Cryptography Part 1 **Fall 2017**

Lecture 1: Introduction to Cryptography

Lecturer: Dan Boneh *Scribe: Lily Li*

1.1 Introduction

This is just an introductory video going over various applications of Cryptography — of which there are many — and security.

I found the Secure Socket Layer (aka TLS) used by HTML to be of particular interest. This protocol consists of two parts.

1. Handshake Protocol: which establish a shared secret key k using public-key cryptography.
2. Record Layer: one the secret key has been established, transmitting data using the key. The key here is confidentiality and integrity.

The first part will be the focus later on in the course. For now, we focus on the second part. For this purpose we will first need to know a bit about symmetric encryption.

Definition 1.1 ***Symmetric encryption*** *takes two* publicly known *encryption algorithms E and D. If Alice wishes to send a message m to Bob, she encrypts m using E and the secret key k producing the cypher test $E(k, m) = c$. Then Bob decrypts the cypher text c using D and the secret key such that $D(k, c) = m$.*

There are two variants of symmetric encryption: the **single use key** and **multi-use key**. As their names implies, the former only uses k to encrypt one piece of information while the later uses the same k to encrypt multiple files. As it turns out the later needs extra machinery to ensure the security of the encrypted data.

Definition 1.2 *A **secure multi-party computation** is sort of a game between n parties. Each person i chooses some value x_i and everyone wants to know the value of function $f(x_1, ..., x_n)$ for some function f without learning any other information.*

Note that elections are private auctions are examples of such computations.

As it turns out that: any computation that can be performed with a trusted authority can also be performed without. Thus it is possible to perform a secure multi-party computation by only passing data between the interested parties.

Security proofs in cryptography will take place in three steps. Together these amount to something similar to reductions in complexity.

1. Specify threat model: what will the attacker do and what is their goal?
2. Propose a construction: or define the protocol.
3. Prove that breaking the construction amounts to solving an underlying *hard problem*.

1.2 Discrete Probability

The following is going to be a brief overview of the important concepts in discrete probability. For more information please refer to:

`https://en.wikibooks.org/wiki/High_School_Mathematics_Extensions/Discrete_Probability`

Let the universe $\mathcal{U}$ be a finite set, usually $U = \{0,1\}^n$. Further, a **probability distribution** P over $\mathcal{U}$ is a function $P : \mathcal{U} \to \{0,1\}$ such that

$$\sum_{x \in \mathcal{U}} P(x) = 1$$

Definition 1.3 ***Union Bound****: given two events A_1 and A_2 (which are subsets of some universe):*

$$\mathsf{Pr}[A_1 \cup A_2] \leq \mathsf{Pr}[A_1] + \mathsf{Pr}[A_2].$$

Equality occurs when A_1 and A_2 are disjoint.

Definition 1.4 *Two events A and B are* ***independent*** *if*

$$\mathsf{Pr}[A \text{ and } B] = \mathsf{Pr}[A] \cdot \mathsf{Pr}[B]$$

Theorem 1.5 *Let Y be a random variable over $\{0,1\}^n$ and X be an independent uniform variable on $\{0,1\}^n$. Then $Z = Y \oplus X$ is a uniform variable on $\{0,1\}$.*

Proof: First consider the case where $n = 1$. Construct the truth table for the distribution. This idea can be extended for $n > 1$. ■

1.3 Stream Ciphers

Definition 1.6 *A* ***cipher*** *is a pair of efficient algorithms (E, D) defined over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$, $\mathcal{K}$ is the set of all possible keys, $\mathcal{M}$ is the set of all possible message and $\mathcal{C}$ is the set of all cipher texts, such that $E : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$, $D : \mathcal{K} \times \mathcal{C} \to \mathcal{M}$ and the consistency equation holds:*

$$\forall m \in \mathcal{M}, k \in \mathcal{K} : D(k, E(k, m)) = m$$

1.3.1 One Time Pad

This is the first instance of a provably secure cipher. The only down side is that the required key is the same length as the message.

Definition 1.7 *A cipher (E, D) over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$ has* ***perfect secrecy*** *if $\forall m_0, m_1 \in \mathcal{M}$ with $|m_0| = |m_1|$ and $\forall c \in \mathcal{C}$*

$$\mathsf{Pr}[E(k, m_0) = c] = \mathsf{Pr}[E(k, m_1) = c]$$

Where k is uniform in $\mathcal{K}$, denoted $k \xleftarrow{r} \mathcal{K}$.

