Cristina Nita-Rotaru

CS355: Cryptography

Lecture 2: Shift cipher, substitution cipher.

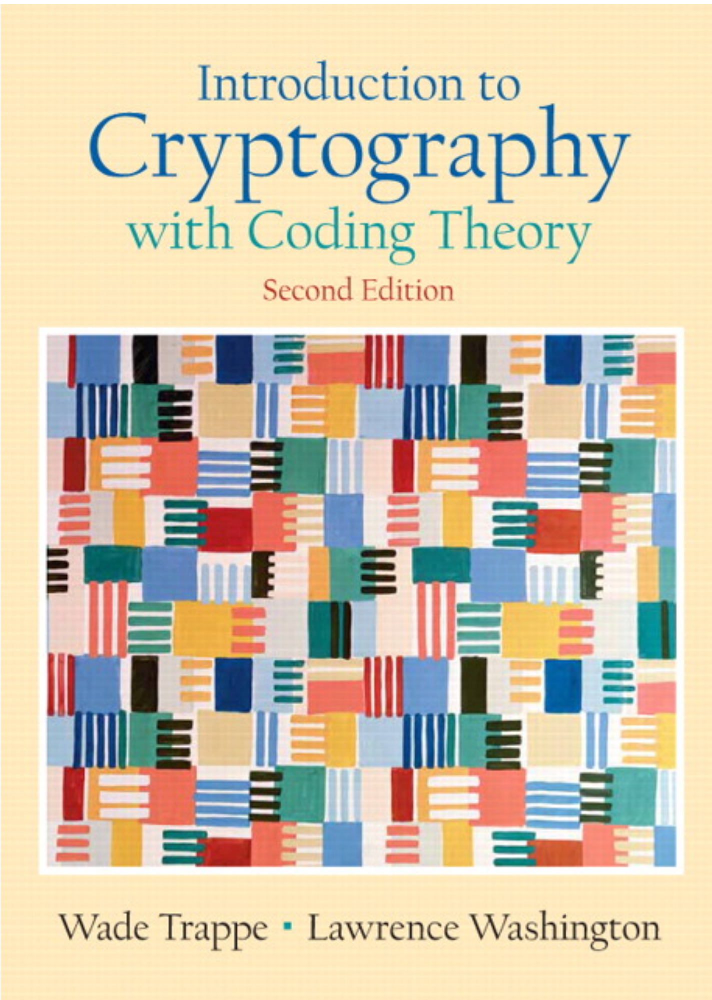
Course overview (1)

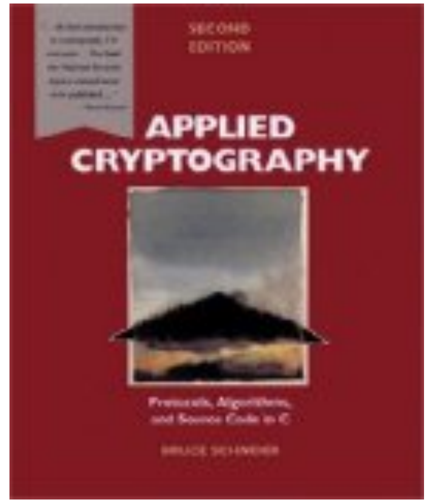
- Concepts and principles of cryptography: security services, attacks and mechanisms.
- Classical cryptographic systems: shift cipher, Vigenere and Vernam ciphers, Jefferson wheel cipher and the Enigma machine.
- Block ciphers: DES, Blowfish, RC5, IDEA, AES.
- Stream ciphers: SEAL, RC4.
- Public-key encryption: RSA, ElGamal, Rabin.
- Data integrity: hash functions, MD5, SHA1, HMAC.
- Digital signatures: RSA, ElGamal, DSA, Schnorr.
- Authentication protocols, data and entity authentication. One time passwords, Lamport's scheme, challenge-response schemes, Kerberos.

