文章编号：2095-4980(2023)03-0371-07

多信道数据碎片化传输安全性证明

陈世康[1]，郭　爽[1]，唐　晋[1]，袁　健[1]，林维涛[*2]，刘　丹[2]

(1.中国电子科技集团公司 第三十研究所，四川 成都 610000；2.电子科技大学 电子科学技术研究院，四川 成都 611731)

摘　要：基于信息数据碎片化的多道隔离传输，是在信息传输过程中保证其安全性的常用方法之一，但其缺乏可证明安全性理论和测评方法。本文形式化地定义了一个多道碎片化传输系统，包括数据的加密、碎片切分多道传输和重组功能。从多信道传输过程中数据泄露概率分析，定义多道碎片化传输技术的可证明安全等级。建立一套多信道数据碎片化传输的安全等级评估方法，从理论角度对于安全等级评估方法的有效性加以验证，为多信道数据碎片化传输技术的实现提供理论指导，同时为多信道数据碎片化传输系统的安全评测提供借鉴。该方法也适用于数据碎片化存储及传输相关领域的安全性分析。

关键词：网络安全；可证明安全性；多信道传输系统；碎片化技术

中图分类号：TP393.08　　**文献标志码：**A　　**doi：**10.11805/TKYDA2021179

Multi-channel data fragmentation transmission security proofs

CHEN Shikang[1]，GUO Shuang[1]，TANG Jin[1]，YUAN Jian[1]，LIN Weitao[*2]，LIU Dan[2]

(1.The 30th Research Institute of China Electronics Technology Group Corporation，Chengdu Sichuan 610000，China；
2.Research Institute of Electronic Science and Technology，University of Electronic Science and Technology of China，
Chengdu Sichuan 611731，China)

Abstract：Multi-channel isolation transmission based on fragmentation of information data is one of the common methods to ensure the security of information transmission，but it lacks the theory of provable security and evaluation methods. A multi-channel fragmentation transmission system is formally defined firstly，including data encryption，fragmentation multi-channel transmission and recombination functions. Then from the data leakage probability analysis in the process of multi-channel transmission，the provable security level of multi-channel fragmentation transmission technology is defined. A set of security level evaluation methods for multi-channel data fragmentation transmission is established to verify the effectiveness of the security level evaluation method from the theoretical point of view. This paper provides theoretical guidance for the realization of multi-channel data fragmentation transmission technology and guidelines for the security evaluation of multi-channel data fragmentation transmission system. The proposed method can also be applied to the security analysis of data fragmentation storage and transmission.

Keywords：cyber security；provable security；multi-channel；fragmentation technology

为实现大数据量的通信，多信道技术已经得到广泛应用，这些技术包括多输入多输出(Multiple-In Multiple-Out，MIMO)、利用光的轨迹角动量(Orbital Angular Momentum，OAM)进行多路复用和通过OAM进行直接调制等[1]。然而，随着攻击方式日益多样和系统结构越发复杂，使用传统方法保护信息可能代价高昂，需要新的视角和新的理论基础。

目前，已有大量研究通过对传输与存储的数据进行分片，以求获得更高的传输效率及安全性。文献[2]提出了基于数据分片多径传输(Data Fragmentation Multipath Transmission，DFMT)的安全自由空间光通信系统，将消息切分为碎片，通过不同的信道随机传输，但是并未阐述如何进行碎片切分。802.11提供了一种高效的分片方案[3]，旨在减少信道错误对于完整性的影响，但其缺点在于分片的大小固定，造成传输效率降低。在多信道传输

收稿日期：2021-04-26；**修回日期：**2021-06-09

[*]**通信作者：**林维涛　email:522138778@qq.com

的前提下，文献[4]建议将信息进行分段，以便截获少量片段的攻击者不能够获得任何有效信息。现有的数据碎片化技术旨在增强数据操作过程，减少处理时间，优化存储，增加操作灵活性，促进数据分发和传输，但没有专门设计数据安全的思想[5]。可证明安全是目前多数安全协议的设计手段，定义合适的安全目标、建立适当的敌手模型是讨论可证明安全性的前提条件[6]。在建立敌手模型时，将碎片化传输过程中的攻击者根据实际攻击类型划分为协作与非协作两类。信息熵由 Shannon 在文献[7]中提出，作为其拓展，Hastad 等在文献[8]中使用不可区分性对于伪熵进行定义。根据伪熵与伪随机性的关系和不可预测性与伪随机性等价[9]，将方案的安全目标定义为敌手通过已知信息预测出其他信道信息的概率，并且分析它们之间的强弱关系。

