A Qualitative Approach to Information Availability

This work has been supported in part by the Greek National Secretariat for Research and Development, under Programme YPER'97.

Theodore TRYFONAS, Dimitris GRITZALIS, Spyros KOKOLAKIS
Dept. of Informatics, Athens University of Economics and Business
76 Patission St., GR-10434, Athens, HELLAS
{tryfonas, dgrit, sak}@aueb.gr

Key words: Information Security, Availability, Confidentiality, Integrity, Security Models

Abstract: During the last fifty years Information and Communication Technology (ICT) has contributed to almost all sectors of organized societies. As a result, information security is fundamental for several social and business processes that rely on ICT. One dimension of information security concerns availability of information and computational resources. It is essential for a system's correct operation and its acceptance from end-users to respond with proper reaction times to authorized requests. But whereas other security parameters have been studied and analysed very well, availability has not. Throughout this paper this fundamental parameter of ICT security is under study through a qualitative perspective. We aim at providing the basis for a consequent formalistic foundation of information availability. Approaches like these may be useful for the conceptual description of the problem domain, whilst this conceptualisation may also help in the realization of the guidelines, which are essential for the development of secure information systems.

1. INTRODUCTION

During the last fifty years Information and Communication Technology (ICT) has contributed to almost all sectors of science and industry and there are cases that its contribution is rather critical. In proportion, ICT security becomes fundamental for several services that rely on ICT. There are many definitions for *information security* but most of them resemble to *"the preservation of confidentiality, integrity and availability of a system, or its parts, from potential threats"*.

ITSEC defines IT security as the preservation of [ITSEC91]:

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35515-3_53

S. Qing et al. (eds.), *Information Security for Global Information Infrastructures*

- ξ Confidentiality: prevention of the unauthorized disclosure of information;
- ξ Integrity: prevention of the unauthorized modification of information;
- ξ Availability: prevention of the unauthorized withholding of information or resources.

That triplet is often called *the CIA security model.* In other bibliographical resources other security parameters, like non-repudiation, authenticity, validity, or utility, are often referred to as definitional ones [PARK95].

Whilst the other security parameters have been well studied and analysed, availability has not. A coarse survey of the literature would suffice to show that information availability has attracted the less attention by security researchers. For example, in 1998's IFIP/SEC Conference the keyword 'Availability' was totally missing from the keyword list. In all 64 accepted papers, availability was not referred to, either to the title or as a keyword [PaPo98]. One of the main reasons for this may be the absence of a rigorous, comprehensive and unambiguous model of information availability. Whilst there are numerous formal models for confidentiality and integrity, there is no model for availability that would define the concept with rigor [KOKO96].

The aim of this paper is therefore twofold. On the one hand, to demonstrate those characteristics of availability, which substantiate the need for it to be considered equally important to confidentiality and integrity, for it is often treated as less significant. Preserving information availability is a technical challenge and a prerequisite for stable Information Systems' operation. On the other hand, we shall try to establish a conceptual basis for consequent formalistic foundations of availability; the paper will attempt to enlighten several aspects of the issue, so as to foster the rigorous description and conceptual clarification of it.

2. AVAILABILITY AND ITS RELATION TO THE OTHER SECURITY PARAMETERS

In ITSEC [ITSEC91] there is a significant difference in the definition of the ICT security parameters. Availability is the only parameter where the prevention of the unauthorized withholding pertains both to information and to resources. Reference to the resources indicates that availability could not be considered in isolation, but one should always consider it in the context of a technology system that could comprise computational resources, processes and humans except from information itself.

In such a context, the availability of any component alone cannot ensure the uninterrupted flow of information. Moreover, while confidentiality and integrity can both be supported by preventing illicit access to information (access control) within a computer system, availability cannot be. As a consequence, we cannot obtain the same level of assurance for availability, as for confidentiality and integrity, which can be enforced by an appropriate reference monitor. This is because one cannot be sure that the system will meet any particular level of availability due to the existence of a fundamental distinction between *confidentiality and integrity* and *availability*. Both former concepts can be characterized in terms of properties that are precisely defined, global and persistent; they can be specified for a particular system in a way that allows one to know, beyond doubt, whether or not the system enforces those properties [BrSc95].

