Experiences of Using a Public Key Infrastructure to Access Patient Confidential Data Over the Internet

D. W. Chadwick[3], C. Carroll[1], S. Harvey[3], J. New[1], A. J. Young[2],
[3]*IS Institute, University of Salford, The Crescent, Salford M5 4WT*
[2]*School of Sciences, University of Salford, The Crescent, Salford M5 4WT*
[1]*Diabetes and Endocrinology, Salford Royal Hospitals NHS Trust, Hope Hospital, Stott Lane, Salford M6 8HD*
d.w.chadwick@salford.ac.uk

Abstract

A project to enable health care professionals (GPs, practice nurses and diabetes nurse specialists) to access, via the Internet, confidential patient data held on a secondary care (hospital) diabetes information system, has been implemented. We describe the application that we chose to distribute (a diabetes register); the security mechanisms we used to protect the data (a public key infrastructure with strong encryption and digitally signed messages, plus a firewall); the reasons for the implementation decisions we made; the validation testing that we performed and the results of the first set of user trials.

From a user acceptance perspective, we conclude that perceived usefulness and perceived ease of use on their own, are insufficient to guarantee that a new application will be used extensively in its new environment. Other domain specific factors, such as the compatibility and integration of the new computing system with the old, the working practices of the clinicians, the costs of using the new system compared to the old, and the actual location of the computing equipment all need to be taken into account when establishing untried information technology in 'real world' settings.

1. Introduction

The effective management of chronic disease, such as diabetes and cardiovascular disease, are increasingly dependent upon information technology. Traditionally these information systems have been developed within secondary care (hospitals), and have only been accessible from within the home institution. Many dozens (if not hundreds) of these information systems exits at hospitals around the UK. Unfortunately the majority of care is provided by primary care (GP's) and nurses, who do not have easy, real time access to the information recorded in these centralised hospital information systems. This can result in inefficient health care provision e.g. duplication of investigations.

The aim of this research was to develop a methodology to convert standalone, hospital based information systems, into highly secure distributed applications running on the Internet. This would enable geographically dispersed health care professionals to have access to the information held within the (previously centralised) information system. The main research question was therefore "Can we connect a highly confidential information system to remote users via a highly insecure network such as the Internet, without compromising the security or the usability of the system, and can the design be sufficiently general that it is easily applicable to other such systems?". As we were converting a pre-existing information system into a distributed system, we were not primarily interested in answering questions about the system's functionality, or its perceived usefulness, as these were taken for granted. Our aim was to see if we could make the distributed system almost as easy to use as the centralised system, when strong security was added to it. Since adding security to an object usually makes it more difficult to use (consider for example adding a lock to a window, or an alarm to a building, or an immobiliser to a car), we were interested in finding out how much impact the addition of strong security might have on the usability of the system, and how we might minimise this.

After some discussions with the clinicians at our local hospital, Hope Hospital in Salford, Greater Manchester, they decided that their Diabetic Information System (DIS) would be a useful application from which to build and test our methodology.

2. The centralised application

Salford is a health care district in Greater Manchester, UK, with a population of 230,510 of whom 5395 are known to have diabetes. The DIS was introduced in January 1992, and holds details about all the known diabetes patients. The DIS is used by all primary care and hospital diabetes

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

services. Records based upon the UK Diabetes dataset [20] are updated and verified during the annual structured preventative care review. Briefly this contains information regarding their type of diabetes, how it is treated, the presence of any diabetes related complications and biochemical indicators used to assess metabolic control.

The DIS is accessible directly via a built-in user interface, and very recently a SQL based fat client has been added to allow access via the hospital LAN. (The client is fat, as it performs authorisation decisions based on the user's login identity. By comparison a thin client would let the database perform the authorisation function.) All clinicians are given a username/password pair for login to the application, and the username determines one's privileges for accessing the data (i.e. hospital consultants and general practitioners can only access data relating to their own patients). As the hospital LAN is a trusted network, with no network connections to the outside world, no security is provided for the data whilst it is in transit between the client and server.

Once a year, paper printouts of each patient are produced and posted to all local GPs who then call the patients in for an annual diabetes examination. The completed pro-formas are posted back to the hospital for manual entry into the DIS, a process that can take several weeks (or months) to complete and produces additional problems. Data are double entered, once onto the pro-forma and once into the database, thus giving rise to potential transcription errors. If patients visit their GP before their annual check up, no up-to-date data is available. The pro-formas can, and do, get lost or misplaced, and the whole process is time consuming.

3. Addressing the security and usability concerns

The first half of the research question was "Can we connect a highly confidential database to remote users via a highly insecure network such as the Internet, without compromising the security or the usability of the system?" To make the application accessible over the Internet, without compromising the security or usability of the system, a number of security and usability issues need to be addressed. It is well-accepted that the Internet is a highly insecure public network [10][14], and that connecting an organisation's network to the Internet can significantly increase the vulnerability of the organisation's network [21]. Most importantly, therefore, we had to protect the confidentiality of patient data as it was transferred across the Internet, and ensure that the hospital network was not compromised by connecting it to the Internet. Further, we had to ensure that only appropriate health care professionals with a "need to know" could access the patient data, as clinicians have a legal duty of care to ensure that patient details are kept confidential [6].

