An International Governmental Mailing System: A Requirement To Prevent Web-enhanced Terrorism

Sérgio Tenreiro de Magalhães[1], Henrique M. D. Santos[1], Paulo Viegas Nunes[2]

[1]University of Minho
Department of Information Systems
Campus de Azurém
4800-058 Guimarães, Portugal
{psmagalhaes, hsantos} @dsi.uminho.pt

[2] Military Academy Research Center (CINAMIL)
Military Academy
Rua Gomes Freire
1169-203 Lisboa, Portugal
pfv@net.sapo.pt

Abstract: E-mail systems are essential, but they also provide terrorists with an opportunity to impersonate public officials and, with the legitimacy and authority of the alleged sender, obtain collaborations or spread misinformation in critical situations, where the urgent need to responses relegates identity confirmations to a lower priority. This paper presents an e-mail system architecture that uses Public Key Infrastructures and behavioural biometrics, namely Keystroke Dynamics, to guarantee that only their legitimate users use governmental domains, automatically confirm their identity and encript/decript messages exchanged between public officers. The biometric components of the system can also contribute to a distributed database destined to identify anonymous e-mail senders.

Keywords: Security, biometrics, communications, e-mail architecture

1. Introduction

The e-mail is now a common and fundamental tool for communication in which we depend on, but the most used protocols used to implement it are unsecured and do not guarantee neither the confidentiality of the message or the identity of the sender. The first can only be achieved by encryption techniques, well known but rarely implemented by common e-mail systems, while the later is not ensured at all. In fact, the normal procedure to identify the sender of the message is to identify the alleged e-mail of the sender, but neither the e-mail server nor the e-mail client verify even if that e-mail address exists. On the other hand, providing more services implies new risks and the webmail services are a good example of that.

Computer network attacks can be considered as Electronic Means of Mass Disruption (Bayles: 2001) or even as Weapons of Mass Destruction (case of Russian Government Officials – 1995 - and of the Director of the USA's National Security Agency - 1998) (Clemmons: 1999) and we know that some terrorist groups have been using the Internet to collect information on targets, to communicate between cells and to plan attacks. They are also using tools available online to disguise their identities and they can use it, for instance, to spread disinformation or to collect money (Thomas: 2003).

For the time being we can expect some security only from more complex tools but, even then, we need that both users (sender and receiver) agree on the technology to be used and to do more then install it: to use it. If all this is achieved we still have a problem once those technologies for digital encryption and/or signature require a level of trust in the used certificates, which is often not suitable for use in official matters. While allowing users to send and receive messages on a browser without any previous configuration of an account, we are spreading the use of the e-mail technology but we are also allowing its use in public spaces that provide Internet access to their customers. These vulnerabilities provide terrorists with an opportunity to impersonate public officials and, with the legitimacy and authority of the alleged sender, obtain collaborations or spread misinformation in critical situations, where the urgent need to responses relegates identity confirmations to a lower priority.

Being so, we need to understand the vulnerabilities of the most common e-mail systems and we must bring the existing technologies together to create an architecture that can provide our governments and governmental structures with an acceptable level of security.

2. The e-mail systems vulnerabilities

Once the e-mail systems are mainly focused on utility, not on security, we can find several vulnerabilities in the most commonly used protocols that make them unsuitable for public use. Common e-mail systems do not implement cryptographic methods, so the message is sent without any protection that can prevent bad intentioned people from accessing to, at least, private information. In Figure 1 we can see an e-mail message sent by an e-mail managing program using the popular SMTP (Simple Mail Transfer Protocol) and the corresponding network packets as they can be captured by a public tool like *Ethereal*. By visualizing the packets, one can easily find the e-mail of the sender, the e-mail of the receiver, the subject of the message and the message itself. Using SMTP, POP (Post Office Protocol, commonly used for downloading e-mails) or HTTP (Hiper Text Transfer Protocol, used in unsecured webmail services) the user also compromises his username and password, once they are sent in the open through the network. This is even more important when Virtual Private Networks (VPNs) are deployed using the same password's file used by the mailing system so, once having the username and password of the e-mail of a worker, one can also access by VPN to the information system.

