doi:10.11959/j.issn.1000-436x.2015064

基于多维正交载体的可裂解流指纹方案

雷程[1,2]，张红旗[1,2]，孙奕[1]，杜学绘[1]
（1. 解放军信息工程大学 三院，河南 郑州 450001；2. 河南省信息安全重点实验室，河南 郑州 450001）

摘 要：针对流交换中流源身份不可知、流交换范围不可控和流路径不可追踪问题，提出了基于多维正交载体的可裂解流指纹方案。利用 2 种相互正交的载体提高指纹信息的容量，并通过时间间隔重心载体特性和基于隐马尔科夫模型的解码技术实现可裂解性，提高方案的健壮性。分析了基于重心属性值随机选取载体的反制多流攻击能力，以及不同条件下指纹重心标记算法和基于隐马尔科夫模型解码技术的正确率。最后通过实验对算法的健壮性和隐蔽性进行了探讨。

关键词：流交换；多维正交载体；可裂解性；指纹重心；指纹字段
中图分类号：TP393 **文献标识码**：A

Cracking-resistance net-flow fingerprint scheme based on multi-dimensional orthogonal carriers

LEI Cheng[1,2], ZHANG Hong-qi[1,2], SUN Yi[1], DU Xue-hui[1]
(1. The Third College, PLA Information Engineering University, Zhengzhou 450001, China;
2. Henan Provincial Key Laboratory of Security Information, Zhengzhou 450001, China)

Abstract: Aimed at problems of unknown source identity, uncontrollable net-flow exchange and untraceable flow- exchanging paths, it proposes cracking-resistance net-flow fingerprint scheme based on multi-dimensional orthogonal carriers. It uses two mutually orthogonal carriers so as to improve the capacity of fingerprint information. It achieves cracking-resistance by interval centroid carrier characteristics and hidden Markov model based decoding technique. Ultimately, it improves the robustness of scheme. What's more, it analyzes the resistance of multi-flow attack ability by using centroid attribute value to select embedding carriers randomly. Besides, it analyzes accuracy of fingerprint centroid algorithm and quantization index modulation decoding technique based on HMM under different cases. Finally, its invisibility and robustness is evaluated by experiments.

Key words: net-flow exchange; multi-dimensional orthogonal carrier; cracking resistance; fingerprint centroid; fingerprint field

1 引言

随着互联网电子政务的数据中心呈现分布式架构特点、数据规模呈现扩大化趋势，流数据安全交换要求高效率、高安全和高可靠。已有针对数据中心服务器的攻击无论其原理[1~3]和手段如何，大都结合跳板和匿名通信技术以实现对攻击源和攻击路径的隐藏。仔细分析不难发现：这类问题产生的根源在于流交换中流源身份不可知，以及它导致的流交换范围不可控和流路径不可追踪。此外，由于政府、金融等部门使用的交互系统多具有数据分组加密、对时延敏感、多流交汇和资源有限的特点，

收稿日期：2013-11-13；**修回日期**：2014-01-08
基金项目：国家重点基础发展计划(“973”计划)基金资助项目(2011CB311801)；国家高技术研究发展计划(“863”计划)基金资助项目 (2012AA012704)；郑州市科技领军人才基金资助项目(131PLKRC644)
Foundation Items: The National Basic Research Program of China (973 Propram) (2011CB311801); The National High Technology Research and Development Program of China (863 Program)(2012AA012704); Zhengzhou Science and Technology Talents (131PLKRC644)

所以流指纹方案[4]不同于流量分析，它除了要快速准确地进行流关联，还要能不受加密影响、在多流交汇的情况下准确地提取出流源身份信息。如图 1 所示，被加密的流数据从目标区域流出，嵌入方将流身份相关信息嵌入到指纹重心标记和指纹字段部分。在流数据传输过程中，流身份信息会由于网络抖动或攻击遭到一定程度的破坏。当到达边界网关处，解码方解码指纹重心标记和指纹字段信息，提取流身份信息。依据网关设定的规则，阻止不符合规则的流数据流出。同时，记录并反馈给主服务器该条流的源身份信息，以形成流交换路径。

已有方案均选用单一流量特征作为载体，可分为基于间隔和基于分组时延 2 类。其中基于间隔的流指纹方案所利用的载体主要有时间间隔(IC) 和时间间隔重心(ICC)。基于间隔的方案[5,6]通过在嵌入方调节选定间隔内的分组数量以嵌入信息。解码方通过计算相应间隔内的分组数量、流速率从而提取相应的信息。这种载体具有可裂解性，可以有效抵御分组丢失、流分离等问题，具有较强的健壮性。但是其易受到具体流量特征的影响，所以在此基础上，文献[7]提出了基于间隔重心的方案。它通过调节选定间隔的分组时间从而改变整个间隔的平均时间以嵌入信息。解码方通过计算间隔重心提取嵌入信息。但是这 2 种方法需要成组地调制数据分组，在多流交汇时存在流间干扰，难以实现多流追踪。此外，由于它的隐蔽性差，易遭到多流攻击(MFA)[8]；针对这 2 个问题，文献[9]提出了利用 DSSS 解决的方法，但是如果不同的流利用相同的 PN 码，会遭到自相似(MSAC)攻击[10]，它利用相同 PN 码间的自相似性，检测和恢复 PN 码序列。针对这个问题，文献[11]提出利用相互正交的 PN 码加以解决的方法，但在长数据流条件下，这种方法很难实现。针对第 2 个问题，文献[12]提出了通过随机选取嵌入位置，削弱具有相同信息的流间依赖关系的解决思路。但是这种方法增加了检测方的计算复杂度和误报率。所以，文献[13]提出了基于信息重排序的多流攻击反制思想，它通过随机种子实现相同信息序列随机化排序，从而在不增加计算复杂度和误报率的同时，有效降低具有相同信息的多流间依赖关系。基于分组时延的方案[14,15]通过调制分组到达时间从而嵌入信息，因为这种方式无需成组地调制数据分组，所以可以有效避免 MFA 攻击。但是由于载体可裂解性差，易导致由于流变换引起的解码错误率上升问题。此外，由于流指纹方案嵌入的指纹信息具有信息量大、易被干扰和破坏的特点，选用单一流量特征作为载体存在容量受限和健壮性差的问题。

针对以上问题，本文设计了一种基于多维正交载体的可裂解指纹方案(MDOCRS)。它利用相互正交的 ICC 和 IPD 双载体，由指纹重心(fingerprint centroid)和指纹字段(fingerprint field)2 部分组成。其中指纹重心部分又分为指纹重心属性段和指纹重心标记段。重心属性段用于选取流指纹的嵌入位置，指纹重心标记段和指纹字段部分用于嵌入流源身份信息。指纹重心部分使用具有可裂解性的 ICC 作为载体，通过改进嵌入算法，提高解码的正确率。同时，利用重心属性段随机选取嵌入位置的方法，抵御 MFA 攻击。指纹字段部分使用 IPD 作为载体，利用基于银马尔科夫模型(HMM)的量化索引调制(QIM)[16]解码技术，增加指纹字段的可裂解性并提高解码的正确率。

2　预备知识

2.1　流指纹方案设计原则

流指纹方案是在流交换过程中通过嵌入和提取流源身份信息，实现检测内部或者外部攻击者通过跳板或匿名通信等方式进行的非法数据交换，其应用场景如图 1 所示。它的设计原则如下。