The security concerns were addressed in the following ways. It is well established that confidentiality of data in transit is best provided by symmetric encryption using a strong algorithm and large key size [18]. The strongest algorithms are those that have been exposed to public scrutiny for many years without flaws being found in them [19]. The DES algorithm and its variants (e.g. Triple DES) readily pass this test. The time that it takes to mount a brute force attack on enciphered data is directly proportional to the key size used to encipher the data. Whilst the time depends upon the hardware being used, nevertheless, it was estimated that a brute force attack on a key size of 128 bits, using multi-trillion dollar specialised hardware, would still take 10^{11} years in 1995 [18]. We thus concluded that a key size of at least 128 bits would be sufficient to protect the confidentiality of the patient data, and so we chose to use the triple DES algorithm that has an effective key size of 168 bits.

In order to ensure that a remote user is genuinely who they say they are, and thus has a "need to know", strong authentication of the users is required. Strong authentication prevents unauthorised people from being able to access the patient data. Many different strong authentication mechanisms exist. We decided to use a public key infrastructure (PKI) [1][17], as this provides strong authentication via digital signatures, as well as strong encryption. Since recent European legislation [7] recognises certain types of electronic signatures are equivalent to hand written signatures, using a PKI can provide irrefutable proof of who is accessing the confidential data should we need it. We were already using an Entrust PKI (see http://www.entrust.com) for other research projects so it was natural to choose this for the DIS application, but if we had not had access to Entrust, there are a number of other commercial PKI vendors and CA service providers to choose from.

Maintaining the privacy of the hospital LAN by keeping out undesirable users and only letting in bona-fide users is a difficult task. The commonly accepted way of providing this functionality is to place a firewall [3][4] between the hospital LAN and the Internet. The Telecommunications Branch of the NHS Information Authority oversees the connection of hospital LANs to the Internet in the UK, and maintains strict guidelines for this. The hospital chose to use FireWall-1 from Checkpoint (see http://www.checkpoint.com) and their implementation and management of this firewall was approved by the NHA IA.

Validation tests were devised to check the security and integrity of the implemented solution. Specifically we tested: if unauthenticated users could access the system, if authenticated users could access the system and retrieve unencrypted data (thereby compromising its security in transit), and if the received encrypted data was exactly the

same after decryption as the original data in the DIS (data integrity).

Turning next to system usability. Ease of use comprises several factors such as initial time to learn to use the system, simplicity of everyday use, and adequate performance. Tom Gilb [9] suggests that usability comprises four factors: the demands made on a user's human attributes such as intelligence, sight, dexterity etc., the time needed to learn to use the system, how productive the system is during normal operation, and how well the users like the system. We devised a series of validation tests and a questionnaire that would help us to quantitatively and qualitatively determine each of these aspects of usability.

Whilst system usability was our primary concern, it is not the only factor affecting a systems success. IS research suggests that user/client acceptance, top management support, good communications, good technical staff and project managers, and client consultation comprise the critical success factors of IT projects [13][15][16][23]. We endeavoured to maximise these factors so that they did not unduly impact upon the usability. Specifically, we were ensured top management support as the hospital consultants in charge of the DIS, and the GPs who owned their own practices, willingly joined the project. We had good technical staff and project management within our research team. We had regular project meetings with the clinicians and the technical staff. User acceptance, according to the Technology Acceptance Model (TAM) [5] is determined by "perceived ease of use" and "perceived usefulness of the application", the latter being the more important factor of the two. This was found to be especially so for physicians using telemedicine applications [11]. As our clinicians had already suggested that the DIS application was the most useful one for us to distribute to the GPs, we assumed that the application had the required level of perceived usefulness. Our task therefore was to maximise the ease of use of the distributed DIS so as to maximise the client acceptance.

4. Optimising the design of the application

We had the choice between building a special purpose user friendly interface, or using an existing well known user friendly interface such as a Web browser. A special purpose interface gives the most flexibility in terms of design and capability, but at the cost of significant development effort plus time invested by the user to learn how to use the new application. The World Wide Web on the other hand is ubiquitous, and most (if not all) computer users already know how to use one of the popular browsers. We recognised that health care professionals are extremely busy with very limited time to spend on learning new applications, so we felt that Web browsers would give us the most chance of success with the primary carers. They probably already knew how to use Web browsers, and if not, it should not take much time to teach them. Furthermore, as Web clients are available for all the computers most likely to be used by primary carers, we would have no portability issues. Finally, in order to maximise acceptance of the interface by the primary carers, we chose to design the main web page to look like the paper form that the clinicians were already used to seeing. This should minimise the demands made on their intelligence and on their time to learn to use the electronic system, and should maximise their productivity and contribute to them liking the system.