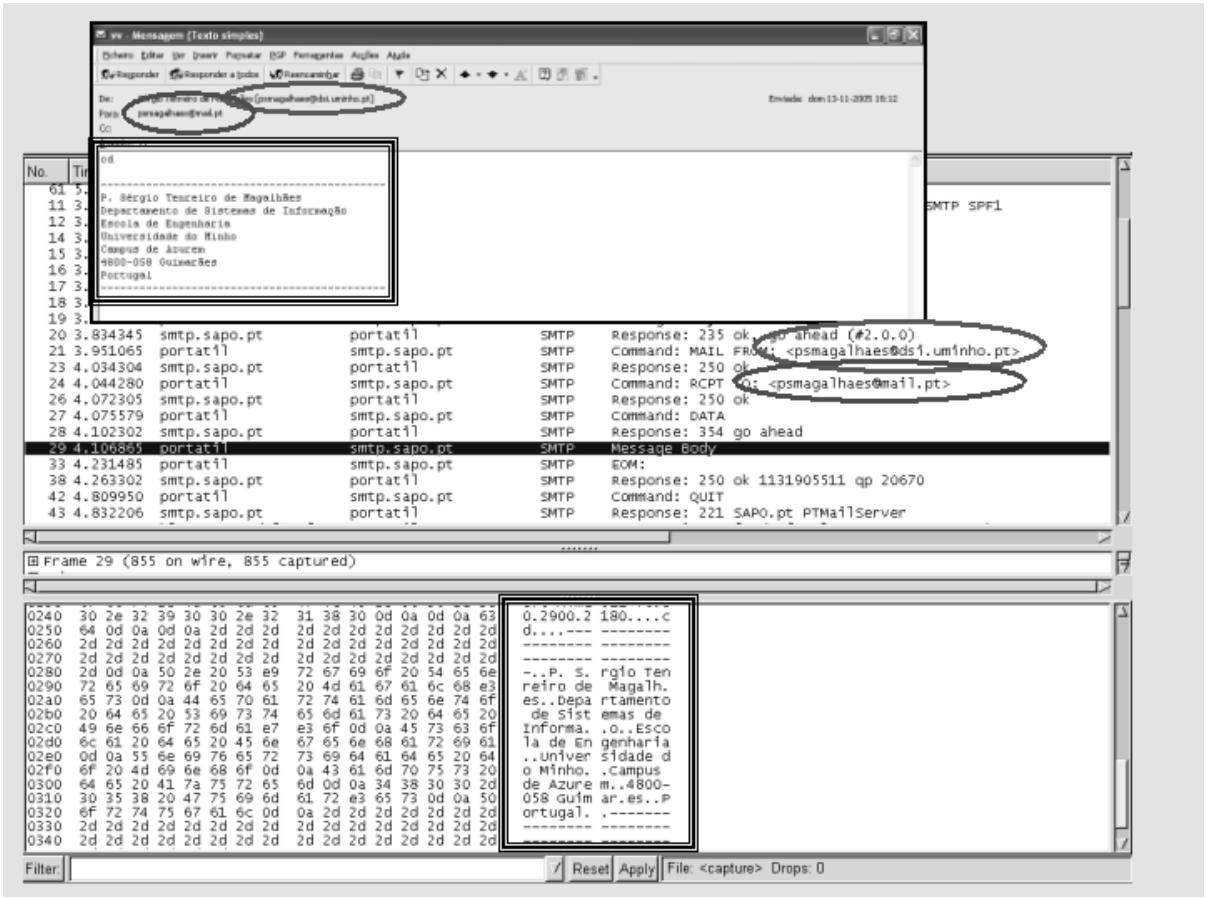

Figure 1 – Translation of the packages that go through the network when the e-mail is sent

The process of intercepting the network packets became more complex with the widespread of segmented networks, but on the other hand it became easier with the increasing ubiquity of our information systems. We now have wireless communications that can support an effective use of the network, even in a cell phone. So the processes change but the challenges remain. Anyway, it is always possible to program a switch to copy all the traffic into one port and that can be done by technologically hacking the switch from the outside, by converting the local network administrator to

the perpetrator's political/religious beliefs or by recruiting in the universities future network administrators.
Another security issue that is critical in the common e-mail systems is the accountability. How can we be sure that the alleged sender was, in fact, the creator of the message? This is even more critical when we are getting used to receive hierarchical directives through e-mail and, being so, many won't hesitate to follow an instruction received from a legitimate e-mail address without checking its legitimacy. In an emergency situation this can create chaos and be used by terrorists to aggravate the results of an attack or to create advantages from natural catastrophes. The Internet has downloadable software that allows a sender without technological knowledge to impersonate a user with an existing mail address or even with a non-existing one. This is the case, for instance of *anonyMail*. For those with programming skills, a small javascript can do the trick and, using a vulnerable SMPT server and a free webpage housing service, allow a terrorist to send a forged message from any public Internet post.

