中山大学学报 (自然科学版)
第 37卷 第 5期 ACTA SCIENTIARUM NATURALIUM Vol. 37 No. 5
1998年 9月 UNIVERSITATIS SUNYATSENI Sep. 1998

一种基于概率分析的密码系统保密测度*

欧振猛 余顺争
(中山大学 无线电电子学系, 广州 510275)

摘 要 在网络数据传输过程中,为了数据安全性的考虑,均采用密码系统来实现. 但任何一种密码系统都不可能是不可破译的,需要在对现有的密码系统不断分析过程中进行改进. 本文在理论上分析了一种抽象的密码系统的破译概率,经过推导得出了最终表达式.

关键词 密码系统 唯一解码量 保密测度
分类号 TN 918

在信息时代,计算机网络的发展为数据通信事业开辟了广阔的前景,同时也对信息的安全保密传输提出了更苛刻的要求,通过将密码技术引入计算机网络中,从而确保信息的保密和系统的完备,这就构成了当前的数据保密系统. 密码技术中的核心部分[1]是设计一个能使数据保密的密码算法. 实际上,任何一种密码算法都不可能使数据完全保密,这使得衡量密码系统的保密程度成为一项重要工作. 本文是选用传统密码系统 (密码系统中的加密密钥和脱密密钥是相同的) 作为分析对象,根据概率测度的方法,在给定密文量的条件下,分析密文可被破译的概率.

1 密码分析的假设

通过给密钥和报文赋予概率,并对加脱密变换、报文、密文以及密钥作某些简单的假设,就可以对密码进行数学分析 (图 1).

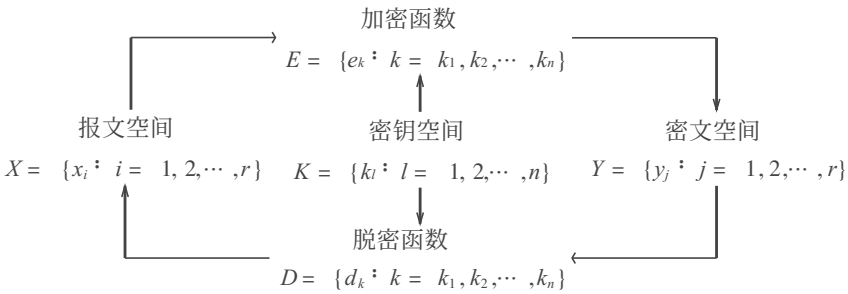

图 1 一种密码系统的结构示意图

Fig. 1 The graph of certain kind of cryptosystem structure

* 收稿日期: 1998-01-30 欧振猛, 男, 23岁, 研究生

(1) 报文空间 ($X=\{x_i: i=1,2,\cdots,r\}$)是由 r 个具有出现概率分别为 $p(x_1),p(x_2),\cdots,p(x_r)$的报文 x_1, x_2, …, x_r 组成的有限集合，报文 (X) 空间被分成两个集合: ① 由 s 个意义明显的，或有意义的报文组成的集合，用 X' 来表示. ② 由 $r-s$ 个无意义的报文组成的集合，用 X'' 来表示.

为了数学分析上的简单起见，假设分析人员预先不知道报文的内容，X' 中的每一份报文都具有相等的概率: $p(x)=1/s$, 对于 X' 中的每一个 x.

(2) 密钥空间 ($K=\{k_l: l=1,2,\cdots,n\}$)是由 n 个具有出现概率分别为 $p(k_1),p(k_2),\cdots,p(k_n)$的密钥 k_1, k_2, …, k_n 组成的有限集合，假设密钥是随机选取的，或者选取的过程是未知的，那么分析人员就要给每一个密钥赋予相等的概率: $p(k)=1/n$, 对 K 中的每一个 k.

(3) 加密函数 ($E=\{e_k: k=k_1,k_2,\cdots,k_n\}$)是一个从报文空间 ($X$)映射到密文空间 ($Y$)的所有一对一单映射函数集合中随机地、独立地挑选出来而组成的有限加密函数集合.

(4) 脱密函数 ($D=\{d_k: k=k_1,k_2,\cdots,k_n\}$)是一个从密文空间 ($Y$)映射到报文空间 ($X$)的一对一单映射脱密函数的有限集合，并且每一个 d_k 是相应的 e_k 的逆函数.

