Perfectly Secure Cipher

Presentation by:
V. Balasubramanian
SSN College of Engineering

Objective

- Principles of Modern Cipher
- Security Definitions
- Cipher text only attack
- Proof for Security

Introduction

- “Heuristic” constructions; construct, break, repeat, …
- Can we *prove* that some encryption scheme
is secure?
- First need to *define* what we mean by “secure” in the first place…

Historically

- Cryptography was an *art*
  - Heuristic design and analysis
- This isn’t very satisfying
  - How do we know when a scheme is secure?

Modern Cryptography

- In the late ‘70s and early ‘80s, cryptography began to develop into more of a *science*

- Based on three principles that underpin most crypto work today

Core principles of modern crypto

- Formal definitions
  - Precise, mathematical model and definition of what security means
- Assumptions
  - Clearly stated and unambiguous
- Proofs of security
  - Move away from design-break-patch

Importance of definition

- *If you don't understand what you want to achieve, how can you possibly know when (or if) you have achieved it?*

Importance of definitions -- analysis

- Definitions enable meaningful analysis, evaluation, and comparison of schemes
  - Does a scheme satisfy the definition?
  - What definition does it satisfy?
    - Note: there may be multiple meaningful definitions!
    - One scheme may be less efficient than another, yet satisfy a stronger security definition

Importance of definitions -- usage

- Definitions allow others to understand the security guarantees provided by a scheme
- Enables schemes to be used as *components* of a larger system (modularity)
- Enables one scheme to be substituted for another if they satisfy the same definition

Assumptions

- With few exceptions, cryptography currently requires *computational assumptions*
  - At least until we prove P $\neq$ NP (and even that would not be enough)

- Principle: any such assumptions should
  be made explicit

Importance of clear assumptions

- Allow researchers to (attempt to) *validate* assumptions by studying them
- Allow meaningful *comparison* between schemes based on different assumptions
  - Useful to understand minimal assumptions needed
- Practical implications if assumptions are wrong

- Enable proofs of security

Proofs of security

- Provide a rigorous proof that a construction satisfies a given definition under certain specified assumptions
  - Provides an iron-clad guarantee (relative to your definition and assumptions!)

- Proofs are crucial in cryptography, where there is a malicious attacker trying to “break” the scheme

Limitations?

- Cryptography remains partly an *art* as well

- Given a proof of security based on some assumption, we still need to *instantiate* the assumption
  - Validity of various assumptions is an active area of research

Limitations?

- Proofs given an iron-clad guarantee of security
  - …relative to the definition and the assumptions!

- Provably secure schemes can be broken!
  - If the definition does not correspond to the real-world threat model
    - I.e., if attacker can go “outside the security model”
    - This happens a lot in practice
  - If the assumption is invalid
  - If the implementation is flawed
    - This happens a lot in practice

Nevertheless…

- This does not detract from the importance of having formal definitions in place
- This does not detract from the importance of proofs of security

