
Cryptography Introduction/Refresher

- Brief introduction to make sure everyone's is on the same page
- Important concepts:
 - Symmetric ciphers
 - Public key encryption
 - Digital signatures
 - Cryptographic hash functions
 - Message Authentication Codes (MACs)
 - Certificates

What is a Cryptosystem?

- $K = \{0, 1\}^l$
- $P = \{0, 1\}^m$
- $C' = \{0, 1\}^n, C \subseteq C'$
- $E : P \times K \rightarrow C$
- $D : C \times K \rightarrow P$
- $\forall p \in P, k \in K : D(E(p, k), k) = p$
- It is infeasible to find $F : P \times C \rightarrow K$

Let's start again, in English...

What is a Cryptosystem?

A cryptosystem is pair of algorithms that take a *key* and convert *plaintext* to *ciphertext* and back.

Plaintext is what you want to protect; ciphertext should appear to be random gibberish.

The design and analysis of today's cryptographic algorithms is highly mathematical. Do *not* try to design your own algorithms.

Properties of a Good Cryptosystem

- There should be no way short of enumerating all possible keys to find the key from any reasonable amount of ciphertext and plaintext, nor any way to produce plaintext from ciphertext without the key.
- Enumerating all possible keys must be infeasible.
- The ciphertext must be indistinguishable from true random values.

Kerckhoffs' Law (1883)

There must be no need to keep the system secret, and it must be able to fall into enemy hands without inconvenience.

In other words, the security of the system must rest entirely on the secrecy of the key.

Keys

- Must be strongly protected
- Ideally, should be a random set of bits of the appropriate length
- Ideally, each key should be used for a limited time only
- Ensuring that these properties hold is a major goal of cryptographic research and engineering

Cipher Strengths

- A cipher is no stronger than its key length: if there are too few keys, an attacker can enumerate all possible keys
- DES has 56 bits — arguably too few in 1976; far too few today.
- Strength of cipher depends on how long it needs to resist attack.
- No good reason to use less than 128 bits
- NSA rates 128-bit AES as good enough for SECRET traffic; 256-bit AES is good enough for TOP-SECRET traffic.
- But a cipher can be considerably weaker! (A monoalphabetic cipher over all bytes has a 1684-bit key, but is trivially solvable.)

CPU Speed versus Key Size

- Adding one bit to the key doubles the work factor for brute force attacks
- The effect on encryption time is often negligible or even free
- It costs *nothing* to use a longer RC4 key
- Going from 128-bit AES to 256-bit AES takes (at most) 40% longer, but increases the attacker's effort by a factor of 2^{128}
- Using triple DES costs $3\times$ more than DES to encrypt, but increases the attacker's effort by a factor of 2^{112}
- Moore's Law favors the defender

Block Ciphers

- Operate on a fixed-length set of bits
- Output blocksize generally the same as input blocksize
- Well-known examples: DES (56-bit keys; 64-bit blocksize); AES (128-, 192-, and 256-bit keys; 128-bit blocksize)

Stream Ciphers

- Key stream generator produces a sequence S of pseudo-random bytes; key stream bytes are combined (generally via XOR) with plaintext bytes
- $P_i \oplus S_i \rightarrow C_i$
- Stream ciphers are very good for asynchronous traffic
- Best-known stream cipher is RC4; commonly used with SSL
- Key stream S must *never* be reused for different plaintexts:

$$\begin{aligned}C &= A \oplus K \\C' &= B \oplus K \\C \oplus C' &= A \oplus K \oplus B \oplus K \\ &= A \oplus B\end{aligned}$$

