A very brief introduction to Cryptography

Charles Cusack

Hope College

October 4, 2017

Elements of a Cryptosystem

- $\mathcal{P}$ = the set of possible plaintexts
- $\mathcal{C}$ = the set of possible ciphertexts
- $\mathcal{K}$ = the keyspace (the set of possible keys)
- For each $K \in \mathcal{K}$, there is an encryption function
  $$e_K : \mathcal{P} \to \mathcal{C}$$
  and a decryption function
  $$d_K : \mathcal{C} \to \mathcal{P}$$
  satisfying
  $$d_K(e_K(x)) = x \text{ for all } x \in \mathcal{P}.$$
- Note that, for all $K \in \mathcal{K}$, e_K must be an injective function, i.e.
  $$x_1 \neq x_2 \;\Rightarrow\; e_K(x_1) \neq e_K(x_2).$$

Modular Arithmetic

- Let m be a positive integer. Given integers a and b, we say $a \equiv b \pmod{m}$ if $b - a$ is divisible by m.
- Every integer a is equivalent $\pmod{m}$ to precisely one element r of $\{0, 1, \ldots, m-1\}$, and we refer to this element r as $a \bmod m$.
- We set $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$, and we note that addition and multiplication can be defined as operations on $\mathbb{Z}_m$.
- For example, working in $\mathbb{Z}_{26}$ we have

  $$14 + 20 = 34 \equiv 8 \pmod{26},$$

  and

  $$5 \cdot 7 = 35 \equiv 9 \pmod{26}.$$

  Therefore, in $\mathbb{Z}_{26}$, $14 + 20 = 8$ and $5 \cdot 7 = 9$.
- Under these operations, $\mathbb{Z}_m$ satisfies the properties required to be an Abelian group (and, in fact, a commutative ring).

Shift Cipher

- $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}$ (or $\mathbb{Z}_m$)
- For all $K \in \mathbb{Z}_{26}$ and $x \in \mathcal{P}$, define

$$\begin{aligned} e_K(x) &= x + K \mod 26 \\ d_K(x) &= x - K \mod 26 \end{aligned}$$

- We use the following text-to-numeric conversion:

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

- Suppose $K = 7$. The plaintext 'ihavefoundthekey' will yield

x	8	7	0	21	4	5	14	20	13	3	19	7	4	10	4	24
$e_K(x)$	15	14	7	2	11	12	21	1	20	10	0	14	11	17	11	5

  which yields the ciphertext 'POHCLMVBUKAOLRLF'.

- To decrypt the message, we subtract 7 from each character (mod 26, of course). Alternatively, you can add 19 to each character (again, mod 26).

Cryptanalysis

- Cryptanalysis attempts to answer the question, “If Oscar is allowed to see the ciphertext, can he use this information to discover the plaintext?”
- For the shift cipher on $\mathbb{Z}_{26}$, there are only 26 possible keys, and therefore the decryption key can be found quickly using an exhaustive key search.
- The number of possible keys (i.e. the cardinality of the keyspace $\mathcal{K}$) becomes an important factor in constructing secure cryptosystems.

Substitution Cipher

- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$
- $\mathcal{K} =$ the set of all permutations of $\{0, 1, \ldots, 25\}$.
- $K \in \mathcal{K}$ is chosen at random and used as the encryption function.
- For example, one possible key is

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

- Given this key, the string “thisistheinput” is encrypted as “MQBHBHMQNBSUCM”

Security of Substitution Cipher

- There are $26! \approx 4.03 \times 10^{26}$ possible keys, making an exhaustive key search infeasible.
- Unfortunately, simple techniques that do not need to exhaust the keyspace can be employed to decrypt this cipher so it is *not* secure.
- For instance, in English, **e** is the most common character, followed by **t, a, o, i, n, s, h, r**. Thus, if **Q** is the most common character in a ciphertext, it is probably the letter **e**.
- Further, the most common digrams in English are **TH, HE, IN**, etc. and the most common trigrams are **THE, ING, AND**, etc.
- This is precisely why the substitution cipher used in many of the puzzles in newspapers.

