CYBERNATION:

PB99-140170

THE AMERICAN INFRASTRUCTURE IN THE INFORMATION AGE

A Technical Primer on Risks and Reliability

Executive Office of the President
Office of Science and Technology Policy

Reproduced from
best available copy.

About the Office of Science and Technology Policy:

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization and Priorities Act of 1976. OSTP's responsibilities include advising the President on policy formulation and budget development on all questions in which science and technology are important elements; articulating the President's science and technology policies and programs; and fostering strong partnerships among Federal, State, and local governments, and the scientific communities in industry and academe.

To obtain additional information regarding the Office of Science and Technology Policy, contact the OSTP Administrative Office at 202-395-7347.

This report was produced by the National Security and International Affairs Division, 202-456-2894.

April 1997

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	3. PB99-140170
4. Title and Subtitle **Cybernation: The American Infrastructure in the Information Age**			5. Report Date **April 1997** 6.
7. Author(s) **Office of Science and Technology Policy**			8. Performing Organization Rept. No.
9. Performing Organization Name and Address **Office of Science and Technology Policy** **Old Executive Office Building** **Washington, DC 20502**			10. Project/Task/Work Unit No. 11. Contract © or Grant (G) No. (C) (G)
12. Sponsoring Organization Name and Address **same as above**			13. Type of Report & Period Covered 14.

15. Supplementary Notes

none

16. Abstract (Limit: 200 words)

The infrastructure on which American society depends, in sectors such as transportation, finance, energy, and telecommunications, is becoming increasingly automated as advances in information technology open up new possibilities for improved service, lower cost, and greater efficiency. The automation - or *cybernation* - of the infrastructure has come about largely because it offers unmistakable economic and performance benefits. As a result, however, the United States has become a wired nation, with implications that are not fully understood.

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
none	**none**	**34**
	20. Security Class (This Page) **none**	22. Price

(See ANSI-Z39.18) *See Instructions on Reverse* OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

CYBERNATION:

The American Infrastructure in the Information Age

A Technical Primer on Risks and Reliability

PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED.
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE

At a glance...

The automation – or *cybernation* – of the domestic infrastructure of the United States, in the transportation, finance, energy, and telecommunications sectors, which has been building for decades, has accelerated dramatically in recent years as advances in computers and information networks open up new possibilities for improved service, lower cost, and greater efficiency. As a result, the United States has become a wired nation, a condition with implications that are not fully understood. [p. 9]

The importance to the nation of infrastructure services makes attention to the reliability of their underlying information networks a necessity. The question is whether the marketplace will adequately anticipate and mitigate reliability deficiencies, or whether the nation will have to endure a major infrastructure problem in order to mobilize and act. [p. 9]

The infrastructure of the United States has historically been very reliable. For most Americans, infrastructure disruptions have been more a nuisance than a nightmare. However, nothing guarantees that future disruptions will be similarly limited in national impact as past disruptions. [p. 15]

Three current trends raise concerns about the reliability of the automated infrastructure:

- Infrastructure services are becoming increasingly dependent on complex information networks which are potentially vulnerable to failure or disruption.
- The business environment is changing with deregulation, downsizing, increasing competition, and the entry of new companies into the market for providing infrastructure services.
- Infrastructure information networks are potentially becoming more accessible even as computer intrusions, already quite common, become increasingly sophisticated. [p. 13]

Network failures can be classified in terms of their *causes* and the *mechanisms* by which they are manifested.

- *Causes* range from natural phenomena such as weather, natural disasters, and other acts of God to deliberate destructive acts by persons intent on doing damage.
- *Mechanisms* range from chain reactions, in which small faults propagate and result in widespread disruptions, to the direct, independent failure of key components that in themselves represent major disruptions. [p. 21]

From a technical standpoint, it is not practical to focus exclusively on any one reliability threat. Like the interactions of prescription drugs, the remedy for one problem can interfere with the remedy for another. A holistic methodology for making the unavoidable tradeoffs is called for. [p. 26]

A technical agenda for addressing the reliability of infrastructure information networks consists of three steps:

1. Develop an analytical understanding of the specific reliability, vulnerability, and threat environment.
2. Establish a system engineering process which treats reliability as a primary parameter.
3. Maintain constant vigilance and continual learning to enhance reliability. [p. 23]

Neither the private sector nor the government can completely address infrastructure reliability alone. Developing consensus on the problem, as well as finding effective long term solutions, will require the sustained engagement of industry, utilities, the public, and government at all levels. [p. 11]

Areas for increased public policy attention include: [p. 32]

- Achieving consensus on what the minimum levels of reliability should be, what the threats are, what risks are acceptable, what protective measures should be taken, and how the costs should be met.
- Enhancing government/industry cooperation for identifying and characterizing reliability challenges, from weather and natural disaster prediction to intelligence collection on the threat of hostile attack.
- Focusing government and industry on the joint development of technical standards and methods to measure and certify reliability.
- Enhancing Federal/State government interaction to ensure consistent and appropriate attention is placed on infrastructure reliability.
- Defining the government research and development investment portfolio for network reliability.
- Working with other countries to develop compatible international legal regimes in cyberspace.
- Clarifying missions, responsibilities, and authorities of Federal Departments and Agencies in cyberspace.

THE WHITE HOUSE

WASHINGTON

The domestic infrastructure which underpins the economic life of our society increasingly depends on electronic networks for the flow of essential information. In sectors such as transportation, finance, energy, and telecommunications, computer networks have become indispensable in providing essential services that we take for granted. Air traffic data for the safe conduct of thousands of flights per day, financial transactions worth many millions of dollars daily, and control signals for operation of power distribution grids, railroads, pipelines, and the telephone system itself, all travel over electronic networks. Electronic networks have truly become the "nerves" of our infrastructure in yet another manifestation of the proliferation of information technology that characterizes the world of today.

How reliable are these networks? How can we ensure they are reliable enough? These pressing questions are not easily answered. Our critical infrastructure information networks face many reliability challenges, from natural disasters to human error, and from equipment failure to terrorists and computer hackers. From a technical standpoint, these are not different problems; they are different parts of the same problem. A systematic sector-by-sector analysis of threats and vulnerabilities, and a sustained system engineering process that emphasizes reliability are the technical ingredients of a successful approach to managing these risks.

As powerful a tool as technology can be, it is not the whole answer. Technology, and especially information technology, is best understood in its societal context. *People* represent both the strongest and the weakest links in the reliability chain. We should therefore not lose sight of the human element as we focus on the technical challenges of assuring infrastructure reliability. Instilling a culture of vigilance in the community responsible for the infrastructure is the most fundamental step in preventing reliability problems.

Deciding how much reliability we need for our infrastructure, against what threats, and at what cost are questions of public policy that will require the sustained consideration of stakeholders throughout society. This report seeks to promote a common understanding of the network reliability challenge in the technical and policy communities in private industry, public utilities, and government at all levels. The efforts of these diverse players, through the broad dialogue of democracy, will be necessary to effectively respond to this long-term challenge.

John H. Gibbons

John H. Gibbons
Assistant to the President
for
Science and Technology

Contents

EXECUTIVE SUMMARY 1

INTRODUCTION 5

INFORMATION NETWORKS AND THE DOMESTIC INFRASTRUCTURE 9

THE TECHNICAL DIMENSION 13

FINDING SOLUTIONS 23

CONCLUSION 31

Executive Summary

The infrastructure on which American society depends, in sectors such as transportation, finance, energy, and telecommunications, is becoming increasingly automated as advances in information technology open up new possibilities for improved service, lower cost, and greater efficiency. The automation – or *cybernation* – of the infrastructure has come about largely because it offers unmistakable economic and performance benefits. As a result, however, the United States has become a wired nation, with implications that are not fully understood.

The widespread application of information technology presents new challenges in what has historically been a highly reliable infrastructure. A changing public utilities business environment characterized by deregulation, corporate downsizing, increased competition, and new entrants to the market potentially places stress on the reliability of the national infrastructure. In addition, the nearly unconstrained application of computer technology in infrastructure control systems raises questions about the reliability of complex systems and their vulnerability to hostile intruders. Whether the forces of the marketplace will continue to provide infrastructure services with acceptable reliability in this environment remains to be seen. The importance to the nation of infrastructure services makes attention to the reliability of their underlying information networks a necessity.

Reliability challenges stem from both natural and manmade sources. To date, most of the national experience with major service interruptions caused by problems with infrastructure-related information networks comes from natural causes, accidents, or human shortcomings in design or operation. Infrastructure information networks will always be subject to these kinds of failures. However, as computer hackers increase in number and grow in sophistication, the threat

of purposeful attacks by hostile actors looms increasingly large. Addressing this dual challenge in a measured way will be a long-term public policy priority.

This report describes the technical problem and sets forth a technical agenda for addressing network reliability. It uses the term *reliability* in its simplest sense — flawless, dependable operation, from the consumer's perspective, despite any reasonable challenge. It first defines a conceptual framework for characterizing network failures in terms of their causes and the mechanisms by which failures are manifested. It then outlines a technical agenda within this framework consisting of three steps: (1) developing an analytical understanding of the existing reliability, vulnerability, and threat environment; (2) establishing a system engineering process that treats reliability as a primary parameter; and (3) fostering a commitment to vigilance and a process of continual learning to enhance reliability.

