Cryptography - Session 2

O. Geil, Aalborg University

November 18, 2010

Random variables

Discrete random variable $\mathbf{X}$:

1. Probability distribution on finite set $\mathcal{X}$.
2. For $x \in \mathcal{X}$ write $\Pr(x) = Pr(\mathbf{X} = x)$.

$\mathbf{X}$ and $\mathbf{Y}$ are independent:

- $\forall x, y : \Pr(x, y) = \Pr(x)\Pr(y)$
- $\forall x, y : \Pr(x|y) = \Pr(x)$.

Perfect secrecy

Given $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ assume in the beginning of the talk that K is used only for ONE encryption.

A cryptosystem has perfect secrecy if for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$ it holds that

$$\Pr(x|y) = \Pr(x).$$

One, observation of cipher text does not reveal anything.

Theorem 2.3: The shift cipher is perfect

Proof:

$$
\begin{aligned}
\Pr(y) &= \sum_{K \in \mathbf{Z}_s} \Pr(K)\Pr(\mathbf{X} = d_K(y)) \\
&= \sum_{K \in \mathbb{Z}_s} \frac{1}{s}\Pr(\mathbf{X} = d_K(y)) \\
&= \frac{1}{s}\sum_{K \in \mathbb{Z}_s} \Pr(\mathbf{X} = y - K) \\
&= \frac{1}{s}\sum_{x \in \mathbb{Z}_s} \Pr(x) = \frac{1}{s}.
\end{aligned}
$$

$$
\Pr(y|x) = \Pr(\mathbf{K} = y - x) = \frac{1}{s}.
$$

Due to Bayes' formula (which holds for all y with $\Pr(y) > 0$):

$$
\Pr(x|y) = \frac{\Pr(x)\Pr(y|x)}{\Pr(y)} = \frac{\Pr(x)\frac{1}{s}}{\frac{1}{s}} = \Pr(x).
$$

Some necessary conditions for perfect cipher

Assume $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ provides perfect secrecy. That is, $\Pr(x|y) = \Pr(x)$ for all $x \in \mathcal{P}, y \in \mathcal{C}$.

Assume $\Pr(x), \Pr(y) > 0$ for all x, y. From Bayes' formula we get $\Pr(y|x) = \Pr(y) > 0$.

Hence, for any fixed x there exists for any y a K such that $e_K(x) = y$.

In conclussion: $|\mathcal{P}| \leq |\mathcal{C}| \leq |\mathcal{K}|$.

Theorem 2.4
Assume $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$. The cryptosystem provides perfect secrecy iff:

1. $\forall K \in \mathcal{K} : \Pr(K) = \frac{1}{|\mathcal{K}|}$.
2. $\forall x \in \mathcal{P}\ \forall y \in \mathcal{C}\ \exists! K \in \mathcal{K} : e_K(x) = y$.

Proof:
To see the $\Uparrow$-part adapt proof of Th. 2.3.
Assume perfect secrecy. As already noted for every x, y there exists a K with $e_K(x) = y$. In other words

$$\mathcal{C} = \{e_K(x) : K \in \mathcal{K}\}.$$

But $|\mathcal{C}| = |\mathcal{K}|$ and therefore no two different keys map x to same y.

Proof cont.

Write $\mathcal{K} = \{K_1, \dots, K_n\}$ and $\mathcal{P} = \{x_1, \dots, x_n\}$. Given fixed y assume w.l.o.g.

$$e_{K_1}(x_1) = y, \dots, e_{K_n}(x_n) = y.$$

From Bayes' formula we get

$$\begin{aligned} \Pr(x_i|y) &= \frac{\Pr(x_i)\Pr(y|x_i)}{\Pr(y)} \\ &= \frac{\Pr(x_i)\Pr(K_i)}{\Pr(y)}. \end{aligned}$$

That is, for any fixed y $\Pr(K_i) = \Pr(y)$ for $i = 1, \dots, n$.

But then $\Pr(K_i) = \frac{1}{n}$.

One-time pad

Let $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_2^n$.

Let $K \in \mathcal{K}$ be chosen equiprobable.

For $K = (K_1, \ldots, K_n)$ and $x = (x_1, \ldots, x_n)$ define

$$e_K(x) = (x_1 + K_1, \ldots, x_n + K_n) \mod 2.$$

Decoding similar.

Entropy

Given $\mathbf{X}$ with $\mathcal{X} = \{x_1, \ldots, x_n\}$ the entropy is

$$H(\mathbf{X}) = -\sum_{x \in \mathcal{X}} \Pr(x) \log_2 \Pr(x).$$

The entropy is a meassure for the uncertainty of the outcome of $\mathbf{X}$.

Theorem: $0 \leq H(\mathbf{X}) \leq \log_2 n$.