For the One Time Pad (OTP), $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0,1\}^n$. The key is a random n bit string. Then $c = D(k,m) = E(k,m) = k \oplus m$. It is easy to see that the consistency equation holds since

$$D(k, E(k,m) = D(k, k \oplus m) = k \oplus (k \oplus m) = 0 \oplus m = m.$$

Lemma 1.8 *OTP has perfect secrecy.*

Proof: Suppose $E(k,m) = c$. Then $k \oplus m = c$ so $k = c \oplus m$. Thus there is only one $k \in \mathcal{K}$ such that $E(k,m) = c$ for all m, c. $\mathsf{Pr}[E(k,m_0) = c]$ is a constant for messages so OTP is secure. ■

From the lemma we know that a cipher text only attack can never break the one time pad. Unfortunately, Shannon proved that any cipher which has perfect secrecy must have the length of the key at least the length of the message.

The stream ciphers seek to make the OTP more practical by using a pseudorandom, instead of a random, key. Generally a pseudo random generator (PRG) is a function $G : \{0,1\}^s \to \{0,1\}^n$ such that $n \gg s$. Note that G is deterministic and efficiently computed.

Definition 1.9 *A PRG $G : K \to \{0,1\}^n$ is **predictable** if*

$$\exists A, 1 \le i \le n-1 : \underset{k \xleftarrow{R} K}{\mathsf{Pr}} [A(G(k))|_{1,\ldots,i} = G(x)|_{i+1}] \ge \frac{1}{2} + \epsilon$$

where A is some efficiently computable algorithm and ϵ is non-negligible.

Note: definitions of negligible and non-negligible depends on the community. In practice $\epsilon \ge 1/2^{30}$ is non-negligible while $\epsilon \le 1/2^{80}$ is negligible. However, theoretically, $\epsilon : \mathbb{Z}_{\ge 0} \to \mathbb{R}_{\ge 0}$ is a function which is non-negligible if $\exists d : \epsilon(\lambda) \ge 1/\lambda^d$ infinitely often.

1.3.2 Attacks on the Stream Cipher

1. Two time pad: if the same pseudo-random key is used more than once to encrypt messages it becomes insecure. Suppose that $c_1 = m_1 \oplus PRG(k)$ and $c_2 = m_2 \oplus PRG(k)$. Then $c_1 \oplus c_2 = m_1 \oplus c_2$. Since English has enough redundancy, we can decipher the x-or of two plain-text messages.
2. Lacks Integrity (malleable): if an attacker intercepts the cipher text on its way to the intended target, it can be modified in a predictable and undetectable way.

1.4 PRG Security Definitions

Stream ciphers cannot have perfect secrecy since the length of the key is shorter than the length of the message. Thus we need another definition of security to analyze stream ciphers.

Definition 1.10 *A **statistical test** on $\{0,1\}^n$ is an algorithm A such that $A(x)$ outputs 0 if x is not random and 1 if x is random.*

Consider the following statistical test:

$$A(x) = 1 \iff \textit{max-run-of-0}(x) \le 10 \cdot \log_2(n)$$

this test checks to make sure that the runs of zeros is not too long. Note however, if a run of ones is long the test will still say that it is random.

Next we need a way to evaluate the quality of a statistical test.

Definition 1.11 *Let $G : K \to \{0,1\}^n$ be a PRG and A a statistic test on $\{0,1\}^n$. The advantage of A on G, denoted* Adv *is*

$$\mathsf{Adv}_{PRG}[A,G] = \left| \Pr_{k \xleftarrow{R} \mathcal{K}}[A(G(k)) = 1] - \Pr_{r \xleftarrow{R} \{0,1\}^n}[A(r) = 1] \right| \in [0,1].$$

Observe that if Adv *is closer to one, then A can distinguish G from a truly random sequence and if* Adv *is closer to zero then A cannot.*

Finally we can define a secure PRG.

Definition 1.12 *$G : K \to \{0,1\}^n$ is a* ***secure PRG*** *if for all efficient statistic test A, $\mathsf{Adv}_{PRG}[A,G]$ is negligible.*

Unfortunately it is not know if there are any probably secure PRGs, as per the above definition — since a prove of a secure PRG implies that $\mathsf{P} \neq \mathsf{NP}$.

Theorem 1.13 *A PRG is secure if and only if it is unpredictable.*

Proof: We will prove the forward direction using contra-position. Suppose that a PRG G is predictable. Then there exists a predictor algorithm A that given the first i bits of the input, A can predict the $i+1$th bit with probability bounded away from $1/2$ by a non-negligible amount ϵ. Using this algorithm, we will design a statistical test to distinguish G from a truly random sequence. We will simply output one if the $i+1$th bit is the same as the prediction of A. Since the probability of getting the value right for a random string is $1/2$, $\mathsf{Adv}_{PRG}[A,G] = \epsilon$.

