使用主动延迟建立攻击连接链的关联*)

李 强 林 雁 刘 琨 鞠九滨
(吉林大学计算机科学与技术学院 长春 130012)

摘 要 网络攻击者通常在攻击最后目标前使用双向交互式连接一系列中间主机(跳板机)掩盖其真实攻击路径。本文提出一个主动干扰跳板机连接中数据包间隔延迟的方法,在攻击者所能干扰跳板机连接数据包延迟间隔允许范围之内,通过分析指定时间窗口内连接数据包的活跃程度,在网络出口检测数据包的到达延迟的周期特征,确定攻击连接链的关联。这个方法可以有效地减少连接关联计算量,提高跳板机检测的有效性。

关键词 反向追踪,连接链,主动延迟

Constructing Correlations in Attack Connection Chains Using Active Delay

LI Qiang LIN Yan LIU Kun JU Jiu-Bin
(School of Computer Science and Technology, Jilin University, Changchun 130012)

Abstract Usually network attackers conceal their real attacking paths by establishing interactive connections along a series of intermediate hosts (stepping stones) before they attack the final target. We propose a method for detecting stepping stones by actively perturbing inter-packet delay of connections. Within the attacker's perturbation range, the method can construct correlations in attack connection chains by analyzing activity degree of correlation window and monitoring period characteristic of packets delay. The method can reduce the complexity of correlation computations and improve the efficiency of detecting stepping stones.

Keywords Traceback, Connection chain, Active delay

1 引言

基于网络的攻击已经成为当前网络信息系统的严重威胁。网络攻击者经常通过各种匿名技术隐藏真实身份,在攻击最后目标前通过一系列中间主机(跳板机)[1]连接。这些连接链攻击技术很容易实现和使用,使得网络攻击源的追踪成为最困难的网络安全问题之一。

例如,登录在主机 A 处的攻击者通过 Telnet 或 SSH 登录到主机 B,进而向主机 C 发起攻击。如果在主机 C 处简单的分析流量只能判断出攻击发自主机 B。只有经过仔细的内容分析,才能得知攻击实际来自主机 A。攻击者从自己的主机上发出的命令可以沿着攻击链中的不同主机不经修改的传递,而使得受害者认为攻击来自链上最后一个主机。如果不能有效而快速地发现攻击源,那么就不能制止有可能继续发生的攻击并惩治攻击者。

追踪者为了寻找攻击源,可以执行一个复杂的反向追踪过程,从链中最后一个主机开始,按照每个主机上的日志逐个回溯。此方法由于攻击者通常销毁踪迹而失效。追踪者也可以采用被动监控的方法,在网络中安装一个跳板机流量监控器,分析网络中每个主机的出入流量,发现其中的关联连接。在使用连接链反向追踪方法确定攻击者用来进行攻击的主机的方法中,其关键问题是解决在中间节点进行输入和输出连接的关联(流匹配)[1,2]。由于使用加密、压缩引起内容改变或系统延迟引起的定时变化,使得匹配操作更加复杂,并且网络入侵在高速网络上进行,所以在中间节点上的关联处理必须是迅速的。基于定时的方法是能够关联加密连接的有效方法之一,但以往基于定时的方法过分依赖于数据包的时间特性,而攻击者却非常容易通过跳板机干扰数据包的时间特性,并且基于被动的定时方法需要收集全部网络流量进行关联匹配,计算量很大。

在文本中,我们提出一个主动干扰跳板机连接中数据包间隔延迟的方法,在攻击者所能干扰跳板机连接数据包延迟间隔允许范围之内,通过分析指定时间窗口内连接数据包的活跃程度,在网络入口处改变部分数据包的到达延迟,使每段时间内的跳板机连接具有单一的周期特征。在网络出口检测此种周期特征,确定攻击连接链的关联。此方法在被动监视网络出口的基础上,采用主动干扰的关联算法,在适应攻击者使用加密连接和干扰连接时间特性的情况下,这个方法可以有效地减少连接关联计算量,提高跳板机检测的有效性。