Given the decision to use a Web based interface, a number of other decisions flowed from it. Firstly, we had to decide how to convert the http requests once they arrived at the web server, into SQL requests for access to the DIS. Secondly, we had to decide between using the SSL protocol [8] and its X.509 certificates [12], or Entrust formatted X.509 certificates and their proprietary protocol. We decided to use CGI scripts to solve the former problem, as they are an already proven way of converting web traffic into application specific queries. The latter decision was somewhat more difficult to make. SSL has a number of benefits, namely: it is a de-facto standard, all web servers and clients support it, and all PKI vendors and service providers can issue certificates in this format. However, SSL as implemented in 1998, had a number of severe disadvantages:

- we only had immediate access to 40 bit encryption, which was inadequate for confidential patient data [18]. We managed to get a plugin to Netscape Communicator for 128 bit encryption (from Fortify see http://www.fortify.net), but not for Microsoft Internet Explorer as this would require a US export licence. (Note that Internet Explorer 5.5 includes 128 bit encryption, but this only became available in 2000).
- SSL browsers and servers did not support the automatic retrieval of certificate revocation lists (CRLs), and this is an important factor to consider, especially from the server side.
- Trust management has to be performed by the users of the browsers and the administrator of the web server. By this we mean that deciding which root CA public keys to trust and which not to trust has to be performed by the users. As these people are typically not security specialists (and additionally in the case of GPs they simply don't have time to do this) we thought that this placed too much burden on them. It is a role that should be carried out by the security officers of the organisations.
- Certificate renewal after expiry is a manual process, and some browsers will still continue to use expired certificates.

Entrust Direct on the other hand, acts as proxies for both web clients and servers. Http requests from standard

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

web clients are intercepted by the Entrust Direct client proxy running in the user's machine, strongly encrypted and digitally signed using the user's private key, before being sent to the web server. The Entrust Direct server proxy, running in the hospital firewall intercepts the traffic for the web server, decrypts it, verifies the signature, and, if the user is trusted, forwards the http request to the web server. Both the Entrust Direct client and server automatically retrieve CRLs and process them, they will not accept expired certificates. They also have automatic mechanisms for certificate renewal (without user involvement) and they have all their trust decisions made for them by the security administrator of the PKI. Whilst a Canadian export permit was required for using 128 bit or stronger encryption, this was merely a formality as Entrust had already been granted blanket approval for the UK. (Since 1998 of course, the export situation has changed somewhat, in that the US has now altered its approach and has granted blanket approval for 128 bit encryption to be exported to commercial organisations in approximately 45 countries.) Given the above advantages of Entrust Direct, especially with the security naïve set of busy users that we had, it became the natural choice for us. The chosen architecture is shown in Figure 1.

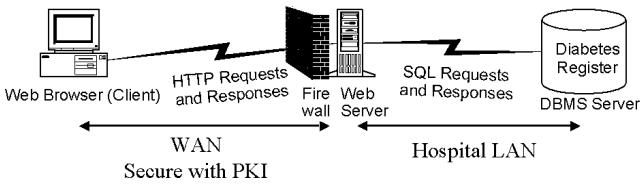

Figure 1. The Chosen Architecture

5. Implementation

In order to operate a PKI, a number of components are needed. The Certification Authority (CA) is the central server that responds to certification requests from the user, and signs the certificates of the users. It also issues the revocation lists. The CA will not issue new certificates or add certificates to revocation lists without being asked to do so by an authorised party. Usually a signed message from a trusted Registration Authority (RA) is sufficient for this. The Registration Authority Agent (RAA) is the administrative client used by a RA to issue signed certification or revocation requests to the CA. In the case of Entrust, after receiving a valid certification request from the RA, the CA returns secret authorisation information to the RA, which the RA gives to the user after authenticating him/her. The user's Client uses the secret information to establish an authenticated encrypted link with the CA, and then sends a certification request to the CA. The possession of this secret information by the user proves to the CA that the user has been authenticated by the RA, and the CA is therefore willing to issue the user certificate. All of the above are provided by Entrust as part of their basic PKI product line. All communications between the CA server and the other PKI entities are secured, but as a further precaution, we placed a Linux IPFWADM firewall between our CA server and the Internet, in order to block any improper traffic from untrusted third parties.

One additional component needed by a PKI is a directory service in which to publish the certificates and CRLs that have been issued by the CA. LDAP [22] is the Internet standard protocol for accessing directory services, and this protocol is used by the CA server to write to the directory, and by clients to retrieve certificates and CRLs from it. We used the directory server from MessagingDirect (see http://www.messagingdirect.com) to publish our certificates and CRLs.

The final components of our PKI are the Entrust Direct proxies that intercept the http traffic between the clinicians and the DIS. As stated previously, they interact with the PKI to fetch CRLs and check the signatures of incoming messages. They also digitally sign all outgoing messages. In the case of the client proxy it uses the private key of the clinician; in the case of the server proxy it uses the private key allocated to the DIS. The mode of operation is that the clinician starts the Entrust Direct client proxy instead of his usual browser, enters his password to gain access to his private key, and then the proxy starts up the browser. The user then interacts with the Web browser in the normal way, with no visible further interference from the client proxy. At the server end, the administrator of the firewall must enter the password protecting the DIS's private key every time the server proxy is started. Thus every interaction with the DIS is digitally signed and may be audited if required.

Whilst the PKI provides the strong authentication function, our CGI scripts needed to provide the authorisation function. The privileges must be administered in exactly the same way as the SQL fat client that they replace. The SQL client administered privileges by holding a table in the DIS containing the username, password and permissions of each registered user. The SQL client had super-user privileges to the DIS, and when a user logged in, it would compare the password with that stored in the DIS, and if correct, would retrieve the appropriate permissions for the user and act accordingly. We built the same functionality into our CGI scripts, but with an added enhancement. The first time a user accessed the DIS, he had to provide his DIS username and password. The CGI script checked that these were correct, and then stored the LDAP distinguished name of the user (obtained from the digital signature) in a new field in the DIS's table. On subsequent login attempts, the CGI script could simply

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

retrieve the clinician's registered username by looking up the one equivalent to the LDAP DN in the table.

