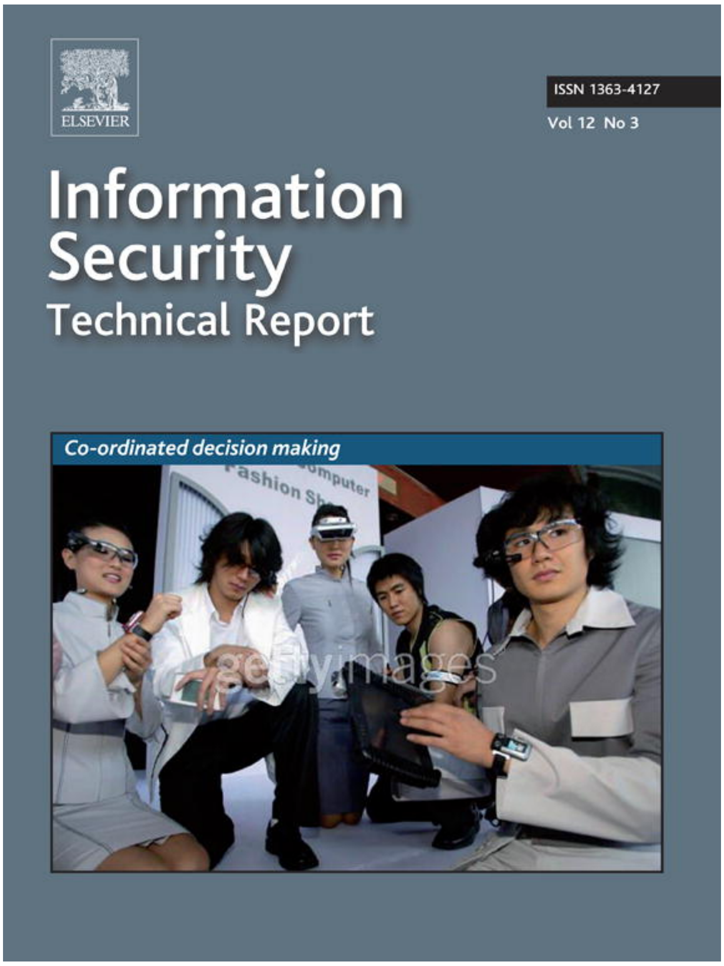

ELSEVIER
ISSN 1363-4127
Vol 12 No 3
Information Security Technical Report
Co-ordinated decision making
gettyimages

INFORMATION SECURITY TECHNICAL REPORT 12 (2007) 172–178

available at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodinf.htm

Information Security Technical Report

Ubiquitous security for ubiquitous computing

Zia Hayat[a,*,1], *Jeff Reeve*[a], *Chris Boutle*[b]

[a]*University of Southampton, United Kingdom*
[b]*BAE Systems*

A B S T R A C T

The potential for rapid and diverse interconnectivity through devices utilising heterogeneous communications interfaces has enabled a truly ubiquitous computing environment. However, this has resulted in equally ubiquitous security risks due principally to the number and complexity of services being run over such networks. As technology advances towards the realisation of a ubiquitous computing environment, what impact does this have on the traditional information security triangle, of preserving the confidentiality, integrity and availability of information? And how does this influence, future information security requirements, particularly in light of always-on business processes which require real-time information sharing? This paper describes research conducted into answering these questions. Emphasis is placed on the need for risk management, and how this may be achieved through context-based access control mechanisms and pro-active threat assessment techniques.

1. Introduction

One of the major changes in computing networks over the last few years is the growing ability to conveniently form agile business processes through paradigms based upon wireless ad hoc communications, service oriented architectures (SOA) and in the longer term grid (or utility) computing. This has resulted in rapid and complex interconnections between devices previously unimaginable. Such complex interconnections result in open and dynamic connectivity across traditional networking boundaries not only at a personal or home and small office level, but increasingly in the constrained environment of the larger enterprise network, where sensitive information and services must be protected. In order to secure such open networks a risk management approach must be adopted, enabling pragmatic protection based upon business requirements.

