Chapter 2

Availability

2.1 Introduction

Given the threats to Information Security [79 and 80], Denial of Service attack continues to be a threat today in the form of much bigger and destructive DDoS. The main target of DoS attacks as we know is rendering an information resource unavailable or in simpler terms the main target is *Availability* of Information. Given the threat of DoS, there is a demand to study, research and analyse *Availability* for a better understanding of "Availability as a security attribute" and also given the fact that Confidentiality and Integrity are the most researched and studied attributes of Information Security [76]. The Paradigm needs to change and needs a shift from a state of sustainable Information Availability to a state of providing complete Availability, as Un-Availability is not an option in today's context, given the heavy dependence of modern world on information resources and the demand for expected delivery of services in a timely and a reliable manner. To achieve this, security practitioners need a much better understanding of Information Availability and study the factors that determine Availability and can influence it under certain conditions (i.e. DoS attack). This will help security practitioners analyse the impact of each factor within the context of their enterprises and determine the changes if necessary, that will achieve the goal of Availability of the organization's critical resources (logical and physical IT resources).

Availability of information, as already mentioned is the least discussed and researched attribute of Information Security [81]. But this does not certainly mean that it is the least important attribute of Information security. In fact, it plays an important role in determining the other attributes of information security (confidentiality and integrity) because these two attributes are directly dependent upon the Availability.

The CIA triad comprising of confidentiality, integrity and availability is the heart of information security [82]. Everything in information security revolves around these three security attributes. CIA triad is the basic model of Information Security and there exist other models that have the attributes of the CIA triad in common [83]. In Figure 2.1 two versions of the CIA model of Information Security are given, the

first one is the good old CIA triad that we see everywhere in theory and practice. The second one that we are vouching for is the more realistic one and it tells about the dependence of Confidentiality and Integrity on the Availability. The classic CIA triad raises a serious question in the mind i.e. it treats all the three attributes of security as equal while as in practice there exists a dependence. The dependence is that we can have Availability even if we don't have Confidentiality and Integrity; however we cannot have Confidentiality and Integrity if we do not have the information available

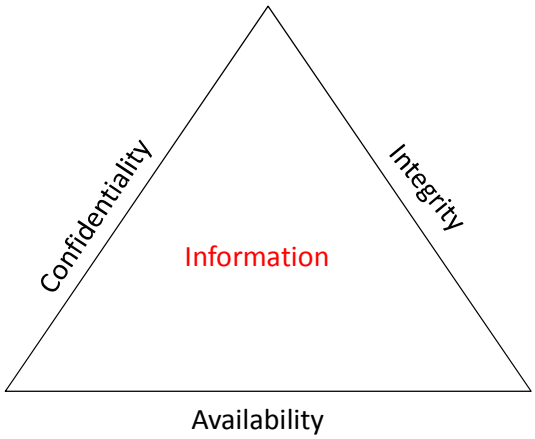

Figure 2.1 (a): Classical CIA triad of Information Security.

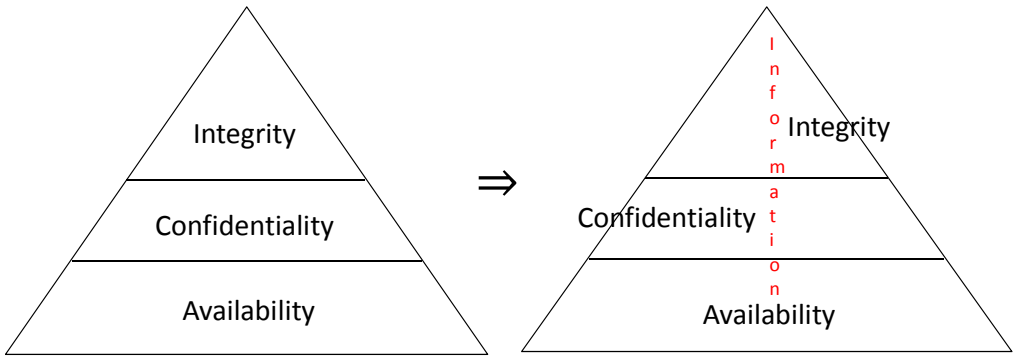

Figure 2.1 (b): Modified and more realistic CIA triad of Information Security.

