ABSTRACT

Authentication: Can Mobile Environments be Secured? (April 1999)

Charles Austin Cropper

Faculty Advisor: Dr. Udo W. Pooch

A mobile system is defined as a network in which one or more of the interconnection links is a wireless medium. Wireless media include but are not limited to, cellular or radio transmissions, satellite services, and wireless computer networks.

The fundamental operations of storage, processing, and transmission of information are undergoing such rapid improvement that the application of securing mobile systems cannot keep up with the rate of advance. This research analyzes security problems and investigates possible solutions that stem from the absence of a "fixed" link between the user and service provider in mobile systems.

This research approaches all security issues from the authentication standpoint, i.e. the process of reliably verifying the identity of two parties in a communication channel. Once identities have been verified, the channel authenticity must be maintained.

Mobile communication systems that utilize three systems, symmetric ciphers, public key systems, and zero-knowledge techniques, are shown to be highly secure. The level security is not degraded due to the absence of a "fixed" link between the user and service provider.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Wireless mobile communications is one of the fastest growing sectors of the Information Technology (IT) industry. Estimations predict that by the year 2000 there will be more than 110 million mobile users, such as cellular subscribes, compared to 22 million that existed in 1995 [15]. As wireless mobile communication systems grow, data security hazards increase. Increasing information transmittal, done via wireless mobile systems, is assisted by new inventions. The fundamental operations of storage, processing, and transmission of information are undergoing such rapid improvement that securing of mobile systems, i.e. guarding against signal interception verses allowing communication access, cannot keep pace with the rate of advance [9]. With the rapid rate of advance come data security issues such as fraud, espionage, or privacy violation.

Vedder [27] cautioned that, because there is no "fixed" link between the user and a service provider, which could serve to "identify" the user for routing and billing purposes, security problems exist. Since the identity of the user must be verified over an air electromagnetic interface, authentication by means of cryptography must be established if impostors are to be precluded from taking on the identity of another individual and "transferring" or falsely attributing calls and charges.

Farmer [11] stated that despite the many practical benefits of mobile systems, activities that transpire using mobile technology result in significant new security threats. The primary complication stems from processes or transactions occurring through multiple channels, or

---

Communications of the ACM was used as a style guide for this manuscript.

domains, each with their own respective degree or level of security. Differing levels of security can adversely impact functionality and are recognized as critical to the acceptability of mobile communication systems. The "weakest" channel through which an information transaction occurs determines the degree of security.

As modern technology has led to the introduction of new methods of speech transmission, the susceptibility of communications to interception and/or intentional corruption also increases. It is now possible to intercept satellite transmissions from a secluded safe location whereas in the past an intervenor would have been obligated to dig up cables or at least maintain a presence at a particular location to intercept a message. Mobile systems, in essence, allow interception without active physical intervention. The development of mechanisms that provide privacy and high levels of security in modern communications is increasing in importance [6].

A. Statement of the Problem

Mobile communication environments do not require a physical link in the form of a copper wire, fiber optic glass, or cable between the user and the receiver. Security of the mobile communication system, therefore, becomes a *sine qua non* of the system [27]. Because messages transmitted via electronic form can be modified and/or monitored without the knowledge of the sender or receiver, security issues are considered critical to the acceptability of distributed mobile systems [11]. The purpose of this paper is to examine the literature concerning authentication, between a user and a service provider, with respect to mobile communication environments and deduce which forms of authentication may be best suited for mobile environments.

B. Definitions of Terms

Authentication. Determining whether or not a particular message, is most likely delivered to its intended recipient from its claimed source passing through a possible arbitrator [25].

Cipher. An encryption scheme, i.e. a set of transformations that change readable text (plaintext) into secret text (ciphertext) [19].

Cryptography. The art of secret writing [17]. The branch of mathematics based on the transformation of data [26].

Mobile System. A communication system requiring no physical link between the user and a service provider [27].

Protocols. Consists of a set of procedures by which activities transpire.

Smart Card. Resembles a credit card in appearance. Smart cards have embedded, single-chip microcomputers specifically designed to deliver information (stored data, computation results) and/or modify its contents (data storage, event memorization) when exposed to an outside element [14]. The smart card must be able to ascertain that a uniquely identifiable person is utilizing the card. Some smart cards inhibit use after a certain number of utilizations. Memory sizes and computation speeds are utilized together in more sophisticated cards; therefore, more and more detailed mechanisms become available ensuring increased security.

Smart cards have been used in banking systems as advanced debit cards, where the card stores pertinent information such as a person's balance. Some access cards are smart

cards. Some telephone systems use smart cards as a simple pre-payment system for phone calls and deduct the cost of a phone call from the card upon completion.

Zero-knowledge. A cryptographic technique by which possession of information can be verified without any part of that information being revealed to the verifier or to any third party [12].