Although the nation is truly fortunate that major, sustained infrastructure outages have not occurred, this good fortune makes foreseeing and forestalling presumptive threats to infrastructure networks more difficult.

Network failures can be classified in terms of their causes and the mechanisms by which they are manifested. Broadly stated, *causes* range from purely natural phenomena such as weather, natural disasters, and other acts of God to deliberate destructive acts by persons intent on doing damage. Between these two extremes lies a wide range of accidental or unintended occurrences with varying degrees of human involvement and varying human motivations.

Similarly, the *mechanisms* by which failures come about vary between two extremes as well. On one hand, a localized failure can become widespread through a chain reaction, in which a subsystem or component failure induces other failures, ultimately propagating through the network until overall performance is significantly degraded. The power outage that affected a large region in the western United States in summer 1996, for example, was attributed to a downed power line in Oregon which caused control system reactions that took generators in several States off-line in succession. On the other hand, major disruptions can occur that are not the result of chain reactions, but rather the independent disablement of critical subsystems. The destructive power of the Northridge, California earthquake in 1994, which interrupted electrical power and telephone service for millions of people, for example, caused outright failure of networks and did not depend on a chain reaction. The first failure mode can be likened to the domino effect; the second is more akin to upsetting the game table and knocking the dominoes to the floor. Of course, a great many events that have some qualities of each fall within these bounds.

Many of the recognized threats to the information networks supporting the domestic infrastructure have not actually been experienced. Although the nation is truly fortunate that major, sustained infrastructure outages have not occurred, this good fortune makes foreseeing and forestalling presumptive threats to infrastructure networks more difficult.

The first step in improving network reliability is to understand the existing reliability, vulnerability, and threat environment. This requires a detailed examination of the network architecture, physical layout, hardware and software, communications links, human factors, and operations. The findings of such an examination can guide a reliability engineering process.

The second step is to establish reliability as a primary tradeoff parameter in a system engineering process. Network configuration changes should not be implemented without a careful assessment of the tradeoffs involved. A

strategy designed to counter one threat may increase vulnerability to another. Strategies narrowly focused on one aspect of reliability may introduce new vulnerabilities. The likelihood and severity of a particular problem may not justify the cost of a proposed solution. Additionally, the cost and performance implications of every strategy need to be understood. A structured methodology for making the unavoidable design tradeoffs between such primary factors as performance, cost, and reliability is essential.

The third essential step of the technical agenda is to foster a commitment to vigilance and a culture of continual learning to enhance reliability. As critical infrastructure information networks grow in size and complexity, there is an urgent need for institutionalized methods for capturing and applying the lessons of experience. Tools and procedures for detecting, reporting, and reacting to network problems all need to be developed and strengthened. An equitable, institutional means, within clear statutory limits, for the timely two-way flow of relevant intelligence information and incident data between government and the public utilities, which protects business-sensitive data as well as sources and methods, would do much to clarify the threat environment and allow for an effective response.

Although industry has a vested interest in assuring the reliability of the infrastructure, the federal government has an indispensable role as well. Neither the private sector nor the government can completely address infrastructure reliability alone. The national interest can only be served with the sustained engagement of industry, utilities, the public, and government at all levels.

Introduction

One of the great engineering marvels of the ancient world was the infrastructure of roads and aqueducts developed by the Romans. The road network, begun in about 300 BC, ultimately consisted of some 50,000 miles of hard-surfaced highway, some of which still survives today. Motivated by military needs, it also facilitated trade, agriculture, mail delivery, and made possible the establishment and administration of Roman rule in the far reaches of the empire. The aqueducts, ambitious projects even by today's standards, brought water to the city for public and private consumption, supplied baths and fountains, and provided for irrigation and sanitation. The well-developed Roman infrastructure contributed immeasurably to the prosperity and economic vitality that are among the hallmarks of ancient Roman civilization. But for all its advantages, this infrastructure also created new and serious vulnerabilities, providing attacking hordes easier access to Roman cities and becoming a target of direct attack itself. As a consequence, the Romans invested heavily in the construction of walls and other fortifications along their highways and around their cities, and they enacted laws and decrees aimed at protecting the structures associated with the water supply.

Today, the infrastructure of the United States is itself an engineering marvel. On a daily basis, the domestic telephone system carries hundreds of millions of calls, domestic and global financial networks conduct trillions of dollars worth of transactions, and our electrical grid serves hundreds of millions of consumers with a total generating capacity measured in the hundreds of millions of kilowatts of power. In sectors such as transportation, finance, energy, and telecommunications, our infrastructure is the machinery behind the American way of life.

However, the national infrastructure today is more than just a larger, more modern and complex version of the Roman road and aqueduct system. Today's infrastructure has a fundamental, indeed momentous, distinguishing characteristic: *it is automated.* From the routing of telephone calls to the distribution of electrical power, from the separation of aircraft to the electronic transfer of funds, the domestic infrastructure operates through automatic information networks. In all sectors, computer networks are an integral part of infrastructure operations – controlling processes, conducting transactions, dynamically adjusting capacity in response to usage, mediating communications among distributed components, and conveying information to human operators. This trend towards cybernation has been building for decades, but it has accelerated dramatically in recent years.

Today's infrastructure has a fundamental distinguishing characteristic: it is automated.

In today's dynamic business environment, increased reliance on automation makes good business sense. Automation with information technology enables new and better service offerings, more efficient operations and use of resources, and the potential for competitive advantages through greater responsiveness to customer demand. Information networks can provide managers with remote access for overseeing and managing their systems and can make it possible for them to tailor infrastructure services to specific customers.

It is worth considering whether the rapid and widespread adoption of information technology, for all its benefits, might also introduce vulnerabilities that could reduce the dependability that the public expects of the infrastructure. Everyday experience shows that when complicated computer systems fail, the failure is often both sudden and complete. "The computer is down" is a familiar lament for all who have endured the temporary inconvenience of computer problems in the workplace, the supermarket, the ticket counter, or the rental car agency. Such aggravations, however, are dwarfed by the potential problems that lie in the wake of comparable malfunctions in the computer networks supporting the nation's infrastructure.

The complex computer networks on which infrastructure operations increasingly depend are subject to failures just as any other manmade system. Human failings in design, construction, or operation, along with the effects of nature, aging, and natural disaster all have clear reliability implications. Any of these unavoidable problems could result in major disruptions of infrastructure services. Even more sobering is the possibility that remote access capabilities, so beneficial for customer service, could also allow computer terrorists – latter day barbarians at the gate – to deliberately disrupt infrastructure services by interfering with the information networks of the underlying control systems. In the extreme, disruptions or failures could threaten the well-being of society and undermine national security.

This report focuses on the reliability of the information networks that support the domestic infrastructure. It uses the term reliability in its simplest sense – flawless, dependable operation, from the consumer's perspective, despite any reasonable challenge. This report seeks to foster a common understanding among the technical and policy communities on the nature of the challenges to network reliability and the means to confront them. In doing so, it considers two questions:

- *How reliable are the critical infrastructure information networks?* An understanding of the technical problems and a sense of proportion about threats and vulnerabilities are essential to ensuring that the right priority is placed on addressing them.

- *How can society be certain that critical infrastructure information networks are reliable enough?* Achieving consensus on the appropriate levels of reliability, and on approaches for meeting them with acceptable costs, is an enduring challenge as the automated infrastructure evolves.

A framework for seeking answers to these questions is laid out in the pages that follow. Although this report concentrates on technical issues, technology alone is not a sufficient response to the reliability challenge. A complete approach must include operating procedures, training and awareness, personnel practices, and organizational factors in combination with technology. Indeed, technical solutions already available are not always effectively implemented. Furthermore, a framework for addressing infrastructure reliability as a public policy challenge is a prerequisite for delineating the issues and reaching consensus on the problems and the range of appropriate solutions.

Information networks and the domestic infrastructure

Electronic control systems are a common, though generally uncelebrated, feature of modern life. Familiar examples abound, from the simple thermostat to automobile cruise controls to automatic cameras. Each purposefully regulates a physical system to achieve a performance objective: if too hot, turn off the heater; too slow, add gas; too much light, reduce the aperture and increase the shutter speed. The technical discipline of automatic control is called *cybernetics*. It employs *feedback* – the use of measurements of present *output* to influence the next *input* – to converge on a desired operating condition. Electronic signals representing these measurements and control inputs are the coin of the cybernetic realm.

THE CYBERNETIC INFRASTRUCTURE

Broadly speaking, the control of any physical process with electronic signals is an application of cybernetics. In this sense, computer networks are the cybernetic control system of the domestic infrastructure. The "cyber" components of the infrastructure, including the computer hardware, software, communication links, and the abstract information embodied in them, make up the nervous system of the infrastructure on which the American public depends.

