Introduction to Modern Cryptography[1]

http://tau-crypto.wikidot.com/

Instructor: Benny Chor
http://www.cs.tau.ac.il/∼bchor

Teaching Assistant: Rani Hod

School of Computer Science
Tel-Aviv University

Fall Semester, 2009–10

[1] Lecture notes 1, October 19, 2009

Travel Advisory

- Based on the feedback forms from 2007-8, most students thought the course was way too hard.
- A few thought it was worth the effort.
- This leaves the lecturer puzzled as to why so many students (approx. 85) insisted on taking it to the (bitter) end...
- Caveat emptor!

Recommended Prerequisites

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

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

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.
- Exam on January 24th, 2010 (Moed B on March 5th).
- 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 *Sage/Maple/Wolfram Alpha* programs).
- Homework submition in groups of size one or two (but not three or more).
- Submissions after the deadlines will not be considered.
- 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.
- Office hours (both Benny & Rani): By e-appointment.
- E-mails: benny AT cs.tau.ac.il , ranihod AT tau.ac.il
- Course site: http://tau-crypto.wikidot.com/

Major Changes from 2007-8 Course

- Added a weekly recitation (but no recitation on first week).
- Symbolic math software: Switched from Maple to Sage.
- This open source package should eliminate bottlenecks of running Maple at a central TAU machine.
- Sage syntax may be simpler than Maple (this remains to be seen).

Collaboration on Assignments, etc.

- Preparing homework assignments independently is a key ingredient for understanding the material (and, consequently, a successful exam :-). So it is highly recommended you and your partner make a serious effort to solve the problems on your own.
- You may collaborate with people from other groups on the problem sets, but your solutions must be written up independently, by you and your partner only.
- You are encouraged to consult online and offline sources for your solutions, but you are (a) expected to give clear cites of your references, and (b) use a write up of your own.
- Recall that Google is a two sided sword.

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.

Bibliography

- Text Books:
  - J. Katz and Y. Lindell, Introduction to Modern Cryptography, Chapman & Hall/CRC Press, 2007. (Its intro chapter is available online.)
  - 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. Available 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.

Course Outline (very optimistic)

- Encryption (private and public key systems)
- Elementary algebra (groups, rings, finite fields)
- Elementary number theory
- Data integrity
- Authentication and identification
- Digital signatures
- Cryptographic hash functions
- Randomness and pseudo-randomness
- Secret sharing
- Cryptographic protocols (e.g. SSL)

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 65% of lectures will be made available on the course site in the form of pdf files (generated using LaTeXBeamer package).
- The remaining 35%, 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)

- Doug Stinson course at Waterloo.
- Mihir Bellare course at University of California, San Diego.
- Benny Pinkas course at Haifa University.
- Eli Biham course at the Technion.