CHAPTER II

REVIEW OF THE LITERATURE

This review focuses on the issues related to the security requirements and security features of a mobile communication system. Literature pertaining to authentication related issues which include security goals and threats, cryptography, and applications of authentication in heterogeneous and homogenous mobile environments is also discussed. Finally, a rationale for a multi-component system based on symmetric ciphers, modern public key algorithms, together with techniques from zero-knowledge proof algorithms will be proffered as the preferred method of security.

A. Issues Related to Securing Mobile Environments: An Overview

First and foremost, the crux of the following discussion involves the absence of a physical connection between two parties in correspondence.

Since the turn of the century, almost all information conveyance, of the electronic form, has been relayed over wires or cables. Only now as the twenty-first century arrives is information flying around the globe aided solely by satellites and the occasional base station. The presence of a "wire" created a system by which people could restrict who received or participated in conversations. Albeit, interlopers could "tap" into wires and steal private information; however, the intruders had to understand how to "tap" into the wires and, more importantly, know the precise location of the wires. With mobility, wires disappear. The air becomes the "wires" of mobile communication systems; therefore, these new "wires" lie pray to vandals, as they are everywhere. The location of the wires, once a security asset, now becomes a liability.

Of the three main operations carried out in wireless mobile environments, (1) storage, (2) processing and (3) transmission, the transmission of information carries the maximum security risks. Since it is difficult to make a widespread network physically secure, many security measures depend on information masking techniques such as cryptography. Designing a system for security means analyzing an adversary problem where both the designer and the opponents are each independently thinking out their respective strategies. This type of "game theory" is extremely difficult to master and merely serves to illustrate both the underlying complexity of the problem and the inadequacy of a naive "risk analysis" approach [9].

Mobile authentication transforms the information arena, as well as the access points to information into a fluid, amorphous environment. An individual utilizing a cellular phone must be able to verify his identity from multiple locations, all of which involve differing degrees of security. Some locations considered "hostile" might eavesdrop on conversations or in extreme cases, cause financial harm. "The increasing and increasingly diverse demand for security by users, operators, and regulatory bodies calls for more advanced security features in third-generation systems such as UMTS [Universal Mobile Telecommunications System, The European Cellular and Satellite System]" [13].

Current efforts in Europe and also by the ITU (International Telecommunication Union) have focused on standardizing access systems to wireless networks as well security protocols. Advanced Communications Technologies and Services (ACTS) has also launched a project, ASPeCT (Advanced Security for Personal Communications Technologies), whose

goal is to "specify such advanced [security] features and verify their feasibility and acceptability" [13].

B. Security Goals and Threats

Our society is dependent upon the ability to transmit spoken messages quickly and accurately. Both business and leisure activities depend upon mobile systems for information. Numerous channels for the transmission of information are available for a mobile system of communication — the worldwide telephone network and a large number of private and military radio communication systems. The number of individuals utilizing these mobile systems of communication increases each year.

As the number of mobile systems of communication users increase each year Moreau [20] estimates that the mobile communication industries will lose millions due to fraud. Prevention and detection of fraud and fraudulent activities, therefore, are important and desirable goals for mobile system users. With many individuals utilizing mobile systems for multiple purposes, security and privacy are an ever-increasing issue. The need to conceal and protect the content of one's message is important; therefore, the user and service provider must consider: (1) amount of protection required, (2) authentication techniques, and (3) cost [6].

From the initial development of mobile systems, security was considered important. Firewalls, gateways controlling access between one network and other networks or between a single computer (host) and a network, developed in the early 1990s, became first-line defenses for security. While this single-dimensional (first generation) security measure was adequate during the early stages of mobile communication, the industry of today has shifted its security efforts to multi-dimensional security techniques [4].

Multi-dimensional, multi-layer security systems are currently being developed for additional security. This second, and developing third generation, utilizes multiple methods and mechanisms to create as secure a security system as possible. This security system involves devices specifically designed for (1) prevention, (2) detection, and (3) response [4].

Preventive security tools include firewalls, Virtual Private Networks, application-level encryption, user authentication, and content screening software. These mechanisms are designed to prevent break-ins, tampering, or unwanted access.

Detection devices read a source of data, i.e. network traffic, system logs, or audit trail information, and take appropriate action upon detection. Some detection devices include network and system scanners, misuse and anomaly detectors, content screening and antiviral software.

Response devices feature response capabilities in addition to their detection features. Often referred to as "adaptive defense mechanisms," response devices sound alarms, send email messages, transmit messages to a pager, shut down a user account, shunt connections from an attacker's address, and replace damaged files [4].

Global System for Mobile communications (GSM) and Digital Enhanced Cordless Telecommunications (DECT), second generation systems, included for the first time, security features based on cryptographic techniques. These cryptographic techniques proved successful in preventing fraud. The increasingly diverse demand for security by users, operators, and regulatory bodies requires more advanced security features in third generation systems, such as UMTS [18].

ASPeCT presents several approaches to identify fraudulent behavior. Through the implementation of the rule-based approach, both the absolute and differential usage are verified against certain rules. This fraud detection approach worked best when user profiles contained explicit information, such as fraud criteria referred to as rules. The implementation of this approach is based on an existing rule-based tool for audit trail analysis known as PDAT (Protocol Data Analysis Tool). PDAT is a rule-based tool for intrusion detection developed by Siemens ZFE (Corporate Research and Development). PDAT has the possibility of online analysis and works in heterogeneous environments [20].