- Guess at A and see if B makes sense; repeat for subsequent bytes

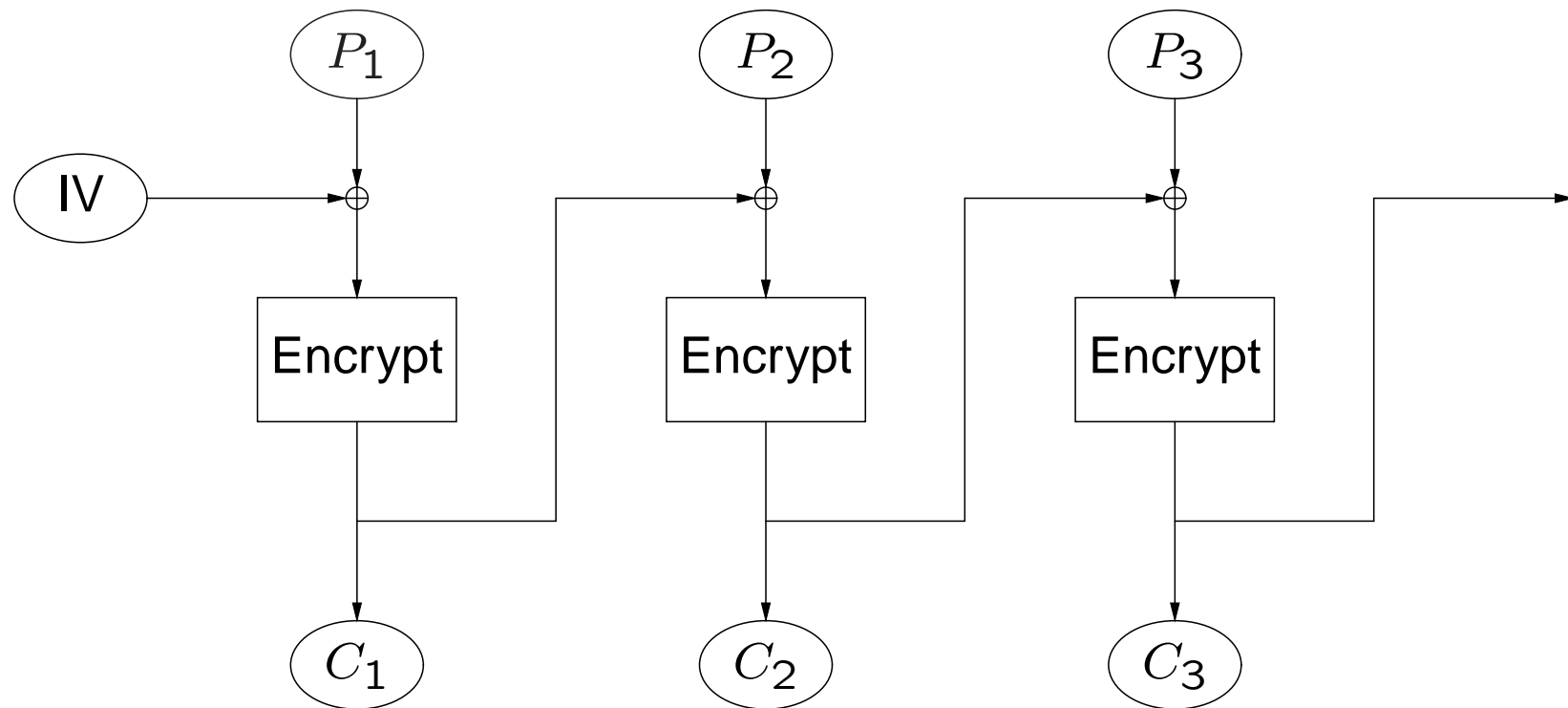
Basic Structure of (Most) Block Ciphers

- Optional key scheduling — convert supplied key to internal form
- Multiple *rounds* of combining the plaintext with the key.
- DES has 16 rounds; AES has 9-13 rounds, depending on key length

Modes of Operation

- Direct use of a block cipher is inadvisable
- Enemy can build up “code book” of plaintext/ciphertext equivalents
- Beyond that, direct use only works on messages that are a multiple of the cipher block size in length
- Solution: five standard *Modes of Operation*: Electronic Code Book (ECB), Cipher Block Chaining (CBC), Cipher Feedback (CFB), Output Feedback (OFB), and Counter (CTR).
- All modes of operation except ECB require an extra block known as the *Initialization Vector* (IV)