(5) 密文空间 ($Y=\{y_j: j=1,2,\cdots,r\}$)对于每一个密钥 ($K$ 中的 k)和报文 (X 中的 x_i)，都具有一个使 $e_{k_j}(x_i)=y_j$ 的密文 (Y 中的 y_j). 这样，密文空间 (Y) 可以分为两个集合: ① 可能的密文的集合 (Y'，那些至少可以从一份有意义的报文产生出来的密文)，用 Y' 来表示，$Y'=\{e_k(x): k\in K, x\in X'\}$; ② 不可能的密文的集合 ($Y''$，只能从无意义的报文中产生出的密文)，$Y''=Y-Y'$. 其中，$Y-Y'$ 定义为在 Y 中而不在 Y' 中的那些元素.

2 密码测度概率的推导

在上面的假设之后，令 y 代表一份截获的密文，即是集合 Y' 中的一个元素，再定义一些定量描述保密测度的变量.

M 是一个随机变量，为脱密一份截获密文 (y_j)，并使之成为有意义报文的所试用的所有密钥的数量. M' 也是一个随机变量，为除原先用来产生给定密文的那个密钥之外，其它可以脱密所截获的密文并使之成为有意义报文的密钥的数量 ($M'=M-1$).

M 的概率分布是分析人员最感兴趣的东西，因为 M 可以从中估算出成功地求解正确密钥的概率，记为 $p(sk)$ (其中，s 代表成功，k 代表密钥). 定义 M' 的目的是为了计算出 M 的概率分布.

假定将用 m 个不同的密钥来脱密一份截获的密文，并使之产生出有意义的报文. 由于这 m 个密钥中任何一个都可能是正确的密钥 (即原先用来产生给定密文的那个密钥)，从这个集合中猜出或随机地选出正确密钥的概率为: $p(sk|M=m)=1/m$

由于 m 可以是 1, 2, …, n 的数 (n 是全部密钥的总数，见图 1)，所以获得正确密钥的概率为

$$p(sk)=\sum_{m=1}^{n}p(sk|M=m)p(M=m)=\sum_{m=1}^{n}(1/m)p(M=m) \tag{1}$$

式中，$p(M=m)$ 是 $M=m$ 的概率. 下面讨论 $p(M=m)$ 的表达式.

根据上述，y 为截获的密文，它是由一个未知的有意义报文 x 加密而来的. 截获了 y 的

敌手仅知道 x 是中 X' 的一个元素, y 是 Y' 中的一个元素. 假定 k_j 是原先用来将 x 加密为 y 的那一个特定密钥, 可利用 k_j 来脱密 y 以得到有意义报文的概率为 1.

在上面的假设中, 密码系统的加密函数是随机、独立选择出来的, 因此用 $(n-1)$ 个不正确的密钥: $k_1, k_2, \cdots, k_{j-1}, k_{j+1}, \cdots k_n$ 中的每一个来脱密 y 的过程, 可以考虑是 $(n-1)$ 次的伯努利试验[2]的随机过程, 其中 s/r 代表一个密钥成功地将 y 脱密成有意义报文的概率, $1-(s/r)$ 是一个密钥不能将 y 脱密有意义报文的概率: 对于所有不等于 j 的 i, $p[d_{ki}(y) \in X'] = s/r$, $p[d_{ki}(y) \in X''] = 1-s/r$, 这样, M' 就具有二项分布的特性

$$p(M' = m') = \binom{n-1}{m'} (s/r)^{m'} (1-s/r)^{n-1-m'} \qquad m' = 0, 1, \cdots, n-1 \tag{2}$$

由于 m' 等于 $m-1$, 故有

$$p(M = m) = \binom{n-1}{m-1} (s/r)^{m-1} (1-s/r)^{n-m} \qquad m = 1, 2, \cdots, n \tag{3}$$

将 (3) 代入 (1) 得

$$p(sk) = \sum_{m=1}^{n} \frac{1}{m} \binom{n-1}{m-1} \left(\frac{s}{r}\right)^{m-1} \left(1-\frac{s}{r}\right)^{n-m}$$

结合二项式定理做适当变换可得

$$p(sk) = \frac{r}{ns}\left[1-(1-s/r)^n\right] \tag{4}$$

这就是成功地获得正确密钥的概率的值, $p(sk)$ 的精确结果取决于 n, r, s.

