P2P网络行为安全监控研究综述

蒋卓明[1,3] 周 旭[1] 许榕生[2]
(中国科学院声学研究所高性能网络实验室 北京 100080)[1]
(中国科学院高能物理研究所计算中心 北京 100049)[2] (中国科学院研究生院 北京 100049)[3]

摘 要 P2P网络行为监控是近年来研究和应用的热门课题。在分布式监控系统的基础上,从P2P网络外部行为特征和内在行为特征两个方面对学术界的研究成果进行总结归纳。主要涉及了流量分析、拓扑挖掘、用户认证和信任模型、传播建模等方面的理论,从现象分析解决方案,并指出不足和改进,重点提出了域发现协议的应用和信任模型的原则。此外,还就产业界建设P2P监控的安全平台方面做了简要介绍。

关键词 P2P网络,行为监控,域发现协议
中图法分类号 TP393.08 **文献标识码** A

Review of P2P Network Behavior Security Monitoring Research

JIANG Zhuo-ming[1,3] ZHOU Xu[1] XU Rong-sheng[2]
(High-performance Network Laboratory, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China)[1]
(Computing Center, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China)[2]
(Graduate School, Chinese Academy of Sciences, Beijing 100049, China)[3]

Abstract P2P network behavior monitoring is a hot subject recently. Based on distribution monitoring system, this paper summarized and analysed the academia research about two sides of P2P network external behavior characteristics and internal behavior characteristics. The theory research mainly involves traffic analysis, topology mining, authentication and trust model, propagation model and so on. All the research above was analysed from phenomenon to solution, and pointed the disadvantage or improvement, mainly including domain discovering protocol and trust model principle. In addition, it also introduced that industry circles build the security platform for P2P network monitoring.

Keywords Peer-to-Peer network, Behavior monitoring, Domain discovering protocol

1 引言

诞生于自治系统的P2P网络,面临着管理混乱、安全问题突出的现状,目前对于P2P用户行为安全监控方面的研究有着特定的需求和方法,对其进行综述分析,一方面可以完善传统的网络安全监控模型,另一方面对P2P网络的良性发展起到保障和促进作用。

传统的网络安全监控模型一般包括数据采集、数据存储、协议还原和分析、应用层数据挖掘、内容和行为关联分析、主动安全控制和安全策略管理等模块,在实际中部署的网络安全监控系统还可能与防火墙、IDS、防病毒软件等联动。本文所调研的资料显示,针对P2P网络行为的安全监控是在此基础上演化发展来的。

2 基础分布式监控模型

由于P2P网络的分布式特性,其监控系统模型也应当是分布式系统。文献[1]根据P2P文件共享应用设计的系统可视为一个基础分布式安全监控模型的代表,实现的功能归纳为:

(1)对内部局域网内各个主机上P2P文件共享软件的使用进行监控。当有文件传输行为时,依据具体的安全策略允许或拒绝传输。

(2)安全策略由管理员集中定义,但实施则分散到各个主机监控终端执行。系统防护不仅仅依赖于网络边界单个检查点,消除了网络性能的瓶颈,容易实施针对特定协议的包过滤,实现对特定应用程序的安全策略。

(3)对P2P文件传输行为提供日志和审计功能。系统记录安全日志,并且可以对日志进行查询统计,生成报告供管理员审计。

由以上可以看出,该P2P分布式监控系统继承了传统网络监控的特点,主要的改进在于安全策略的集中定义和分散执行,以实现协同监控。但监控对象主要仍是针对P2P局域网络中的内容传输,没有涉及到P2P骨干网络中更加突出的安全问题。

3 P2P网络外部行为特征分析

3.1 流量分析

到稿日期:2008-07-28 返修日期:2009-02-03 本文受863国家高技术研究发展计划基金项目(2006AA01Z410)资助。
蒋卓明(1981-),男,博士生,主要研究方向为P2P网络安全等,E-mail:zmjiang@ustc.edu;周 旭(1977-),男,助理研究员,博士,主要研究方向为分布式网络存储等、P2P网络;许榕生(1947-),男,研究员,博士生导师,主要研究方向为网络安全、反黑客技术等。