本文第 2 部分是方法的定义和假设;第 3 部分提出了一个使用主动延迟建立攻击连接链关联的方法;第 4 部分对方法进行测试;第 5 部分分析了相关工作;最后进行了总结。

2 定义与假设

给出一系列计算机主机:$H_1, H_2, \cdots, H_n (n>2)$,当请求依次地将 H_i 连接到 H_{i+1} 上时,称此连接序列$\langle H_1, H_2, \cdots, H_n\rangle$为连接链或链式连接(链路)。连接链的反向追踪问题是:给出某条连接链上的 H_n,要求识别出 $H_{n-1}, H_{n-2}, \cdots, H_1$。在一个连接链上的每个中间主机叫做跳板机。如果一对网络连接都是一个连接链的组成部分,则称这对网络连接为跳板机连接对。

假设每一个主机所在的内部网络都存在一个出入口处,

*)本研究受国家自然科学基金资助(90204014)。

所有流入和流出这个网络的流量必须经过这个节点。我们在这个节点上通过观察和干扰流量，检测属于同一个攻击连接链的出入此网络的连接，完成连接关联过程。

对于一个连接 A，其数据包到达网络出口的时间序列为 $T=\{t_1,t_2,\cdots,t_{n-1},t_n\}$，其数据包间延迟为 $D=\{d_1,d_2,\cdots,d_{n-1},d_n\}$，其中 $d_i=t_{i+1}-t_i$。考察一个连接中相邻的 w 个数据包，定义长度为 w 的窗口，按到达序读取该连接上个数为窗口长度的子序列，称为窗口序列，记为 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$。

若 $d_i=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，则标记最大包间延迟出现位置 $M=i$，且任意时刻有 $M\in(1..w)$。如果有概率 $P(M)$服从均匀分布，那么称该时段连接无周期特征。

如果我们在某连接窗口 W 中的 j 位置主动周期性地加入延迟，使得对任意窗口序列 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，有 $d_{k+j}=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，那么说明最大数据包间延迟出现位置 M 的概率 $P(M)$将近似地服从某一正态分布，其期望 $\theta(M)=j$，称该连接有周期特征，周期信号为 M。

假定攻击连接中的数据包在经过跳板机之后仍保持原有顺序，没有丢包或乱序发生，只考虑攻击者不改变数据包数目的情况。每个连接中指定时间内到达数据包的快慢称为窗口内平均速度。对于任意两个连接 i,j，其窗口内平均速度为 $\overline{V_i},\overline{V_j}$。令活跃相关系数 $\sigma=\left(\frac{\overline{V_i}}{\overline{V_j}}-1\right)^2$。当时 $\sigma=0$，称这两个连接具有同等的活跃程度；当 $\sigma\in(0,\gamma]$时，称这两个连接具有相近的活跃程度(其中 γ 为一常数)。当 $\sigma\in(0,+\infty)$时，称这两个连接具有相异的活跃程度。

3 使用主动延迟建立攻击连接链关联

3.1 主动延迟的加入

通过文[3]的分析，攻击者通过对一个连接扰乱定时或插入多余数据包的方法所造成的影响具有一个理论上限。我们在这里范围之内，通过分析指定时间窗口内连接数据包的活跃程度，在网络入口处改变部分数据包的到达延迟，使每段时间内的跳板机连接具有单一的周期特征。

对于某一连接，依次读取其窗口序列得 $W=\{W_1,W_2,\cdots,W_k,\cdots\}$。对任一窗口序列 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，若 $\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}+\tau<T_{out}$(其中 τ 为常数，T_{out} 为最大延迟时间)，则在 w 位置主动加入延迟，使 $d_{k+w}=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}+\tau$；若 $\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}+\tau\geqslant T_{out}$ 则 $d_{k+w}=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$。