The majority of enterprise users today utilise computing devices with constrained communications for a limited set of well-defined services, however, in the near future these devices will be used to deliver much more varied and flexible services using the most functionally and cost effective communications technique available. This can be seen in the desktop computer which is no longer only enabled for communications over the wired enterprise backbone, but is increasingly capable of communicating through a multiplicity of interfaces such as Infrared and Bluetooth. In addition to the desktop other devices such as the personal digital assistant (PDA) are also being introduced enabling ubiquitous networked computing capabilities through mobile and heterogeneous (WiFi, GPRS, WiMAX, etc.) communications.

From a business perspective ubiquitous computing is facilitating always-on processes, where users can access any information-based service at any time. Therefore ever more critical decision making is based upon information delivered over computing networks. From an information security point-of-view the confidentiality and integrity of such information must be preserved. Availability is the third dimension

[*] *Corresponding author*. Present address: Accenture Consulting, London, UK.
E-mail address: zia.hayat@accenture.com (Z. Hayat).
[1] Sponsored by BAE Systems.

1363-4127/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.istr.2007.05.002

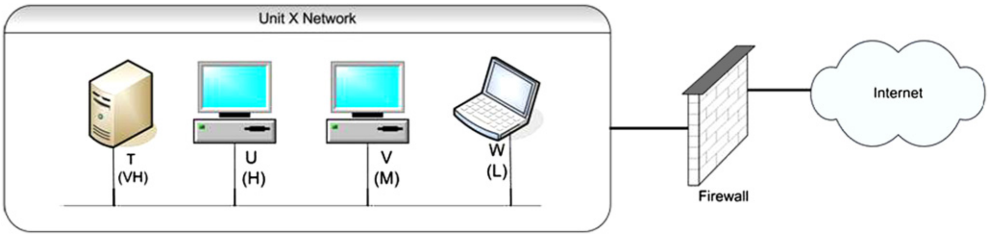


Fig. 1 – Traditional networking scenario.

of information security and has traditionally related to robustness. However, in an always-on business environment it is argued that availability necessitates *timeliness* on a par with robustness, where timeliness implies that all relevant information is delivered to authorised entities at the correct time. Timeliness is essential in enabling processes which require real-time information sharing.

An example of how the traditional internetworking environment is evolving is given in Figs. 1 and 2, respectively. From Fig. 2 it can be seen that a much more complex and meshed environment is developing, where every device has the potential to become a gateway to external networks. The complexity introduced by such extended connectivity adds to the threat vector in information technology (IT) networks. Although the business benefits of using a more flexible approach to internetworking are clear, the complexities introduced pose significant technical challenges, not least from a security point-of-view. The Jericho forum (Bleech, 2005) addresses such challenges in its vision of de-perimeterised security solutions. De-perimeterisation aims to provide tailored security services according to the diverse needs of individual entities (users and their devices). This can already be seen to be taking place, where anti-virus, intrusion detection and firewalls are embedded on individual devices, supplementing current centralised security solutions.

Principally de-perimeterisation acknowledges that the monolithic security solution, that is the traditional organisational firewall, is no longer appropriate, due to the mobility of user devices as well as the complexity of protocols being run. Therefore information can simply traverse and therefore subvert such a solution, by going around it in the case of mobile devices or through it using an increasing number of complex protocols which may be tunnelled via authorised techniques such as hyper text transfer protocol (HTTP) and simple mail transfer protocol (SMTP). Hence the concept of de-perimeterisation aims to evolve security practices in-line with the evolution of computing and communications networks. As a result just as computing capabilities have moved towards a distributed model, de-perimeterisation advocates the need for a distributed security model. Ultimately it is felt that such a model will provide increasingly agile, dynamic and secure information solutions. For example the excessive overheads associated with tasks such as deep packet analysis mean that only a subset of the total traffic can be scanned by centralised security solutions, otherwise users would be subject to significant latencies as described by Network Appliances Inc. (2005). Therefore a de-perimeterised methodology can be seen to support the requirement for more timely security solutions.

