LONG-TERM SECURITY VULNERABILITIES OF ENCRYPTED DATA

Jerald Hughes, The University Of Texas Pan American, jhughes1@utpa.edu
Joseph N. Roge’, The University Of Texas Pan American, jroge@panam.edu

ABSTRACT

Large amounts of digital data continuously move across the Internet, much of it traveling under a wide variety of encryption and security protections. Such schemes are intended to secure this data against information theft as it traverses the various nodes along the path to its intended destination. While such methods, when properly utilized, may be considered secure in the short-term, many widely employed encryption schemes may not meet consumer expectations over longer periods of time. Historically, such protection schemes tend toward failure; they are broken over time. Given a specific security protocol, increases in computing power and availability, along with advances in hacking methods, tend to produce reduced levels of information content protection. In the long term, even encryption methods considered highly secure today are likely to be far less secure in the future. Therefore, we highlight a little-noted vulnerability that exists for certain classes of information that require longer-term security. Encrypted information, captured and stored today, may be decrypted at some point in the future as more powerful computers and more sophisticated methods become available. This paper provides a first description and analysis of this gap in data security practices. A practical method for auditing and addressing data security vulnerabilities of this type is presented, along with a brief demonstration of its use.

Keywords: security, encryption

INTRODUCTION

As the power of computers and networks has increased, concerns about securing information systems have also grown. As a result, information security has become a topic of central importance in information systems (IS) research and practice (16, 19). Administrators attempting to protect their systems face an increasing number of attack vectors. For example, they must guard against electronic intrusion, vandalism, information theft, eavesdropping, denial of service attacks, and various virus-like infections. Responses have included a wide variety of technological protections, such as anti-virus software, intrusion detectors, hardware and software firewalls, password and biometric access control, and encryption.

Overall, considerable progress has been made on the general question of information security. More specifically, efforts have lead to the development of information security frameworks (14) and best practices standards such as ITSEC\(^1\), ISO 17799\(^2\) and ISO 13335\(^3\). However, it is also true that the problems of information system security are far from resolved (15). IS security may be viewed as an arms race that shows no signs of de-escalation. We should not assume that system defenders have the upper hand (6). Security specialists must continually scan for new threats and devise economically feasible means of blocking them.

Still, encryption remains as one of the strongest weapons in the defender’s arsenal. It can be used to prevent data thieves from accessing the information encoded in the binary digits of various computer formats. This information may be obtained from numerous sources, including but not limited to, portable magnetic media, intra-organizational wireless broadcasts, and electronic data packets being transmitted across the Internet. Public-key encryption can be used to protect such information. It allows for the secure transmission of data across unsecured network channels and is a fundamental technology underlying E-commerce, e-mail, and Internet access control in general. Data transmission protocols such as Secure Sockets Layer (SSL) and Transport Layer Security (TLS) have also employed encryption to secure confidential information that must be sent via TCP/IP across multiple Internet nodes. Private-key encryption, which employs far simpler algorithms than public-key method, is also critical to the function of secure websites, and may be used to protect stored data. However, despite its remarkable algorithmic power, the use of encryption for data security is no magic bullet. Its effectiveness in realizing the security protections for which it is employed depends upon

---

the proper application of the method to the problem. Issues already identified as critical here include precisely what data to encrypt, at what level of the system, and by which encryption methods. Other critical issues are related to key management, scalability and performance, risk assessment, and security policies (17). Overlooking these and other issues, or failing to carry out best practices in the implementation of encryption solutions, makes sensitive data vulnerable to misappropriation. Therefore, realizing the ultimate goal of information security requires us to be fully aware of both the strengths and the weaknesses of the methods employed under the varying conditions presented by business operations in the modern IT environment. It is within this context that we move forward.

The purpose and primary contributions of this paper are to explicate and draw attention to a somewhat obscure threat to the security of data that has ostensibly been locked up by means of powerful encryption schemes and to suggest some possible approaches to mitigating this threat. The remainder of this paper is organized as follows: section 2 will provide a brief description of the technical means by which encryption schemes operate and the means by which they may be broken; section 3, the heart of the paper, provides a full description of the threat model which leads to long-term security vulnerabilities of encrypted data, and identifies the specific categories of data for which this particular threat is most salient; section 4 presents a discussion of the security threat and some possible responses.

