INFORMATION & SECURITY

An International Journal

Volume 18, 2006

Cybercrime and Cybersecurity

Edited by
Petya Ivanova

ProCon Ltd., Sofia 2006

Content *I&S*

SECURING CYBERSPACE

Advances in information and communications technologies are transforming our economies and societies. They have formed the basis for global economic growth and an increase in the standard of living. Profound changes can be seen in industry, government and trade, in work, education and leisure. In all aspects of our life we rely on information technology.

At the same time, the information and communications infrastructures have become a critical part of national economies. The other critical infrastructures—those assets, systems, and functions vital to national security, economic need, or public health and safety—encompass a number of sectors, including many basic necessities, such as food, water, public health, emergency services, energy, transportation, banking and finance, and postal services and shipping. All of them increasingly rely on information and communications infrastructures for their operation. Many of the infrastructures' networks are also connected to the Internet.

Unfortunately, the information and communications infrastructures have their own vulnerabilities and offer new opportunities for criminal activity and indeed new forms of crime. The criminal activities may take a large variety of forms and may cross many borders. Criminals, terrorists, and malicious users have exploited the anonymity and global reach of the Internet to launch attacks on the information infrastructure; put harmful and illegal content on the Internet; perform reconnaissance for physical attack; steal money, identities, and secrets; conduct hostile information operations; etc.

The information infrastructure is unique among the critical infrastructures because it is owned primarily by the private sector, it changes rapidly with the fast changes in the information technology area, and, as already mentioned, it is the backbone for many other critical infrastructures. The increasing reliance of critical infrastructures on networks and the Internet has increased the risk of cyber attacks that could harm the infrastructures. In many respects the threat of cyber attacks is escalating due to the increased availability of automated tools for malicious actions, the complexity of the technical environment, and the increased dependence of our society on interconnected systems. Therefore, protection of the information and communications infrastructures is vitally important.

Cybersecurity refers to the defense against attacks on information infrastructure. Cybersecurity has been a major concern of both governments and private sector for many years already. Worldwide, agencies such as the European Network and Information Security Agency, the US Department of Homeland Security, the Council of European Cybercrime Convention, Australia's Critical Infrastructure Protection Group, and the Asia-Pacific Economic Cooperation have set forth policies and initiatives to enhance the information security within their regions. Despite their continuous efforts, the new information and communications technologies has given rise even to more forms of computer-related crime, which pose threats not only to the confidentiality, integrity, or availability of computer systems, but also to the security of critical infrastructure. The same technologies that enabled this unusual growth and now underpin our economy and way of life also increase the vulnerability of the information and communications infrastructure.

Certainly, there is scope for action both in terms of preventing criminal activity by enhancing the security of information infrastructures and by ensuring that the law enforcement authorities have the appropriate means to act. With the annual costs of digital attacks and financial losses rising to billions of dollars, what measures should be taken to minimize the impact these cyber breaches will have on the global economy? What practices should be implemented to address the daily occurrences of spam, spyware, DOS attacks, online extortion, IP theft, viruses, worms, and physical and cyber data breaches?

With the intention to highlight the importance of cybersecurity and the fight against cybercrime, the Editorial Board of *Information & Security: An International Journal* (I&S) decided to prepare a special I&S issue on vulnerabilities of modern information and communications infrastructures and the search for higher levels of cybersecurity. The objective was to promote research and development to understand and reduce vulnerabilities and to stimulate the dissemination of know-how.

The first article in this volume tries to provide a theoretical foundation for the subsequent discussion. Based on the relativity of the concept of cybersecurity, Xingan Li analyzes the economic impact of cybersecurity breaches. The author then argues that cybersecurity is a private good and should be provided mainly by the private sector. Regarding cybersecurity as a public good would discourage the private sector to invest in security provision, comments the author. However, in terms of prevention of security breaches, law enforcement can play an important role in establishing and enforcing liability mechanisms. Although it is still controversial whether and how cybersecurity players should be held liable for their activities, every step made in this direction will bring benefits, Xingan Li concludes.

Any payment system is characterized by a high level of risk in its different domains caused by great volume and number of operations, a lot of complex relations between clients and increasing speed of data transmission. Nowadays, one of the most important and challenging problems for the financial institutions becomes credit card fraud. Fraudulent electronic transactions have already been a significant problem that grows in importance with the expansion of modern information and communications technologies and the growth of the Internet into a global economic force. Fraud prevention and detection methods are being continuously improved; however, banks are losing billions of dollars worldwide each year.

Not only banks lose money because of credit card fraud. Cardholders also pay for the loss through higher interest rates, higher membership fees, and reduced benefits. Hence, it is in the interest both of the banks and cardholders to reduce illegitimate use of credit cards. And to the financial cost to cardholders one should also add the personal cost in time, inconvenience and frustration while an incident is investigated.

Credit card fraud continues to be a growing problem for Internet businesses. It is in a company and card issuer's interest to prevent fraud or, failing this, to detect fraud as soon as possible. Otherwise consumer trust in both the card and the company decreases and revenue is lost, in addition to the direct losses made through fraudulent sales.

Further, fraud on credit and debit cards has also a high cost to society as the proceeds are often used to fund serious organized crime such as drug trafficking and terrorism.

Considering the importance of the problem, the next group of articles discusses the nature of fraud and mechanisms for its detection.

The aim of Krzysztof Woda's article is to analyze the possible modern techniques for money laundering and terrorism financing, which can be carried out with electronic payment systems. The article identifies the most important characteristics of the particular payment systems, which predetermine such systems as especially suitable for illicit activities. Finally, Krzysztof Woda presents solutions to reduce the risk of illicit money transfers with electronic payment systems and puts a special emphasis on the development of reliable methods for detecting illegal money operations and financial computer crime.

Although introducing techniques for prevention is the most efficient way to reduce fraud, fraudsters are adaptive and, given time, will typically find ways to circumvent such measures. Methodologies for the detection of fraud are of great importance once fraud prevention has failed.

The artificial intelligence community constantly provides new technologies and solutions for fraud detection that have been already applied successfully to detect illegal

activities. Fraud detection, however, requires a tool that is intelligent enough to adapt to criminals' strategies and ever changing tactics to commit fraud. To address this need, the second article in this group proposes a new approach to transaction monitoring and credit card fraud detection. Vladimir Zaslavsky and Anna Strizhak attempt to develop a framework for unsupervised fraud detection based on the neural network technology. The authors apply the Self-Organizing Map algorithm to create a model of typical cardholder's behavior and to analyze the deviation of transactions, thus finding suspicious transactions. It enables automated creation of transaction monitoring rules in a learning process and makes possible their continuous improvement in an environment of dynamically changing information in an automated system.

In the ongoing quest for cybersecurity, cryptography plays an increasingly important role. Given this truth, the following two articles in this special issue of *Information and Security* address various security challenges, proposing novel efficient techniques.

With the rapid development of communications and information technologies, Oblivious Transfer (OT) has been widely applied in numerous applications and has become an important cryptography tool. The mechanism of the t-out-of-n OT protocol is a novel and significant version of the general OT protocol. In 2004, researchers proposed a new secure t-out-of-n OT protocol, which after a thorough analysis has shown to lack efficiency. In their article, Jung-San Lee and Chin-Chen Chang propose a novel t-out-of-n OT protocol based on the Generalized Chinese Remainder Theorem. As demonstrated by the authors, the proposed OT protocol not only satisfies the three essential properties of the general OT protocols, but also has better performance than related protocols. The authors further claim that the proposed t-out-of-n OT protocol is secure and efficient enough to be applied in real-world applications.

Advances in networking technologies and the continued growth of the Internet have also triggered another trend – towards outsourcing data management and information technology needs to external service providers. As a recent manifestation of this trend, there has been growing interest in outsourcing database services in both the commercial world and the research community. However, this introduces numerous security research challenges. Despite a large number of research activities done for securing outsourced databases and removing unencrypted data from exposure to the external server and other intruders, no work has been able to radically secure outsourced databases with associated indexes during the query execution. By exploiting such indexes and with relevant available knowledge, attackers can infer confidential information from the outsourced encrypted data. The article by Tran Khanh Dang discusses potential attacks in such situations and introduces two security protocols for outsourcing database services. The main contributions focus on solutions to the problem of data privacy/confidentiality and user privacy. The theoretical analyses

show that the proposed protocols can effectively protect outsourced data and its associated indexes as well as the clients against various sophisticated attacks.

Nowadays, we can transfer money electronically and shop using e-commerce applications. It could be predicted that, with network support, more activities will be performed without the need for a face to face contact. Hence, authentication has become one of the most significant and challenging issues in Internet applications. The following two articles provide a comprehensive treatment of this very important subject – remote authentication.

Password-based remote authentication is one of the most commonly used authentication techniques due to its simplicity and effectiveness. Authentication schemes generally use a password/ verification table stored at the server side. This stored-table system can easily suffer from verifier-stolen or modification attacks. Clearly, a more secure way to verify user legitimacy is required. Therefore, ID-based authentication schemes have been proposed to remove the requirement of having a password/ verification table stored on the server.

Among numerous schemes for protection, the remote password authentication schemes using smart cards are regarded as very efficient. As a result, smart-card based authentication schemes has become a popular research topic in recent years. In 2000, Hwang and Li proposed a new remote authentication scheme using smart cards based on ElGamal's cryptosystem. The main advantage is that a password table is not required to verify a user's legitimacy. Unfortunately, several security flaws have been identified in their method.

As cryptanalysis has evolved, however, a series of modifications that improve the known security flaws have been made subsequently. The article by Tzung-Her Chen, Du-Shiau Tsai, and Gwoboa Horng deals with a security problem found in a latest modification and improves it in order to construct a more secure version. The article also highlights a feature, mutual authentication, between a server and users found in many authentication protocols but seldom found in the considered series of modifications. Compared with other related schemes, the proposed schemes provide higher security. The authors have demonstrated that the proposed schemes are reliable and secure.

The next article by Chin-Chen Chang and Jung-San Lee proposes a novel practical and secure remote password authentication scheme that also overcomes the security weaknesses of Hwang-Li's scheme. The proposed scheme provides mutual authentication between the remote system and the user such that the server spoofing attack cannot have an effect. Mutual authentication is an essential requirement in remote password authentication schemes. Besides, the scheme allows the user to choose and change passwords at will and can resist the replay attack without sophisticated con-

current mechanisms. Since the computational load of both the smart card and the whole system is quite low, the remote authentication scheme proposed by Chin-Chen Chang and Jung-San Lee is efficient, secure, and user-friendly to be applied in practice; moreover it could be employed on imbalanced networks as well.

Finally, this special issue provides also a comprehensive, up-to-date list with on-line resources on cybersecurity related forums, organizations, research groups, events, as well as some important publications.

The reader will not find answers to all related questions in this issue. We believe, though, that this I&S issue will stimulate the useful analysis and discussion on how to better address the issues of corporate, individual, and national cybersecurity between all the interested parties (law enforcement agencies, Internet Service Providers, telecommunications operators, software vendors, individual and organizational users, consumer representatives, data protection authorities, etc.), with the objective to enhance mutual understanding and cooperation. This issue attempts to raise the awareness of the risks posed by criminals on the Internet, to stimulate the research on cybersecurity, to identify effective counter-crime tools and procedures to fight cybercrime and to encourage further development of early warning and crisis management systems and mechanisms.

Information & Security

CYBERSECURITY AS A RELATIVE CONCEPT

Xingan LI

Abstract: Based on the relativity of the concept of cybersecurity, this article analyzes the economic impact of cybersecurity breaches, identifies cybersecurity as a private good that should be provided mainly by the private sector. However, public provision is also necessary when severe security breaches occur and liability mechanisms should be triggered.

Keywords: Cybersecurity, Illegal Behavior, Economic Analysis.

Introduction

The Internet has become a critical infrastructure for both public and private sectors and has brought new levels of productivity, convenience, and efficiency. The increasing incidents of Internet attacks representing examples of how vulnerable the information systems are, how far the offensive technology outpaces the defensive technology, how easy various malicious programs are created and how smart they can spread all over the Internet rapidly, have started to impact the practical facets of our lives. At the same time, the attackers are able to conceal their attacks by disabling logging facilities or modifying event logs, so their activity goes undetected. Even worse, some automated programs have been designed to specifically disable anti-virus software or penetrate firewalls. The security violations have multi-dimensional impacts on both consumers and businesses, including time, human resources, monetary losses and psychological losses.

The Internet and the larger information infrastructure are not secure.[1] McCormick identified five reasons why Internet is vulnerable: failing to enforce policies, ignoring new vulnerabilities, relying too much on technology, failing to thoroughly investigate job candidates, and expecting too much from technical skills.[2] These risks cause serious insecurity problems in the information society.[3]

While the governments have made efforts to better secure their own computer networks to prevent terrorists from hacking into computer systems, the governments have been increasingly concerned that the private sector is vulnerable to cyberterror-

INFORMATION & SECURITY. An International Journal, Vol.18, 2006, 11-24.

ism. The question being asked is whether private businesses provide enough cybersecurity, or some form of government involvement is justified. Many empirical studies examined the economic impact of cybersecurity breaches. Theories diversify in regarding the cybersecurity as an externality,[4] a public good,[5] or a private good.[6]

Based on the concept of relative cybersecurity, this paper analyzes the economic impact of cybersecurity breaches, whether cybersecurity is a public good or a private good. It also establishes liability mechanism for cybersecurity breaches.

Impact of Cybersecurity Breaches

Increasing Investment of Users in Cybersecurity

The users' investment in cybersecurity takes on the tendency of increasing. Although exact statistics on these expenditures is unavailable, the add-up of global users' financial costs will reach a surprising figure. According to a survey conducted by the Computer Security Institute (CSI) and the Federal Bureau of Investigation (FBI), nearly all of the companies surveyed in 2005 used anti-virus software, firewall, and some measures of access control. Besides the hardware and software, the organizational users also have to employ security personnel or institutions to maintain their systems. These measures induce the increase of the investment of network users. But in fact, security measures can hardly ever be a perfect assurance against damage and accidents. Absolute security becomes too expensive to be reasonable.[7]

Frequent Occurrence of Cybersecurity Breaches

Although the investment in cybersecurity is increasing year by year, the breaches still occur frequently. The potential for information security breaches, as well as the magnitude of potential losses associated with such breaches, has been confirmed by empirical studies.

The annual surveys on information security breaches have pointed out that cybersecurity breaches are ubiquitous. The 2005 survey conducted by CSI and FBI revealed that 56 percent of the surveyed 693 U.S. computer security practitioners acknowledged unauthorized use of a computer in their organization in the last 12 months.[8] CERT Coordination Center reported that the computer security vulnerabilities increased nearly 35-fold during one decade with 171 separate holes reported in 1995 and 5,990 reported in 2005.[9] In the recent years, the publicly disclosed virus attacks are billing the global computer users in an accelerated speed, even though many of the users are unaware of, or unwilling to report the losses.

Increasing Costs of Cybersecurity Breaches

As a consequence of the frequent occurrence of cybersecurity breaches, the losses of these breaches are increasing as well. The losses can be divided into direct and indirect, tangible and intangible, and short-term and long-term. Neumann stated that costs of cybercrime are difficult to measure; however, these costs are reasonably substantial and growing rapidly.[10] Scholars proposed various models to try to measure the costs of security breaches, such as in the Forrester Research. Howe and colleagues' analysis indicated that, if the perpetrators were to unlawfully transfer $1 million from an online bank, the financial influence to the bank would reach $106 million.[11]

The direct losses are those directly involved in the attacks, including interruption of business, destruction of software and hardware, expenditure on recovering the systems, installation and update of security means, recruiting security personnel, etc. The indirect losses are losses indirectly related to the attacks, such as reduction of consumers, decrease of stock prices, etc. The other kinds of losses are also easy to emerge.

The 2005 CSI/FBI survey noted that, of the 639 respondents that were willing and/or able to estimate losses due to security breaches, such breaches resulted in losses close to $130 million.[12] On the other hand, Lukasik claims that cybercrime costs are essentially doubling each year.[13] The problem becomes even more complex when one considers the "black figure" of these crimes. Ullman and Ferrera mentioned that, according to FBI estimates, only 17 percent of computer crimes are reported to government authorities.[14]

Relativity of the Cybersecurity Concept

There are various answers to the question "What is cybersecurity?" Cybersecurity is a comparative concept. On one hand, it includes the comparison between security and attack techniques. On the other hand, it includes comparison between different security techniques and measures. Considering the comparison between the techniques for security and attack, it is publicly well recognized that the attack techniques develop faster than the security techniques, regardless of the reasons. In other words, the hardware, the software, or the other information system components are always vulnerable and this fact can be exploited. We could call this the absolute level of security. Considering the comparison between the different security techniques, the existence of different environments, the possession of different hardware, software and other equipment, and the adoption of different security techniques, all this leads to difference in the level of security. Therefore, each of the individual or organizational users has a different security level.

Some viewpoints regard cybersecurity as an externality.[15] Camp and Wolfram point out that if a company does a poor job at cybersecurity, other companies may be affected negatively. Thus, the cost is an externality to the owner of the infected machine.[16] However, if we identify cybersecurity as an externality, it is inevitable that to the extent investments in computer security create positive externalities, too little will be provided.

Security is not the reason that drives the attackers to violate security and launch attacks, nor the condition that facilitates the attacks, but the target that the attacks aim at. In fact, there is no clear boundary between security and insecurity. Security and insecurity have only quantitative difference, but no quality distinction. Neither absolute security nor complete insecurity exists. That is to say, security and insecurity should be considered as security between zero percent and 100 percent. Therefore, security is a relative concept. The security of a higher level is security, while the security of a lower level is insecurity.

Although the information systems on the Internet all have a similar framework, they lack any central control system and are uncontrollable. Not only the physical system, but also the operational process is uncontrollable. Thus, to a great extent, the security of the Internet depends on the security measures taken by the end users, either individuals or organizations. However, the security measures of individual and organizational users are widely different due to the difference in hardware, software, and human resources.

The level of security of the end users on the network is different; an absolute value of security does not exist. Security is just a comparison of relative values. It is both the result of comparison between users and the comparison between past and present, i.e., horizontal and vertical comparison. Due to the large number of network users and the rapid change in the network environment, the result of this comparison changes constantly. In general, a higher level security will change quickly into a lower level security (insecurity) with transformation of techniques and the environment. Therefore, the cybersecurity measures have to be updated and renewed timely, frequently, and efficiently.

If the cybersecurity measures cannot be updated and renewed in a timely, frequent and efficient manner, vulnerabilities might occur. Vulnerability is not the security or insecurity themselves, but a factor that makes it impossible to realize perfect security, and an extra loophole caused by the external factors in the investor's production of the expected complete security. It is the natural adversary of the security product, i.e., flaws that can be detected and exploited by the potential attackers to commit harm and cause loss.

Table 1: Classical Division of Goods in Economy.[17]

Classic Division of Goods in Economy		*Exclusion from Consumption*	
		YES	*NO*
Competition in Consumption	*YES*	Private Good: Food, Clothing, Toys, Furniture, Cars…	Common Good: Natural Environment
	NO	Club Good: Private Schools, Cinemas, Clubs…	Public Good: National Security (Army and Police Forces)

Provision of Cybersecurity as a Private Good

In economics, goods are traditionally classified into four categories as listed in Table 1.

Besides other issues, private good and public good can be generally regarded as a pair of opposites. The main features of the private good are excludability and rivalry. According to Samuelson,[18] public good is a good that produces a positive externality and which is characterized by non-rival consumption and non-excludability. The private provision of private goods, or public provision of public goods are not the unique ways in providing these two kinds of goods (let us not consider the other kinds of goods here). The ways of provision of these two kinds of goods can be illustrated as shown in Table 2.

The public good is usually confronted with the problem of being underprovided or not being provided when it is put on the private market. Such a problem appears in providing cybersecurity. Generally, a higher level of cybersecurity would benefit both the individual or organizational owner and users other than the owner. Because insecure computers are vulnerable to be manipulated to launch attacks against other computers, it is reasonable to assume that if an owner maintains a higher level of cybersecurity, the other users' computers may experience a lesser risk of being attacked. Then the other users would have the good reason to reduce their investment in security protection. The computer users' security provision only diminishes the probability of the others' computers being attacked. However, since individuals are not generally liable for the damage caused when a hacker uses their computer, they do not benefit from the increased security.[19] And because users with ability to provide security do not benefit, they will fail to provide it. The same applies to other computer

Table 2: Ways of Provision of Private Goods and Public Goods.

Different Ways of Provision of Different Goods	*Private Provision*	*Government Provision*	*Mixed Provision*
Private Goods	Clothes, Food, Cars, Private Housing	Food Supply as in Communist China in the End of 1950s	Transportation, Medical Care
Public Goods	Foreign Aid	National Defense	Pollution Reduction

owners, and, therefore, everybody is in a worse situation than would be if everyone provided the security that would have spillover benefits for everyone else.

As we have seen, cybersecurity is both excludable and rivalrous. Cybersecurity has neither territorial boundary nor industrial limit. In the global village, all individuals and organizations are confronted with risks of the same level. In this environment, the security of individuals or organizations' systems matters firstly to themselves. Only in some accidental situations are others involved, such as in the case of DOS attacks.

Powell provides evidence from the financial services industry to prove that cybersecurity is hardly a public good.[20] Individuals and organizations have excludability in cybersecurity. The excludability of cybersecurity roots in the three characteristics of cybersecurity, i.e., confidentiality, integrity and availability, among which confidentiality fully expresses the excludability of cybersecurity. We could see the situation this way: if security is available to one user, it is unavailable to other users, and if others enjoy security, ones' security does not exist any longer. Unsurprisingly, cybersecurity is characterized as preservation of confidentiality, ensuring that information is accessible only to those authorized to have access; integrity, safeguarding the accuracy and completeness of information and processing methods; availability, ensuring that authorized users have access to information and associated assets when required. The users' security is enjoyed solely by themselves. Any sharing entails that systems become insecure. In fact, hackers are precisely the exploiters and sharers of insecure systems. Therefore, cybersecurity has more excludability than any private good.

On the other hand, the cost of expanding security to others is not zero, but enormously high. If one user enjoys a higher level of security, the level of security of the others will relatively decrease. As mentioned above, there is no perfect security. Security and insecurity are relative concepts that exist in comparisons. If one enjoys a higher level of cybersecurity, the level of security of the others will decrease to insecurity. The competition between the security measures is the reason that causes increase of the difference between the relative securities. Of course, the enhancement of the total security level benefits from that competition.

Katyal's study stresses that to some extent private security measures may increase crime.[21] The basic assumption behind this argument is that, if one household locks its door, the thief will turn to the neighbor whose doors are left unlocked. Therefore, locking of one's own door breaks the reciprocity and mutual trust in the neighborhood. If we consider the fact that currently nearly all households, companies, and even government agencies "lock their own doors," we can easily conclude that this assumption is absurd. Only when every household, company and governmental agency is convinced not to take such "inefficient" measures is such an assumption significant. The author believes that such an assumption ignores the dual value of locking in the prevention of crime: on one hand, locking protects from damage and harm, making the potential criminals shrink back at the sight, or taking criminals more time before suffering losses; on the other hand, locking increases the potential criminals' time consumption and material costs in looking for new victims, and even making it impossible for them to find one. If none of the households and organizations locks their doors, potential criminals can easily find possible targets. Therefore, the difficulty of crime will decrease, and the efficiency will increase. The potential criminals are indifferent about costs, benefits, likelihood of success.

This pertains particularly to cybersecurity. If every computer owner is encouraged not to use security control, the computer will be more vulnerable to attacks. Assuming that the environment and the potential of all individual and organizations' computers are the same and the risk of being attacked is also approximately similar, then only when the benefits related to cybersecurity are equal could the provision of public cybersecurity be efficient. But this situation rarely exists in reality. Therefore, an unlimited public cybersecurity would be excessive for some individuals and organizations and insufficient for others. The situation of abundance is economically inefficient, while the situation of insufficiency is inefficient in terms of security. Hence, both ways, the public cybersecurity control cannot function optimally. In result, if cybersecurity is provided in the mode of public good, it is impossible to be more beneficial than as a private good.

Kobayashi notes that cybersecurity is different from traditional security.[22] To discourage crime *ex ante* in the general criminal context, the government could implement sufficient level of punishment to deter the crime from accruing. In the case of cybercrime, the likelihood of detecting is so low that the penalty imposed would have to be of considerable magnitude to deter cybercrime. In what follows, the author will explore the possibility of establishing liability for the different participants in the process of cybersecurity provision.

Public Provision of Cybersecurity: Liability Mechanisms

Even if it were technically feasible to keep all systems 100% secure, the costs would have been so prohibitive as to render such an approach an economic prescription for disaster. The government can neither provide cybersecurity nor manipulate the systems. Naturally, one of the Ernst & Young survey's key findings was that only 11% deemed government security-driven regulations as being highly effective in improving their information security posture or in reducing data protection risks.[23] However, any argument stating that the governments can play no role in the field of cybersecurity is over skeptical. The governments can play a necessary role in deterring the attackers, but they are by no means helpless in the maintenance of an adequate level of cybersecurity. Their roles are to impose penalty through legislation and deter crime by means of *ex post* law enforcement. Providing cybersecurity as a public good is confronted with greater difficulties in international cooperation than as a private good. Even if some countries can convince their taxpayers to pay for the expenses involved in the public provision of cybersecurity, if you cannot simultaneously convince all countries to do so, it will not be cost-efficient. In this section, the author will analyze the characteristics of the possible liability of various players in the field of cybersecurity.

Liability of Hackers

Ballon argues that the major benefits of holding the hacker liable for the damage he causes is that the target has more choices and control in applying the law against hackers.[24] Compared to a criminal action, the liability of hackers can be justified by that it grants the plaintiff "greater control over the litigation and potentially better long-term relief;" that it encourages attack reporting;[25] and that a target will have the motive to recover losses at the same time of punishing the perpetrator.[26]

The disadvantage of tort liability of hackers lies in two aspects: on one hand, the plaintiff has to pay a significant amount of money before receiving any compensation; on the other hand, most hackers have had and will have greater incentives to be judgment-proof.[27] If a hacker has little to lose under tort liability mechanism, his most rational choice will be to hide more secretly himself and his assets.[28] In the networked world, tracking a hacker or finding his money will need more energy, time, and costs, and will even prove to be an impossible task. As a result, the hacker would carry out the act more judgment-proof. Even worse, the hacker might be forced by the civil actions to commit other money-harvesting offences to support his actions.

Currently, dozens of countries have enacted domestic law against cybercrime. In addition, there have also been successful international legal actions, such as the Convention on Cybercrime (2001) and other domestic provisions.[29] Although the legisla-

tion is already there, the practical effects are doubtful. There are many hackers but the detection probability is quite low and the application of legislation is rare.

Liability of Internet Service Providers

Internet Service Providers (ISP)'s tort liability plays an important role in the following two cases: first, a lower level of ISP's security standard might be exploited by hackers; and second, the ISP's vicarious liability for its employee's security breach makes it easier to recover the target's losses.[30] To justify the first aspect, an important economic consideration is that the ISP's cost to improve its security level is lower compared to the hackers' high potential cost to society, and with the security standard the security condition becomes more certain and reliable.[31] This would be expected to lower the overall cost of the Internet service, provide incentive for Internet participation, and increase the value of the network to society.[32] There is no theoretical obstacle in applying the tort liability to cybersecurity breaches.

The only problems in applying tort liability to all ISPs is that there is no uniform standard; that it would be difficult to provide such a standard; and that dual or multiple standard would surely motivate some ISPs to maintain a lower level of security due to economic reasons. The result of this dilemma will be that no deterrence functions on hackers.

Liability of Security Problems Publishers

The security (holes) publisher has two aspects of gain from the publication, one is that the publication can prevent some harm suffered by the general public, the other is that the publication realises more economic or other benefits. However, it takes great risk resulting in users' losses in case hackers exploit the publicized loopholes. In addition, the users have to invest in improving their security protection when they know the new publicized loopholes.

According to Coarse's general principle,[33] whether the publisher should be held liable for his publication is a question of whether the gain of both the general public users from stopping the potential harm and the publisher himself from obtaining a higher confidence value is greater than the losses that the users suffer from the attacks launched exploiting the publicized loopholes and from the extra investment in preventing such attacks. In different cases, the cost effectiveness is different, and is hard to prove. Finally, as Preston and Lefton put it:

> The question is not whether an individual publication causes more harm than good; it is whether a particular rule of liability governing computer security publications causes more harm than good.[34]

Liability of Security Providers

The rapid growth of the computer security industry leads people to consider whether security providers should be held liable when their products and services fail to protect against hackers. Developing higher security level of products and providing high security level of services are costly, but work to prevent hacking from taking place. Security providers' liability will create incentives for them to provide products or services of at least a standard level. The products and services containing security holes take great risks of product liability if their advertisements stated that they are "hack-proof."[35]

The problem with holding security providers liable is that goods and services are usually provided subject to contract or licensing agreements, making tort liability inappropriate because the parties have bargained to allocate the risk between them.[36] The reasonable way in which the agreements are concluded is that neither of the two parties wants to bear more risk. But in general, the party of product or service users might have the greater discretion in choosing with more guarantees and less expenses. The security providers will be generally worse-off.

Liability of Software Vendors

Most of the security holes come from the bad design of software (and sometimes hardware). The software vendors control the only key to solve this problem through fixing their software. However, this work also consumes human resources and investments in terms of money. Therefore, vendors generally do not have the incentive to do so. A way to incorporate their better work into their best interests is to raise the risk of liability, which will raise the cost of their products. If software vendors have liability costs, they will pass those on to users. In turn, the vendors might as well pay to fix the problems.

Liability of Software Authors

Since the authors of software (the programmers) have the biggest opportunity to prevent problems, it seems appropriate to focus on making them responsible for the security of their products.[37] Nonetheless, there are some unique aspects of computer software that make it challenging to apply traditional notions of product liability.

Under such circumstances, if we impose liability on the authors, it is impossible, because the author gets no income to pay the compensation; it is inefficient, because the author would be discouraged from contributing; and it is also unfair, because the users use the software for free and voluntarily.

Liability of System Owners

Systems can be both targets and tools in attacks. For example in a Distributed Denial of Service attack, the attacks are launched from numerous manipulated computers. The owners of such systems, who use software written and sold by third parties, cannot fully secure their systems, cannot stop unforeseeable outsiders' exploitation, and have no way to reduce the risks. In order to hold the system owners liable, two prerequisites are necessary to be in place: the establishment of a security standard, and the mechanism of insurance. The latter was discussed by Fisk in analogy to vehicle operators who are often legally required to carry insurance against accidents.[38]

Conclusion

This article argues that cybersecurity is a private good and should be provided mainly by the private sector. Regarding cybersecurity as a public good would discourage the private sector to invest in security provision. From this standpoint, an early government intervention would reduce the effectiveness and efficiency of cybersecurity. However, in terms of prevention of security breaches, law enforcement can play an important role in establishing and enforcing liability mechanisms. Although it is still controversial whether and how cybersecurity players should be held liable for their activities, every step made in this direction will bring benefits to the private sector to achieve their goals.

Acknowledgement

The author wishes to express his appreciation to Jenny and Antti Wihuri Foundation, the Department of Law at the University of Joensuu, and the Finnish Cultural Foundation, for their generous financial support for his current research. He also wishes to thank the Finnish Economic Education Foundation for supporting him in the early stage of this research. Certainly, the responsibility for the contents is the author's.

Notes:

[1] National Research Council, *Cryptography's Role in Securing the Information Society* (Washington, DC: National Academy Press, 1996).

[2] John McCormick, "Five Reasons You're not Secure," 5 April 2005, <insight.zdnet.co.uk/internet/security/0,39020457,39193819,00.htm> (14 Dec. 2005).

[3] Dan Farmer, "Shall We Dust Moscow?: Security Survey of Key Internet Hosts & Various Semi-Relevant Reflections," November-December 1996, <http://www.trouble.org/survey/> (14 Dec. 2005).

[4] Jennifer A. Chandler, "Security in Cyberspace: Combating Distributed Denial of Service Attacks," *University of Ottawa Law & Technology Journal* 1 (2003-2004): 231-261, <http://www.uoltj.ca/articles/vol1.1-2/2003-2004.1.1-2.uoltj.Chandler.231-261.pdf> (14 Dec. 2005).

[5] Christopher Coyne and Peter Leeson, "Who Protects Cyberspace?" Working Paper 24 (George Mason University, Department of Economics, Global Prosperity Initiative, 2004), <http://www.mercatus.org/pdf/materials/616.pdf> (14 Dec. 2005).

[6] Benjamin Powell, "Is Cybersecurity a Public Good? Evidence from the Financial Services Industry," Working Paper Number 57 (The Independent Institute, 15 March 2001), <http://www.independent.org/pdf/working_papers/57_cyber.pdf> (14 Dec. 2005).

[7] Torgeir Daler, Roar Gulbrandsen, Birger Melgrd, and Torbjorn Sjølstad, *Security of Information and Data* (Ellis Horwood, January 1989), 15.

[8] Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, and Robert Richasrdson, *Tenth Annual CSI/FBI Computer Crime and Security Survey* (Computer Security Institute, 2005), 11, <http://i.cmpnet.com/gocsi/db_area/pdfs/fbi/FBI2005.pdf> (14 Dec. 2005).

[9] CERT Coordination Center, *CERT/CC Statistics 1988-2005* (2005), <http://www.cert.org/stats/cert_stats.html> (14 Dec. 2005).

[10] Peter G. Neumann, "Information System Adversities and Risks" (paper presented at the Conference on International Cooperation to Combat Cyber Crime and Terrorism, Stanford, CA: Hoover Institution, 1999).

[11] Carl Howe, John C. McCarthy, Tom Buss, and Ashley Davis, "The Forrester Report: Economics of Security," (February 1998).

[12] Gordon, Loeb, Lucyshyn, and Richasrdson, *Tenth Annual CSI/FBI Computer Crime and Security Survey*, 15.

[13] Stephen J. Lukasik, "Protecting the Global Information Commons," *Telecommunication Policy* 24, no. 6-7 (2000): 519-531.

[14] Robert L. Ullman and David L. Ferrera, "Crime on the Internet," *Boston Bar Journal*, no. 6 (November/December 1998).

[15] L. Jean Camp and Catherine Wolfram, "Pricing Security," in *Proceedings of the CERT Information Survivability Workshop* (Boston, Massachusetts, 24-26 October 2000), 31-39, <www.ljean.com/ files/isw.pdf> (14 Dec. 2005).

[16] Camp and Wolfram, "Pricing Security."

[17] Source: "Good (Economics and Accounting)," Wikipedia, the free encyclopedia <http://en.wikipedia.org/wiki/Good_%28economics%29> (15 Dec. 2005).