A key factor in achieving de-perimeterisation is the ability to effectively control access to distributed and ad hoc services which are being introduced by ubiquitous computing capabilities. For example if a user is invited to a meeting, then access to the meeting room itself as well as other users' devices for presentation material should be automatic and seamless. The authorisations required to enable such access control must be context-based in order to preserve the principle of least-privileges. This would provide automatic and fine-grained access control which is currently achieved through human intervention. Such human management of increasingly complex networks is proving to be ineffective as described by Chao (2006), where it is stated that the cost of managing ubiquitous networks is now exceeding equipment costs by a ratio of up to 18:1.

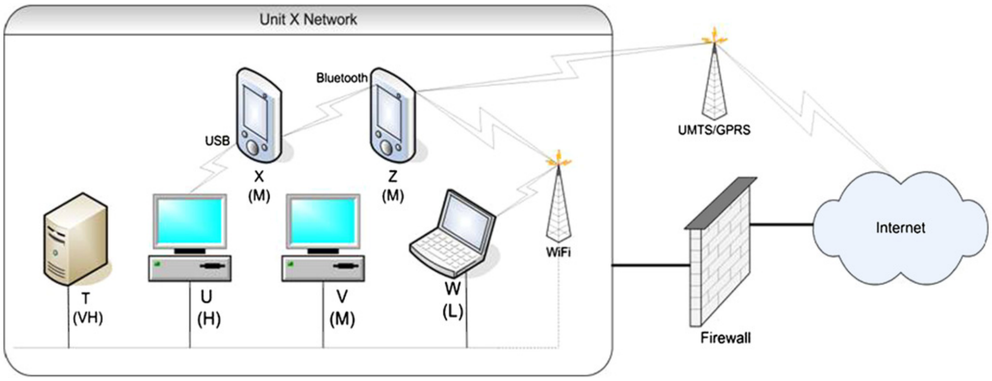


Fig. 2 – Evolving networking scenario.

It is envisaged that security issues will still exist in networks which adopt new security paradigms such as deperimeterisation. One such issue arises due to the shift towards an SOA, which requires open and diverse communications resulting in increased security risks. This is due principally to the fact that many of the exposed services are likely to be based upon common components (services) resulting in a software monoculture. From a security point-of-view this introduces an inherent risk by reducing the potential for defence-in-depth. Therefore if a flaw is identified in a particularly popular service then all instances of that application will be vulnerable to exploitation potentially causing significant disruption to critical business operations.

The next section highlights the use of context-based (context-aware and context-dependant) access control technologies and the state of the art in this field. Section 3 details the compromise path threat analysis technique used to identify and analyse risks in ubiquitous networks, with a corresponding risk model. A brief summary is then given in Section 4.

2. Context-based access control

Access control is based upon the subject–privilege–object methodology, where a subject (user) has an associated privilege (read, write, etc.) with respect to some object (information service). Traditionally access privileges have been based upon identity, where a list of user identities authorised to access a service is pre-defined, the most prevalent access control model is known as an access control list (ACL). In a ubiquitous and SOA computing environment, access control based upon identity alone inappropriately reflects users' needs and authorisation to access services. This is due to the increasing use of plug and play computing services, which are context-based. Contextual attributes such as a user's temporal and geographic location as well as available bandwidth may be used to more accurately control access to services. The potential for context-based access control has previously been identified in Zhang and Parashar (2003), Hu and Weaver (2004), Toninelli et al. (2006) and Yokoyama et al. (2006) where it has been proposed to make dynamic access control decisions based upon measurements of contextual attributes associated with users, such as location.

