CS 556 – Computer Security
Fall 2012

Dr. Indrajit Ray
Email: indrajit@cs.colostate.edu

Department of Computer Science
Colorado State University
Fort Collins, CO 80523, USA
Cryptography

Historical Cryptography

Basic Concepts

Attacking Crypto

Breaking Ciphers - Some Intuitions

Cryptography
Cryptology

- Comes from the Greek words
  - kryptos: hidden
  - logos: word
- Combines the disciplines of Cryptography and Cryptanalysis
Cryptography

- Originally – the art or science encompassing the principles and methods of transforming an intelligible message into one that is unintelligible, and then retransforming that message back to its original form
- Currently – study of techniques and applications that depend on the existence of difficult problems
Cryptography

- Provides a tool for
  - secrecy
  - integrity
  - authentication
  - non-repudiation

- In the face of
  - passive and
  - active attacks

- Not intended to solve inference problem
Attacks

- Passive attacks
  - observe but do not modify information
  - threat for confidentiality

- Active attacks
  - delete, add, modify and replay information
  - threat for confidentiality, integrity, authentication and non-repudiation
Cryptanalysis

- How to compromise cryptographic mechanisms
  - benevolent intention: to judge the strength of cryptographic techniques and improve upon them
  - malevolent intention: to breach security
HISTORICAL CRYPTOGRAPHY
A Brief History of Crypto

- Have a history of at least 4000 years
- Ancient Egyptians enciphered some of their hieroglyphic writing on monuments
- Ancient Hebrews enciphered certain words in the scriptures
- 2000 years ago Julius Caesar used a simple substitution cipher, now known as the Caesar cipher
A Brief History of Crypto

- Roger Bacon described several methods in 1200s
- Geoffrey Chaucer included several ciphers in his works
- Leon Alberti devised a cipher wheel, and described the principles of frequency analysis in the 1460s
- Blaise de Vigenère published a book on cryptology in 1585, & described the polyalphabetic substitution cipher
Thomas Jefferson developed it around 1790; comprised of 36 disks, each with a random alphabet; order of disks was key; message was set, then another row became cipher
Machine Cipher - Wheatstone Disc

- originally invented by Wadsworth in 1817, but developed by Wheatstone in 1860’s; comprised two concentric wheels used to generate a polyalphabetic cipher
Heavily used during World War II; comprised of a series of rotor wheels with internal cross-connections providing a substitution using a continuously changing alphabet.
Basic Concepts
Encryption - the process of coding information such that the meaning is concealed

- encode / encipher are synonyms

Decryption - the process of transforming an encrypted information back to the original form

- decode / decipher are synonyms

Cryptosystem - a system for encryption and decryption
**Terminology**

- **Plaintext** - information in its original form (also called cleartext)
- **Ciphertext** - information in the encrypted form
- **Cipher** - an algorithm for transforming an intelligible message into one that is unintelligible
- **Key** - some critical information used by the cipher, together with the plaintext to generate the ciphertext
Model of Conventional Cryptosystems

Alice

Message Source

Encryption Algorithm

plaintext

key

ciphertext (encrypted plaintext)

Decryption Algorithm

plaintext (decrypted ciphertext)

Charlie

cryptanalyst

Bob

Message Destination

Dr. Indrajit Ray, Computer Science Department

CS 556 - Computer Security - © 2012 Colorado State University – 17 / 41
**Notation**

- $C = E[M,K]$
- $M = D[C,K]$
- $K$: Key
- $E$: Encryption algorithm
- $D$: Decryption algorithm
- $M$: Original message (plain text)
- $C$: Encrypted message (cipher text)
Cryptographic Technologies

- Secret-key cryptosystem
  - Also known as single key / shared key / symmetric key cryptosystem
  - Same key used for encryption and decryption
  - Sometimes the decryption key may be derived from the encryption key
    - However the encryption and decryption keys are not commutative
Public-key cryptosystem

- Also known as two key / asymmetric key cryptosystem
- Different keys used for encryption and decryption
- The encryption and decryption keys are typically commutative

Strong mathematical relation exists between the two keys
ATTACKING CRYPTO
Cryptanalysist is assumed to know E and D

- Objective of the cryptanalyst is to discover the key
- Real objective may be to discover the plaintext message M, but
  - this is generally assumed to be equivalent to discovering the key
  - it is more rewarding for the cryptanalysis to discover the key
Attack Models for Cryptanalysis

- Ciphertext only or Known ciphertext
- Plaintext only or Known plaintext
- Chosen plaintext
- Chosen ciphertext
Ciphertext Only

- Cryptanalyst assumed to have access to a subset of ciphertexts
- Attack is successful if corresponding plaintexts can be deduced
  - Any information about underlying plaintext is also considered success
    - For example, is it information about salary.
- Must be able to guess when we have plaintext
  - Sometimes the statistics of the ciphertext provide insight and can lead to a break
**Known Plaintext**

- Cryptanalyst knows (or suspects) some plaintext-ciphertext pairs
  - We have some, or even large, amount of matching plaintext and ciphertext. The goal is to extract the key.
- Knowledge of text properties can be used to simplify attack
  - For example, if plaintext is known to be ASCII, as well as ciphertext, then only $2^8$ keys can produce the result.
Cryptanalyst has the capability to choose arbitrary plaintexts to be encrypted and obtain the corresponding ciphertexts.

