On the Security Controls of Portable Computing Devices in Healthcare Environments

J.S.Briggs, V.Katos and A.Bhaludin

School of Computing, University of Portsmouth, UK.
e-mail: {jim.briggs, vasilios.katos, azzri.bhaludin}@port.ac.uk

Abstract

This paper reports on the findings of a study relating to the use of portable computing devices in hospitals. The investigation used as a vehicle a recently purpose developed mobile computing infrastructure, but the findings and conclusions have been generalized and could be applied to most healthcare contexts. The two main factors relating to the higher risk of the introduction of a wireless network infrastructure are the uncontrolled boundaries of the corporate (wireless) network, and the elimination of the physical security assumption concerning the mobile computing devices. The methodology adopted in this paper considers the new risks attributed to the threat vectors capable of exploiting vulnerabilities of a mobile network and proposes policy controls and considerations in order to diversify the risks of these threats being successful.

Keywords

Medical Informatics Computing; Personal Digital Assistant

1. Introduction

The benefits of a wireless over a wireline network are well acknowledged by an increasing number of corporations (Fichman and Cronin, 2003). Flexibility and pervasiveness enable an employee to continually access corporate information; on the other hand, the employee is able to efficiently transfer information to the fulfilment centre of the organisation, contributing to improved and timely decision making and resulting in a higher added value overall.

In the context of a health care environment, the above generic benefits are realised by considering the health care processes and tasks of the various roles such as nurses, doctors, and so on. An increasing number of hospitals (and in some cases, individual clinicians) are employing portable wireless devices to deliver clinical or clinically-related information services (Lu *et al.*, 2005). The devices involved may be personal digital assistants (PDAs), laptop or tablet personal computers, mobile phones and other similar electronic devices. The applications to which they may be put include entirely clinical ones (e.g. electronic patient records, decision-support systems), personal communications (e.g. email and web access), and auxiliary ones (e.g. ordering patients' meals). As expected, the adoption of these devices is not without challenges (see for example the research by Li *et al.* (2005) for a list of identified critical adoption factors).

This paper considers the use of PDAs or similar devices with a "cradling" feature in a health care environment, and explores the need for a security policy that addresses the increased risks associated with the introduction of these PDAs. We argue that the consideration of a cradling feature is a mandatory requirement, as this is used for creating a "security point of reference", in order to compensate for the loss of the assumption of physical security. We use the term Portable Wireless Device (PWD) to refer to a PDA type of device with the following characteristics:

- The PWD has wireless capabilities such as IEEE 802.x
- Bluetooth or infra-red connectivity is not present (or can effectively be disabled)
- The PWD and associated servers may be "aware" when the PWD is connected to a cradle
- There are location-aware services, such as GPS or WiFi-based location establishment.

These characteristics capture the main points arising from an extended dialogue with the developer of a portable wireless system designed for hospital use, together with our personal experience of PDA use, the literature on the security of wireless devices, and our knowledge (first and second hand) of typical hospital practices.

The rest of the paper is structured as follows. In Section 2, risks to confidentiality, integrity and availability are put into the healthcare environment context, identifying the relevant vulnerabilities introduced when considering these PWDs. Section 3 presents security policy components, which should be adopted to address the risks associated to the identified vulnerabilities. Section 4 presents the conclusions to this paper.

2. Security implications

A high level diagram of the healthcare information system used as a research vehicle in this study is shown in Figure 1. The PWDs connect wirelessly to the Clinical Information System layer, which can be thought of as a staging server which in turn interfaces with the backend hospital and legacy systems. This study focuses primarily on the PWD environment and the Clinical Information System components.

In terms of security, the devices are vulnerable to many threats which are common to conventional (i.e. non-portable) applications and devices in a hospital setting, and some which are due to the portable nature of the devices. Many of the security risks in this context can be considered as being of "high risk", partly due to the sensitivity of the data being processed, and partly due to the environment in which the systems operate. This paper reports on the outcome of the risk assessment and classification process on vulnerabilities with respect to the devices operating in a healthcare environment. In Figure 2, a taxonomy of potential vulnerabilities is presented. The taxonomy can be mapped to threat vectors by starting from a leaf node and working

up towards security or privacy goals. For instance, a threat exploiting access to multiple records would result in a breach of privacy. These vulnerabilities are further explained in Table 1.