Exercise: Decrypting a Substitution Cipher

- Attempt to decode the following message that was encoded using a substitution cipher. When you have decrypted it, tell me what the plaintext message was. **onndtcn nk udikud nbd lkxxkawzp tdggopd nbon aog dzikudu jgwzp o gjqgnwnjnwkz iwcbdh abdz mkj bovd udihmcndu wn ndxx td abon nbd cxowzndfn tdggopd aog** Get the text here
- You may find this website helpful: http://substitution.webmasters.sk/simple-substitution-cipher.php
- Check with me once you think you have it, but don't spoil it for anyone else by telling them the answer!

Vigenère Cipher

- $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_{26}^m$
- The plaintext is arranged into 'vectors' of length m, and then the chosen vector K is added to each vector.
- With $m = 5$ and $K = (17, 7, 5, 10, 11)$ (or **rhfkl**), the plaintext 'ihavefoundtheproblem' will produce

x	8	7	0	21	4	5	14	20	13	3	19	7	4	15	17	14	1	11	4	12
+	17	7	5	10	11	17	7	5	10	11	17	7	5	10	11	17	7	5	10	11
$e_K(x)$	25	14	5	5	15	22	21	25	23	14	10	14	9	25	2	5	8	16	14	23

  which yields the ciphertext 'ZOFFPWVZXOKOJZCFIQOX'.
- As with the substitution cipher, there are well-known techniques to crack a Vigenère cipher so it is not secure.

One-Time Pad

- A one-time pad is exactly like a Vigenère cipher except that the length of the key is the same as the length of the plaintext.
- As long as the key is not compromised, it is unbreakable (*unconditionally secure*, provides *perfect secrecy*) since every possible plaintext is equally likely to be correct.
- The downside is that a very lengthy key needs to be exchanged and kept secret.
- **Example:** If you can decode this message, you will get an A in this course: **lcncwyyfhia**.

Public-key Cryptography

- So far, the cryptosystems we have seen use a secret key K that is shared between those who wish to communicate.
- Another way to think about them is that if you know how a message was encrypted, then you have enough information to decrypt it.
- These are called **private-key** or **symmetric** cryptosystems.
- In **public-key (or asymmetric) cryptography**, the full details of the encryption function e_K can be known publicly. The cryptosystem is designed so that it is computationally infeasible to determine d_K from e_K without additional information.
- The clear advantage is that no key needs to be shared between two people in order for them to communicate securely.

One-way Functions and Trapdoors

- An injective function $f : \mathcal{P} \to \mathcal{C}$ is called a **one-way function** if it is easy to compute $f(x)$ for all x, but, given y it is hard to find x such that $f(x) = y$.
- For public-key cryptography, we need the encryption function e_K to be a a one-way function with a trapdoor.
- A **trapdoor** consists of secret information that makes inversion of a one-way function easy.
- Thus, what is needed for public-key cryptography is a **trapdoor one-way function**.

Facts about $\mathbb{Z}_n$

- For a positive integer n, $\phi(n)$ is defined as the number of integers in $\{0, 1, 2, \ldots, n-1\}$ that are relatively prime to n.
- An element $a \in \mathbb{Z}_n$ is **invertible** under multiplication if and only if $\gcd(a, n) = 1$.
- The **inverse** of an element $a \in \mathbb{Z}_n$ is the number $b \in \mathbb{Z}_n$ such that $a \cdot b \bmod n = 1$. We denote the inverse of a as a^{-1}.
- Let
$$\mathbb{Z}_n^* = \{a \in \mathbb{Z}_n \,:\, a^{-1} \text{ exists in } \mathbb{Z}^n\}.$$
Then $\mathbb{Z}_n^*$ is a group under multiplication, and $|\mathbb{Z}_n^*| = \phi(n)$.