(i.e. Availability) whenever and wherever we need it. Imagine if the authorized users of information cannot have access to it and cannot use it, who needs confidentiality

and integrity then? How can we apply the advanced methods of data encryption? Or the methods of access controls when there is either no access or delayed access to information? All come into picture only when authorized users have access to information and related resources when they need it. Therefore in spite of the fact that *Availability* is the most ignored and least researched part of information security, it is as important and necessarily required as a component of information security as are Confidentiality and Integrity and in fact it forms the ground for other security attributes and without it no security attribute can exist.

2.2 Understanding Availability

The meaning of Availability is diverse in Computer Science, Information Technology and Applications. Over the past few decades, it has been distinctly studied and researched much in the context of functionality and performance (i.e. computer networks, information processing systems, databases, file systems and data storage etc.) and comparatively has been acknowledged very little as a security requirement [84]. A few decades ago during the dawn of Information Security, a security criteria and evaluation document [85] of the US Department of Defence[1] (the stakeholders of networking and security at that time) was being framed, unfortunately, Availability was not considered important and worth of being a part of that document at that time. A spark was provided in the paper [87] that got attention among the security stakeholders and practitioners towards the fact that Availability cannot be left out. Some standards [88 and 92] from the stakeholders of security (i.e. government) followed and some studies [91], acknowledged the fact that Availability cannot be left out and is a key pillar for providing Information Security. All of them vouched for the information security attribute "Availability" and agreed to a common goal (some well-known and most referred definitions from various stakeholders and practitioners of Information Security):

[1] US Department of Defense's (DoD) agency DARPA is responsible for development of many technologies like computer networking and NLS ("oN-Line System," the first hypertext system) [86].

"Enable access to authorized information or resources to those who need them"[88].

Or

"Timely, reliable access to data and information services for authorized users" [89].

Or

"An authorized party should not be prevented from accessing objects to which he, she, or it has legitimate access" [90].

Or

A "requirement intended to assure that systems work promptly and service is not denied to authorized users" [79].

All the definitions and the explanations regarding Availability revolve around the definitions given above. Furthermore in [76] the Author puts Information Availability as "*the ability to make information and related physical and logical resources accessible as needed, when they are needed, and where they are needed*". With the above mentioned facts, by the year 1991 the major attributes of information security were established and identified as *Confidentiality, Integrity* and *Availability*, although some went a step ahead and recommended some extensive frameworks [75]. In the paper [87], the author explains the concept of Availability uniquely by defining denial of service if the following manner: let's take two groups of authorised users, A and B of specified services S_1 and S_2 respectively. A is said to deny service to B, if A makes the specified service S_1 unavailable to the group B for a time which exceeds the predetermined and agreed upon *service Maximum Waiting Time (MWT)*. The author has also mentioned and expressed the needfulness of *detection* and *recovery* to Availability. One more author in the paper [93] vows for detection and recovery and takes it forward and says that availability problem forces the issue of how unusual actions of legitimate users or the abnormal behaviour of their software can be detected, and how can an information system be recovered from such unusual actions. In [84], the author puts a different perspective about Availability not being a safety, a

system property or a liveness property under certain conditions explained in the same paper. The author further advocates about the malicious behaviour of the user being monitored and addressed by the information system itself and the information system should have the capability of self-detection and recovery from the mistakes or from any system misuse by the users. In the paper [94] the author puts forward the behavioural aspect of Security and terms Availability as a behavioural concept. The author explains Availability as the capability of the information system to bear its services to the legitimate users. In [76] the author explains Availability as the capability of an information system to make Information available including all the logical and physical resources reachable and accessible wherever and whenever they are needed. The author further highlights some critical issues related to Availability which often are ignored while analysing Availability like a security policy which is often found to be woeful and subsequently proves to be disastrous for an organization. The security policy should be an effective one so that issues related to Information Security in an organization are handled affectively. The author in [95] puts a slightly different view about Availability than what we have been seeing until now. The concept of Information Availability is defined as the protection against malevolent concealment of information. That means, no users are permitted to hide the information, of which another user may be having the access rights.