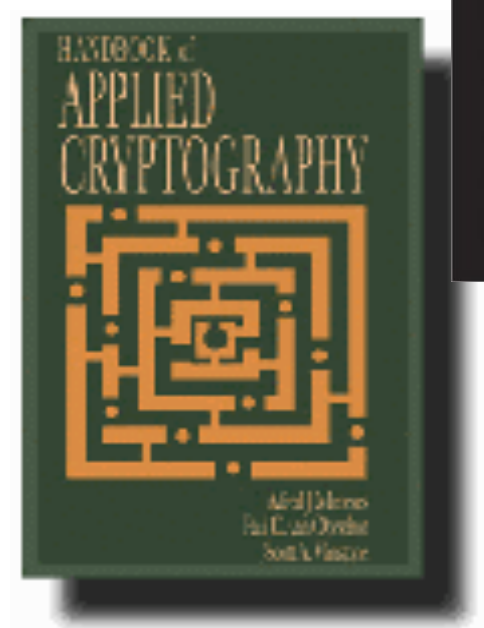
Course overview (2)

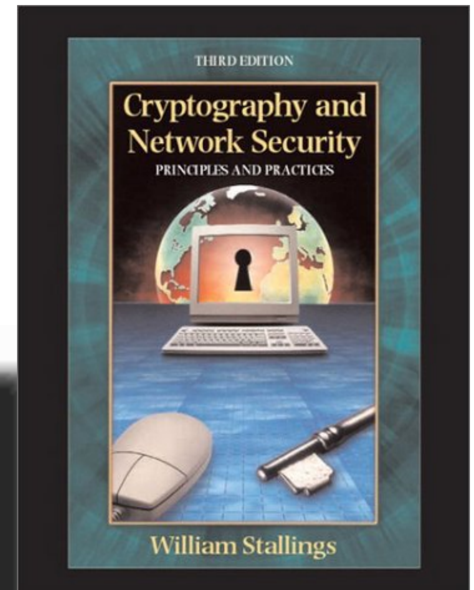
- Key management: two-party key exchange and group key management protocols.
- Digital rights.
- Zero-knowledge proofs.
- Identity-based cryptosystems.
- Notions of threshold cryptography.
- Biometrics

Reference material









Course information

- Meetings
  - MWF 12:30-1:20 LWSN 1106
  - Make up class: Monday at 6 pm in same room, TU 6 pm in LWSN B 146. We will use the lab for exercises in class.
- Professor contact info:
  - Office: 2142J
  - Email: crisn@cs
  - **Office hours: by appointment**
- TA: Denis Ulybyshev
  - Email: dulybysh@purdue.edu
- Class webpage
- http://homes.cerias.purdue.edu/~crisn/courses/cs355_Fall_2012/
- Use Piazza for questions and postings

Class attendance

- Slides will be available before lecture, **class attendance is recommended**
- Email me if you must miss lectures
- If you miss a lecture it is your responsibility to find out what happened in class
- Monitor class website and piazza to know what's going on in the class
- If you can not attend Monday make-up classes, I can meet with you and discuss the missed lecture

Grading policy

Written Assignments (~5)	20%
Projects (~4, 3+final project)	25%
Midterm Exam	20%
Final Exam	25%
Class/Piazza Participation	10%

Extra days

- **Every student has 5 extra days** for all the written assignments and **5 extra days** for individual programming projects
- YOU DECIDE HOW TO USE THEM
- Email me and the TA with name and number of extra days used for an assignment. 1 minute late counts as 1 extra day
- After using your extra days, no late homework or project will be accepted

Homework

- Homework must by TYPED. IF IT'S NOT TYPED, WE DO NOT GRADE IT
- Homework is due by 12:30 by email (me and the TA). Use PDF format. If you plan to use any extra day you have to email me and the TA by 12:30 to let us know
- Homework will be returned in class.
- You **must work alone** on the written homework, write everything in your own words

Exams

- Midterm - proposed date week Oct. 3 in class (before Fall break week)
- Final – check university web page
- We will have a review of the material before midterm and final
- Final covers all the material studied all semester
- Exam problems are similar with homework and test also what you learn through the programming projects
- You will receive a practice final