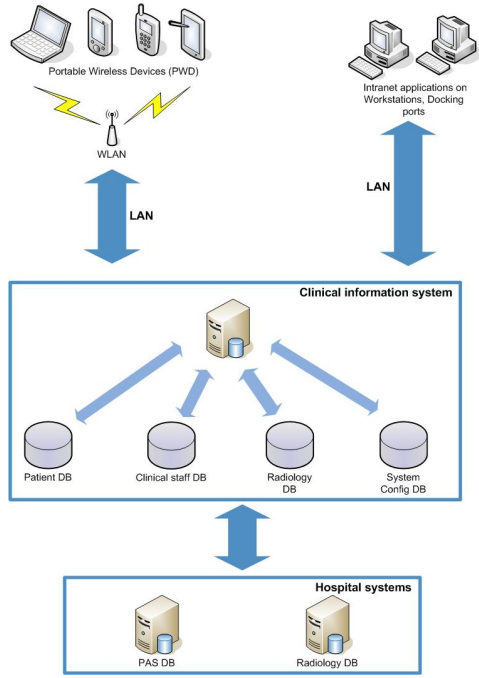

Figure 1: Main components of the Healthcare Information System

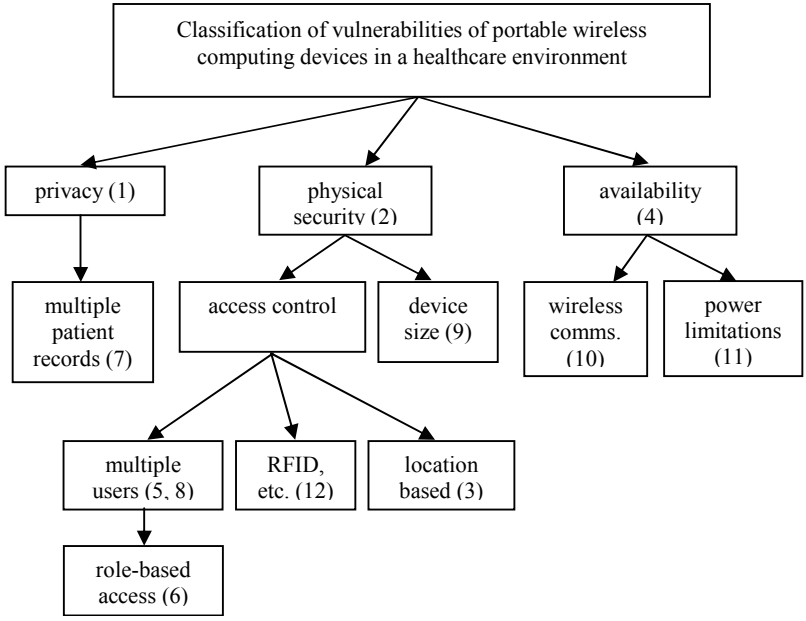

Figure 2: A taxonomy of potential vulnerabilities

1	Because the information collected and stored is about an identifiable individual's physical or mental health or condition, it constitutes "sensitive personal data" within the meaning of directive 95/46/EC of the European Parliament, the UK Data Protection Act 1998, the US Health Insurance Portability and Accountability Act (HIPAA) 1996 and similar legislation in many other countries.
2	The portable device may be operating in an environment occupied not only by doctors, nurses and other hospital employees, but also by members of the public in the form of patients and visitors.
3	The environment may be somewhat limited in that the devices are only intended for use within a particular hospital or designated part of a hospital (e.g. a particular ward). There may be no requirement for the device to be usable outside its designated zone, and indeed there may be a requirement for it not to be.
4	Because the information processed within a system may be used for clinical decision-making, the risks arising from its non-availability or inaccuracy are significant.
5	Each device may be used by several users and therefore the system may need to block access to cached or other data created and used by the previous users.
6	The users of the devices may have different roles (e.g. doctor, nurse, or administrative staff).
7	Each device may be used to process data on several patients.
8	In some locations where portable devices are used, there may be a high proportion of temporary and/or new staff.
9	Often the devices are (by their nature) small and portable. They can be easily concealed in clothing or a container.
10	Communication between the devices and the rest of the system is normally primarily by wireless network, though a second mode of use through a cradle or docking station (or similar physical connection) may be available.
11	The battery in the device has a limited charge. Each device needs to be returned to a recharging point periodically.
12	There are an increasing number of complementary technologies that portable wireless devices can integrate with. One example of this is RFID (radio frequency identification).

