实时IP反向追踪的研究方法*)

李 强 朱弘恣 鞠九滨
(吉林大学计算机科学与技术学院 长春130012)

摘 要 实时性对于在DoS或DDoS网络攻击中发送假源地址包的主机进行IP反向追踪非常重要。实时的IP反向追踪可在洪水源头处阻止攻击,是建立对网络攻击的网络范围的有效、快速、自动响应的基础。本文分析了IP反向追踪的模型和分类,在比较当前关于提高IP反向追踪实时性研究的基础上,针对其计算复杂性、路由器开销、误报率等,提出实时IP反向追踪需解决的关键问题。为了说明IP反向追踪实时性的重要性和可行性,建立了一个随机包标记算法测试环境。

关键词 攻击源,IP反向追踪,实时性,DDoS

Research Methods of Real-Time IP Traceback

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

Abstract Real-time is very important to IP tracebacking hosts that send spoofed source address packets in DoS or DDoS attacks. The real-time IP traceback can be used to prevent attack near attacking sources and is the netwok-wide fundation of effective, rapid and automatic responding to network attacks. This paper analyzes the model and classification of IP traceback and presents several key problems concerning computing complexity, router overhead and false positive rate, based on the current research on improving real-time in IP traceback. To illustrate importance and feasibility of real-time IP traceback, we establish a PPM algorithm test environment.

Keywords Source of attack, IP traceback, Real-time, DDoS

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阻碍,特别是近两年来的DoS及DDoS[1,2]攻击引起网络安全研究者极大关注。现存的网络安全机制如IDS、防火墙和IPSEC以及容忍攻击技术都还没有考虑到网络攻击源的检测和追踪问题,即使检测到攻击也缺乏网络范围的自动响应。建立对网络攻击的有效响应的一个主要问题是缺少对攻击源的辨认。如果没有有效的源追踪,受到攻击的主机在抵御网络攻击时是盲目的,不可能采取有效的对策如封锁和牵制等,网络攻击直到发现攻击源为止不可能被有效地击退。网络攻击源反向追踪成为入侵检测和响应系统的必要组成部分。对攻击源的实时反向追踪是建立对网络攻击的网络范围的有效、快速、自动响应的基础,对于及时处理大规模网络的DDoS攻击和网络总体的抗打击能力具有重要意义。

由于各种网络攻击使用不同的匿名技术,全面解决网络攻击追踪问题是很复杂的,例如DDoS攻击通常是从多个以前被入侵过的从机上发出、在一个远程主机控制下产生的。发自从机的单向洪泛信息流通常来自伪造的源IP地址,使得即使追踪从机也很困难。IP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。Lee和Shields给出一个对攻击源的全程反向追踪模型[1],指出两个基本研究问题:识别包的中间源(它们可能通过IP地址欺骗被伪装)和确定一个主机的到达包和发出包的因果关系。前者又叫IP反向追踪问题,后者又叫连接链反向追踪问题。如图1所示,模型中包括攻击者、跳板机、从机、反射器和受害者几个部分。从机就是一个隐藏包地址的主机,在继续攻击之前对攻击包进行彻底的变换或引入延迟。反射器属于未被攻击者控制的主机,它按照自己正常的功能协助了攻击。

IP攻击源反向追踪可以分为实时和非实时两种类型。实时的追踪可以在攻击进行时,以最快的速度找到攻击者;非实时的是在攻击结束后分析数据和状态来判断攻击者。当前流行的网络攻击速度异常迅速,即经常'打完就跑',依靠人工处理只能发现其中的一小部分,给攻击者较多的掩盖罪证时间和逃跑时间。如果被入侵主机能近似实时地确定攻击包的路径,将非常容易地、迅速地制止攻击,即使找到局部路径信息也可以在远端路由器制止该攻击。实时性不仅能使我们在接近其源处阻止网络入侵,还能帮助我们较好地保护主机不被入侵成为从机,从而阻止DDoS攻击。