Defining secure encryption

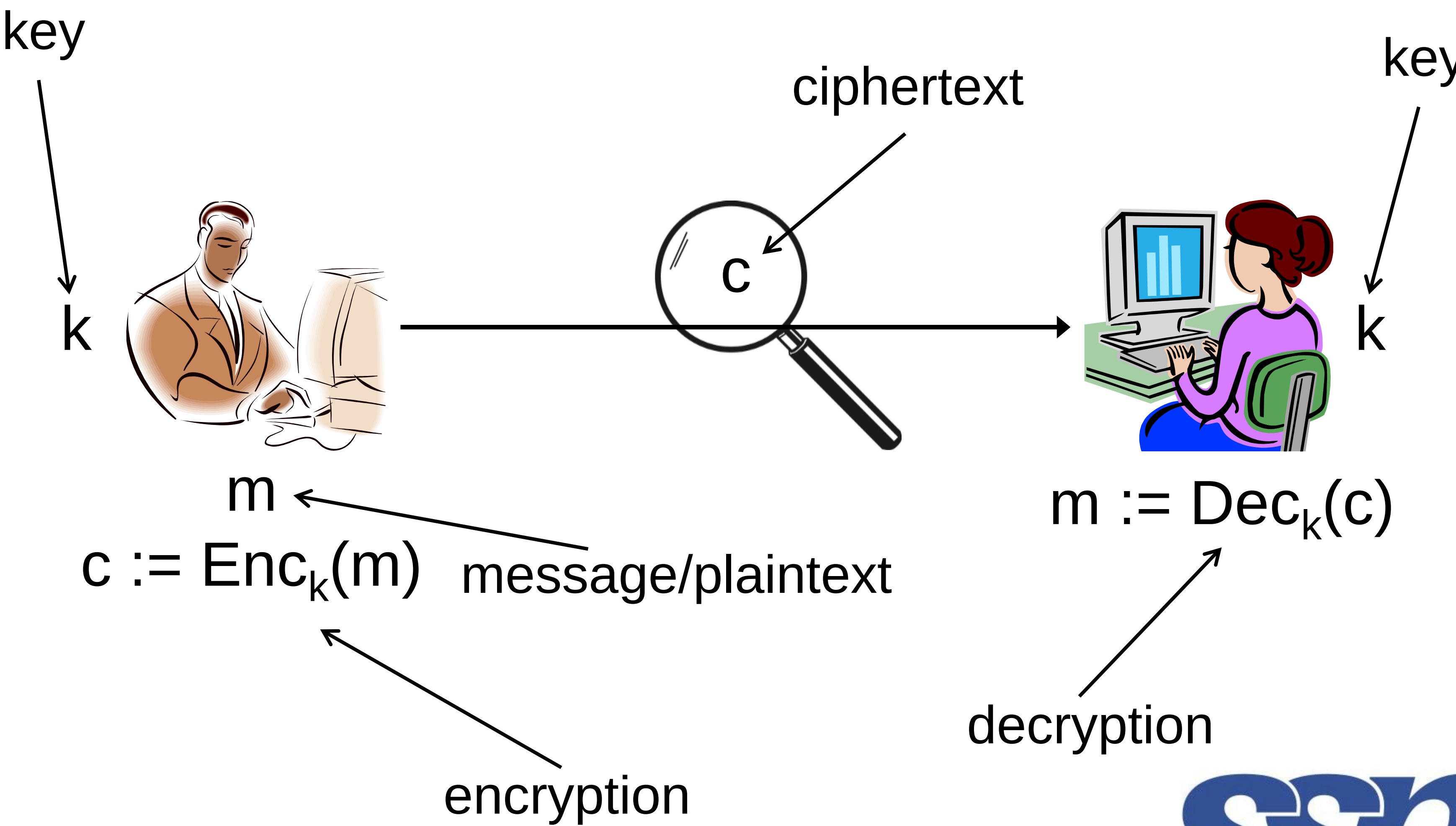
Crypto definitions (generally)

- Security guarantee/goal
  - What we want to achieve and/or what we want to prevent the attacker from achieving
- Threat model
  - What (real-world) capabilities the attacker is assumed to have

Recall

- A *private-key encryption scheme* is defined by a message space $\mathcal{M}$ and algorithms (Gen, Enc, Dec):
  - Gen (key-generation algorithm): generates k
  - Enc (encryption algorithm): takes key k and message
    $m \in \mathcal{M}$ as input; outputs ciphertext c
    $$c \leftarrow \mathrm{Enc}_k(m)$$
  - Dec (decryption algorithm): takes key k and ciphertext c as input; outputs m.
    $$m := \mathrm{Dec}_k(c)$$

Private-key encryption



Threat models for encryption

- Ciphertext-only attack
  - One ciphertext or many?
- Known-plaintext attack
- Chosen-plaintext attack
- Chosen-ciphertext attack

Type of Attack	Known to Cryptanalyst
Ciphertext Only	■ Encryption algorithm ■ Ciphertext
Known Plaintext	■ Encryption algorithm ■ Ciphertext ■ One or more plaintext–ciphertext pairs formed with the secret key
Chosen Plaintext	■ Encryption algorithm ■ Ciphertext ■ Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key
Chosen Ciphertext	■ Encryption algorithm ■ Ciphertext ■ Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key
Chosen Text	■ Encryption algorithm ■ Ciphertext ■ Plaintext message chosen by cryptanalyst, together with its corresponding ciphertext generated with the secret key ■ Ciphertext chosen by cryptanalyst, together with its corresponding decrypted plaintext generated with the secret key

Goal of secure encryption?

