计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2016.20150066
53(6):1314-1324,2016

基于种群动力学的 P2P IPTV 系统内容污染扩散模型

王海舟　陈兴蜀　杜　敏　王文贤
(四川大学计算机学院　成都　610065)
(whzh.nc@scu.edu.cn)

A Modeling Framework with Population Dynamics for Content Pollution Proliferation in P2P IPTV System

Wang Haizhou, Chen Xingshu, Du Min, and Wang Wenxian
(*College of Computer Science, Sichuan University, Chengdu* 610065)

Abstract　With the growing maturity of peer-to-peer (P2P) technology, IPTV (Internet protocol television) applications based on it have gained great commercial success, which has received more and more attentions from both global industry and academia. However, the characteristics of distribution, anonymity, rapid propagation, and large-scale users in P2P IPTV systems make them more likely to be vulnerable targets. In this paper, the procedure of content pollution attack in P2P IPTV systems is presented firstly, and then the various user behaviors under the pollution attack are analyzed. With that knowledge, a modeling framework with population dynamics of content pollution in P2P IPTV systems is proposed. Different from the existing content pollution propagation models, this model considers the impact of user behaviors in the content pollution attack of real world P2P IPTV systems, such as continuous watching behavior, arrival behavior, departure behavior. Theoretical analysis and simulation experiments show that the proposed model is a feasible and efficient tool to analyze the content pollution propagation in real world P2P IPTV systems, especially for the dynamic evolution of participating users and the user departure behavior.

Key words　P2P (peer-to-peer) technology; IPTV (Internet protocol television) system; content pollution attack; modeling; population dynamics

摘　要　近年来,随着 P2P(peer-to-peer)技术的日益成熟,基于 P2P 技术的 IPTV(Internet protocol television)系统逐渐流行并取得了巨大成功,吸引了来自全球产业界和学术界的密切关注.但是,由于 P2P IPTV 系统自身特有的分布性、匿名性、快速传播性、用户规模大等特点使得其更容易成为攻击的目标.在对 P2P IPTV 内容污染攻击过程和节点在内容污染攻击中的行为特征进行详细分析的基础上,建立了一个具有种群动力学特征的 P2P IPTV 系统内容污染扩散模型,该模型充分考虑了在真实世界的 P2P IPTV 系统内容污染攻击中用户的持续观看行为、用户的加入和离开行为对频道用户规模动态演变的影响.实验结果表明,提出的 P2P IPTV 内容污染扩散模型可以有效和准确地刻画真实内容污染攻击场景中污染攻击对频道的在线节点规模动态演变和节点离开行为的影响情况.

关键词　P2P 技术;IPTV 系统;内容污染攻击;建模;种群动力学

中图法分类号　TP393.08

收稿日期:2015-01-22;**修回日期**:2015-05-05

基金项目:国家科技支撑计划基金项目(2012BAH18B05);国家自然科学基金项目(61272447)
This work was supported by the National Key Technology Research and Development Program of China (2012BAH18B05) and the National Natural Science Foundation of China (61272447).

在 P2P(peer-to-peer)技术因其具有良好的自组织性、可扩展性、健壮性等诸多优势被广泛应用于主流的互联网应用,其中以 PPTV[1](原 PPLive),PPStream[2],UUSee[3]等为代表的 P2P 网络电视系统(P2P IPTV)取得了巨大的商业成功,凝聚了极其庞大的用户群体,深刻影响着互联网的用户行为,吸引了来自全球的产业界和学术界的密切关注.据中国互联网络信息中心(CNNIC)发布的最新统计报告显示[4]:截止 2013 年 12 月底,中国网络视频用户数量增至 4.28 亿,网络覆盖率达到 69.3%.

P2P 技术在给互联网用户带来各种便利的同时,因其自身特有的分布性、匿名性、快速传播性、用户规模大等特点更容易受到攻击且造成的破坏更加严重,影响更加广泛,给互联网带来了巨大的安全隐患和新的挑战.有研究指出,占据了互联网视频直播市场超过 60%市场份额[1]的 PPTV 网络电视面临内容污染攻击的威胁[5-6],该攻击会极大地降低 P2P IPTV(Internet protocol television)系统的性能和服务质量,严重影响用户的正常收视行为.基于 P2P 技术的 IPTV 系统内容污染问题已经成为了 P2P 研究领域的一个重要研究方向,受到了来自 P2P IPTV 系统运营商、互联网服务提供商和信息安全领域科研人员等的广泛关注.

提出并建立能准确反映 P2P IPTV 系统内容污染传播的数学模型,一方面为深入研究 P2P IPTV 覆盖网络中内容污染的扩散特征、建立有效的抵御内容污染攻击的防御方法提供理论依据;另一方面,为 P2P IPTV 系统的安全评估提供参考,为 P2P IPTV 系统相关安全标准的制定奠定基础,为 P2P IPTV 系统监管技术的发展提供技术支持.

1 相关研究背景及本文创新点

内容污染攻击起初源于 P2P 技术引起的版权纠纷.P2P 文件共享系统给用户分享资源带来极大便利的同时,也因用户未经授权发布并传播正版资源导致版权行业遭受巨大的经济损失.为了保护合法权益,版权所有者雇佣专业的污染公司(如美国 Overpeer 公司、法国 RetSpan 公司等)采取技术手段对 P2P 文件共享系统实施内容污染攻击以尽可能地阻止用户获得原始的资源文件拷贝.但是内容污染攻击技术却没有从根本上解决盗版纠纷问题,反而给恶意攻击者提供了破坏 P2P 文件共享系统和传播恶意病毒程序的新思路,对 P2P 文件共享系统带来了巨大的危害.

2005 年,Liang 等人[7]首次对 P2P 文件共享系统 KaZaA 的覆盖网络实施了内容污染测量研究.他们设计了一个爬行平台采集原始数据并利用一个自动处理程序判定该数据是否被污染.研究结果表明针对流行歌曲的内容污染普遍存在,超过 50%的歌曲副本已经遭受污染.随后,Liang 等人[8]利用主动测量方法对非结构化的 P2P 文件共享系统 FastTrack 和基于 DHT(distributed Hash table)网络的 Overnet 进行了索引污染和内容污染评估.研究发现这 2 种系统均极易遭受污染攻击.Lee 等人[9]认为 P2P 文件共享系统中的用户行为是影响污染攻击动态的重要因素,通过实验发现系统中的用户对污染攻击并不敏感,他们通常在资源刚下载完成或者很长时间之后进行资源检查.在此基础上,他们建立了一个数学模型用以理解污染攻击的动态演变,分析结果显示通过降低用户对污染的意识可以进一步提高污染攻击的效果.Christin 等人[10]为了研究索引污染攻击和内容污染攻击在对 P2P 文件共享系统内容可用性影响的差异,采用主动、被动相结合的测量方法对 4 款非常流行的 P2P 文件共享网络(eDonkey,Overnet,FastTrack,Gnutella)的污染态势进行了对比研究,结果表明遭受攻击的目标 P2P 文件共享网络拓扑结构属性与内容可用性存在紧密联系.Dhungel 等人[11]采用主、被动测量相结合的方法对 BitTorrent 进行了伪造数据块攻击和不合作节点攻击的研究.研究结果表明这 2 类攻击可以延缓用户下载文件的平均耗时,但不超过 50%,因此不足以严重影响到 BitTorrent 用户的正常使用.此外,史建焘等人[12]提出了一种通过对网络环境和节点行为进行抽象的随机模型来定量分析内容污染攻击对 BitTorrent 下载节点以及共享网络的影响,并通过广域网实验分析了污染攻击的危害性.相比于 P2P 文件共享系统,针对 P2P IPTV 系统内容污染攻击测量工作起步相对较晚,由于 P2P IPTV 系统拥有的巨大学术研究价值,正吸引来自全球学术界越来越多的关注.

