

基于改进 IPD 质心的 Tor 网络流水印检测方法

杜捷, 何永忠, 杜晔

(北京交通大学计算机与信息技术学院, 北京 100044)

摘要: Tor 是一种为隐藏流量源提供服务的匿名网络机制, 但其存在入口流量特征明显、易被识别的问题。obfs4 等网桥协议为解决此问题应运而生, 由此带来的新挑战尚未攻克, 因此, 提出一种 IPD 质心方案, 利用 k -means 的聚类特性将原方案进行改进, 使加入的流水印在 obfs4 网桥 3 种模式上均能被高效地检测出。实验结果表明, 改进后的算法有更高的检测率和识别率, 且应对不同的网络环境有较强的适应能力, 有利于良好安全网络环境的构建。

关键词: 主动流量分析; 网络流水印; 匿名通信; Tor

中图分类号: TP391.1

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2019041

Improved method of Tor network flow watermarks based on IPD interval

DU Jie, HE Yongzhong, DU Ye

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract: Tor is an anonymous network mechanism that provides services for hiding traffic sources, but it has the problem that the entry traffic flows of Tor are clearly identifiable. Bridge protocols such as obfs4 come into being to solve this problem, which brings new challenges that have not yet been overcome. An IPD interval scheme is proposed, which uses the clustering characteristics of k -means to improve the original scheme, so that the added flow watermark can be detected efficiently in the three modes of obfs4 bridges. The results of experiments show that the improved algorithm has higher detection rate and recognition rate, and has good adaptability to variable netflow traffic, which is conducive to the construction of a nice secure network environment.

Key words: active traffic analysis, network flow watermarks, anonymous communication, Tor

收稿日期: 2018-11-28; 修回日期: 2018-12-26
通信作者: 杜捷, 418382002@qq.com

论文引用格式: 杜捷, 何永忠, 杜晔. 基于改进 IPD 质心的 Tor 网络流水印检测方法[J]. 网络与信息安全学报, 2019, 5(4): 91-98.

DU J, HE Y Z, DU Y. Improved method of Tor network flow watermarks based on IPD interval [J]. Chinese Journal of Network and Information Security, 2019, 5(4): 91-98.

1 引言

当今网络环境中存在各种各样的攻击、非法交易等情况, 确认攻击者、非法交易人员的身份和流量源, 对于打击此类违法行为、维护安全和谐的网络环境有十分重要的意义。毫无疑问, 敌手会想方设法隐藏自身的身份和流量源, 如使用假冒的 IP 地址作为采取攻击的主机 IP 地址。而追踪假冒 IP 地址之后真正的流量源头, 被称作 IP 追踪 (traceback) [1-3]。

目前有众多可以隐藏身份和流量源的匿名网络机制为网络攻击者、非法交易者提供服务, 如 Tor (the onion route, 洋葱路由)、I2P、JAP 等。作为用户使用量最大的洋葱路由, Tor 可以十分有效地隐藏它们的身份。Tor 是众多匿名网络机制中的一员, 因其开源、极易搭建、延迟低和服务稳定的特点, 成为当前用户量绝无仅有、受欢迎程度史无前例的一种匿名通信机制[4]。Tor 网络运用多层路由的手段, 使追踪到源头——即流量发送者成为一个难题, 因此 Tor 连接时常会遭受限制。由于其入口流量特征较明显, 极易被识别出来, 流向已公开 Tor 入口节点 IP 和端口的流量会被高度重视, 其匿名性受到重创。

Tor 团队针对这一不利条件发布了可插拔传输协议 (pluggable protocol), 如 MEEK、obfs4 等, 该协议也被称作网桥 (bridges)。它运用多种传输协议, 将连接到 Tor 入口节点的流量伪装成友好的、不引人注目的流量。因此用户可通过该协议接入原生 Tor 网络, 经其混淆再转发送至 Tor 入口节点进行匿名操作。

目前, 没有一种可行的方法来追踪在 obfs4 协议上传输流的现状亟待解决, 主要面临的困难有: 1) Tor 为了保证匿名性, 每 10 min 更换一次节点线路, 不同的节点线路由于网络延迟和传输距离千差万别, 传输的速率不会一成不变, 要求

算法具有极高的稳健性; 2) obfs4 存在 3 种模式, 各自对于数据包大小和时间有着独特的处理方式, 甚至会无规律地对传输的数据包产生延迟以干扰传统的流水印方案。