Neural networks, another technique utilized to identify fraudulent behavior, utilizes flexible and adaptive protocols to implement pattern recognition problems. Neural networks are systems of elementary decision units that are adapted by training in order to recognize and classify arbitrary patterns. Neural networks are currently being utilized in telephone networks throughout the world [5], [28].

In sum, there are four contemporary goals considered in information security, confidentiality of data, integrity of data, availability of data, and legitimate use of data [12]. Nearly all four goals have been described in the scenarios presented previously. Some projects focus more heavily on certain goals than others in their respective views of mobile systems.

C. Cryptography

Cryptography is formed from two Greek words, κρυπτο meaning secret and γραφη meaning writing. Coincidentally, cryptography is art of secret writing [17].

Swanson [26] stated that cryptography is a branch of mathematics based on the transformation of data. Cryptography provides an important tool for protecting information

and provides mobile system users the ability to send information between participants in such a way that third parties cannot understand the message. The most common form of cryptography represents information as numbers and mathematically manipulates the numbers. The message, in its original form, is known as plaintext. The encrypted text is known as ciphertext. As the message is received, it is converted from ciphertext into plaintext.

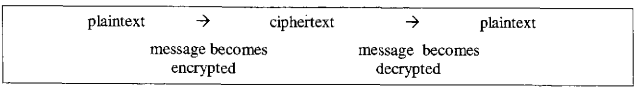

Figure 1 General Encryption Scheme

Algorithms and secret values form the foundation of cryptographic systems. The secret is known as the "key". When a key is combined with an algorithm, it becomes more difficult to develop new algorithms allowing for reversible scrambling of information. The key is analogous to the combination for a combination lock, although the concept of a combination lock is well known, one cannot open the lock without the combination [17].

Cryptography can be called upon to provide four major "services" [19].

1. *Confidentiality.* Information is kept in context with those who are authorized to access and/or use it.

2. *Integrity.* Maintenance against message modification during transit.

3. *Nonrepudiation.* Guard against denial of message transmittal when a sender truly sent a particular message.

4. *Authentication.* Ascertain a message origin as well as the identification of the sending party.

In order to ensure security, the ciphertext must not reveal any possible information about the plaintext, except possibly its length. The key should be reasonably long, and normally, a key should be not reused. By "reasonably," a key should be long enough not to degrade the strength of an algorithm. A good cryptographic algorithm keeps this information to a minimum and utilizes a compression program to reduce the size of the text before encrypting it. Compression reduces the redundancy of the message as well as the volume of work required to encrypt and decrypt [24].

In recent years, typical cipher based cryptography has been augmented to include public key algorithms. Number theory created public key algorithms stemming from difficulty in factoring extremely large numbers. Public key cryptography is useful because of its ease of configurability and implementation in network settings [17]. The gist of public key systems is the existence of a biparte key split into public and private halves. When a sender wishes to encrypt using a public key algorithm, the sender encrypts his message with the receiver's public key and only the receiver with the paired private key can decrypt the message. The mathematics involved is fairly complicated and not necessary to the understanding of the underlying concepts; thus they will not be discussed further.

Modern cryptography is used to provide numerous security devices such as digital signatures, encrypted tunnels (for Virtual Private Networks), and normal symmetric ciphers. Several important issues should be considered when designing, implementing, and integrating cryptography in a mobile system [26]:

1. *Select Design and Implementation Standards.* Managers and users of mobile systems must select among various criteria when deciding to use cryptography. Their selection

should be based on cost-effectiveness analysis, trends in the standard's acceptance, and interoperability requirements. In addition, each criterion should be carefully analyzed to determine if it is applicable to the organization and the desired application.

2. *Decide on Hardware vs. Software Implementations.* The trade-offs among security, cost, simplicity, efficiency, and ease of implementation need to be studied by managers acquiring various security products. Cryptography can be implemented in either hardware or software; each has related costs and benefits.

3. *Manage Keys.* All keys need to be protected against modification and secret keys and private keys need protection against unauthorized disclosure. Key management involves the procedures and protocols, both manual and automated, used throughout the entire life cycle of the keys. This includes the generation, distribution, storage, entry, use, destruction, and archiving of cryptographic keys.

4. *Secure Cryptographic Modules.* The proper functioning of cryptography requires the secure design, implementation, and use of the cryptographic module. Actions required include protecting the module against tampering. Cryptography is typically implemented in a module of software, firmware, hardware, or some combination thereof. This module contains the cryptographic algorithm(s), certain control parameters, and temporary storage facilities for the key(s) being used by the algorithm(s).

5. *Comply with Export Rules.* Users must be aware that the U.S. Government controls, via civil and criminal penalties, the export of cryptographic implementations. The rules governing export can be quite complex, since they consider multiple factors. In addition, cryptography is a rapidly changing field, and rules may change from time to time. Questions

concerning the export of a particular implementation should be addressed to appropriate legal counsel.