The sectors of the domestic infrastructure, as well as the cybernetic systems that support them, all have distinctive features, including:

- The energy sector provides power to meet the needs of the public in all aspects of modern life. It delivers energy in the form of electricity, oil, and natural gas, and has significant

physical plant consisting of production facilities, distribution networks, substations, and rights-of-way. In this sector, the underlying cybernetic networks are made up of supervisory control and data acquisition equipment and associated communications links which control switches, relays, pumps, and valves throughout the distribution system. Its communications links often use the same distribution lines and rights of way as the infrastructure itself.

- Electronic transactions within the financial services infrastructure underpin the entire national economy, as well as the operations of the other infrastructure sectors. This sector depends on communications links and geographically distributed computer data bases, and uses electronic networks for the transfer of funds for consumer and business-to-business transactions, inter-bank transfers, stock, bond, and commodities markets, and government-to-government financial transactions.

- The transportation infrastructure, in addition to providing the mobility of personal travel for people, also delivers the manufactured goods and agricultural products that are the lifeblood of commerce. In this sector, information networks are used for traffic control, navigation, and separation in the air, on the sea, on coastal and inland waterways, and on the ground.

- The telecommunications infrastructure is unique in that it not only is designed to deliver a service – the ability to communicate – but it also often comprises the signal channels on which the other sectors rely for the flow of their own cybernetic information. For the public telephone network itself, the cybernetic network is the signaling and call routing system, and the switches and signals that control individual connections. More broadly, it includes the internal corporate information networks and computer data bases that telecommunications companies use to support the operations, administration, maintenance, and provisioning of the wide range of services they offer, from "plain old telephone service" to digital, wireless, broadband, and customized subscriber services.

Besides these individual characteristics, the different sectors of the domestic infrastructure have much in common, including:

- The sectors serve a wide variety of customers throughout society. Major interruptions in the services of any sector could have serious and widespread health, safety, and national security implications.

- There are numerous interconnections and mutual dependencies among the infrastructure sectors and among the information networks that support them. The public telephone network, for example, relies in part on the power grid, the power grid on transportation, and all of the sectors on telecommunications and the financial infrastructure. Most sectors employ the public telephone network for at least some of their cybernetic channels. Most control networks also have some connection to public networks, many to the Internet. Additionally, there are shared rights-of-way in many locations throughout the country.

- The infrastructure is inherently regional, national, and even global in scope. All sectors have components distributed over wide geographic areas.

- The infrastructure sectors are owned and operated predominantly by private industry, with various sector-specific interfaces with Federal, State, and local governments.

- Varying degrees of coordination exist among providers within a sector, but there is no complete central authority within or among sectors. Approaches to reliability vary by sector, ranging from voluntary self-regulation to various forms of partnership between the private sector and the govern-

ment. Some degree of government regulation is the norm within all sectors, despite a general trend towards deregulation.

NETWORK RELIABILITY AND PUBLIC POLICY

The domestic infrastructure, although largely in private hands, is of such importance to the well-being of the nation that the government has an abiding responsibility in seeing that it best serves the national interest. For decades, government and industry have worked together to establish minimum levels of service, fair prices, and equitable access to infrastructure services for the American people. As the vulnerabilities of the automated infrastructure become understood, reliability takes its place as an explicit public policy objective as well. Simply put, the reliability objective is to provide, at reasonable societal cost, flawless, dependable infrastructure services that can withstand foreseeable challenges without interruption. Pursuing this goal will be a long-term focus of public policy.

The concern for today is whether the marketplace will adequately anticipate and mitigate reliability deficiencies, or whether the nation will have to endure a major infrastructure problem in order to mobilize and act.

The private sector has historically taken the lead in setting and meeting reliability goals in most infrastructure sectors. This approach has been highly successful. Because reliability deficiencies affect the corporate bottom line, either by disrupting revenue-producing services or eroding customer confidence and loyalty, industry can be expected to continue responding to credible reliability threats in the future. However, the industry is undergoing rapid changes in all sectors. Corporate reengineering, downsizing, the entry of new service providers due to deregulation, and the almost unconstrained application of computer network technology are all putting new pressures on formerly staid public utilities. The concern for today is whether the marketplace will adequately *anticipate* and *mitigate* reliability deficiencies in this highly dynamic environment, or whether the nation will have to endure a major infrastructure problem in order to mobilize and act.

The traditional policy tools available to the government for working with the marketplace to achieve national objectives – including legislation, regulation, licensing, tax and rate-setting regimes, and other inducements – all offer important options in the reliability arena. However, none of them can be effective unless and until there is consensus on what the minimum levels of reliability should be, what the threats are, what risks are acceptable, what protective measures should be taken, and how the costs should be met. At the present time, these questions are far from settled. Even the terms in which reliability thresholds should be expressed in the various sectors are open to debate. Additional policy emphasis can help address these difficult questions, but ultimately societal consensus will be a product of the broad democratic process. Part of the government's responsibility is to stimulate the public discourse and provide avenues for it to reach fruition in public policy. Government-sponsored forums, such as the Federal Communications Commission's Network Reliability and Interoperability Council in the telecommunications sector, can be instrumental in this regard by bringing together service providers, equipment manufacturers, standards-setting organizations, and consumer organizations in a public forum to develop recommendations for enhancing network reliability.

The Federal government, to create a climate that encourages infrastructure reliability through private sector initiative, can play a vital role by fostering innovation and commercialization, working with industry to develop technical standards, encouraging the development of

methods to measure and certify reliability, and making it easy for industry to make reliability improvements. The government can also contribute by applying Federal capabilities to identify and characterize reliability challenges, from weather and natural disaster prediction to intelligence collection on the threat of hostile attack. Identifying appropriate activities in these areas, and putting them into practice, calls for continuous public policy attention.

Federal government interaction with State governments, and with State and regional regulatory commissions, is also essential to framing the question of how much reliability is needed in the infrastructure and to ensure that consistent and appropriate attention is placed on the reliability of the associated information networks. The Federal government can be helpful in providing uniform guidance to the States where appropriate, and in coordinating and focusing the resources of the Federal departments and agencies.

Public policy extends to government investment in research and development as well. Federal investment in science and technology – conducted at federal laboratories, universities, and in industry – has been instrumental in the unfolding information revolution. The advanced development of the integrated circuit, creation of the Internet, development of a global communications infrastructure, and other products of government research have done much to enable the automation of the infrastructure. Wise federal investments today in areas such as intrusion detection and prevention, robust network architectures, configuration management, and secure communications will pay similar dividends in the future.

Reliability is also served by laws that delineate the boundaries of permissible behavior. As an international arena that disregards national boundaries, cyberspace presents unique legal difficulties. Distance and geography do not impede the trespasser, bandit, or terrorist in this realm, and questions about jurisdictions, sovereignty, and the applicability of laws frequently arise. The Federal government has an obligation to work with other countries to develop compatible cyberspace legal structures and to foster worldwide cooperation among law enforcement agencies.

Similarly, cyberspace threats to the infrastructure are challenging existing boundaries between the national defense, intelligence, law enforcement, and regulatory roles of the U.S. government. Clarification of missions, responsibilities, and authorities in this new context are needed, and will necessarily involve all three branches of government, Executive, Legislative, and Judicial.

Throughout the policy realm, the sustained engagement of an informed public will ensure that the choices made will best serve the national interest.

The technical dimension

The information networks which support the domestic infrastructure are complicated, evolving systems. Unlike Minerva, the mythological goddess who sprang fully grown from the head of Jupiter, they are not generally the product of a single top-down design. Rather, they have developed over time and continue to grow and change with the addition of new technology, new features, and new capabilities. This section identifies some of the main technical challenges inherent in networks of this type, and presents a conceptual framework for considering possible failure scenarios.

Most of today's cybernetic networks are actually combinations of networks, interconnected and interdependent. Interactions among these subsystems are critical to overall network performance, indeed they are the essence of network performance. Because the system also interacts with the real-world environment, the interactions among subsystems are not necessarily predictable and sequential, like the steps of an assembly line process, but can be essentially random, unsychronized, and even unanticipated. Many transactions are generated automatically by computers pursuing the logic with which they are programmed. The term "intelligent" is sometimes used to describe these networks, and it can be a fitting characterization, given the relatively autonomous nature of control systems.

Computer controlled subsystems often have little tolerance for variations – in sequence, content, or timing, for example – in the transactions they undertake with other subsystems. Small margins, like highway tailgaters, are vulnerable to dangerous chain reactions. Subsystems which have small margins – often called "tightly coupled" subsystems – are a reality within complex computer networks.

By their nature, infrastructure cybernetic systems must operate continuously, controlling physical

components in real-time and ultimately serving human customers. They must deal with essentially random simultaneous inputs from a great many sources – countless telephone calls and personal financial transactions, the energy usage of legions of subscribers, the flight paths of thousands of aircraft, as well as the control inputs of managers throughout the networks. Control systems, optimized for statistically-important usage profiles, must also be able to handle unusual and even unlikely inputs. Telephone companies, for example, are well familiar with (and prepared for) the "Mother's Day" phenomenon, in which predictable usage spikes are prompted by national events. Nevertheless, the full range and diversity of input conditions is difficult to define before-hand. Experience teaches that the problem of unanticipated input conditions is formidable when designing for high reliability.