To date the most advanced work in the area of context-based access control is that undertaken the mobile workers' secure business applications in ubiquitous environments (MOSQUITO) programme (MOSQUITO, 2005). In MOSQUITO a comprehensive framework has been specified to enable context-aware access control. The framework incorporates various technologies which define the format of user credentials, security policy and the process of security decision making. User credentials are implemented using EureCA certificates which are based upon the concept of an attribute certificate (AC), as introduced by Farell and Housley (2002). An AC is a digital certificate based mechanism which can be used for both authorisation as well as authentication purposes. The EureCA certificate uses an extensible mark-up language (XML) schema to describe a public key infrastructure (PKI) based certificate, which must be digitally signed by an authority trusted by the service provider. A snippet of an example AC is given in Fig. 3, where it can be seen that as well as including the public key for authentication of identity, additional attributes such as a user's current location are provided as additional constraints for accessing a specified service. In the MOSQUITO programme security policies are defined using extensible access control mark-up language (XACML).

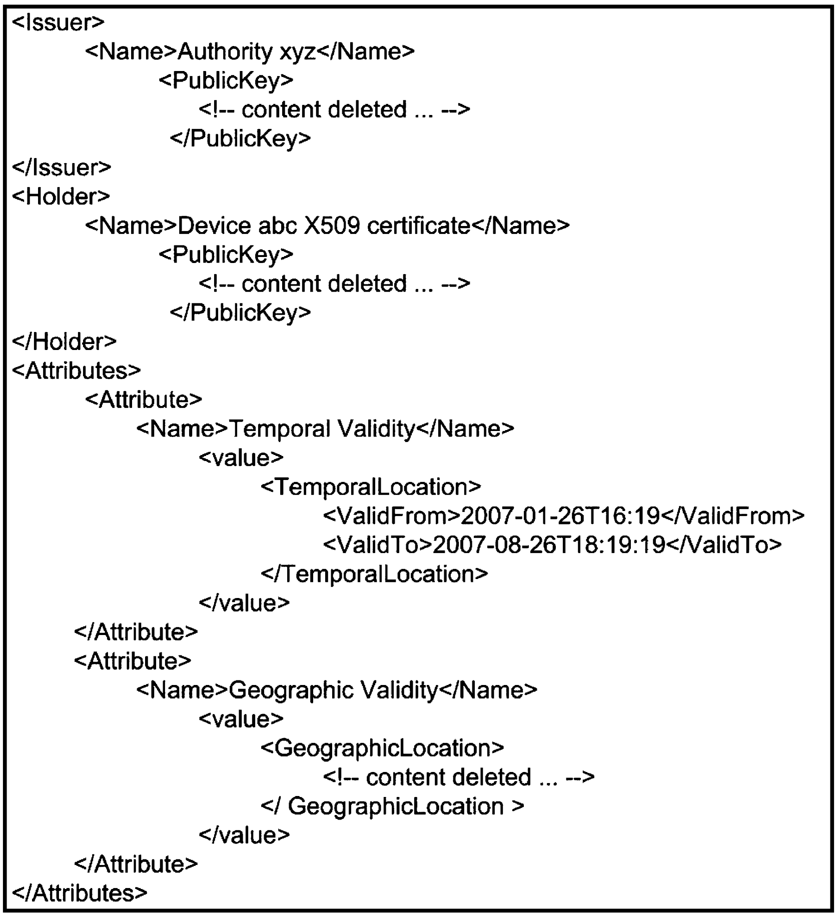
In context-aware access control systems context information (CI) such as a device location must be asserted by a trustworthy source. This would enable CI to be used by service providers to make security decisions, such as not displaying secret content in a public area. Therefore security policies incorporating context must be associated with each service. In the MOSQUITO programme a user's CI is delivered to decision points through a context-aware trust and security (CATS) mechanism, which consists of three components: CATS engine, context information broker, and context information handler. All three CATS components are installed on all mobile devices. When a service is requested the service provider (or designated security decision maker) requests necessary credentials and CI to make an authorisation decision. A high-level overview of CI elements in the MOSQUITO framework is given in Fig. 4, where it can be seen that all CI is managed by the various components of the CATS architecture. Other systems for gathering trustworthy CI have been developed such as the context toolkit by Abowd and Dey (1999).