现有研究表明，给定一个随机序列，将其切分为 k 个不同的随机序列，而不损失每个碎片的熵，此任务困难的程度等价于猜测 k 位的停机问题[10]，考虑这个问题的弱化版本，即对于伪熵可加性的研究。熵描述了序列的信息量，通过对比整体序列的熵与其片段的熵，可以获知不同片段信息之间的独立关系，即通过某一片段信息无法预测剩余片段信息，那么是否能够通过伪熵的可加性判断此类问题。已知碎片化方案的安全目标通过信息的不可预测性进行定义，提出碎片化技术三级安全标准，利用伪熵的可加性对于前 2 级进行定义，最高级则使用伪随机性定义，并从理论角度对其有效性进行验证。

本文提出了一个碎片化系统，系统能够实现对于已加密数据的切分和重组功能。根据实际攻击类型将攻击者分为协作与非协作 2 类，提出了隔离安全性的定义，以敌手通过已知信息预测其他信道信息的概率大小区分隔离安全的强弱。提出 3 级碎片化技术安全评估等级，并对其有效性进行测评。

1 基础知识

一个从非负整数映射到非负实数的函数 $\varepsilon(n)$，若对于任意整数 $c\geqslant 1$，均存在 $n_0>0$，使得对于所有 $n>n_0$ 都有 $\varepsilon(n)<1/n^c$，称其为可忽略函数。

设 X 和 Y 是取值在有限集合 S 中的随机变量，定义 X 和 Y 的统计距离为：

$$\mathrm{Dist}(X,Y)=\frac{1}{2}\sum_{s\in S}|\Pr[X=s]-\Pr[Y=s]| \tag{1}$$

同样地，

$$\mathrm{Dist}(X,Y)=\max_{S'\subset S}|\Pr[X\in S']-\Pr[Y\in S']| \tag{2}$$

若 $\mathrm{Dist}(X,Y)\leqslant\varepsilon$，称 X 和 Y 是 ε–close 的。

设 $X=\{X_n\}_{n\geqslant 0}$ 与 $Y=\{Y_n\}_{n\geqslant 0}$ 为随机变量序列，对于每个 $n>0$，X_n 和 Y_n 取有限集合 S_n 中的值。如果 $\mathrm{Dist}(X,Y)$ 是可忽略的，则称 X 和 Y 是统计不可区分的(statistic indistinguishable)。出于计算目的，通常将集合 S_n 编码为多项式长度的位串。对于任意输出 0 或 1 的概率算法 A，将 A(关于 X 和 Y)的优势(advantage)定义为函数：

$$\mathrm{Dist}_A^{X,Y}(n)=\left|\Pr[A(1^n,X_n)=1]-\Pr[A(1^n,Y_n)=1]\right| \tag{3}$$

如果对于所有概率多项式时间算法 A, $\mathrm{Dist}_A^{X,Y}(n)$ 可忽略，则说 X 和 Y 是计算不可区分的(computational indistinguishable)。

若存在分布 $X=\{X_n\}_{n\geqslant 0}$ 与均匀分布 $\mathcal{U}=\{\mathcal{U}_n\}_{n\geqslant 0}$，对于所有概率多项式时间算法 A, $\mathrm{Dist}_A^{X,\mathcal{U}}(n)$ 可忽略，则称分布 $X=\{X_n\}_{n\geqslant 0}$ 是伪随机的(pseudo–random)。

$$\mathrm{Dist}_A^{X,\mathcal{U}}(n)=\left|\Pr\left[A\left(1^n,X_n\right)=1\right]-\Pr\left[A\left(1^n,\mathcal{U}_n\right)=1\right]\right| \tag{4}$$

定义 s 是不可预测的，若对于任意概率多项式时间算法 A 和任意正整数 n，存在可忽略函数 $\varepsilon(n)$ 满足：

$$\left|\Pr\left[A\left(s[1]\|s[2]\|\ldots\|s[n-1]\right)=s[n]\right]-\frac{1}{2}\right|\leqslant\varepsilon(n) \tag{5}$$

