

Chapter 2 | Perfectly-Secret Encryption

Wednesday, April 19, 2023 10:39 AM

Written on

« Wednesday, April 19, 2023

Story

- In this chapter, we look at perfectly secret encryption schemes, i.e. schemes that are secure against adversaries with unbounded computational power.
- The material is more "classical cryptography" than "modern cryptography" because
 - Material developed in the mid-70s and early-80s
 - Constructions don't need unproven assumptions (only use principle 1 and 3 from § 1.4)
 - This approach has inherent limitations (that we prove); therefore good basis to lay groundwork and proceed to the later chapters

§ 2.1 Definitions and Basic Properties

Recall: An encryption scheme is defined by

- three algorithms, Gen, Enc, Dec, and
- specification of a message space \mathcal{M} (where $|\mathcal{M}| > 1$; if there's only one message, don't need to communicate (let alone encrypt))
- Also, \mathcal{K} is the set of keys that can be output by Gen.

Syntax for Enc:

Enc may be a probabilistic algorithm and to emphasise this, we use

$$c \leftarrow \text{Enc}_k(m)$$

to denote the probabilistic process by which the message m is encrypted using key k to give the ciphertext c

When Enc is deterministic, we may emphasise this

$$\text{as } c := \text{Enc}_k(m).$$

- Let \mathcal{C} denote the set of all ciphertexts that can be output by Enc for all choices of messages and for all choices of keys (and also any intrinsic randomness in the algorithm)

Syntax for Dec:

Dec takes as input a key $k \in \mathcal{K}$ and a ciphertext $c \in \mathcal{C}$ and outputs a message $m \in \mathcal{M}$.

Assumption:

Perfectly Correct:

$$\forall k \in \mathcal{K}, m \in \mathcal{M}$$

$$\text{and any ciphertext } c = \text{Enc}_k(m),$$

$$\text{it holds that } D_k(c) = m \text{ w.p. } 1.$$

This implies that we can take Dec to be deterministic without loss of generality.
(because Dec must give the same answer every time it is run)

(Using the notation above)
We write $m := \text{Dec}_k(m)$.

Notation:

- In discussions later, we refer to **probability distributions** over \mathcal{K}, \mathcal{M} & \mathcal{C}
- The distribution over \mathcal{K} = distribution defined by Gen
 - i.e. for $k \in \mathcal{K}$ we let

$$\Pr[K = k]$$

random variable corresponding to the key.

denote the probability that Gen outputs k .

- Similarly, for $m \in \mathcal{M}$, we let

$$\Pr[M = m]$$

denote the probability that the message sent equals m .

Discussion:

- Assuming that there is a distribution over messages (rather than the message being fixed)
 - is meant to model the fact that
 - from the POV of the adversary,
 - different messages may have different probabilities of being sent.

Written on

« Thursday, April 20, 2023

- E.g. The adversary may know the message is either "attack tomorrow" or "don't attack".
The adversary may even know (by other means) that
 $\Pr[\text{"attack tomorrow"}] = 0.7$ and $\Pr[\text{"don't attack"}] = 0.3$

NB: The distributions over \mathcal{K} and \mathcal{M} are independent—
the key and message are chosen independently

Justification:

The distribution of \mathcal{K} is determined by Gen
(i.e. it is fixed by the encryption scheme itself),
while the distribution over messages is determined by
the parties using the encryption scheme.

Notation:

- $\Pr[C = c]$ is used to denote the probability that the cipher text is c ,
given Enc, the distribution over \mathcal{K} and \mathcal{M} .

The Actual Definition

Story: We have introduced all the concepts we need to formally define
the notion of perfect secrecy.
Let's motivate the definition.

