信源定位方案中基于 Bloom Filter 存储的概率日志记录方法研究

薛开平[①]　洪佩琳[①]　郭　婵[①]　卢汉成[①]　骆连合[②]
[①](中国科学技术大学电子工程与信息科学系　合肥　230027)
[②](中国电子科技集团公司第 54 研究所　石家庄　050081)

摘　要：该文在信源定位方案中提出了一种基于 Bloom filter 存储的概率采样日志记录方法。该方法对经过路由器的所有数据实现概率采样，存储采用了高效的 Bloom filter 存储结构，使得采样信息能够在一定时间内存储在内存中便于查找。基于此方法该文提出信源定位服务器的概念，从而使得核心网络路由器除了路由转发功能之外，只需要完成对数据包的概率采样即可。文中还对相关参数的选择进行了理论分析，从理论上分析了信源定位服务的存储开销以及信源定位有效性，方案具有存储开销小、效率高的特点，从而为进一步的实际网络部署提供了理论依据。

关键词：信源定位；日志记录；Bloom filter；概率采样
中图分类号：TP393　　**文献标识码**：A　　**文章编号**：1009-5896(2009)11-2738-06

Study of Probabilistic Logging Based on Bloom Filter for Source Tracing

Xue Kai-ping[①]　Hong Pei-lin[①]　Guo Chan[①]　Lu Han-cheng[①]　Luo Lian-he[②]
[①](*Department of Electronic Engineering and Information Science*, *University of Science and Technology of China*, *Hefei* 230027, *China*)
[②](*The* 54*th Research Institute of China Electronic Technology Group Corporation*, *Shijiazhuang* 050081, *China*)

Abstract: This papar presents a probabilistic logging scheme based on Bloom filter for source tracing. The scheme makes probabilistic sampling of all packets through each router, and uses efficient Bloom filter for storage. The sampling information can stored in memory, which make it easier to find. This paper introduces first the concept of source locating server. Besides forwarding packets, the routers in the core network only need probabilistic sampling of packets. In addition, this paper gives theoretical analysis of the choice of the relevant parameters. In theory, This paper analyzes the cost of storage in probabilistic logging scheme and the validity of source location. The proposed scheme has the characteristics of small storage costs and high efficiency, which provides a theoretical basis for further actually deplyment.

Key words: Source tracing; Logging; Bloom filter; Probabilistic sampling

1　引言

Internet 的开放性是其能够获得成功的重要因素之一，但这也为攻击者们提供了便利。现在网络攻击呈现出大规模、多样化、隐蔽性强、智能化和综合化等特点，新的攻击形式层出不穷。这使得网络安全的威胁日益严重，尤其是在攻击使用伪造 IP 地址的情况下，使被攻击者很难确定攻击源的位置，从而不能有针对性的实施保护策略，也阻碍了入侵反击和恢复功能的发挥。因此，信源定位技术就显得尤为重要。它采取主动出击的策略帮助被攻击者定位攻击源头。根据信源定位的结果，不仅可以有针对性地在更适当的位置部署相关防御措施(如流量限速器)，而且可以从攻击的源头有效地扼制攻击流对中间传输网络和被攻击者的影响。同时信源定位技术还是计算机犯罪取证的重要手段。

信源追踪与定位技术的研究从 20 世纪九十年代就已开始，发展到现在已经有着很多的方案，但现有的方案在现实构建中缺少可行性，要么需要路由器的过多参与，在路由功能的基础上增加太大的负担；要么修改网络协议，改造数据包，这往往会带来新的安全问题。基于上述原因，本文提出一种全新的基于 Bloom filter 存储的概率采样日志记录法(PSL-BF：Probabilistic Sampling Logging based on Bloom Filter)。其主要思想是在核心网络每个路由器并置一“镜像”服务器(本文称之为“信源定位服务器”)，路由器不直接提取和存储日志信息，而是将数据包概率采样并“镜像”给信源定位服务器，由信源定位服务器来完成数据包日志信息的提取和记录工作。所有的信源定位服务器构成信源定位系统。在信源定位过程中则完全由分布式的信源定位