本文将分析IP反向追踪的模型和分类,在比较现有对IP反向追踪实时性的研究的基础上,针对其计算复杂性、路由器开销等,提出实时IP反向追踪需解决的关键问题。

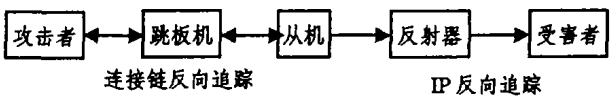

图1 全程反向追踪模型

*)国家自然科学基金项目(No.90204014)。李 强 博士生,研究方向为网络安全。鞠九滨 教授,博士生导师,研究方向为分布式系统与网络软件。

2 IP 反向追踪问题

在受害者发现的攻击包中，一个最简单的标识攻击者方法是提取攻击包源 IP 地址，而大多数攻击者使用一个或多个技术掩盖真正的攻击包源地址。IP 反向追踪是识别使用伪造源地址发送 IP 包(如 DoS 攻击)的主机或网络设备。IP 反向追踪的目标是在不中断正常流量的情况下，在大量洪水包中找到攻击源，必须尽可能独立使用而不过多依赖于其它系统，必须兼容现有的设备、减小路由器的改变，只引入一小部分网络复杂性，要求路由器处理速度快、代价和部署时间最少。

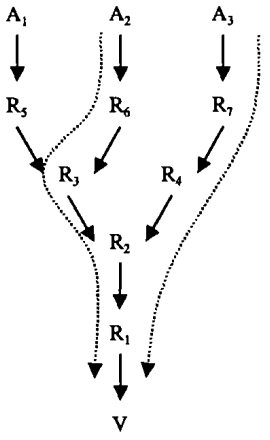

图2 IP 反向追踪示意图

图2表示从受害者角度观察到的网络拓扑，其中 V 表示一个受害者主机或网络设备，每个可能的攻击者 A_i 视为以 V 为根的树的叶子节点。路由器 Ri 位于攻击者 A_i 和受害者 V 之间路径上。假定从任一个攻击者 A_i 到 V 的路径是唯一的。按发现攻击者的精确程度分为准确反向追踪(exact traceback)和近似反向追踪(approximate traceback)。前者是确定攻击路径和每一个关联的攻击者，由于 IP 地址欺骗和攻击者发送混淆数据而异常复杂；后者是发现每个攻击者的包括真正攻击路径部分的候选路径。如果一个没有参加攻击的路由器或主机出现在受害者恢复出来的攻击路径中称为误报(false positive)，则参加攻击的路由器或主机没有出现在受害者恢复出来的攻击路径中称为漏报(false negative)。

当前的 IP 反向追踪主要有以下方法正在使用：

a、在包中标记路由信息。Savage 等[3]使用随机包标记(PPM)的方法，即随机地在 IP 包头部中加入路由和距离信息的有效编码。PPM 方法允许一个受害者在无需和 ISP 进行交互式操作鉴别攻击流量的网络路径，而且追踪还可以在攻击进行完之后进行。相对于 PPM 法在 IP 头部存储路径信息，Doeppner 等[4]使用 IP 可选项来随机存放追踪信息，分确定性标记和随机性标记两种方式。

b、Bellovin[5]从路由器向包的目的地随机地以一个较低的概率对每个它所处理的包发送一个 ICMP 追踪消息包到目的主机。在目的主机中，其内容和源地址可以帮助用于恢复一个从源到目的的路由，确认伪造的 IP 包的真实路径。

c、入口(出口)过滤[6]阻止基于网络的拒绝服务攻击。边界网关的边缘过滤模块用于限制 IP 源地址欺骗、路由器分辨和锁定非法地址包，它通常是边缘网络或 ISP 使用的主要方法之一。此外还有使用主动网络技术防止地址欺骗进行主动过滤研究的有 MIT 的 Van[7]、UCLA 的地址过滤协议[8]、Park 等提出的基于路由的分布式包过滤 DPF[9]技术。

d、连接测试。现有的大多数反向追踪技术开始于离受害者最近的路由器，并交互式地测试上游连接直到发现流量产生者。主要有两类：第一类是 Ferguson 和 Senie 等提出的输入调试[10]，根据输出端口判断输入端口和流量，进而确定上游路由器；第二类是 Burch 和 Cheswick 提出的有控制的洪水[11]，通过有选择的突发有记号的网络负载，观察对攻击者流量造成的干扰情况决定攻击者的源路径。