分布族 $D=\{D_n\}_{n\geqslant 0}$ 是伪随机的，当且仅当对任意字符串 $s\leftarrow D_n$，均是不可预测(unpredictable)的。

YAO Andrew C 在文献[9]中证明了不可预测性(unpredictability)和伪随机性对于位序列是等价的。若一个多项式时间算法 $G:\{0,1\}^n\rightarrow\{0,1\}^{l(n)}$ 为伪随机数生成器，当且仅当 $\forall n:l(n)>n$ 且 $\{G(U_n)\}_{n\in\mathbb{N}}$ 是一个伪随机分布序列。

根据 Shannon 在文献[7]中提出的信息熵(entropy)概念，设 x 为事件 $\mathbb{X}$ 的离散随机变量，x 发生的概率函数为 $p(x)=\Pr\{X=x\},x\in\mathbb{X}$。离散随机变量 X 的信息熵 $H(X)$ 定义为：

$$H(X)=-\sum_{x\in\mathbb{X}}p(x)\log p(x) \tag{6}$$

作为信息熵的推广，Hastad等[8]使用不可区分性定义了伪熵(pseudo-entropy)，假设存在2个分布$X=\{X_n\}_{n\geqslant 0}$和$Y=\{Y_n\}_{n\geqslant 0}$，分布X的熵为n。如果对于任意概率多项式时间算法A，$\mathrm{Dist}_A^{X,Y}(n)$可忽略，则不存在多项式区分算法能够区分分布X和Y，X的伪熵为n。

2　碎片化技术形式化描述

在数据外包的背景下，加密与数据碎片化的联合使用第一次被Aggarwal等提出[11]，提议通过在2个独立的数据库服务器上拆分信息和在必要时加密信息从而保证隐私要求。但此提议的主要限制在于，数据隐私依赖于2个服务器之间完全没有通信，在现实环境中，这个假设过于强大，并且服务器与用户之间的勾结很容易造成隐私被侵犯。Ciriani[12]提出的解决方案克服了Aggarwal等方案的局限性，首先将信息分割为不同片段，然后进行数据存储，最后将加密密钥提供给需要访问信息的授权用户。当前关注数据碎片的安全性方面最先进的方法依赖加密来确保数据的机密性[5]。故将碎片化技术形式化地抽象为数据的加密与切分重组过程，加密过程保证信息的机密性，碎片化过程专注于在单点数据泄漏的情况下限制对整体安全性的影响。

定义 1　数据碎片化方案。设有3个概率多项式时间的图灵机组成一个碎片化非对称密钥加密方案$\Psi=<\mathrm{Gen},\mathrm{Fra},\mathrm{Rec}>$，其中Fra, Rec分别为碎片化算法和碎片重组算法，Gen表示密钥生成算法，满足：

Fra和Rec利用密钥生成算法Gen和安全参数λ生成了密钥k：

$$\mathrm{Gen}:(1^{\lambda})\longmapsto k \tag{7}$$

Fra利用加密密钥k和消息$c\in\{0,1\}^*$作为输入，输出碎片化后的密文$(c_1,c_2,\cdots,c_l)$：

$$\mathrm{Fra}:(k,1^l c)\longmapsto(c_1,c_2,\cdots,c_l) \tag{8}$$

Rec利用解密密钥k和碎片化密文$(c_1,c_2,\cdots,c_l)\in\{0,1\}^*$作为输入，输出原始数据$c$：

$$\mathrm{Rec}:[k,(c_1,c_2,\cdots,c_l)]\longmapsto c \tag{9}$$

一个碎片化系统是正确的，如果对任意原始数据$c\in\{0,1\}^*$，经过加密后的待传输数据c再通过解密后应当与原始数据输入相同，即：

$$\mathrm{Rec}_k\left(\mathrm{Fra}_k\left(1^l,c\right)\right)=c \tag{10}$$

或

$$\Pr\left[\mathrm{Rec}_k(\mathrm{Fra}_k(1^l,c))\right]=1 \tag{11}$$

简记第i个碎片的碎片化算法$\mathrm{Fra}_{k,i}:c\longmapsto c_i$，给定$i,k\in\mathbb{N}$，定义分布$\mathrm{Fra}_{k,i}(\bullet)$的熵(entropy)为$H\left(\mathrm{Fra}_{k,i}\right)=H\left(c_i\right)$。碎片化算法的伪熵(pseudo-entropy)$PH\left(\mathrm{Fra}_{k,i}\right)=PH\left(c_i\right)\geqslant h$，如果存在一个分布$F$，满足$H(F)=h$，则$\mathrm{Fra}_i$和$F$不可区分，即：