Invariably, human beings are key elements of infrastructure cybernetic systems. Nothing can replace human judgment, and no network is designed to be completely "hands off." However, the possibility for human error in the operation of complex systems is ever-present. The complexity and speed of interaction among network elements can easily exceed the ability of operators to assess and respond to problems. The alerts, indications, and displays available, along with the input actions allowed and the time available to make them, will frame any human intervention. Human beings also often provide the linkage path between two otherwise independent subsystems, creating unexpected feedback paths and opening up new possibilities for unanticipated subsystem interactions. Finding the right balance between human and machine control is a technical problem that goes to the heart of the network reliability challenge.

The physical integrity of hardware components and communications links is a basic requirement for a reliable network. Exposure to the elements, continuous duty, and simple aging impose unavoidable stresses on infrastructure hardware. The cybernetic system must ultimately interface with electromechanical devices – the switches, relays, valves, and motors that make the infrastructure function – and must be robust enough to accommodate performance variations in these components and continue to function despite degradation.

One of the motivations for employing information networks in infrastructure control systems in the first place is to make the control systems more accessible and responsive to management inputs. However, accessibility once afforded is not easy to constrain. For example, many networks are migrating to common technologies – especially Internet technology (never designed to be highly secure) – for reasons of cost and efficiency. The more common a networking technology becomes, the more widely known and exploitable are its weaknesses. Additionally, the move to fewer, more standardized configurations simplifies a potential attacker's problem. The use of public network as bridges between internal corporate networks is also becoming increasingly common. Lacking security discipline, employees sometimes place unauthorized and unprotected dial-up modems on internal networks, creating potentially serious and unrecognized vulnerabilities. Deregulation, particularly in telecommunications and energy, allows new entrants to legitimately gain access to control networks that were previously proprietary or carefully protected. Whether by design or not, infrastructure control systems are increasingly accessible to outsiders.

Designers must therefore be concerned about the full range of possible intrusions into infrastructure cybernetic systems. The potential perpetrators of such intrusions include the cyberspace equivalent of the graffiti artist as well as the hardened cyber-terrorist. Hackers are growing in number and sophistication and have increasingly advanced tools. They are also organized, freely exchanging tools, techniques, and information on vulnerabilities worldwide through the Internet. Deliberate measures to protect against this threat are a necessity, but its

specific measure-countermeasure nature makes development of broadly applicable defenses extremely difficult. It is usually necessary to find specific defenses against specific attacks. These defenses, in turn, become targets for future attack. Presently there is much about this threat that is not known.

Finally, infrastructure information networks are inherently dependent on software. Ensuring the reliability of software-based systems is among the most difficult of engineering challenges. The cost of exhaustive testing to validate software intended for complex real-time environments often proves to be prohibitive, when it is even technically possible. Complicating this challenge is the fact that companies today are increasingly contracting with others, often in other countries, for the development of software. Such "outsourcing" can leave system integrators with little insight into the development and validation of critical control software. Commercial off-the-shelf technology is also being adopted more frequently, even though software design details are usually not available and reliability may be uncertain. Long term maintenance of software is made difficult by changing preferences in programming languages and lack of support tools for obsolete or orphaned systems. Generational differences in aging equipment can give rise to insidious software incompatibilities. Competition can pressure developers to rush software to market without sufficient testing. It is even possible for malicious code, deliberately and surreptitiously included in critical software during production, to go undetected in installation. Additionally, every software performance enhancement carries the possibility of introducing logical errors, undoing previous algorithm corrections, changing software timing performance, and even introducing coding errors, all of which can increase system vulnerabilities. A misplaced bit in a data structure or an almost-correct algorithm may work most of the time, but when these errors manifest themselves, the result can be a dramatic transition from normalcy to catastrophe.

An approach that seeks to manage risks by weighing cost, performance, and reliability tradeoffs along with the likelihood and severity of specific threats is the most sensible.

The information systems that have brought automation to the infrastructure are inherently complex. Complexity by itself, however, is not a good indicator of reliability. Human beings are capable of producing very complicated creations that are nevertheless quite reliable, such as jet engines, microchips, skyscrapers, and pharmaceuticals, to cite just a few examples. For information networks, complexity might make a network fragile and prone to failure, or it might be a source of robustness. In any case, these terms are relative and not precise enough for making engineering or policy judgments. The susceptibility of a network to major disruptions can only be gauged by carefully assessing a great many technical and operational factors in the context of the total threat environment. In the end, an approach that seeks to manage risks by weighing cost, performance, and reliability tradeoffs along with the likelihood and severity of specific threats is the most sensible.

HISTORICAL EXPERIENCE

Over the years, there have been many incidents in which network-related faults caused infrastructure problems, including these prominent representative cases:

- A widespread electrical power blackout affected 15 States in the western United States, as well as some regions of Canada and Mexico, in July 1996. The outages affected some two million people, and caused airport delays and subway breakdowns from Denver to San Francisco. The cause of the outage was traced to an over-

heated 500,000-volt transmission line in northern Oregon which sagged into tall trees, short circuited, and shut down. Two other 500,000-volt lines subsequently became overloaded and shut down, followed shortly thereafter by shutdown of the Pacific Intertie, the main power artery between the Northwest and California. The disturbance rippled as safety systems automatically shed load to try to keep the system in balance. Increased power demand caused generators in California to shut down; generators in other regions shut down because they suddenly had too much power and nowhere to send it. A similar blackout affecting the same region was experienced the following month.

- A series of breakdowns that disrupted local telephone service for some 16 million customers in Los Angeles, Baltimore, San Francisco, and Pittsburgh in June and July 1991 was attributed to a defect in a few lines of computer code in critical algorithms of the signaling system. The manufacturer traced the problem to a recent upgrade in its software which had not been put through its customary thorough testing because the change entailed only a few lines of new code.

- In September 1991, an internal power failure at a Manhattan telephone switching center cut off approximately half of the long distance traffic of the nation's largest long distance carrier into and out of New York City. This incident had a particularly serious impact on air traffic because the affected switching center also carried some 90% of the communications of the New York air traffic control center. Although no aircraft accidents were attributed to the outage, about 400 flights were canceled at the three major New York airports and tens of thousands of passengers were inconvenienced over an eight-hour period. The outage was blamed on "a combination of equipment failure and human failure." Under an agreement with the local power company, the telephone company had adopted the practice of turning off city power and relying on its own generators in periods of high electrical demand. In this event, however, workers failed to follow established procedures and confirm proper operation of the generators. Unluckily, failed rectifiers prevented the generators from delivering power, leaving the switching system to draw power entirely from its backup batteries. Alarm bells and warning lights went unheeded for six hours and the batteries became depleted.

- In July 1994, a software upgrade to the computers of the over-the-counter Nasdaq marketplace caused that system to go down for over two hours, cutting volume for the day by about one third, and affecting stock exchanges, trading desks, and stock-index mutual funds throughout the country. Nasdaq, a stock exchange with no trading floor, relies on a nationwide computer network for a trading volume of hundreds of millions of shares daily. The software problems were manifested directly on the mainframe computers located in Connecticut. The backup system in Rockville, Maryland, which was being upgraded at the same time to maintain compatibility, also failed.

- The Northridge, California earthquake of 1994 caused long distance telephone service outages for about two million people for approximately eight hours as two major switching facilities at Sherman Oaks failed. Thirty-five cellular sites were also out of service. However, while the earthquake damaged many telephone exchange buildings, most continued in operation. Customers unable to access long distance service still had dial tone and could call numbers within their local dialing area including local emergency response organizations.

- In January 1990, a piece of interface equipment in one of the telephone toll switching systems of the nation's largest

long-distance carrier in New York City developed a minor hardware problem. The control software, which had recently been upgraded network-wide, entered its fault recovery routine, suspending new call processing briefly. However, a flaw in the software effectively prevented the switch from coming back into service and disabled backups in the process. The problem cascaded through the network with switches throughout the system also going out of service. The result was blockage of some fifty percent of all switched traffic for that carrier nationwide for a period of seven hours. Of approximately 148 million call attempts made, 83 million were completed. Ironically, the software that caused the problem was intended to speed restoration of call processing after suspension.

- In 1965, and again in 1977, the Northeastern United States experienced massive and costly electrical power outages. In both cases, the problem came about due to a cascading series of events in which operators and automated components followed the logic with which they were trained or programmed, shutting down or disconnecting generators as a protective measure in response to anomalous conditions.