2008-12-01 收到，2009-04-28 改回
国家自然科学基金(60602018，60772033)，中瑞国际合作项目(2008DFA11950)和安徽省高等学校优秀青年人才基金重点项目(2009SQRZ004ZD)资助课题

服务器协同交互完成，而无需路由器的参与，从而保证了路由器功能的单一性和安全性。此外在信息的存储上本文采用了高效的 Bloom filter 存储结构，使得采样信息能够在一定时间内存储在内存中而便于查找。由于 Bloom filter 会随着插入信息的增多，而增加查找误差，PSL-BF 方案将会设定一定的时间阈值，周期性将当前的 Bloom filter 结构信息转而存储到硬盘或者其他存储介质中，并将当前的 Bloom filter 结构清空。从而可以保证信源定位过程的有效性和查找的快捷性。

本文组织如下，第 2 节概述相关工作，第 3 节介绍 PSL-BF 方案，第 4 节给出性能分析，第 5 节为结束语。

2　相关工作

日志记录法的基本思想是在路由器中记录下与报文有关的日志，然后采用数据挖掘技术得到报文传输的路径，其实质是对日志记录的查询。攻击发生时，受害者提取攻击数据包的相关日志信息，依据该信息询问上游路由器，查看该数据包是由哪些路由器转发的。上游路由器接到查询时，检查自己的数据库中是否有该日志的摘要，如果有，则判断攻击流经过了该路由器，并依此法逐级向上直至查出数据包源头。

最早的日志记录法直接记录数据包的日志信息，由 Chang 等人提出[1]。但文献[1]中的日志记录法被认为是不切实际的，路由器往往无法承受大量的数据包日志记录的存储开销。Snoeren 等人在文献[1]基础上进行了改进，提出一种只记录报文摘要(Hash)的源路径隔离引擎(Source Path Isolation Engine，SPIE)[2]，SPIE 采用 Bloom filter 的数据结构存储报文摘要。进一步地，文献[3]在文献[2]的基础上提出了对工程实现的讨论，并且在一定程度上解决了文献[3]中错误结果返回的问题。但是在现实中，路由器处理数据包日志的速率必须与数据包的到达速率相当，文献[2,3]要求每个具有追踪功能的路由器对经过它的每个报文都进行若干次散列函数计算来说，是难以实现的。Lee 等[4]提出在日志中记录流的审计信息，而非每个报文的审计信息，以减少日志的产生量，这种方法虽然可以在一定程度上减少日志量，但并不能从本质上减少内存开销，并且还同时提高了路径重构的误判率。Gong 等[5]则提出将日志法和包标记法交替使用以减少日志量和计算量。类似地，Takurou H 等所提出方案[6]也将日志记录法和包标记法结合使用，并在 SPIE 的基础上随机采样数据包进行日志记录。但 Gong 提出方案以及文献[6]中的方案，与传统的包标记方法一样，需要对数据包格式进行定义或者特定域的复用。此外所有以上方案，在信源定位过程中需要有路由器的参与，从而分布式定位出攻击源，这与普遍意义上需要对路由器实现单一化功能的目标是相背离的。

本文提出的基于 Bloom filter 存储的概率采样日志记录法，并基于此提出了信源定位服务器的概念，不会给路由器带来太多额外的开销，能够取得很好的效果。

3　方案描述

3.1　信源定位服务器的布置和分布式互联

如图 1 所示，PSL-BF 方案在每个路由器上并置一个信源定位服务器，并在路由器上使用特定的数据包采样技术将本地的转发流量进行概率采样，通过特定的端口“镜像”到与其并置的信源定位服务器上，由这些信源定位服务器处理并记录所需的相关信息。在信源定位过程中通过这些服务器的分布式交互来完成信源定位功能。路由器除了必要的数据“镜像”功能之外，只负责高速的路由转发而无需对信源定位功能进行特殊的处理。信源定位服务器之间可以采用 IP 协议的方式进行互联，无需重新布置专网。每个信源定位服务器知道相邻路由器并置的信源定位服务器的可访问 IP 地址信息。在信源定位过程中，这些信源定位服务器组成分布式系统，通过有效的交互查找机制，定位出真实的攻击源或者最大程度地反向恢复出攻击路径。