$$\left|\Pr[\mathcal{D}(\mathrm{Fra}_{k,i}(c))=1]-\Pr[\mathcal{D}(F(c))=1]\right|\leqslant\varepsilon(\lambda) \tag{12}$$

同理，说碎片化算法的伪熵$PH\left(\mathrm{Fra}_{k,i}\right)<h$，如果对任意满足$H(F)=h$的函数$F$，$\mathrm{Fra}_i$和$F$是可区分的，即存在正整数$c$和一个多项式时间的区分器$\mathcal{D}$，使得：

$$\left|\Pr[\mathcal{D}(\mathrm{Fra}_{k,i}(c))=1]-\Pr[\mathcal{D}(F(c))=1]\right|>\lambda^{-c} \tag{13}$$

显然有$PH\left(c_i\right)\geqslant H\left(c_i\right)$成立，因为概率分布和自身一定是不可区分的。并且根据定义，若$\mathrm{Fra}_{k,i}$是伪随机的，那么其伪熵达到最大值，即$PH\left(c_i\right)=n$。

考虑到敌手的实际攻击方式，故将攻击者划分为2类：一类为可协作攻击者；另一类则为不可协作攻击者。另外考虑到截获片段的攻击者通过获得的片段计算剩余数据的能力，即在单点数据泄漏的情况下对整体安全性的影响，提出以任意攻击者通过以获取信息计算出其他任意信道上数据的概率为度量，以此界定信道隔离安全的强弱。

定义 2 多信道的隔离可证明安全。对于一个 l–信道数据传输上的碎片化方案 $\Psi=<\mathrm{Gen},\mathrm{Fra},\mathrm{Rec}>$，设信道中的数据分别为 $(c_1,c_2,\cdots,c_l)$，假设每个信道监听者 A_i 只能够获取对应信道的全部数据 c_i。定义监听者的可协作性：

1) 可协作攻击者：攻击者间可以分享监听到的数据；

2) 不可协作攻击者：攻击者间不可分享监听到的数据；

3) 进而定义攻击者的获取目标，即信道隔离安全的强弱。

信道隔离安全：任意攻击者 A_k 的优势

$$Adv_{\Psi,A_k}(\lambda)=\left|\Pr\left[A_k(c_k)=c_i\right]-1/2^{H(c_i)}\right| \tag{14}$$

表示计算出其他任意信道上的数据的概率与 $1/2^{H(c_i)}$ 的差。其中 $H(c_i)$ 是指数据 c_i 的信息熵。

若对于任意攻击者，其优势可忽略，则称其是信道隔离安全。强信道隔离安全：任意攻击者 A_k 的优势

$$Adv_{\Psi,A_k}(\lambda)=\left|\Pr\left[A_k(c_k)=c_i\right]-1/2^{|c_i|}\right| \tag{15}$$

表示计算出其他任意信道上的数据的概率与 $1/2^{|c_i|}$ 的差。其中 $|c_i|$ 表示数据 c_i 的长度。若对于任意攻击者，其优势可以忽略，则称其是信道强隔离安全。

一个 l–信道数据传输在不可协作攻击下是(强)安全的，如果满足对于任意常数 $p=1/2^{H(c_i)}$（$p=1/2^{|c_i|}$），任意非空信道数据 c_i 和任意攻击者 $A_k(k\neq i)$，均有：

$$\left|\Pr\left[A_k(c_k)=c_i\right]-p\right|\leqslant\varepsilon(\lambda) \tag{16}$$

称一个 l–信道数据传输在可协作攻击下是(强)安全的，如果满足对于任意常数 $p=1/2^{H(c_i)}$（$p=1/2^{|c_i|}$），任意非空信道数据 c_i 和任意攻击者 A_k（$k\neq i$），均有：

$$\left|\Pr\left[A_k(c_{-1})=c_i\right]-p\right|\leqslant\varepsilon(\lambda) \tag{17}$$