Haridasan 等人[13]首次对 P2P IPTV 系统进行了内容污染攻击研究和评估.作者设计了一款基于拉模式架构的 P2P IPTV 直播系统—Secure-Stream.仿真实验表明,Secure-Stream 可以抵御系统中有限比例数量的恶意节点,在审计组件的帮助下该系统能够在更大规模的攻击时提供令人满意的服务质量.Lin 等人[14]提出了一种基于真实 P2P IPTV 系统的仿真系统 SPoIM,研究了基于该系统对不同网络

配置情况下的污染攻击对 P2P IPTV 系统的影响情况.作者提出了一种新的污染攻击方式“伤寒广告攻击”(typhoid adware attack),这类攻击方式以修改传播的媒体内容为目的,比如在流媒体中植入广告信息.通过仿真实验研究表明,污染攻击的效果和效率并不依赖于网络的规模,而与网络的稳定性和污染者的访问带宽密切相关. Vieira 等人[15-16]通过仿真实验对基于网状结构的 P2P IPTV 系统进行了污染攻击损害评估.作者在对 P2P 文件共享系统中的信誉系统进行修改的基础上提出了一种用于 P2P IPTV 系统的信誉系统——StRepS,并评估了该系统抵御内容污染传播的效果.研究结果显示少量的污染者可以对整个 P2P IPTV 系统产生巨大影响,而作者提出的信誉系统可以快速识别和孤立污染者,避免了攻击者的共谋行为. Hu 等人[17]提出了一个信任管理系统用以识别攻击者并阻止他们对 P2P 网络电视系统实施进一步的污染攻击.模拟实验表明,作者提出的信任管理系统能够有效地检测污染攻击者,可以帮助用户接收更多的清洁数据并提高 P2P IPTV 系统的整体性能. Hu 等人[18]随后提出了一个同时具备污染数据检测和攻击者识别的联合系统,并对其进行了性能分析,研究了权衡污染防御性能和系统开销的折中方案.

以上的研究工作均是基于搭建仿真 P2P IPTV 系统的基础上展开,但是仿真平台很难模拟真实系统中的用户复杂的动态行为(如用户的节目持续观看行为、用户的加入和离开行为、用户规模的动态变化等)和拓扑结构动态演变.因此与在真实网络环境的 P2P IPTV 系统采集的污染攻击数据相比,模拟实验的数据和结论很难得到有效性验证. Dhungel 等人[5,19]首次对一款非常流行的商用 P2P IPTV 系统 PPLive 实施了内容污染攻击主动测量研究,提出了一种有效的针对 P2P IPTV 系统进行内容污染的方法. Dhungel 等人的研究工作首次通过真实的测量实验证明了当时的商用 P2P IPTV 系统存在遭受内容污染攻击的风险,具有里程碑意义,不过文中所述的内容污染攻击方法建立在 PPLive 系统早期明文通信版本的基础上,对后续已经实施了加密通信的这类系统的攻击效果尚不可知.此外,作者仅研究了 PPLive 在遭受污染攻击时节点数量变化情况和视频块上传、下载情况 2 类指标,没有更加深入地评估内容污染攻击对 P2P IPTV 系统的影响.随后, Wang 等人[6]在 Dhungel 等人研究工作的基础之上对采用了加密机制的 PPLive 新版本进行了内容污染攻击主动测量研究,并分别从节点数量动态演变、用户生命周期特性、用户连接特性、污染数据块动态演变和污染率动态演变 5 个方面对 PPLive 网络进行了更为详细的污染攻击评估.

目前,利用数学建模方法对内容污染攻击在 P2P IPTV 系统的传播进行研究的相关工作较少. Yang 等人[20-21]首次对 P2P IPTV 系统进行内容污染攻击的形式化分析,作者建立了一个概率模型去捕获内容污染的进程,利用了模拟实验的方式对内容污染进行了评估.作者借鉴了 P2P 文件共享系统中内容污染攻击模型的相关研究成果,首次提出了 P2P IPTV 系统的内容污染数学模型,对于进一步开展相关的研究工作具有重要研究价值. Li 等人[22]首次提出了一个分布式的防御和检测 P2P 网络电视系统中内容污染攻击节点的数学模型,该模型考虑了单个污染者和多个污染者 2 种攻击场景.仿真结果表明该方法可以有效和快速地检测出恶意节点.作者随后将该方法推广到了无线网状网络(WMNs)和在线社交网络(OSNs)[23]. Kang 等人[24]在研究多种内容污染攻击特性的基础上提出了一个防御 P2P IPTV 系统内容污染攻击的信任管理系统数学模型,该模型同时考虑直接和间接信任 2 种场景,利用一个动态置信因子来调节直接和间接信任的权值.实验研究表明,该信任管理系统可以有效识别污染者并阻止其进一步扩散污染数据块.

相比已有的研究工作,本文主要的创新点包括:1)相比已有的研究工作中利用仿真 P2P IPTV 系统来验证模型的准确性,本文首次将内容污染扩散模型与真实污染环境中的测量数据进行对比和验证,确保了数学模型在真实污染环境中的可用性和准确性.对比结果表明,本文提出的 P2P IPTV 内容污染扩散模型可以有效地模拟商用 P2P IPTV 在遭受内容污染攻击时的污染扩散情况和用户动态变化情况.2)用户的节目观看行为(如:用户会对感兴趣且播放效果良好的节目持续观看,否则会切换到其他节目)、用户的加入和离开行为是真实世界中 P2P IPTV 覆盖网络搅动(churn)[25]特性的具体体现,在真实世界中用户的这些行为显然会受到内容污染攻击的影响从而呈现出特有的行为特征.已有的研究工作很少甚至没有考虑到这些因素对仿真系统测量结果和内容污染模型准确性的影响.本文首次利用了种群动力学的思想来考虑 P2P IPTV 内容污染扩散模型的建立,将用户的加入行为(人口出生)、正常离开行为(人口正常死亡)、因污染攻击离开行为(人

口因病死亡)、切换频道行为(受疾病感染的人群获得暂时免疫力)等引入到模型的建立中,从而建立更接近真实污染环境的内容污染扩散数学模型.