**Encryption**

A full explanation of current encryption technologies would fill multiple volumes. However, a succinct overview, excerpted from Brenton (3) is available online. Additionally, *Cryptography Decrypted* (10) provides a complete description of the modern encryption techniques central to SSL and electronic encryption generally. The intent of our current discussion is only to point out the characteristics of a few widely used encryption methods which are relevant to the problem presented here, and thus provide a minimal technical basis for understanding the threat.

**Encryption Methods**

Cryptographic methods widely used to securely encrypt data today are algorithmic: the bits which constitute the data requiring protection are subjected to mathematic transformations which effectively scramble the information, making it unreadable for anyone lacking the keys to unscramble it. The security of information protected by such algorithmic methods is premised on the practical mathematical difficulty of discovering the key that is required to unlock such a mass of encrypted data. The keys are themselves relatively large numbers and discussions of the difficulty in cracking encryption schemes refer to, among other considerations, the length of time required to find a particular number among possible key candidates. Such keys are dozens, hundreds, or even thousands of digits long and longer keys are related to longer search times. Before the 1970s, single-key encryption methods (among others) were used, meaning that the same key could be used to encrypt and decrypt the data via reversible calculations.

The field of cryptography was revolutionized in the 1970s by “public-key” cryptography (4). Developed in practical form (11), it employs ingenious calculations which allow someone wishing to receive a secure message to publicly release a one-way mathematical key. Notably, a one-way key can be used to scramble data; however, it cannot be used to unscramble it. Upon receipt of the scrambled message, the receiver must use a separate but mathematically related number to unscramble the message. This protocol removes the risk that would be incurred if a single key were used to both encrypt and decrypt messages since a single key would have to be transported from one end of the transmission channel to the other before a message could be scrambled. If intercepted, unauthorized use could compromise messages by unlocking them after encryption.

Public-key cryptography allows secure transmissions since the public key can be distributed at large without compromising the security of scrambled incoming messages. This includes messages from persons unknown to the message receiver since the key used to encrypt a message cannot be used to unlock the message. This characteristic is what makes Public-Key Infrastructure (PKI) particularly useful to E-commerce: online businesses and their customers are able to quickly create secure channels of data exchange with each other, even if they have never communicated before.

Internet Web PKI currently uses a hybrid system employing both public-key and single-key methods. Because public-key cryptography is computation-intensive, it is typically used only to establish the
secure channel, rather than to provide encryption for an entire data exchange. A single-key number is generated, then encrypted using public-key encryption. Once this first message is received and decrypted, both ends of the message channel possess a single-key that can be used to scramble data for efficient transmissions, back and forth. The actual process involves many more steps and precautions; however, the basic idea at the heart of all electronic encryption schemes is the mathematical security of the methods employed. Of course, it is not enough to simply scramble the data; it must be scrambled by a key that is difficult for a data thief to guess. The current state of the art can be considered as a needle-in-a-haystack problem: even when the data thief knows that a key exists, and knows both how and where to look for it, the scope of possible candidates is so vast that the chances of actually finding the correct key are miniscule. This is accomplished by making both public and single keys very large numbers.

All other things being equal, an encryption scheme with a long key is more secure than one with a shorter key. Keys of 40 bits, 56, 80, 112, 128, 256, 512, 1024 bits, among others, have been used in various encryption schemes. Brute-force methods of breaking encryption, which look at every possible key candidate number, thus face increasingly longer times to recover the key as bit-length increases. A recent innovation in encryption, elliptic-curve cryptography, provides greater security than previous methods obtained with a key of identical length, but the same principle of “greater key length equals greater security” still applies. Table 1 below illustrates the geometric growth of the key search space as the length increases.

### Table 1 – Typical Encryption Keys

<table>
<thead>
<tr>
<th>Key Length</th>
<th>Used In</th>
<th>Number of Possible Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 bits</td>
<td>early SSL</td>
<td>$1.1 \times 10^6$</td>
</tr>
<tr>
<td>56 bits</td>
<td>DES</td>
<td>$72.1 \times 10^6$</td>
</tr>
<tr>
<td>128 bits</td>
<td>current SSL</td>
<td>$3.4 \times 10^{38}$</td>
</tr>
<tr>
<td>168 bits</td>
<td>Triple-key DES</td>
<td>$3.7 \times 10^{50}$</td>
</tr>
<tr>
<td>256 bits</td>
<td>AES-256</td>
<td>$1.2 \times 10^{77}$</td>
</tr>
</tbody>
</table>

Finally, it is important to note that once encrypted information has been obtained, its level of protection is fixed.

