Information & Security, Volume 15, Number 2

Secure e-Mail

e-Voting

e-Assistance

I&S Monitor

SECURE E-MAIL APPLICATION SOFTWARE FOR GOVERNMENT IN INDONESIA

Kridanto SURENDRO and Setiyo CAHYONO

Abstract: Exchanging information using e-mail brings a good deal of vulnerability that can be exploited by an unauthorized third party for individual or organizational purposes. This is quite probable since e-mail systems are designed to provide a straightforward and fast way of information delivery without considering the security of information. Prior to applying any specific security solution, an organization has to consider system characteristics and the existing problems through evaluation of security needs and faced risks. An approach that can be used to determine the security needs of an organization is risk management. Risk analysis can aid the organization in identifying the risks, why there can be a risk, to determine priorities and create prevention strategy to reduce the risks. In this article, the authors discuss the development of secure e-mail software. E-mail protection is accomplished using Secure Socket Layer (SSL) to protect the communication between the web server and the local computer, encrypting e-mail messages with combination of public and symmetric key encryption, dynamic encryption key and adding a digital signature. The experimental results show that the software can be used to protect information exchange and can reduce such security threats as eavesdropping, identity theft, false message, message modification and repudiation. Using encryption expands the size of the e-mail message to 161.96% from the actual size and the time required for encryption process is increased with 3.68%.

Keywords: Risk Analysis, Cryptography, Secure E-Mail, e-Government.

E-Government is being developed in Indonesia as a means to provide electronic-based services to citizens, to improve the quality of public services in an efficient and effective manner. Indonesian government has issued national policy and strategy for e-Government development[1] through President Instruction No.3 from 2003 as a foundation and a framework for the whole process of e-Government development. However, a consistent and supportive regulation, a standard, and a guidance are still needed, to conduct the e-Government development in a systematic and an integrated way.

E-mail is one component of the e-government services that will be implemented in local and central government organizations. Therefore, the information sent through e-mails has to be secure; it could contain confidential and urgent information. The implementation of security mechanisms in the e-mail system will reduce the risk an unauthorized party to use the information improperly for individual, group, or even one nation to another interests.

This article will discuss several aspects that can be used in determining e-mail security requirement standard for the Indonesian government. The design of an e-mail application that can assure and secure data exchange will also be presented by the authors. Hopefully, the results will become data transfer and exchange standard application for the government offices.

Basic Concept of Cryptography

Cryptography[2] is a branch of cryptology (science or study about cryptography) that deals with algorithmic design for encryption and decryption, ensuring privacy and authentication of a message. Encryption is a transformation process from a plain text (clear text and data) to some meaningless forms (cipher texts). Decryption is a reverse process to encryption. Encryption algorithms are employed for the encryption and decryption processes. The resource used for encryption and decryption is called a crypto-system.

The general encryption and decryption operations can be described as follows[3]:

$$Y = E_{KE}(X) \qquad \text{(encryption)}$$

$$X = D_{KD}(Y) \qquad \text{(decryption)}$$

where X is plain text, Y is cipher text, KE and KD are encryption and decryption keys, respectively. A general crypto system is illustrated in Figure 1.

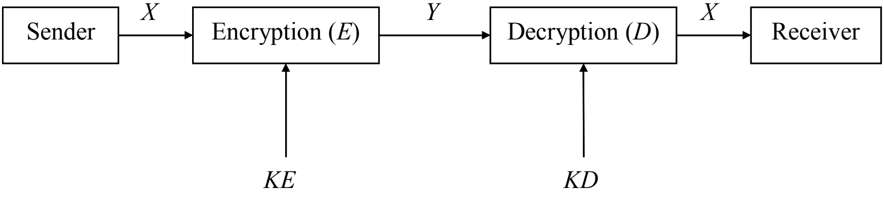


Figure 1: General Concept of a Cryptosystem.[4]

Data Encryption Standard

Data Encryption Standard (DES)[5] is a block-type cryptographic key algorithm. It takes 56 bits length key. The key is usually a 64 bits number, where every 8-th bit is used for a parity bit. Basic building block for DES is a combination of a substitution technique followed by permutation of the text, according to the key. This is known as a *round*, and DES consists of 16 rounds.

DES operates on 64 bits plain-text blocks. After initial permutation, each block is divided into two parts, a left part and a right part, with 32 bits in each part. Then 16 rounds follow in which similar operations/functions (called *f*) are performed, and data are combined with the key. After the final 16-th round, left and right parts are joined and a final permutation (reverse process to initial permutation) is performed. On each round, the bit key is shifted and 48 bits out of the 56-bits key are chosen. The left parts are expanded to 48 bits by expansion permutation, the shifted 48 bits and the permutated key are combined using XOR, resulting into 8 Sbox new 32 bits, and permutated again. Those four operations form the *f* function. The result after the *f* function is then combined again with the left parts using XOR, to produce new right parts. Old left parts then become new left parts. And these operations are repeated 16 times.

Rivest-Shamir-Adleman (RSA)

Rivest-Shamir-Adleman (RSA) is a public key cryptographic algorithm, invented by Ron Rivest, Adi Shamir, and Leonard Adleman from MIT in 1977. Difficulty to factoring big numbers is the major advantage of RSA. Public and private keys are a pair of big prime number (100-200 digits or even bigger).

A pair of keys can be produced considering the following procedure[6]:

1. Choose randomly two big prime numbers (assume p and q)
2. Calculate $n = p * q$
3. Choose randomly encryption key (e), where e and $(p-1)*(q-1)$ are relatively prime.
4. Calculate $d = e^{-1} \mod((p-1)*(q-1))$.

Numbers e and n are public keys, numbers d and n are private keys. The sender and receiver have to tell the n value. The sender knows the e value, and the receiver knows the d value.

Before encryption, the plain text M is divided into blocks, where each block has a binary value less than n. Algorithmically, encryption and decryption look like:

$$C = M^e \mod n \quad \text{(encryption)}$$

$$M = C^d \bmod n \quad \text{(decryption).}$$

Radix-64 conversion (Base 64)

Radix-64 conversion or Base 64 is used to convert binary input to printable characters. The conversion form has several characteristics, as stated below[7]:

- The range of the conversion is sets of characters that universally represent the whole set, not specific to some character sets. As such, those characters can be converted into a form required by the system. For example, the "E" character can be represented as 45 hexadecimal (in ASCII-based system) and CS (in EBCDIC-based system).
- The character set consists of 65 printable characters; one of them can be used as a padding character. With 2^6=64 characters available, each character can be used to represent a 6-bits input.
- No control characters are included. It means that a message can be converted into radix-64 form by the mail handling system that scans data stream for character control.
- The character "-" is not available. This character has its own meaning in RFC822 format and has to be avoided.

Security Requirements and Risk Analysis

Before implementing a security solution to an organization, security requirements and risks have to be evaluated thoroughly. One approach is to adopt an organization's perspective and identify what needs to be secured, why there could be a risk, and to propose a solution. This can be achieved using a risk management approach. Risk management is a continuous process to identify the risks and to implement a relevant plan to overcome it.[8]

Risk management consists of several activities, including[9]:

- Identification of risk to information security
- Risks analysis and determination of priorities
- Planning for improvement and reducing risks by developing security strategy
- Planning how to implement security strategy and reducing risk by developing complete activity plans. This activity including cost-benefit analysis is based on strategy and activities.
- Implementation of the chosen activity plan
- Monitoring of plan's improvement and effectiveness

- Control of variations in plan execution by conducting corrective action if necessary.

Risk Identification

Threat Identification

Common threat sources can be divided as[10]:

- *Natural Threats.* Floods, earthquakes, tornadoes, landslides, avalanches, electrical storms, and other similar events.
- *Human Threats.* Events that are either enabled by or caused by human beings, such as unintentional acts (inadvertent data entry) or deliberate actions (network-based attacks, malicious software upload, unauthorized access to confidential information).
- *Environmental Threats.* Long-term power failure, pollution, chemicals, liquid leakage.

This article will only discuss human threats, with the assumption that the remaining two threat sources have been reduced by the organization.

Identification of Vulnerability

Vulnerability[11] is a lack or weakness of the system either from security, design, implementation, or internal control procedures that will likely be exploited un/intentionally and will cause security system failure or violation. Mail vulnerabilities can be listed as follows[12]:

- Sending e-mail plainly; Sending e-mail from client to server and vice versa is conducted in plain form.
- Saving e-mail plainly; E-mail messages are saved in plain form to all SMTP servers, backup can be done from server and client computer. Backup of mail server is completed with copy of the message, and saved messages backup can be maintained for a certain period.
- Open port; Use of port 25 for SMTP and 110 for POP might trigger vulnerability to mail server.
- User authentication to mail server plainly; when using e-mail service, user authentication is conducted by sending username and password in plain form to the mail server.

Risk might occur, when existing vulnerability to system/application can be exploited by threat. Table 1 describes risk that is likely to occur when existing vulnerability can be exploited by threat.

Table 1: Vulnerability That Can Be Exploited by Threat.

No	Vulnerability	Threat
1	Sending e-mail plainly	Eavesdropping Message modification False Message Repudiation
2	Saving e-mail plainly	Message modification
3	Open port	Denial of service Spam Virus
4	User authentication to mail server plainly	Identity theft

Risk Analysis

Risk-level matrix has to be obtained in order to measure and scale the risk. Hence, a 3x3 matrix is constructed considering threat (high, medium, low) and impact due to its occurrence (high, medium, low). A qualitative approach is used for measurement. Risk scale is determined using likelihood, impact, the risk-level matrix, and the necessary action according to risk scale.

A survey on e-mail security has been conducted in order to understand the likelihood level of risk due to existence of threat that exploited vulnerability.[13] The results are shown in Table 2 and Table 3.

Table 2: Interview Results of Likelihood Level.

No	Vulnerability	Threat	Likelihood
1	Sending e-mail plainly	Eavesdropping	High
2	Sending e-mail plainly	Message modification	High
3	Sending e-mail plainly	False message	Medium
4	Sending e-mail plainly	Repudiation	Medium
5	Saving e-mail plainly	Message modification	Medium
6	Open port	Spam	Low
7	Open port	Virus	Low
8	Open port	Denial of service	Low
9	User authentication plainly	Identity theft	High

Table 3: Interview Result of Impact Level.

No	Vulnerability	Threat	Impact
1	Sending e-mail plainly	Eavesdropping	High
2	Sending e-mail plainly	Message modification	High
3	Sending e-mail plainly	False message	High
4	Sending e-mail plainly	Repudiation	Medium
5	Saving e-mail plainly	Message modification	High
6	Open port	Spam	Low
7	Open port	Virus	Medium
8	Open port	Denial of service	High
9	User authentication plainly	Identity theft	High

Based on the results of the survey, the risk scale can be measured using risk-level matrix (see Table 4).

Table 4: Risk Scale Measurement Result.

No	Vulnerability	Threat	Likelihood	Impact	Risk Scale
1	Sending e-mail plainly	Eavesdropping	High	High	High
2	Sending e-mail plainly	Message modification	High	High	High
3	Sending e-mail plainly	False message	Medium	High	Medium
4	Sending e-mail plainly	Repudiation	Medium	Medium	Medium
5	Saving e-mail plainly	Message modification	Medium	High	Medium
6	Open port	Spam	Low	Low	Low
7	Open port	Virus	Low	Medium	Low
8	Open port	Denial of service	Low	High	Low
9	User authentication	Identity theft	High	High	High

Security Strategy to Reduce Risk

A strategy to overcome risks can be conducted after risk scale measurement. Security strategy is only implemented to any risks that have high or medium level. The security strategy proposed in this article to reduce risk based on measurement result is described in Table 5.

Table 5: Security Strategy.

No	Vulnerability	Threat	Security	Control type
1	Sending e-mail plainly	Eavesdropping	Encryption of sent and saved e-mail message in storage media	Prevention
2	User authentication plainly	Identity theft	Securing transmission line between client and server using Secure Sockets Layer/ Transport Layer Security (SSL/ TLS)	Prevention
3	Sending/ saving e-mail plainly	Message modification	Adding message digest or finger print to e-mail message	Detection
4	Sending e-mail plainly	False message	Adding digital signature to e-mail message	Detection
5	Sending e-mail plainly	Repudiation	Adding digital signature to e-mail message	Detection

E-mail Security Strategy

Determining the e-Mail Application

The authors choose to implement a web-based e-mail application based on the following criteria:

- E-mail could be accessed from any computer anywhere.
- Users do not need to configure and install any specific e-mail application on their local computer. The application needs to be installed only on the web server.

- Local computer specifications will not pose restrictions to run the application since it locally resides at the server side.
- No restrictions exist about the operating system on the local computer. The application requires only a web browser.

In order to increase the security of a web-based application, security requirements between the local computer and the web server have to be confirmed. Communication security has to be conducted using SSL.[14] An advantage using SSL is that the users can be sure that they have access to the right server. Another is that using SSL in a web-based e-mail will ensure that all the communication between the local computer and the web server is encrypted, thus someone may not easily eavesdrop the original data. SSL will only secure the transmission line between the local computer and the web server, and the data will not be encrypted after being sent to another server.

Secure e-Mail Message Format

There are two formats in an e-mail message: a message header and a message body. To separate the message header from the message body, an empty line is used. The message header contains such information about an e-mail as sender's e-mail address, destination e-mail address, and subject. The message body contains the message itself.

The information contained in the message header is used by the mail server to deliver the e-mail to its destination. It explains why the encryption process can only be applied to the message body.

The encryption and decryption processes would have been much easier if session key, profile-id, digital signature, bit check, date, and time sent were added to the message body. Bit check is the first four characters of fingerprint and is used to check whether decryption to digital signature is successful or not. Fingerprint is created using hash function from sender's e-mail address, time sent, and message content. Figure 2 illustrates the e-mail message format that will be used.

Session Key Generation

Symmetric key cryptography applying the same key for the encryption and decryption processes of a message is used. These keys have to be maintained properly in order to prevent the access of an unauthorized person. To increase security, the key should have the following characteristics: to be random, hard to predict, and to be used only once. A key that is used only for one encryption is called a session key. Session key generation will become a problem if the user has to insert a key each time a message has to be encrypted. The system will handle the creation of session key.

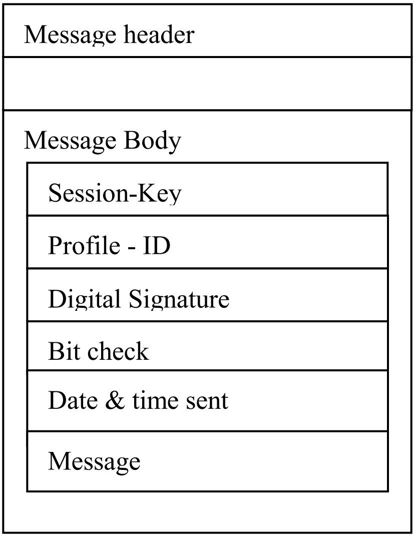


Figure 2: Secure e-Mail Message Format.

The session key will be generated as follows:

- Generate random numbers in the range from 0 to 255; as many as 100 numbers. The random number generator can be facilitated using a function inside the programming language;
- Convert these random numbers into ASCII characters;
- Calculate hash value with function hash MD5 algorithm to produce 32 hexadecimal characters;
- Convert hash value with base64 to produce 44 ASCII characters.

Using the above procedure, the session key will be different each time it is generated. Thus, different encrypted messages will be produced from the same message.

E-mail Compatibility

The e-mail system is limited to allow only printable ASCII characters. The encrypted message could consist of ASCII characters from 0 to 255, hence it cannot be directly transmitted to the e-mail server system. To overcome such a limitation, the encrypted message needs to be converted into printable ASCII character form. Base64 conversion may be used to produce printable ASCII characters. This is enabled due to the fact that the characters in base64 consist of the letters A-Z, a-z, the digits 0-9, "+", "/", "=", with no character control.

Digital Signature Generation

Digital signature is used as a proof that the original sender has composed the e-mail. This is one way to avoid denial of e-mail from both the sender and the receiver.

The process of digital signature generation can be described as follows:

- Create a fingerprint based on date/time of creation and the message using a hash function;
- Create digital signature by encrypting the fingerprint with cryptography public key using sender's private key;
- Add the digital signature to the message.

Generating a fingerprint based on date/time of creation and the message is performed to ensure that the digital signature is only valid for that message and the date/time when the e-mail has been created. Adding a digital signature to the message guarantees that the sender will not be able to deny the message. This is due to the fingerprint of the encrypted with cryptography public key message that highly depends on private key. On the other hand, the private key belongs only to the sender.

Encryption and Decryption Processes

A specific algorithm (whether encryption, a hash function, or a digital signature) will not be applied in the development of secure e-mail. This will provide flexibility and it will enable the use or addition of a specific algorithm. Encryption algorithms used for encryption, a hash function, and a digital signature are shown in Table 6.

Table 6: Cryptography Public Key, Symmetric Hash Function, and Digital Signature Algorithm.

No	Algorithm name	Type
1	Rivest Shamir Adelman (RSA)	Cryptography public key Digital signature
2	Data Encryption Standard	Cryptography Symmetric key
3	Triple DES	Cryptography Symmetric key
4	Rijndael	Cryptography Symmetric key
5	Blowfish	Cryptography Symmetric key
6	Message Digest 5 (MD5)	Hash function
7	Secure Hash Algorithm (SHA)	Hash function

The e-mail encryption process can be conducted as follows:

1. Generate a session key.

2. Generate a fingerprint from date/time of creation and message content using a hash function.
3. Generate a check bit by taking the first four characters from the fingerprint.
4. Generate a digital signature by encrypting the fingerprint with cryptographic public key using sender's private key.
5. Encrypting the check bit, date/time of creation and message content with a cryptographic symmetric key using the session key.
6. Encrypting sender's profile-id and the session key with a cryptographic symmetric key using sender's public key.
7. Combining the encrypted session key, the digital signature, and the encrypted message, and then converting it using base64.
8. To distinguish the encrypted e-mail, an identifier tag is added, "----BEGIN SECURE E-MAIL MESSAGE----" at the beginning of the message body and "----END SECURE E-MAIL MESSAGE----" at the end of the message body.

The e-mail decryption process can be conducted as follows:

1. If the message body begins with "----BEGIN SECURE E-MAIL MESSAGE----," then it is assumed that the e-mail is encrypted.
2. Extract the message body from "----BEGIN SECURE E-MAIL MESSAGE----" to "----END SECURE E-MAIL MESSAGE----"
3. Convert the message body into ASCII format using base64.
4. Split the message body into encrypted session key, digital signature, and encrypted message.
5. Decrypt the encrypted session key with cryptographic public key using the receiver's private key to obtain the session key and sender's profile-id.
6. Decrypt the encrypted message with cryptographic symmetric key using the session key to obtain the check bit, date/time of creation and message.
7. Create a fingerprint from date/time of creation and message using a hash function.
8. Decrypt the digital signature with cryptographic public key using sender's public key.
9. Compare the first four characters from the decrypted fingerprint, the digital signature, and the check bit. If they coincide, then the decryption to a digital signature has been successful.
10. Compare the fingerprint resulting from the 7-th and 8-th processes. If they coincide, then it is said to be a valid fingerprint.

E-mail Security Architecture

All processes required for e-mail management, including the encryption and decryption processes, are conducted at the server side. The web server where the secure e-mail prototype is installed is completed with SSL. If the user accesses uniform resource locator from the prototype, then SSL will be automatically activated to communicate with the web server by the browser application. In order to obtain maximum security, all the communication between the local computer, the web server, and the mail server has to be accomplished using SSL. However, in this article, SSL only applies to secure the communication between the web server and the local computer.

The process of retrieving/reading e-mail is:

1. The receiver logs on to the secure e-mail application.
2. The web server accesses the receiver's mail server, downloading e-mail and saving the messages into a database.
3. To read the e-mail, the web server retrieves the e-mail from the database.
4. If an encrypted e-mail is found, then the web server will first perform decryption, and subsequently will pass the result to receiver's computer.

The process of sending e-mail looks like:

1. The sender logs on to the secure e-mail application.
2. The sender composes a new e-mail massage and sends it to the web server.
3. If the destination address is registered into the database, then the web server will first perform encryption. On the contrary, if destination address does not exist, then encryption will not be performed.
4. The web server will send the e-mail to the receiver's mail server.

Secure E-Mail Software

Requirement Analysis

Software requirements analysis is performed using Unified Modeling Language (UML).[15] The process involves determining actors and use cases. Actor is someone or something that interacts with the system being developed. An actor can be a human being, hardware or another system. Use cases are services or functions provided by the system to its users. A use case describes the behavior of the system, including the interaction between the actor and the system.

The actors inside the secure e-mail software are:

- The web user, the actor that uses the software to compose, read, send, and receive e-mail;

- The administrator, the actor who is in charge with administering the server;
- The mail server, the actor that represents the mail server and performs sending and receiving e-mail.

The functions that are required from the software are listed as follows:

- Login, to authenticate the user who uses the system;
- System maintenance, to list accounts, to create, and register request to create an account;
- E-mail encryption/decryption, to perform e-mail encryption/decryption process;
- Registering Account, to register new user;
- Composing mail, to create an e-mail message;
- Reading mail, to read an e-mail massage;
- Sending mail, to send an e-mail message to the mail server using SMTP protocol;
- Retrieving mail, to retrieve e-mail from the mail server using POP3 protocol;
- Managing the address book, to add or delete e-mail addresses from the address book;
- Managing an account, to view or change a user account.

Based on the above-described functions, the use case diagram shown in Figure 3 can be created.

Testing

Testing[16] is conducted to see the effect of encryption on using e-mail. The size of the e-mail and the time needed for encryption become factors to be observed. Based on the testing process, encryption will increase the average e-mail size to 161.96%, with average time for encryption about 3.68 seconds.

Conclusion

The research described in this article concludes that:

- The Risk Management approach is useful to determine the security requirements that will be implemented in an organization. By performing risk analysis, an organization can identify risks, determine why risks occur, decide priority based on the risk, and create a security strategy to reduce the risks.

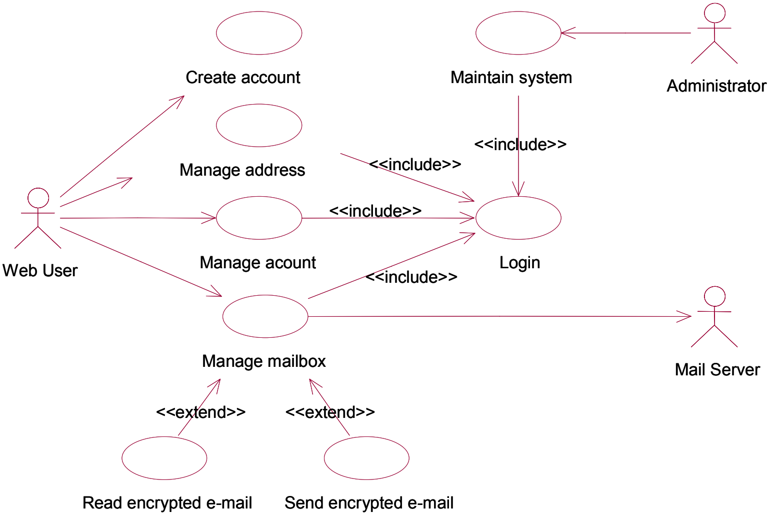


Figure 3: Use Case Diagram.

- The proposed security strategy can reduce the risks of eavesdropping, identity theft, message modification, false message, and denial of service.
- The encryption process will increase the average e-mail size to about 161.96%, with an average time for encryption about 3.68 seconds.

In accordance to the developed software, further progress needs to be done in order to reduce all possible risks.

Notes:

[1] *National Policy and Strategy for e-Government*, Indonesian's President Instruction No. 3 (2003), <http://www.ri.go.id/produk_uu/produk2003/ip2003/ip3'03.htm> (10 May 2004).

[2] Anthony Ralston, Edwin D. Reilly, and David Hemmendinger, eds., *Encyclopedia of Computer Science*, 4th edition (John Wiley & Sons, June 2000).

[3] Alferd J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, *Handbook of Applied Cryptography* (CRC Press, October 1996).

[4] Menezes, van Oorschot, and Vanstone, *Handbook of Applied Cryptography*.

[5] Bruce Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C* (John Wiley & Sons, 1995).

[6] Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*.

[7] William Stallings, *Network and Internetwork Security: Principles and Practice* (Prentice Hall, 1995).

[8] Chistopher Alberts and Audrey Dorofee, *Managing Information Security Risks: The OCTAVE Approach* (Addison Wesley Professional, 2002).

[9] Alberts and Dorofee, *Managing Information Security Risks*.

[10] Alberts and Dorofee, *Managing Information Security Risks*.

[11] Gary Stoneburner, Alice Goguen, and Alexis Feringa, *Risk Management Guide for Information Technology Systems* (National Institute of Standards and Technology, July 2002), <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf> (25 November 2004).

[12] Stoneburner, Goguen, and Feringa, *Risk Management Guide for Information Technology Systems*.

[13] Setiyo Cahyono, *Development of Secure e-Mail Prototype*, Master Thesis in Information System (Informatics Department, Institute of Technology Bandung-Indonesia, 2004).

[14] Stephen A. Thomas, *SSL and TSL Essentials: Securing the Web* (John Wiley & Sons, 2000).

[15] Wendy Boggs and Michael Boggs, *Mastering UML with Rational Rose 2002* (Sybex Inc., January 2002).

[16] Roger S. Pressman, *Software Engineering: A Practical Approach* (McGraw Hill International, 1997).

KRIDANTO SURENDRO is Head of the Information Systems Laboratory Department of Informatics Engineering Institute of Technology, Bandung, Indonesia. He graduated from the Institute of Technology, Bandung, in 1987, and received a Ph.D. degree in Computer Science from Keio University, Japan, in 1999. He is one of the nationally recognised experts in Information Systems in Indonesia. *Address for correspondence*: Information System Laboratory, Department of Informatics Engineering, Institute of Technology Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia; *Phone:* 62-22-250 8135; *E-mail:* endro@itb.ac.id.

SETIYO CAHYONO is a M.Sc. student at the Information Systems Department of Informatics Engineering Institute of Technology Bandung. He is expected to graduate his Master programme in 2004. *Address for correspondence*: Information System Laboratory, Department of Informatics Engineering, Institute of Technology Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia.

DECISION MODEL ANALYSIS FOR SPAM

Agustin ORFILA, Javier CARBO, and Arturo RIBAGORDA

Abstract: One of the security challenges in e-Government is to offer a smooth dialogue with citizens, which guarantees the availability, confidentiality and integrity of the information interchanged. Spam jeopardizes the survival of electronic mail as a communication means. Many approaches to tackle the problem with spam have been proposed. This paper shows the necessity of studying the real value of spam filters. Contrary to common belief, false positive rate and false negative rate do not completely reveal to what extent a junk filter is worth using. Very important parameters like the hostility of the environment (summarized by the probability of receiving spam) or the error costs associated with the filter play a decisive role.

Keywords: Network Security, Spam Detection, Decision Analysis, Blacklists, False Positives, False Negatives.

There is no single definition of the term spam.[1,2] The American Heritage Dictionary of the English Language[3] defines it as: "unsolicited e-mail, often of a commercial nature, sent indiscriminately to multiple mailing lists, individuals, or newsgroups." The first problem in identifying spam is that people do not agree on the meaning of the word spam. The barrier between legitimate and illegitimate email is not always clear and depends, in certain cases, on the recipient of the message. For instance, an employee could receive an email from the trade unions in order to sign some proposal. Although this is an unsolicited message, it can be considered legitimate (and even important) for the employee. In fact, Bayesian filters make use of this idea to classify email taking into account the spam concept of the user. However, most users would agree that messages about commercial items sent out in bulk should be filtered.

According to Lambert[4], MSN Hotmail stated in an e-mail sent to all its users in May 2003 that they were blocking more than 2 billion spam e-mails every day. Ferris Research reported that in 2003 the corporate user would waste 15 hours deleting e-mail, compared to 2.2 hours in the year 2000. Gartner Inc. reported that in 2002, 25% of all e-mails were qualified as spam. Brigthmail Inc.[5] points out that in April 2004 64% of total internet email was spam and that these figures are increasing month by month.

There are different techniques to detect spam. Combination of blacklists and whitelists, header analysis, content analysis, Bayesian filters, challenge filters or honeypots.[6,7,8] This article attempts to show the necessity of analyzing the value of spam filters rigorously. Some open questions studied in the paper are the influence of the spam frequency on the value of these tools and how to measure to what extent these filters are worth deploying.

The remainder of the article is organized as follows. The next section reviews related work. The decision model proposed for solving the problem with spam is described after that. Results based on the proposed model and main conclusions finalize the article.