在破译实验中, 随着截获的密文字符长度的增大, 可以在数量为 r 的总报文空间中排除更多的无意义报文, 也就是有意义报文的数量 s 在减小, 从而获得正确密钥的概率越来越大, 即密文字符长度、破译概率、s/r 之间有密切的关系, 由于 s/r 是一个很小的变量, 对于定量讨论不便, 所以让它和一个大常数相乘, 而 n 是密钥总数, 是一个非常大的常数. 因此引入参量 λ ($\lambda = ns/r$) 以方便讨论.

香农[3]将密码的唯一解码量定义为使 λ 等于 1的 N 值, N 是从截获的密文解码成功得到原来报文所需的密文的字符长度. 针对一种具体的密码系统, 可以通过信息论的知识和马尔科夫概率过程计算出唯一解码量, 计算过程比较复杂, 限于篇幅, 这里不作分析.

当达不到唯一解码量, 对密钥的攻击一定不会成功; 超过了唯一解码量, 对密钥的攻击就可能会成功. 而现在从网络上获得的密文经常能突破唯一解码量的限制, 也就是说对于任何一种密码系统都能超过唯一解码量, 任何一种密码系统都可能被攻击破译, 而在满足唯一解码量条件下的攻击成功概率为:

对于 $p(sk)$ 来说, 由于 $\lambda/n = s/r \ll 1$, 且 $\lim_{x\to 0}(1+x)^{1/x} = e$, 有

$$\lim_{\lambda/n\to 0}(1-\lambda/n)^n = \lim_{-\lambda/n\to 0}\left[\left(1+(-\lambda/n)\right)^{-n\lambda}\right]^{-\lambda} = e^{-\lambda},$$

将 $\lambda = ns/r$ 代入 (4)式得 $p(sk)$ 的近似结果

$$p(sk) \approx (1/\lambda)(1-e^{-\lambda}) \tag{5}$$

可见它只与 λ 有关, 即成类似反比关系.

根据香农的唯一解码量的定义, 令 $\lambda = 1$, 此时成功破译密文得到密钥的概率为 0.6321. 这是一个分界点, 当密文长度大于唯一解码量时, 成功破译概率就很快逼近 1; 反之, 当密文长度小于唯一解码量时, 成功破译概率就很快逼近 0.

在英语简单代替[3]中, 有 $n = 26!$ 种方法可以将 26个字母的报文字母表转换成 26个字

母的密文字母表, 报文的总量 $r = 26^N$ (N 为接收的报文长度). 即在英语简单代替中, 唯一解码量为 22.2个字符, $s = 2^{2.02N}$.

根据(5)式和 λ 参量, 计算唯一解码量时的 $p(sk)$ 值数据(表 1).

表 1 在 N 接近唯一解码量时的 $p(sk)$ 值

Tab. 1 The value of $p(sk)$ when N approaches unicity distance

N	18.9	20.0	21.1	22.2	23.3	24.4	25.6
$p(sk)$	0.002 0	0.015 6	0.125 0	0.632 1	0.940 0	0.992 2	0.999 0

从表 1可见, 这种理论推导出来的分界点破译概率与实际公布的分界点破译概率数据是比较相符的, 而且也符合香农的唯一解码量的定义.

参 考 文 献

1 David J S, Sylvia M. 计算机网络安全奥秘. 程佩青等译. 北京: 电子工业出版社, 1994. 290~293

2 盛骤, 谢式千, 潘承毅, 等. 概率论和随机过程. 北京: 高等教育出版社, 1994. 36~39

3 H贝克, F派普尔. 密码体制通信保护. 四川: 通信保密编辑部, 1983. 102~104, 329~334

A Measure of Security in Cryptosystem Based on Probability Analysis

Ou Zhenmeng * *Yu ShunZheng*

Abstract Cryptosystem is widely used in the data transmission through network in order to protect the security of data. Since any cryptosystem may be broken, continuously enhancing the robustness of cryptographic algorithm is necessary. The authors analyses the abstract probability of a decrypted cryptosystem and final expression is derived after careful derivation.

Keywords cryptosystem, unicity distance, cryptologic measure

* Department of Radio and Electronics, Zhongshan University, Guangzhou 510275, China