The converse is Yao's theorem. ■

Definition 1.14 *Consider two distributions P_1, P_2 over $\{0,1\}^n$. P_1, and P_2 are computationally indistinguishable, denoted $P_1 \approx_p P_2$ if for all efficient statistic tests A,*

$$\left| \Pr_{x \xleftarrow{P_1} \{0,1\}^n}[A(x) = 1] - \Pr_{x \xleftarrow{P_2} \{0,1\}^n}[A(x) = 1] \right| < \textit{ negligible}$$

1.5 Semantic Security

We define semantic security, for a one-time key, in-terms of two experiments one for each $b \in \{0,1\}$. We have two parties in the experiment: the challenger and the adversary (these are analogs of the PRG and statistic test respectively). The adversary passes the challenger two messages $m_0, m_1 \in M$ with $|m_0| = |m_1|$. The challenger choses a key $k \leftarrow K$ and passes the encrypted message $c = E(k, m_b)$ back to the adversary. The adversary outputs $b' \in \{0,1\}$ which is guess of the index of the encrypted message.

Let $W_b =$ [event that $EXP(b) = 1$] where $EXP(0)$ sets $b = 0$ and $EXP(1)$ sets $b = 1$. Then the advantage is defined as

$$\mathsf{Adv}_{ss}[A,E] = |\mathsf{Pr}[W_0] - \mathsf{Pr}[W_1]|$$

Definition 1.15 *The encryption scheme E is **semantically secure** if for all efficient adversaries A, $\mathsf{Adv}_{SS}[A, E]$ is negligible. Or more symbolically, for all explicit $m_0, m_1 \in M : \{E(k, m_0)\} \approx_p \{E(k, m_1)\}$*

Theorem 1.16 *Stream ciphers are semantically secure. More formally: given a secure PRG $G : K \to \{0,1\}^n$, the stream cipher E derived from G is semantically secure.*

Proof: The proof is by contra-position. Suppose that there exists a semantic adversary A which breaks E. Then we will use A to construct a PRG adversary B such that

$$\mathsf{Adv}_{SS}[A, E] \leq 2 \cdot \mathsf{Adv}_{PRG}[B, G]$$

Since B is secure and thus $\mathsf{Adv}_{PRG}[B, G]$ is negligible then $\mathsf{Adv}_{SS}[A, E]$ must be negligible as well.

Formally we recall the definition of W_0 and W_1 which are the events that the adversary outputs 1 when it receives m_0 and m_1 encrypted with the pseudo-random key respectively. Further let R_0 and R_1 be similar events now with the messages encrypted with a truly random string.

First note that $|\mathsf{Pr}[R_0] - \mathsf{Pr}[R_1]| = \mathsf{Adv}_{SS}(A, OTP) = 0$. That is to say the advantage of the adversary A against the OTP is zero.

Next, we claim that there exists a statistic test B such that $|\mathsf{Pr}[W_b] - \mathsf{Pr}[R_b]| = \mathsf{Adv}_{PRG}[B, G]$ for $b = 0, 1$. Since $\mathsf{Pr}[R_1] = \mathsf{Pr}[R_0]$ from the above, we have that

$$\mathsf{Adv}_{SS}[A, E] = |\mathsf{Pr}[W_0] - \mathsf{Pr}[W_1]| \leq 2 \cdot |\mathsf{Pr}[W_b] - \mathsf{Pr}[R_b]| = 2 \cdot \mathsf{Adv}_{PRG}[B, G]$$

The statistical test B is actually dead simple to construct. We use the adversary A as a subroutine in B. B takes as input a random string $y \in \{0,1\}^n$ and the two messages m_0 and m_1 output by A. B sets $c = m_0 \oplus y$ and feeds it to A outputting the output of A.

It remains to calculate the advantage of B. Reading off the advantage from the definition

$$\mathsf{Adv}_{PRG}[B, G] = \left| \underset{r \leftarrow \{0,1\}^n}{\mathsf{Pr}}[B(r) = 1] - \underset{k \leftarrow \mathcal{K}}{\mathsf{Pr}}[B(G(k)) = 1] \right| = |\mathsf{Pr}[R_0] - \mathsf{Pr}[W_0]|$$

we have exactly what we wanted. ■