no protection. The enhanced technologies will have to provide strong "end-to-end" guarantees in spite of this.

Challenge 3: Develop Accurate Risk Analysis for Cybersecurity

Even the best technology for enabling trustworthy computing will be ineffective if it is not deployed and used in practice. Executives, corporate boards of directors and chief information officers are jointly responsible for balancing investments and risks, and use many metrics for determining whether or not a projected return on investment (ROI) justifies the investment. Spending for new information security measures is such an investment for most organizations.

Figure 2.

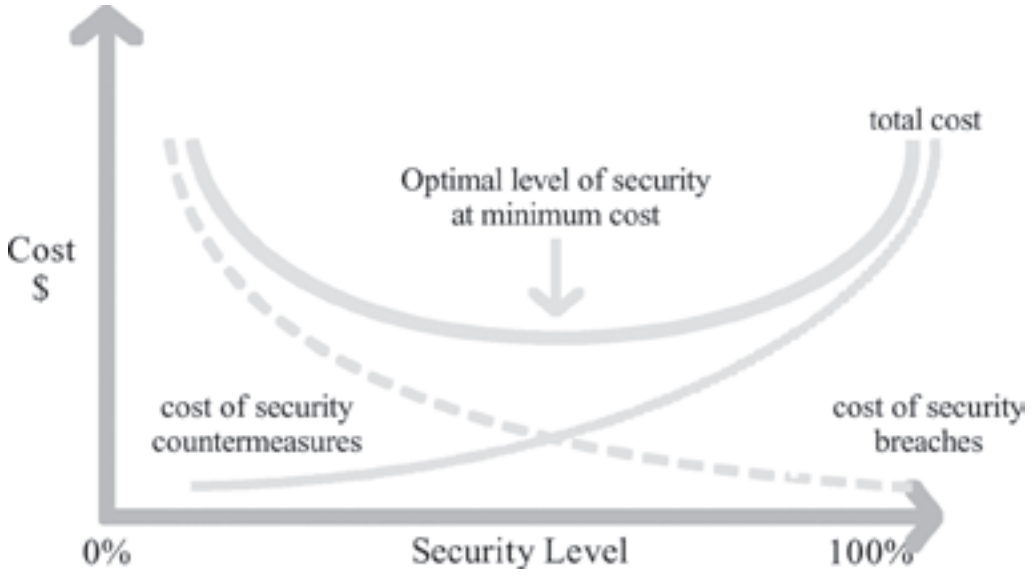


Figure 2 illustrates the kind of ROI analysis that is in common use today. By investing more money to increase the "level of security" an organization can decrease the costs or risk associated with security breaches. A wise manager chooses an effective tradeoff between risks and investment costs.

ROI analysis has proved to be remarkably ineffective in spurring effective investments in information technology. Despite CERT/CC data and daily newspaper headlines that document the increasing costs of attacks and vulnerabilities of the computing base, investments for information security in all types of organizations, expressed as a percentage of overall IT spending, has actually decreased since September 11, 2001.

A major reason for this seemingly irrational behavior on the part of decision-makers is the lack of effective models of risk. In other words, ROI analysis is only valid if there is some assurance that increasing spending on security measures actually increases security. In this regard, trustworthy computing is still in its infancy and models are needed that put information security on the same level as financial systems with regard to accurate risk modeling.

The challenge of the research community is to develop, within ten years, quantitative IT risk management that is at least as effective as quantitative financial risk management. The February 2005 PITAC report placed quantitative benefit-cost modeling at number nine on its list of "Cyber Security Research Priorities."[10]

The Difficulty of Assessing Risk

This challenge is especially difficult because we do not yet understand the full nature of what constitutes IT risk. In fact, as systems become more complex and interconnected, emergent behavior (i.e., unanticipated, complex behavior caused by unpredictable interactions between systems) of global systems exposes emergent vulnerabilities.

Meeting this challenge requires new mathematical and statistical models. One example of how difficult it will be to arrive at the correct models is the difficulty of modeling failures in networked systems. In most systems of engineering interest, failures are statistically independent. In other words, like the flipping of a fair coin—the appearance of heads or tails on one flip does not affect the outcome of the second flip of the coin—a component failure in one part of the system does not affect the failure of another similar component in another part of the system. This leads to especially beautiful and useful models of system failure that are effectively applied thousands of times a day by working engineers. In networked systems, however, failures are not independent. Components of networked systems share information, which leads to dependencies between system components. A failure in one component, rather than being an isolated failure, may correlate highly with the behavior of other system components, leading to massive, catastrophic system failure.