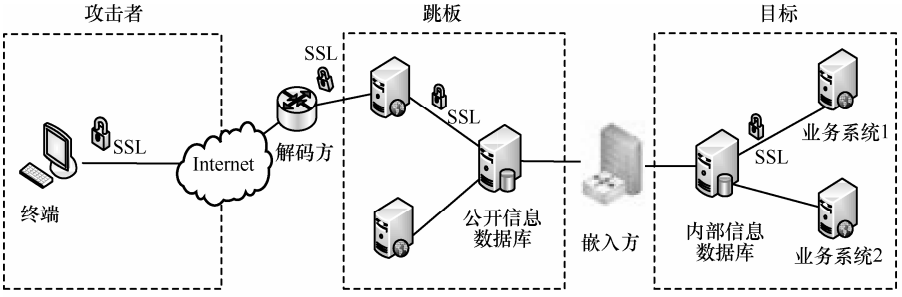


图 1　流指纹方案应用场景

1) 普适性。流指纹方案不应受到通信系统异构性的影响，而应将其作为“黑盒”处理。同时，不依靠系统间的通信协议和流数据的内容获取流源身份等信息。

2) 实时检测。由于流具有动态性、无限性和单次扫描的特性，且电子政务系统多具有对时延敏感、多流交汇和资源有限的特点。所以流指纹方案需要在有限的时间和资源条件下，准确检测流源身份信息。

3) 正确率。流指纹算法应该从理论上证明指纹信息编码解码技术和载体调制解调技术的正确性，并且在流变换条件下具有较高的解码和解调正确率。

4) 健壮性。流在受到不同干扰和攻击的情况下，可正确解码的概率。流指纹方案应该能够抵御网络抖动、分组加密问题，并且能够在一定程度上抵御流变换问题。可裂解性就是载体具有可以抵御流变换中流分离、分组丢失和分组合并问题的一种特性。

5) 隐蔽性。又称不可见性，要求嵌入指纹前后的流数据在数据分组分布上差异小，嵌入的信息应该对正常用户是透明的、让攻击者很难辨别。

2.2 隐马尔科夫模型

隐马尔科夫模型（HMM）[17,18]是一个二重马尔科夫随机过程，它包括可观测层和隐藏层。其中隐藏层是一个不可观测的、用状态转移概率描述的马尔科夫链；可观测层是一个与隐藏层关联的用输出观测值概率描述的随机过程。HMM 由 $\{N, M, A, B, \pi\}$ 五元组组成，N 表示状态的有限集合；M 表示观测值的有限集合；A 表示状态转移概率；B 表示观测值概率分布；π 表示初始状态概率分布。它有 3 个基本假设。

假设 1 马尔可夫性假设（状态构成一阶马尔可夫链）$P(q_i \mid q_{i-1}\cdots q_1) = P(q_i \mid q_{i-1})$。

假设 2 不动性假设（状态与具体时间无关）$P(q_{i+1} \mid q_i) = P(q_{j+1} \mid q_j)$，对任意 i，j 成立。

假设 3 输出独立性假设（输出仅与当前状态有关）$p(O_1,\cdots,O_T \mid q_1,\cdots,q_T) = \Pi p(O_t \mid q_t)$。

HMM 主要解决以下 3 个基本问题。

1) 评估问题：对于给定模型 λ，求某个观察值序列 O 的概率 $P(O \mid \lambda)$。

2) 解码问题：对于给定的观察值序列 O 和模型 λ，求可能性最大的状态序列 $\max_Q\{P(Q \mid O,\lambda)\}$。

3) 学习问题：对于给定的一个观察值序列 O，调整参数 λ，使得观察值出现的概率 $P(O \mid \lambda)$ 最大。

通过观测 QIM 解码器提取的序列 y_i，利用 HMM 解决学习问题，再利用向前−向后算法求出概率 $\Pr(y^{M'} \mid w_i^S)$，最后使用最大似然解码函数得到满足最优化准则的指纹序列，可增加指纹字段的可裂解性。本文构造的 HMM 模型如下。

状态的有限集合：

$N = \{(x_1,O_1),\cdots,(x_i,O_i),\cdots\}$，其中，$x_i$ 为嵌入序列，O_i 为序列在解码方相对嵌入方的位置偏移量。

观察值的有限集合：

$M = \{y_{1+O_1},\cdots,y_{i+O_i},\cdots\}$，其中，$y_{i+Oi}$ 为 QIM 提取的序列。

状态转移概率：

$A = \{a_{(i-1)i}\}, a_{(i-1)j} = \Pr[(x_i,O_i) \mid (x_{i-1},O_{i-1})]$

观察值转移概率：

$B = \{b_{(i-1)i}\}, b_{(i-1)i} = \Pr[y_{i-1+O_{i-1}}^{i-1+O_i} \mid (x_{i-1},O_{i-1})]$

初始状态概率分布：

$\pi = \{\pi_i\}, \pi_i = \Pr(x_i,O_i)$

2.3 量化索引调制

量化索引调制(QIM,quantization index modulation)依据待嵌入信息的不同，将选定的原始载体数据量化到不同的量化区间，在检测时根据所属的量化区间来识别嵌入的指纹信息。量化索引调制的一个典型应用就是量化抖动调制，它根据要嵌入的信息，选用特定结构的量化器量化载体系数。文献[16]提出了一类用量化技术嵌入水印信息的标准量化索引调制算法。它具有实现简单，消除载体干扰和增加嵌入信息健壮性的特点。

假设待嵌入的信息都可以表达为二元的比特序列。在二元标量 QIM 中，基本过程是根据二元水印值 $b \in \{0,1\}$，选择相应的量化器量化载体系数 s。一个步长为 Δ 的标量均匀量化器定义为 $Q_S = \Delta\left\lfloor \frac{s}{\Delta} \right\rfloor$。使用量化器可以产生 2 个抖动量化函数：$Q_b(s) = Q(s - d_b) + d_b$，$b = 0,1$，$d_0 = -\Delta/4$；$d_1 = \Delta/4$。当信号传输过程中受到噪声干扰且小于 $\Delta/4$ 时，信息可无差错地被提取。所以 QIM 技术可以增加嵌入信息的健壮性。

3 MDOCRS 方案

如图 2 所示，MDOCRS 利用相互正交的 ICC 和 IPD 双载体，由指纹重心和指纹字段 2 部分组成。

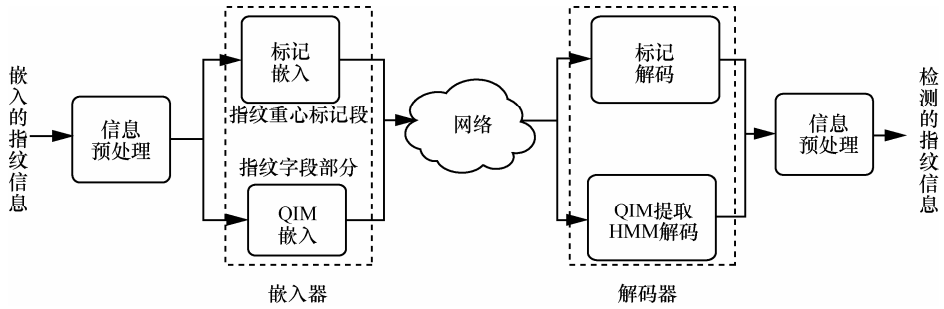


图 2　MDOCRS 流程

由于使用多种载体共同承载流信息，所以需要保证载体之间彼此相对正交。所谓相互正交的载体就是每种载体的改变都几乎不会影响其他载体的特征值。MDOCRS 通过选定合适参数，保证 ICC 与 IPD 这 2 种载体调制后互不干扰，实现载体的正交。由于 ICC 相较于 IPD 具有更好的健壮性，借鉴文献[19]思想，在嵌入信息时先嵌入指纹重心标记字段，再嵌入指纹信息。虽然嵌入的指纹信息对指纹重心标记字段有一定程度的影响，但是由于 ICC 对分组延时不敏感，所以不会影响检测方对指纹重心标记段信息的正确提取。