针对以上问题, 本文提出了一种基于跨包延迟 (IPD, inter-packet delay) 的流水印方案来追踪 obfs4 上的流量, 该方案可有效追踪使用 obfs4 混淆协议的 Tor 流量, 并且兼容其提供的 3 种保护模式。

2 技术背景

截止到 2018 年 10 月, Tor 的全球用户量达到 225 万, 其中通过网桥访问的用户量超过 15 万, 每秒在 Tor 上产生的流量约为 300 GB, 但相较于前统计的数据[5], Tor 的中继节点和网桥节点数量呈现下降趋势, 用户量的上升、流量的增长导致现存 Tor 节点上的环境更为复杂。

Tor 网络实现匿名的核心是重路由技术, 经过多层路由转发, 每层路由只知晓上一级节点, 故难以追踪到流量源头, 加之网桥协议会将其源头伪装成其他数据流量, 不让第三方从报文消息内容分辨出, 问题十分严重。针对 obfs4 上 Tor 流量的追踪尚无文献提及, 本文对现有检测方法进行了调研, 作为主动流量分析方法的代表——流水印方案, 相较于被动的流量分析, 后者需要建立在长时间流量特征观察的基础上, 虽然拥有较高的检测率, 但是需要极大的流量长时间分析, 以减少巨大的错误率、误报率, 因此流水印方案更适用于对 obfs4 上 Tor 流量进行追踪。

流水印方法的核心是将一种水印模式加入流中, 使其在传递到目的地址之前都是可以被追踪到。运用在数据包时间上的流水印方案主要分为两类。一类是基于 IPD 展开的研究。Wang 等[1]提出了一种基于 IP 的方案, 该模式使用 QIM 框架调制水印以改变 IPD 的平均值, 达到较好的检测效果。随后, Wang 等[6]在点对点的 VoIP 流上进行了实

验, 在 Skype 这种较为稳定的持续性流量上, 只要增加冗余值 r 到足够大时, 可达到 100% 的检测率, 该方法具有极强的隐蔽性。另一类基于 IP 的方案^[7]是 RAINBOW 方法, 这种方法具有非盲属性, 在检测结果时需要原始的未加入水印的数据包。

另一类流水印方案是基于时隙 (interval) 的, 在固定的时间间隔内被编码成一批分组的水印模式同样可以被检测出来。Wang 等^[8]提出了一种基于时间质心的攻击方案, 这是首次提出针对匿名系统的攻击, 能在很短的时间内 (约 5 500 个包) 检测出水印。还有一个类似的设计^[9]是将水印模式被嵌入数据包的时间间隔内。基于区间的方案由于成批量的移动数据包, 容易被敌手发现并采取相应的对策, 如将发出的数据包调整为一个稳定的速率以消除潜在的水印攻击。

选择 Wang 等^[6]的方案在 obfs4 上的 Tor 流量进行实验, 该方案在其特定的实验环境下可达到良好效果, 但面对为数据包时间提供保护的 obfs4 协议, 其刻意制造的类似水印的时间间隔会对检测造成极大干扰, 原水印模式无法应对, 将难以区分原始流量和加水印流量; 此外, 该方法建立在概率算法的基础上, 需要极其众多的数据包才能验证结果, 大量持续稳定的数据包在实际网络环境中难以采集。针对以上存在的问题, 本文在其基础上加以改进, 得到了良好的检测结果。

3 IPD 质心方案

本文采用基于跨包延迟的算法对 obfs4 上 Tor 流量进行研究。在 Wang 等^[6]方案的基础上采用聚类算法, 得到类似时隙的时间质心, 这里把它叫作 IPD 质心。分析原始数据流调制 IPD 质心达到水印嵌入的目的, 从而使其可以被追踪。