- How would you define what it means for encryption scheme (Gen, Enc, Dec) over message space $\mathcal{M}$ to be secure?
  - Against a (single) ciphertext-only attack

Secure encryption?

- "Impossible for the attacker to learn the key"
  - The key is a *means to an end*, not the end itself
  - Necessary (to some extent) but not sufficient
  - Easy to design an encryption scheme that hides the key completely, but is insecure
  - Can design schemes where most of the key is leaked, but the scheme is still secure

Secure encryption?

- "Impossible for the attacker to learn the plaintext from the ciphertext"
  - What if the attacker learns 90% of the plaintext?

Secure encryption?

- "Impossible for the attacker to learn any character of the plaintext from the ciphertext"
  - What if the attacker is able to learn (other)
    partial information about the plaintext?
    - E.g., salary is greater than $75K
  - What if the attacker guesses a character correctly?

Perfect secrecy

Perfect secrecy

- "Regardless of any *prior* information the attacker has about the plaintext, the ciphertext should leak no *additional* information about the plaintext"
  - The right notion!
  - How to formalize?

Probability review

- *Random variable (r.v.):* variable that takes on (discrete) values with certain probabilities

- Probability distribution for a r.v. specifies the probabilities with which the variable takes on each possible value
  - Each probability must be between 0 and 1
  - The probabilities must sum to 1

Probability review

- *Event*: a particular occurrence in some experiment
  - Pr[E]: probability of event E

- Conditional probability: probability that one event occurs, *given that* some other event occurred
  - Pr[A | B] = Pr[A and B]/Pr[B]

- Two r.v.'s X, Y are *independent* if
  for all x, y: Pr[X=x | Y=y] = Pr[X=x]

Probability review

- Law of total probability: say $E_1, \ldots, E_n$ are a *partition* of all possibilities. Then for any A:

  $$\Pr[A] = \Sigma_i \Pr[A \text{ and } E_i] = \Sigma_i \Pr[A \mid E_i] \cdot \Pr[E_i]$$

Notation

- $\mathcal{K}$ (key space) – set of all possible keys

- $\mathcal{C}$ (ciphertext space) – set of all possible ciphertexts

Probability distributions

- Let M be the random variable denoting the value of the message
  - M ranges over $\mathcal{M}$
  - This reflects the likelihood of different messages being sent by the parties, given the attacker's prior knowledge
  - E.g.,

$$\Pr[M = \text{"attack today"}] = 0.7$$
$$\Pr[M = \text{"don't attack"}] = 0.3$$

Probability distributions

- Let K be a random variable denoting the key
  - K ranges over $\mathcal{K}$
- Fix some encryption scheme (Gen, Enc, Dec)
  - Gen defines a probability distribution for K: Pr[K = k] = Pr[Gen outputs key k]

Probability distributions

- Random variables M and K are *independent*
  - I.e., the message that a party sends does not depend on the key used to encrypt that message

Probability distributions

- Fix some encryption scheme (Gen, Enc, Dec), and some distribution for M
- Consider the following (randomized) experiment:
    1. Choose a message m, according to the given distribution
    2. Generate a key k using Gen
    3. Compute $c \leftarrow \mathrm{Enc}_k(m)$
- This defines a distribution on the ciphertext!
- Let C be a random variable denoting the value of the ciphertext in this experiment

Example 1

- Consider the shift cipher
  - So for all k $\in$ {0, …, 25}, Pr[K = k] = 1/26
- Say Pr[M = ‘a’] = 0.7, Pr[M = ‘z’] = 0.3
- What is Pr[C = ‘b’] ?
  - Either M = ‘a’ and K = 1, or M = ‘z’ and K = 2
  - Pr[C=‘b’] = Pr[M=‘a’]·Pr[K=1] + Pr[M=‘z’] ·Pr[K=2]

$$= 0.7 \cdot (1/26) + 0.3 \cdot (1/26)$$
$$= 1/26$$

Example 2

- Consider the shift cipher, and the distribution Pr[M = ‘one’] = ½, Pr[M = ‘ten’] = ½
- Pr[C = ‘rqh’] = ?