3. The Public Key Infrastructure (PKI) and the confidence chain

Most of the threats directly related with the communication infrastructures can be mitigated by cryptography. From all the tools and techniques available, those based on public key algorithms are special interesting since they avoid the exchange of secret keys in an open environment, like the Internet (Schneier: 1996) (Kaufman: 2002). This way we transfer the security problem from the message itself to a key pair (public key and private key), which is a simpler entity to take care of.
In fact, to securely communicate with someone, besides the required tools (a lot of them in public domain), all we need is to give our public key to everyone that wants to communicate with us, and get the public key of those we want to communicate with. Information encrypted by one of the keys can only be decrypted by the other. Processing the message with a hash function and encrypting the result is a perfect way to implement an electronic signature. This is easy to deploy within small groups of users which know each others. However, if we try to scale the solution to larger groups of distant and unknown users, the (old) trust issue about the presumed owner of a certain public key arises.
To solve this problem we can use a certificate, which is a data structure containing the public key, cryptographically signed by a trustable third party, a Certificate Authority (CA). It is assumed that we have the CA's public key and we trust this entity to what concerns its ability to certificate every pair user/public-key. Again, we are trying to transfer the security problem from a larger uncontrolled domain to just one key, but this time assuming its owner has a tremendous power to handle a huge number of public keys. Due to economical and geographic distribution issues, a CA typically relies on local Registration Authorities (RA) to verify user authentication during the certification requesting phase. Finally, a CA must keep actualized information about good and compromised or non-valid keys and provide a way to store and give access to public keys (key management). This kind of structure is called a PKI (Public Key Infrastructure).
PKIs have been around for several years, but there are few examples of well succeeded deployments, especially in large scale organizations where people do not know each other. Application integration is another obstacle. Most of the solutions available require external applications and some specific training to use them. Today e-mail clients are an exception and it is now possible to find one that supports a cryptographic public key algorithm. With these recent achievements it is not a surprise to find some recommendations about how to deploy large scale PKIs, particularly in the public sector (Gritzalis: 2005). However, there are some issues we need to take care of. For instances, the PKI security depends on the CA's Security Police, which could be incompatible with the Security Police of each organization/department that is using it. But even worst, at a public sector scale, the users' know-how is very different and there is always the possibility to compromise a private key or to fraudulently use a public key. To mitigate this lasts vulnerabilities a strong authentication is desirable

4. Keystroke Dynamics

The use of biometric technologies to increase the identification and authentication efficiency of a computer system has become a widely discussed subject. While governments and corporations are pressing for a deeper integration of these technologies with common security systems (like passports or identity cards), human rights associations are concerned with the ethical and social implications of its use. This situation creates a challenge to find biometric algorithms that are less intrusive, easier to use and more accurate.

The precision of a biometric technology is usually measured by its False Acceptance Rate (FAR), the permeability of the algorithm to attacks, by its False Rejection Rate (FRR), the resistance of the algorithm to accept a legitimate user, and by its Crossover Error Rate (CER), the interception point of the FAR curve with the FRR curve, that indicates the level of usability of the technology. Typically, when we force an algorithm to be more restrictive, its FAR gets lower but its FRR gets higher; usually the FAR and FRR values are defined by the system administrator, according to the security requirements – normally an outcome of the risk evaluation. The threshold can also be, in theory, dynamic and defined with the help of an Intrusion Detection System.

Establishing the error rates of a biometric technology is a complex issue. Studies have been made to normalize that evaluation, but the results are strongly dependent of the number of individuals involved in the process and, what is worst, of their characteristics. This means that, even with a large amount of data collected, the results can be very different if we change the group evaluated. This happens because it is very difficult to obtain a sample representative of the population, since we do not know how to characterize the population. A good example of this disparity are the results of the Fingerprint Verification Competition (FVC) 2004, where the best CER achieved was 2,07% (Maio D. et al: 2004), compared with the results of the FVC 2002, where the best CER achieved was 0,19% (Maltoni D. et al: 2003). Some international companies were present in both contests and the only justification for the disparity of the results is the difference in the sample data used to test the algorithms. This means we can only compare two algorithms using the same test data. The results also vary according to the final use: a system used to identify an individual is less accurate that a system used to just authenticate him/her.

Biometric technologies are usually classified as behavioural (e.g. voice recognition) or physical (e.g. retinal recognition), according to the human being characteristics used. But they can also be classified as collaborative, if they require the user to know about its existence and to participate in the process, or as stealth technologies, if they can be used without the knowledge of the person being authenticated or identified (Magalhães P. S. and Santos H. D.: 2003).

