Availability Policies in an Adversarial Environment

Hilary H. Hosmer
Data Security, Inc.
58 Wilson Road
Bedford, MA 01730
(617) 275-8231
Hosmer@dockmaster.ncsc.mil

ABSTRACT

Availability of our information systems is crucial. Yet availability policies have received a disproportionately small share of INFOSEC attention. This paper explores current assumptions about availability and proposes a new availability policy paradigm operating in an adversarial cyberspace environment. In the proposed paradigm, threats are social as well as technical, and content availability is as important as system availability. Availability measures may take on negative values.

SECTION 1

INTRODUCTION

President Clinton in July of 1996 created a task force to study the preservation of our national infrastructure in the face of information warfare attacks on the computers that help manage it. At risk are the complex and distributed systems required for electric power, water, telecommunications, transportation, and banking. Availability is clearly a critical national issue.

What do we do to anticipate and forestall a major information warfare attack on our national infrastructure? What policies would be appropriate if the information system controls for a major airport, dam, or hospital are attacked? What if the entire telephone switching system is placed under siege so that no calls can be made?

Many threats to availability are well-known, ranging from natural disasters, like lightening, to malicious attacks, such as flooding a system with spurious messages or producing an electromagnetic pulse that wipes out all electromagnetically stored information. Yet INFOSEC research into availability and availability policies is disproportionately meager. This is due, I believe, to several misconceptions, including the notions that "preventing denial of service requires ensuring the complete functional correctness of a system - something unlikely to be done in the foreseeable future"[1] and that conflicting goals (e.g. availability and confidentiality) can not both be active.[2]

KEY POINTS

This paper makes six key points.

1. This is an appropriate time for a new availability paradigm.

2. A set of availability policies (or response scenarios) is usually necessary to handle normal, stressed, and catastrophic conditions.

3. Although availability is thought by many to be unfailingly desirable, (e.g. an airline system that is up 99.8% of the time, for example, is superior to one that is up 80.5% of the time.), in some situations *less* availability is better.

4. Availability is multifaceted and context-sensitive.

5. Balancing conflicting values is often necessary to integrate availability policies with policies to achieve other goals, such as confidentiality, safety, and survivability.

6. A more comprehensive availability paradigm is needed.

WHY NOW?

This is an appropriate time for a new availability paradigm because:

- The current paradigm is inadequate for the problems we have to solve. It won't scale upward or respond flexibly to the heterogeneous world we live in. Popular assumptions about availability, such as requiring complete provable correctness, may actually hinder progress.

- Risk management rather than perfect security has been proposed as the new DOD paradigm.[3] This allows more realistic expectations for availability and freedom from dependence upon "the complete functional correctness" of the underlying software and hardware.

- New technology that greatly impacts availability is now in place, including:
  Fault-tolerant hardware;
  The Internet and the World Wide Web;
  Intrusion detection and vulnerability detection software.

- All of our infrastructure is becoming so computerized, common cross-industry patterns and assumptions are emerging.

SECTION 2

AVAILABILITY IN NORMAL, STRESSED, AND CATASTROPHIC SITUATIONS

The real-world scenarios below challenge some common assumptions about availability:

(1) "A high level of availability is always better than a low level;"
(2) "The more availability the better;"
(3) "Fastest response times are best."
(4) "Availability is assured service for authorized users."

Scenario 1

A command and control center is about to be overrun by the enemy. As the commanding officer and his staff escape, they blow up the computers and equipment to be left behind so that neither the technology nor the data will be available to the invading forces.

When the enemy leaves and the original force retakes the destroyed position, none of the systems are available, although the officer and his troops are authorized users.

Scenario 1: *No* availability may be preferable.

When no availability is desired, other goals, like confidentiality, are preeminent.

Scenario 2

For two days after the January 1994 California earthquake, the telephone companies blocked all telephone calls coming into California in order to keep remaining telephone lines available for out-going calls, for disaster relief organizations, and for residents needing medical help. [4]

Scenario 2: In some situations *less* availability is better.

An effective availability policy may actually deny service to authorized users! Policies for catastrophic situations often redistribute system availability so that emergency workers get the resources they need. For example, in Norway, certain individuals in each town, including the fire chief, the mayor, the director of public works, and the doctors, are identified as people who will continue to receive telephone service in a catastrophe.[5]

Reduced availability (whole or partial shutdown) may be appropriate for a system under remote attack.