2 P2P IPTV 内容污染攻击简介

2.1 P2P IPTV 节点发现及视频分发架构

由于涉及到商业隐私,主流的 P2P IPTV 系统均未公开其通信协议,也尚未发布任何的相关技术文档,对了解其系统架构和实施测量研究带来了巨大挑战. Hei 等人[26]首次利用协议分析技术对当时最为流行的 PPLive 网络电视系统通信报文进行了分析,并首次披露了 PPLive 的节点发现和视频分发架构. 姜志宏等人[27]的研究结果表明,主流的商用 P2P IPTV 系统,如 PPLive,PPStream 和 UUSee 等,均采用了类似的系统架构. 如图 1 所示,描述了一个典型 P2P IPTV 系统的节点发现和视频分发过程:

1) 当 P2P IPTV 客户端软件启动后,首先向频道列表服务器(channel-list servers)发送频道列表文件更新请求,如果可用就下载到本地磁盘并更新已有文件,否则就使用本地保存的上次成功更新的频道列表文件.

2) 当用户点击选择某个频道之后,客户端随即注册并加入 P2P IPTV 的覆盖网络(overlay network)成为一个网络节点,随后向多个节点列表服务器(peer-list servers)发送节点列表请求获取观看当前节目的初始节点列表,列表中的节点信息通常包括节点的 IP 地址和当前使用的 TCP 和 UDP 端口号. P2P IPTV 系统的覆盖网络以频道为基本单位,每一个直播或点播频道均拥有一个覆盖网络,每个节点最多只能同时加入一个覆盖网络. (注:在 P2P IPTV 系统中,一个频道通常由多个节目记录组成,如:CCTV-6 被称为频道、正在播放的电影《阿凡达》被称为一条节目记录.)

3) 客户端然后向初始节点列表中的节点发送节点列表请求以获得更多的节点信息并将这些新节点加入到初始节点列表. 当拥有了一定数量观看当前节目的用户节点后,客户端将与这些节点中的存活节点进行视频数据交换,实现指定节目的播放.

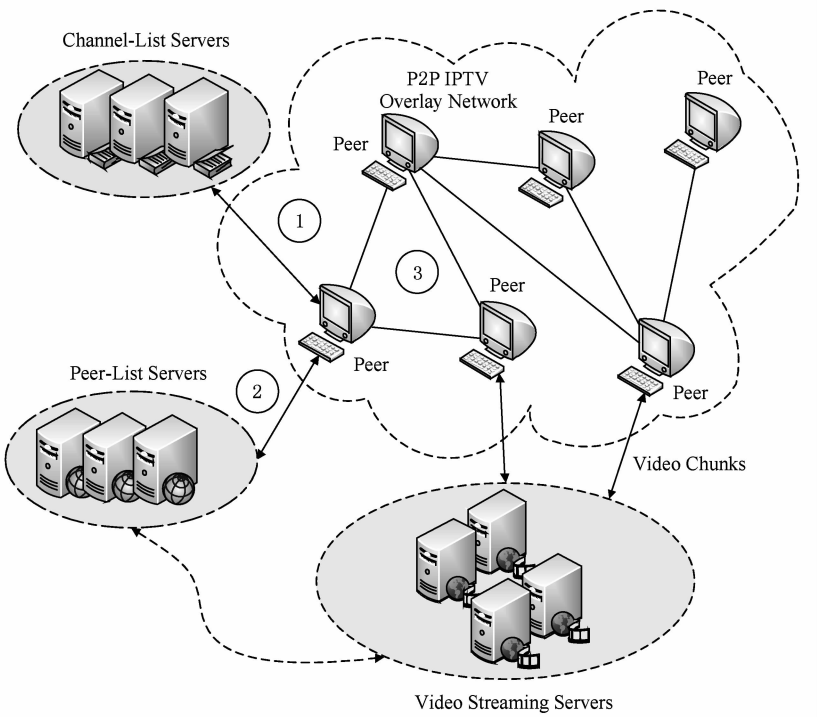


Fig. 1 Architecture of peers discovery and video distribution for P2P IPTV system.

图 1 P2P IPTV 系统节点发现和视频分发架构

2.2 P2P IPTV 内容污染攻击架构

Dhungel 等人[5,19]首次对非常流行的商用 PPLive 网络电视系统实施了内容污染攻击测量研究,提出了一种有效的针对 P2P IPTV 系统进行内容污染攻击的方法. Wang 等人[6]在 Dhungel 等人研究工作的基础之上,针对最新版本的 PPLive 系统,从节点数量动态演变、用户生命周期特性、用户连接特性、污染数据块动态演变和污染率动态演变 5 个方面对

其进行了更为详细的内容污染攻击测量.利用以上文献的研究结果,本文绘制了针对 P2P IPTV 系统的典型内容污染攻击架构,如图 2 所示.在 P2P IPTV 系统的内容污染攻击中,攻击者首先将自己注册到目标频道覆盖网络中,然后向正常节点广播其拥有大量可用的视频块(video chunk)来吸引受害节点下载.由于受害节点没有任何抵御这类污染攻击的防御机制不会对已经遭受污染的视频块做任何校验操作,将收到的污染数据又转发给观看当前节目的其他邻居节点,而这些节点则会将污染数据转发给更多的节点.最终,污染数据不断扩散并蔓延到整个 P2P IPTV 系统的频道覆盖网络.研究表明内容污染攻击会极大地降低 P2P IPTV 系统的性能和服务质量,严重影响用户的正常收视行为[5-6,19].

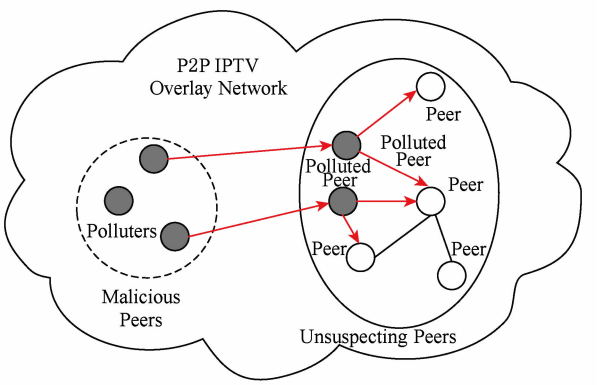


Fig. 2 Pollution attack for P2P IPTV.