The overheads associated with executing secure CI architectures such as CATS on mobile devices may prove to be a limiting factor of context-aware access control. Therefore in Hayat et al. (2006a) it is proposed to use context-dependant rather than context-aware access control mechanisms, for ubiquitous computing applications such as decentralised information sharing in a mobile ad hoc network (MANET). Context-dependant access control assumes that a user's real-time CI does not have to be assessed, instead authorisations asserting the context in which they may operate (e.g. location, time and bandwidth consumption) can be pre-defined. This can be seen in potential location based services such as road traffic information, which may be purchased by motorists in advance to access the latest information on a given road(s) for some period of time.

In order to achieve efficient and effective business processes based upon ubiquitous computing, automated and fine-grained context-based access control is required. Credentials such as the EureCA certificate described previously can be used to deliver such automated authorisation mechanisms. This would reduce the need for manual processes which are currently used in many ad hoc computing applications, where users distribute external media devices such as USB to share necessary data. Such manual processes limit accounting and audit trails whilst crucially reducing the timeliness of information sharing in real-time business environments.

3. Compromise path threat analysis

If one assumes that in the future services will be locked down (i.e. only available to a well-defined set of users) by distributed security solutions such as embedded firewalls, then the rate of spread of malware may be significantly reduced, as only



```
<Issuer>
        <Name>Authority xyz</Name>
                <PublicKey>
                    <!-- content deleted ... -->
                 </PublicKey>
</Issuer>
<Holder>
        <Name>Device abc X509 certificate</Name>
                <PublicKey>
                    <!-- content deleted ... -->
                 </PublicKey>
</Holder>
<Attributes>
        <Attribute>
              <Name>Temporal Validity</Name>
                     <value>
                            <TemporalLocation>
                                   <ValidFrom>2007-01-26T16:19</ValidFrom>
                                   <ValidTo>2007-08-26T18:19:19</ValidTo>
                            </TemporalLocation>
                     </value>
       </Attribute>
       <Attribute>
              <Name>Geographic Validity</Name>
                     <value>
                            <GeographicLocation>
                                   <!-- content deleted ... -->
                            </ GeographicLocation >
                     </value>
       </Attribute>
</Attributes>
```

Fig. 3 – Snippet from an attribute certificate.

a limited set of authorised (or subverted) entities will be capable of exploiting any vulnerabilities. However, malware may still be able to compromise all devices by using indirect as well as direct routes. The restricted connectivity between devices due to policies enforced by embedded firewalls, Virtual Local Area Network's (VLAN) and other directory services such as Windows Active Directory, can be seen to provide a limited and computationally feasible search space for threat