Programming projects

- Three programming projects
- One final project
- You must work alone
- More information will come
- Purpose of the projects is to offer a glimpse of what is means to design and implement secure protocols
- Submission will be via turn in, make sure you have a CS account of you're not a CS major

Academic integrity

- Purdue University Academic Integrity:

http://www.purdue.edu/ODOS/osrr/conductcode.htm

- Class policy

http://www.cerias.purdue.edu/homes/spaf/cpolicy.html

Phases in cryptography's development

- Cryptography is driven by computing and communication technology
- First stage:
  - paper and ink based scheme
- Second stage:
  - use cryptographic engines
- Third stage, modern cryptography:
  - relying on mathematics and computers
  - information-theoretic security
  - computational security

Shift cipher

- A substitution cipher
- The Key Space:
  - [0 .. 25]
- Encryption given a key K:
  - each letter in the plaintext P is replaced with the K'th letter following corresponding number (shift right)
- Decryption given K:
  - shift left

History: K = 3, Caesar's cipher

Shift cipher: An example

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

P = CRYPTOGRAPHYISFUN

K = 11

C = NCJAVZRCLASJTDQFY

C → 2; 2+11 mod 26 = 13 → N

R → 17; 17+11 mod 26 = 2 → C

…

N → 13; 13+11 mod 26 = 24 → Y

Shift cipher: Cryptanalysis

- Can an attacker find K?
  - YES: exhaustive search, key space is small (<= 26 possible keys).
- Once K is found, very easy to decrypt

Mono-alphabetical substitution cipher

- The key space: all permutations of $\Sigma = \{A, B, C, \ldots, Z\}$
- Encryption given a key (permutation) π:
  - each letter X in the plaintext P is replaced with $\pi(X)$
- Decryption given a key π:
  - each letter Y in the cipherext P is replaced with $\pi^{-1}(Y)$

Example:

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

BECAUSE $\rightarrow$ AZDBJSZ

Cryptanalysis of mono-alphabetical substitution cipher

- Exhaustive search is infeasible
  - key space size is $26! \approx 4*10^{26}$
- Dominates the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then, until …. frequency analysis