  = Pr[C = ‘rqh’ | M = ‘one’] · Pr[M = ‘one’]
  + Pr[C = ‘rqh’ | M = ‘ten’] · Pr[M = ‘ten’]

  = 1/26 · ½ + 0 · ½ = 1/52

Perfect secrecy (informal)

- "Regardless of any *prior* information the attacker has about the plaintext, the ciphertext should leak no *additional* information about the plaintext"

Perfect secrecy (informal)

- Attacker's information about the plaintext = attacker-known *distribution* of M
- Perfect secrecy means that observing the ciphertext should not change the attacker's knowledge about the distribution of M

Perfect secrecy (formal)

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

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

- i.e., the distribution of M does not change conditioned on observing the ciphertext

Example 3

- Consider the shift cipher, and the distribution Pr[M = ‘one’] = ½, Pr[M = ‘ten’] = ½
- Take m = ‘ten’ and c = ‘rqh’

- Pr[M = ‘ten’ | C = ‘rqh’] = ?

  = 0

  ≠ Pr[M = ‘ten’]

Bayes Theorem

- $Pr[A \mid B] = Pr[B \mid A] \cdot Pr[A]/Pr[B]$

Example 4

- Shift cipher;
  Pr[M='hi'] = 0.3,
  Pr[M='no'] = 0.2,
  Pr[M='in']= 0.5

- Pr[M = 'hi' | C = 'xy'] = ?

  = Pr[C = 'xy' | M = 'hi'] · Pr[M = 'hi']/Pr[C = 'xy']
- Pr[C = 'xy' | M = 'hi'] = 1/26
- Pr[C = 'xy']

  = Pr[C = 'xy' | M = 'hi'] · 0.3 + Pr[C = 'xy' | M = 'no'] · 0.2
  + Pr[C='xy' | M='in'] · 0.5
  = (1/26) · 0.3 + (1/26) · 0.2 + 0 · 0.5
  = 1/52

Contd…

- $\Pr[M = \text{'hi'} \mid C = \text{'xy'}] = ?$

  $= \Pr[C = \text{'xy'} \mid M = \text{'hi'}] \cdot \Pr[M = \text{'hi'}]/\Pr[C = \text{'xy'}]$

  $= (1/26) \cdot 0.3/(1/52)$

  $= 0.6$

  $\neq \Pr[M = \text{'hi'}]$

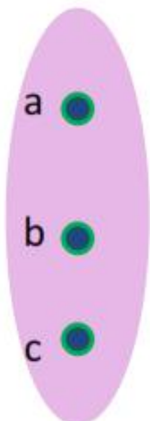
Conclusion

- The shift cipher is not perfectly secret!
  - At least not for 2-character messages

Example 5

Plaintext Distribution

- Let **X** be a discrete random variable over the set $\mathbb{P}$
- Alice chooses x from $\mathbb{P}$ based on some probability distribution
  - Let *Pr[**X** = x]* be the probability that x is chosen
  - This probability may depend on the language