Until now we discussed Availability w.r.t Information and Information System/Resource. But on the other side of this exists many views of Availability w.r.t an organization, which are presented in the figure 2.2[2]**.** The figure gives us a glimpse of the fact, how important is Availability of all the other assets in an organization from the information (and related assets) perspective. The things that this thesis is concerned with is what is happening to assets at level 4 and level 2 and the point of mentioning the other assets is that those assets can impact the availability of assets at level 4 and 2 [76]. Needless to mention in today's world these two levels are the backbone of an organization in driving the workforce forward. A detailed explanation of all these assets about their availability is given in [98].

[2] Inspired from work in [98].

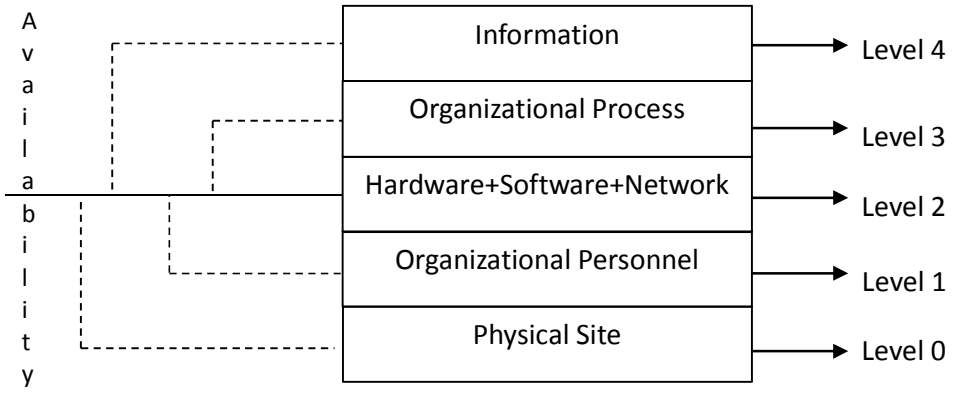

Figure 2.2: A panorama of Availability in an organization.

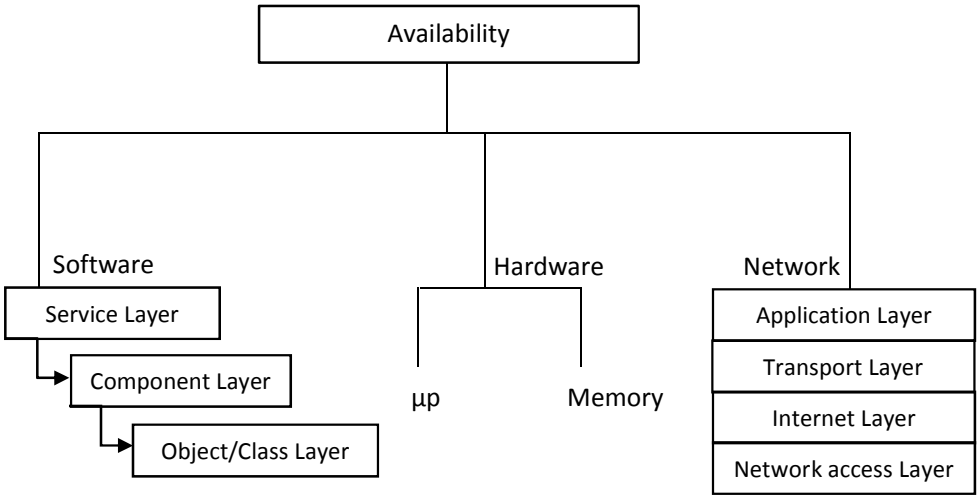

Figure 2.3: Factors that can affect Availability in an IS.

Since the working of Organizations in today's world is heavily dependent upon the use of Information and Communications Technology, the critical resource from security point of view without a doubt is the Information System and the network to which the Information System is connected. The components of an

Information System, where the dependence of *Availability* can be exhibited, are classified as shown in the figure 2.3, which are:

1. *Software*
2. *Hardware*
3. *Network*

1. Software: This is the most critical factor of the three mentioned above. The other two factors i.e. hardware and network are driven by their respective operating codes and code is what is exploited by the malicious users during an attack. All the security attacks and the corresponding remedies are addressed w.r.t the software or the operating code. So if one has to keep an information system secure one has to primarily think about safeguarding the software part first, then the rest.