3.1　指纹重心算法

给定流 F_N，由定义可知，它可以看作持续时间是 t，具有 n 个有序数据分组的集合。令每个时间间隔长度为 T，则共有 $\left\lfloor \frac{t}{T} \right\rfloor$ 个时间间隔，其中最后一个时间间隔不进行任何操作。类似于 ICBW[7]，它要求嵌入方与解码方共享以下信息，如表 1 所示。

表 1　　指纹重心部分共享的参数

系统参数	秘密参数
时间间隔长度 T	指纹重心属性值 C_{FP}
最大时延 a $(0<a<T)$	秘钥 K_S
量子化乘数 q	—
间隔数目 $s(s=\left\lfloor \frac{t}{T} \right\rfloor -1)$	—

1) 获得指纹重心属性值：令 t_0 是第一个时间间隔 T_0 内第一个数据分组的到达时间，$\Delta t_j^{B2} = t_j^{B1} - t_0^{B1}$，$C_{FP}$ 为

$$C_{FP} = \frac{1}{m}\sum_{j=1}^{m} \Delta t_j^{FP} \tag{1}$$

借鉴文献[11,13]的思想，利用式(2)选择函数，得到[0, s]上基重心属性值 S，随机选取待嵌入的 ICC，实现载体选择的随机化，以消除嵌入相同信息的流间依赖性，防止多流攻击。如果简单地将重心从[0, T]映射到[0, s]上，由于重心会与流速率和分组时延有关，所以 S 在[0, s]上会集中于某一段区间，易造成选取相同的载体集合，从而遭到 MFA[20]攻击。通过采用量子化乘数 q，使 S 近似于均匀分布。保证不同流的指纹重心是随机选取的，从而可有效抵御 MFA 攻击。

$$S = \lfloor qsC_{FP}/T \rfloor \bmod s,\ q>1 \tag{2}$$

2) 标记选取：如图 3 所示，利用 S 对应的随机数 $\pi_j^{(S)} \in K_S$ $(j=0,\cdots,k-2,k-1)$，从 $\left\lfloor \frac{t}{T} \right\rfloor -1$ 个间隔中随机选出 k 个$(0<k<\left\lfloor \frac{t}{T} \right\rfloor -1)$，其中，$K_s$ 为秘钥。将前 $2N$（$2N<k$）个间隔作为指纹重心标记部分，按顺序分为 N 组。

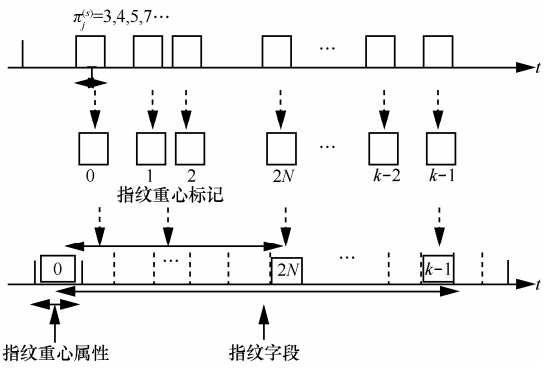
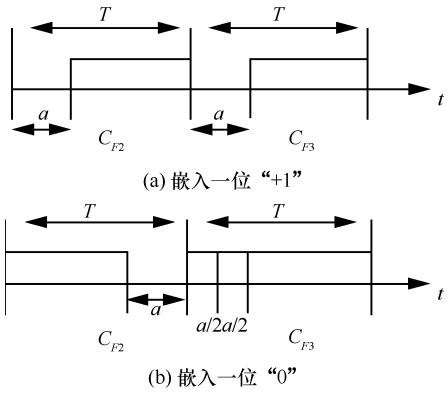


图 3　随机选取与初始偏移添加流程

3) 标记嵌入：对一组内的 2 个时间间隔重心分别进行调制。设一组内 2 个间隔的 ICC 分别为 C_{F2} 和 C_{F3}，取 $C_{F2} = \frac{1}{m}\sum_{j=1}^{m} \Delta t_j^{F2}$，其中，$\Delta t_j^{F2} = \Delta t_j^{F2} - \Delta t_0^{F2}$；$C_{F3} = \frac{1}{n}\sum_{j=1}^{n} \overline{\Delta t_j^{F3}}$，其中，$\overline{\Delta t_j^{F3}} = T - (\Delta t_j^{F3} - \Delta t_0^{F3})$。

令 $Y_F = C_{F2} - C_{F3}$，通过改变 Y_F 实现对指纹重心标记部分的嵌入与解码。图 4 说明了如何改变 Y_F 从而嵌入“+1”或者“0”。



(a) 嵌入一位“+1”

(b) 嵌入一位“0”

图 4 嵌入标识信息流程

若嵌入“+1”，通过增加每个数据分组的时延从而增加 C_{F2}、减少 C_{F3}。对 C_{F2} 和 C_{F3} 中的每个 Δt_j^{F2} 和 Δt_j^{F3} 做出如下调整

$$\Delta t_j^{F*'} = a + ((T-a)\Delta t_j^{F*})/T \quad (*=2,3)$$

由此，可以计算出 $\mathrm{E}(C_{F2}) = \dfrac{T+a}{2}$；$\mathrm{E}(C_{F3}) = T - \dfrac{T+a}{2} = \dfrac{T-a}{2}$。则有

$$\mathrm{E}(Y_{+1}^F) = \frac{T+a}{2} - \frac{T-a}{2} = a \tag{3}$$

若嵌入“0”，通过将 C_{F2} 中的数据分组部分转移到 C_{F3} 中，从而减少 C_{F2}、增加 C_{F3}。对 C_{F2} 中的数据分组[0, T−a]上的数据分组不做变换；[T−a, T]上的数据分组进行如下变换：$\Delta t_j^{F2'} = (a + \Delta t_j^{F2})/2$；对 C_{F3} 中的数据分组进行如下变换：对于[0, a]上的数据分组变换为：$\Delta t_j^{F3'} = (a + \Delta t_j^{F3'})/2$；对于[$a$, T]上的数据分组变换为：$\Delta t_j^{B3'} = \dfrac{a}{2} + \left(\left(T - \dfrac{a}{2}\right)\Delta t_j^{B3}\right)/T$。由此，可以计算出 $E(C_{F2}) = \dfrac{T-a}{2}$；$E(C_{F3}) = \dfrac{T+a}{2} - \dfrac{a^2}{(T+a)}$。则有

$$E(Y_{-1}^F) = \frac{T-a}{2} - \left(\frac{T+a}{2} - \frac{a^2}{(T+a)}\right) = -\frac{a^2}{(T+a)}a \tag{4}$$

4) 标记解码：由于指纹重心标记部分中嵌入的“0”、“+1”是等概率的，所以依据式(3)和式(4)可得判断阈值 V_{th} 为

$$V_{\mathrm{th}} = \frac{\mathrm{E}(Y_{+1}) + \mathrm{E}(Y_{-1})}{2} = \frac{a^2}{2(T+a)} \tag{5}$$