The biggest problem we had to overcome was how to store the DIS super-user name and password securely. As each CGI script needs access to these, we could not expect the firewall administrator to have to type them in each time. Therefore we had to rely on obscurity rather than security, and hide the username and password somewhere where the scripts could find them (details withheld for obvious reasons!). A future enhancement is to build a Session Manager that is a permanently running service that sits between the CGI scripts and the DIS. This will have to have the super-user name and password entered at start up time by the firewall administrator. The complete system with all its components is shown in Figure 2.

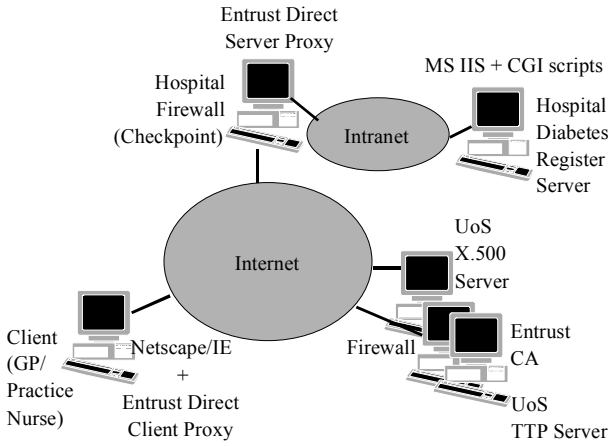

Figure 2 The System Components

6. Validation Testing

Once the system was built, members of the project team and closely associated colleagues in the university and hospital (7 testers in all) performed a series of validation tests on the system. These were carried out prior to the implementation with the GP pilot users, as we wanted to be sure that the system was user friendly, reliable, and fast enough in performance before we gave it to busy GPs in their working environment. These tests provided us with quantitative evidence of both the system's usability and security. Each database access test lasted between twenty and thirty minutes, to simulate what a clinician in the field would probably do under normal clinical situations. The following areas were examined under timed conditions:

- Logging onto the Application securely
- Searching for a patient's record and retrieving the details (4 times per testing session)
- Updating one patient's data with new information and retrieving it
- Closing the Application down and logging off
- Attempting to log on in an insecure manner (for security and performance comparison purposes)

In addition we logged the availability of the underlying hardware, software and networking infrastructure throughout the testing period. We measured the network availability by issuing Pings at timed intervals to various stations on the university LAN, the university's Internet gateway, University College London (a well known reliable site on the Internet), Demon Internet (to test LINX connectivity for dialup users), the hospital gateway, and the hospital's LAN. We measured data integrity by asking the testers to Email a copy of their retrieved record to a researcher, who compared this bit-wise with a copy he had previously downloaded from the diabetic register. We checked that the encryption was there by placing a line monitor on the connection between the user and the DIS. Finally we measured the time it took to install each user with the Entrust PKI. We revoked one user to see if this worked satisfactorily, and we reinstalled another user who pretended to forget his password to his private key. The complete set of tests and results are shown in Table 1.

6.1. Test results

We decided that 15 minutes was a reasonable time in which to install a new user with the client components of the Entrust PKI system. Out of the 8 users who volunteered to perform the tests, 1 was already installed prior to the testing, but we only managed to install 3 of the remaining 7 users within the allotted 15 minutes. 3 more were successfully installed within 24 hours of starting the installation, but we could not install the final user due to her PC's limited capabilities and the interactions with her Internet Service Provider (which limited outgoing TCP/IP connections). Clearly client installation is not as straightforward as it should be, even given that we were working in a very heterogeneous environment of PCs (different memories, CPUs, manufacturers, operating systems, configurations etc.)[1].

The 7 installed users completed 32 series of tests. This was less than we had planned, but never the less we felt it was sufficient to perform a meaningful analysis of the results. The overall success rate measures the percentage of attempted tests that were not completed successfully first time for one reason or another (e.g. unexpected result, network failure etc.). The purpose of this test was to attempt to estimate a "real-world" failure rate if this system had been rolled out to GPs in its current state. 70% of the recorded failures were due to a design bug in the system. The record retrieved after update was not the

[1] A subsequent project with opticians, using a later version of Entrust, provided to be much more reliable during user installation.

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

Test Category	Required Performance	Actual Performance
User Testing	59 Tests allocated to testers	36 Tests Attempted 4 tests were completely aborted 32 Tests Completed 31 Tests Completed and Considered Analysable
Overall first-time success rate	99%	25% (FAIL)
PKI Functionality		
User installation	99% success within 15 minutes	43% successfully completed within 15 minutes 57% within 30 minutes 86% within 24 hours (FAIL)
User revocation	100% success	100% (PASS)
User re-installation	100% success	100% (PASS)
Hardware components Availability		
Unix workstation running X.500 directory	99% availability	100% (PASS)
NT4 PC running Entrust CA infrastructure	99% availability	100% (PASS)
Database server running diabetic register	99% availability	100% (PASS)
Software components Availability		
LDAP directory server	99% availability	100% (PASS)
Entrust CA server	99% availability	100% (PASS)
Entrust Direct Server Proxy	99% availability	66% (FAIL)
IIS server + Scripts	99% availability	100% (PASS)
Network Uptime	99% availability	93% (FAIL)
System Security		
Unauthenticated access	100% failure	100% (PASS)
Data Integrity	100% accurate	99% (PASS - see text)
Non confidential retrieval	100% failure	100% (PASS)
Ease of use and Performance		
Time spent to learn to use the system	Less than 15 minutes	Between 0 and 8 minutes (PASS)
Time to launch the application after connection to the Internet has been opened	Less than 60 seconds, or 150% of the time taken to launch the insecure version	Between 45 and 118 seconds. Compared to insecure access: 158% (FAIL)
Time spent to initiate a request	Less than 30 seconds	Between 2 and 32 seconds (both operations combined). (PASS)
Time for reply to be received	Less than 1 minute	
Keying errors/mistakes in use	<1%	0.02% (PASS)

