

Layered Protection of Availability

Jussipekka Leiwo and Yuliang Zheng

Peninsula School of Computing and Information Technology

Monash University

McMahons Road, Frankston, Vic 3199, AUSTRALIA

Phone +61-(0)3-9904 4287, Fax +61-(0)3-9904 4124

E-mail: {skylark,yuliang}@fcit.monash.edu.au

Abstract

Of the three common objectives of information security, availability is far less understood than confidentiality and integrity of information. As threats against availability are too numerous to be identified, a different approach is required for the protection. In this paper, a layered protection of availability shall be established, and the cost of protection shall be justified at each layer. A very large scope shall be taken, where violations of availability cover both intentional and accidental threats originating from both external and internal sources. Different layers required reflect the different nature of these threats.

1 Introduction

In formal studies, availability has typically been considered as a protection against denial of service attacks, where an entity is not granted access to the information that it is, based on its clearance, authorized to. From a wider point of view, protection of availability can be considered as specification of measures that guarantee that both information and information processing services are protected against different threats, whether accidental or intentional, so that every request of an authorized user may be satisfied always. As the wide perspective causes some problems, mainly because it is not possible to identify all causes of violations of availability in the large scale [5, 11, 17], some sources have suggested availability shouldn't be an objective of information security at all. The approach taken within this paper is to study availability from the wide point of view, and to establish a comprehensive framework for the protection of availability of information.

Other general objectives of information security, confidentiality and integrity, has been studied widely and several models exist for their protection (for example [7, 8,

9, 10, 12, 14, 16]), but availability is not as well understood. Different evaluation criteria also have very little requirements on availability. European ITSEC sets a requirement of continuity-of-service, and the Canadian CTCPEC has availability as an empty mark holder. In 1985 DoD Workshop on Network security [1] concluded that no generic, mission independent, denial-of-service conditions can be identified. In the 1990 CTCPEC workshop on availability [13], the result was that the difference should be made to the loss of availability due to malicious actions by a user, for example Trojan horse, and random failures affecting the functionality.

On very high levels, the only protection measure is transformation of responsibility by signing service provision agreements with external parties. These parties take the responsibility of providing continuous service, for example electricity or data communication service, with a specified cost to the organization. On the organizational level, administrative routines are considered an effective measure for recovering and correcting, or to reduce the probability of a violation of availability [21]. On lower levels, not very many formal models has been introduced. One method to enforce acceptable response times requires an additional layer on top of Trusted Computing Base (TCB), called Denial-of-Service Protection Base (DPB) as suggested by Millen [19].

These three layers lead to the identification of fundamental layers of comprehensive protection of availability, as studied in section 3. It is, anyhow, essential to study threats against availability before specifying protection measures. Therefore, the paper shall be started by a study of threats in section 2. The cost of comprehensive protection, based on the layered approach shall be analysed and alignment with the value of information shall be studied in section 4. Once the cost of protection has been established, section 4.3 analysis essential factors in the alignment of cost of protection to different risks. Conclusions shall be drawn, and directions for future work highlighted in section 5.

2 Threats and countermeasures

When compared to other goals of protection, availability is very different from the nature in terms of predictability of the effect of protection measures. When integrity and confidentiality are concerned, the effect of protection can be measured in terms of difficulty of breaking them, and the effort required to break protection can be calculated. If, for example, confidentiality of information transferred in an encrypted form is improved by encryption, it is obvious that the effect of different encryption schemes and key lengths can be approximated (if assumed that no back doors exist). Anyhow, since violations of availability can result of an uncontrolled amount of sources, it is not easy to say whether the result of not having violations is the result of protection measures or lack of violation attempts. Therefore, objectives for availability must be

slightly different from different nature that from confidentiality and integrity.

The scope towards concept availability shall be very wide. In formal studies the concept is usually specified to be very narrow, but a different approach shall be taken within this paper. Protection of availability shall be seen as protection against any threat that may cause an authorized request to any service of information fail. As threats can be intentional or accidental, internal or external, active or passive, semantic or syntactic, static or dynamic [2, 6, 18, 24], and definitions of availability are usually very wide, not all threats against availability can be identified [5, 11, 17]. Millen [19] has divided threats into two: threats through resource allocation and threats through resource destruction. For the scope of this paper, the classification into three shall be provided, where threats are classified into three as follows. The lowest level, technical threats, are mostly threats by resource allocation, and upper level threats through resource destruction.