This small sampling of the historical evidence shows that there have been many "major" interruptions of infrastructure services. Nevertheless, on almost any meaningful scale, the automated infrastructure of the United States has been highly reliable. In recent decades, few infrastructure disruptions have had large-scale effects on the population at the national level. Disruptions and outages that have occurred have generally been selective, affecting subscribers by region, for example, or by choice of provider, or by some other discriminator particular to the situation. When outages have threatened the safety of large numbers of people, such as when electrical power has been interrupted during major winter storms, the conditions created were addressable by local emergency services, government disaster assistance programs, and the dedicated emergency crews of the utility companies. For most Americans, infrastructure disruptions have been more a nuisance than a nightmare.

In some cases, small problems have snowballed into major disruptions. In others, the abject failure or destruction of key components brought about major service interruptions directly.

Besides illustrating the range of network-related infrastructure problems, the examples cited here also underscore the critical importance of the mutual dependencies among infrastructure sectors. Telecommunications, energy, transportation, and finance are all bound together – literally as well as figuratively given that they often depend on the same fiber optic bundle. It is also important to realize that each of these incidents, and many others like them, prompted remedial actions – engineering, procedural, and policy changes – to help avoid their recurrence. This provides a significant institutional legacy on which to build for the future protection of the infrastructure.

Nothing guarantees that future disruptions will be similarly limited in national impact as past disruptions. However, past experience certainly does provide insight into *how* networks are likely to fail in the future. These examples show that in some cases, small problems have snowballed into major disruptions. In others, the abject failure or destruction of key components brought about major service interruptions directly.

The propagating "chain reaction" failure mechanism is characteristic of complex systems with tightly coupled subsystems. This mechanism depends on the dynamics of the system itself, in that seemingly inconsequential events trigger multiple failures through an unanticipated domino effect. Further, the

problems can be exacerbated by the very features and procedures intended to protect against failures. For this failure mechanism, the real problem is not the triggering event itself, but the interaction of anomalous operating modes among subsystems that it sets in motion. Like Mrs. O'Leary's cow and the Great Chicago Fire, the event which triggers a cascading system-wide catastrophe may have no intended nor apparent connection with the ultimate outcome.

On almost any meaningful scale, the automated infrastructure of the United States has been highly reliable. For most Americans, infrastructure disruptions have been more a nuisance than a nightmare.

Not all failures are the result of chain reactions, however. The direct, independent failure of certain components, or combinations of components, can result in the same level of disruption. Natural disasters are the most familiar cause of this type of network problem, but systemic design flaws can have the same apparent result.

For actual events, the failure mechanism depends both on properties of the system in question and on the particular circumstances of the incident. The two broad mechanisms described here represent the conceptual extremes. In real life, failures usually have characteristics of each and fall somewhere between the two.

MALICIOUS ATTACKS

Albert Einstein once observed, "The Lord God is subtle, but malicious he is not." The same cannot be said of man. To fully understand network reliability challenges, it is therefore necessary to also consider the possibility of deliberate attacks.

Although the illicit penetration of computer systems is an increasingly familiar occurrence, deliberate acts of computer intrusion or sabotage are most often associated with the Internet rather than with infrastructure cybernetic systems. This distinction, blurred even now, will be less clear in the future. Further, as personal on-line computer access is further woven into the fabric of society, it too will one day be considered an essential infrastructure service.

To date, although there is little historical experience in which cyber attacks of any kind caused serious disruptions of infrastructure services, experience that has accrued with computer hackers is helpful in understanding how this threat might be manifested.

- In 1988, the Internet "worm," a computer program designed to consume the memory and resources of computers, was deliberately released on the Internet. Thousands of computers were affected before the worm was brought under control. This incident is representative of a large class of attacks which involve the introduction of self-replicating malicious code. By their nature, such intrusions are not selective in their destructive effect.

- In 1994, more than 150 intrusions were made to the Air Force's Rome Laboratory by two hackers using specialized software that allowed their intrusions to masquerade as legitimate transactions. The attackers were able to seize control of Rome's support systems for several days, establish links to foreign Internet sites, copy and download critical data, and successfully attack systems at other government facilities, defense contractors, and private sector organizations. The Air Force, which did not even recognize the attack for at least three days, estimated the cost to the government at over $500,000, not including the value of the information that was stolen. This class of hacker intrusion is characterized by the use of sophisticated tools and techniques to seize control of a network and take actions normally

reserved for trusted system managers. A spate of intrusions to government World Wide Web sites in 1996 serve as another example of this class of attack.

- In 1996, several Internet service providers were victims of deliberate attacks from unidentified sources using sophisticated software to overload servers with hundreds of messages per second. The messages, called synchronization requests, contained false return addresses, which confused and rapidly tied up the servers and rendered them unable to handle legitimate transactions.

- The World Trade Center bombing in New York City in 1993, and the Oklahoma City bombing in 1995, although not attacks on infrastructure *per se*, reveal the potential for infrastructure disruption that domestic terrorism holds. Destructive physical attacks on network components could cripple an infrastructure sector's cybernetic system and cause major service interruptions that would be difficult to alleviate.

The wide deployment of information technology potentially puts destructive capability, once the province of nations, into the hands of individuals.

These examples, by no means exhaustive, give an idea of the broad classes of malicious attacks that could be mounted in cyberspace. At least three distinct types of cyber attack can be postulated which could result in serious infrastructure disruptions:

Computer hackers. Small-scale, even unsophisticated, intrusions into information networks by "cyberspace joyriders" could induce major network problems if they altered or destroyed data, overloaded input circuits, or caused locally degraded operations. An intruder in this category is motivated by curiosity, technical challenge, mischief-making, or the aim of stealing services, but could nevertheless trigger a cascading failure with widespread effect.

Anarchist attacks. An attack on infrastructure cybernetic elements could be mounted to help achieve a broader criminal purpose, or by anarchists who may not have a clear objective or a cogent strategy other than to disrupt and destroy. The perpetrator in this category is a purposeful actor intent on doing damage, but who has probably not attempted a careful assessment of the precise effects the attack would have. Attacks on critical network components such as communications links, control nodes, or switching stations could render equipment inoperative either through physical damage, or through corruption of software or data. In this case, the affected component could itself be important enough that disabling it could cause a major disruption, or it could initiate a cascading failure that results in a major disruption.

Coordinated cyber attacks. Attacks in this category are focused, organized, and carefully calculated to yield a specific outcome. The perpetrator of this type of attack is motivated by strategic political goals, and may employ the same types of tools used by the anarchist described above, although in a more sophisticated manner. Destructive Trojan horses or "logic bombs" could conceivably be placed in operating systems software to induce a systemic failure. An insider could use specialized knowledge to maliciously attack a network and induce major disruptions through normal management controls. A coordinated attack on multiple vital components or communications links could directly disable the cybernetic system without the dynamic effect of cascading subsystem failures. Since this type of attack is carefully planned, it is more likely to employ tools with direct, decisive outcomes than to rely on a relatively unpredictable chain reaction.

Attacks in each of these three categories could cause the same level of disruption, the

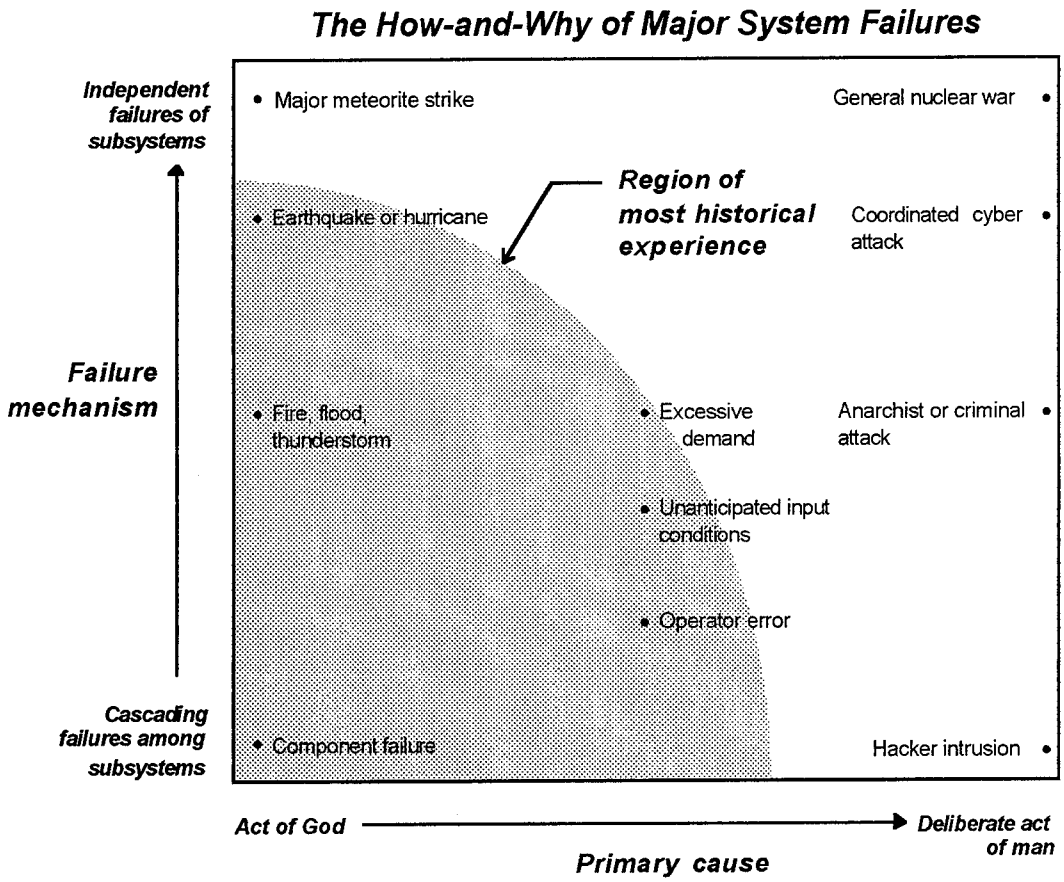

Figure 1. The failure scenarios chart. *This chart notionally depicts some possible scenarios in which major infrastructure disruptions are brought about by events affecting the supporting cybernetic system. It illustrates the* ***how*** *(failure mechanism) and the* ***why*** *(primary cause) of major system failures, and can be a useful tool for placing hypothetical scenarios in context. The specific details of a scenario, as well as properties of the particular network, determine the precise placement of incidents on this chart.*

