What will we cover in this lecture?

- One-way functions
- Average-case complexity
Most forms of cryptography depend on $P \neq NP$

- Whenever there is a private key with the property that an encoded message can be decoded efficiently with the private key, this is an NP problem.

- So if $P = NP$, breaking the cryptographic scheme can be done in polynomial time.
One-way functions (OWFs)

Definition (one-way functions)

A polynomial-time computable function \( f : \{0, 1\}^* \rightarrow \{0, 1\}^* \) is a one-way function if for every polynomial-time probabilistic TM \( M \) there is a negligible function \( \epsilon : \mathbb{N} \rightarrow [0, 1] \) such that for every \( n \in \mathbb{N} \):

\[
P_{x \in \{0,1\}^n} \left[ M(y) = x' \text{ such that } f(x') = y \right] < \epsilon(n)
\]

where a function \( \epsilon : \mathbb{N} \rightarrow [0, 1] \) is negligible if \( \epsilon(n) = \frac{1}{n^{\omega(1)}} \), that is, for every \( c \) and sufficiently large \( n \), \( \epsilon(n) < \frac{1}{n^c} \).

- Conjecture: there exist one-way functions (implying \( P \neq NP \))
- OWFs can be used to create private-key cryptography
Consider the function $f_U$ that is defined as follows. If there exists any OWF $f$, then $f_U$ is also an OWF.

- Treat the input $x$ as a list $x_1, \ldots, x_{\log n}$ of $n / \log n$ bit long strings.
- Output $M_{1}^{n^2}(x_1) \ldots M_{\log n}^{n^2}(x_{\log n})$.
- Here $M_{i}^{t}(y)$ denotes the output that the $i$th TM $M_i$ gives on input $y$, or $0^{|y|}$ if $M_i$ takes more than $t$ steps on input $y$.

Main idea:

- If there is an OWF, then there is one that runs in time $n^2$—using padding.
- The function that concatenates the output of several (polynomial-time computable) functions $f_1, \ldots, f_k$ is an OWF if and only if at least one of $f_1, \ldots, f_k$ is an OWF.
- Whenever $n$ gets large enough, there is some $M_i$ that is an OWF that runs in time at most $n^2$, and so therefore is $f_U$. 
Definition

An encryption scheme is a pair \((E, D)\) of algorithms, each taking a key \(k\) and a message \(x\), such that \(D_k(E_k(x)) = x\).

The scheme is perfectly secret, for messages of length \(m\) and keys of length \(n\), if for every pair \(x, x' \in \{0, 1\}^m\) of messages, the distributions \(E_{U_n}(x)\) and \(E_{U_n}(x')\) are identical.

The scheme is computationally secure if for every probabilistic polynomial-time algorithm \(A\), there is a negligible function \(\epsilon : \mathbb{N} \to [0, 1]\) such that

\[
P \left( k \in_R \{0, 1\}^n \right. x \in_R \{0, 1\}^m \left[ A(E_k(x)) = (i, b) \text{ s.t. } x_i = b \right] < \frac{1}{2} + \epsilon(n).
\]

- Suppose that OWFs exist. Then for every \(c \in \mathbb{N}\) there exists a computationally secure encryption scheme \((E, D)\) using \(n\)-length keys for \(n^c\)-length messages.
A problem $L \subseteq \{0, 1\}^*$ can be solved in worst-case running time $T(n)$ if there exists an algorithm $A$ that solves $L$ and that halts within time $T(|x|)$ for each $x \in \{0, 1\}^*$.

In other words, the worst-case running time $T(n)$ is the maximum of the running times for all inputs of size $n$. 
A *distributional problem* $\langle L, D \rangle$ consists of a language $L \subseteq \{0, 1\}^*$ and a sequence $D = \{D_n\}_{n \in \mathbb{N}}$ of probability distributions, where each $D_n$ is a probability distribution over $\{0, 1\}^n$. 
The class \( \text{distP} / \text{avgP} \)

**Definition (distP)**

\( \langle L, \mathcal{D} \rangle \) is in the class \( \text{distP} \) (also called: \( \text{avgP} \)) if there exists a deterministic TM \( M \) that decides \( L \) and a constant \( \epsilon > 0 \) such that for all \( n \in \mathbb{N} \):

\[
\mathbb{E}_{x \in R \mathcal{D}_n} \left[ \text{time}_M(x)^\epsilon \right] \text{ is } O(n).
\]

- The \( \epsilon \) is there for technical reasons—to invert a polynomial to \( O(n) \).
Definition (P-computable distributions)

A sequence $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ of distributions is \textit{P-computable} if there exists a polynomial-time TM that, given $x \in \{0, 1\}^n$, computes:

$$
\mu_{D_n}(x) = \sum_{y \in \{0, 1\}^n, y \leq x} \mathbb{P}[y],
$$

where $y \leq x$ if the number represented by the binary string $y$ is at most the number represented by the binary string $x$. 

Definition (P-samplable distributions)

A sequence $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ of distributions is P-samplable if there exists a polynomial-time probabilistic TM $M$ such that for each $n \in \mathbb{N}$, the random variables $M(1^n)$ and $D_n$ are equally distributed.
The class $\text{distNP}$ and $\text{sampNP}$

<table>
<thead>
<tr>
<th>Definition (distNP)</th>
<th>A problem $\langle L, D \rangle$ is in $\text{distNP}$ if $L \in \text{NP}$ and $D$ is P-computable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition (sampNP)</td>
<td>A problem $\langle L, D \rangle$ is in $\text{sampNP}$ if $L \in \text{NP}$ and $D$ is P-samplable.</td>
</tr>
</tbody>
</table>

- The questions “$\text{distNP} \neq \text{distP}$” and “$\text{sampNP} \neq \text{distP}$” are average-case analogues of the question “$\text{NP} \neq \text{P}$”
Definition (zero-error heuristics)

A zero-error heuristic $H$ for $f$ is a probabilistic polynomial-time algorithm that for each $x \in \{0, 1\}^*$, when given $x$ as input, it outputs either $f(x)$ or “?”.

Definition (zero-error average-case hardness)

Let $\alpha : \mathbb{N} \rightarrow [0, 1]$ be a function. A function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is zero-error $\alpha$-hard-on-average if for all zero-error heuristics $H$ for $f$ and all sufficiently large $n \in \mathbb{N}$, it holds that:

$$\mathbb{P}_{x \in \{0, 1\}^n} [ H(x) = "?" ] \geq \alpha(n).$$
One-way functions

Average-case complexity