Information Theory

Definition of perfect secret security, why you need as many keys as plaintexts to have perfect security. Definition of entropy, and proof of some of the inequalities properties it satisfies. Unicity distance (but be careful, this may take a lot of time, so test this beforehand)

The "Elevator Pitch" (Synthesis)

> Explain **perfect secrecy** to a peer in 3 sentences. Focus on WHY we use it in cryptography, not just how.

Core Vocabulary & Syntax (Total Recall)

Shift Cypher

> The Shift Cipher (Caesar substitution) works on an alphabet containing n symbols that we number from 0 to $n-1$, corresponding to the elements in $\mathbb{Z}_n$. For instance, in the English alphabet, a would correspond to 0, b to 1...

> We set $K = P = C = \mathbb{Z}_n$ and the system works as follows:
> Key Generation: Output k chosen uniformly at random from $\mathcal{K}$.
> Encryption Set: $E_k(x) = (x + k) \mod n$
> Decryption Set: $D_k(y) = (y - k) \mod n$

> This cipher was used by encrypting each letter in a message using the same key. This clearly leads to a very insecure solution, but it's not necessarily the case if we vary the key more often.

Perfect security

> A cryptosystem has perfect security if for all $x \in P$ and $y \in C$, it holds that $P[x|y] = P[x]$ The definition simply says that seeing the ciphertext does not help the adversary at all: the distribution on the plaintext remains the same whether you are given the ciphertext or not.

> Using Bayes' theorem ($P[A|B] = \frac{P[A,B]}{P[B]} = \frac{P[B|A]P[A]}{P[B]}$), we can see that the definition implies that **A cryptosystem has perfect security if and only if** $P[y|x] = P[y]$ This is often easier to test (~ if we do not know the key, we cannot predict what the ciphertext will be)

Entropy

> Shannon's notion of Entropy gives us a way to assign a number to a random variable with a given probability distribution. This number is called the uncertainty or the entropy of the variable, and it corresponds exactly to the numbers of bits we need to encode the result. If an

> event A occurs with probability p and you are told that A occurred, then you have learned $log_2(\frac{1}{p})$ bits of information.

> Let X be a random variable that takes values $x_1, \dots, x_n$ with probabilities $p_1, \dots, p_n$. Then the entropy of X is defined to be
>
> $$H(X) = \sum_{i=1}^{n} p_i \log_2(1/p_i)$$

> $0 \geq H(X) \geq log_2(n)$. Furthermore, $H(X) = 0$ if and only if one value X has probability 1 (and the others 0). $H(X) = log_2(n)$ if and only if it is uniformly distributed, i.e., all probabilities are $1/n$. Proof: lower bound trivial because of sum, upper bound because log is a concave function and Jensen's inequality is a something. As expected: $H(X)$ is non-negative, the higher it is the more uncertain you are (but there's an upper limit based on how many possibilities there are).

> For any random variables X, Y, we have $H(X, Y) = H(Y) + H(X|Y)$. Furthermore, $H(X|Y) \leq H(X)$ with equality if and only if X and Y are *independent*. (Less otherwise, because you can't become more uncertain with more info.)

> For any cryptosystem with deterministic encryption function, it holds that $H(K|C) = H(K) + H(P) - H(C)$. $H(K|C)$ is the remaining uncertainty about which key was used after observing the ciphertext.
>
> Starting uncertainty: You begin with $H(K)$ bits of uncertainty about the key and $H(P)$ bits about the plaintext.
>
> Information gained: Observing the ciphertext provides $H(C)$ bits of information.
>
> Remaining uncertainty: The equation shows **how much uncertainty about the key remains after seeing the ciphertext**.
>
> **For perfect secrecy, you need** $H(K \mid C) = H(K)$, meaning the ciphertext reveals nothing about the key, which requires $H(K) \geq H(P)$, meaning for perfect security your keyspace must be at least as big as your plaintext space (to have enough entropy).

> What's important that entropy (by design) behaves as intuitively expected from a quantity describing uncertainty of a random variable.