differences being in the motives and methods of the perpetrators and the mechanisms by which the failure is manifested. The wide deployment of information technology potentially puts destructive capability, once the province of nations, into the hands of individuals. At the same time, the growth and increasing interconnectedness of computer networks offers additional possibilities for outsiders to break into what had previously been closed, internal systems. Increasingly advanced hackers and the proliferation of sophisticated tools are forewarning that any of these scenarios could be dangerously real.

A UNIFIED FRAMEWORK

The engineering challenge of building highly reliable networks is extremely broad. *Causes* of network failures cover the complete range from Mother Nature to human nature – from natural occurrences to the deliberate actions of hostile persons. Failure *mechanisms* also vary. Tightly coupled systems often fail through a cascading process in which a subsystem failure propagates in a chain reaction. In other cases, the independent failure of critical components is the dominant mechanism.

Figure 1 puts these two variables together in one framework that captures the *how* and the *why* of major network failures. This framework – conceptual rather than quantitative – is useful for comparing actual events with hypothetical scenarios, postulating and assessing scenarios to uncover potential vulnerabilities, and analyzing the risks and benefits of strategies for dealing with reliability deficiencies. Figure 1 makes the essential point that while challenges to reliability exist throughout the total space, historical experience to date is concentrated in one corner of it. Much about the network reliability challenge has not been experienced.

All of the events depicted by the failure scenarios chart represent network failures of potentially *equal severity*. An important discriminator among scenarios is their *likelihood*. Two probabilities need to be considered: the probability that the event will occur, and the probability, assuming the event did occur, that it will result in a major failure. The overall probability, which is the product of the two, is an indicator of how seriously the threat should be taken.

For example, component failures are a certainty, but the probability that the loss of a single transformer, switching device, or sensor, for example, would trigger a chain reaction that disrupts a major portion of the network (although it does happen) is usually quite small. A meteorite strike, on the other hand, is a very unlikely event, but if it were to occur, there would be a high probability that it would destroy and disable infrastructure networks. This type of analysis can aid in understanding the relative priority of reliability threats.

There are many unanswered questions about how to characterize the cyber threat. Hacker intrusions are themselves an accepted reality, but the probability that a hacker could initiate a chain reaction that causes significant disruption of a network, even inadvertently, needs careful, detailed analysis. The probability of a coordinated cyber attack, and the probability that it would be effective if launched, also need examination. Producing a specific desired outcome by intruding into a complex network might be as sure as setting the hands of a precision timepiece, or it could be like expecting a Rube Goldberg invention to work as designed. A careful analysis of specific cases can shed light on vulnerabilities and can help identify the actors with the greatest chance of success. An effective response to cyber attack depends on understanding the specific attack methods, probabilities, and network failure modes.

Finding solutions

Addressing all the reliability challenges captured by the failure scenarios chart calls for a comprehensive technical methodology. The basic steps of such a technical approach include: first, developing an analytical understanding of the existing reliability, vulnerability, and threat environment; second, establishing a system engineering process which treats reliability as a primary parameter; and third, fostering a commitment to vigilance and a process of continual learning to enhance reliability. These steps, listed in Table 1, are discussed separately in the sections that follow.

A Technical Agenda for Network Reliability

1. *Develop an analytical understanding of the specific reliability, vulnerability, and threat environment.*
2. *Establish a reliability engineering process.*
3. *Maintain constant vigilance and continual learning to enhance reliability.*

Table 1. *These three steps form the basis of a technical agenda for network reliability.*

STEP 1: UNDERSTAND THE RELIABILITY ENVIRONMENT

Two metaphors give insight into the network reliability problem. On one hand, a network can be compared to a house of cards for which mutual dependencies are so great that removing one component can cause the whole precarious structure to collapse. On the other hand, a network can be compared to a chain-link fence which, although also made up of interdependent parts, is flexible and robust over a wide range of inputs. A breach in one section of the mesh does not cause complete failure of the fence. And once breached, a fence can be repaired – it does not need to be reconstructed in its entirety.

Neither of these metaphors is perfect, but both are illuminating. Clearly the chain-link fence metaphor is the more desirable of the two. But whether or not it applies in a given infrastructure sector is largely a function of the engineering and operational decisions made in the course of the network's life. Fundamentally, reliability is about the extent to which the operation of individual components affects the overall performance of the system. Robust systems are robust precisely because overall system operation is tolerant of component and subsystem faults. Since information networks already exist in every infrastructure sector, the first step is to characterize these networks and develop a full appreciation of their strengths and weaknesses. This can be aided by a thorough assessment in the following areas:

Since information networks already exist in every infrastructure sector, the first step is to characterize these networks and develop a full appreciation of their strengths and weaknesses.

Operational concept of the infrastructure sector. A detailed description of what the information network is used for, including a cataloging of the options for external entry, ability to execute commands remotely, and the nature and range of computer control available, is essential. How the system responds to failures, disruptions, or data corruption is also critically important information.

Network architecture and information flows. A detailed technical description of the network is needed, including the physical and logical layout, the flow of information, and the major nodes and interconnections. Particular attention should be given to how subsystems at all levels interact with each other under normal and degraded conditions, and how tolerant they are of faults, delays, malfunctions, or errors in other subsystems. This analysis would reveal how tight the coupling is among network subsystems. Formal mathematical methods, computer modeling and simulation, and transaction analysis are useful tools for describing and analyzing architectures and information flows.

Network components. Operating limitations or design flaws in supervisory control and data acquisition components, gateways, firewalls, routers, servers, or other critical nodes – especially those controlled by software – could constitute weaknesses that undermine overall network reliability. All components in the system should be examined from a reliability perspective for the applications in which they are used. Migration to common components and off-the-shelf equipment also potentially increases the likelihood of exploitable security weaknesses.

Signal protocols and transmission methods. Signals of interest can range from simple analog signals on dedicated circuits to complex digital protocols multiplexed onto high capacity data channels including fiber optic cable and satellite links. Signals may or may not be encrypted for transmission, and may be susceptible in varying degrees to monitoring, interception, interference, spoofing, or jamming. The design choices made have direct bearing on the network's vulnerability to physical and electronic disruption.

Human factors. People are at the heart of virtually all information networks, as system managers, operators, engineers, and technicians. The human interface – through controls, displays, alert and warning indications, and equipment layout – brings human judgment into the control loop. Operator carelessness, inattention, or procedural error are ever-present hazards. Well-intentioned workarounds of established procedures and system configurations can undermine reliability. The personal reliability of key people is critical to reliability of the network. A complete understanding of the

human factors environment is vital to an overall assessment of system vulnerabilities.

Existing security environment. Reliability is heavily influenced by the existing network security environment. An understanding of the priorities and tradeoffs that led to the current configuration is necessary. There may be known vulnerabilities or methods of intrusion for which security measures have been designed. The security of password files, access to supervisory features, the integrity of access logs, and the ability of a system administrator to detect intrusions are all relevant factors. Security cannot be considered separately from analysis of the threat environment. Tools and techniques available to hostile intruders must be understood if they are to be effectively countered. Security provisions in particular are very dependent on implementation factors. For this reason, security discipline in all aspects of network operation should be assessed.

When compiled and combined with operational experience, the findings in these areas make up a comprehensive network reliability data base that can serve as the foundation of a sound system engineering process. Such compilations, however, should be carefully safeguarded to prevent access by those who would misuse such information to exploit system weaknesses.

STEP 2: ESTABLISH A RELIABILITY ENGINEERING PROCESS

In an ideal world, an information network could be designed "right" from the start and it would then function "correctly". However, in the real world, networks evolve and change with time and technology, often growing more by accretion than by design. The demands of the marketplace also evolve. In this dynamic environment, a fixed design is not likely to be sustained in the long run. The "perfect" network unfortunately cannot exist – there is a constant need to engineer solutions for specific problems.