[18] Paul A. Samuelson, "The Pure Theory of Public Expenditure," *Review of Economics and Statistics* 36 (November 1954): 387-389.

[19] Hal R. Varian, "System Reliability and Free Riding," in *Proceedings of the First Workshop on Economics and Information Security* (University of California, Berkeley, 16-17 May 2002), <http://www.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/49.pdf> (14 Dec. 2005).

[20] Powell, "Is Cybersecurity a Public Good? Evidence from the Financial Services Industry."

[21] Neal Kumar Katyal, "The Dark Side of Private Ordering for Cybersecurity," in *The Law and Economics of Cybersecurity*, ed. Mark F. Grady and Francesco Parisi (Cambridge University Press, November 2005).

[22] Bruce H. Kobayashi, "An Economic Analysis of the Private and Social Costs of the Provision of Cybersecurity and Other Pubic Security Goods," *Supreme Court Economic Review* 14 (2005), <http://law.bepress.com/gmulwps/gmule/art26> (15 Dec. 2005).

[23] Earnest & Young, *Global Information Security Survey 2004*, BYG No. FF0231, <http://www.ey.com/global/download.nsf/International/2004_Global_Information_Security_Survey/$file/2004_Global_Information_Security_Survey_2004.pdf> (14 Dec. 2005).

[24] Ian C. Ballon, "Alternative Corporate Responses to Internet Data Theft," in *17th Annual Institute on Computer Law* 737, 744 (PLI Patents, Copyrights, Trademarks & Literary Prop. Course, Handbook Series No. 471, 1997).

[25] David L. Gripman, "The Doors Are Locked but the Thieves and Vandals Are Still Getting in: A Proposal in Tort to Alleviate Corporate America's Cyber-Crime Problem," 16 *J. Marshall J. Computer & Information Law*, 167 (1997): 174-176.

[26] Michael Hatcher, Jay McDannel, and Stacy Ostfeld, "Computer Crimes," *American Criminal Law Review* 36, 397 (1999): 406.

[27] James Brooke, "Calm Scene Isn't Really, Police Say," *New York Times*, 22 April 2000, C1.

[28] Mary M. Calkins, "They Shoot Trojan Horses, Don't They? An Economic Analysis of Anti-Hacking Regulatory Models," *Georgia Law Journal* 89, no. 171 (November 2000): 214-217.

[29] Convention on Cybercrime 2001.

[30] David Icove, Karl Seger, and William VonStorch, *Computer Crime: A Crimefighter's Handbook* (O'Reilly and Associates, Inc., August 1995): 427.

[31] Icove, Seger, and VonStorch, *Computer Crime.*

[32] Marc D. Goodman, "Why the Police Don't Care about Computer Crime," *Harvard Journal of Law and Technology* 10, no. 3 (1997): 465-494.

[33] Ronald H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (1960): 1-44.

[34] Ethan M. Preston and John Lofton, "Computer Security Publications: Information Economics, Shifting Liability and the First Amendment," *Whither Law Review* 24, no. 71 (Fall 2002): 130.

[35] Natalee Drummond and Damon J. McClendon, "Cybercrime – Alternative Models for Dealing with Unauthorized Use and Abuse of Computer Networks," (Summer 2001), <http://gsulaw.gsu.edu/lawand/papers/su01/drummond_mcclendon/> (14 Dec. 2005).

[36] E. Gabriel Perle, Mark A. Fischer, and John Taylor Williams, "Electronic Publishing and Software," Part A, *Computer Law* (January 2000).

[37] Mike Fisk, "Causes and Remedies for Social Acceptance of Network Insecurity" (paper presented at Workshop on Economics and Internet Security, University of California, Berkeley, 16-17 May 2002), 3, <http://www.sims.berkeley.edu:8000/resources/affiliates/workshops/econsecurity/econws/35.pdf> (14 Dec. 2005).

[38] Fisk, "Causes and Remedies for Social Acceptance of Network Insecurity."

XINGAN LI, born in 1967, LLB (1989), LLM (1994), is Associate Professor at Inner Mongolia University Law School. He was a visiting scholar at Kyushu University (2000-2001), and researcher at the University of Lapland and the University of Joensuu. His research interests are criminology, criminal psychology, criminal law and criminal procedural law, and particularly, cybersecurity and cybercrime. His publications include the books "Criminal Law of England and Wales," "Principles of Criminal Law," and several papers on economic crime, cybersecurity and cybercrime. *Address for Correspondence*: Sepänkatu 15 C 57, 80110 Joensuu, Finland; *Phone*: +358 044 910 7632; *E-mail*: li@cc.joensuu.fi.

Dealing with Fraud

- Money Laundering Techniques with Electronic Payment Systems
- Credit Card Fraud Detection Using Self-Organizing Maps

MONEY LAUNDERING TECHNIQUES WITH ELECTRONIC PAYMENT SYSTEMS

Krzysztof WODA

Abstract: In case of terrorist actions, the main emphasis lies on the security aspect. Nevertheless, the financial aspect, as for example the collection, transfer and withdrawal of money from the payer to the payee plays an equally important role for preparation of terrorist actions. Besides, the techniques of money laundering are often used to conceal or disguise the origin, nature, source, location, disposition or ownership of assets financing terrorist actions. The Financial Action Task Force (FATF) organization has identified many activities and typologies used for money laundering or terrorist financing, e.g. with shell companies and nominee, through unofficial money transfer systems, "smurfing" practices (through dividing of large payment sums into multiple deposits under the threshold value), transfers, money smuggling, etc. Many of these activities are carried out by means of electronic payment systems, which transfer the monetary values over telecommunication networks. Customer instruments such as Internet software (electronic purse), smart cards, mobile devices, debit and credit cards could be applied as payment instruments. The suitability of an electronic payment system for financing of illegal activities depends to a great extent on such supported characteristics as anonymity, mobility, etc. It can be furthermore assumed that payment systems differ with regard to their suitability for a single money laundering phase (and thereby for terrorist financing). The multi-process character of money laundering is typical for terrorist financing and often contains a series of transactions from collection of money to money withdrawal in order to conceal the origin, nature or disposition of money.

Keywords: Money Laundering, Terrorist Actions, Electronic Payment Systems.

Introduction and Definitions

Money laundering is an intentionally committed offense that signifies the conversion and transfer of assets of an illicit origin. The objective of this action consists of disguising the true origin, location, nature, disposition, movements and transfer of assets that are derived from illegal activities.[1] Participation, support or facilitation of the realization of illegal activities, such as transfer of money of illicit origin to several

INFORMATION & SECURITY. An International Journal, Vol.18, 2006, 27-47.

bank accounts and afterwards their conversion into legal financial products, are regarded also as money laundering actions.

Different goods are considered as potential assets for money laundering; they can to a different degree fulfill the functions defining money ("medium of exchange," "store of value," and "unit of account"). The assets with distinctive "medium of exchange" function include highly liquid funds like cash, sight deposits, checks and electronic currencies, as for example prepaid bearer payment instruments (e-money) and virtual gold currencies. The gold currencies support particularly the "store of value" function, while they are based on real gold reserves with current market value. Besides, gold acts as an internationally exchangeable property with simply calculated prices in various currencies ("unit of account") as well as a direct payment instrument (bearer instrument) without identification features as for example credit card number or bank account number. Typical finance products and real estates as well as products and services of certain commercial branches could also be considered as other assets for money laundering, such as those coming from restaurants, casinos or shops in e-commerce.

The transformation of liquid funds or cash into finance products is often a crucial moment for detection of illegal activities.[2] This could be accomplished often as checking the business activities of suspects by state authorities (e.g., related to tax payments), audit companies and banks (following some money laundering guidelines); however, the origin and disposition of assets must be determinable in case of legal activities.[3] The money launderer can use the anonymous funds or generally accepted money currencies functioning as medium of exchange for any intended purpose (also terrorism financing) only if this money has left the business cycle successfully, i.e. without disclosure of its illicit origin. Such stepwise introduction, movement through several cash transaction systems and business cycles and, in the following, legal use of the laundered money is called in general *placement*, *layering* and *integration* of assets in the money laundering cycle.[4] In the placement phase, assets from criminal activities are mostly deposited (on a bank account, e.g., against checks), invested (in finance products) or smuggled (cash, diamonds). The layering phase consists of transfers of the placed assets between several accounts in different institutions and other business participants with the purpose to conceal the identity of the true owner or the trading person. Besides, the illegally trading persons try to avoid different legal restrictions, as for example the report obligation for transaction amounts above legally defined level (currency change, owner of foreign bank accounts). In the integration phase, the legal as well as the illegal assets are combined together and integrated into the business cycle. The already inseparably booked assets will be often transferred back to the owner, now as legalized possessions.

The money laundering techniques and the laundered money are often used for terrorist financing. The planning, logistics and acquisition of objects for terrorist actions often require a cross-border transfer of funds to the country of destination. Direct importing of cash will be avoided for the reason of strict border control; more sophisticated techniques will rather be applied for quick and mostly complex transfer of funds through existing legal and illegal transfer systems and financial instruments. Electronic payment systems are generally characterized by high performance and mobility; some of them have also such important features as for example anonymity of transfer, cross-border payment possibility, cost efficiency, as well as high security of communication (confidentiality) due to the use of cryptographic procedures. Therefore, the electronic payment systems are quite suitable for the conduct of money laundering operations in every phase of the money laundering cycle.

The aim of this article is to analyze the possible techniques for money laundering and terrorism financing, which can be carried out with electronic payment systems. Furthermore, the article will identify the most important characteristics of the particular payment systems, which predetermine such systems as especially suitable for illicit activities. Finally, solutions will be presented to reduce the risk of illegal money transfers with electronic payment systems.

Money Laundering Techniques with Electronic Payment Systems

The money laundering techniques involve direct use of electronic payment systems for terrorism financing or their use only as a transporting instrument in one of the three phases of the money laundering cycle. In what follows, the typical techniques for money laundering involving the use of electronic payment methods and identified by the Financial Action Task Force (FATF) organization will be presented.

Transfers

Money wire transfers can be characterized as the easiest transfer method within the money laundering activities. Transfers are financial transactions by which value unities are transported from the payer to the payee electronically over telecommunication networks. In case of money laundering, often the sender and the receiver of the transferred money is one and the same person who tries to conceal the origin of money by several money movements (transfers).[5] In general, there exist legal transfer systems and illegal ones, often called parallel bank transfer systems. The legal transfer systems comprise, in the private customer area above, all electronic banking and, in the corporate clients area, the international large value payments through SWIFT or TARGET. The illegal transfers take place through systems such as Hawala for example, which are based on informal or trust connections and effect money transfer without using official bank accounts. The legal transfers can be well traced due to ar-

chiving of the transaction data by the financial institutions. Exceeding some legally fixed threshold values for money transfers or deposits, as for example 15,000 Euros in the EU (Directive 2001/97/EC, Article 3) or $10,000 in the U.S. (Section 326, the U.S. Patriot Act), automatically triggers reporting and checking of the origin of money by bank employees and then, in suspicious cases, by supervisory authorities. Even opening an account in a bank or a financial institution (on-line broker, mutual funds, and insurance companies) requires an extensive customer identification (e.g., the Customer Identification Program in the U.S.) and investigation, as for example in the U.S. through comparison with the lists of private individuals or organizations whose assets were frozen by the Treasury Department.[6]

Despite the electronic tracking, investigation possibilities and identification requirements, the transfers are an efficient money transferring method for terrorist financing. The following cases can be considered as possible abuse cases:

- *The use of falsified or false identities* (front men, letterbox companies). Customers with good reputation, often on the basis of ethnic, religious or cultural affiliation with the money launderers or terrorists, let the transfer of money through their official bank accounts internationally. The legal account holders can further guarantee to the money launderers access to their bank accounts by disclosing their PIN or password. Therefore, it would be difficult to distinguish bank accounts for a suspicion of money laundering (assuming that the transfers would not contain extraordinary large amounts of money), due to the fact that they combine legal transfer(s) with the illegal ones. Also, email accounts are often opened with false identity at public places (e.g., library, university) to conceal the real identity.[7]
- *Structured payments* also called "*smurfing*." Large payments are split and transferred always in smaller sums that lie under the legal threshold value required for checking on suspicion of money laundering. Such payments are conducted through several channels (phones, online-banking, chip cards with electronic purse function, mobile payment systems) to complicate the detection of structured payments.
- *Transfers through banks in offshore countries* with customer identity protected from jurisdiction. The opening of bank accounts as well as the transfers can occur only on the Internet, so the investigation have to rely solely on electronic evidence like the visited IP addresses, the computer cache, as well as the information about the conducted transactions stored on the server of the Internet Service Provider (ISP). However, the Internet Service Provider would often be chosen from countries without restrictive bank regulation or countries not-cooperating with organizations like FATF that prevent cooperation between service providers, banks, and the authorities. Furthermore,

the relevant information about the transactions could be encrypted symmetrically or asymmetrically by cryptographic software programs (containing known algorithms as RSA, AES, Triple-DES, etc.) and, therefore, it can be confidentially exchanged between the participants (e.g., terrorists) without risk of disclosure by the authorities. A report of the U.S. Treasury points to another technique for hiding information – the technique of email drafts.[8] All the terrorists receive the password and the user name for a given email account. If one of them writes a draft, without sending it, then this draft remains on a provider server for this email account. All the terrorists could then access the account and read the draft.

- *Transfers as a result of criminal actions*, as for example hacking of bank systems or attacking private computers (the man-in-the-middle-attack). Such attacks can be used mainly for fundraising for terrorism financing, but not for money laundering since money laundering is a multistage and often a long-term process that goes also through legal transaction systems and business cycles with the objective to legalize the funds.
- *Informal money transfer systems*, such as Hawala, enable a special category of transfers in the Far and Middle East. Money is deposited at a Hawala representative in the country of the payer and is paid by another Hawala representative in the country of the payee. The Hawala representatives calculate their demands and liabilities often mutually or balance the difference through their bank accounts. The payments in the Hawala system are predominantly cash-based, while the communications and the payment confirmation occur often electronically (e.g., email, fax, chat). Hence, the Hawala transfer system combines the advantages of the traditional cash systems (anonymity, no registration of the transactions, and transferability to other private individuals (person-to-person)) with the advantages of electronic communication (high speed and cost efficiency).

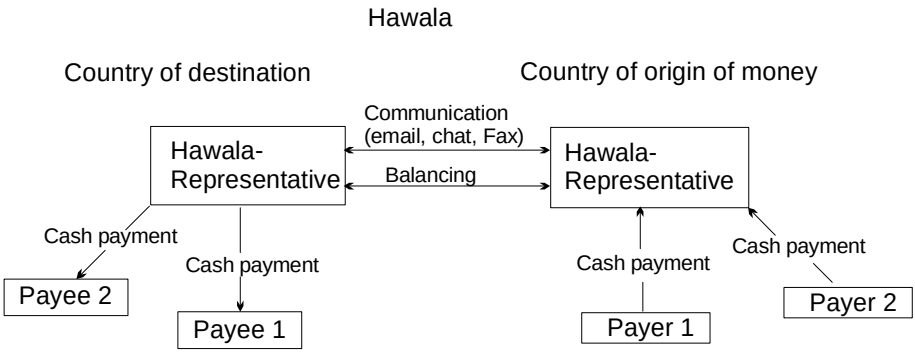


Figure 1: The Hawala System.

The electronic wire transfers are the easiest method for terrorism financing, nevertheless, with high risk for detection. Prior to the terrorist attacks in America on September 11, 2001, the transfers were considered less suitable for terrorist financing because of the risk for electronic tracking (the money for performing the attacks on the WTC were transferred mostly with legal wire transfers and illegal Hawala transfers).[9] The possible combination of wire transfers with the Hawala system as well as the use of the Internet for communication determine the transfers as ideal instruments for short-term illegal activities such as terrorist actions. The transfers are also suitable for money laundering, however, only to a limited degree, mainly during the layering phase for multiple movements of money between accounts.

Shell Companies, Offshore Corporations, Nominees, and Charities

While the wire transfers have a one-time character and a high risk of detection mainly when transferring large sums, different business activities are used for long-term money laundering and sophisticated terrorist financing. Registered companies or non-profit organizations like charities can transfer larger amounts over borders than private individuals. The charities often operate in crisis areas where also the terrorists are active. Even if the enterprises or organizations are supervised by the authorities (e.g. bank supervisory or tax authorities), it leads mostly to no indications for terrorist financing. Also, the on-line check of the WWW contents or the check of logs can often deliver no direct indication of suspicious cases. In such cases, it should be rather looked for the list of charities' contributors; however, the contributors commit no crime if they have no idea of the illegal financing activity of the charity. Raising money for humanitarian purposes could generally not be distinguished from the financing of illegal activities. Therefore, key role in detecting illegal financing activity of a charity may play the use of the means or the history of the conducted transfers. Unfortunately, the charities or the other money raising organizations often operate on informal basis in quite closed groups, such as ethnic or religious groups, and afterwards can transfer money also informally, e.g. through the Hawala system, so there is no track of transfers or fundraising in official transaction systems.

The charities can themselves operate as informal transfer system if, for example, their employees are in several countries (also in the countries where terrorists operate), and carry out illegal transfers. Then transfers take place without transferring the money physically; rather internal accounts are balanced without leaving a track in the legal bank system. In addition, the charities are often linked organizationally with each other or are represented by the same persons what complicates at last the later investigation of the cases.

The use of the so-called shell and offshore corporations (or from offshore territories) offers another method for big money transfers. The shell corporations are enterprises

without usual business activity, assets, and liabilities. They are rather used only as an intermediary for the transfer of capital. They become often linked or pyramided to disguise the track of moved and laundered money. Generally, the shell companies have merely an address (e.g., a letterbox company), a manager (nominee like attorney or manager of the offshore corporation) and often many bank accounts. The funds are sent electronically through bank accounts and between different places worldwide. Hence, shell corporations can function as ideal camouflage for illicit money mainly in the layering phase of the money laundering process.[10]

An indication of potential illegal activity of a shell corporation provides the owner's structure and activity profile of the company. The owners of the shell corporations often become the actual owner of the bearer shares or unregistered stock, thus no private person or shareholder is registered in the commercial register.[11] Such omitted regulation appears in some offshore centers with restrictive bank secrecy law, also for enterprises, allowing the owners of shell corporations not to be identified. Other advantages from the offshore location of a shell corporation are the possible tax exemptions and the strict protection of customer privacy by the attorney-client relationship, e.g., in case of investigation by supervisory authorities of another country.

Shell corporations in offshore countries are well suited for the integration phase of money laundering. After the money has been moved to an offshore center (e.g., by structured payments through several bank accounts, charities or by transfer of virtual gold currencies), it can return to the owner already in a legal form. A known method is the loan-back schema. The placed money is transferred in form of a loan, e.g., from a shell corporation or an offshore bank to the domestic company (furthermore, no taxation on the loan or laundered money is due). The loan is paid back with laundered money and officially appears as an origin of the money to the borrower (money launderer). In reality, often in the money laundering case, the borrower as well as the credit grantor is the same physical person who conceals his real identity using companies (shell corporations), nominees, etc.

Other schemas include manipulations with invoices for delivered or ordered goods and services. In case of under-invoicing, the beneficiary (seller) gets a very low amount of money, far below the usual market price for a delivered product (e.g., computers, cars); while the payer (buyer) can resell the product and can thereby register big profits (laundered money).[12] The seller pays in reality, e.g., an illicit goods delivery (drugs, weapons), or acts as a shell corporation which conceals the identity of the contractor (in money laundering cases the seller and the buyer is one and the same physical person). With the over-invoicing schema, exorbitant prices for goods and services are paid by contractors resulting in extraordinarily high profits (laundered money) for the seller. Such transactions are often characterized by fictitious deliveries of goods and services or have hardly determinable values (e.g. market

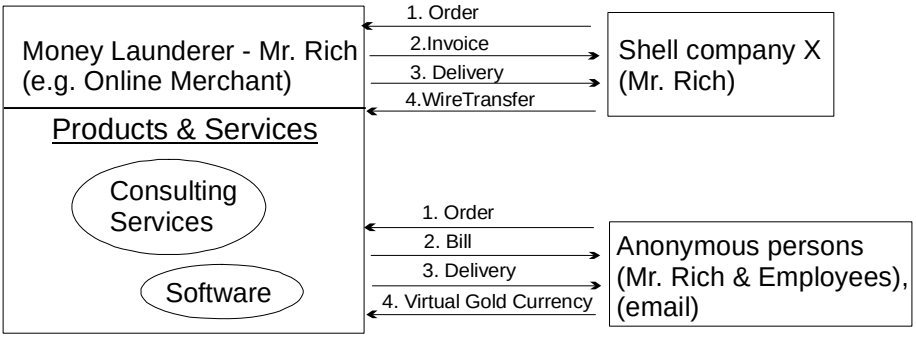


Figure 2: Over-Invoicing on the Internet

analysis, consulting services, and software products on the Internet). The transactions with software programs on the Internet, for example, can be carried out by shell corporations from offshore territories and anonymous persons (merely with email address or IP number identifiable), or with anonymous electronic payment systems, which makes the retracing from the payee to the payer practically impossible. Thus, the transferred money for goods or services is legalized and can be reintegrated in the business cycle (e.g., real estate purchase or business activities of a legal enterprise).

The role of the electronic payment systems in the traditional money laundering techniques with companies is very important. Most transfers take place electronically using electronic banking (e.g. electronic fund transfers (EFT)), through SWIFT, TARGET or by the involvement of virtual currencies such as gold currencies on the Internet. Many providers of virtual gold currencies are located in offshore centers whereby the transfers in the layering phase of money laundering may be anonymous.[13] The transfers are often final (no chargeback risk) and leave no information at the providers of private currencies or transfer systems. Without electronic payment systems, the money laundering models with shell corporations (or from offshore countries) would be not so effective since it would require transporting of the physical cash by a courier to an offshore centre in the placement phase of money laundering (with the concomitant high risk of detection).

Financial Products

The financial products are ideal instruments in the integration phase of the money laundering process given the size of the market, the easy access and the availability as well as the diversity of products. Primarily, insurance policies are bought to legalize the laundered money, e.g., through income from life insurance policies.[14] The advantage of the insurance products for possible money laundering transactions lies in the

free access to such products given that broker and external agents are rather interested in profits than in the due diligence of the customers. In addition, another party could benefit from payment of insurance policies. Detecting money laundering activities is clearly complicated if the insurance policies were bought in offshore-centers or through trustees or nominees.

Other financial products could also become investment objective of money launderers as, for example, share certificates, bonds or not-standardized derivatives such as swaps and futures. Not-standardized securities could mainly be used in deals between a money launderer and a shell corporation that belongs to him and could be the origin of high-level-speculative profits or losses. Such high-level-speculative trading between the members of a criminal network can be conducted especially in the placement phase of money laundering. Nevertheless, security trading requires the involvement of a licensed security broker; however, such trading can be carried out on the Internet and by offshore brokers quickly and efficiently. Furthermore, the securities bought can act as security for loans in the loan-back schema for money laundering. In this way, the cycle of money laundering activities closes (eliminating possible suspicions) with the appearance to have only legitimate transfers of money and securities in circulation.

Money Laundering with Electronic Payment Systems

Electronic payment systems are characterized by large diversity in terms of access methods (off-line or on-line), transport methods (access to bank account, to card with electronic chip or to virtual currency account, etc.), charging methods (flat, pay per use, variable fees dependent on volume), and time of the payment settlement (prepaid, post paid, pay now). Differentiation criteria such as access to payment application often determine the other features of a payment system. During payment, the on-line payment systems are characterized by an on-line connection to the system provider who authorizes the transaction. Relevant access data such as passwords as well as currencies are stored on a central server of the system provider. Hence, requests (e.g., for PIN) and answers (authorization) between the participants (client, server of system provider) are exchanged on a real-time basis. In case of off-line procedures, the end devices of users guarantee not only data transportation and authorization, but they often act also as storage medium for currencies. Therefore, the currencies or payment orders flow directly to the receivers without involving the bank or the transaction system of the system provider. Authorization of an off-line transaction is not possible on a real-time basis (only, e.g., at the end of a specified period of time if the turnovers were submitted to the bank (acquirer)), and bears a high risk of fraud. The on-line as well as the off-line payment systems can be distinguished by certain special features that sometimes position the systems against each other in a conflict. These

conflicts are often irresolvable and, hence, determine the character of a payment system as well as its potential use for illicit actions.

Key Features of Electronic Payment Systems

Many empirical studies have shown that the most important reasons for the use of an electronic payment system from the point of view of the end users are its flexibility (e.g., the use of existing end devices for new payment systems), convenience (e.g., an easy entry to the payment system, possibly without registration, as well as quick settlement), and security.[15] Low availability, high transaction costs as well as a nescience in dealing with the electronic payment systems were named as the most important reasons that obstruct the adoption of such systems. High significance is often given to the security of these systems; however, this is rather based on an uncertainty or on nescience than on acknowledgement of the concrete risks and security mechanisms.

The most important reasons for the adoption or the refusal to use electronic payment systems often coincide with the most important reasons to adopt or refuse the use of such systems by the money launderers. Indeed, the money launderer has specific goals of use and strategies (e.g., for disguising the origin of the money); however, the choice of concrete technologies and in this case of a suitable payment system for a money laundering operation often depends on the features supported by an electronic payment system as well as on the knowledge and ease of handling of the system by the money launderers. The money launderer would have to decide on a specific profile of the supported features, taking in consideration the fact that between the features already mentioned irresolvable conflicts exist. And to date no electronic payment system supports even approximately the majority of the desired features by the consumers (or in the narrower sense by the money launderers).

Conflict between Security and Cost Efficiency

High security is guaranteed mainly to centralized on-line payment systems applying many cryptographic mechanisms. Costly mathematical procedures are used for encoding and decoding of data, which should satisfy the well-known requirements of security: confidentiality of the data and transfer; integrity of the data as protection from manipulations (by generating hash sums); authentication of the communication partners on the basis of passwords, PINs, digital signatures, etc.; and non-repudiation, e.g., owing to digital fingerprints. The cryptographic operations are very costly especially in the case of on-line payment systems with high security level since they are based on asymmetric procedures (long keys), with multiple on-line communication generating high costs (for computational costs, on-line connections, storage capacities). Contrary to the on-line procedures, the off-line payment systems, as for example

Millicent, use only the highly cost-effective check sums (hash value) and digital signatures.

Transaction costs represent principal costs in the total cost of a transaction. In general, the electronic payment systems are more efficient than the traditional paper- or account-based payment systems owing to the lower transportation and storage costs (no significant costs for transport or insurance policies, only communication and storage space costs). The settlement costs often vary as a function of the value of a transaction, the risks of non-repudiation as well as the costs for the infrastructure, e.g., for clearing and verification of turnovers by payment system provider (e.g., costs for the card evidence and control centers of the German GeldKarte chip card-based payment system).[16] For the customers, searching costs appear in the form of search efforts for acceptance places, virtual brokers, banks, change agents, etc. For money launderers, these searching costs are often an integral part of the layering phase of money laundering, while money circulates through several accounts, shell companies or offshore centers. Some payment systems also offer certain profit possibilities (besides lower transaction costs), as, for example, virtual gold currencies in case of increasing market prices of gold.

Conflict between Anonymity and Non-Repudiation

Anonymity means secrecy of customer identity as well as hiding of transaction data. The primary purpose of anonymity is protection of customer's private area in order to prevent the creation of customer profile for unauthorized marketing actions. Besides, no relationships/ associations would be possible to be established between customers, traders and related data. As back as in 1987 had David Chaum referred to the possibility of linking computer data of different organizations with the help of certain key data for unauthorized actions and as a response suggested the use of unique digital pseudonyms for each transaction.[17] The idea of Chaum, which is in the form of an anonymous digital signature (the so-called blind signature), was used in an electronic money system known as Ecash (the identification characteristics, as for example the serial number of an electronic coin, are covered with a blinding factor before sending to the issuing bank for digital signing. Perfect anonymity is guaranteed only if the customer does not identify, e.g., by giving information about his physical address for delivery).[18] Other mechanisms—such as dual signatures with SET (the clearing centre as well as the trader gain access only to the data part relevant to them, e.g., the payment information or the order data), special alias number instead of telephone number for mobile payment systems (e.g., Paybox), anonymous prepaid telephone cards or coupon cards (e.g., Paysafecard)—guarantee anonymity of transactions to a large extent. Such anonymous technologies are also used by the money launderers helping them to avoid the identification procedures (e.g. resulting to the "know your cus-

tomer" rule in the banks). Furthermore, other Internet technologies are also used, such as IP-Spoofing (modifying packet headers of Internet messages to make them appear to have originated from a trusted site) or generating anonymous IP by anonymous hosting (also from offshore countries).[19]

Anonymity is supported only by a small number of electronic payment systems. Nearly all electronic payment systems are characterized by extremely short payment circulations (unique coins or transfers), no transferability of the coins or the checks to private persons without the involvement of a bank (protocol and verification of the payment data) and the existence of many identity characteristics (as for example serial number of a coin, bank details, telephone number for a mobile transaction, shadow accounts for cards with electronic chip as for example GeldKarte, etc.). Excluding anonymity is often done in order to maintain consistency of the payment systems in case of unauthorized copying of the electronic coins or checks (the so-called double spending) as well as in case of potential use of such anonymous payment systems for illegal actions such as money laundering, tax evasion, and terrorism financing. For this reason, authentication mechanisms will be adopted for unique customer identification (e.g., PIN for debit cards, transaction random numbers for on-line transfers, telephone number or alias for Paybox and Mpay, digital signatures and fingerprints for Ecash, card and terminal serial numbers for GeldKarte, "passphrase" for e-gold account of e-gold Ltd., etc.) which should guarantee non-repudiation at the same time.

Conflict between Convenience, Mobility, and Security

Convenience for the users of an electronic payment system implies time and location freedom, access possibilities to other integrated applications as for example electronic banking and brokerage, free transferability of the assets between private users as well as an easy and comfortable use, possibly without any registration (physically in the branch, on-line on Internet or WAP).[20] The requirement for convenience is clearly in the interest of money launderers who use, for example, person-to-person transfers primarily in the layering phase. Besides, protocol of transactions and hence traceability would be avoided. Electronic payment systems would rather reach in such cases cash functionality and be suitable, therefore, for terrorist financing.

For money laundering activities, mobility of payments is of great importance as another characteristic bringing convenience. In the ideal case for money laundering, transfers will be carried out trough countries with bank secrecy laws, with allowed anonymous accounts and customer identity protecting policy, generally with the help of notaries and attorneys (most of all during the layering phase). Mobility of payments is predictable considering the international character of e-commerce (e.g., sup-

plier's ordering systems or auctions) and, therefore, meets the requirement of money laundering operations for cross-border transfers.

Many electronic payment systems (e.g., ecash, CyberCoin, SET, Paybox in Germany) have not been accepted by the customers due to insufficient convenience, flexibility, and mobility. Also, in spite of the developed mature security technologies (digital, blind signatures) for confidentiality and authentication of the customer, the systems could not reach the critical mass of customers. Instead, payment methods spread as, for example, transfers and debit procedures per SSL protocol, which does not guarantee non-repudiation of the transaction or authentication of the payer (potential money launderer) (simply the server of the payment provider will be authenticated with appropriate certificates), even though it allows to carry out the transfer fast, easy, and without additional software (only Internet browser) or hardware.

Suitability of Electronic Payment Systems for Money Laundering

Choosing an electronic payment system for illegal activities such as money laundering or terrorist financing depends on many factors as, for example, the duration of the operation, the amount of money to be transferred, the international or local character of the transfer and, furthermore, it also depends on such individual preferences of the money launderers or smugglers as their attitude to new technologies, risk aversion to electronic payment methods, and many others. Therefore, an electronic payment system cannot be analyzed by means of a universal pattern or matrix containing specific features, but rather intuitively and considering the definition of money laundering (concealing or disguising the origin, location, use, nature of assets, etc.). Hence, a number of payment systems were selected for the analysis, which posses such characteristics that make possible to carry out with them international money transfers in a convenient, fast, and flexible way, through anonymous accounts.

Virtual Gold Currencies

Virtual gold currencies (e.g., e-gold, Goldmoney, e-Bullion, AnonymousGold) are account-based electronic payment systems whose value is backed by 100%-golden deposits in a physical form (bullions, bars or specie). The gold reserves are in a private storage of the system provider who often operates from an offshore country (e.g., e-gold Ltd., Nevis Corporation).[21] In the case of gold currencies, only certain weights of gold are booked to accounts of receivers. While the possession of gold reserves changes constantly, the gold in the treasury vault remains untouched.

For exchange or purchase of gold currencies, the user opens an account for a virtual gold currency at a system provider. The identification requirements are negligible in comparison to opening a bank account and are often limited only to a request for information such as name, email address and occasionally physical address to which

then a "verification code" is sent. Furthermore, also a copy of an ID could be required to be faxed or sent. Nevertheless, such verification will be often omitted if, for example, the transactions do not exceed the value of 15,000 euros or $15,000.[22] Structured payments for money laundering or terrorist financing can be made by opening several accounts at a system provider or accounts at many different providers without the need for identification (only the email address). Similar verification obligations are also required by the exchange agents who exchange gold currencies for national currencies worldwide (e.g. Gold-cash.biz informs that in the case of a foundation of an offshore enterprise in Delware no copy of the real ID is required).[23]

Person-to-person transfers between users of virtual gold currencies are allowed (which shows suitability for the layering phase); the transfers are conducted very fast worldwide and with no chargeback risk. For cash withdrawal and the integration phase of money laundering, the special payment cards for Automated Teller Machines (ATM) are very appropriate. Such anonymous ATM cards are often issued by offshore banks without name, addresses or credit investigation (any addressing information is accepted) and can be used worldwide for cash withdrawal from ATM from a gold currency account.[24]

In special cases, the golden bars can be transferred physically to the customer (redemption) or are sent by the customer to the provider also physically (bailment, exchange, e.g., by e-gold).[25] Nevertheless, the exchange of gold mostly occurs against a central bank currency through an exchange agent on the Internet. The new customer pays an amount in national central bank currency on the account of a change agent (on the Internet) who credits an amount on the customer account at a provider of gold currencies (after deduction of a commission). Besides credit cards or bank wire transfers, other payment methods are also accepted by many exchange agents, which are hardly controllable again by the supervisory authorities – cash payment, money or postal transfer orders. Some exchange agents try to avoid possible risks of transfers for money laundering or terrorist financing and determine clear policy for the transfers (e.g., E-forexgold.com accepts no checks and money orders for the purchase of virtual gold currencies. Also the payments directly on the accounts from e-forexgold.com are rejected before confirmation of the payment form. E-forexgold.com stresses thereby its role as exchange agent and not as a trustee or nominee for anonymous transfers. The Paybox customers have to be registered before participation in the system by SSL, sending their bank account details to the system provider. Mpay of Vodafone requires no registration; nevertheless, the customers are identified by the contract with the mobile phone company).