P2P 网络中最为显著直观的特征就是其巨大的流量，这也是带来安全隐患的一个重要因素。在理论研究中经常用到信息论来分析流量，比如利用一些分布变化过程来衡量流量的某个特征的信息量。在传统互联网上的经典流量模型包括泊松过程、马尔科夫过程等。例如 Web 上的 Http/Mail/Ftp 等流量满足以上过程，它的突发性将随着采样时间间隔的增加而逐步平缓。

然而一些实际 P2P 应用的流量均显示，在加大采样时间间隔后，流量的突发性并未发生明显的变化，而且不同的时间间隔内的流量具有相类似的特征，因此这些实际的流量特性在统计学上被称为自相似性[2]。自相似性中最重要的方面之一就是流量的突发性没有一个特征尺度，结果导致多个流量源的叠加也并不在某个时间度量尺度上面出现平滑的迹象，而满足泊松到达模型的流量在逐渐加大时间度量后，呈现平滑的趋势。

因此，流量的突发性在一个较广的时间度量范围内不发生变化是 P2P 网络自相似过程的重要特征。为了利用该特性对 P2P 应用的流量进行建模评估，可以对一定时间段内的 P2P 网络上下行流量以及持续时间进行统计，违反相应自相似性的 P2P 网络行为很有可能是异常行为。文献[3]中具体应用了网络流量自相似分析方法：聚集方差法、R/S 分析法、周期图法和 Whittle 法。提出基于网络流量自相似分析的网络异常检测，采用正常流量模型、对网络流量自相似性参数 Hurst 及其时变函数 $H(t)$ 进行分析，对网络流量进行实时限幅及使用数据库统计，通过检测自相似性变化，判断网络流量是否异常。文献中通过分布式拒绝服务攻击试验表明，这种流量建模的方法比传统的基于特征匹配的网络异常检测法在识别精度与实时性上有较大提高。

3.2 拓扑挖掘

P2P 网络拓扑与传统网络有着明显区别的特征，在学术界提出了几种典型的 P2P 网络拓扑模型：主要包括随机图模型[4]、小世界模型(Small-World)[5]、无标度网络模型(Scale-Free)[6]，现概述如下。

随机图模型是最为简单直观的一种网络拓扑，它从网络图的连通性、度数分布和网络直径等拓扑属性上使用随机算法，在数学分析中是精确可解的，而且也接近社会网络的基本属性，因此较早被提出并为人们所接受。

小世界模型则更接近真实的 P2P 网络，它比起随机图有着几乎相同的网络直径，但又同时具有密集、局部集群结构。小世界网络中一些少量的随机边显著地减少了平均路径长度，形成一定意义上的“短路”。

无标度网络是为了应对更为复杂的网络拓扑而构建的模型，也称幂率网络。它可以描述大规模 P2P 网络乃至整个 Internet 随时间而增长的过程，其中节点的行为与网络的规模是无关的，节点获得一条新边的概率正比于在每一时间步一个节点所具有的度，类似于数学中的分形理论，所形成的整体拓扑像是局部拓扑的复制。在大多数 P2P 网络中，该模型指出随机故障率很可能发生在具有低连接度的节点上，这些节点的异常不会对网络的连通性发生太大影响。

以上几种模型都是描述复杂网络理论中的一部分，文献[7]在 P2P 网络中利用这些网络拓扑属性(节点度数、网络直径、连通性等)建模，发现重点监控对象和监控域。

首先对骨干网络拓扑进行挖掘，可以从不同的粒度上在错综复杂的网络中发现需要重点监控的对象，它的过程主要包括重要节点发现、虚拟骨干网重构和挖掘粒度调整 3 个步骤。首先依据网络的拓扑特征，如节点度数、节点介数、边介数和集聚系数等，发现重要节点；其次将挖掘出的节点连接重构为新的虚拟骨干网；最后通过对虚拟骨干网的性能进行评估，并根据需求和评估结果重新调整挖掘粒度，直到获得所需粒度的虚拟骨干网络模型。