Related work

The related work that tries to evaluate and compare proposals about spam filters uses parameters such as false positive rate and false negative rate. Blacklists[9] and Bayesian filters could be mentioned as examples.[10,11] The authors of this publication only know about an attempt to include the error costs of the filter in the analysis of its performance.[12] Androutsopoulos and co-workers define a quantity called True Cost Ratio (TCR) as a measure of how much time is wasted to delete manually all spam messages when no filter is used, compared to the time wasted to delete manually any spam messages that passed the filter plus the time needed to recover from mistakenly blocked legitimate messages. However they do not take into account the probability of receiving spam in their analysis.

The Receiver Operating Characteristic (ROC) of a detector[13] is a plot of detection probability versus false alarm probability. ROC curves have been extensively used in the characterization of communication systems and radar systems,[14] and in sciences such as Psychology[15] or Meteorology.[16] In all these fields there is a need of decision making under uncertainty in order to classify a data set. Depending on the ROC curves, ROC analysis can be incomplete because it does not take into account neither the error costs nor the hostility of the environment. In Computer Science field, a decision model analysis has been proposed for Intrusion Detection Systems (IDS) establishing that the best operating point for an IDS inherently depends on the error costs and the probability of having an intrusion.[17] They have also demonstrated that for the comparison of IDS' effectiveness these parameters should be taken into account.

Decision Model Analysis for Spam

Independently of its internal nature, we can model a spam filter as a detector. Each email that arrives to the inbox can be considered to be in one of these two states: Spam (S) or Non spam (NS). A spam detector can classify a message as Legitimate

(L) or Illegitimate (I). The main parameters of this binary detector are: the probability of classifying as illegitimate a spam email $P(I \mid S)$, also called hit rate (H), and the probability of classifying as illegitimate a non spam message $P(I \mid NS)$, also known as false positive rate (F).

Let us consider the error costs of misclassifying an email. If the message is spam and the detector does not classify it as illegitimate, there is an associated cost (L) for the recipient that consists of recognizing it and deleting it. Otherwise, if the detector classifies a non spam message as spam then it causes a cost C for the recipient that depends on the kind of filter's response. In general, these costs are asymmetric being the concrete figures related to the consequences of each error. In general, C is greater than L since losing a non spam mail is usually worse than not to filter a spam one. Most email clients with filtering capabilities do not just wipe the messages marked as spam but put them into a trash folder for later deletion. So the user would be able to rescue a badly classified email. C directly depends on the action taken over the filtered mail and the user awareness of the limitations of the current technology. If the trash folder is never consulted each false positive may cause great losses. The costs associated with the response made based on the detector's classification are given in the decision-model contingency matrix shown in Table 1.

The decision model that this article presents combines the detector's report, the response taken, the real email condition and the consequences in order to make the best decision. The best decision should be the one that minimizes expenses.

Figure 1 shows the decision tree with the sequence of actions (squares) and uncertain events (circles) that describes the detector's operation and the response actions (mark the email, move it to a spam folder, erase it, etc.) that can be taken. The costs shown correspond to the consequences of the action taken.

Decision nodes or action nodes (squares) are under the control of the decision system, which will choose which branch to follow. Event nodes (circles) are not under the control of the decision system, but depend on uncertainty. A probability distribution represents the uncertainty about which branch will follow an event node.

Table 1: Decision-Model Contingency Matrix. If a spam email comes and it is not filtered then a cost L is incurred. Otherwise, if a non spam message arrives and the detector filters it then a loss C is incurred.

		Occurs	
		Non spam	spam
Take action	No	0	L
	Yes	C	0

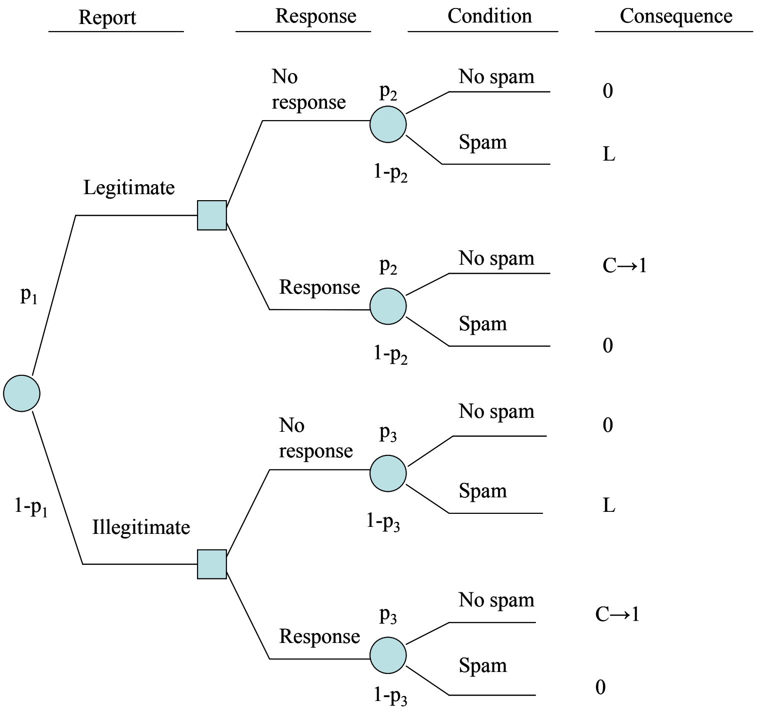


Figure 1: Decision Tree for the Spam Detector. The squares represent the possible actions that can be taken by the decision-making system and the circles describe the uncertain events that can take place.

Each combination of actions and events is characterized by its cost. There is a probability of occurrence associated to each uncertain event. There are three probabilities specified in the tree:

- p_1 : the probability that the detector's report is "legitimate"
- p_2 : the conditional probability of occurring non spam given that the detector's report is "legitimate"
- p_3 : the conditional probability of occurring non spam given that the detector's report is "illegitimate."

The last two probabilities account for both filter errors – falsely reporting that a message is spam (p_3) and falsely reporting that an email is not spam ($1-p_2$).

The expected cost is determined for event nodes by taking the sum of products of probabilities and costs for all of the node's branches. The expected cost at a decision node is the lowest expected cost from among the node's branches. The process is repeated until expected values are determined for all nodes.

Let F and H be the false alarm rate and the hit rate, respectively. An operation point is defined by a pair (F , H). Analyzing Figure 1, it is possible to determine the expected cost associated with an operating point and a specification of the best response decision to make conditional on the detector's report. Let us represent by p the prior probability that a received email is spam. The probabilities of the filter's report are calculated by applying the formula of total probability:

$$p_1 = P(L) = P(L \mid NS)P(NS) + P(L \mid S)P(S) = (1-F)(1-p) + (1-H)p$$
$$1 - p_1 = P(I) = P(I \mid NS)P(NS) + P(I \mid S)P(S) = F(1-p) + Hp \qquad (1)$$

The probabilities of the system's state conditional on the detector's report are calculated by applying Bayes' Theorem[18]:

$$p_2 = P(NS \mid L) = P(L \mid NS)P(NS)/P(L) = (1-F)(1-p)/p_1$$
$$p_2 = (1-F)(1-p)/((1-F)(1-p) + (1-H)p) \qquad (2)$$

$$1 - p_2 = P(S \mid L) = P(L \mid S)P(S)/P(L) = (1-H)p/p_1$$
$$1 - p_2 = (1-H)p/((1-F)(1-p) + (1-h)p) \qquad (3)$$

$$p_3 = P(NS \mid I) = P(I \mid NS)P(NS)/P(I) = F(1-p)/(1-p_1)$$
$$p_3 = F(1-p)/(F(1-p) + Hp) \qquad (4)$$

$$1 - p_3 = P(S \mid I) = P(I \mid S)P(S)/P(I) = Hp/(1-p_1)$$
$$1 - p_3 = Hp/(F(1-p) + Hp) \qquad (5)$$

The expected cost of each response conditional on the detector's response is calculated by taking the sum of the products of the probabilities and costs for the node following the response. The results of the expected costs are shown in Table 2.

Table 2: Expected Costs of Responses Conditional on the Detector's Report.

Detector's report	No response	Response
Legitimate	$L(1-p_2) = L(1-H)p/p_1$	$Cp_2 = C(1-F)(1-p)/p_1$
Illegitimate	$L(1-p_3) = LHp/(1-p_1)$	$Cp_3 = CF(1-p)/(1-p_1)$

The expected cost given the detector's report is the expected cost of the least costly response given the report. So the expected cost given legitimate report from the detector is:

$$Min\{L(1-H)p, C(1-F)(1-p)\}/p_1 . \tag{6}$$

Similarly, the expected cost given a report showing illegitimate behaviour is:

$$Min\{LHp, CF(1-p)\}/1-p_1 . \tag{7}$$

The expected cost M of operating at a given point (F, H) is the sum of the products of probabilities of the detector's report and the expected costs conditional on the reports. The expected cost of operating at an operating point is:

$$\begin{aligned} &M = p_1 Min\{L(1-H)p, C(1-F)(1-p)\}/p_1 + \\ &+(1-p_1)Min\{LHp, CF(1-p)\}/(1-p_1) \\ &M = Min\{L(1-H)p, C(1-F)(1-p)\} + Min\{LHp, CF(1-p)\} \end{aligned} \tag{8}$$

Without loss of generality we can rescale costs by defining the cost ratio $L' = L/C$. This substitution results in costs of 1 and L as shown in Figure 1. With this substitution

$$M = Min\{L(1-H)p, (1-F)(1-p)\} + Min\{LHp, F(1-p)\} \tag{9}$$

It is important to mention that this formulation includes the possibility that a decision is made to take action or not regardless of the detector's report. If the expected cost of going against the detector's report is lower than following its indication, the decision maker will decide contrary to the detector. This makes this model stronger than other approaches to decision analysis.[19,20]

For a perfect deterministic forecast $H = 1$, $F = 0$, hence

$$M_{per} = 0. \tag{10}$$

A detector that works based on the knowledge of the spam probability would base its decision on never protecting ($H = 0$ and $F = 0$) or always protecting ($H = 1$ and $F = 1$), which will incur an expected cost of

$$M_{fre} = Min\{Lp, (1-p)\}. \tag{11}$$

We define the value of a spam detector as a measure of the reduction in M over M_{fre}, normalized by the maximum possible reduction associated with a perfect deterministic forecast, i.e.

$$V = (M_{fre} - M)/(M_{fre} - M_{per}). \tag{12}$$

For a detection system which is no better than the one based on the probability of having spam, $V = 0$; for a perfect detector $V = 1$.

This metric is very useful because it includes all the relevant parameters involved in the evaluation of spam detector effectiveness. The decision maker has to respond in order to maximize V.

Results

The proposed decision model analysis for spam detectors can help us to understand the consequences of deploying spam detection technology in any particular scenario. The following paradigmatic examples try to show the inherent relationship between the spam detector effectiveness, the error costs, and the probability of receiving spam.

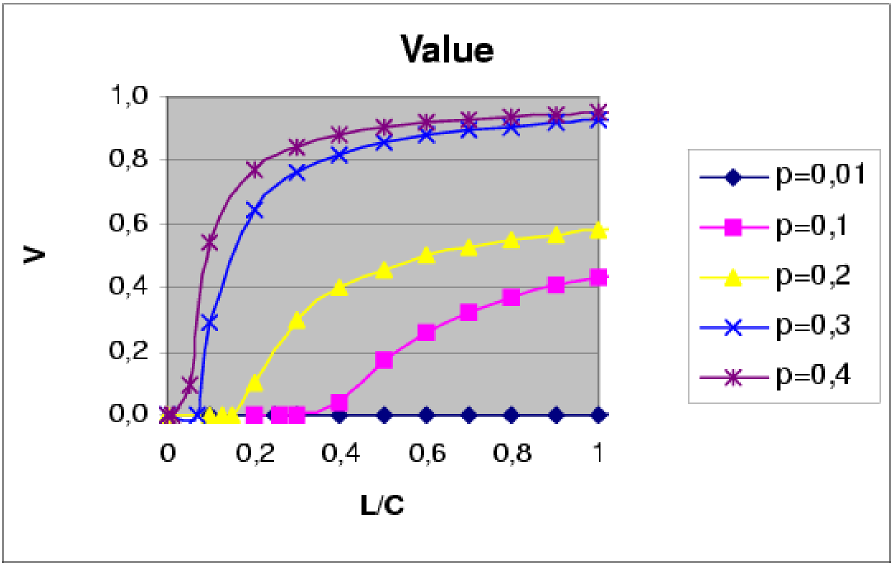


Figure 2: Decision Maker Value as a Function of L/C Relationship. Different curves are plotted for the different probabilities of receiving spam shown at the right part of the figure.

First, let us consider the case where $H = 0.995$ and $F = 0.03$ achieved by a Bayesian filter proposed by Graham.[21] Let us vary the probability of receiving spam. We focus on the cases where $L < C$ because it is more realistic to think that losing non spam email is worse than not to filter spam messages. Results are shown in Figure 2.

The spam filter is more valuable as the probability of receiving spam increases. The range L / C where it has some value is also wider as this probability increases. The value is greater as L becomes closer to C.

It is important to note that for a probability under 1% the proposed filter is worthless. In such a case it is better to base the decision just on the frequency of receiving spam. Furthermore, in the case study, if the error cost of filtering a non spam email is 5 times the cost of not to filter a spam one, then for probabilities of receiving spam under 10% the spam detector would be worthless. But for instance, for $p = 0.4$ the detector would have a 77% of the value of the perfect detector.

Let us now analyze the situation where no false positives are present. Results are shown in Figure 3 and we can see that, in the range of study ($L < C$), the value of the detector is numerically the same with the hit rate H . So, if the cost of losing an email is greater or equal than not filtering a spam one, then a detector that is able to tune it-

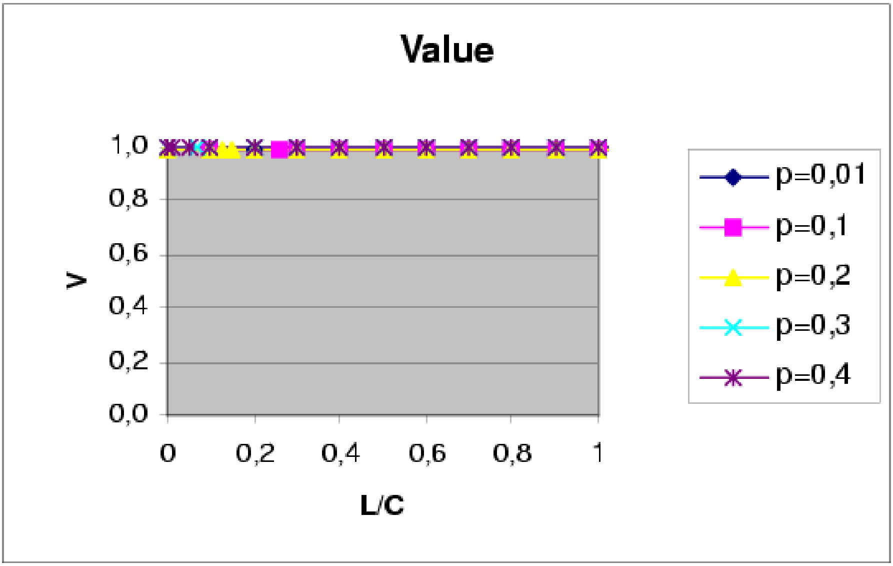


Figure 3: Decision maker value for an ideal detector that does not produce any false positives. The value is numerically equal to the hit rate for L / C range between 0 and 1.

self in order to avoid false positives is clearly superior. In theory, such a detector could exist nowadays. It is the one based on the combination of blacklists and whitelists. The whitelists always have priority over the blacklists, so a zero false positive rate could be achieved. False negatives come from addresses that are not yet in the blacklists or from spammers that have forged the sender address with one that is on the whitelist. The reasons why this is not the definite technology in real world are two. First, blacklisting a single mail address is not useful because sender addresses use to be forged but blacklisting an entire domain usually leads to a set of false positives. Second, there is an operational cost of upgrading both lists that is not considered in the presented decision model.

Conclusions

This paper proposes a new approach to the analysis of the effectiveness of spam filters. As far as the authors are aware, the parameters that have been used to analyze the different techniques have just been based on the number of false positives and false negatives. The inherent relationship between the error costs, the hostility of the environment and the filter capabilities demands a new look. Decision analysis has demonstrated very good performance in other fields and we have pointed out the advantages of applying it to spam detection. The simple decision model proposed has revealed how important would be to ignore false positives and to consider the inherent relationship between the probability of receiving spam and the value of a spam detector. As this probability increases the decision maker's opinion becomes more valuable for the usual range of error costs. The model also increases its value as the cost difference between losing a non spam email gets closer to not filtering an illegitimate one. Whatever spam filter is used and whatever the costs of misclassifying a message are, the decision maker gives a probabilistic measure of the decision value. This article has shown that in order to decide whether to deploy a solution it is not only important to softly increase the hit rate or decrease the false alarm rate but to analyze what the frequency of having spam is and to estimate the expected costs in the particular scenario. The decision model proposed would give a quantitative measure of how valuable a spam filter would become for a specific case.

Notes:

[1] *Definition of Spam*, <http://www.mail-abuse.com/spam_def.html> (26 November 2004).

[2] *Spam Defined*, <http://www.monkeys.com/spam-defined/> (26 November 2004).

[3] *The American Heritage Dictionary of the English Language*, Fourth Edition (Boston: The Houghton Mifflin Company, 2000), <http://www.bartleby.com/61/> (26 November 2004).

[4] Anselm Lambert, *Analysis of Spam*, M.Sc. Dissertation (Department of Computer Science, Trinity College Dublin, September 2003).

[5] *Spam Statistics* (Brightmail Inc., 2004), <http://www.brightlight.com/spamstats.html> (May 2004).

[6] Lambert, *Analysis of Spam*.

[7] Ion Androutsopoulos, Georgios Paliouras, Vangelis Karkaletsis, Georgios Sakkis, Constantine D. Spyropoulos, and Panagiotis Stamatopoulos, "Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach," in *Proceedings of the workshop "Machine Learning and Textual Information Access*," 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-2000), ed. H. Zaragoza, P. Gallinari and M. Rajman (Lyon, France, September 2000), 1-13.

[8] J. Davila, "La Nueva Plaga del Spam," *Newsletter SIC* 57 (November 2003).

[9] Lambert, *Analysis of Spam*.

[10] Patrick Pantel and Dekang Lin, "SpamCop: A Spam Classification & Organisation Program," in *Proceedings of AAAI-98 Workshop on Learning for Text Categorization* (Madison, Wisconsin, 1998), 95-98, <http://www.isi.edu/~pantel/Download/Papers/aaai98.pdf> (26 November 2004).

[11] Paul Graham, *A Plan for Spam* (August 2002), <http://www.paulgraham.com/spam.html> (26 November 2004).

[12] Androutsopoulos, Paliouras, Karkaletsis, Sakkis, Spyropoulos, and Stamatopoulos, "Learning to Filter Spam e-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach."

[13] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I: Detection, Estimation, and Linear Modulation Theory (John Wiley & Sons, Inc., 2001).

[14] James P. Egan, *Signal Detection Theory and ROC-Analysis*, Series in Cognitition and Perception (New York: Academic Press, 1975).

[15] John A. Swets, Robyn M. Dawes, and John Monahan, "Psychological Science Can Improve Diagnostic Decisions," *Psychological Science in the Public Interest* 1, no. 1 (May 2000): 1-26.

[16] Richard W. Katz and Allan H. Murphy, "Forecast Value: Prototype Decision-Making Models," in *Economic Value of Weather and Climate Forecasts*, ed. R.W. Katz and A.H. Murphy (Cambridge, UK: Cambridge University Press, 1997), 183–217.

[17] Jacob W. Ulvila and John E. Gaffney, Jr., "A Decision Analysis Method for Evaluating Computer Intrusion Detection Systems," *Decision Analysis* 1, no. 1 (March 2004): 35-50.

[18] Morris H. DeGroot, *Optimal Statistical Decisions* (New York, NY: McGraw-Hill Book Co., 1970).

[19] John C. Hancock and Paul A. Wintz, *Signal Detection Theory* (New York: McGraw-Hill Book Co., 1966).

[20] Alvin Martin, Mark Przybocky, George Doddington, and Douglas Reynolds, "The NIST Speaker Recognition Evaluation– Overview, Methodology, Systems, Results, Perspectives," *Speech Communications* 31 (2000), 225-254.

[21] Paul Graham, *Better Bayesian Filtering* (January 2003), <http://www.paulgraham.com/better.html> (26 November 2004).

AGUSTIN ORFILA is currently junior lecturer at the Computer Science Department of the Carlos III University of Madrid, Spain. He is a member of the Information Security Group of this department. He obtained a degree in Physics in 1999 and a M.Sc. degree in Computer Science in 2003. Mr. Orfila has several publications in international conference proceedings. His interests and his PhD research are focused on Intrusion Detection Systems and Decision Analysis. *Address for correspondence:* Despacho 22A22, Depto. Informatica, Univ. Carlos III, Av. Universidad 30, Leganes 28911 Madrid (SPAIN); *Phone*: +34 916249422; *E-mail:* adiaz@inf.uc3m.es.

JAVIER CARBO is currently Associate Professor at the Computer Science Department of the Carlos III University of Madrid, Spain. He is a member of the Research Group of Applied Artificial Intelligence at this department. Dr. Carbó has published over 30 papers in international journals and conference proceedings. He has acted as a reviewer of more than 15 international conferences; he has also been invited speaker at several research seminars and has organized 3 workshops and a special session on AI and Information Security. He has participated in a United Nations founded research project, 2 ESPRIT programs and other national projects. His interests include trust issues, automated negotiations, multi-agent systems, electronic payments and fuzzy logic. *Address for correspondence:* Despacho 21B21, Depto. Informática, Univ. Carlos III, Av. Universidad 30, Leganés 28911 Madrid (SPAIN); *Phone*: +34 916249105; *E-mail:* jcarbo@inf.uc3m.es.

ARTURO RIBAGORDA is Professor at the Computer Science Department of the Carlos III University of Madrid, Spain, and he is its current director. He leads the Information Security Group at this department and has over 50 papers in national and international journals. Dr. Ribagorda is also involved in several European and Spanish research projects and his interests are related to Network Security. *Address for correspondence:* Despacho 22A27, Depto. Informática, Univ. Carlos III, Av. Universidad 30, Leganés 28911 Madrid (SPAIN); *Phone*: +34 916249463, *E-mail:* arturo@inf.uc3m.es.

PRIVACY-PRESERVING ELECTRONIC VOTING

Justin ZHAN, Stan MATWIN, and Li-Wu CHANG

Abstract: Privacy is an important issue in electronic voting. Several electronic voting schemes have been developed in the past and some of them provided methods for dealing with privacy protection in the electronic voting system. To further enhance the privacy level, in this paper, we propose a new approach to tackle the privacy problem inherent in the electronic voting system. A privacy measure is proposed and extensive privacy analysis is conducted for the proposed scheme. It is shown via experiments that the proposed method is effective in electronic voting systems.

Keywords: Privacy, Security, Randomization, Electronic Voting.

Introduction

Election is a very important process that allows citizens to choose their government representatives. Paper-form elections have been traditionally used. People normally go to certain places, show their valid identification cards to obtain ballots, then write down their preferences and finally cast their ballots into voting boxes. In this process, election officers have to know the real identification of each person who participates in the vote. To prevent a person from voting multiple times, the officers record all persons' IDs. When new people come to vote, the officers have to check whether they have already voted. This process is not only very inefficient but it also violates the voter's privacy.

The rapid developments in computer and network technologies alter the whole election process. Electronic voting has been developed, where computers are used in the election process enabling greater effectiveness. Especially, on-line voting is the trend in voting system development. Voters do not need to go to a central place to cast their ballots; instead they can stay at home or at any other place with Internet connection to cast their ballots. Obviously, electronic voting or on-line voting[1] is desirable since it is convenient for voters and speeds up the whole election process. However, privacy and security requirements may hinder this desirable election approach from being implemented.

A list of privacy and security requirements were proposed by Neumann.[2] In this paper, we mainly deal with privacy requirements. As pointed out by Neumann,[3] voting privacy means that neither election authorities nor anyone else can link any ballot to the voter who has cast it, and all ballots remain secret while the voting is not completed. Voter anonymity can be achieved by hiding the identity of each voter from her ballot and reverse engineering cannot be made. Although various anonymous schemes have been proposed in the past, no completely satisfactory scheme exists to guarantee voter's anonymity in current electronic voting system. On the other hand, the disclosure of ballots after voting is finished also impairs voter's privacy. Therefore, a stronger privacy requirement not only guarantees the voter's privacy before voting is completed but also after its completion.

Aiming at enhancing voter's privacy, we propose a new scheme. The basic idea of our scheme is to somehow mask the vote before each voter casts her ballot, so that even if an election officer can match the identity of each voter with the actual vote, and divulge the votes after voting is finished, the election officer will still not exactly know the real vote of each voter.

The remainder of this paper is organized as follows. First related work is presented. Next, the authors present their privacy enhancing approach. Then, in the section that follows, the privacy achieved by the proposed method is measured. Experimental results are presented in a further section. This is followed by a discussion of other randomization schemes. Finally, conclusions and future research directions are outlined.

Related Work

In the early work on electronic voting, Chaum proposed the first multi-party secure election protocol, where a technique based on public key cryptography was presented.[4] However, it cannot prevent someone able to break RSA from tracing ballots back to particular voters. In another publication, Chaum proposed a protocol where a voter's privacy can be ensured if all other voters do not cooperate;[5] voters can guarantee that their ballots can be counted; and voters wishing to disrupt an election can cause only a limited delay before being disenfranchised, unless RSA is broken. Iversen[6] presented a cryptographic scheme that fully conforms to the requirements of holding large scale general elections. By ensuring independence between the voters in that they do not have to be present at the same time or go through several phases together, the scheme preserves the privacy of the voters against any subset of dishonest voters and against any proper subset of dishonest candidates, including the government. Robustness is ensured in that no subset of voters can corrupt or disrupt the election. Benaloh and Tuinstra proposed the first verifiable secrete-ballot election protocols in which participants are unable to prove to others how they voted.[7] Sako and Kilian[8] then proposed a receipt-free voting scheme based on mix-type anonymous channel,[9] with an assumption that there exists a private channel through which the center can send the voter a message without fear of eavesdropping.

Recently, Nguyen, Naini and Kurosawa proposed a formal model for security of verifiable shuffles.[10] The model is general and can be extended to mix-nets and verifiable shuffle decryption. A new efficient verifiable shuffle system based on Paillier encryption scheme was developed and its security was proved. Acquisti presented a voting protocol that protects voters' privacy and achieves universal verifiability, receipt-freeness, and uncoercibility without ad hoc physical assumptions or procedural constraints.[11] The proposed scheme allows voters to combine voting credentials with their chosen votes applying the homomorphic properties of certain probabilistic cryptosystems. A cryptographic randomized response technique[12] is developed by Ambainis, Jakobsson, and Lipmaa[13] to guarantee unconditional privacy for respondents to polls.

In this paper, we propose a flexible randomization scheme for multiple candidate elections.

Privacy Enhancing Approach

Since its introduction, electronic voting has received a great deal of attention. It is believed to be the major voting method in electronic government. Briefly, electronic voting is the process where voters submit their electronic ballots at a certain location or via Internet; the ballots are transmitted to a back-server where they are collected. After obtaining all valid ballots or the voting day has passed, the back-server counts the number of votes for each candidate. Finally, the candidate who receives certain sufficient amount of votes wins the election.

Problem

We consider the case where there are n political parties participating in the election campaign and there are totally N voters who will cast their vote. Without loss of generality, let us assume that there is only one candidate from a particular party. In other words, there are n candidates and N voters. Since voters are concerned about their ballot's privacy, they do not want to reveal their real votes to anyone including the back server. It is desirable to use some technique to mask the real vote of each voter, but we can still compute the accurate number of counts for each candidate. Based on the above requirement, we propose an estimation scheme. The basic idea of the proposed approach is that each voter randomizes the vote before sending it to the back-server. After the back-server receives the randomized votes, the number of total votes can be estimated with sufficient accuracy.

Two-Candidate Randomization Scheme

In this scheme, let us assume that there is an even number of candidates in the campaign. We randomly separate all the candidates into $n/2$ groups where each group is composed of two candidates. For example, suppose there are four candidates: C_1, C_2, C_3, and C_4. We then randomly partition them into two groups, e.g., C_1 and C_4 in one group; and C_2 and C_3 in the other group. When voters cast the ballots, they will keep their original votes with a certain probability θ; they will alter their original vote to the other candidate in the same group with probability of $1-\theta$. For example, if Alice wants to vote for C_1, she generates a random number; if the number is not greater than a certain value θ, she sends vote for C_1 to the back-server; if the number is greater than θ, she then sends vote for C_4 to the back-server. If Alice wants to vote for C_2, she generates a random number again, if the number is not greater than θ, vote for C_2 will be sent to the back-server; otherwise vote for C_3 will be sent to the back-server.