广域网路由器的“镜像”功能能够对数据包只需要进行采样并“镜像”到并置的信源定位服务器上，由信源定位服务器来完成包记录算法的剩余操作。这样可以减少路由器的负担，在不影响路由器

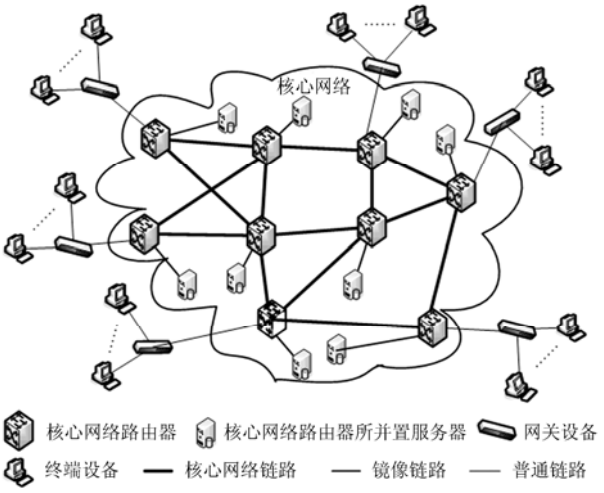

图 1 信源定位系统部署示意图

的转发速度的前提下实现高效的信源定位。“镜像”功能可以通过在路由器上嵌入相关硬件芯片来实现，如 sFlow 技术[7]、Netflow 技术[8]和 Port Mirroring[9]等都具有类似的功能。

路由器上的 “镜像”模块对路由器所有端口进入的流量都要进行操作。但由于是基于概率的方式，同时该模块不需要对数据包进行处理，只需依概率采样数据包并将其通过特定的物理端口“镜像”给服务器即可。对于路由器而言，添加的功能能够以较低的代价处理较高速的转发流量的“镜像”功能，而不会影响到实际的路由转发能力。具体流程如图 2 所示。

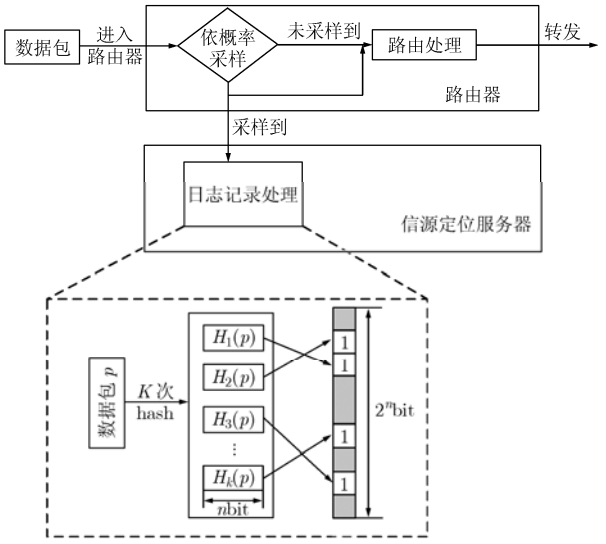

图 2 PSL-BF 中数据包的记录操作流程

3.2 基于Bloom filter的日志记录法

信源定位服务器独立地对采样的 IP 报文进行数据包特征信息提取并将该特征信息加以存储。具体地，特征信息定义为数据包 IP 头的不变域和数据载荷中的前 8 字节，这类似于文献[3]所采用的方法。

