Introduction to Modern Cryptography

Lecture 1, November 1st, 2016

Instructor: Benny Chor
Teaching Assistant: Orit Moskovich

School of Computer Science
Tel-Aviv University

Fall Semester, 2016–17
Tuesday 12:00–15:00
Venue: Rosenberg 001

Course site: http://tau-crypto-f16.wikidot.com/

Recommended Prerequisites

- Linear Algebra
- Probability
- Algorithms ("official" prerequisite)
- Computational Models
- "Mathematical Maturity" (most important)

Interested students lacking some prerequisites, pls talk to instructor (soon).

Small Craft Advisory

Parts of the material in this course deal with a variety of mathematical objects. These include, among others

- (finite) groups, rings, and fields.
- greatest common divisor of polynomials and integers.
- irreducible polynomials and prime numbers.
- computer algebra implementations.

The mathematical properties of these objects play a crucial role in many of the subjects studied in this course.

All these notions will be explained during the course. However, if you suspect that digesting them may cause you sea sickness (or any other inconvenience or difficulty), you may consider taking an alternative course.

Administrative Details

- Intended for both 3rd year undergrads and grad students.
- Grade determined by exam (70-80%) and homework (30-20%). In order to pass the course, you must pass the exam.
- Recitation (tirgul) on Wednesday 11-12 or 12-13. Same material.
- Exam on January 30th, 2017 (Moed B on February 27th).
- Exam is closed book except for 2 double sided pages.

Administrative Details (2)

- 4-5 assignments, each with both "dry" and "wet" component (latter involve writing and running short *Python* and *Sage* programs).
- Homework submition in groups of size one or two (but not three or more).
- Submissions are expected on time with the exception of 5 day total accumulative delay.
- If one member of a pair has a valid reason for late submission, the other member is still expected to meet the deadline on his/her own.
- Appeals/missing grade issues (and other HW grading issues) should be sent to the grader: `bdikacs AT gmail.com`
- Office hours (both Benny & Orit): By e-appointment.
- E-mails: benny AT cs.tau.ac.il , orit.mosko AT gmail.com
- Course site: http://tau-crypto-f16.wikidot.com/

Collaboration on Assignments, etc.

- Cases of plagiarism that will be detected will be dealt with severely. (For example, reducing grades for the whole course, not just the relevant assignment, and/or reporting the incident to the appropriate university authority.)
- If we suspect Alice had copied from Bob, both will be regarded as cheaters.

Course Outline (very optimistic)

- Encryption (private and public key systems)
- Elementary algebra (groups, rings, finite fields)
- Elementary number theory
- Authentication and digital signatures
- Cryptographic hash functions
- Randomness and pseudo-randomness
- Secret sharing
- Secure multi party computation
- Zero knowledge
- Blockchains and bitcoin

Another (positive, we believe) side effect of the course is the exposure to symbolic mathematical software (specifically, open source Sage).

Class Notes and Course Site

- About 75% of lectures will be made available on the course site in the form of pdf files (generated using LaTeXBeamer package).
- The remaining 25%, mostly the number theory and algebra parts, will be given in old fashion style, whiteboard (or even blackboard, depends) presentations. Consequently they will not be available on the course site.
- Announcements, assignments, and the like will be primarily disseminated through the course web site. Please take a look at it often. We will usually not use email for announcements.

Other Introductory Crypto Courses with Online Lectures (a very partial list)

- Mihir Bellare course at University of California, San Diego, 2016.
- Benny Pinkas course at Haifa University, 2006.
- Eli Biham course at the Technion, 2015.
- Dan Boneh (Stanford) Cryptography I course on Coursera.

Bibliography

- Text Books:
  - J. Katz and Y. Lindell, Introduction to Modern Cryptography, Chapman & Hall/CRC Press, 2014 (first edition 2007).
  - D. Stinson, Cryptography Theory and Practice, CRC Press, 2005.
  - V. Shoup, A Computational Introduction to Number Theory and Algebra (Version 1), 2005. Available online at http://www.shoup.net/ntb/ntb-v1.pdf
- Other Relevant Books:
  - M, Bellare and P. Rogaway, Introduction to Modern Cryptography. Text and great slides vailable online at http://www-cse.ucsd.edu/users/mihir/cse207/classnotes.html
  - A. Menezes, P. Van Oorschot, S. Vanstone, Handbook of Applied Cryptography, CRC Press, 2001. Available online at http://www.cacr.math.uwaterloo.ca/hac
  - B. Schneier, Applied Cryptography, John Wiley & Sons, 1996.
  - P. Giblin, Primes and Programming: An Introduction to Number Theory with Computing, Cambridge University Press, 1993.