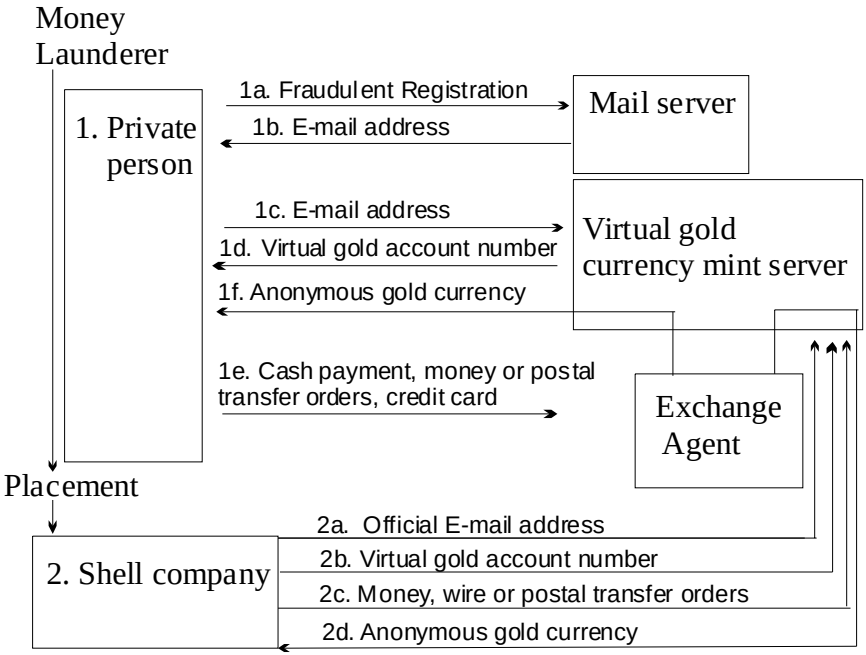


Figure 3: Gold Currencies and Shell Corporations.

The gold currencies can be an attractive electronic payment system for money laundering and terrorist financing taking into consideration their features and their suitability for realization of each phase of money laundering. In the placement phase, the laundered money can be deposited with the help of an exchange agent on the Internet, shell corporations, a trustee or a private transfer system (e.g., Hawala) in form of postal orders, checks, charity payments or payments of other already existing anonymous users of, for example, virtual gold currencies (person-to-person transfers). The network of numerous providers and agents on the Internet creates ideal conditions for the movement of money between private persons and corporations, also from offshore countries (layering phase). Cash withdrawals with the so-called Gold ATM cards, re-exchange with gold currencies and central bank currencies, or paid dividends, e.g., from an e-commerce (business) activity of the money launderer, allows reintegration of the laundered money back into the legal transaction systems (integration).

Prepaid Cards

Prepaid customer smart cards with a rewriteable memory storing electronic currencies are a multipurpose payment instrument that can be used in stationary trade, for person-to-person transfers as well as for e-commerce (using special devices connected to

the computers). Most card products are suitable for money laundering only to a limited extent due to the maximum fixed loading amount for a card (e.g., 200 euros for the GeldKarte) or the account-based character of payments (in this case the customer becomes clearly authenticated). Hence, for terrorist financing the account-unlinked products (with just e-purse functionality) are more attractive; they allow loading of currencies, e.g., in exchange of cash. Example of such a card is the so-called "white GeldKarte" primarily developed for young people without bank account. In spite of the guarantee for perfect anonymity for the users, only approximately 5 to 6 % of all loading processes in the GeldKarte system were conducted with cash.[26] The marginal acceptance of the "white GeldKarte" prevents the wide use of these payment instruments for money laundering since fast increase of transactions with the account-unlinked GeldKarte could immediately trigger a suspicion alert for illegal activities. Nevertheless, in the future such prepaid cards (assuming their wide acceptance) could be used by the money launderers as a possible placement instrument.

Other prepaid cards, such as the coupon card Paysafecard that can be purchased in stationary trade in Austria as anonymous prepaid telephone card, are also suitable for the placement phase of money laundering.[27] With Paysafecard the customer can pay in e-commerce applications for goods or services, while the system provider merely checks the credit balance of the card with the help of the 16-figure PIN (it is printed on the card and is rubbed off by the customer for payments). It would be more difficult to carry out the other money laundering phases, i.e. the layering and the integration phases, because the on-line merchants have to be registered at the system provider and no person-to-person transfers are permitted.

Other prepaid cards with electronic chip, which function as electronic purse, are characterized only by a few features suitable for money laundering – only for a single phase. The transactions are stored by the system provider and can be linked together on the basis of certain identification features such as serial number of payment unities, card number, account number or terminal number. The projects with electronic payment systems are often only of a national character, making them not suitable for international transfers. Nevertheless, to a limited extent, certain characteristics of the electronic payment systems can be used mainly in the layering phase as for example the possibility of person-to-person transfers in the off-line mode with Mondex or interoperability of Proton and VisaCash systems.

Mobile Payment Systems

In general, mobile payment systems are characterized by high flexibility in many application domains (mobile commerce, electronic commerce, stationary trade, as well as for person-to-person transfers) and by high reach-possibility both for customers and on-line suppliers. Other characteristics, such as anonymity and convenience, are

system-specific and often depend closely on the transportation medium used in the system (direct debit or prepaid cards), the charging methods (e.g., telephone bill) or the access methods of the payment application (off-line versus on-line). The on-line and server-based payment systems have payment applications on the server of the system provider and require authentication of the users on a real-time basis. In the off-line systems, the data is stored in the mobile end device (chip card). Therefore, the off-line payment systems are usually not account-based, with prepaid character (payment guarantee for receiver) and anonymous (anonymous prepaid cards; payment occurs when currency units are transferred from the customer prepaid card to the receiver card on a real-time basis as in IrDA, Bluetooth, etc.). The flexibility of mobile cards with stored payment applications guarantees that the cards can be widely exchanged between any different mobile end devices or, in general, between card readers (e.g., dual slot or dual chip end devices).

Nowadays, the off-line mobile payment systems with integrated payment functions in the mobile end devices are supported mainly by many initiatives for development of a universal payment system (e.g., Mobey Forum, Mobile Electronic Transaction). The systems, which are already operating on the market, represent mainly the server-based solution with registered user accounts and with on-line authorization of the transactions.

Furthermore, the on-line payment systems have features that attract the money launderers as for example suitability for cross-border transfers. Paybox has already been implemented in several countries (Austria, Spain, and the Middle East);[28] Simpay plans its first implementation in Belgium, Great Britain, and Spain. Simpay has the highest potential among all mobile payment systems to achieve quickly the critical mass of users and thereby to become target of illegal money activities. The customers of Simpay are identified by their signed contract with a mobile phone company. However, the complex structure of the Simpay chain value with many participants could make difficult a future investigation of money laundering activities. This complex structure is based on the business connections between the Simpay's joint company with Orange, Telefonica Moviles, T-Mobile and Vodafone, the merchants and the new intermediary authority, Mobile Merchant Acquirer (MMA), which can operate after the positive certification by Simpay as a real service provider.[29] The MMA receives the request for payment of the content provider and passes it on through Simpay to the customer. As a response, the payment authorization of Simpay is sent through MMA, after payment confirmation by the customer, to the content provider. The MMA acts as an important administration and contacting center in the Simpay system, which in addition guarantees a high scalability. Account debiting is done through the customer telephone bill (post-paid; theoretically no money laundering risk, only through the involvement of shell companies) or directly by debiting cus-

tomer prepaid card (high risk of money laundering). Later, the wallet function will also be added to the user phone card with several payment options (debit and credit card) which again increases the exchange and disguising combinations for the potential money launderers in the layering phase.

Solutions

The actions/ measures for limiting and combating money laundering can be categorized as organizational, legislative, and technical. The legislative measures include the regulatory measures at national and international level. Many standards and recommendations were already established worldwide for national legislation (e.g., Forty Recommendations of FATF, Risk Management Principles for Electronic Banking of the Basel Committee on Banking supervision), which define the concrete measures against money laundering or for detection of suspicious activities. Nevertheless, the regulatory solutions remain ineffective so far as countries or territories exist without regulation of the money laundering activities. The FATF organization publishes at certain time intervals a list of the not-cooperating countries and territories, which act as a shelter for many criminal networks.[30] The list of the not-cooperating countries is an important indication for the supervisory authorities investigating possible involvement of suspected persons in criminal activities. Unfortunately, the fact that a country is in the list is often only informative and does not limit the dimension of money laundering (shell companies in offshore territories). A regulatory solution for this problem could be an administrative restriction of the economic relations with a non-cooperating country. However, it seems unrealistic that such restrictions or sanctions will be efficient due to the unlimited communication possibilities of the Internet (closing of one location is followed by opening of another location through Internet).

The organizational measures include different methods for testing (checking all transactions that exceed a threshold value) or some calculation methods as for example the net value (worth) measuring the difference between assets and liabilities of a suspected person (its increase has to be a result of legal income), supervision systems (e.g., Suspicious Activity Reports filed by financial institutions in the U.S. and EU), and early warning systems of potential fraud risks according to the risk management methodology.[31] Record keeping of transaction and customer data at the system providers also belongs to the organizational measures and solutions. Traceability of transfers and net traffic often present a significant problem for the supervisory authorities due to the fact that the data of the customers and consequently their privacy are protected by the Internet Service Providers (the ISP can also operate from not-cooperative countries). Other postulated measures against potential money laundering activities, such as setting a maximum loading value for a prepaid card, demand for bank involvement in each transaction (no peer-to-peer transfers) or restriction of the

use to national level (no international payments), are neither realistic (e-Commerce) nor innovative or efficient.

The technical solutions represent the most interesting and, at the same time, efficient part of the solution approaches regarding authentication of users or traceability of transactions, for example. Many technical solutions that could limit money laundering have already been developed:

- Digital signatures and certificates based on the Public Key Infrastructure (PKI) – This is a hierarchical certification technology based on asymmetrical cryptography for authentication, confidentiality and integrity of data, as well as for non-repudiation of electronic transactions. The generation of key pairs and the confidential distribution of public keys with certificates are also important for combating money laundering conducted by means of different electronic payment systems due to the fact that it secures practically a worldwide authentication of the transaction participants (worldwide interoperability).
- Special cryptographic protocols such as for example the off-line payment protocol of Chaum, Fiat and Naor, developed on the basis of the "blind signature" for an anonymous off-line payment procedure. In contrast to the online payment procedures, where the verification processes and payments should be processed between the merchants and banks in real time, the electronic checks and coins are collected in off-line payment protocols by the merchant first and then are submitted in aggregated form at the bank of the merchant in the end of a given period. A fraudulent case (e.g., double spending) enhances enormously the probability of disclosure of customer identity.

Conclusion

Based on the supported features, the gold currencies are often favored for potential application in money laundering. Other electronic payment systems have characteristics attractive to money laundering to a different degree. Being suitable for a single phase of money laundering, a combination of different payment systems would enhance their suitability for the whole money laundering process – prepaid cards for the placement phase, mobile payment systems for the layering phase and virtual gold currencies for the integration phase. The number of possible combinations for illegal activities increases enormously when other traditional techniques are involved (e.g., transfers through Hawala, intermediation of shell companies and nominees or investment in legal financial products). Hence, money laundering is a complex and continuously changing process; however, the dimension of illegal money activities can be limited by suitable measures and approaches (primarily technical solutions) or detected by early warning and supervision systems.

Notes:

[1] "Directive 2001/97/EC of the European Parliament and of the Council of 4 December 2001 amending Council Directive 91/308/EEC on prevention of the use of the financial system for the purpose of money laundering," *Official Journal of the European Communities*, <http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/l_344/l_34420011228en00820082.pdf> (05 September 2005).

[2] John Madinger and Sydney A. Zalopany, *Money Laundering. A Guide for Criminal Investigators* (Boca Raton, FL: CRC Press, 1999).

[3] Madinger and Zalopany, *Money Laundering. A Guide for Criminal Investigators*.

[4] Madinger and Zalopany, *Money Laundering. A Guide for Criminal Investigators*.

[5] *Report on Money Laundering and Terrorist Financing Typologies 2004-2005*, FATF-XV (Financial Action Task Force on Money Laundering (FATF), 10 June 2005), <http://www.fatf-gafi.org/dataoecd/16/8/35003256.pdf > (05 December 2005).

[6] *2003 National Money Laundering Strategy* (U.S. Treasury: The Office of Terrorism and Financial Intelligence (TFI), 2003), <http://www.treas.gov/offices/enforcement/publications ml2003.pdf> (05 September 2005).

[7] *2003 National Money Laundering Strategy*.

[8] *2003 National Money Laundering Strategy*.

[9] *2003 National Money Laundering Strategy*.

[10] Madinger and Zalopany, *Money Laundering. A Guide for Criminal Investigators*.

[11] Madinger and Zalopany, *Money Laundering. A Guide for Criminal Investigators*.

[12] Madinger and Zalopany, *Money Laundering. A Guide for Criminal Investigators*.

[13] e-gold Ltd. is a Nevis Corp. and the bullion backing the e-metal currencies is held by the e-gold Bullion Reserve Special Purpose Trust in the Iceland of Bermudas, <http://www.e-gold.com/contracts/egold-spt-111899.htm> (05 September 2005).

[14] *Report on Money Laundering and Terrorist Financing Typologies 2003-2004*.

[15] Karl-Heinz Ketterer, *Internet-Zahlungssysteme aus der Sicht der Verbraucher – Ergebnisse einer Online-Umfrage IZV6* (Erhebung des Instituts für Wirtschaftspolitik und Wirtschaftsforschung der Universität Karlsruhe, May 2003), 1-13, <http://www.iww.uni-karlsruhe.de/izv/pdf/izv6_auswertung.pdf> (14 July 2004).

[16] Eberhard Stickel and Krzysztof Woda, "Electronic Money," in *E-Finance*, ed. Erhard Petzel (Hrsg.) (Gabler Verlag, 2005), 831-860.

[17] David Chaum, "Sicherheit ohne Identifizierung: Scheckkartencomputer, die den Großen Bruder der Vergangenheit angehören lassen - Zur Diskussion gestellt," *Informatik-Spektrum* 10, no. 5 (1987): 262-277.

[18] Chaum, "Sicherheit ohne Identifizierung: Scheckkartencomputer, die den Großen Bruder der Vergangenheit angehören lassen."

[19] Anonymous hosting (e.g. <http://www.katzglobal.com/hosting/hosting.html> (05 September 2005)).

[20] Susanne Leist and Krzysztof Woda, "Analyse der Erfolgsfaktoren mobiler Zahlungssysteme," (Frankfurt (Oder): Europa-Universität Viadrina, Arbeitsbericht Nr. 217, July 2004), 1-23.

[21] e-gold, "What is e-gold?" <http://www.e-gold.com/unsecure/qanda.html> (05 September 2005).

[22] Gold-cash.biz., FAQ, <http://www.gold-cash.biz/faq.php> (05 September 2005).

[23] Gold-cash.biz., FAQ.

[24] Gold-ATM, Debit and Prepaid Cards for Digital Currencies Users, <http://www.gold-atm.biz/cards.php> (05 September 2005).

[25] e.gold, "e-gold Account User Agreement," (last modified on 20th December 2003), <http://www.e-gold.com/unsecure/terms.htm> (05 September 2005).

[26] Stickel and Woda, "Electronic Money."

[27] Stickel and Woda, "Electronic Money."

[28] Paybox Austria AG, FAQ, <http://www.paybox.at/238.php#> (05 September 2005).

[29] Simpay, FAQ's, <http://www.simpay.com/faqs.php> (05 September 2005).

[30] The Financial Action Task Force on Money Laundering (FATF), "Annual Review of Non-Cooperative Countries or Territories," 2 July 2004, <http://www.fatf-gafi.org/dataoecd/3/52/33922473.PDF> (05 September 2005).

[31] *2003 National Money Laundering Strategy*.

KRZYSZTOF WODA received his Ph.D. degree in Economics from the Viadrina University in Frankfurt (Oder), Germany, in 2003. His current research area interests include modern techniques for money laundering and terrorism financing, the role of electronic payment systems in supporting illicit money transfers as well as the development of quantitative methods for detecting illicit money transfers and financial computer crime. *Address for Correspondence*: Dept. of Information Systems, Viadrina University, Postfach 1786, 15207 Frankfurt (Oder), Germany; *E-mail*: kwoda@euv-frankfurt-o.de

CREDIT CARD FRAUD DETECTION USING SELF-ORGANIZING MAPS

Vladimir ZASLAVSKY and Anna STRIZHAK

Abstract: Nowadays, credit card fraud detection is of great importance to financial institutions. This article presents an automated credit card fraud detection system based on the neural network technology. The authors apply the Self-Organizing Map algorithm to create a model of typical cardholder's behavior and to analyze the deviation of transactions, thus finding suspicious transactions.

Keywords: Payment System, Transaction, Fraud Detection, Self Organizing Map.

Introduction

Any payment system (PS) is characterized by a high level of risk in its different domains caused by great volume and number of operations, a lot of complex relations between clients and increasing speed of data transmission. In order to manage risks, PS should develop and use mathematical models to determine suspicious/ risky situations, establish scenarios for its development and evaluate consequences of their realization.

Nowadays, one of the most important and challenging problems for PS and its members becomes credit card fraud – the illegal use of credit cards by third parties. Fraudulent electronic transactions have already been a significant problem that grows in importance as the number of access points grows, especially when transactions are fully enabled on the Internet for electronic commerce.[1] Fraud detection and prevention methods are being continuously improved; however, banks all over the world lose millions of US dollars each year. Experts from Visa International predict annual growth of some fraud types up to 65%.[2] According to the Association for Payment Clearing Services, fraud losses per one credit card are expected to increase up to $11 by year 2008.

Credit card fraud is perpetrated in various ways and, generally, it is based on unauthorized write-off of funds from accounts of banks' clients – cardholders.[3] Credit card

fraud can be broadly categorized as application, 'missing in post,' stolen/ lost card, counterfeit card and 'cardholder not present' fraud.[4,5] The number of different variants of fraud is great enough, they change continuously, and new ways of fraud appear as far as protection of credit cards is improved. In the past, banks—members of PS—had solved fraud prevention problems by means of organizational measures: limits on number and amounts of cardholder's operations, monitoring of transactions in high risk countries, use of various methods for card verification, etc.[6,7] According to the theory and practice of risk management, each bank has to implement special measures in order to detect and prevent fraud in time. International payment systems, such as Visa International and MasterCard International, demand from their banks-members implementation of various measures in order to reduce the number of fraudulent operations in PS and they recommend turning from reaction methods to proaction methods for dealing with fraudulent operations with cards.[8]

In order detection and prevention of fraud to be effective, banks should develop and use in their practice special fraud detection systems targeted to reveal among stream of transactions the fraudulent ones and thus to prevent banks as well as their clients from the illegal activities of fraudsters.[9] One should develop special rules for analysis, models and methods that can describe fraudulent behavior, rules and methods of fraud prevention and generation of different decision alternatives in risky situations. Mathematical models and algorithms for classification and pattern recognition problems could be considered as a basis for such systems.

In this article, models and algorithms for detection of fraudulent operations in PS are proposed.

Problem Definition

Banks-members of PS keep databases (DB) of all their cards issued in PS. For each card, the database holds card number, account number, operational limits, current state of account (account balance) and some other data about the cardholder. Let $C_n = \{c_1, \ldots, c_k, \ldots, c_{k_n}\}$ be a set of records in DB that contains information about all cards used in PS; $c_k = (c_1^k, c_2^k, \ldots, c_s^k)$ is a record in DB, which contains information about the card c_k and its component c_1^k is a unique card number.

The processing centre of PS constantly receives information about operations carried out by cardholders (such as cash withdrawal, balance statement, purchase, etc.). The information about an operation is represented in the form of transaction message (in accordance with ISO 8583) that includes various operation parameters: card number, amount of transaction, date and time of transaction, type of operation, number of terminal, retailer identifier, etc.

Let $X_n = \{x^1,\ldots,x^i,\ldots,x^n\}$ be the set of transactions carried out in PS up to some moment t_n, where $x^i = (x_1^i,\ldots,x_j^i,\ldots,x_m^i)$ is the message about i-th transaction. Each component x_j^i holds numerical (for example transaction amount) or symbolic information (operation type, retailer code, terminal, city, etc.). An analogue (numerical) component $x_j^i \in R$. A symbolic component x_j^i (which are a majority) takes its values from some discrete set $x_j^i \in T_j = \{\tau_j^1,\ldots,\tau_j^s,\ldots,\tau_j^{s_j}\}$, where τ_j^s – s-th unique value of x_j^i. For example, component x_j^i – "terminal type" may take its values from the set T_j ={'ATM', 'POS'}, where 'ATM' means that transaction was initiated on ATM and 'POS' means that transaction was carried out in Point-Of-Sale. A symbolic field may contain from at least two values (e.g. the type of credit card) up to several hundred thousand values (as merchant code, for instance).

As time goes by the size of the set X_n grows as new transactions are executed in PS. Let us suppose that the transactions executed after moment t_n up to t_{n+k} are new ones and denote them as $x^{n+1}, x^{n+2},\ldots, x^{n+k}$.

Let $X_{c_k} = \{x^i \mid x_1^i = c_k, x^i \in X_n\}$ be the set of transactions $X_{c_k} \subseteq X_n$, executed in PS using card $c_k \in C_n$ up to moment t_n.

The problem of detection of fraudulent transactions in PS lies in classifying a new transaction $x^{n+1} = (x_1^{n+1};\ldots;x_j^{n+1};\ldots;x_m^{n+1})$ using the information about transactions X_n performed earlier and the appropriate record C_n in DB. To classify means to determine the class (fraudulent or legal) to which a transaction belongs.

Problem Analysis

A variety of methods can be applied to solve the presented problem. The simplest method used in the earliest transaction monitoring systems was control of transaction parameters $x^{n+1} = (x_1^{n+1};\ldots;x_j^{n+1};\ldots;x_m^{n+1})$; these transaction variables were compared to established levels/ thresholds.[10] The thresholds $L_n = \{l_1,\ldots,l_s,\ldots,l_{s^*}\}$ are the critical levels for the most significant parameters $x_{j_1}^{n+1},\ldots,x_{j_s}^{n+1},\ldots,x_{j_{s^*}}^{n+1}$ set by domain experts based on their experience and knowledge of the data domain. For example, if $x_{j_s}^{n+1} \geq (\leq) l_s$ for some s, then transaction x^{n+1} is classified as fraudulent.

Another approach is to apply some set of rules $R = \{R_1,\ldots,R_i,\ldots,R_{i^*}\}$ for verification of transaction x^{n+1}. Such rules describe fraudulent behavior and should be de-

fined by experts on the basis of analysis of wide range of transactions.[11] Each rule $R_i,\ i = 1,\ldots,i^*$ is a structure "$R_i : IF\ < Condition_i >\ THEN\ Transaction\ x^{n+1}\ is\ fraudulent$," which means that transaction x^{n+1} is considered to be fraudulent if it satisfies the condition of some rule $R_i \in R$.

The described methods are rather simple; however, they suffer from the following shortcomings: they detect only fixed suspicious situations established beforehand and do not take into account the variable nature of fraud; they do not consider the individual characteristics of cardholders' behavior; the control of such rule-based system is rather complex task for the expert.[12]

The authors argue that a more efficient way is to use methods such as neural networks, fuzzy logic, theory of probability, statistics and other data mining methods for automatic creation of fraudulent transaction patterns on the basis of transactions' history, its constant update and checking of all new transactions for deviation.[13,14]

In this article, the authors propose to use a type of neural network algorithm—the Self Organizing Map (SOM)—for transactional data analysis and detection of fraudulent behavior.

Principles of Transaction Classification

The described fraud detection task can be considered as pattern recognition or classification problem.[15] The set X_n of all transactions in PS is divided into two disjoint subsets: legal transactions $X_n^l \subseteq X_n$ and fraudulent ones $X_n^f \subseteq X_n$, $X_n^l \cap X_n^f = \varnothing$. If we assume that the numerical images (i.e., points in some multidimensional space) of fraudulent and legal transactions belong to different areas in this space, then it is possible to make a decision about the image of a new transaction x^{n+1}.[16]

The following two hypotheses are considered as a basis for such classification.

- *Hypothesis* H_l: Transaction $x^{n+1} = (x_1^{n+1},\ldots,x_m^{n+1})$ on card c_k is similar to all previous transactions from the set X_{c_k}, which were carried out earlier by the cardholder. If hypothesis H_l is confirmed for transaction x^{n+1}, then the transaction x^{n+1} is classified as legal and included into the set X_n^l.
- *Hypothesis* H_f: Transaction $x^{n+1} = (x_1^{n+1},\ldots,x_m^{n+1})$ is similar to earlier executed fraudulent transactions $X_n^f = \{x^i - considered\ fraudulent \mid x^i \in X_n\}$.

If hypothesis H_f is confirmed for transaction x^{n+1}, then transaction x^{n+1} is classified as fraudulent and included into the set X_n^f.

It seems reasonable to use neural network techniques for clustering and classification in order to check the proposed hypotheses H_l and H_f. The main idea is to create (and later recognize) pattern of "legal cardholder" and pattern of "fraudster" on the basis of neural network "learning" from the transactions X_n executed earlier and to develop "rules" of cardholder's behavior and fraudster's behavior. Learning algorithms allow the system to follow the cardholder's behavior and self-adapt to changes in it. If a transaction does not correspond to the pattern of "legal cardholder" or is similar to the "fraudulent" pattern it is classified as suspicious for fraud.

One of the most suitable methods of data analysis for the described problem is the Self-Organizing Map (SOM), an unsupervised neural network.[17] Neural networks of this type are often used to solve a great variety of problems from recovery of missing data to data analysis and retrieval of patterns.

Transaction Analysis with SOM

SOM Main Principles

SOM is a neural network with feed-forward topology and an unsupervised training algorithm that uses a self-organizing process to configure its output neurons according to the topological structure of the input data.[18] The self-organizing process is based on competitive training and consists in tuning the weights $w^i = (w_1^i, w_2^i, \ldots, w_q^i)$, $i = \overline{1;d}$ (q is the dimension of the input vector $a^j = (a_1^j, a_2^j, \ldots, a_q^j)$) by a method of progressive approximation using weights' values from the previous iteration [19]: $w^i(t+1) = w^i(t) + h(t) \cdot (a^j(t) - w^i(t))$; here t is the iteration number and $h(t)$ is function of the radius of the considered neighborhood. As a result from the learning process, a matrix of weights of the input connections of neurons is obtained, which allows to group subsets of input data and form prototypes (profiles):

$$W = \begin{pmatrix} w_1^1 & w_1^2 & \cdots & w_1^d \\ w_2^1 & w_2^1 & \cdots & w_2^1 \\ \cdots & \cdots & \cdots & \cdots \\ w_q^1 & w_q^1 & \cdots & w_q^1 \end{pmatrix}.$$

Testing the Hypotheses H_l and H_f

The authors propose to use the SOM for testing the hypotheses H_l and H_f in the following way.

Testing the hypothesis H_l for transaction x^{n+1} on card c_k includes the following steps:

1. Create a typical cardholder's behavior model (pattern) W_{c_k} on the basis of past transactions $X_{c_k} \in X_n^l$ executed earlier with the card c_k. This model W_{c_k} is represented as a SOM, which is the cardholder's profile.
2. Determine the similarity rate $\delta(x^{n+1}, W_{c_k})$ of transaction x^{n+1} to profile W_{c_k}.
3. Hypothesis H_l is accepted if the similarity rate $\delta(x^{n+1}, W_{c_k})$ satisfies the condition $\delta(x^{n+1}, W_{c_k}) \le \varepsilon_l$, where ε_l is some parameter.

Testing the hypothesis H_f for transaction x^{n+1} is performed in a similar to the previous scheme way:

1. Create a typical fraudster's behavior model (pattern) W_f on the basis of fraudulent transactions X_n^f executed earlier in PS and determined as fraudulent. This model W_f is also represented as a SOM, which is the fraudster's profile.
2. Determine the similarity rate $\delta(x^{n+1}, W_f)$ of transaction x^{n+1} to profile W_f.
3. Hypothesis H_f is accepted if the similarity rate $\delta(x^{n+1}, W_f)$ satisfies the condition $\delta(x^{n+1}, W_f) \le \varepsilon_f$, where ε_f is some parameter.

The authors describe the algorithm of profile creation and calculation of similarity rate $\delta(x^{n+1}, W_{c_k})$ for hypothesis H_l below. (The scheme for testing of the hypothesis H_f is similar.)

Creation of Cardholder's Profile

Cardholder's profile W_{c_k} is a typical cardholder behavior model, which represents a generalized pattern of the transactions executed earlier by the holder of card c_k. This model is a special structure neural network trained by the SOM algorithm on the basis of the set of transactions $X_{c_k} \in X_n^l$ and is able to recognize typical transactions of a legal cardholder.

In the process of building the self-organizing map, the authors suggest not to use $x^i \in X_{c_k}$ directly, but rather vectors $p^i = (p_1^i, \ldots, p_m^i, \ldots, p_M^i) \in P_{c_k}$, $i = \overline{1;v}$, obtained from the vectors $x^i \in X_{c_k}$ and the parameters of the current state of the card account $c_k = (c_1^k, \ldots, c_s^k)$. To build the set P_{c_k}, the authors apply a function $\varphi : X_{c_k} \to P_{c_k}$ that is a composition of functions $\varphi_0, \varphi_1, \ldots, \varphi_{M-m}$ described later.

The components of the vector $p^i \in P_{c_k}$ can be divided into two groups:

1) The characteristics $p_1^i, \ldots, p_m^i$ of the current transaction $x^i \in X_{c_k}$, which are in fact the values of the appropriate components $x^i \in X_{c_k}$ to which a function φ_0 is applied:

$$p_j^i = \varphi_0(x_j^i) = \begin{cases} x_j^i, \text{if } x_j^i \text{ is a numeric characteristic} \\ I(x_j^i), \text{if } x_j^i \text{ is a symbolic characteristic} \end{cases}, \quad j = \overline{1;m}.$$

The function $I(x_j^i)$ is built using a statistics-based indexing method. Each symbolic value is associated with a numeric index according to its frequency in the training set, which is later used in the training of the neural network as described below.

- The frequency $F(\tau_j^s)$ in the training set X_{c_k} of each unique value $\tau_j^s \in T_j$ of a symbolic parameter x_j^k is calculated as follows:

$$F(\tau_j^s) = \sum_{k=1}^{v} \chi_k(\tau_j^s), \text{ where } \chi_k(\tau_j^s) = \begin{cases} 1, \text{if } x_j^k = \tau_j^s \\ 0, \text{if } x_j^k \neq \tau_j^s \end{cases}, s = \overline{1;s_j}.$$

- The set of unique values $T_j = \{\tau_j^1, \ldots, \tau_j^s, \ldots, \tau_j^{s_j}\}$ is ordered according their decreasing frequency $F(\tau_j^s)$, $s = \overline{1;s_j}$, $F(\tau_j^1) \geq F(\tau_j^2) \geq \ldots \geq F(\tau_j^{s_j})$.

- Each unique symbolic value $\tau_j^s \in T_j$ is associated with a numeric index $I_{\tau_j^s}$:

$$I_{\tau_j^1} = 1; I_{\tau_j^s} = I_{\tau_j^{s-1}} + 1, s = \overline{2; s_j} .$$

- Then, the function $I(x_j^i)$ is defined as:

$$I(x_j^i) = I_{\tau_j^s} \text{ when } x_j^i = \tau_j^s .$$

Such an indexation allows maintaining the initial relative importance of the unique values and the correlation between them.

Examples of characteristics $p_1^i, \ldots, p_m^i$ are transaction amount, transaction time, transaction type, terminal number, terminal city, etc.

2) The characteristics $p_{m+1}^i, p_{m+2}^i, \ldots, p_M^i$ of the transaction history on card c_k, calculated using the functions $\varphi_0, \varphi_1, \ldots, \varphi_{M-m}$ on the basis of the set of transactions X_{c_k}, executed earlier with card c_k up to moment t_n:

$$p_{m+1}^i = \varphi_1(x^1, x^2, \ldots, x^i), \quad p_{m+2}^i = \varphi_2(x^1, x^2, \ldots, x^i), \quad p_M^i = \varphi_{M-m}(x^1, x^2, \ldots, x^i) .$$

Examples of characteristics $p_{m+1}^i, p_{m+2}^i, \ldots, p_M^i$ are: number of transactions carried out during a period of D hours, cumulative amount of transactions during D hours, number of terminals used by the cardholder during D hours, etc.

The resultant set $P_{c_k} = \{p^1 = (p_1^1, \ldots, p_M^1), \ldots, p^v = (p_1^v, \ldots, p_M^v)\}$ is the training set used for creating cardholder's profile W_{c_k}.

As a result of SOM learning [20] with the training set P_{c_k} a matrix of neuron weights of the trained map is obtained, which is actually the cardholder's profile for card c_k: $W_{c_k} = \left\| w_k^s \right\|_{\substack{s=\overline{1;d} \\ k=\overline{1;M}}}$. The weight vectors $w^i = (w_1^i, \ldots, w_M^i), i = \overline{1;d}$ specify the most typical values of the components of vector $p^i = (p_1^i, \ldots, p_M^i)$, which are present in the training set P_{c_k}.

In result, for each transaction $x^i \in X_{c_k}$ there is a certain j-th cell on the SOM such that $\left\| x^i - w^j \right\| = \min_{k=1;2;\ldots;d} \left\| x^i - w^k \right\|$.

Calculation of Transaction Similarity Rate to Profile

Once the neural network learning process is over, every new transaction x^{n+1} on card c_k is checked for similarity to profile W_{c_k}.