More about $\phi(n)$

- If p is prime, then $\phi(p) = p - 1$.
- If p is prime and $e \geq 2$, then $\phi(p^e) = p \cdot \phi(p^{e-1})$.
- If p is prime and $e \geq 1$, then $\phi(p^e) = p^e - p^{e-1}$.
- If p and q are relatively prime, then $\phi(pq) = \phi(p)\phi(q)$.
- If $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ is a prime factorization,

$$\phi(n) = \prod_{i=1}^{k} (p_i^{e_i} - p_i^{e_i - 1}).$$

- We will need this later:
  Theorem: If $b \in \mathbb{Z}_n^*$, then $b^{\phi(n)} \equiv 1 \pmod{n}$.

Euclidean Algorithm

- The Euclidean Algorithm solves two problems:
  - Given positive integers a and b, find $\gcd(a, b)$.
  - Given positive integers b and n with $\gcd(b, n) = 1$, find b^{-1} in $\mathbb{Z}_n$.
- **Example:** Compute $\gcd(70, 26)$.

$$\begin{aligned} 70 &= 2 \times \mathbf{26} + \mathbf{18} \\ 26 &= 1 \times \mathbf{18} + \mathbf{8} \\ 18 &= 2 \times \mathbf{8} + \mathbf{2} \\ 8 &= 4 \times \mathbf{2} + \mathbf{0} \end{aligned}$$

When the remainder is 0, the GCD is the number on the right side of the $\times$. Thus, $\gcd(70, 26) = 2$.

Euclidean Algorithm to find a^{-1}

Example: To find 11^{-1} in $\mathbb{Z}_{26}$.

- First, computer $gcd(26, 11)$ using Euclidean algorithm.

$$
\begin{aligned}
26 &= 2 \times \mathbf{11} + \mathbf{4} \\
11 &= 2 \times \ \ \mathbf{4} + \mathbf{3} \\
4 &= 1 \times \ \ \mathbf{3} + \mathbf{1} \\
3 &= 3 \times \ \ \mathbf{1} + \mathbf{0}
\end{aligned}
$$

- Next, do substitutions backwards to find a^{-1}.

$$
\begin{aligned}
1 &= 4 - (1 \times \mathbf{3}) \\
1 &= 4 - (1 \times (11 - 2 \times \mathbf{4})) = 3 \times \mathbf{4} - 11 \\
1 &= 3 \times (26 - 2 \times \mathbf{11}) - 11 = 3 \times 26 - 7 \times \mathbf{11}
\end{aligned}
$$

Thus, $-7 \times 11 \equiv 1 \bmod 26$, so
$11^{-1} \bmod 26 = (26 - 7) = 19$.

RSA

- Let p and q be distinct odd primes. Let $n = pq$.
- We have $\phi(n) = (p-1)(q-1)$.
- $\mathcal{P} = \mathcal{C} = \mathbb{Z}_n$.
- $\mathcal{K} = \{(n, p, q, a, b) : ab \equiv 1 \pmod{\phi(n)}\}$.
- For $x \in \mathcal{P}$ and $y \in \mathcal{C}$, define

$$e_K(x) = x^b \mod n,$$

and

$$d_K(y) = y^a \mod n.$$

- Public key: n and b.
- Private information: p, q, a.

Why RSA works

- $e_K(x) = x^b \mod n$ where $d_K(y) = y^a \mod n$ and $ab \equiv 1 \pmod{\phi(n)}$.
- We need to show that decryption "works,", i.e. that for all x, $d_K(e_K(x)) = x$. This amounts to showing that

$$(x^b)^a \equiv x \pmod{n} \qquad \text{for all } x \in \mathbb{Z}_n.$$

- If $x \in \mathbb{Z}_n^*$, then

$$(x^b)^a \equiv x^{ab} \equiv x^{\phi(n)t+1} \equiv (x^{\phi(n)})^t x \equiv 1^t x \equiv x \pmod{n}.$$

- If $x \in \mathbb{Z}_n \setminus \{\mathbb{Z}_n^* \cup 0\}$ (That is, x has either p or q as a factor), then it can also be shown that $(x^b)^a \equiv x \pmod{n}$, but it is more complicated.