### Breaking Encryption

In actual practice, a data thief need not test every possible key available for a given number of bits. The algorithms used in public-key cryptography are relatively well known. The widely used RSA method requires the cryptanalyst to factor a large number, which means that only prime numbers need be searched in order to find the key. For long key lengths, this is actually not much help, since the space of prime numbers that are possible factors of, say, a 256-bit number, is still very large. Therefore, given appropriate key lengths, the long-term threat from the mathematics side of the question lies with the unpredictable development of innovative approaches. Several approaches more efficient than simply trying all the available primes already exist.5 Furthermore, there is no guarantee that some ingenious mathematician will not develop an even faster method tomorrow. What we can observe as historical fact is that the mathematical means of understanding the kinds of numbers important in encryption have become more powerful over time and that no meaningful indication that such progress is at an end exists.

Another equally important consideration is related to the speed and power of available computer resources. The application of faster, more powerful computing assets tends to reduce the time required to crack encryption schemes. Therefore, judgments concerning the security level provided by specific encryptions methods should be adjusted as available computer resources rise. Moore’s Law suggests that we should expect future processors to be far more powerful than current ones. Parallel computing architectures, by allowing the application of thousands of processors to a task, multiplies system power by orders of magnitude beyond what would otherwise be available. Distributed computing holds out the possibility that not just thousands but perhaps millions of computers could pool their processing power to solve calculation-intensive problems such as cracking encryption schemes. Of particular concern is that these distributed networks might conceivably include “botnets” of internet-connected “zombie” computers, which have been compromised by hackers.

---

5 For example, Dixon’s algorithm, Continued fraction factorization, Quadratic sieve, General number field sieve, Shank’s square forms factorization.
In the long run, fundamentally different computer hardware, such as DNA-based and quantum computers, hold out the promise of extremely large leaps forward in computing power. While such systems are generally not considered to have moved beyond early experimental stages, the potential threat they pose to security has not been lost on encryption experts. Bone and Castro (2) suggested that a quantum computer could break RSA-140, which uses a 140-bit key length, in a matter of seconds.

A history of successful attacks resulting in long-key encryption schemes being cracked demonstrates that such methods tend to have shorter effective lifetimes than expected. When public-key methods first became available, they were typically predicted to require millions of years to break (5). Instead, developers of encryption schemes have seen once-secure keys cracked earlier than expected (9) and, as a result, have responded by moving toward longer and longer keys. Consider the following examples. Digital Versatile Disk (DVD) encryption, based on a 40-bit key, was broken shortly after the introduction of the format. It may be that the far stronger encryption protection for HD-DVD and Blu-Ray disks has been broken as well. The 56-bit DES single-key cipher, adopted in 1978, was cracked with homemade equipment in 1998. Given that it is now considered generally inadequate for most purposes (13), it only achieved an effective lifetime of about twenty years. The SHA-1 hash function, used to authenticate messages, was introduced in 1995. Given a feasible attack for present-day supercomputers, an effective lifetime for the scheme in the range of ten years is likely. In 2002, it was predicted that the AES-256 encryption standard should remain secure until “at least 2010”—at the time of this writing, just three years away. Overall, the implication is that predictions of encryption strength should be modest, and suspicion of claims of mathematical and computer-based insolvability high.

But this caution alone is not enough, even if exercised with great skepticism. The wisdom of choosing an encryption scheme for a particular security context is a function of more than the strength of the mathematical and computer-based methods used. The level of security required to prevent the unauthorized decryption of a given piece of information is dependent on three related dimensions:

1. The power of the encryption method (the typical focus of security policy)
2. The required security period of the specific data being protected
3. The resources available to likely attackers

The following section will describe our first contribution, the precise nature of the gap in security that this analysis implies.

LONG-TERM SECURITY VULNERABILITY

The basic model of data theft posited here is long-term in nature: the attacker collects information today, for example, data packets from encrypted data streams being transmitted via accessible channels; notes the likely encryption methods applied; and then stores the information for future use. When the technical means for cracking those schemes becomes available, possibly some years later, the data is unscrambled and read.