Scenario 3

> Have you ever wondered if the telephone company deliberately lets it take more time to speak to a directory assistance operator than it takes to look up a number in a phone book?

Scenario 3: *Less* availability may be most effective
in normal circumstances in the long run.

Fast response time may not always be in a corporation's best interest! J.W. Forrester wrote a book on system dynamics illustrating that social systems often produce the opposite of what was intended -- a fast response time may ultimately result in an overloaded and slow system.[6] Ruth Nelson recommends slow release as a deterrent to information collectors and intruders.[7]

These three examples illustrate that availability policies involve more than keeping systems up and running, or providing as much service as possible as fast as possible while preventing denial of service attacks.

SECTION 3

AVAILABILITY IS MULTIFACETED

Availability often means different things in different application contexts. To the telephone company "readiness for use" means that users get a "dial tone" when they pick up the handset. To a global airline, hotel, or auto reservation service, it means that the computer reservation network is "up" and functioning. To the power company, it means customers can get the electricity that they pay for when they need it. In the military, "availability" implies that a ship, platoon, computer, missile or tank, etc. is ready for deployment. To a bank, it can mean that automated tellers are functioning and enough money is on hand. Basic assumptions, terminology and emphasis often differ from one infrastructure industry, such as telecommunications, to another, like power. At the end of the paper, we summarize common definitions.

Availability often means different things under different circumstances. As we saw in the scenarios above, external conditions may vary from normal, to stressed, to catastrophic.

Availability policies may involve distributing capability and deciding who should get service and who shouldn't. Policy alternatives are usually thought out in advance, and implementation mechanisms are installed ahead of time. Responsibility for selecting the appropriate policy is often shared with government representatives as well as knowledgeable volunteers.

- When Robert Morris Jr.'s Internet worm brought down about 6000 UNIX systems overnight, the initial response was *ad hoc*. Since then coordinated emergency response teams have been set up to deal with malicious software before it can assume catastrophic proportions.

ATTRIBUTES OF AVAILABILITY POLICIES

Like any security policy, availability policies include objectives to be met, threats and vulnerabilities to be countered, risks to estimate, security mechanisms to be used, real-world constraints and measures of effectiveness. Threat scenarios can help identify availability requirements and required responses.

Availability *objectives* are usually stated in terms that can be measured, such as degree of readiness, average response time and percentage up-time.

> The readiness objective of the rebel farmers of the American Revolution was to be ready to fight in "a minute," hence their name "Minutemen."

Scenario 4: Availability objectives are measurable,

Threats to information system availability include loss of power, denial-of-service attacks, loss of keys to encrypted data, physical damage to equipment (via accidents, sabotage, terrorism, natural disasters, war), magnetic erasure (from electromagnetic pulse, electric current, magnets), to name just a few.

Security mechanisms to support availability include extra capacity, backups (for power, data, operations staff, air conditioning, etc.), redundant systems, and limits on repetitious behavior.

Risks provide an estimate of the likelihood of a threat and its potential impact, both financial and loss of life. *Risk analysis* justifies the cost of security mechanisms as well as deployment of security actions that may enhance or reduce availability, such as "increasing audit logging, selectively disabling remote services, and disabling selected accounts."[8]

Vulnerabilities are weak points which need to be addressed. Availability vulnerabilities include single points of failure and hogging resources.

Constraints include time, cost, ease-of-use, bandwidth, capacity, safety and endurance.

Internal and external instruments *measure* availability as system up-time, system response time, and degree of readiness. They include analytical tools to compute averages, minimums, maximums, deviations from the norm, etc.

Negative Measures of Availability

In an adversarial situation, the degree of readiness may be a negative number because equipment or personnel have been rendered inoperable, requiring time and money to repair or replace. In war, keeping one's own systems available while denying availability of the enemy's to the enemy is a key strategy.

The English privateers of the 16th century sacked Spanish ships, sinking them or rendering them inoperable for months.

Scenario 5: How long does it take to restore a galleon to operability?

In our own day, "The essence of Information Warfare is to destroy the enemy's communications system before he destroys yours."[9] This implies that degree of readiness has a negative axis, as roughly illustrated below.