Another difference between *confidentiality* and *integrity and availability* this time, relies on the type of threats each one is exposed to. Confidentiality is endangered when a malicious entity is interested in trusted information. On the other hand, integrity or availability can suffer either because of faults (that might have random/accidental cause) or because of malevolent action (e.g. Denial-of-Service attacks).

As categories of potential threats against availability one could consider:

- The excessive workload on a system, which harms the ability of it to provide information or services to authorized requests (e.g. a heavily used web server).
- The system malfunction, which may be due to physical errors (e.g. a hard disk failure).
- IS or IT-components erroneous design (e.g. omission of proper administrative procedures).

In real-time applications, the response time to a request is important. In this respect, availability is similar to acceptable *latency*. The acceptable response time of an application depends on the application itself, but is assessed according to the demands and expectations of the operational context. An example could be an intensive-care application that supports operations; the demanded response time of that application is different during an operation, where immediate response is required, than that during a case study [GRIT94]. Hence, availability is heavily subjective and therefore any quantification of it can be done only in context.

3. SECURITY TRADE-OFFS AND POLICIES

3.1 Conflicts and trade-offs between security parameters

When designing a security policy, one has to make architectural trade-offs between security parameters, in order to establish a suitable control environment. In a given context a designer has to compromise with the more critical ones. In this section trade-offs that one should be aware of will be addressed. Conflicts are usually resolved in favour of the dominant security parameter.

Whenever the environment is loosened from strong confidentiality requirements, most of the preventive measures could cover both integrity and availability. For example duplication of resources/information (e.g. RAID, backups), error corrective codes (CRC, etc.) and similar practices can be used for the preservation of both availability and integrity. Of course, despite the fact that redundancy favours availability, integrity may suffer severe problems due to data replication. Update or disposal of multiplied information, its distribution and similar procedures could seriously affect integrity in the case they are not carefully deployed. The consistency of cloned information is a major subject to several areas, e.g. computer architecture (the cache coherence problem) and distributed databases (replicated entities management).

3.2 Policies with regard to availability

Depending on the environment, the policies set forth for preserving availability may vary. Availability policies rely on the dependency of the system and the environment's security requirements. Many times, availability requirements are often in conflict with others, set forth in the same security policy; for instance, by restricting availability, confidentiality could strengthen.

Availability is critical for control systems (e.g. nuclear reactors), and one should not expect to find such a system open and exposed to public access. A regulatory system is usually isolated, therefore its confidentiality downgrades to physical access control. On the other hand, payments over the Internet must ensure the preservation of confidentiality and integrity of transmitted data.

Therefore, in an environment, not threatened particularly by malicious actors or software, integrity downgrades to validity of data and can be protected by the generic mechanisms that could protect availability as well. On the contrary, the more open the system is, the more the other security parameters are strengthening. Hence, one cannot evaluate availability in

terms of universalising statements like "the quicker the better". Other security parameters may be vital as well, reducing the value of timing. In an open and inborn adversarial environment constraints of availability are required, while in isolated systems requirements for other security parameters are loosened.

4. AVAILABILITY PERCEPTIONS, TEMPORAL CONSIDERATIONS AND A TAXONOMY OF IS

4.1 Availability perceptions and derived definitions

Due to its definition, availability refers to the efficiency of a system to provide information and services "permanently" and "on-time" [KeUl94]:

***Permanence*:** Information and human and technical resources are basic actors for the operation of an IS. Ensuring availability requires a three-factored scheme, capable of protecting all of these actors. The first factor concerns the protection of *physical resources*. The second deals with the *human resources* and the third is concerned with the *procedures* specified. Therefore, permanence guarantees macroscopically uninterrupted operation.

***On-timeness*:** This dimension is related with the real-time operation of a system. There is always a demand for a service, involving either *data* or *media*, that we expect to be satisfied within an acceptable amount of time.

Figure 1 depicts a classification of availability. *Real-time availability* is fuzzy, for it depends on the environment (context-based). *Macroscopic availability*, on the contrary, primarily relies on the architecture of a system and the procedures selected and established for guaranteeing continuity of operations. As a result, availability can be thought of as a multi-faceted security parameter [HOSM96].

