Chapter 7. Cryptography

7.1 Definition: **Cryptography** is the study of secret codes. When we convert a message from a normal language, say English, to a secret code, we say that we **encrypt** (or **encipher**) the message, and the coded word is called the **ciphertext**. When we convert the ciphertext back into normal language, we say that we **decipher** (or **decrypt**) the ciphertext to obtain the original message.

7.2 Example: One of the simplest encryption methods is a **Caesar cipher**. Suppose Alice wants to send a secret message to Bob using a Caesar cipher. Alice and Bob agree in advance on a number n between 1 and 25. Alice encrypts the message by replacing each letter in the message by the letter which follows it by n positions (modulo 26) in the English alphabet. For example, if $n = 4$ then the letter P would be replaced by the letter T (which follows P by 4 positions), and the message PONY would be replaced by the ciphertext TSRB. Bob can easily decrypt the ciphertext by replacing each letter by the letter which precedes it by n positions.

7.3 Example: A slightly more secure encryption method is a **substitution cipher**. Suppose that Alice wants to send a secret message to Bob using a substitution cipher. Alice and Bob agree in advance on a permutation p of the letters of the English alphabet. Alice enciphers the message by replacing each letter by the letter which corresponds to it under the permutation p. For example, if the permutation p is given as follows

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
V G S C F U Q L A P I D X N W T H Y O J K Z B E R M

then the letter H would be replaced by the letter L and the message HORSE would be replaced by the ciphertext LWYOF.

7.4 Definition: A far more secure encryption system, which is commonly used by modern computers, is the **RSA scheme**. The letters R, S and A stand for Rivest, Shamir and Adleman. who first described this encryption system. The RSA scheme is a **public key** encryption system, which means that when a person, say Alice, wishes to receive a secret message, she makes her encryption rules publicly known so that anyone can encipher a message and send it to Alice and yet, although everyone knows the encryption rules, only Alice knows the decryption rules and can decipher the ciphertext.

Suppose that Alice wishes to receive a secret message using the RSA scheme. Alice chooses two large prime numbers p and q (in practice, p and q would have over 100 decimal digits) and calculates $n = pq$ and $\varphi = \varphi(n) = (p-1)(q-1)$. Then Alice chooses a positive integer $e < \varphi$ with $\gcd(e, \varphi) = 1$ and calculates $d = e^{-1} \mod \varphi$. The number e is called the **encryption key** and the number d is called the **decryption key**. Then Alice makes the numbers n and e publicly known. Suppose that Bob wishes to send a message to Alice. Bob converts his message to a positive integer m with $m < n$ (if his message is too long then he breaks it into shorter messages). Bob calculates the ciphertext $c = m^e \mod n$ which he sends to Alice. Note that since $ed = 1 \mod \varphi$, we have $c^d = (m^e)^d = m^{ed} = m^1 = m \mod n$ by the Euler Fermat Theorem, and so Alice can recover the original message m by calculating $m = c^d \mod n$.

7.5 Note: Alice can save some time if, instead of calculating $\varphi = (p-1)(q-1)$ and $d = e^{-1} \bmod \varphi$, she instead calculates $\psi = \operatorname{lcm}(p-1, q-1)$ and $d = e^{-1} \bmod \psi$. Verify that when $c = m^e \bmod n$ we have $c^d = (c^e)^d = c^{ed} = c^1 = m \bmod n$.

7.6 Note: The reason that the RSA scheme is practical and secure is that there do exist efficient (polynomial time) algorithms which can be used to find p, q, n, φ, e and d and to calculate $c = m^e \mod n$ and $m = c^d \mod n$, but there is no known efficient algorithm which can be used to determine m from n, e and c. In particular, there do exist efficient algorithms which can be used to determine whether a given positive integer n is prime, but there is no known efficient algorithm which can determine a prime factor of n in the case that n is composite.

There do, of course, exist inefficient algorithms which can determine a prime factor of n. For example, we can use the Sieve of Eratosthenes to list all primes p with $1 < p \leq \sqrt{n}$ and then test each such prime p to determine whether it is a factor of n. But when the prime factors of n are over a hundred digits long, this algorithm is too slow (if a computer could list 10^{10} prime numbers each second then it would take about 10^{80} years to list all the prime numbers p with $p < 10^{100}$).

7.7 Example: The calculation of $d = e^{-1} \mod \varphi$ can be performed using the Euclidean Algorithm, which is efficient.

7.8 Example: When n, e and m are all large, we can calculate $c = m^e \mod n$ efficiently as follows. Express e in base 2, say $e = \sum_{i=1}^{\ell} 2^{k_i}$ with $0 \leq k_1 < k_2 < k_3 < \cdots$, calculate the residues $m^1, m^2, m^4, m^8, \cdots, m^{2^{k_\ell}} \mod n$, then calculate $c = m^e = \prod_{i=1}^{\ell} m^{2^{k_i}} \mod n$. This algorithm is known as the **Square and Multiply Algorithm**.