可以使得加入的延迟更合理。对于某一连接，依次读取其窗口序列 $W=\{W_1,W_2,\cdots,W_k,\cdots\}$。对任一窗口序列 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，主动注入延迟使 $d_{k+w}=\frac{\Gamma}{w-1}\sum_{i=1}^{w-1}d_{k+1}$($\Gamma\in[1..3]$(为常数系数)。

3.2 主动延迟的检测

在入口加入主动延迟后，可在网络出口检测此种周期特征，确定攻击连接链的关联。对于某一连接，依次读取其窗口序列 $W=\{W_1,W_2,\cdots,W_k,\cdots\}$。对任一窗口序列 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，及人为加入最大延迟值的统计数组 $S=\{s_1,s_2,\cdots,s_{w-1},s_w\}$，初始值 $s_i=0,i\in\{1,\cdots,w\}$，检测其周期信号 M 出现位置。

令 $s_i=\alpha s_i+\lambda$，当且仅当 $d_{k+i}\geqslant\mathrm{MAX}\{d_{k+i-(w-1)},d_{k+i-(w-2)},\cdots,d_{k+i-2},d_{k+i-1}\}$；否则 $s_i=\alpha s_i$。其中 α 为衰退常数系数，λ 为增长常数系数，且 $\alpha\in(0..1)$。

在一定时间 T_{period} 内，依次计算统计数组 $S=\{s_1,s_2,\cdots,s_{w-1},s_w\}$，若对于统计数组 $S=\{s_1,s_2,\cdots,s_{w-1},s_w\}$，有 $\frac{\mathrm{MAX}\{s_1,s_2,\cdots,s_{w-1},s_w\}}{\sum_{i=1}^{w}s_i}\geqslant\kappa$，$\kappa\in(0..1)$为周期识别常数系数，则认为该连接在 T_{period} 内为有周期连接，或在 T_{period} 内存在周期信号。

对于加入的另外一种延迟，令 $s_i=\alpha s_i+\lambda$，当且仅当 $\left|d_{k+i}-\frac{\Gamma}{w-1}\sum_{j=1}^{w-1}d_{k+i-j}\right|\leqslant\sigma'$，$\sigma'$为一足够小的常数；否则 $s_i=\alpha s_i$。(其余同上)

3.3 分析

通过以上方法可以看出，能否建立关联的关键在于选取合适的相异活跃程度常数系数 r，排除相异的干扰连接。通过实验合理选取衰退常数系数 α，增长常数系数 λ，周期识别常数系数 $\kappa\in(0..1)$识别周期信号，选取适当的辅助常数系数 $\Gamma\in[1..3]$，σ'和 τ 确保工作性能。

假设已加入延迟周期信号的输入窗口序列为 $W_k=\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，经攻击者改变包间延迟后的输出窗口序列为 $W'_k=\{d'_{k+1},d'_{k+2},\cdots,d'_{k+w-1},d'_{k+w}\}$，并且攻击者随机地改变包间延迟。改变过程描述为 $d'_i=\mu d_i$。假设 μ 在区间(0.5,1.5)上服从均匀分布，并且取衰退常数系数 $\alpha=\frac{n}{n+1}$，增长常数系数 $\lambda=\frac{1}{n+1}$，其中 n 为当前接收到的该连接的数据包的总数，初始时为 0。则当 d_{k+i} 满足周期信号条件时，$s'_i=\frac{n}{n+1}s_i+\frac{1}{n+1}$；当 d_{k+i} 不满足周期信号条件时，$s'_i=\frac{n}{n+1}s_i$。且有 $\sum_{i=1}^{w}s_i=1$。

如果没有主动延迟加入，称之为无周期连接。考虑周期信号检测中，由于周期信号出现位置 M 满足平均分布，则将有 $E(s_i)=\frac{1}{w}$，简化时令 $w=10$，有期望 $E(s_i)=0.1$。