The extreme cases being:

- $H(\mathbf{X}) = 0$ iff $\Pr(x_i) = 1$ for some i.
- $H(\mathbf{X}) = \log_2(n)$ iff $\Pr(x_1) = \cdots = \Pr(x_n) = \frac{1}{n}$.

Given random variables **X** and **Y** then $(\mathbf{X}, \mathbf{Y})$ is also a random variable.

Theorem:

$$H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) + H(\mathbf{Y})$$

with equality iff **X** and **Y** are independent.

Uncertainty is maximal iff **X** does not reveal anything about **Y** and vice versa.

Given $y \in \mathcal{Y}$ consider $\Pr(x_1|y), \ldots, \Pr(x_n|y)$. The corresponding entropy is

$$H(\mathbf{X}|y) = -\sum_{x \in \mathcal{X}} \Pr(x|y) \log_2 \Pr(x|y)$$

which is the uncertainty of **X** given the information that $\mathbf{Y} = y$ holds.

The average of $H(\mathbf{X}|\mathbf{Y})$ taken over all y is

$$H(\mathbf{X}|\mathbf{Y}) = -\sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \Pr(x|y) \log_2 \Pr(x|y)$$

which is the average uncertainty of **X** when **Y** is observed.

Theorem: $H(\mathbf{X}, \mathbf{Y}) = H(\mathbf{Y}) + H(\mathbf{X}|\mathbf{Y})$,

Theorem: $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ (it does not hurt to know **Y**.)

Given $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ then $H(\mathbf{K}|\mathbf{C})$ meassures the uncertainty of the key when the cipher text is observed.

If $H(\mathbf{K}|\mathbf{C}) = 0$ then the cipher text always reveals the key.

Theorem: $H(\mathbf{K} \mid \mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C})$.

Proof:

- $H(\mathbf{K}, \mathbf{P}, \mathbf{C}) = H(\mathbf{C}|\mathbf{K}, \mathbf{P}) + H(\mathbf{K}, \mathbf{P})$ by theorem above.
- $H(\mathbf{C}|\mathbf{K}, \mathbf{P}) = 0$ as $y = e_K(x)$.
- $H(\mathbf{K}, \mathbf{P}) = H(\mathbf{K}) + H(\mathbf{P})$ as independent.

Hence, $H(\mathbf{K}, \mathbf{P}, \mathbf{C}) = H(\mathbf{K}) + H(\mathbf{P})$.

- $H(\mathbf{P} \mid \mathbf{K}, \mathbf{C}) = 0$ as $x = d_K(y)$.

Hence, $H(\mathbf{K}, \mathbf{P}, \mathbf{C}) = H(\mathbf{K}, \mathbf{C})$.

In conclussion:

$$\begin{aligned} H(\mathbf{K}|\mathbf{C}) &= H(\mathbf{K}, \mathbf{C}) - H(\mathbf{C}) \\ &= H(\mathbf{K}, \mathbf{P}, \mathbf{C}) - H(\mathbf{C}) \\ &= H(\mathbf{K}) + H(\mathbf{P}) - H(\mathbf{C}). \end{aligned}$$

Knowing the language

Recall, crypto analysis is about revealing K.

Assume plaintext is a natural language $\mathcal{L}$.

If Oscar sees cipher text then from knowledge about the language he may rule out some keys. The left keys, except the correct one, are called spurious keys.

English language: $H(\mathbf{P}) \simeq 4.19$.

But some digrams, trigrams (or even books) are more common than others.

$\mathbf{P}^n$ (text of length n).

$$H_{\mathcal{L}} = \lim_{n\to\infty} \frac{H(\mathbf{P}^n)}{n}$$

is called the entropy of language $\mathcal{L}$.

A language with letters distributed equiprobable would have entropy $\log_2 |\mathcal{P}|$. Hence, the fraction of redundancy in $\mathcal{L}$ is

$$R_{\mathcal{L}} = \frac{\log_2 |\mathcal{P}| - H_{\mathcal{L}}}{\log_2 |\mathcal{P}|} = 1 - \frac{H_{\mathcal{L}}}{\log_2 |\mathcal{P}|}.$$

Study of English text yields $1.0 H_{\mathcal{L}} \leq 1.5$.

Assuming $H_{\mathcal{L}} = 1.25$ gives fraction of redundancy $R_{\mathcal{L}} \simeq 0.75$.

This means that using Huffman coding one could compress English text by a factor four.

Estimating number of spurious keys

Probability distribution on $\mathcal{K}$ and $\mathcal{P}^n$ induces probability distribution on $\mathcal{C}^n$.