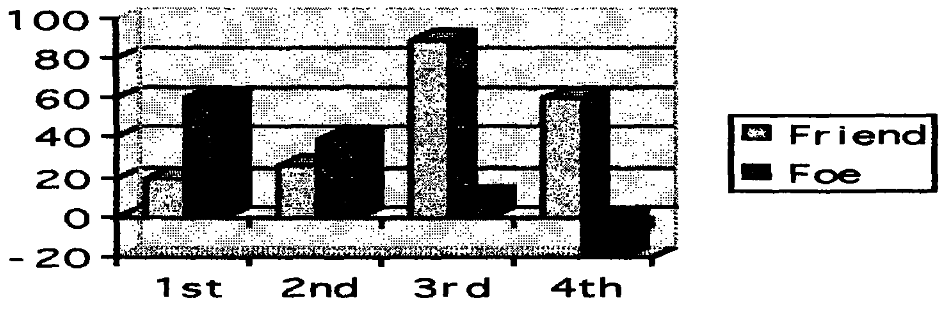
Communications Availability

As Friend builds up to attack Foe's territory, Friend increases its communications capability over the 2nd and 3rd quarters.

As Foe's communications are attacked, the availability measure goes down over the 2nd, 3rd, and 4th quarters, until it goes below zero to negative numbers, reflecting the time, money, and effort needed to restore communications.

	Time	*1*	*2*	*3*	*4*	*Quarter*
Friend		20	27	90	40	Readiness measure
Foe		60	40	10	-20	Readiness measure

(Illustrated on next page)



Scenario 6: Availability's Negative Axis Appears in Adversarial Situations

Larger negative numbers imply longer times to recover.

INTERACTIONS WITH OTHER POLICIES

Availability policies interact with other system policies, sometimes synergistically, sometimes antagonistically. Part of availability's complexity results from such interactions.[10]

Synergistic Interrelationships

"Capacity" supports availability. An increase in capacity can alleviate certain classes of availability problems.

"Robustness," "reliability," and "survivability" support availability by supporting readiness for use. Robustness emphasizes inherent strength, while reliability focuses on endurance over time, and survivability focuses on endurance through dangerous and damaging conditions.

> In the Persian Gulf War it took longer than expected to bring down Saddham Hussein's American-built communications systems because they were programmed for reliability and sought out alternate communications paths whenever one node was knocked out.
>
> The reliability and assured service American firms had built into their telecommunications products worked against America in war.

Scenario 7: Information warfare may alter normal availability interactions.

Availability supports "safety" by keeping critical systems going. Backup systems, alternative routes, graceful degradation, readiness testing and other strategies provide for both safety and availability.

Availability supports “access control” by enabling the operation of the system making access control decisions and keeping information quickly available to those permitted to see or operate on it.

Conflicting Interrelationships

Privacy policies conflict with the computerized availability of personal data.

Availability often interferes with confidentiality. For example, terminals at nurses’ work stations or airline ticket desks are routinely left connected for long periods of time because of the inconvenience of repeatedly identifying and authenticating users.

Fee-for-service resource allocation (availability) policies may conflict with safety policies. The regressive effects of fee-for-service telephone service may be counterbalanced, for example, by emergency phone service offered at no charge to the house-bound elderly or infirm.

To promote availability to all segments of the population, the federal government is placing computers in libraries, hospitals, and other public places. Opening the Internet to a larger, global group of people encourages widespread use, but makes the network and the information in it more vulnerable to attack.

Destroying one’s own position in anticipation of an enemy overrun in battle preserves confidentiality, but intentionally destroys integrity and availability.

Policy Differences

It has often been stated that availability policies differ from integrity and confidentiality policies because availability operates on a continuum and is usually measured rather than counted.[11] However, integrity and confidentiality both rely on assurance which has always been continuum-based. Furthermore, new continuum-based paradigms for authentication,[12] confidentiality, and integrity[13] suggest that other security goals can be measured.

SECTION 6

EVOLVING AVAILABILITY PARADIGMS

The Single Computer Paradigm

This paradigm centered around a computer system (mainframe, minicomputer, workstation or personal computer) that was crucial to the organization. System up-time and responsiveness was valuable, permitting superior customer service and a competitive edge.