Availability can also be addressed through side concepts such as the *mean time to failure*, MTTF, and the *mean time to repair*, MTTR. We will present those concepts in brief as to demonstrate how related concepts can be realised in simple mathematical relations. Hypothetical time intervals that a system operates properly and, on the other hand, those that the system cannot meet the required objectives with regard to availability can be defined formally, as we shall see.

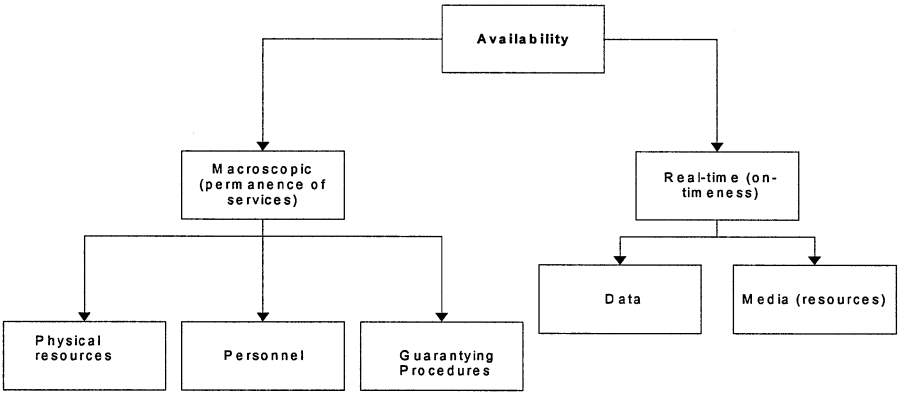

Figure 1. Classification of availability

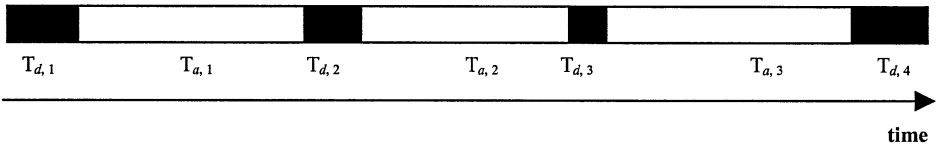

Figure 2. Active/inactive time intervals of a system's operation (white tabs are active periods)

Suppose we observe the operation of a system for some period. Say that during that observation period it was "alive" N_a times and non-working N_d times. We set $N_a + N_d = N$. We can easily verify that $|N_a - N_d| = 1$, because it is always either $N_a = 1 + N_d$, or $N_a + 1 = N_d$ (see figure 2). *Mean time to failure* is the quantity:

$$MTTF \quad \frac{\sum_{i=1}^{N_a} T_{a,i}}{N_a}$$

Respectively, *mean time to repair* is the quantity:

$$MTTR \quad \frac{\sum_{j=1}^{N_d} T_{d,j}}{N_d}$$

If the total observation time is T_{obs}, then (for "adequately large" values of N_a, N_d, which means $N >> 1$), this is equal to:

$$T_{obs} = N_a \cdot MTTF + N_d \cdot MTTR \approx (N/2) \cdot (MTTF+MTTR) \quad \textbf{(Eq. 1)}.$$

Hence, we expect a system to have large MTTF in order to be stable from a continuity of operations perspective. Under that point of view the authorization of requests is not addressed as a concern, because such an approach refers to the non-functional aspect of response time.

Let us now define *dependency* qualitatively; a system is considered to be dependent when its infallible operation and the preservation of the CIA parameters rely, not only on its architectural design, but also on its context.

In the paragraphs to follow we will categorize systems with criteria: (a) their dependency level, and (b) the expectation for their availability through the continuity of operations perspective.

4.2 Temporal concerns for protection

It is interesting to discuss why availability is fuzzy and context-dependent. As we addressed, the main qualitative concept of availability is *time*. It is meaningful to define a position in space without time, but there is no meaning in concepts like speed, latency, or throughput without it. Since availability can be thought of as *"acceptable latency to authorized requests"*, qualities of time, such as fuzziness, are also qualities of availability. Unavailability incidents are clearly detectable but it is not possible to correct them immediately; when such incidents *happen*, they are considered historical.