- Feasible when attacker has access to the encryption hardware or software
- Can use knowledge of algorithm structure to attack
- Batch chosen plaintext: Cryptanalyst chooses all the plaintexts before any of them are encrypted.
- Adaptive chosen plaintext: Cryptanalyst makes a series of interactive queries and choosing subsequent plaintexts based on the information from the previous encryptions.
**Chosen Ciphertext**

- Cryptanalyst knows some plaintext-ciphertext pairs for ciphertext of the cryptanalyst's choice
- Cryptanalyst submits arbitrary encrypted messages to be deciphered and see the resulting plaintext
- Feasible when attacker has access to the decryption hardware or software
General Attack Techniques

- Brute Force (also called Exhaustive search)
  - Try to decipher ciphertext under every possible key until readable messages are produced.
  - Given enough time all cryptosystems can be broken by brute-force.
  - Question remains “What is readable?”

- Divide and Conquer to make brute-force easier
  - Isolate small components or aspects so they can be solved separately
General Attack Techniques

- **Dictionary**
  - Form a list of the most likely keys, then try those keys one-by-one (a way to improve brute force)

- **Codebook**
  - Develop or collect a lookup table of transformations. Each plaintext has one or more ciphertexts in the table
  - Match a plaintext-ciphertext pair against the codebook
General Attack Techniques

★ Birthday Attacks

✦ Use the birthday paradox; the idea is that it is much easier to find two values which match than it is find a match to some particular value
✦ Typically birthday attacks are used to break message digest algorithms

★ Replay Attacks

✦ Record and save some ciphertext blocks or messages (especially if the content is known) then resend those blocks when useful
✦ Very common technique to bypass authentication protocols (More later)
General Attack Techniques

- **Differential Cryptanalysis**
  - Find a statistical correlation between key values and cipher transformations (typically the XOR of text pairs); then use sufficient defined plaintext to develop the key
  - Typically used against symmetric key cryptosystem that is iterative in structure

- **Algebraic Coding**
  - From the cipher design, develop equations for the key in terms of known/ chosen plaintext, then solve those equations
**General Attack Techniques**

- **Related Key**: Specify a change in any particular key bit, or some other relationship between keys and observe the effect on the ciphertext, specially patterns of non-randomness
  - 10 round AES-256 has been broken this way within practical complexity

- **Timing**: Measure the duration of ciphering operations and use that to reveal information about the algorithm, key or data

- **Fault Analysis**: Induce random faults into the ciphering system and use those to expose the key
General Attack Techniques

- Man-in-the-middle: Subvert the routing capabilities of a network and pose as the other side to each end of the communications
  - Have been used to break Double DES
  - Diffie-Hellman is susceptible to this kind of attacks
- Use pitfalls in protocol design (More in the discussion)
- Use bugs in crypto implementation
Strength of Encryption

- There is no theory which guarantees strength for any conventional cipher
- Ciphers traditionally are considered strong when they have been used for a long time with “nobody” knowing how to break them
- While cryptanalysis can prove “weakness” for a given level of effort, cryptanalysis cannot prove that there is no simpler attack
Two fundamentally different ways ciphers may be secure

- Unconditional security
  - No matter how much computer power is available, the cipher cannot be broken

- Computational security
  - Given limited computing resources (e.g. time needed for calculations is greater than age of universe), the cipher cannot be broken
Breaking Ciphers - Some Intuitions
Breaking Ciphers

- Ciphertext is known, want to find the plaintext
  - Know what keys are possible
  - Want to try out all key
  - Stop at the point where we have found the key

- Two questions
  - How many keys do we have to try?
  - How likely is it that some key will give us a message by chance that is actually wrong but looks as if it might be correct?
How Many Keys to Try?

- Worst case – If keys are k-bit numbers and all numbers are equally likely then by trying $2^k$ we will find the correct key.
- Frequently, the possible keys are not all equally likely.
  - If probability of a bit being 1 is $p < 0.5$ then 0s are more likely than 1s in the key.
  - Try all zero keys first, then all keys with a single 1 and so on.
- Do we gain by following such a strategy?
  - How many keys do we have to try in this order until the probability that we have tried the actual key reaches some value say 0.5 or 0.75 or 0.99?
Brief Intro to Entropy

- Entropy of a message is a measure of the information content of a message as measured in bits
- Defined as the minimum number of binary digits necessary to encode a message
- If $s_n$ is the number of occurrences of symbol $s$ in message $S$, $S_m$ is the total number of symbols in $S$ with the symbols drawn from a set $T$ whose cardinality is $|T|$, then
  - Entropy for a given symbol $s$ is $-\log_{|T|}(\frac{s_n}{S_m})$
  - Entropy for a collection of like symbols $s$ in $S$ is $-\log_{|T|}(\frac{s_n}{S_m}) \cdot (s_n)$
  - Entropy of $S$ is calculated as the sum of the entropies for all collections of like symbols in $S$
Assume probabilities of individual binary digits are independent; probability of a bit being 1 is $p$; probability of a bit being 0 is $(1 - p)$; number of bits = $n$

- Probability for a key (string of 1’s and 0’s) is the product of the probabilities of the bits.
- $\log(\text{probability for a key}) = \sum \log(\text{probability of individual bits})$
  - $\log(p) \times \text{number of 1’s} + \log(1 - p) \times \text{number of 0’s}$
- For key of length $n$ bits, expected number of 1’s = $n \times p$ and expected number of 0’s = $n \times (1 - p)$
- Expected value of the logarithm of the probability for key is $n(p \times \log(p) + (1 - p) \times \log(1 - p))$
Entropy of a single binary digit = 
\[ p \times (-\log(p)) + (1 - p) \times (-\log(1 - p)) \]

Example - If \( p = 0.1 \) and we use base 2 logarithm then the entropy of 1 bit is \( H(p) = 0.469 \) and entropy of a 128 bit key is about 60

Probability of keys around the average is roughly \( 2^{-60} \)