Listen

Number Theory and the One-Time Pad

Number Theory

Modular arithmetic underlies many ciphers and is an extremely important concept in modern cryptography. It is used throughout the course. Modular arithmetic is used to study how remainders behave after a number is divided by a fixed base n. The Euclidean algorithm states that, for any integer d, there is a unique remainder r, an integer between 0 and $n - 1$, such that $d = m^*n + r$ for some integer m. This allows association to any number d, its remainder r after division by n. This is the modulus of d by n, and it is written as $d = r \pmod{n}$. The particularly interesting fact about integers mod n is that addition, subtraction, and multiplication can still be performed with them. Division does not always work. In certain cases, however, particularly when $n = p$ is a prime number, division always works too. Letters can be considered as the numbers 0, . . . , 25, and this collection can be identified as integers mod 26. Then the Caesar and affine ciphers can be written as equations mod 26, and methods for attacking these algorithms consequently become more obvious.

Prime number and factoring composite numbers is another important number theory concept for cryptography. This module addresses the mathematics underlying several public key algorithms that we'll later encounter. Elementary properties of integers are reviewed. Every integer N can be written uniquely as a product of prime numbers $p_1, p_2, \ldots, p_r$ with integer exponents $e_1, \ldots, e_r$: $N = p_1^{e1} \ldots p_r^{er}$. This property is called unique factorization. This is the starting point of elementary number theory on which most of modern cryptography is based. Properties of primes and the Euclidean algorithm for finding the greatest common divisor of two integers are also fundamental for cryptography.

Last, Fermat's little theorem is also introduced. Fermat's theorem results from number theory that works in a more general context; however, it is particularly useful for a new algorithm to be studied in a future module. Fermat's little theorem is a property describing exponentiation modulo a prime p. In particular, it allows efficient computations for future cryptosystems based on factorizations of integers.

One-Time Pads

The first concept analyzed in this module is number representations. Usually, a number is represented in base 10. The number of a base refers to the number of allowable digits; in base 10, there are 10 digits used (0, 1, 2, . . . , 9). Computers use base 2. All numbers in this case are represented in terms of the two digits 0 and 1. Other common bases include base 16, which uses the digits 0, 1, 2, . . . , 9, A, B, C, D, E, F, or base 26, which represents numbers using the 26 "digits" of A, B, . . . , Z. These systems are called number representations because every number can be uniquely represented in each system. In fact, there is a one-to-one function that, for a given number in base 10, associates the unique representation in base 2 and vice versa. One can quickly conclude that the more digits a

representation has, the fewer the digits needed to represent a number. For example, the number 45 in base 10 is represented by two digits: 4 and 5. In base 2, the same number is represented by 101101; that is, six digits are needed. To compute representations base *b*, one must analyze remainders under division by powers of *b*. This is most easily seen in base 10:

415 = 400 + 10 + 5 = 4*100 + 1*10 + 5*1 = 4*102 + 1*101 + 5*100

Now that binary representation is understood, the concept of a one-time pad can be studied. The one-time pad encryption method involves representing a message as a sequence of 0s and 1s and using a random key of the same length as the message. The encryption process involves adding the key to the message using an exclusive or operation. Decryption uses the same key by subtracting the key from the ciphertext. This encryption is unbreakable for a ciphertext-only attack, as it gives no information about the plaintext except its length.

However, generating a truly random key can be challenging, and once used, the key should not be reused. This comes at a disadvantage in that the scheme requires a long key, making it costly to use. Various methods that provide an approximation to a one-time pad are used in practice.

Theoretically, one-time pads provide perfect secrecy. Perfect secrecy means that the ciphertext provides no information about the plaintext, except for its length, and that even with unlimited computational resources, it is impossible to decipher the original message without the key. Perfect secrecy ensures that no matter how much ciphertext is collected or how advanced the decryption methods are, the original message remains completely secure and undecipherable without the specific one-time pad key. However, as noted above, the key must be truly random and as long as the message, making it expensive and challenging to implement in practice.