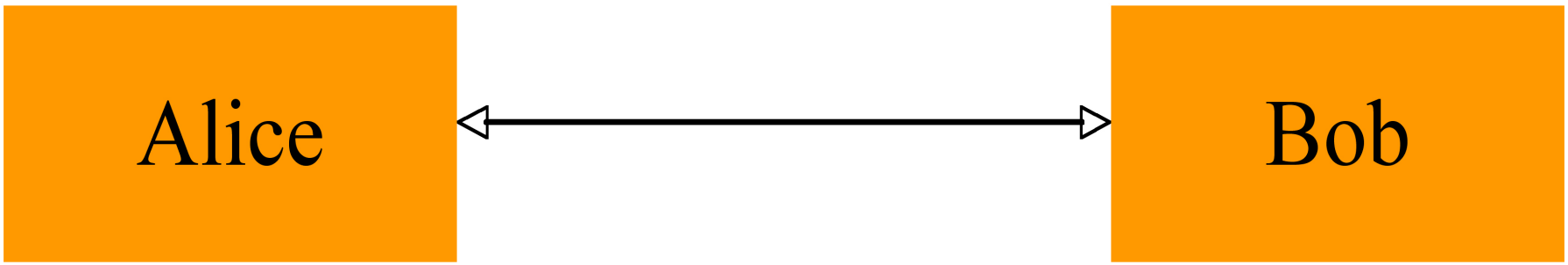
Encryption

Notations and Definitions

- Encryption function (& algorithm): E.
- Decryption function (& algorithm): D.
- Encryption key k_1.
- Decryption key k_2.
- Message space (usually binary strings, either of certain block length or unlimited stream), $\mathcal{M}$.
  Remark: Block length typically tied to key length.
- Consistency requirement: For every message $m \in \mathcal{M}$ and matching pair of keys k_1, k_2: $D_{k_2}(E_{k_1}(m)) = m$.
- So far, no requirement of secrecy.

Communication Model

Let us welcome the two major players in this field, Alice and Bob (claps!).



1. Two parties – Alice and Bob
2. Reliable communication line
3. Shared encryption scheme: E, D, k_1, k_2
4. Goal: send a message m confidentially

Security Goals

There are some different goals we may be after

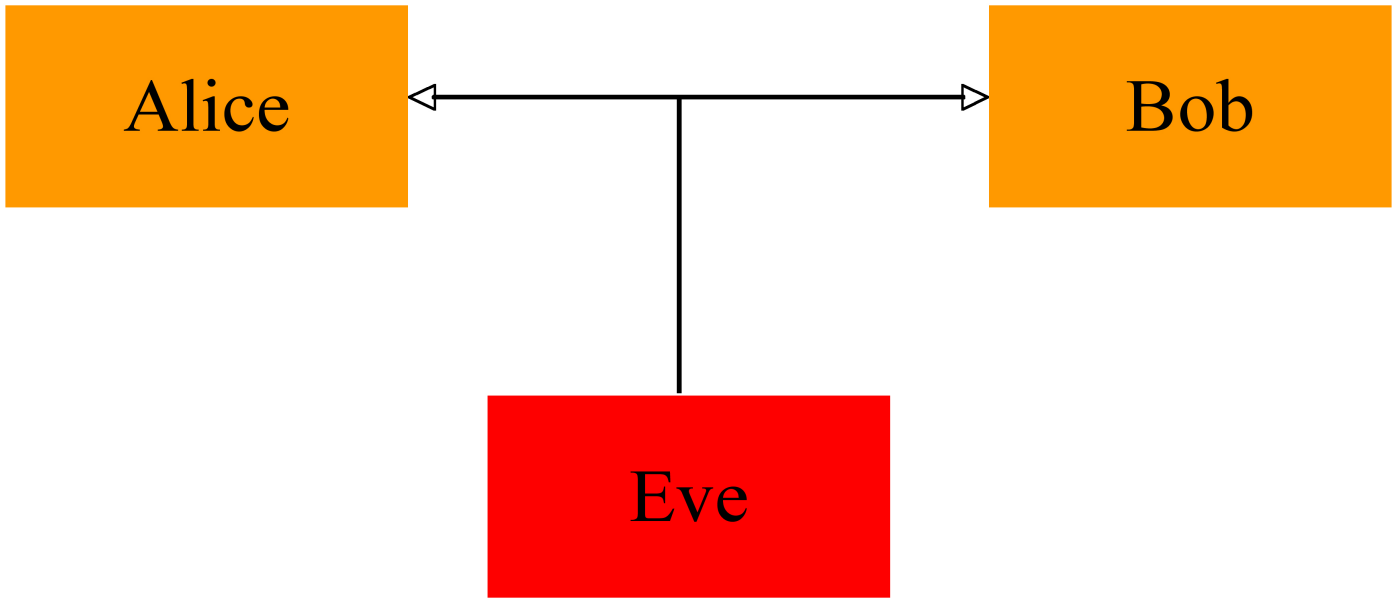
- No adversary can determine m
- No adversary can determine any information about m
- No adversary can determine any meaningful information about m.