3.1 计算 IPD

本文算法需要使用 Wang 等^[6]对 IPD 的描述: 假设已获取数据包 $(P_1, P_2, \dots, P_n)$, 每个数据包到达

的时刻记为 $(t_1, t_2, \dots, t_n)$, 则一定有 $t_1 < t_2 < \dots < t_i < t_j < \dots < t_n$ 成立。取一个分组交叉间隔 d , 独立且随机地选取 $2r$ 个数据包作为处理对象。记选取出来的 $2r$ 个数据包为 $(P_{s_1}, P_{s_2}, \dots, P_{s_{2r}})$, 则它们对应的包到达时刻记为 $(t_{s_1}, t_{s_2}, \dots, t_{s_{2r}})$ 。根据之前选取的包交叉间隔 d , 计算分组的跨包延迟 IPD。

$$ipd(s_i, d) = t_{s_i+d} - t_{s_i}, i \in (1, 2r)$$

3.2 计算延迟距离

此时得到的 $ipd(s_i, d)$ 序列为 $2r$ 个数据包与包交叉间隔 d 之后的 $2r$ 个数据包的跨包延迟。再将这 $2r$ 个 $ipd(s_i, d)$ 序列随机分成数目相同的两组:

$ipd_1(s_i, d)$ 和 $ipd_2(s_i, d)$, $j \in (1, r)$ 。

计算包的延迟距离, 记作 $D(s_j, d)$, 则

$$D(s_j, d) = |ipd_1(s_j, d) - ipd_2(s_j, d)|, j \in (1, r)$$

延迟距离 $D(s_j, d)$ 存在的极端值 (由网络延迟、丢包重传、obfs4 保护机制等造成) 会影响后续聚类结果。为了减少极端值的影响, 移除四分位距 (IQR, interquartile range) 范围标记的异常值, 这是用于检测噪声测量数据中异常值的标准度量^[10]。由于聚类是对 $D(s_j, d)$ 的绝对值处理的, 因此只考虑剔除 $D(s_j, d)$ 中的极大值。四分位差 IQR 表示第三四分位数和第一四分位数的距离。

$$|IQR| = Q3 - Q1$$

取一个控制极端值剔除的门限 η , 则定义极端值 M 为

$$M > Q3 + \eta |IQR|$$

3.3 计算 IPD 质心

为了对延迟距离进一步处理, 采用了 k -means 算法, 该算法是聚类算法中以划分方法为核心的经典算法, 对于二维及多维数据都有良好的聚类效果。 k -means 主要的缺陷在于: 1) k 值的估计困难; 2) 初始选取的聚类中心对结果影响重大。本文方案根据延迟距离的分布在加入水印后会呈现两个峰值, 确定 k 的取值为 2; 众多靠近峰值的数据保证即使初始选择的聚类中心为异常值, 也

会通过欧氏距离的均值调整至峰值附近；此外，数据维度只有一维，效率更快、准确率更高。取 k 值为 2，对 $|D(s_j, d), j \in (1, r)|$ 进行 k -means 聚类，算法伪代码如下。

算法 1

输入: $D(s_j, d) | j \in (1, r)$

输出: Cluster center C_0, C_1 with the number of

n_0, n_1

1) close two points C_0, C_1 randomly from $D(s_j, d)$

2) repeat

$$3) \quad d_{0j} \leftarrow \sqrt{(D(s_j, d) - C_0)^2}$$

$$4) \quad d_{1j} \leftarrow \sqrt{(D(s_j, d) - C_1)^2}$$

5) if $d_{0j} < d_{1j}$ then

6) Put $D(s_j, d)$ into C_0

7) else

8) Put $D(s_j, d)$ into C_1

9) end if

$$10) \quad C_{0j} \leftarrow \frac{1}{n_0} \sum_{k=1}^{n_0} C_0$$

$$11) \quad C_{1j} \leftarrow \frac{1}{n_1} \sum_{k=1}^{n_1} C_1$$

12) until C_0, C_1 do not change OR exceed limited times

在未加入水印时，得到两个聚类中心 C_0, C_1 ，将其称为 IPD 质心， C_0 代表近 0 端中心， C_1 代表远 0 端中心，两个聚类中心的内聚类数分别为 n_0, n_1 。质心 C_0 中的数量远高于质心 C_1 ，即 $n_0 \gg n_1$ ；随后根据质心 C_1 选择相近的值 ω ，以 ω 作为水印大小加入，且水印占比为 μ ，由此重复算法步骤进行聚类，得到两个 IPD 质心 C'_0, C'_1 ，两个聚类中心的内聚类数分别为 n'_0, n'_1 ，IPD 质心 C'_1 等于加入的水印大小 ω ，IPD 质心 C'_0 中数量与 IPD 质心 C'_1 中数量的比接近 $\frac{1-\mu}{\mu}$ ，即 $\frac{n'_0}{n'_1} \approx \frac{1-\mu}{\mu}$ 。