Accessing Encrypted Data

The first step in this process is to acquire the scrambled information to be accessed. Careful information security policies, such as those described in ISO-17799, can aid in preventing theft of data from the storage devices of installed information systems. Still, data snooping from authorized but unethical users and system intrusion by skilled hackers is always a present threat. In fact, there are numerous ways in which encrypted data might be released from organizations. While space considerations don’t allow us to cover all of these ways here, we briefly discuss Internet connections as an example since very large volumes of confidential data traverse it daily. As this encrypted data moves toward its destination it passes through various nodes whose identity is determined ad hoc by routing algorithms. The task for a determined data thief is to identify the best point at which to collect packets, in order to obtain valuable information. The already wide and still increasing adoption of wi-fi IT makes this task even easier. Packet sniffing programs are simple to obtain and implement. Once operational,
they enable the data thief to eavesdrop at the endpoint of a message exchange channel, where one may be more confident of obtaining most or all of the data packets transporting a complete message. Wi-fi systems are notoriously insecure, due in part to the facts that 1), any casual wi-fi-capable laptop may constitute a vulnerable message recipient, and 2), many wi-fi access points are being set up by consumers with virtually no understanding of how to implement effective security (7).

Decrypting Data

As indicated by the analysis above, we believe that most encryption methods, but especially schemes still in wide use today, will inevitably become vulnerable in time. Thus, the question of being able to actually read misappropriated but encrypted data packets is a matter of not “if”, but “when”. Let us suppose that the time required to break an encryption scheme, as history has demonstrated, is a matter of years: a couple of decades, a dozen years, perhaps less. By the time the data thief is able to see the content of the stolen packets, the value of much of it will have declined to zero. The question of interest, then, is the relationship between the time value of the information and the effective lifetime of the encryption employed to protect it. While it is not possible to definitively establish firm values for either of these numbers, one may begin by positing categories of confidential information commonly transported across the Internet with greater, or lesser, required periods of protection.

Secrets regarding processes, formulas, and methods of accomplishing highly valuable or highly dangerous goals are likely to be of highest concern. For example, top secret information must now be encrypted via AES-192 or AES-256 — methods which, as we have already noted, experts believe may be broken as soon as 2010, although some estimates do range as high as 2031 for AES-256 (8). What does this bode for industry, government, or military secrets which may still be highly sensitive in ten years, or twenty? We suggest that the only prudent course in such cases would be to restrict such data from ever traversing open networks.

Personal identity information may be equally long-lived. Someone currently in their 20’s with online accounts and various financial instruments may require secure identity protection for 60, 70, even 80 years — an unthinkable period of effectiveness for any mode of encryption now available. Yet, E-commerce is well established and perhaps more used by this age group. Still, it cannot, or at least would not likely, be halted even on the basis of an apparently near-certain data vulnerability. In this case, technical protections must be combined with prudent policies to insure that personal identity is not compromised in the future due to the limitations of current protection schemes. The move away from the typical “mother’s maiden name” password confirmation step in many online systems, towards questions custom-made by the customers themselves, may be a first step in the right direction.

Medical and legal confidentiality is a matter of law and thus poses a difficult problem for electronic records since the period of confidentiality extends to at least the lifetime of the client. As our health system makes the transition to digital platforms, insuring long-term confidentiality will require the strongest possible encryption methods available. Still, proposed technological means of remote diagnosis and even treatment will pose significant risks to patient information. These risks include data packet capture and, thus, vulnerability to future decryption by data thieves.

A shorter protection period may be required for other types of information. These would include, but not be limited to, information that is simpler to update than one’s identity. A specific credit card number, for example, may have a lifetime of a few years before being replaced with a new number. Here, altering such periods is only a matter of company policy and the procedural steps involved, such as mailing out new cards. In this case, a 128-bit system expected to last until 2020 is probably sufficient to protect a credit card issued in 2007 with a planned lifetime of no more than five years—unless the attacker has the capacity to bring supercomputer-level processing resources to bear on the problem—or unless a significant breakthrough in mathematics suddenly and dramatically reduces the key-search problem. Unfortunately, such occurrences are always a possibility. On the other hand, experiments are being carried out in the consumer market on the issuance of single-use credit card numbers intended specifically for Internet use. Since the lifetime of the number is a matter of mere minutes, or at most a few days, confidence in typical encryption protection can be relatively high. However, such revisions of business methods cannot constitute a total solution, since the issuance of single-use numbers themselves requires authentication of the identity of the person using the number, which brings us back to the very long-lived personal identity problem.