将解码后的值与阈值进行比较，如果 Y_i 大于 V_{th}，那么嵌入的标识值为“+1”，否则为“0”。

3.2 指纹字段算法

该部分基于指纹重心属性值，在选定的载体上嵌入流指纹信息[4]，由于流指纹信息量大，所以选择的载体容量应尽可能大。在理想条件下， IC 载体的最大容量为 $L_{\mathrm{Cap}} = \dfrac{t}{rT}$；ICC 载体的最大容量为 $L_{\mathrm{Cap}} = \dfrac{t}{2rT}$；IPD 载体的最大容量为 $L_{\mathrm{Cap}} = \dfrac{n-1}{r}$，其中，$r$ 为指纹信息的冗余度。由此可知 IPD 载体的容量最大。但是，IPD 存在自同步性不强的弱点，如传输过程中出现的流变换，会对解码造成影响。针对此问题，本文利用稀疏化调制技术，增加指纹信息的冗余程度作为指纹信息自同步的指示标识[21]，以提高信息的健壮性和自同步性。该部分需要共享的参数如表 2 所示。

表 2 指纹重心部分共享对参数

系统参数
量化步长Δ ($2q_s$)
秘密参数
稀疏因子 n
指纹信息 w
伪随机二进制序列 r_M

1) 指纹字段选取。如图 2 所示，依据基重心属性值选取指纹重心，通过随机选取嵌入指纹的数据分组，增加指纹信息的隐蔽性。

2) 指纹稀疏化。它将每个嵌入的等概率“0”“+1”序列变为概率系数为 f_w 的伯努利分布，从而增加信息冗余，提高自同步性。选定一个稀疏因子 n，将指纹序列中每位信息都稀疏成长度为 n 的序列。那么指纹信息 w^l 稀疏成为一个长为 M 的序列 w_M，有 $M = nl$。将 M 位随机序列和稀疏化后的指纹信息进行与运算，有 $w_i^s = r_i^M \oplus w_i^M$，$(i = 1,2,\cdots,M)$。令 f_w 为稀疏后指纹序列中“+1”的比重，f_w 为

$$f_w = \frac{\sum_{i=1}^{M} w_i^S}{M} \tag{6}$$

3) QIM 嵌入与检测。分别计算 T 内 IPD，

$ipd_{ij}=t_{i(j+1)}-t_{ij}$，$i\in[0,k-1],(j=1,2\cdots(r-2))$，最后一个分组时延不进行处理。由于 IPD 理论上是一个连续值，所以首先要将其量化。利用标准均匀量化函数 round(x)，将 x 量化为其最近的整数，设定量化步长为 $2q_s$>0，量化函数如下

$$f_q(ipd,q_s)=\mathrm{round}(ipd/q_s) \tag{7}$$

由式 (7) 可知，$\forall i\in R, y\in(-q_s/2,q_s/2]$，$f_q(iq_s,q_s)=f_q(iq_s+y,q_s)$。假设流指纹信息位来源于{0,1}，因为分组时延只能增加不能减少，所以，如图 5 所示，为了保证经过 f_E 函数运算后的值至少是 ipd，嵌入函数利用 $ipd+q_s/2$ 而非 ipd。嵌入指纹信息后，得到每个时间间隔的分组时延(IPD^F)。通过调节 q_s 使增加的时延足够小，从而可以让正常用户认为是网络抖动引起，以保证指纹信息的隐蔽性。嵌入函数如下

$$f_E(ipd,q_s,w)=[f_q(ipd+q_s/2,q_s)+\Delta]q_s \tag{8}$$

其中，$\Delta=(2+w-f_q(ipd+q_s/2,q_s)\bmod 2)\bmod 2$。

提取函数 f_x 如下

$$y_i=f_x(ipd^F,q_s)=f_q(ipd^F,q_s)\bmod 2 \tag{9}$$

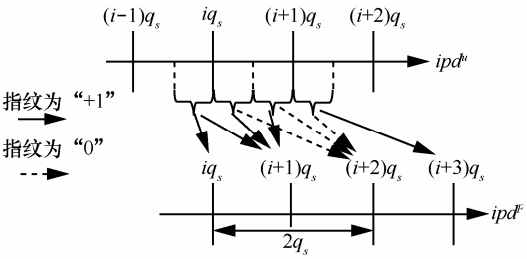


图 5　嵌入指纹信息前后对应关系

4) 基于 HMM 的解码。将传输过程建模成 HMM，对指纹信息的解码问题即为 HMM 中的学习问题。利用向前—向后算法求出概率 $\Pr(y^{M'}\mid w_i^S)$，再利用最大似然解码函数得到满足最优可能的指纹序列，解码函数 f_D 为

$$w_i^R=\arg\max_{w_i^S}[\Pr(y^{M'}\mid w_i^S)] \tag{10}$$

其中，$w_i^S\in\{0,1\},i=1,2,\cdots,M$。

5) 去稀疏化。利用相同的伪随机序列将解码得到的指纹序列异或，每 n 位序列对应为一位指纹信息。

4　理论分析

由于网络中数据分组服从独立、相同的均匀分布，所以可推出，在时间间隔 T 内的数据分组均匀分布于(0,T)内。指纹重心中的 Δt_j^{F2} 和 Δt_j^{F3} 服从均匀分布；指纹字段中的 IPD 服从均匀分布。对于流变换，由文献[22]可知，实际条件下流变换的分布如图 6 所示。由正态分布和拉普拉斯分布的特性[23,24]可知，拉普拉斯分布相较于正态分布尾部更长、更加平坦，所以将流变换分布近似为拉普拉斯分布更适合，同时在分析正确率上更加保守。

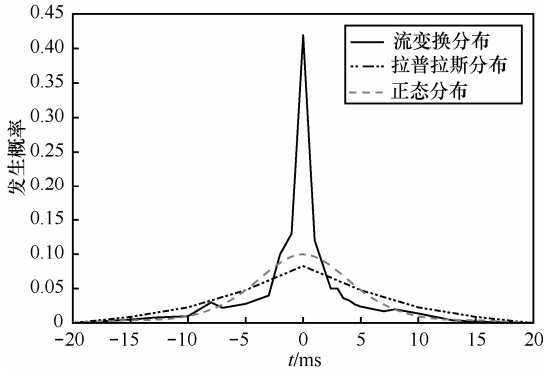


图 6　流变换与相应的拉普拉斯和正态分布

4.1　基于重心属性值随机选取方法的 MFA 反制能力分析

MFA 攻击的主要原理是利用多条嵌入相同信息流之间的依赖关系，检测指纹信息的存在性并恢复相关参数，以达到检测和破坏信息的目的。假设给定 m 条流数据，它们选取相同间隔的概率为 $\left(\frac{1}{C_S^{2N}}\right)^m$，当 s、N 不变，随着 m 的增加呈指数减少。依据鸽巢原理[25]，从 $C_S^{2N}+1$ 条流中，总能够找到选取相同时间间隔的 2 条流，计算复杂度随着 C_S^{2N} 呈超指数增长。所以，在不知道参数的情况下利用穷举的方法几乎不可行。因此，基于重心属性值随机选取的方法可以有效抵御 MFA 攻击。