在得到虚拟骨干网络模型之后，可以依据网络规模、安全威胁性质活动状态，从相应粒度上确定防护对象，使得防护策略具有自适应性和针对性，提高防护效率，降低防护成本。例如，当威胁较为轻微时，用细粒度挖掘出一般骨干网，利用现有安全资源进行普通级别的防护。反之，当面对高危性的安全威胁时，首先用粗粒度挖掘出核心骨干网并集中资源进行重点防护，最大程度地保护重要网络的运行和重要任务的完成。

基于小世界等网络模型的拓扑挖掘，还可以发现网络中节点的组团状态。该方法根据网络节点间边的作用关系，将其划分为若干域，使不同域间的节点彼此连接最少，从而发现网络的结构特征，确定监控的边界。在具体操作中一种有效的方法是基于边介数的分割法，即依据边不属于域的程度逐步把不属于任何域的边(即域之间连接的边)删除，直到产生指定数目的域为止。

3.3 域发现协议(Domain Discovering Protocol)

在网络流量分析和拓扑挖掘的基础上，可以观察到 P2P 网络的分域管理的必要性和可行性，按照 P2P 网络自身的拓扑特性来进行自治域的层次化划分和管理，使得监控的性能更加高效。事实上，这正是对基础分布式监控系统的集中管理、分散执行策略的改进。在 P2P 标准化的研究中，一种基于 DDP 域发现协议的监控模型被提出，如图 1 所示。

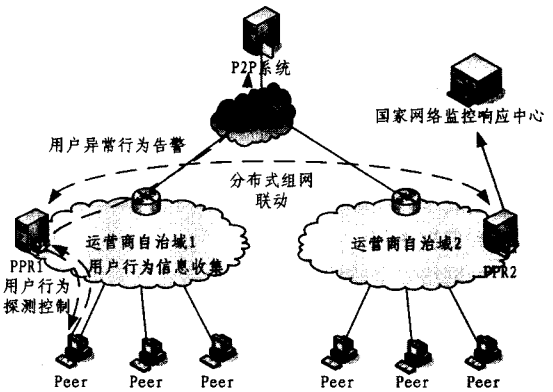

图 1 基于 DDP 协议的监控模型

首先结合运营商的实际网络基础设施参数(例如子网划分、网络量度等)和拓扑挖掘的结果，对网路拓扑参数进行模式化分析，得出适当的自治域划分方案；其次需要研究当前主流的 P2P 应用协议，总结其共性和不足，制定标准化的 DDP 用户行为监管协议。

DDP 系统部署的核心实体称为 PPR(流量重定向实体)，部署在一个自治域内，收集域内节点的 IP、端口、访问内容索引、协议类型等信息，实现流量引导的功能。在此基础上，通过标准化协议的扩展，再收集部分附加的用户行为信息，如用户宣告和申请的资源、用户连接的对端 peer 信息、用户发起

请求的频度、访问成功/失败的比例、上传下载的流量情况等。所有这些信息在单域内的 PPR 上汇总并记录之后，通过一定的数据聚合算法，并结合多域的分布式 PPR 联动，监控整个 P2P 网络的用户行为。此外，域内还可以部署 P2P 流量的缓存实体，结合 PPR 的实时监控，对网络行为进行事后审计。

4 P2P 网络内在行为特征分析

4.1 行为特征归纳

以上是从 P2P 网络的宏观角度对用户行为进行监控的理论方法，但是难以从微观角度把握单个或群体用户的行为特征。从 P2P 终端用户的内在行为表现的角度，也可以提炼出多种监控手段来完善已有的方案。首先列举 P2P 网络中常见的行为特征：

(1)连接 IP 的固定性。一些 P2P 终端客户会固定去探测某些服务器地址[3]，例如 Skype 用户在首次登录时将 ping 美洲/亚洲/欧洲/大洋洲等 4 个固有的服务器，以选择合适的登录区域，利用这种特性，既可以监视此类私有 P2P 协议的存在，又可以了解其是否工作异常。

(2)在用户端出现大量的包复制行为，以及大量不同的目的 IP/端口行为，可能意味着蠕虫类恶意代码的爆发或是 DDOS 攻击的前奏[8]。

(3)大量无效 IP 地址和无效服务请求[8]。用户的恶意行为为了在网络中迅速传播和扩散，其目标的选择具有盲目性。信息的收集和探测都会导致大量无效 IP 地址的产生，由于目标 IP 地址的无效性，因此相应的服务请求也得不到应答。