Keystroke dynamics is a behavioural biometric technology that can be used with the collaboration of the user or in stealth mode, and that allows a high precision level, both in authentication and in identification. Furthermore it does not require any special device since it works by analysing the user keystroke patterns, as he types (a password, a passphrase or general text) on a keyboard. Due to the possible integration level, these algorithms can also adjust their parameters to adapt themselves to evolutions of the user typing patterns.

As in many other problems, there have been two different approaches to the challenge of finding an algorithm for keystroke dynamics that minimizes the CER: machine-learning and deterministic algorithms. Deterministic algorithms are applied to keystroke dynamics since the late 70's. In 1980 (Gaines R. et al, 1980) Gaines presented a report of his work to study the typing patterns of seven professional typists. The small number of volunteers and the fact that the algorithm is deducted from their data and not tested in other people later, results on a lower confidence on the FAR and FRR values presented. But the method used to establish a pattern was a breakthrough: a study of the time spent to type the same two letters (digraph), when together in the text. Since then, many algorithms based on Algebra and on Probability and Statistics have been presented. Joyce (Joyce R. and Gupta G.: 1990) presented an algorithm to calculate a value that represents the distance between acquired keystroke latency times and correspondent times previously stored. Monrose (Monrose F. and Rubin A. D.: 1997) uses the Euclidean Distance and probabilistic calculations based on the assumption that the latency times for one digraph exhibits a Normal Distribution. Later, in (Monrose F. and Rubin A. D.: 2000), he also presents an algorithm for identification, based on the similarity models of Bayes, and in (Monrose F. et al: 2001) he presents an algorithm that uses polynomials and vector spaces to generate complex passwords from a simple one, using the keystroke pattern. In 2005, Revett (Revett et al: 2005) presented the results of applying the Rough Sets theory do keystroke Dynamics and obtained 97% classification accuracy. Magalhães (Magalhães et al: 2005) presented an algorithm that has several possible levels of accuracy (figure 2) that can be established according to the moment's need for both security and comfort of use. This study involved 170.391 attempts to crash 143 patterns of legitimate users and 251 legitimate logins and returned an equal error rate under 5% living in open the possibility to be more demanding on the users and taking the false acceptance rates to the zero region, or improving the levels of comfort and provide the users with a false rejection rate near zero.

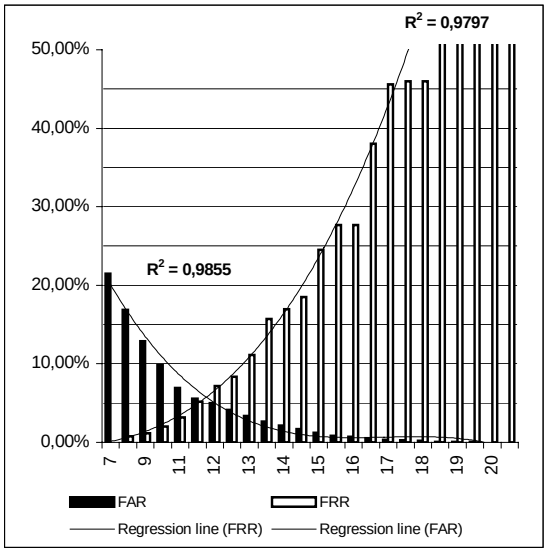

Figure 2 - false acceptance rate and false rejection rate for several possible thresholds and estimation of the crossover error rate for a keystroke dynamics algorithm

The algorithms cited are a small example of many approaches used to find adequate keystroke dynamics algorithms with a convenient CER. Many others could also be referred, all with different evaluation methods. One thing is certain: keystroke dynamics is a mature technology with highly acceptable levels of precision and without any inconvenience to the user, making it suitable for large scale use in environments where many aren't trained for complex uses of technologies.

5. A secure e-mail architecture for public institutions

The proposed architecture intends to guarantee the confidentiality, the integrity and also the accountability of the e-mail systems used under governmental domains or others controlled by governmental institutions. The main idea is to implement a PKI that is biometrically assisted and that uses the existing hierarchical chain of command to guarantee the necessary trust to the used certificates. On the root of the several systems we find certificates that are not only well known, but also guaranteed by all the other roots (figure 3).