The similarity rate $\delta(x^{n+1}, W_{c_k})$ of transaction $x^{n+1} = (x_1^{n+1}, \ldots, x_m^{n+1})$ to profile W_{c_k} can be determined as the deviation of the vector $p^{n+1} = \varphi(x^{n+1})$ from the nearest cell of the map W_{c_k}, or in other words as the minimum of the distances between the vector $p^{n+1} = (p_1^{n+1}, \ldots, p_M^{n+1})$ and the vectors of neurons' weights $w^1, \ldots, w^d$:

$$\delta(x^{n+1}, W_{c_k}) = \min_{i=1;2;\ldots;d} \left\| p^{n+1} - w^i \right\|.$$

The most commonly used type of distance measure is the Euclidean distance:

$$\left\| p^{n+1} - w^i \right\| = \sqrt{\sum_{k=1}^{M} (p_k^{n+1} - w_k^i)^2}.$$

However, in some applications more complex distance measures are required. It depends mainly on specific characteristics of the data space and the expected results:

- *Squared Euclidean distance*: $\left\| p^{n+1} - w^i \right\| = \sum_{k=1}^{M} (p_k^{n+1} - w_k^i)^2$. This distance measure place progressively greater weight on objects that are further apart;
- *Manhattan distance*: $\left\| p^{n+1} - w^i \right\| = \sum_{k=1}^{M} \left| p_k^{n+1} - w_k^i \right|$. In most cases, this distance measure yields results similar to the simple Euclidean distance. However, the effect of single large differences (outliers) is dampened;
- *Chebychev distance*: $\left\| p^{n+1} - w^i \right\| = \max_{k=1,\ldots,M} \left| p_k^{n+1} - w_k^i \right|$. This distance measure may be appropriate in cases when one wants to define two objects as "different" if they are different on any one of the dimensions/ coordinates;
- *Power distance*: $\left\| p^{n+1} - w^i \right\| = \left(\sum_{k=1}^{M} \left| p_k^{n+1} - w_k^i \right|^p \right)^{1/r}$. Sometimes one may want to increase or decrease the progressive weight that is placed on dimensions on which the respective objects are very different;

- *Percent disagreement*: $\left\| p^{n+1} - w^i \right\| = \frac{1}{l} \sum_{k=1}^{M} \psi(p_k^{n+1}; w_k^i)$, where

$$\psi(p_k^{n+1}; w_k^i) = \begin{cases} 1, \text{if } p_k^{n+1} \neq w_k^i \\ 0, \text{if } p_k^{n+1} = w_k^i \end{cases}$$, which is useful for categorical features.

Algorithm

The proposed method for transaction analysis is represented as a block diagram in Figure 1. The process of transaction monitoring consists of three stages: data accumulation, training (building of cardholder's profile) and control of transactions.

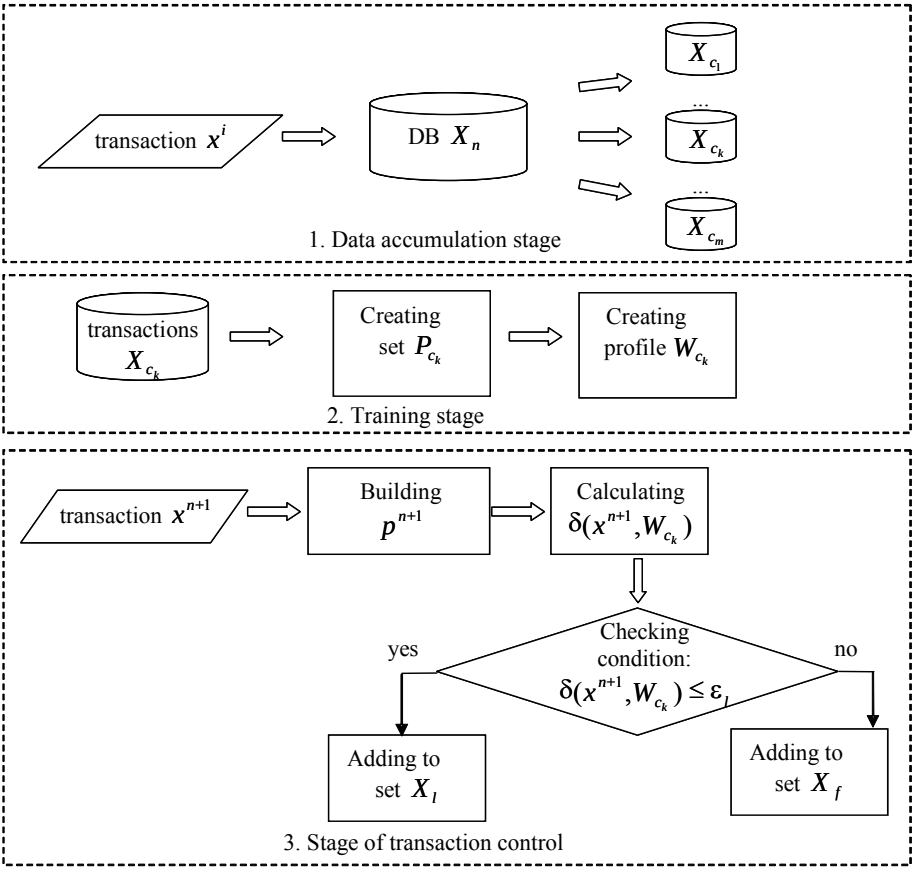


Figure 1: Block Diagram of Transaction Monitoring Algorithm.

At the stage of data accumulation, the data about the transactions on card c_k are collected in the database DB. If the size of X_{c_k} exceeds some predefined level, sufficient to build an adequate profile, then the monitoring process goes to stage two.

At stage two, the training stage, the cardholder's profile W_{c_k} is created as follows:

- The set P_{c_k} is built using the function φ;
- The neural network is trained on the basis of set P_{c_k};
- The profile $W_{c_k} = \left\| w_k^s \right\|_{\substack{s=\overline{1;d} \\ k=\overline{1;M}}}$ is built as a result of training.

After the training stage, the process goes to the stage of transaction control, which consists of the following:

- The vector p^{n+1} is built applying the function φ to every new transaction x^{n+1}: $p^{n+1} = \varphi(x^{n+1})$;
- The deviation of the current transaction x^{n+1} from the profile W_{c_k} (created at the training stage) is calculated: $\delta_0 = \delta(x^{n+1}, W_{c_k})$;
- The value δ_0 is compared with the threshold ε_l fixed for the profile W_{c_k} (ε_l is a boundary value for the degree of similarity of the transactions on card c_k to the profile W_{c_k}. It makes it possible to cut off transactions that deviate from the early established norm and to control the accuracy of fraud detection.);
- If $\delta_0 \le \varepsilon_l$ then transaction x^{n+1} is considered typical/ legal and the vector x^{n+1} is added to the set $X_l = X_{c_k}$;
- If $\delta_0 > \varepsilon_l$ then transaction x^{n+1} is considered suspicious for fraud and is added to the set X_f for further expert analysis.

Example

This section will illustrate the proposed approach to fraud detection. Transactional data is confidential information; therefore, the initial data set was simulated (a list of real transaction parameters and a range of their values were used). The following characteristics (features) were chosen to analyze the transactions: p_1 – transaction amount, p_2 – transaction type, p_3 – terminal identifier, p_4 – city, p_5 – country,

p_6 – number of transactions over the last 48 hours, p_7 – accumulated amount of transactions over the last 48 hours, and p_8 – number of terminals used in the last 48 hours.

A number of credit cards with different characteristics of cardholder's behavior were examined in order to explore the dependence of the constructed model of cardholder's behavior on the transaction similarity degree and to define the required minimum number of transactions in the training set (see Table 1). A small number of transactions (100) in the training set were used intentionally considering the specificity of the Ukrainian credit card market. Most of the cards are characterized by a low number of transactions per month and thus poor transaction history.

Table 1: Characteristics of Credit Cards.

Card #	*Total Amount of Transactions* [21]	*Type of Cardholder Behavior*
Card #1	100+10	All transactions are similar
Card #2	100+10	Most of the transactions are similar, but rare atypical transactions appear
Card #3	90+10	Various transactions

Several models of typical cardholder behavior were built using different number of transactions in the training set. Computational results are given in Table 2.

In the table, ε denotes the average error in the set and $\varepsilon_{\max}$ – the maximum error in the set.

As could be seen from Table 2, the accuracy of detection of fraudulent and legal transactions (test and validation sets) increases with the increase of the number of transactions in the training set. For Card #1, acceptable recognition accuracy (ε=0.0068) has already been reached with 30 transactions in the initial set; for Card #2 and Card #3, with more heterogeneous cardholders behavior, similar recognition accuracy is reached with 60 and 90 transactions, respectively.

Two-dimensional Kohonen maps were built for cardholder behavior model. Figure 2 depicts the distance matrix and the clusters formed for the model of cardholder behavior for Card #3.

The clusters on the map show that cardholder's behavior is characterized by three pronounced types, which were named "Typical ATM transactions," "Typical POS transactions," and "Rare/ anomalous transactions." After processing the anomalous transactions for Card #3, it was observed that their deviation from the model of typi-

cal behavior greatly exceeded the error of recognition of legal transactions (as illustrated in Figure 3). Moreover the more anomalous a transaction is, the greater its deviation from the model. So, this characteristic can be used as degree of suspiciousness of a transaction.

Conclusion

This article has proposed a new approach to transaction monitoring and credit card fraud detection using the Self-Organizing Map algorithm. It enables automated creation of transaction monitoring rules in a learning process and makes possible their continuous improvement in an environment of dynamically changing information in an automated system.

Table 2: Results from Experiments.

Card #	*Set*	*Number of Transactions in the Initial Set*					
		30+10		*60+10*		*90+10*	
		ε	$\varepsilon_{\max}$	ε	$\varepsilon_{\max}$	ε	$\varepsilon_{\max}$
1	Training	7.25E-10	8.71E-9	5.18E-10	9.16E-9	0.0013	0.0263
	Test	0.0068	0.0340	0.0039	0.0395	0.0034	0.0339
	Validation (legal)	0.0068	0.0909	0.0015	0.0395	0.0030	0.0250
	Validation (fraud)	0.9190	1.8136	0.8916	1.7192	0.8917	1.7192
2	Training	0.0007	0.0099	0.0042	0.0351	0.0027	0.0404
	Test	0.0299	0.1495	0.0049	0.0404	0.0381	0.1575
	Validation (legal)	0.1069	0.2257	0.0084	0.1575	0.0456	0.0923
	Validation (fraud)	0.8334	1.7198	0.7840	1.6757	0.7235	1.6757
3	Training	0.0136	0.0754	0.0202	0.0879	0.0224	0.1502
	Test	0.0659	0.3432	0.0508	0.3066	0.0576	0.1371
	Validation (legal)	0.0536	0.3951	0.0321	0.1796	0.0345	0.1252
	Validation (fraud)	0.6930	1.4015	0.6458	1.4852	0.6575	1.5122

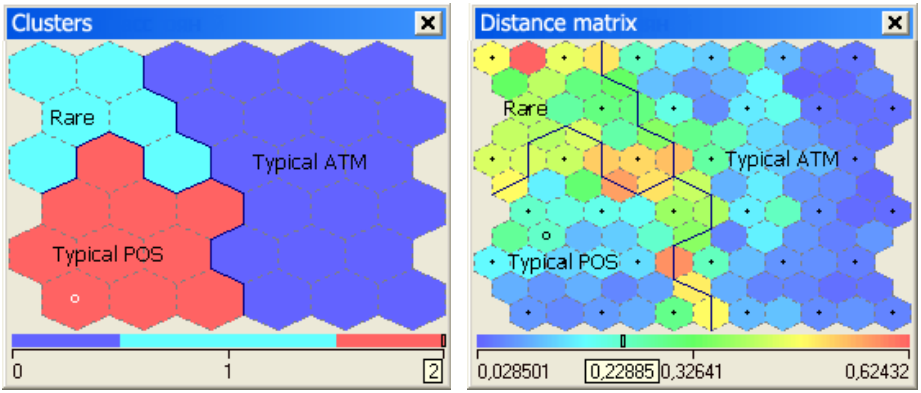
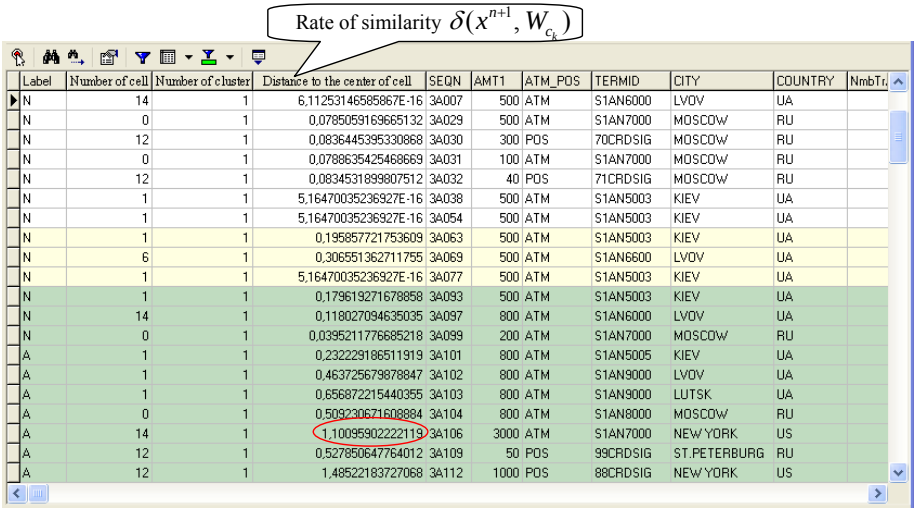


Figure 2: Cardholder's Behavior Model (Card #3).

The advantages of the proposed approach are: the success of the algorithm does not depend on statistical assumptions about data distribution; it deals successfully with noisy data; the method allows modification of the model as new transactions are added and it does not require *a priori* information besides some set of transactions performed by the cardholder; the achieved accuracy of the produced rules is stable (in contrast to the changing concentration and attention of the experts, for example as a



Label	Number of cell	Number of cluster	Distance to the center of cell	SEQN	AMT1	ATM_POS	TERMID	CITY	COUNTRY	NmbTr
N	14	1	6,112531465858676E-16	3A007	500	ATM	S1AN6000	LVOV	UA	
N	0	1	0,0785059169665132	3A029	500	ATM	S1AN7000	MOSCOW	RU	
N	12	1	0,0836445395330868	3A030	300	POS	70CRDSIG	MOSCOW	RU	
N	0	1	0,0788635425468669	3A031	100	ATM	S1AN7000	MOSCOW	RU	
N	12	1	0,0834531899807512	3A032	40	POS	71CRDSIG	MOSCOW	RU	
N	1	1	5,16470035236927E-16	3A038	500	ATM	S1AN5003	KIEV	UA	
N	1	1	5,16470035236927E-16	3A054	500	ATM	S1AN5003	KIEV	UA	
N	1	1	0,195857721753609	3A063	500	ATM	S1AN5003	KIEV	UA	
N	6	1	0,306551362711755	3A069	500	ATM	S1AN6600	LVOV	UA	
N	1	1	5,16470035236927E-16	3A077	500	ATM	S1AN5003	KIEV	UA	
N	1	1	0,179619271678858	3A093	500	ATM	S1AN5003	KIEV	UA	
N	14	1	0,118027094635035	3A097	800	ATM	S1AN6000	LVOV	UA	
N	0	1	0,0395211776685218	3A099	200	ATM	S1AN7000	MOSCOW	RU	
A	1	1	0,232229186511919	3A101	800	ATM	S1AN5005	KIEV	UA	
A	1	1	0,463725679878847	3A102	800	ATM	S1AN9000	LVOV	UA	
A	1	1	0,656872215440355	3A103	800	ATM	S1AN9000	LUTSK	UA	
A	0	1	0,509230671608884	3A104	800	ATM	S1AN8000	MOSCOW	RU	
A	14	1	1,10095902222119	3A106	3000	ATM	S1AN7000	NEW YORK	US	
A	12	1	0,527850647764012	3A109	50	POS	99CRDSIG	ST.PETERBURG	RU	
A	12	1	1,48522183727068	3A112	1000	POS	88CRDSIG	NEW YORK	US	

Figure 3: Anomalous Transactions on Card #3.

result of tiredness); the simple visualization of data (even in the case of a large number of transactions); and the possibility to detect isolated data structures.

The methodology described in this article is an early stage of research aimed to produce a framework for unsupervised fraud detection. The objective is to improve and implement in detail the proposed method for accurate and fast fraud detection. Furthermore, it would be interesting to compare the results obtained with the proposed in this article method with results obtained with other methods for fraud detection.

The application of the proposed method for transaction analysis is not restricted to the problem described in this article. It could also be used to create a profile of typical activity of Point-Of-Sale, profile of "good" potential clients, general profile of "good" and "bad" transactions, etc.

Notes:

[1] *Managing Risk in the 21st Century. Strategies for Issuing & Acceptance* (Visa Int., October, 2000), 156.

[2] *Managing Risk in the 21st Century. Strategies for Issuing & Acceptance*.

[3] A.I. Ginsburg, *Plastic Cards* (Saint Petersburg: Peter, 2004), 128 (in Russian).

[4] M.C. Vertusaev, Ya.Yu. Kondrat'ev, S.E. Pugachev, A.M. Yurchenko, "Crime Methods Utilizing Bank Cards," *Information Technologies for Information Protection* 3, no. 1 (1999): 50-67.

[5] Tej Paul Bhatla, Vikram Prabhu, and Amit Dua, "Understanding Credit Card Frauds," *Cards Business Review* 1 (Tata Consultancy Services, June 2003).

[6] Bhatla, Prabhu, and Dua, "Understanding Credit Card Frauds."

[7] K. Chikin and I. Shlyik, "Countering Illegal Transactions in Internet Purchasing Systems," *World of Cards* 7 (2002): 15-21.

[8] *Managing Risk in the 21st Century. Strategies for Issuing & Acceptance*.

[9] Ginsburg, *Plastic Cards.*

[10] Bhatla, Prabhu, and Dua, "Understanding Credit Card Frauds."

[11] Bhatla, Prabhu, and Dua, "Understanding Credit Card Frauds."

[12] Bhatla, Prabhu, and Dua, "Understanding Credit Card Frauds;" Chikin and Shlyik, "Countering Illegal Transactions."

[13] Rüdiger W. Brause, T. Langsdorf, and M. Hepp, "Credit Card Fraud Detection by Adaptive Neural Data Mining," Internal Report 7/99 (J.W. Goethe-University, Computer Science Department, Frankfurt, Germany, 1999), <http://www.cs.uni.frankfurt.de/fbreports/07.99.ps.gz> (12 Dec. 2005); Bhatla, Prabhu, and Dua, "Understanding Credit Card Frauds."

[14] Richard J. Bolton and David J. Hand, "Unsupervised Profiling Methods for Fraud Detection," Technical Report (Department of Mathematics, Imperial College, London, 2002).

[15] Philip D. Wasserman, *Neural Computing: Theory and Practice* (New York: Van Nostrand Reinhold, 1990), (Translation into Russian, Moscow, Mir, 1992): 192; Brause, Langsdorf, and Hepp, "Credit Card Fraud Detection by Adaptive Neural Data Mining."

[16] Chikin and Shlyik, "Countering Illegal Transactions."

[17] Teuvo Kohonen, "The Self-Organizing Map," *Proceedings of the IEEE* 78, no. 9 (September 1990): 1464–1480.

[18] Wasserman, *Neural Computing: Theory and Practice*.

[19] A. Gorbunov, "Application of the Self-Organizing Map in Business and Finance," Bank Technologies 4 (1999): 34-40 (in Russian); Wasserman, *Neural Computing: Theory and Practice*.

[20] Kohonen, "The Self-Organizing Map."

[21] For every card 100 typical transactions and 10 anomalous ones were generated.

VLADIMIR A. ZASLAVSKY is Head of Department Mathematical Methods for Ecological and Economic Research, Faculty of Cybernetics, National Taras Schevchenko University of Kiev. He was Vice Dean of the Faculty of Cybernetics in the period 2000-2004. Dr. Zaslavsky received his PhD in Mathematics from the Faculty of Cybernetics, in 1984. He has been Associate Professor since 1992 and has more than 100 publications in the areas of system analysis of complex systems, risk analysis, reliability optimization and redundancy, and decision support systems. He is a member of IIASA Society and President of the AFCEA – Ukraine Chapter. *E-mail*: zas@unicyb.kiev.ua.

ANNA A. STRIZHAK obtained a M.S. degree in Systems Analysis and Theory of Decision Making from National Taras Shevchenko University of Kiev in 2004. She is currently pursuing her Ph.D. degree in Systems Analysis from the Faculty of Cybernetics, National Taras Shevchenko University of Kiev, Ukraine. Simultaneously, she works as system analyst at "UkrCard" company (International Payment System "UkrCard") and is actively involved in a research project called "Development of software for transactional risk analysis and evaluation in PS UkrCard." Her current research interests include systems analysis, analysis and evaluation of risks in payment systems based on card technology, application of neural technologies for card fraud detection, development of automated fraud detection and prevention systems. *E-mail*: st-anna@ukr.net

Novel Security Protocols

- Choosing t-out-of-n Secrets by Oblivious Transfer
- Security Protocols for Outsourcing Database Services

CHOOSING T-OUT-OF-N SECRETS BY OBLIVIOUS TRANSFER

Jung-San LEE and Chin-Chen CHANG

Abstract: Oblivious Transfer (OT) has been regarded as one of the most significant cryptography tools in recent decades. Since the mechanism of OT is widely used in many applications such as e-commerce, secret information exchange, and games, various OT schemes have been proposed to improve its functionality and efficiency. In 2001, Naor and Pinkas proposed a secure 1-out-of-*n* OT protocol, in which the sender has *n* messages and the chooser can get one of these *n* messages in each protocol run. What is more, the sender cannot find which message has been chosen by the chooser and the chooser knows only the correct message. In 2004, Wakaha and Ryota proposed a secure *t*-out-of-*n* OT protocol, which is an extension of the 1-out-of-*n* OT protocol proposed by Naor and Pinkas. Wakaha and Ryota's *t*-out-of-*n* OT protocol allows the chooser to get *t* messages from the sender simultaneously in each protocol run. Besides, the sender cannot know what the chooser has chosen and the chooser can only know the exact *t* messages. However, getting deep understanding of Wakaha and Ryota's protocol, it could be concluded that it still lacks efficiency such that it is hard to be applied in real-world applications. In this article, a secure and efficient *t*-out-of-n OT protocol based on the Generalized Chinese Remainder Theorem is proposed, in which the chooser can securely get *t* messages from the sender simultaneously in each protocol run. The efficiency of the proposed *t*-out-of-*n* OT protocol is higher than that of Wakaha and Ryota's protocol in terms of practical application.

Keywords: Oblivious Transfer, Generalized Chinese Remainder Theorem, Communications, Secrets Exchange.

Introduction

Recently, numerous Oblivious Transfer (OT) protocols have been applied in many applications such as e-commerce, secret information exchange, games, and others. Therefore, OT has become an important cryptography tool. In 1981, Rabin first proposed the concept of oblivious transfer.[1] People can think of Rabin's OT protocol as a game between two participants, Alice and Bob, where Alice is the sender and Bob is the chooser. Alice sends one bit to Bob, and Bob will get either nothing with prob-

ability 1/2 or the same bit with the same probability. What is more, Alice cannot know which event has happened to Bob. Rabin's idea of OT has attracted a lot of attention; it has become a popular research topic since it was proposed.

An extended concept is one-out-of-two OT protocol, denoted as (OT_1^2), in which Alice sends two bits to Bob, b_1 and b_2. Besides, Bob can choose to get either b_1 or b_2 and can receive one of the two bits with the same probability 1/2. However, Alice cannot know which bit Bob has chosen in this protocol run. Later, a more significant version *1*-out-of-*n* OT protocol, denoted as (OT_1^n), was proposed, in which Alice possesses *n* messages and Bob can get one of them in each protocol run. Similarly to the OT_1^2-protocol, Alice cannot know which message Bob has received, and Bob can get nothing else than the correct message.[2,3,4,5,6]

In general, many OT protocols have been proposed with the objective to improve efficiency or functionality. The most recent research on OT protocols is the *t*-out-of-*n* version, denoted as (OT_t^n), in which Alice possesses *n* messages and Bob can get *t* out of these *n* messages simultaneously in each protocol run. Besides, Alice cannot find out which messages Bob has received, and Bob can know nothing other than the correct *t* messages. However, the majority of these improvements are either based on parallel computing or need heavy computation.[7,8,9,10,11,12,13,14] In 2004, Wakaha and Ryota proposed a secure *t*-out-of-*n* OT protocol, which is an extension of the 1-out-of-*n* OT protocol proposed by Naor and Pinkas. Although Wakaha and Ryota's *t*-out-of-*n* OT protocol allows the chooser to get *t* messages from the sender simultaneously in each protocol run, getting understanding of Wakaha and Ryota's *t*-out-of-*n* OT protocol shows that it still lacks efficiency.[15,16]

In this article, the authors propose a secure and more efficient version of the *t*-out-of-*n* OT protocol based on the Generalized Chinese Remainder Theorem (GCRT).[17] The proposed OT protocol can meet the following requirements, which are considered as the most important ones for the general OT protocols.[18,19,20]

- *Requirement 1*: *Correctness* – If both the sender and the chooser follow the *t*-out-of-*n* OT protocol, the chooser will receive the correct *t* messages after executing the protocol with the sender.
- *Requirement 2*: *Privacy of the chooser* – After the OT protocol is performed with the chooser the sender cannot know which messages are chosen by the chooser.
- *Requirement 3*: *Privacy of the sender* – After the OT protocol is performed with the sender the chooser can get nothing else except these *t* messages.

The rest of this paper is organized as follows. The next section will review the 1-out-of-*n* OT protocol proposed by Naor and Pinkas and the *t*-out-of-*n* OT protocol proposed by Wakaha and Ryota. Some preliminaries are described afterwards, followed by description of the proposed protocol. Some discussions and analyses of the proposed protocol and comparisons between the proposed OT_t^n protocol and other related works are given next. Finally, the last section gives some conclusions.

Review of Related Work

This section introduces the 1-out-of-*n* OT protocol proposed by Naor and Pinkas and the *t*-out-of-*n* OT protocol proposed by Wakaha and Ryota.

Review of Naor and Pinkas's 1-out-of-n OT Protocol

Prior to describing Naor and Pinkas's 1-out-of-*n* OT protocol, the authors define some notations used in their protocol. Let g be a generator of a multiplicative group with a prime order q. Alice is the sender, while Bob is the chooser. $M_1, M_2, \ldots, M_n \in < g >$ are the n messages kept by the sender Alice. G is a large prime. M_c is the choice of the chooser Bob, where c is the serial number of the chosen message and $1 \le c \le n$. The details of the protocol proposed by Naor and Pinkas are given below.

Step 1: Bob constructs a polynomial $f(x)$ as follows

$$f(x) = x - c,$$

and then he chooses a and b randomly from Z_q. Next, Bob generates

$$f'(x) = f(x) + ab = x + (ab - c),$$

and sets

$$e = ab - c.$$

Finally, Bob computes

$$A = g^a \bmod G,$$

$$B = g^b \bmod G, \text{ and}$$

$$E = g^e \bmod G,$$

and sends the messages $\{A, B, E\}$ to Alice.

Step 2: After receiving the messages sent by Bob, Alice computes

$$Y_i = E\, g^i \bmod G, \text{ for } i = 1, 2, \ldots, n.$$

Then, for $i = 1, 2, \ldots, n$, Alice selects s_i and r_i randomly from Z_q and computes

$$H_i = A^{s_i} g^{r_i} \bmod G,$$

$$K_i = Y_i^{s_i} B^{r_i} \bmod G, \text{ and}$$

$$F_i = K_i * M_i \bmod G.$$

Finally, Alice sends all pairs of (H_i , F_i) to Bob, where $i = 1, 2, \ldots, n$.

Step 3: When Bob receives the messages sent by Alice, he computes

$$K_c = H_c^b \bmod G,$$

and then reveals the demanded message as follows

$$M_c' = F_c / K_c \bmod G.$$

Review of Wakaha and Ryota's t-out-of-n OT Protocol

In this subsection, the authors introduce Wakaha and Ryota's OT_t^n protocol which is an extension of Naor and Pinkas's 1-out-of-*n* OT protocol. The authors begin with definition of the notations used in the OT_t^n protocol proposed by Naor and Pinkas. Let g be a generator of a multiplicative group with a prime order q. G is a large prime number. The sender Alice possesses *n* messages M_1, M_2, …, and M_n, where M_1, M_2, …, $M_n \in < g >$. M_{c_1}, M_{c_2}, …, and M_{c_t} are the t choices of the chooser Bob, where M_{c_1}, M_{c_2}, …, and $M_{c_t} \in \{M_1, M_2, \ldots, M_n\}$ and $1 \le c_1, c_2, \ldots, c_t \le n$. M_{c_t} denotes the c_t-th message chosen by Bob. The details of the proposed by Wakaha and Ryota *t*-out-of-*n* OT protocol are presented below.

Step 1: Bob constructs a polynomial $f(x)$, where

$$f(x) = (x - c_1)(x - c_2)\ldots(x - c_t).$$

Then, Bob chooses a, b_0, b_1, …, and b_{t-1} randomly from Z_q, and generates another polynomial $f'(x)$, where

$$f'(x) = f(x) + a(b_0 + b_1 x + \cdots + b_{t-1}x^{t-1}).$$

Let e_0, e_1, …, e_t denote the coefficients of $f'(x)$, that is,

$$e_0 = (-c_1)(-c_2)\ldots(-c_t) + ab_0,$$

$$\vdots$$

$$e_{t-1} = (-c_1 - c_2 - \cdots - c_t) + ab_{t-1}, \text{ and}$$

$$f'(x) = e_0 + e_1 x + \cdots + e_{t-1}x^{t-1} + x^t.$$

Next, Bob computes

$$A = g^a \bmod G,$$

$$B_j = g^{b_j} \bmod G, \text{ where } 0 \le j \le t, \text{ and}$$

$$E_j = g^{e_j} \bmod G, \text{ where } 0 \le j \le t,$$

and then sends the messages $\{A, B_0, B_1, \ldots, B_{t-1}, E_0, E_1, \ldots, E_{t-1}\}$ to Alice.

Step 2: Receiving the messages sent by Bob, Alice computes

$$Y_i = E_0 \, E_1^i \, E_2^{i^2} \ldots E_{t-1}^{i^{t-1}} \, g^{i^t} \bmod G, \text{ where } i = 1, 2, \ldots, n.$$

For $i = 1, 2, \ldots, n$, Alice chooses r_i and s_i randomly from Z_q and computes

$$H_i = A^{s_i} g^{r_i} \bmod G,$$

$$K_i = Y_i^{s_i} (B_0 \, B_1^i \, B_2^{i^2} \ldots B_{t-1}^{i^{t-1}})^{r_i} \bmod G, \text{ and}$$

$$F_i = K_i * M_i \bmod G.$$

Next, Alice sends all pairs of (H_i , F_i) to Bob, where $i = 1, 2, \ldots, n$.

Step 3: Upon receiving the messages sent by Alice, for $i \in \{c_1, c_2, \ldots, c_t\}$ Bob computes K_i' as follows:

$$K_i' = H_i^{b_0 + b_1 i + \cdots + b_{t-1} i^{t-1}} \bmod G.$$

Finally, Bob can retrieve those t messages that he really wants to know as follows:

$$M_i' = F_i / K_i' \bmod G, \text{ for } i \in \{c_1, c_2, \ldots, c_t\}.$$

Preliminaries

This section introduces the exact definition of the proposed *t*-out-of-*n* OT protocol and the Generalized Chinese Remainder Theorem.

Definition of the t-out-of-n OT Protocol

The *t*-out-of-*n* OT protocol is a two-party protocol in which the sender has *n* messages, $\{a_1, a_2, \ldots, a_n\}$, and the chooser can securely get t of these messages simultaneously in each protocol run. Nevertheless, the sender cannot find which t messages are chosen by Bob, and the chooser knows only these t messages. In the following, the authors introduce the three essential properties of the general *t*-out-of-*n* OT protocol.

- *Property 1*: *Correctness* – If both the sender and the chooser follow the *t*-out-of-*n* OT protocol, the chooser will get the correct t messages after executing the protocol with the sender.
- *Property 2*: *The privacy of the chooser* – After the *t*-out-of-*n* OT protocol is executed with the chooser, the sender cannot find out which *t* messages are chosen by the chooser.
- *Property 3*: *The privacy of the sender* – After the *t*-out-of-*n* OT protocol is executed with the sender, the chooser cannot learn anything else but these t messages.

Generalized Chinese Remainder Theorem

This subsection presents the Generalized Chinese Remainder Theorem (GCRT) followed by an example.

Definition of the Generalized Chinese Remainder Theorem [21]

Let d_1, d_2, …, and d_n denote n positive integers and let form the modulus set, where d_i and d_j are relatively prime in pairs for $i, j = 1, 2, \ldots, n$ and $i \neq j$. a_1, a_2, …, a_n are any n positive integers. $D = k * d_1 * d_2 * \ldots * d_n$, where k is a positive integer that satisfies $\text{Max}\{a_1, a_2, \ldots, a_n\} < k < \text{Min}\{d_1, d_2, \ldots, d_n\}$. $D_i = k * d_1 * d_2 * \cdots * d_n / d_i$ for $i = 1, 2, \ldots, n$. $N_i = \lceil a_i * d_i / k \rceil$ for $i = 1, 2, \ldots, n$. Then the following congruences have the same unique solution,

$$\lfloor X / d_1 \rfloor \equiv a_1 (\text{mod}\, k),$$

$$\lfloor X / d_2 \rfloor \equiv a_2 (\text{mod}\, k),$$

$$\vdots$$

$$\lfloor X / d_n \rfloor \equiv a_n (\text{mod}\, k).$$

The reader may ask: how do we compute X from k, d_1, d_2, …, d_n, a_1, a_2, …, and a_n? Since d_1, d_2, …, and d_n are relatively prime in pairs, for $i = 1, 2, \ldots, n$ we have $(d_i, D_i) = 1$. Therefore, there should exist an integer y_i such that $(D_i)\, y_i \equiv k \,(\text{mod}\ k * d_i)$ for $i = 1, 2, \ldots, n$. Besides, $(D_i)\, y_i \equiv 0 \,(\text{mod}\, k * d_j)$, where $j \neq i$. This is due to the fact that (D / d_i) is h times of d_i, where $h \in N$. Let $X = (D_1) y_1 N_1 + (D_2) y_2 N_2 + \cdots + (D_n) y_n N_n \bmod D$. Then X is the unique solution of the above congruence system.

An Example of GCRT

Task: Find a positive integer X for the RNS (2, 3, 4) with the moduli set (6, 7, 11) and the general modulus $k = 5$.

Solution: For the numbers D, D_1, D_2, D_3, we obtain

$$D = 5 * 6 * 7 * 11 = 2310,$$

$$D_1 = (D / d_1) = (2310/6) = 385,$$

$$D_2 = (D / d_2) = (2310/7) = 330, \text{ and}$$

$$D_3 = (D / d_3) = (2310/11) = 210.$$

Therefore solving $385\ y_1 \equiv 5 \pmod{6*5}$, we get $y_1 = 5$,

$330\ y_2 \equiv 5 \pmod{7*5}$, we have $y_2 = 5$, and

$210\ y_3 \equiv 5 \pmod{11*5}$, we have $y_3 = 5$.