Table 1 - Potential vulnerabilities list

3. Security policies

This study adopts a three-tiered view of the organization. More specifically, the following layers are considered:

- Infrastructure view. This consists of the network infrastructure components, including the PWDs, network access points and servers. The security policies at this layer typically describe system administrator tasks and responsibilities, as well as monitoring and logging.

- Application view. This layer relates to the user experience and the user's interaction with the system via PWD user interfaces. Security policies in this level deal with (technical) user authentication issues.

- Organization view. This layer deals with non-technical security processes such as acceptable use policies, user awareness programs, and so on.

In the following subsections we describe the policy components that need to be in place in relation to the above views.

3.1. Infrastructure level policies

The introduction of wireless connectivity into the network infrastructure, as well as the use of the vulnerable PWDs, effectively waives the physical security assumption. This is because the boundaries of a wireless network are not clearly defined (the network perimeter is expected to extend outside the physical boundaries of the hospital building), and the PWDs are inherently vulnerable to loss or theft. As such, the security policies need to enforce encryption of the wireless traffic and stronger access control policies.

3.1.1. Re-cradling

Cradling attempts to create a point of reference in order to claim back some of the benefits and advantages of physical security and support confidentiality.

In the event that a device would not be returned to an appropriate base within an appropriate time, the existence or not of the wireless network could dictate the course of action. If there is a wireless signal, the device could report the event, and the server could "page" the device with the request. That could trigger a timeout for a return to base. Upon expiration of that timer, an incident response process could be triggered. For a more relaxed policy, upon the server's paging of the device, if there is an authentication token (e.g. an RFID tag built in to the ID card belonging to an authorised user) detected as being present, the user could be logged out and challenged to re-authenticate, in order to "buy" some time. This could be repeated a finite number of times, indicated by the policy. If there is no wireless signal and no authentication token present then there is a probability that the device is either misplaced or in the possession of an unauthorised person, and a data deletion process should be initiated.

Re-cradling is a key aspect of the security of the PWD infrastructure and is further justified in the security states section below.

3.1.2. Location based access control and location awareness

For portable devices that contain GPS components or for WiFi devices with defined access points (and also GSM-based devices where it is possible to estimate the location of the device), access control policies should incorporate the location where the device is being used.

The relevant policy should state which users should legitimately use the PWD in certain locations. For example, a nurse may only be allowed to use a device when on the hospital ward that is his/her usual place of work.

If the device finds itself in an inappropriate location, a similar policy to the one suggested for re-cradling could be employed. If the location is known but not legitimate (e.g. it is connected to a different wireless access point), then a simple message to the user asking them to take the device "home" might be sufficient. If the device is not returned within a specified time, or is in an unknown location, a data deletion process may be initiated.

3.1.3. Data deletion

As a security precaution, the portable device could be programmed to delete data should it find itself in unreliable states. As described in the sections above, these may be because it finds itself in an unrecognised place or without authenticated use for a period of time. There should be a policy describing a hierarchy in the deletion of data, depending on the state of the device and the conditions in which the incident response was initiated, in order to allow data recovery to a certain extent if it is proved that no attack has taken place. Table 2 shows a progressive deletion alert level policy.

Level 1	Delete the short-term security context data. This involves deletion of the session keys, and in general the security information which relates to accessing sensitive data. Recovery of the data would only be possible if the device reconnected with the server.
Level 2	Level 1 plus delete the long-term security context data. This involves the user data. From this point and onwards there is no option for a user login. The sensitive data can still be accessed if the device is returned to base (re-cradled).
Level 3	Level 2 plus deletion of the sensitive data. At this stage the device would require a hard reset and re-initiation which can only be performed when it is returned to base.

Table 2: Deletion levels

The timings of these levels should be specified by the policy and may differ between re-cradling policy violations or the switch-on policy violations. The concept of the various “alert levels” where a device performs security related tasks in a proactive manner is commonly found in security frameworks for mobile devices (see for example Clarke and Furnell, 2006).

3.1.4. Tamper resistance

Ultimately the security of information stored on a portable device (in terms of confidentiality) relies on the degree of resistance it can provide to an attacker who steals it, as physical security cannot be assumed. The nature of portable devices often lends them to easy concealment and removal from their normal environment.

If by physically possessing a device, an attacker can read its memory (which is theoretically the case to someone with the appropriate hardware equipment), then steps must be taken to protect the data held therein. Cryptographic solutions will often solve the first level problem of protecting the data itself (though if by copying

it a brute force attack becomes possible, this may still carry theoretical risks), but there may be a second level problem of how to protect the cryptographic keys themselves from disclosure. For that, it may be necessary to employ a device that contains a special tamper resistant component that is designed to thwart direct attempts to access them. This can be used to store the cryptographic keys, and the rest of memory is protected by using those keys to encrypt information stored in general memory.

