Lecture 2: Some Basic Vocabulary of Computer and Network Security and a Brief Review of Classical Encryption Techniques

Assoc. Prof. Yıldıran YILMAZ (yildiran.yilmaz@erdogan.edu.tr)

2025

Goals

- To introduce the rudiments of the vocabulary of computer and network security and that of encryption/decryption.
- To trace the history of some early approaches to cryptography and to show through this history a common failing of humans to get carried away by the technological and scientific hubris of the moment.
- Simple Python and Perl scripts that give you pretty good security for confidential communications. Only good for fun, though.

Table of Contents

1. Some Basic Vocabulary to Get Us Started
2. Building Blocks of Classical Encryption Techniques
3. Caesar Cipher
4. The Swahili Angle ...
5. Monoalphabetic Ciphers
6. The All-Fearsome Statistical Attack
7. Multi-Character Encryption: Playfair Cipher
8. Hill Cipher
9. Polyalphabetic Ciphers: Vigenere Cipher
10. Transposition Techniques
11. Secure Communications for Fun
12. Homework Problems

Basic Vocabulary: Encryption and Decryption

- **plaintext**: This is what you want to encrypt
- **ciphertext**: The encrypted output
- **enciphering or encryption**: The process by which plaintext is converted into ciphertext
- **encryption algorithm**: The sequence of data processing steps that go into transforming plaintext into ciphertext.

Secret Key

- A **secret key** is used either to set some or all of the various parameters used by the encryption algorithm or for directly mixing with the plaintext.
- **Important**: In classical cryptography, the same secret key is used for encryption and decryption.
- This is why classical cryptography is also referred to as **symmetric key cryptography**.
- In modern cryptographic algorithms, the encryption and decryption keys are different, and one of them is placed in the public domain.
- Such algorithms are commonly referred to as **asymmetric key cryptography**, **public key cryptography**, etc.

More Vocabulary

- **deciphering or decryption**: Recovering plaintext from ciphertext
- **decryption algorithm**: The sequence of data processing steps that go into transforming ciphertext back into plaintext.
- **cryptography**: The many schemes available today for encryption and decryption
- **cryptographic system**: Any single scheme for encryption and decryption
- **cipher**: Means the same thing as a "cryptographic system"

Block Cipher vs Stream Cipher

- **block cipher**: Processes a block of input data at a time and produces a ciphertext block of the same size.
- **stream cipher**: Encrypts data on the fly, usually one byte at a time.

Cryptanalysis

- Means “breaking the code”.
- Relies on knowledge of the encryption algorithm and some knowledge of the possible structure of the plaintext.
- Goal: Partial or full reconstruction of the plaintext from ciphertext, and infer the key for decryption of future messages.
- Methods depend on:
  - Whether the attacker has just ciphertext, or plaintext-ciphertext pairs
  - How much structure is possessed by the plaintext
  - How much of that structure is known to the attacker

Key Space and Brute-Force Attack

- **key space**: The total number of all possible keys that can be used in a cryptographic system.
- Example: DES uses a 56-bit key, so key space size is $2^{56} \approx 7.2 \times 10^{16}$.
- **brute-force attack**: Trying every possible key on a piece of ciphertext until an intelligible translation into plaintext is obtained.

Codebook Attack

- In general, a codebook is a mapping from plaintext symbols to ciphertext symbols.
- In a codebook attack, the attacker tries to acquire as many mappings as possible.
- In modern times, a codebook is the mapping between plaintext bit blocks and ciphertext bit blocks.
- If you collect mappings between all possible plaintext bit-blocks and their ciphertext bit-blocks, you have broken the code regardless of the secret key used.

Algebraic Attack and Time-Memory Tradeoff

- **algebraic attack**: Express plaintext-to-ciphertext relationship as a system of equations. Try to solve for the encryption key.
- **time-memory tradeoff**: Trading off memory for time to devise more effective attacks.
- Example: Rainbow tables in password cracking (Lecture 24).