其中 $c_{-1}=\left(c_1,c_2,\cdots,c_{i-1},c_{i+1},\cdots,c_l\right)$。

若一个多信道安全传输是隔离安全的，则称其为多信道隔离安全传输。

3 安全等级定义及安全性证明

注意到，对熵的可加性研究已有定论，故对其弱化版本，即伪熵的可加性研究产生了兴趣，将碎片化方案安全评估等级分为安全、强安全和伪随机级别，前两级由伪熵的可加性进行定义，根据第 2 节所定义的安全目标和敌手模型，对于所提出的安全评估等级的有效性进行了验证。

定义 3 碎片化方案安全评估等级。设 $\Psi=<\mathrm{Gen},\mathrm{Fra},\mathrm{Rec}>$ 是一个碎片化系统，k 为由密钥生成算法 Gen 生成的密钥。称 Ψ 是安全的，若对于任意的数据 $c\in\{0,1\}^*$，任意正整数 $i,j\in\mathbb{N}$，$(i\neq j)$，有：

$$PH\left(c_i\right)+PH\left(c_j\right)=PH\left(c_i\|\,c_j\right) \tag{18}$$

称 Ψ 是强安全的，若对于任意的数据 $c\in\{0,1\}^*$，有：

$$\sum_{i=1}^{l}PH(c_i)=PH\left(\mathrm{Fra}_k(c)\right)=PH\left(c_1\|\,c_2\|\,\cdots\|\,c_l\right) \tag{19}$$

安全碎片化系统的定义使用伪熵来刻画，要求不同碎片拼接以后的伪熵，不比每个碎片的伪熵求和小。对于任意 2 个分布族 X,Y，满足：

$$PH(X)+PH(Y)\geqslant PH\left(X\|\,Y\right) \tag{20}$$

若存在分布 Z 与 $X\|\,Y$ 不可区分，且满足 $H(Z)=PH\left(X\|\,Y\right)$，注意到仅仅有可忽略概率下分布 Z 的样本长度和分布 $X\|\,Y$ 的不相等。令 $Z=Z_X\|\,Z_Y$，那么 $H(Z)=H\left(Z_X\|\,Z_Y\right)\leqslant H\left(Z_X\right)+H\left(Z_Y\right)$。显然有 X 与 Z_X 不可区分，而 Y 与 Z_Y 不可区分。从而

$$PH\left(X\|\,Y\right)=H(Z)\leqslant H\left(Z_X\right)+H\left(Z_Y\right)\leqslant PH(X)+PH(Y) \tag{21}$$

因此安全碎片化方案要求碎片化的过程不增加伪熵。

定理 1 若一个碎片化系统 Ψ 是强安全的，那么一定是安全的。

需要证明若：

$$\sum_{i=1}^{l} PH(c_i) = PH\left(\mathrm{Fra}_k(c)\right) \tag{22}$$

成立，那么对任意 $i,j \in \mathbb{N}$，$(i \neq j)$，有：

$$PH\left(c_i\right) + PH\left(c_j\right) = PH\left(c_i \| c_j\right) \tag{23}$$

不妨设

$$PH\left(c_1\right) + PH\left(c_2\right) > PH\left(c_1 \| c_2\right) \tag{24}$$

那么 $PH\left(\mathrm{Fra}_k(c)\right) \leqslant PH\left(c_1 \| c_2\right) + \sum_{i=3}^{l} PH(c_i) < \sum_{i=1}^{l} PH(c_i)$ (25)

给出一个更强的碎片化方案 Ψ 定义，利用伪随机性来刻画。事实上，不需要也不一定保证碎片中出现冗余的字段，哪怕在计算上是可以简单找到的，但各个碎片出现冗余字段并不一定影响其信道隔离安全性。

定义 4 伪随机性碎片化方案。设 $\Psi = <\mathrm{Gen}, \mathrm{Fra}, \mathrm{Rec}>$ 是一个碎片化系统，设 k 为密钥生成器 Gen 生成的密钥，对于任意的数据 $c \in \{0,1\}^*$，若 $\mathrm{Fra}_k(c)$ 是伪随机的，则称 Ψ 是伪随机的。