Besides, N_1, N_2 and N_3 can be derived as follows

$$N_1 = \lceil 2*6/5 \rceil = 3,$$

$$N_2 = \lceil 3*7/5 \rceil = 5, \text{ and}$$

$$N_3 = \lceil 4*11/5 \rceil = 9.$$

Using the equation $X = (D_1)y_1N_1 + (D_2)y_2N_2 + \cdots + (D_n)y_nN_n \bmod D$, we obtain

$$X = 385*5*3 + 330*5*5 + 210*5*9 = 375 \ \bmod\ 2310.$$

Therefore, $X = 375$ is the solution of the example.

Verification:

$$\lfloor 375/6 \rfloor \bmod 5 = 62 \ \bmod\ 5 = 2,$$

$$\lfloor 375/7 \rfloor \bmod 5 = 53 \ \bmod\ 5 = 3,$$

$$\lfloor 375/11 \rfloor \bmod 5 = 34 \ \bmod\ 5 = 4.$$

The Proposed Protocol

This section presents the proposed *t*-out-of-*n* OT protocol based on the Generalized Chinese Remainder Theorem. The flowchart of the proposed OT protocol is shown in Figure 1. First, the authors summarize the notations used in their *t*-out-of-*n* OT protocol as follows.

- Alice is the sender;
- Bob is the chooser;
- e is the public key of the sender Alice;
- d is the private key of the sender Alice;
- G is a large prime number;

- a_1, a_2, …, a_n are the n messages held by Alice, where $a_i \in N$ and $i = 1, 2, \ldots, n$;
- d_1, d_2, …, d_n are n positive integers that are relatively prime in pairs, where $d_i > a_i$ for $i = 1, 2, \ldots, n$;
- k that satisfies $\text{Max}\{a_1, a_2, \ldots, a_n\} < k < \text{Min}\{d_1, d_2, \ldots, d_n\}$ is a positive integer;
- ID_i is the identity of the message a_i, where $i = 1, 2, \ldots, n$;
- T_1, T_2, …, T_n are the average values for the chooser enabling him/her to retrieve the demanded messages, where $T_i = d_i^e \bmod G$ for $i = 1, 2, \ldots, n$;
- D is the value of $k * d_1 * d_2 * \cdots * d_n$;
- D_i equals D / d_i for $i = 1,2,\ldots,n$;
- N_i equals $\lceil a_i * d_i / k \rceil$ for $i = 1,2,\ldots,n$;
- b_1, b_2, …, b_t are the t messages that Bob wants to know, where $b_j \in \{a_1, a_2, \ldots, a_n\}$ with the corresponding item (ID_j, T_j) for $j = 1, 2, \ldots, t$.

In what follows, the authors provide the details of the proposed *t*-out-of-*n* OT protocol based on GCRT.

Step 1: Receiving the request sent by Bob, for all messages a_1, a_2, …, a_n, Alice selects n positive integers, d_1, d_2, …, d_n, that are relatively prime in pairs for this protocol run, where $d_1 > a_1$, $d_2 > a_2$, …, and $d_n > a_n$. Then Alice generates a positive integer k that satisfies $k > \text{Max}\{a_1, a_2, \ldots, a_n\}$ and $k < \text{Min}\{d_1, d_2, \ldots, d_n\}$ and computes

$$D = k * d_1 * d_2 * \cdots * d_n,$$

$$D_i = D / d_i, \text{ for } i = 1, 2, \ldots, n,$$

$$N_i = \lceil a_i * d_i / k \rceil, \text{ for } i = 1, 2, \ldots, n;$$

then she constructs the following congruence system:

$$\lfloor X / d_1 \rfloor \equiv a_1 (\bmod k),$$

$$\lfloor X / d_2 \rfloor \equiv a_2 (\bmod k),$$

$$\vdots$$

$$\lfloor X / d_n \rfloor \equiv a_n (\bmod k).$$

Next, Alice computes X as follows:

$$X = (D_1)y_1N_1 + (D_2)y_2N_2 + \cdots + (D_n)y_nN_n \bmod D \text{ by GCRT,}$$

where $(D_i)y_i \equiv k(\bmod d_i * k)$, for $i = 1, 2, \ldots, n$.

Afterwards, Alice computes

$$T_1 = d_1^e \bmod G,$$

$$T_2 = d_2^e \bmod G,$$

$$\vdots$$

$$T_n = d_n^e \bmod G,$$

by using the public key e. Next, Alice transmits X, k and all pairs of (ID_i, T_i) to Bob for $i = 1, 2, \ldots, n$.[22,23]

Step 2: Receiving the messages sent by Alice, Bob selects t pairs (ID'_j, T'_j), for $j = 1, 2, \ldots, t$, and generates t corresponding random numbers r_1, r_2, …, r_t, for each pair (ID'_j, T'_j). Next, Bob computes

$$\alpha_1 = r_1^e * T'_1 \bmod G,$$

$$\alpha_2 = r_2^e * T'_2 \bmod G,$$

$$\vdots$$

$$\alpha_t = r_t^e * T'_t \bmod G,$$

by using Alice's public key e. Then, Bob sends the computational result $\{\alpha_1, \alpha_2, \ldots, \alpha_t\}$ to Alice.

Step 3: Upon receiving the messages sent by Bob, Alice computes

$$\rho_1 = \alpha_1^d \bmod G,$$

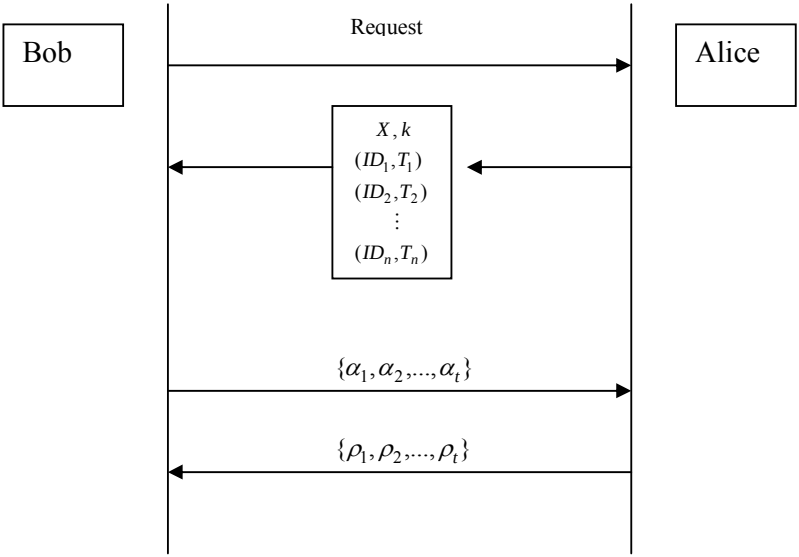


Figure 1: Flowchart of the Proposed *t*-out-of-*n* OT Protocol.

$$\rho_2 = \alpha_2^d \bmod G,$$
$$\vdots$$
$$\rho_t = \alpha_t^d \bmod G,$$

using her private key d, and then she sends the computational results $\{\rho_1, \rho_2, \ldots, \rho_t\}$ to Bob.

Step 4: Receiving the messages sent by Alice, Bob computes

$$d_1' = r_1^{-1} * \rho_1 \bmod G,$$
$$d_2' = r_2^{-1} * \rho_2 \bmod G,$$
$$\vdots$$
$$d_t' = r_t^{-1} * \rho_t \bmod G,$$

Finally, Bob can successfully use X and k to compute the required t messages as follows,

$$b_1 = \lfloor X / d_1' \rfloor \bmod k,$$

$$b_2 = \lfloor X / d_2' \rfloor \bmod k,$$

$$\vdots$$

$$b_t = \lfloor X / d_t' \rfloor \bmod k.$$

Discussion and Analysis

This section demonstrates that the proposed protocol satisfies the essential requirements of the general *t*-out-of-*n* OT protocols mentioned in a previous section and presents some comparisons between the protocol and other related work.

Analysis of the Essential Requirements

This subsection demonstrates that the OT protocol presented in this article meets the three essential requirements of the general OT protocols.

Requirement 1: Correctness

At the beginning, it is assumed that both the sender Alice and the chooser Bob are honest. After they both, Alice and Bob, execute the *t*-out-of-*n* OT protocol, Bob will get *t* messages $\{b_1, b_2, \ldots, b_t\}$. For $j = 1, 2, \ldots, t$, b_j will be equivalent to one of the n messages $\{a_1, a_2, \ldots, a_n\}$ possessed by Alice. It is due to the fact that the T_js are computed as follows:

$$T_1 = d_1^e \bmod G,$$

$$T_2 = d_2^e \bmod G,$$

$$\vdots$$

$$T_n = d_n^e \bmod G,$$

and chosen by Bob from these n messages sent by Alice. Furthermore, X is generated by the following congruence system:

$$\lfloor X / d_1 \rfloor \equiv a_1 (\bmod k),$$

$$\lfloor X / d_2 \rfloor \equiv a_2 (\bmod k),$$

$$\vdots$$

$$\lfloor X / d_n \rfloor \equiv a_n (\mathrm{mod}\, k).$$

That is,

$$X = (D_1) y_1 N_1 + (D_2) y_2 N_2 + \cdots + (D_n) y_n N_n \bmod D \text{ by GCRT},$$

where $(D_i) y_i \equiv k (\mathrm{mod}\, d_i * k)$, $N_i = \lceil a_i * d_i / k \rceil$, for $i = 1, 2, \ldots, n$.

As a result, for each selected item (ID'_j, T'_j), where $j = 1, 2, \ldots, t$, Bob can reveal one corresponding message that he really wants to know by the following derivation,

$$\alpha_j = r_j^e * T'_j \bmod G,$$

$$\rho_j = \alpha_j^d \bmod G,$$

$$d'_j = r_j^{-1} * \rho_j \bmod G, \text{ and}$$

$$b_j = \lfloor X / d'_j \rfloor \bmod k.$$

Consequently, the proposed protocol can satisfy this requirement.

Requirement 2: Privacy of the Chooser

First, it is assumed that these t pairs (ID'_j, T'_j) are the messages that Bob chooses from the n messages sent by Alice, where $j = 1, 2, \ldots, t$. Bob generates a random number r_j for each pair (ID'_j, T'_j), where $j = 1, 2, \ldots, t$. Even if Alice computes $\rho_j = \alpha_j^d \bmod G$ instead of Bob to reveal ρ_j by using her private key d, Alice cannot find d'_j yet. The reason is that d'_j is computed as:

$$d'_j = r_j^{-1} * \rho_j \bmod G.$$

That is, d'_j is protected by r_j^{-1}. However, only Bob knows r_j. Therefore, without knowing r_j Alice cannot reveal which t messages are chosen by Bob. Certainly, Alice may select a set of $\{b'_1, b'_2, \ldots, b'_t\}$ to try guessing which events have happened to Bob. Considering that for each event that Alice guesses the correct choice of Bob is independent, the probability that Alice guesses the correct choices is estimated as follows:

$$\Pr(b_j = b'_j \mid j = 1, 2, \ldots, t) = 1/n^t,$$

where t is the number of the messages that Bob wants to know and n is the total number of the messages kept in Alice's database. Generally speaking, the number of the messages stored in the sender's database is not less than ten thousand. While t is not less than five, the probability that Alice guesses the correct messages chosen by Bob is estimated as follows:

$$\Pr(b_j = b'_j \mid j = 1,2,\ldots,t) \le 1/10^{20}.$$

In other words, the probability that Alice can reveal which t messages are chosen by Bob is quite small. And, therefore, the proposed *t*-out-of-*n* OT protocol can conditionally ensure the privacy of Bob's choices. Thus, this requirement is also met by the *t*-out-of-*n* OT protocol proposed in this article.

Requirement 3: Privacy of the Sender

First, it is assumed that the sender Alice can be trusted. After the proposed protocol is performed by both Alice and Bob, Bob can get nothing else than the chosen *t* messages. It is due to the fact that Alice only computes

$$\rho_j = \alpha_j^d \bmod G,$$

for $j = 1, 2, \ldots, t$, by using her private key d. Without knowing Alice's private key d, Bob cannot decrypt the needed ρ_j to retrieve d'_j by computing

$$d'_j = r_j^{-1} * \rho_j \bmod G.$$

As a result, Bob cannot know b_j for $b_j \notin \{b_1, b_2, \ldots, b_t\}$. Consequently, Bob can know nothing else than these t messages that he really wants to know and the presented protocol can meet this requirement.

Comparison between the Proposed t-out-of-n OT and Other Related Work

Protocol

This subsection presents some comparisons between the protocol presented in this article and other related OT protocols described above. The authors begin with description of the notations used in Table 1. As usual, Alice is the sender, while Bob denotes the chooser. n denotes the number of the messages kept in Alice's database. t denotes the number of the messages that Bob wants to know. *Exp* denotes exponential computation operation.

Table 1: Comparison between the Proposed Protocol and Other Related Work.

Members / Protocols	Alice	Bob
Naor and Pinkas's Protocol	$4(t*n)\ Exp$	$4t\ Exp$
Wakaha and Ryota's Protocol	$4n\ Exp$	$(3t+1)\ Exp$
The Proposed Protocol	$(n+t)\ Exp$	$t\ Exp$

Usually, the computation complexity of an OT protocol depends mainly on the number of the exponential computation operations. Therefore, only the number of the exponentiation computation operations of the proposed OT protocol and the other related protocols is considered in Table 1. Furthermore, repeating a 1-out-of-*n* OT protocol t times still can achieve the functionality of executing a *t*-out-of-*n* OT protocol only once. The computational load of Naor and Pinkas's 1-out-of-*n* OT protocol presented in Table 1 is obtained repeating the 1-out-of-*n* OT protocol t times.

Considering the sender's side, since t is very much less than n, the computational load needed in the proposed protocol is about $4*t$ times lighter than that of Naor and Pinkas's protocol and it is nearly a quarter of the load needed in Wakaha and Ryota's protocol. On the other hand, considering the chooser's side, the computational load required in Naor and Pinkas's and Wakaha and Ryota's protocol are four and three times heavier than that of the protocol presented in this article, respectively. The figures shown in Table 1 clearly demonstrate that the performance of the proposed *t*-out-of-*n* OT protocol is better than that of the related protocols both from sender's and chooser's perspective.

Conclusions

With the rapid development of communication and information technologies, Oblivious Transfer (OT) is widely applied in numerous applications. And, therefore, OT has become an important cryptography tool. The mechanism of the *t*-out-of-*n* OT protocol is a novel and significant version of the OT protocol. In 2004, Wakaha and Ryota proposed a secure *t*-out-of-*n* OT protocol that allows the chooser to get *t* messages from the sender simultaneously in each protocol run. Unfortunately, getting better understanding of Wakaha and Ryota's *t*-out-of-*n* OT protocol, it becomes clear that it still lacks efficiency.

In this article, a secure and more efficient *t*-out-of-*n* OT protocol based on the Generalized Chinese Remainder Theorem (GCRT) is proposed. As analyzed in the article, the proposed OT protocol not only satisfies the three essential properties of the general OT protocols, but also has better performance than that of other related protocols. Therefore, the proposed *t*-out-of-*n* OT protocol is secure and efficient enough to be applied in real-world applications.

Notes:

[1] Michael O. Rabin, "How to Exchange Secrets by Oblivious Transfer," Technical Report TR-81 (Harvard University: Aiken Computation Laboratory, 1981).

[2] Narn-Yih Lee and Tzonelih Hwang, "On the Security of Fair Blind Signature Scheme Using Oblivious Transfer," *Computer Communications* 22, no. 3 (1999): 287-290.

[3] Wen-Guey Tzeng, "Efficient 1-Out-of-n Oblivious Transfer Schemes with Universally Usable Parameters," *IEEE Transactions on Computers* 53, no. 2 (February 2004): 232-240.

[4] Mihir Bellare and Silvio Micali, "Non-Interactive Oblivious Transfer and Applications," in *Proceedings of Advances in Cryptology - CRYPTO'89*, volume 435 of Lecture Notes in Computer Science (Springer-Verlag, 1990), 547-557.

[5] Moni Naor and Benny Pinkas, "Efficient Oblivious Transfer Protocols," in *Proceedings of the 12th Annual Symposium on Discrete Algorithms* (Washington, DC, USA, 7-9 January, 2001), 448-457.

[6] Bill Aiello, Yuval Ishai, and Omer Reingold, "Priced Oblivious Transfer: How to Sell Digital Goods," in *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques: Advances in Cryptology* (Aarhus, Denmark, 21-23 May 2001), volume 2045 of Lecture Notes in Computer Science, 119-135.

[7] Christian Cachin, "On the Foundations of Oblivious Transfer," in *Advances in Cryptology: EUROCRYPT'98* (Espoo, Finland, May/June 1998), volume 1403 of Lecture Notes in Computer Science, ed. Kaisa Nyberg (Springer-Verlag, 1998), 361-374.

[8] Giovanni Di Crescenzo, Tal Malkin, and Rafail Ostrovsky, "Single Database Private Information Retrieval Implies Oblivious Transfer," in *Advances in Cryptology: EUROCRYPT'00: International Conference on the Theory and Application of Cryptographic Techniques* (Bruges, Belgium, 14-18 May 2000), volume 1807 of Lecture Notes in Computer Science, ed. Bart Preneel (Springer 2000), 122-138.

[9] Yan Zong Ding, "Oblivious Transfer in the Bounded Storage Model," in *Proceedings of Advances in Crypto'01* (Santa Barbara, California, USA, August 2001), volume 2139 of Lecture Notes in Computer Science, 155-170.

[10] Juan A. Garay and Philip D. MacKenzie, "Concurrent Oblivious Transfer," in *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science FOCS 2000* (Redondo Beach, California, USA, 12-14 November) (IEEE Computer Society Press, 2000), 314-324.

[11] Yevgeniy Dodis and Silvio Micali, "Lower Bounds for Oblivious Transfer Reductions," in *Advances in Cryptology: Proceedings of Eurocrypt'99*: *International Conference on the Theory and Application of Cryptographic Techniques* (Prague, Czech Republic, 2-6 May 1999), volume 1592 of Lecture Notes in Computer Science, ed. Jacques Stern (Springer Verlag, 1999), 42-55.

[12] Moni Naor and Benny Pinkas, "Distributed Oblivious Transfer," in *Advances in Cryptology: ASIACRYPT 2000, 6th International Conference on the Theory and Application of Cryptology and Information Security* (Kyoto, Japan, 3-7 December 2000), volume 1976 of Lecture Notes in Computer Science, ed. Tatsuaki Okamoto (Springer 2000), 205-219.

[13] Moni Naor and Benny Pinkas, "Oblivious Transfer and Polynomial Evaluation," in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing* (Atlanta, Georgia, USA, 1-4 May 1999) (ACM, 1999), 245-254.

[14] Shimon Even, Oded Goldreich, and Abraham Lempel, "A Randomized Protocol for Signing Contracts," *Communications of the ACM* 28, no. 6 (June 1985): 637-647.

[15] Wakaha Ogata and Ryota Sasahara, "k-out-of-n Oblivious Transfer without Random Oracles," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 87-A, no. 1 (January 2004): 147-151.

[16] Naor and Pinkas, "Efficient Oblivious Transfer Protocols."

[17] Yeu-Pong Lai and Chin-Chen Chang, "Parallel Computational Algorithms for Generalized Chinese Remainder Theorem," *Computers and Electrical Engineering* 29, no. 8 (November 2003): 801-811.

[18] Moni Naor and Benny Pinkas, "Oblivious Transfer with Adaptive Queries," in *Advances in Cryptology – CRYPTO'99*: *19th Annual International Cryptology Conference* (Santa Barbara, California, USA, 15-19 August 1999), volume 1666 of Lecture Notes in Computer Science, ed. Michael J. Wiener (Springer-Verlag, 1999), 573-590.

[19] Naor and Pinkas, "Oblivious Transfer and Polynomial Evaluation."

[20] Even, Goldreich, and Lempel, "A Randomized Protocol for Signing Contracts."

[21] Lai and Chang, "Parallel Computational Algorithms for Generalized Chinese Remainder Theorem."

[22] George I. Davida, David L. Wells, and John B. Kam, "A Database Encryption System with Subkeys," *ACM Transactions on Database Systems* 6, no. 2 (June 1981): 312-328.

[23] Taher ElGamal, "A Public-Key Cryptosystem and a Signature Scheme based on Discrete Logarithms," *IEEE Transactions on Information Theory* IT-31, no. 4 (1985) 469-472.

JUNG-SAN LEE received a BS degree in Computer Science and Information Engineering from the National Chung Cheng University, Chiayi, Taiwan, in 2002. He is currently pursuing his Ph.D. degree in Computer Science and Information Engineering from the National Chung Cheng University, Chiayi, Taiwan. His current research interests include electronic commerce, information security, cryptography, and mobile communications. *Address for correspondence*: Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan, 621, R.O.C.; *Fax*: 886-5-2720859; *E-mail*: ljs@cs.ccu.edu.tw.

CHIN-CHEN CHANG received a BS degree in Applied Mathematics in 1977 and a MS degree in Computer and Decision Sciences in 1979, both from the National Tsing Hua University, Hsinchu, Taiwan. He received his Ph.D. in Computer Engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1982. Since February 2005, he has worked as a chair professor at the Department of Information Engineering and Computer Science at Feng Chia University, Taichung, Taiwan. His current research interests include database design, computer cryptography, image compression and data structures. Dr. Chang is a fellow of the IEEE, a fellow of the IEE, a research fellow of the National Science Council of R.O.C., and a member of the Chinese Language Computer Society, the Chinese Institute of Engineers of the Republic of China, the International Association for Crypto-logic Research, the Computer Society of the Republic of China, and the Phi Tau Phi Honorary Society of the Republic of China. Dr. Chang was the chair and is the honorary chair of the executive committee of the Chinese Cryptography and Information Security Association. *Address for correspondence*: Department of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan, 40724, R.O.C.; *E-mail*: ccc@cs.ccu.edu.tw.

SECURITY PROTOCOLS FOR OUTSOURCING DATABASE SERVICES

Tran Khanh DANG

Abstract: Advances in networking technologies and the continued growth of the Internet have triggered a new trend towards outsourcing data management and information technology needs to external service providers. As a recent manifestation of this trend, there has been growing interest in outsourcing database services in both the commercial world and the research community. Although the outsourced database service model is emerging as an efficient replacement solution for traditional in-house database management systems, its clients, however, have to store their private data at an external service provider, who is typically not fully trusted, and so it introduces numerous security research challenges. To ensure data confidentiality, the outsourced data is usually encrypted and querying is then carried out with the support of trusted client front-ends or secure coprocessors. Despite a large number of research activities done for securing outsourced databases and removing unencrypted data from exposure to the external server and other intruders, no work has been able to radically secure outsourced databases with associated indexes during the query execution. By exploiting such indexes and with relevant available knowledge, attackers can infer confidential information from the outsourced encrypted data. This article discusses potential attacks in such situations and introduces two security protocols for outsourcing database services. The main contributions focus on solutions to the problem of data privacy/ confidentiality and user privacy. The theoretical analyses show that the proposed protocols can effectively protect outsourced data and its associated indexes as well as the clients against various sophisticated attacks.

Keywords: Outsourced Database Services, Data/ User Privacy, Private Information Retrieval/ Storage, Tree-Based Index Structure, Untrusted Server, Encrypted Data.

Introduction

Advances in networking technologies and the continued growth of the Internet have triggered a new trend towards outsourcing data management and information technology needs to external service providers. As a recent manifestation of this trend, there has been growing interest in outsourcing database services in both the commer-

INFORMATION & SECURITY. An International Journal, Vol.18, 2006, 85-108.

cial world and, especially, the research community.[1,2] In the outsourced database service (ODBS) model, clients rely upon external servers and experts for the storage, maintenance, and retrieval of their data. The possibility of outsourcing such database services has generated wide interest in organizations because such a model alleviates their needs to purchase expensive hardware and software, or to pay for professionals to deploy, maintain, and upgrade the system, which are now taken over by the service provider. However, this ODBS model also introduces numerous research challenges and thus has rapidly become one of the most active topics in the research community.[3,4,5,6,7,8,9,10]

As mentioned, in the ODBS model, a client stores its private data at an external service provider who is typically not fully trusted. On the other hand, in this digital age, for most clients, databases take a critical role related directly to their existence and development. Therefore, ensuring clients' data confidentiality is obviously one of the foremost challenges in this model. The question "how is clients' private data protected against sophisticated attackers?" has got much attention from researchers.[11,12] Sophisticated attackers here mean both intruders and insiders, including operators of the external server. Notably, with these malicious insiders, traditional database security techniques [13,14] are useless.

Basically, regardless of the untrusted server at the provider's side, the ultimate goal that clients want is that they can use the outsourced database service as an in-house one. This includes a requirement that clients can operate on their outsourced data without worrying about leak of their sensitive information. This requirement in turn poses several additional challenges related to privacy-preserving for client's queries as well as for the outsourced data during the execution of operations at the untrusted server. Overall, although security requirements are different between real-world applications, the following requirements are most noteworthy:

- *Data confidentiality*: Outsiders and even the server's operators (database administrators) are not able to see the client's outsourced data contents in any case (including when the client's queries are performed on the server).
- *User privacy*: Clients do not want the server to know about their queries and the returned results.
- *Data privacy*: Clients are not allowed to get more information than what they are querying on the server.
- *Authentication and data integrity*: Clients must be ensured that data returned from the untrusted server has originated from the data owner and has not been tampered with.

The above security requirements are different from the traditional database security issues and will in general influence the performance, usability and scalability of the

ODBS model. Among the four, the last security objective (i.e. authentication and data integrity) is out of the scope of this article and we refer interested readers to some recent publications [15,16] for more details. In this article, the author concentrates on addressing the first three security objectives for the outsourced databases that come together with tree-based index structures as discussed below.

To ensure data confidentiality in the ODBS model, outsourced data is usually encrypted before being stored at the external server and querying the data is then carried out with the support of trusted client front-ends [17] or secure coprocessors.[18] This approach can protect the data from outsiders as well as the server, but it introduces difficulties in the querying process. It is hard to protect the user and data privacy as performing queries over encrypted data while still maintaining an acceptable query processing performance. We will elaborate on this issue in the next section with concrete examples.

Although several research activities have been conducted on securing the outsourced database and removing the plaintext (unencrypted data) from exposure to the external server and other intruders,[19,20,21,22,23,24,25] no work has been done to radically secure very large outsourced databases with associated indexes, which are used to accelerate the process of data retrieval. Very large databases augmented by sophisticated and efficient indexes, especially tree-based indexes, are very popular in modern database application domains such as image processing, geographical information systems (GISs), time-series databases, CAD/CAM, and so on.[26] Moreover, not only for such very large databases, the problem of protecting tree-based indexes in traditional RDBMSs and random access files from potential attacks is also important.[27] Basically, the index structures help clients improve the query performance in terms of CPU-, memory-, and IO-cost.[28] By exploiting such (encrypted) indexes and with relevant available knowledge malicious users can infer confidential data/ information from the outsourced *encrypted* data. Some approaches have been recently proposed to deal with this problem.[29] Nevertheless, none of them gives a complete solution to the problem. This article will discuss potential attacks in such situations and introduce two extreme security protocols for outsourcing database services. The proposed novel security protocols employ the state-of-the-art private/ repudiative information retrieval (PIR/RIR) protocols in order to secure both the encrypted data and the associated tree-based indexes to be outsourced against a variety of attacks.

The rest of this article is organized as follows. The next section briefly introduces and discusses related work that has been done or ongoing. Specifically, various approaches to securing the outsourced data will be introduced that resort to both software- and hardware-based solutions, and their weaknesses will be discussed. Next, two new security protocols for outsourced encrypted data with associated tree-based indexes will be introduced. After that, the author discusses and presents possible

changes to these new security protocols in order to balance security and performance. Later, open research issues relevant and indispensable to the real-world application systems are presented. And finally, concluding remarks and future work are given in the last section.

Related Work and Discussions

Consider the following real-life scenario: An organization *M* has a DNA database containing patterns about various diseases. *M* stores these DNA patterns on a database server *DB* and allows a client *A* to access the database to get information with respect to *A*'s DNA sequence. This scenario poses several security issues as follows:

- If *DB* is an untrusted external server, *M* then has to protect its data contents, i.e. the DNA patterns, from being accessed and analyzed by *DB* and other intruders. This security issue is referred to as data confidentiality in the previous section.
- Whenever *A* accesses *DB*, s/he does not want *M* or even *DB*'s operators to know exactly what she is concerned about, both the query and its result. In other words, *A* is concerned about her privacy (the user privacy issue).
- Client *A* is not allowed to get more information other than what s/he is querying on *DB*. This is an important aspect in the real-world scenarios because *A* may have to pay for what she can get from *DB* and *M* does not allow her to get more than what she has paid for or even *A* does not want to get what she does not need from *DB* and *M* (e.g., because *A* is using a low bandwidth connection, limited memory/ storage devices). This security issue is referred to as data privacy (see the introduction).

The need of data confidentiality, data or user privacy depends on particular scenarios in the ODBS model and this must be considered carefully. For example, if *DB* is hired just for *M* to use, i.e. client *A* is *M* itself and *M* is outsourcing its database services only to make use of the advantages of the ODBS model, then, although the data privacy is unnecessary in this case, neglecting the user privacy as mentioned above may potentially lead to expose the outsourced data to danger, even if they have been encrypted. We will detail this problem later.

In general, protecting outsourced data mainly relates to the three security issues as mentioned above [30] and we now briefly introduce and discuss related work done or ongoing in addressing these issues. Figure 1 below sketches the general service provider models that will be discussed:

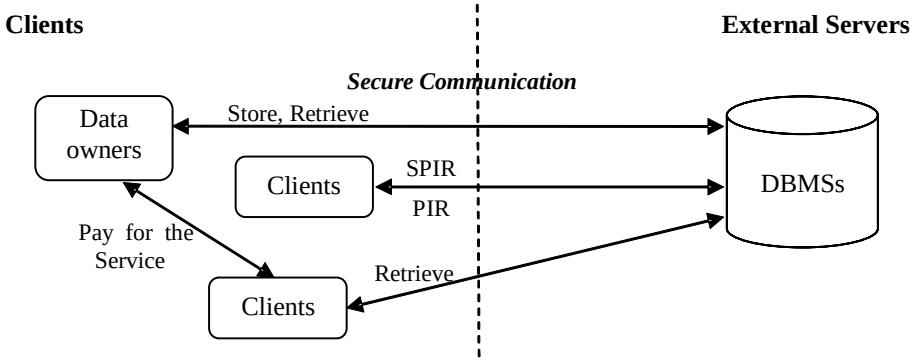


Figure 1: An Overview of Service Provider Models.

As shown in Figure 1, there are four main service provider (SP) models based on the client/ server architecture:[31]

- *UP-DP model*: Data owners are also the SPs. They sell information and charge clients for using their services. The sold information is important and thus the SP is concerned about the data privacy. In this model, the client is concerned about the user privacy. We, therefore, call this outsourcing model the UP-DP model (UP-DP stands for User Privacy – Data Privacy).
- *UP-nDP model*: Similarly to the UP-DP model, data owners here are also the SPs and they charge clients only for using their services, but the stored data is *public*. In this model, the client is also concerned about the user privacy, but the SP is *not* concerned about the data privacy. We, therefore, call this outsourcing model the *UP-nDP* model (nDP stands for not data privacy).
- *DC-UP model*: Data owners are also *unique* clients and their data is outsourced to the external database server. In this model the data owner (also the client) is only concerned about the data confidentiality and the user privacy. We thereafter refer to this outsourcing model as the *DC-UP* model (DC stands for data confidentiality).
- *DC-UP-DP model*: Data owners outsource their data and charge clients for using their data/ information. This is the most complex model in terms of security issues. The data owner is concerned about both the data confidentiality and data privacy with respect to both the external database server and its clients. The client, in turn, is concerned about the user privacy with respect to both the data owner and the server. Moreover, the data owner also takes the client role when accessing its outsourced data on the server and, in this case, the data owner is concerned about the user privacy as well. We, therefore, call this outsourcing model the *DC-UP-DP* model.

We can easily realize that each SP model requires different security objectives and thus different security techniques/ protocols have been invented to satisfy these objectives. In Table 1 we summarize security techniques/ protocols related to the SP models being discussed.

In the UP-DP model, it is not necessary the stored data to be encrypted because the data owner (also the SP) can employ traditional database security techniques to protect their data more efficiently. To satisfy the user privacy requirement, private information retrieval (PIR)-like protocols are employed.

The PIR protocol was first introduced by Chor and colleagues [32] and it has been investigated as well as improved by many researchers thereafter.[33,34] In principle, the PIR protocol allows a client to access a database without revealing to the server both the query and the returned result. More specifically, using the PIR protocol, clients have the possibility of retrieving the *i*-th record of an N-record database without revealing the value *i* to the server. In other words, the server does not know what data the client is querying or getting, hence the user privacy is satisfied. In addition, it is easy to observe that security objectives in the UP-nDP model can be solved simply by using any PIR-like protocol.

Notably, Ostrovsky and Shoup [35] have developed the PIR protocol so that it can also support the writing operations privately. Their new protocol is named the private information storage (PIS). Recently, Asonov and Freytag [36] proposed a repudiative information retrieval (RIR) protocol, which is a modified version of the PIR one, to preserve the user privacy but with a better IO-cost for preprocessing before answering a query. The new IO-cost complexity is reduced from $O(N \log N)$ to $O(\sqrt{N})$, where N is the number of records in the database. The main idea of the RIR protocol is the relaxation of the privacy requirement in which some information on the record identity is allowed to be revealed. However, the information revealed should not be enough to indicate definitely if it was record 1, or 2, …, or N.

Nevertheless, PIR/ RIR protocols cannot satisfy the data privacy objective, which should also be dealt with in the UP-DP model. Aiming to address this issue, some research work has been carried out. Specifically, Gertner and colleagues have developed a protocol called symmetrically private information retrieval (SPIR) protocol that can be built on the basis of any PIR protocol with the aim to satisfy both user and data privacy requirements.[37] In addition, it should be pointed out here that all approaches developed for the UP-DP model have not been designed to secure tree-structured data against potential attacks. As stated by Du and Atallah, this is not a trivial task and needs much more research.[38] Specifically, whenever applying approaches developed for the UP-DP model where the data are indexed using some tree-based indexing technique, the data privacy will not be satisfied because the com-

parison at a node of the outsourced search tree will give information about the data which is associated with that node. We will consider an example with a B+-tree later.

Table 1: Security Techniques and Protocols Related to the Outsourced Database Service Model.