The primary threats were mechanical or human accidents, like loss of air conditioning or dropping a disk pack. A number of security measures, including off-site backups, fire extinguishers, field service contracts, and duplicated components assured a reasonable amount of availability. Cost/benefit analysis determined which measures were appropriate.

Availability was expressed in terms of the percentage of time the system was up and running. High availability systems had goals like 99.9% availability, while low availability systems had goals like 60% availability. To accomplish these availability goals, system designers anticipated maximum loads and planned for them, incorporated fail-safe and fault-tolerant components, duplicated all or portions of their systems, provided alternate communication routes, and designed back-up procedures.

The Network Paradigm

Networks of computers, enhanced by redundancy, distributed processing and distributed databases, increased availability by orders of magnitude. Switching, multiplexing, alternate routing, fault-tolerant equipment, high bandwidth (communications satellites, asynchronous transfer mode (ATM), fiber optics), and other many other technical enhancements improved sharing of network resources.

Availability was still measured primarily by percentage up-time and responsiveness, and designers assumed that more availability and smaller response times are better. The major threats to availability in the network paradigm were: 1) loss of an unreplicated focal point; and 2) lack of interoperability between systems and networks. Protocols like TCP/IP made the Internet possible, but this network of networks was only available to government and research institutions.

In this paradigm resistance to denial-of-service attacks became increasingly important. New threats included malicious human intent, implemented via mechanisms like viruses, worms, and Trojan horses. Countermeasures include intrusion detection, virus eradicators, and laws and regulations.

The Cyberspace Paradigm

Cyberspace[14] is the electronic information world built upon computers that emphasizes information availability as much as system and network availability. Ease of use and global connectivity make computer power and information available to many more people while multiple media expand the kinds of information available. Critical enabling factors include:

- Opening of the Internet to individuals, commercial organizations, and other countries;
- Development of the hypertext-based World Wide Web;
- Next generation networks and software.

In Cyberspace threats are social[15] as well as technical.[16] Intellectual property laws, market-based distribution of service, and security requirements restrict availability. In additional to traditional adversaries like enemies, terrorists, hackers, and competitors, new threats to availability include:

- Censors: Eliminators of offensive, illegal, or age-inappropriate information;
- Marketers: Producers of electronic junk mail;
- Poverty: Cyberspace isn't available to those without resources;
- System limitations: Graphics, films, and animation require much more bandwidth, for example, interfering with phone service.

Policies for Cyberspace

Availability policies must cover a wider range, from assured service for oneself to assured non-service for one's enemies, incorporating gaming strategies and other techniques developed for adversarial conditions.

Responsiveness may be highly desirable (the more the better), or it may depend on circumstances. Slow responsiveness may be the policy of choice in a variety of situations. Fuzzy logic is appropriate for modelling qualities that can be measured along a continuum.

Availability may cover information content as well as system availability. Parents may be able to restrict their children from seeing certain types of programs or playing certain types of games. Dictators may be able to restrict critical or seditious programming. Censorship may be necessary during war, to prevent the enemy from knowing where force build-ups are taking place. The degree of restriction may vary, depending upon threat conditions.

Availability may conflict with other goals, like confidentiality. Strategies for dealing with conflicting goals, such as defining response scenarios and metapolicies,[17] are appropriate.

Measures

In addition to the traditional measures such as up-time and responsiveness, the new availability paradigm measures:

1. The time, effort, and money it would take to destroy an adversary's system availability, or rebuild it afterward;

2. The time, effort, and money it takes to destroy one's own system before an adversary overrun, or rebuild it afterward;

3. The usefulness and choice of information;

4. The desired and actual rate of release of information;

5. The degree of conflict with other policies.

CONCLUSION

Although "confidentiality, integrity and availability" are the three pillars of information security, INFOSEC researchers and lawmakers have given availability policies a disproportionately small share of attention. This paper begins to address this imbalance, focusing on availability policies in adversarial situations. The paper uncovered some shortcomings in availability policy theory and proposed a more balanced view. In the process, it corrected four popular misconceptions:

1. Preventing denial of service requires ensuring the complete functional correctness of a system;
2. Availability is the same as assured service;
3. Availability is always desirable;
4. More availability is always better.