e、路由器记录流量。Sager 和 Stone 提出在关键路由器上记录数据包[12,13]并使用数据挖掘技术在攻击后决定攻击路径，它需要大量处理资源和数据存储。BBN 的 SPIE[14,15]使用路由器硬件生成流量审计摘要，能够在一段时间内找到一个单独数据包的信息。

f、主动查询路由设备的流量处理情况，如美国 NAI 实验室的 IDIP 和 CITRA[16~18]。需要知道真正包源地址的目的主机发送一个查询包给它的路由设备，这个路由设备转发查询到它所连接的任意路由设备上，根据反馈响应构造数据包路径。

此外还有其它用于阻塞控制和追踪流量的方法，如 ACIRI 使用基于流量聚合的拥塞控制和回推[19]用于攻击源认定，认为分布式拒绝服务攻击是由预先定义好的流量组合，而不是单个流。它主要通过两个途径来实现，一个是在本地路由器上通过 ACC 代理探测和控制高带宽总计，另一个是路由器请求上游协作回推机制。DECIDUOUS[20]系统利用 IPSec 协议的验证头进行确定性隧道标记。MIT 的 Thomer 等[21]在每个网络设备上建立一个 MULTOPS 数据结构使用在线包统计监控流量。

3 影响实时性的因素

以上各种方法为了能够实现实时 IP 反向追踪，都分别采取了不同的途径，也存在着相对应的问题，这些问题也是解决反向追踪问题的关键点，下面将按照不同的影响因素进行分析。

3.1 IP 反向追踪的准备阶段

上述方法中为了实现反向追踪，需要在上游路由器或主机上进行修改数据包或者调试端口的操作，这些操作是为了最终获取 IP 攻击数据包源头的准备阶段工作。准备阶段中路由器操作算法、策略和通信方式直接决定了追踪的实时性完成程度的大小。例如，PPM 需要在攻击者和受害者之间的路由器上小概率地将路由器信息标记在 IP 数据包头部的指定字段上；ICMP 法也需要在路由器或主机上以小概率生成数据包的路径信息发送给受害者；入口/出口过滤法需要路由器具有极高的处理能力来分析每个包，在追踪执行前掌握合法流量和非法流量的地址知识；单包追踪的 SPIE 需要路由器平时记录流经数据包的摘要信息。

3.2 IP 反向追踪的执行阶段

执行阶段是体现实时性的标志。IP 反向追踪中用于自动恢复攻击路径的算法是追踪反应速度的关键，最终计算攻击路径的时间决定实时性的优劣。现有的所有反向追踪算法，只有极少数支持实时。PPM 法中根据受害者接收到的标记包进行解码和路径推测的计算，虽然后继工作改进的随机包标记[22]和代数法包标记[23]努力提高了标记方法，使用快速的路径推测算法，减少用于路径推测计算的边标记组合数目来满足实时需要，但都以各种假设和条件为基础，制约了算法的实际实现范围。基于 IP 散列的 SPIE 方法的推测路径算法比

PPM 简单，但它受到查询处理和通信可靠性的制约。连接测试法由于不能全部自动响应，也不能在短时间内完成追踪。Mankin 等[24]分析 ICMP 法的特点和缺点，说明对于 DDoS 中每个攻击源只发送少量攻击包或以固定概率生成 ICMP 追踪包并不容易发现攻击路径，并且容易受到多层 DDoS 的反向人为流量背景干扰造成误报，提出根据路由器的路由表和包转发表增填意向位，在感兴趣的路由器间进行追踪；但它也受到概率标记的制约而必须有足够时间接收到全部需要的数据包，且 ICMP 包经常被一些路由器过滤掉，而追踪包的产生依赖于路由器的入口过滤能力。连接测试法开始于离受害者最近的路由器，并交互式地测试上游连接直到发现流量产生者。这种方式需要攻击必须持续一段时间，主要有 Ferguson 和 Senie 等提出的输入调试、Burch 和 Cheswick 提出的有控制的洪水法，它们主要依靠具有过滤、输入调试功能的路由器在网络内的普及程度，大部分工作依赖于管理员手工操作，必须在攻击时才能响应，所以不具备实时性条件。