Security Techniques and Protocols	***Security Objectives***			***Indexing Support***	***References (Can be used for)***
	Data Confidentiality	Data Privacy	User Privacy		
PIR/RIR, PIS			x		see Chor, Goldreich, Kushilevitz, and Sudan;[39] Asonov,[40] Chor, Gilboa, and Naor;[41] Ostrovsky and Shoup;[42] Asonov and Freytag [43] (UP-nDP model)
SPIR		x	x		see Gertner, Ishai, Kushilevitz, and Malkin;[44] and Du and Atallah;[45] data owners also host the server (UP-DP model)
Untrusted 3rd parties, Secure coprocessors	x	x	x		see Smith;[46] Du and Atallah;[47] and Smith and Safford [48] (DC-UP-DP model)
Index of range, Hash-based methods	x			x	see Damiani, Vimercati, Jajodia, Paraboschi, and Samarati;[49] Hacigümüs, Iyer, Li, and Mehrotra;[50] data owners are also clients (not pay-as-you-use service)
User anonymity			x		see the papers by Reiter and Rubin [51,52] (identity hiding)
Extreme protocol, Secure coprocessors, Access redundancy and node swapping	x		x	x	This article and Dang;[53] Smith and Safford;[54] Lin and Candan;[55] and Smith [56] (DC-UP model)
Extreme protocol	x	x	x	x	This article (DC-UP-DP model)

Besides, user privacy in some context also requires user anonymity,[57] which means that not only the user's query and its result are of a concern, but also the user's identity itself needs to be hidden (see Table 1). However, user anonymity solutions still have a lot of limitations in both technical and social aspects.[58] More importantly, even when such user anonymity solutions are employed, the outsourced data is still in danger due to sophisticated attacks as will be discussed below. Furthermore, the UP-DP and UP-nDP models are not of main consideration with respect to the ODBS model. In this article, the DC-UP and DC-UP-DP models are in fact of greater interest.

There are some recent approaches related to the data confidentiality requirement for the ODBS model.[59] Among them, actually, solutions resorting to special hardware equipment have also been investigated and developed.[60] Although these hardware-based solutions may satisfy security objectives in several applications (see Table 1), there are still a matter of controversy.[61,62,63] For the security protocols that will be introduced in this article, it is assumed that such a special hardware is not needed and we rely solely on the available software/ hardware infrastructure. Several recent noteworthy approaches not employing any special hardware were also introduced.[64]

Du and Atallah have introduced protocols for secure remote database access with approximate matching.[65] The problem of answering similarity and approximate queries has been extensively studied by many researchers,[66] but not for outsourced data. The original problem is to search a data repository for some data items that are close to a user's query. The closeness is measured using some metric (e.g., Euclidean metric). Du and Atallah have also proposed solutions to four different e-commerce models, which are quite similar to the presented above four SP models. Their solutions can be used for securing the outsourced data with respect to data confidentiality, data and user privacy where appropriate according to the involved model. Contrary to other related approaches, Aggarwal and coworkers have proposed an approach to outsourcing database services without having to encrypt *all* data fields.[67] This approach needs two non-colluded servers to store the outsourced data and can be used for the DC-UP model. However, all of the above described solutions fail to protect the outsourced data as well as the user privacy in case tree-based index structures are used to access the data more efficiently. Such indexes are an indispensable component to large and high-dimensional databases, which are appearing in many modern database applications as mentioned in a previous section.

Nowadays, there are two approaches aiming to protect the data confidentiality for outsourced indexed data.[68] Both approaches protect the outsourced data from intruders and the server's operators through some encryption method. To process queries over encrypted data, two different solutions have been introduced. Hacigümüs and colleagues have proposed storing, together with the encrypted data, additional indexing information.[69] This information can be used by the untrusted server to select

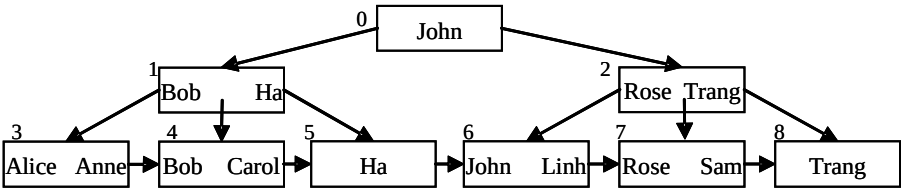


Figure 2: An Example of B+-tree on Attribute *CustomerName*.

the data in response to a user's query. The main idea to process a query in this scheme is to split the original query into: (1) a corresponding query over encrypted relations to run on the untrusted server; and (2) a client query for post-processing the results returned from the server query. The major challenge in this scenario is how to compute and represent index information. Particularly, the relationship between indexes and data should not open the door to inference and linking attacks that can compromise the protection granted by encryption. However, as stated by Damiani and colleagues,[70] although the *index of range* technique proposed by Hacigümüs and team, which relies on partitioning of the domains of client tables' attributes into sets of intervals, is suitable for both exact match and range queries, it introduces difficulties in managing the correspondence between intervals and the actual values present in the database as well as some limitations in such a protection. Similarly to the work of Hacigümüs and colleagues, Damiani and team have also introduced a method to query a tuple-level encrypted database but with a better security level for the outsourced data. For exact match queries, they have analyzed some potential inference and linking attacks and proposed a hash-based indexing method. In order to execute interval-based (range) queries in the ODBS model, they have proposed a solution employing B+-trees typically used in DBMSs.[71] Unfortunately, both of the above approaches do not meet the requirements for user and data privacy (see Table 1). This fact has also been confirmed by Damiani and team. And it can be exploited to carry out inference and linking attacks as will be shown below.

Figure 2 illustrates an example of the B+-tree on an attribute *CustomerName* with sample values. Assume a client/ user is querying all customers whose name is *Ha* on this B+-tree. Following the approach proposed by Damiani and colleagues, the trusted front-end will produce a sequence of queries that will access in sequence nodes 0, 1, and 5. In this case, during the querying process, the user will get more information showing that there are at least two other customers named John and Bob in the database [72] so the data privacy requirement cannot be satisfied. In addition, the server also realizes that the user was accessing nodes 0, 1, and 5, and node 0 is the root, node 1 is an internal node, and node 5 is a leaf node of the tree. Using such information collected gradually, together with statistical methods and data mining tech-

niques, the server can rebuild the whole tree and infer sensitive information from the encrypted database.[73] In this example, this could happen because the user privacy was not protected during the query process.

To protect the user privacy in such cases, there is a recent approach proposed by Lin and Candan.[74] The authors have introduced new techniques to access outsourced tree nodes, called access redundancy and node swapping. The access redundancy technique can be viewed as a computational security version of computational PIR-like protocols, in which information-theoretical security objectives are traded off against performance. Their approach, however, can only be used for the DC-UP model (see Table 1) and has critical limitations, which have been overcome in a recent work conducted by the author.[75] This approach will be discussed in more details in the section on balancing security and efficiency.

As we can see from the analyses above, all introduced approaches that do not employ special security hardware equipment have not dealt radically with potential sophisticated attacks made by exploiting outsourced tree-based index structures. In the next section, general, simple and effective security protocols for securing the outsourced encrypted data with such associated tree-based indexes will be introduced.

Two Extreme Security Protocols

In this section, we consider two ODBS models as mentioned above: the DC-UP model and the DC-UP-DP model. For these two models, we assume that data of an organization *M* is outsourced to some untrusted external database server *DB*. Moreover, to manage the storage and retrieval of data efficiently, assume that the outsourced data is indexed using tree-based index structures that are the most popular technique and play a fundamental and important role in both traditional and modern database application domains.

B+Table

NID	*Node*
0	(1,John,2,-,-1)
1	(3,Bob,4,Ha,5)
2	(6,Rose,7,Trang,8)
3	(Alice,Anne,4)
4	(Bob,Carol,5)
5	(Ha,-,6)
6	(John,Linh,7)
7	(Rose,Sam,8)
8	(Trang,-,-1)

B+EncryptedTable

NID	*EncryptedNode*
0	D0a1n2g3Kh75nhs&
1	T9&8ra§ÖÄajh³q91
2	H&$uye”µnÜis57ß@
3	L?{inh*ß²³&§gnaD
4	Wh09a/[%?Ö*#Aj2k
5	j8Hß}[aHo$§angµG
6	#Xyi29?ß~R@€>Kh
7	~B³!jKDÖbd0K3}%§
8	T-§µran&gU19=75m

Figure 3: The Corresponding Plaintext and Encrypted Table Used to Store the B+-Tree at the External Server.

Similarly to other previous approaches, in order to protect the outsourced data from possible intruders we encrypt the data prior to outsourcing. In line with the work of Damiani and colleagues,[76] we choose to encrypt each tree node as a whole since protecting a tree-based index by encrypting each of its fields would disclose to *DB* the ordering relationship between the index values. Moreover, the unit of storage and access in the described approach is also a tree node. Each node is identified by a unique node identifier (NID). The original tree is then stored in *DB* as a table with two attributes: NID and an encrypted value representing the node content. Let us have a look at an example: Figure 3 shows the corresponding plaintext and encrypted table used to store the B+-tree in Figure 2 at the external server. As we can see, the B+-tree is stored at the external server as a table over the schema *B+EncryptedTable = {NID, EncryptedNode}*. A client then retrieves a node from the server by sending a request including the NID of the node.

To ensure the private information storage (PIS) in the future (refer to the previous section), the NID can be assigned arbitrarily by the trusted front-end as a node is inserted. Obviously, to make this feasible, a small amount of meta-data should be kept at the client side. Based on the above settings, in the next two sub-sections general protocols will be succinctly presented in order to meet the security objectives for the two considered ODBS models.

The DC-UP Model

In this ODBS model, the data owner is also the unique client so the data privacy objective as mentioned before is not important and could be ignored. As we can observe from the example presented in the previous section (illustrated in Figure 2), even if the data has been encrypted, potential attacks are still possible due to the lack of user privacy-preserving during the querying process. Therefore, in this model, we can simply employ any PIR-like protocol [77] for the client's queries (in this case *M* is also the client) in order to satisfy the user privacy objective. The following formula can be given to ensure the data confidentiality and the user privacy for this model:

$$DC + UP = Encryption + PIR\ protocol \quad (1)$$

Now let us again consider the example from the previous section and the same situation: *M* is querying all customers whose name is *Ha* using the B+-tree as shown in Figure 2 (note that in this model *M* is not concerned about data privacy). Due to the fact that the PIR protocol is employed, the server *DB* does not know which nodes *M* is accessing. The tree information and structure are, therefore, kept secret and no inference and linking attacks are possible.

However, if the data is sometimes changed and *M* needs to update its data on *DB* to reflect the changes, i.e. *M*'s outsourced database is dynamic,[78] we then also need to extend the user privacy requirement so that the server *DB* will not be able to see what have been changed and updated. This is critically important because, with all tree-based index structures, such update operations may lead to node splits.[79] When the split nodes are updated in *DB* and if the server knows this information, which can be collected gradually, it is not difficult to reconstruct the whole tree structure. Then, in this case, the problem of potential inference and linking attacks comes back. To avoid such situations, we need PIS-like protocols in order to protect *M*'s privacy in both reading and writing operations from and to *DB*, respectively. Therefore, we obtain the following formula for this ODBS model:

$$DC + UP = Encryption + PIS\ protocol\ * \quad (2)$$

(* for private reading and writing operations)

The correctness and effectiveness of the proposed protocol could be proved by the theoretical analysis performed in the previous section. Specifically, as demonstrated by Damiani and team,[80] even if the attacker is aware of the distribution of plaintext values in the original database, the outsourced data that has been *encrypted* and *indexed* will still be secure against inference and linking attacks if the index information has been kept secret. In a later section we will further elaborate on the efficiency of this protocol and propose possible changes/ improvements.

The DC-UP-DP Model

In this model, assume that *M* is selling its data stored in *DB* and a client *A* is paying for this service. For each query *Q* sent from *A*, both *M* and *DB* should not get any information about *Q* (user privacy) and, in turn, *A* should not get more data/ information from *DB* other than the results of *Q* (data privacy). Note that, in this case, *DB* can even become a client of *M* and it could compromise the privacy of the database by conducting a number of queries and discovering the way the database is encrypted or disguised. A security protocol should defend against this type of active attack. As far as the author is aware, there has been no solution to this model for outsourced tree-structured data. Relying on the solid protocol that has been just proposed for the DC-UP model above, the article proposes a protocol to meet the security requirements of the DC-UP-DP model resorting to a *trusted* third-party, namely *K*. The use of a trusted third-party aims to turn this ODBS model, which is very hard to deal with directly, into the well-behaved DC-UP model.

The assumption for this protocol is that *K* will not collude with *M*, *A*, or *DB* in any way.[81] Furthermore, *K* may send queries to *DB* on behalf of *M* when allowed and up

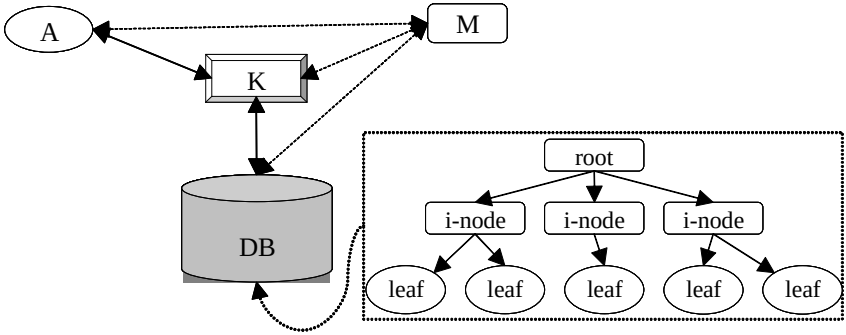


Figure 4: A Security Protocol for the DC-UP-DP Model.

to the level *A* is registered to use *M*'s service s/he can send queries to *K*. It means that *A* can access *M*'s outsourced data indirectly by sending her/his requests to *K*. This protocol is illustrated in Figure 4.

As shown in Figure 4, the outsourced encrypted tree nodes, including the root, the internal nodes (i-nodes), and the leaves, are stored in *DB*. The client *A* pays for the service to the data owner *M* and could query *DB* indirectly via *K*. Thanks to *K*, the third party that both *A* and *M* trust implicitly, the user and data privacy requirements are satisfied. The general steps necessary to perform a query *Q* from *A* are as follows:

1. Client *A* sends a query *Q* to *K*.
2. Upon receiving *Q*, *K* informs *M* (for billing, for example) and waits for approval from *M* in order to access *DB*.
3. Once the information from *K* is received, *M* informs *DB* so that *K* can query *M*'s outsourced database on behalf of *M*. After receiving *DB*'s acknowledgement, *M* informs *K*.
4. From this time on, *K* assumes the role of *M* in the DC-UP model as discussed above, and it accesses *DB* using the security protocol presented in Equation 1 (note that *A* is only able to retrieve information from *DB*, not to update *M*'s outsourced data in *DB*).
5. Finally, *K* filters and returns to *A* only the results of the query *Q*. Obviously, *K* has been informed by *M* what *A* is able to get from the database, but *M* will not be informed what information *A* has got regarding any queries.

In the presented protocol, the trusted third party *K* acts similarly to the secure coprocessors described by some authors.[82,83] This is also the only weakness of the protocol. With the assumption that we could establish such a trusted third party, it is easy to prove that the above protocol ensures all the security objectives for the ODBS model, i.e. data confidentiality, and user and data privacy. Eliminating *K* from this protocol,

while still ensuring all security objectives for the ODBS model, is an open research question and also one of the biggest challenges for future research.

Another point worth mentioning here is that it has been implicitly assumed that the *real* data of the outsourced database is all kept in the tree's leaf nodes. This is, however, not always true with multidimensional access methods (MAMs). With complex data objects to be indexed, the leaf nodes contain only identifiers (IDs) of the data objects, and the data objects are usually kept at a separate place.[84] For such types of index structures, i.e. structures without real data kept in their leaf nodes, we can still apply the same scheme for encryption and storage at the server as follows: (1) the tree nodes are encrypted as a whole and stored in the server as described above (refer to Figure 3), and (2) each real data page that contains the real data objects is encrypted as well, and they will also be stored on the server with additional IDs similarly to the tree nodes. Therefore, both the server and intruders are not able to differentiate between tree nodes and data pages. This uniform storage creates more difficulties in compromising the database confidentiality.

In real-world applications, the extreme security protocols proposed for the two ODBS models above can be modified so as to reduce their communication and computation costs, whereas an acceptable security level is still maintained. In the section below, the author will discuss issues related to the efficiency of the proposed protocols and present some possible modifications.

Balancing Security and Efficiency

In this section, the efficiency of the proposed security protocols will be discussed. The efficiency in the context here can be interpreted in terms of CPU-, IO- and memory-cost. For many other approaches supporting user privacy,[85] the costs are linear in database size. Obviously, this is an undesirable situation due to the difficulty in deploying such cost-inefficient protocols in real-world applications.

As observed in the section on the DC-UP model, the main factor influencing database access efficiency in the DC-UP model is the efficiency of the employed PIR protocol (note that the PIS protocol that ensures both the private information retrieval and storage is also built on a certain PIR protocol – consider equation (2)). Specifically, as pointed out by Chor and team,[86] the information-theoretic PIR protocol will become prohibitively expensive when only one server is employed to host the outsourced data. This indicates that the proposed in this article protocols (for the two ODBS models) will also become prohibitively expensive if there is no replication of the outsourced data and an information-theoretic PIR protocol is employed. The main question is "*How will the client's queries be performed effectively, efficiently and obliviously over encrypted data without revealing any information about both data*

and queries to unauthorized people?" It has motivated the author to look for possible modifications of the protocols in order to make them more practical. Below, such possible modifications for the two protocols introduced for the DC-UP and DC-UP-DP models will be presented.

Modifications for the DC-UP Model

As has been already introduced and discussed, the RIR protocol is a modified and improved version of the PIR protocol in terms of cost reduction. In fact, the RIR protocol is a computational PIR protocol, in which the extreme security requirements are relaxed in order to gain a better query performance. Therefore, we can employ the RIR protocol instead of the information-theoretic PIR protocol to reduce the costs for database access. Formula 3 below reflects this change in the presented protocol for the DC-UP model:

$$DC + UP = Encryption + PIR\ protocol \tag{3}$$

Besides, similarly to the PIS protocol, we can also build repudiative information storage (RIS) protocol based on the RIR protocol to support both reading and writing operations privately. The RIS protocol is better than the PIS protocol in terms of IO-cost. Therefore, one can employ the following modified formula for the DC-UP model with *dynamic* outsourced databases and associated tree-based index structures:

$$DC + UP = Encryption + RIS\ protocol\ * \tag{4}$$

(* for repudiative reading and writing operations)

Also, in order to support the oblivious search on a single outsourced search tree (i.e., the replication of the outsourced database as in some PIR/ RIR-like protocols is unnecessary), Lin and Candan present a protocol based on two new techniques: access redundancy and node swapping.[87] With these two techniques and some additional settings, their protocol can be used for the DC-UP model. The two techniques are briefly summarized in what follows.

Access Redundancy

This technique requires that whenever a client accesses a node, called target node, she asks for a set of m-1 randomly selected nodes in addition to the target node from the server. By this access redundancy, the probability that the server can guess the target node is 1/m. Here, m is an adjustable security parameter.

As mentioned, the access redundancy technique can be viewed as a *computational* PIR-like protocol with a better performance, but a worse security level compared with the information-theoretic PIR-like protocols. This technique is also different from the

one presented by Damiani and colleagues,[88] where only the target node is retrieved (this may reveal the tree structure as shown above).

Apart from redundancy in node access, this technique bears also another weakness: it may lead to leak of information about the target node. This could be easily observed: multiple access requests for the root node will reveal its position by simply calculating the intersection of the redundancy sets of the requests. If the root node position is disclosed, there is a high risk that its child nodes (and also the whole tree structure) may also be revealed. This shortcoming could be overcome by secretly changing the target node's address each time it is accessed.

Node Swapping

Each time a client requests to access a node from the server, it asks the server for a redundancy set of *m* nodes consisting of at least one *empty* node together with the target one. The client then (1) decrypts the target node; (2) manipulates its data; (3) swaps it with the empty node; and (4) re-encrypts the nodes in the redundancy set and writes them back to the server. As proven by the authors of this technique, with it, the possible position of the target node is randomly distributed over the data storage space on the untrusted server, and thus the weakness of the access redundancy technique is overcome. Note that, in order to prevent the server from differentiating between read and write operations, a read operation is always followed by a write operation for all nodes in the redundancy set back to the server.

Although these techniques are applicable to searching outsourced search trees with sound experimental results reported, it has several limitations and weaknesses. As has been elaborated by the author in a recent publication,[89] this solution can not be applied to dynamic outsourced search trees where data items may be inserted into, removed from, or modified. More importantly, it has also been pointed out that applying this solution directly to such dynamic trees may lead to leak of information about the queries and the tree structure, and so the security objectives are compromised. The author has presented solutions to overcome these limitations and weaknesses as well as to deal with privacy-preserving basic operations (including both search and updates) on outsourced search trees.

Modifications Related to the DC-UP-DP Model

First, it is easy to realize that all possible modifications for the DC-UP model can also be *suitably* applied to the DC-UP-DP model (refer to step 4 in the protocol proposed for the DC-UP-DP model). However, it should be noted that the client *A* is only allowed to retrieve the outsourced data but not to update the data. Only the data owner *M* is able to update its outsourced data on the server *DB*. Therefore, all possible modifications presented in the previous subsection can be used for the relevant

operations between *M* and *DB*, while only the protocol as shown in Formula 3 and the access redundancy and node swapping techniques are needed for the operations between the trusted third party *K* and *DB*.

Furthermore, with the DC-UP-DP model, in order to reduce communications complexity, we could store meta-data of the outsourced tree, namely its root and internal nodes, at the trusted third party *K* instead of storing them all on the server *DB*. In this case, there is just a slight change for *DB* compared to the DC-UP model's settings: *DB* now stores only leaf nodes of the tree (and may be real data pages – refer to the section devoted to the DC-UP-DP model). It is not necessary to encrypt the meta-data stored at *K*. Afterwards, in step 4 of the security protocol, *K* first processes *Q* using the meta-data of the database, i.e. the root and leaf nodes of the tree. *K* will access *DB* if it finds it necessary to do so. Similarly, from this time on, *K* takes the role of *M* in the DC-UP model, i.e. *DB* will not know *K*'s queries as well as their results. Note that, in this case *K* will not let *M* know whether it needs to access *DB*. This prevents *M* from inferring that the information that the client *A* needs is currently very likely in its database. Some variants of this approach can also be employed, for example: storing just a part of the tree's meta-data on *K*, but not the root and *all* the internal nodes.

Nevertheless, there is a flaw in the modified scheme just presented: consider the case that *K* has checked its meta-data relevant to the query *Q* and has found that it is unnecessary to access *DB*, and if *K* does so, the data owner *M* is able to discover it. Basically, *M* can carry out this "attack" in various ways. A simple scenario is as follows: *M* colludes with *DB* to see whether *K* will access its outsourced database after *M* has informed *DB* as shown in step 3 of the protocol. As described above, *K* will not access the database if unnecessary and, in that case, both *M* and *DB* will know that what *A* is trying to get does not exist in *M*'s outsourced database (i.e., *Q*'s result set is empty). Therefore, the user privacy is partially not preserved. To resist this kind of attack, *K* will still have to perform some dummy accesses to *DB* even if it has found out that this is no longer necessary to answer *Q*.

Open Research Questions

Obviously, the first very important question is "Do the proposed security protocols open the door for criminals to carry out fraudulent actions more confidentially?" Computer criminal-related problems have been increasingly growing and now, if we provide clients with the means of hiding their identifiers (e.g., the CROWDS model as introduced in previous sections), their queries, or what they have taken away (the user privacy), how can we protect other clients and organizations (including the service providers) from malicious actions?

The second question, which is somewhat related to the first one, is "How can DBMSs conduct auditing activities in systems provided with such extreme security protocols (without employing special hardware equipment)?" The DBMS may not know who is accessing the system, what they are asking for, and what the system returns to the client, how can it tackle the accountability or develop intrusion detection systems? The goals of privacy-preserving and accountability appear to be in contradiction and an efficient solution to balance the two is still open. More discussions about this topic can be found in a recent publication.[90]

Besides, as already mentioned, avoiding the use of a third party in the DC-UP-DP model is an interesting and challenging problem as well. All of these questions/ problems (and many others) are still open and require future research.

Conclusions and Future Work

In the ODBS model, the private data is stored at an external service provider, who is typically not fully trusted. Therefore, dealing with security issues in the ODBS model has rapidly become one of the most active topics in the research community. In general, to protect the outsourced data from malicious users, three major issues need to be dealt with radically: data confidentiality (DC), user privacy (UP), and data privacy (DP). For each particular ODBS model, we need to address different security objectives. In this article, the main contributions lie in the following: (1) We have summarized, discussed, and classified different service provider models as well as security techniques and protocols related to them; (2) Two security protocols for the two most popular ODBS models have been introduced, namely the DC-UP and DC-UP-DP models, as well as possible modifications/ improvements have been proposed so that they can scale well to different real-world application domains; and (3) This work has presented important open research directions, which are relevant and necessary for real-world applications.

The proposed protocols for the two ODBS models support outsourced encrypted tree-structured data. This is an important aspect because tree-based index structures have taken a fundamental and crucial role in both traditional and modern database application domains. Especially, the two proposed security protocols have proven to be extreme security protocols for the corresponding ODBS models. They can protect user's data with associated tree-based index structures against various sophisticated attacks from intruders as well as insiders, including the server's operators. To the best of our knowledge, these are among the advanced solutions to the problem of radically securing outsourced data with associated indexes.

Last but not least, considering the fact that the proposed protocols are rather theoretical, the future work will be focused on the open research directions as mentioned, to-

gether with implementing and evaluating the efficiency and effectiveness of these protocols on different real-world application domains. In particular, comparing the efficiency of the protocol for the DC-UP model using some efficient *computational* PIR protocol with the one introduced by the author in another publication will be of particular interest.[91] This will enable the evaluation of the practical value of PIR-like protocols.

Notes:

[1] Einar Mykletun, Maithili Narasimha, and Gene Tsudik, "Authentication and Integrity in Outsourced Databases" (paper presented at the 11th Annual Network and Distributed System Security Symposium –NDSS04, California, USA, February 2004).

[2] Tran Khanh Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees" (paper presented at the International Workshop on Privacy Data Management – PDM05, in conjunction with ICDE05, IEEE Computer Society, Tokyo, Japan, April 2005).

[3] Wenliang Du and Mikhail J. Atallah, "Protocols for Secure Remote Database Access with Approximate Matching" (paper presented at the 7th ACM Conference on Computer and Communications Security, the 1st Workshop on Security and Privacy in E-Commerce, Athens, November 2000).

[4] Sean W. Smith and Dave Safford, "Practical Server Privacy with Secure Coprocessors," *IBM Systems Journal* 40, no. 3 (2001): 683-695.

[5] Hakan Hacigümüs, Bala R. Iyer, and Sharad Mehrotra, "Providing Database as a Service" (paper presented at the 18th International Conference on Data Engineering, San Jose, February-March 2002), 29-40.

[6] Luc Bouganim and Philippe Pucheral, "Chip-Secured Data Access: Confidential Data on Untrusted Servers" (paper presented at the 28th International Conference on Very Large Data Bases, Hong Kong, August 2002), 131-142.

[7] Mykletun, Narasimha, Tsudik, "Authentication and Integrity in Outsourced Databases."

[8] Ping Lin and K. Selçuk Candan, "Hiding Traversal of Tree Structured Data from Untrusted Data Stores" (paper presented at the 2nd International Workshop on Security in Information Systems-WOSIS04, Porto, Portugal, April 2004), 314-323.

[9] Richard Brinkman, Jeroen Doumen, and Willem Jonker, "Using Secret Sharing for Searching in Encrypted Data" (paper presented at the Workshop on Secure Data Management in a Connected World, Toronto, Canada, August 2004), 18-27.

[10] Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees".

[11] Ernesto Damiani, Sabrina De Capitani di Vimercati, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs" (paper presented at the 10th ACM Conference on Computer and Communication Security, USA, 27-30 October 2003), 93-102.

[12] Hakan Hacigümüs, Bala R. Iyer, Chen Li, and Sharad Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model" (paper presented at the ACM SIGMOD International Conference on Management of Data, USA, June 2002), 216-227.

[13] Silvana Castano, Mariagrazia G. Fugini, Giancarlo Martella, and Pierangela Samarati, *Database Security* (Addison-Wesley and ACM Press, 1994).

[14] Amjad Umar, *Information Security and Auditing in the Digital Age. A Practical and Managerial Perspective* (NGE Solutions, December 2003).

[15] Mykletun, Narasimha, Tsudik, "Authentication and Integrity in Outsourced Databases."

[16] Wenbo Mao, *Modern Cryptography: Theory and Practice* (Prentice Hall PTR, 1st Edition, July 2003).

[17] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[18] Smith and Safford, "Practical Server Privacy with Secure Coprocessors."

[19] Yan-Cheng Chang and Michael Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data" (Cryptology ePrint Archive: Report 2004/051), <http://eprint.iacr.org/2004/051> (20 Dec. 2005).

[20] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency."

[21] Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data."

[22] Bouganim and Pucheral, "Chip-Secured Data Access: Confidential Data on Untrusted Servers."

[23] Smith and Safford, "Practical Server Privacy with Secure Coprocessors."

[24] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching."

[25] Dawn Xiaodong Song, David Wagner, and Adrian Perrig, "Practical Techniques for Searches on Encrypted Data" (paper presented at the IEEE Symposium on Security and Privacy, Berkeley, CA, USA, May 2000), 44-55.

[26] Tran Khanh Dang, *Semantic Based Similarity Searches in Database Systems (Multidimensional Access Methods, Similarity Search Algorithms)* (PhD Thesis, FAW-Institute, University of Linz, Austria, May 2003).

[27] Rudolf Bayer and J.K. Metzger, "On the Encipherment of Search Trees and Random Access Files," *ACM Transaction on Database Systems* 1, no. 1 (March 1976): 37-52.

[28] See note 24 for more details about index structures and related algorithms.

[29] Lin and Candan, "Hiding Traversal of Tree Structured Data from Untrusted Data Stores;" Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs;" Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model."

[30] We note that the authentication and data integrity issue can be solved independently and separately from these three ones.

[31] Our classification of SP models is quite similar to the one presented in note 3.

[32] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan, "Private Information Retrieval" (paper presented at the 36th Annual IEEE Symposium on Foundations of Computer Science, USA, 1995), 41-50.

[33] Dmitri Asonov, "Private Information Retrieval – An Overview and Current Trends" (paper presented at the ECDPvA Workshop, Informatik 2001, Austria, September 2001), 889-894.

[34] Benny Chor, Niv Gilboa, and Moni Naor, "Private Information Retrieval by Keywords," (Technical Report, CS0917, Technion: Israel Institute of Technology, Department of Computer Science, 1997).

[35] Rafail Ostrovsky and Victor Shoup, "Private Information Storage" (paper presented at the 29th ACM Symposium on Theory of Computing, Texas, USA, May 1997), 294-303.

[36] Dmitri Asonov and Johann-Christoph Freytag, "Repudiative Information Retrieval" (paper presented at the ACM Workshop on Privacy in the Electronic Society, Washington, DC, USA, Nov. 2002), 32-40.

[37] Gertner, Ishai, Kushilevitz, and Malkin, "Protecting Data Privacy in Private Information Retrieval Schemes."

[38] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching."

[39] Chor, Goldreich, Kushilevitz, and Sudan, "Private Information Retrieval."

[40] Asonov, "Private Information Retrieval – An Overview and Current Trends."

[41] Chor, Gilboa, and Naor, "Private Information Retrieval by Keywords."

[42] Ostrovsky and Shoup, "Private Information Storage."

[43] Asonov and Freytag, "Repudiative Information Retrieval."

[44] Yael Gertner, Yuval Ishai, Eyal Kushilevitz, and Tal Malkin, "Protecting Data Privacy in Private Information Retrieval Schemes" (paper presented at the 30th Annual ACM Symposium on Theory of Computing, Dallas, Texas, USA, May 1998), 151-160.

[45] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching."

[46] Sean W. Smith, "Secure Coprocessing Applications and Research Issues" (Los Alamos Unclassified Release LA-UR-96-2805, Los Alamos National Laboratory, 1996).

[47] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching."

[48] Smith and Safford, "Practical Server Privacy with Secure Coprocessors."

[49] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[50] Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model."

[51] Michael K. Reiter and Aviel D. Rubin,"Crowds: Anonymity for Web Transactions," *ACM Transactions on Information and System Security* 1, no. 1 (June 1998): 66-92.

[52] Michael K. Reiter and Aviel D. Rubin, "Anonymous Web Transactions with Crowds", *Communications of the ACM* 42, no. 2 (February 1999): 32-38.

[53] Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees."

[54] Smith and Safford, "Practical Server Privacy with Secure Coprocessors."

[55] Lin and Candan, "Hiding Traversal of Tree Structured Data from Untrusted Data Stores."

[56] Smith, "Secure Coprocessing Applications and Research Issues."

[57] Reiter and Rubin,"Crowds: Anonymity for Web Transactions;" Reiter and Rubin, "Anonymous Web Transactions with Crowds."

[58] Reiter and Rubin, "Anonymous Web Transactions with Crowds."

[59] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching;" Smith and Safford, "Practical Server Privacy with Secure Coprocessors;" Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs;" Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model;" Smith, "Secure Coprocessing Applications and Research Issues."

[60] Smith and Safford, "Practical Server Privacy with Secure Coprocessors;" Smith, "Secure Coprocessing Applications and Research Issues".

[61] Oded Goldreich and Rafail Ostrovsky, "Software Protection and Simulation on Oblivious RAMs," *Journal of the ACM* 43, no. 3 (May 1996): 431-473.

[62] Smith, "Secure Coprocessing Applications and Research Issues".

[63] Gagan Aggarwal, Mayank Bawa, Prasanna Ganesan, Hector Garcia-Molina, Krishnaram Kenthapadi, Rajeev Motwani, Utkarsh Srivastava, Dilys Thomas, and Ying Xu, "Two Can Keep a Secret: A Distributed Architecture for Secure Database Services" (paper presented at the 2nd Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 2005), 186-199.

[64] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching;" Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs;" Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model;" Aggarwal, Bawa, Ganesan, Garcia-Molina, Kenthapadi, Motwani, Srivastava, Thomas, and Xu, "Two Can Keep a Secret: A Distributed Architecture for Secure Database Services."

[65] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching."

[66] Dang, *Semantic Based Similarity Searches in Database Systems.*

[67] Aggarwal, Bawa, Ganesan, Garcia-Molina, Kenthapadi, Motwani, Srivastava, Thomas, and Xu, "Two Can Keep a Secret: A Distributed Architecture for Secure Database Services."