数据包的特征信息以 Bloom filter 的数据结构存储在信源定位服务器上。Bloom filter[10]是一种空间效率很高的随机数据结构，它用一个定长的位串简洁地索引一个集合，有效的节省了存储空间，并能快速判断一个元素是否属于这个集合。Bloom filter 使用 k 个独立的 hash 函数对每个数据包进行 hash，得到 k 个 hash 值，每个 hash 值长度为 n。然后以长为 2^n 比特的位串来索引这些 hash 值，每个 hash 值映射到其中的一位上，即以 hash 值为索引将位串中对应比特置“1”(初始时，Bloom filter 所有位的值均为“0”)。如图 2 中所示，对数据包 p，信源定位服务器将其 k 个值 $H_1(p),H_2(p),\cdots,H_k(p)$ 映射到 Bloom filter 的 k 个不同的位置。

检测一个元素是否在 Bloom filter 中时，先计算出该元素的 k 个 hash 值，再检查这 k 个值在 Bloom filter 中对应的位是否全为 1，如果是，则表明该元素在 Bloom filter 中被索引，我们称为一次命中。但这种方式存在误检情况，我们称为假命中。并且随着数据包数目 N 的增大，假命中的概率也会增大。为了减少假命中的概率，在 PSL-BF 方案中采用周期性更新 Bloom filter 的机制。即路由器并置的服务器每隔时间 ΔT_R 将 Bloom filter 置位信息写入磁盘中存储，同时标注上该 Bloom filter 的记录时间间隔(所有信源定位服务器实现精确性要求不高的时间同步)，然后把 Bloom filter 全部置零重新进行映射记录。

3.3 信源定位流程描述

信源定位系统中所有信源定位服务器形成一个整体。信源定位过程由被攻击端或者特定的检测设备发起，主要的信源定位工作则由分布式的信源定位系统完成。一次完整的信源定位过程可以分为以下 3 个部分。

(1)被攻击端或者检测设备发起信源定位过程：当被攻击段检测到攻击需要进行信源定位时，其首先提取部分攻击数据包(称为样本攻击包)的特征信息，并分别计算出这些特征信息相对应的 hash 值(Bloom filter 所使用的 k 个)，然后被攻击端将 hash 值发送给上游的几个信源定位服务器发起查询请求。

(2)信源定位服务器进行信源定位：中间的信源定位服务器接到查询请求时，首先检查自己的数据结构中是否存储这些特征信息中的一个或者多个。如果没有数据包被命中，则认为攻击流没有经过该路由器，返回结果。反之，如果至少有一个数据包被命中，则认为攻击流经过了对应的路由器，并进一步上上游路由器发起查询请求。重复以此法逐级向上直至查出攻击端或者查询终止于上游路由器所并置服务器。

(3)信源定位服务器向攻击端或者检测系统返回定位结果：定位到某个边缘路由器并置的信源定位服务器在确认存在存储后向攻击端或者检测设备返回信源定位结果。

以图 3 为例(路由器和其并置的信源定位服务器统一由圆圈表示)，当受害者 A 发现被攻击时，采样攻击数据包并向上一级路由器绑定的信源定位服务器 $R1$ 发出查询请求，请求被接受后，由 $R1$ 代理进行信源定位过程。$R1$ 继续向其上一级路由器绑定的信源定位服务器 $R2$，$R3$ 发送查询请求，$R2$，$R3$

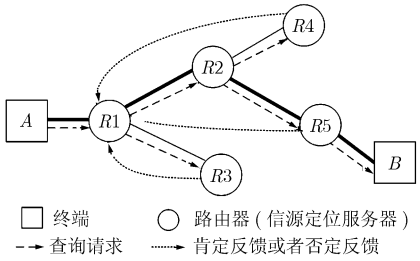

图 3 日志记录法追踪过程示意图

分别查询自己数据结构中是否包含有一个或者多个受害者提供的信息摘要结果，如果有，则继续上上游信源定位服务器发起查询请求。$R2$ 有相应的结果命中，则继续向 $R4$，$R5$ 发起查询请求。$R3$ 没有相应的结果命中，则返回否定应答。如此类推，直到找到攻击者 B。