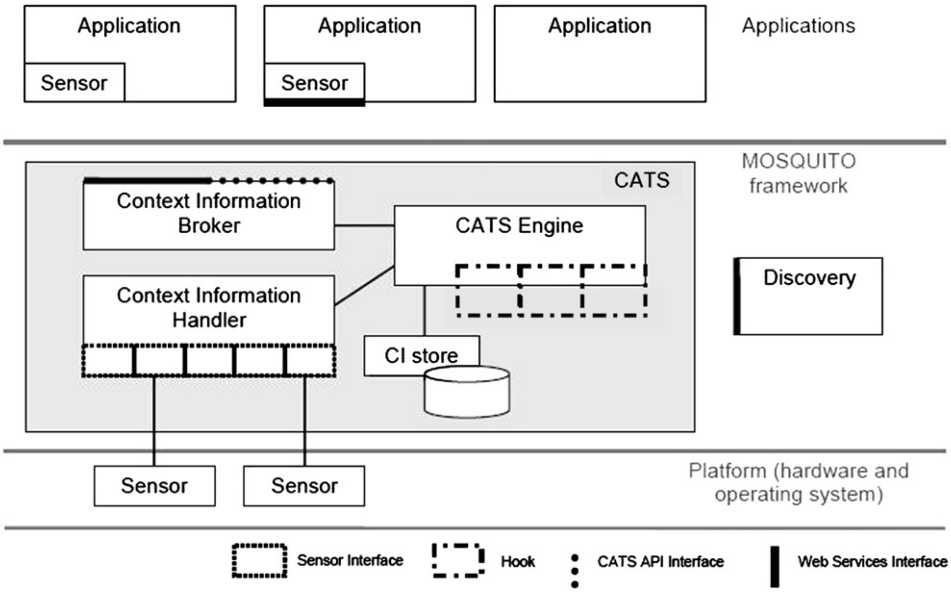


Fig. 4 – Overview of CI provision through the CATS architecture in the MOSQUITO framework.

analysis purposes. This is unlike the completely connected graph of internal networks of the past. In previous networking models all entities on the 'inside' of the organisational firewall are assumed to be equally trusted and all external entities are perceived to be equally 'untrusted'. Such a binary security model not only increases security risks due to insiders but also limits communications and therefore business effectiveness with external entities.

In Hayat et al. (2006b) the compromise path threat analysis is introduced, where the aforementioned restrictions on connectivity between devices are used to limit compromise connectivity due to vulnerabilities in services. If an attacker wants to compromise a given device(s) (victim) using a worm for example, then they must launch an attack from a device(s) which is authorised to connect to the victim otherwise the attempt to connect will be rejected. We call such compromise connections as compromise paths, where a compromise path consists of one or a series of compromise connections. This is illustrated in Fig. 5, where the link between devices 'T' and 'U' is a compromise connection and the chain of connections between devices 'T' and 'Z' (T–U–X–Z) is known as a compromise path.

The malware considered here is that which spreads truly automatically (i.e. through the *backdoor*), without any user intervention. Therefore malware transported using manual (floppy disks, USB tokens, etc.) and social engineering (i.e. e-mail requiring a user to open an attachment or navigate to spoof web sites) techniques is not considered. Compromise connections may be systematically[2] exploited by a sophisticated attacker using proprietary exploitation tactics or even a less sophisticated attacker (script kiddie) using 'off the shelf' exploits. A series of compromise connections (one or more) between a victim and attacker device is called a compromise path. Numerous other techniques (Schneier, 1999; Salter et al., 1998; Surdu et al., 2003; Moore et al., 2001) exist for the identification, analysis and ranking of risks in a network scenario. However, a major limitation with all of these techniques is that they assume an exhaustive search of the problem space (i.e. identify individual risk(s) to device(s) and then quantify these based upon the attacker model), which can be vast and complex in the case of information security vulnerabilities in modern networks.

In Ammann et al. (2005) the authors describe a technique commonly used by penetration testers, which is analogous to compromise path threat analysis identifying and analysing the chain of exploits an attacker may pursue to circumvent the security of a victim device. The chain is categorised as one of two levels (normal or super) according to the subverted user's privileges, where the maximal compromise is assumed. This means that if an attacker can gain both normal and super user privileges on a victim device the super user exploit is associated with that chain. We add to this by further distinguishing between compromise (exploit) paths by considering the impact of an attack on the overall business process. This is achieved by assigning a criticality level to individual devices, which is subsequently used to assess the likelihood and impact of an attack.

In previous techniques such as that described in Ammann et al. (2005) all information regarding the type of attack is discarded, as it is not deemed useful when considering the maximum potential impact of a chain of exploits between two (victim and attacker) specified devices for penetration testing purposes. However, the vast majority of malware such as viruses and worms exploit a finite set of vulnerabilities to infect and spread, hence when modelling the potential impact of a specific worm it is important to consider the connectivity between devices due to the vulnerabilities it exploits. This is illustrated in Fig. 5 where the potential connectivity between devices based upon authorisations to access vulnerable distributed services is highlighted.