External threats are those that are out of the direct control of organization. For example, communication services are usually leased from external service providers. Therefore, even if the organization is dependent on the availability of these services, in the case of violation, all action must be taken by the service provider. Typical examples are different acts of God, like floods, falling trees etc. The amount of sources can not be identified.

Administrative threats are those that can be covered by good administrative routines. This is also where most of the research is carried out in the field. Threats include loss of information due to system failures and accidental loss of files etc. Typically, administrative routines like proper backup routines, file system duplication, and physical security are very adequate measures against these threats.

Technical threats can be considered as intentional violations of availability, that is denial of service attacks. This is where the lack of formal protection emerges. Availability is usually considered as an enforcement of finite response time policy, where the protection model should guarantee that systems provide the user with a response in a given maximum waiting time. Obviously, before such policies taking effect, it must be guaranteed that there are enough resources to satisfy such requests.

These threats can be further divided into subclasses, and different types of protection measures can be specified for each layer. In the remaining of this paper, the classification shall be followed where technical protection measures shall be studied on three levels and administrative measures on two levels. Technical measures shall include first adequate HW resources, where action is taken to guarantee that that actual resources of system are adequate to satisfy the response time policy. On top

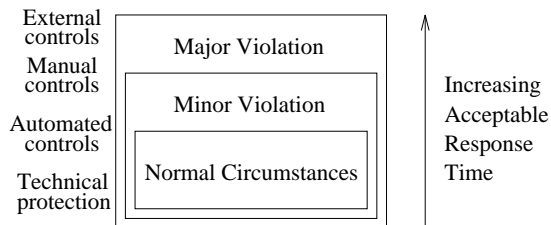


Figure 1: Violations of availability

of this is the allocation of resources in the form of operating system (OS) design. Adequate resources must have suitable OS level allocation schemes to make sure that resources are allocated in an effective and fair manner. The third component of technical level is the actual Trusted Computing Base (TCB) that is a layer on top of operating system to provide protection. All these components are targetted on protection against intentional logical attacks that might violate availability.

Administrative level shall be divided into two. First, there are physical protection measures that include structural protection of computer centres, monitoring of traffic and other such measures. The scope of protection is widened to cover also physical attacks and accidental violation of availability. Physical attacks can be protected with good physical security, that requires adequate administration and administrative routines including backup practises and other sych measures. The significant feature is, that these measures protect only information and services that can be controlled b the organizations. To protect also services that are external from the organization, different service delivery agreements are required with external parties to control the provision of such services and to specify rights and responsibilities in the case of violations.

As the fundamental goal of availability is to guarantee acceptable response times in all circumstances, the violations can also be classified into layers according to the severity. The protection measures, studied in detail in section 3 have a significant feature, that lowest layer is needed to provide normal circumstances, that is prevention of violations, whereas administrative and external layer are needed mostly to recover from a violation that has already occured. To match violations with types of threats, as illustrated in figure 1, the assumption shall be made, that the more severe the violation, the higher level measures are needed for recover, and the more the acceptable response time must increase.

At the top levels, where the actual correction is carried out in the case of a violation, required measures can be further divided into autoamted and manual controls. Violations of lesser severity can be recovered using automated measures, whereas some violations may require manual correction. For example, in the case of a disk crash, automated actions may be taken to switch into a duplicate of the disk, or to recover

<i>Duration of interrupt</i>	<i>% of Organizations</i>	<i>Cumulative</i>
0.5 days	3 %	3%
1 days	25 %	28 %
2 days	37 %	65 %
3 days	11 %	76 %
4-10 days	19 %	95 %
>10 days	5 %	100 %

Table 1: Operation times of organizations if information or information processing services are unavailable

the disk from parity disks. Other possibility is to manually install a new disk and return the information from back up tapes. The justification of the chosen method should be the cost of protection, that shall be studied in detail in section 4.

3 Layered protection of availability

Availability of information and information processing services is a critical survival factor for most organizations. As illustrated in table 1, most organizations are not capable of operating if the interrupt in information processing services lasts more than three days. One third or organizations can not operate if the interrupt is longer than one day [22]. Dependence of information and information processing services makes the question on how to guarantee that information and services are available always upon an authorized request, that is protection of availability, essential.

We shall approach the problem by studying protection measures at each layer: external measures, administrative measures, and technical measures. These shall be studied in detail in sections 3.1, 3.2 and 3.3, respectively. Once different layers are identified, the relationship of layers, threats and protection measures shall be studied in section 3.4.