D. Applications of Authentication

Protocols abound in human society. A nominal protocol consists of a set of procedures by which activities transpire. For example, the protocol of a restaurant is to seat the customers, take their orders, prepare the orders, serve their orders, and so forth. At junctions in a protocol, decisions occur which determine whether acts, or orders, will transfer from a sending party, such as the restaurant, to a receiving party, such as customers waiting on their food. In the heart of authentication are protocols.

Authentication is nothing more than determining whether or not a particular message, or food delivery in the case of the restaurant, is most likely delivered to its intended recipient from its claimed source passing through a possible arbitrator [25]. Scenarios which typify non-authentic activities include, i) the restaurant bringing a customer a wrong order, ii) the customer ordering from a bus-boy instead of a waiter, or iii) the manager (arbitrator) voiding an order which may prevent customers' eating. All three cases keep two tasks from completing, the customers' eating and the restaurant making a sale. The case where an order is placed with the proper restaurant personnel, cooked correctly, delivered correctly to the waiting customer, and in which the customer satisfactorily consumes and pays for the meal yields an authentic event. The subset of all scenarios that yield successful transactions is determined by an authentication protocol.

Authentication can be either unilateral or mutual; i.e. is one or are both parties in a message transaction being authenticated [12]. Arbitrators can exist, like the manager of the

restaurant, and are assumed to exist only as an unbiased judge during message conveyance. Situations, as in the restaurant example, arise when either party, henceforth known as the claimant and verifier, dispute or otherwise reject each other's claim and/or identity. Cases also exist where a third party, henceforth an eavesdropper, can disrupt the flow of a message. An example, using the restaurant analogy, of an eavesdropper would be a thief that steals a customer's wallet before the customer pays for his meal.

Throughout the rest of this paper, the claimant and verifier are treated as mutually mistrusting or always adversarial. The arbitrator will act only in accordance with the rules established by an authentication system, and eavesdroppers will always be taken into account.

Generalized authentication protocols fall into three categories delineated by the mechanism they employ in verification. The basis of identification may be determined by something known, something possessed, or something inherent [9]. The attribute known, possessed, or inherent need not belong to a human but may also be determined by something non-human.

Beginning with known facts, verification can be determined through devices such as passwords, combinations, or any profusion of knowledge, which can be delivered upon request to a verifier. An individual's mother's maiden name exemplifies knowledge based verification system.

Possessed articles can also lend to verification systems. Common objects people use on a daily basis are keys. They are physical in nature and require possession in order to be useful to an individual. The situation where someone is locked out of their automobile brings

to mind authentication via possession except the required possession is inside the car rather than being within the owner's hands.

Inherent information is a dichotomy of immutable properties, including physical characteristics and involuntary actions. Examples of physical attributes include retinal patterns, fingerprints, and hand geometries. Involuntary actions include handwritten signatures and voice patterns. Both types of inherent information stem from an individual and, typically, are immutable. These attributes can be used to verify someone's identity in an authentication protocol [19].

At this stage, it must be stressed that *all* authentication protocols have a probability of failure; therefore, the goal of a "good" authentication system is to minimize the probability of deception [25]. The gist in minimizing this probability arises from restricting the set of messages which culminate in authentication. In other words, only certain events, in a certain order, will yield an authentic message. Cryptography is often employed to define this restricted subset of instances yielding authentic events.

Authentication and entity identification form another critical foundation of computer security. These components form the basis for most types of access control and for establishing user accountability. Authentication protocols utilizing cryptographic designs are based on the principle of convincing a verifier that, because a claimant knows a secret key, that claimant is the true principal. Authentication and identification form technical measures that prevent unauthorized people (or unauthorized processes) from entering a mobile system. Access control usually requires that the system have the capability of identifying and differentiating among users. For example, access control refers to the granting to users of only

those accesses minimally required to perform their respective duties. User accountability requires the linking of activities on a mobile system to specific individuals and, therefore, requires the system to identify users [12].

The following should be considered when requiring authentication [26]:

1. *Require users to authenticate.* An organization should require users to authenticate their claimed identities on IT systems. It may be desirable for users to authenticate themselves with a single log-in. This requires the user to authenticate themselves only once and then be able to access a wide variety of applications and data available on local and remote systems. A side-affect exists, though, because requiring only a single log-in creates a single point of failure. Once logged-in, an imposter can have free rein in a system. User authentication should, therefore, be augmented to include multiple levels of identification checks scattered through a user's session.

2. *Restrict access to authentication data.* An organization should restrict access to authentication data. Authentication data should be protected with access controls and unidirectional encryption to prevent unauthorized individuals, such as hackers from obtaining the data.

3. *Secure transmission of authentication data.* An organization should protect authentication data transmitted over public or shared data networks. When authentication data, such as a password, is transmitted to a mobile environment, it can be electronically monitored. This can happen on the network used to transmit the password or on the mobile system itself. Simple encryption of a password that will be used again does not solve this problem because

encrypting the same password will create the same ciphertext; the ciphertext becomes the password.