具体发送的查询请求的数据结构如图 4 所示，“所有摘要信息完整性校验和”域是为了确认摘要信息在传输中过程中产生误码。该域计算和校验方式与 IP 协议中的校验和域方式一致。“路径节点标识域”为在每个命中查询的信源定位服务器处将本身所并置的路由器标识添加后末尾，并且将路径长度加 1。

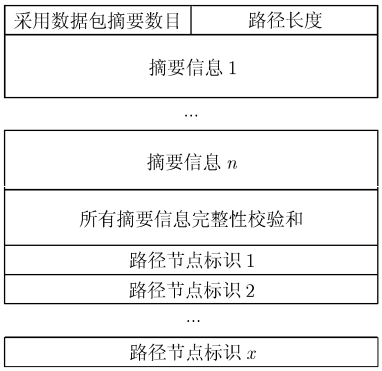

图 4 查询消息结构

4 理论分析

为了讨论方案的有效性，这里将从两个方面进行理论分析，分别为信源定位服务器存储开销以及定位有效性。这两者是反映信源定位方案有效性的两个重要因素。

4.1 信源定位服务器存储开销

为了保证 Bloom filter 假命中率尽可能小，必须使得存储开销变大来做保证。假设 Bloom filter 采用 2^n 长度的存储空间，使用 k 个 Hash 函数作为索引函数。可以推算出在对 N 个不同的数据包特征信息进行索引之后，该 Bloom filter 的假命中的概率 f 为

$$f=\left(1-\left(1-1/2^n\right)^{kN}\right)^k \tag{1}$$

令 $(1-1/2^n)^{kN}\approx e^{-(kN/2^n)}$，则式(1)可进一步得到 $f\approx(1-e^{-(kN/2^n)})^k$。固定 n 和 N，并令 $f=e^g$，其中 $g=k\ln(1-e^{-(kN/2^n)})$，则求 g 的最小值的过程即是求假命中率 f 最小值的过程，求导可得

$$\frac{\mathrm{d}g}{\mathrm{d}k}=\ln\left(1-e^{-(kN/2^n)}\right)+\frac{kN}{2^n}\cdot\frac{e^{-(kN/2^n)}}{1-e^{-(kN/2^n)}} \tag{2}$$

令 $\mathrm{d}g/\mathrm{d}k=0$，则可知 $k=(\ln 2)\cdot(2^n/N)$ 时误检率 f 取最小值，为

$$f_{\min}=e^{-(\ln 2)^2\cdot(2^n/N)} \tag{3}$$

假设核心网络路由器对流量的转发速度为 G bit/s，每个信源定位服务器上 Bloom filter 设定的更新周期为 ΔT_R。取一个 IP 包的大小为 1500 Byte(这是以太网环境中的典型值)。可得到路由器转发包的速度为 $G/(1500\times 8)\approx G\cdot 10^{-4}$ packets/s。在采样概率为 P_R 的情况下，对于信源定位服务器而言，被随机采样到的包的到达率即为 $P_R\cdot G\cdot 10^{-4}$ packets。那么 Bloom filter 在一个更新周期 ΔT_R 里面里应当存储的包个数即为 $P_R\cdot G\cdot\Delta T_R\cdot 10^{-4}$ packets。取 $N=P_R\cdot G\cdot\Delta T_R\cdot 10^{-4}$，当 $k=(\ln 2)\cdot(2^n/(P_R\cdot G\cdot\Delta T_R\cdot 10^{-4}))$ 时，假命中概率 f 达到理论上的最小值：

$$f_{\min}=e^{-(\ln 2)^2\cdot\left(\frac{2^n}{P_R\cdot G\cdot\Delta T_R\cdot 10^{-4}}\right)}=e^{-(\ln 2)^2\cdot\left(\frac{2^n/\Delta T_R}{P_R\cdot G\cdot 10^{-4}}\right)} \tag{4}$$