One-time pad

> The one-time pad cryptosystem was invented for secure transmission between teletype machines. It is essentially the Shift cipher using $n = 2$, and following the rule that every key is used only once. Observing that addition and subtraction modulo 2 is the same operation, both equal to the xor-operation. The one-time pad is the most practical perfectly secure cryptosystem: it is very easy to implement and decryption is the same operation as encryption. But of course, it cannot get around the fundamental limitation that the key must be at least long as the message.

Unicity distance

> Unicity distance is the minimum amount of ciphertext needed so that, after trying all possible keys in a brute-force attack, only one decryption makes sense.

> When trying every possible key on some ciphertext, with a short ciphertext, you'll get many **"spurious keys"** - incorrect keys that still produce seemingly valid plaintext. As the ciphertext grows longer, these spurious solutions disappear. The unicity distance is when the expected number of spurious keys drops to zero.

> The unicity distance is calculated as: (assuming uniformly chosen keys) $n_0 \approx \frac{H(K)}{D}$, where D is the redundancy of the language per character.
>
> $$n_0 \approx \frac{log_2(|K|)}{R_L \cdot log(\mid P \mid)}$$
>
> **Redundancy of a language** (R_L) measures "unused" information capacity per character: $R_L = 1 - \frac{H_L}{\log(|P|)}$, where H_L is actual entropy per letter and $\log(|P|)$ is maximum entropy. $H_L \approx 1.25$ bits per letter in English. For English, $R_L \approx 0.75$ (highly redundant)—this lets us understand partial speech and lets cryptanalysts distinguish meaningful plaintext from random strings.
>
> For example, a substitution cipher with a 26-letter alphabet has about 88 bits of key entropy, and English has roughly $3.5 \text{ bits/character}$ redundancy, giving a unicity distance of about 25 characters.

Logic Flow / Mechanism (Process)

Perfect Security Condition

Derive the equivalence from Definition 5.1 using Bayes' formula:

1. Assume $P[x|y] = P[x]$ for all $x \in P$, $y \in C$.
2. [$P[x|y] = \frac{P[x,y]}{P[y]} = \frac{P[y|x]P[x]}{P[y]} = P[x]$] (Bayes' formula).
3. [$\frac{P[y|x]}{P[y]} = 1 \text{ (if x isn't 0)} \Rightarrow P[y|x] = P[y]$].
4. Conclude the cryptosystem has perfect security if and only if $P[y|x] = P[y]$ and $P[x|y] = P[x]$.

Theorem 5.4: Key Size Requirement for Perfect Security

Prove $|K| \geq |C| \geq |P|$:

1. For correct decryption, $|C| \geq |P|$, since [two plaintexts cannot be sent to the same ciphertext].
2. [Given $x \in P$, using different keys every $y \in C$ must be a legit encryption, otherwise the attacker could learn some info from y alone. That means there must exist at least one key for each $x \Rightarrow y$ mapping].
3. Therefore, $|K| \geq |C| \geq |P|$.

Exam Simulation

Oral Exam Question 1 (Presentation, 5 min): "Present and prove Theorem 5.3: why the Shift cipher (one-time use) achieves perfect security. Derive$P[y] = P[y|x]$explicitly."

Oral Exam Question 2 (Presentation + Q&A, 7 min): "Walk me through Theorem 5.4 and 5.5. Prove why perfect security demands $|K| \geq |P|$, and characterize the minimal case. Include the one-time pad as example—what is its encryption/decryption rule?"

Oral Exam Question 3 (Deep Dive, 6 min): "Define entropy per Shannon's convention. Explain its role in cryptography (link to perfect security limitations). Contrast IC for English vs. uniform—derive why IC breaks Vigenère block size."

Source data missing for: Full entropy definition/proof (source truncated at 5.3.1); explicit unicity distance formula/proof; Jensen's inequality application to entropy properties.

Source Map

- CryptographyV6.pdf | Page 39-41 | Covers: Perfect security definition, equivalence lemma, Shift cipher proof, Theorems 5.3-5.5, one-time pad.
- CryptographyV6.pdf | Page 42-43 | Covers: Entropy introduction (partial), Shannon convention.
- CryptographyV6.pdf | Page 30-34 | Covers: Frequency analysis, IC definition/computation for Vigenère.
- CryptographyV6.pdf | Page 38 | Covers: Chapter 5 outline (unicity distance mentioned).