[68] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs;" Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model."

[69] Hacigümüs, Iyer, Li, and Mehrotra, "Executing SQL over Encrypted Data in the Database-Service-Provider Model."

[70] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[71] Note that none of those two approaches supports outsourced multidimensional access methods (MAMs).

[72] The user even gets more information: pointers to nodes 2, 3, and 4, which can be used to access their contents.

[73] See note 11 for more details of possible inference and linking attacks in such situations.

[74] Lin and Candan, "Hiding Traversal of Tree Structured Data from Untrusted Data Stores."

[75] Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees."

[76] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[77] Asonov, "Private Information Retrieval – An Overview and Current Trends."

[78] One disadvantage of the approach introduced by Lin and Candan (see note 8) is that it does not support basic tree operations such as modifications, insertions, and deletions in dynamic outsourced databases.

[79] Dang, *Semantic Based Similarity Searches in Database Systems.*

[80] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[81] This is quite similar to an important assumption for PIR-protocols using multiple replication servers: these servers are not colluded with each other. It is also quite similar to the assumption for the secure multi-party computation (SMC) problem, where the parties participating in a computation want to preserve the privacy of their inputs. However, these two problems are essentially different from the problem we are addressing in the DC-UP-DP model. More discussions about the SMC problem can be found in: Wenliang Du and Mikhail J. Atallah, "Secure Multi-Party Computation Problems and Their Applications: A Review and Open Problems" (paper presented at the New Security Paradigms Workshop, USA, September 2001).

[82] We should note that, as introduced by Smith and Safford (see note 4), secure coprocessors cannot be used for the DC-UP-DP model with tree-structured data, but they can only be used for the DC-UP model with or without associated indexes and the DC-UP-DP model without tree-based indexes (see Table 1).

[83] Smith and Safford, "Practical Server Privacy with Secure Coprocessors;" Smith, "Secure Coprocessing Applications and Research Issues."

[84] Dang, *Semantic Based Similarity Searches in Database Systems.*

[85] Du and Atallah, "Protocols for Secure Remote Database Access with Approximate Matching;" Smith and Safford, "Practical Server Privacy with Secure Coprocessors;" Song, Wagner, and Perrig, "Practical Techniques for Searches on Encrypted Data."

[86] Chor, Goldreich, Kushilevitz, and Sudan, "Private Information Retrieval."

[87] Lin and Candan, "Hiding Traversal of Tree Structured Data from Untrusted Data Stores."

[88] Damiani, Vimercati, Jajodia, Paraboschi, and Samarati, "Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs."

[89] Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees."

[90] Mike Burmester, Yvo Desmedt, Rebecca N. Wright, Alec Yasinsac, "Accountable Privacy" (paper presented at the 12th International Workshop on Security Protocols, Cambridge, UK, 2004).

[91] Dang, "Privacy-Preserving Basic Operations on Outsourced Search Trees."

Tran Khanh DANG has been working as a lecturer and researcher in the School of Computing Science, Middlesex University in London (UK) since August 2003. He received his BEng. Degree in Information Technology from the IT Faculty in HCMC University of Technology (Vietnam) in 1998. He achieved the medal awarded for the best graduation student. From 1998 till 2000 he worked as a lecturer and researcher in the IT Faculty. He got a PhD scholarship from the Austrian Exchange Service (OeAD) for the period 2000-2003, and finished his PhD degree (Dr.techn.) in May 2003 at the FAW Institute, Johannes Kepler University of Linz (Austria). Dr. Dang's research interests include database and information security, similarity search and flexible query answering systems, modern information systems and applications, and distributed systems and parallel processing. *Address for Correspondence*: The Burroughs, Hendon, London NW4 4BT, United Kingdom, *E-mail*: k.dang@mdx.ac.uk.

Remote Authentication

- Secure User-Friendly Remote Authentication Schemes
- An Efficient and Secure Remote Authentication Scheme Using Smart Cards

SECURE USER-FRIENDLY REMOTE AUTHENTICATION SCHEMES

Tzung-Her CHEN, Du-Shiau TSAI, and Gwoboa HORNG

Abstract: Recently, Hwang and Li proposed a remote user authentication scheme that does not require a password table to verify the legitimacy of a legal user.[1] This method uses smart cards. To benefit from this advantage, other research works have explored adding such features as reducing the computational cost, adopting user-friendly passwords, making it easier to change user passwords, etc. However, as cryptanalysis has evolved, a series of modifications that improve the known security flaws have been made subsequently. This article deals with a security problem found in a latest modification and improves it in order to construct a more secure function. The article also highlights a feature, mutual authentication between a server and users, found in many authentication protocols but seldom found in the considered series of modifications.

Keywords: Mutual Authentication, Remote Authentication, Smart Card, User Impersonation.

Introduction

Nowadays, we can transfer money and shop using e-commerce applications. It could be predicted that, with network support, more activities will be performed without the need for a face to face contact. Hence, authentication has become one of the most significant and challenging issues in Internet commerce.

Remote authentication is the process through which one proves and verifies certain information over networks. Password-based remote authentication is one of the most commonly used authentication techniques due to its simplicity and effectiveness.

Authentication schemes generally use a password/ verification table stored at the server side. This stored-table system can easily suffer from verifier-stolen or modification attacks. Clearly, a more secure way that requires no password/ verification in the server to verify user legitimacy is required. Therefore, ID-based [2] authentication schemes [3,4] have been proposed to remove the requirement of having a password/ verification table stored on the server.

INFORMATION & SECURITY. An International Journal, Vol.18, 2006, 111-121.

Hwang and Li have proposed a remote user authentication scheme using smart cards.[5] The main advantage is that a password table is not required to verify a user's legitimacy. Unfortunately, several security flaws have been identified in their method.[6,7,8,9]

Inspired by the scheme proposed by Hwang and Li, Sun [10] has presented an efficient scheme with two main advantages: (1) low communication and computation costs and (2) no password table required. Unfortunately, the user password is computed by the server and is too lengthy to be memorized easily. Hwang [11] and Chien [12] have independently proposed hash-function-based schemes with much lower computation than before. However, Hwang's scheme does not provide mutual authentication, while the scheme proposed by Chien provides mutual authentication but suffers from the parallel session attack.[13] A scheme proposed by Wu and Chieu [14] in 2003, focuses on user friendliness, allowing the users to freely choose and change their passwords.

This article first determines the security flaw in the Wu-Chieu's scheme and then eliminates it to form a new approach (named shortly Method 1). Second, the authors propose another scheme (Method 2) to highlight a feature, mutual authentication between a server and remote users, found in many authentication protocols but seldom addressed in the considered approaches.

The remainder of this paper is organized as follows. In the next section, the Wu-Chieu's scheme is briefly described and its security flaw outlined. Two enhancements are proposed afterwards. This is followed by a discussion and security analysis. Conclusions are given in the last section.

Review and Weakness of Wu-Chieu's Scheme

The Wu-Chieu's password authentication scheme consists of three phases [15]: registration, login, and authentication phases.

Registration Phase

Step R.1 $U \rightarrow S : ID, PW$

In the registration phase, a user U sends his/her identity ID and password PW to the server S through a secure channel.

Step R.2 $S \rightarrow U$: smart card $\{ID, A, B, h(.), p, q\}$

Upon receiving the registration request, the server computes the following values:

- $A = h(ID, x)$, where x is the server's secret key and $h(.)$ is a collision resistant one-way function.

- $B = g^{A \cdot h(PW)} (\text{mod p})$ where p is a large prime number, and g is a primitive element in GF(p) .

The server then writes $\{ID, A, B, h(.), p, q\}$ into the user's smart card and releases it to the user.

Login Phase

Step L.1 $U \rightarrow S : ID, B^*, C, T$

In the login phase, user U inserts his/her smart card into a login device and enters his/her PW^* . The smart card will calculate the following values:

- $B^* = g^{A \cdot h(PW^*)} (\text{mod p})$
- $C = h(T \oplus B)$, where T is the current date and time and $\oplus$ is the exclusive-OR operation.

The client then sends the login message $\{ID, B^*, C, T\}$ to the remote system.

Authentication Phase

Upon receiving the login message at time T' , the server performs the following operations:

- Checks the format of ID .
- Checks the time interval validation between T and T' ; whether $(T' - T) \geq \Delta T$ (the expected valid time interval for transmission delay). If the time interval is invalid the system rejects the login request.
- Computes $C^* = h(T \oplus B^*)$ and checks if C^* is equal to the received C . If it holds, it implies that the input PW^* is equal to PW and the user is authenticated; otherwise, this login request is rejected.

Security of Wu-Chieu's Scheme

In the Wu-Chieu's scheme, a user requesting login is authenticated if his/her login message $\{ID, B^*, C, T\}$ satisfies $C = h(T \oplus B^*)$. No one is able to forge $C = h(T \oplus B) = h(T \oplus g^{A \cdot h(PW)})$ due to the fact that C has to be derived from PW and A , calculated from the server secret key x .

The authors will demonstrate below that an attacker can impersonate a legal user and pass server authentication by successfully forging the login message using the following methods:

Attack 1

Assume that an attacker has intercepted the last login message $\{ID, B^*, C, T\}$. Now, s/he calculates $C^{**} = h(T^{**} \oplus B^*)$, where T^{**} is the current date and time. Then s/he sends $\{ID, B^*, C^{**}, T^{**}\}$ to the remote server.

After receiving the login request, the server computes $C^{***} = h(T^{**} \oplus B^*)$ and checks if C^{***} is equal to the received C^{**}. Unfortunately, it holds and the attacker is authenticated.

Attack 2

An even simpler attack can be launched by first computing $c = h(T \oplus b)$, where b is a randomly selected number and t is the current date and time. Then $\{ID, b, c, t\}$ is sent to the remote server as a login message.

After receiving the login message, the server computes $C = h(T \oplus b)$ and checks if C is equal to the received c. Unfortunately, this will hold and an attacker will be authenticated.

The Proposed Schemes

In this section, a secure enhanced scheme is proposed as Method 1 to improve the security of the Wu-Chieu's scheme. The seldom addressed issue, mutual authentication, will be highlighted and solved in Method 2.

The Proposed Security Enhancement (Method 1)

First, an improved version of the Wu-Chieu's scheme is described below.

The *registration phase* goes as follows.

Step R.1 $U \rightarrow S : ID, h(PW)$

A user U sends his/her identity ID and the hash value of the password PW to the server S in a secure way.

Step R.2 $S \rightarrow U$: smart card $\{ID, A, B, h(.), p, q\}$

Upon receiving the registration request, the server computes the following values:

- $A = h(ID, x)$.
- $B = g^{A \cdot h(PW)} (\text{mod p})$.

Then the server writes $\{ID, A, B, h(.), p, q\}$ into the user's smart card and releases it to the user.

The *login phase* goes as follows.

Step L.1 $U \rightarrow S : ID, B^*, C, T$

The user U inserts his/her smart card into a login device and enters his/her password PW^*. The smart card will calculate the following values:

- $B^* = g^{A \cdot h(PW^*)} (\text{mod p})$.
- $C = h(T \oplus B \oplus A)$, where T is the current date and time.

The client then sends the login message $\{ID, B^*, C, T\}$ to the remote server.

Upon receiving the login message at time T', the server performs the following operations:

- Checks the format of ID and the validation of the time interval between T and T'.
- Computes $C^* = h(T \oplus B^* \oplus h(ID, x))$, and checks if C^* is equal to the received C. If it holds, it implies that the input PW^* is equal to PW and the user is authenticated; otherwise, this login request is rejected.

In case the user wants to change his/her password, the following operations are performed.

- The user inputs his new password PW^*.
- The smart card computes new $B = g^{A \cdot h(PW^*)} (\text{mod p})$ and updates it.

The Proposed Scheme with Mutual Authentication (Method 2)

Although Method 1 enhances the security of the Wu-Chieu's scheme, it does not provide mutual authentication. Method 2 addresses this issue. The registration phase is the same as that of Method 1 and will be omitted here. The login phase goes as follows.

Step L.1 $U \rightarrow S : ID, B^*, C_1, C_2, T$

U inserts his/her smart card into a login device and enters his/her PW^*. The smart card will calculate the following values:

- $B^* = g^{A \cdot h(PW^*)} (\text{mod p})$.

- $C_1 = h(T \oplus B \oplus A)$, where T is the current date and time.
- $C_2 = r \oplus A$, where r is a random number as a challenge for the remote server.

Step L.2 $S \rightarrow U : h(r)$

Upon receiving the login message at the time T', the server performs the following operations:

- Checks the format of ID and the validation of the time interval between T and T'.
- Computes $C_1^* = h(T \oplus B^* \oplus h(ID, x))$, and checks if C_1^* is equal to the received C_1. If it holds, it implies that the input PW^* is equal to PW and the user is authenticated; otherwise, this login request is rejected.
- Extracts r from C_2 using $h(ID, x)$.
- Computes and sends $h(r)$ to the user as a response.

Upon receiving $h(r)$, the user checks the validation of $h(r)$. If it holds, the server is authenticated. The user is convinced that the server, which he is going to communicate with, is a regular one.

The change of the user password is the same as in Method 1 and will not be described here.

Discussion and Security Analysis

In Method 1, the proposed scheme replaces $C = h(T \oplus B)$ in the login phase of the Wu-Chieu's scheme with $C = h(T \oplus B \oplus A)$. Hence, an attacker has no efficient way to forge C, which equals $C = h(T \oplus B \oplus A)$, without knowing A, protected by the smart card.

In Method 2, the identity of the user is verified by checking if the hash value of $T \oplus B^* \oplus h(ID, x)$ is equal to the received C_1 in Step L.1. The identity of the server is verified by checking if the server possesses the secret key x to generate $A = h(ID, x)$ and uses A to extract r from C_2. If the user receives the server response $h(r)$, it implies that authentication for the server is indirectly proved since only both the legal user and the regular server know A or $h(ID, x)$.

The advantages of the proposed schemes in terms of adding important functions can be summarized as follows.

- *Reducing computation cost*: In the login and authentication phases, the proposed schemes need one modular exponential operation (similarly to the Wu-Chieu's scheme), while five operations are required in the scheme proposed by Hwang and Li.[16] The methods proposed by Chien, Jan, and Tseng,[17] Hwang, Lee, and Tang,[18] and Sun [19] do not require modular exponential operations other than one-way hash functions. The computation cost for a secure one-way hash function is not yet addressed clearly; however, it is widely believed that one-way functions exist [20] and its computation cost is lower than that of a modular exponential operation. In fact, no function has been found that is really a one-way yet. Modular exponentiation is well-regarded as a candidate for a one-way function.
- *Eliminating verification table*: The proposed schemes do not require a password/ verification table to verify the users. Hence, it provides higher security level of the system and reduces cost of maintaining the sensitive tables on the server.
- *Securing password from server*: If the server knows the user password, it is possible that the server will impersonate a legal user, an especially sensitive issue in such applications as electronic accounting, electronic transfer, etc. In the proposed schemes, the user sends in the registration phase $h(PW)$ to the server but not PW. It reduces the possibility of revealing PW to the server.
- *Choosing friendly password*: Passwords are useful if kept secret, in providing additional security protection in case the smart card is lost. If a password is not convenient to use or not friendly, it will not be used at all or it will be used incorrectly. Hence, the proposed schemes provide also this feature.
- *Changing password easily*: The proposed schemes offer this alternative function to facilitate simplicity, friendliness and effectiveness.
- *Providing mutual authentication*: Most authentication schemes provide only unilateral authentication, making it is possible for an attacker to impersonate the server to fool the legal user into divulging security information. In some situations, mutual authentication is an important feature with a higher security level.

Table 1 compares the proposed schemes with several related schemes.

Remote authentication schemes could be attacked from the client side, in the transmission channel, and from the server side. This is in addition to password-guessing attacks. In the transmission channel, an attacker can intercept or modify the login message between the user and the server and pretend that s/he is the user or the server.

Table 1: Functional Comparisons among a Series of Related Remote User Authentication Schemes.

	Hwang-Li [21]	*Sun* [22]	*Wu-Chieu* [23]	*Tang-Lee-Hwang* [24]	*Chien-Jan-Tseng* [25]	*Proposed Method* 2
Computation	Medium	Extremely Low	Low	Extremely Low	Extremely Low	Low
Verification table	No	No	No	No	No	No
Server - know password	Yes	Yes	Yes	No	Yes	No
Friendly password	No	No	Yes	Yes	Yes	Yes
User - change password	No	No	Yes	Yes	No	Yes
Mutual authentication	No	No	No	No	Yes	Yes

From the client side, an attacker could impersonate a legal user to login to the server (user impersonation attacks) or merely replay the intercepted login message (replay attacks).

From the server side s/he may impersonate the server to fool a legal user (server impersonation attacks); or modify the authentication message to cheat the server. Of course, an attacker may find other ways to steal the password/ verification table stored on the server (verifier-stolen attacks) to guess a password, perform impersonation operations or just modify the table (modification attacks) to deny legal users from being able to successfully login.

To demonstrate the work of the proposed schemes, in what follows the authors will present the possible attacks against password authentication scheme.

Password Guessing Attacks

The login message is $B^* = g^{A \cdot h(PW^*)}$. If an attacker intercepts B^*, it is not possible to guess the user password without knowing A since s/he has no feasible way to determine the correct password.

User Impersonating Attacks

In Method 1, an attacker may impersonate a legal user by forging a login request $\{ID, B^*, C, T\}$. Due to the fact that the server checks the $T \oplus B^* \oplus h(ID, x)$ hash value an attacker must have $h(ID, x)$ or A to compute $C = h(T \oplus B^* \oplus h(ID, x))$ in Step L.1 so as to pass authentication. However, s/he has no idea about the server's

secret key x to obtain $h(ID, x)$. The attacker will have no efficient way to find A from $C = h(T \oplus B^* \oplus A)$ or $B^* = g^{A \cdot h(PW^*)}$ due to the NP-hardness of the problem of breaking one-way hash functions and solving discrete logarithm.

Similarly, in Method 2, the attacker cannot extract A from $C_2 = r \oplus A$. He faces the same challenge of not knowing the server's secret key x.

Replay Attacks

The login message is refreshed for each login phase by introducing a timestamp. Hence, in both Methods 1 and 2, an attacker cannot login to the remote server by replaying a previous login message.

Server Impersonating Attacks

Method 1 focuses on how to verify the identity of a user for a server, but not on verifying the legality of a server. This attack is not discussed here.

In Method 2, if an attacker attempts to impersonate the remote server successfully, he must send the exact $h(r)$ to the client (see Step L.2 in Method 2). The client will compute the hash value of r and compare it with the received $h(r)$. If equal, server authentication will be successful. This implies that the attacker has to extract r from C_2. He cannot extract r for the reason that he faces an NP-hard computation problem.

Verifier-Stolen Attacks & Modification Attacks

Because no password/ verification table is stored on the server, verifier-stolen and modification attacks are not possible.

Conclusions

Friendly passwords are very useful, if kept secret, in protecting from theft by providing another defense line. To avoid the risk of revealing any sensitive information from the password/ verification table, it is a better strategy to eliminate the sensitive table from the server. Two solutions have been proposed in this paper. The first proposed password authentication scheme removes the sensitive table security flaw from the server. The second proposed scheme adds a mutual authentication feature. Compared with other related schemes, the proposed schemes provide higher security. The authors have demonstrated that the proposed schemes are reliable and secure.

Notes:

[1] Min-Shiang Hwang and Li-Hua Li, "A New Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 1 (February 2000): 28-30.

[2] Shigeo Tsujii and Toshiya Itoh, "An ID-Based Cryptosystem Based on the Discrete Logarithm Problem," *IEEE Journal on Selected Areas in Communications* 7, no. 4 (May 1989): 467-473.

[3] Chin-Chen Chang and Shin-Jia Hwang, "Using Smart Cards to Authenticate Remote Passwords," *Computers and Mathematics with Applications* 26, no. 7 (1993): 19-27.

[4] Chin-Chen Chang and Tzong-Chen Wu, "Remote Password Authentication with Smart Cards," *IEE Proceedings – Part E* 138, no. 3 (1991): 165-168.

[5] Hwang and Li, "A New Remote User Authentication Scheme Using Smart Cards."

[6] Chi-Kwong Chan and L.M. Cheng, "Cryptanalysis of a Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 4 (November 2000): 992-993.

[7] Chin-Chen Chang and Kuo-Feng Hwang, "Some Forgery Attacks on a Remote User Authentication Scheme Using Smart Cards," *Informatica* 14, no. 3 (2003): 289-294.

[8] Kai-Chi Leung, L.M. Cheng, Anthony S. Fong, and Chi- Kwong Chan, "Cryptanalysis of a Modified Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 49, no. 4 (November 2003): 1243-1245.

[9] Jau-Ji Shen, Chih-Wei Lin, and Min-Shiang Hwang, "A Modified Remote User Authentication Scheme Using Smart Cards," IEEE *Transactions on Consumer Electronics* 49, no. 2 (May 2003): 414-416.

[10] Hung-Min Sun, "An Efficient Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 4 (November 2000): 958-961.

[11] Yuan-Liang Tang, Cheng-Chi Lee, and Min-Shiang Hwang, "A Simple Remote User Authentication Scheme," *Mathematical and Computer Modelling* 36 (2002): 103-107.

[12] Hung-Yu Chien, Jinn-Ke Jan, and Yuh-Min Tseng, "An Efficient and Practical Solution to Remote Authentication: Smart Card," *Computers & Security* 21, no. 4 (2002): 372-375.

[13] Chien-Lung Hsu, "Security of Chien et al.'s Remote User Authentication Scheme Using Smart Cards," *Computer Standards and Interfaces* 26, no. 3 (2004): 167-169.

[14] Shyi-Tsong Wu and Bin-Chang Chieu, "A User Friendly Remote Authentication Scheme with Smart Cards," *Computers & Security* 22, no. 6 (2003): 547-550.

[15] Wu and Chieu, "A User Friendly Remote Authentication Scheme with Smart Cards."

[16] Hwang and Li, "A New Remote User Authentication Scheme Using Smart Cards."

[17] Chien, Jan, and Tseng, "An Efficient and Practical Solution to Remote Authentication: Smart Card."

[18] Hwang, Lee, and Tang, "A Simple Remote User Authentication Scheme."

[19] Sun, "An Efficient Remote User Authentication Scheme Using Smart Cards."

[20] Shafi Goldwasser, "The Search for Provably Secure Cryptosystems," in *Cryptology and Computational Number Theory, Proceedings of Symposia in Applied Mathematics* 42, ed. C. Pomerance (Washington: American Mathematical Society, 1990), 89-113.

[21] Hwang and Li, "A New Remote User Authentication Scheme Using Smart Cards."

[22] Sun, "An Efficient Remote User Authentication Scheme Using Smart Cards."

[23] Wu and Chieu, "A User Friendly Remote Authentication Scheme with Smart Cards."

[24] Tang, Lee, and Hwang, "A Simple Remote User Authentication Scheme."

[25] Chien, Jan, and Tseng, "An Efficient and Practical Solution to Remote Authentication: Smart Card."

TZUNG-HER CHEN was born in Tainan, Taiwan, Republic of China, in 1967. He received his B.S. degree from the National Taiwan Normal University (Department of Information & Computer Education) in 1991 and his M.S. degree from Feng Chia University (Department of Information Engineering), in 2001. In 2005, he obtained his Ph.D. degree from the National Chung Hsing University (Department of Computer Science). He has been Assistant Professor in the Department of Computer Science and Information Engineering at the National Chiayi University since August 2005. His research interests include information hiding, multimedia security, digital rights management, and network security. He is an honorary member of the Phi Tau Phi Scholastic Honor Society. *Address for Correspondence*: Department of Computer Science and Information Engineering, National Chiayi University, 300 University Road, Chia-Yi City, Taiwan 600, R.O.C.; *Fax*: 886-5-2717741; *E-mail*: thchen@mail.ncyu.edu.tw.

GWOBOA HORNG received his B.S. degree in Electrical Engineering from National Taiwan University in 1981 and his M.S. and Ph.D. degrees from University of Southern California in 1987 and 1992 respectively, all in Computer Science. Since 1992, he has been on the faculty of the Institute of Computer Science at National Chung-Hsing University, Taichung, Taiwan, R.O.C. His current research interests include artificial intelligence, cryptography and information security. *Address for Correspondence*: Department of Computer Science, National Chung Hsing University, 250 Kuo-Kuang Road, Taichung, Taiwan 402, R.O.C.

DU-SHIAU TSAI received his B.S. degree from the Providence University (Department of Computer Science and Information Management), Taiwan, in 1996 and his M.S. degree from the National Chung-Hsing University (Institute of Computer Science), Taiwan, in 2003. He is currently pursuing his Ph.D. degree in the Institute of Computer Science, National Chung-Hsing University. His research interests include cryptography, information security and digital watermarking. *Address for Correspondence*: Department of Computer Science, National Chung Hsing University, 250 Kuo-Kuang Road, Taichung, Taiwan 402, R.O.C. and Department of Information Management, Hsiuping Institute of Technology, 11, Gongye Rd., Dali City, Taichung County, Taiwan 412, R.O.C.

AN EFFICIENT AND SECURE REMOTE AUTHENTICATION SCHEME USING SMART CARDS

Chin-Chen CHANG and Jung-San LEE

Abstract: Security of data communication becomes a crucial challenge due to the rapid development of computer and information technologies. To ensure security of resource transmission, engineers have proposed numerous schemes for protection. Among them, the remote password authentication schemes using smart cards are regarded as very efficient. As a result, smart-card based authentication schemes has become a popular research topic in recent years. In 2000, Hwang and Li proposed a new remote authentication scheme using smart cards based on ElGamal's cryptosystem. Unfortunately, the scheme Hwang and Li propose suffers from some security flaws. In this article, the authors propose a practical and secure version providing mutual authentication, increasing the authentication efficiency, and allowing the user to choose and change his/ her password at will.

Keywords: Communications, Remote Authentication, Smart Cards, Resource Protection.

Introduction

The control of the access to remote resources has become a crucial challenge due to the rapid development of computer and information technologies. Generally, most of the resources provided on Internet are not free for all users. Often, some services on the remote servers are paid. In other words, the providers of the services/ facilities have to put their resources under appropriate protection. The password authentication schemes are usually considered as the most efficient and practical method to achieve the goal of protecting remote resources. In the password authentication schemes, the user sends his/ her identity (ID) and password (PW) to the server in order to get authentication when he/ she wants to access facilities on the remote system. ID and PW are issued to the user by the remote server in the registration phase. If the user is authenticated successfully, the user will be authorized to access the facilities provided by the remote system; otherwise, the access request will be rejected.[1,2,3,4,5,6,7,8,9,10]

INFORMATION & SECURITY. An International Journal, Vol.18, 2006, 122-133.

In 1981, Lamport proposed a remote authentication scheme for communication through insecure channels.[11] Lamport's scheme can withstand the replay attack and it needs a verification table to verify the legality of the login user. Nevertheless, the verification table makes Lamport's scheme suffer from the stolen-verifier attack if somehow an intruder succeeds to get the stored verifier. Recently, a great deal of password authentication schemes using smart cards has been proposed to improve the traditional authentication schemes. The smart card is used to authenticate the legality of the user such that it is unnecessary for the remote server to store the verification table. Hence, the stolen-verifier attack can be resolved.[12,13,14,15,16,17,18,19] What is more important, a timestamp is often used to resist the replay attack. However, sophisticated hardware is needed to support the concurrent mechanism.

In 1994, Chang and Liao proposed a remote authentication scheme based on ElGamal's signature scheme.[20] Later, Wu proposed an efficient remote authentication scheme based on a simple geometric approach.[21] Wu's scheme makes possible the user to choose his/ her password at will. Unfortunately, Hwang has found and outlined weaknesses in Wu's scheme.[22] In 2000, Hwang and Li proposed a novel remote authentication scheme (Hwang-Li's scheme) using smart cards based on ElGamal's cryptosystem.[23] Shortly afterwards, Chan and Cheng demonstrated that the scheme of Hwang and Lee is insecure as well.[24] Besides, in 2003, Shen, Lin, and Hwang presented another attack on Hwang-Li's scheme.[25] Furthermore, mutual authentication between the remote server and the user is essential to ensure the security of the data transmitted over the insecure networks. But in most of these schemes, only the user is authenticated.[26,27,28] Therefore, in this article the authors propose an efficient and secure authentication scheme. The proposed scheme not only provides mutual authentication between the remote server and the user but also increases the efficiency of authentication. Furthermore, the attacks that Hwang-Li's scheme cannot resist do not affect the proposed scheme.

The rest of the paper is organized as follows. The article first reviews the scheme proposed by Hwang and Li. Some attacks on Hwang-Li's scheme are presented afterwards. A novel scheme is then proposed. Next, the security of the proposed scheme is analyzed, followed with discussions. Finally, the authors give some conclusions and outline directions for future research.

Review of Hwang-Li's Scheme

This section reviews the scheme proposed by Hwang and Li and then presents possible attacks on that scheme.

Review of Hwang-Li's Scheme

In this subsection, the authors briefly review the remote password authentication scheme of Hwang and Li.[29] The security of Hwang-Li's scheme is based on the ElGamal's public key cryptosystem. Hwang-Li's scheme is divided into three phases: registration phase, login phase and authentication phase. The details of the three phases are shown below.

Registration Phase

Let x be a secret key maintained by the system and $H(\cdot)$ be the public one-way hash function. A new user U_i needs to submit his/her identity ID_i to the system first for registration. The remote system chooses a large prime number P and computes U_i's password PW_i as follows:

$$PW_i = (ID_i)^x \bmod P,$$

The registration center then issues a smart card containing $H(\cdot)$ and P and sends PW_i to U_i through a secure channel.

Login Phase

When the user U_i decides to access data from the remote site, he/she has to insert his/her smart card into the input device and type his/her identity ID_i and password PW_i first. The smart card then executes the following procedure:

- *Step 1*: Generates a random number r.
- *Step 2*: Computes $C_1 = (ID_i)^r \bmod P$.
- *Step 3*: Computes $t = H(PW_i \oplus T) \bmod (P-1)$, where T is the current time-stamp of the input device and $\oplus$ denotes an exclusive "or" operation.
- *Step 4*: Computes $K = (ID_i)^t \bmod P$.
- *Step 5*: Computes $C_2 = K(PW_i)^r \bmod P$.
- *Step 6*: Sends the message $M = \{C_1, C_2, ID_i, T\}$ to the remote server.

Authentication Phase

Upon receiving M from U_i, the system authenticates U_i by performing the following procedure:

- *Step 1*: Checks the validity of ID_i. If the format of ID_i is not correct, rejects the access request; otherwise, the procedure goes to the next step.
- *Step 2*: Checks the validity of the time interval between T and T'. If $(T - T') \geq \Delta T$, rejects the access request; otherwise, performs the next step. Here T' is the current timestamp of the system and ΔT is the acceptable time interval of the transmission delay.
- *Step 3*: Checks if $C_2(C_1^x)^{-1} \bmod P = (ID_i)^{H(PW_i \oplus T)}$. If it holds, the access request is accepted; otherwise, terminates the connection.

Review of Chan-Cheng's Attack

In 2000, Chan and Cheng [30] pointed out that Hwang-Li's scheme cannot resist the masquerade attack. This attack could be described as follows. Suppose that a user U_c wants to counterfeit other legal users to access facilities on the remote system. U_c submits his/her ID_c to the remote system for registration. The server then issues a smart card and the corresponding password PW_c to him/her after the identity is verified. Now, U_c can generate a legal user identity ID_f and the corresponding password PW_f by computing

$$ID_f = (ID_c \cdot ID_c) \bmod P, \text{ and}$$

$$PW_f = (ID_f)^x = (PW_c \cdot PW_c) \bmod P.$$

Therefore, U_c can successfully login to the remote system to access facilities with the counterfeit (ID_f, PW_f).

Review of Shen-Lin-Hwang's Attack

In 2003, Shen, Lin, and Hwang presented another attack on Hwang-Li's scheme.[31] Their attack could be described as follows. Suppose that a user U_c wants to impersonate a legal user U_h to access the facilities on the remote system. Because ID_h is public, U_c can choose his/her $ID_c = (ID_h)^z \bmod P$, where z is a random integer chosen by U_c and $\gcd(z, \phi(P)) = 1$. The user U_c then submits his/her ID_c to the remote system for registration. Upon receiving the registration request from U_c, the remote server will verify the identity attached to the registration request and compute $PW_c = (ID_c)^x \bmod P$. The registration center then sends PW_c and issues a

smart card containing $H(\cdot)$ and P to U_c. As a result, U_c can derive U_h's password PW_h as follows:

$$\begin{aligned}(PW_c)^{-z} \bmod P &= ((ID_c)^x)^{-z} \bmod P \\ &= (((ID_h)^z)^x)^{-z} \bmod P \\ &= (ID_h)^x \bmod P \\ &= PW_h \bmod P.\end{aligned}$$

Then, U_c successfully obtains U_h's password to access the remote system as U_h.

The Proposed Scheme

Due to the fact that the computational ability of the smart card is usually low, it is quite important to reduce the computation load of the smart card to increase the efficiency of authentication. In addition, mutual authentication between the remote system and the user is also an essential requirement in remote password authentication schemes as mentioned above. Therefore, the authors propose an efficient password authentication scheme that not only ensures mutual authentication between the remote system and the user to enhance security, but also increases the efficiency of authentication. Besides, the proposed scheme also overcomes the security weaknesses present in Hwang-Li's scheme. In the proposed scheme, in order to make the password authentication scheme more user-friendly, the user is permitted to choose his/her password at will.

The scheme proposed in this article also consists of three phases: registration phase, login phase and authentication phase. Figure 1 presents the flowchart of the login and authentication phases of the scheme. The three phases could be described as follows.

Registration Phase

Let p and q be the secret keys maintained by the system and $H(\cdot)$ be the one-way hash function, where p and q are large prime numbers. g is a primitive element in $GF(n)$, where $n = p \cdot q$. First, for registration, a new user U_i has to submit his/her identity ID_i and the password PW_i chosen by himself/ herself to the system using a secure channel. After receiving the registration request, the remote system performs the following operations:

- *Step 1*: Computes $H(PW_i)$.
- *Step 2*: Computes *temp* as follows: $temp = g^{ID_i^{-1}} \bmod n$.