Cryptology and Systems Security Vocabulary

- **cryptology**: Cryptography and cryptanalysis together.
- Systems security vocabulary sources:
  - Google's Android Security Reports (2018)
  - Android White Papers (2021)
  - Cisco cybersecurity reports

Backdoor

- A backdoor allows an intruder to get inside a networked device without user authentication credentials.
- Created by:
  - Malware installed through phishing attacks
  - Exploiting vulnerabilities in security protocols

Commercial Spyware

- Any application that transmits sensitive information off the device without user consent and does not display a persistent notification.
- Legitimate uses: Parental tracking
- Illegitimate uses: Tracking without knowledge or permission

Denial of Service

- Goal: Prevent legitimate users from accessing a network resource.
- Malware may turn a machine into a device for mounting a denial-of-service attack.

Hostile Downloader

- An application that is not in itself potentially harmful, but downloads other potentially harmful apps.
- Example: A gaming app that displays a misleading “Security Update” link that installs harmful apps.

Mobile Billing Fraud

- An application that charges the user in an intentionally misleading way.
- Types:
  1. **sms fraud**: Charges users to send premium SMS without consent
  2. **call fraud**: Makes calls to premium-rate telephone numbers without user consent
  3. **toll fraud**: Tricks users to subscribe or purchase content via their mobile phone bill

Phishing

- An application that pretends to come from a trustworthy source, requests user's authentication credentials and/or billing information, and sends the data to a third party.
- Common targets: Banking credentials, credit card numbers, social network accounts.

Mobile Unwanted Software (MUwS)

- Any application that collects without user consent:
  - Information about installed applications
  - Information about third-party accounts
  - Names of files on the device

Privilege Escalation

- An application that compromises the integrity of the system by:
  - Breaking the application sandbox
  - Changing or disabling access to core security-related functions
- Can allow an app to steal credentials from other apps and prevent its own removal.
- Example: Meltdown and Spectre processor vulnerabilities.

Ransomware

- Malware that makes your computer unusable by encrypting all your files.
- Types:
  - Locks user out of device and demands money to restore control
  - Encrypts data and demands payment for decryption
  - Leverages device policy manager features and cannot be removed by user

Rooting

- A privilege escalation app that roots the device.
- **Non-malicious rooting apps**: Inform user in advance, do not execute other harmful actions.
- **Malicious rooting apps**: Do not inform user, or inform but execute other harmful actions.

Spam and Spyware

- **spam**: Unsolicited, unwanted, and frequently annoying email messages.
- **spyware**: An application that transmits sensitive information off the device without disclosure.
- Examples of sensitive information:
  - Contact list, photos, email content, call log, SMS log, web history, etc.

SSL/TLS and TCP/IP

- **SSL/TLS**: Certificate-based client and server authentication that makes e-commerce possible.
- SSL = Secure Socket Layer, TLS = Transport Layer Security.
- **TCP/IP**: Foundational protocols for internet communication.
- IP: Specifies how hosts send data packets using IP addressing.
- TCP: Adds handshaking to ensure data packets are received.

Tor and Trojan

- **tor**: A route anonymizing protocol for accessing censored websites.
- **trojan**: An application that appears benign but engages in undesirable behavior.
- Example: A tic-tac-toe game that sends premium SMS messages in the background.

VPN

- **VPN (Virtual Private Network)**: An overlay network that allows hosts to communicate confidentially using IPSec.
- IPSec is a secure version of the IP protocol.

Useful Websites

- Krebs On Security Blog: `https://krebsonsecurity.com/`
- Digital Attack Map: `https://www.digitalattackmap.com/`
- Honeypots are used to detect attacks and capture malware.

Building Blocks

- Two building blocks: **substitution** and **transposition**.
- **substitution**: Replacing an element of plaintext with an element of ciphertext.
- Substitution rule may be applied uniformly or vary by position.
- **transposition**: Rearranging the order of appearance of plaintext elements.
- Also called **permutation**.
- Modern algorithms use multiple rounds of both.