3.1.5. Access logging

A standard security policy is to keep records of who uses a system and what transactions they perform. These records are typically kept in log files, and may contain one or more entries for every transaction (including authentication) performed by a user. The log can be examined in the event of a security breach being detected or suspected. It may also be used for the purposes of accountability to demonstrate a particular user's action at a particular time.

In a system using PWDs, it is important that the log files are available to the system as a whole. It is therefore advisable to log both to the local device and to the central server. If the network connection between the portable device and the central server is not permanent, the log files will need to be transferred along with other data at times when the connection is live.

3.1.6. Power management

Since a battery powered device becomes useless when its reserves run out, it is important to ensure either that the power is kept topped up or that alternative sources of power are available. Policies about recharging devices need to be established, and users made aware of the typical length of time the device can be used for without charge. Most portable devices are able to warn their user when the battery charge has fallen below a pre-specified level. It may also be appropriate to include a battery charge level display in the user interface to allow users to monitor its status.

3.2. Application level policies

3.2.1. Role based access control

Role-based access control provides a useful and effective model for managing security in a system. In any context, it is important that the roles set up within the system correspond closely with the real-world roles and responsibilities held by the user. An important distinction is between "normal" user roles, those of "supervisors" or "managers", and the role of the system administrator. In conventional systems, system administrators have total access, but this may not be desirable when confidential data about patients is concerned. The ability for the system administrator to create new user accounts may be necessary, but the ability to create new patient records may not be desirable. Further, for the purposes of accountability (especially important from the medico-legal perspective), the ability of any user to edit or delete information should be severely restricted. These functions can be replaced by the ability to annotate existing information and mark information as no longer required –

neither of them being a destructive process. Such a state can be achieved if access control is extended to include the administrator.

3.2.2. User authentication

The security policy should describe the *additional* authentication required compared to the authentication to existing desktop systems in a physically controlled environment, as means of diversifying the new risks. Two-factor authentication or the composite authentication solution proposed by Clarke and Furnell (2006) is suggested.

Although the capabilities of portable computing devices are improving at a rather impressive rate, there are known usability and acceptability issues with respect to user authentication technologies (Braz, 2006; Stajano, 2006). Biometric authentication, in particular, has received a fair amount of attention, and has been extensively criticised as not a solution that could meet the requirements of all subscribed users (Daugman, 2000; Cambier, 2000). For an overview on biometric authentication in relation to portable devices, the reader is referred to the work by Clarke and Furnell (2007).

Authentication tokens such as smart cards and security dongles are prone to the risk of theft and/or replication. These threats must be identified and addressed – the most common approach being disabling the acceptance of keys that have been reported as lost or stolen, and replacing them with new issue ("changing the locks"). Carrying keys or other tokens can be inconvenient – a particular problem in an environment where the risk of disease transfer via such objects can be high, or where the token has to be inserted into the portable device to be recognised. Tokens that work on a proximity basis (RFID tags are the best known current example) may prove to be more convenient in these cases. However, RFID tags are theoretically easy to duplicate (in the same way that keys are) because the information in them is effectively public. Other means of mitigating this risk (e.g. keeping keys out of sight and RFID tags away from potentially adversarial readers) need to be adopted. Alternative means of providing key information (e.g. the user typing in an RFID code rather than reading it wirelessly) should be avoided since they may provide a backdoor into the application.

3.2.3. Security states

Strictly speaking, the definition of the security states could be classified as an infrastructure type of activity. However, the purpose of defining a device's acceptable security states and processes to be carried out in each state, is to make sure all security mechanisms are in place prior to entering the user session phase, opening up interactivity with the user.

Risks of misuse can be reduced by permitting certain functions to take place only when the device is in certain states. Such states might be physical (e.g. while attached to a docking station) or logical (e.g. while connected to a particular wireless access point). They may require the attachment of some authentication token, such as a dongle, to the device.