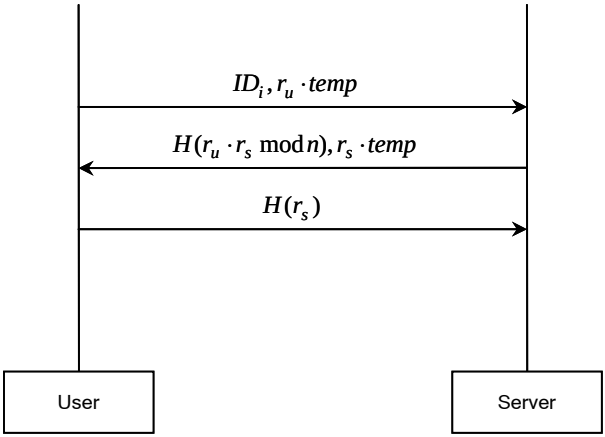


Figure 1: Flowchart of Login and Authentication Phases of the Proposed Scheme.

- *Step 3*: Computes $H(PW_i) \oplus temp$.
- *Step 4*: Issues a smart card containing $\{ (H(PW_i) \oplus temp)$, n, $H(\cdot) \}$ to U_i.

Login Phase

When the user U_i wants to access the facilities offered by the remote system, he/she has to insert the smart card into the input device and enter ID_i and PW_i. The smart card then executes the following procedure:

- *Step 1*: Generates a random number r_u, which is used only once.
- *Step 2*: Computes $H(PW_i)$.
- *Step 3*: Retrieves $temp$.
- *Step 4*: Computes $H(PW_i) \oplus (H(PW_i) \oplus temp)$.
- *Step 5*: Computes $r_u \cdot temp$.
- *Step 6*: Sends the message $M = \{ID_i, r_u \cdot temp\}$ to the remote system.

Authentication Phase

After sending M to the system, the system and the user U_i execute the following procedure to authenticate each other.

- *Step 1*: While receiving M, the remote server generates a random integer r_s, then computes $(((temp)^{-1} \bmod n) \cdot (r_u \cdot temp)) \bmod n$ to retrieve r_u, where r_s is used only once.
- *Step 2*: The system then computes $H(r_u \cdot r_s \bmod n)$ and $r_s \cdot temp$ and sends the message $M' = \{H(r_u \cdot r_s \bmod n),\ r_s \cdot temp\}$ to U_i.
- *Step 3*: Upon receiving M' sent from the remote system, the smart card retrieves r_s and computes $((temp)^{-1} \bmod n) \cdot (r_s \cdot temp) \bmod n$.
- *Step 4*: The smart card then computes $H(r_u \cdot r_s \bmod n)$ and compares it with the received one. If they are not equivalent, the connection is terminated; otherwise, the remote system is authenticated successfully by the user U_i, and the phase continues.
- *Step 5*: The smart card computes $H(r_s)$ and sends the result to the remote system.
- *Step 6*: After receiving the transmitted message sent from U_i, the remote server computes $H(r_s)$ and compares it with the received one. If they are not equal, the authentication fails; otherwise, U_i is authenticated successfully by the remote server.

Security Analysis

This section will demonstrate that the proposed remote password authentication scheme is secure with the following attack scenarios.

The Replay Attack

The replay attacks cannot work on the proposed scheme. That is, retransmitting neither the login message $M = \{ID_i, r_u \cdot temp\}$ in the login phase nor the response message $M' = \{H(r_u \cdot r_s \bmod n), r_s \cdot temp\}$ in the authentication phase will succeed since the validity of M and M' can be checked with the random numbers r_u and r_s, respectively. Even if an intruder eavesdrops M successfully in the login phase and replays M to fool the remote server, he/she will fail in Step 3 of the authentication phase. The reason is that only the legal user can retrieve r_s chosen by the remote system, and r_s is used only once. On the other hand, if an intruder eavesdrops M' in the authentication phase and retransmits it to fool the login user, he/she will fail in Step 4 of the authentication phase. The value of $H(r_u \cdot r_s \bmod n)$ is not equal to the replayed

one since the random number r_u is used only once and it is absolutely different from the replayed one.

Deriving the Secret Key of the Remote Server

When an intruder attempts to derive the secret key of the remote system, p and q, from n, he/she will fail. The reason is that he/she will encounter the difficulties of solving the factoring problems.

Losing the Smart Card

An intruder will also fail in the case when he/she steals the smart card of the legal user U_i and wants to use it to access the facilities provided by the remote server. The reason is that the intruder can not retrieve $temp$ without knowing PW_i.

The Server Spoofing Attack

When a masqueraded server intends to cheat an innocent user, it cannot fake a response message $M'' = \{H(r_u \cdot r_s \bmod n), r_s \cdot temp\}$. First, it is due to the fact that the intruder cannot retrieve $temp$ as stated before. Second, without knowing $temp$, r_u, used only once, also cannot be obtained successfully.

The Stolen-Verifier Attack

In the scheme proposed in this article, no verification table is needed. That is, it is impossible for an attacker to mount the stolen-verifier attack on it.

Discussion

The security flaws of the remote password authentication scheme proposed by Hwang and Li do not affect the scheme proposed here. On the other hand, when a legal user U_i wants to change his/her password, he/she only needs to submit his/her smart card to the remote system together with a new password PW_i' through a secure channel. The remote system then re-computes $H(PW_i')$ and $(H(PW_i') \oplus temp)$. Then, the server uses $(H(PW_i') \oplus temp)$ to replace the original one stored in U_i's smart card. After the replacement, the user U_i can use the new password PW_i to login to the remote server. Therefore, it is easy for the user to change the password and to remember it. In other words, the proposed scheme enables the users to choose and change their passwords at will.

Table 1 provides comparison between the characteristics of the proposed scheme and the scheme of Hwang and Li. Among the considered characteristics, mutual authenti-

Table 1: Comparison between the Proposed Scheme and Hwang-Li's Scheme.

	Proposed Scheme	*Hwang-Li's Scheme*
Security Flaw	No	Yes
Choose and Change Password at Will	Yes	No
Verification Table	No	No
Concurrent Mechanism	No	Yes
Mutual Authentication	Yes	No

cation is only present in the proposed scheme to enhance security as elaborated in a previous section. In addition, the proposed scheme can withstand the replay attack without employing the timestamp as shown above. In other words, the scheme presented in this article does not require an additional sophisticated concurrent mechanism to resist the replay attack.

What is more important, in what follows the authors present efficiency comparisons between the proposed scheme and Hwang-Li's scheme in order to demonstrate that the scheme proposed here is not only more secure but also more efficient. Nowadays, most of the servers have high computational power, while the smart cards are still with low computation ability. The key point in improving the efficiency of the authentication processes is reducing the computation load of the smart card. Therefore, the authors have made the computation load of the smart card as light as possible in the scheme they propose.

First, the authors define the symbols used in Table 2. e denotes the exponential computation operation. h denotes the computation operation of the one-way hash func-

Table 2. Comparison of the Efficiency between the Proposed Scheme and Hwang-Li's Scheme.

	Proposed Scheme		*Hwang-Li's Scheme*	
	Smart Card	*System*	*Smart Card*	*System*
Computation of RP	0	$1e$, $1h$, $1\oplus$	0	$1e$
Computation of LP	$1h$, $1\oplus$	0	$3e$, $1h$, $1\oplus$	0
Computation of AP	$2h$	$1e$, $1h$	0	$3e$, $1h$, $1\oplus$

tion. ⊕ denotes XOR. RP means Registration Phase. LP means Login Phase. AP means Authentication Phase. Considering the several computation operations, the exponential operation is of main concern for the whole computation load since the computation load of the exponential operation is far heavier than that of the other operations. According to Table 2, it is evident that the proposed scheme is more efficient than Hwang-Li scheme. It is because only one exponential operation is needed for mutual authentication.

Conclusions

In this article, the authors propose a novel remote password authentication scheme that overcomes the security weaknesses of Hwang-Li's scheme. The proposed scheme provides mutual authentication between the remote system and the user such that the server spoofing attack cannot have an effect. Besides, the scheme allows the user to choose and change passwords at will and can resist the replay attack without sophisticated concurrent mechanisms.

As mentioned above, the proposed scheme is more efficient, secure and user-friendly than Hwang-Li's scheme. Since the computation load of both the smart card and the whole system is quite low, the proposed scheme is efficient, secure, user-friendly to be applied in practice; moreover it could be employed on imbalanced networks as well.

Notes:

[1] Jau-Ji Shen, Chih-Wei Lin, and Min-Shiang Hwang, "A Modified Remote User Authentication Scheme Using Smart Cards," IEEE *Transactions on Consumer Electronics* 49, no. 2 (May 2003): 414-416.

[2] Min-Shiang Hwang and Li-Hua Li, "A New Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 1 (February 2000): 28-30.

[3] Shyi-Tsong Wu and Bin-Chang Chieu, "A User Friendly Remote Authentication Scheme with Smart Cards," *Computers & Security* 22, no. 6 (2003): 547-550.

[4] Hung-Min Sun, "An Efficient Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 4 (November 2000): 958-961.

[5] Suguru Yamaguchi, Kiyohiko Okayama, and Hideo Miyahara, "Design and Implementation of an Authentication System in WIDE Internet Environment," in *Proceedings of the 10th IEEE Regional Conference on Computers and Communication Systems* (Hong Kong, 24-27 September 1990), 653-657.

[6] Chin-Chen Chang and Tzong-Chen Wu, "Remote Password Authentication with Smart Cards," *IEE Proceedings-E* 138, no. 3 (1991): 165-168.

[7] Cheng-Chi Lee, Li-Hua Li, and Min-Shiang Hwang, "A Remote User Authentication Scheme Using Smart Cards," *ACM Operating Systems Review* 36, no. 4 (2002): 23-29.

[8] Yuan-Liang Tang, Cheng-Chi Lee, and Min-Shiang Hwang, "A Simple Remote User Authentication Scheme," *Mathematical and Computer Modelling* 36 (2002): 103-107.

[9] Tzong-Chen Wu and Hung-Sung Sung, "Authenticating Passwords over an Insecure Channel," *Computers & Security* 15, no. 5(1996): 431-439.

[10] Kaijun Tan and Hongwen Zhu, "Remote Password Authentication Scheme Based on Cross-Product," *Computer Communications* 22, no. 4 (March 1999): 390-393.

[11] Leslie Lamport, "Password Authentication with Insecure Communication," *Communications of the ACM* 24, no. 11 (November 1981): 770-772.

[12] Shen, Lin, and Hwang, "A Modified Remote User Authentication Scheme Using Smart Cards."

[13] Chin-Chen Chang and Wen-Yuan Liao, "A Remote Password Authentication Scheme Based upon ElGamal's Signature Scheme," *Computers & Security* 13, no. 2 (April 1994): 137-144.

[14] Min-Shiang Hwang, "Cryptanalysis of a Remote Login Authentication Scheme," *Computer Communications* 22 (1999): 742-744; Hwang and Li, "A New User Authentication Scheme Using Smart Cards."

[15] Chin-Chen Chang and C. S. Liah, "Comment on Remote Password Authentication with Smart Cards," *IEE Proceedings-E* 139, no. 4 (1992): 372-372.

[16] Hung-Yu Chien, Jinn-Ke Jan, and Yuh-Min Tseng, "An Efficient and Practical Solution to Remote Authentication: Smart Card," *Computers & Security* 21, no. 4 (2002): 372-375.

[17] Debbie McElroy and Efraim Turban, "Using Smart Cards in Electronic Commerce," *International Journal of Information Management* 18, no. 1 (1998): 61-72.

[18] Chin-Chen Chang and Shin-Jia Hwang, "Using Smart Cards to Authenticate Remote Passwords," *Computers and Mathematics with Applications* 26, no. 7 (1993): 19-27.

[19] Shiuh-Jeng Wang and Jin-Fu Chang, "Smart Card Based Secure Password Authentication Scheme," *Computers & Security* 15, no. 3 (1996): 231-237.

[20] Tan and Zhu, "Remote Password Authentication Scheme Based on Cross-Product."

[21] Tzong-Chen Wu, "Remote Login Authentication Scheme Based on a Geometric Approach," *Computer Communications* 18, no. 12 (1995): 959-963.

[22] Min-Shiang Hwang, "Cryptanalysis of a Remote Login Authentication Scheme," *Computer Communications* 22, no. 8 (1999): 742-744.

[23] Hwang and Li, "A New User Authentication Scheme Using Smart Cards."

[24] Chi-Kwong Chan and L.M. Cheng, "Cryptanalysis of a Remote User Authentication Scheme Using Smart Cards," *IEEE Transactions on Consumer Electronics* 46, no. 4 (November 2000): 992-993.

[25] Shen, Lin, and Hwang, "A Modified Remote User Authentication Scheme Using Smart Cards."

[26] Wu and Chieu, "A User Friendly Remote Authentication Scheme with Smart Cards."

[27] Sun, "An Efficient Remote User Authentication Scheme Using Smart Cards."

[28] Chien, Jan, and Tseng, "An Efficient and Practical Solution to Remote Authentication: Smart Card."

[29] Hwang and Li, "A New User Authentication Scheme Using Smart Cards."

[30] Chan and Cheng, "Cryptanalysis of a Remote User Authentication Scheme Using Smart Cards."

[31] Shen, Lin, and Hwang, "A Modified Remote User Authentication Scheme Using Smart Cards."

JUNG-SAN LEE see page 84

CHIN-CHEN CHANG see page 84

I&S Monitor

- Cybersecurity Related Internet Sources
- Managing Cybersecurity Resources: A Cost-Benefit Analysis

CYBERSECURITY RELATED INTERNET SOURCES

USEFUL SITES, PORTALS AND ORGANIZATIONS

CEO CyberSecurity Resource Center

http://www.technet.org/cybersecurity/

For many CEOs, information security risk management is a new area of responsibility. This web site helps CEOs come up to speed on information security risk management issues and quickly assess their own company's cyber preparedness.

Carnegie Mellon CyLab

http://www.cylab.cmu.edu

The CyLab in Carnegie Mellon University attempts to create a public-private partnership to develop new technologies for measurable, available, secure, trustworthy, and sustainable computing and communications systems and to educate individuals at all levels. CyLab is a university-wide, multidisciplinary initiative involving more than 200 faculty, students, and staff that builds on more than two decades of Carnegie Mellon's leadership in information technology. CyLab works closely with the CERT® Coordination Center, a leading, internationally recognized center of Internet security expertise. Through its connection to the CERT/CC, CyLab also works closely with US-CERT—a partnership between the Department of Homeland Security's National Cyber Security Division (NCSD) and the private sector—to protect the U.S. information infrastructure.

CyberSecurity Institute

http://www.cybersecurityinstitute.biz/

The Cybersecurity Institute is a world leader for digital forensics training. The more technical definition that the CyberSecurity Institute uses to describe computer forensics or forensic computing in the vein of computer crime or computer misuse is as follows: "The preservation, identification, extraction, interpretation, and

documentation of computer evidence, to include the rules of evidence, legal processes, integrity of evidence, factual reporting of the information found, and providing expert opinion in a court of law or other legal and/or administrative proceeding as to what was found."

Institute for Information Infrastructure Protection

http://www.thei3P.org/

The Institute for Information Infrastructure Protection (I3P) is an U.S. consortium of leading cybersecurity research and development organizations including universities, federally funding labs and non-profit organizations. The goals of the I3P are to address research and policy-related aspects of the vulnerabilities inherent in the information infrastructure, bring experts together to identify and mitigate threats aimed at the U.S. information infrastructure, and promote collaboration and information sharing among academia, industry, and government.

Institute for Security Technology Studies at Dartmouth College

http://www.ists.dartmouth.edu/

The Institute for Security Technology Studies (ISTS) at Dartmouth College and its core program on cybersecurity and information infrastructure protection research serve as a principal U.S. center for counter-terrorism technology research, development, and assessment. The institute is dedicated to pursuing interdisciplinary research and education for cybersecurity and emergency response technology. ISTS is also a member of the Institute for Information Infrastructure Protection (I3P). The site includes a large bibliography and description of current research.

CyberSecurity and Emergency Preparedness Institute

http://csepi.utdallas.edu/

The CyberSecurity and Emergency Preparedness Institute at the University of Texas at Dallas (UTD) focuses primarily on performing innovative digital forensics, information assurance and emergency preparedness research in areas including network survivability, rapidly deployable networks, sensor networks, reconfigurable hardware, self healing software, anti-piracy methods, signal processing, data mining, high assurance systems engineering, emergency response information systems and others. Importantly, the researchers have continually demonstrated their ability to deliver comprehensive, practical solutions at the device, system and network level.

Internet Crime Complaint Center (IC3)

http://www.ic3.gov/

The Internet Crime Complaint Center (IC3) is a partnership between the Federal Bureau of Investigation (FBI) and the U.S. National White Collar Crime Center (NW3C). IC3's mission is to address fraud committed over the Internet. The IC3 gives the victims of cybercrime a convenient and easy-to-use reporting mechanism that alerts authorities of suspected criminal or civil violations. For law enforcement and regulatory agencies at the federal, state, local and international level, IC3 provides a central referral mechanism for complaints involving Internet related crimes.

The United States Computer Emergency Readiness Team

http://www.us-cert.gov/

The United States Computer Emergency Readiness Team (US-CERT) is a partnership between the Department of Homeland Security and the public and private sectors. Established in 2003 to protect the nation's Internet infrastructure, US-CERT coordinates defense against and responses to cyber attacks across the U.S. US-CERT interacts with federal agencies, industry, the research community, state and local governments, and others to disseminate reasoned and actionable cybersecurity information to the public. Information is available from the US-CERT web site, mailing lists, and RSS channels. US-CERT also provides a way for citizens, businesses, and other institutions to communicate and coordinate directly with the United States government about cyber security.

CERT Coordination Center

http://www.cert.org/

The Computer Emergency Readiness Team (CERT) is a center of Internet security expertise, located at the Software Engineering Institute, a federally funded research and development center operated by Carnegie Mellon University. The Center studies Internet security vulnerabilities; research is done on long-term changes in networked systems; and information and training is developed to help in improving security.

Australian Computer Emergency Response Team (AusCERT)

http://www.auscert.org.au/

AusCERT is the national Computer Emergency Response Team for Australia and a leading CERT in the Asia/Pacific region. As a trusted Australian contact within a

worldwide network of computer security experts, AusCERT provides computer incident prevention, response and mitigation strategies for members, a national alerting service and an incident reporting scheme.

U.S. National Computer Crime Squad, Federal Bureau of Investigation

http://www.emergency.com/fbi-nccs.htm/

The FBI's National Computer Crime Squad (NCCS) investigates violations of the Federal Computer Fraud and Abuse Act of 1986. These crimes cross multiple state or international boundaries. Violations of the Computer Fraud and Abuse Act include intrusions into government, financial, most medical, and Federal interest computers. Federal interest computers are defined by law as two or more computers involved in a criminal offense, which are located in different states. Therefore, a commercial computer which is the victim of an intrusion coming from another state is a "Federal interest" computer.

The Federal Bureau of Investigation (FBI) has a Computer Crime Squad Web page that contains contact information for the Squad.

The Purdue University Center for Education and Research in Information Assurance and Security (CERIAS)

http://www.cerias.purdue.edu/

The Purdue University Center for Education and Research in Information Assurance and Security (CERIAS) is currently viewed as one of the world's leading centers for research and education in information assurance and security. CERIAS is unique in its multidisciplinary approach to the problems, ranging from purely technical issues (e.g., intrusion detection, network security, etc) to ethical, legal, educational, communicational, linguistic, and economic issues, and the interactions and dependencies among them. The following areas summarize the research focus areas for the faculty involved with the center: Risk Management, Policies, and Laws; Trusted Social and Human Interactions; security awareness, education, and training; assurable software and architectures; enclave and network security; incident detection, response, and investigation; identification, authentication, and privacy; and cryptology and rights management.

The Computer Security Resource Center of the U.S. National Institute of Standards and Technology (NIST)'s Computer Security Division

http://csrc.nist.gov/

The Computer Security Division (CSD) is one of eight divisions within NIST's Information Technology Laboratory. The mission of NIST's Computer Security Division is to improve information systems security by: (1) Raising awareness of IT risks, vulnerabilities and protection requirements, particularly for new and emerging technologies; (2) Researching, studying, and advising agencies of IT vulnerabilities and devising techniques for the cost-effective security and privacy of sensitive systems; (3) Developing standards, metrics, tests and validation programs; and (4) Developing guidance to increase secure IT planning, implementation, management and operation.

The SANS Institute

http://www.sans.org/

The SANS (SysAdmin, Audit, Network, Security) Institute was established in 1989 as a cooperative research and education organization. Its programs now reach more than 165,000 security professionals, auditors, system administrators, network administrators, chief information security officers, and CIOs who share the lessons they are learning and jointly find solutions to the challenges they face. At the heart of SANS are the many security practitioners in government agencies, corporations, and universities around the world who invest hundreds of hours each year in research and teaching to help the information security community. The SANS Institute is the most trusted and by far the largest source for information security training and certification in the world. It also develops, maintains, and makes available at no cost, the largest collection of research documents about various aspects of information security, and it operates the Internet's early warning system – Internet Storm Center.

Many SANS resources, such as the weekly vulnerability digest (@RISK), the weekly news digest (NewsBites), the Internet's early warning system (Internet Storm Center), flash security alerts and more than 1,200 award-winning, original research papers are free to all.

Forum for Incident Response and Security Teams (FIRST)

http://www.first.org/

FIRST is a leading organization in incident response. It brings together a variety of computer security incident response teams from government, commercial, and educational organizations. FIRST aims to foster cooperation and coordination in incident prevention, to stimulate rapid reaction to incidents, and to promote information sharing among members and the community at large. Membership in

FIRST enables incident response teams to more effectively respond to security incidents – reactive as well as proactive.

FIRST also provides value added services, such as access to up-to-date best practice documents, technical colloquia for security experts, hands-on classes, annual incident response conference, publications and web-services and special interest groups

At present FIRST has more than 170 members, spread over the Americas, Asia, Europe and Oceania.

ON-LINE PUBLICATIONS

Cybercrime and Cybersecurity Communication

http://europa.eu.int/ISPO/eif/InternetPoliciesSite/Crime/CrimeCommEN.html

This is a communication from the Commission of the European Communities to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions on creating a safer information society by improving the security of information infrastructures and combating computer-related crime.

Perspective: The Left and Right Hands of Cybersecurity

http://news.com.com/2010-1009_3-6036384.html

This is a commentary by Eric J. Sinrod, partner at Duane Morris. A report from the National Association of State Chief Information Officers (NASCIO) criticizes the Department of Homeland Security (DHS) for failing to coordinate with state and local law enforcement against cyberthreats. The report finds that state and local agencies would rather work with DHS than with the private sector, which has proven detached from local interests. The NASCIO recommend adding cybersecurity to DHS's State Homeland Security Assessment and Strategy process. State and local governments need better training in best practices, cybersecurity, risk assessment, and continuity of operations. The report also finds that DHS needs to deliver information in a timelier manner, arguing that "more emphasis needs to be placed on external-directed attacks, and internal ineptitude and maliciousness." State and local agencies also need better academic and educational opportunities.

Convention Committee on Cybercrime (T-CY)

http://www.coe.int/T/E/Legal_affairs/Legal_co-operation/Combating_economic_crime/6_Cybercrime/T-CY/

Owing the dangers of cybercrime and the need for common minimum technical and legal standards to fight such crime at a global level, the Convention on cybercrime (ETS N° 185) was prepared by Council of Europe member States and Canada, Japan, South Africa and the United States. It entered into force on 1 July 2004. Its Additional Protocol concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems (ETS N° 189) will enter into force on 1 March 2006.

The Convention is the only binding international instrument dealing with cybercrime. It has received widespread international support and is open to all states. The Convention provides for consultations of the parties.

Cyber Criminals Stepping up Targeted Attacks

http://news.zdnet.com/2100-1009_22-6046606.html

Symantec Internet Security Threat report says that during the second half of 2005, attackers continued to move away from broad attacks seeking to breach firewalls and routers and are now taking aim at the desktop and Web applications.

The report said threats such as viruses, worms, and trojans that can unearth confidential information from a user's computer rose to 80 percent of the top 50 malicious software code threats from 74 percent in the previous six months.

The Truth about Cyberterrorism

http://www.cio.com/archive/031502/truth.html

"Since September 11, threats once considered digital aggravations have been tagged cyberterrorist provocations. What cyberterrorism really means, according to the National Infrastructure Protection Center, is an act perpetrated through computers that results in violence, death and/or destruction, and creates terror for the purpose of coercing a government to change its policies. To qualify as cyberterrorism, an act must fulfill two criteria: a political motivation and a destructive result. Most computer attacks satisfy only the first criterion. It's far less likely than the media would have us believe that cyberterrorists could cause destruction, especially to the nation's physical infrastructure. More credible is the danger to critical data: a cyberterrorist who hacks critical computer systems to steal or irreversibly damage vital data, such as the Social Security database. The good news for CIOs is that protecting against any security threat protects against cyberterrorism."

Cybersecurity for the Homeland

http://hsc.house.gov/files/cybersecurityreport12.06.04.pdf

This report discusses the activities and findings of the Chairman and Ranking member of the House Subcommittee on Cybersecurity, Science, and Research & Development of the Select Committee on Homeland Security. This report addresses the following areas: case for action, role of the Department of Homeland Security, subcommittee oversight, and cybersecurity roadmap for the future.

VerySign's White Paper on Cybersecurity

http://www.verisign.com/static/005567.pdf

Today, national security and the concomitant need to protect the nation's critical infrastructure and maintain the continuity of government and financial services are equally important drivers. To meet the requirements for protecting national security and to address internal business requirements for online security, cyber systems must be able to share data securely; ensure the continuous availability of critical services; interoperate across federal, state, and local systems; and comply with federal consumer-privacy regulations. All this has been discussed in VerySign's white paper on cybersecurity.

Creating a National Framework for Cybersecurity: An Analysis of Issues and Options

http://www.thecre.com/pdf/secure/20050404_cyber.pdf

This Congressional Research Service (CRS) report (February 2005) discusses: (1) what is Cybersecurity; (2) where are the major weaknesses in cybersecurity; (3) what are the major means of leverage; and (4) what roles should government and the private sector play.

Information Security: Emerging Cybersecurity Issues Threaten Federal Information Systems (May 2005)

http://www.au.af.mil/au/awc/awcgate/gao/d05231.pdf

This report by the Government Accountability Office (GAO), United States, discusses (1) the potential risks to federal information systems from emerging cybersecurity threats such as spam, phishing, and spyware; (2) the 24 Chief Financial Officers Act agencies' reported perceptions of these risks and their actions and plans to mitigate them; (3) government and private-sector efforts to address these emerging cybersecurity threats on a national level, including actions to increase consumer

awareness; and (4) government-wide challenges to protecting federal information systems from these threats.

Guide for Developing Performance Metrics for Information Security

http://csrc.nist.gov/publications/drafts/draft-sp800-80-ipd.pdf

NIST's Computer Security Division has completed the initial public draft of Special Publication 800-80, Guide for Developing Performance Metrics for Information Security. This guide is intended to assist organizations in developing metrics for an information security program. The methodology links information security program performance to agency performance. It leverages agency-level strategic planning processes and uses security controls from NIST SP 800-53, Recommended Security Controls for Federal Information Systems, to characterize security performance. To facilitate the development and implementation of information security performance metrics, the guide provides templates, including at least one candidate metric for each of the security control families described in NIST SP 800-53.

Critical Infrastructure Protection: Department of Homeland Security Faces Challenges in Fulfilling Cybersecurity Responsibilities

http://www.gao.gov/new.items/d05434.pdf

This report by the Government Accountability Office (GAO), United States, discusses (1) Department of Homeland Security (DHS)'s roles and responsibilities for cyber critical infrastructure protection, (2) the status and adequacy of DHS's efforts to fulfill these responsibilities, and (3) the challenges DHS faces in fulfilling its cybersecurity responsibilities.

Assessing the Risks of Cyber Terrorism, Cyber War and Other Cyber Threats

http://www.shaneland.co.uk/ewar/docs/dissertationsources/institutionalsource1.pdf

This is a report by James A. Lewis, published by the Center for Strategic & International Studies (CSIS). The author discusses issues related to cyber-terrorism and cyber attacks on critical infrastructure and their implications for national security.

Special Issue on Cybercrime of the International Journal of Communications Law and Policy (IJCLP)

http://www.ijclp.org/Cy_2004/index.html

The International Journal of Communications Law and Policy and the Yale Journal of Law and Technology published in autumn 2004 Issue 9 on Cybercrime in two parts. It features the following articles:

- Architectural Regulation and the Evolution of Social Norms (by Lee Tien)
- Transborder Search: A New Perspective in Law Enforcement? (by Nicolai Seitz)
- The Fourth Amendment Unplugged: Electronic Evidence Issues & Wireless Defenses - Wireless Crooks & the Wireless Internet Users Who Enable Them (by Tara McGraw Swaminatha)
- Launch on Warning: Aggressive Defense of Computer Systems (by Curtis E. A. Karnow)
- Real World Problems of Virtual Crime (by Beryl A. Howell)
- Distributed Security: Moving away from Reactive Law Enforcement (by Susan W. Brenner)
- The Price of Restricting Vulnerability Publications (by Jennifer Stisa Granick)
- Technology, Security and Privacy: The Fear of Frankenstein, the Mythology of Privacy and the Lessons of King Ludd (by Kim A. Taipale)
- Surfing While Muslim: Privacy, Freedom of Expression & the Unintended Consequences of Cybercrime Legislation (by Jason M. Young)
- Privacy vs. Piracy (by Sonia K. Katyal)
- Characteristics of a Fictitious Child Victim: Turning a Sex Offender's Dreams into His Worst Nightmare (by James F. McLaughlin).

Defending against Cybercrime and Terrorism: A New Role for Universities

http://www.fbi.gov/publications/leb/2005/jan05leb.pdf
http://www.fbi.gov/publications/leb/2005/jan2005/jan2005.htm#page14

This is an article by Tony Aeilts, published in FBI-Law Enforcement Bulletin, Washington, vol. 74, no.1 (January 2005), p.14-20. The article discusses the need to include college and university resources in the fight against cybercrime and the threat of terrorism.

Reducing Opportunities for e-Crime

http://www.eurim.org.uk/activities/ecrime/reducingops.doc

"The paper focuses on the need for industry and law enforcement to work together to produce practical, plain English guidance for users at all levels, but most especially small firms and consumers, on what to do to protect themselves and what to do when they suspect they have been victimized. That guidance needs to include material on identifying and assessing risk and what to do about it. Similar guidance is needed for large organizations because un-prioritized governance paperchases, to meet the demands of regulators, can serve to increase vulnerability by diverting resources and attention from practical action."

MANAGING CYBERSECURITY RESOURCES: A COST-BENEFIT ANALYSIS

A book by Lawrence A. Gordon and Martin Loeb

The book was just published by McGraw-Hill. A fundamental argument throughout the book is that the proper use of economic concepts will allow organizations (in both the private and public sectors) to achieve a higher level of cybersecurity than otherwise possible. This argument is developed by providing an economic framework for:

- Determining the appropriate amount for an organization to invest in cybersecurity, and
- Procedures for efficiently allocating such resources to particular cybersecurity activities.

The book also provides chapters on The Business Case for Cybersecurity, Risk Management and Cybersecurity, Cybersecurity Auditing, and Cybersecurity's Role in National Security.

More information about the book can be found at:

www.rhsmith.umd.edu/faculty/lgordon/cybersecuritybook.htm.

Address for Correspondence:

Lawrence A. Gordon, Ph.D.
Ernst & Young Alumni Professor of Managerial Accounting and Information Assurance
Director, Ph.D. Program
Affiliate Professor in University of Maryland Institute for Advanced Computer Studies
Robert H. Smith School of Business

3359 Van Munching Hall
University of Maryland
College Park, MD 20742-1815
(301) 405-2255 TEL
(301) 314-9611 FAX
lgordon@rhsmith.umd.edu
http://www.rhsmith.umd.edu
http://www.rhsmith.umd.edu/faculty/lgordon/

INSTRUCTIONS TO CONTRIBUTORS

Editorial Policies

Information & Security: An International Journal publishes articles on scientific and technical issues related to national and international security in the Information age, information operations, information warfare, command and control warfare, critical information technologies, computer aided exercises, simulators, and information security. It is published four times per year both in electronic form in Internet and on paper. All papers will be in English. Articles of exceptional quality in Bulgarian or Russian are also accepted for publication, in which case we shall publish an extended abstract in English.

Manuscripts are usually reviewed by at least two referees for significance and scholarliness. Every effort is made to inform authors within three months whether papers are accepted, require revision prior to possible publication, or are rejected.

All accepted manuscripts are edited for adherence to Journal format and style, clarity, syntax, and punctuation. Authors must transfer copyright in writing to the publisher when an article is accepted.

The Journal also publishes book reviews and documents, lists of recent relevant articles and Internet addresses, and readers' comments. It presents scholars, researchers, research centers, companies and products.

Manuscript Preparation and Submission

Submitted articles should be no longer than 20 double-spaced typewritten pages, including double-spaced endnotes (or no more than 5,000 words overall). They should be accompanied by a cover letter giving the paper's title and the name, mailing address, e-mail address, and telephone number of the corresponding author. They should be also accompanied by an abstract of 200-300 words and a brief statement summarizing the author's present affiliation, publishing career, and research interests. It is recommended, when possible, that translation of the abstract in the other two languages is included. DO NOT indicate authors' names on manuscript pages. DO NOT reveal authors' identity through references in the text or in any other way.

Send two copies to Dr. Todor Tagarev, Managing Editor, *Information & Security*, Mladost 4, POBox 16, Sofia 1715, Bulgaria. *Include a diskette* with all files prepared with commonly used word-processing software.

Alternatively, you may send all computer files to**:** infosec@mbox.digsys.bg

Number endnotes consecutively; these numbers must correspond to those in the text. Endnote should follow *The Chicago Manual of Style*. Examples:

Book: 1. Carl von Clausewitz, *On War*, Anatol Rapoport, editor (London: Penguin Books, 1968), 164-67.

Article: 2. William Owens, "The Emerging System of Systems," *Military Review* 75, 3 (May-June 1995), 15-19.

Chapter: 3. David Alberts, "The Future of Command and Control with DBK," in *Dominant Battlespace Knowledge*, ed. Stuart E. Johnson and Martin C. Libicki (Washington: National Defense University, 1996), 67-88.

Subsequent shortened citations should read as follows:

1. Clausewitz, *On War*, 31.
2. Owens, "System of Systems," 17.
3. Alberts, "Future of C2," in *DBK*, 73.

NOTE: It is understood that submitted articles have not been previously published and are not currently under review for publication elsewhere.

.

ISSN 1311-1493

Information & Security

An International Journal

Volume 18, 2006

This volume is from the I&S publication plan for 2005.

Publisher: ProCon Ltd

Mladost 4, POBox 16
Sofia 1715, Bulgaria
Fax: (+359 2) 946 8355

E-mail: infosec@procon.bg

Internet Issue:
http://infosec.procon.bg

.