如果有主动延迟加入，称之为周期连接。考虑最大值的加入方式，$d_{k+m}=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}$，周期信号加入位置为 w。$d_{k+w}=\mathrm{MAX}\{d_{k+1},d_{k+2},\cdots,d_{k+w-1},d_{k+w}\}+\tau$。简化令 $\tau=0$。则在输出序列中检测周期信号时，d_{k+m}和 d_{k+w}被作为周期信号的可能性均约等于 0.5(粗略估计)。又 d_{k+m}出现的位置是均匀分布的，则 d_{k+m}作为周期信号的可能性为 $0.5+\frac{0.5}{w}$，其余位置的可能性为 $\frac{0.5}{w}$。进一步简化，令 $w=10$，则有 $P(M=w)=0.55$，$P(M=i)=0.05(i\neq w)$。且易知 $E(s_w)=0.55$，$E(s_i)=0.05(i\neq w)$。

4 实验

我们使用 NLANR[4] 中 Bell 实验室的追踪数据，从 4900 万个数据包头部中抽取连接数据，选取其中 121 个较活跃的 SSH 流。每个 SSH 流至少包含 600 个数据包和至少 120 秒长的连接。用这 121 个 SSH 连接数据测试主动延迟的方法。

首先选取两个 SSH 连接，选择窗口 $w=10$，$\kappa=0.5$，在第一个连接中不加入周期信号，在第二个连接加入周期信号。检测所得两个连接的 K 值如图 1 所示，其中横坐标为周期个数，纵坐标为 K 值。我们发现主动加入周期的连接，其周期

特性明显；而没有加入周期信号的连接，其 K 值趋于平均分布约为 0.1。

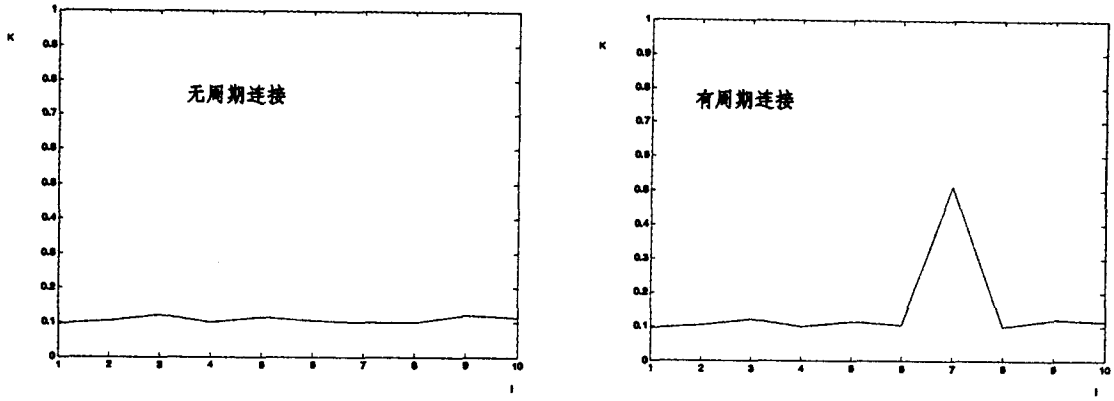
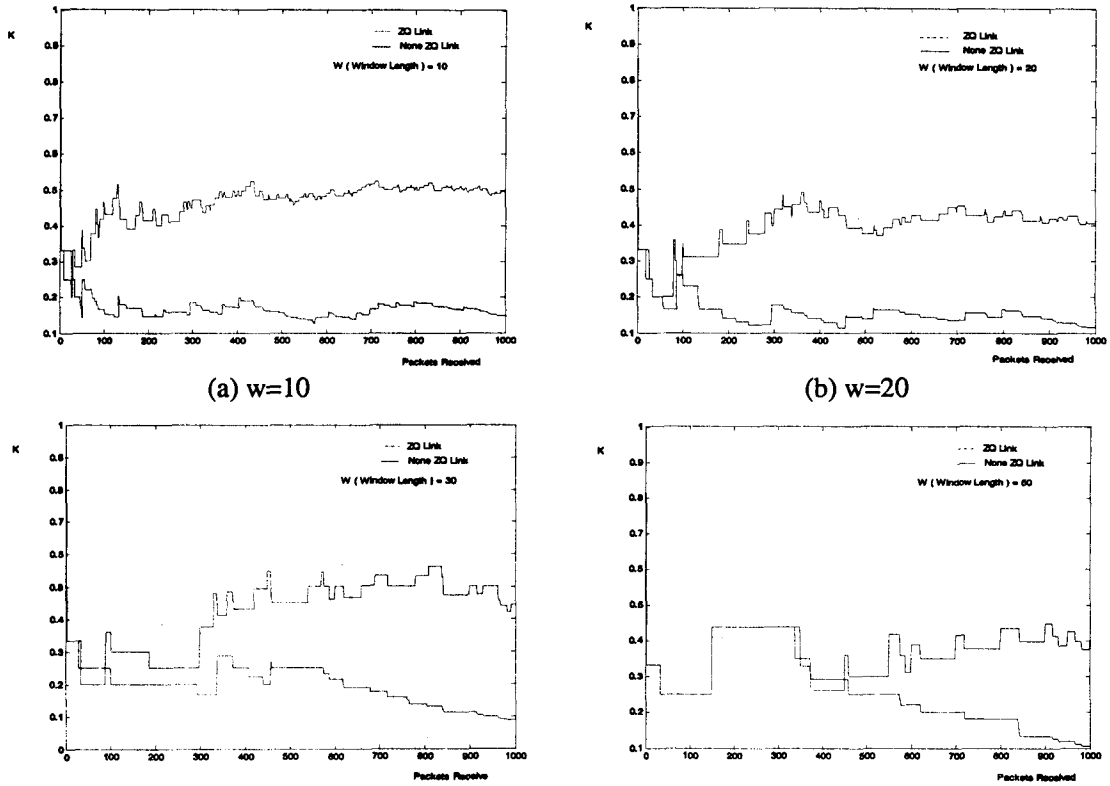