As an example, imagine a PDA being used to collect patient data in a hospital ward and send that data via a wireless network to a central server. A session key is required before data will be transmitted from a device. In this scenario, there are a number of phases that could be identified, including:

PDA setup phase (i.e. when the PDA is prepared for deployment)	The PDA's private/public key pair should be generated at this time and the tamper resistant module of the PDA updated with the private key. The server should add the PDA related information to the "white list" of recognised network devices. The server should also send its public key certificate to the PDA.	This phase could be restricted to being performed while the PDA is docked in its cradle, and may use the cable connection as a means of communication, rather than any wireless connection.
Security context synchronisation phase	During this phase, the PDA should (re-) authenticate to the server and, upon successful authentication, the server should send an update for the PDA's user database and other security policy updates (for the PDA's policy database). The PDA should send in return the access logs to the server.	This could be restricted to when the PDA is docked for its battery to be recharged. This creates a *reference point* for the desirable level of security, as this phase relates to the assumption of physical security that is offered by docking
PDA terminal session initialisation phase	Depending on the policy specifications, a terminal session relates to the period during which midlife keys (such as the session key) are valid. During this phase, the session key is updated. However, the older session key should be kept in the keys database for a period specified by the security policy.	Session initialisation could be restricted to while there is an authenticated connection between the PDA and the server. It should not be allowed when PDAs are operating in peer-to-peer mode.
User session phase	This phase refers to the period during which the user is logged on to the PDA, following a successful authentication.	Provided a valid session is open, data transmission can take place

The rationale for the security context synchronization above stems mainly from physical security requirements and is an attempt to effectively diversify some of the risk. The philosophy behind this phase is that some critical (security) updates to the device should be only performed in an environment where physical security is higher. This assumes that there are adequate physical access controls in place for the room where the cradles or docking stations are. Obviously, this arrangement would impose restrictions from a usability perspective and therefore the security related data which are to be updated only through the cradle should be carefully selected and balanced with the overall benefits. For example, according to the arrangement proposed in this report, if a user lost their password and requested a password reset, the policy could be that the new password would not be effective until that user returned the device to the cradle.

3.3. Policies for lost or stolen devices

Portable computing devices are particularly susceptible to loss or theft. By the nature of their portability, and often small size, they can also be mislaid accidentally. Therefore, in developing a technical security solution, one also needs to take into consideration the physical threats posed, in the light of "standard" operating tasks and scenarios, and develop a suitable risk mitigation strategy.

4. Concluding remarks

This paper presented an overview of the additional security controls and directions that should be considered when introducing portable devices into a healthcare environment. We argue that the use of a physically secure and controlled cradling room can to some extent restore the loss of physical security which is inherent in mobile computing devices. The cradle approach (when combined with the concept of alert levels to proactively respond to security events) seems to be a viable security control. An initial investigation showed that this approach can potentially reduce associated risks to acceptable levels (depending on the relevant stakeholder view), as it can effectively address a selection of vulnerabilities included in the taxonomy presented in this paper. Both the cradle approach and the alert levels are technically and economically feasible controls. Further analysis and study of the threat vectors that may manifest in this context is an ongoing area of research.

5. References

Braz, C., (2006) "Security and usability: the case of the user authentication methods". *Proceedings of the 18th international conference on Association Francophone d'Interaction Homme-Machine*, Montreal, Canada: ACM Press, pp. 199-203.

Cambier, J. L. (2000), "Biometric Identification in Large Populations". *Information Security Bulletin*. March 2000;5(2):17-26.

Clarke, N. and Furnell, S. (2006), "A Composite User Authentication Architecture for Mobile Devices". *Journal of Information Warfare*, vol. 5, no. 2, pp11-29,

Clarke, N. and Furnell, S. (2007). "Advanced user authentication for mobile devices". *Computers & Security*, vol. 26, no. 2, pp109-119.

Fichman, R. and Cronin, M. (2003), "Information-rich commerce at a crossroads: business and technology adoption requirements", *Communications of the ACM,* 46(9), pp. 96-102.

Li, Y.C., Chang, I.C., Hung, W.F. and Fu, H.K., (2005) "The Critical Factors Affecting Hospital Adoption of Mobile Nursing Technologies in Taiwan", *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 6*, p. 157.

Lu, Y., Xiao, Y., Sears, A. and Jacko J., (2005) "A review and a framework of handheld computer adoption in healthcare". *International Journal of Medical Informatics*, 74(5), pp. 409-422.

Stajano, F., (2006) "One user, many hats; and, sometimes, no hat - towards a secure yet usable PDA". In: Christianson B, Crispo B, Malcolm JA, Roe M, editors. *Proceedings of Security Protocols, 12th International Workshop* Cambridge: Springer-Verlag; pp. 51-64.