4.2　无流变换条件下的解码

4.2.1　指纹重心部分的解码

以解码“+1”为例，在没有流变换的理想条件下，由切比雪夫不等式可得，文献[7]的解码错误率上限为

$$\begin{aligned}&\Pr[Y_{+1}<0]=\Pr[Y_{+1}-\mathrm{E}(Y_{+1})<-\mathrm{E}(Y_{+1})]\\&=\frac{1}{2}\Pr[|Y_{+1}-\mathrm{E}(Y_{+1})|>-\mathrm{E}(Y_{+1})]\leqslant\frac{T^2+(\mathrm{T}-\mathrm{a})^2}{6a^2}\end{aligned} \tag{11}$$

指纹重心标记段的解码错误率上限为

$$\Pr[Y_{+1}<V_{\text{th}}]=\Pr[Y_{+1}-\mathrm{E}(Y_{+1})<V_{\text{th}}-\mathrm{E}(Y_{+1})]$$
$$=\frac{1}{2}\Pr[|Y_{+1}-\mathrm{E}(Y_{+1})|>V_{\text{th}}-\mathrm{E}(Y_{+1})]\leqslant\frac{(T-a)^2(T+a)^2}{3a^2(a+2T)^2} \quad (12)$$

比较式(11)和式(12)可知，式(11)的概率值更大，所以MDOCRS基于ICC的嵌入算法解码正确率更高。

4.2.2 指纹字段部分的解码

对于指纹字段解码，f_E和f_D函数性质保证了指纹信息解码的正确性，具体性质如下。

性质 1 $\forall ipd>0,q_s>0,\mathrm{w}\in\{0,\}$，$f_X(f_E(ipd,q_s,\mathrm{w}),q_s)=y=w$

证明 由式(7)得

$\exists b\in\mathbf{Z}$,使$ipd+q_s/2=bs+c,-q_s/2<c\leqslant q_s/2$。

由式(8)和式(9)可知，$f_X(f_E(ipd,q_s,w),q_s)$可展开得到

$$y=f_q\{[b+((w-b)\bmod 2+2)\bmod 2]q_s\}\bmod 2$$
$$=[b+((w-b)\bmod 2+2)\bmod 2]\bmod 2$$
$$=w$$

性质 2 $\forall ipd>0,q_s>0,w\in\{0,1\}$，$0\leqslant f_E(ipd,q_s,w)-ipd<2q_s$

证明 由式(7)可得，如果$\mathrm{round}(ipd/q_s+1/2)=i$，那么$ipd/q_s+1/2\in(i-1/2,i+1/2]$，可得$ipd\in((i-1)q_s,(i+1)q_s]$。将$ipd$区间的2个极值代入到式(8)中，可得$ipd\leqslant f_E(ipd,q_s,w)<ipd+2q_s$，所以$0\leqslant f_E(ipd,q_s,w)-ipd<2q_s$成立。

由f_E和f_D性质可知，指纹字段的最大可容忍时延为$q_s/2$，所以$\forall D_S\in(-q_s/2,q_s/2]\Rightarrow d(ipd+D_s,q_s)=d(ipd,q_s)$。在没有流变换的条件下，只有网络抖动，由文献[26]可假设，IPD最大改变量为D。那么在$[-D,-q_s/2)\cup(q_s/2,D]$区间上，解码会出现错误，概率为$\Pr(ipd_i>\frac{q_s}{2})=\Pr(Y_i>\frac{1}{2})$。由切比雪夫不等式可知，$\Pr(Y_i>\frac{1}{2})=\Pr[Y_{+1}-\mathrm{E}(Y_{+1})>\frac{1}{2}-\mathrm{E}(Y_{+1})]$。由式(6)可知，嵌入的信息位累积分布服从二项分布，那么$\mathrm{E}(Y_{+1})=Mf_w$；$\mathrm{Var}(Y_{+1})=Mf_W(1-f_W)$，所以指纹字段的错误率上限为$\Pr(Y_i>\frac{1}{2})\leqslant\frac{2Mf_W(1-f_W)}{(1-2Mf_W)^2}$。

4.3 流变换条件下的解码

4.3.1 指纹重心部分的解码

对于指纹重心标记段的解码，因为指纹重心本身具有可裂解性，所以流变换中分组丢失、分组合并及流分离等问题对间隔重心没有影响。由于只有分组分离、多流交汇和虚假数据分组添加问题对间隔重心有影响，所以，在流变换情况下分析解码正确率只需要考虑分组分离、多流交汇和虚假数据分组添加问题。而这3个问题对流的影响都是由于添加了新的数据分组从而可能造成间隔重心的改变。

假设C_2和C_3间隔增加的数据分组个数与原有数据分组个数的比重为R_2和R_3，由大数定理可知，令$R_2\approx R_3=R$。以解码“+1”为例，如果有敌手刻意地添加数据分组时，考虑最极端的情况，即敌手添加的数据分组重心值为“0”。考虑添加的数据分组对解码标记位的影响，利用切比雪夫不等式可得

$$\Pr[Y'_{+1}<0]=\Pr[Y'_{+1}-\mathrm{E}(Y'_{+1})>-\mathrm{E}(Y'_{+1})]$$
$$=\frac{1}{2}\Pr[|Y'_{+1}-\mathrm{E}(Y'_{+1})|>-\mathrm{E}(Y'_{+1})]\leqslant\frac{\mathrm{Var}(Y'_{+1})}{2\mathrm{E}(Y'_{+1})^2} \quad (13)$$

由此可知，文献[7]中的错误率上限为$\Pr[Y'_{+1}<0]\leqslant\frac{(1+2R)T^2+(T-a)^2}{6a^2}$；解码指纹重心标记段的错误率上限为$\Pr[Y'_{+1}<V_{\text{th}}]\leqslant\frac{2(1+R)(T^2-a^2)^2}{3Na^2[a-(2+\frac{R}{N})T]^2}$。对比可知，MDOCRS方案错误率更低。

4.3.2 指纹字段部分的解码

对于指纹字段的解码，由于IPD健壮性相较于ICC差。此外，指纹字段可裂解性差，分组丢失、流分离对指纹字段的解码有较大影响，所以MDOCRS利用HMM方法，提高在分组丢失、流分离条件下指纹重心部分的可裂解性。假设嵌入方发送第$i-1$个数据分组后，解码方从对应的第$j-1$个数据分组提取出信息位x_{i-1}。在嵌入方发送第i个数据分组后，解码方接收数据分组的可能有4种情况。

1) 第i个数据分组在传输过程中丢失，同时没有数据分组添加。

2) 第i个数据分组在传输过程中丢失，同时又出现了丛发性数据分组添加。

3) 第i个数据分组被接收，同时又出现了丛发性数据分组添加。

4) 只有第i个数据分组被接收。该情况已在理想条件下讨论，不再赘述。

流变换导致了前 3 种情况的发生，它对流指纹信息影响可归结为合并（丢失）、替换和添加问题。

对于合并（丢失）问题，假设第一个数据分组不会丢失，其他每个数据分组丢失的概率为 α。当第 i 个数据分组没有在传输过程中丢失，第 i+1 个数据分组丢失，那么对于 ipd_i 与 ipd_{i+1} 可以认为 ipd_i 被合并为 $ipd_i' = ipd_i \oplus ipd_{i+1}$，$ipd_{i+1}$ 丢失。在流指纹信息传输过程中每个流指纹信息位发生合并问题的概率为

$$\mathrm{Pr}_D = 1 - f(\alpha) = \frac{1}{2}\mathrm{e}^{\frac{-\sqrt{2}\alpha}{\sigma}} \tag{14}$$