为了判定该算法的检测率，设定一组容错值

(α, β) ，IPD 质心 C'_1 与加入水印大小 ω 的差值绝对值应小于 α ，取得结果的占比 μ' 与加入水印占比 μ 的绝对值差应小于 β ，即 $|C'_1 - \omega| < \alpha$ ， $|\mu - \mu'| < \beta$ ，二者都满足的情况下认为本次检测结果存在水印。

4 实验及结果

实验在美国德克萨斯州达拉斯市、英国伦敦和日本东京部署了 3 台服务器，其中，东京的服务器配备了完整的 Tor 网络和 obfs4 网桥，以作为 Tor 的入口；伦敦服务器作为出口节点；达拉斯市的服务器作为访问的目标服务器。通过连接东京服务器的 obfs4 网桥进入 Tor 网络，匿名访问达拉斯市服务器，每次实验以连续的 1 000 个含有负载的 TCP 数据包作为研究对象，前 500 个 TCP 分组无水印，后 500 个 TCP 分组加入水印，且水印占比 μ 为 50%。由于 obfs4 本身会自发地额外增加时间间隔，相当于增加为“1”位的水印，因此实验中采用的水印位均为“1”。

为了防止 Tor 更换线路对实验造成影响，多次测试并根据源码进行分析，得出如下结论：若当前线路能够正常且良好地工作，10 min 内新产生的 TCP 流都被 Tor 安排使用该线路；一旦线路发生问题，Tor 立即开辟一条新的路由节点线路。对实验计时，确保每次实验均在 10 min 内完成，且保持 TCP 连接不中断，实际采集到的数据流的出口节点保持不变，即 Tor 没有更换线路。随后的实验均以该方式实施。

随机取出 $2r$ 个数据包使取样率 p 小于“90%”，目的是在使用尽可能多数据包的前提下保证取样的随机性，此时 $n=500, r=226$ 。计算出未加水印的原始实验数据 IPD 如图 1 所示，未加水印的原始延迟距离如图 2 所示，IQR 的门限值 η 取 0.8，进行异常值筛选，筛选后的概率分布如图 3 所示。进行 k -means 聚

类后,得到的结果如表1所示。为避免个别聚类结果有偏差,所示数据均为计算100次后取得数学期望值,本文后续所呈现两种模式的数据均为期望值。

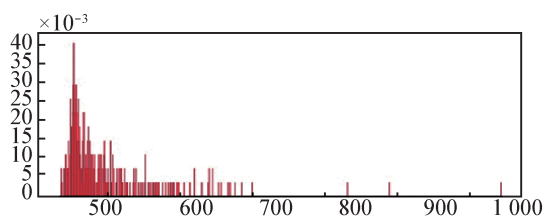


图1 跨包延迟 IPD 的概率分布函数

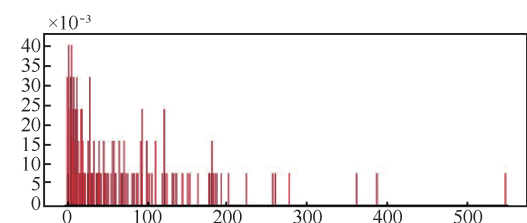
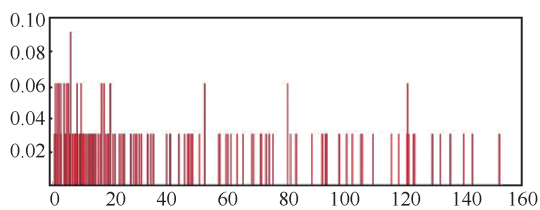
图2 延迟距离 $D(s_j, d)$ 的概率分布函数图3 IQR 异常值识别后 $D(s_j, d)$ 概率分布函数

表1 原始数据聚类结果

名称	聚类中心 C_0	聚类中心 C_1
中心值/ms	19.650 164 83	85.742 117 29
中心数目	140	57

从远离0点的聚类中心 C_1 的大小来设置水印 ω 。由表1可得聚类中心 C_1 的大小约为86 ms, 本次实验以较接近的值100 ms为 ω 的大小。加入水印后计算出 IPD, 并用 IQR 筛选剔除6.13个极端值, 保留原始数据97.13%, 再次利用改进算法计算出聚类结果如表2所示。