Let us assume that candidates C_i and C_j belong to the same group. For convenience, we use the following notation:

- Let Pr(i) denote the real proportion of votes for candidate *i*;
- Let Pr(j) be the real proportion of votes for candidate *j*;
- Let $\Pr^+(i)$ be the proportion of votes for candidate *i* in terms of randomized votes;
- Let $\Pr^+(j)$ be the proportion of votes for candidate *j* in terms of randomized votes.

$\Pr^+(i)$ is contributed by $\Pr(i)$ with a probability θ and by $\Pr(j)$ with a probability $1-\theta$. $\Pr^+(j)$ is contributed by $\Pr(j)$ with a probability θ and by $\Pr(i)$ with a probability $1-\theta$.

We can obtain then the following estimation model:

$$\begin{cases} \Pr^+(i) = \Pr(i) * \theta + \Pr(j) * (1-\theta) \\ \Pr^+(j) = \Pr(j) * \theta + \Pr(i) * (1-\theta) \end{cases} \tag{1}$$

What we want to compute from this estimation model is $\Pr(i)$ and $\Pr(j)$. We know that θ, $\Pr^+(i)$ and $\Pr^+(j)$ can be calculated from the randomized votes. Solving the above equations, we can obtain $\Pr(i)$ and $\Pr(j)$. Once we get $\Pr(i)$ and $\Pr(j)$, the number of votes for C_i and C_j, denoted by $Vote\_count(C_i)$ and $Vote\_count(C_j)$ respectively, can be computed as follows:

$$\begin{cases} Vote\_Count(C_i) = \Pr(i) * N \\ Vote\_Count(C_j) = \Pr(j) * N, \end{cases}$$

where N is the total number of voters.

To get $\Pr(i)$ and $\Pr(j)$, we have to apply the estimation model presented by Equation 1. How close are the estimated probabilities $\Pr(i)$ and $\Pr(j)$ to the original ones is critical for the election process. In a consequent section, a set of experiments will be conducted to test the proposed scheme.

Measuring Privacy

In a two-candidate scheme, even if an election officer somehow knows the vote (e.g., that it is for candidate C_i) of a particular voter (e.g., Alice), he/she is not sure that the true vote of Alice is for candidate C_i and only knows that Alice votes for C_i with probability of θ. In this section, we develop a privacy measure for the proposed two-candidate scheme. To preserve a fair treatment of all groups, the same θ values are used for all groups, and the privacy measure will be the same for different groups.

For a vote, the original value can be for candidate *i*, (C_i), or candidate *j*, (C_j); the randomized vote can be for candidate *i*, (C_i), or for candidate *j*, (C_j), as well. The privacy comes from the uncertainty about each voter's original vote given a randomized vote. There are four possible randomization results:

- Original vote is for C_i, but the vote after randomization is for C_i;
- Original vote is for C_i but the vote after randomization is for C_j;
- Original vote is for C_j but the vote after randomization is for C_i; and
- Original vote is for C_j, however, the vote after randomization is for C_j.

Let us adopt the following notation:

- Let X_m be the original vote;
- Let Y_m stand for the vote after randomization;
- Let W_m be the probability that the original vote is C_i, that is $\Pr(X_m = C_i)$. The probability that the original vote is C_j will be $(1 - W_m)$, that is $\Pr(X_m = C_j) = 1 - W_m$.

The privacy measure for a two-candidate scheme denoted by P_two can be derived as follows:

$$\begin{aligned} P\_two = {} & \Pr(X_m = C_i) * \Pr(Y_m = C_i \mid X_m = C_i) * \Pr(X_m = C_j \mid Y_m = C_i) \\ & + \\ & \Pr(X_m = C_i) * \Pr(Y_m = C_j \mid X_m = C_i) * \Pr(X_m = C_j \mid Y_m = C_j) \\ & + \\ & \Pr(X_m = C_j) * \Pr(Y_m = C_i \mid X_m = C_j) * \Pr(X_m = C_i \mid Y_m = C_i) \\ & + \\ & \Pr(X_m = C_j) * \Pr(Y_m = C_j \mid X_m = C_j) * \Pr(X_m = C_i \mid Y_m = C_j) \end{aligned}$$

$$= Component_1 + Component_2 + Component_3 + Component_4$$

The first component contains three parts:

- $\Pr(X_m = C_i)$ is the real probability that a voter votes for C_i, which is W_m.
- $\Pr(Y_m = C_i \mid X_m = C_i)$ is the probability that a randomized vote is for C_i given the original vote is for C_i, which is θ.
- $\Pr(X_m = C_j \mid Y_m = C_i)$ is the probability that an original vote is for C_j given that the randomized vote is for C_i. Applying Bayes' rule, we obtain $\Pr(Y_m = C_i \mid X_m = C_j) * \Pr(X_m = C_j) / \Pr(Y_m = C_i)$. $\Pr(X_m = C_j)$ is the probability that the original vote is for C_j, which is $1 - W_m$. $\Pr(Y_m = C_i \mid X_m = C_j)$ is the probability that a randomized vote is for C_i given that the original vote is for C_j, which is $1 - \theta$. As for $\Pr(Y_m = C_i)$, we can expand this term and details for that are shown below:

$$Component_1 = W_m * \theta * \frac{\Pr(Y_m = C_i \mid X_m = C_j) * \Pr(X_m = C_j)}{\Pr(Y_m = C_i)}$$

$$= \frac{\theta * (1-\theta) * W_m * (1 - W_m)}{\Pr(Y_m = C_i \mid X_m = C_i) * \Pr(X_m = C_i) + \Pr(Y_m = C_i \mid X_m = C_j) * \Pr(X_m = C_j)}$$

$$= \frac{\theta * (1-\theta) * W_m * (1 - W_m)}{\theta * W_m + (1-\theta) * (1 - W_m)}$$

Similarly, the other components can be computed as follows:

$$Component_2 = \frac{\theta*(1-\theta)*W_m*(1-W_m)}{\theta*(1-W_m)+(1-\theta)*W_m}$$

$$Component_3 = \frac{\theta*(1-\theta)*W_m*(1-W_m)}{\theta*W_m+(1-\theta)*(1-W_m)}$$

$$Component_4 = \frac{\theta*(1-\theta)*W_m*(1-W_m)}{(1-\theta)*W_m+\theta*(1-W_m)}$$

We then obtain:

$$\begin{aligned} P\_two &= \frac{\theta*(1-\theta)*W_m*(1-W_m)}{\theta*W_m+(1-\theta)*(1-W_m)} + \frac{\theta*(1-\theta)*W_m*(1-W_m)}{\theta*(1-W_m)+(1-\theta)*W_m} \\ &+ \frac{\theta*(1-\theta)*W_m*(1-W_m)}{\theta*W_m+(1-\theta)*(1-W_m)} + \frac{\theta*(1-\theta)*W_m*(1-W_m)}{(1-\theta)*W_m+\theta*(1-W_m)} \qquad (2) \\ &= \frac{2\theta*(1-\theta)*W_m*(1-W_m)}{\theta*W_m+(1-\theta)*(1-W_m)} + \frac{2\theta*(1-\theta)*W_m*(1-W_m)}{(1-\theta)*W_m+\theta*(1-W_m)} \end{aligned}$$

From Equation (2) can be seen that P_two is determined by two parameters: a control parameter θ and the original vote distribution W_m. What else we can observe from Equation (2) is that P_two is symmetric with respect to $\theta = 0.5$ and $W_m = 0.5$. For instance, for a given θ, P_two when $W_m = 0$, 0.1, 0.2, 0.3, and 0.4 is the same as the privacy when $W_m = 1$, 0.9, 0.8, 0.7, and 0.6; for a given W_m, P_two when $\theta =$ 0, 0.1, 0.2, 0.3, and 0.4 is the same as the privacy when $\theta = 1$, 0.9, 0.8, 0.7, and 0.6.

To get a better idea of our privacy measure, we conducted a set of experiments on the original data with various distributions. Specifically, we conducted experiments when $W_m = 0.1$, 0.2, 0.3, 0.4, and 0.5. For each data distribution, we compute privacy value for the cases when $\theta = 0$, 0.1, 0.2, 0.3, 0.4, and 0.5.

As we can see from the results (see Figure 1):

- For a given W_m, when $\theta = 0$, the original votes are all changed to the other value in the same group. The original votes are entirely disclosed since an adversary can change all the randomized votes to the original ones. Privacy value is 0; when θ is away from 0 and approaches 0.5, the randomization probability increases. The level of privacy enhances.

- For a given θ, the privacy level increases with the distribution of the original vote approaching 0.5. The privacy level is at its highest point when $W_m = 0.5$.

We see how privacy changes with varying θ and W_m. In practice, an important issue is how to select a proper value for θ. It cannot only be determined by the privacy level. Accuracy of the results is another critical factor for choosing θ. In the next section, we will conduct a set of experiments and show the relationship between the accuracy and θ.

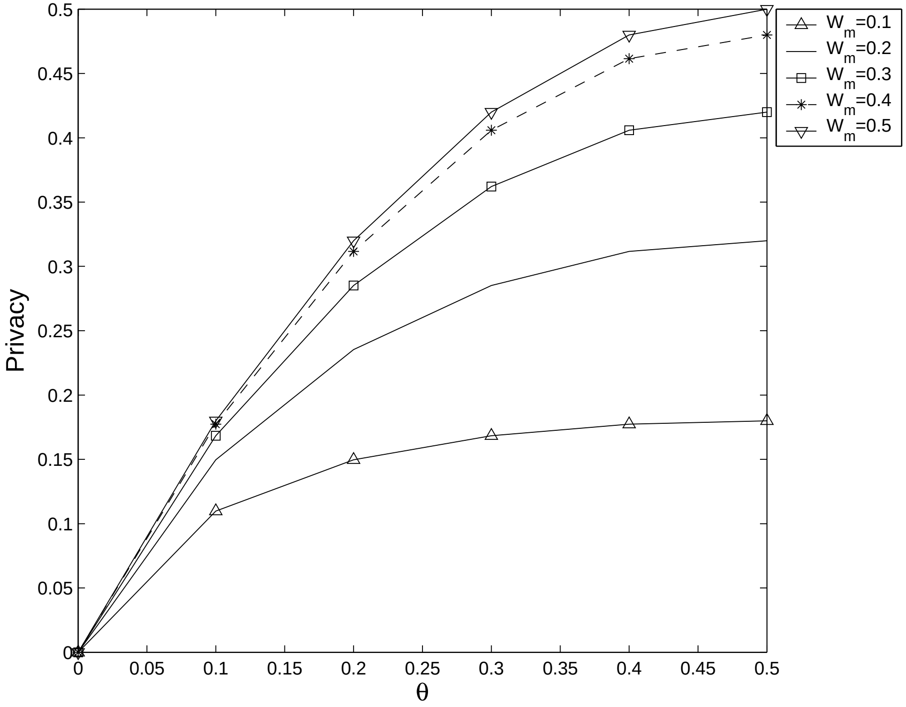


Figure 1: Results from Experiments on the Original Data Set.

Experimental Results

To evaluate the effectiveness of the proposed scheme, we have conducted a number of experiments on data sets with various distributions. The data sets are randomly generated according to different distributions. In the two-candidate scheme, there are only two candidates C_i and C_j in one group. What we want to know is how close is the estimated proportion of votes to the true proportion of votes for each candidate.

Experimental Steps

1. Data Generation

The data sets used in the experiments are randomly generated according to the given distribution. The authors evaluate the proposed scheme on nine data sets whose distributions are as follows:

Data Sets	Pr(C_i)	Pr(C_j)
D_1	0.1	0.9
D_2	0.2	0.8
D_3	0.3	0.7
D_4	0.4	0.6
D_5	0.5	0.5
D_6	0.6	0.4
D_7	0.7	0.3
D_8	0.8	0.2
D_9	0.9	0.1

2. θ Selection

For $\theta = 0$, 0.1, 0.2, 0.3, 0.4, 0.51, 0.6, 0.7, 0.8, 0.9, and 1, the following steps are performed on data set D_1:

- *Step I. Randomization*. For each value in D_1, we generate a random number r $(0 \le r \le 1)$ according to an uniform distribution. If $r \le \theta$, the value remains the same; otherwise, the value of *r* will be changed to its opposite. For example, assume that the original value is $C_i(C_j)$, if $r \le \theta$, the value after randomization will still be $C_i(C_j)$; otherwise, the value after randomization will be changed to $C_j(C_i)$. We perform this randomization step for all the values in D_1. The data set after the randomization process is denoted by G_1.
- *Step II. Estimation*. The model shown by Equation 1 is estimated on the randomized data set G_1. We then obtain Pr(C_i) and Pr(C_j). Due to the fact

that $\Pr(C_i) = 1 - \Pr(C_j)$, we only record $\Pr(C_i)$ in the resultant tables.

- *Step III. Repeating.* Steps I and II are repeated 50 times, and 50 $\Pr(C_i)s$ are obtained.
- *Step IV. Mean, Variance and Error.* The mean, variance and error percentage of these 50 $\Pr(C_i)s$ are computed.

3. Compute Mean, Variance and Error for Other Data Sets

Step 2 is performed on D_2, D_3, …, and D_9.

Results Analysis

Tables 1, 2, ..., and 9 show the experimental results on data sets D_1, D_2, D_3, …, and D_9, respectively. First, the results are presented followed by a detailed analysis.

Table 1: Results on Data Set D_1.

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.1	0.1	0.1	0.1	0.1	0.0995	0.0998	0.1	0.1	0.1	0.1
Var $(*10^{-3})$	0	0	0.0001	0.0004	0.0016	0.1489	0.0015	0.0004	0.0001	0	0
Error (%)	0	0	0	0	0	0.05	0.02	0	0	0	0

Table 2: Results on Data Set D_2.

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.2	0.2	0.2	0.2	0.2	0.1996	0.2	0.2	0.2	0.2	0.2
Var $(*10^{-3})$	0	0	0.0001	0.0003	0.0012	0.1264	0.0013	0.0002	0.0001	0	0
Error (%)	0	0	0	0	0	0.04	0	0	0	0	0

Table 3: Results on Data Set D_3.

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.3	0.3	0.3	0.3001	0.3	0.2995	0.2999	0.3001	0.3	0.3	0.3
Var $(*10^{-3})$	0	0	0.0001	0.0002	0.0012	0.1244	0.0012	0.0002	0.0001	0	0
Error (%)	0	0	0	0.01	0	0.05	0.01	0.01	0	0	0

Table 4: Results on Data Set D_4 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.4	0.4	0.4	0.4001	0.4001	0.3988	0.4	0.4	0.4	0.4	0.4
Var $(*10^{-3})$	0	0	0.0001	0.0002	0.0010	0.1498	0.0010	0.0002	0.0001	0.0001	0
Error (%)	0	0	0	0.01	0.01	0.12	0	0	0	0	0

Table 5: Results on Data Set D_5 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.5	0.5	0.5	0.5	0.4999	0.5010	0.5	0.5	0.5	0.5	0.5
Var $(*10^{-3})$	0	0	0.0001	0.0002	0.0011	0.1316	0.0013	0.0003	0.0001	0	0
Error (%)	0	0	0	0	0.01	0.1	0	0	0	0	0

Table 6: Results on Data Set D_6 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.6	0.6	0.6	0.6	0.5998	0.6020	0.6001	0.6	0.6	0.6	0.6
Var $(*10^{-3})$	0	0	0.0001	0.0004	0.0021	0.2469	0.0017	0.0003	0.0001	0	0
Error (%)	0	0	0	0	0.02	0.2	0.01	0	0	0	0

Table 7: Results on Data Set D_7 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.7	0.7	0.7	0.6999	0.6999	0.7015	0.7	0.7001	0.7	0.7	0.7
Var $(*10^{-3})$	0	0.0001	0.0001	0.0003	0.0018	0.144	0.0016	0.0004	0.0001	0	0
Error (%)	0	0	0	0.01	0.01	0.15	0	0.01	0	0	0

Table 8: Results on Data Set D_8 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.8	0.8	0.8	0.8	0.7999	0.8018	0.8	0.8	0.8	0.8	0.8
Var $(*10^{-3})$	0	0	0.0001	0.0002	0.0010	0.1498	0.0010	0.0002	0.0001	0.0001	0
Error (%)	0	0	0	0	0.01	0.18	0	0	0	0	0

Table 9: Results on Data Set D_9 .

θ	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.51*	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*
Mean	0.9	0.9	0.9	0.9	0.9	0.9010	0.9	0.9	0.9	0.9	0.9
Var $(*10^{-3})$	0	0	0.0001	0.0004	0.0020	0.2500	0.0022	0.0004	0.0001	0.0001	0
Error (%)	0	0	0	0	0	0.1	0	0	0	0	0

Analysis of Mean and Variance

It can be seen from the tables with the results that when $\theta = 0$ and $\theta = 1$, the estimated proportion of votes is exactly the same as the original proportion, and the variance is 0. This is due to the fact that the randomization process for these two cases does not hide the original votes. When θ deviates from 1 and 0 and approaches 0.5, the level of randomization increases, and as a result the original information is better disguised. Therefore, the mean of the estimated proportion may deviate from the original proportion and the variance has a trend of becoming larger. Note that when $\theta = 0.5$, the estimation model cannot be applied since Equation (1) does not have a solution in this case. Therefore a value $\theta = 0.51$, instead of 0.5, is used in the experiments.

Error Rate Analysis

Estimation error is an important factor in electronic voting. The estimated proportion should not differ very much from the true proportion. Otherwise, the proposed scheme cannot be applied in real electronic voting. Let us use the upcoming (for the time of writing) presidential elections in the United States for illustration. Assume that the true vote proportion for Kerry is 55% and he should be elected as the new president, but the estimated proportion is 45% and he loses the campaign in result. Obviously, it leads to a serious problem. Therefore, the estimated proportion has to be very close to the original proportion. And the most desirable case is when they are the same. As we can see from the results, for most of the θ values, the estimated proportion is the same as the true proportion. From accuracy point of view, the proposed two-candidate scheme is efficient in these cases.

There are also some limitations in the proposed scheme. Before election, a threshold needs to be agreed among all candidates. Since there are error rates for some cases, we need to find a threshold that is significantly greater than the possible error rates. The error rate reaches its highest point when $\theta = 0.51$ for all data sets. As we can see the highest error rate for D_1 is 0.05%, for D_2 is 0.04%, for D_3 is 0.05%, for D_4 is 0.12%, for D_5 is 0.1%, for D_6 is 0.2%, for D_7 is 0.15%, for D_8 is 0.18%, and for

D_9 is 0.1%. The threshold has to be significantly higher than the highest possible error rate, e.g., 0.2%.

Accuracy and privacy are complementary goals. Given W_m, the best privacy is achieved when the control parameter θ is 0.5; however, the accuracy will be worst in this case. The best accuracy is attained when the control parameter θ is 0 or 1, however, the privacy is at its lowest level then. Trade-offs are also applied when θ has a value between 0 and 1. In practice, how to select θ depends on our primary goal. If we want the results to be very precise, we have to choose values near 1 or 0; in contrast, if privacy is the primary goal, we choose values near 0.5.

Extension of the Randomization Scheme

Three-Candidate Scheme

In the two-candidate scheme, the randomization process is applied between two candidates since each group contains two candidates. The number of candidates within each group can be increased. In this section, the case when the number of candidates within each group is three is considered. Without loss of generality, let us assume that *n* can be perfectly divided by 3. We randomly separate all the candidates into *n*/3 groups where each group contains three candidates. Suppose that C_i, C_j and C_k belong to the same group. When voters cast their ballots, they keep their original votes with a probability θ_1, they alter their original vote to other candidate with probability θ_2 and θ_3, respectively. For instance, if Alice's original vote is for C_i, instead of directly sending the vote for C_i to the back-server, she generates a random number r_1. If $r_1 \le \theta_1$, she sends a vote for C_i to the back-server. If $\theta_1 \prec r_1 \le \theta_2$, she sends a vote for C_j to the back-server. If $r_1 \succ \theta_2$, she sends a vote for C_k to the back-server. In other words, she keeps her original vote for C_i with a probability θ_1, modifies her original vote to C_j with a probability θ_2 and to C_k with a probability $\theta_3 = (1-\theta_1-\theta_2)$.

Let us use the following notation:

- $\Pr(i)$ is the real proportion of votes for candidate i (C_i).
- $\Pr(j)$ is the real proportion of votes for candidate j (C_j).
- $\Pr(k)$ is the real proportion of votes for candidate k (C_k).
- $\Pr^+(i)$ is the proportion of votes for candidate *i* in randomized votes.
- $\Pr^+(j)$ is the proportion of votes for candidate *j* in randomized votes.

- $\Pr^+(k)$ is the proportion of votes for candidate *k* in randomized votes.

$\Pr^+(i)$ is contributed by $\Pr(i)$ with probability θ_1, $\Pr(j)$ with probability θ_3, and $\Pr(k)$ with probability θ_2. $\Pr^+(j)$ is contributed by $\Pr(j)$ with probability θ_1, $\Pr(i)$ with probability θ_2, and $\Pr(k)$ with probability θ_3. $\Pr^+(k)$ is contributed by $\Pr(k)$ with probability θ_1, $\Pr(j)$ with probability θ_2, and $\Pr(i)$ with probability θ_3.

The estimation model can be built as follows:

$$\begin{cases} \Pr^+(i) = \Pr(i) * \theta_1 + \Pr(j) * \theta_3 + \Pr(k) * \theta_2 \\ \Pr^+(j) = \Pr(i) * \theta_2 + \Pr(j) * \theta_1 + \Pr(k) * \theta_3 \\ \Pr^+(k) = \Pr(i) * \theta_3 + \Pr(j) * \theta_2 + \Pr(k) * \theta_1 \end{cases} \tag{3}$$

In the above model θ_1, θ_2 and θ_3 are known. $\Pr^+(i)$, $\Pr^+(j)$ and $\Pr^+(k)$ can be computed directly from the randomized votes. We can then solve the above model and obtain $\Pr(i)$, $\Pr(j)$ and $\Pr(k)$. The total number of votes for candidate *i* is $\Pr(i) * N$, for candidate *j* is $\Pr(j) * N$ and for candidate *k* is $\Pr(k) * N$.

n-Candidate Scheme

In general, we can treat all the candidates in one group. We call it n-candidate scheme. Voters keep their true votes with a probability θ_1 and alter the vote to the other candidates with probabilities θ_2, θ_3, …, and θ_n, respectively.

Let us assume the following notation:

- $\Pr(C_m)$ $(m = 1, 2, \cdots, n)$: the real proportion of votes for candidate *m* (C_m).
- $\Pr^+(C_m)(m = 1, 2, \cdots, n)$: the proportion of votes for candidate *m* in randomized votes.

Then the estimation model will look as follows:

$$\begin{cases} \Pr^+(C_1) = \Pr(C_1) * \theta_1 + \Pr(C_2) * \theta_n + \Pr(C_3) * \theta_{n-1} + \Pr(C_4) * \theta_{n-2} + \ldots + \Pr(C_n) * \theta_2 \\ \Pr^+(C_2) = \Pr(C_1) * \theta_2 + \Pr(C_2) * \theta_1 + \Pr(C_3) * \theta_n + \Pr(C_4) * \theta_{n-1} + \ldots + \Pr(C_n) * \theta_3 \\ \Pr^+(C_3) = \Pr(C_1) * \theta_3 + \Pr(C_2) * \theta_2 + \Pr(C_3) * \theta_1 + \Pr(C_4) * \theta_n + \ldots + \Pr(C_n) * \theta_4 \\ \qquad\qquad\qquad\qquad \cdots\cdots \\ \Pr^+(C_n) = \Pr(C_1) * \theta_n + \Pr(C_2) * \theta_{n-1} + \Pr(C_3) * \theta_{n-2} + \Pr(C_4) * \theta_{n-3} + \ldots + \Pr(C_n) * \theta_1 \end{cases}$$

In the above estimation model, we can estimate $\Pr(C_m)$ $(m = 1, 2, \cdots, n)$ since θ_m $(m = 1, 2, \cdots, n)$ are known and $\Pr^+(C_m)$ $(m = 1, 2, \cdots, n)$ can be directly computed from the collected randomized votes. The total number of votes for candidate *m* is then $\Pr(C_m) * N$ $(m = 1, 2, \cdots, n)$.

We see that the number of equations in the estimation model of the n-candidate scheme is equal to the number of candidates participating in the campaign. Although the two-candidate scheme is much simpler, other candidate schemes can certainly be used. In practice, different schemes can be combined together. For instance, since the number of candidates may not be perfectly divided by two or three, we can combine two- and three-candidate schemes. One possibility is to separate all the candidates into *g* groups with *g-1* groups containing two candidates and with 1 group containing three candidates.

Conclusions and Future Work

Electronic voting is an efficient way of performing governmental elections. Especially, the controversial year 2000 elections in the United States of America made people realize the importance of electronic voting. Although electronic voting or online voting systems can significantly improve the efficiency of voting, security and privacy violations may prevent them from being implemented. To preserve privacy in electronic voting, a privacy protection method has been introduced in this paper. In the proposed technique, voter's vote is randomized before sending it to back-server. The performed experiments have illustrated that the election results are still very accurate although the original votes have been hidden. A privacy measure has been developed and a privacy analysis conducted. Trade-offs between privacy and accuracy have been discussed. In the future, the authors intend to combine the proposed technique with other privacy and security protection techniques for electronic voting. The approach will also be extended to other e-government services.

Notes:

[1] Aviel D. Rubin, "Security Considerations for Remote Electronic Voting over the Internet" (22 June 2001), <http://avirubin.com/e-voting.security.pdf> (15 November 2004).

[2] Peter G. Neumann, "Security Criteria for Electronic Voting," (paper presented at the 16th National Computer Security Conference, Baltimore, Maryland, September 1993), 20-23.

[3] Neumann, "Security Criteria for Electronic Voting."

[4] David L. Chaum, "Untraceable Electronic Mail, Return Address, and Digital Pseudonyms," *Communication of the ACM* 24, no. 2 (1981): 84-88.

[5] David L. Chaum, "Elections with Unconditionally-Secret Ballots and Disruption Equivalent to Breaking RSA," in *Advances in Cryptology - EUROCRYPT '88* (Berlin, 25-27 May

1988), *Lecture Notes in Computer Science* 330, ed. Christoph G. Gunther (Springer-Verlag, 1988), 177-182.

[6] Kenneth R. Iversen, “A Cryptographic Scheme for Computerized General Elections,” in *Advances in Cryptology -- CRYPTO '91* (11-15 August 1991), *Lecture Notes in Computer Science* 576, ed. J. Feigenbaum (Berlin: Springer-Verlag, 1992), 405-419.

[7] Josh D. Benaloh and Dwight Tuinstra, “Receipt-Free Secret-Ballot Elections (extended abstract),” in *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing STOC’94* (New York, May 1994), 544-553.

[8] Kazue Sako and Joe Kilian, “Receipt-Free Mix-Type Voting Scheme – a Practical Solution to the Implementation of a Voting Booth,” in *Advances in Cryptology - EUROCRYPT’95* (Saint-Malo, France, 21-25 May 1995), *Lecture Notes in Computer Science* 921, ed. Louis C. Guillou and Jean-Jacques Quisquater (Berlin: Springer-Verlag, 1995), 393-403.

[9] Choonsik Park, Kazutomo Itoh, and Kaoru Kurosawa, “All-Nothing Election Scheme and Anonymous Channel,” in *Advances in Cryptology - EUROCRYPT’93* (Lofthus, May 1993), *Lecture Notes in Computer Science* 765, ed. T.Helleseth (Berlin: Springer-Verlag, 1993), 248-259; Chaum, “Untraceable Electronic Mail, Return Address, and Digital Pseudonyms.”

[10] Lan Nguyen, Reihaneh Safavi-Naini, and Kaoru Kurosawa, “Verifiable Shuffles: A Formal Model and a Paillier-Based Efficient Construction with Provable Security,” in *Applied Cryptography and Network Security: Second International Conference ACNS 2004* (Yellow Mountain, China, 8-11 June 2004), *Lecture Notes in Computer Science* 3089, ed. Markus Jakobsson, Moti Yung, and Jianying Zhou (Heidelberg: Springer-Verlag, 2004), 61-75.