For the purposes of simulation and testing semantic network modelling is used as described in Ammann et al. (2005) and Monahan (2005) to represent compromise network connectivity, where a compromise connection (undirected edge) represents the potential to exploit a flaw(s) by one device (vertex) in another therefore enabling the spread of malware such as viruses and worms. For example if a device hosts a particular web service application, which is subsequently found to have a flaw (e.g. buffer overflow), then all external devices authorised to access the application are considered to have a compromise connection to the device hosting the vulnerable application. Detailed analysis of how an attacker may subvert a particular service such as snooping on unencrypted password transmissions for services such as **telnet** or **ftp** is not considered, in our current model. However, such detail on the *how* of exploitation may be considered in future work to more faithfully assess the risk of compromise based upon difficulty.

A criticism of using such a model driven approach for the description of a network architecture is that the model may quickly become out-of-date particularly where mobile and ad hoc devices are used, which have the ability to offer and consume services much more dynamically, therefore creating and destroying potential compromise links. However, we consider the maximal potential connectivity (i.e. assume full connectivity) within a network of devices of interest. This is a straightforward requirement due to the relatively static nature of firewall access control rules. As well as this it is possible to identify dynamic compromise connectivity using a number of mature standards such as the simple network management protocol (SNMP) and Common Information Model (CIM) with corresponding tools such as **Cheops-ng**, **Nmap**, **Nessus**, **Retina** and **HPOpenView**. Such tools provide semantically rich network topology information in real-time. These standards are capable of providing descriptions in a number of formats, which may be used by techniques such as that described here for subsequent processing and analysis.

Overall the aim of our compromise threat analysis technique is to achieve pragmatic risk management in environments, where services are offered and consumed by processes in order to fulfil business requirements. This is unlike a risk prevention methodology which would result in limited interactions and therefore reduced business effectiveness. Hence different risk models may be adopted by

[2] An attack may use a number of exploits to compromise the security of a specific device in a specific manner; therefore we consider the compromise connectivity of more than one exploit simultaneously.

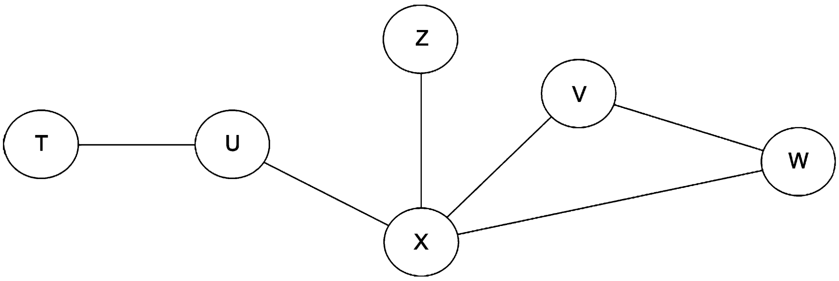


Fig. 5 – Compromise connectivity due to vulnerability in a networked application for the network given in Fig. 2.

individual risk owners according to their risk appetite, this results in differing levels of security as well as communication and therefore business process efficiency.

The principal use of the compromise threat analysis technique has been for prioritising the order in which individual devices receive security servicing in the event of a flaw being identified in a service; this is advocated in Monahan (2005) and Rogers and Allen (2002). Prioritisation is achieved in [iee_pri] by associating a valuation and risk with individual assets (devices). Following this threats and their potential impact can be analysed, based upon the risks to the devices under consideration. It is believed that a prioritisation strategy is essential in light of the rapidly decreasing time between vulnerabilities being discovered and maliciously exploited by malware. Currently it is left to the systems administrators' discretion to choose in which order to protect individual devices. However, such an arbitrary method introduces an unacceptable level of risk to the security of those devices, which are critical to business processes. In Hayat et al. (2006b) the concept of a Vulnerability Period (VP) is introduced. The VP for each susceptible device is the time between a vulnerability first being reported to the time at which that device is made secure from such a vulnerability. The VP is defined as follows:

$$\mathrm{VP} = P\tau + K \tag{1}$$

where P is the priority (integer value starting from '1', which is the highest priority) assigned to a given device, τ is the average time taken for a service to be successfully performed per device, τ is also variable due to differences in: individual services, network latencies and dynamic characteristics of individual devices; K is the time taken for the developers to provide a solution to fix the vulnerability from the time of its discovery. Therefore from Eq. (1) it can be seen how prioritisation unlike an arbitrary technique can aid the owners of devices to control the risk exposure to a given device by reducing or increasing its VP accordingly. Without prioritisation the VP can be any time between 1 and n (number of devices under consideration) times the average time (τ) taken to service a single device.