Caesar Cipher

- Earliest known substitution cipher.
- Each character is replaced by a character three positions down.
- Example:
  - plaintext: are you ready
  - ciphertext: DUH BRX UHDGB
- Mathematical representation:

$$c = E(3, p) = (p + 3) \mod 26$$

General Caesar Cipher

- General version with any shift k:

$$c = E(k, p) = (p + k) \mod 26$$

- Decryption:

$$p = D(k, c) = (c - k) \mod 26$$

- k is the secret key.

The Swahili Angle

- A simple substitution cipher may seem weak if plaintext structure is known.
- But if plaintext is in an unknown language or format, it becomes harder.
- Example: Base64 encoding for email attachments.
- Base64 encodes 6 bits at a time into printable characters.
- Set: A-Z, a-z, 0-9, '+', '/'.

Internet Communications

- Slogan: "All internet communications are character based."
- Technically wrong (TCP/IP handles all files), but practically important.
- Non-printable characters in data files can cause corruption if not properly encoded.

A Seemingly Strong Monoalphabetic Cipher

- Caesar cipher is a monoalphabetic cipher.
- Consider a random permutation of the 26 letters:
  - plaintext: a b c d e f . . . . . .
  - substitution: t h i j a b . . . . . .
- Key: the random permutation.
- Key space size: $26! > 4 \times 10^{26}$.

Large Key Space But ...

- Key space is huge: $26! \approx 4 \times 10^{26}$.
- Much larger than DES key space ($2^{56} \approx 7.2 \times 10^{16}$).
- Brute-force attack seems infeasible.
- But... it's still breakable! Read on.

Statistical Attack

- Any monoalphabetic substitution cipher can be broken with statistical attack.
- Compare frequency distributions:
  - Single characters
  - Digrams (pairs)
  - Trigrams (triples)
- Compare with known statistics for the language (e.g., English).