3.3 IP 反向追踪的环境

现有的反向追踪技术根据需求和使用技术不同，依赖于各种系统环境和网络环境。如入口过滤和连接测试可以在传统路由器上使用路由器功能模块执行，ICMP 法、PPM 法需要新型路由器的支持，SPIE 法等需要专用路由器存储大量数据，基于代理的 IDA 系统等需要移动代理的执行环境，基于水印和主动查询法需要主动路由器节点的执行环境。NAI 实验室的 IDIP 和 CITRA 使用主动网络技术[25]实现流量控制。这种方案会产生额外的网络通信量，需要记录一个有效时间段内经过路由设备的所有流量信息，然后采取级连查询的方式推测路径，这就要求追踪必须是实时进行的。但实时性受到路由设备存储信息能力的限制，查询时间也受到设备部署情况和网络规模的影响。反向追踪系统在网络内的部署情况和运行的系统环境将严重影响反向追踪实时性能。

3.4 精确性和准确性的要求

反向追踪算法的准确性就是对攻击结果的误报率和漏报率。造成误报和漏报的原因有多个：如攻击者尽量混入干扰正常计算的虚假流量和处理延迟；网络传输的非对称、延迟及阻塞丢包；推测路径算法的健壮性等。现有的各种方法都尽力提高算法的准确度，如 PPM 法在 IP 头部存储固定长度的信息防止单独传输标记包的丢失，包记录法在关键路由器上全部存储数据包来最大可能地满足查询，在边缘路由器上充分使用过滤模块来识别攻击流量。入口过滤在各种 IP 追踪技术中由于直接对路由器进行端口过滤，因此具有最好的精确度。PPM 法确定攻击路径于最远收到标记包的路由器，所以精确度不高。用 ICMP 产生追踪包的方法由于追踪包在传输时的不确定性，也具有精确度不高的问题。影响精确度的原因可能是网络地址的划分和连接方式，恢复计算的复杂性太高，必须简化计算。

追踪的实时性要求是影响准确度和精确度的重要因素。有时在短时间内得到一个大致结果比在长时间内得到一个准确结果对追踪更有帮助，所以有时只要求计算出攻击者所处的子网或大致区域而不是具体地址。片面地追踪准确度和精确度的计算也增加了计算复杂性而需要过多的处理开销。

3.5 开销

在 IP 反向追踪中不可避免地增加了路由器的处理开销。PPM 法需要按照小概率随机处理 IP 头部的地址边采样和距离信息。ICMP 法需要足够的地址知识，在路由器端口随机采样数据包生成相对应的追踪包。入口过滤一般依赖于路由器的过滤模块对每一个数据包进行处理。连接测试更加依赖于路由器的调试功能。包记录法需要路由器能够记录一段时间内的所有包信息。如果路由器也负责连接链连接匹配的话，那么需要它具有更高的处理功能和速度。实时处理需要路由器额外的处理能力，而攻击后处理一般需要有路由器额外的存储能力。

攻击包不完整、网络阻塞引起的包丢失、动态路由、攻击者不同程度的欺骗以及主机和路由器的信任程度等原因对反向追踪的结果产生了不同的影响。攻击者将尽可能减少其攻击包在源处被捕获的概率（通过其它方向的洪水达到混淆目的）。如 PPM 法必须在头部保存连接信息，存在标记域欺骗问题。Song 等改进时给出攻击路径验证，包括利用网络拓扑顺序验证路由器顺序和用基于时间变化的密钥加密标记。这些路由器处理开销和认证开销严重影响了反向追踪实时处理。但一定的处理开销有时是不可避免的，所以在开销和实时处理之间存在一个折衷。