图 1　无周期连接 K 值图(左)和有周期连接 κ 值图(右)

接下来，对这 121 个 SSH 连接测试不同 w 对检测过程中 κ 值变化的影响，并确定最优 w 的值。改变窗口大小，观察 κ 值变化的曲线图。如图 2 所示，其中横坐标为收到的数据包个数，纵坐标为 κ 值，表示 w 不同值时有周期和无周期连接的 κ 值。我们发现选择窗口 w=10，能够较快地区分出周期连接和无周期连接，可以用来区分主动延迟。



(a) w=10　(b) w=20

(c) w=30　(d) w=50

图 2　窗口长度 w 对 κ 值变化的影响图

最后观察不同 κ 值对正确率和误报率的影响，并确定最优 κ 的值。图 3 和图 4 是选取不同的 κ 值，分别为 κ_1=0.3，κ_2=0.4，κ_3=0.5，对 121 个 SSH 连接关联匹配计算。发现 κ_1=0.3 时，其正确率最高，误报率最低。而 κ_3=0.5 其正确率最低，误报率最高。但是当接收的数据包数目达到 50 以上的时候，正确率和误报率都趋近于 0。

5　相关工作

现有连接链反向追踪工作按追踪对象可以主要分为两类[5]：基于主机的连接链反向追踪[1,6,7]和基于网络内容监控的连接链反向追踪[2,5,8~10]。这两类方法在追踪过程中在追踪范围、性能开销、追踪准确性等方面各有特点。