定理 2 若一个碎片化方案 Ψ 是伪随机的，则 Ψ 是强安全的。

证明：若一个碎片化系统 Ψ 是伪随机的，根据定义，其伪熵 $PH\left(\mathrm{Fra}_k(c)\right) = \left|\mathrm{Fra}_k(c)\right|$。又由于一个伪随机串的子串也是伪随机的，有 $PH\left(c_i\right) = |c_i|$，因此：

$$PH\left(\mathrm{Fra}_k(c)\right) = \sum_{i=1}^{l} PH(c_i) \tag{26}$$

即该碎片化系统是强安全的。

注意到若一个串是伪随机的，那么将该串各比特位做一个随机置换也是伪随机的，因此碎片在各信道的分配不影响其安全性。

方案实现的安全目标和敌手的攻击方式定义了一个密码方案的安全性，将隔离安全与强隔离安全作为方案所需要实现的安全目标，将协作攻击与非协作攻击作为敌手的攻击方式，接下来将对碎片化方案的 3 个安全级别进行证明。

定理 3 若一个多信道的碎片化系统是安全的，则该多信道数据传输在非协助攻击下是隔离安全的。

证明：考虑 l – 多信道，设第 i 信道中存在一个监听者为 A_i，$i \in (1,2,\cdots,l)$。假设 A_i 只能获取所在信道上的所有数据，记作 c_i，其中 $\mathrm{Fra}_k(1^l, c) \to (c_1, c_2, \cdots, c_l)$。

由于碎片化系统 Ψ 是安全的，即对于任意的数据 $c \in \{0,1\}^*$ 和任意正整数 $i,j \in \mathbb{N}$，有：

$$PH\left(c_i, c_j\right) = PH\left(c_i\right) + PH\left(c_j\right) \tag{27}$$

那么存在分布族 $Z = X_Z \| Y_Z$，满足：

$$H(Z) = H\left(X_Z \| Y_Z\right), H\left(X_Z\right) = PH\left(c_i\right), H\left(Y_Z\right) = PH\left(c_j\right) \tag{28}$$

并且 c_i 与 X_Z 不可区分，c_j 与 X_Y 不可区分，$c_i \| c_j$ 与 Z 不可区分。

以往证明在使用了安全的碎片化系统 Ψ 下，对于任意 $i \in \{1,2,\cdots,l\}$，A_i 无法获知其他的非空数据 $c_j (j \neq i)$，即：

$$\Pr\left[A_i\left(c_i\right) = c_j\right] \leqslant \frac{1}{2^{PH(c_j)}} + \varepsilon(\lambda) \leqslant \frac{1}{2^{H(c_j)}} + \varepsilon(\lambda) \tag{29}$$

使用反证法，若结论不成立，则存在算法 B 和多项式 $p(\cdot)$ 使得：

$$\Pr\left[B\left(c_i\right) = c_j\right] = \frac{1}{2^{PH\left(c_j\right)}} + \frac{1}{p(\lambda)} \tag{30}$$

那么可以利用算法 B 构造区分器 $\mathcal{D}$。设对于输入 $x \| y$，若 $B(x)=y$，则 $\mathcal{D}$ 输出 1，否则输出 0。那么

$$\Pr\left[\mathcal{D}\left(X_Z \| Y_Z\right)=1\right] \leqslant \frac{1}{2^{H(Y_Z)}} \tag{31}$$

此时对任意可忽略函数 $\varepsilon(\cdot)$，都有：

$$\left|\Pr\left[\mathcal{D}\left(c_i \| c_j\right)=1\right]-\Pr\left[\mathcal{D}\left(X_Z \| Y_Z\right)=1\right]\right|=\left|\frac{1}{2^{PH\left(c_j\right)}}+\frac{1}{p(\lambda)}-\frac{1}{2^{H\left(Y_Z\right)}}\right|=\frac{1}{p(\lambda)} \tag{32}$$