3.6 多攻击者和多攻击路径

现有的攻击都是多攻击者处于不同地理位置和网络范围，在一段时间内向同一个或一组网络主机发起的进攻。由于 Internet 路由的特性，攻击包并不是按照同一路径传输，不同的攻击源洪水和传输路径交织在一起，因此能同时反向追踪多个攻击者和一个攻击源的多个攻击路径成为反向追踪系统必备的功能之一。在现有方法中，改进的各 PPM 法、ICMP 法可以在受害者处利用不同长度的计算时间在恢复单个攻击路径的基础上恢复多个路径。现有方法对于一个攻击源的多传输路径问题都没有解决的方法。多攻击源的路径推测计算复杂性与单攻击源恢复计算的比例关系将严重影响反向追踪实时性。

4 IP 反向追踪测试环境

Song、Dean、Tao Peng[26]、Adler[27]等都为减少 IP 头部存储标记位数、提高执行速度、降低误报率和提高精确度等方面作了努力。表1是几种 PPM 法在标记概率为0.04、攻击者距离受害者距离为25个路由器的一个初步比较。

表1 PPM 方法特性比较

方法	空间	单攻击受害者推测路径需收到的包	复杂性	误报率	可追踪攻击源数量
Savage	16bit	3000	$O(K^8)$	高	15
Song I	16bit	375	$O(K)$	低	50
Song II	16bit	2000-3000	$O(K)$	低	500-1500
Dean	13bit	-	$O(K^4)$	低	—
TaoPeng	16bit	最少1100	$O(K)$	—	—

在 IP 可选项中加入随机标记会引起路由器的额外开销，包转发处理时间变长。而在 IP 头部存储随机标记由于位数有限，受害者需要较多的时间接收足够数量的 IP 包。虽然算法有许多改进，但算法本身运算时间和受攻击主机数目增多而增加的计算时间也不容乐观。Park 和 Lee 分析了这种方法的有效性[28]，认为包标记算法的概率选择与攻击者的进攻是一个互相影响的约束最小优化问题，建议在标记概率、推测路径长度和接收到的包之间进行权衡，可应用到任何随机标记方案。为了防止欺骗，路由器间的验证也增加了额外的处理时间，可标记包的路由器部署规模、标记概率、包阻塞丢失和要

求误报率的高低也影响着实时处理的快慢。但总的来说，PPM 法基本上能近似实时地恢复攻击者的路径。

基于此，我们在反向追踪以往工作的基础上初步建立了一个随机包标记算法测试环境。此环境也可以作为其它 IP 反向追踪算法的测试和验证平台，对算法的执行效果和性能进行评价。该环境主要由攻击主机、模拟路由器、检测模块、受害者主机、监控和测试主机等组成，也是一个攻击模拟、追踪检测的验证环境。其中，受害者主机作为树型网络的根，攻击主机为叶节点，模拟路由器为中间路径节点。

在后续工作中可以通过以下参数对 IP 反向追踪方法进行测试：1. 一个全路径推测需要的包数量；2. 推测路径的计算开销；3. 大规模 DDoS 下的算法的健壮性；4. 部署开销。

结束语 到现在为止，已有的研究都没有全部解决 IP 反向追踪问题，我们在总结以往 IP 反向追踪在实时性方面的工作之后，认为提高反向追踪实时性可以从以下几点深入研究：

1. *设计新的或改进原有的追踪算法* 实时性是保证反向追踪能够成功的重要因素，主要工作是解决追踪算法的实时性。随着各种新奇的攻击模式不断出现，利用系统漏洞、反射器的拒绝服务攻击层出不穷。新的攻击技术可能突破了现有反向追踪算法的限制，针对追踪技术变化出更不易被发现源头的攻击。所以要设计出不易被欺骗和破坏的追踪算法，也可以在原有算法如随机包标记、包记录、连接关联等的基础上通过降低计算复杂性，减少处理开销，增强健壮性等方面来提高实时性。可以看出如果要求反向追踪实时处理，采用的算法必须满足以下几个条件：所使用的算法必须可以在收集到一定攻击流量后计算得出结果；用于路径计算的算法必须满足具有一个计算复杂性上限；算法要求的数据量最小。