This includes both the operating system code and the information processing application software of the Information system. The Software can be further divided into three levels, as mentioned in [108] while describing the architectural levels of a software. The levels are:

- *Service/System Level*
- *Component Level*
- *Object/Class Level*

Service/System Level: This is the topmost level in the application software architecture. This is the first place where the Attackers/Hackers start to exploit a system. At this level the hackers are always looking for some open services, so that they can get into the system and bring havoc. This is the external view of the system i.e. In case of Interaction between different applications programs (client server scenario on a network), a program (client agent) may interact with the services of a different program (server agent) in order to get the required task done. Now from the security perspective more the number of services open in an application program more vulnerable the system is to risks/threats from other calling programs, because it's not that hard for a malicious user to get hold of services of an application program through a legitimate way and then after getting access to the program

exploit it for misuse in a number of illegitimate ways. Thus in order to maintain a sustainable state of Availability it is necessary for an information system to have a check on the various services being accessed by various calling programs.

Component Level: This is the second level of abstraction. Components are accessed via interfaces and can provide service to any client program that provides the necessary interface required. An important thing here to consider here from a security perspective is who can access a particular component if the client programs have the necessary interface required for the access to the component. There is a need of a strong authentication and access control mechanism between the various interacting components, so that only authorised client programs can access their respective counterparts. With the access to components provided via their explicit interfaces, what is actually being achieved here is information hiding or better known as *encapsulation*. While designing a system using the component modelling approach one must not keep only security into mind, because it may affect the overall performance of the system. In [109] the author advocates the importance of addressing the issues of scalability and performance as functions for security solutions of a software system.

Object/Class Level: This is the lowest and the most fine-grained level of the software architecture. It encompasses the concepts of Object Oriented Paradigm i.e. encapsulation, information hiding, classes and objects etc. in such a way that results in efficient performance and secure functioning of the software system. Since there exists a dependency and association between all these layers, it is important if there exists a bad design or some logical loophole in the design at the lowest level it will be exploited step by step by gaining access from the above layers. Therefore an efficient and secure software system should take care of the design right from the lowest levels up-to all the layers that exist.

2. Hardware: when the application software is compromised, it may have adverse effects on the operating system or the system software of the information system, given that the ultimate goal of the attacker/ hacker is to bring havoc to the receiving

information system and to render the system unavailable for its normal functioning. Now if this is the case then the Hardware component of the system is the one that has to face the real music, since hardware is wholly and solely controlled by the software (including system software) running above it and whatever way the software behaves the hardware will simply follow. The ultimate aim of the attacker is to consume the resources of the receiving system i.e. the processing capabilities of the system may get exhausted, both the Microprocessor and the RAM may get exhausted by repeatedly giving them needless jobs in huge quantities that they cannot handle. Such jobs can be created by malicious programs that the attackers install on the target machines or can also be done in a live manner through the network i.e. DOS attack using a tool called Sockstress [110] eats up the resources to crash a service or the whole machine. This situation can be averted if we have redundancy in hardware components i.e. make use of fault tolerant systems and in such a system if a hardware component fails, alternative component takes over the required task. *High Availability* systems employ this scheme.

3. Network: when an attacker fails to intrude the target machine, the simplest thing that can be done in order to render the services of the target machine unavailable, is to attack the network to which the target system is connected. Most common network attack is to flood the network with illegitimate traffic, most lethal lately being the deadly DDOS attack, which can generate data of the magnitude of 600 Gbps [23]. Also an attacker can send invalid data to network services that may result in abnormal behaviour of services or applications. Also an attacker may block all the traffic resulting in loss of access to the network based resource to the legitimate users. The network stack i.e. the *TCP/IP* model has 4 layers as shown in the figure 2.3. An attacker can exploit protocols (security holes in network hardware as well as software) at any level in the network and thus render services unavailable to the legitimate users. A logic DoS attack for example in the Internet Protocol (IP) packet, may modify the pay load data size which may result in crashing the target OS due to a fault in the OS software.