Letter Frequencies in English

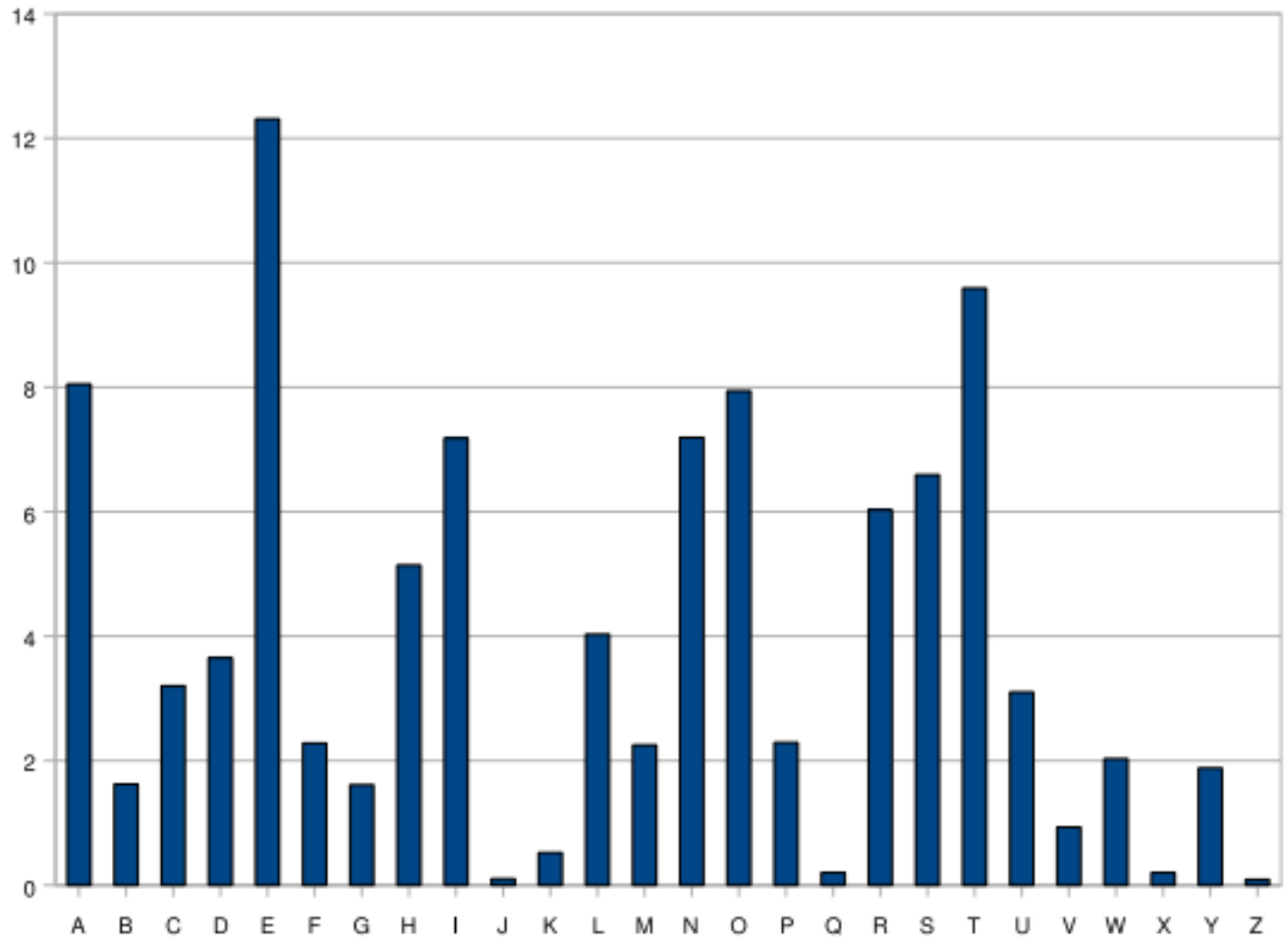

Figure: Relative frequencies of letters in English text.

Digram and Trigram Frequencies

- **Digrams**: Pairs of adjacent characters.
- **Trigrams**: Triples of characters.
- Most frequent trigrams: the, and, ent, ion, tio, for, nde, ...

Digram Frequencies Table

digram	freq	digram	freq	digram	freq	digram	freq
th	3.15	to	1.11	sa	0.75	ma	0.56
he	2.51	nt	1.10	hi	0.72	ta	0.56
an	1.72	ed	1.07	le	0.72	ce	0.55
in	1.69	is	1.06	so	0.71	ic	0.55
er	1.54	ar	1.01	as	0.67	ll	0.55
re	1.48	ou	0.96	no	0.65	na	0.54
es	1.45	te	0.94	ne	0.64	ro	0.54
on	1.45	of	0.94	ec	0.64	ot	0.53
ea	1.31	it	0.88	io	0.63	tt	0.53
ti	1.28	ha	0.84	rt	0.63	ve	0.53
at	1.24	se	0.84	co	0.59	ns	0.51
st	1.21	et	0.80	be	0.58	ur	0.49
en	1.20	al	0.77	di	0.57	me	0.48
nd	1.18	ri	0.77	li	0.57	wh	0.48
or	1.13	ng	0.75	ra	0.57	ly	0.47

Table: Digram frequencies in English text.

Playfair Cipher

- Maps multiple characters at a time to ciphertext.
- Uses a 5×5 matrix filled with a key and the alphabet (I/J share a cell).
- Example key: "smythework"

$$\begin{bmatrix} S & M & Y & T & H \\ E & W & O & R & K \\ A & B & C & D & F \\ G & I/J & L & N & P \\ Q & U & V & X & Z \end{bmatrix}$$

Substitution Rules for Playfair

Scan plaintext in pairs:

1. Same row: Replace with letters to the right (circular).
2. Same column: Replace with letters below (circular).
3. Otherwise: Replace with letters in same row but in column of the other letter.

Insert filler letter (e.g., 'x') between repeating letters.