从两次聚类的结果可以明显看出: 加入水印后的远0端聚类中心 C_1 内聚集数, 相比原始远0端聚类中心 C_1 发生了变化, 如图4所示, 且占比

约等于水印占比 μ 。移除水印后, 得到的聚类结果恢复如表1所示。由此可以得出, 该方案注入的水印有较稳定的特征, 水印加入与否对比明显, 检测结果清晰。

表2 加入水印聚类结果

项目	聚类中心 C'_0	聚类中心 C'_1
中心值	23.265 414 24	99.079 305 76
中心数目	115	91

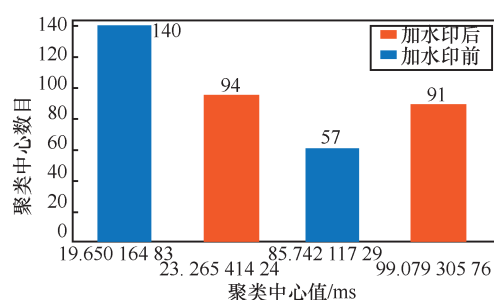


图4 加水印前后结果对比

由于 obfs4 存在有3种模式, iat-mode 可取值有0、1、2, 官方声明后两种模式会对数据包的时间和大小采取进一步处理, 以干扰潜在的分析机制检测出 obfs4 的流量。下面分别对这3种模式进行实验, 将上下行分组大小和时间进行了统计, 得到的结果如下, 图5是固定大小、持续发送的数据包上下行统计结果, 图6是持续增长大小、持续发送的数据包上下行统计结果。

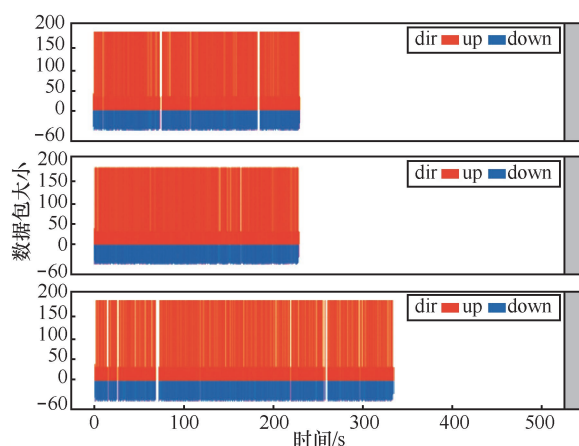


图5 固定大小、持续发送的数据包上下行统计

在统计中可以发现, $iat-mode=1$ 时, 包大小和包速率稳定的情况下, 可选保护机制不会产生作用, 传输时间甚至比模式 $iat-mode=0$ 时还短; 而当保护机制启动时, 可以明显看出延长了整体的传输时间, 刻意增加了许多较大的时间间隔以干扰流水印等方式的检测效果。对两种模式采用相同的实验步骤取得的结果如图 7 所示, 水印值分别设置为 50 ms 和 300 ms。实验结果表明, 本文方案能够应对 obfs4 上 Tor 流量的 3 种不同模式。

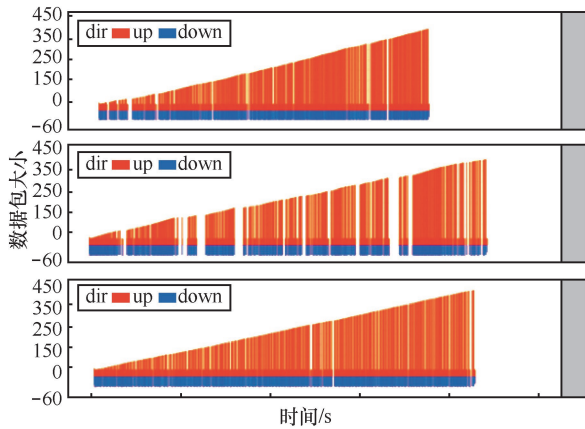
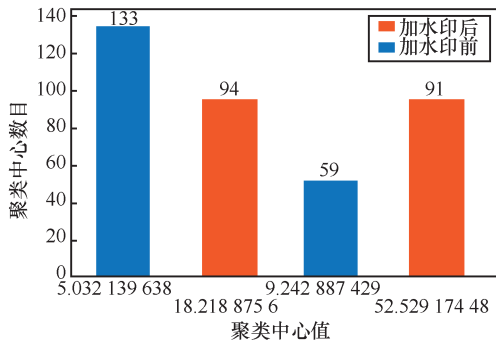
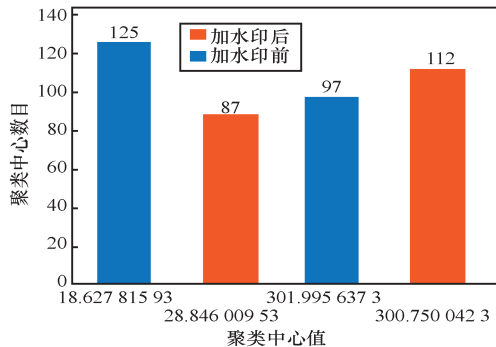