图 2 P2P IPTV 内容污染攻击示意图

3 P2P IPTV 内容污染传播建模

3.1 模型假设及参数说明

在建立 P2P IPTV 内容污染扩散模型之前,本文做 5 点假设:

1) 在内容污染攻击期间,节点在单位时间内以常数 A 加入被污染频道的覆盖网络中,而污染攻击者一旦加入被污染频道的覆盖网络后则不会离开直至污染攻击结束.

2) 节点在观看正常的节目过程中以概率 δ 离开 P2P IPTV 系统,其中 $0<\delta<1$;当遭受内容污染时,节点因系统提供的播放质量下降导致离开 P2P IPTV 系统的概率为 ω,$0<\delta<\omega<1$.

3) 频道覆盖网络节点的总数 $N(t)$ 是一个变量,随时间变化而变化,单位时间内有节点加入,也有节点离开该覆盖网络,$N(t)=G(t)+P(t)$.

4) 任何时刻,节点只能处于以下 4 种状态中的一种:健康状态 G(good)、污染状态 P(polluted)、滞留状态 R(remaining)和离开状态 Q(quit);处于污染状态的节点不会再次遭受污染,处于滞留状态的节点可能会再次切换回被污染的频道,而离开的节点再次加入系统被认定为新的节点;节点的状态转移在一个时间单元(time unit)内完成.

5) 当一个健康节点的邻居节点中存在至少一个污染节点时,该节点在下一个时间单元将被污染而成为污染节点.为了便于描述本文提出的内容污染传播模型,将建模时用到的各类参数和符号列举在表 1 中.

Table 1 Definition and Explanation of Notations

表 1 模型中的符号定义及说明

Notation	Explanation
$G(t)$	The number of good peers at unit t
$P(t)$	The number of polluted peers at unit t
$R(t)$	The number of remaining peers at unit t
$Q(t)$	The number of exited peers at unit t
$\varphi(t)$	Probability that a good peer is polluted at unit t
A	The number of peers which join in overlay per time unit
δ	Probability that a good peer leaves system per time unit
α	Probability that a remaining peer goes back to polluted overlay per time unit
β	Probability that a polluted peer replays polluted channel per time unit
ω	Probability that a polluted peer leaves system per time unit
μ	Probability that a polluted peer switches to other channels per time unit
N_p	The number of polluters
d	The average number of neighbors per peer
k	Correction factor

3.2 内容污染扩散模型的建立

3.2.1 节点所处状态说明

1) 健康状态(G).当用户打开 P2P 流媒体系统并点击播放某个频道时,节点处于健康状态,此时流媒体系统获取到的视频块均为干净数据,未被污染.

2) 污染状态(P).内容污染攻击开始后,节点因接收到攻击者上传的污染数据而进入到污染状态.

3) 滞留状态(R).当被污染的节点感受到系统服务质量正在恶化时,选择切换到另一个频道等待被污染频道的服务质量恢复正常,即进入滞留状态 R.

4) 离开状态(Q).节点关闭 P2P IPTV 客户端软件,离开系统即进入离开状态.

3.2.2 节点状态转移分析

根据对 P2P IPTV 内容污染攻击过程和节点在

内容污染攻击中的行为进行分析,本文得出了如图 3 所示的节点状态转移图.

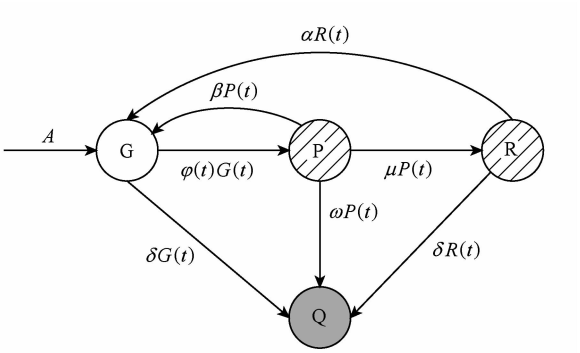


Fig. 3 The transition diagram for the peers of P2P IPTV system.

图 3 P2P IPTV 节点的状态转移图

首先,在内容污染攻击开始后,健康节点因接收到攻击者上传的污染数据而成为污染节点,由状态 G 进入污染状态 P. 根据模型的假设,只要健康节点接收到一个污染的视频块即认定该节点被污染而进入状态 P.

随着污染数据在 P2P IPTV 覆盖网络中不断蔓延,受污染的视频数据将极大地降低被污染节点的视频播放质量,严重影响用户的正常体验. 当受污染节点无法忍受低下的视频播放质量时,节点将可能执行以下操作中的一种:1)不执行任何操作,等待被污染频道的服务质量恢复正常;2)直接离开 P2P IPTV 系统,进入状态 Q;3)再次点击播放按钮重新加入到频道覆盖网络中恢复成为健康状态 G;4)切换到其他正常频道,等待被污染频道的服务质量恢复正常,进入滞留状态 R.

处于滞留状态 R 的节点由于并未真正离开 P2P IPTV 系统而是进入了另一个频道覆盖网络中,因此该状态不同于离开状态 Q. 如果被污染的频道非常流行且足够吸引用户,大部分用户愿意耐心等待系统恢复正常并重新尝试播放该节目,进入健康状态 G.

3.2.3 模型的建立

1) 健康节点的变化率

时刻 t 健康节点的数量为 $G(t)$,被污染的节点数量为 $P(t)$,污染攻击者的数量为 N_p,每个节点的平均邻居节点数为 d. 根据模型假设,一个健康节点只要连接到一个污染节点即被感染,而时刻 t 一个健康节点连接的任何 1 个邻居节点为健康节点的概率为 $\frac{G(t)}{G(t)+P(t)+N_p}$,那么 d 个邻居节点全部是健康节点的概率为 $\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+P(t)+N_p-i}$,因此可计算得时刻 t 一个健康节点因接收到污染数据而成为污染节点的概率 $\varphi(t)$ 为

$$\varphi(t)=1-\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+P(t)+N_p-i}. \quad (1)$$

由于被污染节点之间可以存在满连接关系,即污染节点的邻居节点也全部是污染节点,这将导致可供健康节点连接的污染节点数减少,且随着污染节点数量的不断增加,存在满连接的污染节点数量也将增加. 因此,对式(1)进行修正得:

$$\varphi(t)=1-\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_p)-i}, \quad (2)$$

其中,k 为修正系数,$0<k<1$. 因此,单位时间里共有 $\varphi(t)G(t)$ 个健康节点成为污染节点.