Motivation:

- Imagine an adversary knows the distribution over \mathcal{M} .
- Suppose the adversary eavesdrops and learns a ciphertext.
- Ideally, we want the knowledge of this cipher text to have no effect on the knowledge of the adversary about which message was sent i.e. the distribution over \mathcal{M} (for the adversary) should stay the same
- And, this should hold irrespective of the computational power of the adversary

Definition 2.1 *distⁿ over messages doesn't change*

An encryption scheme (Gen, Enc, Dec) over a message space \mathcal{M} is **perfectly secret** if for every probability distribution over \mathcal{M} , every message $m \in \mathcal{M}$, and every ciphertext $c \in \mathcal{C}$ (for which $\Pr[C = c] > 0$), it holds that

$$\Pr[M = m | C = c] = \Pr[M = m]$$

Remarks:

- The requirement that $\Pr[C = c] > 0$ is technical; ensures there's no division by zero when we consider conditional probabilities
- Another interpretation of Definition 2.1 (perfect secrecy): the distribution over messages and cipher texts are independent.

Convention:

For ease of presentation, we restrict to probability distributions over \mathcal{M} and \mathcal{C} that assign non-zero probabilities to all messages and ciphertexts (resp.). This is not a fundamental limitation—the results hold in general.

An equivalent formulation

Story: The following is equivalent to Definition 2.1

Lemma 2.2 *distⁿ over ciphertexts doesn't change*

An encryption scheme (Gen, Enc, Dec) over a message space \mathcal{M} is **perfectly secure** iff for every probability distribution over \mathcal{M} , for every message $m \in \mathcal{M}$ and every ciphertext $c \in \mathcal{C}$, it holds that

$$\Pr[C = c | M = m] = \Pr[C = c].$$

conditions

Proof

Idea: Use Bayes' Theorem.

Conditions \Rightarrow Perfectly Secure

Fix a distribution over \mathcal{M} and select an arbitrary $m \in \mathcal{M}$ and $c \in \mathcal{C}$.

Given: $\Pr[C = c | M = m] = \Pr[C = c]$

NB: $\Pr[C = c | M = m] \cdot \Pr[M = m] = \Pr[M = m]$

Given: $\Pr[C=c | M=m] = \Pr[C=c]$

$$\text{NB: } \frac{\Pr[C=c | M=m] \cdot \Pr[M=m]}{\Pr[C=c]} = \Pr[M=m]$$

$$\text{Recall: } \Pr[A|B] \Pr[B] = \Pr[B|A] \Pr[A]$$

$$\text{NB2: } \Pr[M=m | C=c] \cdot \frac{\Pr[C=c]}{\Pr[C=c]} = \Pr[M=m]$$

Perfectly Secure \Rightarrow Conditions

go backwards, starting from NB2.

□

Remark:

In the proof above, we divide by probabilities assuming they are non zero. Thus, the convention.

Perfect Indistinguishability

Story:

- We now use Lemma 2.2 to obtain another equivalent and useful formulation of perfect secrecy.
- This formulation, termed **perfect indistinguishability** states that the distributions over cipher-texts $\mathcal{C}(m_0)$ and $\mathcal{C}(m_1)$ are identical for any two messages $m_0, m_1 \in \mathcal{M}$.
- It is immediate that Lemma 2.2 implies perfect indistinguishability—the other direction also holds.

Lemma 2.3 (Perfect Indistinguishability)

An encryption scheme (Gen, Enc, Dec) over a message space \mathcal{M} is **perfectly secret** iff the scheme satisfies **perfect indistinguishability**, i.e.

for every probability distribution over \mathcal{M} , and
for every $m_0, m_1 \in \mathcal{M}$ and
for every $c \in \mathcal{C}$, it holds that

$$\Pr[C=c | M=m_0] = \Pr[C=c | M=m_1].$$

Proof.

Perfectly Secret \Rightarrow Perfect Indistinguishability

Assume:

the encryption scheme is perfectly secret and
fix $m_0, m_1 \in \mathcal{M}$ and $c \in \mathcal{C}$.