Being able to accurately model and ultimately predict such failures is the ultimate goal of this challenge. Without a widely accepted system of risk measurement, effective management of investment and risk is quite hopeless. Without an effective model, decision-makers will either over-invest in security measures that do not pay off or will under-invest and risk devastating consequences.

Two aspects of measurement that will need the most determined attention are:

1. Measuring the wrong thing is ultimately worse than not measuring anything at all.
2. Because choices and decisions need to be made by many organizations over long periods of time, the measures need to be consistent, unbiased and unambiguous.

Box 6. Why Does it Matter?

Lord Kelvin (William Thompson) wrote:

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science."

[10] "Cyber Security: A Crisis of Prioritization," op. cit., p. 30.

Why is Progress Possible?

Much progress has been made in similar, related areas, which gives us encouragement that progress is possible here. The mathematics of investment risk, epidemiology, public health, accelerated failure time testing, software assurance and other endeavors is undergoing a resurgence of development and there is much that can be learned from these fields.

Perhaps even more importantly, the researchers are collecting data in sometimes prodigious amounts. These data will serve as the basis for new experimental and theoretical work.

Lastly, in an attempt to protect against terrorist attacks, policy- and decision-makers, as well as the citizenry in general, have become accustomed to rough models of risk that can be used to adjust resources and behavior.

Barriers to Success

Successfully meeting this challenge will result in predictable outcomes for security investments. ROI analysis will be a practical reality.

Aside from the purely technical barriers of picking the right measure and gathering the right data, there are a number of significant social and cultural barriers that will have to be overcome.

Data-gathering itself presents some severe challenges. In many organizations, these kinds of data are either proprietary or closely held. There is no "first mover" advantage in disclosing the data, so finding willing partners for research collaborations will be difficult. In many organizations, there are perceived risks involved in releasing data that expose vulnerabilities. Executives cannot hide behind "plausible deniability" when the data are open to inspection. Some organizations would regard the release of such data as an admission of negligence or wrongdoing. Systems will have to be devised that address such issues.

Data-gathering will also require data-sharing. Standards and common terminology do not exist in any useful form today, so cooperation among sometimes competing organizations will be required to address this grand challenge.

Challenge 4: Secure the Ubiquitous Computing Environments of the Future

The fourth and final grand challenge is to protect our future technological base. For the dynamic, pervasive computing environments of the future, we will give computer end-users security they can understand and privacy they can control. Technology can easily outrun comprehensibility, and a trustworthy computing base should not make this worse. By the same token, identity will be many-faceted and ubiquitous in a world of pervasive computing, and individuals should be able to maintain control of it.

The future is a grand challenge because it looms before us. The pace of technology change continues unabated. Instant access to information is becoming a reality. IT is being exploited everywhere, but especially in consumer-oriented environments in which convenience, safety and empowerment are central issues.

As has been pointed out several times already, the risk of leaving these concerns for later is unacceptably high. Building security into our design for the future world is a necessity.

Getting it Right from the Start

Experience teaches us that it is important to treat security as a driving concern from the earliest stages of system design. Also, our experience with the adoption of the Internet is evidence that information security has to reflect the sensibilities of the underlying social systems as opposed to simple technological systems. If the systems of the future are deployed without adequate security, or if our privacy is compromised in the name of technological change, we may not be able to regain it. It is important to issue a call to the security community to assert a leadership role now.

Why is Progress Possible?

We are at a unique moment in history in which there is widespread concern in many segments of society. There is a new awareness that trust and cybersecurity require a broader view of needs and the groundwork is being laid to respond to this challenge (see Box 7).

Barriers to Success

We do not yet live in the future, so we can only make educated guesses about the nature of user needs and acceptance for new security methods and mechanisms. In addition to the pure technology challenges posed by dynamic, changing environments and the wide variety of IT devices that will be required to deliver services to end-users, there will be multiple, competing stakeholders.

We know that user needs will be much broader than traditional security models. IT has traditionally been focused on the underlying mechanisms. In the future, IT will be much more human-centered than it has been in the past, and it will be a significant challenge to bridge this gap between users and underlying mechanisms. Another barrier will be to reconcile privacy with security.

We will know we have succeeded if the future is filled with IT that is accepted by society, in which users are in control.

Meeting this challenge will enable the emergence of a world in which ubiquitous computing and communications are simple and easy to use, dependable and reliable and non-intrusive. It will be a world of trustworthy computing.