5. *Limit log-on attempts.* Organizations should limit the number of log-on attempts. Many operating systems can be configured to lock a user ID after a set number of failed log-on attempts. This helps to prevent guessing of authentication data.

6. *Secure data as it is entered.* Organizations should protect authentication data as it is entered into the mobile system, including suppressing the display of the password as it is entered and orienting keyboards away from view.

7. *Administer data properly.* Organizations should carefully administer authentication data and tokens including procedures to disable lost or stolen passwords or tokens and monitoring systems to look for stolen or shared accounts.

A *perfect* authentication system as defined by [25], is a system in which the probability of an intruder's success is equal to the uncertainty introduced into the system due to encoding. Realistic systems, however, do not have the luxury of pure coding theory, i.e. unlimited memory, unlimited time, and so forth. All mobile systems will be constrained in some way, and thus must be analyzed as being a non-perfect system.

E. Symmetric Ciphers

Common encryption systems like the Data Encryption Standard (DES) are symmetric ciphers. Symmetric ciphers are highly advanced forms of a simple Caesar cipher. A Caesar cipher is a monoalphabetic cipher, i.e. textual mapping of one letter for another [17]. With inclusion of mathematical functions and keys, monoalphabetic ciphers quickly transform into modern day symmetric ciphers.

One of the first influential devices that used a symmetric cipher was the Enigma coding machine during World War II. It enabled Adolph Hitler's communications to occur in secret during the War.

Modern symmetric ciphers are mathematical algorithms that take a message (plaintext) together with a key and perform some type of operation on the plaintext in corroboration with the key to yield the ciphertext. The operation usually consists of mathematical and/or logical manipulation

The Data Encryption Standard (DES) created the public realm of encryption in the mid-1970's when it was fully specified and released into the public domain. The Data Encryption Standard was adopted as a federal standard on November 23, 1976 and authorized for use on all unclassified government communications. It is defined under the American Federal Information Processing Standard (FIPS) 46-2 [21].

Never, before 1977, had a fully defined encryption algorithm been published and released into the public domain. The Data Encryption Standard was the first National Security Administration (NSA) sponsored algorithm released in such a fashion. The NSA thought DES was going to be used in hardware implementations only. The standard mandated a hardware implementation, but the National Bureau of Standards (NBS) now the National Institute of Standards & Technology (NIST) published enough details as to allow DES implementation in software [24].

DES is a block cipher algorithm; it encrypts data in 64-bit blocks. A 64-bit block of plaintext (normal text) enters one end of the algorithm and a 64-bit block of ciphertext

(encrypted text) exits. Since DES is symmetric, the same key and algorithm are used for both encryption and decryption.

DES uses a 56-bit key; actually a 64-bit key with 8 odd parity bits, one per byte within the key. The key can be any 56-bit number. A few combinations of weak keys & semi-weak keys exist which should be avoided as they destroy the cryptographic power of DES. All security rests within the key.

At the simplest level, the algorithm is nothing more than a combination of the two basic techniques of encryption: confusion and diffusion [24]. A combination of a substitution with a permutation, i.e. bit shuffling, constitute each of sixteen steps in DES known as a round. Rounds involve rearranging the bits within the input text in an effort to hide the original information. Just like shuffling a deck of cards, there is an optimal number of rounds (shuffles). Once the cards are sufficiently randomized, extra shuffles just waste time. The crux of a cryptographic system is determining the optimal number of rounds that sufficiently secure the data while not degrading efficiency by using too many rounds.

In modern times, more symmetric ciphers are being created and used. The prior description of DES gives a basic description of how most symmetric ciphers work. Modern algorithms include RC5, RC6, Blowfish and Twofish. The main premise, however, remains that the cryptographic system is bi-directional, i.e. encryptable and decryptable using the same key.

Symmetric ciphers, specifically DES, have been the primary references when mentioning cryptography in last twenty years. More recently, though, a system has been introduced that "splits" the key. These systems are public key systems.

F. Public Key Algorithms

Public key cryptography takes the "key" from symmetric ciphers and splits it into two different pieces. It was invented in 1976 by Diffie and Hellman [10]. The logic behind the strength of public key algorithms lies in the ability to factor a large number which is a composite of large prime numbers.

Given a public and private key pair, an individual establishes a rudimentary public key system by disseminating their public key to a key distribution center. Any person wishing to deliver a message to a specific person takes that person's respective public key and encrypts their message with the recipient's public key prior to sending it. Only the matched private key can decrypt the message. The reader is referred to [17] or [19] if desiring more information concerning mathematical underpinnings.

Public key systems, albeit a strange system of encryption, provide for more than simple message passing. Public key systems also form the foundation for digital signatures, session key establishment, as well as standard encryption like that provided in Pretty Good Privacy (PGP) [17].

Digital signature or digital certificate can be viewed as a cyberspace identification card like a driver's license [12]. The signature validates the sources of a particular public key. The algorithm used today for most digital signatures is RSA.