2.3 Realm of Availability

Availability as a system property has been categorized into three [97] domains:

1) *Basic Availability:* This type of Availability is related to a standalone system that is developed with the necessary components (software and hardware) in order to achieve its functional requirements. It is as a whole a single component and has a single point of failure i.e. it will provide services to the users as long as there is no DoS attack or any maintenance procedure. In case of a failure system is not capable to oblige to any user requests.
2) *High Availability:* Can be achieved with a system that has redundancy in components (software and hardware) in order to achieve its functional requirements. The paper [96] explains High Availability w.r.t fault-tolerant systems. Fault-tolerant system has redundancy in its components and high availability systems also employ this system component redundancy [101]. If a component fails in such a system, alternate component will take over the required task i.e. if there happens to be a DoS attack on a web-server that has redundancy with many servers available, in case of a failure alternate server will take over the functioning and handling of the user requests, thus ensuring much higher Availability than a single server, which often is a single hardware component and hence has a single point of failure. Thus we can say that with fault-tolerant or redundant systems Availability can be achieved at higher rates. There are two different high Availability [100] models:
    i. *Active/Standby*: In this model, for one active component in a system one redundant stand-by component is used. It follows the fail-over model, when an active component fails; the idle redundant component takes over the working. Now the replaced idle component becomes the active one and based on the level of high availability there may be more idle components waiting for the newly made active component to go defunct.
    ii. *Active/Active*: This model of high availability contains more than one active redundant system component. Unlike Active/Standby this model does not waste any resource by keeping them idle. The number of redundant components active depends upon the level of high availability

employed. A more detailed discussion on High Availability models and further classification of the two models is given in [100].

3) *Continuous Availability:* In such a system the features of a High Availability system and Availability during planned outages are both present or in other words it extends the definition of High Availability. It makes use of the "Masking" strategy, which means masking a fault to shield it from any external observation i.e. in simpler terms making things appear as if no failure has occurred. This is achieved by means of replication of the appropriate component (hardware and software) and when a failure occurs the appropriate redundant component takes over. Masking strategy is basically used in High Availability systems. Its use is derived here for masking of planned outages like maintenance.

When we talk about Availability of an information resource (talking as a *metric)* we say it can be either 0 or 1 (0% or 100%), 0 means no-Availability and 1 means any acceptable and meaningful state of Availability (any form from the above three domains but in theory and practice we mostly refer to 1 as continuous availability). we always expect an information system to show Continuous Availability i.e. whenever we purchase something from an online store, we always expect the website to be up and never expect it be down even if we are purchasing things at 3:30 in the morning lying down in the bed. In some cases High Availability is also acceptable to a large extent. The only difference that separates the two domains is availability during maintenance, else nothing much separates them. Basic Availability is the least form of availability preferred and it can also show good performances (under favourable conditions like no attack, less concurrent users, very little maintenance etc., which is a rare thing and only possible if the information system does not house any critical resources and thus does not attract too many users).

2.4 Availability and Other Security Attributes

The security attributes *Confidentiality, Integrity* and *Availability* are distinctly defined in the context of Information Security. While *Confidentiality* and *Integrity* are defined w.r.t Information, *Availability* on the other hand is usually defined w.r.t Information

and Information System/Resource and some other domains as explained in section 2.2. Availability cannot be considered in isolation i.e. availability of information only, it must always be considered within the walls of an Information System and related assets , that may include processing resources (Hardware, Software and Networks), system and organizational processes, humans and including the information itself. In short there is no Availability of Information if the Information System is not available. Despite of the fact that there is no mention of Information System in the definitions of Confidentiality and Integrity, yet they cannot be achieved in isolation without the presence of the Information Resource or better put as, there is no Confidentiality and Integrity without the presence of Availability, while as there can be Availability without Confidentiality and Integrity.

Both Confidentiality and Integrity can be guaranteed by enforcing certain restrictions on malicious access to Information within an Information System while as Availability cannot be. As a result the security levels for Availability cannot be achieved at par with what is being achieved for Confidentiality and Integrity.

