International Journal of Engineering Education Vol. 28, No. 6, pp. 1380–1387, 2012
Printed in Great Britain

0949-149X/91 $3.00+0.00

A Novel Solution to the Secure Exchange of Environmental Engineering Education Data*

ROBERTO SEPÚLVEDA LIMA
Superior Polytechnic Institute José Antonio Echeverría—Informatics Engineering Faculty ISPJAE, Calle 114 No. 11901. e/Ciclovía y Rotonda, Marianao, La Habana 19390, Cuba. E-mail: sepul@ceis.cujae.edu.cu

CORNELIO YÁÑEZ MÁRQUEZ
National Polytechnics Institute—Center for Computing Research IPN CIC, Av. Juan de Dios Bátiz, Esq. Miguel Othón de Mendizábal, CIC Building, Mexico City 07738, Mexico. E-mail: cyanez@cic.ipn.mx

ITZAMÁ LÓPEZ YÁÑEZ
National Polytechnics Institute—Interdisciplinary Professional Unit on Engineering and Advanced Technologies IPN UPIITA, Av. Instituto Politécnico Nacional No. 2580, Mexico City 07340, Mexico. E-mail: ilopezy@ipn.mx

OSCAR CAMACHO NIETO
National Polytechnics Institute—Center for Technological Design and Development in Computer Science IPN CIDETEC, Av. Juan de Dios Bátiz, Esq. Miguel Othón de Mendizábal, CIDETEC Building, Mexico City 07738, Mexico. E-mail: oscarc@cic.ipn.mx

The rise and evolution of the knowledge society has brought evident advantages for humans; however, at the same time, new risks and treaths to the integrity of the information that flows constantly from one community to another have appeared too. This situation leaves at risk a very important aspect of contemporary life: knowledge transmission. Thus, the search for methods that enable secure information exchange has become a relevant topic of current scientific and engineering education research, which is closely related to the ethic values of modern society. In the current paper, a novel solution to the secure exchange of engineering education data, in the context of environmental pollution research at Mexico City, is presented.

Keywords: secure data exchange; environmental engineering education; data transmission

1. Introduction

The application of cryptographic techniques is considered to be a fundamental feature in achieving the protection of data exchanged through insecure channels, regardless of the content, topic or intention of said data. On the other hand, security incidents are reported throughout literature, regarding the lack of protection of a particular kind of information, apparently inoffensive, but of which informatic delinquents take advantage for their attacks [1]. In this sense, many private and governmental organizations are interested in a better protection for the data they store and transmit, in order to avoid these valuable resources from being consulted or even altered by non authorized personnel [17–20].

The Federal District Institute of Science and Technology (ICyTDF, *Instituto de Ciencia y Tecnología del Distrito Federal* in Spanish) of México, through the projects PIUTE10-77 and PICSO10-85, has been developing methods to predict air contaminants concentrations by applying several of the Alpha-Beta Associative Models. Such methods were selected given their competitive performance exhibited in previous experiments, regarding the prediction of air quality. Such knowledge regarding future foreseen levels of air quality in Mexico City is of particular value for both authorities and the population, since it enables and improves decision making related to health and environmental engineering knowledge management, particularly for selecting emerging measures before exceptional situations [2, 21].

An inadequate management of such knowledge can lead to excessive alarms, inaction, or increased costs and spending, making said knowledge of high sensitivity for the Mexico City government. This is why the current security solution has been developed, based on the use of cryptographic protocols and algorithms.

The rest of this paper is organized as follows: section 2 is dedicated to explaining the cryptographic concepts used by the method included in the proposed solution: the replicated sets protocol. The third section describes the proposed solution, as well as some of it's the architectural and design aspects. Then, the implementation is further detailed, while the experimental results are discussed in section 5. Conclusions are presented in section 6, and finally the references are presented.

2. Cryptographic preliminaries

This section is dedicated to introducing and discussing the cryptographic methods that give the pro-

* Accepted 2 August 2012.