By Lemma 2.2, it follows that

$$\Pr[C=c | M=m_0] = \Pr[C=c] = \Pr[C=c | M=m_1].$$

Perfect Indistinguishability \Rightarrow Perfectly Secret

Assume:

for every distribution over \mathcal{M} ,

every $m_0, m_1 \in \mathcal{M}$ and
every $c \in \mathcal{C}$ it holds that

$$\Pr[C = c | M = m_0] = \Pr[C = c | M = m_1].$$

Observe that

$$\begin{aligned} \Pr[C = c] &= \sum_{m \in \mathcal{M}} \underbrace{\Pr[C = c | M = m]}_{\gamma} \Pr[M = m] \\ &= \gamma \cdot \sum_{m \in \mathcal{M}} \Pr[M = m] \\ &= \gamma = \Pr[C = c | M = m] \end{aligned}$$

NB: γ is the same, irrespective of which value m is picked

Since this holds for all m ,
the equation above is equivalent to the "conditions" in Lemma 2.2 and
therefore equivalent to perfect secrecy.

□

Adversarial Indistinguishability

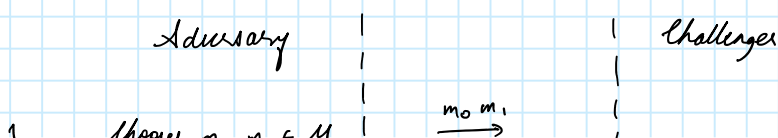
Story:

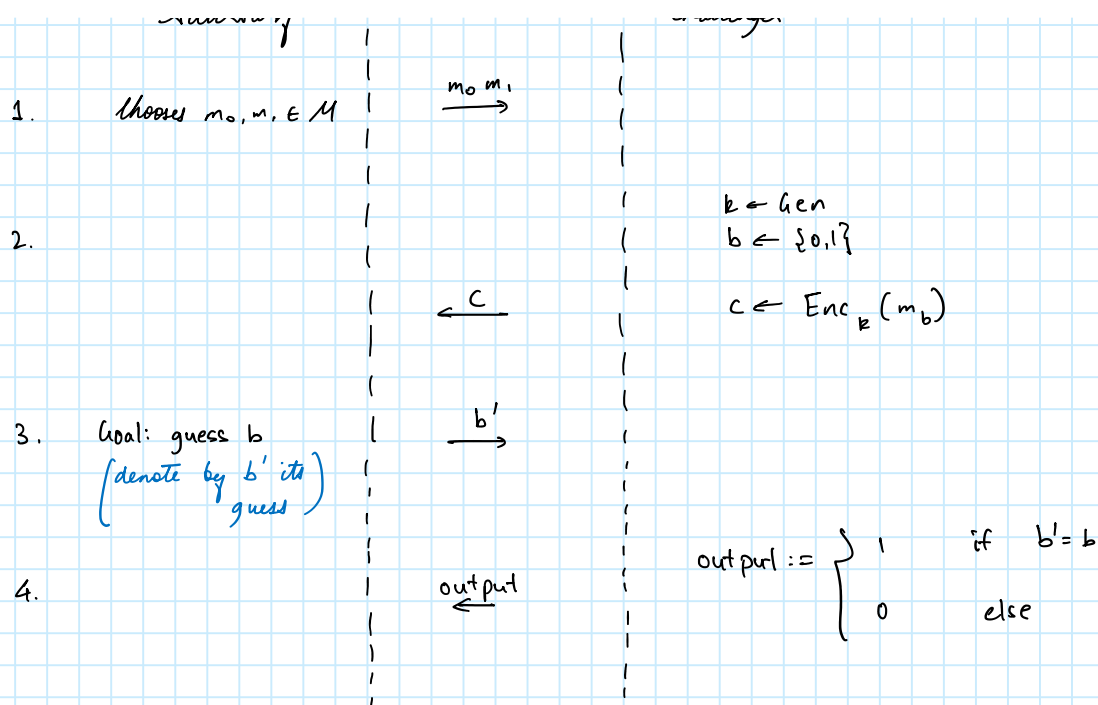
- We conclude the section by presenting one more equivalent definition of perfect secrecy.
- This definition is a *game* (or experiment) where an adversary \mathcal{A} must distinguish encryption of one plain-text from that of another
- It is therefore called **adversarial indistinguishability**
- This definition will be our starting point when we consider the notion of computational security (next chapter)