7.9 Example: Alice wishes to receive a message. She chooses $p = 13$ and $q = 17$ and calculates $n = pq = 221$. She also chooses $e = 35$ and makes the numbers n and e public. Bob wishes to secretly send Alice the letter T. Bob converts the letter T to the number $m = 20$ (since T is the 20th letter in the English alphabet) and sends the cyphertext $c = m^e \bmod n$. As an exercise, calculate $c = m^e \bmod n$ and calculate $\psi = \operatorname{lcm}(p-1, q-1)$ and $d = e^{-1} \bmod \psi$, then directly calculate $c^d \bmod n$ to verify that $c^d = m \bmod n$.

7.10 Definition: Let us describe a simple test for primality which is called the **Fermat Primality Test**. Suppose that we are given an integer $n > 2$. Choose an integer a with $1 < a < n$. By Fermat's Little Theorem, if n is prime then we must have $\gcd(a, n) = 1$ and $a^{n-1} = 1 \mod n$, so we use the Square and Multiply Algorithm to calculate $a^{n-1} \mod n$. If $a^{n-1} \neq 1 \bmod n$ then we can conclude that n is composite while if $a^{n-1} = 1 \mod n$ then we can conclude that n is probably prime.

7.11 Example: Unfortunately, given $n, a \in \mathbf{Z}^+$ with $1 < a < n$, if $a^{n-1} = 1 \bmod n$ then it does not necessarily follow that n is prime. For example, verify that $2^{340} = 1 \bmod 341$ but $341 = 11 \cdot 31$. As another example, verify that $3^{90} = 1 \bmod 91$ but $91 = 7 \cdot 13$.

7.12 Definition: Let $n, a \in \mathbf{Z}^+$ with n composite and $1 < a < n$. If $a^{n-1} \neq 1 \bmod n$ then we say that a is a **Fermat witness** for the compositeness of n. If $a^{n-1} = 1 \bmod n$ then we say that a is a **Fermat liar** and that n is a **Fermat pseudoprime** to base a.

7.13 Note: We can improve the reliability of the above test simply by repeating it. Given $n \in \mathbf{Z}^+$, we choose a finite set S of integers a with $1 < a < n$. For each $a \in S$ we calculate $a^{n-1} \bmod n$. If we find some $a \in S$ such that $a^{n-1} \neq 1 \bmod n$ then we know that n is composite. If we find that for every $a \in S$ we have $a^{n-1} = 1 \bmod n$ then we can conclude that n is probably prime.

7.14 Example: Unfortunately, if if $a^{n-1} = 1 \bmod n$ for every a with $1 < a < n$ and $\gcd(a, n) = 1$ then it does not necessarily follow that n is prime. For example, show that when $n = 3 \cdot 11 \cdot 17 = 561$ we have $a^{n-1} = 1 \bmod n$ for all $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$.

7.15 Definition: For $n \in \mathbf{Z}^+$ we say that n is a **Carmichael number** when n is composite and $a^{n-1} = 1 \bmod n$ for every $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$.

7.16 Theorem: *(Carmichael Numbers) Let $n \in \mathbf{Z}^+$.*
(1) If $n = p_1 p_2 \cdots p_l$ where $\ell \geq 2$ and the p_i are distinct primes which satisfy $(p_i - 1)\big|(n-1)$ for all indices i, then n is a Carmichael number .
(2) If $n = p_1 p_2 \cdots p_l$ where $\ell \geq 2$ and the p_i are distinct primes which satisfy $(p_i - 1)\big|(n-1)$ for all indices i (so that n is a Carmichael number, by Part (1)) then n is odd and $\ell \geq 3$.

Proof: Suppose that $n = p_1 p_2 \cdots p_l$ where the p_i are distinct primes with $(p_i - 1)\big|(n - 1)$. Let $a \in \mathbf{Z}^+$ with $\gcd(a, n) = 1$. Fix an index i. Since $\gcd(a, n) = 1$ we have $p_i \not| \, a$ and so $a^{p_i - 1} = 1 \mod p_i$ by Fermat's Little Theorem. Since $a^{p_i - 1} = 1 \mod p_i$ and $(p_i - 1)\big|(n-1)$, we also have $a^{n-1} = 1 \mod p_i$. Since $a^{n-1} = 1 \mod p_i$ for every index i, it follows from the Chinese Remainder Theorem that $a^{n-1} = 1 \mod n$. Thus n is a Carmichael number, so we have proven Part (1).

Let us prove Part (2). Since $l \geq 2$, at least one of the primes p_i is odd, say p_k is odd. Since $p_k - 1$ is even and $(p_k - 1)\big|(n - 1)$, it follows that $(n - 1)$ is even and so n is odd.

Suppose, for a contradiction, that n is a Carmichael number of the form $n = pq$ where p and q are primes with $p < q$ and we have $(p - 1)\big|(n - 1)$ and $(q - 1)\big|(n - 1)$. Note that $n - 1 = pq - 1 = p(q - 1) + (p - 1)$. Since $(q - 1)\big|(n - 1)$ we have $(q - 1)\big|(n - 1) - p(q - 1)$, that is $(p - 1)\big|(p - 1)$. But this implies that $q \leq p$ giving the desired contradiction.