$\mathbb{P}$

Plaintext set
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

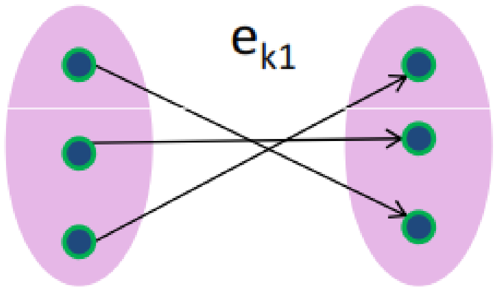
Note : Pr[a] + Pr[b] + Pr[c] = 1

Example 5

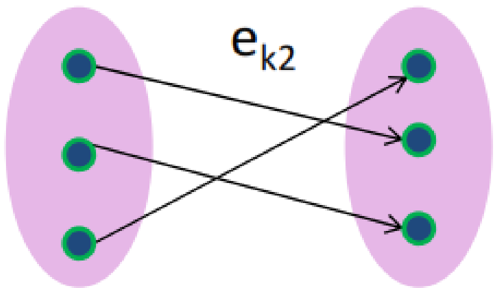
Key Distribution

- Alice & Bob agree upon a key k chosen from a key set $\mathbb{K}$
- Let ***K*** be a random variable denoting this choice

keyspace
Pr[**K**=k_1] = ¾
Pr[**K**=k_2] = ¼



There are two keys in the keyset
thus there are two possible encryption
mappings



Cipher Text Distribution

- Let **Y** be a discrete random variable over the set $\mathbb{C}$
- The probability of obtaining a particular ciphertext y depends on the plaintext and key probabilities

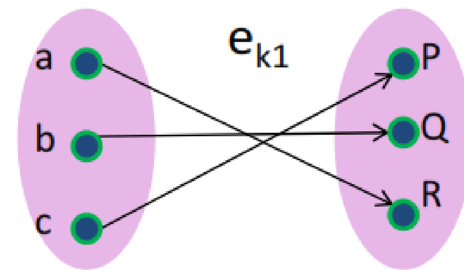
$$\Pr[Y = y] = \sum_k \Pr(k)\Pr(d_k(y))$$

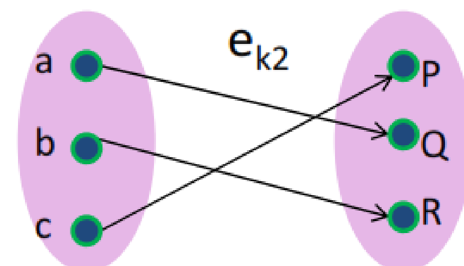
Pr[Y = P] = $\Pr(k_1) * \Pr(c) + \Pr(k_2) * \Pr(c)$
= $(3/4 * 1/6) + (1/4 * 1/6) =$ **1/6**

Pr[Y = Q] = $\Pr(k_1) * \Pr(b) + \Pr(k_2) * \Pr(a)$
= $(3/4 * 1/3) + (1/4 * 1/2) =$ **3/8**

Pr[Y = R] = $\Pr(k_1) * \Pr(a) + \Pr(k_2) * \Pr(b)$
= $(3/4 * 1/2) + (1/4 * 1/3) =$ **11/24**

Note: Pr[Y=P] + Pr[Y=Q] + Pr[Y=R] = 1





plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = ¾
Pr[**K**=k_2] = ¼

Attacker's Probability

- The attacker wants to determine the plaintext ***x***
- Two scenarios
  - Attacker does not have y (a priori Probability)
    - Probability of determining ***x*** is simply ***Pr[x]***
    - Depends on plaintext distribution (eg. Language charcteristics)
  - Attacker has y (a posteriori probability)
    - Probability of determining x is simply ***Pr[x|y]***

Posteriori Probability

- How to compute the attacker's a posteriori probabilities? $\Pr[X = x \mid Y = y]$
  - Bayes' Theorem

$$\boxed{\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}}$$

probability of the plaintext

probability of this ciphertext

?