用户的加入和离开行为是 P2P 流媒体系统覆盖网络 Churn 特性的具体体现,为了更加准确地描述用户在内容污染攻击中呈现的行为特性,建立接近真实世界的内容污染场景,本文首次在建立内容污染模型时引入了用户的加入和离开行为. 根据已有的测量研究结果表明,在用户规模为 3 528 的一个 PPLive 流行频道中,未遭受内容污染攻击时,单位时间(1 min)内加入频道覆盖网络的平均用户数约为 150 人[6]. 因此,在模型中本文假定在内容污染攻击期间(90 min),用户以恒定的速率加入频道的覆盖网络,用常量 A 表示. 单位时间内,用户的正常离开系统的概率为 δ,因此时刻 t 离开的健康节点数量为 $\delta G(t)$. 与此同时,单位时间内有 $\alpha R(t)$ 个滞留节点切换回被污染频道并播放节目而再次成为健康节点,有 $\beta P(t)$ 个污染节点通过再次点击播放按钮重新加入到频道覆盖网络中恢复为健康节点. 因此,健康节点的变化率为

$$\frac{dG(t)}{dt}=A-\varphi(t)G(t)-\delta G(t)+\alpha R(t)+\beta P(t); \quad (3)$$

将式(2)带入式(3)得:

$$\frac{dG(t)}{dt}=A-\left(1-\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_p)-i}\right)\times G(t)-\delta G(t)+\alpha R(t)+\beta P(t). \quad (4)$$

2) 污染节点变化率

单位时间内有$(1-\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_p)-i})\times G(t)$个健康节点因接收到的污染数据成为污染节点. 与此同时,用户正常离开系统的概率为 δ,用户因污染攻击离开系统的概率为 ω,因此时刻 t 离开

系统的污染节点数量为$\omega P(t)$. 此外,单位时间内有$\beta P(t)$个污染节点通过再次点击播放按钮而离开污染状态,有$\mu P(t)$个污染节点切换到另一个频道而离开污染状态,等待被污染频道服务质量恢复正常. 所以,污染节点的变化率为

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t}=\left(1-\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_{\mathrm{p}})-i}\right)G(t)-\omega P(t)-\mu P(t)-\beta P(t). \tag{5}$$

3) 滞留节点变化率

从图 3 可知,当污染节点感受到系统服务质量恶化时,如果被污染频道有足够的吸引力(如非常流行的节目资源),则节点就会以概率μ暂时切换到其他频道并进入滞留状态,单位时间切换的节点数量为$\mu P(t)$. 在等待被污染节目恢复正常的过程中,节点仍然可能会离开系统,单位时间离开的节点数量为$\delta R(t)$,更多的节点会选择切换回被污染频道并尝试播放该节目,单位时间内有$\alpha R(t)$个滞留节点转变为健康节点. 此外,当滞留节点观看的节目比被污染频道的节目更吸引用户,他们则将以$1-\delta-\beta$的概率继续观看. 于是,滞留节点的变化率为

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t}=\mu P(t)-\delta R(t)-\alpha R(t). \tag{6}$$

4) 离开节点变化率

根据以上的分析可知,时刻t单位时间内离开系统的健康节点数量为$\delta G(t)$,离开系统的污染节点数量为$\omega P(t)$,离开的滞留节点为$\delta R(t)$. 因此,离开系统节点的变化率为

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t}=\delta G(t)+\omega P(t)+\delta R(t). \tag{7}$$

综合以上对建模变量和参数的设置以及 4 类节点变化率的分析,可用如下的非线性微分方程组描述 P2P IPTV 系统的内容污染传播模型:

$$\begin{cases}\dfrac{\mathrm{d}G(t)}{\mathrm{d}t}=A-\left(1-\displaystyle\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_{\mathrm{p}})-i}\right)\times G(t)-\delta G(t)+\alpha R(t)+\beta P(t);\\ \dfrac{\mathrm{d}P(t)}{\mathrm{d}t}=\left(1-\displaystyle\prod_{i=0}^{d-1}\frac{G(t)-i}{G(t)+k(P(t)+N_{\mathrm{p}})-i}\right)\times G(t)-\omega P(t)-\mu P(t)-\beta P(t);\\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t}=\mu P(t)-\delta R(t)-\alpha R(t);\\ \dfrac{\mathrm{d}Q(t)}{\mathrm{d}t}=\delta G(t)+\omega P(t)+\delta R(t).\end{cases} \tag{8}$$

3.3 内容污染扩散模型的验证

在已有的针对 P2P IPTV 系统的测量和模型化研究中,研究人员要么利用已有的对 P2P IPTV 系统架构[26]的先验知识搭建一个仿真的 P2P 网络电视平台并开展相关的测量工作,如 Secure-Stream[13], SPoIM[14], StRepS[15-16],要么利用仿真的 P2P IPTV 系统来验证建立的数学模型的准确性[20-21]. 由于 Dhungel 等人[5,19]的测量工作只关注了污染攻击时节点规模变化情况和视频块上传、下载情况 2 类指标,因此本文采用了文献[6]针对 PPLive 网络电视系统实施内容污染攻击获得的真实测量数据与本文模型得出的结果数据进行对比,以验证本文提出的 P2P IPTV 内容污染扩散模型在真实污染攻击环境中的有效性和准确性.

根据文献[6]的测量结果设置本文模型的相关参数如下:初始的健康节点数$G(0)=3\,528$,初始的污染节点数$P(0)=0$,初始的滞留节点数$R(0)=0$,初始离开系统的节点数量$Q(0)=0$,初始污染攻击者数$N_{\mathrm{p}}=1$. 根据文献[6]测量结果显示,在整个污染攻击期间,节点的加入率比较稳定,单位时间平均加入频道覆盖网的节点数量约为 150,因此模型中设置参数$A=150$. 根据文献[28]对 PPLive 的一个流行频道进行拓扑结果测量的结果显示,平均节点度正比于网络的规模,3 500 左右规模的频道的平均节点度约为 16,因此设置模型中参数$d=16$. 对于模型的其余参数,如受污染节点的离开概率ω,利用主动测量研究无法获得受污染节点的数量因此无法获知ω的取值. 对于模型中的这类参数,需通过不断地调整参数值以拟合真实测量的目标曲线,当将相应的参数值设置如下时,模型曲线与测量结果曲线达到了很好的拟合效果:$\delta=0.03$, $\alpha=0.8$, $\beta=0.325$, $\omega=0.325$, $\mu=0.3$, $k=0.145$. 此外,节点的状态转移的时间单元(time unit)设置为 1 min.

对于非线性的微分方程组,直接求其解析解极其困难甚至很可能不存在解析解,因此本文利用 Matlab 工具求解该模型的数值解. 如图 4 所示为本文提出的内容污染扩散模型与真实的污染攻击实验测量结果的对比情况. 其中,Model 曲线描述了利用 Matlab 求解模型获得的频道用户规模$N(t)$的变化趋势,$N(t)=G(t)+P(t)$;Measurement 曲线来源于文献[6]提供的真实测量数据.