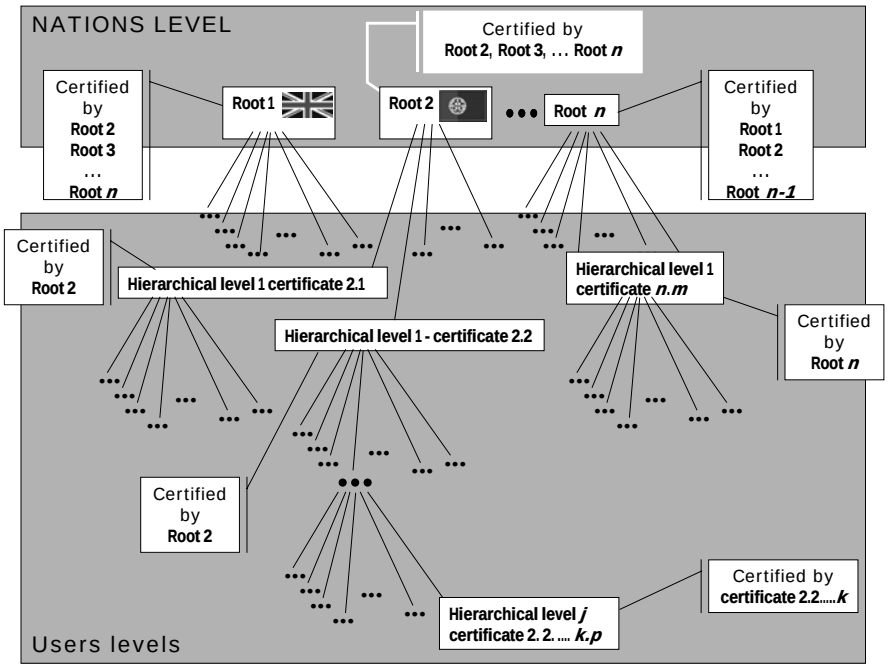

Figure 3 – Certificates confidence chain

This PKI gives some level of trust to the system but some things must be ensured to guaranty the success of the implementation:

1. Transparency: the system must work in a transparent way so that the user doesn't have to understand the principals associated to PKI. It must verify the signature of all mails received/ sign every message destined from/to a governmental address and also decript/encript those messages. If something turns out wrong an alert must be issued to the user and to the corresponding authority.
2. Authentication of the user: for one to be able to trust the system we must ensure that even if a perpetrator has access to a private key of a user and to the corresponding password he won't be able to use it. To ensure that, we'll use a keystroke dynamics biometric system to generate the key pair, to verify the password and to certify that the message was written by the legitimate owner of that e-mail address. This latter function of the biometric components returns not a binary answer but a level of trust in the presented pattern.

In this way, we ensure that all communication between public officials are encrypted, signed and, in fact, authentic. In order to allow the access of the governments to all the information exchanged within the system, all the messages are encrypted to the receiver's private key (with the corresponding public key) and to the governmental private key (with the root's public key). In this way, each government has absolute control over the corresponding subsystem, having no direct access to other government's subsystems, but the safety of its system depends on the existence of other countries subsystem that will certify the root's public key.

The biometrical data and the public keys of all users of the system are stored in a hierarchical tree with as many roots as many the countries included allowing each government to have full access to all the data generated by the system and by its use (Figure 4). If the authentication processes based on keystroke dynamics makes its way into the non-governmental e-mail providing systems, a government will be able to identify probable senders of any suspect e-mail sent and if the operating systems adopt this technology we might, like it happens today with fingerprints databases, solve many cybercrimes by comparing the patterns logged with those stored.