1、基于主机的连接链反向追踪，如美国加州大学 Davis 分校的分布式入侵检测系统 DIDS[5]，每个 DIDS 域内的监视主机收集审计形迹和发送审计摘要到一个中心 DIDS 进行指导。CIS(Caller Identification System)[6]中每个登录链使用链式记录所有信息，只有当所有登录查询主机信息匹配，才能获取下一个主机登录权利。Caller ID[5]按照入侵者入侵的路径反向入侵回去。Yung[7]等利用主机日志，分析请求和响应的时间差来检测交互式终端会话的长连接链。基于主机的方法都采用主机作为信息采集点，受到主机处理能力的制约。由于在主机间过多的进行认证和通信交互导致处理时间过长，不能达到实时性要求。

2、基于网络内容监控的连接链反向追踪，开创性的工作

是 Staniford-Chen 等的拇指纹方法[8]，使用少量信息表示连接并能唯一地确定关联连接。这个方法将流分割为离散的时间间隔并创建每个间隔的包摘要。通过比较两个流的摘要计算相似程度。Zhang 和 Paxson 使用定时方法而不是基于内容检测跳板机[2]。它在网关上配置网络嗅探器，记录所有网络连接的建立时间和拆除时间，根据主机输入和输出连接的结束时刻进行关联。Yoda 和 Etoh 使用偏差法[9]同时考虑定时特性和 TCP 顺序号，通过比较输入和输出连接的序列号偏差，判定两个连接是否关联。Wang 等[10]利用连接链中各个数据包间延迟近似的方法对加密连接链进行关联，只需要利用滑动窗口内的数据包，而不是全部连接信息。基于网络的方法的主要问题是必须在网络内部署若干能记录全部流量的采集点，能将尽可能长时间内的连接信息汇总分析。记录主机的全部并发的输入和输出连接(即使不需要追踪时)，对所有输入连接和输出连接进行关联计算。由于与攻击无关的流量浪费了大部分计算时间，分析连接需要较长时间的采集和计算。

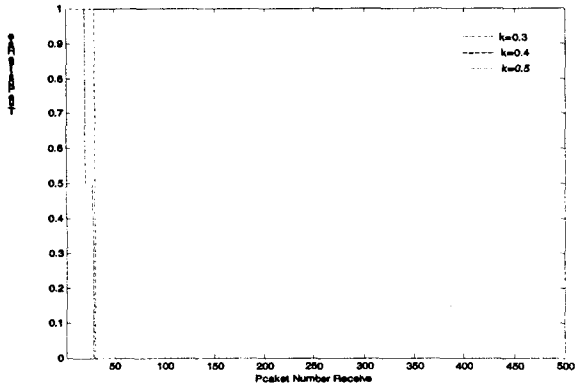


图 3　正确率实验

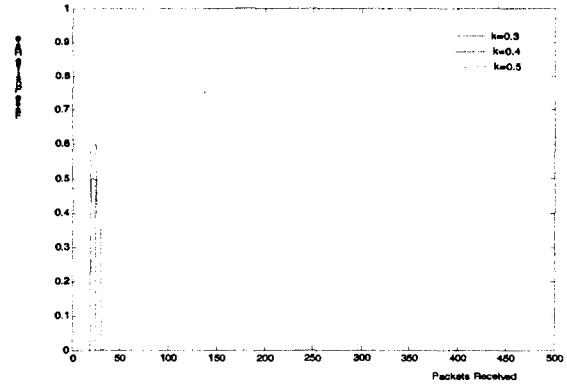


图 4　误报率实验

和以上被动方法不同，主动方法可以积极干扰追踪流量或连接，借此分析连接关联，减少追踪时间和开销。AN-IDR 项目[11]提出各个主动节点在连接内容后面依次添加连接护卫来追踪入侵者，它需要路径上的主机全部部署为主动节点并且追踪速度缓慢。Wang 等使用主动休眠水印追踪 SWT 关联技术[5]，它通过目标主机向网络中入侵链注入一个水印，唤醒自身追踪功能并和中间路由器配合，但它需要在数据包内容里加入水印，不能在加密连接上使用，并且需要整个网络的配合。