Example of RSA

- Suppose $n = 98069$ and $b = 36119$.
- If the plaintext is $x = 76111$, then

$$e_K(x) = 76111^{36119} \mod 98069 = 91332.$$

- With the additional information that $n = 98069 = 281 \cdot 349$, Bob can compute $\phi(n) = 280 \cdot 348 = 97440$, and then compute

$$36119^{-1} \mod 97440 = 839.$$

  Then

$$d_K(91332) = 91332^{839} \mod 98069 = 76111.$$

RSA Excercise

- Assume you know the following

$$\begin{aligned} n &= 42876092449717 \\ b &= 33389740312697 \quad \text{(encryption key)} \\ e_k(x) &= 37247990695057 \quad \text{(cipher text)} \end{aligned}$$

  Find the plaintext x. ▸ Get the values here
- Hint: Factor n, compute $\phi(n)$, compute $a = b^{-1} \pmod{\phi(n)}$, and finally compute x.
- You may use ▸ WolframAlpha or similar tool for your computations.

Security of RSA

- RSA is believed to be secure for large primes p and q.
- $e_K(x) = x^b \mod n$ is believed to be a one-way function.
- The trapdoor is the factorization of n as pq.
- If someone knows p and q, they can compute $\phi(n) = (p-1)(q-1)$, and thereby compute a using the extended Euclidean algorithm.

Implementation

- The primes p and q must be chosen large enough so that factoring n is computationally infeasible. For safety, p and q are typically primes that require 512 bits to represent them in binary.
- Let n be a k-bit integer. RSA requires
  - modular addition and subtraction mod n (takes $O(k)$ time),
  - modular multiplication mod n (takes $O(k^2)$ time), and
  - modular inversion mod n (takes $(O(k^3))$ time).
- Computing $x^c \mod n$ can be done using $c - 1$ modular multiplications, but this is very inefficient if c is large.
- Instead, we use the SQUARE AND MULTIPLY ALGORITHM, which runs in time $O(k^2 \log c)$.

Repeated Squaring

- Computing $x^c \mod n$ using square and multiply algorithm is pretty straightforward.
- Intuitively, we express c in binary as $c_{\ell-1}c_{\ell-2}\cdots c_1c_0$, then compute $x^c \mod n$ by computing

$$x^{c_0}(x^{c_1}(x^{c_2}(\cdots(x^{c_{\ell-1}})^2\cdots)^2)^2)^2.$$

- For example, to compute $3^{57} \mod 7$, we write $57 = 111001_2$. Then

$$3^{57} = 3^{32}3^{16}3^{8}3^{1} = 3(((3(3(3)^2)^2)^2)^2)^2.$$

From this, we can see that $3^{57} \mod 7 = 6$.

RSA Implementation and Parameter Generation

- Choose two large primes p and q.
- Set $n = pq$ and $\phi(n) = (p-1)(q-1)$. This can be done in time $O((\log n)^2)$.
- Choose a random b with $\gcd(b, \phi(n)) = 1$, and compute $a = b^{-1} \pmod{\phi(n)}$. This can be done in time $O((\log n)^2)$ using the EXTENDED EUCLIDEAN ALGORITHM.
- RSA encryption and decryption using the SQUARE AND MULTIPLY ALGORITHM each take time $O((\log n)^3)$.

Related Topics

- Hashing (How do you store passwords so that they cannot be retrieved?)
- Digital Signatures (How can you authenticate the sender of a message?)
- Key Distribution (How do you exchange private keys over a public channel?)
- Identification Schemes (How do you prove you are who you say you are?)
- Secrete Sharing Schemes (How do you require that (for instance) two of three people be present to open a safe?)
- Zero Knowledge Proofs (How do you convince someone that a statement is true without revealing *any* information beyond the fact that the statement is true?)