Notation:

- The game is denoted by $\text{PrivK}^{\text{eav}}$
(since it considers the private-key encryption and an eavesdropping adversary—it receives c and tries to learn something about the plain text)
- To define the scheme, fix
 - an encryption scheme $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$ over a message space \mathcal{M}
 - and any adversary \mathcal{A} .
- Let $\text{PrivK}_{\mathcal{A}, \Pi}$ denote an execution of the game where the game is defined as follows.

Definition: $\text{PrivK}_{\mathcal{A}, \Pi}^{\text{eav}}$ — **adversarial indistinguishability game**





NB.

The adversary \mathcal{A} can always succeed with probability one half
(by simply making a random guess)

The question is
can \mathcal{A} do better than half?

Story:

- The alternate definition we give basically requires that no such adversary \mathcal{A} succeeds with probability greater than half

Definition 2.4 (Adversarial Indistinguishability)

An encryption scheme $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$ over a message space \mathcal{M} is adversarially indistinguishable if
for every adversary \mathcal{A} it holds that

$$\Pr[\text{PrivK}_{\mathcal{A}, \Pi}^{\text{eav}} = 1] = 1/2.$$

Proposition 2.5 (Adversarial Indistinguishability \Leftrightarrow Perfectly Secret)

Let $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$ be an encryption scheme over messages \mathcal{M} . Then,
 Π is perfectly secret (Def. 2.1) $\Leftrightarrow \Pi$ is adversarially indistinguishable (Def. 2.4)

Proof idea.

Perfect secrecy \Rightarrow Adversarial indistinguishability
(Recast the LHS in Def 2.1 as the probability of winning;
just take \mathcal{M} to be the specific distribution over m_0 and m_1)

$$\begin{aligned}
 & \Pr[\mathcal{A} \text{ wins}] \\
 &= \sum_{\text{ciphertext}} \Pr[\mathcal{A} \text{ guesses } M \text{ correctly} \mid \text{ciphertext}] \Pr[\text{ciphertext}] \\
 &= \Pr[M = m_0 \mid C = c_0] \Pr[C = c_0] + \Pr[M = m_1 \mid C = c_1] \Pr[C = c_1]
 \end{aligned}$$

$\stackrel{\text{cipher text}}{=} \sum_{c \in \mathcal{C}} \Pr[C=c] \Pr[M=m_0 | C=c] + \Pr[M=m_1 | C=c] \Pr[C=c]$

$$= \Pr[M=m_0 | C=c_0] \Pr[C=c_0] + \Pr[M=m_1 | C=c_1] \Pr[C=c_1]$$

$$= \frac{1}{2} \left[\Pr[M=m_0] + \Pr[M=m_1] \right] \quad (\because \text{Perfect secrecy } \Pr[M=m_0 | C=c_0] = \Pr[M=m_0])$$

$$= \frac{1}{2} \cdot 1$$

$\Pr[C=c_0] = \frac{1}{2}$ by construction of the game

Written on

« Monday, April 24, 2023

– Perfect indistinguishability \Rightarrow – Adversarial indistinguishability

(There is at least one c in Def 2.3 s.t. $\Pr[c|m] > \Pr[c|m']$)

The strategy would be to ask either m or m' to be encrypted; if c is returned, predict m .

This strategy succeeds with probability greater than $1/2$.)

§ 2.2 The One-Time Pad (Vernam's Cipher)

Story:

- In 1917, Vernam patented a cipher, claiming it had perfect secrecy.
- The notion of perfect secrecy hadn't even been invented yet!
- Approximately, 25 years later, Shannon formalised this notion (i.e. of perfect secrecy) and proved Vernam's cipher (aka the One-Time Pad) indeed provides perfect secrecy.