The probability that y is obtained given x depends on the keys which provide such a mapping

$$\Pr[y \mid x] = \sum_{\{k \,:\, d_k(y) = x\}} \Pr[k]$$

P[Y|X]

$\Pr[P|a] = 0$

$\Pr[P|b] = 0$

$\Pr[P|c] = 1$

$\Pr[Q|a] = \Pr[k_2] = 1/4$

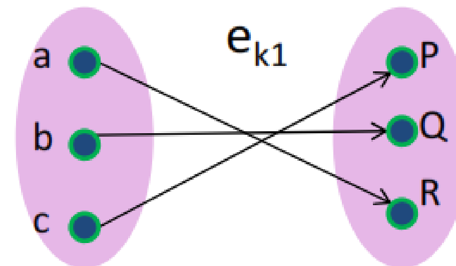
$\Pr[Q|b] = \Pr[k_1] = 3/4$

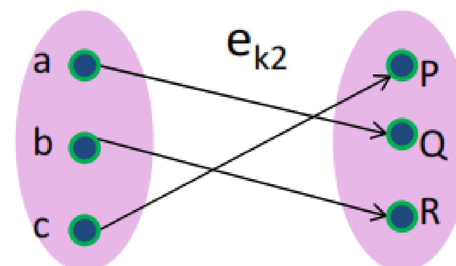
$\Pr[Q|c] = 0$

$\Pr[R|a] = \Pr[k_1] = 3/4$

$\Pr[R|b] = \Pr[k_2] = 1/4$

$\Pr[R|c] = 0$





keyspace
$\Pr[\mathbf{K}=k_1] = 3/4$
$\Pr[\mathbf{K}=k_2] = 1/4$

Computing Posteriori Probbaility

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

ciphertext
Pr[**Y**=P] = 1/6
Pr[**Y**=Q] = 3/8
Pr[**Y**=R] = 11/24

Pr[y|x]
Pr[P|a] = 0 Pr[P|b] = 0 Pr[P|c] = 1
Pr[Q|a] = ¼ Pr[Q|b] = ¾ Pr[Q|c] = 0
Pr[R|a] = ¾ Pr[R|b] = ¼ Pr[R|c] = 0

Pr[a|P] = 0	Pr[b|P] = 0	Pr[c|P] = 1
Pr[a|Q] = 1/3	Pr[b|Q] = 2/3	Pr[c|Q] = 0
Pr[a|R] = 9/11	Pr[b|R] = 2/11	Pr[c|R] = 0

If the attacker sees ciphertext ***P*** then she would know the plaintext was ***c***
If the attacker sees ciphertext ***R*** then she would know ***a*** is the most likely plaintext
Not a good encryption mechanism!!

Perfect Secrecy

Perfect secrecy achieved when

a posteriori probabilities = a priori probabilities

$$\boxed{\Pr[x \mid y] = \Pr[x]}$$

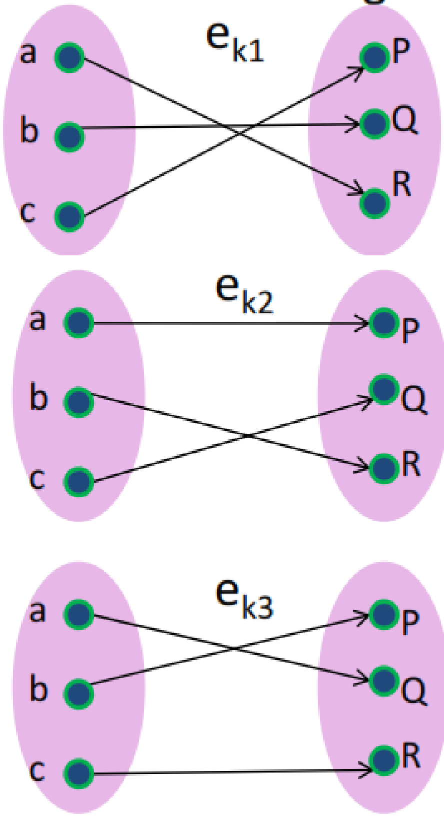
i.e the attacker learns nothing from the ciphertext

Example 6

- Find the a posteriori probabilities for the following scheme
- Verify that it is perfectly secret.

plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = 1/3
Pr[**K**=k_2] = 1/3
Pr[K=k_3] = 1/3