[11] Alessandro Acquisti, “Receipt-Free Homomorphic Elections and Write-in Voter Verified Ballots,” Technical Report CMU-ISRI-04-116 (Carnegie Mellon University, School of Computer Science, April 2004), <citeseer.ist.psu.edu/663984.html> (15 November 2004).

[12] Stanley L. Warner, “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias,” *Journal of American Statistical Association* 60, no. 309 (March 1965): 63-69.

[13] Andris Ambainis, Markus Jakobsson, and Helger Lipmaa, “Cryptographic Randomized Response Techniques,” in *Public Key Cryptography – PKC 2004: 7th International Workshop on Theory and Practice in Public Key Cryptography* (Singapore, 1-4 March 2004), *Lecture Notes in Computer Science* 2947, ed. Feng Bao, Robert H. Deng, and Jianying Zhou (Heidelberg: Springer-Verlag, 2004), 425-438.

JUSTIN ZHAN is a part-time professor at the School of Information Technology and Engineering, University of Ottawa, Canada. His research interest is privacy and security issues in data mining. *E-mail:* zhizhan@site.uottawa.ca.

STAN MATWIN is a professor at the School of Information Technology and Engineering, University of Ottawa, Canada. His research is in machine learning, data mining, and their applications, as well as in technological aspects of Electronic Commerce. *E-mail:* stan@site.uottawa.ca.

LI-WU CHANG is a research scientist at Center for High Assurance Computer Systems of Naval Research Laboratory, USA. *E-mail:* lchang@itd.nrl.navy.mil.

IMPROVED ANONYMOUS SECURE E-VOTING OVER A NETWORK

Chou-Chen YANG, Ching-Ying LIN, and Hung-Wen YANG

Abstract: In a democratic country voting is one of the most important activities. However, many eligible voters do not exercise their right simply because they do not want to visit a public booth where they can vote. In 1981, David Chaum first introduced the concept of electronic voting and attempted to overcome the problems associated with the traditional voting environment. Hereafter, Mu and Varadharajan proposed in 1998 an anonymous secure e-voting scheme over a network. They claimed that the proposed scheme is not only capable of preventing double voting but it can also protect the privacy of voters. However, many researchers afterwards have discovered that Mu and Varadharajan's scheme is not secure. Attackers can easily forge a valid ballot and can vote more than once. In this paper, an e-voting scheme based on Mu and Varadharajan's scheme is proposed that meets the following e-voting requirements: democracy, accuracy, anonymity, mobility and efficiency.

Keywords: Anonymity, Blind Signature, ElGamal Public Key Crypto-System, e-Voting.

In a democratic society voting is one of the most important activities. In such a traditional voting environment, the voting process sometimes becomes quite ineffective and inconvenient due to the fact that the eligible voters have to visit a voting booth to cast their votes. Very often the eligible voters do not exercise their right to vote simply because they do not have time to visit a voting booth. Besides, the traditional voting method requires more expenses and involves more social resources, social cost and human resources. Therefore, the effort of many researchers has been put into finding solutions to the problems inherent in traditional voting environment. The development of computer networks and the elaboration of the cryptographic techniques facilitate the implementation of electronic voting.

Electronic voting is very convenient for the voters due to the fact that they can cast their ballots through a network. Even if the voters do not have time to go to the voting booth they can still cast they vote through a computer with Internet connection. Thus, they can exercise their right to vote. Furthermore, this electronic voting method can reduce the expenses and avoid errors. The first electronic election scheme was proposed by David Chaum;[1] then the same author suggested the concept of blind signatures.[2] Later on, many researchers have worked and proposed various developments in e-voting.[3,4,5,6,7,8,9,10,11,12,13] The existing e-voting systems can be divided into two types. One type of systems is based on homomorphic functions,[14] and the other is based on blind signatures.[15] The e-voting systems based on blind signatures are preferable in practice in comparison with those based on homomorphic functions due to the fact that the system of homomorphic functions is less flexible and it only provides YES or NO function on the ballot. The other system based on blind signatures provides more flexibility on the ballot. In other words, it can allow more formats on the ballot. The general requirements of e-voting are listed in Table 1.

Table 1: Requirements of e-Voting.

Property	*Definition*
Democracy	Only eligible voters can participate in election.
Convenience	Voters can cast their ticket easily and quickly.
Mobility	No restrictions imposed on the location where voters can cast their ballots.
Efficiency	The voters cast their ballot in a reasonable amount of time and he/she is not required to wait for others to complete the process.
Robustness	No one can disrupt or disturb the election because of the independence of the voting processes.
Anonymity	No one can trace the identity of the voter from the ballot.
Authentication	The authorities and voter should verify each other during the process of voting.
Validation	The authorities are able to check whether the ballots are valid or not.
Uniqueness	No voters can vote more than once.
Completeness	An eligible voter is always accepted by the authorities.
Fairness	The authorities are prohibited to cheat, even if they attempt to collude.

In 1998, Mu and Varadharajan proposed an anonymous secure electronic voting scheme,[16] which is based on ElGamal's digital signature algorithm.[17] In essence, the authors claimed that their scheme can ensure the secrecy of voters and prevent the occurrence of double voting. However, Lin, Hwang, and Chang[18] and Chien, Jan, and Tseng[19] pointed out that there are flaws in Mu and Varadharajan's scheme. They outlined problems such as a voter can vote more than once without being detected, the voter's identity can be revealed by the authorities, and the valid ballot can be forged without being authenticated.

To enhance the security of Mu and Varadharajan's scheme, in 2003 Lin, Hwang, and Chang[20] proposed an improved scheme. Later, Chien, Jan, and Tseng[21] also pointed out some limitations existing in Mu and Varadharajan's scheme. In turn, the authors of the current paper have observed that the improved scheme reported by Lin, Hwang, and Chang cannot resist to the attack proposed by Chien, Jan, and Tseng. Therefore, this paper proposes an improved scheme that can overcome the deficiencies existing in Mu and Varadharjan's scheme.

The paper is structured as follows. First, the environment and the scheme and process of general e-voting will be introduced. Then, Mu and Varadharajan's scheme will be briefly reviewed. The section that follows shall present the general act of attacking of an e-voting scheme and the Chien, Jan, and Tseng's scheme. After that, an improved scheme that overcomes the weakness of Mu and Varadharajan's scheme will be proposed, followed by a security analysis. Conclusions and future research directions are given in the last section.

E-Voting Environments and Processes

E-voting Environments

This section introduces the environments for e-voting. In general, there are five parties in anonymous e-voting environments,[22] which are listed as follows:

- *Voter*: Citizens who are qualified to vote.
- *Authentication Server (AS)*: AS is responsible for authenticating the voters and granting voting tickets.
- *Voting Servers (VS)*: It collects voting tickets from voters.
- *Tickets Counting Server (TCS)*: Responsible for tallying the votes.
- *Certificate Authority (CA)*: Provides a certificate service provider for all voters enrolled.

The e-voting activities are:

(1) 〔Voter→CA〕: Before voting, all voters should register at CA as valid voters.

(2) 〔CA→Voter〕: After a voter completes the enrollment, the CA signs and issues a certificate to the voter.

(3) 〔Voter→AS〕: Voter sends a request to AS for a voting ticket.

(4) 〔AS→Voter〕: AS sends a blindly signed ticket to voter.

(5) 〔Voter→VS〕: Voter sends the voting ticket via network to VS.

(6) 〔VS→TCS〕: VS sends voting tickets to TCS once the voting box is full. The counting of tickets is done by TCS.

Figure 1 provides an illustration of the anonymous e-voting environment and its process.

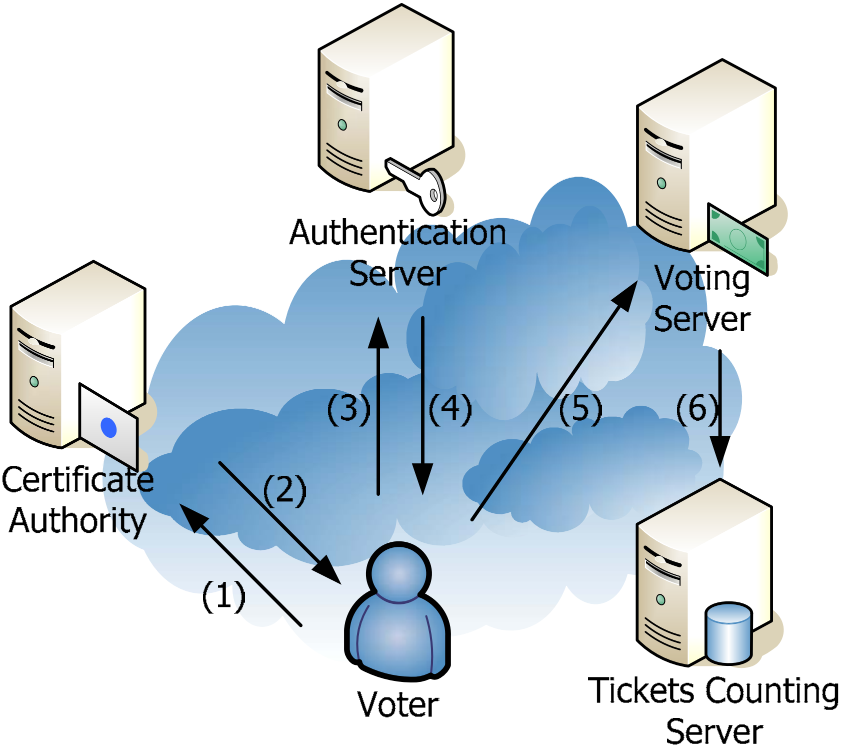


Figure 1: The Anonymous e-Voting Environment and its Process.

e-Voting Processes

In general, the e-voting process involves at least three phases: a registration phase, a voting phase, and a counting phase.

In 1998, Mu and Varadharajan proposed two anonymous secure electronic voting schemes[23] based on ElGamal's digital signature algorithm.[24] Both schemes have five parties involved and three processes as described above. However, the first scheme proposed by Mu and Varadharajan assumes that the Authentication Server (AS) is trustworthy; the second scheme assumes that the trusted AS is unnecessary. The second scheme is more efficient, secure and practical as compared with the first scheme.

However, as Lin, Hwang, and Chang[25] and Chien, Jan, and Tseng[26] showed the second scheme of Mu and Varadharajan is insecure. The following section provides a brief review.

First, we define the notations used is this paper.

- ID_x : The identity of X .
- $Cert_x$: The certificate of X , which includes X 's identity, public key, serial number, valid period, and CA's signature.
- (d_x, e_x) : The RSA secret/public key pair of X .
- p : A large prime number.
- n_x : A product of two large prime numbers.
- g : A generator for Z_p*.
- t : The current timestamp.
- $||$: Concatenation of bits.

Registration Phase

The registration phase includes two procedures: getting a voting certificate from CA and obtaining a voting ticket from AS. The procedures are shown in Figure 2.

1. Getting a Voting Certificate:

A voter generates a key pair first, and then s/he sends it to CA. In order to achieve the e-voting requirements of democracy and completeness, CA has to check the identity of the voter and issue a certificate.

2. Obtaining a Voting Ticket:

Step 1: Before voting, the voter has to send a request to AS for getting a voting ticket. In order to meet the requirement of anonymity, the voter chooses a blind factor b and two secret parameters r and $k_1 \in Z_{p-1}$. Then $a \equiv g^r \bmod p$, $x_1 \equiv gb^{e_{AS}} \bmod n_{AS}$, $x_1' \equiv g^{k_1} b^{e_{AS}} \bmod n_{AS}$, and $x_2 \equiv ab^{e_{AS}} \bmod n_{AS}$ are computed. After that, the voter sends $\{ID_v, ID_{AS}, Cert_V, (x_1 \| x_1' \| x_2 \| t)^{d_v} \bmod n_v\}$ to AS in order to request a voting ticket.

Step 2: When AS receives the message from the voter, the voter's signature has to be verified first; then a random parameter $k_2 \in Z_{p-1}$ is chosen; $x_3 = (k_2 \| t)^{e_v} \bmod n_v$ and $x_4 \equiv (x_1^{k_2} x_1' x_1^{2k_2} x_1' x_2)^{d_{AS}} \bmod n_{AS}$ are computed; and k_2 is stored in a database. Finally, the AS sends the message

$\{ID_{AS}, ID_V, x_3, (x_4 \parallel t)^{e_v} \bmod n_v\}$ to the voter.

Step 3: When the voter receives the message from AS, he uses his secret key to decrypt x_3 and get k_2. Then, the parameters k, k', y_1, and y_2 are computed for the voter, where $k = k_1 + k_2$, $k' = k_1 + 2k_2$, $y_1 = g^k$, and $y_2 = g^{k'}$. Moreover, the blind factor $b^{3(k_2+1)}$ can be removed in order to obtain $x_4' \equiv (y_1 y_2 a)^{d_{AS}} \bmod n_{AS}$, and compute $s_1 = k^{-1}(ma - r) \bmod p\text{-}1$ and $s_2 = k'^{-1}(ma - r) \bmod p\text{-}1$. Finally, the voting ticket is composed as $T \equiv \{a \parallel g \parallel y_1 \parallel y_2 \parallel x_4' \parallel s_1 \parallel s_2 \parallel m\}$.

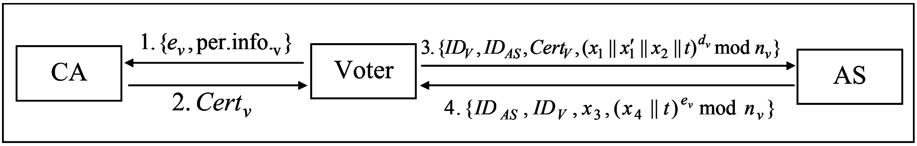

CA
1. $\{e_v, \text{per.info.}_v\}$
2. $Cert_v$
Voter
3. $\{ID_V, ID_{AS}, Cert_V, (x_1 \parallel x_1' \parallel x_2 \parallel t)^{d_v} \bmod n_v\}$
4. $\{ID_{AS}, ID_V, x_3, (x_4 \parallel t)^{e_v} \bmod n_v\}$
AS

Figure 2: Registration Phase of Mu and Varadharajan's Scheme.

Voting Phase

On the voting day, the voter sends a voting ticket $\{ID_{VS}, (T \parallel t)^{e_{VS}} \bmod n_{VS}\}$ to the Voting Server, where *T* is the ticket and *t* denotes the real timestamp. After VS receives the voting ticket, it decrypts $(T \parallel t)^{e_{VS}} \bmod n_{VS}$ using its secret key d_{VS} to obtain *T*. Then VS verifies whether the signature of AS is valid or not. If valid, then VS verifies the voter's signatures s_1 and s_2 by $y_1^{s_1} a = g^{ma} \bmod p$ and $y_2^{s_2} a = g^{ma} \bmod p$. If valid, VS casts the voting ticket into voting boxes.

Counting Phase

After voting is over, TCS executes the procedures of counting tickets and checking whether double voting has occurred by examining whether (a, g, y_1, y_2) are used more than once. In the case of double voting, the AS could compute $k = k_1 + k_2$ and $k' = k_1 + 2k_2$ to get k_2 and trace the voter's identity back.

Attacks

In general, the attacks on the e-voting systems can be divided into three types:

- *Forge ticket*: The malicious attacker can forge a valid ticket without being detected.

- *Trace voter's identity*: The identity of a voter can be traced from the cast voting ticket.
- *Double voting*: An eligible voter can vote more than once.

We have reviewed the Mu and Varadharajan's e-voting scheme in a previous section. Unfortunately, their scheme is vulnerable to attacks as mentioned above. In 2003, Chien, Jan, and Tseng[27] pointed out the weaknesses of Mu and Varadharajan's e-voting scheme as follows.

Attack 1: Forge Ticket

Assume that an attacker chooses three random numbers k_1, k_2, and r (let $k_1, k_2, r < p$), and computes the following equations:

$$\begin{aligned} &a = g^{re_{AS}},\ y_1 = g^{k_1 e_{AS}},\ y_2 = g^{k_2 e_{AS}}, \\ &s \equiv_{n_{AS}} g^{k_1 + k_2 + r} \bmod n_{AS}, \\ &s_1 = (k_1 + k_2)^{-1}(ma - r) \bmod np - 1, \\ &s_2 = (k_1 + 2k_2)^{-1}(ma - r) \bmod p - 1 \end{aligned}$$

Then, the attacker can construct the ticket $T = \{a \parallel g \parallel y_1 \parallel y_2 \parallel s \parallel s_1 \parallel s_2 \parallel m\}$, and VS and TCS cannot detect whether the ticket is forged. This attack can also be found in the work of Lin, Hwang, and Chang.[28]

Attack 2: Trace Voter's Identity

In Mu and Varadharajan's scheme, the parameter k_2 indicates a unique identity for each voter. However, TCS can trace the voter's identity back from the voting ticket even the situation of double voting has not occurred. To get the parameter k_2 TCS computes the following equation:

$$\frac{y_2}{y_1} \equiv \frac{g^{k_1+2k_2}}{g^{k_1+k_2}} \equiv g^{k_2} \bmod p$$

Thus, the Mu and Varadharajan's scheme cannot meet the requirement of anonymity.

Attack 3: Double Voting

The definition of double voting is the situation where a voter or an attacker can vote more than once in the same election process. In Mu and Varadharajan's scheme, the eligible voter can perform the voting process correctly and compose the voting ticket as follows. First, the voter chooses $g = g'^k$, r, and k_1; then the voter sends $\{ID_v, ID_{AS}, Cert_V, (x_1 \| x_1' \| x_2 \| t)^{d_v} \bmod n_v\}$ to AS for obtaining a voting ticket. In this way, the voter can construct the voting ticket $T = \{a \| g \| y_1 \| y_2 \| s \| s_1 \| s_2 \| m\}$, where $a = g^r$, $y_1 = g^{k_1+k_2}$, $y_2 = g^{k_1+2k_2}$, and $s \equiv (y_1 y_2) a^{d_{AS}} \bmod n_{AS}$. Moreover, the voter can construct another valid voting ticket as follows. The voter lets $a' = g^{r+k_1+k_2}$, $y_1' = g^{k_1}$, and $y_2' = g^{2k_2}$; it becomes obvious that $\{a', g', y_1', y_2', s\}$ can still satisfy $s \equiv_{n_{AS}} (y_1' y_2' a')^{d_{AS}}$, and the voter can compose a new ticket using these parameters as $T' = \{a' \| g' \| y_1' \| y_2' \| s \| s_1' \| s_2' \| m\}$. VS and TCS are not able to detect the double voting.

Improvements

In this section, a new improved scheme is presented to enhance the security and prevent the above-mentioned attacks. The proposed scheme consists also of three phases and there are five parties involved, very much the same as in Mu and Varadharajan's scheme. The voting process will be described in what follows.

Registration Phase

The registration phase involves two procedures: requesting a voting certificate from CA and obtaining a valid voting ticket from AS. The voter should provide his identity when requesting a voting ticket from AS. The procedures are shown in Figure 3.

1. Requesting a Voting Certificate:

Each voter generates a key pair (d_v, e_v) and a large number n_v to request a voting certificate from CA. After checking the identity of a voter, CA issues a certificate to the voter to achieve the e-voting requirement of democracy and completeness.

2. Obtaining a Voting Ticket

Step 1: The voter has to send a request to AS for getting a voting ticket. In order to achieve the e-voting requirement of anonymity and authentication, the voter chooses a blind factor b_1 and four random numbers $b_2, q,\ r$ and $k_1 \in Z_{p-1}$. With these parameters, a, x_1, x_2 and x_2 can be computed for the voter using the following equations:

$$\begin{aligned} a &= g^{r} \bmod n, \\ x_1 &= g b_1^{e_{AS}} \bmod n, \\ x_1' &= g^{2k_1+q} b_1^{e_{AS}} \bmod n, \\ x_2 &= a b_2^{e_{AS}} \bmod n \end{aligned}$$

where $g \in Z_p^*$ is the system's public parameter. Finally, the voter sends the message $\{ID_V,\ ID_{AS},\ Cert_v,\ (x_1 \| x_1' \| x_2 \| t)^{d_v} \bmod n_v\}$ to AS for getting parameters to compose the voting ticket.

Step 2: Upon receiving the message, for authentication, AS verifies whether the signature $(x_1 \| x_1' \| x_2 \| t)^{d_v} \bmod n_v$ is valid or not. If valid, AS chooses a random number k_2, that is different for each voter in order to meet the property of uniqueness and to compute the following equations:

$$\begin{aligned} x_3 &= (k_2 \| t)^{e_v} \bmod n_{AS}, \\ x_4 &= (x_1^{k_2}\ x_1'\ x_1^{2k_2}\ x_1'\ x_2)^{d_{AS}} \bmod n_{AS} \\ &= (y_1\ y_2\ a)^{d_{AS}} b_1^{3k_2+2} b_2 \bmod n_{AS} \end{aligned}$$

where $y_1 = g^{k_1+2k_2+2q}$, and $y_2 = g^{3k_1+k_2}$. In order to achieve the e-voting requirement of validation, AS will store k_2 in the database, compute the value of $h(x_2)$ and publish it so it can detect double-voting. Finally, AS sends the message $\{ID_{AS},\ ID_V,\ x_3,\ (x_4 \| t)^{e_v} \bmod n_V\}$ to the voter.

Step 3: When the voter receives the message, x_3 is decrypted to get k_2 and the blind factor $b_1^{3k_2+2}\ b_2$ is removed to obtain the signature (s'); then the following equations are computed:

$$\begin{aligned} s' &= (y_1\ y_2\ a)^{d_{AS}} \bmod n_{AS}, \\ s_1 &= (k_1 + 2k_2 + 2q)^{-1}(ma - r) \bmod p - 1, \\ s_2 &= (3k_1 + k_2)^{-1}(ma - r) \bmod p - 1. \end{aligned}$$

Finally, the voter can compose the voting ticket as follows: $T = \{a \| g \| y_1 \| y_2 \| s' \| s_1 \| s_2 \| m \| b_2\}$.

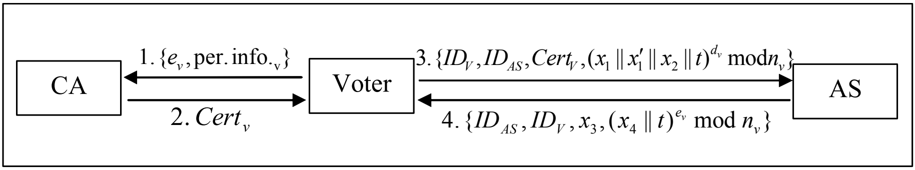

CA
Voter
AS
$1. \{e_v, \text{per. info.}_v\}$
$2. Cert_v$
$3. \{ID_V, ID_{AS}, Cert_V, (x_1 \| x_1' \| x_2 \| t)^{d_v} \bmod n_v\}$
$4. \{ID_{AS}, ID_V, x_3, (x_4 \| t)^{e_v} \bmod n_v\}$

Figure 3: Registration Phase of the Proposed Scheme.

Voting Phase

After obtaining a valid voting ticket from AS, the voter can send the voting ticket. The procedure is shown in Figure 4.

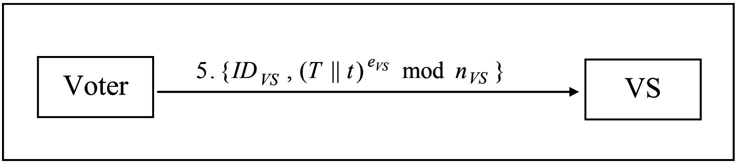


Figure 4: Voting Phase of the Novel Scheme.

Step 1: The voter sends $\{ID_{VS}, (T \| t)^{e_{VS}} \bmod n_{VS}\}$ to VS, where T is the voting ticket and t denotes the current timestamp.

Step 2: VS decrypts $(T \| t)^{e_{AS}} \bmod n_{VS}$ with its secret key (d_{AS}) to achieve the e-voting requirement of authentication. VS verifies the validity of a, y_1, and y_2 using AS's signature from the equation $(s')^{e_{AS}} = (y_1 y_2 a) \bmod n_{AS}$. If the result is true, then VS checks the correctness of s_1 and s_2 using voter's signatures from the following equations:

$$y_1^{s_1} a = g^{ma} \bmod p$$

$$y_2^{s_2} a = g^{ma} \bmod p.$$

If the results are true, VS can be sure that the ticket is valid. Finally, VS stores all the voting tickets and sends them back to TCS via a network.

Counting Phase

After voting is over, TCS checks whether the parameters (a, g, y_1, y_2) have been used more than once and checks $h(ab_2^{e_{AS}}) \bmod n_{AS} = ?\, h(x_2)$ to detect double voting, where $h(x_2)$ is published by AS to be used to confirm the authentication of the ticket. If the results are positive, then TCS will calculate the valid ballot and announce the result obtained from this electronic election.

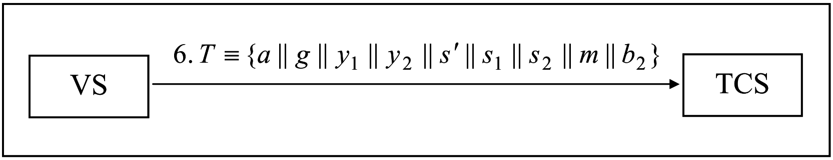

VS
$6. T \equiv \{a \| g \| y_1 \| y_2 \| s' \| s_1 \| s_2 \| m \| b_2\}$
TCS

Figure 5: Counting Phase of the Proposed Scheme.

Security Analysis

In this section, it will be illustrated that the proposed scheme can enhance the security and overcome the limitations of Mu and Varadharajan's scheme.

Prevent Tracing the Voter's Identity

Chien, Jan, and Tseng's work[29] points out that the authorities can obtain a voting ticket from the public network and compute the following equation to get the value of k_2.

$$\frac{y_2}{y_1} \equiv \frac{g^{k_1+2k_2}}{g^{k_1+k_2}} \equiv g^{k_2} \bmod p$$

Thus, the authorities can easily trace the identity of the voting ticket. However, in our improved scheme, no one can trace the identity from the voting ticket. The authorities cannot employ the above equation to obtain k_2 since the parameters y_1 and y_2 contain the secret value k_1 and q chosen by the voter. In consequence, in the proposed scheme, attacking using the trace from the voter's identity as reported by Chien, Jan, and Tseng cannot be successful. In what follows it will be demonstrated why:

$$y_1 = g^{k_1+2k_2+2q},$$
$$y_2 = g^{3k_1+k_2},$$
$$\frac{y_2}{y_1} \equiv \frac{g^{k_1+2k_2+2q}}{g^{3k_1+k_2}} \bmod p$$

It is obvious that:

$$\left(\frac{y_2}{y_1} \bmod p \right)$$

cannot give k_2. Therefore, the authorities cannot find the identity from the voting ticket.

Prevent Forging Ticket and Double Voting

In the improved scheme the authors propose, the public value of $h(x_2)$ is used to detect forged tickets and double voting, where $(x_2 = ab_2{}^{e_{AS}} \bmod n)$ and b_2 are chosen by the voter. In the registration phase, AS has to confirm the identity of the voter first, and then publish $h(x_2)$, where x_2 is received from the voter. However, in the counting phase, in order to detect forged tickets, TCS will check $h(ab_2{}^{e_{AS}}) \bmod n_{AS} = ?\, h(x_2)$. If TCS computes a value of $h(ab_2{}^{e_{AS}}) \bmod n_{AS}$ which is not in the list published by AS, TCS can determine whether the ticket is valid or not. So, if the voter wants to forge another ticket, he has to send another x_2 to AS. Obviously, it is not possible, since the CA will not issue another certificate to the same voter. So, if the voter cannot get another certificate, then AS will not accept the identity of the voter and will not publish the value of $h(x_2)$.

On the other hand, if an attacker attempts to forge $h(x_2)$ in the list published by AS, then he/she will discover that this is not possible due to the property of the hash function, i.e. that there is no way to find the value of x_2 from $h(x_2)$. Besides, it is difficult to find other parameters such as a' and b_2' that can allow through $h(a'b_2{}'^{e_{AS}} \bmod n) = h(x_2)$ passing the validation of signature. As a result, the proposed scheme can resist forged attack and prevent double voting.

Conclusion

In this paper, the authors have introduced an anonymous secure e-voting environment, together with its processes and common attacks. Furthermore, the authors have also proposed an improved scheme that can enhance the security of Mu and Varadharajan's scheme.

Acknowledgement

This research was partially supported by the National Science Council, Taiwan, R.O.C., under contract No. NSC92-2213-E-324-005.

Notes:

[1] David L. Chaum, "Untraceable Electronic Mail, Return Address and Digital Pseudonyms," *Communictions of the ACM* 24, no. 2 (1981): 84-88.

[2] David L. Chaum, "Blind Signatures System," in *Advances in Cryptology*, *CRYPTO'83* (1983), 153– 156.

