Chapter 1

Reinforcement

Problem R-1.12

Compare and contrast symmetric encryption with public-key encryption, including the strengths and weaknesses of each.

Solution Scalability: with public-key encryption, multiple users can send encrypted messages to Alice using her public key and these messages can be decrypted only by Alice; thus, a linear number of public-private key pairs need to be established, distributed and protected to allow pairwise confidential communication between any two users; instead, symmetric encryption requires a quadratic number of secret keys. Efficiency: existing symmetric encryption methods are much faster and use much shorter keys than existing public-key encryption methods. Usability: symmetric-key encryption is easier to understand by an non-expert than public-key encryption.

Problem R-1.14

Suppose the author of an online banking software system has programmed in a secret feature so that program emails him the account information for any account whose balance has just gone over $10,000. What kind of attack is this and what are some of its risks?

Solution This is a Trojan horse, since it has a hidden malicious action that goes with a useful service.

Problem R-1.16

Give an example of the false sense of security that can come from using the "security by obscurity" approach.

Solution There are many examples. One possibility would be to use a weak encryption algorithm, like the Caesar cipher and try to keep secret the type of algorithm that you are using, in addition to keeping the key secret. The problem with this approach is that if someone guesses you are using such an algorithm or is able to reverse engineering your software, then they will discover your algorithm. From there it is a simple matter to break your weak encryption scheme.

Problem R-1.17

The English language has an information content of about 1.25 bits per character. Thus, when using the standard 8-bit ASCII encoding, about 6.75 bits per character are redundant. Compute the probability that a random array of t bytes corresponds to English text.

Solution Since each byte has 8 bits, the total number of t-byte arrays is $T = \left(2^8\right)^t = 2^{8t}$. Given that the information content of English text is 1.25 bits per character, the number of t-byte arrays corresponding to English text is $E = \left(2^{1.25}\right)^t = 2^{1.25t}$. Thus, the probability that a random array of t bytes corresponds to English text is given by $E/T = 2^{-6.75t}$.

Problem R-1.18

Suppose that a symmetric cryptosystem with 32-bit key length is used to encrypt messages written in English and encoded in ASCII. Given that keys are short, an attacker is using a brute-force exhaustive search method to decrypt a ciphertext of t bytes. Estimate the probability of uniquely recovering the plaintext corresponding to the ciphertext for the following values of t: 8, 64, and 512.

Solution Brute-force decryption generates 2^{32} candidate plaintexts, one for each possible key value. Each plaintext has probability $2^{-6.75t}$ of being English text. Thus, the attack is expected to produce $2^{32-6.75t}$ candidate English plaintexts. Since this number is less than one for the given values of t, the attack is expected to always recover the plaintext.

Problem R-1.19

Suppose you could use all 128 characters in the ASCII character set in a password. What is the number of 8-character passwords that could be constructed from such a character set? How long, on average, would it take an attacker to guess such a password if he could test a password every nanosecond?

Solution There are 128^8 possible passwords with 8 ASCII characters. Guessing a password will take on average $\frac{1}{2}128^8 10^{-9}$ seconds. This is $9,223,372,037$ seconds or about 417 days.

Creativity

Problem C-1.2

Describe an instance of a file that contains evidence of its own integrity and authenticity.

Solution Take a file and concatenate a digital signature on that file from the owner of that file or from another trusted authority. Don't forget to include the digital certificate of the signer.

Problem C-1.3

Suppose an Internet service provider (ISP) has a voice over IP (VoIP) telephone system that it manages and sells. Suppose further that this ISP is deliberately dropping 25% of the packets used in its competitors VoIP system when those packets are going through this ISP's routers. Describe how a user could discover that his ISP is doing this.

Solution Suppose the user bought both VoIP solutions. He could then do a set of simple end-to-end performance tests to see if one had degraded throughput with respect to the other in terms of packet delivery. If, say, in 10 tests, one is 25% worse than the other, then it is highly likely that this is due to a deliberate packet dropping strategy on the part of the ISP.

Problem C-1.5

Suppose that you are a computer virus writer; hence, you know that you need to store a copy of the code for your virus inside the virus itself. Moreover, suppose you know that a security administrator is also aware of this fact and will be using it to detect the presence of your virus in operating systems files, as described in the previous problem. Explain how you can hide the embedded copy of your virus so that it is difficult for the security administrator to find it.