2. *充分利用新的网络技术* 移动代理和主动网络技术是近年来出现的新型网络平台技术。移动代理的追踪技术更适合主机系统的参加，而主动网络追踪一般需要网络设备中少许边缘路由器参与，不需要修改核心路由器。在主动网络中协议和追踪算法的实现和修改更容易和灵活。使用新技术可以弥补传统网络诸如源地址欺骗等的缺陷，并且可以在尽可能小的代价下满足反向追踪的技术需求，可以只在需要反向追踪的区域部署，适应网络环境和攻击特征的改变。由于主动网络侧重包转发处理的定制特性，因此主动网络技术适合于对包处理的反向追踪方法，无须更改主机上的应用程序或服务员软件，也不依赖于受害者主机脆弱的反应，在距离受害者最近的主动路由器上实现追踪启动。同时用主动网络技术解决实时性的方法用于现行网络，例如可以在现行网络中适当增加部分可编程路由器或交换机。

3. *多个方法的合成* 虽然当前反向追踪中的不同算法只能解决其中一些问题，但通过合成多个方法，可以更加高效实时地解决问题。如 ICMP 法和随机包标记法配合，在无攻击时不用随机包标记法，ICMP 法负责收集网络拓扑等情况，减小开销；有攻击时通过 ICMP 法或主动网络控制随机包标记的启动，主动网络和随机包标记法结合解决 IP 反向追踪等。

例如，Xu 等[29]让受害者参与过滤，在标记大部分包的同时让一小部分包中携带 IP 追踪信息，在推测攻击路径的同时过滤攻击流量。Yarr 等[30]为主动过滤攻击流量而用路径信息来推测路径。

4. *建立实时全程模型* 各种不同阶段的反向追踪算法是否能够紧密协同、自动连接是达到实时追踪的关键。设计一个满足实时要求的、可用于 IP 反向追踪和连接链反向追踪的统一综合算法，以及建立一个网络攻击源的实时全程反向追踪模型，是反向追踪技术发展的需要。它不仅可以在体系结构上保证追踪的连续性，而且可以融入各种先进追踪技术保证追踪的可靠性。现有的所有反向追踪算法，不管是用于 IP 反向追踪的包标记、ICMP、连接测试、包记录等，还是用于连接链反向追踪的拇指纹、定时、偏差、水印等，都是不完全的'部分解'，即只解决了全过程的局部问题，没有把这两个部分紧密结合起来成为全程的'全解'。只有实现全程的反向追踪算法才能自动实时地对攻击进行反击和遏制。统一的包标记法是现有各种方法中唯一能同时用于 IP 反向追踪和连接链追踪的算法。首先使用包标记算法追踪 DoS/DDoS 攻击源头，然后在洪水包源头中使用包标记算法追踪连接链，发现'跳板机'。它与其它算法比较有很多优点：不需要与 ISP 交互操作，避免了很大的输入调试管理开销；不需要大量额外网络信息流；可用于追踪多重攻击；在路由器上实现开销很小；IP 反向追踪和连接链反向追踪都可使用。

5. *实时性的评价* 虽然现有的各种反向追踪技术都针对某一问题提出解决方案，在现有理论证明基础之上，和入侵检测技术的评价相比，还缺少能针对不同网络环境和攻击模式进行实时性评价的标准、环境和测试工具[31]，这将是以后反向追踪技术研究的一个重要方面。

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 Moore D, Voelker G, Savage S. Inferring Internet Denial-of-Service Activity. In: 10th USENIX Security Symposium, Washington, D. C., USA, Aug. 2001

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

4 Doeppner T, Klein P, Koyfman A. Using router stamping to identify the source of IP packets. In: 7th ACM Conf. on Computer and Communications Security, Athens, Greece, ACM, 2000

5 Bellovin S M. ICMP Traceback Messages. Internet Draft: draft-bellovinitrace-00. txt, Mar. 2000

6 Ferguson P, Senie D. Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing. RFC 2827, May 2000

7 Van V C. A Defense Against Address Spoofing Using Active Networks. Master Thesis of MIT, 1997

8 Li J, Mirkovic J, Wang M, Reiher P, Zhang L. SAVE: source address validity enforcement protocol. INFOCOM 2002

9 Park K, Lee H. On the Effectiveness of Route-Based Packet Filtering for Distributed DoS Attack Prevention in Power-Law Internets. In: Proc. ACM SIGCOMM '01, 2001