图 6 持续增长大小、持续发送的数据包上下行统计结果



(a) $iat-mode=1$, 水印大小 $\omega=50$ ms



(b) $iat-mode=2$, 水印大小 $\omega=300$ ms

图 7 加水印前后聚类中心分布

原算法^[6]在这一实验中计算出的原始流量独立同分布 $\overline{Y_{r,d}}$ 对称中心十分不稳定, 按照算法说明对称中心大于 0 表明嵌入水印位“1”; 对称中心小于 0 表明嵌入水印位“0”。令 $r=14$, 检测 $iat-mode=0$ 水印, 加水印前后计算 1 000 次结果如图 8 所示, 存在很高的误差。由于实验加入水印位均为“1”, 得出原算法在 obfs4 $iat=0$ 上 Tor 流量的检测率为 51%。在 $iat-mode$ 为 1、2 时检测率分别为 50.6%和 49.5%。

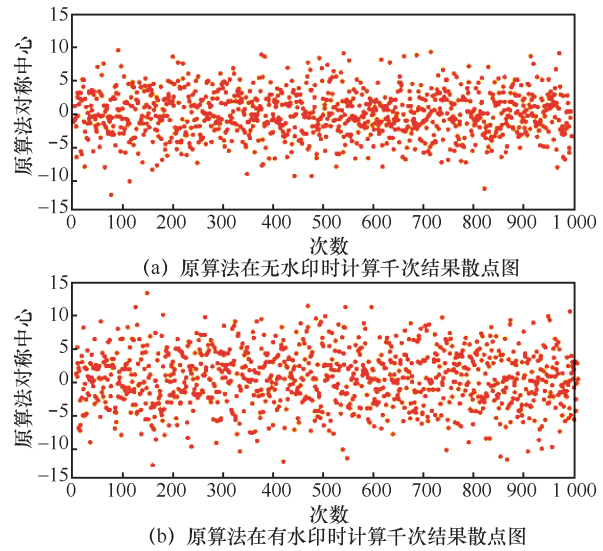


图 8 原算法加水印前后计算结果

分别对 3 种模式实验数据进行 1 000 次 IPD 质心计算, 调整容错值 (α, β) 的大小, 实验结果如图 9 所示, 其中, 横轴表示 α , 颜色的深浅表示 β , 纵轴表示检测率, $\alpha \in [5, 9]$, $\beta \in [6, 10]$ 。在 3 种模式下, 微调容错值 α 和 β 都能有效地提高算法检测的检测率, (α, β) 取值均为 (7,7) 以上可以达到至少 94.0% 的检测率, 与原算法相比检测率有明显提升, 但意味着此时仍有 6% 的可能性检测不出加入水印的流量。

模拟相对真实的环境以测试算法对水印的识别率, 在伦敦配置的 Tor 出口节点服务器上抓取传送出的流量, 427.49 s 的时间 (约 7 min 7 s) 产生了 2.91 GB 大小的 Pcap 数据包。根据往返 IP

地址将其划分为 65 349 个连接, 对其进行 IPD 质心水印算法的检测, 取 (α, β) 为 (7,7), 其中满足 α 的目标有 30 个, 满足 β 的目标有 22 个, 二者同时满足的流量唯一且为目标流量。