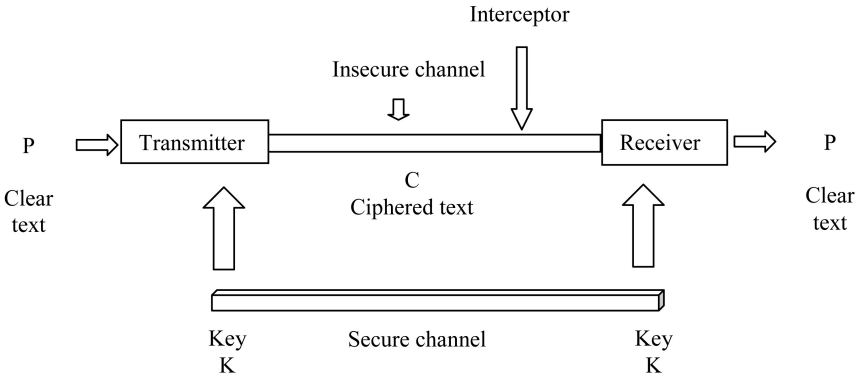


Fig. 1. Ideal model for a secret communications system.

posed solution its security and strength. First, a short introduction to cryptography and secure data transmission is given; in order to later present the particular technique to be used: the method of replicated sets.

2.1 Cryptography and its premises

The ideal model for a secret communications system was proposed by Claude Shannon in 1946 [3]; a simplification of that model is presented in Fig. 1.

The model reflects the exchange of ciphered information by means of an algorithm, which allows a transmitter to convert a clear text P into a ciphered text C, by means of a set of complex operations involving a key K, increasing the degree of confusion and diffusion of the symbols storing the information. The same algorithm decodes the ciphered text into the original clear text, with help of the known key K, which is secretly transferred through a secure communications channel. In practice, a channel is considered to be insecure if it has present an interceptor capable of capturing the ciphered texts, and which in many cases knows the algorithm that allows it to code and decode clear and ciphered texts [4].

The proposal of a perfect secure system includes [5]:

- The length of the key and the clear text to be ciphered coincide.
- The basic characters making up the key are selected randomly.
- The key is used for a single session of messages exchange.

Practice forces many restrictions to the total satisfactions of the former requirements, thus giving rise to many proposals trying to substitute the secure communications channel by means of exploiting the insecure channel to exchange the cipher keys, usually making use of asymmetric cryptography [6]. The following attributes of security are generally recognized [4, 7–10]: confidentiality of the exchange; data integrity; authentication of information, users or services; auditability; availability; and no repudiation.

It is generally accepted that Cryptography contributes substantially to achieving these attributes in several contexts, and has become in later years an essential tool for information security solutions [10]. Cryptography is based on some peculiar premises, which distinguish it from other approaches to information security. Especially, it presupposes the presence of an interceptor which has access to the communications channel and, more often than not, with explicit knowledge of the cipher algorithm employed (but not the key). Also, the interceptor is able to capture the ciphered text being transmitted on the communications channel. Thus, cryptography focuses on making the decoding of the captured information as complex as possible.

2.2 The replicated sets method

The method of the replicated sets [11] enables the exchange of cipher keys through an insecure communications channel, deriving said key from the position of an element in a secret set which both transmitter and receiver share, previous to the first exchange between them. This method exhibits the following general characteristics:

- It is a two-part protocol [4]; in other words, is based only in establishing rules of exchange between transmitter and receiver.
- The protocol is composed by three phases: key exchange and cipher transmission; transmitter authentication; and receiver authentication.
- Instead of asymmetric algorithms (which are slower in run-time for many cases), it uses a generic symmetric cipher algorithm, which allows a proper selection, tuned to each particular application.
- Uses of fundamental cryptographic functions, such as hash functions [12–14].

- The foreseen operations for modifying the content of the shared set are elemental and configurable, which facilitates their implementation either by software, hardware, or a hybrid approach.

Now, let us see the operation of the method in its three phases in greater detail.

2.2.1 Key exchange, cipher transfer and transmitter authentication phases

As mentioned before, both transmitter and receiver share some secret information previous to the exchange of the keys. This secret is the set $Q^{(Z)}$ of cardinality N (sufficiently large), made up by randomly generated integer numbers. Then, the keys exchange, cipher transfer, and the transmitter authentication are done by following the steps described below.