Box 7. Description of NSF Cybertrust Initiative

The cybersecurity center led by the University of California, Berkeley, will investigate key issues of computer trustworthiness in an era of increasing attacks at all levels on computer systems and information-based technologies. The Team for Research in Ubiquitous Secure Technology (TRUST) will address a parallel and accelerating trend of the past decade--the integration of computing and communication across critical infrastructures in areas such as finance, energy distribution, telecommunications and transportation.

"The overlapping and interacting trends force us to recognize that trustworthiness of computer systems is not an IT (information technology) issue alone," say center leaders. They explain that the center will lead development of new technologies based on findings from studies of software and network security, trusted platforms and applied cryptographic protocols. The center will also look at systems problems through modeling and analysis, development of secure, embedded systems, and integration of trusted components and secure information management software. The center will merge these efforts with investigations of social science questions involving economics, public policy and societal challenges, human-computer interfaces and privacy, among other issues.

The TRUST center will also have an education and outreach component to K-12 schools, undergraduate students and institutions serving underrepresented populations. These education programs will lay the groundwork for training new scientists and engineers, who, center leaders say, will develop the next generation of trustworthy systems. Students will also benefit in their future roles as users, consumers and beneficiaries of these systems. The overall project involves several major partners.

NSF established the Science and Technology Center program in 1987, responding to a presidential commitment to fund important fundamental research activities that also create educational opportunities. The program was designed to encourage technology transfer and provide innovative approaches to interdisciplinary research challenges. In 1997, the STC program was modified to emphasize the contributions of partnerships.

Source: NSF press release 05-053 http://www.nsf.gov/news/news_summ.jsp?cntn_id=103178&org=NSF&from=news

Appendix: Conference Attendees

Virgilio A.F. Almeida	University of Minas Gerais, Brazil
Annie I. Antón	North Carolina State
Dirk Balfanz	PARC
Terry V. Benzel	ISI, University of Southern California
Andrew Bernat	Computing Research Association
Luis Bettencourt	Los Alamos National Lab
William J. Caelli	Queensland University of Technology
David D. Clark	MIT
Kay Connelly	Indiana University
Steve Crocker	Shinkuro Inc.
Richard A. DeMillo	Georgia Tech
David Evans	University of Virginia
David Farber	Carnegie Mellon University
Scott Flinn	NRC of Canada
Simson L. Garfinkel	MIT
Dan E. Geer, Jr.	TheWorld
Anup K. Ghosh	DARPA ATO
Helen Gill	National Science Foundation
Virgil D. Gligor	University of Maryland
Peter Harsha	Computing Research Association
James Jay Horning	Network Associates Labs
Cynthia Irvine	Naval Postgraduate School
Doug Jacobson	Iowa State University
Anita Jones	University of Virginia
Kamal Karlapalem	IIIT Centre for Data Engineering, Hyderabad, India
Jay Lala	Raytheon
Susan Landau	Sun Microsystems Labs
Carl Landwehr	National Science Foundation
Ruby B. Lee	Princeton University
Doug Maughan	Potomac Institute for Policy Studies
Dana Neill	Computing Research Association
Clifford Neuman	ISI, University of Southern California
Ralph C. Merkle	Georgia Tech
Peter G. Neumann	SRI
Cristina Nita-Rotaru	Purdue University
Kenneth G. Olthoff	NCSC
David Pager	University of Hawaii
David A. Patterson	UC Berkeley
Adrian Perrig	Carnegie Mellon University
John Richardson	Intel Corp.
M. Angela Sasse	University College London, UK
Stefan Savage	UC San Diego
R. Sekar	SUNY Stony Brook
Jonathan S. Shapiro	Johns Hopkins University

Daniel Simon	Microsoft Research
Diana Smetters	PARC
Jean Smith	Computing Research Association
Sean Smith	Dartmouth College
Anil Somayaji	Carleton University, Ottawa
Eugene H. Spafford	Purdue University
David Stork	Ricoh Innovations, Inc.
Stephen L Squires	Hewlett-Packard Co.
Bhavani Thuraisingham	National Science Foundation
Ravi Sundaram	Northeastern University
Peter Wayner	Author
Jeannette M. Wing	Carnegie Mellon University
William Wulf	University of Virginia and National Academy of Engineering
Suguru Yamaguchi	Nara Institute of Science and Technology, Japan
Taieb Znati	National Science Foundation

1100 17th Street, NW, Suite 507, Washington, DC 20036-4632
URL: http://www.cra.org • E-mail: info@cra.org
Tel: 202-234-2111 • Fax: 202-667-1066