Important questions:

- What does the adversary know or seen before?
- What are the adversary's computational resources?

Adversarial Model: Passive Eavesdropper

Enters our third major player, Eve (claps again!).



- Eve attempts to discover information about m
- Eve knows the algorithms E, D
- Eve knows the message space
- Eve has intercepted $E_{k_1}(m)$
- Eve does not know k_1, k_2

Additional Definitions

- Plaintext – the message prior to encryption ("attack at dawn", "sell MSFT at 57.5")
- Ciphertext – the message after encryption ("$\Im\partial\text{Æ}\perp\xi\varepsilon\beta\Xi\Omega\Psi\mathring{A}$", "jhhfo hjklvhgbljhg")
- Symmetric cryptosystem – encryption scheme where $k_1 = k_2$ (classical cryptography)

Encryption: Conceivable Attacks

- Eavesdropping (only ciphertexts known)
- Known plaintext (could sometime infer from reactions)
- Chosen plaintext
- Chosen ciphertext
- Adaptive chosen text attacks
- Physical access
- Physical modification of messages

Examples – (Weak) Symmetric Ciphers

- Shift cipher
- Conclusion – large key space required
  (this can be formalized in information theoretic terms)

- Substitution cipher
- Large key space, still "easy" to break

- Vigenère cipher (poly-alphabetic shift)
- Larger key space, took much longer to break

Next, we will briefly discuss these systems.

Substitution Ciphers

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

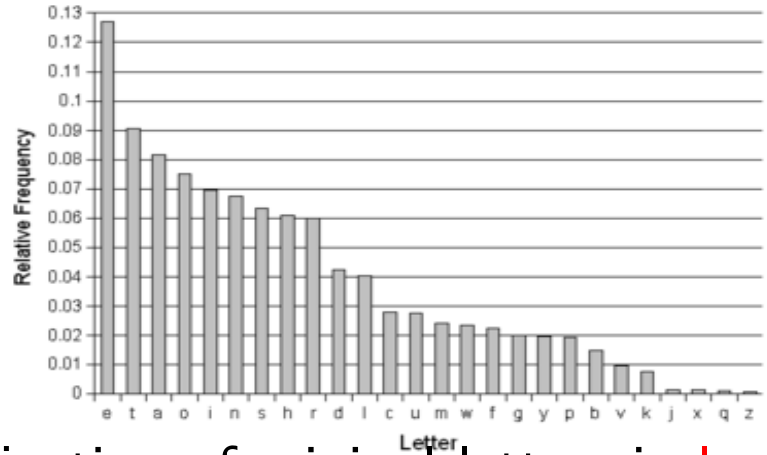
Example:
Plaintext: `attack at dawn`
Ciphertext: `waawoq wa vwmk`

Size of key space is $26! = 403291461126605635584000000 \approx 4 \cdot 10^{27}$.

This is large enough space to prevent exhaustive search for key (at least for old machines, and probably even today). Yet easily breakable due to known (and very non uniform) statistics of single letters, pairs of letters, triplets, etc., in all natural languages.

Natural Languages: Non Uniform Statistics

Distribution of single letters in natural languages' texts is highly non uniform.



This enables identification of original letters in long enough ciphertext, encrypted by a substitution ciphers.

Additional helpful clues follow from the distribution of pairs of letters, triplets, etc., which are also very non uniform. For example, in English q is always followed by u (well, more precisely this is almost always, *e.g.* some of you may have flown Qantas).

In addition, $\Sigma_{i=1}^{26} p_i^2 \approx 0.065$ (for English), while distributions close to uniform have $\Sigma_{i=1}^{26} p_i^2 \approx 1/26 = 0.038$. This discrepancy is useful in breaking some simple ciphers (correctness of tentative key).

Substitution Ciphers (ReVisited)

- Single letter frequencies in natural languages' texts typically have a substantial variance (from one text to another).
- Thus if we simply decipher by assigning "highest frequency letter" in ciphertext to "highest frequency letter" in the language, we will typically not retrieve the plaintext.

- Employing statistics of two letters usually suffice to fully resolve the ambiguities.
- You will get a hands on chance at this in Assignment 1 (on either Hebrew or English text).

Vigenère Cipher

Example (from Katz and Lindell). Secret key is `beads`

t	h	e		m	a	n		a	n	d		t	h	e		w	o	m	a	n
b	e	a		d	s	b		e	a	d		s	b	e		a	d	s	b	d
V	M	F		Q	T	P		F	O	H		M	J	J		X	S	F	C	S

It is an interesting exercise (fully resolved in KL though) to ponder how to break this cipher.