[3] Josh D. Cohen and Michael J. Fischer, "A Robust and Verifiable Cryptographically Secure Election Scheme," in *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science (FOCS)* (Portland, OR, 21-23 October 1985), (New York, USA: IEEE Computer Society, 1985), 372-382.

[4] Josh C. Benaloh and Dwight Tuinstra, "Receipt-Free Secret-Ballot Elections (extended abstract)," in *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing STOC'94* (Montreal, Quebec, Canada, 23-25 May 1994), (New York, USA: ACM, 1994), 544-553.

[5] Josh C. Benaloh and Moti Yung, "Distributing the Power of a Government to Enhance the Privacy of Voters," in *Proceedings of the 5th Annual ACM Symposium on Principles of Distributed Computing (PODC)* (Calgary, Alberta, Canada, August 1986), (New York, USA: ACM, 1986), 52-62.

[6] Hung-Yu Chien, Jinn-Ke Jan, and Yuh-Min Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes," *Applied Mathematics and Computation* 139, no. 2-3 (July 2003): 525-530.

[7] Atsushi Fujioka, Tatsuaki Okamoto, and Kazuo Ohta, "A Practical Secret Voting Scheme for Large-Scale Elections," in *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques: Advances in Cryptology - AUSCRYPT'92* (Gold Coast, Queensland, Austrailia, 13-16 December 1992), *Lecture Notes in Computer Science* 718, ed. Jennifer Seberry and Yuliang Zheng (Berlin: Springer-Verlag, 1993), 244-251.

[8] Martin Hirt and Kazue Sako, "Efficient Receipt-Free Voting based on Homomorphic Encryption," in *Advances in Cryptology, EUROCRYPT'00*, *Lecture Notes in Computer Science* 1807, ed. B. Preneel (Berlin: Springer-Verlag, 2000), 539-556.

[9] Wen-Shenq Juang and Chin-Laung Lei, "A Secure and Practical Electronic Voting Scheme for Real World Environments," *IEICE Transactions on Fundamentals* E80-A, no. 1 (January 1997): 64-71.

[10] Horng-Twu Liaw, "A Secure Electronic Voting Protocol for General Elections," *Computers and Security* 23, no. 2 (March 2004), 107-119.

[11] Iuon-Chang Lin, Min-Shiang Hwang, and Chin-Chen Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network," *Computer Standards & Interfaces* 25, no. 2 (May 2003): 131-139.

[12] Yi Mu and Vijay Varadharajan, "Anonymous Secure e-Voting over a Network," in *Proceedings of the 14th Annual Computer Security Applications Conference (ACSAC'98)* (Scottsdale, AZ, USA, 7-11 December 1998), (IEEE Computer Society, 1998), 293–299.

[13] Kazue Sako and Joe Killian, "Secure Voting Using Partially Compatible Homomorphisms," in *Advances in Cryptology--CRYPTO'94*, *Lecture Notes in Computer Science* 839, ed. Yvo G. Desmedt (Berlin: Springer-Verlag, 1994), 411-424.

[14] Benaloh and Fischer, "A Robust and Verifiable Cryptographically Secure Election Scheme;" Benaloh and Tuinstra, "Receipt-Free Secret-Ballot Elections;" Benaloh and Yung,

"Distributing the power of a government to enhance the privacy of voters;" Hirt and Sako, "Efficient Receipt-Free Voting based in Homomorphic Encryption;" Sako and Killian, "Secure Voting Using Partially Compatible Homomorphisms."

[15] Fujioka, Okamoto, and Ohta, "A Practical Secret Voting Scheme for Large-Scale Elections;" Juang and Lei, "A Secure and Practical Electronic Voting Scheme for Real World Environments;" Liaw, "A Secure Electronic Voting Protocol for General Elections;" Lin, Hwang, and Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network;" Mu and Varadharajan, "Anonymous Secure e-Voting over a Network."

[16] Mu and Varadharajan, "Anonymous Secure e-Voting over a Network."

[17] ElGamal, "A Public-Key Cryptosystem and a Signature Scheme based on Discrete Logarithms."

[18] Lin, Hwang, and Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network."

[19] Chien, Jan, and Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes."

[20] Lin, Hwang, and Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network."

[21] Chien, Jan, and Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes."

[22] Mu and Varadharajan, "Anonymous Secure e-Voting over a Network."

[23] Mu and Varadharajan, "Anonymous Secure e-Voting over a Network."

[24] ElGamal, "A Public-Key Cryptosystem and a Signature Scheme based on Discrete Logarithms."

[25] Lin, Hwang, and Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network."

[26] Chien, Jan, and Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes."

[27] Chien, Jan, and Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes."

[28] Lin, Hwang, and Chang, "Security Enhancement for Anonymous Secure e-Voting over a Network."

[29] Chien, Jan, and Tseng, "Cryptanalysis on Mu-Varadharajan's e-Voting Schemes."

CHOU-CHEN YANG received his B.S. in Industrial Education from the National Kaohsiung Normal University in 1980. He received his M.S. in Electronic Technology from the Pittsburg State University in 1986, and his Ph.D. in Computer Science from the University of North Texas in 1994. He is an associate professor at the Department of Management Information Systems at National Chung Hsing University. His current research interests include network security, mobile computing, and distributed system. *Address for correspondence:* Department of Management Information System, National Chung Hsing University, 250 Kuo Kuang Road, 402 Taichung, Taiwan, R.O.C.; *E-mail:* cc.yang@nchu.edu.tw.

CHING-YING LIN received her B.S. in Information Management from Chaoyang University of Technology in 2004. She is pursuing her M.S. in Networking and Communication Engineering from Chaoyang University of Technology. Her current research interests include information security and mobile communications. *Address for correspondence:* Graduate Institute of Networking and Communication Engineeing, Chaoyang University of Technology, 168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan 413, R.O.C.; *E-mail:* s9330601@mail.cyut.edu.tw.

HUNG-WEN YANG received his B.S. in Information Management from Taichung Healthcare and Management University in 2003. He is pursuing his M.S. in Information Management from Chaoyang University of Technology. His current research interests include information security and mobile communications. *Address for correspondence:* Department of Information Management, Chaoyang University of Technology, 168 Gifeng E. Rd., Wufeng, Taichung County, Taiwan 413, R.O.C.; *E-mail:* s9214605@mail.cyut.edu.tw.

A LICENSE PLATE RECOGNITION SYSTEM IN E-GOVERNMENT

Hsien-Chu WU, Chwei-Shyong TSAI, and Ching-Hao LAI

Abstract: In recent years, there has been a lot of research in license plate recognition, and many license plate recognition systems have been proposed and used. In the past, the license plate recognition system was a key element in many applications such as traffic control and parking lots access control. In this article, the authors present current research in license plate recognition, and elaborate on the use of license plate recognition in e-Government. License plate recognition involves three main components: plate detection, character segmentation and character recognition. The authors consider all three components and thoroughly deal with the issue how to apply license plate recognition in e-Government in order to improve performance.

Keywords: e-Government, License Plate Recognition, Plate Detection, Character Segmentation, Traffic Control.

Nowadays, license plate recognition is extensively used in traffic management to identify a car whose owner has violated traffic laws or to find stolen vehicles. It is also applied to parking lot access control. Or, in other words, parking lots do not need human resources. When a car enters a parking lot, a computer equipped with a sensor and a license plate recognition system can recognize the license plate number of the car, and record the car data and the entry time. When the car leaves the parking place, the computer can automatically compute the parking cost. The license plate recognition system is automated and convenient. In addition, it is cost efficient due to the fact that less human resources are needed.

The license plate recognition system could also be used in e-Government. E-Government can apply license plate recognition to car identification. In recent years, a lot of cameras have been installed at the intersections by government or police. These cameras record a lot of images. If license plate recognition systems are installed at police offices, networks can be used to connect the cameras and the police offices. The cameras capture images regularly. When a camera captures an image, it can be transferred to the police office and the license plate number can be identified by the license plate recognition system. This license plate number can be used to acquire any information about the car and record the license plate number, time, place and so on. This information can aid the police in seizing car thieves and in traffic control. In addition, there have been numerous parking lots in the city areas. Usually, more than one person has to be employed at a given parking lot to record license plate number and to compute parking costs. If license plate recognition systems are installed at the parking lots, the license plate numbers of all cars parked in parking lots can be identified and the parking costs computed automatically. If license plate recognition systems are installed at government parking lots, the government can employ fewer employees and save more costs on employment.

License plate recognition systems provide to e-Government convenience and efficiency. In this article, recent research on license plate recognition is presented and some directions how to apply the methods of license plate recognition and to enable license plate recognition systems to work efficiently in an e-Government environment are given.

Overview

License plate recognition consists of three main phases: license plate detection, character segmentation and character recognition. The role of license plate detection is to extract the license plate regions from a vehicle image. Sometimes, there may be more than one license plates in an image. Some research in the field of license plate detection deals with the issue how to extract multiple license plates from an image. It is a very important subject in license plate recognition. After extracting all license plates from an image, all characters in a license plate have to be segmented at the character segmentation stage of license plate recognition. If the obtained character segment is good, character recognition will be more accurate. The final stage completed by any license plate recognition system is character recognition. At this stage, all characters have to be recognized. This stage is the most important one in license plate recognition. It has a substantial effect on success. In this section, the authors survey recent research on license plate recognition systems.

License Plate Detection

There has been a lot of research in license plate detection. There exist two principal types of image sources in plate detection. One is a video image and the other is a static image. If the image source is video, one has to extract the single license plate in real time. Due to the fact that it is necessary to extract the license plate in a very short time, it is extremely important that we extract the license plate quickly. If the source image is static, extracting multiple license plates from an image with a complex background is the main issue.

In the method proposed by Kim and Chien, Generalized Symmetry Transform (GST) and edge detection are used for license plate detection and normalization.[1] Park and co-workers use a neural network for license plate detection.[2] These methods are applied to images with only one car and without complex background.

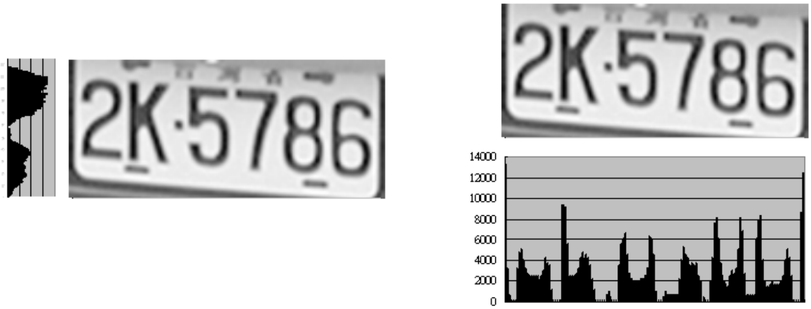
There are many works that deal with images with a complex background. It is very difficult to extract many license plates from an image with a complex background. Gao and Zhou proposed a method to extract the license plates from an image with a complex background.[3] They use histogram equalization to find a threshold to enhance the license plate image. After image enhancement, the authors compute the variance of the pixels in every large region. There would be some noise in these regions, and there may not be license plate regions. Gao and Zhou then smooth the edges by image dilation. Finally, they check the number of pixels, the proportion of width and height, and the proportion of black pixels in the region. Their method then eliminates the regions that are not license plate regions and outputs the images of license plates. Gao and Zhou assume that the license plates are near the center of the images. When the license plate is in the margins of the image, the license plate would not be extracted easily. The accuracy of Gao and Zhou's method is 80.7%.

Maro, Chacon and Zimmernam use Pulse Coupled Neural Network (PCNN) to find the candidate regions from a static image that consist of only one car.[4] Then, they use statistics and edge detection to find the regions of license plates and output them. There are three features in this technique. First, the images are static. Second, only one license plate from an image is extracted. Third, it is difficult to extract license plate from an image with a complex background.

Character Segmentation

In license plate recognition, when all license plate images are extracted from an image, one has to find all characters in the license plate and recognize them. Character segmentation is applied to the license plate in order to outline the individual characters. It affects greatly the accuracy of recognition. If the contours of the characters are inaccurate, it might lead to errors in the recognition stage or even to failed recognition stage.

Majority of current research deals with the segmentation of only one type of license plates. The methods proposed use prior knowledge to help in character segmentation. Prior knowledge includes information such as the size of the license plate, the size of a character, the size of the interval between the characters, and the number of all characters within a license plate. If there is only one type of license plates, it is useful



(a) (b)

Figure 1: (a) Vertical Projection; (b) Horizontal Projection.

to make use of prior knowledge in character segmentation to achieve better results.

The character segmentation methods proposed in the majority of works are quite straightforward. They are typically based on the projection method.[5] The projection method uses vertical and horizontal projections to perform vertical and horizontal segmentation, respectively. Horizontal segmentation eliminates the unnecessary top and bottom parts of the license plate image. And vertical segmentation divides each character that is in the same license plate image. The projection method analyzes the vertical and horizontal projections (shown in Figure 1). How to locate all the valleys is a very important part of the whole process due to the fact that the valleys usually occur in the intervals between the characters within a license plate image. The process of locating valleys in the projection is called “projection analysis.” Segmentation lines are created at these valleys. The projection method is the simplest, the most traditional, and most commonly used method for character segmentation. Rahman, Badawy and Radmanesh use only the projection method to segment all characters.[6] However, if the projection method is used alone, the problems of rivet, rotation, and illumination variance can be observed. When these problems appear, there will be many errors in character segmentation. Zhang, Yang and Wang use also projection method to segment characters.[7] In addition, they use prior knowledge and Hough transform to help in segmenting the characters. Prior knowledge is the information about license plate in China, and Hough transform can help in deleting wrong horizontal segmentation lines and combining the correct lines. Before the segmentation, the authors divide the license plate image into several parts and perform horizontal and vertical segmentations, respectively, for each part. Zhang-Yang-Wang’s method can solve the problems of rivet, rotation, and illumination variance. Their method could be used in the license plate recognition systems in e-Government.

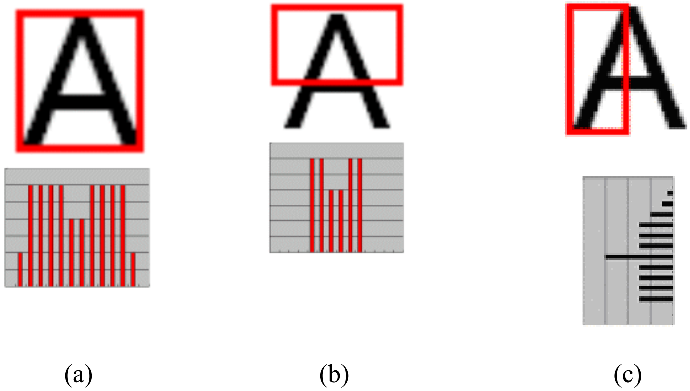


Figure 2: (a) Horizontal and Full-Size Histogram; (b) Upper-Half Histogram; (c) Left-Half Histogram.

Character Recognition

The final and the most important stage of license plate recognition is character recognition. At this stage, the character images that are extracted from the license plate image have to be recognized. The research on character recognition distinguishes between several main methods: methods using character features, methods based on neural networks, methods using pattern mapping, etc. Prior to formally applying a neural network, one has to generate many samples of characters to train the network. It is essential to know how to form the inputs to the neural network and how to train it. Pattern mapping is a traditional method. Before recognizing characters, the patterns have to be created. There are several types of patterns: images of each character, of different sizes or at different angles, projections of each character, and others.

The method proposed by Rahman and colleagues is based on pattern mapping.[8] There are fifteen different types of projection histograms for each character in a pattern database before applying character recognition. These fifteen histograms are created using different parts of the characters[9]: (1) full-size (horizontal); (2) lower-half (horizontal); (3) upper-half (horizontal); (4) lower-one-third (horizontal); (5) upper-one-third (horizontal); (6) lower-one-forth (horizontal); (7) upper-one-forth (horizontal); (8) upper-two-third (horizontal); (9) full-size (vertical); (10) left-half (vertical); (11) right-half (vertical); (12) left-one-third (vertical); (13) right-one-third (vertical); (14) left-one-fourth (vertical); (15) right-one-fourth (vertical). Figure 2 illustrates horizontal and full-size histogram (Figure 2(a)), upper-half histogram (Figure 2(b)) and left-half histogram (Figure 2(a)). Histogram mapping is performed until a single character is recognized.

Koval and co-workers use a neural network with the objective to take advantage of the network capability to solve the problems with noise and different positions of characters on a license plate.[10] They use images of characters to train the neural network. The structure of a general neural network is shown in Figure 3. The neural network is multi-layered and has 663 elements. The experimental results demonstrate accuracy of 95% for plate images having noise with 50% density.

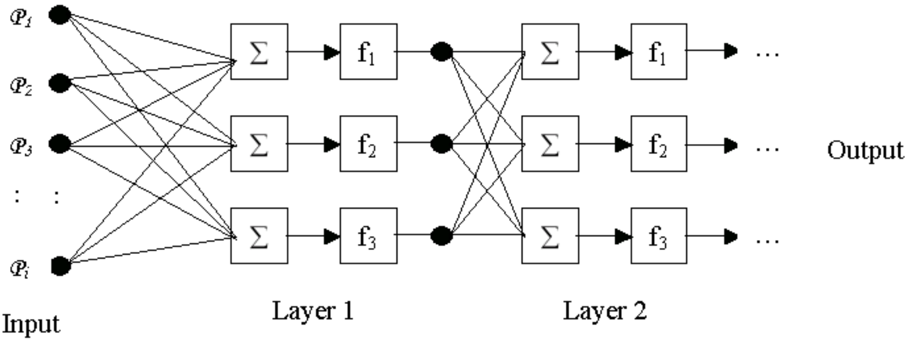
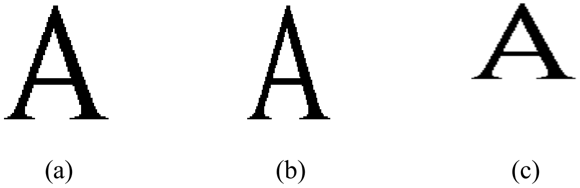


Figure 3: General Structure of a Neural Network.

Dai, Ma, Liu and Li divide character recognition into three stages: preprocessing, feature extraction, and classification.[11] There are three steps in preprocessing: noise removal, character normalization and changing all pixel values to binary values. These steps can help in improving the quality of character images and they are beneficial for feature extraction. There are three features to be extracted including projection feature (CPF), perimeter distance feature (PDF) and direction contribution density (DCD). The extracted features are used then to classify the characters. All techniques in the Dai-Ma-Liu-Li's method are very simple, and the performance is satisfactory. The accuracy is above 97% in daytime and 90% at night. The recognition time is between 0.1 and 1.5 seconds.

Wang and Lee extract two features before character recognition to help in recognizing the characters.[12] These features are "crossing-count features (CCFs)" and "peripheral background area features (PBAFs)." In the preceding phase, the input character image has to be normalized in a coordinate system. Please note that the images of the same character may be of different type. Three different types of the character "A" are given in Figure 4. In principle, CCFs and PBAFs are very useful in character recognition. The accuracy of Wang and Lee's method is 98.6%, and it is very accurate.



(a) (b) (c)

Figure 4: Different Types of Character “A”.

Discussion

Many methods for license plate recognition have been proposed. However, different methods satisfy different requirements and have different applications. In this section, the authors will elaborate on the issue which methods are suitable for achieving high performance in e-Government.

As mentioned above, the license plate recognition system can be used in e-Government for traffic control, parking lots access control and so forth. However, the license plate recognition system has to work in real time. In order to develop an on-line, high speed license plate recognition system in e-Government, cameras have to be installed on roads or at the entrances of parking lots outdoors, as well as license plate recognition system has to be installed at control centers. Networks can be used to connect each camera and control center. When a camera captures a car image, the image can be transferred to the control center. The license plate recognition system installed at the control center can recognize the image and output the license plate number. It can record the license plate number, time, location, etc. These records are very useful. At the police offices, they can be used to obtain the traffic flow. The police can also access information about a stolen car and arrest the car thieves quickly. At parking lots, these records can be used to compute the costs for parking.

License Plate Detection

In this section, the authors propose a method for license plate detection in e-Government. In e-Government, license plate recognition has to be fast and very accurate. License plate detection is used for finding and extracting license plate regions in a car image. In general, the pixel values in the license plate region are very different due to the fact that the characters and the background of the license plate have a big difference in pixel values. We can find the region that has the biggest difference in pixel values, or the license plate region in the image. This method is very common. To aid

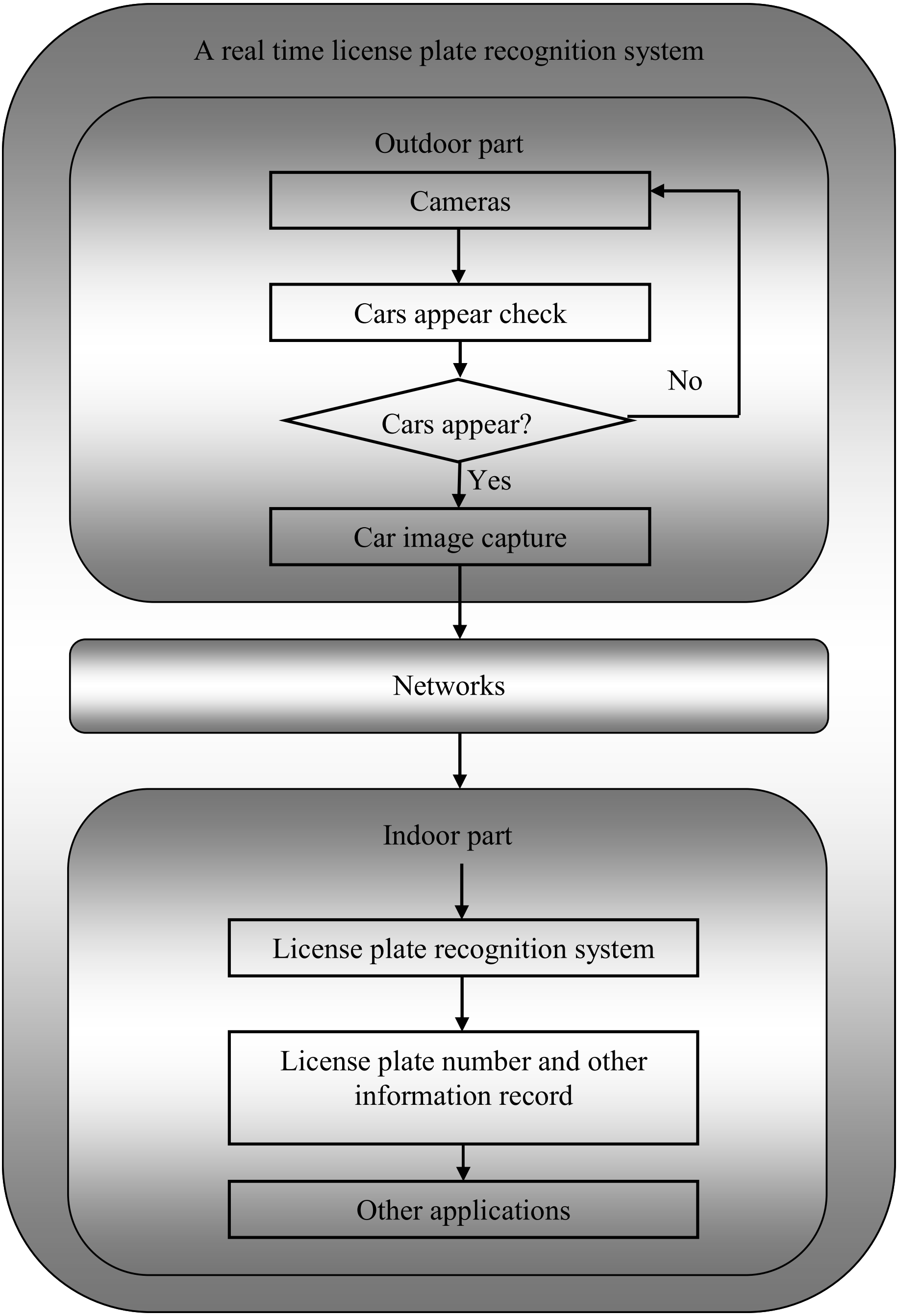
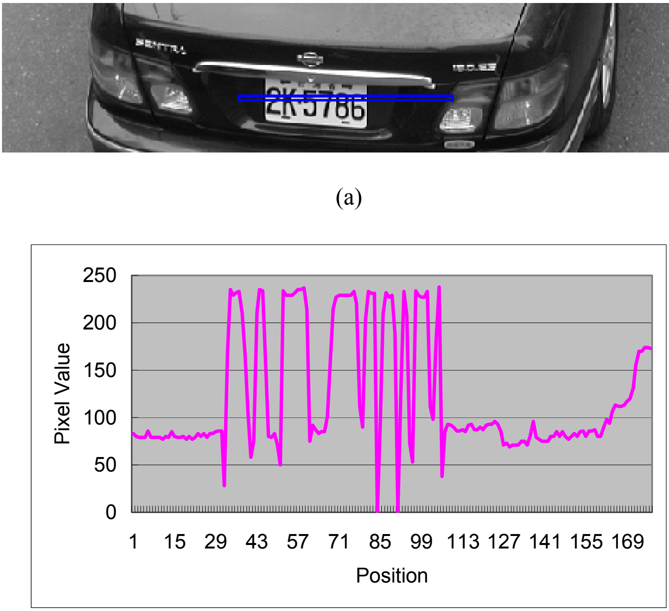

A real time license plate recognition system
Outdoor part
Cameras
Cars appear check
No
Cars appear?
Yes
Car image capture
Networks
Indoor part
License plate recognition system
License plate number and other information record
Other applications

Figure 5: Structure of the Proposed License Plate Recognition System.



(a)

(b)

Figure 6: (a) The License Plate; (b) Horizontal Histogram.

license plate recognition in e-Government, the authors propose the following modification: the department issuing license plates could put a special logo on the license plate. When the license plate recognition system extracts license plates, the system can locate the special logo first. The method of finding license plates will be simpler and faster.

Character Segmentation

In general, the research on character segmentation often uses some prior information, such as the size of characters, the intervals between characters, and the size of the license plates. Using this prior information, the characters in the license plate can be segmented quickly and precisely.

Considering the fact that the characteristics of the license plates are useful in segmenting the characters, the standard for license plates have to be unified. If all license plates have a common standard, the license plate recognition system can segment characters derived from different types of license plates.

As elaborated above, a simple and fast method for character segmentation is the projection method. Using horizontal and vertical projection histograms, one can locate segmentation points. These points may be incorrect, and one can use the characteristics of the license plates to eliminate such points. Horizontal and vertical projection histograms are shown in Figure 1. We recommend this method, because it is rather simple and fast.

Character Recognition

Pattern mapping is a common and simple method for character recognition. In a nutshell, many patterns for each character are created in the database. Once a character image is extracted and segmented, it is compared with all patterns in the database. There are several types of patterns, including images of each character, projection histograms[13] of each character, etc. Using projection histograms to perform character recognition is quicker. The authors recommend using projection histograms to produce patterns in the database.

In applying pattern mapping methods, the size and the font of the characters are very important. In order to get better recognition results, the patterns have to be created using the same font of the characters in the license plates. The size of characters is also very important. Before proceeding with the mapping work, all character images have to be normalized, and the size has to be the same as that of the pattern.

Some letters look similar to numbers, for example: B and 8; I and 1; O and 0; S and 5; etc. If letters and numbers have designated positions on the license plate—for example, there are 7 characters on a license plate, and the first three characters contain only letters and the remainder contains numbers—letters and numbers will not be mixed. Therefore, the authors recommend that letter and number positions be divided on a license plate.

Conclusions

License plate recognition systems can help government in traffic control, in seizing car thieves and in managing no-man parking lots automatically. Therefore, in order to develop an efficient e-Government, the license plate recognition system will be very useful.

In this article, the authors have presented the state-of-the-art research on license plate recognition. There has been a great number of methods proposed for license plate detection, character segmentation and character recognition. In order to be applied in e-Government, the license plate recognition system has to be a quicker and more accurate real-time system. It is recommended that the methods of license plate recognition be simple.

There is no doubt that the license plate recognition system will be used by the government more and more in the future and, therefore, it is very important to design improved and more efficient systems for application in an e-Government environment.

Notes:

[1] Dong-Su Kim and Sung-I. Chien, "Automatic Car License Plate Extraction Using Modified Generalized Symmetry Transform and Image Warping" (paper presented at the IEEE International Symposium on Industrial Electronics ISIE 2001, Pusan, Korea, 12-16 June 2001), Volume 3, 2022-2027.