Design principles

From the standpoint of reliability, three design principles should guide this ongoing engineering process:

Reduce the possibility of disruptions occurring. This "prevent defense" principle focuses on the static and external features of the design. It suggests strategies aimed at minimizing the possibility of component and subsystem failures, such as controlling the system's environment, thoroughly validating components, physical and cyber security, and personnel practices. It also fosters the recognition and avoidance of systemic weaknesses and single point failures.

Minimize the effect of disruptions which do occur. This principle emphasizes the dynamics of failure within a network. It focuses on the details of cascading failures, the interaction of subsystems and their failure modes, the detection of anomalies, and system tolerance to degraded components. In real-world occurrences, cascading network failures ultimately are contained at some level. An appreciation for "what stops the snowball," that is, what constrains cascading failures, gleaned from operational experience, may give valuable design insights, and should also be encouraged.

Design for efficient recovery. This principle focuses on operations and highlights the importance of recovery procedures, equipment re-initialization sequences, and the human interface in the aftermath of a disruption in which backups fail, prove inadequate, or become irrelevant. Providing the ability to restore normal operations from any of a huge theoretical number of failure states requires concerted attention at all stages of the engineering process. Recovery operations also present the greatest need to improvise; anticipating this need may suggest design features that do not affect design integrity but that could make a significant difference if reconstituting ever became necessary. For example, in the aftermath of the 1965 northeast power blackout, a US Navy destroyer, the USS

Bristol, sailed from the Brooklyn Naval Shipyard and provided ship's power to one of the electric company's substations. It is doubtful that this particular need could have been foreseen in its details, but it is not unrealistic to ask designers to consider the general need for expedient or provisional repairs.

Design tradeoff studies

In implementing these design principles, strategies relevant to each region of the failure scenarios chart of Figure 1 are pursued. For example, to address cascading failures, increasing the margins among interacting subsystems so as to reduce the possibilities for unsuccessful transactions may be appropriate. To reduce the possibility that independent failures would result in major disruptions, measures such as distributing key components geographically, adding redundancy, and diversifying the operating software may all be worthwhile. Improving physical security or strengthening barriers against network intrusions could help prevent failures caused by deliberate attacks.

Like the interactions of prescription drugs, the remedy for one problem can interfere with the remedy for another.

However, many strategies, useful singly, are at odds with each other and with network performance objectives. For example, diversifying the software by employing subsystems with different operating systems may avoid certain single-point failures, but it may also impose serious cost and performance penalties. Furthermore, protection features however well-intentioned may introduce additional vulnerabilities and new failure modes.

To illustrate, Figure 2 depicts a single example strategy (increasing component redundancy) overlaid on the failure scenarios chart. This strategy, pursued to reduce the effect of independent failures of critical components, may actually be harmful in other scenarios because it increases subsystem interactions and provides additional paths by which a failure can propagate. Deploying the redundant components may also create more intrusion options, thus exacerbating physical security concerns.

The crucial point is that any strategy, if pursued in isolation, may actually reduce overall reliability rather than enhance it. A focus on one type of threat or one region of the failure scenarios chart can increase the vulnerability to, and the effects of, another. Like the interactions of prescription drugs, the remedy for one problem can interfere with the remedy for another. Clearly, a holistic methodology for making the unavoidable tradeoffs is called for. The cost, performance, and reliability effects of technical options should all be weighed in the context of the likelihood and severity of the threat. Table 2 lists some of the questions which design tradeoff studies should address, including whether a proposed solution will work, whether it can be implemented, what the marketplace effects are, and how the national interest is met.

Design tradeoff studies are the centerpiece of the system engineering process. It is critically important that reliability be included along with such principal tradeoff parameters as cost and performance when any engineering change is considered.

It is not erroneous to talk of a system engineering process for networks that have no single owner who possesses complete configuration authority. The prerequisite is that there be a firm foundation of technical standards and practices so that decisions made at what is inevitably the subsystem level are not far from optimum for the total system. Industry coordinating committees, national and international standards organizations, and technical dialogue are all essential for getting designers, equipment manufacturers, service providers, and installers on the same page of the textbook.

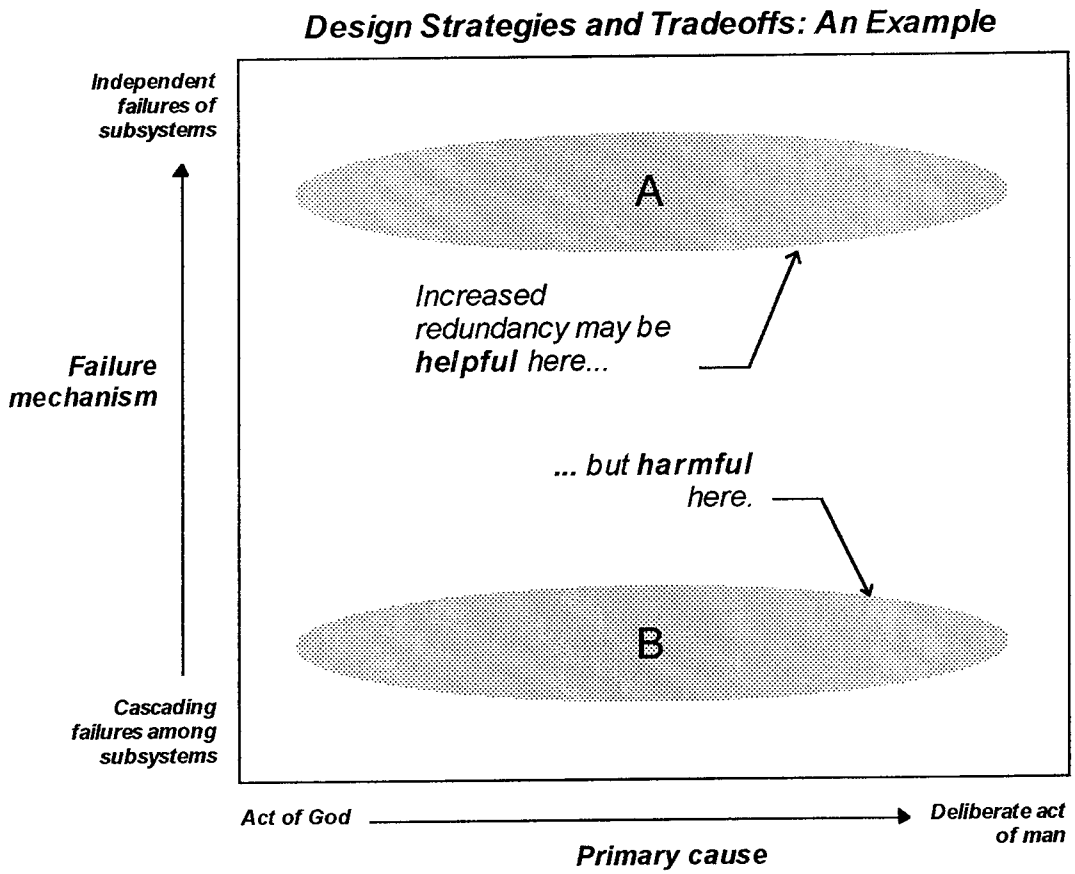

Figure 2. An example of the need for tradeoffs. *This chart illustrates the challenge of reliability engineering. In this example, a strategy for enhancing reliability that is helpful for scenarios which fall into region A might be harmful for scenarios in region B. Additionally, it might conflict with another strategy, impose cost or performance penalties, or introduce additional vulnerabilities. A methodology for making informed analysis and tradeoffs is essential.*

STEP 3: MAINTAIN CONSTANT VIGILANCE AND CONTINUAL LEARNING

The third essential step of the technical agenda is to foster a commitment to vigilance and a culture of continual learning to enhance reliability. Although this is easy to agree with in principle, it is often neglected in practice.

Specific needs include:

Early warning of hostile activities. Early warning of potential and actual hostile activity directed at the public infrastructure would allow both the service providers and the law enforcement community to take measures to prevent or minimize the disruption. An equitable, institutional means, within clear statutory limits, for the timely two-way flow of

Assessing Technical Options for Improving Network Reliability

Effectiveness
How well does the proposed solution address the region of interest in the failure scenarios chart?
What effects does the proposed solution have in regions of the failure scenarios chart other than the region for which it is designed?
Is the proposed solution practical? available? usable?
What effect does the proposed solution have on network performance? on reconstitution after failures?
What new vulnerabilities does the proposed solution introduce?
Will the proposed solution work in the long run?
Implementation approach
What legal, regulatory, or other regime is necessary to implement this solution?
What is the schedule for implementation?
How will the proposed solution figure in the long term management of the system?
Marketplace effects, costs, and benefits
Do the likelihood and severity of the problem justify the cost of solution?
Who bears the cost of adopting the proposed solution?
To what extent does the effectiveness of the proposed solution depend on its bringing about changes in the marketplace?
Will changes in the marketplace negate the effectiveness of the proposed solution?
Government and the national interest
Will its political implementation negate the effectiveness of the proposed solution?
How, if at all, does the government interact with the private sector to implement, operate, and administer the proposed solution?
Does the proposed solution affect government national security or emergency preparedness services?