1. The transmitter generates a random sequence ω of w positions (w<<N) in order to concatenate enough symbols to generate a key K, to be applied in the corresponding cipher algorithm.
2. The transmitter ciphers file P (in clear text) using the corresponding cipher algorithm and the key K. That is:

$$C = f_K(P) \tag{1}$$

3. The transmitter computes a hash of the clear text P.

$$H = h(P) \tag{2}$$

4. The transmitter sends over the insecure communications channel the sequence ω, the ciphered text C, and the value H.
5. The receiver obtains from the channel the sequence ω', the ciphered text C′, and the value H′, generating the key K′ by concatenating the values in the secret set, occupied by the positions given by ω'. Notice that each element received on the insecure channel may have been altered in transit (accidentally or consciously), thus they are denoted as ω' instead of ω, for instance.
6. The receiver executes the following operations:

$$P' = f'_K(C') \tag{3}$$

$$H'' = h(P') \tag{4}$$

7. If H′ = H″ then:

$$P' = P \tag{5}$$

The operation described in step number 7 constitutes the phase of transmitter authentication before the receiver, which s based on the success of recovering a clear text whose hash value corresponds to the original text. The latter implies that:

$$C' = C \tag{6}$$

$$P' = P \tag{7}$$

$$K' = K \tag{8}$$

Equations (6), (7), and (8) indicate that the interceptor was unable to modify in a harmful manner the messages exchanged over the insecure channel.

2.2.2 Receiver authentication phase

Once the receiver has successfully done the operations of steps 5 and 6 mentioned in the former phase, its own authentication takes place. This phase is characterized by a previous agreement between both parties about a specific set of operations, proposed by the receiver, which will be applied to both sets $Q^{(Z)}$ for each session Z of key exchange and ciphering.

Let $\varphi = \{o_1, o_2, \ldots, o_m\}$ be a set of m binary operations whose arguments are two elements of the set $Q^{(Z)}$ and whose result is stored in position l of the set $Q^{(Z)}$. Given $X_i, X_j \in Q^{(Z)}$, the operations in φ may be expressed as:

$$X_{l=0_r}(X_i, X_j);\ 1 \le r \le m \tag{9}$$

Notice that $X_l \in Q^{(Z+1)}$, meaning that the selected operation modifies the element X_l of the set. The latter set $Q^{(Z+1)}$ will be used in the following session of key exchange.

Let $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_q\}$ with q < m be a sequence of operations taken randomly from φ. Then, the receiver authentication phase is made up by the following steps:

1. The receiver selects randomly the positions i, j, and l, in order to apply (9) to each operation in sequence λ.
2. The receiver applies the modifying operations on set $Q^{(Z)}$ and calculates the hash value of the obtained set.

$$Hr = h(Q^{(Z+1)}) \tag{10}$$

3. The receiver send the transmitter the cipher CA, composed by the concatenation of the sequence λ; the operators i, j, l for each operation in λ (which will be denoted as $\lambda + \{i, j, l\}_z$); and the hash value of $Q^{(Z+1)}$. The same key recovered from the exchange with the transmitter is used, calculating

$$CA = f_k(\lambda + \{i, j, l\}_z) \tag{11}$$

4. The transmitter receives Hr′ and CA′, applying the processes expressed in (12) and (13).

$$\lambda + \{i, j, l\}'_Z = f_k(CA) \tag{12}$$

5. The receiver applies the modifying operations on $Q^{(Z)}$, obtaining $Q^{(Z+1)}$ and its hash value.

$$Hr' = h(Q^{(Z+1)'}) \tag{13}$$

6. If Hr′ = Hr then:

$$Q^{(Z+1)'} = Q^{(Z+1)} \tag{14}$$

In step 6, the receiver concludes that both sets have been modified and that there is a high probability that they contain identical data, therefore the exchange of session Z+1 can take place successfully. The actions expressed in equations (12), (13), and (14) make up the receiver authentication phase.