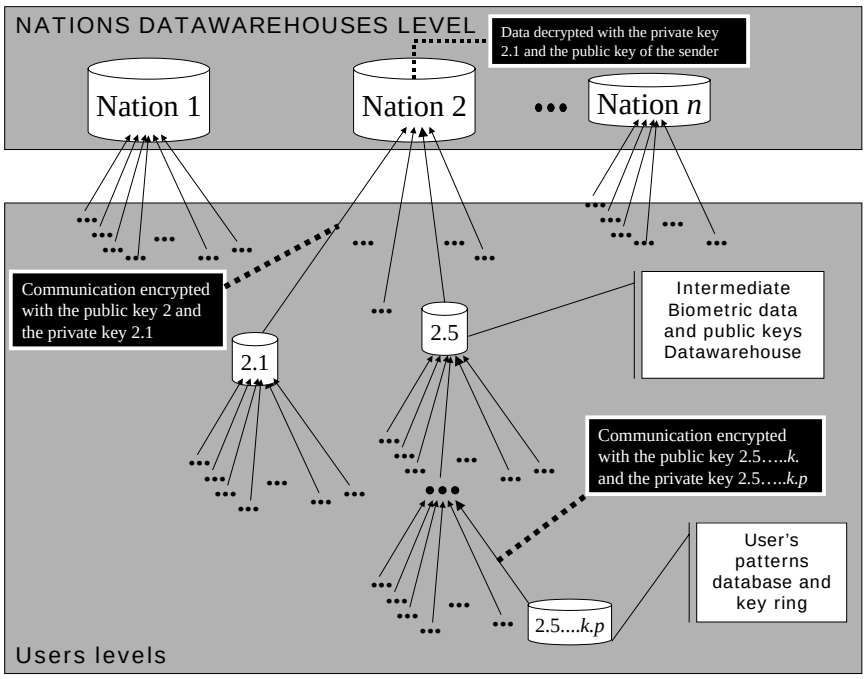

Figure 4 – Data storage of both biometric data and public keys/certificates

In order to obtain an integrated system, the cooperating governments must define diplomatic actions that will allow a query on another country's data.

6. Conclusions

This paper shows some of the vulnerabilities of the e-mail systems that are used everyday by millions all around the world, including official workers, and that can be exploited by terrorists in order to spread the panic, generate chaos or even to mislead hundreds into a certain death. In order to prevent public instructions to be given by others that those that were empowered to do so, we present

an architecture that automatically signs and encrypts every official message and, simultaneously, guarantees that the message was wrote by the owner of the corresponding certificate and that an e-mail sent using a governmental, or governmental dependent, domain can always be traced back to the person that wrote it. More, if the used biometrical technology is spread into the private systems that are supplying e-mail addresses in the Internet, one can easily identify anyone that has produced an illegal message, for instance threatening someone or cyberplanning a crime.
Once again, international cooperation between governments and between governments and private corporations, is the key that opens all of these new doors.

7. Bibliography

Thomas, T. (2003): *Al Qaeda and the Internet: The Danger of "Cyberplanning",* Parameters – U.S. Army War College Quarterly, Vol. XXXIII, No. 1, pp112-123.

Bayles, W.J. (2001): *The Ethics of Computer Network Attack,* Parameters – U.S. Army War College Quarterly, Vol. XXXI, No. 1, pp44-58.

Clemmons, B. Q. (1999): *Cyberwarfare: Ways, Warriors and Weapons of Mass Destruction,* Military Review – September-October, pp35-45.

Magalhães, S. T. and Santos, H. D. (2005): *An Improved Statistical Keystroke Dynamics Algorithm*, Proceedings of the IADIS Virtual Multi Conference on Computer Science and Information Systems, pp223-227.

Monrose, F. et al, 2001. Password Hardening based on Keystroke Dynamics. *International Journal of Information Security*.

Monrose, F. and Rubin, A. D., 1997. Authentication via Keystroke Dynamics*. Proceedings of the Fourth ACM Conference on Computer and Communication Security*. Zurich, Switzerland.

Monrose, F. and Rubin, A. D., 2000. Keystroke Dynamics as a Biometric for Authentication*. Future Generation Computing Systems (FGCS) Journal: Security on the Web*.

Gaines, R. et al, 1980. Authentication by keystroke timing: Some preliminary results. *Rand Report R-256-NSF.* Rand Corporation, Santa Monica, CA.

Joyce, R. and Gupta, G., 1990. Identity authorization based on keystroke latencies. *Communications of the ACM*. Vol. 33(2), pp 168-176.

Schneier, B., 1996, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*: John Wiley & Sons, Inc.

Kaufman, C., Perlman, R., and Speciner, M., *Network Security,* 2002*: Private Communication in a Public World*, Second Edition ed. Upper Saddle River, NJ 07458: Prentice Hall PTR.

Gritzalis, S., 2005: "A good-practice guidance on the use of PKI services in the public sector of the European Union member states," *Information Management & Computer Security*, vol. 13, pp. 379-398.

Magalhães, S. T., Revett, K., Santos, H. D., 2005: Password Secured Sites - Stepping Forward With Keystroke Dynamics, *IEEE International Conference on Next Generation Web Services Practices*, IEEE CS Press.

Revett, K., Magalhães, S. T., Santos, H. D., 2005: Developing a keystroke dynamics based agent using rough sets, *International Workshop On Rough Sets And Soft Computing In Intelligent Agent And Web Technologies*, Compégne: University of Technology of Compiègne.