对比结果表明,本文提出的内容污染扩散模型与真实测量数据的整体吻合效果良好,可以有效和准确地反映真实的内容污染攻击场景中污染攻击对用户规模的影响. 在污染攻击的前 4 min 里,节点的规模并没有因污染攻击受到影响,而是缓慢的增加.

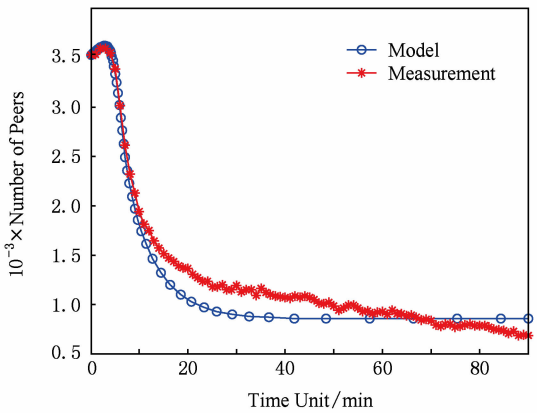


Fig. 4 Model verification: the evolution of overall peers during pollution.

图 4 模型的验证:内容污染攻击对节点规模的影响

这是由于在污染攻击的初始阶段,遭受污染的节点数量较少,健康节点减少得并不明显. 此外,遭受污染的节点首先会选择尝试重新播放该节目或者暂时切换到其他频道等待被污染节目恢复正常,因此该阶段的节点规模没有减少反而增加. 但是,随着污染数据的不断扩散,大量的用户因无法忍受低下的视频播放质量而选择切换到其他节目或者离开系统,因此节点规模呈现快速的减少趋势. 最后,由于用户的持续稳定的加入频道覆盖网络,虽然污染攻击过程一直在持续,但是节点规模在污染攻击期间始终未低于 500 并维持在一个稳定的水平[5-6,19].

为了进一步验证本文提出的内容污染扩散模型的准确性,本文还对比了 Matlab 模拟的节点离开率变化趋势与文献[6]对污染攻击期间用户离开率的真实测量结果,如图 5 所示. 注:模型中处于滞留状态和离开状态的节点均未在频道的覆盖网络中,因此模型中的节点离开率可通过统计单位时间内处于滞留状态和离开状态的新增节点进行计算.

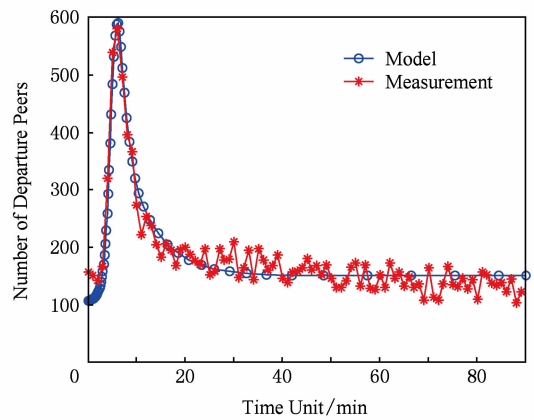


Fig. 5 Model verification: the evolution of departure peers during pollution.

图 5 模型验证:内容污染攻击对节点离开行为的影响

对比结果表明,本文提出的内容污染扩散模型与真实测量数据的整体吻合效果良好,可以有效和准确地反映真实的内容污染攻击场景中污染攻击对用户离开行为的影响. 在内容污染攻击的前 4 min 里,用户的离开行为并未受到内容污染攻击的影响. 从第 5 分钟开始,用户的离开行为受污染攻击的影响开始显现,用户离开率开始快速增加,在大约第 7 分钟时到达峰值,1 min 内约 590 个节点同时离开了该覆盖网络. 之后,用户离开率逐渐降低并趋于稳定.

4 实验及分析

为了更加深入了解 P2P IPTV 内容污染攻击的扩散过程,本文利用 Matlab 软件分别对内容污染攻击中健康节点规模动态演变、污染节点规模动态演变、滞留节点规模动态演变和污染节点比例动态演变 4 个污染指标进行了模拟.

4.1 污染攻击对健康节点规模的影响

如图 6 所示为内容污染攻击中的健康节点规模的变化趋势.

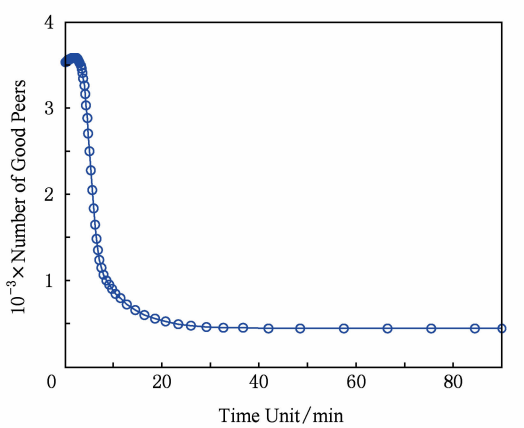


Fig. 6 The evolution of good peers during pollution.

图 6 内容污染攻击对健康节点规模的影响

在污染攻击的前 4 min 里,健康节点的规模同样没有因污染攻击而受到影响,相反节点数量在缓慢地增加. 这是由于在污染攻击的初期,遭受污染的节点数量较少,并且单位时间里遭受污染的节点数量和离开系统的节点数量之和低于加入频道覆盖网络的节点数量. 因此在该阶段,健康节点的规模没有减少反而增加. 随着污染数据在 P2P IPTV 覆盖网络中不断地扩散,大量的健康节点因接收到污染数据而成为污染节点,健康节点的规模呈现快速减少趋势. 实施内容污染攻击 8 min 后,健康节点的数量

减少了约 65%. 之后,健康节点数量下降趋势逐渐放缓,并最终维持在约 440 的稳定水平.

4.2 污染攻击对污染节点规模的影响

如图 7 所示为内容污染攻击中的污染节点规模的变化趋势. 从图 7 可知,在污染攻击的前 8 min 里,健康节点被污染的数量呈现指数增长趋势,在第 8 分钟时达到峰值 1 280 左右;而随后,污染节点的数量呈现快速减少趋势. 这是因为,在污染攻击的前 7 分钟里,存在大量的可供污染的健康节点,污染数据可以快速地蔓延,导致污染节点数量快速增加. 随后由于大量的污染节点因无法忍受系统服务质量的恶化选择了离开系统,致使污染节点的数量逐渐减少. 最终,在污染攻击 30 min 后,污染节点的数量趋于平稳,直至污染攻击结束均保持在 410 左右的规模.