As we explained, the definition of *mean time to failure* (MTTF) is governed by time and can be evaluated in relation with the expectations of the subjects that participate to the system. Tolerance mechanisms with regard to availability usually enforce methods that are applied to the system, when an incident has occurred. Hence, unavailability tolerance means the ability to quickly reconstruct lost, corrupted, or held-back data/equipment. More difficult to be deployed solutions apply essential preventive actions after a forecast of the system's behaviour, so as to avoid potential denial of service.

4.3 An availability-focused taxonomy of IS

A taxonomy of information systems based on their requirements for availability will be introduced in this section and may help developers to realize particular security needs of the IS, from the early stages of its development. Such taxonomies could also be used as tools for the integration of IS security requirements with the other functional and non-functional ones. We set:

Criterion 1: Dependency. A *{low, medium, high}* scale will be used. Low dependency characterizes a system the operation of which does not rely "very much" upon the environment that uses it. Medium dependency characterizes systems where "several" of their features are dominated by their operational context. High dependency characterizes systems the behaviour of which is "totally" dependent upon their environment.

Criterion 2: Expectation for continuity of services. A *{low, normal, advanced, strict}* scale will be used to map requirements for availability.

The above categorization scheme makes it possible for systems of several types to be described in terms of their availability. There are systems for example that depend on the integrity of their data because they are used for *analysis and statistical processing*; in such systems the major concern is the validity of data. Other systems, which are used to *support decisions*, depend on both the validity of their data and their on-time process; in such systems the expectation for availability is considerably greater. *Regulatory systems* (e.g. reactor control mechanisms) depend heavily on availability; there uninterrupted operation is expected, even in the case of corrupted data that can cause false alarms. Table 1 presents in detail what has been discussed here.

5. EXISTING AND EMERGING WAYS OF UNDERSTANDING AND APPROACHING AVAILABILITY

5.1 Establishing a framework

For its correct operation it is essential for an IS to be designed with regard to proper response-times to authorized requests. ITSEC realizes that requirement by relevant *functionality classes* that are presented and discussed in that standard. Availability there is approached under the perspective of continuity of services through proper architectural design and system's maintenance. Thus, ITSEC states a framework for an approach to availability, an *availability paradigm* [HOSM96]. In the following paragraphs a number of availability paradigms will be briefly presented and commented upon.

5.2 The Single Computer

This paradigm is the first that evolved; i.e. the preservation of the uninterruptible operation of a single computer system that provides specific services. The availability of such a system is usually measured through the *mean time to failures* (MTTF). A series of physical protection measures (fire protection, physical access control, maintenance contracts etc.), in accordance with technical safeguards (duplicated parts, backups) may ensure the efficient operation of the system and form a satisfactory control environment. This paradigm applied wherever a single computer was the basis of information processing services [HOSM96].

Table 1. Availability-focused IS taxonomy

System type	Depend-ency	Expectation for availability	High requirements for	System example
Information analysis, statistical processing	Low	Low	Integrity	Market research support
Quality management systems	Medium	Normal	Integrity	ISO 9002 implementation computerized systems
Decision support systems	High	Advanced	Integrity, availability	Business applications, ERP
Safety management systems	High	Advanced	Availability	Vessel navigation assistants, fire alarm controls
Regulatory systems - Critical application support systems (I)	High	Strict	Availability	Reactor arbitration control, intensive care support applications
National security - Critical application support systems (II)	High	Strict	Integrity, availability, confidentiality	Military communications, governmental sites

5.3 Computer Networks

Communication capabilities enlarged considerably the limited computing capabilities of the single computer. Advanced computational and network equipment, satellite technologies, ATM and fibre optics provide a framework where architectures like distributed databases and proxies, fit in and assure higher availability. In this context availability is equalized with *latency*. The main threats in this paradigm are unrecoverable damages, **interoperability issues** and in addition, the interconnection capability induced another class of threats, i.e. the malicious attacks (e.g. viral software). Interconnection of networks and the confrontation of the threats stated above, gradually introduced the next paradigm.