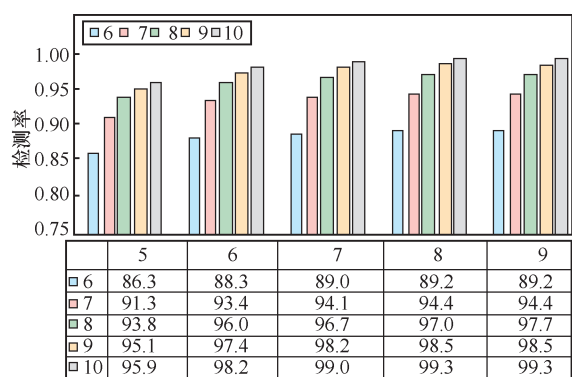
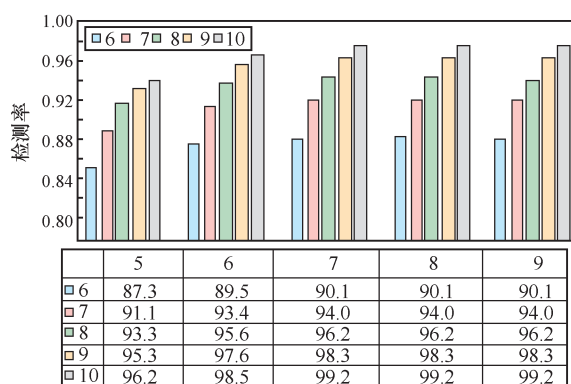
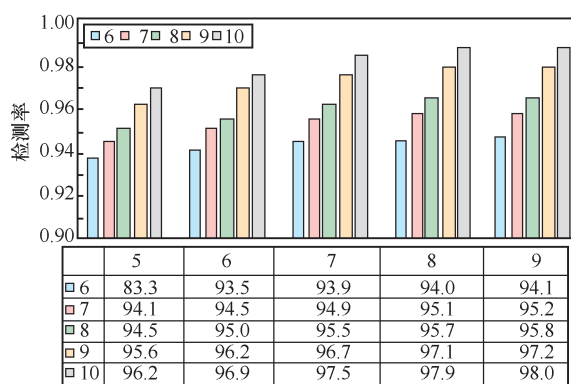
(a) $iat-mode=0$ (b) $iat-mode=1$ (c) $iat-mode=2$

图9 3种不同容错值下的检测率

5 结束语

综上所述, 本文方案具有较高的检测率、识别率和稳健性, 体现在: 定性分析得到的图像足以判

断加入水印与否, 定量分析为其提供了更加准确的结果判定, 检测率均能达到 94% 以上; 真实复杂的流量环境中能够唯一识别目标流量; 水印特征保持稳定不会被轻易移除, 对 obfs4 上不同模式的 Tor 流量都能得到良好的结果。相比原始方法, 采用统计算法聚类代替概率算法, 区分原始流量与加水印流量效果明显, 所使用数据包数量较小, 检测率有了显著提高, 且运行时 CPU 速度无显著增加。

由于不同实验环境的数据流会根据当前情况调制水印, 网络的延迟较大对应水印值也大, 检测效果会动态地变化, 而 IPD 质心方案采用半盲的方式, 先对原始流量进行分析以确定水印大小, 从容应对各种环境。无论水印的大小 ω 如何选择, 在搭载 obfs4 的 Tor 流量上总会出现与加入水印大小 ω 相同的延迟距离。不同的 obfs4 模式也会干扰性地增加额外延迟。本文方案使用 $\eta=0.8$ 的 IQR 进行极端值排查, 过于异常的间隔时间被剔除, 基于 k -means 聚类算法的特性, 小幅增加时间间隔对聚类中心的影响甚微, 靠近峰值的大量数据会使其欧氏距离包含其中, 对水印占比造成的影响很小, 不足以导致结果产生决定性误差。

向 bridges@torproject.org 发送电子邮件从而获取了 492 个 obfs4 网桥, 其中 458 个网桥采用的是 $iat-mode=0$, 34 个 obfs4 网桥采用的是 $iat-mode=1$, 没有 $iat-mode=2$ 的情况。鉴于 Tor 的连接速度相对缓慢, 多数用户为提高传输效率不采用后两种模式, 尽管本文方法在 $iat-mode=0$ 的情况下具有较好的使用效果, 但对比后两种方案的稳定性略显不足, 从实际应用的角度出发应进一步提高这种模式的算法稳定性。

相比原方案, 本文方案在水印隐蔽性和水印容量上均存在不足。流水印技术的隐蔽性和健壮性无法同时满足, 从与当前广泛使用的 ANFW 技术进行对比^[11]可以看出。Lei 等^[12]运用最优化理论模型证明了这个非此即彼的问