2.2.3 Security of the replicated sets protocol

The method of replicated sets approximates Shannon ideal model in the following aspects:

- The selection of the elements which form the key is made randomly.
- The set $Q^{(Z)}$ changes notoriously for each exchange session, thus reducing the probability of deriving a key identical to that of the previous session, even when the same positions are selected.
- The security of the protocol is sustained by the security of both the symmetrical cipher and the hash function used, making this method scalable, flexible, and able to incorporate new algorithms.

On a different track, the described protocol follows the usual course of events in communication. The actual implementation, then, should take appropriate measures of contingency to inform any or both parties that some process has been aborted once started. In such situation, the exchange would continue using the secret set from the previous session.

The most promising approach to attack such protocol lies in supplanting the privileges of either the transmitter or the receiver by the interceptor through the capture of the set $Q^{(z)}$. Therefore, it may be advisable to keep this set as a ciphered file, obtaining during run-time a particular element of it.

3. Proposed solution

The basic architecture of the proposed secure communications system for transmitting air pollutants concentration knowledge follows the client-server model, in order to take advantage of well-known and tested technologies, such as the TCP/IP protocol stack [15]. In this sense, a deployment model [16] consisting of two generic components is proposed; the two components are a server module and a client module.

The client component is physically located near an environmental data acquisition station, and has a data processing login including:

- A symmetric cipher algorithm.
- A key generation algorithm, based on the use of the secret replicated set.
- A hash function.
- An implementation of the replicated steps protocol, related to the transmitter.

On the other hand, the server component is located in a central site with physical security, in which the implementation of the replicated sets algorithm related to the receiver is included. Thus, the server is expected to receive ciphered data, verifying them by means of the previously described processes of hash functions, cipher, replicated set modifying functions and operations, as well as communicating with the client.

Also, the server should follow a process of configurable consult to the clients for the environmental data exchange and its preprocessing before delivery to the Alpha-Beta associative model for its Pattern recognition processing [2, 21].

Additionally, both client and server modules include a graphical user interface (GUI) which allows them to present any information of interest to the user. Fig. 2 illustrates the deployment of the different nodes of the proposed system. Notice that communication is done only from clients to server,

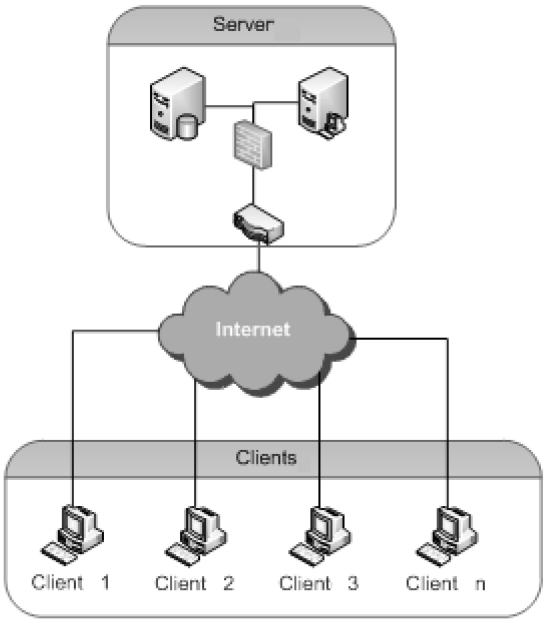


Fig. 2. Deployment of the nodes in the system.

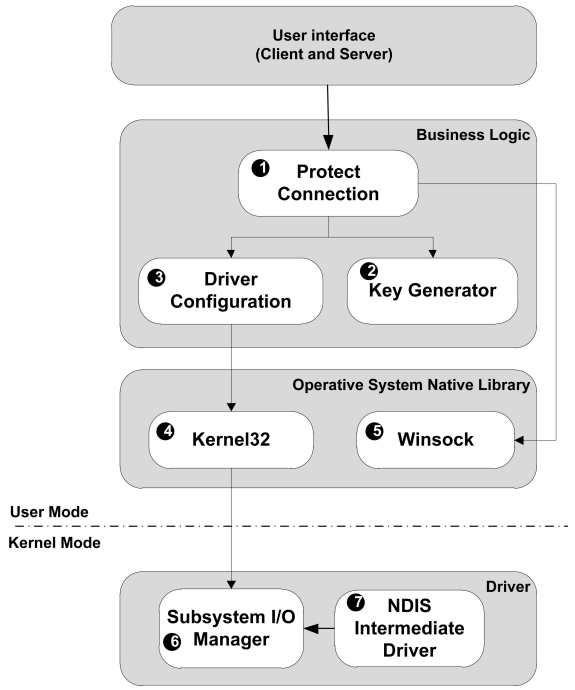


Fig. 3. System architecture

thus ensuring no exchange of information takes place between the clients through the server.