RSA, named after its inventors, Rivest, Shamir, and Adleman, is a public key algorithm that does encryption as well as digital signatures. It functions by algebraic properties of exponentiation, i.e. $S^{xy} = S^{yx}$ with x and y being numbers derived from the public and private keys. It functions identically for signature production.

The security of RSA together with most other public key cryptosystems is not known. It has been shown that *if* factoring large integers can be accomplished easily, then breaking the RSA encryption system is easy. After a decade of research into the strength of RSA, no easier method than factoring is known to break RSA [8]. The advantages of public key systems has been shown for years but they have their drawbacks.

"Throughput rates for the most popular public key encryption systems are several orders of magnitude slower that the best known symmetric key systems" [19]. The key sizes used in public key cryptography are substantially larger than their counterparts in symmetric key cryptography, by substantially larger, compare 126 bits to 1024 bits. The length of the keys in public key systems result in slow operation of both encryption and decryption.

To ameliorate the operational speed of public key systems, a technique is used which mixes symmetric key cryptography and public key cryptography. A "session" key is randomly chosen from the initiator of an information transaction and encrypts the random session key with the public key system. The initiator then sends this "temporary" key to the recipient. All following encrypted communications will occur with the session key using a symmetric key system. Since the key was randomly chosen and only utilized during a particular communication "session" it provides high levels of security without the high time consuming side-affects of public key cryptography [17].

Public key dissemination is another more pertinent problem, especially when applied to mobile communication environments. Deciding who and where public keys will be stored is a question left open to debate. The access becomes increasingly difficult when applied to multinational communication systems, i.e. when communication occurs across national

boundaries. The central key agencies are given the name Trusted Third Parties (TTP). The Jefferies, Mitchell, & Walker (JMW) scheme proposed by [16] created a system for TTP services. These services were originally intended to be used in conjunction with escrow systems like the Clipper system [22]. Now, the JMW system is applied to public key dissemination as well as escrowed keyed systems. One of the major areas of analysis in ASPeCT is the creation of TTPs [3].

G. Zero-Knowledge Protocols

Zero-knowledge is one form of authentication. Zero-knowledge is an *interactive* proof system that involves having the verifier issue a number of challenges to the claimant. This interactive proof system is successful if it succeeds in proving the desired statements and nothing else. In essence, the verification system occurs without aid from digital signatures, or public key encryption [19].

This latter form of cryptographic authentication shows great capacity for future application in authentication protocols. A zero-knowledge technique is a means by which "possession of information can be revealed without any part of that information being revealed, either to the verifier or to any third party" [12]. For example, zero-knowledge involves having the verifier issue a number of challenges to the claimant. The claimant responds to the responses, and the verifier is able to establish a satisfactorily high level of confidence that the claimant does possess the secret information, although no part of the secret information is actually disclosed in the responses.

Stronger zero-knowledge cryptographic techniques vary in the number of challenge-response pairs required and the intricacy of the problem solving required at both ends. Guillou

and Quisquater [14] proposed using a smart card to establish zero-knowledge cryptology. The smart card resembles a credit card, but in actuality, it is a multipurpose, tamper-resistant security apparatus. A microcircuit is placed in the plastic base of the smart card. The microcircuit is adhered to a circuit board and connected to electrical contacts on the board. The purpose of the microcircuit is to deliver stored data and computation results and/or to modify its contents when interfaced with an outside element.

"The zero-knowledge property implies that a prover executing the protocol ... does not release any information ... not otherwise computable in polynomial time from public information alone. Thus participation does not increase the chances of subsequent impersonation" [19]. Within some assumptions, [7] has shown that everything provable is provable in zero-knowledge.

The restrictions that comprise assumptions made in zero-knowledge proofs concern level of zero-knowledge, i.e. computational or perfect zero-knowledge. "A protocol is computationally zero-knowledge if an observer restricted to probabilistic polynomial-time tests cannot distinguish real from simulated transcripts. For perfect zero-knowledge, the probability distributions of the transcripts must be identical" [19]. A transcript is an information transaction, so for any information transaction to exist in a perfect zero-knowledge protocol, each interaction neither increases nor decreases the likelihood of impersonation.

Quisquater and Guillou [23] explained zero-knowledge with an interesting litany concerning a cave. The cave had a secret passage, in the case of our cave, between C and D. Alice knows the secret to the cave and wants to "prove" to Bob that she knows it. To prove to Bob the following procedure occurs:

Bob stands at A while Alice goes to C or D. Once Alice is in position, she yells to Bob who then moves to B. Bob then shouts to Alice to either come from the C or D side of the cave. Alice, using the secret of the cave, comes out either side Bob requests. This procedure occurs the number of times that satisfies Bob statistically that Alice must know the secret because the probability of her not knowing it would be increased by a factor of two each iteration. If Bob asked Alice twenty times then the probability that she would come out of the correct side every time up to twenty, assuming randomness, would be $\frac{1}{2}^{20}$ or 0.00009%. Stated differently, if Alice did not know the secret of the cave but only resorted to random chance, i.e. she would pick the same side Bob would later choose, she would have a 0.00009% chance of picking the same side twenty times consecutively. Alice obviously knows the secret of the cave.