7.17 Exercise: Find distinct primes p and q such that $145\,p$ and $145\,q$ are both Carmichael numbers.

7.18 Theorem: *(The Miller-Rabin Test Theorem) Let n be an odd prime number and let $a \in \mathbf{Z}$ with $\gcd(a, n) = 1$. Write $n - 1 = 2^s d$ where $s, d \in \mathbf{Z}^+$ with d odd. Then*

$$\textit{either} \;\; a^d = 1 \bmod n \;\; \textit{or} \;\; a^{2^r d} = -1 \textit{ for some } 0 \leq r < n.$$

Proof: First we remark that since n is prime, $\mathbf{Z}_n$ is a field, so for all $x \in \mathbf{Z}_n$ we have

$$x^2 = 1 \Longleftrightarrow x^2 - 1 = 0 \Longleftrightarrow (x - 1)(x + 1) = 0 \Longleftrightarrow x = \pm 1\,.$$

By Fermat's Little Theorem, we have $a^{n-1} = 1 \bmod n$, that is $a^{2^s d} = 1 \bmod n$. By the above remark $\big($using $x = a^{2^{s-1} d}\big)$ it follows that $a^{2^{s-1} d} = \pm 1 \bmod n$. If $a^{2^{s-1} d} \neq -1$ then $a^{2^{s-1} d} = 1$ so, by the above remark again, it follows that $a^{2^{s-2}} d = \pm 1$. Similarly, if $a^{2^{s-1} d} \neq -1$ and $a^{2^{s-2} d} \neq -1$ then $a^{2^{s-2} d} = 1$ and hence $a^{2^{s-3}} d = \pm 1$ and so on. Repeating the above argument we find that if $a^{2^{s-1} d} \neq -1$, $a^{2^{s-2} d} \neq -1$, $\cdots$, $a^{2^2 d} \neq -1$ and $a^{2d} \neq -1$ then $a^{2d} = 1$ and hence $a^d = \pm 1$.

7.19 Definition: Using the above theorem we obtain the following test for primality, called the **Miller-Rabin Primality Test**. Given an odd integer $n \in \mathbf{Z}^+$ write $n-1 = 2^s d$ and choose an integer a with $1 < a < n$. By the above theorem, if $a^d \neq 1$ mod n and $a^{2^r d} \neq -1$ mod n for all $0 \leq r < n$ then we can conclude that n is composite. If, on the other hand, we find that either $a^d = 1$ mod n or $a^{2^r d} = -1$ mod n for some $0 \leq r < s$ then we can conclude that n is probably prime.

7.20 Example: Unfortunately, given $n = 1 + 2^s d$ where $s, d \in \mathbf{Z}^+$ with d odd, and given $a \in \mathbf{Z}$ with $1 < a < n$, even if it is true that either $a^d = 1$ mod n or $a^{2^r d} = -1$ for some $0 \leq r < s$, it does not necessarily follow that n is prime. For example, verify that when $n = 221 = 13 \cdot 17$ and $a = 174$ we have $s = 2$ and $d = 55$ and $a^{2d} = -1$ mod n.

7.21 Definition: Let $n, a \in \mathbf{Z}^+$ where n is an odd composite number and $1 < a < n$. Write $n - 1 = 2^s d$ where $s, d \in \mathbf{Z}^+$ with d odd. If $a^d \neq 1$ and $a^{2^r d} \neq -1$ for all $0 \leq r < s$ then we say that a is a **Miller-Rabin witness** (or a **strong witness**) for the compositeness of n. If either $a^d = 1$ or $a^{2^r d} = -1$ for some $0 \leq r < n$ then we say that a is a **Rabin-Miller liar** (or a **strong liar**) and that n is a **Rabin-Miller pseudoprime** (or a **strong pseudoprime**).

7.22 Note: As with the Fermat primality test, we can make the Miller-Rabin test more reliable simply by repeating it. Given an odd positive integer n, write $n - 1 = 2^s d$ with $s, d \in \mathbf{Z}^+$ and d odd. Choose a finite set S of integers a with $1 < a < n$. For each $a \in S$, calculate $a^{2^r d}$ mod n for $0 \leq r < s$. If we find some $a \in S$ for which $a^d \neq 1$ mod n and $a^{2^r d} \neq -1$ for all $0 \leq r < s$ then we know that n is composite. If, on the other hand, we find that for every $a \in S$, either $a^d = 1$ mod n or $a^{2^r d} = -1$ mod n for some $0 \leq r < s$ then we can conclude that n is probably prime.

7.23 Note: Recall that repeating the Fermat primality test does not make the test become completely reliable because of the existence of Carmichael numbers. The situation is different with the Miller-Rabin primality test. It has been proven that for every composite positive integer n, at least $\frac{3}{4}$ of the numbers a with $1 < a < n$ are strong witnesses for the compositeness of n. It follows that, given an odd composite number n, if we choose m integers a with $1 < a < n$, the probability that none of the numbers a is a strong witness is at most $\frac{1}{4^m}$.