我们的主动方法不需要修改数据包负载，并且可以应用到加密连接，不需要跳板机进行配合。FootFall 项目[12]采用在连接延迟中加入主动水印的方法，通过水印的编码和提取检测可以有效地完成跳板机的检测。和我们方法的弱点一样，为防止攻击者有针对的干扰，都存在水印参数或延迟参数不能被攻击者所了解的弱点。但我们的方法主动干扰输入连接部分处理过程简单，需要观测的窗口可以根据实时性要求适当缩放，只需要改变观测窗口中的一个数据包延迟，数据包处理过程较少。

总结　在文本中，我们提出一个主动干扰跳板机连接中数据包间隔延迟的方法，在攻击者所能干扰跳板机连接数据包延迟间隔允许范围之内，通过分析指定时间窗口内连接数据包的活跃程度，在网络入口处改变部分数据包的到达延迟，使每段时间内的跳板机连接具有单一的周期特征。在网络出口检测此种周期特征，确定攻击连接链的关联。此方法在被动监视网络出口的基础上，采用主动干扰的关联算法，在适应攻击者使用加密连接和干扰连接时间特性的情况下，可以有效地减少连接关联计算量，提高跳板机检测的有效性。进一步的工作是把通过 IP 数据包检测跳板机和 IP 反向追踪[13]结合为一个全程的反向追踪模型，通过全程模型中各模块的协作完成对一个整体攻击的反向追踪过程。

参 考 文 献

1　Lee S C, Shields C. Tracing the Source of Network Attack: A Technical, Legal and Societal Problem. In: Proc. of the 2001 IEEE Workshop on Information Assurance and Security, June 2001

2　Zhang Y, Paxson V. Detecting Stepping Stones. In: Proc. of 9th USENIX Security Symposium, Aug. 2000

3　Donoho D, Flesia A G, Shanka U, et al. Multiscale Stepping Stone Detection: Detecting Pairs of Jittered Interactive Streams by Exploiting Maximum Tolerable Delay. In: Proc. of the 5th International Symposium on Recent Advances in Intrusion Detection (RAID 2002), Springer Verlag Lecture Notes in Computer Science, #2516. Oct. 2002

4　NLANR Trace Archive. http://pma.nlanr.net/Traces/long/

5　Wang X, Reeves D, Wu S F, Yuill J. Sleepy Watermark Tracing: An Active Network-Based Intrusion Response Framework. In: Proc. of IFIP Conf., on Security, Mar. 2001

6　Jung H, et al. Caller Identification System in the Internet Environment. In: Proc. of 4th USENIX Security Symposium, 1993

7　Yung K H. Detecting Long Connection Chains of Interactive Terminal Sessions. In: Proc. of the RAID 2002 Conf. Oct. 2002

8　Staniford-Chen S, Heberlein L T. Holding Intruders Accountable on the Internet. In: Proc. of IEEE Symposium on Security and Privacy, 1995

9　Yoda K, Etoh H. Finding a Connection Chain for Tracing Intruders. In: F. Guppens, Y. Deswarte, D. Gollamann, M. Waidner, eds. 6th European Symposisum on Research in Computer Security - ESORICS 2000 LNCS -1985, Toulouse, France, Oct. 2000

10　Wang X, Reeves D, Wu S F. Inter-Packet Delay Based Correlation for Tracing Encrypted Connections Through Stepping Stones. In: Proc. of European Symposium on Research in Computer Security (ESORICS 2002)

11　Active Network Intrusion Detection and Response project. http://www.pgp.com/research/nailabs/adaptive-network/active-networks.asp, 2001

12　Wang X, Reeves D S. Robust Correlation of Encrypted Attack Traffic Through Stepping Stones by Manipulation of Interpacket Delays. In: Proc. of ACM Conf. on Computer and Communications Security (CCS 2003), Oct. 2003

13　Savage S, Wetherall D, Karlin A, Anderson T. Network Support for IP Traceback. ACM/IEEE Transactions on Networking, June 2001, 9(3): 226～237