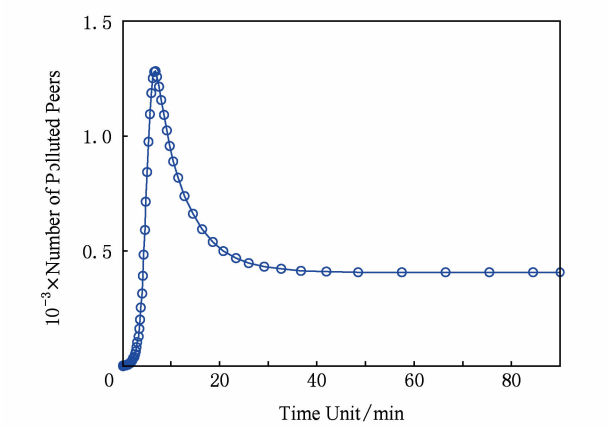


Fig. 7 The evolution of polluted peers during pollution.

图 7 内容污染攻击对污染节点规模的影响

4.3 污染攻击对滞留节点规模的影响

如图 8 所示为内容污染攻击中的滞留节点规模的变化趋势. 从图 8 可知,污染攻击中滞留节点规模的变化趋势类似于污染节点规模的变化,在前 8 min 里同样呈现指数增长趋势. 但是,滞留节点数量峰值的到达时间比污染节点更晚且规模更低,约为 410. 内容污染攻击开始 9 min 后,滞留节点的数量开始逐渐减少并最终稳定在 150 左右.

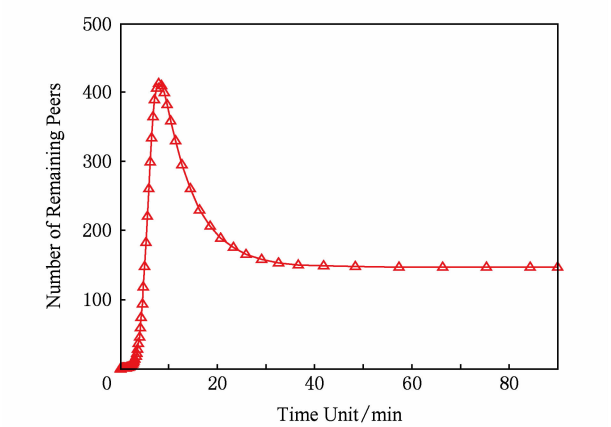


Fig. 8 The evolution of remaining peers during pollution.

图 8 内容污染攻击对滞留节点规模的影响

4.4 污染攻击对污染节点比例的影响

如图 9 所示为内容污染攻击中的污染节点比例的变化趋势. 由图 9 可知,在污染攻击开始的前 8 min 里污染节点的比例快速增加,在约第 9 分钟时到达峰值. 此刻,目标频道覆盖网络中约 52%的健康节点遭受到了内容感染. 随后,受污染节点的比例缓慢下降,并始终保持在 47%左右的水平直至内容污染攻击结束. 这一结果不同于已有的研究工作中污染节点的比例最终将达到 100%的结论[14,21],这是由于他们在建立内容污染扩散模型和搭建模拟的 P2P IPTV 仿真平台式时并没有考虑到用户的加入和离开行为、用户的频道切换行为、用户的重播行为对频道节点规模的影响,而是将频道的节点规模设置为固定数值.

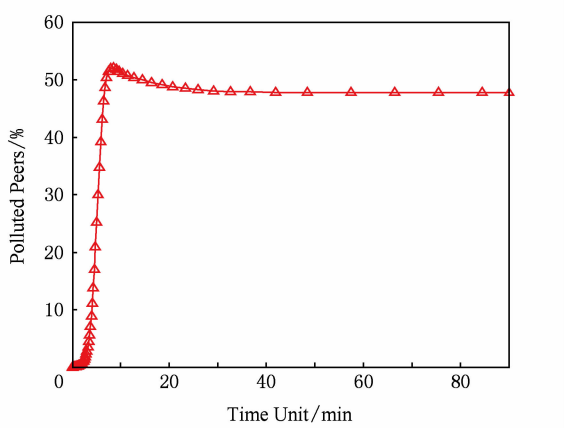


Fig. 9 The evolution of the proportion for polluted peers during pollution.

图 9 内容污染攻击对污染节点比例的影响

5 结 论

伴随着 P2P 技术的发展和日益成熟,基于该技术的 P2P IPTV 应用取得了巨大的商业成功,凝聚了极其庞大的用户群体,深刻影响着互联网的用户行为,吸引了来自全球的产业界和学术界的密切关注. 但是,由于 P2P IPTV 系统自身特有的分布性、匿名性、快速传播性、用户规模大等特点使得其更容易受到攻击且造成的破坏更加严重. 为了深入研究 P2P IPTV 系统在遭受内容污染攻击时呈现的扩散

特征及提出抵御内容污染攻击的有效防御方法,提出并建立一个能准确反映 P2P IPTV 系统内容污染扩散的数学模型具有重要的意义.

本文在对 P2P IPTV 内容污染攻击过程和节点在内容污染攻击中的行为进行详细分析的基础上,建立了一个具有种群动力学特性的 P2P IPTV 系统内容污染扩散模型,该模型充分考虑了在真实世界的内容污染攻击中用户的节目观看行为、用户的加入和离开行为对用户规模动态变化的影响,使得该模型能更好地模拟真实的内容污染场景.为了验证本文提出的 P2P IPTV 内容污染扩散模型在真实污染攻击环境中的有效性和准确性,本文首次将内容污染扩散模型与真实污染场景中的测量数据进行对比.研究结果表明,本文提出的 P2P IPTV 内容污染扩散模型可以有效和准确地刻画真实内容污染攻击场景中污染攻击对频道的节点规模动态演变和节点离开行为的影响情况.此外,为了深入了解 P2P IPTV 内容污染攻击的扩散过程,本文分别对内容污染攻击中健康节点规模动态演变、污染节点规模动态演变、滞留节点规模动态演变和污染节点比例动态演变 4 个方面进行了 Matlab 模拟和详细分析.利用提出的内容污染扩散模型分析 P2P IPTV 系统中增加了防御机制后的内容污染扩散情况以及防御效果将是本文的下一步研究工作.