Solution If the embedded virus code is stored in encrypted form and only decrypted just before it is replicated in another operating system file, then it would be difficult to see the repeated pattern when the virus is at rest inside the infected file.

Problem C-1.9

Benny is a thief who tried to break into an Automated Teller Machine (ATM) using a screwdriver, but was only able to break five different keys on the numeric keypad and jam the card reader, at which point he heard Alice coming, so he hid. Alice walked up, put in her ATM card, successfully entered her 4-digit PIN, and took some cash. But she was not able to get her card back, so she drove off to find help. Benny then went back to the ATM, and started entering numbers to try to discover Alice's PIN and steal money from her account. What is the worst-case number of PINs that Benny has to enter before correctly discovering Alice's PIN?

Solution Since Benny broke 5 different keys and Alice was still able to enter her PIN, it must only use the 5 remaining keys. So the total number of possible keys is now $5^4 = 625$. In the worst case, Benny will have to enter 625 before he enters the correct one.

Problem C-1.10

As soon as Barack took office, he decided to embrace modern technology by communicating with cabinet members over the Internet using a device that supports cryptographic protocols. In a first attempt, Barack exchanges with Tim brief text messages, encrypted with public-key cryptography, to decide the exact amounts of bailout money to give to the largest 10 banks in the country. Let p_B and p_T be the public keys of Barack and Tim, respectively. A message m sent by Barack to Tim is transmitted as $E_{p_T}(m)$ and the reply r from Tim to Barack is transmitted as $E_{p_B}(r)$. The attacker can eavesdrop the communication and knows the following information:

- Public keys p_B and p_T and the encryption algorithm, such that there is exactly one ciphertext for each plaintext.
- The total amount of bailout money authorized by congress is \$900B
- The names of the largest 10 banks
- The amount each bank will get is a multiple of \$1B
- Messages and replies are terse exchanges of the following form:

  Barack: How much to Citibank?
  Tim: \$144B.

Barack: How much to Bank of America?
Tim: $201B.
...

Describe how the attacker can learn the bailout amount for each bank even if he cannot derive the private keys.

Solution The attacker performs a dictionary attack. Since the message format is fixed and there are 10 possible banks and 900 possible bailout amounts, the attacker encrypts the 10 candidate messages from Barack (one for each bank) using public key p_B, and the 900 candidate responses from Tim (one for each bailout amount), using public key p_T. The attacker then matches the ciphertexts exchanged by Barack and Tim with the precomputed ones and determines the corresponding plaintexts. Note that the attacker does not need access to the private keys used by Barack and Tim.

Problem C-1.11

As a result of the above attack, Barack decides to modify the protocol of Exercise C-1.10 for exchanging messages. Describe two simple modifications of the protocol that are not subject to the above attack. The first one should use random numbers and the second one should use symmetric encryption.

Solution In the first case, Barack can add a random value with b bits to his message, which increases the number of possible messages by a factor of 2^b. In the second case, Barack can first encrypt a (random) key K for a symmetric encryption scheme, and then send the encrypted version of K along with an encryption of his actual message using key K and the symmetric cryptosystem .

Problem C-1.12

Barack often sends funny jokes to Hillary. He does not care about confidentiality of these messages but wants to get credit for the jokes and prevent Bill from claiming authorship of or modifying them. How can this be achieved using public-key cryptography?

Solution Barack digitally signs his jokes and sends each joke together with its signature.

Problem C-1.13

As public-key cryptography is computationally intensive and drains the battery of Barack's device, he comes up with an alternative approach. First, he shares a secret key k with Hillary but not with Bill. Next, together with a joke x, he sends over the value $d = h(k||x)$, where h is a cryptographic hash function. Does value d provide assurance to Hillary that Barack is the author of x and that x was not modified by Bill? Justify your answer.

Solution Value d is a message authentication code (MAC), which gives Hillary assurance of the authorship and integrity of Barack's jokes. The reason is that a cryptographic hash function is one-way, Bill cannot recover the key k from value d, Thus, Hillary knows that only Barack could have computed value d from joke x. Also, if Bill replaces joke x with a joke of his, x', it would infeasible for Bill to compute the MAC value corresponding to x'.