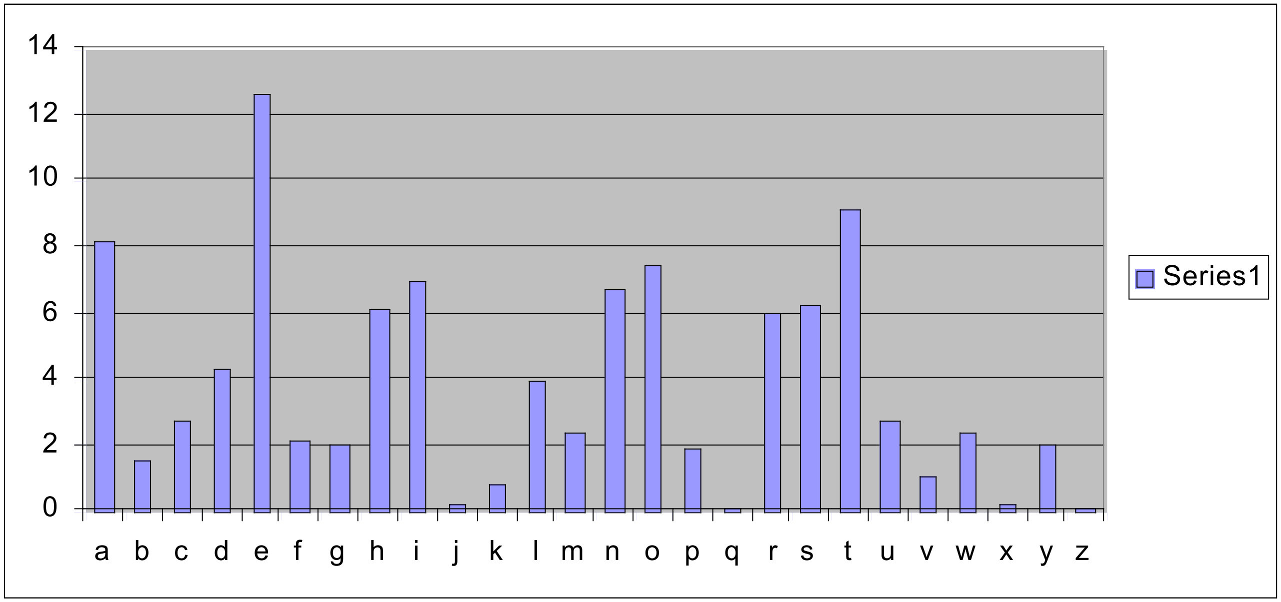
History of frequency analysis

- Discovered by the Arabs
  - Earliest known description of frequency analysis is in a book by the ninth-century scientist Al-Kindi
- Rediscovered or introduced from the Arabs in the Europe during the Renaissance
- Frequency analysis made substitution cipher insecure

Frequency analysis

- Each language has certain features: frequency of letters, or of groups of two or more letters
- Substitution ciphers preserve the language features
- Substitution ciphers are vulnerable to frequency analysis attacks

Frequency of letters in English



Other languages

French

E	16.7%	T	7.3%	C	3.5%	G	1.1%	J	0.3%
S	8.2%	O	5.8%	P	3.0%	Q	1.1%	Y	0.2%
A	8.0%	U	5.5%	M	2.9%	B	0.7%	Z	0.2%
N	7.9%	L	4.9%	V	1.4%	X	0.6%	K	0.1%
I	7.6%	D	3.9%	F	1.2%	H	0.5%	W	0.0%
R	7.4%								

German

E	18.0%	T	5.7%	G	3.2%	F	1.6%	P	0.8%
N	10.6%	D	5.4%	O	2.7%	W	1.5%	J	0.3%
I	8.1%	U	4.6%	C	2.7%	K	1.3%	Y	0.0%
R	7.2%	H	4.1%	M	2.3%	Z	1.1%	X	0.0%
S	6.9%	L	3.3%	B	1.7%	V	0.9%	Q	0.0%
A	6.0%								

Other frequency features of English

- Vowels, which constitute 40 % of plaintext, are often separated by consonants
- Letter A is often found in the beginning of a word or second from last.
- Letter I is often third from the end of a word.
- Letter Q is followed only by U
- And more …

Frequency analysis in action

- The number of different ciphertext characters or combinations are counted to determine the frequency of usage
- The cipher text is examined for patterns, repeated series, and common combinations
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics

Example

- KYZMZ BMZ HGIZ KMZZS BVC KYMZZ HXTPZMS GV KYZ LZBCTP TH CMZBLS GV KZFBS

Solving with frequency analysis

- KYZMZ BMZ HGIZ KMZZS BVC KYMZZ HXTPZMS GV KYZ LZBCTP TH CMZBLS GV KZFBS
- Most frequent: Z = 13; M= 6, K = 5, B = 5
- Guess Z is E
- KYEME BME HGIE KMEES BVC KYMEE HXTPEMS GV KYE LEBCTP TH CMEBLS GV KEFBS

Guess: K is T and Y is H

THEME BME HGIE TMEES BVC THMEE HXTPEMS GV THE LEBCTP TH CMEBLS GV TEFBS

Most frequent, E, T, A , try B is A

THEME AME HGIE TMEES AVC THMEE HXTPEMS GV THE LEACTP TH CMEALS GV TEFAS

Obvious M is R

THERE ARE HGIE TREES AVC THREE HXTPERS GV THE LEACTP TH CREALS GV TEFAS

FINISH UP THE EXERCISE ON YOUR OWN

Solution

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

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

Improve the security of substitution cipher

- Using nulls
  - e.g., using numbers from 1 to 99 as the ciphertext alphabet, some numbers representing nothing are inserted randomly
- Deliberately misspell words
  - e.g., "Thys haz thi ifekkt off diztaughting thi ballans off frikwenseas"
- Homophonic substitution cipher
  - each letter is replaced by a variety of substitutes
- These make frequency analysis more difficult, but not impossible

Take home lessons

- Shift ciphers are easy to break using brute force attacks, they have small key space
- Substitution ciphers preserve language features and are vulnerable to frequency analysis attacks