Solution

- Given

plaintext
Pr[**X**=a] = 1/2
Pr[**X**=b] = 1/3
Pr[**X**=c] = 1/6

keyspace
Pr[**K**=k_1] = 1/3
Pr[**K**=k_2] = 1/3
Pr[K=k_3] = 1/3

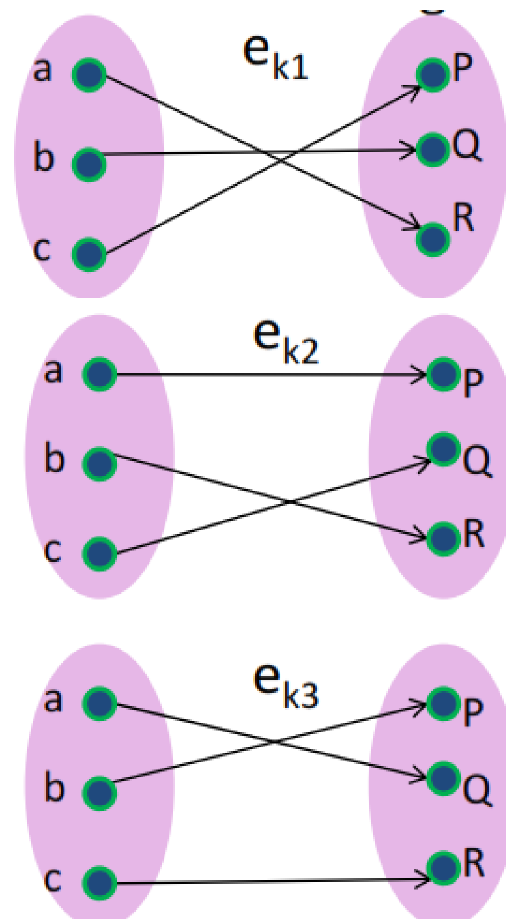
Cipher Text Distribution

$$P_r[Y = y] = \sum_k P_r(k).P_r(d_k(y))$$

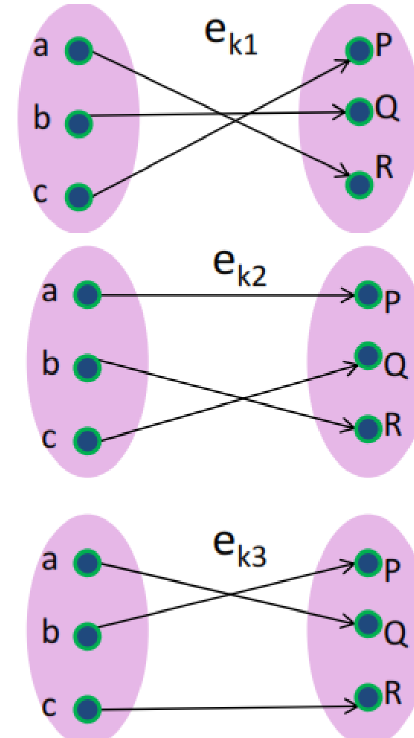
$$P_r[Y = P] = P_r(k_1).P_r(c) + P_r(k_2).P_r(a) + P_r(k_3).P_r(b) = \frac{1}{3} * \frac{1}{6} + \frac{1}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{3} = \frac{1}{18} + \frac{1}{6} + \frac{1}{9} = \frac{1+3+2}{18} = \frac{1}{3}$$

$$P_r[Y = Q] = P_r(k_1).P_r(b) + P_r(k_2).P_r(c) + P_r(k_3).P_r(a) = \frac{1}{3} * \frac{1}{3} + \frac{1}{3} * \frac{1}{6} + \frac{1}{3} * \frac{1}{2} = \frac{1}{9} + \frac{1}{18} + \frac{1}{6} = \frac{2+1+3}{18} = \frac{1}{3}$$

$$P_r[Y = R] = P_r(k_1).P_r(a) + P_r(k_2).P_r(b) + P_r(k_3).P_r(c) = \frac{1}{3} * \frac{1}{2} + \frac{1}{3} * \frac{1}{3} + \frac{1}{3} * \frac{1}{6} = \frac{1}{6} + \frac{1}{9} + \frac{1}{18} = \frac{3+2+1}{18} = \frac{1}{3}$$