On the other hand, Fig. 3 shows the main elements of the system architecture, both of the client and server modules. The business logic associated to the keys and agreements exchange of the cryptographic algorithms is executed in user mode of the operating system. Meanwhile, the rest of the logic related to the information cipher and decipher is run in the kernel mode of the operating system.

The connection between server and clients is done using the TCP/IP protocol stack by means of the *Winsock* subsystem (5). Once the connection has been established, the *Protect Connection* subsystem (1) begins the process of protecting the connection. For this, the key and cryptographic algorithm agreement messages are exchanged between server and clients. The *Key Generator* subsystem (2) is in charge of generating the keys in both sides of the connection, parting from the exchanged information. Then, the configuration information (keys and algorithms) is sent to the *NDIS Intermediate Driver* (7) through the *Driver Configuration* subsystem (3). For the information transmission, it is necessary to establish a communication between the user mode and the kernel mode of the operating system. In this case, *Kernel32* (4) is used as a bridge to reach the *Subsystem I/O Manager* (6). It is the latter which notifies NDIS of the presence of new data, as well as the time for ciphering and deciphering the exchanged information.

The NDIS Intermediate Driver is found right on top of the physical network layer, above the Network Interface Card (NIC), and below the transport layer [15]. At this level, the NDIS can cipher and decipher the information regardless of the application which uses it. This circumstance allows a future extension of the current implementation to support other applications in a transparent manner [19, 20].

4. System implementation

Figure 4 shows the server GUI, which supports the following functionalities:

- Administration of the files containing the replicated sets for different sessions.
- Configuration of the request plan for the files

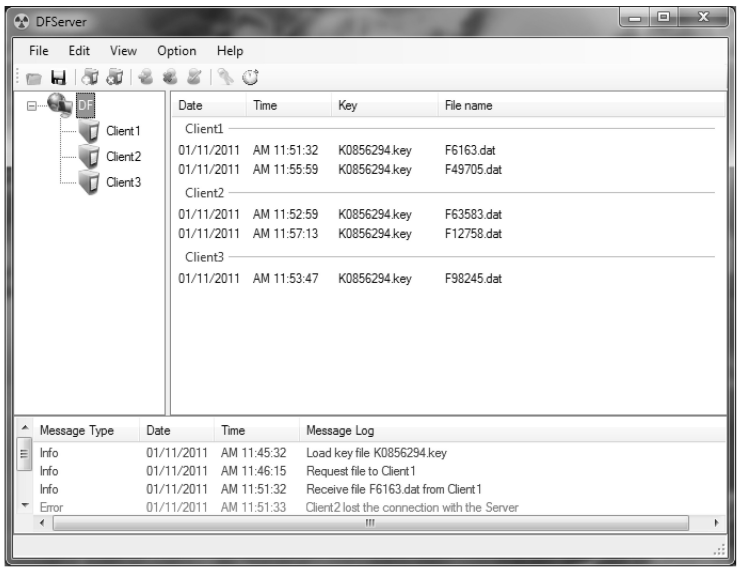


Fig. 4. Server graphical interface.