[2] Sung Han Park, Kwang In Kim, Keechul Jung, and Hyung Jin Kim, "Locating Car License Plates Using Neural Networks," *IEE Electronics Letters* 35, no. 17 (August 1999): 1475-1477.

[3] Da-Shan Gao and Jie Zhou, "Car License Plates Detection from Complex Scene" (paper presented at the International Conference on Signal Processing, Beijing, 21-25 August 2000), 1409-1414.

[4] I. Maro, M. Chacon, and S. Alejandro Zimmerman, "License Plate Location based on a Dynamic PCNN Scheme" (paper presented at the IEEE International Symposium on Computational Intelligence in Robotics and Automation, 16-20 July 2003), 972- 976.

[5] Yungang Zhang and Changshui Zhang, "A New Algorithm for Character Segmentation of License Plate" (paper presented at the IEEE Intelligent Vehicles Symposium, Beijing, 9-11 June 2003), 106-109.

[6] Choudhury A. Rahman, Wael M. Badawy, and Ahmad Radmanesh, "A Real Time Vehicle's License Plate Recognition System" (paper presented at the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03), Miami, Florida, 21-22 July 2003), (IEEE Computer Society, 2003), 163-166.

[7] Zhang, Yang, and Wang, "A New Algorithm for Character Segmentation of License Plate."

[8] Rahman, Badawy, and Radmanesh, "A Real Time Vehicle's License Plate Recognition System."

[9] Rahman, Badawy, and Radmanesh, "A Real Time Vehicle's License Plate Recognition System."

[10] Valeriy N. Koval, Volodymyr Turchenko, V. Kochan, Anatoly Sachenko, and George Markowsky, "Smart License Plate Recognition System based on Image Processing Using Neural Network" (paper presented at the Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing System: Technology and Applications, 8-10 September 2003), 123-127.

[11] Yan Dai, Hongqing Ma, Jilin Liu, and Langang Li, “A High Performance License Plate Recognition System based on the Web Technique” (paper presented at the IEEE Intelligent Transportation Systems Conference, Oakland, CA, USA, 25-29 August 2001), 325-329.

[12] Shen-Zheng Wang and Hsi-Jian Lee, “Detection and Recognition of License Plate Characters with Different Appearances” (paper presented at the Sixth IEEE Intelligent Transportation Systems, Shanghai, China, 12-15 October 2003), 979-984.

[13] Rahman, Badawy, and Radmanesh, “A Real Time Vehicle’s License Plate Recognition System.”

HSIEN-CHU WU was born in Tainan, Taiwan, Republic of China, on 26 October 1962. She received B.S. and M.S. degrees in Applied Mathematics in 1985 and 1987, respectively, from the National Chung Hsing University, Taichung, Taiwan. She received her Ph.D. in Computer Science and Information Engineering in 2002 from the National Chung Cheng University, Chiayi, Taiwan. From 1987 to 2002, she was a lecturer of the Department of Information Management at National Taichung Institute Technology, Taichung, Taiwan. Since August 2002, she has worked as an associate professor at the Department of Information Management of National Taichung Institute Technology, Taichung, Taiwan. Her research interests include image authentication, digital watermarking and image processing. *Address for correspondence:* Department of Information Management, National Taichung Institute of Technology, 129 Sec. 3, San-Min Road, Taichung, Taiwan 404, R.O.C. *FAX:* 886-4-22196610; *Email:* wuhc@ntit.edu.tw.

CHWEI-SHYONG TSAI was born in Changhua, Taiwan, Republic of China, on 3 September 1962. He received a B.S. degree in Applied Mathematics from National Chung Hsing University, Taichung, Taiwan, in 1984. He also received a M.S. degree in Computer Science and Electronic Engineer from National Center University, Chungli, Taiwan, in 1986. He received his Ph.D. degree in Computer Science and Information Engineering in 2002 from National Chung Cheng University, Chiayi, Taiwan. From August 2002 until 2004, he was an associate professor at the Department of Information Management at National Taichung Institute of Technology, Taichung, Taiwan. Since August 2004, he has been an associate professor at the Department of Management Information System at National Chung Hsing University, Taichung, Taiwan. His research interests include image authentication, information hiding, and cryptography. *Address for correspondence:* Department of Management Information System, National Chung Hsing University, 250, Kuo-Kuang Road, Taichung, Taiwan 402, R.O.C. *Email:* tsaics@ntit.edu.tw.

CHING-HAO LAI was born in Taichung, Taiwan, R.O.C., in 1982. He received a B.B.A. degree from the Department of Information Management of National Taichung Institute of Technology, Taichung, Taiwan, R.O.C., in 2004. He is currently working towards a M.S. degree at the Institute of Computer Science and Information Technology of National Taichung Institute of Technology, Taichung, Taiwan, R.O.C. His research interests include image processing and pattern recognition. *E-mail:* s18933111@ntit.edu.tw.

A SECURE ONLINE MEDICAL INFORMATION SYSTEM IN DISTRIBUTED AND HETEROGENEOUS COMPUTING ENVIRONMENT

Muhammad Nabeel TAHIR

Abstract: The objective of this article is to analyze the importance and the role of Information Security in online medical information systems. Healthcare organizations have to protect private information pertaining to the individuals they serve. As more and more healthcare organizations implement computer-based ERPs, telemedicine, EDI, data warehouses and other network-based information systems, information security in healthcare gains importance more than ever before. Possible questions and issues related to information security requirements might be: "How do the healthcare professionals protect the information in their EPR systems?", "How can network data exchange and transfer over the Internet be accomplished without being tampered by hackers and other unauthorized individuals or groups?" To answer these questions, players in the healthcare chain (providers, physicians) are turning to computerized solutions. As one might recall, turning to computerized medical record systems was the solution for healthcare organizations some years ago. Now we are dealing with the problems that those computerized systems may bring.

Keywords: Medical Information Systems, Information Security, Distributed Computing.

Background

Information technology has become increasingly important in improving the quality and in lowering the costs in healthcare. Attempts to protect patient's privacy have to centre, therefore, on finding ways to protect the sensitive electronic healthcare information in a computerized environment rather than on opposing to the use of information technology in healthcare organizations.

As time progresses, ensuring information security in medical institutions becomes a burden for the Chief Information Officers (CIO) and other healthcare professionals. The technology evolves rapidly and keeping up with the technology is a tough task. For this and other not so important reasons priorities seem to be fluctuating. Healthcare information security professionals face a variety of issues at the forefront of information systems planning; however, it is a must for organizational success.

Problem Definition: Research Problems and Research Questions

Recent research has identified an increase in the awareness of the need of location-based services to support the application of secure E-cure.

Implementing security plans and technologies to protect electronic medical records systems is the paramount health data security issue today. This implication is backed up by the report of the National Research Council in 1997 entitled "Protecting Electronic Health Information." The conclusion of the report consists of seven items, which can be summarized as follows: healthcare organizations need to take a more aggressive approach to improving the security of health information systems to better protect electronic health information.

Healthcare organizations have been slow in adopting strong security practices, largely due to a lack of strong management and organizational incentives. No major breach of security has occurred that has catalyzed such efforts. Thus, the information technology vendor community has not found a market for providing security features in health information systems.

Patients have important roles to play in addressing privacy and security concerns. The greatest concerns regarding the privacy of health information derive from widespread sharing of patient information throughout healthcare industry and the inadequate federal and state regulatory framework for systematic protection of health information.

At the level of individual organizations, electronic health information is vulnerable to both authorized users who misuse their privileges to perform unauthorized actions (such as browsing through patient records) and outsiders who are not authorized to use the information systems, but break in with the intent of malicious and damaging action.

Adequate protection of healthcare information depends on both technology and organizational practices for privacy and security.

For IT professionals, the first real issue regarding integrity and security of electronic information came when the shift from mainframe to the client-server computing took place. In the mainframe model, no information was "leaking" to the outside world, the systems were of the single closed systems type. With the emergence of client-server architectures, distributed computing, heterogeneous computing environment and mostly with that of networking, information integrity and security became a serious concern as their use is not limited to desktop PCs but also to wireless technologies like mobiles, PDA's, laptops, etc.

In distributed computing systems (e.g. client-server systems or today's multi-tier systems) data is stored on many computers across an enterprise and even outside of the enterprise. The results of data being stored in and transmitted to numerous places (e.g. laptop, PDA, mobile phones) are massive and information system professionals recognized that data security would be an issue that had to be addressed with new solutions and new technologies.

When in the late 1980s the PC revolution took place and inevitably networking came on the scene, today's data security concerns started to appear. Before that, many healthcare information systems vendors and pioneering healthcare providers were installing IT products with the motto "Make it work first, then think about the security." But with the changes in the computer world and the health data security issues, securing systems is becoming a priority. Contrary to popular belief, security does not simply involve protecting the confidentiality of information. There are three basic elements of data security, all of which should be considered. These are *confidentiality*, *integrity* and *availability*.

Data confidentiality is most commonly associated with security and is easy to understand. Healthcare organizations must protect all confidential data so that it does not get disclosed either accidentally or maliciously. There have been several instances when health information about a patient has been leaked. Such disclosure, whether of a celebrity's health data or a private citizen's health data, can ruin a person's career, insurability or even his life. Thus, using technology and policies to protect the confidentiality of electronic health care information is a must.

Data integrity is not always associated with security, particularly in the eyes of the general public. Protecting data integrity means ensuring stored information is correct and not in any way corrupted. In the late I980s, a hacker group in Milwaukee known as The 414 Gang broke into several organizations, including Memorial Sloan-Kettering Cancer Centre in New York. They got into the provider's computer systems and were fooling around, doing nothing really malicious though. But this example clearly shows how data integrity is a crucial part of data security efforts. And in healthcare, if you corrupt patient records, that could cause a lot of problems, perhaps even the death of a patient.

The third and last major aspect of data security is data and system *availability*. Computer systems and electronic data must be available to users whenever they need it.

One of the factors that have slowed progress toward greater health data security is a lack of understanding of all three primary aspects of the security issue. There long has been an understanding that the only thing security precautions are supposed to do is to protect data confidentiality. While that is the truth, it is not the whole truth.

The idea of, “Make it work first, then secure it,” is a step towards organizational suicide. Experts agree on the point that it is a huge mistake to install security after the fact. Some observers say that healthcare professionals are beginning to think about security first. These experts say that security is starting to be viewed as an enabling technology rather than as an inhibiting addition.

People have never really thought about these other aspects of security because they want to believe in the common good. They think there are not companies out there tracking your every movement and sending you catalogue for components when they have found out that you have broken your leg. People do not want to believe this incredible networked environment exists.

The confidentiality aspect of security has been around since Hippocrates. But when it comes to aspects like integrity, a member of the public would not know the appropriate way to cross something out on a medical record. The people on the streets would not even think about things like that; the only thing they think about is the confidentiality of their medical information. They presume its integrity, which can be a fatal mistake.

In short, success of an individual, business, government agency and health organization increasingly depends on the ability to securely communicate around the world in real time. The advent of widespread connectivity via the Internet and an array of ubiquitous and powerful mobile devices have changed the face of computing and communications. With the vast benefits of increased connectivity, however, a multitude of new risks has emerged, risks on a scale which few in the industry have anticipated.

Conclusions

Many security professionals have worked hard to secure the patients’ medical records but a lot of work is required to be done in this field to achieve a secure medical information system due to the use of wireless heterogeneous technology. The objective of the future research work in this area should be to identify the major security weaknesses in the currently available hospital/clinic management information systems built for both online and offline client-server distributed and heterogeneous computing environments. The future is for the secure information technology and the same applies to health information as well. A distributed, wireless and heterogeneous client-server medical information system is a necessity nowadays. Security is a major factor when we deal with client-server environments or distributed and heterogeneous computing and addressing these problems is a big issue.

References:

1. Ross J. Anderson, "A New Family of Authentication Protocols," <http://www.secinf.net/authentication_and_encryption/A_New_Family_of_Authentication_Protocols.html> (10 November 2004).
2. *A History of Hacking*, <http://www.sptimes.com/Hackers/history.hacking.html> (8 November 2004).
3. Haio Roeckle, Gerhard Schimpf, and Rupert Weidinger, "Process-Oriented Approach for Role-Finding to Implement Role-Based Security Administration in a Large Industrial Organization," in *Proceedings of the 5th ACM Workshop on Role-Based Access Control (RBAC'00)* (Berlin, Germany, 2000), 103-110.
4. Amit P. Sheth and James A. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computing Surveys* 22, no. 3 (September 1990): 183-236.
5. Konstantin Beznosov, "Requirements for Access Control: US Healthcare Domain," in *Proceedings of the 3rd ACM Workshop on Role-Based Access Control* (Fairfax, Virginia, USA, 22-23October 1998), 43.
6. D.J. Thomsen, Richard C. O'Brien, and C. Payne, "Napoleon: Network Application Policy Environment," in *Proceedings of the 4th ACM Workshop on Role-Based Access Control* (Fairfax, Virginia, USA, 28-29 October1999), 145-152.
7. Salem Benferhat, Rania El Baida, and Frédéric Cuppens, "A Stratification-Based Approach for Handling Conflicts in Access Control," in *Proceedings of the Eighth ACM symposium on Access Control Models and Technologies*, ACM 1-58113-681-1/03/0006 (Como, Italy, 2-3 June 2003), 189-195.
8. George W. Dinolt, Lee A. Benzinger, and Mark G. Yatabe, "Combining Components and Policies," in *Proceedings of the Computer Security Foundations Workshop VII*, ed. Joshua Guttman (Los Alamitos, CA: IEEE Computer Society Press, IEEE Computer Society, June 1994), 22-33.
9. Jonathan D. Moffett and Morris S. Sloman, "Policy Conflict Analysis in Distributed Systems Management," *Journal of Organizational Computing* 4, no. 1 (1994): 1-22.

MUHAMMAD NABEEL TAHIR works as a Lecturer in Multimedia University Melaka-Malaysia. He is pursuing his PhD from Multimedia University Malaysia and his area of interest is Object Oriented Software Analysis & Design, Information Security in Medical Information Systems. He holds a Master of Science Degree in Computer Science and teaches to Bachelors Degree in Computer Science students at Multimedia University. *Address for Correspondence:* Multimedia University, Ayer Keroh P.O. Box 75450, Melaka-Malaysia; *Phone:* 0060-126823495, 0060-6-2523422; *Fax:* 0060-6-2318840, *Email:* m_nabeeltahir@hotmail.com.

A BTC-BASED WATERMARKING SCHEME FOR DIGITAL IMAGES

Shu-Fen TU and Ching-Sheng HSU

Abstract: This article presents a novel watermarking scheme for digital images based on Block Truncation Coding (BTC). Unlike other watermarking schemes, the proposed method does not alter the original image during the process of watermark casting. Instead, an ownership share is constructed as a key to reveal the watermark without resorting to the original image. Moreover, it is possible to register multiple watermarks for a single host image in the proposed scheme. During watermark casting, the feature of the host image, which is extracted by means of BTC, is combined with the watermark to generate an ownership share. When revealing the watermark, the author can address his/her ownership share to extract the watermark. Even though the host image has been attacked, the extracted watermark is still perceivable. Altogether, the method has many other applications besides copyright protection. For example, it can be used to cover the transmission of confidential images. The experimental results of this study show the robustness of the novel scheme against several common attacks.

Keywords: Copyright Protection Scheme, Block Truncation Coding, Intellectual Property Right, Watermarking.

The development of Internet promotes communication of digital multimedia, such as image, audio, and video. All digital data possess several common features that can be used to harm the intellectual property right. For example, they are easy to falsify, to counterfeit, to snoop, and to duplicate. Nowadays, many techniques have been developed to protect the intellectual property right for digital images. Digital watermarking, a kind of such technique, is designed to insert a meaningful signature, called a watermark, directly into a digital host image to register the ownership. Then, the watermark can be extracted when the ownership of the image needs to be identified. The watermark in the watermarked image can be either visible[1] or invisible.[2,3,4,5] This article focuses on the invisible watermarks. Generally, the watermark should meet certain requirements, such as:

- Imperceptibility to human eyes,

- Robustness to common image processing operations and malicious attacks,
- Unambiguousness to ownership and copyright identification,
- Security and keys against unauthorized parties, and
- Capacity for embedding maximum information.[6,7,8,9]

Some of these requirements may conflict with each other and thereby introduce many technical challenges. For example, imperceptibility and capacity may conflict with robustness. Therefore, a reasonable compromise is required to achieve better performance for the intended applications. Current watermarking methods can be grouped into two categories. One is the spatial-domain approach,[10,11,12] and the other is the transform-domain approach.[13,14,15,16] Most related techniques have to alter the original image to embed watermark. Therefore, if multiple watermarks need to be registered for a single digital image, it is impossible for such methods to embed a subsequent watermark without destroying the former ones. In addition, when the ownership of the image needs to be identified, some of the methods require the aid of the original image to extract the watermark.

Recently, Chang, Hsiao, and Yeh[17] utilized visual cryptography and discrete cosine transformation (DCT) to design a copyright protection scheme that allows registering multiple ownerships. In essence, their model comprises ownership share construction and watermark revelation phases. During the ownership share construction phase, the DC coefficients of the different DCT blocks are extracted from the host image to form a master share, then an ownership share obtained by combining the master share and the watermark is constructed as a key to reveal the watermark without resorting to the original image. Since their method does not actually embed the watermark into the image, the host image will not be altered. However, their method requires the size of the watermark to be much smaller than that of the host image. For example, if the size of the original image is $M_1 \times M_2$, then the size of watermarks in their method should be at most $M_1/12 \times M_2/12$ for four colors, $M_1/20 \times M_2/20$ for 13 colors, and $M_1/92 \times M_2/92$ for gray-level and 256 colors.[18] Besides, a transformation of the image from the spatial domain to the frequency domain has to be performed so that the master share can be extracted.

In this article, a digital watermarking scheme without resorting to the host image is proposed. The authors apply Block Truncation Coding (BTC) to transform a gray-level host image into a binary image, which preserves the features of the host image. Then, the binary image is combined with the watermark to construct the ownership share with the aid of the XOR operation. During the watermark casting process, a pseudo-random key is used to permute the host image to enhance the robustness of the proposed scheme. When the rightful ownership needs to be identified, the authors again transform the image to be identified to a binary one and then combine it with the ownership share to reveal the watermark.

In summary, the new method has several advantages. First, there is no need to alter the original image; hence, the image quality will not be degraded and the risk of deliberately detecting or erasing the watermark from the host image can be avoided. Second, the proposed method can identify the ownership without resorting to the original image. Third, multiple watermarks are allowed to be registered for a single image without causing any damage to other hidden watermarks. Fourth, the method can achieve the requirement of robustness for digital watermarking due to the fact that the features of the image can not be easily changed by many attacks. Fifth, the security of the scheme is ensured by the ownership share kept by a trusted third party and the secret key held secretly by the copyright owner. Altogether, the new method has many other applications than copyright protection. For example, it can be applied to protect the transmission of confidential images.

The rest of the article is organized as follows. The next section presents a brief description of Block Truncation Coding (BTC). Then, the authors demonstrate how to utilize BTC in the method to construct a copyright protection scheme with a binary watermark. Experimental results, which prove the robustness of the proposed method, are given after that. And finally, discussion and conclusions are presented in the last section.

Block Truncation Coding

Block truncation coding is a lossy compression technique for gray-level images proposed by Delp and Mitchell.[19] The image is divided into blocks of m pixels and each block is processed separately. The mean value (μ) and the standard deviation (σ) are calculated for each block and the first two sample moments are preserved in the compression. The original block is encoded into a bit plane (B), where pixels with values less than the mean value are set to ‘0’, and those with values greater than or equal to the mean value are set to ‘1’. The block is decompressed according to the triple (μ, σ, B). The bit ‘0’ of B is set to a, and the bit ‘1’ of B is set to b, where a and b are computed according to Equations (1a) and (1b), and q stands for the number of bits ‘1’ in B.

$$a = \mu - \sigma \cdot \sqrt{\frac{q}{m-q}} \tag{1a}$$

$$b = \mu + \sigma \cdot \sqrt{\frac{m-q}{q}} \tag{1b}$$

2	9	12	15
2	11	11	9
2	3	12	15
3	3	4	14

$\mu = 7.94$

$\sigma = 4.91$

(a) Original

0	1	1	1
0	1	1	1
0	0	1	1
0	0	0	1

$q = 9$

(b) Bit-plane B

2	12	12	12
2	12	12	12
2	2	12	12
2	2	2	12

$a = 2.3$

$b = 12.3$

(c) Reconstructed

Figure 1: An Example of BTC by (μ , σ , B).

Figure 1 is an example taken from the article of Fränti, Nevalatinen, and Kaukoranta.[20] Figure 1(a) is an original block of 16 pixels, and the mean value μ and standard deviation σ of the block are 7.94 and 4.91, respectively. The block is encoded into a bit plane B with nine bits '1' as shown in Figure 1(b). The block can be reconstructed according to triple (μ , σ , B) and Equations (1a) and (1b). Although the compression ratio (i.e. a bit rate) is not low enough, this coding method gives good reconstructed image since it preserves local characteristics of blocks of the image important to the human observer. Besides, the process of compression and decompression is very simple and fast.

In this article, the authors utilize BTC to construct the master share of the original image; therefore, the master share can preserve the features of the original image. Together, the master share and the watermark are used to construct the ownership share. Even though the original image is attacked, the features of the image can not be changed much. Since the ownership share is generated according to the features of the original image, the authors can enhance the robustness of their scheme. Readers interested in BTC can refer to the work of Delp and Mitchell[21] and Fränti, Nevalatinen, and Kaukoranta[22] for further reading.

The Proposed Scheme

This section demonstrates how to cast a binary watermark into a gray-level host image and how to detect the watermark from a gray-level test image to identify the ownership. The whole process in the proposed scheme can be partitioned into two phases: one is the watermark casting phase; the other is the watermark detection phase. In the watermark casting phase, the host image is divided into equal-sized blocks of 3×3 pixels. Then, BTC is used to extract the features of each block. Finally, the ownership share is constructed according to the features of the host image and the pixels of the watermark. In the watermark detection phase, the test image and the ownership share are divided into equal-sized blocks of 3×3 pixels. Then, each block of the test image is transformed into a bit plane. Finally, the pixels of the watermark are set according to the bit planes of the test image and the corresponding block of the ownership share. Since the ownership share is the key to identify the ownership of the image, the copy-

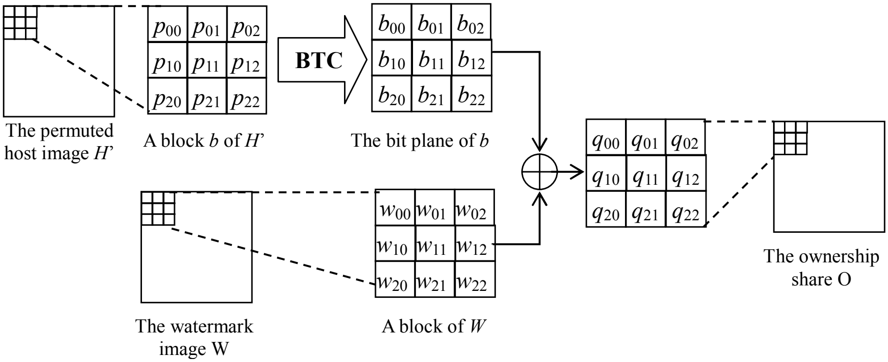

p_{00} p_{01} p_{02}
p_{10} p_{11} p_{12}
p_{20} p_{21} p_{22}
BTC
b_{00} b_{01} b_{02}
b_{10} b_{11} b_{12}
b_{20} b_{21} b_{22}
The permuted host image H'
A block b of H'
The bit plane of b
q_{00} q_{01} q_{02}
q_{10} q_{11} q_{12}
q_{20} q_{21} q_{22}
w_{00} w_{01} w_{02}
w_{10} w_{11} w_{12}
w_{20} w_{21} w_{22}
The ownership share O
The watermark image W
A block of W

Figure 2: Illustration of Watermark Casting.

right owner has to send it to a trusted third-party for further authentication. Combined with BTC, the proposed scheme is shown to be robust for detecting the watermark. Note that the host image has to be permuted by a pseudo random number generator before BTC is applied to it.

Watermark Casting

Presume that a binary watermark W of $N \times M$ pixels is cast into a gray-level host image H of $N \times M$ pixels. First, H has to be permuted by a pseudo-random number generator seeded by a key k to enhance the robustness of the scheme. Then the permuted host image H' is divided into blocks of 3×3 pixels, each of which corresponds to a pixel of W. Each block of H' is transformed into a bit plane according to the mean value μ of the pixels in the block; that is, pixels with values greater than or equal to μ are set to '1' (i.e. black), and those with values less than μ are set to '0' (i.e. white). Meanwhile, the corresponding pixel of W is divided into blocks of 3×3 pixels as well. The two bit planes are used to create the corresponding 3×3 block of the ownership share O through XOR logic. Figure 2 depicts the whole process of watermark casting, and the algorithm of watermark casting is shown below.

Algorithm *Watermark Casting*

Input: A gray-level watermark image W of $N \times M$ pixels
A gray-level host image H of $N \times M$ pixels
A seed k of the pseudo random number generator

Output: An ownership share O of $N \times M$ pixels

Step 1. Permute H by pseudo-random numbers seeded by k. Denote the permuted image as H'.

Step 2. Divide H' into equal-sized blocks of 3×3 pixels to derive a set of blocks $\{H'_{ij}\}$, where $i = 0 \sim (N/3-1)$ and $j = 0 \sim (M/3-1)$. Each pixel of H'_{ij} is denoted as p^{ij}_{mn}, where $m = 0..2$ and $n = 0..2$, and p^{ij}_{00} is located at $(3i, 3j)$ of H'.

Step 3. For each block H'_{ij}, calculate the mean value μ_{ij} of the pixels. Set p^{ij}_{mn} to '1' if its value is greater than or equal to μ_{ij}; otherwise, set it to '0'. Hence, H' is transformed into a binary image containing $(N/3) \times (M/3)$ bit planes $\{B_{ij}\}$ of 3×3 pixels b^{ij}_{mn}, where $i = 0..(N/3-1)$, $j = 0..(M/3-1)$, $m = 0..2$ and $n = 0..2$.

Step 4. Divide W into equal-sized blocks of 3×3 pixels to derive a set of blocks $\{W_{ij}\}$, where $i = 0..(N/3-1)$ and $j = 0..(M/3-1)$. Each pixel of W_{ij} is denoted as w^{ij}_{mn}, where $m = 0..2$ and $n = 0..2$, and w^{ij}_{00} is located at $(3i, 3j)$ of W.

Step 5. Generate the pixel q_{uv} of O through XOR logic; i.e. $q_{uv} = b^{ij}_{mn} \oplus w^{ij}_{mn}$, where $u = 3i + m$, $v = 3j + m$, $i = 0..(N/3-1)$, $j = 0..(M/3-1)$, $m = 0..2$ and $n = 0..2$.

Step 6. Perform Step 5 repeatedly until all pixels of O are generated.

Figure 3 is an example of watermark casting. Suppose that Figure 3(a) is a 3×3 block of H' and the corresponding bit plane of W is shown in Figure 3(c). The mean value μ of the pixels in b is 69.33; therefore, pixels with values greater than or equal to 69.33 are set to '1'; otherwise, they are set to '0' as shown in Figure 3(b). Figure 3(d) is the corresponding block of the ownership share, where each pixel is the XOR result of the corresponding pixels of the two bit planes. After completing watermark casting, the copyright owner has to send the ownership share to a trusted third-party for authentication. In addition, the key to permute the host image has to be kept by the owner.

55	120	70
25	36	6
18	94	200

(a) A block *b* of H'
μ = 69.33

0	1	1
0	0	0
0	1	1

(b) The bit plane of (a)

1	1	1
1	1	1
0	0	0

(c) The corresponding bit plane of *W*

1	0	0
1	1	1
0	1	1

(d) The XOR result of (b) and (c)

Figure 3: An Example of Watermark Casting.

Watermark Detection

If a gray-level image G is suspected to be a piracy copy, the owner can resolve the dispute about the ownership by detecting the existence of a watermark. First, the test image G is permuted by a pseudo-random number generator seeded by the same key k. Then, both the permuted image G' and the ownership share O are divided into blocks of 3×3 pixels. Each block of G' is transformed into a bit plane by means of BTC. Then the owner has to address his/her authenticated ownership share O. Note that G has to be permuted with the same pseudo-random key before starting the process of master share construction. Let G' denote the permuted image. With G' and O we can reveal the watermark W', which may be different from the original one W. The process of watermark revelation is described as follows.