两边取 ln 可得：

$$2^n/\Delta T_R=\frac{\ln(1/f_{\min})}{(\ln 2)^2}\cdot P_R\cdot G\cdot 10^{-4}\,\mathrm{bit/s} \tag{5}$$

其中 $2^n/\Delta T_R$ 表示了信源定位服务器每秒钟需要多少存储空间来进行日志记录，即信源定位服务器在进行日志记录操作时的存储空间消耗速率，对于信源定位系统而言，从时效性角度，过于长时间之前的信息通常情况下不再需要保存。

图 5 和图 6 分别显示为信源定位服务器的存储消耗随着路由器转发速率和 Bloom filter 误检率的变化而变化的情况。图 5 显示为在记录概率 P_R 取 1 时，假命中率分别为 0.1%，0.5%，1%和 5%时，信源定位服务器存储损耗速率与路由器转发速率直接的关系。由图 5 可知，当误检率控制在 0.1%时，处理能力为 10 Gbps，40 Gbps 和 320 Gbps 的路由器在全概率采样的情况下服务器每秒消耗的存储空间约为 14.37759 Mbps，57.51035 Mbps 和 460.0828 Mbps。当进行概率随机采样时，存储损耗速率将按 P_R 的比例缩小。由此我们可以得出基于 Bloom filter 的随机采样日志记录法的存储开销是能够接受的。图 6 显示为理论上当路由器转发速率分别为 10

Gbps，40 Gbps 和 320 Gbps 时服务器存储空间消耗速率与误检率 f 的变化关系。从图结果可以分析出对于一个转发速率为 40 Gbps 的路由器，当误检率保持在 0.1%时，所并置的信源定位服务器存储空间消耗速率为 57.51035 Mbps，而当误检率为 5%时其值仅为 24.9409 Mbps，即随着假命中概率 f 的降低存储空间消耗速率上升的并不快。同样，当进行概率随机采样时，存储损耗速率将按 P_R 的比例缩小。

4.2 基于Bloom filter的随机采样日志记录法信源定位的有效性分析

攻击分为单源攻击和多源攻击两种，多源攻击实际上是单源攻击的组合。这里首先考虑单源攻击的情况。首先考虑在单个信源定位服务器上正确判定攻击流是否经过路由器的概率。当路由器随机采样的概率为 P_R，并在终端或者特定的检测设备采用 m 个样本攻击包作为样本进行信源定位时，能够正确判断攻击流是否经过该路由器的概率应为(在不考虑 Bloom filter 的误检率的情况下)：

$$P = 1-(1-P_R)^m \tag{6}$$

由式(6)可知，在要求判定概率 P 达到一定精度的情况下，如果 P_R 越大，则 m 越小。本文分别取 P 的值为 0.9，0.99，0.999 和 0.9999，计算范围为 1%到 50%的 P_R 值下，进行信源定位时提取样本攻击包的个数 m，结果如图 7 所示。由图可见，当随机采样的概率 P_R 为 0.1，样本攻击包的个数 m 为 100 时，P 的值可以达到 0.99976，精确度较高。即使在本方案中加上 Bloom filter 的误检率，仍然具有较大概率成功定位到攻击源。在单个信源定位服务器上的正确判定攻击流是否经过该路由器的概率 P 将直接影响最终信源定位成功的概率。下面分别讨论在单源攻击和多源攻击情况下，基于 Bloom filter 的随机随机采样日志记录法信源定位的有效性进行分析。