Table 1. The Validation Tests and Results

correct one, because the code had been designed to simply search for the latest date and not the latest date and time. (In an operational environment retrieval on date alone might have been acceptable, as the clinicians had never expected a patient to visit them multiple times in one day. But during testing many updates were performed on the same record in one day.) This bug was fixed after the validation testing was completed. The remaining failures were due to the Entrust Direct server proxy being unavailable from 4pm onwards every day. We never did find out the true cause of this bug, but when we changed the CRL publication time from 4 hours to 24 hours it seemed to go away.

The system proved to be very reliable throughout the testing period, with most components scoring 100% availability. Only the Entrust Direct server was temporarily unreliable, at 66% availability. Network availability, at 93%, was also poor, but unfortunately this

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

was out of our control. The poor performance was almost entirely accounted for by failures in the university LAN, and not in the Internet.

We were particularly encouraged by the time taken to learn to use the system (less than 8 minutes) plus the performance once the application was installed and the connection made to the database (between 2 and 32 seconds to search for and retrieve a patient record). However, the time taken to launch the application and make the connection to the database (up to 2 minutes) was longer than we would have hoped for. Certainly we felt that this time would be prohibitive for a GP during a patient consultation, especially if the time to initiate an Internet connection had to be added onto this.

Data integrity measured 99% and not 100% as expected. However, the 1% failure was found to be due to a documented bug in Windows NT. Of the 72 sets of results received from the testers, all were identical except for one where the dates were in American format rather than European format. A workaround has since been applied to the application.

User revocation and user re-installation both worked correctly with no hickups. A revoked user was unable to access the system, as was an unauthenticated user (one without a trusted public key certificate). Furthermore, it was not possible for a genuine user to access the system without first being authenticated to it, or after authentication for the user to retrieve unencrypted information from the server. This was because the Entrust Direct server proxy rejected every request that was not digitally signed, and because the proxies intercepted every message between the user and the server and digitally signed and encrypted them all.

Our overall conclusions from the validation testing were that the application was very secure, easy to use, and performed well after launching, but that application installation was problematical and application launching was too time consuming.

7. User installation

The results of the installation saw a repeat of the difficulties that we found during the validation testing, only this time the environments were even more heterogeneous, with a combination of different PCs, some connected by LANs and some stand alone, using a variety of different ISPs. Consequently we experienced an even greater number of unforeseen technical problems during installation.

The original plan was to pilot the application with 12 doctors and practice nurses located in 4 different surgeries. Surgery A is split over 2 physical locations and has 2 GPs and a practice nurse that work in both locations. Surgery B has 4 GPs and 2 practice nurses (but only 1 of the latter agreed to participate in the pilot). Surgery C has 1 GP and 1 practice nurse, and surgery D has a diabetic specialist in a tertiary care unit.

Surgery A had no suitable PC at either location, and so the project lent them a suitably configured PC and modem. This was placed in one of their locations on a worktop away from the GPs desks, and so was not readily accessible during consultations. Installation proved difficult due to the fact that their ISP, one of the numerous free ones in the UK, demanded to know the calling telephone number before giving them complete Internet access (i.e. access to all protocols and port numbers. We were using non-standard ports for our LDAP service, and Entrust uses non-standard protocols and ports). As most GPs in the UK are ex-directory, they usually withhold their telephone number from outgoing calls. We had to try 10 different free ISPs before we found one that did not demand to know the calling number during login. Halfway through the trial our logging indicated that this surgery had not used the system once. When queried about this they admitted that they needed the PC to be on a mobile trolley rather than on a worktop, so that it could be moved to where the GPs were working. They would also like to have a printer available to them so that they could print out the patient's details. These were provided for the second half of the trial.

Surgery B has a LAN installed, with PCs on every GPs desk and a central server which automatically dials the Internet when needed. Though this dialup can take a minute to connect, it means that the Internet is always there when needed and no explicit action is needed on the part of the GPs. However, despite making several visits to the surgery, after one month we had only installed the software with 2 GPs and the practice nurse. The problems were all of a technical nature, but were compounded by the limited time available to the GPs. Therefore if the installation did not proceed as planned within the allotted time we usually had to book another appointment. Problems arose from the LAN configuration, and this made installation difficult. One of the GPs had volunteered to use a smart card to hold his private key, but after an hour of trying to install the card and reader he decided to use software based keys instead. (A report of some of the difficulties we experienced with smart cards is given in [2].) However, after installation our log files showed that they were using the system.