10 Systems C. Characterizing and tracing packet floods using Cisco routers. Aug. 1999

11 Burch H, Cheswick B. Tracing Anonymous Packets to Their Approximate Source. Usenix LISA, Dec. 2000

12 Stone R. Centertrack: An IP overlay network for tracking DoS floods. In: Proc. of 9th USENIX Security Symposium, Aug. 2000

13 Sager G. Security fun with OCxmon and cflowd. Nov. 1998, Presentation at the Internet 2 Working Group

14 Snoeren A C, Partridge C, Sanchez L A, et al. Hash-Based IP Traceback. In: Proc. of the ACM SIGCOMM 2001 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communication, Aug. 2001

15 Sanchez L A, Milliken W C, Snoeren A C, et al. Hardware Support for Hashed-Based IP Traceback. In: Proc. of the 2nd DARPA Information Survivability Conf. and Exposition (DISCEXII), June 2001

16 Rowe J. Intrusion Detection and Isolation Protocol: Automated Response to Attacks. Presentation at *RAID'99*, Sep. 1999

17 Schnackenberg D, Djahandari K, Strene D. Infrastructure for Intrusion Detection and Response. In: Proc. of DISCEX, Jan. 2000
18 Schnackenberg D, Djahandari K, Strene D, Holiday H, Smith R. Cooperative Intrusion Traceback and Response Architecture (CITRA). In: Proc. of the 2nd DARPA Information Survivability Conf. and Exposition (DISCEXII), June 2001
19 Mahajan R, et al. Controlling High Bandwidth Aggregates in the Network. AT&T Center for Internet Research at ICSI (ACIRI), DRAFT, Feb. 2001
20 Chang H Y, et al. Deciduous: Decentralized Source Identification for Network-based Intrusions. In: 6th IFIP/IEEE International Symposium on Integrated Network Management, IEEE Communications Society Press, May 1999
21 Gil T M, Universiteit V, Poletto M. MULTOPS: A Data-Structure for Bandwidth Attack Detection. In: 10th USENIX Security Symposium, Washington, D.C., USA, Aug. 2001
22 Song D X, Perrig A. Advanced and Authenticated Marking Schemes for IP Traceback. In: Proc. of InfoCom 2001
23 Dean D, Franklin M, Stubblefield A. An algebraic approach to IP traceback. ACM Trans. Information and System Security, May 2002
24 Massey D, Wu C L, Wu S F, Zhang L. On Design and Evaluation of Intention-Driven ICMP Traceback. In: Proc. of IEEE Intl. Conf. on Computer Communications and Networks, 2001
25 Active Network Intrusion Detection and Response project. http://www.pgp.com/research/nailabs/adaptive-network/active-networks.asp, 2001
26 Peng T, Leckie C, Kotagiri R. Adjusted Probabilistic Packet Marking For IP Traceback: [Technical Report of Department of Electrical and Electronic Engineering]. University of Melbourne, Australia, 2001
27 Adler M. Tradeoffs in Probabilistic Packet Marking for IP Traceback. In: Proc. of 34th ACM Symposium on Theory of Computing (STOC) 2002
28 Park K, Lee H. On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack. In: Proc. of IEEE INFOCOM '01, 2001. 338～347
29 Sung M, Xu J. IP Traceback-based Intelligent Packet Filtering: A Novel Technique for Defending Against Internet DDoS Attacks. In: 10th IEEE Intl. Conf. on Network Protocols(ICNP), Paris, France, Nov. 2002
30 Yaar A, Perrig A, Song D. Pi: A Path Identification Mechanism to Defend against DDoS Attacks. In: Proc. of the IEEE Symposium on Security and Privacy 2003
31 Kuznetsov V, Simkin A, Sandström H. An evaluation of different ip traceback approaches. In: Proc. of the 4th Intl. Conf. on Information and Communications Security. Dec. 2002

(上接第36页)