并非可忽略，这与假设矛盾。

定理 4 的证明本质上与定理 3 的证明相同。

定理 4 若一个多信道的碎片化系统是强安全的，则该多信道数据传输在协作攻击下是隔离安全的。

证明：考虑 $l-$多信道，设第 i 信道中存在一个监听者为 A_i，$i \in(1,2,\cdots,l)$。假设 A_i 只能获得所在信道上的所有数据，记作 c_i，其中 $c_{-i}=c_1 \| c_2 \| \cdots \| c_{i-1} \| c_{i+1} \| \cdots \| c_l$ 且 $(c_1,c_2,\cdots,c_l) \leftarrow \mathrm{Fra}_k(1^l,c)$。

由于碎片化系统 Ψ 是强安全的，即对于任意的数据 $c \in\{0,1\}^*$ 和任意正整数 $i,j \in \mathbb{N}$，有：

$$PH\left(c_{-i} \| c_j\right)=PH\left(c_{-i},c_j\right)=PH\left(c_{-i}\right)+PH\left(c_j\right) \tag{33}$$

那么存在分布族 $Z=X_Z \| Y_Z$，满足：

$$H(Z)=H\left(X_Z,Y_Z\right),H\left(X_Z\right)=PH\left(c_{-i}\right),H\left(Y_Z\right)=PH\left(c_j\right) \tag{34}$$

并且 c_i 与 X_Z 不可区分，c_j 与 X_Y 不可区分，$c_i \| c_j$ 与 Z 不可区分。

要证明在使用了强安全的碎片化系统 Ψ 下，对于任意 $i \in\{1,2,\cdots,l\}$，A_i 无法获知其他的非空数据 $c_j(j \neq i)$，

$$\Pr\left[A_i\left(c_{-i}\right)=c_j\right] \leqslant \frac{1}{2^{PH(c_j)}}+\varepsilon(\lambda) \leqslant \frac{1}{2^{H(c_j)}}+\varepsilon(\lambda) \tag{35}$$

使用反证法，若结论不成立，则存在算法 B 和多项式 $p(\cdot)$ 使得：

$$\Pr\left[B\left(c_{-i}\right)=c_j\right]=\frac{1}{2^{PH\left(c_j\right)}}+\frac{1}{p(\lambda)} \tag{36}$$

那么可以利用算法 B 构造区分器 $\mathcal{D}$。设对于输入 $x \| y$，若 $B(x)=y$，则 $\mathcal{D}$ 输出 1，否则输出 0。那么：

$$\Pr\left[\mathcal{D}\left(X_Z \| Y_Z\right)=1\right] \leqslant \frac{1}{2^{H(Y_Z)}} \tag{37}$$

此时对任意可忽略函数 $\varepsilon(\cdot)$，都有：

$$\left|\Pr\left[\mathcal{D}\left(c_{-i} \| c_j\right)=1\right]-\Pr\left[\mathcal{D}\left(X_Z \| Y_Z\right)=1\right]\right|=\left|\frac{1}{2^{PH\left(c_j\right)}}+\frac{1}{p(\lambda)}-\frac{1}{2^{H\left(Y_Z\right)}}\right|=\frac{1}{p(\lambda)} \tag{38}$$

并非可忽略的，这与假设矛盾。

基于伪随机性与不可预测性的等价性定理，若一个碎片化系统是伪随机的，那么攻击者在已知部分碎片条件下是不能预测其他碎片的。

定理 5 若一个多信道的碎片化系统是伪随机的，则该多信道数据传输在协助攻击下是强隔离安全的。

证明：考虑 $l-$多信道，设第 i 信道中存在一个监听者为 A_i，$i \in(1,2,\cdots,l)$。假设 A_i 只能获取所在信道上的所有数据，记作 c_i。要证明在使用了安全的碎片化系统 Ψ 下，对于任意 $i \in\{1,2,\cdots,l\}$，A_i 无法获知其他信道传输的非空数据 $c_j(j \neq i)$，即：

$$\Pr\left[A_i\left(c_{-i}\right)=c_j\right] \leqslant \frac{1}{2^{PH(c_j)}}+\varepsilon(\lambda) \tag{39}$$

其中 $c_{-i}=c_1 \| c_2 \| \cdots \| c_{i-1} \| c_{i+1} \| \cdots \| c_l$ 且 $(c_1,c_2,\cdots,c_l) \leftarrow \mathrm{Fra}_k(1^l,c)$。