In surgery C the GP uses the Internet frequently, and we were specifically asked not to alter any of his Internet settings. However, the GP was using an ISP provided dialer, rather than Windows Dial-Up Networking, and this forced the web browser to use that ISPs default proxy settings. These settings stopped Entrust/Direct from working. Fortunately the GP did have another ISP already installed, and so we had to instruct him to use that ISP, but even this caused him some disruption. The GP also agreed to pilot the use of a smart card, and this increased

the installation time to 1.5 hours. However he had problems subsequently when using the smart card, and eventually reverted to using software based keys.

Surgery D had no suitable PC, and so the project lent the specialist a laptop and a modem and set her up with an ISP. She was only able to dial the ISP from one of her 3 consulting rooms as (a) the telephone connection in one room barred external calls starting with a zero and (b) the telephone connection in another room had a broken dialling tone which the modem did not recognise (we had to dial the "9" for an outside line manually to get a normal dialling tone!). However, once installed, this user used the application very frequently from the third consulting room (several times a day in fact).

8. Analysis of users' experiences with the pilot system

At the end of a four-month trial period, users were contacted in person or by telephone, depending on their availability and underwent a short, highly structured interview based on a questionnaire. A total of 6 users were successfully interviewed, 4 doctors and 2 practice nurses. The interviews covered the following usability aspects:

- Satisfaction with the existing paper-based system
- Reasons for agreeing to join the trial
- Installation issues
- Interface design issues
- Usage issues
- Security issues
- Impact of Security on Usability

8.1. Satisfaction with the paper-based system

Although the paper-based diabetes register was regarded as slow, frustrating, and time consuming to update, it is appreciated in that it provides general practices with relatively up to date information about their patients and in particular is valued as an audit tool by them. Practice nurses have been given the responsibility of providing Hope Hospital with updated database information. However, all those interviewed were keen to have access to an electronic version of the database as this was perceived as being quick, easy to use and as providing practices with real-time information about their patients. Perceived usefulness was therefore initially high.

8.2. Reasons for agreeing to join the trial

Users agreed to join the trial for a variety of reasons. These included:

- To improve access to the database
- To gain experience in electronic networking
- To gain access to network professionals with a view to gaining Internet experience / training
- To help out colleagues
- To gain insights into the long-term potential of Internet technology.

During the trial period, as the database is constantly used for patient care both in primary and secondary care establishments, it was necessary to update the database both electronically and manually, using the paper format, until the reliability and robustness of the computerised version could be guaranteed. All users were aware of this at the outset of the trial, but were willing to continue as they could appreciate the long-term potential of such applications.

8.3. Installation issues

As already described installation procedures were not straightforward for the majority of users (4/6). For those four, this was a source of frustration, especially for the two doctors given smart cards, as neither of these worked easily. The actual problems encountered have already been described. However, once they were set up no one experienced any specific problems regarding password use and all thought the application itself was intuitively easy to use and understand.

8.4. Interface design issues

Everybody liked the actual interface design. As mentioned previously this was purposely designed to mimic the paper format of the database. Users appreciated this and it certainly enabled them to learn to use application almost immediately and to quickly familiarise themselves with it. There were no complaints about the actual locations of patient data. However, the fact that it was necessary to scroll down to gain access to all the data, caused some irritation and frustration to a couple of users. This was particularly the case for surgery D, where the user was supplied with a laptop with a very small screen. In her case, it contributed significantly to her dislike of the system.

8.5. Usage issues

After they had been installed, few users used the application routinely. This was due to a number of reasons:

- The computer was in an inaccessible place,
- Problems were experienced logging into the security software, where sometimes it took several attempts for users to gain access to the system

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

- Problems were experienced with passwords not being accepted,
- Problems were encountered with access to the Internet (lines being busy),
- The application was not inherently useful,
- Most people thought it made their job harder not easier,
- Clinicians were still forced to use the paper system at the same time as the Internet version.

One of the major problems with the pilot system as a whole was that it was duplicating work that had to be carried out anyway in a paper format. Unfortunately, we could not forego the paper system until the Internet version was proven to be robust. The pilot system was also an extra IT system that was not compatible with the GPs' in-house patient care systems, and therefore, in 2 of the 3 practices, required a separate designated computer. Hence, nobody relied on the pilot system to provide them with up to date results, as they already had access to reasonably current results, either on their own in-house patient care systems or in a paper format from the hospital. Also the Internet version of the database was regarded as being too slow for routine use during clinic consultations with the patients present, since the mode of use was to login to the Internet as and when needed, rather than to leave the connection open all day.

Another factor was the working practices of the GP's. Most are used to looking through the patients paper notes for letters and results when consulting with patients and the computer would not supersede this activity, especially as it only provided the same information as the paper notes. Therefore, most people who actually used the pilot system did so at the end of clinics where they updated everyone's record who had been seen that day. The only exception to this was the consultant ophthalmologist, who used the computer system with patients to obtain up to date information about her patients' diabetes status and blood pressure, both of which impact upon the progression of diabetic retinopathy. She did not routinely have access to this information, so for her, the computerised database provided her with extra information and was therefore inherently useful.

Although most users tried to use the system on more than one occasion, nobody used it routinely, and only 3 people accessed it after month two of the trial. The main reason for this is that it takes too much effort to gain access to the system as 'logging on' procedures recurrently fail or are too slow. The consultant ophthalmologist also stopped using the system for the same reason, even though she had the most to gain from the system. Other usability factors include the fact that clinicians had to scroll down to 'see' all the relevant sections of the information sheet. This was not a problem for those with large computer screens, but was a major factor for the consultant ophthalmologist as she only had a laptop with a small screen.