Example: Cipher Block Chaining



$$\{P_i \oplus C_{i-1}\}_k \rightarrow C_i$$
$$\{C_i\}_{k-1} \oplus C_{i-1} \rightarrow P_i$$

Things to Notice About CBC

- Identical plaintext blocks do not, in general, produce the same ciphertext. (Why?)
- Each ciphertext block is a function of all previous plaintext blocks. (Why?)
- The converse is not true, but we won't go into that in this class

Alice and Bob

- Alice wants to communicate securely with Bob
- (Cryptographers frequently speak of Alice and Bob instead of A and B ...)
- What key should she use?

Pre-Arranged Key Lists?

- What if you run out of keys?
- What if a key is stolen?

“Why is it necessary to destroy yesterday’s [key] . . . list if it’s never going to be used again?”

“A used key, Your Honor, is the most critical key there is. If anyone can gain access to that, they can read your communications.”

(trial of Jerry Whitworth, a convicted spy.)

- What if Alice doesn’t know in advance that she’ll want to talk to Bob?

The Solution: Public Key Cryptography

- Allows parties to communicate without prearrangement
- Separate keys for encryption and decryption
- Not possible to derive decryption key from encryption key
- Permissible to publish encryption key, so that anyone can send you secret messages
- All known public key systems are very expensive to use, in CPU time and bandwidth.
- Most public systems are based on mathematical problems.

RSA

- The best-known public key system is RSA.
- Generate two large (at least 512 bit, probably more) primes p and q ;
let $n = pq$
- Pick two integers e and d such that $ed \equiv 1 \pmod{(p-1)(q-1)}$.
Often, $e = 65537$, since that simplifies encryption calculations.
- The public key is $\langle e, n \rangle$; the private key is $\langle d, n \rangle$.
- To encrypt m , calculate $c = m^e \pmod n$; to decrypt c , calculate $m = c^d \pmod n$.
- The security of the system relies on the difficulty of factoring n .
- Finding such primes is relatively easy; factoring n is believed to be extremely hard.

Classical Public Key Usage

- Alice publishes her public key in the phone book.
- Bob prepares a message and encrypts it with that key by doing a large exponentiation.
- Alice uses her private key to do a different large exponentiation.
- It's not that simple. . .

Complexities

- RSA calculations are *very* expensive; neither Bob nor Alice can afford to do many.
- RSA is too amenable to mathematical attacks; encrypting the wrong numbers is a bad idea.
- Example: “yes”³ is only 69 bits, and won’t be reduced by the modulus operation; finding $\sqrt[3]{503565527901556194283}$ is easy.
- We need a better solution

A Realistic Scenario

- Bob generates a random key k for a conventional cipher.
- Bob encrypts the message: $c = \{m\}_k$.
- Bob pads k with a known amount of padding, to make it at least 512 bits long; call this k' .
- k' is encrypted with Alice's public key $\langle e, n \rangle$.
- Bob transmits $\{c, (k')^e \bmod n\}$ to Alice.
- Alice uses $\langle d, n \rangle$ to recover k' , removes the padding, and uses k to decrypt ciphertext c .
- In reality, it's even more complex than that. . .

Who Sent a Message?

- When Bob receives a message from Alice, how does he know who sent it?
- With traditional, symmetric ciphers, he may know that Alice has the only other copy of the key; with public key, he doesn't even know that
- Even if he knows, can he prove to a third party — say, a judge — that Alice sent a particular message?

Digital Signatures

- RSA can be used backwards: you can encrypt with the private key, and decrypt with the public key.
- This is a *digital signature*: only Alice can sign her messages, but anyone can verify that the message came from Alice, by using her public key
- It's too expensive to sign the whole message. Instead, Alice calculates a *cryptographic hash* of the message and signs the hash value.
- If you sign the plaintext and encrypt the signature, the signer's identity is concealed; if you sign the ciphertext, a gateway can verify the signature without having to decrypt the message.