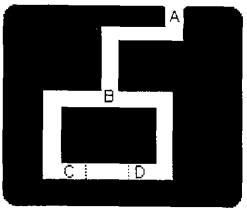

Figure 2 Zero-Knowledge Cave

Twisting the cave story to include a camcorder, if Bob recorded everything that happened previously, he would have a problem later trying to convince someone else that Alice knows the secret. The age-old cliche of "seeing is believing" applies. A third party could simply argue that the tape of Alice had been edited so that she came out the correct side of the cave each time. Thus, the property of a zero-knowledge proof exists; the real-life demonstration cannot be distinguished from the videotaped, and Bob couldn't learn anything from the real-life demonstration; therefore, Bob cannot learn anything from the taped

demonstration. "Zero" knowledge gets exchanged between Alice and Bob, *except* that Alice knows the secret to the cave.

Zero-knowledge is still in its infancy and only heavily investigated in Europe. No good physical implementations of zero-knowledge exist, whether in wired or wireless systems.

CONCLUSION

A rationale for a multi-component system based on symmetric ciphers with public key cryptography along with techniques from zero-knowledge proof algorithms will now be presented as the preferred method of security in future mobile systems.

All three of the systems listed previously have had differing levels of utilization in the last twenty years. Only within the last five years has zero-knowledge been heavily analyzed by academia. Public key systems have been in heavy use since RSA's inception, and symmetric ciphers have been used since the beginning of cryptography.

Modern mobile communication systems exhibit properties that call for at minimum, symmetric ciphers with enhancements public key cryptography such as session keys. They also stipulate partial zero-knowledge proof characteristics such as challenge-response algorithms [2].

Modern cryptography, through information theory and complexity theory, has created several protocols by which messages can be transmitted via an electronic medium and be nearly as secure as their physical counterparts. Mathematics allows many variables to be used when creating an authentication protocol.

It is the author's opinion that current mobile systems will be secured via symmetric ciphers including some form of public key cryptography. As zero-knowledge ages, it will become more easily implementable and usable in mobile systems as a means of authentication. Mobile systems will be authenticatable and highly secure. The following example will elucidate a general authentication system in a mobile communication system.

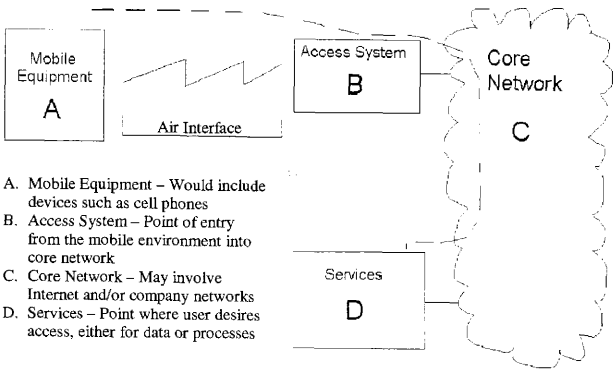

Figure 3 Authentic Mobile Network

The dashed line in figure 3 represents the data link between A and D. It is established in the following manner. First, the mobile user (A) and access system (B) establish a connection using a challenge-response system over the air interface – the best place for a zero-knowledge protocol. Second, the user from A establishes a session to D through C and B using a public key protocol. The session is established by sending a symmetric cipher key encrypted using the services' public key, and subsequently the service (D) sending an acknowledgement back to A encrypted using the session key with a symmetric cipher. All subsequent communication is authentic from point A to D including the air interface.

Current trends of society are pushing electronic media from standard phone systems or otherwise physically "wired" systems towards media which are "wireless". The cellular phone

networks that exist today typify "wireless" communication networks. "The increasing interest in wireless communications have made mobile systems possible. Mobile environments will extend the usability of current distributed systems by allowing users of mobile devices access to a large pool of networked resources ... from almost anywhere in the world" [1].

REFERENCES

1. Abdul-Rahman, A. and Hailes, S., "Security Issues in Mobile Systems." Department of Computer Science: University College London, November 1995.
2. Advanced Security for Personal Communications Technologies, "Report on the use of UIMs for UMTS," Advanced Communications Technologies & Services project AC095 report #D04, Aug. 1996, Available at http://www.esat.kuleuven.ac.be/cosic/aspect/.
3. Advanced Security for Personal Communications Technologies, "Trusted third parties: evaluation report," Advanced Communications Technologies & Services project AC095 report #D14, June 1997, Available at http://www.esat.kuleuven.ac.be/cosic/aspect/.
4. Avolio, F., "A Multi-Dimensional Approach to Internet Security: Putting It All Together," *netWorker*, Vol. 2, No. 2, April/May, 1998, pp. 15-22.
5. Barson, S., Davey, N., McAskie, G., and Frank, R., "The Detection of Fraud in Mobile Phone Networks", *Neural Network World*, Vol. 6, No. 4, July/August 1995, pp. 477-484.
6. Beker, H.J., and Piper, F.C., *Secure Speech Communications*, Academic Press, New York, 1984.
7. Ben-Or, M., Goldreich, O., Goldwasser, S., Håstad, J., Kilian, J., Micali S., and Rogaway, P., "Everything Provable is Provable in Zero-Knowledge," *Advances in Cryptology – CRYPTO '88* (Santa Barbara, California, Aug. 21-25, 1988), Springer-Verlag, pp. 37-56.
8. Cormen, T., Leiserson, C., and Rivest, R., *Introduction to Algorithms*, McGraw-Hill, New York, 1990.