Given $\vec{y} \in \mathcal{C}^n$ let

$$K(\vec{y}) = \{K \in \mathcal{K} : \exists \vec{x} \in \mathcal{P}^n \text{ with } \Pr(\vec{x}) > 0 \text{ and } e_K(\vec{x}) = \vec{y}\}.$$

If $\vec{y}$ is observed then the number of spurious keys are $|K(\vec{y})| - 1$.

Average number of spurious keys when plain text is n long is called $\bar{s}_n$.

$$\begin{aligned}
\bar{s}_n &= \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y})(|K(\vec{y})| - 1) \\
&= \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y})|K(\vec{y})| - \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y}) \\
&= \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y})|K(\vec{y})| - 1.
\end{aligned}$$

- $H(\mathbf{K}|\mathbf{C}^n) = H(\mathbf{K}) + H(\mathbf{P}^n) - H(\mathbf{C}^n)$ (Th. 2.10)
- $H(\mathbf{P}^n) \simeq nH_{\mathcal{L}} = n(1 - R_{\mathcal{L}}) \log_2 |\mathcal{P}|$ (Definition of $H_{\mathcal{L}}$.)
- $H(\mathbf{C}^n) \leq \log_2 |\mathcal{C}|^n = n \log_2 |\mathcal{C}|$.

Hence, if $|\mathcal{C}| = |\mathcal{P}|$ then

$$H(\mathbf{K}|\mathbf{C}^n) \geq H(\mathbf{K}) - nR_{\mathcal{L}} \log_2 |\mathcal{P}|. \tag{1}$$

$$\begin{aligned} H(\mathbf{K}|\mathbf{C}^n) &= \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y}) H(\mathbf{K}|\vec{y}) \\ &\leq \sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y}) \log_2 |K(\vec{y})| \\ &\leq \log_2 \Big(\sum_{\vec{y} \in \mathcal{C}^n} \Pr(\vec{y}) |K(\vec{y})| \Big) \\ &= \log_2(\bar{s}_n + 1). \end{aligned} \tag{2}$$

If keys are chosen equiprobable then $H(\mathbf{K}) = \log_2 |\mathcal{K}|$. Eqs. (1) and (2) then give

$$\bar{s}_n + 1 \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_{\mathcal{L}}}}.$$

For n big enough this is taken as an estimate.

Substitution cipher applied to English text: If $n \simeq 25$ the approximately 0 spurious keys.

Product of crypto systems

Given
$S_1 = (\mathcal{P}, \mathcal{C} = \mathcal{P}, \mathcal{K}_1, \mathcal{E}_1, \mathcal{D}_1)$,
$S_1 = (\mathcal{P}, \mathcal{C} = \mathcal{P}, \mathcal{K}_2, \mathcal{E}_2, \mathcal{D}_2)$.

Define
$S_1 \times S_2 = (\mathcal{P}, \mathcal{C} = \mathcal{P}, \mathcal{K}_1 \times \mathcal{K}_2, \mathcal{E}, \mathcal{D})$
with
$e_{(K_1,K_2)}(x) = e_{K_2}(e_{K_1}(x))$.

If $S \times S = S$ then called idempotent (NOT interesting).

Examples of idempotents are: Shift ciphers, Hill ciphers, affine ciphers, substitution ciphers, Vigenére ciphers, permutation ciphers.

BUT, combinations of two DIFFERENT of the above ciphers may be interesting.

Iterated cipher

Consider cypto system $(\mathcal{P}, \mathcal{C} = \mathcal{P}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ which is not idempotent (can itself be a product of two different idempotents).

Given a "key" construct from this a key schedule $K^1, \ldots, K^{Nr} \in \mathcal{K}$.

Write $g(x, K) = e_K(x)$ and encode as follows:

$$\begin{array}{rcl}
w^0 & \leftarrow & x \\
w^1 & \leftarrow & g(w^0, K^1) \\
w^2 & \leftarrow & g(w^1, K^2) \\
 & \vdots & \\
w^{Nr-1} & \leftarrow & g(w^{Nr-2}, K^{Nr-1}) \\
w^{Nr} & \leftarrow & g(w^{Nr-1}, K^{Nr}) \\
y & \leftarrow & w^{Nr}.
\end{array}$$

Decoding: Start from the bottom.

SPN

g is build up by substitution, permutation and XOR with key (from key schedule).

Example: $\mathcal{P} = \mathcal{C} = \mathbb{Z}_2^{16}$.

S-box: Divide block of size 16 into four blocks of size four. Each block is modified by applying the substitution $\pi_S : \mathbb{Z}_2^4 \to \mathbb{Z}_2^4$.

Permutation (of positions in entire block): Apply the permuation $\pi_P : \mathbb{Z}_2^{16} \to \mathbb{Z}_2^{16}$.

Initialization: $w^0 = (x_1, \ldots, x_{16})$.