They're Not Like Real Signatures

- Real signatures are strongly bound to the person, and weakly bound to the data
- Digital signatures are strongly bound to the data, and weakly bound to the person — what if the key is stolen (or deliberately leaked)?
- A better term: digital signature algorithms provide *non-repudiation*

Cryptographic Hash Functions

- Produce relatively-short, fixed-length output string from arbitrarily long input.
- Computationally infeasible to find two different input strings that hash to the same value (“collision”)
- Computationally infeasible to find any input string that hashes to a given value (“pre-image”)
- Computationally infeasible to find any input string that hashes to the same value as the hash of a given input (“second preimage”)
- Strength roughly equal to half the output length
- Best-known cryptographic hash functions: MD5 (128 bits), SHA-1 (160 bits), SHA-256/384/512 (256/384/512 bits)
- 128 bits and shorter are not very secure for general usage

The Birthday Paradox

- How many people need to be in a room for the probability that two will have the same birthday to be $> .5$?
- Naive answer: 183
- Correct answer: 23
- The question is not “who has the same birthday as Alice?”; it’s “who has the same birthday as Alice or Bob or Carol or . . .” assuming that none of them have the same birthday as any of the others

The Birthday Attack

- Alice can prepare lots of variant contracts, looking for any two that have the same hash
- More precisely, she generates many trivial variants on m and m' , looking for a match between the two sets
- This is much easier than finding a contract that has the same hash as a given other contract
- As a consequence, the strength of a hash function against brute force attacks is approximately half the output block size: 64 bits for MD5, 80 bits for SHA-1, etc.

Message Integrity

- We need a way to prevent tampering with messages
- We can use a key and a cryptographic hash to generate a *Message Authentication Code* (MAC).
- Simpler solutions don't work
- One bad idea: append a cryptographic hash to some plaintext, and encrypt the whole thing with, say, CBC mode

$$\{P, H(P)\}_K$$

- This can fall victim to a *chosen plaintext attack*

HMAC

- Build a MAC from a cryptographic hash function
- Best-known construct is HMAC — provably secure under minimal assumptions
- $\text{HMAC}(m, k) = H(\text{opad} \oplus k, H(\text{ipad} \oplus k, m))$ where H is a cryptographic hash function
- Note: authentication key *must* be distinct from the confidentiality key
- Frequently, the output of HMAC is truncated

Cryptography and Authentication

- Some way to use a cryptographic key to prove who you are
- (Much more on that next class)
- Can go beyond simple schemes given above
- Can use symmetric or public key schemes
- Most public key schemes use *certificates*

What are Certificates

- How does Alice get Bob's public key?
- What if the enemy tampers with the phone book? Sends the phone company a false change-of-key notice? Interferes with Alice's query to the phone book server?
- A certificate is a digitally-signed message containing an identity and a public key — prevents tampering.

Why Trust a Certificate?

- Who signed it? Why do you trust them?
- How do you know the public key of the *Certificate Authority (CA)*?
- Some public key (known as the *trust anchor*) must be provided out-of-band — trust has to start somewhere.

Certificate Authorities

- Who picks CAs? No one and every one.
- Your browser has some CAs built-in — because the CA paid the browser vendor enough money. Is that grounds for trust?
- Matt Blaze: “A commercial certificate authority can be trusted to protect you from anyone from whom they won’t take money.”

What Else is in a Certificate?

- Technical information, such as algorithm identifiers
- More identification information — company, location, etc.
- Expiration date
- Logos
- Certificate role

Cryptographic Protocols

- Combine various cryptographic primitives in a series of messages
- Many different types, for many different goals
- Simplest example: “realistic” public key encryption message discussed earlier: $\langle \{m\}_k, (\text{pad}(k))^e \bmod n \rangle$
- Very common goal: Alice and Bob must agree on a key
- Very subtle; very hard to get right. Don't try it yourself

Recommended Primitives

- Block cipher: AES
- Stream cipher: RC4?
- Hash function: SHA-256 (perhaps SHA-1)
- Public key, digital signature: RSA with 2048-bit modulus (or Elliptic Curve Cryptography if patents aren't an issue and performance is)