Cost is also an important consideration for general practitioners, who have access to a free postal service to their local hospital (in this case Hope Hospital). Internet access requires them to pay the cost of a local telephone call, and if it is accessed for prolonged periods of time, this will have implications for their practice budget. Interestingly, all the doctors but neither of the nurses mentioned cost as a factor to be taken into consideration. This probably reflects the fact that doctors are responsible for practice budgets and not nurses.

8.6. Security issues

Users were aware that security software was being used, but had little idea how it worked or how efficacious it was. In fact, they placed their trust in the computing personnel who set them up as users, and relied on the latter telling them the truth. They mentioned that they had no means of checking the validity of the latter's claims that the security software was bone-fide, but also that they were happy to place their trust in them. This is perhaps not as surprising as it may at first seem. When faced with decisions in a realm that is unfamiliar to us, most people will tend to ask for the advice of domain experts and will follow it. It would be unreasonable to expect doctors to utilise line monitors and software diagnostic tools to ensure that the data was being encrypted as it traversed the Internet, so trusting the computer experts is a rational alternative. After all, we routinely trust our health and our lives to our doctors, so trust is part of their working culture.

On the whole, most doctors thought the information system was more trustworthy than the paper-based system, as gaining access to paper and envelopes was thought to be easier than gaining unauthorised access to the database.

This raises a number of issues about the legal liability of doctors should breaches of the medical data occur. Since we did not prove our claims to the doctors, other than that the NHS Information Authority (Telecommunications Agency) had authorised us to use the system, if a problem had occurred, it is likely that the doctors would have had to bear some or most of the responsibility for this.

The two nurses interviewed, however, were more sceptical than the doctors, and unsure as to the reliability of the security software. They placed more trust in the paper-based system. The reasons for this seemed to be multiple. Firstly nurses are not as computer literate as a lot of doctors and may not use the computer as much. They therefore may be more suspicious of computers *per se*. Secondly these nurses were much 'closer' to the database than the doctors, as they were directly

responsible for updating the database information and also used its information much more actively in their patient care. Therefore, they were more aware of the database limitations and inaccuracies compared with doctors. Finally, they also experienced significant problems trying to access the system. This latter problem seems to have influenced the attitudes of both the nurses and some of the doctors as to the overall security of the system. Their logic was if there are so many problems gaining access to the system, there may be problems with the actual security of the system as well. Hence, the problems experienced with logging onto the system has reduced the users confidence in its other aspects such as its security.

8.7. Impact of Security on Usability

From the interviews with the users it was clear that the only evidence of added security was the use of their password to access their private key, and otherwise it was transparent to them. Once users had gained access to the system, no-one thought the security software was an imposition, as they did not realise it was there after they had successfully logged on. This fact is advantageous from a usability perspective, in that the security software does not impose a further burden on clinicians using the system, but it is disadvantageous from a security perspective, as they cannot tell whether the data is secure or not as they access the system. (In fact the software does display a turning key icon in the Windows system tray when it is encrypting and decrypting data, but most of the users were oblivious of this fact.)

9. Limitations/Future Research

A number of factors impeded our research. Firstly because the DIS was a pilot system, the existing paper based system had to be run in parallel with it, and this caused extra work for and resistance from the users. Also the availability of the paper print out tended to undermine the inherent usefulness of the electronic system to the doctors, and therefore provided a disincentive to use it. Finally because the Web access to the DIS was a separate application to the GPs existing electronic patient management system, it meant that any data retrieved from the hospital was not automatically incorporated into their existing system, and so remained isolated from it. All these factors conspired against the users making frequent use of the DIS, and coupled with the problems the users experienced in gaining access to the DIS, the costs clearly outweighed the benefits.

The research was further limited in that only 12 users took part, and only half of these were successfully interviewed at the end of the trials. This small sample makes it difficult to produce generalizable results. The research would benefit from being repeated with different user groups, in different organisational contexts, particularly with those who do not currently receive paper output from the DIS, so that the electronic interface will be their only means of access to the diabetic information. Opticians, for example, are one such candidate group, and a follow on project is currently underway with them.

Concerning the methodology of obtaining the user feedback, we relied solely upon structured interviews based upon a questionnaire. No direct observations by either a researcher or video camera took place, and this could have provided additional unbiased valuable user feedback. Thus the user's experiences that we have captured are the ones verbally given to us after a time for reflection, and they may have been intentionally or unintentionally modified during this time.

The scope of the research was purposely limited by its primary research question focussing on security and usability, and we did not address the wider issues of information systems success. Additional research questions could be posed concerning the success of the distributed system, for example: which organisational and contextual issues lead to success, how maintainable and modifiable is the chosen system, what is the commercial viability of system deployment and operation, how is the quality of the database affected by online updates as opposed to paper based ones, and what are the demographic and educational-related user acceptance criteria.

Finally in order to demonstrate that the method described here is generically useful, it would be good to repeat the experiment with several other health care applications.