The new view incorporates gaming strategies and other techniques developed for adversarial conditions. The paper demonstrated that availability measures need a negative axis.

A new availability paradigm will enable us to more successfully model security in the real world, and is essential to providing security on the nets.

1996 ACM New Security Paradigm Workshop Lake Arrowhead CA
Copyright 1997 ACM 0-89791-878-9 96 09 ..$3.50

GLOSSARY

Policy

A “policy” is a set of rules for a domain, set by a domain authority.

Availability

A popular dictionary defines generic "availability" as the state of being "ready for use, usable, readily obtainable, accessible, or having sufficient power or efficacy,"[18]

The 1996 *ISSO Glossary of INFOSEC and INFOSEC-related Computer Terms* defines INFOSEC “availability” as “ensuring that computer resources are available to authorized users when they need them.”

“Availability” is an evolving term in the INFOSEC community.

- It does not appear in the Orange Book,[19] published in 1985, where the required “explicit and well-defined security policy” refers to mandatory and discretionary access control.

- By 1987 the *Trusted Network Interpretation* of the TCSEC and other proposed standards include "denial of service," defined as "the prevention of authorized access to system assets or services, or the delaying of time-critical operations."

- The awkward double negative “preventing denial-of-service” was rephrased in positive terms to parallel confidentiality and integrity. By 1991 *confidentiality, integrity*, and *availability* had become the major goals of information security, although some proposed more comprehensive frameworks.[20]

END NOTES

[1] Gasser, Morrie, *Building a Secure Computer System," Van Nostrand Reinhold, New York, 1988.*

[2] Hosmer, Hilary, "The Multipolicy Paradigm", *Proceedings of the 15th National Computer Security Conference,* 1992.

[3] Joint Security Commission to the Secretary of Defense and the Director of the CIA, *Redefining Security,* Washington, 1994.

[4] *Newsweek,* January 31, 1994.

[5] Josang, Audun, Norwegian University of Science and Technology, Trondeim, Norway.

[6] Forrester, J.W. *The Counterintuitive Behavior of Social Systems,* MIT Sloane School of Management, 1970.

[7] Nelson, Ruth, "Unhelpfulness as a Security Policy or It's about Time," *Proceedings of the 1995 ACM SIGSAC New Security Paradigms Workshop,* Little Compton, R.I., IEEE Press, 1995.

[8] SBIR 1997.1 DARPA SB971-009.

[9] Jordan, Wesley, Rear Admiral, Ret., AFCEA meeting, Hanscom AFB, MA, 1995.

[10] Jonathan Millen of MITRE and Virgil Gligor of the University of Maryland have explored these issues in more depth.

[11] Nelson, Ruth, private conversation.

[12] Yesberg, John, and Mark Anderson, "QuARC: Expressive Security Mechanisms," *Proceedings of the 1995 ACM SIGSAC New Security Paradigms Workshop,* Little Compton, R.I., IEEE Computer Society Press, 1995.

[13] Hosmer, Hilary, "Security is Fuzzy! Applying the Fuzzy Logic Paradigm to the Multipolicy Paradigm," *Proceedings of the 1993 ACM SIGSAC New Security Paradigms Workshop,* Little Compton, R.I., IEEE Computer Society Press, 1993.

[14] A term coined by novelist William Gibson.

[15] Miller, Steven, *Civilizing Cyberspace: Policy Power, and the Information Superhighway,* ACM Press, 1996.

[16] Dr. Frederick Cohen, "50 ways to Attack Your World Wide Web Site", *Computer Security Alert,* CSI, December 1995.

[17] Hosmer, Hilary H. "The MultiPolicy Paradigm" and "Metapolicies II," *Proceedings of the 15th National Computer Security Conference,* Baltimore, MD, 1992.

[18] *The Random House College Dictionary, Revised Edition,* Jess Stein, ed., New York, 1984

[19] DoD, *Trusted Computer System Evaluation Criteria* (TCSEC), 1985.

[20] Parker, Donn, "Restating the Foundation of Information Security," *Proceedings of the 14th National Computer Security Conference,* October 1-4, 1991, Washington, D.C. Proposes a framework for information security that includes utility, authenticity, and possession, as well as confidentiality, integrity, and availability.