3.1 External protection measures

The controls to protect against violations of availability on the large scale are not only internal to the system or organization. As identified by Baskerville [6], both internal and external controls are required. These concepts are also relative to the level or protection to be established. In the normal circumstances, where the scope is on the allocation of adequate resources, external controls refer to the administrative controls to support that operation, whereas from the violation point of view internal controls refer to these controls to detect, correct, and recover by organizational routines, and external controls are those that must take place outside the organizations direct

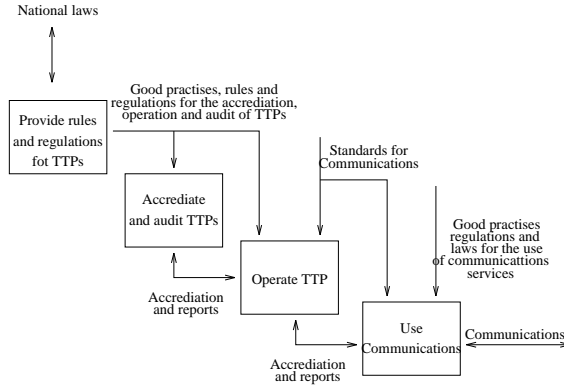


Figure 2: Establishment and operation of TTP

control.

As the protection measures are outside of the organizations control, they can only be affected by contracts with external parties where responsibilities and rights of each party in the case of a violation are set. For example, most of the organizations lease their WAN solutions from telecommunication or other network operators. In the case of external and accidental threats, there is nothing the organization can do, but assume that the system is recovered within an acceptable time. The organization using the service transfer the responsibility of operation to the third party, and according to the availability requirements, agrees with service continuity on a specified cost.

When security is concerned, it is essential that third parties, that is providers of communication between groups of service users, are trusted [4]. Trusted Third Parties (TTP) are responsible for the provision of secure communication services for different organizations. Establishment and operation of a TTP consists of four phases as illustrated in figure 2. First is the provision of good practises, rules and regulations for the accreditation and operation of TTPs. Second comes the accreditation, re-accreditation, and audit of TTPs, third TTP functions themselves, and finally the use of communication provided by TTP. TTP's are responsible for providing the service with respect to national and international laws and agreements, as well as other regulations including contracts with customers.

3.2 Administrative protection measures

On the large scale, major protection measures against violations of availability include [21] proper backup practices, good access control, multiple naming of files, utility programs, and shadowed or mirrored files. This does, anyhow, easily lead to the loss of integrity of information, caused by, for example, when different versions of a file are stored by different names. Administrative routines can also not alone remove threats

unless supported by technical measures. In the ITSEC terminology [3], availability affects both systems and information, whereas confidentiality and integrity are properties of only information. The administrative practices listed above do not remove the threats but do significantly reduce the risk of service being unavailable. For example, RAID technology employing ten disks, of which two are for parity control, can reduce the failure rate compared to large disks from 2-3 years to 90 years [23].

As listed before, availability can be protected using good access control, backups, multiple naming and other such measures, it is very difficult to predict the effect of these measures. The severity of a violation, that is the loss caused by realization of a specific risk, can be reduced, but prevention is more complex. Therefore, an objective of availability on a large scale must be more concerned on the recovery and correction than on prevention. A feasible objective could be, for example, that if a file on a given level of importance is lost, the backup file can be found so, that at maximum of 24 hours amount of work is lost, that is backups from files of that importance are taken and stored every 24 hours. Of course, then a question can be transformed to the protection of those backup files and stores. On the lower scale, a goal can be established that a response time of some particular service, when accessed by an authorized entity, must not exceed a time specified in the security policy.

It is not clear where the line between administrative and technical protection measures goes. As layered security policies require, abstract policies are refined step by step, so the differences are not too clear. Anyhow, the major difference is that all measures in the administrative category - when availability is concerned - are focused on actions taken when a violation has occurred rather than prevention. Backups, duplicated files, multiplication of services and other typical measures are intended to maintain the information or service, so once a violation has occurred, the cost of recovery and correction can be reduced.

3.3 Technical protection measures

On the large scale, availability can be taken as in the ISO standard [2] as “The property of being accessible and usable upon demand by an authorized entity”. Due to ambiguity of this definition, a different approach shall be taken in [20]. Reliable operation of a (distributed) system includes the integrity of the system, that means protection against corruption and loss of data. Protection of availability requires the protection of availability whether a hardware or software malfunction or any illegal activity occurs. As these definitions are very broad, a more narrow approach must be taken.