10. Conclusions

We have shown that it is possible to provide highly secure remote access to a hospital information system via the Internet, using commercial products and tailor made CGI scripts. By carefully designing the user interface around a Web browser, it is possible to build a system that is extremely easy to use by both doctors, specialists and practice nurses. The time to learn to use the system is minimal, and the high security does not impose any significant burdens on the users. The response time is also adequate, though would benefit from an improvement. However the time taken to launch the application is an inhibiting factor, especially when using dial-up access via an ISP. Initial user logon (application launching) thus proved to be the biggest problem to the users during normal operation. The largest problem we faced was installing the users as members of the public key infrastructure. Installation was plagued with unforeseen technical difficulties, and the smart cards proved so time

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

consuming to install and problematical to use that all our users stopped using them during the trials.

Despite the fact that the users experienced significant problems with the system during the pilot trial, none the less they valued the experience and could see the potential benefits of such a system, providing it can be made quick enough and robust enough, and integrated with their existing systems. Perceived usefulness therefore remained high. However, from a user acceptance perspective, other considerations such as the compatibility and integration of the new computing system with the old, the working practices of the clinicians, the costs of using the new system compared to the old, and the actual location of the computing equipment all need to borne in mind when establishing untried information technology in 'real world' settings. We conclude that perceived usefulness and perceived ease of use on their own, are insufficient to guarantee that a new application will be used extensively in its new environment. Other domain specific factors need to be taken into account.

We further conclude that our method for securely distributing health care applications does ensure high security for the data whilst in transfer and does not significantly impede the usability of the application, although the installation and application launching steps still need significant improvement.

11. Acknowledgements

This research was funded by: the European Commission IV Framework Programme Trusthealth 2 (Contract No: HC 4023) and ICE-CAR (Contract No: RE 4006) projects and the UK EPSRC under grant number GR/L60548. The authors would also like to thank Entrust Technologies for making their security software available to the university on preferential terms.

12. References

[1] Adams, C., Lloyd, S. (1999). "Understanding Public-Key Infrastructure: Concepts, Standards, and Deployment Considerations". Macmillan Technical Publishing, 1999

[2] Chadwick, D.W. (1999). "Smart Cards aren't always the Smart Choice", IEEE Computer, December 1999, Vol. 32, No. 12, pp 142-143

[3] Chapman, D.B., Zwicky, E.D. (1995). "Internet Security Firewalls", O'Reilly & Associates, Sebastapol, CA

[4] Cheswick, W.R., Bellovin, S.M. (1994). "Network Firewalls", IEEE Communications Magazine, September, pp 50-57.

[5] Davis, F.D. (1989). "Perceived usefulness, perceived ease of use, and user acceptance of information technology", MIS Quarterly, 13, 3 (September), pp319-340

[6] Department of Health (1994). "Confidentiality, Use and Disclosure of Personal Health Information", DoH Health Care (Administration) Division 4D, London

[7] European Community (1999). Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures (available from http://europa.eu.int/comm/internal_market/en/media/sign/99-915.htm)

[8] Frier, A., Karlton, P., Kocher, P. (1996). 'The SSL 3.0 Protocol', Netscape Communications Corp., Nov 18.

[9] Gilb, T. (1988) "Principles of Software Engineering Management", Addison-Wesley Publishing Company. pp 371-372.

[10] Hawkins, S., Yen, D.C., Chou, D.C. (2000). "Awareness and challenges of Internet security", Information Management and Computer Security, 8/3, 133-143

[11] Hu, P.J., Chau, P.Y.K, Liu Sheng, O.R., Kar Yan Tam. (1999). "Examining the Technology Acceptance Model Using Physician Acceptance of Telemedicine Technology", Journal of Management Information Systems, Fall, Vol 16, No.2, pp 91-112.

[12] ITU-T (1997) Recommendation X.509: 'The Directory - Authentication Framework'.

[13] Lee, S. M., Kim, Y. R., Lee, J. (1995). "An empirical study of the relationships among end-user information systems acceptance, training, and effectiveness." Journal of Management Information Systems, Autumn (12/2), pp189-203

[14] Lichtenstein, S. (1998). "Internet Risks For Companies", Computers & Security (UK), Vol 17 No 2

[15] Markus, M. L., Keil. M. (1994). "If we build it they will come: designing information systems that people want to use". Sloan Management Review, (USA), Summer 94 (35/4) pp 11-25

[16] Pinto, J.K., Prescott, J.E. (1988). "Variations in Critical Success Factors Over the Stages in the Project Life Cycle", Journal of Management, Vol.14, No.1 pp5-18.

[17] Rivest, R.L., Shamir, A., Adleman, L. (1978). "A Method for Obtaining Digital Signatures and Public Key Cryposystems", Communications of the ACM, 21 (2), February, pp120-126

[18] Schneier, B. (1996). "Applied Cryptography", 2nd edition, John Wiley & Sons, 1996, pp152-154

[19] ibid, p 214

[20] Vaughan, N.J., Home, P.D. (1995). 'The UK Diabetes Dataset: a standard for information exchange.' Diabetes Audit Working Group of the Research Unit of the Royal College of Physicians. British Diabetic Association. *Diabet Med*; 12: 717-22.

[21] Yasin, R. (1998). 'Hackers to Users, Feds: Internet is 30 Minutes from Disaster', Internet Week, 22 May, p 1

[22] Yeong, W., Howes, T., Kille, S. (1995). "Lightweight Directory Access Protocol". RFC1777, March

[23] Robert, D.W. (1997). "Creating an environment for project success". Information Systems Management, (USA), Winter, 14/1, pp73-77