Another contrast between the security attributes is the type of threats and risks that the attributes are exposed to. Confidentiality is threatened when an unauthorised user, process or entity is interested in accessing the information intended for an authorized user. Integrity is threatened when an unauthorized user indulges in improper Information alteration or deletion or simply causes information loss. The deletion part may be some times a system fault (un-intentional) but mostly it is an intentional one. While as Availability on the other hand can be endangered by an un-intentional cause (system fault, accidental damage etc.) or an intentional malicious activity like a DoS attack that can simply render all the services of a system un-available. DoS attack can indirectly threaten Integrity and Confidentiality as well, because of the fact that Confidentiality and Integrity are dependent upon Availability.

There exists a trade-off between the security parameters. While an attribute may be strong enough in an environment, the other may be not so strong in the same environment. E.g. In systems like RAID, the redundancy may favour Availability, while as Integrity may suffer because of data duplication. Therefore we need to strike

the right balance so that all the security attributes stand firm against any threat that they possess.

A generalized security policy about the attributes of security is often conflicting between the security attributes. Since the requirements for Availability differ than those of Confidentiality and Integrity (taking into perspective a common security scenario), it can create a problem if Availability is addressed in the same manner as others in the same security policy [99] i.e. by bolstering Confidentiality or Integrity, Availability could be diminished.

2.5 Availability and the Adversaries

While providing Availability, the information system (or any other system responsible for providing security from outside) may have to go through many adversaries, which are categorized as *Faults, Failures* and *Outages* [100].

A *fault* is any variance from the normal behaviour that the system exhibits. In other words a system has a set of pre-defined functionalities to carry out and while processing, the system produces a distinctly different functionality. Whenever this happens a fault is said to have occurred. While the fault is any unforeseen disagreement within the system, a f*ailure* is a fault that is an expected variance from the normal and is visible clearly to the user. The terms fault and failure are used interchangeably whenever the difference between a visible and a non-visible fault is very minimal. The different types of faults that Availability may face are:

i. *Design Errors:* if the system is not designed correctly, then the system may not show the expected behaviour. This kind of error can be anywhere in hardware or software. E.g. Programming errors.
ii. *Hardware Failure*: failure in a hardware component due to thermal or mechanical stress. It can be a permanent one as in most of the cases hardware is not that easy to repair but this kind of fault can be a once in a lifetime one. E.g. when a network component like a switch crashes or memory crash of the server.
iii. *Overloading the system:* making the system work beyond its specified capabilities can also cause many failures e.g. if a router is capable of handling

data only at the rate of 10 Mbps, but an application is transmitting at the rates of 1Gbps. This will surely create a lots of issues like latency, packet-drop etc.

iv. *Error while execution:* there can be failure while executing a procedure e.g. a procedure on memory allocation may fail to execute if the system has no more further memory for allocation.

There are also the malicious (or byzantine) and the non-malicious (or benign) faults that can bring harm to the system. The malicious fault like a Virus or a Worm can bring a lot of harm to Availability. While as a benign fault such as a disk cash (which also can cause loss of critical data and thus deny availability) is easily detectable and correctable most of the times.

An *outage* is any variance from the specified behaviour of the system [101]. It incorporates all the things that are discussed above. Faults and failures are classified as *unplanned outages* while as in the case of a *planned outage* the system is prevented intentionally from providing services for some time in order to perform the timely mandatory operations like maintenance, software upgrade, hardware upgrade etc.

2.6 Conclusion

While providing Information Security, the security stakeholders should treat the three attributes of security i.e. Confidentiality, Integrity and Availability equally with in their respective domains wherever they are applicable. Availability is more critical of the three coz of the fact that the other two are directly dependent upon it and without *Availability* of the information system/resource as one cannot apply the methods of confidentiality and integrity without the information being available. Since *Availability* is dependent upon software, hardware and the network, the measurements regarding systems Availability in all the three cases can be done both at the system level (as a whole) and the individual component level. Critical Information processing systems are always trying to achieve *continuous availability*, which is very hard to maintain. While as most of the information processing systems settle down for h*igh availability*. *Basic Availability* is the lowest achievable form of *Availability*.