応用数学６ 2010 年 9 月 29 日

秘匿通信問題と現代暗号

講師: 安永憲司

1 秘匿通信問題

Alice と Bob，二者間の通信を考える．二人は，安全でない通信路を使って，秘密にメッセージを伝えたい．安全でない通信路とは，盗聴可能な，公開された通信路のことである．Alice と Bob は，通信内容すべてが Eve に盗み見られたとしても，メッセージを秘匿したいと考える．

どのようにすれば秘匿通信を達成できるだろうか．

2 古典暗号

一つの解決方法として，次のような方法が考えられる．Alice と Bob は，「秘密のコード」をあらかじめ共有しておき，それを通信の際に使う．秘密のコードとは，鍵，暗号化方法，復号方法の三つで構成される．Alice は，メッセージを送りたいとき，そのメッセージを鍵を使って暗号化し，暗号文を Bob に送る．Bob は受け取った暗号文を，鍵を使って復号し，メッセージを復元する．

2.1 秘密鍵暗号方式

上記の方法を，定式化する．暗号化方法とは，アルゴリズム Enc のことであり，復号方法とは，アルゴリズム Dec のことであると考える．さらに，鍵 k は，アルゴリズム Gen によって生成されると考える．

このとき，どの情報を秘密に，そしてどの情報を公開する必要があるのだろうか．古典的な方法では，$\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec}$ そして生成された鍵 k すべてを秘密にする．それは，情報をなるべく敵に与えない方が，安全であるからである．しかし，1883 年に Kerckhoffs によって提唱された，**Kerckhoffs の原理**では，秘密であると仮定するのは鍵 k だけであり，他のすべては公開されるべきだと主張している．なぜこのような考え方をするのだろうか．一つの理由として，暗号化アルゴリズムというのは結果として漏れてしまうことが多いからである．また別の理由として，アルゴリズムが公開されている暗号化方式が未だに破られていなければ，(アルゴリズムが公開されず，少人数の人たちが安全だと主張する方式よりも）その方式が安全であるという信頼を与えることができるからである．

Kerckhoffs の原理から，$\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec}$ すべてを決定的なアルゴリズムにすることはできないことがわかる．もしすべてが決定的アルゴリズムであれば，Alice と Bob のもつ情報すべてを Eve が知ることができ，秘匿通信は実現できない．少なくとも，鍵を生成する Gen は確率的アルゴリズムでなくてはならない．

定義 1 (秘密鍵暗号方式) 三つ組のアルゴリズム $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ がメッセージ空間 $\mathcal{M}$，鍵空間 $\mathcal{K}$ 上の**秘密鍵暗号方式**であるとは，以下を満たすときである．

1. Gen は**鍵生成アルゴリズム**であり，鍵 $k \in \mathcal{K}$ を出力する確率的アルゴリズムである．
2. Enc は**暗号化アルゴリズム**であり，鍵 $k \in \mathcal{K}$ とメッセージ $m \in \mathcal{M}$ を入力とし，暗号文 c を出力する（確率的）アルゴリズムである．
3. Dec は**復号アルゴリズム**であり，鍵 k と暗号文 c を入力とし，メッセージ m を出力する決定的アルゴ

リズムである．

4. 任意の $m \in \mathcal{M}$ に対して，

$$\Pr[\mathsf{Dec}_k(\mathsf{Enc}_k(m)) = m \mid k \leftarrow \mathsf{Gen}] = 1.$$

上の定義では，安全性については触れていない．

2.2 古典暗号の例

古典暗号として，Ceasar 暗号と換字暗号を紹介する．

定義 2 (Ceasar 暗号) Ceasar 暗号は次の $(\mathcal{M}, \mathcal{K}, \mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ で定義される．

$$\begin{aligned}
\mathcal{M} &= \{A, B, \ldots, Z\}^* \\
\mathcal{K} &= \{0, 1, 2, \ldots, 25\} \\
\mathsf{Gen} &= k;\ \text{ただし } k \xleftarrow{R} \mathcal{K} \\
\mathsf{Enc}_k(m_1 m_2 \ldots m_n) &= c_1 c_2 \ldots c_n;\ \text{ただし } c_i = m_i + k \bmod 26 \\
\mathsf{Dec}_k(c_1 c_2 \ldots c_n) &= m_1 m_2 \ldots m_n;\ \text{ただし } m_i = c_i - k \bmod 26
\end{aligned}$$