Updating: For $i = 1, \ldots, 4$ (w^4 is not used)

$$\begin{array}{ccc} u^i & w^{i-1} \oplus K^i & \\ v^i & = & S(u^i) \\ w^i & = & \pi_p(v^i) \end{array}$$

Finalization: $y = v^4 \oplus K^5$

Picture from Stinson's book

Preparing for crypto analysis of SPN

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be independent binary random variables:

$$\begin{aligned} p_i &= \Pr(\mathbf{X}_i = 0) \\ 1 - p_i &= \Pr(\mathbf{X}_i = 1) \end{aligned}$$

Denote by $\epsilon_i = p_i - 0.5$ the bias of the distribution of $\mathbf{X}_i$.

Examples:
If $p_i = 0.5$ then $\epsilon_i = 0$.
If $p_i = 0$ then $\epsilon_i = -0.5$.
If $p_i = 1$ then $\epsilon_i = 0.5$.

Piling-up Lemma: Let $\epsilon_{i_1}, \ldots, \epsilon_{i_k}$ denote the bias of independent binary variables $\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_k}$. The bias of $\mathbf{X}_{i_1} \oplus \cdots \oplus \mathbf{X}_{i_k}$ equals

$$\epsilon_{i_1,\ldots,i_k} = 2^{k-1} \prod_{j=1}^{k} \epsilon_{i_j}.$$

Proof: By induction.

The S-box from our SPN is given in Table 3.1 of Stinson's book.

$\Pr(X_1 \oplus X_4 \oplus Y_2 = 0) = \frac{8}{16}$, that is bias=0.

$\Pr(X_3 \oplus X_4 \oplus Y_1 \oplus Y_2 = 0) = \frac{2}{16}$, that is bias=$-\frac{3}{8}$.

This kind of information will be used in linear attack on SPN.

From Stinson's book

Linear attack

$$
\begin{array}{lcll}
T_1 & = & U_5^1 \oplus U_7^1 \oplus U_8^1 \oplus V_6^1 & \text{bias is } 0.25 \\
T_2 & = & U_6^2 \oplus V_6^2 \oplus V_8^2 & \text{bias is } -0.25 \\
T_3 & = & U_6^3 \oplus V_6^3 \oplus V_8^3 & \text{bias is } -0.25 \\
T_4 & = & U_{14}^3 \oplus V_{14}^3 \oplus V_{16}^3 & \text{bias is } -0.25
\end{array}
$$

The variables T_1, T_2, T_3, T_4 are not independent. Even so, we use the piling lemma. We get that the bias of $T_1 \oplus T_2 \oplus T_3 \oplus T_4$ is $-1/32$.

Rewriting we get

$$\begin{aligned} T_1 \oplus \cdots \oplus T_4 &= X_5 \oplus X_7 \oplus X_8 \oplus U_6^4 \oplus U_8^4 \oplus U_{14}^4 \oplus U_{16}^4 \oplus K_5^1 \oplus K_7^1 \\ &\oplus K_8^1 \oplus K_6^2 \oplus K_6^3 \oplus K_{14}^3 \oplus K_6^4 \oplus K_8^4 \oplus K_{14}^4 \oplus K_{16}^4. \end{aligned}$$

For fixed (unknown key) we get that the bias of

$$X_5 \oplus X_7 \oplus X_8 \oplus U_6^4 \oplus U_8^4 \oplus U_{14}^4 \oplus U_{16}^4$$

is $1/32$ or $-1/32$.
For every guess of a key we can calculate U_i^4 from cipher text (the value of U_i^4 will be correct if we guess the right key).
For a not too small sample of plain text/chipher text estimate the bias of $X_5 \oplus X_7 \oplus X_8 \oplus U_6^4 \oplus U_8^4 \oplus U_{14}^4 \oplus U_{16}^4$ for every combination of values of $K_5^5, K_6^5, K_7^5, K_8^5, K_{13}^5, K_{14}^5, K_{15}^5, K_{16}^5$.

Choose, the combination with bias approximately $1/32$ or $-1/32$.

DES and AES

DES (used to be the standard). AES (becoming the standard).

DES uses the Feistel cipher:
Divide stage u_{i-1} into (L^{i-1}, R^{i-1}).
$(L^i, R^i) = g(L^{i-1}, R^{i-1}, K^i)$ where

$$\begin{aligned} L^i &= R^{i-1} \\ R^i &= L^{i-1} \oplus f(R^{i-1}, K^i) \end{aligned}$$

Note, that f needs not be invertible. In DES the function f involves substituion and permutation.

AES not Feistel cipher. For the substitution we use the inverse in $\mathbb{Z}_{2^8}$. This map is a socalled "almost non-linear map" which protects against differential attacks.

Is AES volnourable to algebraic attacks? No real success with algebraic attacks yet.