(4)某些 P2P 用户对所有请求都进行肯定的应答，对任何资源都声称本地具有，这也有可能是一种欺骗连接，进而传播恶意代码[8]。

(5)用户行为的异常相似性[8]。类似于病毒和蠕虫的感染机制，被感染的用户系统进程调具有相似性，传出数据具有相似性，用户间进一步扩散传播的行为也具有循环相似性。

(6)用户出现异常的失效概率[8]。按照统计规律，某一 P2P 网络的用户应该知道一定范围内波动的加入和退出网络的概率，如果用户频繁出现异常失效，就有可能意味着恶意行为。在此还应区分不同性质的节点，例如一些带有超级节点的系统，其超级节点的失效概率应远小于普通节点。

(7)Sybil 攻击[9]：P2P 网络的自由开放特性使得节点可随意创建身份加入系统，因此容易引发 Sybil 攻击。恶意节点通过不断声明有多重身份，使其更易成为 P2P 路由路径中的节点，然后和其他攻击方法结合使用，达到攻击的目的。

(8)Byzantine 攻击[10]：恶意节点通过某种合谋协议在一定范围内排挤正常节点，达到污染和破坏网络正常运行的目的。解决这个理论上的经典难题有助于提高 P2P 系统的容错性和稳健性。

(9)Free-riding“搭便车”行为[11]：节点在 P2P 网络中只请求资源而不贡献资源，该行为不会造成直接的安全威胁，但会影响 P2P 网络的长远健康发展，造成资源枯竭。

4.2 用户认证和信任模型

针对以上 P2P 网络用户内在的行为特征，特别是异常和恶意行为的监控，学术界引入了用户认证和信任模型[12,13]，用于确保交互双方的真实可信和激励节点更好的表现行为，隔离恶意节点，提高交互成功率。在实际的应用系统中，例如 eBay[14]电子商务系统、Maze[15]文件共享系统等，将用户的身份与信任(声誉)值相结合，根据用户的表现行为为其进行选举评分，由一个管理服务器维护节点的身份证书和信任值。

一般地，建立信任模型应当满足以下原则：

(1)信任随着传递削弱，即一个节点的信任值推演到另一个节点的可信度必然会降低，需要引入信任链的传播演化机制；

(2)低的信任值会更大地影响可信度的计算。这是与社会学原理吻合的，即良好信任的建立和维护是缓慢而困难的，而破坏信任则相对容易；

(3)信任值需要进行历史累计修正，近期的表现会更大地影响节点当前的信任值；

(4)为了防止 P2P 节点在执行恶意行为后通过漂白身份来欺骗监控系统，还需要合理的信任值初始化机制。

4.3 传播行为建模

从传播行为的动态角度对用户行为特征进行分析，文献[16]中提到了根据用户行为特征值和分布建立，在 P2P 网络中建立双因素(Two-Factor)传播行为监控模型，通过解析方法获得 P2P 用户恶意行为传播趋势的动态变化。其数学模型见式(1)。

$$\begin{cases} dR(t)/dt=\gamma I(t) \\ dQ(t)/dt=\mu S(t)J(t) \\ \beta(t)=\beta_0[1-I(t)/N]^{\eta} \\ N=S(t)+I(t)+R(t)+Q(t) \\ dS(t)/dt=-\beta(t)S(t)I(t)-dQ(t)/dt \end{cases} \tag{1}$$

假设 P2P 网络中有用户传播恶意代码，其中 $R(t)$ 表示时刻 t 感染后被免疫的主机数；$I(t)$ 表示具有感染性的主机数；$Q(t)$ 表示时刻 t 被恶意代码感染前就作了免疫处理的主机数；$S(t)$ 表是时刻 t 易感染的主机数；$J(t)$ 表示时刻 t 已被恶意代码感染的主机数，$J(t)=R(t)+I(t)$；$\beta(t)$ 表示时刻 t 的感染率；γ,μ 和 β_0 为常量。