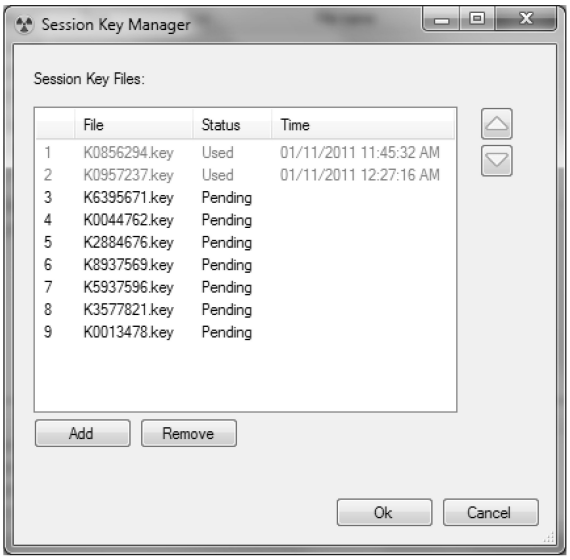


Fig. 5. Configuration of the replicated sets files.

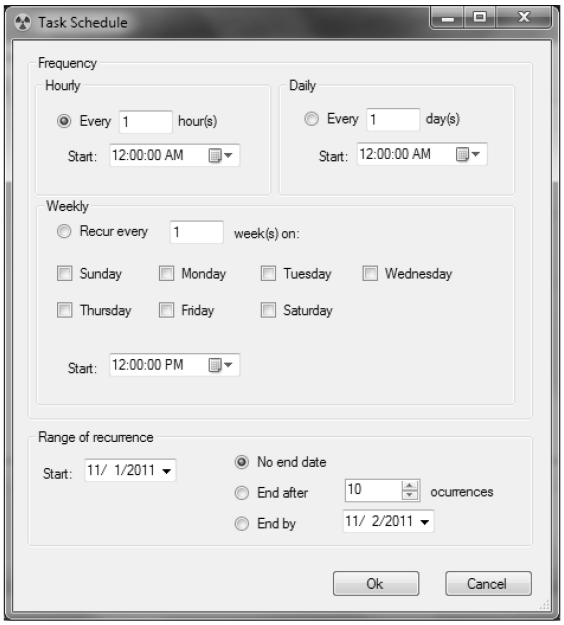


Fig. 6. Configuration of the request plan.

containing the environmental engineering information, stored in the clients.
- Maintain a register of the environmental data files received, both in general and from each client.
- Log the traces generated by the application.
- Administration of clients, users, logs, algorithms configurations, networks and connections rules, among other configurations.

Initially, the server requires the establishment of the replicated sets on independent files. It is through the configuration interface (see Fig. 5) that the user can configure the replicated sets files for each session.

Also, the user may increase the degree of automation by programming a request plan for the files containing the environmental information, which shall be delivered by the different clients, as shown in Fig. 6.

The user can program the requests of file form the client with different frequencies (hours or days) and for a particular period of time, either until the request set has been cleared or a specific date has been reached.

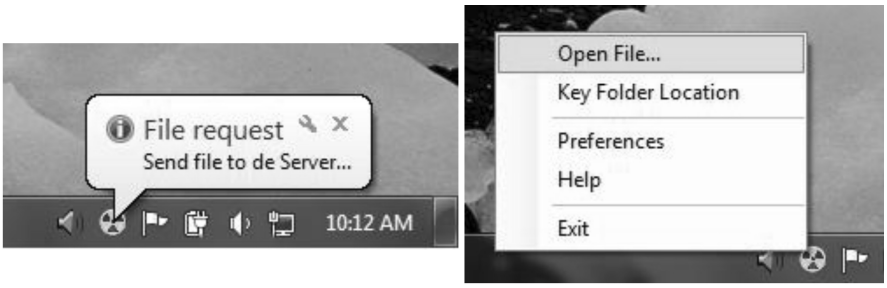
On the other hand, the client application is a lot lighter (being even deployed as a system tray application), supporting the following functionalities:

- Receiving notifications from the server (see Fig. 7a).
- Configuration of the path of the environmental information files to be sent (see Fig. 7b).
- Configuration of the path of the replicated sets files.
- Configuration of supported cryptographic algorithms, network connections, among others.

At the beginning of each session, the server agrees with the client the following information:

- Which file containing the replicated sets is to be used for the current session.
- Which cryptographic algorithms are to be used, along with their particular configurations.

The latter enables a dynamic updating of the main elements which take part of the protocol, without affecting directly the implementation of said protocol.



(a) **(b)**

Fig. 7. Client graphical interface.