4. Conclusion

Modern computing networks have become increasingly ubiquitous, this has enabled an always-on information society needing to utilise information-based services in real-time. In order to sufficiently protect such services a shift towards deperimeterised security solutions has been advocated. However, a key requirement of any such solutions is the need to adhere to changing user requirements. Such changes can be seen to impact the traditional information security triangle, where the aspect of availability has evolved to necessitating timeliness on a par with robustness. This paper has discussed context-based access control and compromise threat analysis as two candidate information security risk management technologies, which aim to contribute to changing information security requirements. It is believed that this form of information security solution will ultimately enable business processes.

REFERENCES

Abowd GD, Dey AK. The context toolkit: aiding the development of context-aware applications. In: Human factors in computing systems. ACM Press; 1999. p. 434–41.

Ammann P, Pamula J, Ritchey R, Street J. A host-based approach to network attack chaining analysis. In: Proceedings of the 21st annual computer security applications conference; 2005.

Bleech N. Visioning white paper, what is Jericho forum? Available from: <http://www.opengroup.org/projects/jericho/uploads/40/6809/vision_wp.pdf>; 2005.

Chao L. Autonomic computing. Intel Technol J 2006;10(4) [Preface].

Farell, S, Housley R. An Internet attribute certificate profile for authorization. RFC 3281; 2002.

Hayat Z, Reeve J, Boutle C, Field M. Information security implications of autonomous systems. In: Proceedings of the 25th IEEE MILCOM conference; 2006.

Hayat Z, Reeve J, Boutle C. Prioritisation of network security services. IEE Inform Secur 2006b;153(2):43–50.

Hu J, Weaver AC. A dynamic, context-aware security infrastructure for distributed healthcare applications. In: Proceedings of the first workshop on pervasive privacy security, privacy, and trust; 2004.

Monahan B. Infrastructure security modelling for utility computing. HP Laboratories, Technical Report HPL-2005-04; 2005.

Moore A, Ellison R, Linger R. Attack modelling for information security and survivability. Technical Note CMU/SEI-2001-TN-001, Carnegie Mellon University; 2001.

MOSQUITO Programme. Specification of context-sensitive security infrastructure; 2005.

Network Appliance Incorporated. Antivirus scanning best practices guide. Network Appliance Incorporated; 2005. Technical Report TR 3107.

Rogers L, Allen J. Securing information assets: security knowledge in practice. CrossTalk – Defense Software Eng J, <http://www.stsc.hill.af.mil/crosstalk/2002/11/rogers.html> 2002 [US Air Force].

Salter C, Saydjari O, Schneier B, Wallner J. Toward a secure system engineering methodology. In: Proceedings of new security paradigms workshop; 1998. p. 2–10.

Schneier B. Attack trees: modelling security threats. Dr. Dobb's J, <http://www.schneier.com/paper-attacktrees-ddj-ft.html> 1999.

Surdu J, Hill J, Dodge R, Lathrop S, Carver C. Military academy attack/defense network simulation. In: Proceedings of advanced simulation technology symposium. Military, government, and aerospace simulation; 2003.

Toninelli A, Montanari R, Kagal L, Lassila O. A semantic context-aware access control framework for secure collaborations in pervasive computing environments. In: Proceedings of the international conference on semantic web; 2006.

Yokoyama S, Kamioka E, Yamada S. An anonymous context-aware access control architecture. In: Proceedings of the international workshop on managing context information and semantics in mobile environments; 2006.

Zhang G, Parashar M. Dynamic context-aware access control for grid applications. In: Proceedings of IEEE/ACM international workshop on grid computing; 2003.