Vigenère Cipher

Suppose length of plaintext is ℓ and length of secret key is k.

- If $\ell \leq k$ then this is exactly an instance of one time pad, so cannot decipher ciphertexts (but system is not too practical).
- Even if $\ell > k$, Vigenère cipher obliterates all "short range" statistics (non-uniformity of pairs of letters, triplets, etc.). Breaking system for moderate values of ℓ/k may still be impossible.
- However, if $\ell \gg k$, then viewing cypher (a kosher variant of cipher, according to the dictionary) as k disjoint shift ciphers allows *efficient* deciphering using single letters statistics.

Formulating Encryption

- A finite message (plaintext) space, $\{M_1, \ldots, M_n\}$.
- Each plaintext is associated with an a-priori probability $p_i = Pr(M_i)$.
- Important: Whether $\mathcal{M}$ contains ancient messages in Sanskrit, plans for a hydrogen bomb, or reconnaissance photos, the a-priori plaintext probabilities, $Pr[plaintext = P]$, are almost never uniform.
- In addition, these probabilities are often hard to estimate.
- A finite key space, $\{K_1, \ldots, K_m\}$.
- Each key is associated with an a-priori probability. These probabilities are often uniform.
- A finite cipher text (encrypted messages) space, $\{C_1, \ldots, C_\ell\}$.

- A simple combinatorial observation: The number of ciphertexts must be at least as large as the number of plaintexts, namely $\ell \geq n$. You will be asked to prove this easy fact as part of HW1.

Formulating Encryption, take 2

- A finite cipher text (encrypted messages) space, $\{C_1, \ldots, C_\ell\}$.
- The probability that a certain ciphertext, C_j, is produced, is determined by the probabilities associated with the plaintexts and the keys: $Pr(C_h) = \sum_{E_{K_i}(M_j)=C_h} Pr(M_j) \cdot Pr(K_i)$.
- Consider the conditional probability that the plaintext equals M_j, given that ciphertext C_h was sent, *i.e.* $Pr(M_j|C_h)$.
  Note that by Bayes' rule
  $Pr(C_h)Pr(M_j|C_h) = Pr(M_j \wedge C_h) = Pr(M_j)Pr(C_h|M_j)$.

Perfect Cipher

- We say that a cipher is perfect if the following holds: Given a ciphertext, C, the probability that $D_{k_2}(C) = M$ for any plaintext M is equal to the apriori probability that M is the plaintext.
- Probability over what?
- Over the key space $\{k_2\}$ and the message space $\mathcal{M}$
- In a probabilistic language:

$$Pr[plaintext = M \mid C] = Pr[plaintext = M]$$

- In daily language: Knowing the ciphertext gives absolutely no information towards knowing the plaintext.

Example – One Time Pad

- Plaintext space – $\{0,1\}^n$
- Key space – $\{0,1\}^n$. The key k is chosen at random and indep. of P.
- The scheme is symmetric, $\oplus$ stands for bit-wise XOR:
  $E_k(P) = C = P \oplus k$
  $D_k(C) = C \oplus k = P$

Pros and Cons, One Time Pad

- Claim: One time pad is a perfect cipher.
- Problem: Size of key space.
- Theorem (Claude Shannon): If a cipher is perfect, then the size of its key space is at least as large as the size of its message space.
- This is bad news. Perfect ciphers are only practical for fairly small message spaces.

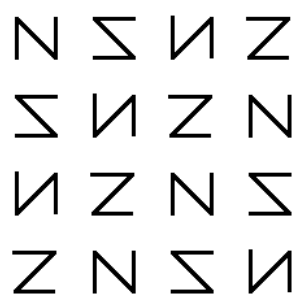
From Perfect to Computational Security

We strive to have perfect (information theoretic) security for encryption and most other cryptographic tasks.
In particular, this would imply resistance to an adversary of unlimited resources.

Our adversaries usually have only limited resources, thus in practice resilience to computationally bound adversaries is good enough.

Group theory (whiteboard presentation)

- Definitions (group's axioms - not necessarily commutative)
- Examples: $(\{0,\ldots,m-1\},+ \bmod m)$ for any positive integer m, $Z_p^* = (\{1,\ldots,p-1\},\cdot \bmod m)$ for any prime p,
  $(S_n,\circ)$: permutations over $\{1,\ldots,n\}$ under composition (non commutative).
  $(SL(n,R),\cdot)$: n-by-n matrices with determinant 1, over the reals, under matrix multiplication (non commutative).
- Cosets aH wrt any subgroup H partition the group G to equivalence classes ($aH = bH$ or $aH \bigcap bH = \emptyset$).
- Lagrange theorem: If $(H,\cdot)$ is a subgroup of finite group $(G,\cdot)$, then $|H|$ divides G.



(image from http://www.ics.uci.edu/∼eppstein/junkyard/sym.html)