Let $a \oplus b$ denote the bitwise exclusive-or (XOR) of two binary strings a and b

(i.e. if $a = a_1 \dots a_\ell$ and

$b = b_1 \dots b_\ell$ then

$$a \oplus b = (a_1 \oplus b_1, a_2 \oplus b_2, \dots, a_\ell \oplus b_\ell)$$

)

Definition: One-time pad encryption:

1. Fix an integer $\ell > 0$.

Suppose the message space, key space and ciphertext space are all equal to

$$\mathcal{M} = \mathcal{K} = \mathcal{C} = \{0,1\}^\ell$$

2. Gen: Choose a string from $\mathcal{K} = \{0,1\}^\ell$ uniformly at random
3. Enc: Given $k \in \mathcal{K}$ and $m \in \mathcal{M}$, output $c := k \oplus m$
4. Dec: Given $k \in \mathcal{K}$ and $c \in \mathcal{C}$, output $m := k \oplus c$

NB: Correctness: It is immediate that $\text{Dec}_k(\text{Enc}_k(m)) = k \oplus k \oplus m = m$.

Story:

- Intuition for security:
 - Given a cipher text c , there's no way an adversary can know which message m it corresponds to.
 - This is because for every possible message m , there is a key k s.t. $c = \text{Enc}_k(m)$

i.e. use $k = m \oplus c$.

- Each key is chosen with uniform probability—so no key is more likely than any other.
- We now prove this formally

Theorem 2.6 The one-time pad (Vernam's cipher) is a perfectly secret encryption scheme.

Proof.

- For some reason, the book uses the original definition of perfect secrecy
- However, using "Perfect indistinguishability",
 - it is almost immediate that the scheme is secure because
 - it holds that $\Pr[c|m] = \Pr[c|m'] = 2^{-\ell}$.

□

Discussion:

- One time pad encryption has a number of drawbacks
 - **Key** is required to be **as long as the message**
 - Limits ability to send very long messages
 - also requires us to know an upper bound on the message in advance
 - NB: **Cannot use the key more than once**—used more than once leaks information about the message
 - e.g. $c = m \oplus k$ and $c' = m' \oplus k$ allows the adversary to compute $c \oplus c' = m \oplus m'$ (i.e. it allows for some function of the messages to be learnt)
 - may not sound like much but it does rule out any claims of perfect encryption
 - using frequency analysis (if the encryption language is known), one may be able to recover the message as well
 - **Not secure against known-message attacks**—when the adversary is allowed to request encryption of messages it likes
 - e.g. if an adversary learns the encryption $c = m \oplus k$ of any known message m , then clearly, using $c \oplus m$, it can recover the key!

§ 2.3 Limitations of Perfect Secrecy

Story:

- Could it be that no scheme satisfying perfect secrecy can achieve the points listed above?
- The answer turns out to be yes.
- We prove the third point, i.e.
 - we prove that *any (or, if you prefer, every)* perfectly secret encryption scheme must have a key space that is at least as large as the message space.

Theorem 2.7 Let $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$ be a **perfectly secret encryption** scheme (over a message space \mathcal{M}), and let \mathcal{K} be the key space (determined by Gen). Then, **$|\mathcal{K}| \geq |\mathcal{M}|$** .

Proof.

Goal: We show that if $|\mathcal{K}| < |\mathcal{M}|$, then the scheme is not perfectly secret.

Idea: To show that the number of messages corresponding to a given ciphertext is $< |\mathcal{K}|$ and then use this cipher text to argue $\Pr[m|c] \neq \Pr[m]$.