From the technical point of view, an assumption is made that information is available if the response to an authorized request is provided within a given time limit [15]. This approach further lead to the identification of a key concept within this paper, a denial-of-service protection base [19]. A violation of availability is seen

as an event that results in an unacceptable long response time of an access to any resource, whether information or service. It should be noted, that no difference is made between availability and utility or usefulness of information, as suggested by [21]. It is assumed that information is not available even though the raw data from which - if able to correctly interpret - might be available. Therefore, availability is within this report concerned with the loss of either data or the interpretation of it, or the service that does the interpretation and further processing.

The actual resource allocation on hardware and operating system level are not actually considered as protection measures, but rather system planning activities. Threats of inadequate resource, or invalid resource allocation, can be overcome by proper planning, but this is not usually considered as information security enforcement but is a more general planning. On technical sense, Trusted Computing Base (TCB) is the major target of security design. DPB is a layer on top of TCB where the technical resource allocation policy is enforced. There is also indirect cost of DPB, called user agreement. As always, security increases complexity, and difficulty of use, and in this case the cost is that some operations shall be denied in order to protect against denial of service attacks.

3.4 Relationship of layers, threats and protection methods

As availability is defined such widely, several layers of protections are required, as illustrated in table 2. Even if hardware resources are adequate, they may not be efficiently utilized by the operating system, and problems may occur. On top of operation system, is the actual Denial-of-Service Protection base, that is intended to protect the system against malicious action of users (either legitimate or illegal). As that layer, that is the major focus of this paper, is concerned with the logical violations of availability, the next layer is physical protection, where the system is to be protected against physical attacks. The next layer is the administration layer, where threats are numerous, and the focus is also on prevention, and minimization of effect. Protection measures at this layer are proper administrative routines, like back-up practices, duplication of files, etc. The top layer is labeled external layer. Here are threats that are outside of the direct control of organization, like physical attacks against telecommunication lines. Basically, the organization can not do much against these threats, and the only protection method is to make service availability contracts with service providers.

It should be noted in table 2, that external controls are only modified by contracts with service providers. Therefore, only administrative layers, physical protection and administration, are where the question of automated and manual controls is required. Any measure can be automated or left manual, and as will be studied in section 4, the cost may be very different. The final decision making is not within the scope of this paper, but the cost of protection should be aligned with the cost of resources.

<i>Layer</i>	<i>Threats</i>	<i>Protection methods</i>
HW	Inadequate HW resources	Proper planning
OS	Improper allocation of resources	Well designed resource allocation algorithms
DPB	Logical attacks, malicious SW	DPB
Physical	Physical attacks	Physical security
Administration	both intentional and accidental losses of information or service	Proper administrative procedures
External	Attacks outside of the direct control of organization	Contracts with service providers

Table 2: Layered protection of availability

4 Cost of protection

In this section, the cost of layered protection of availability shall be analysed with a comparison to the cost of information. First, a formal notation for a high level response time policy shall be given in section 4.1. After this, the cost of protection shall be estimated in section 4.2. Once the cost has been estimated, it can be aligned with the protection requirements and risks, as studied in section 4.3.

4.1 Layered availability requirements

As illustrated in figure 1, the more severe the violation, the longer the response time. Therefore, when preparing to the abnormal circumstances, the acceptable response time policy should be prepared to pay attention to the severity of violations. Assume that assets to be protected consists of Information I and services S . Then, assets A can be simply specified as a combination of these $A = \{i_n | i_n \in I \cup \{s_m | s_m \in S\}$. The specification of acceptable response time for each asset $a_i \in A$ is $T_i = \{t_1, t_2, t_3\}$ where each t_j represents an acceptable time in the case of different severity violation. The notation can be simplified by assumint that t_1 represents requirement in normal circumstances, t_2 in the case of minor violations, and t_3 in the case of major violation.