5.4 Internetworks

Multimedia eased end-user access to huge amounts of information. Raw data enhanced with images, voice, animation and video, disengaged end-users from the overhead of particularly interpreting them. The 3W, using that powerful ease-of-use of multimedia, opened the Internet to the public. New problems, not solely technical, evolved. Social-related availability problems arose like the public's ability to access the Net. The technical framework

changed from the concept of the idealistic distributed processing, to the selective distribution of services. A typical example of that trend is an IP network, where services are issued trough a number of servers. Existing technical threats are a combination of the potential threats against either a single computer system and/or a network. Table 2 summarizes the characteristics of each paradigm.

Table 2. Availability paradigms

	Computational philosophy	*Technical availability threats*	*Social availability concerns*
Single Computer	Centralized, batch processing	Mainly physical and maintaining problems	Limited
Computer Networks	Distributed (networks, multiprocessors, distributed DBMSs)	Physical, maintenance, latency, distributed authentication and access control	Indications of emerging potential threats
Internetworks	Distribution of services through a number of dedicated nodes (client/server, ORB)	Physical, maintenance, latency, security management	Substantial

6. SUMMARY AND CONCLUSIONS

Throughout the paper, availability was under study through multiple perspectives and we addressed several aspects that require further research. The ideas offered may prove to be useful for its rigorous description and conceptual clarification. In fact, there are currently very few formal approaches to information availability because there seems to be a lack of understanding on the nature of it. Availability's particular (and sometimes unique) characteristics include the following:

- It is the only parameter where the prevention of an unauthorized action (withholding) pertains both to *information* and to *resources*.
- Its policies often conflict with confidentiality and integrity ones.
- We cannot obtain the same level of assurance for it, as for confidentiality and integrity, for we lack a precise definition and a rigorous, comprehensive and unambiguous model for it.

- Its conceptualisation involves consideration of subjective judgements regarding the *acceptable* or *desired* system's reaction times.
- It is threatened both by *internal* and *environmental* threats.

An attempt to provide formal foundation of availability should consider:

- That it should be defined both from its macroscopic perspective and its real-time point of view.
- That it is heavily subjective and should always be considered in its context. An attempt to quantitatively model availability should consider approaches that can represent subjective parameters (e.g. fuzzy logic).

7. REFERENCES

BrSc95, Brinkley, D., Schell, R. "Concepts and Terminology for Computer Security". M. Abrams, S. Jajodia, H. Podell (Eds.), *Information Security: An integrated collection of essays*, IEEE Computer Society Press, 1995.

GRIT94, Gritzalis, D. *Information Systems Security in Dependable Environments*, Ph.D. Dissertation, University of the Aegean, 1994.

HOSM96, Hosmer, H. "Availability policies in an adversarial environment". In *Proc. of the New Security Paradigms Workshop*, pp. 105-117, USA, 1996.

ITSEC91, Information Technology Security Evaluation Criteria (ITSEC), Commission of the European Communities, Brussels, June 1991.

KeUl94, Keus, K. and Ullman, M. "Availability: Theory and fundamentals for practical evaluation and use". In *Proc. of the 10th Annual Computer Security Applications Conference*, pp. 258-264, USA 1994.

KOKO96, Kokolakis, S. "Is there a need for new information security models?" In P. Horster (Ed.), *Communications and Multimedia Security II.* Chapman & Hall, 1996.

LeZh97, Leiwo, J. and Zheng, Y. *Layered Protection of Availability*, Pacific-Asia conference on information systems (PACIS '97), Australia, April 1997.

LYU96, Lyu, M. (Ed.), *Handbook of software reliability engineering.* McGraw-Hill, 1996.

PaPo98, Papp, G., Posch, R. (Eds.), *Global IT Security: Proceedings of the IFIP TC11 14th International Conference on Information Security*. Chapman & Hall, 1995.

PARK95, Parker, D.B. "A new framework for information security to avoid information anarchy". In J. Ellof, S. von Solms (Eds.), *Information Security - the Next Decade.* Chapman & Hall, 1995.