对于分组替换问题，假设数据分组替换的概率为 γ，在流指纹信息传输过程中每个流指纹信息位发生替换问题的概率为

$$\mathrm{Pr}_s = 1 - f(\gamma) = \frac{1}{2}\mathrm{e}^{\frac{-\sqrt{2}\gamma}{\sigma}} \tag{15}$$

对于添加问题，由网络丛发性分组的特性可知，添加产生的新 IPD 值可以认为其对应的信息是相同的（0 或者 1）。为便于说明，以下分析假设对应的信息位为“0”。其中每个信息位所添加的信息位数服从参数为 p 的几何分布；每个信息位在传输过程中发生添加问题的概率为 β，它服从拉普拉斯分布。添加问题的概率为

$$\mathrm{Pr}_I = 1 - f(\beta) = \frac{1}{2}\mathrm{e}^{\frac{-\sqrt{2}\beta}{\sigma}} \tag{16}$$

其中，$\beta = p^{k-1}(1-p)$。

基于以上分析，可得 x_i、O_i、y_{i+O_i} 的相互关系和状态转移概率 AB。令 O_i 为嵌入方发送的第 i 个数据分组在解码方的位置偏移量。如果给定发送的第 i−1 个数据分组的位置偏移量 O_{i-1}，解码方接收数据分组的 1）、2）、3）这 3 种情况可归纳为如下 2 种情况

$$O_i \begin{cases} O_{i-1} - 1, & \mathrm{Pr}_D(1-\mathrm{Pr}_I) \\ O_{i-1} + k, & \mathrm{Pr}_D\,\mathrm{Pr}_I^{k+1}(1-\mathrm{Pr}_I) + (1-\mathrm{Pr}_D)\mathrm{Pr}_I^{k}(1-\mathrm{Pr}_I) \end{cases} \tag{17}$$

1) 没有接收到新的数据分组。即发送的第 i−1 个数据分组丢失，同时没有新的数据分组添加。

2) 总共接收了 k 个数据分组。这又分为发送的第 i−1 个数据分组丢失同时添加了数据分组，或者发送的第 i−1 个数据分组接收后又添加了数据分组。

在没有丛发性分组添加的情况下，当第 i 个数据分组被接收后，IPD 是第 i 个数据分组和解码方之前接收的数据分组得到的。如果第 i−1 个数据分组被解码方接收，那么 IPD=x_i；如果第 i−1 个数据分组没有被接收，那么得到的 IPD 是 x_i 和之前的 IPD 值 x'_{i-1} 的异或。

$$x' \begin{cases} x_i, & (1-\mathrm{Pr}_D)(1-\mathrm{Pr}_I) \\ x_i \oplus x'_{i-1}, & \mathrm{Pr}_D(1-\mathrm{Pr}_I) \end{cases} \tag{18}$$

根据指纹稀疏化过程，可以将式(18)扩展如下。

$$x'_i \begin{cases} m_i, & (1-f)(1-\mathrm{Pr}_D(1-\mathrm{Pr}_I)) \\ m_i \oplus 1, & f(1-\mathrm{Pr}_D(1-\mathrm{Pr}_I)) \\ m_i \oplus x'_{i-1}, & (1-f)\mathrm{Pr}_D(1-\mathrm{Pr}_I) \\ m_i \oplus x' \oplus 1_{i-1}, & f\,\mathrm{Pr}_D(1-\mathrm{Pr}_I) \end{cases} \tag{19}$$

当传输过程中只出现替换问题，那么第 i 位信息 x'_i 被替换成为 $x'_i \oplus 1$。具体如下。

$$x'_i \begin{cases} x'_i, & 1-\mathrm{Pr}_S \\ x'_i \oplus 1, & \mathrm{Pr}_S \end{cases} \tag{20}$$

若 $i' = i + O_i$，结合式(17)可将式(20)推广如下。

$$y_{i'} \begin{cases} x'_i, & 1-\mathrm{Pr}_S \\ x'_i \oplus 1, & \mathrm{Pr}_S \end{cases} \tag{21}$$

由以上式(17)、式(19)和式(21)式可得，转移概率 $AB = \mathrm{Pr}[y_{i-1+O_{i-1}}^{i-1+O_i}, x_i, O_i \mid (x_{i-1}, O_{i-1})]$。

$$\begin{cases} x'_i = x'_{i-1} \oplus m_i, O_i = O_{i-1} - 1, y_{i-1+O_{i-1}}^{i-1+O_i} = \phi \\ \qquad\qquad (1-f)\mathrm{Pr}_D(1-\mathrm{Pr}_I) \\ x'_i = x'_{i-1} \oplus m_i \oplus 1, O_i = O_{i-1} - 1, y_{i-1+O_{i-1}}^{i-1+O_i} = \phi \\ \qquad\qquad f\,\mathrm{Pr}_D(1-\mathrm{Pr}_I) \\ x'_i = m_i, O_i = O_{i-1} + k, y_{i-1+O_{i-1}} = x'_{i-1} \\ \quad (1-f)(1-\mathrm{Pr}_S)(1-\mathrm{Pr}_I)(\mathrm{Pr}_D\,\mathrm{Pr}_I^{k+1} + (1-\mathrm{Pr}_D)\mathrm{Pr}_I^{k}) \\ x'_i = m_i \oplus 1, O_i = O_{i-1} + k, y_{i-1+O_{i-1}} = x'_{i-1} \\ \quad f(1-\mathrm{Pr}_S)(1-\mathrm{Pr}_I)(\mathrm{Pr}_D\,\mathrm{Pr}_I^{k+1} + (1-\mathrm{Pr}_D)\mathrm{Pr}_I^{k}) \\ x'_i = m_i, O_i = O_{i-1} + k, y_{i-1+O_{i-1}} = x'_{i-1} \oplus 1 \\ \quad (1-f)\mathrm{Pr}_S(1-\mathrm{Pr}_I)(\mathrm{Pr}_D\,\mathrm{Pr}_I^{k+1} + (1-\mathrm{Pr}_D)\mathrm{Pr}_I^{k}) \\ x'_i = m_i \oplus 1, O_i = O_{i-1} + k, y_{i-1+O_{i-1}} = x'_{i-1} \oplus 1 \\ \quad f\,\mathrm{Pr}_S(1-\mathrm{Pr}_I)(\mathrm{Pr}_D\,\mathrm{Pr}_I^{k+1} + (1-\mathrm{Pr}_D)\mathrm{Pr}_I^{k}) \end{cases} \tag{22}$$

其中，$y_{i-1+O_{i-1}}^{i-1+O_i}$ 表示从 $(i-1+O_{i-1})$ 到 $(i-1+O_i)$ 的 IPD 序列异或值。

用向前—向后算法求概率 $\mathrm{Pr}(y^{M'} \mid w_i^S)$，由算法可得

$$\Pr(y^{M'}\left|w_i^S\right.)=\sum_{(i-1)n}^{in}F_{(i-1)n}(x'_{(i-1)n},d_{(i-1)n})F'_{in}(x'_{in},d_{in})\cdot B_{in}(x'_{in},d_{in})$$

1) 对于 $F_{in}(x'_{in},O_{in})=\Pr(y_1^{(i-1)n+d_{in}},\ x'_{in},O_{in})$，其中，$i=1,2,3,\cdots,l$，它的递推关系式为

$$F_{in}(x'_{in},O_{in})=\sum_{x'_{(i-1)n},O_{(i-1)n}}F_{(i-1)n}(x'_{(i-1)n},O_{(i-1)n})\cdot \Pr(y_{(i-1)n+d_{(i-1)n}}^{(i-1)n+d_{in}},x'_{in},O_{in}\mid x'_{(i-1)n},O_{(i-1)n})$$