P(Y|X)



The probability that y is obtained given x depends on the keys which provide such a mapping

$$\Pr[y \mid x] = \sum_{\{k \,:\, d_k(y)=x\}} \Pr[k]$$

$$P(P|a) = P_r(k_2) = \frac{1}{3} \quad P(P|b) = P_r(k_3) = \frac{1}{3} \quad P(P|c) = P_r(k_1) = \frac{1}{3}$$
$$P(Q|a) = P_r(k_3) = \frac{1}{3} \quad P(Q|b) = P_r(k_1) = \frac{1}{3} \quad P(Q|c) = P_r(k_2) = \frac{1}{3}$$
$$P(R|a) = P_r(k_1) = \frac{1}{3} \quad P(R|b) = P_r(k_2) = \frac{1}{3} \quad P(R|c) = P_r(k_3) = \frac{1}{3}$$

Computing Posteriori Probability

$$\Pr[x \mid y] = \frac{\Pr[x] \times \Pr[y \mid x]}{\Pr[y]}$$

$$P_r[a|P] = P_r[a] * \frac{P_r[P|a]}{P_r[P]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r}[\mathrm{a}] = \frac{1}{2}$$

$$P_r[a|Q] = P_r[a] * \frac{P_r[Q|a]}{P_r[Q]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r}[\mathrm{a}] = \frac{1}{2}$$

$$P_r[a|R] = P_r[a] * \frac{P_r[R|a]}{P_r[R]} = \frac{1}{2} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{2} = \mathrm{P_r}[\mathrm{a}] = \frac{1}{2}$$

Contd…

$$P_r[b|P] = P_r[b] * \frac{P_r[P|b]}{P_r[P]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r}[\mathrm{b}] = \frac{1}{3}$$

$$P_r[b|Q] = P_r[b] * \frac{P_r[Q|b]}{P_r[Q]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r}[\mathrm{b}] = \frac{1}{3}$$

$$P_r[b|R] = P_r[b] * \frac{P_r[R|b]}{P_r[R]} = \frac{1}{3} * \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{1}{3} = \mathrm{P_r}[\mathrm{b}] = \frac{1}{3}$$

Example

Example 3.3 Let $\mathcal{P} = \{a, b\}$ with $\mathbf{Pr}[a] = 1/4, \mathbf{Pr}[b] = 3/4$. Let $\mathcal{K} = \{K_1, K_2, K_3\}$ with $\mathbf{Pr}[K_1] = 1/2, \mathbf{Pr}[K_2] = \mathbf{Pr}[K_3] = 1/4$. Let $\mathcal{C} = \{1, 2, 3, 4\}$, and suppose the encryption functions are defined to be $e_{K_1}(a) = 1, e_{K_1}(b) = 2$; $e_{K_2}(a) = 2, e_{K_2}(b) = 3$; and $e_{K_3}(a) = 3, e_{K_3}(b) = 4$. This cryptosystem can be represented by the following ***encryption matrix***:

	a	b
K_1	1	2
K_2	2	3
K_3	3	4

Contd…

$$
\begin{aligned}
\mathbf{Pr}[1] &= \frac{1}{8} \\
\mathbf{Pr}[2] &= \frac{3}{8} + \frac{1}{16} &&= \frac{7}{16} \\
\mathbf{Pr}[3] &= \frac{3}{16} + \frac{1}{16} &&= \frac{1}{4} \\
\mathbf{Pr}[4] &= \frac{3}{16}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{Pr}[a|1] &= 1 & \mathbf{Pr}[b|1] &= 0 \\
\mathbf{Pr}[a|2] &= \frac{1}{7} & \mathbf{Pr}[b|2] &= \frac{6}{7} \\
\mathbf{Pr}[a|3] &= \frac{1}{4} & \mathbf{Pr}[b|3] &= \frac{3}{4} \\
\mathbf{Pr}[a|4] &= 0 & \mathbf{Pr}[b|4] &= 1.
\end{aligned}
$$

ssn