9. Davies, D. and Price, W., *Security for Computer Networks*, Wiley & Sons, New York, 1984.

10. Diffie, W. and Hellman, M., "New directions in cryptography," *IEEE Trans. Information Theory*, Vol. IT-22, No. 11, Nov. 1976, pp. 472-492.

11. Farmer, W., Guttman, J., and Swarup, V., "Security for Mobile Agents: Authentication and State Appraisal", *Computer Security – ESORICS '96* (Rome, Italy, Sep. 25-27, 1996), Springer, pp. 118-130.

12. Ford, W., *Computer Communications Security*, Prentice Hall, New Jersey, 1994.

13. Francis, C., Herbrig, H., and Jefferies, N., "Secure Provision of UMTS Services over Diverse Access Networks", *IEEE Communications Magazine*, Vol. 36, No. 2., Feb. 1998, pp. 128-136.

14. Guillou, M., Ugon, M., and Quisquater, J., "The Smart Cart: A Standardized Security Device Dedicated to Public Cryptology," in *Contemporary Cryptology: The Science of Information Integrity*, Ed. Simmons, G., IEEE Press, New York, 1992, pp. 561-613.

15. Horn, G., and Preneel, B., "Authentication and Payment in Future Mobile Systems", *Computer Security – ESORICS '98* (Louvain-la-Neuve, Belgium, Sep. 16-18, 1998), Springer, pp. 277-293.

16. Jefferies, N., Mitchell, C., and Walker, M., "A Proposed Architecture for Trusted Third Party Services," *Cryptography: Policy and Algorithms* (Brisbane, Australia, July 3-5, 1995), Springer, pp. 98-104.

17. Kaufman, C., Perlman, R., and Speciner, M., *Network Security*, Prentice Hall, New Jersey, 1995.

18. Martin, K., Preneel, B., Mitchell, C., Hitz, H., Horn, G., Poliakova, A., and Howard, P., "Secure Billing for Mobile Information Services in UMTS," *Intelligence in Services and Networks: Technology for Ubiquitous Telecom Services: 5th International Conference on Intelligence in Services and Networks, IS&N '98* (Antwerp, Belgium, May 25-28, 1998), Springer, pp. 535-548.

19. Menezes, A.J., Van Oorschot, P.C., and Vanstone, S.A., *Handbook of Applied Cryptography*, CRC Press, New York, 1997.

20. Moreau, Y., Preneel, B., Burge, P., Shawe-Taylor, J., Stoermann, C., and Cooke, C., "Novel Techniques for Fraud Detection in Mobile Telecommunication Networks", ACTS Mobile Summit, Granada, Spain, Nov. 27, 1996, Available at http://www.esat.kuleuven.ac.be/cosic/aspect/.

21. National Bureau of Standards. FIPS Publication 46-2: Data Encryption Standard. 1977, revised December 1993, Available at http://www.itl.nist.gov/div897/pubs/fip46-2.htm.

22. National Institute of Standards and Technology. FIPS Publication 185: Escrowed Encryption System. February 1994. Available at http://www.itLnist.gov/div897/pubs/fip185.htm.

23. Quisquater, J., Guillou, L., and Berson, T., "How to Explain Zero-Knowledge Protocols to Your Children," *Advances in Cryptology – CRYPTO '89* (Santa Barbara, California, Aug. 20-24, 1989), Springer-Verlag, pp. 628-631.

24. Schneier, B., *Applied Cryptography*, Wiley & Sons, Inc., New York, 1996.

25. Simmons, G., "A Survey of Information Authentication," in *Contemporary Cryptology: The Science of Information Integrity*, Ed. Simmons, G., IEEE Press, New York, 1992, pp. 379-419.

26. Swanson, M. and Guttman, B., *Generally Accepted Principles and Practices for Securing Information Technology Systems*, National Institute of Standards and Technology, U.S. Department of Commerce, Washington, D.C., September, 1996.

27. Vedder, K., "Security Aspects of Mobile Communication," *Computer Security and Industrial Cryptography* (Leuven, Belgium, May 21-23, 1991), Springer-Verlag, New York, pp. 193-210.

28. Yuhas, B., "Toll-Fraud Detection", *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications* (Nassau and Princeton, New Jersey, Oct. 16-20, 1993), Eds. Alspector, J., Goodman, R., and Brown, T., Lawrence Erlbaum Associates, 1993, pp. 239-244.