Algorithm *Watermark Detection*

***Input**:* A gray-level test image G of $N \times M$ pixels
An ownership share O of $N \times M$ pixels
A seed k of the pseudo random number generator

***Output**:* A watermark W' of $N \times M$ pixels

Step 1. Permute G by pseudo-random numbers seeded by k. Denote the permuted image as G'.

Step 2. Divide G' into equal-sized blocks of 3×3 pixels to derive a set of blocks $\{G'_{ij}\}$, where $i = 0..(N/3-1)$ and $j = 0..(M/3-1)$. Each pixel of G'_{ij} is denoted as p^{ij}_{mn}, where $m = 0..2$ and $n = 0..2$, and p^{ij}_{00} is located at $(3i,3j)$ of G'.

Step 3. For each block G'_{ij}, calculate the mean value μ_{ij} of the pixels. Set p^{ij}_{mn} to '1' if its value is greater than or equal to μ_{ij}; otherwise, set it to '0'. Hence G' is transformed into a binary image containing $(N/3) \times (M/3)$ bit planes $\{B_{ij}\}$ of 3×3 pixels b^{ij}_{mn}, where $i = 0..(N/3-1)$, $j = 0..(M/3-1)$, $m = 0..2$ and $n = 0..2$.

Step 4. Divide O into equal-sized blocks of 3×3 pixels to derive a set of blocks $\{O_{ij}\}$, where $i = 0..(N/3-1)$ and $j = 0..(M/3-1)$. Each pixel of O_{ij} is denoted as q^{ij}_{mn}, where $m = 0..2$ and $n = 0..2$, and q^{ij}_{00} is located at $(3i,3j)$ of O.

Step 5. Generate pixels w_{uv} of W through performing XOR logic on b^{ij}_{mn} and q^{ij}_{mn};

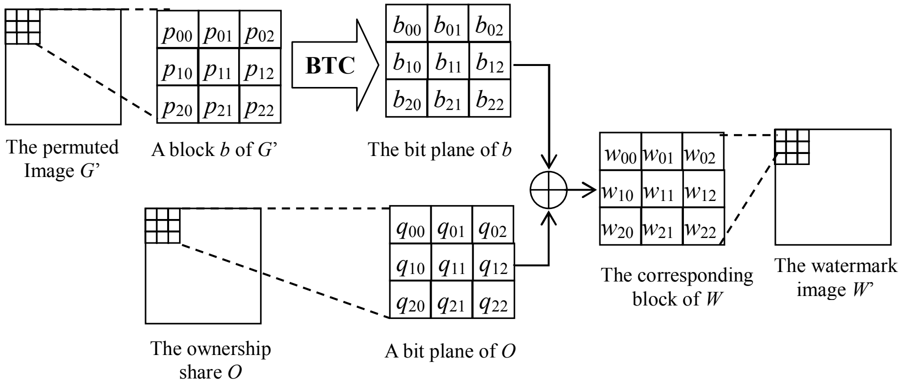

The permuted Image G'
p_{00} p_{01} p_{02}
p_{10} p_{11} p_{12}
p_{20} p_{21} p_{22}
A block b of G'
BTC
b_{00} b_{01} b_{02}
b_{10} b_{11} b_{12}
b_{20} b_{21} b_{22}
The bit plane of b
w_{00} w_{01} w_{02}
w_{10} w_{11} w_{12}
w_{20} w_{21} w_{22}
The corresponding block of W
The watermark image W'
q_{00} q_{01} q_{02}
q_{10} q_{11} q_{12}
q_{20} q_{21} q_{22}
The ownership share O
A bit plane of O

Figure 4: Illustration of Watermark Revelation.

i.e. $w_{uv} = b^{ij}_{mn} \oplus q^{ij}_{mn}$, where $u = 3i + m$, $v = 3j + m$, $i = 0..(N/3-1)$, $j = 0..(M/3-1)$, $m = 0..2$ and $n = 0..2$.

Step 6. Perform Step 5 repeatedly until all pixels of W are generated.

Figure 4 illustrates the process of watermark revelation. The permuted gray-level image G' is divided into equal-sized blocks of 3×3 pixels. We take a block b from G' and transform it into a bit plane by means of the BTC method. Next, we take a corresponding bit plane from O and perform the XOR operation on the two bit planes. Finally, the corresponding block of W' can be generated. The remaining blocks of W' are generated by analogy. Figure 5 is an example of watermark revelation. Figure 5(a) is a block b of G' with mean value μ =50.67. Performing XOR logic on the bit plane of b (Figure 5(b)) and the corresponding bit plane of O (Figure 5(c)), we can get the corresponding block of W' as shown in Figure 5(d).

92	15	100
66	84	10
3	26	60

(a) A block b of G'
μ = 50.67

1	0	1
1	1	0
0	0	1

(b) The bit plane of (a)

0	1	1
1	0	0
0	0	0

(c) The corresponding bit plane of O

1	1	0
0	1	0
0	0	1

(d) The XOR result of (b) and (c)

Figure 5: An Example of Watermark Revelation.

Experimental Results

In this section, we use Figure 6(a) as an experiment to demonstrate the robustness of the proposed scheme against several common attacks, including blurred (Figure 6(b)), brighten (Figure 6(c)), JPEG lossy compressed (Figure 6(d)), cropped (Figure 6(e)), darken (Figure 6(f)), distorted (Figure 6(g)), noised (Figure 6(h)), rescaled (Figure 6(i)), and sharpen (Figure 6(j)). Two common similarity measurements are introduced to evaluate the proposed watermarking scheme. One is the peak signal-to-noise ratio (PSNR) used to evaluate the similarity of two gray-level images and the other is the normalized correlation (NC) used to measure the similarity between two bi-level images. The first measurement, signal-to-noise ratio, is defined as follows:

$$PSNR = 10 \times \log \frac{255^2}{MSE} \tag{2}$$

where

$$MSE = \frac{1}{M_1 \times M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} (c_{i,j} - c'_{i,j})^2 \tag{3}$$

$c_{i,j}$ denotes a pixel color of the original host image, $c'_{i,j}$ denotes a pixel color of the attacked image, and $M_1 \times M_2$ is the image size. The second measurement, normalized correlation, is defined as follows:

$$NC = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \overline{w_{i,j} \oplus w'_{i,j}}}{N_1 \times N_2} \times 100\% \tag{4}$$

where $w_{i,j}$ denotes a pixel color of the original watermark W, $w'_{i,j}$ denotes a pixel color of the revealed watermark W', and $N_1 \times N_2$ is the image size.[23]

The size of the host image is 384×384 pixels, and that of the watermark is 192×192 pixels. Therefore, only half of the host image is needed to construct the ownership share. We use the PSNR as a measurement to express how severe the host image is attacked. The smaller the PSNR is, the more dissimilar the attacked image is. Usually, ‘30’ is a tolerable bottom line. Observing Figure 6, we can see that most PSNR values are less than ‘30’, which means that the host image comes under severe attacks.

Figure 7(a) is the original watermark, and Figures 7(b)-7(j) are extracted watermarks from Figures 6(b)-6(j). These extracted watermarks are compared to the original watermark with the measurement NC. The larger the NC is, the more similar the extracted watermark is. Usually, ‘80%’ is a tolerable bottom line. We can see from Figure 7 that the NC values of the extracted watermark are all greater than 80%.

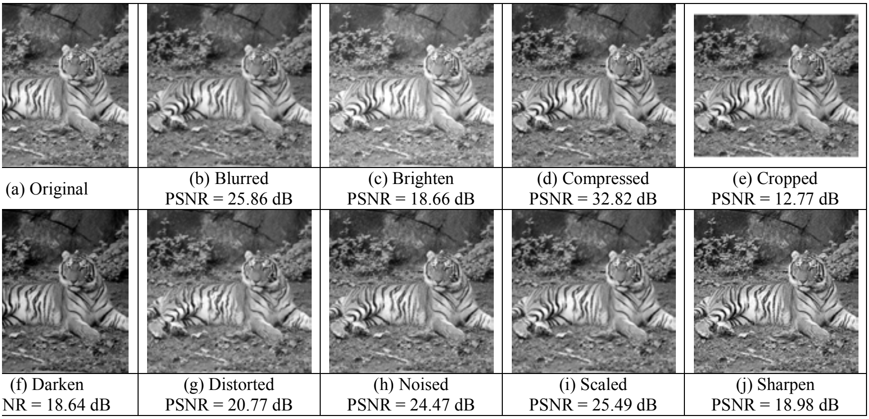
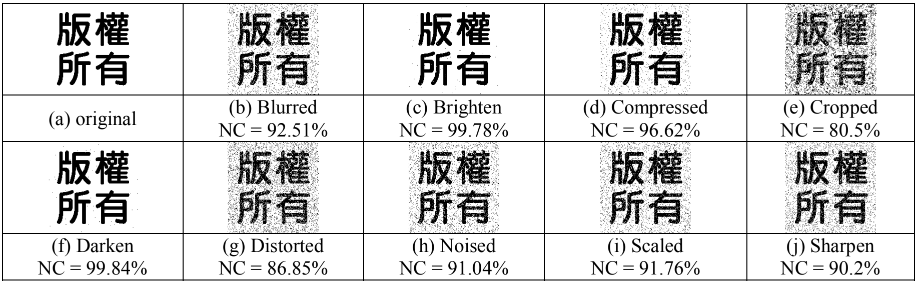

(a) Original
(b) Blurred PSNR = 25.86 dB
(c) Brighten PSNR = 18.66 dB
(d) Compressed PSNR = 32.82 dB
(e) Cropped PSNR = 12.77 dB
(f) Darken NR = 18.64 dB
(g) Distorted PSNR = 20.77 dB
(h) Noised PSNR = 24.47 dB
(i) Scaled PSNR = 25.49 dB
(j) Sharpen PSNR = 18.98 dB

Figure 6: The Host Image and Nine Different Attacks on the Host Image (384 × 384 pixels, 300 dpi).

Besides objective measurement, human eyes are also a good subjective measurement. It is obvious that the words on the extracted watermarks are still distinguishable.

Conclusions

In this article, a novel scheme for digital watermarking based on BTC is proposed. The presented method applies BTC to extract the features of the host image. Then, the ownership share is constructed by combining the features of the host image with the watermark. Since the watermark is cast according to the features of the host image, the authors believe that it is highly possible the watermark to survive under attacks. Actually, the experimental results also show the robustness of the new scheme.



(a) original
(b) Blurred NC = 92.51%
(c) Brighten NC = 99.78%
(d) Compressed NC = 96.62%
(e) Cropped NC = 80.5%
(f) Darken NC = 99.84%
(g) Distorted NC = 86.85%
(h) Noised NC = 91.04%
(i) Scaled NC = 91.76%
(j) Sharpen NC = 90.2%

Figure 7: The Original Watermark and Revealed Watermarks Extracted from Figure 6(b)-6(j) (192 × 192 pixels, 300 dpi).

In summary, the proposed scheme has the following advantages. First, the host image is not altered or damaged by the watermark casting procedure due to the fact that the watermark is not actually embedded into the host image. Hence, the image quality will not be degraded and the risk of deliberately detecting or erasing the watermark from the host image can be avoided. Second, the method can identify the ownership without resorting to the original image. Third, the proposed scheme allows multiple watermarks to be cast into a single host image without causing any damage to other hidden watermarks. Fourth, the security of the scheme can be ensured by the ownership share and the secret key held secretly by the copyright owner. Finally, the scheme seems robust according to the experimental results.

Although the present version of the proposed scheme only deals with bi-level watermarks, it is possible to extend the method to gray-level or color watermarks. For example, each pixel of a gray-level watermark can be encoded into a block of 3 × 3 pixels, where the first eight pixels of the block map to the eight bits of a pixel value of the watermark. To deal with color watermarks, we can employ the last pixel of the block to store the palette of 256 colors. Thus, the first eight pixels can be used to represent the index of colors. In the future, the issue of gray-level and color watermarks will be further studied.

Notes:

[1] Gordon W. Braudaway, Karen A. Magerlein, and Frederick C. Mintzer, "Protecting Publicly-Available Images with a Visible Image Watermark," in *Proceedings of the SPIE International Conference on Electronic Imaging, Science and Technology: Optical Security and Counterfeit Deterrence Techniques* (San Jose, CA, 1-2 February 1996), Volume 2659 (Bellingham, Wa.: SPIE, 1996), 126-133.

[2] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Transactions on Image Processing* 6, no. 12 (1997): 1673-1687.

[3] Stephen H. Low and Nicholas F. Maxemchuk, "Performance Comparison of Two Text Marking Methods," *IEEE Journal on Selected Areas in Communications* 16, no. 4 (May 1998): 561-572.

[4] Kineo Matsui, Junji Ohnishi, and Yasuhiro Nakamura, "Embedding a Signature to Pictures under Wavelet Transform," *IEICE Transactions* J79-D-II, no. 6 (June 1996): 1017-1024.

[5] Ryutarou Ohbuchi, Hiroshi Masuda, and Masaki Aono, "Watermarking Three-Dimensional Polygonal Models through Geometric and Topological Modifications," *IEEE Journal on Selected Areas in Communications* 16, no. 4 (May 1998): 551-560.

[6] Cox, Kilian, Leighton, and Shamoon, "Secure Spread Spectrum Watermarking for Multimedia."

[7] Stefan Katzenbeisser and Fabien A.P. Petitcolas, eds., *Information Hiding Techniques for Steganography and Digital Watermarking* (Norwood, MA: Artech House Inc., January 2000), 101-109.

[8] Eckhard Koch, Jochen Rindfrey, and Jian Zhao, "Copyright Protection for Multimedia Data" (paper presented at the International Conference on Digital Media and Electronic Publishing, Leeds, UK, December 1994), 6-8.

[9] Nikos Nikolaidis and Ioannis Pitas, "Copyright Protection of Images using Robust Digital Signatures," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP-96)*, Volume 4 (Atlanta, Georgia, May 1996), 2168-2171.

[10] Young-Chang Hou and Pei-Min Chen, "An Asymmetric Watermarking Scheme Based on Visual Cryptography," in *Proceedings of the Fifth Signal Processing Conference*, Volume 2 (Beijing, China, 21-25 August 2000), 992-995.

[11] Low and Maxemchuk, "Performance Comparison of Two Text Marking Methods."

[12] Ohbuchi, Masuda, and Aono, "Watermarking Three-Dimensional Polygonal Models through Geometric and Topological Modifications."

[13] Chin-Chen Chang, Ju Yuan Hsiao, and Jyh-Chiang Yeh, "A Colour Image Copyright Protection Scheme based on Visual Cryptography and Discrete Cosine Transform," *The Imaging Science Journal* 50 (2002): 133-140.

[14] Cox, Kilian, Leighton, and Shamoon, "Secure Spread Spectrum Watermarking for Multimedia."

[15] Chiou-Ting Hsu and Ja-Ling Wu, "Hidden Digital Watermarks in Images," *IEEE Transactions on Image Processing* 8, no. 1 (January 1999): 58-68.

[16] Young-Sik Kim, O-Hyung Kwon, and Rae-Hong Park, "Wavelet Based Watermarking Method for Digital Images using the Human Visual System," *Electronics Letters* 35, no. 6 (March 1999): 466-468.

[17] Chang, Hsiao, and Yeh, "A Colour Image Copyright Protection Scheme based on Visual Cryptography and Discrete Cosine Transform."

[18] Chang, Hsiao, and Yeh, "A Colour Image Copyright Protection Scheme based on Visual Cryptography and Discrete Cosine Transform."

[19] Edward J. Delp and O. Robert Mitchell, "Image Compression using Block Truncation Coding," *IEEE Transactions on Communications* 27, no. 9 (September 1979): 1335-1342.

[20] Pasi Fränti, Olli Nevalainen, and Timo Kaukoranta, "Compression of Digital Images by Block Truncation Coding: A Survey," *The Computer Journal* 37, no. 4 (1994): 308-332.

[21] Delp and Mitchell, "Image Compression using Block Truncation Coding."

[22] Fränti, Nevalatinen, and Kaukoranta, "Compression of Digital Images by Block Truncation Coding: A Survey."

[23] Chang, Hsiao, and Yeh, "A Colour Image Copyright Protection Scheme based on Visual Cryptography and Discrete Cosine Transform."

SHU-FEN TU received a BS degree in Management Information System from the National Cheng Chi University, Taiwan, R.O.C., in 1996 and a MS degree in Information Management from the National Chi Nan University, Taiwan, R.O.C., in 1998. From 1998 to 1999, she was a software engineer at The Syscom Group Co., Taiwan, R.O.C. Currently, she is pursuing her Ph.D degree in Information Management at the National Central University, Taiwan, R.O.C. Her research interests include steganography, document protection, and secret sharing. *Address for correspondence*: Department of Information Management, National Central University, P.O. Box 9-277, Jhongli, Taoyuan County 32099, Taiwan, R.O.C.; *Phone:* +886-3-420-4131; *E-mail:* ariel_tu@anet.net.tw.

CHING-SHENG HSU received a BS degree in Management Information Systems from the National Cheng Chi University, Taiwan, R.O.C., in 1994, and a MS degree in Information Management from the National Chi Nan University, Taiwan, R.O.C., in 1998. From 1998 to 1999, he was a software engineer at The Syscom Group Co., Taiwan, R.O.C. Currently, he is pursuing his Ph.D degree in Information Management at the National Central University, Taiwan, R.O.C. His research interests include steganography, copyright protection, cryptography, evolutionary computation, and distance learning. *Address for correspondence*: Department of Information Management, National Central University, P.O. Box 9-236, Jhongli, Taoyuan County 32099, Taiwan, R.O.C.; *Phone:* +886-3-420-3157; *E-mail:* jacketcc@mgt.ncu.edu.tw.

- ♦ I&S Internet Sources

I&S Monitor

.

E-GOVERNMENT AND SECURITY OF INFORMATION INTERNET SOURCES

E-GOVERNMENT AND SECURITY OF INFORMATION RELATED PUBLICATIONS

GENERAL PUBLICATIONS

U.S. e-Government Strategy

http://www.whitehouse.gov/omb/inforeg/egovstrategy.pdf

The U.S. e-Government Strategy was published on 27 February 2002. It presents the federal government's action plan for e-Government. The primary goals for the U.S. President's "Expanding E-Government" initiative are to:

- Make it easy for citizens to obtain service and interact with the federal government;
- Improve government efficiency and effectiveness; and
- Improve government's responsiveness to citizens.

OMB Director Mitchell E. Daniels initiated an interagency E-Government Task Force to identify the action plan for implementing the President's E-Government initiative. The Task Force was made up of about 80 federal employees from across the federal government.

BSI's E-Government Manual

http://www.bsi.bund.de/fachthem/egov/

http://www.e-government-handbuch.de

These sites provide up-to-date information from Das Bundesamt für Sicherheit in der Informationstechnik (BSI) on the subject of "Secure e-Government," and the latest version of the E-Government Manual.

.

Australia – Better Services, Better Government – The Federal Government's e-Government Strategy

http://www.noie.gov.au/publications/NOIE/better_services-better_gov/index.htm

A framework for the next stage of e-government was launched on 11 November 2002 by Minister for Communications, Information Technology and the Arts, Senator Richard Alston. Better Services, Better Government is a high level e-government strategy designed to outline broad directions and priorities for the future of e-Government. It seeks to maintain the momentum of the achievements under the Government Online Strategy.

Canada – Government On-Line: Serving Canadians in a Digital World

http://www.gol-ged.gc.ca/pub/serv-can/serv-can00_e.asp

Government On-Line is the plan that supports the Prime Minister's commitment to make the Government of Canada the most electronically connected government in the world to its citizens by 2004 and provide Canadians with electronic access to federal information and services.

United Kingdom Online Strategy

http://e-government.cabinetoffice.gov.uk/Homepage/fs/en

This Strategy is the UK Government's comprehensive programme to lead the knowledge economy revolution, and how the Strategy is managed.

New Zealand E-government Strategy 2001 and 2003

http://www.egov.vic.gov.au/pdfs/egovt-strategy.pdf

http://www.e-government.govt.nz/docs/e-gov-strategy-june-2003/index.html

http://www.e-government.govt.nz/docs/e-gov-strategy-june-2003/strategy-2003-complete.pdf

On 26 April 2001, the NZ Government launched the e-Government Strategy, which sets out an operational vision for e-government: that New Zealand will be a world leader in e-government.

eEurope 2005 – A Study of the Degree of Alignment of the New Member States and the Candidate Countries

http://www.dree.org/elargissement/RapportsSite/INSEADeEuropeCompar0408.pdf

This study prepared by INSEAD for SAP measures the alignment of the new EU Member States and the candidate countries with the ‘old’ members with regard to the objectives set in the eEurope 2005 Action Plan. This measurement is based on the eEurope 2005 Index, a composite indicator of the level of development of the Information Society incorporating five dimensions: general Internet indicators, modern online public services, dynamic e-business environment, secure information infrastructure, and broadband. Based on their performance in these fields, four categories of countries were identified according to their alignment with the average score of the 15 ‘old’ Member States: ‘Global Leaders’, ‘Totally Aligned’, ‘Somewhat Aligned’, and ‘Development required’.

European Interoperability Framework for pan-European e-Government Services

http://europa.eu.int/idabc/en/document/3761

http://europa.eu.int/idabc/servlets/Doc?id=18952

The EIF is the reference document on interoperability for the IDABC program. It is the result of an extensive consultation process with the Member States and thus represents the highest ranking module for the implementation of European e-Government services. This first version provides a series of recommendations and defines generic standards with regard to organizational, semantic and technical aspects of interoperability, offering a comprehensive set of principles for European co-operation in e-Government. More information about the EIF related activities of the IDABC program can be found at: http://europa.eu.int/idabc/en/document/2319/ 5644.

U.S. Department of Labor e-Government Strategic Plan

http://www.dol.gov/_sec/e_government_plan/p23_security_privacy.htm

This document focuses on development of U.S. Department of Labor’s e-Government security and privacy framework, implementation of PKI, and assessment of the impact of privacy issues related to IT systems.

U.K. e-Government Interoperability Framework

http://www.govtalk.gov.uk/schemasstandards/egif.asp

The e-Government Interoperability Framework (e-GIF) defines the technical policies and specifications governing information flows across government and the public sector. They cover interconnectivity, data integration, e-services access and content management. Version 6.0 contains the high level policy statements, management, implementation and compliance regimes, whilst technical policies and specifications are contained in the Technical Standards Catalogue. This document was created by the U.K. Cabinet Office, e-Government Unit, Technology Policy Team, Interoperability Policy Advisor and published on 19 August 2004.

U.K. e-Service Development Framework

http://www.govtalk.gov.uk/schemasstandards/eservices_document2.asp?docnum=515

The e-Service Development Framework provides a structure for developing semantic specifications and standards for e-Services. An e-Service is any electronic service involving interoperability between computer systems. This is Version 1.0 of the document, published on 12 May 2001.

E-Government Leadership – Realizing the Vision

http://www.accenture.com/xd/xd.asp?it=enweb&xd=newsroom/epresskit/egov/epres_realizing.xml

http://www.accenture.com/xdoc/en/industries/government/egov_april2002_3.pdf

National governments throughout the world significantly improved their online service delivery, increasing the range and sophistication of e-Government services for citizens and businesses alike, according to Accenture's third annual global e-Government study. As governments progress along the e-Government path, they also are demonstrating greater understanding of technology's potential to help fully transform the way they operate – both in terms of service delivery and administrative effectiveness. This Accenture Report looks in detail at findings from 23 countries. Accenture looks at progress made in these countries since its previous study conducted in 2001. The report also shows numerous examples of best practices in several government service sectors.

Papers by Janet Caldow, Director, Institute for Electronic Government, IBM Corporation

http://www-1.ibm.com/industries/government/ieg/library/papers.html

- *Seven E-Government Leadership Milestones* As electronic government comes of age around the world, leadership remains at the core of success, beginning with the definition of e-government itself (http://www-1.ibm.com/industries/government/ieg/pdf/Seven_E-Gov_Milestones.pdf).
- *e-Democracy: Putting Down Global Roots* As governments achieve more sophisticated levels of e-Government, strategy should include progression to

more sophisticated levels of e-Democracy within and beyond national borders.

- *Lessons from around the World* Until leaders are willing to inspire fundamental reform, e-Government will remain unfulfilled – an elusive concept.
- *e-Government Goes Wireless: From Palm to Shining Palm* Just when you thought you could sit back, relax, and let e-government roll its way onto shore, the next big swell has already formed out there on the Internet high seas. Wireless. Need some proof.
- *The Virtual Ballot Box: A Survey of Digital Democracy in Europe* Elected representatives today are serving office during a unique period in history – at the threshold of a technological, social, economic, and political transition the likes of which we haven't seen since the Industrial Revolution.
- *e-Government: A Go-to-Market Strategy* Five years ago, the term "electronic government" was not in our vocabulary. Today, "e-government" has become a battle cry for digital age governments around the world.
- *The Quest for Electronic Government: A Defining Vision* This paper incorporates good advice from four years' research, literature searches, innovative practices from around the world, emerging strategies and future indicators and trends.

E-Government Enterprise Architecture Guidance (Common Reference Model)

http://www.feapmo.gov/resources/E-Gov_Guidance_Final_Draft_v2.0.pdf

The purpose of this document is to provide augmenting architectural guidance to the official direction from the U.S. Federal Enterprise Architecture Program Management Office. It is intended to provide a consistent, industry-aligned approach for defining and communicating about the components needed to cost and plan e-Government programs – both the 24 Presidential Priority e-Government Initiatives and other e-Government Initiatives across the Federal Government.

This document describes a Federal-wide e-Government target conceptual architecture. The architecture is based on the business requirements derived from the initiatives as well as system engineering design best practices. It provides a workable description of the components needed by e-Government Initiatives and business activities to move rapidly into the web service-enabled business transaction environment. This is Version 2.0 of the document, prepared by the FEA Working Group; published on 25 July 2002, and endorsed by the Architecture and Infrastructure Committee, Federal CIO Council.

E-Government: Connecting the Dots?

http://www.accenture.com/xdoc/en/industries/government/eGovernmentA4.pdf

This study by Accenture looks at how Governments in Europe are meeting the challenge of re-inventing the business of Government. The report provides insights into executives' attitudes to e-Government—what private sector executives want from e-Government and the challenges and opportunities public sector executives perceive as they seek to deliver on the promise of e-Government.

Electronic Governance: Re-inventing Good Governance

http://www1.worldbank.org/publicsector/egov/Okot-Uma.pdf

A paper written by Rogers W'O Okot-Uma, Commonwealth Secretariat, London. The paper includes a discussion of different types of e-Government applications and some country experiences.

E-Government Best Practice Examples from Austria, Germany and Switzerland

http://www1.worldbank.org/publicsector/egov/egovbestpractice.pdf

The paper discusses some small service oriented projects that show how value can be delivered to citizens via small pilot projects.

Roadmap for E-government in the Developing World

http://unpan1.un.org/intradoc/groups/public/documents/other/unpan006407.pdf

The paper poses ten questions e-government leaders should ask themselves. It is prepared by the Working Group on e-Government in the Developing World; April 2002.

The Information Society: The Role of Information in the Emerging Global e-Government, e-Governance and e-Democracy Environments

http://www.electronicgov.net/pubs/research_papers/tracking03/IntlTrackingReporttApr03no3.pdf

International Tracking Survey Report '03, Number Three, by Thomas B. Riley, Chair and Executive Director, Commonwealth Centre for Electronic Governance, Visiting Professor, University of Glasgow, President, Riley Information Services. Prepared under the auspices of the Commonwealth Secretariat and co-sponsored by Government Telecommunications and Informatics Services, Public Works and Government Services Canada, April 16, 2003.

Latin America's e-Government Sites

http://www1.worldbank.org/publicsector/egov/LA_EGovSites.doc

The document classifies federal, state, and local government websites in Argentina, Brazil, Colombia, and Mexico along three dimensions: search (general vs. specific), communication (one-way vs. interactive), and type of linkage (between different levels of government or horizontal).

Global e-Government Report 2004

http://www.insidepolitics.org/egovt04int.html

A report by Darrell M. West, Center for Public Policy, Brown University, Providence, Rhode Island 02912-1977, United States. This study highlights how e-Government has developed around the world. It analyses 1 935 government websites in 198 different countries. It also highlights the various developments taking place across specific regions and countries.