To meet the multilevel security (MLS) requirements, assets shall be classified into priority classes. let $CL = \{\lambda_1, \lambda_2, \dots, \lambda_l\}$ be a set of unique priorities. Uniqueness means, that no classes are overlapping, that is $\forall i, j \leq l, \forall a \in A : (a \in \lambda_i \wedge a \in \lambda_j) \Rightarrow i = j$. Based on these definitions, the acceptable response time policy R shall be specified in equation 1. For the policy to make sense, it should be that $\forall i \leq n : t_{i,1} \leq t_{i,2} \leq t_{i,3}$ and for classification to make sense, it should be that $\forall 1 < i \leq n, j = 1, 2, 3 : t_{i,j} < t_{i-1,j}$.

$$P = \{(\lambda_1, T_1), (\lambda_2, T_2), \dots, (\lambda_l, T_l)\} \quad (1)$$

4.2 Cost estimation

To align the cost of protection with the cost of information, that is the output of risk analysis, it is essential to analyse the cost of protection at each layer. Assumptions are as earlier in this paper. The higher level action is required to recover from a violation, the more cost there will be. It shall be shown how the total costs of availability depends on the cost of protection at different layers. An assumptions shall be made, that layers below DPB shall not be analysed, as they shall be seen as costs not caused by information security, but the system in general.

The cost of technical layer shall be specified in equation 2. We assume, that on a give time frame t , there will be n requests from subjects to objects, and the total processing overhead required to enforce the resource aloocation policy shall be composed of cost of policy consultation C_c , policy enforcement C_e , and the cost of requestst denied C_d .

$$C_{tech} = C_d + \sum_{k=1}^n (C_c + C_e) \quad (2)$$

For the automated controls, assume that there are I duplicate components, denoted as D_i where $i = 1, 2, \dots, I$. The cost of duplication includes the cost of components and the cost of their operation. Let C_i^{dc} refer to the cost of a duplicate component D_i , and C_i^{do} to the operational cost of this component. Other costs originate from cost of detection of a corruption C_i^{de} and recovery costs of the duplicate C_i^{re} .

The total expected cost of automated action is specified in equation 3, where the cost of duplications are multiplied with the probabilities of components failing. Assume that each component has a failure probability p_i . Therefore, the probability of the need for the activation of both duplicate D_i and D_j is $p_i \times p_j$. This leads to the cost of protection as stated in equation 3.

$$C_{automated} = \sum_{k=1}^I p_k \times (C_k^{dc} + C_k^{do} + C_k^{de} + C_k^{re}) \quad (3)$$

The cost of manual actions C_{manual} can be calculated as in equation 3. By comparing the estimates of different cost of different components of actions, the simple comparison can be performed to justify the cost. The comparison to higher level availability policies may, anyhow, prevent the solution that is most desirable in terms of cost, if required performance can not be guaranteed. Similarly, in the case of external hazard, the cost of protection becomes simply the cost of a contract C_{co} and the cost of expected violation, that can be calculated as the probability and severity of an interrupt in service, as specified in equation 4.

$$C_{external} = C_{co} + \sum_{k=1}^I p_i \times C_{int} \quad (4)$$

Based on these costs, the total costs of protection can be calculated as in equation 5. Because each component includes the alignment of the severity and probability of a violation, the total costs can be simply calculated by summing the components.

$$C_{total} = C_{tech} + C_{automated} + C_{manual} + C_{external} \quad (5)$$

4.3 Alignment of risks and cost of protection

Based on the risk analysis, it can be assumed that the value of information, and probabilities of violations has been estimated. Therefore, based on the cost of protection, these values and probabilities can be evaluated with respect to the layered availability requirements, as specified in equation 1. Assume, that based in risk analysis, the results R can be organized into triples $R = \{(a, \rho, \sigma)\}$, where, intuitively, the interpretation can be given where the loss of asset a , with severity σ is of value ρ .

As specified in equation 1, the availability requirements (that is availability policy) can be specified as pairs $P = \{(\lambda, T)\}$ where each T is a triple (t_1, t_2, t_3) . Now it is easy to organize the policy with risk analysis as illustrated in equation 6.

$$(a, \rho_a, C_{total}, \sigma, \lambda, t_1, C_{tech}, t_2, C_{automated}, t_3, C_{manual}) | \sigma = \lambda \quad (6)$$

Now, as each cost is known, it is simple to compare different options of protection by altering parameters (that is modeling different protection options) in equations 2, 3 and 4 to find the optimal alternative for protection without violating the availability policy. The limit of optimization should be that in any combination, it shouldn't be that $\rho_a < C_{total}$ that means the value of information should be more than the cost of protection.

5 Conclusions and future work

A method to estimated the cost of availability of information and information processing services at the high level has been studied within this paper. it has been shown, that an estimated of the cost can be provided and that cost can be justified with respect to the value of information, and the availability policy of an organization, where the availability requirements are formally established. As confidentiality and integrity of information are usually much better understood than availability, and as the definitions of availability in different standards and guidelines are much broader than those of confidentiality and integrity, it is important to study measures to protect and understand availability.