Sketch:

- Let c be a cipher text and
- let $\mathcal{M}(c) = \{m: m = \text{Dec}_k(c) \text{ for some } k\}$ be
- Observe that $|\mathcal{M}(c)| \leq |\mathcal{K}|$
 - This is because there can be at most 1 message $m \in \mathcal{M}(c)$ for every key k (recall, $\text{Dec}_k(\cdot)$ is assumed deterministic)
- Thus, simply by counting, we know that
 - there exists an $m' \in \mathcal{M}$ but $m' \notin \mathcal{M}(c)$.
 - i.e. $\Pr[M = m'|C = c] = 0$ but $\Pr[M = m'] \neq 0$ (recall, we restricted to distributions over \mathcal{M} s.t. no m appears with zero probability)
 - But that contradicts the perfect secrecy condition, i.e. $\Pr[M = m'|C = c] = \Pr[M = m']$.

□

Written on

« Tuesday, April 25, 2023

§ 2.4 Shannon's Theorem

Story:

- Shannon's breakthrough work on perfect secrecy was to show the following:
- He characterised perfectly-secret encryption schemes
 - Suppose $|\mathcal{K}| = |\mathcal{M}| = |\mathcal{C}|$
 - Then, the Gen algorithm must choose a secret key *uniformly* (from the set of all keys)
 - And, there exists a *single* key, mapping the plain-text to the cipher text
- Remarks:
 - Interesting by itself
 - Also useful for proving (or contradicting) perfect secrecy of schemes
- Convention: As before, assume no $m \in \mathcal{M}$ and no $c \in \mathcal{C}$ is assigned zero probability
- Remark:
 - $|\mathcal{K}| = |\mathcal{M}| = |\mathcal{C}|$ is the most efficient setting
 - $|\mathcal{K}| \geq |\mathcal{M}|$ (by the theorem above for perfect secrecy)
 - $|\mathcal{M}| \leq |\mathcal{C}|$
 - (otherwise one cannot have perfect correctness;
 - if $|\mathcal{M}| > |\mathcal{C}|$, then two messages must be mapped to a message)

Theorem 2.8 (Shannon's theorem).

Let $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$ be an encryption scheme over \mathcal{M} with $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$.

This scheme is *perfectly secret* iff:

1. Every key $k \in \mathcal{K}$ is chosen with equal probability (i.e. $1/|\mathcal{K}|$) by algorithm Gen.
2. For every $m \in \mathcal{M}$, and every $c \in \mathcal{C}$, there is a single key $k \in \mathcal{K}$ s.t. $\text{Enc}_k(m) = c$.

Proof.

Intuition [I'm actually not sure the intuition matches how the proof goes; but ok]:

- "easy direction" Conditions (1) and (2) imply perfect secrecy
 - Fix any cipher text c .
 - Now c could have resulted from encrypting any message m (because condition (2) says there's a key relating every message and ciphertext).
 - From condition 1, we know that each key is used with equal probability.
 - Proceed as in the one-time pad to show perfect secrecy.
- "hard direction" Perfect secrecy implies conditions (1) and (2)

- It is intuitive that if $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$, then exactly one key k relates any given message m and ciphertext c
 - If some message m is not mapped to a given c at all, it violates perfect secrecy
 - If m is mapped to c by many keys, then it must be that another message m' is not mapped to c and then due to m' , the scheme would violate perfect secrecy
- Thus, condition (2) should hold
- Given this, it must be the case that the keys are chosen with equal probability—else certain ciphertexts would become more likely than others, contradicting perfect secrecy.

Formal arguments:

Aim: "hard direction" Perfectly Secret \Rightarrow Conditions (1) and (2) hold.

We first show (2) holds.

Story: In fact, showing *there is a key* for every m and c is easy (because we argued similarly before).

NB1: $\forall m \in \mathcal{M} \ \& \ c \in \mathcal{C}$
 \exists (at least one) $k \in \mathcal{K}$ s.t.
 $Enc_k(m) = c.$

(otherwise, $P_X[M=m|C=c] = 0 \neq P_X[M=m]$) [violates perfect secrecy]

Story: The new step is showing there is *at most one key* for every m and c .