Security of Playfair Cipher

- Thought unbreakable for decades.
- Used by British Army in WW1, U.S. Army in WW2.
- But easily broken.
- Alters frequencies but not sufficiently.
- Cryptanalysis aided by reversible digrams: if AB $\rightarrow$ XY, then BA $\rightarrow$ YX.

Hill Cipher

- Mathematical approach to multi-letter substitution.
- Assign integers 0-25 to letters 'a'-'z'.
- Encryption key: a 3×3 matrix K of integers.
- Transform three letters at a time:

$$\vec{C} = [K]\vec{P} \mod 26$$

- Decryption:

$$\vec{P} = [K^{-1}]\vec{C} \mod 26$$

Security of Hill Cipher

- Very secure against ciphertext-only attacks (large key space).
- Zero security when plaintext-ciphertext pairs are known.
- Key matrix can be easily calculated from known pairs.

Vigenere Cipher

- Polyalphabetic cipher: substitution rule changes with each character.
- Align encryption key with plaintext (repeat if necessary).
- Each key letter denotes a shifted Caesar cipher.
- Example:
  - key: abracadabra...
  - plaintext: canyoumeetmeatmidnightihavethegoods
  - ciphertext: CBEYQUPEFKMEBK...

Security of Vigenere Cipher

- Multiple ciphertext letters for each plaintext letter masks frequencies.
- Longer key $\rightarrow$ better masking.
- Ideal: key as long as plaintext, purely random $\rightarrow$ "Random polyalphabetic".
- Breaking method: Kasiski Examination.
- Find repeated sequences in ciphertext, use distances to guess key length.

Enigma Machine

- Historically famous polyalphabetic cipher.
- Used by German military in WW2.
- Rotor machines implement polyalphabetic ciphers.
- Broken partly due to predictable message beginnings (e.g., “Heil Hitler!”).

Transposition Cipher

- Pure permutation cipher.
- Write plaintext in rows of a matrix, read ciphertext along columns in key order.
- Example:
  - key: 2 5 3 1 6 4
  - plaintext: meetme atmid night forth egood diesxy
  - ciphertext: ETGTTMDFGXEMHHEMATRDENOOYTTTES
- More secure with multiple rounds.

Differential XORing

- Based on XOR properties:

$$[A \oplus B] \oplus C = A \oplus [B \oplus C]$$
$$A \oplus A = 0$$
$$A \oplus 0 = A$$

- Differential XORing: output for each block depends on previous block.
- Destroys repetitive patterns.

Encryption Script (Python)

Decryption Script (Python)

Perl Versions

- Perl scripts provided for encryption and decryption.
- Similar logic to Python versions.
- Use `Algorithm::BitVector` module.

Homework Problems 1-4

1. Manually construct Base64 for “hello\njello” $\rightarrow$ “advsbd8KamVsbd8=”. What is ‘=’ for?
2. A text file with “hello” has 6 bytes (including \n). Why? How to prevent?
3. Write script to XOR two hex/decimal numbers using BitVector.
4. Write Base64 encoding script (half dozen lines with BitVector).

Homework Problems 5-8

5. All classical ciphers are symmetric key. What does that mean?
6. What are the two building blocks of classical ciphers?
7. True or false: Larger key space $\rightarrow$ more secure? Justify.
8. Example of cipher with large key space, simple algorithm, poor security.

Homework Problems 9-12

9. Difference between monoalphabetic and polyalphabetic ciphers.
10. Main security flaw in Hill cipher.
11. Why is Vigenere more secure than Playfair?
12. After encryption/decryption, `diff` says binary files differ but `cat` shows identical content. Why?

Homework Problems 13-15

13. Write `hist.pl` or `hist.py` for letter frequency histogram.
14. Write `poly_cipher.pl` or `poly_cipher.py` for Vigenere cipher.
15. Cryptanalysis: Break encryption from Lecture 2 scripts (brute-force with BLOCKSIZE=16).

Credits

- Letter frequency data and digram table from:
  `http://jnicholl.org/Cryptanalysis/Data/EnglishData.php`