Currently, there has been a lot of discussion about denial of service attacks exploiting features of communication protocols, like SYN flooding where TCP/IP connections are established, but not acknowledged to cause an overflow of resources and,

hence, denial of the service from authorized users. High level measures can not prevent from technical attacks, but once thi high level concepts are clear, lower level technical protection measures, like advanced Denial of Service Protection Bases, can be developed. The most important area for further research in the field of availability is to develop methods to enforce the protection of availability by different resource allocation models.

References

- [1] *Proceedings of the Department of Defence Computer Security Center Invitational Workshop on Network Security*, New Orleans, LA, USA, March 19-22 1985.
- [2] International standard ISO 7498-2. information processing systems - Open systems interconnection - Basic reference model - Part 2: Security architecture, 1988.
- [3] Information technology security evaluation criteria (ITSEC). provisional harmonized criteria, version 1.2. Comomission of the European Communities COM(92) 298 final, Brussels, Belgium, Sep 1992.
- [4] Green book on the security of information systems. The Council of the European Communities, Oct 18 1993. draft 4.0.
- [5] D. Bailey. A philosophy of security management. In M. D. Abrams, S. Jajodia, and H. J. Podell, editors, *Information Security - An Integrated Collection of Essays*. IEEE Computer Society Press, Los Alamitos, CA, USA, 1995.
- [6] R. Baskerville. *Designing Information Systems Security*. John Wiley & Sons, 1988.
- [7] D. Bell and L. J. LaPadula. Secure computer systems: Mathematical foundations and model. Technical Report M74-244, MITRE Corporation, Bedford, Massachusetts, USA, 1975.
- [8] D. E. Bell. Concerning "modeling" of computer security. In *1988 IEEE Symposium on Security and Privacy*, pages 8–13, 1988.
- [9] K. Biba. Integrity considerations for secure computer systems. Technical Report TR-3153, MITRE Corporation, Bedford, Massachusetts, USA, 1977.
- [10] D. Brewer and M. Nash. The chinese wall security policy. In *1989 IEEE Symposium on Security and Privacy*, 1989.

- [11] D. L. Brinkley and R. R. Schell. Concepts and terminology for computer security. In M. D. Abrams, S. Jajodia, and H. J. Podell, editors, *Information Security - An Integrated Collection of Essays*. IEEE Computer Society Press, Los Alamitos, CA, USA, 1995.
- [12] D. D. Clark and D. R. Wilson. A comparison of commercial and military security policies. In *1987 IEEE Symposium on Security and Privacy*, 1987.
- [13] Communications Security Establishment, Government of Canada. *Proceedings of The 1990 CTCPEC Availability Workshop*, February 6-7 1990.
- [14] D. E. Denning. A lattice model of secure information flow. *Communications of the ACM*, 19(5):236–243, May 1976.
- [15] V. Gligor. A note on the denial-of-service problem. In *1983 IEEE Symposium on Research in Security and Privacy*, 1983.
- [16] M. Harrison, W. Ruzzo, and J. Ullman. Protection in operating systems. *Communications of the ACM*, 19(8):461–471, 1976.
- [17] J. Leiwo and S. Heikkuri. Clarifying concepts of information security management. In *Proceedings of the 2nd Baltic Workshop on DB and IS*, Tallinn, Estonia, Jun 1996.
- [18] J. McDermid and Q. Shi. A formal model of security dependencies for analysis and testing of secure systems. In *The Computer Security Foundations Workshop IV*, Franconia, New Hampshire, USA, Jun 1991.
- [19] J. K. Millen. A resource allocation model for denial of service. In *1992 IEEE Symposium on Research in Security and Privacy*, Oakland, California, May 1992.
- [20] S. Muftic, A. Patel, P. Sanders, R. Colon, J. Heijnsdijk, and U. Pulkkinen. *Security Architecture for Open Distributed Systems*. John Wiley & Sons, 1994.
- [21] D. B. Parker. A new framework for information security to avoid information anarchy. In *Proceedings of the IFIP TC11 11th international conference of Information Security*, Cape Town, South Africa, May 1995.
- [22] A. Reed. Computer disaster: The impact on business in the 1990s. In *Proceedings of the IFIP TC11 8th International Conference on Information Security*, Singapore, May 1992.
- [23] A. Silberschatz and P. Galvin. *Operating System Concepts*. Addison-Wesley, 4th edition, 1994.

- [24] F. Simonds. *Network Security: Data and Voice Communications*. McGraw-Hill, USA, 1996.