如图 8，当攻击为单源攻击时，攻击路径上的每个攻击包都以概率 P_R 被路由器采样。攻击包通过同一条路径由攻击端到达被攻击。在单源攻击的信源定位过程中，攻击路径上每个服务器正确判定攻击流是否经过的概率 P 均为 $1-(1-P_R)^m$。那么，对于一个包含 h 跳的攻击路径来说，正确定位攻击端的概率即为 $(1-(1-P_R)^m)^{h+1}$。图 9 显示了 P 值分别为 0.9，0.99，0.999 和 0.9999 的情况下，攻击路径跳数 h 由 1 到 50 范围内变化情况下，能够成功定位攻击端的概率变化情况。可见，当 P 值为 0.9999 时，正确定位攻击端的概率始终在 98%以上，即使 P 值为 0.999(当 m=100，P_R=8%可使 P 达到该值)时，正确定位攻击端的概率仍能保持在 92%以上。这显然是一种极端的情况，因此基于 Bloom filter 的随机采样日志记录法可以较大概率有效的定位单源攻击的攻击端。

在多源攻击中，因为存在攻击流的汇聚问题，所以如果仍在所有的路由器上使用 m 个样本攻击包来进行判定，则正确判断攻击流是否经过该路由器的概率不再恒为 $1-(1-P_R)^m$。图 10 是一个简单的多源攻击示例。图中 3 个攻击端 $A1$，$A2$，$A3$ 沿着不同攻击路径对同一个终端进行攻击。假设这 3 个攻击端攻击时间和发送的攻击流量是相同的，可以看到链路 1 中的攻击流是由链路 2，链路 3 的攻击流汇聚而成的，也就是说，链路 1 中的攻击流有部分来自链路 2，2/3 来自于链路 3。同样被攻击端收到的攻击包也是 1/3 来自链路 2，2/3 来自于链路 3。如果被攻击端随机提取 m 个包作为样本包在 $R2$ 上判定，则由于这 m 个包只有 1/3 的概率经过了 $R2$，所以他们被记录的概率降为了 $P_R/3$，$R2$ 能够正确判定攻击包经过的概率降为 $1-(1-P_R/3)^m$。以此类推，随着分支路径的增多，每个服务器正确判定的概率也会降低，从而最终导致信源定位正确性的降低。