Publications on Enabling e-Government of the Center for Technology in Government

http://www.ctg.albany.edu/publications/publications?sub=egov

- *Return on Investment in Information Technology: A Guide for Managers* (August 2004) New information technology (IT) systems are serious, and potentially risky, investments for government agencies and nonprofit organizations. This guide is designed to help public sector managers better understand how a return on investment (ROI) analysis can take some of that risk out of their next IT investment.
- *Untangle the Web: Delivering Municipal Services through the Internet* (December 2002) The Web offers people and organizations a new way to interact and communicate. This report provides a framework for helping local governments achieve the benefits of the Web without being overcome by its complexity.
- *Making a Case for Local E-Government* (July 2002) Local and county governments are exploring the best ways to implement e-Government. This report details the strategies, funding, barriers, and benefits brought to bear by several New York State local e-government pioneering initiatives, with insight and advice for their colleagues.
- *And Justice for All: Designing Your Business Case for Integrating Justice Information* (May 2000) Efforts to improve public safety in the United States

are pointing to an increasing need for justice agencies to share information. This guidebook offers a series of lessons and tools justice officials can use to build business cases to win support and funding for integrated justice information systems.

- *Highlights: Exploring the Feasibility of a Digital Government Journal* (July 2004) This project administered an online survey exploring the opinions and preferences of the digital government (DG) research community with respect to the need for, feasibility, and sustainability of a dedicated digital government journal.
- *The New York State-Local Internet Gateway Prototype Project: Current Practice Research* (July 2004) In the fall of 2002, the Center for Technology in Government (CTG) at the University at Albany conducted current practice research to identify and examine existing government to government (G2G) portal projects.
- *Bridging the Enterprise: Lessons from the New York State-Local Internet Gateway Prototype* (May 2004) This project report details the Gateway Prototype project from conceptualization and development to findings and recommendations. The Prototype was developed to create a single point of contact among state and local governments to test and evaluate mechanisms for government-to-government (G2G) business relationships.
- *Creating and Maintaining Proper Systems for Electronic Record Keeping* (December 2002) E-Government is changing the way government conducts business and captures records created during that business. This paper provides a framework for developing new e-government systems that foster electronic records management.
- *The Future of e-Government* (June 2002) This paper is based on testimony presented to the New York City Council on a sustainable definition and model of electronic government.
- *What Citizens Want from e-Government* (October 2000) Governments in the US are using a variety of methods to find out what citizens want from electronic government services. This report presents those methods, and weighs the pros and cons of each of them.
- *Understanding New Models of Collaboration for Delivering Government Services* (Communications of the ACM, Volume 46, Number 1, January 2003, pp. 40-42) More and more government agencies are creating collaborative relationships to improve services they provide. This article presents a summary of an international research project that is studying eleven collaborative partnerships developed to deliver government information.

- *Electronic Government: A Vision of the Future that is Already Here* (Syracuse Law Review, Volume 52, Number 4, 2002, 1243-1251) Though they may be going unnoticed, e-Government initiatives are changing the way that the public sector works. This article introduces a four-faceted vision of e-Government and describes some of the ways that it is already changing government.
- *Realizing the Promise of Digital Government* (IMP Magazine, October 2000) Many of us have already experienced the potential of the Web to change our relationships with other individuals, businesses, and now government. This article discusses the transformation needed before we can realize the promises of electronic government.
- *Implications of Legal and Organizational Issues for Urban Digital Government Development* (Government Information Quarterly, Volume 18, 2001, 269-278) Legal and organizational issues converge when developing digital government in large urban settings. This paper contends that this convergence is a powerful determinant of how these projects develop and how likely they are to succeed.

The UN Global e-Government Survey 2004

http://www.unpan.org/egovernment4.asp

The UN Global e Government Survey 2004 presents a comparative ranking of the countries of the world according to two primary indicators: the state of e-government readiness; and the extent of e-participation. Constructing a model for the measurement of digitized services, the Survey assesses the 191 member states of the UN according to a quantitative composite index of e-Government readiness based on website assessment; telecommunication infrastructure and human resource endowment. As countries progress in both coverage and sophistication of their state-provided e-service and e-product availability they are ranked higher according to a numerical classification corresponding to the five stages of Emerging presence, Enhanced presence; Interactive presence; Transactional presence and Networked presence.

PUBLICATIONS ON SECURITY OF INFORMATION

BSI's IT-Baseline Protection Manual

http://www.bsi.bund.de/gshb

http://www.it-grundschutzhandbuch.de

Standard security measures are published in the BSI's IT Baseline Protection Manual. The online version could be found at the above sites.

CSI/FBI Computer Crime and Security Survey

http://www.gocsi.com/forms/fbi/csi_fbi_survey.jhtml

http://i.cmpnet.com/gocsi/db_area/pdfs/fbi/FBI2004.pdf

The Computer Crime and Security Survey is conducted by the Computer Security Institute (CSI) with the participation of the San Francisco Federal Bureau of Investigation's Computer Intrusion Squad. The survey is conducted by Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn and Robert Richardson. The survey is now in its ninth year and is, according to the authors, the longest-running survey in the information security field. This year's survey results are based on the responses of 494 computer security practitioners in U.S. corporations, government agencies, financial institutions, medical institutions and universities.

ISQ Handbook - Information Security Qualifications

http://www.isn.ethz.ch/researchpub/publihouse/misc/isq/isq_handbook.pdf

A handbook written by Anna Hess Sargsyan and Edgar Danielyan. The Information Security Qualifications Handbook is a joint initiative between the ISN and its Cooperating Partner Danielyan Consulting LLP. The Handbook offers an in-depth coverage of existing information security qualifications with an objective to introduce EAPC/PfP information security professionals to the most appropriate certification programs for a sustainable education in information security. The ISQ Handbook is a follow-up to the first ISN Partnership for Peace Seminar on Information Security held at the ETH Zurich in August 2003. The Handbook is an integral part of the ISN's continuing efforts to promote and enhance information security through training and education within the framework of Switzerland's commitment to fighting asymmetric threats throughout the Euro-Atlantic Partnership Council region.

Implementing Information Security: Risks vs. Cost

http://www.cyberguard.info/news_room/e_newsletter_archives.cfm#

Published in CyberGuard E-Newsletter, June 2004. As a security professional who understands how the business world works, Gideon T. Rasmussen (CISSP, CISM, CFSO, SCSA) wrote this article to convey the imperative need for security professionals and senior management to see eye-to-eye. Being motivated by business, senior management focuses on productivity and the bottom line. It is sometimes difficult to calculate a return on investment for security, but the damage caused by the absence of efficient controls is far greater than the cost of implementing them.

Creating Security for e-Government Services

http://www.checkpoint.com/promoforms/ww/2004/gov2004ww03.html

The site provides free download of "Creating Security for e-Government Services" white paper that helps the reader to learn about U.S. federal laws and guidelines and how to implement stronger security for e-Government services with COTS products.

Protect Privacy and Security: Imperative 5 of "Eight Imperatives for Leaders in a Networked World"

http://www-1.ibm.com/industries/government/ieg/pdf/eightImperative.pdf

A series produced by the Harvard Policy Group on Network-Enabled Services and Government.

E-Government Security for Managers

http://www.state.ga.us/gta/tips/bill_spernow.pdf

By Bill Spernow, CISSP, Georgia Student Finance Commission; December 4, 2001; Introducing INFOSEC at a State Government Agency: Lessons Learned.

Security Concepts and Requirements of e-Government Sites and other Public Electronic Processes

http://falcon.ifs.uni-linz.ac.at/research/phd_hof/

From this site the PhD thesis of Sonja Hof can be downloaded. The thesis describes the concept of e-Government and as an example the idea of one-stop Government. Additionally, some of the newer and more innovative methods to achieve security are presented. After pointing out the security requirements, some security solutions are listed and analyzed whether they are useful for an e-Government implementation.

E-Government Case Studies: United Kingdom: e-Government Security

http://unpan1.un.org/intradoc/groups/public/documents/un/unpan008830.pdf

Security – U.K. e-Government Strategy Framework Policy and Guidelines

http://www.govtalk.gov.uk/documents/security_v4.pdf

This document sets out a framework for the expression of security requirements for the procurement and acceptance of e-Government services and their implementation. It also describes the approach to assuring the presence and proper operation of the security countermeasures put in place to meet the security requirements. This is Version 4.0 of the document, published on 7/11/2002.

Government Security and Authentication Guidelines

http://www.becta.org.uk/page_documents/corporate/events/expert/Sharon_Wiltshire.pdf

By Sharon Wiltshire, Office of the e-Envoy, Becta Security Seminar, 15th October 2002, Coventry.

The ISF's Standard of Good Practice for Information Security

http://www.isfsecuritystandard.com

The Standard of Good Practice for Information Security is designed to help any organization, irrespective of market sector, size or structure, keep the risks associated with its information systems within acceptable limits. The Standard has been produced by the Information Security Forum. The Standard has been developed using a proven methodology to produce the international benchmark for information security. The Standard is updated every two years, refining proven practices and addressing 'hot topics', such as intrusion detection, information privacy, effective security awareness, e-mail, broadband, wireless communications and PDAs.

BOOKS ON E-GOVERNMENT AND SECURITY OF INFORMATION

Building the Virtual State: Information Technology and Institutional Change

http://www.brook.edu/press/books/virtual_state.htm

A book by Jane E. Fountain, Brookings Institution Press, 2001. The book finds that many issues involved in integrating technology and government have not been adequately debated or even recognized. Drawing from a rich collection of case studies, the book argues that the real challenges lie not in achieving the technical capability of creating a government on the web, but rather in overcoming the entrenched organizational and political divisions within the state. Questions such as who pays for new government websites, which agencies will maintain the sites, and who will ensure that the privacy of citizens is respected reveal the extraordinary obstacles that confront efforts to create a virtual state.

Managing Information and Knowledge in the Public Sector

http://www.amazon.com/exec/obidos/tg/detail/-/0415204232/qid=1105969300/sr=1-1/ref=sr_1_1/103-8350796-3556619?v=glance&s=books

A book by Eileen M. Milner. For the public sector, which is globally the largest employer of people and repository of information, managing information and knowledge is an extremely problematic area to address. The essence of both resources is that they are intangible, their impact and value cannot be measured through traditional accounting methods, yet they are also where the greatest value and potential for improvement is located.

Electronic Governance: An International Comparative Study

http://www.amazon.com/exec/obidos/tg/detail/-/0850927676/qid=1105987495/sr=1-2/ref=sr_1_2/103-8350796-3556619?v=glance&s=books

A book by Thomas B. Riley and Rogers Okot-Uma, 2005.

Building a Digital Community: A Leadership Guidebook

http://www-1.ibm.com/industries/government/ieg/pdf/guidelines.pdf

A product of the e-Communities Task Force provides communities with a strategy for utilizing the Internet and communications technologies to improve quality of life and economic vitality.

Delivering the Vision: Public Services for the Information Society and the Knowledge Economy

http://www.amazon.com/exec/obidos/tg/detail/-/0415241553/qid=1105969300/sr=1-2/ref=sr_1_2/103-8350796-3556619?v=glance&s=books

A book by Eileen M. Milner (Editor). The book explores the way in which public service ‘visions’ have developed globally and how successful they have been in contributing to major social and economic change. Contributions focus both on those factors critical to success and on reasons for failure, but a common theme to emerge across all contributions is the requirement for a clear political vision, commitment and leadership if the shift from traditional forms of social and economic organization to high-value, knowledge-intensive economies is to be safely negotiated.

Electronic Governance and Electronic Democracy: Living and Working in the Wired World

http://www.amazon.com/exec/obidos/tg/detail/-/B0006E92DQ/qid=1105987495/sr=1-6/ref=sr_1_6/103-8350796-3556619?v=glance&s=books

A book by Thomas B Riley from Information Technology & Globalization series.

RELATED CONFERENCES, WORKSHOPS AND COURSES

International Conference on e-Government (27-28 October 2005, Lord Elgin Hotel, Ottawa, Canada)

http://www.academic-conferences.org/iceg2005/iceg2005-home.htm

Not only is interest in e-Government continuing to grow and extend within the public service sector, but new issues—such as e-Democracy (including e-Voting) and e-Politics—are evolving. These issues are relevant to government at all levels including local, provincial, central or federal government and also at the supranational level such as the European Union. The interest in e-Government is, at least in part, driven by an agenda to radically transform the delivery of public services through the adoption of advanced information and communications technology (ICT) to make the whole process of government more effective.

5th European Conference on e-Government (16-17 June 2005, University of Antwerp, Belgium)

http://www.academic-conferences.org/eceg2005/eceg2005-home.htm

e-Government is clearly not only about technology – it is about reinventing the way in which public sector service providers and citizens interact. It is about enhancing the democratic processes and also about using new ideas to make lives easier for the citizen by transforming government processes, providing community leadership, enabling economic development and renewing the role of government itself in society.

ECMLG 2005: The European Conference on IS Management, Leadership and Governance (7-8 July 2005, Blackhorse House, Reading University, Reading, UK)

http://www.academic-conferences.org/ecmlg2005/2-ecmlg2005-home.htm

The Conference offers an opportunity for scholars and practitioners interested in the issues related to Management, Leadership and Governance, especially as it relates to the information systems field, to share their thinking and research findings. These fields of study are broadly described as including issues related to the management of information systems resources, the interface between CIOs and CEOs, the formal governance of the information systems function, the appointment and responsibility of the information systems director and the relationship between information systems management and the rest of the organization. This Conference provides a forum for discussion, collaboration and intellectual exchange for all those interested in any of these fields of research or practice.

ECIW 2005: The 4th European Conference on Information Warfare and Security (11-12 July 2005, University of Glamorgan, UK)

http://www.academic-conferences.org/eciw2005/eciw2005-home.htm

The Fourth European Conference on Information Warfare and Security (ECIW) is an opportunity for academics, practitioners and consultants from Europe and elsewhere who are involved in the study, management, development and implementation of systems and concepts to combat information warfare or to improve information systems security to come together and exchange ideas. There are several strong strands of research and interest that are developing in the area including the understanding of threats and risks to information systems, the development of a strong security culture, as well as incident detection and post incident investigation.

IEEE Symposium on Security and Privacy (8-11 May 2005, Berkeley/Oakland, California)

http://www.ieee-security.org/TC/SP-Index.html

Since 1980, the IEEE Symposium on Security and Privacy has been the premier forum for the presentation of developments in computer security and electronic privacy, and for bringing together researchers and practitioners in the field. Papers offer novel research contributions in any aspect of computer security or electronic privacy. Papers may represent advances in the theory, design, implementation, analysis, or empirical evaluation of secure systems, either for general use or for specific application domains.

IDABC: CROSS-BORDER E-GOVERNMENT SERVICES for Administrations, Businesses and Citizens

http://europa.eu.int/idabc/en/chapter/5606

The conference “IDABC: CROSS-BORDER E-GOVERNMENT SERVICES for Administrations, Businesses and Citizens” will take place at Centre Borschette in rue

Froissart in Brussels, Belgium on 17 and 18 February 2005. It will be open to participants from all sectors of society and address in particular associations representing citizens or industry. Participants from the public sector are also encouraged to attend since the objective of the Conference is to increase the awareness and understanding of the services citizens and businesses expect from the European public sector and to help identify the obstacles and challenges that need to be tackled to provide these. The conference discussions will help orientate and prioritise the IDABC programme's activities in support of pan-European e-Government services.

Second Conference on Email and Anti-Spam (CEAS 2005)

http://www.ceas.cc/

The conference will be held on 21 and 22 July 2005 at Stanford University, Palo Alto, CA. It is organized in cooperation with The International Association for Cryptologic Research and The IEEE Technical Committee on Security and Privacy.

18th IEEE Computer Security Foundations Workshop (20-22 June 2005, Aix-en-Provence, France)

http://www.lif.univ-mrs.fr/CSFW18/

For nearly two decades, the Computer Security Foundations Workshop has brought together a small group of researchers to examine foundational issues in information security. Many seminal papers and techniques were first presented at CSFW. The interest is in both new theoretical results in computer security and also in more exploratory presentations that examine open questions and raise fundamental concerns about existing theories. Panel proposals are welcome as well as papers.

CSI 32nd Annual Computer Security Conference

http://www.gocsi.com/annual/

The CSI 32nd Annual Computer Security Conference & Exhibition will be held on 14-16 November 2005 in Washington, D.C. at the Marriott Wardman Park.

25th Annual International Cryptology Conference CRYPTO 2005 (14-18 August 2005, Santa Barbara, California, USA)

http://www.iacr.org/conferences/crypto2005/

CRYPTO 2005 will be held at the University of California, Santa Barbara. The academic program covers all aspects of cryptology. Formal proceedings, published by

Springer-Verlag, will be provided to registered attendees at the conference. CRYPTO 2005 is sponsored by the International Association for Cryptologic Research (IACR), in cooperation with the IEEE Computer Society Technical Committee on Security and Privacy, and the Computer Science Department of the University of California, Santa Barbara.

The NetSec'05 Conference

http://www.lif.univ-mrs.fr/CSFW18/

CSI NetSec is industry network security conference. NetSec combines management topics with a technical focus to bring you cutting edge strategies and solutions. NetSec 2005 will be held on 13-15 June 2005 at The Phoenician in Scottsdale, Arizona.

12th Annual Fast Software Encryption Workshop (ENSTA, Paris, France, 21-23 February 2005)

http://crypto.rd.francetelecom.com/fse2005/

FSE 2005 is the 12th annual Fast Software Encryption workshop, sponsored for the fourth year by the International Association for Cryptologic Research. It focuses on symmetric cryptology and more precisely, on efficient design and cryptanalysis of block and stream ciphers, as well as hash functions and message authentication codes.

EUROCRYPT 2005 (22-26 May 2005, Aarhus, Denmark)

http://www.brics.dk/eurocrypt05/

Eurocrypt 2005 is a scientific conference that focuses on research in cryptology. It is organized by the International Association for Cryptologic Research (IACR) in cooperation with the Computer Science Department of University of Aarhus.

8th Information Security Conference (20-23 September 2005, Singapore)

http://isc05.i2r.a-star.edu.sg/

ISC is an annual international conference covering research and applications on information security. Continuing its past success, the conference aims to attract high quality papers in all technical aspects of information security. The conference proceedings will be published by Springer in its LNCS series. Selected papers in the proceedings will be invited for submission to a special issue of International Journal of Information Security.

Tenth Australasian Conference on Information Security and Privacy (4-6 July 2005, Brisbane, Australia)

http://www.isrc.qut.edu.au/events/acisp2005/

Original papers pertaining to all aspects of information security and privacy are solicited for submission to the 10th Australasian Conference on Information Security and Privacy (ACISP 2005). Papers may present theory, techniques, applications and practical experiences on a variety of topics.

Workshop on Cryptographic Hardware and Embedded Systems (29 August – 1 September 2005, Edinburgh, Scotland)

http://islab.oregonstate.edu/ches/

The focus of this workshop is on all aspects of cryptographic hardware and security in embedded systems. The workshop will be a forum of new results from the research community as well as from the industry. Of special interest are contributions that describe new methods for efficient hardware implementations and high-speed software for embedded systems, e.g., smart cards, microprocessors, DSPs, etc.

IACR Calendar of Events in Cryptology

http://www.iacr.org/events/

The IACR calendar lists events (conferences, workshops, ...) that may be of interest to IACR members or deal with research in cryptology.

RELATED NEWSLETTERS, MAGAZINES AND JOURNALS

E-GOVERNMENT RELATED JOURNALS

Journal of e-Government

http://www.egovjournal.com/

The Journal of e-Government is a new professional journal focusing on the application and practice of e-Government in its broadest sense. The Journal of e-Government is now officially affiliated with the Information Technology and Politics (ITP) section of the American Political Science Association (APSA). Papers may be on any aspect of e-Government, ranging from local to national and international initiatives and developments. Articles from practitioners of e-Government and industry experts as well as researchers in the field are welcomed. The editorial board seeks a broad range of case studies, research articles, reviews, and industry notes relevant to the evolving field of e-Government. Submissions are encouraged from all allied fields, including government and public administration, political science, public health, justice and law enforcement, communications, public finance, and other fields related to the use of information and communication technology in the public sector.

Electronic Journal of e-Government

http://www.ejeg.com/

The Electronic Journal of e-Government aims to publish perspectives on topics relevant to the study, implementation and management of e-Government. The journal contributes to the development of both theory and practice in the field of e-Government. The journal accepts academically robust papers, topical articles and case studies that contribute to the area of research in, and practice of e-Government.

eGovAustralia.com

http://www.egovaustralia.com/

This is a new electronic magazine and information site covering the latest news and progress in Online Government. It features developments within both the Australian government sector and governments globally. It is an independent publication for leaders in policy, management, business development, marketing, media and IT. Subscription to eGovAustralia.com is free.

E-Government Bulletin

http://www.headstar.com/egb/

This is the website of an email service covering electronic government, teledemocracy and the information society in the UK and worldwide. The Bulletin is a free, independent publication, aimed at internet users across government, local government, the social sector and their private sector partners.

PublicTechnology.net – e-Government & IT News for the UK Public Sector

http://www.publictechnology.net/

PublicTechnology.net delivers IT and e-Government news and information for the UK public sector, focusing on IT and e-Government projects, targets, UK and EU regulations, IT best practice and case studies. Plus: UK comprehensive UK public sector directory.

Government Technology Magazine

http://www.govtech.net/

This is the Web site for the magazine. They have an e-Government section. The user can subscribe to a newsletter to receive e-mail updates. Focus is on federal, state, and big cities but they occasionally have articles relevant to small and medium-sized communities.

Government Computing Magazine

http://www.kablenet.com/kgc.nsf/WebPagesFrontPage/gcAbout#?expandtab=lyrKB5

Government Computing magazine is read by public sector decision makers concerned with information management and business process change. It looks at how technology can improve the delivery of public services, rather than concentrating on the technical details of the IT. Government Computing was launched in 1996 and is the longest running monthly public sector IT title. Over 60% of readers are in central and local government, whilst the rest are from executive agencies, justice, education, health and defense. Nearly 50% of its readers are top officials (chief executives, ministers, directors and deputies), 34% are senior managers (heads of department and IT managers) and the remainder is managers. The main topics covered in the magazine include: e-Content, Data protection, Freedom of information, Data storage, Communications channels, Flexible working, CRM, IT security, Integration and migration, GIS, Back office systems, Knowledge management, Business performance management, Training, Document and record management, e-Procurement, and Authentication.

Governing Magazine

http://www.governing.com/

Governing is a monthly magazine whose primary audience is state and local government officials: governors, legislators, mayors, city managers, council members and other elected, appointed and career officials. They are the men and women who set policy for and manage the day-to-day operations of cities, counties and states, as well as such governmental bodies as school boards and special districts.

SECURITY OF INFORMATION RELATED MAGAZINES, NEWSLETTERS, NEWS SITES AND JOURNALS

International Journal of Information Security

http://link.springer.de/link/service/journals/10207/index.htm

The International Journal of Information Security is an English language international journal on research in information security. Information security builds on computer security and cryptography, but also reaches out to other branches of the information sciences. Information security is an important aspect of protecting the information society from a wide variety of threats. In this new century, The International Journal of Information Security will provide prompt publication of important technical work in information security, whether theoretical, applicable, or related to implementation. The scope of the International Journal of Information Security (IJIS) is theory, applications, and implementations of information security. This includes, but is not limited to:

- system security – intrusion detection, secure end systems, secure operating systems, database security, security infrastructures, security evaluation
- network security – Internet security, firewalls, mobile security, security agents, protocols, anti-virus and anti-hacker measures
- content protection – watermarking, software protection, tamper resistant software
- applications – electronic commerce, electronic government, health, telecommunications, mobility
- foundations – privacy, access control, authentication, identification, cryptography, steganography, formal methods in information security

Journal of Cryptology

http://www.iacr.org/jofc/jofc.html

http://link.springer.de/link/service/journals/00145/index.htm

Journal of Cryptology provides a forum for original results in all areas of modern information security. Both cryptography and cryptanalysis are covered, including information theoretic and complexity theoretic perspectives as well as implementation, application, and standards issues. Illustrative topics include public key and conventional algorithms and their implementations, cryptanalytic attacks, pseudo-random sequences, computational number theory, cryptographic protocols, untraceability, privacy, authentication, key management and quantum cryptography. In addition to full-length technical, survey, and historical articles, short notes are acceptable.

ACM Transactions on Information and System Security (TISSEC)

http://www.acm.org/pubs/tissec/

TISSEC is a scholarly, scientific journal that publishes original research papers in all areas of information and system security, including technologies, systems, applications, and policies. Topics of Interest

- Security Technologies: authentication; authorization models and mechanisms; auditing and intrusion detection; cryptographic algorithms, protocols, services, and infrastructure; recovery and survivable operation; risk analysis; assurance including cryptanalysis and formal methods; penetration technologies including viruses, Trojan horses, spoofing, sniffing, cracking, and covert channels.
- Secure Systems: secure operating systems, database systems and networks; secure distributed systems including security middleware; secure web browsers, servers, and mobile code; specialized secure systems for specific application areas; interoperability, and composition.
- Security Applications: threats, system tradeoffs, and unique needs of applications; representative application areas include information systems, workflow, electronic commerce, electronic cash, copyright and intellectual property protection, telecommunications systems, wireless systems, and health care.
- Security Policies: confidentiality, integrity, availability, privacy, usage, and survivability policies; tradeoffs, conflicts and synergy among security objectives.

Journal of Computer Security

http://www.iospress.nl/html/0926227x.html

http://www.csl.sri.com/programs/security/jcs/

The Journal of Computer Security presents research and development results of lasting significance in the theory, design, implementation, analysis, and application of secure computer systems. It also provides a forum for ideas about the meaning and implications of security and privacy, particularly those with important consequences for the technical community. The journal welcomes contributions on all aspects of computer security: confidentiality, integrity, and assurance of service - that is, protection against unauthorized disclosure or modification of sensitive information, or denial of service. Of interest is a precise understanding of security policies through modelling, as well as the design and analysis of mechanisms for enforcing them, and the architectural principles of software and hardware systems implementing them.

Cipher Newsletter

http://www.ieee-security.org/cipher.html

A Newsletter of the IEEE Committee on Security & Privacy provides a wide range of information of current news in the field.

IEEE Transactions on Dependable and Secure Computing

http://www.computer.org/tdsc/

The IEEE Transactions on Dependable and Secure Computing publishes archival research results related to research into foundations, methodologies, and mechanisms that support the achievement—through design, modeling, and evaluation—of systems and networks that are dependable and secure to the desired degree without compromising performance. The focus will also include measurement, modeling, and simulation techniques, and foundations for jointly evaluating, verifying, and designing for performance, security, and dependability constraints.

Computer Fraud & Security

http://www.sciencedirect.com/science/journal/13613723

Computer Fraud & Security focuses on providing practical, usable information to effectively manage and control computer and information security within commercial organizations. Areas regularly covered include: audit and financial control methodologies, data encryption, risk management, network security, contingency planning and disaster recovery, access control, security software and software protection, authentication and validation, virus reports, and e-commerce security.

Computers & Security

http://www.sciencedirect.com/science/journal/01674048

Computers & Security is the official journal of Technical Committee 11 (computer security) of the International Federation of Information Processing. It is one of the most respected technical journals in the IT security field. The journal provides a unique blend of leading edge research and sound practical management advice. It is aimed at the professionals involved with computer security, audit, control and data integrity in all sectors – industry, commerce and academia.

IEEE Security and Privacy Magazine

http://www.computer.org/security/

Denial of service, worms, DNS, and router attacks are increasing. To help in staying one step ahead of these and other threats, the IEEE Computer Society has published a new periodical in 2003, IEEE Security & Privacy magazine. IEEE Security & Privacy will rethink the role and importance of networked infrastructure and help in developing lasting security solutions. Topics covered include: Wireless Security; Securing the Enterprise; Designing for Security; Infrastructure Security; Privacy Issues; Legal Issues; Digital Rights Management; Cybercrime; Intellectual Property Protection, and Piracy; The Security Profession; and Education. The primary objective of IEEE Security & Privacy is to stimulate and track advances in information assurance and security and present these advances in a form that can be useful to a broad cross-section of the professional community-ranging from academic researchers to industry practitioners. It is intended to serve a broad readership.

CyberGuard E-Newsletter

http://www.cyberguard.com/

Journal of Privacy Technology

http://www.jopt.org/

The Journal of Privacy Technology is a refereed online journal published by the Privacy Technology Center within the Institute for Software Research International, a division of the School of Computer Science at Carnegie Mellon University in Pittsburgh, Pennsylvania. The Journal is a forum for publication of current research in privacy technology. It will consider any material dealing primarily with the technological aspects of privacy or with the privacy aspects of technology, which may include analysis of the interaction between policy and technology or the technological implications of legal decisions.

Network Security

http://www.sciencedirect.com/science/journal/13534858

Network Security is devoted to solving network security problems in system-specific detail. Every month Network Security covers: worldwide news, regular industry columns, authoritative news and analysis on the major networks and their operations and the impact on the organizations, in-depth technical feature articles, regular case studies, legal brief and Cybernet features, highlights from recent conferences, book reviews and a comprehensive calendar of events.