5. Experimental results

The replicated sets protocol has been tested experimentally, both on point-to-point connections, as well as local and global networks. The implementation of the protocols main operations in a device driver has enabled the inclusion of the rules defined by said protocol in several applications, achieving secure information transmissions.

A test scenario consisting of three computers communicated through TCP/IP was built. Two of these computers executed the client and server software modules, respectively; while the third computer ran a network packet sniffer, in order to simulate the presence of an interceptor.

The test cases which were run have shown that the operations foreseen in the implementation are executed correctly, accomplishing successfully all of the following phases:

- Cipher keys exchange through the insecure channel.
- Authentication of both transmitter and receiver.
- Update of the replicated sets on both sides.

The tests applied to the current implementation show that the exchange done does not compromises the information to be protected, given that it is completely contained in ciphered packets and hash values. This latter fact supports the relevance of this system in terms of practical information security [4, 10, 17, 18].

6. Conclusions and future work

The proper management of the environmental data justifies the adoption of protective measures to avoid any accidental or malicious alterations of said data. The application of cryptographic techniques to the former task has been adopted, since they offer a good coverage of the security attributes while exploiting available communications networks, which usually are considered to be insecure.

This combination of needs, technology, and requirements support the security proposal discussed in this paper, which is based on the replicated sets technique. Such proposal offers several advantages: solid mechanisms for randomly selecting cipher keys, very low probability of reusing the same key on two successive cipher exchange sessions, and integrity control of the main objects of the system through hash functions and symmetric cipher algorithms.

The basis of the replicated sets protocol allows a flexible configuration of different cryptographic algorithms, either owned or developed by a third party, as well as enabling the development of implementations on software, hardware, or a hybrid (e.g. a combination of hardware on the client and software on the server). Also, the presented solution guarantees seamless collateral mechanisms for the authentication of both transmitter and receiver, without compromising the functional performance of the implementation used.

Acknowledgements—The authors would like to thank the ICyTDF (grants PIUTE10-77 and PICSO10-85), the ISPJAE, the Instituto Politécnico Nacional (Secretaría Académica, COFAA, SIP, CIC, CIDETEC, and UPIITA), the CONACyT, and SNI for their economical support to develop this work.

References

1. F. Stutzman, R. Capra and J. Thompson, Factors mediating disclosure in social network sites, *Computers in Human Behavior*, **27**(1), 2011, pp. 590–598.
2. I. López-Yáñez, C. Yáñez-Márquez, O. Camacho-Nieto and A. J. Argüelles-Cruz, Prediction of air contaminant concentration based on an associative pattern classifier, *Revista Facultad de Ingeniería, Universidad de Antioquia,* **60**, 2011, pp. 411–418.
3. C. E. Shannon, Communication Theory of Secrecy Systems; *The Bell System Technical Journal*, **28**(4), 1949, pp. 656–715.
4. B. Schneier, *Applied Cryptography*, John Wiley & Sons, CA, USA, 1996, pp. 15–30.
5. X. Lai, *On the Design and Security of Block Ciphers*, Hartung-Gorre Verlag, Zurich, Switzerland, 1992, pp. 11–83.
6. R. Kohlas, J. Jonczy and R. Haenni, A New Model for Public-Key Authentication, *Kommunikation in Verteilten Systemen (KiVS)*, Bern, Switzerland, February 26–March 2nd. 2007, pp. 213–224.
7. G. J. Simmons, *Contemporary Cryptology: The Science of Information Integrity*, IEEE Press, NJ, USA, 1992, pp. 1–40.
8. J. Ma and M. A. Orgun, Formalising theories of trust for authentication protocols, *Information Systems Frontiers*, **10**(1), 2008, pp. 19–32.
9. Baldwin and S. Shiu, Enabling shared audit data, *International Journal of Information Security*, **4**(4), 2005, pp. 263–276.
10. A. J. Menezes, P. C. V. Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press Inc., FL, USA, 1996, pp. 582–596.
11. R. Sepúlveda, *Protocolo de enlace de datos para la autenticidad y la seguridad de las comunicaciones PhD Thesis [In Spanish]*, Comisión Nacional de Grados Científicos, La Habana, Cuba, Julio 1998.
12. T. Baicheva, S. Dodunekov and P. Kazakov, On the cyclic redundancy-check codes with 8-bit redundancy, *Computer Communications*. **21**(11), 1998, pp. 1030–1033.
13. Preneel, The State of Cryptographic Hash Functions, *Lectures on Data Security, Modern Cryptology in Theory and Practice, Summer School*, Aarhus, Denmark, July 1998, 1999, pp. 158–182.
14. R. Stinson, Some Observations on the Theory of Cryptographic Hash Functions, *Designs, Codes and Cryptography*, **38**(2), 2006, pp. 259–277.
15. Comer and D. Stevens, *Internetworking with TCP/IP Vol. III: Client-Server Programming and Applications Linux/Posix Sockets Version*, Addison Wesley, Mass., USA, 2001, pp. 64–89.
16. F. Sommerville, *Software Engineering*, Addison Wesley, Mass., USA, 2010, pp. 118–133.
17. P. Korovessis, Information Security Awareness in Academia, *International Journal of Knowledge Society Research*, **2**(4), 2011, pp. 1–17.
18. M. Perez, R. Berry and C. Hollman, Information Technology Security Awareness in Academia, *Issues in Information Systems*, **4**, 2003, pp. 660–666.