题：衡量不可见性的失真函数，与衡量健壮性的距离函数在理想条件下相等，均小于一个阈值 D_s ；该阈值越大，失真越多，不可见性越弱，健壮性则更加顽强；反之依然成立。本文方案水印的嵌入只通过增加时间间隔实施，隐蔽性相对较弱，对于一些针对水印的攻击不能有效地应对，如 MFA 攻击^[13]，该方法可以寻求水印存在的规律从而进行抹平或伪造。如何提高隐蔽性是下一步工作的重点。

原方案用对称中心的位置代表“0”“1”两种状态，可包含的水印容量足，但代价是需要较为持续稳定的数据流量。IPD 质心方案舍弃了水印容量以换取少量数据流量的需求，在实际应用中对追踪目标及时采取应对措施，比增加多样化的水印容量、隐藏自身检测行为更行之有效。

此外，本文方案水印大小的选取依赖正常流量的分析，因此是半盲属性的，主要适用于针对某一特定的流量进行追踪。若应用于大范围、多流量的匿名网络分析，半盲属性会成为整个方案的短板。降低对正常流量的依赖也有待未来的研究。

参考文献：

- [1] WANG X, REEVES D. Robust correlation of encrypted attack traffic through stepping stones by flow watermarking[J]. IEEE Transactions on Dependable and Secure Computing, 2011, 8(3):434-449.
- [2] SNOEREN A, PATRIDGE C, et al. Hash-based IP traceback[C]// Proceedings of ACM SIGCOMM, 2001: 3-14.
- [3] GOODRICH M T. Efficient packet marking for large-scale iptraceback[C]//Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002). 2002: 117-126.
- [4] 吕博, 廖勇, 谢海永. Tor 匿名网络攻击技术综述[J]. 中国电子科学研究院学报, 2017(1): 14-15.
- LV B, LIAO Y, XIE H Y. Survey on attack technologies to Tor anonymous network[J]. Journal of CAEIT, 2017(1): 14-15
- [5] 何永忠, 李响, 陈美玲, 等. 基于云流量混淆的 Tor 匿名通信识别方法[J]. 工程科学与技术, 2017(2): 121-122.
- HE Y Z, LI X, CHEN M L, et al. Identification of Tor anonymous communication with cloud traffic obfuscation[J]. Advanced Engineering Sciences, 2017(2): 121-122.
- [6] WANG X, CHEN S, JAJODIA S. Tracking anonymous peer-to-peer VoIP calls on the Internet[C]//ACM Conference on Computer and Communications Security. 2005: 81-91.
- [7] HOUMANSADR A, KIYAVASH N, BORISOV N. RAINBOW: a robust and invisible non-blind watermark for network flows[C]// Network and Distributed System Security Symposium, San Diego, 2009.
- [8] WANG X, CHEN S, JAJODIA S. Network flow watermarking attack on low-latency anonymous communication systems[C]// IEEE Symposium on Security and Privacy. IEEE Computer Society, 2007: 116-130.
- [9] PYUN Y J, PARK Y H, WANG X, et al. Tracing traffic through intermediate hosts that repacketize flows[C]//IEEE International Conference on Computer Communications. 2007: 634-642.
- [10] PANCHENKO A, LANZE F, ZINNEN A, et al. Website fingerprinting at internet scale[C]// Network and Distributed System Security Symposium. 2016.
- [11] 郭晓军, 程光, 朱琛刚, 等. 主动网络流水印技术研究进展[J]. 通信学报, 2014, 35(7): 178-192.
- GUO X J, CHENG G, ZHU C G, et al. Progress in research on active network flow watermark[J]. Journal on Communications, 2014, 35(7): 178-192.
- [12] LEI C, ZHANG H, LIU Y, et al. Net-flow fingerprint model based on optimization theory[J]. Arabian Journal for Science & Engineering, 2016, 41(8): 3081-3088.
- [13] KIYAVASH N, HOUMANSADR A, BORISOV N. Multi-flow attacks against network flow watermarking schemes[C]//Conference on Security Symposium. USENIX Association, 2008: 307-320.

[作者简介]



杜捷（1993—），男，青海海东人，北京交通大学硕士生，主要研究方向为网络安全、匿名通信。



何永忠（1969—），男，重庆人，博士，北京交通大学副教授、硕士生导师，主要研究方向为网络安全、计算机安全、密码协议。



杜晔（1978—），男，黑龙江哈尔滨人，博士，北京交通大学副教授、博士生导师，主要研究方向为网络安全、态势感知、软件可靠性分析与评估。