**DISCUSSION AND RECOMMENDATIONS**
At the heart of the problem identified here is the time dimension of the protection offered by encryption. It is definitely the case that currently available encryption methods with long keys can protect important information in transit on public networks from immediate theft. By way of example, this is one use of Virtual Private Networks (VPN). Used as such, they are the response of many corporations to the potential threats to security posed by open Internet transmission channels and for many types of content this is no doubt a best practice for remote communications. However, for other types of data a VPN may actually increase one’s vulnerability to exploitation. Consider for example the case in which the vaunted security of a VPN tempts users into transmitting information whose security requirements demand a long period of security protection. In fact, the very existence of a VPN tunnel may draw attention to itself as a channel that is likely to be carrying valuable information due to strength of the protections layered around it. Attackers hoping to exploit the particular security gap discussed in this paper might very well deliberately attempt to capture VPN packets specifically. For example, their reasoning might suggest that the information inside VPN packets has a greater chance of yielding high value later on, even more so than would the information of a casual E-commerce consumer.

Our research leads us to conclude that there are two basic ways to address this potential security gap: 1) by shortening the required lifetime of the information, and 2) by lengthening the effective protection time of the encryption scheme used. Of course, both methods could also be combined as appropriate. The time dimension of security has already received some attention in other respects. The National Institute of Standards and Technology in Barker et al (1) discusses the notion of a “cryptoperiod – the time span during which a specific key is authorized for use.” This work also includes a recommendation that a shorter period generally increases security, given that key distribution methods are secured. To our knowledge, no detailed guidance currently exists regarding selection or retirement of encryption schemes, aside from ballpark predictions, typically represented as some number of years, of how long a given encryption method may remain robust before the practical difficulties of finding keys are overcome.

Our second contribution is therefore to propose a simple metric for making encryption decisions. We put forward the concept of an Encryption Security Ratio (ESR), where ESR is equal to the Effective Security Lifetime (ESL) of the encryption scheme, divided by the Required Protection Time (RPT) of the actual data in question:

$$ESR = \frac{ESL}{RPT}$$

Furthermore, we suggest that this ratio should be considerably in excess of a value of 1. Suppose, for example, that one wishes to use AES-256 to protect customer account numbers which are automatically changed every 4 years. Noting that the NIST predicts that AES-256 will be cracked by 2031, 24 years from the time of this writing, under our rule of thumb, ESR = 24/4 = 6. Therefore, the specific application of this particular encryption scheme to this particular datum could be deemed reasonably secure. On the other hand using the same scheme (AES-256 is approved today for all uses, including Top Secret documents of the government) to encrypt a customer’s Social Security number would, for a customer of age 34, yield a value for RPT of 48 (given average life expectancy of 82). In this case, our resultant ESR indicator would be ESL/RPT = 24/48 = 0.5, pointing to the conclusion that this particular encryption method is insufficient today for this particular datum. Notice that the ESR value for a given data/encryption combination will decrease regularly over time; by 2011, for example, the Effective Security Lifetime of AES-256 will have dropped to 20, yielding an ESR of 5 for the customer account number example above.

Of course, neither ESL nor RPT can be determined as exclusively correct values to be authoritatively applicable to cases. One can easily imagine many scenarios with greater detail in their descriptions. These might suggest larger or smaller values for the required arguments, with a resulting difference in the conclusions drawn from the analysis. What we are proposing here is a quick look heuristic, a rule of thumb, useful because it draws attention to the time dimension of security, and requires consideration of that dimension with respect to the meaning of the data being encrypted, not just to the algorithmic difficulty of the encryption method.

ESR would thus be one among many tools available to persons designing a system that requires some level of security, or perhaps, for auditing such a system for compliance. A systematic application of ESR could involve the following steps:

1. List the types of data requiring protection
2. Determine the RPT of each type
3. List the encryption methods considered feasible for the system
4. Determine the ESL of each type
5. Create a table displaying the resultant ESR matrix of possible choices
6. Discard those choices for which ESR is “too low”, where this is a judgment call based on the likelihood and resources of the expected threat, the losses which would be incurred should the information be compromised, and the level of confidence in the values chosen for ESL and RPT.

7. Choose from among the remaining alternatives based on the desired security level and other considerations such as cost, computational speed, etc.

Finally, we acknowledge that judgments related to encryption security may be considered subjective. Consider, for example, the following questions. How secure is “secure enough”? How strong is “strong encryption”? Just as civil engineers build a considerable margin for error into structures such as bridges and towers, we suggest that information security professionals should build a considerable margin for error into their judgments of the adequacy of encryption methods in order to protect, at least partially, against the dangers of the unknown.

REFERENCES