相应的最小副本数，和前面的实际测试结果比较得到图1。从图1可以看出，当节点在线率为0.6时，期望文件可用性为0.76，AdpReplica 模型返回最小副本数为2。当系统保持2个副本时，实际测试的平均文件可用性为0.69，从而验证了模型的正确性。

为了验证 AdpReplica 的动态性，系统初始保持一份副本，为之设定文件期望可用性为0.6，以一定频率访问文件，在不同时刻测试副本数，图2为副本数随时间的变化图。从图2可以看出，系统根据 AdpReplica 模型动态增加副本，当满足用户期望可用性后，将副本维持在一个合理的水平。

5.2 性能测试

以100个 PC 机器作为节点构建一个 P2P 实验网络。设定可用带宽 $C=10$MB，系统初始保持一份副本，设定文件访问频率阈值为20，表示当文件的访问频率超过20时触发 AdpReplica 模型，逐渐增加文件的访问频率，测试文件的平均访问延迟和平均副本数，得到图3和4。

图3和图4说明，初始阶段，文件只有一个副本，平均访问延迟较大，随着访问热度的提升，副本随之增加，表现在明显降低访问延迟，但副本数达到一定程度后，副本总数和访问延迟维持在一个稳定的水平，副本没有无谓地增加。

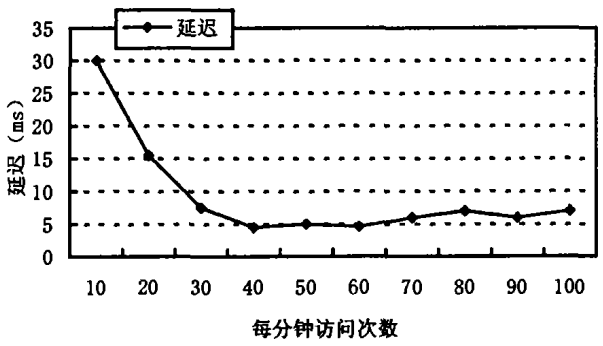

图3 访问延迟随访问频率的变化情况

结束语 本文提出一种高效的自适应副本管理机制—AdpReplica。AdpReplica 将期望可用性和一致性维护开销看成副本数目的函数，并建自适应模型，动态管理系统中的副本数，使文件副本数维护在一个合理的水平，既提高了数据可靠性和访问效率、避免 Hot-Spot 的产生、平衡负载，又减少带宽消耗和系统性能的震荡，保证系统的稳定，并且能适应系统的动态扩展，为用户提供满意的存储 QoS。

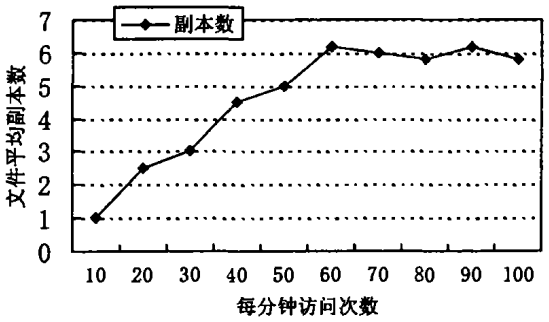

图4 副本数随访问频率的变化情况

参 考 文 献

1 FreeNet. http://freenet.sourceforge.net.
2 Napster. http://www.napster.com.
3 Hanhua. Studies on Internet Oriented Distributed Massive File Storage: [PH.D. Dissertation of Peking University]. June 2002
4 Ranganathan K, Iamnitchi A, Foster I. Improving Data Availability through Dynamic Model-Driven Replication in Large Peer-to-Peer Communities. In: Proc. of the Workshop on Global and Peer-to-Peer Computing on Large Scale Distributed Systems, Berlin, May 2002
5 Acharya S, Zdonik S B. An Efficient Scheme for Dynamic Data Replication. Brown University CS-93-43, 1993
6 Wolfson O, Jajodia S, Huang Y. An adaptive data replication algorithm. ACM Transactions on Database Systems, 1997, 22: 255～314
7 Ranganathan K, Foster I. Identifying Dynamic Replication Strategies for a High Performance Data Grid. In: Intl. Workshop on Grid Computing, Denver, CO, 2001