由式(1)能推导出 $I(t)$ 和时间 t 的关系式，见式(2)，其解析曲线如图 2 所示。

$$dI(t)/dt=\beta(t)[N-R(t)-I(t)-Q(t)]I(t)-dR(t)/dt \tag{2}$$

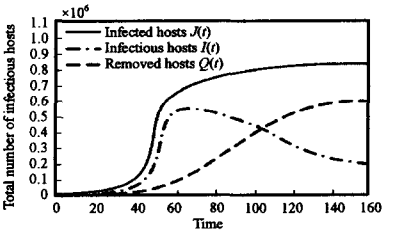

图 2 Two-Factor 模型中网络蠕虫的传播趋势

图中给出了 Two-Factor 模型中恶意代码的传播趋势。图中取结点数 $N=10000000$，$I_0=1$，$\eta=3$，$\gamma=0.05$，$\mu=0.06/N$，$\beta_0=0.8/N$。可以看到，随着 $Q(t)$ 的增长，$I(t)$ 的变化走势趋向于 0。根据此模型揭示的规律，监控系统可以在适当的时间段采取相应的安全措施，例如身份认证、权限限制、连接截断、传播跟踪等手段。

更深入地，文献[17]中对 Two-Factor 模型进行改进，提出 Worm-Anti-Worm(WAW)模型，WAW 考虑传播中出现的对抗因素，并得出更为精确的传播行为。

4.4 行为相关性研究

为了将 P2P 网络宏观外在表现与微观内在行为特征结合起来，文献[18]提出了行为相关性研究，即把用户加入和退出 P2P 网络的活跃时间、发起服务请求以及被请求服务数、用户上传与下载文件分布等与 P2P 网络的带宽、流量综合起来考虑，通过对 Maze 系统的实际运行数据进行相关性分析的结果表明，P2P 用户的流量大小与其在线活跃时间具有很强的相关性，而带宽高低对流量及用户请求次数的影响并没有绝对的关系。从此研究结果出发，可以均衡热点(时间/空间)瓶颈，限制 free-riding 等不良行为。

5 P2P 监控的安全平台

在学术界提出各种 P2P 网络行为监控理论模型的同时，产业界也设计了特定的安全平台，使得监控的理论模型得以在其基础之上实现。

JXTA[19] Project 是 SUN 公司提出的 P2P 开发平台，它为构建跨平台、跨操作系统和跨编程语言的 P2P 应用提供应用程序的底层支持。JXTA 的底层核心包括了节点状态监控和安全子模块，作为支撑整个 P2P 服务和应用的基础，为 P2P 网络中的安全问题提供了一些解决策略，例如：为了保障数据传输的安全性，传输数据的途径由 API 控制；为对等机建立对等机 ID 和对等机组，在分散结构上加强了单点控制；建立了信任机制，在信任的范围内对等机组的成员可以交互信息，实现共享。

PTPTL[20] (Peer to Peer Trust Library)是 Intel 公司提出的一个开放源码项目。它不仅试图构建一个 P2P 安全平台，还希望成为 P2P 应用的安全标准，允许各个不同的 P2P 应用程序之间互相通讯。PTPTL 建造在 OpenSSL 工具包上，提供对数字证书、对等节点认证、安全存储、公私钥加密、对称加密及数字签名等的技术支持，Intel 认为在 PTPTL 平台上可以构建可信的 P2P 网络，以及实现对其运行的安全监控。

结束语 综上所述，目前提出的各种 P2P 用户行为安全监控理论在实践，依然存在着一些问题：

算法过于复杂。例如在非结构化复杂网络拓扑挖掘中，涉及到的算法复杂度一般都在指数级或幂级，随着 P2P 网络规模的扩大，很难在有效的时间和空间内挖掘出异常和恶意行为，也就难于达到实时监控和防范的目的。而结构化的 DHT 算法的维护机制较为复杂，尤其是结点频繁加入退出造成的网络波动会极大增加 DHT 的维护代价。

缺乏统一的监控协议。在分布式异构的 P2P 网络中，不同的应用协议报文和交互时序各不相同，为了提取其行为特征，需要有一种附加的协议来统一处理监控的协议信息。

收集信息量大。在用户行为特征分析中，需要收集大量的终端用户连接信息、发送和接受报文信息、请求和提供服务信息、加入和退出网络信息等，这对于监控服务端的存储和处理能力是一个挑战，因此未来的监控发展需要轻量级和可扩展的信息收集机制。