Table 2. *This table lists some of the key questions relevant to assessing proposed solutions to network reliability challenges. These questions should be addressed in system design and tradeoff studies.*

relevant intelligence information and incident data between government and the public utilities, which protects business-sensitive data as well as sources and methods, would do much to clarify the threat environment and allow for an effective response. Some threats, particularly cyber attacks, may be diffuse and only identifiable when data from a number of sources is aggregated. Numerous small and widely distributed anomalies that escape notice individually may, when compiled and correlated, be indicators of a systemic problem. In these cases, infrastructure network control centers could be important sources of strategic warning for the same reasons that global television news networks have become *de facto* sources of intelligence – on-scene presence and direct observation of real events. What is lacking, however, are the technical and institutional means to synthesize data from many sources and draw meaningful and timely conclusions. This is an area in which government and industry can work together for the benefit of all.

Infrastructure network control centers could be important sources of strategic warning for the same reasons that global television news networks have become sources of intelligence – on-scene presence and direct observation of real events.

Tools to detect and characterize network anomalies as they occur. At the operational level, network intrusions are difficult to detect because they can masquerade as legitimate transactions or go unnoticed in a busy network. In many networks today, successful intrusions are more likely to be detected by their effects rather than by any discernible telltale signature. Trustworthy tools for detecting anomalies early and judging their seriousness would help enable system administrators to take corrective action before major problems develop.

Management controls for reacting to disruptions. A range of system management controls including, for example, such measures as reconfiguration of the network, increased security in response to intrusion attempts, or decreased functionality to limit range of control can contain disruptions and prevent them from worsening. A means for selecting an appropriate response from among such control measures in an actual event is also needed.

Effective internal reporting procedures. Although self-monitoring, self-healing information networks may be on the horizon, ultimately the responsibility for recognizing problems and doing something about them rests with people. Effective internal reporting procedures need to be developed, implemented, and followed within the companies that operate infrastructure systems. Some industry-wide uniformity in reliability reporting, similar in concept to the standards of financial accounting and reporting established by the Financial Accounting Standards Board, can help put discipline into the process of capturing and applying lessons learned.

Mechanisms for sharing reliability information among competitors. Private sector companies involved in providing infrastructure services should be able learn from each other. Mechanisms for sharing information – within legal limits – on vulnerabilities, incident data, technical solutions, and best practices among network managers which protect the confidentiality of proprietary or business-sensitive information are urgently needed. The Network Security Information Exchanges in the telecommunications sector are prominent examples of forums for intercompany sharing within carefully delimited boundaries. Such sharing is also important among sectors, especially where common networking technology is being employed.

Reliability certification. Legitimized, generally accepted, and broadly applicable methods for certifying levels of reliability in components and networks would directly enhance reliability by enabling companies to make more informed tradeoffs. They would also help foster a culture of reliability throughout the technical community. Scientifically-based consensus codes and standards which address network design and installation – as the National Fire Protection Agency's *National Electrical Code* addresses electrical practices – would lessen the danger of unreliable systems. Institutionalized product reliability testing and certification, comparable with the broader safety testing and certification that Underwriters Laboratories performs, would be of immense value in establishing a common yardstick by which products could be compared.

Reliability is ultimately a human enterprise and in most practical situations, human-machine interactions are part-and-parcel of the system.

Measures to strengthen human factors. Reliability is ultimately a human enterprise and in most practical situations, human-machine interactions are part-and-parcel of the system. Measures to ensure operators continually maintain situation awareness and other personal reliability standards need to be stressed at every point.

There are no simple solutions to reliability challenges. There are no permanent solutions. However, the three steps outlined in this section, if pursued with sustained commitment, represent a sensible approach to assuring that the dependability of the domestic infrastructure matches the importance placed upon it.

Conclusion

The automation – or cybernation – of the domestic infrastructure has been motivated largely by the cost, efficiency, and performance benefits offered by information technology. Because reliability is linked to profitability, the private sector will continue responding to credible reliability threats in the future. Today's concern is whether the marketplace will adequately *anticipate* and *mitigate* reliability deficiencies in a highly dynamic business and threat environment, or whether the nation will have to endure a major infrastructure problem in order to mobilize and act. The importance of infrastructure services to the well being of the nation make infrastructure reliability a public policy concern.

Threats to reliability come from both natural and manmade sources, from Mother Nature to human nature. From a technical standpoint, it is not practical to focus exclusively on any one reliability threat. Strategies to counter one threat can exacerbate vulnerabilities to another. Like the interactions of prescription drugs, the remedy for one problem can interfere with the remedy for another. A holistic methodology for making the unavoidable tradeoffs is called for.

Reliable networks, like scientific inquiry, must be based on real data and actual facts – facts about the nature of vulnerabilities, the evolving reliability challenges, and the real-world, real-time environment in which infrastructure information networks operate. There is no substitute for a reasoned, methodical approach to understanding this problem and seeking solutions to it. Cost, performance, and reliability objectives must all be balanced through an engineering process of analysis and informed tradeoffs.

The risks associated with infrastructure threats can never be eliminated entirely. A sound

technical approach is one which recognizes the need to manage risks and keep them at societally acceptable levels. A technical agenda for comprehensively addressing the reliability problem consists of three steps: (1) developing an analytical understanding of the existing reliability, vulnerability, and threat environment; (2) establishing a system engineering process which treats reliability as a primary parameter; and (3) fostering a commitment to vigilance and a process of continual learning to enhance reliability.

Areas for increased public policy attention include:

- *Achieving consensus on the problem and approaches to solutions.* The traditional policy tools available to the government for working with the marketplace to achieve national objectives – including legislation, regulation, licensing, tax and rate-setting regimes, and other inducements – all offer important options in the reliability arena. However, none of them can be effective unless and until there is consensus on what the minimum levels of reliability should be, what the threats are, what risks are acceptable, what protective measures should be taken, and how the costs should be met. At the present time, these questions are far from settled.

- *Enhancing government/industry cooperation for identifying and characterizing reliability challenges.* Government and industry each have unique capabilities to apply in identifying and characterizing reliability challenges. Federal activities, from weather and natural disaster prediction to intelligence collection on the threat of hostile attack, can contribute greatly to industry's understanding of reliability challenges. Additionally, industry network control centers could be important sources of strategic warning if the technical and institutional means to synthesize data from many sources were in place. Identifying appropriate activities in these areas, and putting them into practice, calls for long-term public policy attention.

- *Focusing government and industry on the joint development of technical standards and methods to measure and certify reliability.* Legitimized, generally accepted, and broadly applicable methods for certifying levels of reliability in components and networks would directly enhance reliability by enabling companies to make more informed tradeoffs. They would also help foster a culture of reliability throughout the technical community. Scientifically-based consensus codes and standards which address network design and installation would lessen the danger of unreliable systems. Institutionalized product reliability testing and certification would be of immense value in establishing a common yardstick by which products could be compared. The most progress in this area can be made through government and industry working together.

- *Enhancing Federal/State government interaction to ensure consistent and appropriate attention is place on infrastructure reliability.* Federal government interaction with State governments, and with State and regional regulatory commissions, is also essential to framing the question of how much reliability is needed in the infrastructure and to ensure that consistent and appropriate attention is placed on the reliability of the associated information networks. The Federal government can be helpful in providing uniform guidance to the States where appropriate, and in coordinating and focusing the resources of the Federal departments and agencies.

- *Defining the government research and development investment portfolio.* Federal investment in science and technology has been instrumental in the unfolding information revolution. Wise investments today in areas such as intrusion detection and pre-

vention, robust network architectures, configuration management, secure communications, and other critical areas will pay reliability dividends in the future.

- *Working with other countries to develop compatible international legal regimes in cyberspace.* Reliability is also served by laws that delineate the boundaries of permissible behavior. As an international arena that disregards national boundaries, cyberspace presents unique legal difficulties. Distance and geography do not impede the trespasser, bandit, or terrorist in this realm, and questions about jurisdictions, sovereignty, and the applicability of laws frequently arise. The Federal government has an obligation to work with other countries to develop compatible cyberspace legal structures and to foster worldwide cooperation among law enforcement agencies.

- *Clarifying missions, responsibilities, and authorities of Federal Departments and Agencies in cyberspace.* Threats to the infrastructure are challenging existing boundaries between the national defense, intelligence, law enforcement, and regulatory roles of the U.S. government. Clarification of missions, responsibilities, and authorities in this new context are needed, and will necessarily involve all three branches of government, Executive, Legislative, and Judicial.

Both government and the private sector have responsibilities for delivering dependable infrastructure services that can, at reasonable cost, withstand foreseeable challenges without interruption. Developing consensus on the problem as well as finding effective long term solutions will require the sustained engagement of industry, utilities, the public, and the government at all levels. Together all these stakeholders can assure the reliability of the most capable infrastructure in all of history.

For further information, contact:

Office of Science and Technology Policy
National Security and International Affairs Division: 202-456-2894
Internet: http:/www.whitehouse.gov