Let: $\mathcal{C}(m) := \{c : c = Enc_k(m) \text{ for some } k\}$
 $= \{Enc_k(m)\}_{k \in \mathcal{K}}$

NB: $|\mathcal{C}(m)| \geq |c|$
 \because for every $c \in \mathcal{C}$
 \exists a $k \in \mathcal{K}$ s.t.
 $Enc_k(m) = c$ [using NB1]

NB2: $|\mathcal{C}(m)| \leq |\mathcal{C}|$ because $\mathcal{C}(m) \subseteq \mathcal{C}$ by definition of \mathcal{C} and $\mathcal{C}(m)$

Thus: $|\mathcal{C}(m)| = |\mathcal{K}|$ (because recall, $|\mathcal{K}| = |\mathcal{C}|$)

NB3: "Therefore", there do not exist distinct keys $k, k' \in \mathcal{K}$ s.t.

$$Enc_k(m) = Enc_{k'}(m) = c,$$

i.e. there is *at most one key* $k \in \mathcal{K}$ encrypting m to c .

To see this, note that for each $k \in \mathcal{K}$,

one needs to add a *distinct element* $Enc_k(m)$ to $\mathcal{C}(m)$

(i.e. one can't have even two different k s map m to the same c)

otherwise the count won't add up to $|\mathcal{K}|$.

NB4: Combining NB1 and NB3, we obtain condition (2).

« Wednesday, April 26, 2023

Story: It remains to show condition (1) also holds, i.e.
we show $\Pr[K = k] = 1/|\mathcal{K}|$ for all $k \in \mathcal{K}$.

Let:

$$|\mathcal{K}| = N$$
$$\mathcal{M} = \{m_1 \dots m_N\}$$

Fix: a ciphertext c .

NB: From Condition (2), one can always label the keys $\{k_1 \dots k_N\}$ that map m_i to $\text{Enc}_{k_i}(m_i) = c$

NB2: The key k_i for each m_i is distinct

NB3: Perfect secrecy yields

$$\begin{aligned} \Pr[M=m_i] &= \Pr[M=m_i | C=c] \\ &= \frac{\Pr[C=c | M=m_i] \cdot \Pr[M=m_i]}{\Pr[C=c]} && \text{(conditional)} \\ &= \frac{\Pr[K=k_i] \cdot \Pr[M=m_i]}{\Pr[C=c]} && \Pr(A|B) = \frac{\Pr(A \wedge B)}{\Pr(B)} \\ &&& \text{(Using the mapping b/w the } c \text{ \& } m_i) \end{aligned}$$

$$\Rightarrow \Pr[K=k_i] = \Pr[C=c]$$

Thus, K follows the uniform distribution.

Aim: "Easy direction" Conditions (1) and (2) \Rightarrow Perfect Secrecy

Fix: Any message $m \in \mathcal{M}$

Recall:

Condition (2) tells us that there is exactly one key k mapping m to c .
Condition (1) says every key appears with equal probability

Thus:

$$\Pr[c|m] = 1/|\mathcal{K}| \text{ for any } c \in \mathcal{C}.$$

NB: Further, m was arbitrary, so we have

$$\Pr[c|m] = \Pr[c|m']$$

i.e. our scheme satisfies perfect indistinguishability (and that is equivalent to perfect secrecy).

□

Uses of Shannon's theorem

Remarks:

- [recall] it is of independent interest
- Conditions (1) and (2) do not depend on the distribution over \mathcal{M} .
 - So, to check the conditions, one can use any distribution that is convenient
 - And thus deduce (via the theorem) perfect secrecy without having to prove it for all distributions over \mathcal{M}
- Conditions (1) and (2) are easier to analyse (don't need to worry about probabilities, unlike the definition of perfect secrecy)
 - CAVEAT: must ensure $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$
 - NB: one-time pad is trivially secure using these conditions

Closing Remark

We studied perfectly secure encryptions.
Other cryptographic problems also have "perfect security"

E.g. Message authentication—aim is to prevent an adversary from modifying a message (in an undetectable way) en route from one user to another (see Chapter 4).