19. C. Dede, Emerging technologies and distributed learning, *American Journal of Distance Education*, **10**(2), 1996, pp. 4–36.
20. I. Tuomi, *From periphery to center: emerging research topics on knowledge society*, TEKES, Helsinki, Finland, 2001, pp. 41–49.
21. C. López-Martin, I. López-Yáñez and C. Yáñez-Márquez, Application of Gamma Classifier to Development Effort Prediction of Software Projects, *Applied Mathematics & Information Sciences*, **6**(3), 2012, pp. 411–418.

Roberto Sepúlveda Lima received his PhD degree (1998) on Technical Sciences at Superior Polytechnic Institute 'José Antonio Echeverría' (ISPJAE). He is currently the President of the National Commission for the Informatics Engineering Career and Member of the National Group for the Development of Cryptography, in Cuba. Also, he is a Titular Professor at the Informatics Engineering Faculty in the ISPJAE, of which he is a former Dean. Areas of interest: Cryptography and Information Security, Artificial Intelligence, Software Engineering, and Computer Networks.

Cornelio Yáñez-Márquez obtained his Bachelor degree (1989) on Physics and Mathematics at National Polytechnics Institute (IPN) Physics and Mathematics Superior School. His MSc (1995) and PhD (2002) degrees were received at IPN Center for Computing Research (CIC). Currently a Researcher Professor, Titular C, at IPN CIC. He was granted the Lázaro Cárdenas Award by the President of the Republic. Member of the National Researchers System (SNI). Areas of interest: Associative Memories, Neural Networks, Mathematical Morphology, and Software Engineering.

Itzamá López-Yáñez received his Bachelor degree as Information Systems Engineer (2003) at Monterrey Institute of Technology and Superior Studies (ITESM), while the MSc (2007) and PhD (2011) degrees on Computer Sciences at National Polytechnics Institute (IPN) Center for Computing Research (CIC), both with mention of Honor. He was granted the Lázaro Cárdenas 2012 Award by the President of the Republic. Currently he is both a Professor at, and the President of the Telematics Academy at IPN Interdisciplinary Professional Unit on Engineering and Advanced Technologies (UPIITA). Areas of interest: Associative Memories, Neural Networks, Software Engineering, and Pattern Classification, in particular the Gamma classifier.

Oscar Camacho Nieto obtained his Bachelor degree as Communications and electronics Engineer (1989) at National Polytechnics Institute (IPN) Superior School of Mechanical and Electrical Engineering (ESIME), while the MSc degree (1995) on Computer engineering and the PhD degree (2003) on Computer Sciences, both at IPN Center for Computing Research (CIC). Currently a Researcher Professor, Titular C, at IPN, and Director of the IPN Center for Technological Design and Development in Computer Science (CIDETEC). Areas of interest: Computer Architectures, Associative memories, Microprocessors, Digital Systems, and Neural Networks.