由于碎片化系统 Ψ 是伪随机的，即对于任意的数据 $c \in\{0,1\}^*$ 和任意正整数 $i,j \in \mathbb{N}$，有 $c_1 \| c_2 \| \cdots \| c_l$ 是伪随机的，那么 $PH\left(c_{-i}\right)=|c_{-i}|$ 且 $PH\left(c_j\right)=|c_j|$，得到：

$$\Pr\left[A_i\left(c_{-i}\right)=c_j\right]\leqslant\frac{1}{2^{|c_j|}}+\varepsilon(\lambda) \tag{40}$$

4　结论

攻击方式的多样化与系统的复杂化给碎片化系统的安全性评估带来了一定的挑战。碎片化技术侧重于在单点数据泄漏的情况下限制对整体安全性的影响，根据实际攻击类型划分敌手能力，以敌手通过已知信息预测其他信道信息的概率大小区分隔离安全的强弱。提出3级碎片化技术安全评估等级，并加以有效性验证，建立一套多信道数据碎片化传输的安全等级评估指导方法。该方法也适用于数据碎片化存储和传输相关领域的安全性分析，对于数据碎片化技术的安全性评估和系统设计具有较大的实用价值和现实意义。

参考文献：

[1] KUPFERMAN Judy, ARNON Shlomi. Energy saving for data centers using spatial multichannel optical wireless communication[J]. Journal of Communication and Information Networks, 2017(4):88-99.

[2] HUANG Qingchao, LIU Dachang, CHEN Yinfang, et al. Secure free-space optical communication system based on data fragmentation multipath transmission technology[J]. Optics Express, 2018,26(10):13536-13542.

[3] O'HARA B,PETRICK A. IEEE 802.11 handbook: a designer's companion[M]. Copy URL,USA:IEEE Standards Association, 2005.

[4] PURNIMA Murali Mohan,TENG Joon Lim,MOHAN Gurusamy. Fragmentation-based multipath routing for attack resilience in software defined networks[C]// IEEE 41st Conference on Local Computer Networks(LCN). Dubai,United Arab Emirates:IEEE, 2016:583-586. doi:10.1109/LCN.2016.98.

[5] HUDIC A,ISLAM S,KIESEBERG P,et al. Data confidentiality using fragmentation in cloud computing[J]. International Journal of Pervasive Computing and Communications, 2013,9(1):37-51.

[6] 冯登国. 可证明安全性理论与方法研究[J]. 软件学报, 2005,16(10):1743-1756. (FENG Dengguo. Research on theory and approach of provable security[J]. Journal of Software, 2005,16(10):1743-1756.) doi:CNKI:SUN:RJXB.0.2005-10-007.

[7] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948,27(3):379-423.

[8] HÅSTAD J, IMPAGLIAZZO R, LEVIN L, et al. A pseudorandom generator from any one-way function[J]. SIAM Journal on Computing, 1999,28(4):1364-1396.

[9] YAO Andrew C. Theory and application of trapdoor functions[C]// 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). Chicago,IL,USA:[s.n.], 1982:80-91.

[10] BARMPALIAS G,LEWIS-PYE A. Chapter 3:limits of the kučera-g ά cs coding method[M]. Singapore:World Scientific, 2020: 87-109.

[11] AGGARWAL G,BAWA M,GANESAN P,et al. Two can keep a secret:a distributed architecture for secure database services[C]// 2nd Biennial Conference on Innovative Data Systems Research. Asilomar,CA,USA:[s.n.], 2005:186-199.

[12] CIRIANI Valentina,SABRINA De,CAPITANI Di,et al. Combining fragmentation and encryption to protect privacy in data storage[J]. ACM Transactions on Information and System Security(TISSEC), 2010,13(3):1-33.

作者简介：

陈世康(1970-)，男，学士，研究员级高级工程师，主要研究方向为网络安全方向.email:shikangchen@163.com.

郭　爽(1981-)，男，硕士，高级工程师，主要研究方向为网络空间安全.

唐　晋(1982-)，男，硕士，高级工程师，主要研究方向为网络安全、数据安全等.

袁　健(1978-)，男，学士，高级工程师，主要研究方向为网络空间安全.

林维涛(1998-)，男，在读硕士研究生，主要研究方向为网络空间安全、自然语言处理.

刘　丹(1969-)，男，博士，副教授，主要研究方向为网络空间安全、自然语言处理.