ただし，アルファベット $\{A, B, \ldots, Z\}$ は，整数値 $\{0, 1, \ldots, 25\}$ に，それぞれ対応していると考える．

簡単に述べると，暗号化では，メッセージの各文字を，長さ k だけシフトし，復号のときは，それを元に戻している．

Ceasar 暗号では，鍵 k を知らなければ，文字列がランダムに並べ替わっているように見える．しかし，k として 26 通りすべてを試せば，意味のあるメッセージが復元されたかどうかが確認できてしまう．もしメッセージがある程度長い文書であれば，この方式は簡単に破れてしまう．このような攻撃は，**全数探索攻撃**と呼ばれる．

定義 3 (換字暗号) 換字暗号は次の $(\mathcal{M}, \mathcal{K}, \mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ で定義される．

$$\begin{aligned}
\mathcal{M} &= \{A, B, \ldots, Z\}^* \\
\mathcal{K} &= \{A, B, \ldots, Z\} \text{ 上の置換の集合} \\
\mathsf{Gen} &= k;\ \text{ただし } k \xleftarrow{R} \mathcal{K} \\
\mathsf{Enc}_k(m_1 m_2 \ldots m_n) &= c_1 c_2 \ldots c_n;\ \text{ただし } c_i = k(m_i) \\
\mathsf{Dec}_k(c_1 c_2 \ldots c_n) &= m_1 m_2 \ldots m_n;\ \text{ただし } m_i = k^{-1}(c_i)
\end{aligned}$$

換字暗号では，鍵が 26! 通り存在するので，全数探索攻撃は難しいと考えられる．しかし，英文におけるアルファベットの出現頻度を利用すると，ある程度簡単に破られてしまう．

それでは，次は，どうすればよいのか．暗号は，歴史的に，以下のサイクルに従って発展してきた．

1. ある人が暗号方式を提案
2. 提案者は，既存の攻撃ではその方式を破ることはできないと主張
3. その暗号方式が利用されるようになる
4. 攻撃方法の改良によってその方式が破られる
5. 改良された攻撃に対処するように方式を修正

歴史的には，暗号従事者の主な仕事は，暗号解読であった．暗号解読自体は現在でも重要な研究分野であるが，現代暗号は，「暗号の設計を正しく行えば，暗号解読は必要ない」という理念をもっている．

3 現代暗号

現代暗号では，暗号の設計を，芸術から科学に転換させた．その場しのぎの閃きによって発明されるものではなく，現代暗号は次のパラダイムにもとづいている．

- 安全性の定義を数学的に与える．
- 数学的仮定を正確に述べる（例えば，「素因数分解が難しい」など．もちろん，「難しい」を数学的に定義する必要がある）．
- 安全性の証明を与える．言い換えると，もしある方式が破れたとしたら，それは仮定に反する，ということを証明する．つまり，その仮定が正しければ，その方式は破られることはない．

この講義は，この考え方に従って進めていく．

記法

アルゴリズム

A をアルゴリズムとする．入力が一つの場合，$A(\cdot)$ と書き，二つの場合，$A(\cdot,\cdot)$ と書く．$A(x)$ は，アルゴリズム A に，入力 x を与えたときの出力を表す．

入力が与えられると，出力が一つに決まるアルゴリズムを決定的アルゴリズムと呼ぶ．決定的アルゴリズムの場合，$A(x)$ はある一つの値を表す．確率的アルゴリズムは，アルゴリズムの内部でコインを投げることができ，動作が確率的になる．確率的アルゴリズムでは，出力が確率的であるため，$A(x)$ は，アルゴリズム A に，入力 x を与えたときの出力の確率分布を表すものとする．確率的なアルゴリズムは，入力の他に，一様乱数系列を入力として受け取ることができる決定的アルゴリズムだと解釈することができる．

実験

確率分布 D に対して，D に従ってサンプルして x を得る実験を $x \leftarrow D$ と書く．F が有限集合のとき，$x \xleftarrow{R} F$ は，集合 F から一様ランダムにサンプルして x を得る実験を表す．

確率

表記を簡潔にするため，確率分布と確率変数を区別なく使うことがある．例えば，確率分布 D からサンプルして，x が得られる確率を

$$\Pr[D = x]$$

と表す．この場合，D は確率変数であるとみなす．この確率は，

$$\Pr_{x' \leftarrow D}[x = x']$$

と表すこともできる．また，

$$\Pr_{x \leftarrow D}[f(x) = y],\ \Pr[f(D) = y]$$

はいずれも，確率分布 D からサンプルした値を関数 f で評価した値が，y になる確率を表す．確率的アルゴリズム A に入力として x を与え，その出力値を関数 f で評価した値が z になる確率は，

$$\Pr_{y \leftarrow A(x)}[f(y) = z],\ \Pr_{A}[f(A(x)) = z]$$

のように表す．また，表記を簡潔にするため，Pr の下の添字を省略することもある．その場合，どのように確率を取っているかをできるだけ明記するが，明らかな場合などは省略をすることがある．