面向全局处理开销大。当前的理论模型都是针对全局网络进行建模，在实际处理中会面临大负载下的集中式处理瓶颈。作为改进，需要有灵活的局部化策略，并且能将局部的监控协同联动，以实现对全网的有效监控。

安全平台作用有限。一些厂商提出的 P2P 安全平台缺乏标准化和互通性，目前来看还局限于研究实验的阶段，难于在 P2P 应用层上发挥作用。

总之，在没有统一有效的 P2P 网络行为监控基础设施的情况下，作为 P2P 的用户和厂商还缺少足够的动力来加强自身安全监控和防范，而运营商和管理部门要对全局网络进行集中式的用户行为分析与控制，在实现上难度大，未来还需要进一步研究新的理论框架和工程模型。

参 考 文 献

[1] 吴中伟，王义安，韩进，等. 一种 P2P 文件共享监控系统的实现[J]. 计算机工程，2007，33(14)：139-141

[2] 刘刚，方滨兴，胡铭曾，等. BitTorrent 流量的捕获方法及自相似性的评价[J]. 计算机应用研究，2006，23(5)：205-206，209

[3] 彭维. P2P 客户端行为机制研究[D]. 北京工业大学，2007

[4] Gilbert E. Random Graphs[J]. Annals of Mathematical Statistics，1959，30：1141-1144

[5] Watts D J，Strogatz S H. Collective dynamics of 'small-world' network [J]. Nature，1998，393(6684)：440-442

[6] Bollobás B，Riordan O M. Mathematical results on scale - free random graphs[M] // S. Bornholdt and H. G. Schuster，eds. Handbook of Graphys and Network-From the Genome to the Internet. Wiley-VCH，2003

[7] 第文军，薛丽军，蒋士奇. 运用网络流量自相似分析的网络流量异常检测[J]. 兵工自动化，2003，22(6)：28-31

[8] 周瑛. 基于 P2P 技术的网络蠕虫防御机制研究[D]. 重庆：重庆大学，2007

[9] 王鹏，王琳，祝跃飞. 在 P2P 网络下 Sybil 攻击的研究与防范[J]. 微电子学与计算机，2006，23(4)：162-165

[10] Chen Ying，Hwang Kai. Byzantine Fault Tolerance of Inverse de Bruijn Overlay Networks for Secure P2P Routing[J]. IEEE Transactions on Parallel and Distributed Systems (TPDS)，October 2006

[11] Krishnan R，Smith M D，Tang Zhulei，et al. The impact of free-riding on peer-to-peer networks[C] // System Sciences，2004. Proceedings of the 37th Annual Hawaii International Conference：10

[12] Yu Lan，Susilo W，Safavi-Naini R. X2BT Trusted Reputation System：A Robust Mechanism for P2P Networks[C] // Lecture Notes in Computer Science. Volume 4301/2006

[13] 侯孟书，卢显良，周旭，等. P2P 系统的信任研究[J]. 计算机科学，2005，32(4)：113-115

[14] eBay Feedback Forum[DB/OL]. http：//pages. ebay. com/services/forum/feedback. html? ssPageName=STRK

[15] Yang Mao，Dai Yafei，Li Xiaoming. Bring Reputation System to Social Network in the Maze P2P File-Sharing System[C] // The International Symposium on Collaborative Technologies and Systems. USA，May 2006

[16] Zou CC，Gong W，Towsley D. Code Red worm propagation modeling and analysis[C] // Proc. of the 9th ACM Symp. on Computer and Communication Security. Washington，2002：138-147

[17] 文伟平，卿斯汉，蒋建春，等. 网络蠕虫研究与进展[J]. 软件学报，2004，15(8)：1208-1219

[18] 陈宝钢，许勇，胡金龙，等. P2P 文件共享系统用户行为特性研究[J]. 计算机科学，2007，34(12)：122-125，142

[19] http：//jxta. dev. java. net [DB/OL]

[20] http：//sourceforge. net/projects/ptptl [DB/OL]