为了提高多源定位的准确度，可以在分支对应的路由器并置服务器上查询时提高样本攻击包的个数。随着攻击源的增加，样本攻击包的需求个数也

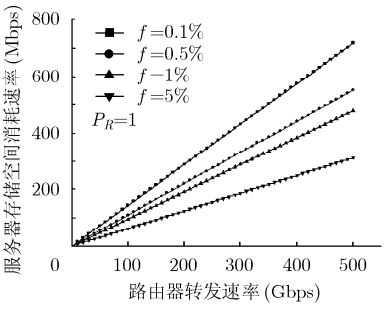

图 5 信源定位服务器存储空间消耗速率随转发速率的变化关系图

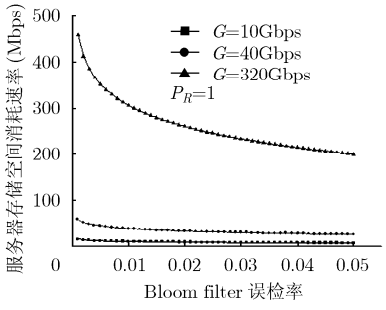

图 6 信源定位服务器存储空间消耗速率随 Bloom filter 误检率的变化关系

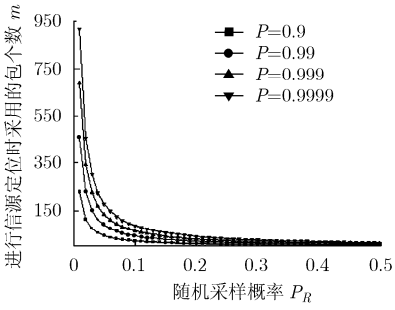

图 7 随机采样概率和攻击包样本数示意图

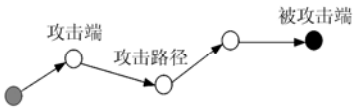

图 8 单源攻击示意图

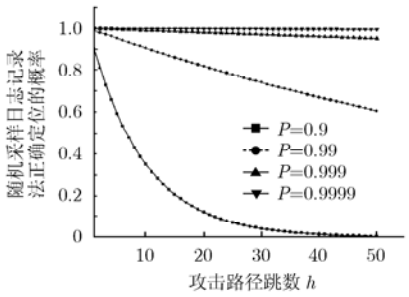

图 9 PSL-BF 信源定位正确性分析

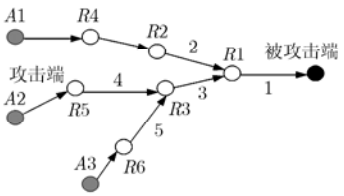

图 10 多源攻击示意图

会成倍增长，可能会导致查询开销过大，降低了信源定位系统的效率。所以，基于概率包记录机制的信源定位方法虽然对于单源攻击可以有效地定位到攻击源，对于多源攻击则具有一定的局限性。但在一定极端情况下即使无法恢复出所有到达攻击源的多路径汇聚树，但至少是该汇聚树的部分，该结果对于信源定位和后续安全布防是有借鉴意义的。

5　结束语

本文提出了使用在信源定位中一种全新的基于 Bloom filter 存储的概率日志记录法，方案具有两个特点：(1)提出并置于路由器的信源定位服务器的概念；(2)对经过路由器的所有数据包实现概率采样，存储由信源定位服务器基于 Bloom filter 实现。文中还基于以上创新描述了信源定位过程。在信源定位过程中，完全由信源定位服务器分布式实现相关功能。方案不会给路由器带来太多额外的开销，实现简单，并且具有维护开销小，效率高的特点。进一步的工作是基于方案实现信源定位服务器功能，并进行规模测试。

参 考 文 献

[1] Chang H Y, *et al.*. DecIDUouS: Decentralized source identification for network-based intrusions. Integrated Network Management, Distributed Management for the Networked Millennium. 1999.

[2] Snoeren A C, *et al.*. Hash-based IP traceback. Proceedings of the ACM SIGCOMM 2001, San Diego, California, USA, August 27–31, 2001: 3–14.

[3] Hilgenstieler E, Duarte E P, Mansfield-Keeni G, and Shiratori N. Improving the precision and efficiency of log-based IP packet traceback. Proceedings of 50th IEEE Global Communications Conference (GLOBECOM'07), Washington D.C., USA, 2007: 1823–1827.

[4] Lee Tsern-Huei, Wu Wei-Kai, Yau Tze, and Huang William. Scalable packet digesting schemes for IP traceback. The 2004 IEEE International Conference on Communications (ICC'04), Paris, France, 2004: 1008–1013.

[5] Gong Chao and Sarac K. A more practical approach for single-packet IP traceback using packet logging and marking. *IEEE Transactions on Parallel and Distributed Systems*, 2008, 19(10): 1310–1323.

[6] Takurou H, Matsuura K, and Imai H. IP traceback by packet marking method with Bloom filters. The 41st Annual IEEE International Carnahan Conference on Security Technology, Ottawa, Canada, 2007: 255–263.

[7] Phaal P, Panchen S, and Mckee N. InMon corporations sFlow: A method for monitoring traffic in switched and routed networks. IETF RFC3176, September, 2001.

[8] Estan C, Keys K, Moore D, and Varghese G. Building a better NetFlow. Technical report, 2004. http://www.caida.org/outreach/papers/2004/tr-2004-03/.

[9] Zhang Jian and Moore A W. Traffic trace artifacts due to monitoring via port mirroring. Proceedings of the Fifth IEEE/IFIP E2EMON, Munich, Germany, 2007: 1–8.

[10] Bloom B H. Space/time trade-offs in hash coding with allowable errors. *Communications of ACM*, 1970, 13(7): 422–426.

薛开平：　男，1980 年生，讲师，研究方向为下一代网络体系结构与网络安全.

洪佩琳：　女，1961 年生，教授，研究方向为下一代网络体系结构与网络安全.

郭　婵：　女，1985 年生，博士生，研究方向为移动网络与网络安全.