其中，$\Pr(y_{(i-1)n+d_{i-1}}^{(i-1)n+d_i},x'_{in},O_{in}\mid x'_{(i-1)n},O_{(i-1)n})$ 由转移概率可得。

对于 $F'_i(x'_j,O_j)=\Pr(y_{(i-1)n+d_{(i-1)n}}^{j-1+d_j},x'_j,O_j\mid x'_{(i-1)n},\ O_{(i-1)n},\ w_{i-1})$，其中 w_{i-1} 稀疏化后可写为 $w_{(i-1)n+1}^{in}$，它的递推关系式为

$$F'_i(x'_j,O_j)=\sum_{x'_{j-1},O_{j-1}}F'_i(x'_{j-1},O_{j-1})\cdot \Pr(y_{(i-1)n+d_{(i-1)n}}^{j-1+d_j},x'_j,O_j\mid x'_{(i-1)n},O_{(i-1)n},w_{(i-1)n+1}^{in})$$

其中，$\Pr(y_{(i-1)n+d_{(i-1)n}}^{j-1+d_j},x'_j,O_j\mid x'_{(i-1)n},O_{(i-1)n},w_{i-1})$ 由转移概率推导可得概率分布如下。

$$\begin{cases}x'_j=x'_{j-1}\oplus m_j\oplus w_j,O_j=O_{j-1}-1,y_{j-1+O_{j-1}}^{j-1+O_j}=\phi & \Pr_D(1-\Pr_I)\\ x'_j=m_j\oplus w_j,O_j\geqslant O_{j-1},y_{j-1+O_{j-1}}=x'_{j-1}\oplus 1 & \Pr_S(1-\Pr_I)(\Pr_D\Pr_I^{d_j-d_{j-1}+1}+(1-\Pr_D)\Pr_I^{d_j-d_{j-1}})\\ x'_j=m_j\oplus w_j,O_j\geqslant O_{j-1}+k,y_{j-1+O_{j-1}}=x'_{j-1} & (1-\Pr_S)(1-\Pr_I)(\Pr_D\Pr_I^{d_j-d_{j-1}+1}+(1-\Pr_D)\Pr_I^{d_j-d_{j-1}})\end{cases}$$

2) 对于 $B_{in}(x'_{in},O_{in})=\Pr(y_{in+d_{in}}^{M}\mid x'_{in},O_{in})$，它的递推关系式为

$$B_i(x'_i,O_i)=\sum_{x'_{i+1},O_{i+1}}\Pr(y_{i+d_i}^{i+d_{i+1}},x'_{i+1},O_{i+1}\mid x'_i,O_i)B_{i+1}(x'_{i+1},O_{i+1})$$

$\Pr(y_{i+d_i}^{i+d_{i+1}},x'_{i+1},O_{i+1}\mid x'_i,O_i)$ 由式(22)可得。

得到概率后 $\Pr(y^{M'}\mid w_i^S)$ 利用式(10)可解码指纹信息。相较于文献[15]，MDOCRS 通过基于 HMM 的解码实现指纹字段信息的可裂解性，可更有效地抵御流变换问题，具有更好的健壮性。

5 实验结果与分析

5.1 实验环境

如图 7 所示，选择 2 个区域分别作为内部数据处理区、公开数据处理区，其余区域作为外部接入区。实验条件下的流有约 2 000 个数据分组；分组速率为 1.68 packet/s；流变换分布服从 Lap(0,2.23) 的分布；分组替换概率 $\Pr_s=0.1\%$。MDOCRS 参数如表 3 所示，检测的指纹重心和指纹字段的组数和分组个数对应关系如表 4 所示。

表 3 MDOCRS 参数

参数名称	参数值
量化步长 Δ	60 ms
最大时延 a	700 ms
稀疏因子 n	8
间隔数目 s	240
量子化乘数 q	1.25

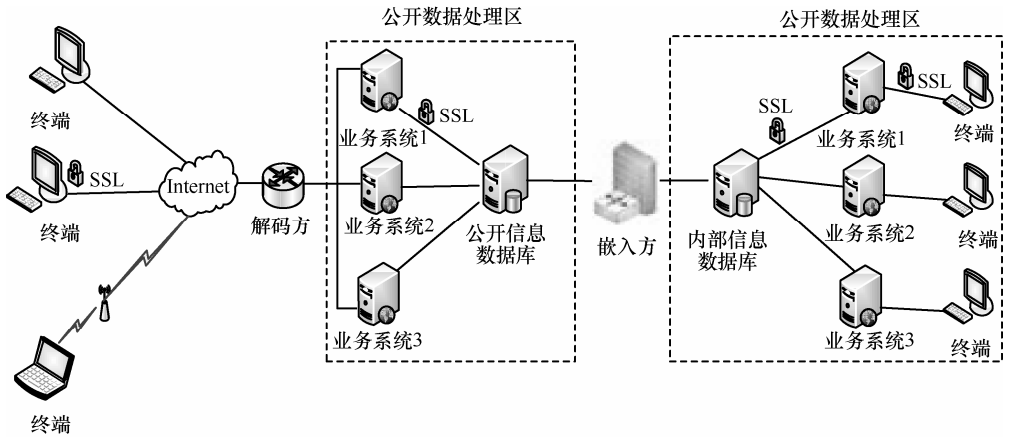


图 7 MDOCRS 实验场景

表 4　MDOCRS 检测组数与分组数量

T/s	指纹重心检测组数/组	指纹重心检测数据分组数/个	指纹字段检测数据分组数/个	指纹字段检测组数/组
3	10	50	2 000	200
6		100		94
9		150		62
12		200		39

5.2　健壮性实验

在 T=3 s、6 s、9 s、12 s 的情况下，检测指纹重心标记字段的正确率如表 5 所示。所以 MDOCRS 指纹重心标记字段在约 150 个数据分组时，正确率在 90%以上；约 200 个数据分组时，正确率可以达到 100%。在相同情况下，文献[7]需要约 240 个数据分组才能达到正确率在 90%以上；需要约 380 个数据分组才能达到 100%正确率。所以与之相比较，MDOCRS 利用改进后的嵌入算法提高了指纹重心标记的健壮性。

表 5　指纹重心标记字段正确率

T/s	正确率/%
3	62.1
6	79.6
9	92.8
12	100.0

实际条件下 $P_I, P_D \in (0.1\%, 3\%)$。利用干扰器改变 P_I 和 P_D 值，在 P_I 和 P_D 分别为 2%、4%、6%、8%、10%的情况下，得到指纹字段的正确率如表 6 所示。由表 6 可知，P_I 和 P_D 在 4%以下，检测正确率可达到 100%。所以，利用基于 HMM 的解码技术，使指纹字段具有良好的可裂解性，可有效抵御流变换问题。

表 6　指纹字段正确率

P_I/%	P_D/%	正确率/%
2	2	100
4	4	100
6	6	97.2
8	8	89.0
10	10	77.4

5.3　隐蔽性实验

利用 K-S 测试检测指纹信息的隐蔽性，假设给定的两条流，分别是嵌入指纹信息的流 A_W 和未嵌入指纹信息的流 B_U，K-S 距离定义为 $\sup_x(|F_A(x)-F_B(x)|)$，其中，$F_A(x)$和 $F_B(x)$是流 A 和 B 的经验分布函数。由文献[27]可知，当 K-S 距离小于 0.036，可以认为两条流是相同的。如表 7 所示，在不同 T 的情况下，嵌入信息的流 A_W 和未嵌入信息流的 B_U 的 K-S 距离。由此可知嵌入的指纹信息具有良好的隐蔽性，对正常用户是透明的。