And Finally, Let's Talk Business

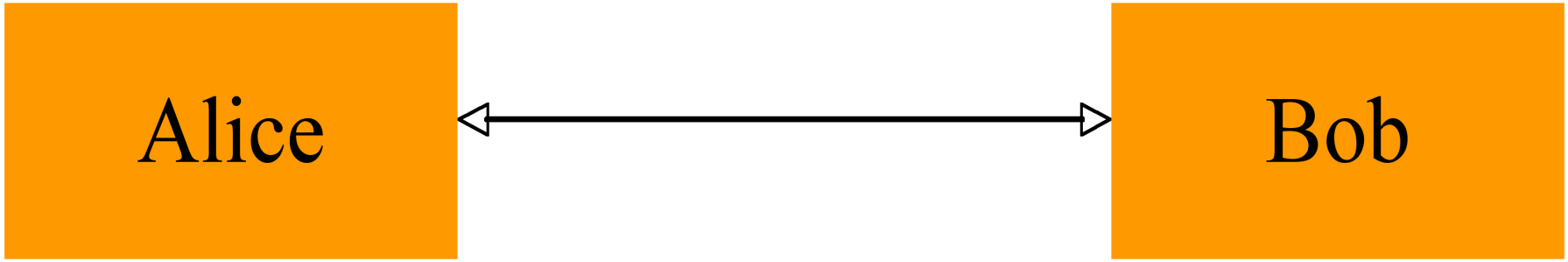
Encryption

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

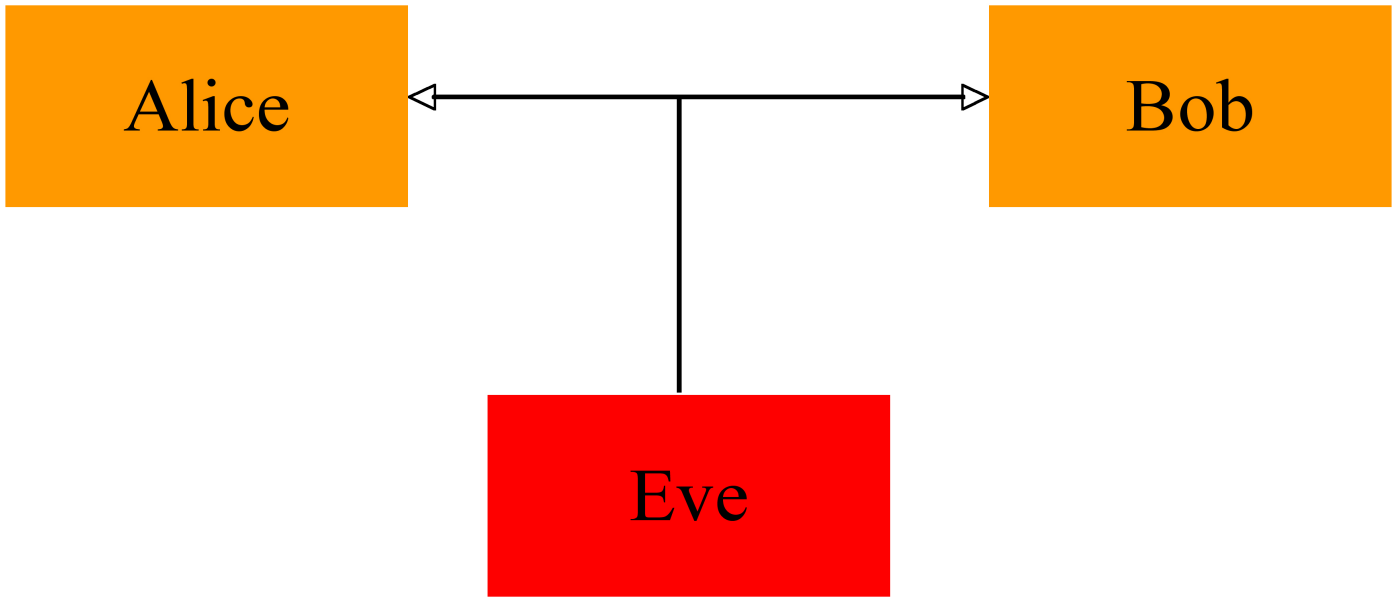
- No adversary can determine m
- No adversary can determine any information about m
- No adversary can determine any meaningful information about m.

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 ("$\Im\partial$Æ$\perp\xi\varepsilon\beta\Xi\Omega\Psi\mathring{A}$", "jhhfo hjklvhgbljhg")
- Symmetric cryptosystem – encryption scheme where $k_1 = k_2$ (classical cryptography)

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

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

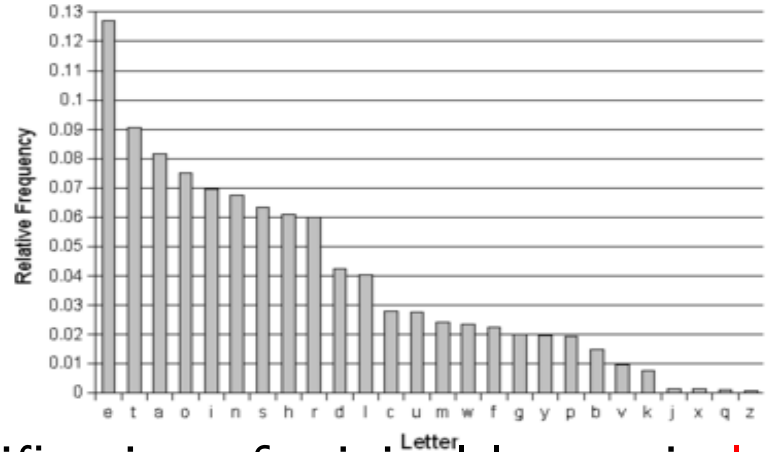
Example:
Plaintext: `attack at dawn`
Ciphertext: `waawoq wa vwmk`

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 a Hebrew text).

Perfect Cipher

- Plaintext (message) space – $\{0,1\}^n$
- 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:

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

- In daily language: Knowing the ciphertext gives absolutely no information towards knowing the plaintext.
- Important: Whether $\mathcal{M}$ contains ancient messages in Sanskrit, plans for a hydrogen bomb, or reconnaissance photos, $Pr[plaintext = P]$ is almost never uniform.

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 (Claud 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.

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 as k disjoint shift ciphers allows *efficient* deciphering using single letters statistics.

Computational Resources

Any serious discussion of cryptography must take into account the computational resources of all parties.
The adversary may have enough information to break a system, but if this requires resources he lacks, the threat is not real.

- Time
- Storage/Memory
- Hardware
- Theoretically: Polynomial vs. non-polynomial (probabilistic) computations
- Practically: 2^{70} steps are (barely) feasible, 2^{100} are not

Modern cryptographical research and modern complexity theory have advanced "hand in hand", often fertilizing the other domain considerably. But this will not be the focus of our course.

Remark: Theory vs. Practice

Following the introduction of public key cryptography, research in the area has to a large extent moved from secretive, military-like organizations to open, academic departments, and (to a lesser extent) to commercial companies.
Starting in the early 80's, theoretical foundations of cryptographic primitives, cryptographic protocols, compositions thereof, etc., were established. Provable security notions led to clearer understanding of many issues, and had far reaching (and highly unexpected) consequences in theoretical Computer Science.

These issues are mostly out of scope for the course, but there is some controversy around them. Neal Koblitz (Univ. of Wahington) and Alfred Menezes (Waterloo) published the article "Another Look at Provable Security", which criticizes several typical provable security results in modern cryptography. In his essay "On Post-Modern Cryptography", Oded Goldreich (Weizmann Inst.) responds to the Koblitz and Menezes arguments.

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