表 7　K-S 实验

T/s	K-S 距离
3	0.012 4
6	0.013 0
9	0.011 6
12	0.010 4

6　结束语

本文针对流交换中流源身份不可知、流交换范围不可控和流路径不可寻问题，提出了基于多维正交载体的可裂解流指纹方案。不同于已有的流量分析方法，它除了能够快速地进行流关联，而且能够让检测点准确地检测出流身份信息，具有不受加密影响、准确率高、误报率低、观测时间短、带宽消耗小、可同时追踪多流等特点。MDOCRS 由指纹重心和指纹字段两部分组成。指纹重心部分采用具有可裂解性的 ICC 作为载体，通过重心属性值随机选取嵌入的间隔，减少嵌入相同信息的流间依赖，有效防止 MFA 攻击。同时，通过改进已有的流重心标记算法提高解码正确率。指纹字段部分采用 IPD 作为载体，通过基于 HMM 解码技术，实现指纹字段的可裂解性，提高解码的正确率。此外，通过选定适合的参数，实现 ICC 和 IPD 载体的相互正交。实验发现，在流变换条件下，指纹重心字段只需约 200 个数据分组，解码正确率即可达到 100%；指纹字段在 P_I 和 P_D 小于 4%的条件下，正确率可达到 100%。隐蔽性测试中，K-S 距离约为 0.013，小于阈值 0.036。所以，MDOCRS 具有高正确率、强健壮性和良好的隐蔽性。

参考文献：

[1] ZHU Z, LU G, CHEN Y, *et al*. Botnet research survey[A]. Computer Software and Applications, 32nd Annual IEEE International[C]. 2008. 967-972.

[2] ZHANG Y Z, XIAO JUN, YUN X C, *et al*. DDoS attack detection and

control methods[J]. Journal of Software, 2012, 23(8): 2058-2072.

[3] SCHUBA C L, KRSUL I V, KUHN M G, *et al.* Analysis of a denial of service attack on TCP[A].1997 IEEE Symposium on Security and Privacy[C]. 1997.208-223.

[4] HOUMANSADR A, BORISOV N. The need for flow fingerprints to link correlated network Flows[A]. Privacy Enhancing Technologies[C]. Springer Berlin Heidelberg, 2013.205-224.

[5] PYUN Y J, PARK Y H, WANG X, *et al*. Tracing traffic through intermediate hosts that repacketize flows[A]. 26th IEEE International Conference on Computer Communications[C]. 2007.634-642.

[6] PYUN Y J, PARK Y, REEVES D S, *et al.* Interval-based flow watermarking for tracing interactive traffic[J]. Computer Networks, 2012, 56(5): 1646-1665.

[7] WANG X, CHEN S, JAJODIA S. Network flow watermarking attack on low-latency anonymous communication systems[A] IEEE Symposium on Security and Privacy[C] 2007.116-130.

[8] KIYAVASH N, HOUMANSADR A, BORISOV N. Multiflow attacks against network flow watermarking schemes[A]. Proceedings of 17th USENIX Security[C]. San Jose, 2008.307-320.

[9] YU W, FU X, GRAHAM S, *et al*. DSSS-based flow marking technique for invisible traceback[A]. IEEE Symposium on Security and Privacy[C]. 2007.18-32.

[10] JIA W, TSO F P, LING Z, *et al*. Blind detection of spread spectrum flow watermarks[J]. Security and Communication Networks, 2013, 6(3):257-274.

[11] HUANG J, PAN X, Fu X, *et al*. Long PN code based DSSS watermarking[A]. 2011 INFOCOM, Proceedings IEEE[C]. 2011. 2426- 2434.

[12] HOUMANSADR A, KIYAVASH N, BORISOV N. Multiflow attack resistant watermarks for network flows[A]. Proceedings of IEEE International Conference on Acoustic, Speech, and Processing[C]. 2009.1497-1500.

[13] 王振兴, 张连成, 郭毅等. 基于水印信息重排序的多流攻击反制方法[J]. 应用科学学报, 2013, 31(3):278-284.
WANG Z X, ZHANG L C, GUO Y, *et al*. Multi-flow attack resistance based on reordering of watermark bits[J]. Journal of Applied Sciences, 2013, 31(3):278-284.

[14] HOUMANSADR A, KIYAVASH N, BORISOV N. Rainbow: a robust and invisible non-blind watermark for network flows[A]. Inndss[C]. 2009.

[15] WANG X, REEVES D S. Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays[A]. Proceedings of the 10th ACM conference on Computer and Communications Security[C]. ACM, 2003.20-29.

[16] CHEN B, WORNELL G W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding[J]. IEEE Transactions on Information Theory, 2001, 47(4): 1423-1443.

[17] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257-286.

[18] EDDY S R. Hidden markov models[J]. Current opinion in structural biology, 1996, 6(3): 361-365.

[19] MINTZER F, BRAUDAWAY G W. If one watermark is good, are more better[A]. IEEE International Conference on Acoustics, Speech, and Signal Processing[C]. 1999.2067-2069.

[20] KIYAVASH N, HOUMANSADR A, BORISOV N. Multi-Flow attacks against network flow watermarks: analysis and countermeasures[J]. arXiv preprint arXiv:1203.1390, 2012.

[21] MATTHEW C. DAVEY,DAVID *et al*. Reliable communication over channels with insertions, deletions, and substitutions[J]. IEEE Transactions on Information Theory, 2001,47:687-698.

[22] BAVIER A C, BOWMAN M, CHUN B N, *et al*. Operating systems support for planetary-scale network services[A]. NSDI[C]. 2004. 19-19.

[23] ELTOFT T, KIM T, LEE T W. On the multivariate Laplace distribution[J]. Signal Processing Letters, IEEE, 2006, 13(5): 300-303.

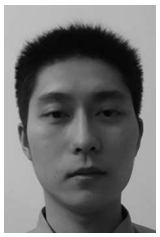
[24] HAZEWINKEL M. Normal distribution[J]. Encyclopedia of Mathematics, 2001, 13(6): 337-342.

[25] GRIMALDI R P. Discrete and Combinatorial Mathematics: An Applied Introduction 5th ed[M]. Massachusetts: Addison Wesley, 2003.

[26] DONOHO D L, FLESIA A G, SHANKAR U, *et al*. Multiscale stepping-stone detection: detecting pairs of jittered interactive streams by exploiting maximum tolerable delay[A]. Recent Advances in Intrusion Detection[C]. Springer Berlin Heidelberg, 2002.17-35.

[27] MASSEY JR F J. The Kolmogorov-Smirnov test for goodness of fit[J]. Journal of the American statistical Association, 1951, 46(253): 68-78.

作者简介：



雷程（1989-），男，北京人，解放军信息工程大学硕士生，主要研究方向为网络与信息安全、数据安全交换。

张红旗（1962-），男，河北遵化人，解放军信息工程大学教授、博士生导师，主要研究方向为网络安全、等级保护。

孙奕（1979-），女，河南郑州人，解放军信息工程大学博士生，主要研究方向为网络与信息安全、数据安全交换。

杜学绘（1968-），女，河南新乡人，解放军信息工程大学教授、博士生导师，主要研究方向为网络与信息安全、计算机应用技术。