参 考 文 献

[1] PPTV. PPTV Homepage [EB/OL]. [2014-09-12]. http://www.pptv.com

[2] PPStream. PPStream Homepage [EB/OL]. [2014-09-12]. http://www.pps.tv

[3] UUSee. UUSee Homepage [EB/OL]. [2014-09-12]. http://www.uusee.com

[4] China Internet Network Information Center (CNNIC). The 34th statistical report on Internet development in China [R/OL]. [2014-09-12]. http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/P020140721507223212132.pdf (in Chinese)
(中国互联网络信息中心(CNNIC). 第 34 次中国互联网络发展状况调查统计报告[R/OL]. [2014-09-12]. http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/P020140721507223212132.pdf)

[5] Dhungel P, Hei X, Ross K W, et al. Pollution in P2P live video streaming [J]. Int Journal of Computer Networks and Communications, 2009, 1(2): 99-110

[6] Wang Haizhou, Chen Xingshu, Wang Wenxian, et al. Understanding pollution dynamics in large-scale peer-to-peer IPTV system [J]. Journal of Central South University, 2012, 19(8): 2203-2217

[7] Liang J, Kumar R, Xi Y, et al. Pollution in P2P file sharing systems [C] //Proc of the 24th IEEE Int Conf on Computer Communications (INFOCOM'05). Piscataway, NJ: IEEE, 2005: 1174-1185

[8] Liang J, Naoumov N, Ross K W. The index poisoning attack in P2P file sharing systems [C] //Proc of the 25th IEEE Int Conf on Computer Communications (INFOCOM'06). Piscataway, NJ: IEEE, 2006: 1-12

[9] Lee U, Choi M, Cho J, et al. Understanding pollution dynamics in P2P file sharing [C] //Proc of the 5th Int Workshop on Peer-to-Peer Systems (IPTPS'06). Berkeley, CA: USENIX Association, 2006: 1-6

[10] Christin N, Weigend A S, Chuang J. Content availability, pollution and poisoning in file sharing peer-to-peer networks [C] //Proc of the 6th ACM Conf on Electronic Commerce (EC'05). New York: ACM, 2005: 68-77

[11] Dhungel P, Wu D, Schonhorst B, et al. A measurement study of attacks on BitTorrent leechers [C] //Proc of the 7th Int Workshop on Peer-to-Peer Systems (IPTPS'08). Berkeley, CA: USENIX Association, 2008: 7-7

[12] Shi Jiantao, Zhang Hongli, Fang Bingxing. Study on the countermeasures of BitTorrent fake block attack [J]. Chinese Journal of Computers, 2011, 34(1): 15-24 (in Chinese)
(史建焘, 张宏莉, 方滨兴. BitTorrent 假块污染攻击的对抗方法研究[J]. 计算机学报, 2011, 34(1): 15-24)

[13] Haridasan M, Renesse R. SecureStream: An intrusion-tolerant protocol for live-streaming dissemination [J]. Computer Communications, 2008, 31(3): 563-575

[14] Lin E, de-Castro D M N, Wang M, et al. SPoIM: A close look at pollution attacks in P2P live streaming [C] //Proc of the 18th IEEE Int Workshop on Quality of Service (IWQoS'10). Piscataway, NJ: IEEE, 2010: 1-9

[15] Vieira A B, Campos S, Almeida J. Fighting attacks in P2P live streaming, simpler is better [C] //Proc of the 28th IEEE Int Conf on Computer Communications (INFOCOM'09). Piscataway, NJ: IEEE, 2009: 1-2

[16] Vieira A B, Almeida R, Almeida J, et al. SimplyRep: A simple and effective reputation system to fight pollution in P2P live streaming [J]. Computer Networks, 2013, 57(4): 1019-1036

[17] Hu B, Zhao H V. Pollution-resistant peer-to-peer live streaming using trust management [C] //Proc of the 16th IEEE Int Conf on Image Processing (ICIP'09). Piscataway, NJ: IEEE, 3057-3060

[18] Hu B, Zhao H V. Joint pollution detection and attacker identification in peer-to-peer live streaming [C] //Proc of the 35th IEEE Int Conf on Acoustics Speech and Signal Processing (ICASSP'10). Piscataway, NJ: IEEE, 2010: 2318-2321

[19] Dhungel P, Hei X, Ross K W, et al. The pollution attack in P2P live video streaming: Measurement results and defenses [C] //Proc of ACM SIGCOMM Peer-to-Peer Streaming and IP-TV Workshop (P2P-TV'07). New York: ACM, 2007: 323-328

[20] Yang S, Jin H, Li B, et al. A modeling framework of content pollution in peer-to-peer video streaming systems [J]. Computer Networks, 2009, 53(15): 2703-2715

[21] Yang S, Jin H, Li B, et al. The content pollution in peer-to-peer live streaming systems: Analysis and implications [C] // Proc of the 37th IEEE Int Conf on Parallel Processing (ICPP'08). Piscataway, NJ: IEEE, 2008: 652-659

[22] Li Y, Lui J C S. Stochastic analysis of a randomized detection algorithm for pollution attack in P2P live streaming systems [J]. Performance Evaluation, 2010, 67(11): 1273-1288

[23] Li Y, Lui J C S. On detecting malicious behaviors in interactive networks: Algorithms and analysis [C] //Proc of the 4th IEEE Int Conf on Communication Systems and Networks (COMSNETS'12). Piscataway, NJ: IEEE, 2012: 1-10

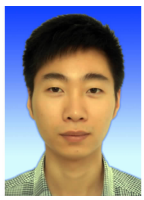
[24] Kang X, Wu Y. Fighting pollution attack in peer-to-peer streaming networks: A trust management approach [C] // Proc of the 27th IFIP Int Information Security and Privacy Conf (SEC'12). Berlin: Springer, 2012: 537-542

[25] Ho C, Chung M, Yen L, et al. Churn: A key effect on real-world P2P software [C] //Proc of the 42nd IEEE Int Conf on Parallel Processing (ICPP'13). Piscataway, NJ: IEEE, 2013: 140-149

[26] Hei X, Liang C, Liang J, et al. A measurement study of a large-scale P2P IPTV system [J]. IEEE Trans on Multimedia, 2007, 9(8): 1672-1687

[27] Jiang Zhihong, Wang Hui, Fan Pengyi. Crawler-based measurement of large scale P2P IPTV systems [J]. Journal of Software, 2011, 22(6): 1373-1388 (in Chinese)
(姜志宏, 王晖, 樊鹏翼. 基于爬行器的大规模 P2P IPTV 测量[J]. 软件学报, 2011, 22(6): 1373-1388)

[28] Chen Xingshu, Hao Zhenghong, Wang Haizhou, et al. Measuring and characterizing topologies of P2P IPTV [J]. Journal of Sichuan University: Engineering Science Edition, 2012, 44(3): 86-94 (in Chinese)
(陈兴蜀, 郝正鸿, 王海舟, 等. P2P 网络电视拓扑测量方法研究与特性分析[J]. 四川大学学报: 工程科学版, 2012, 44(3): 86-94)



Wang Haizhou, born in 1986. PhD and lecturer in the College of Computer Science, Sichuan University, China. His main research interests include peer-to-peer IPTV systems, information security, and network measurement.

Chen Xingshu, born in 1968. Professor and PhD supervisor in the College of Computer Science, Sichuan University, China. Her main research interests include peer-to-peer networks, information security and computer networks.

Du Min, born in 1986. PhD candidate in the College of Computer Science, Sichuan University, China. His main research interests include computer network and pattern recognition, especially in P2P technology.

Wang Wenxian, born in 1978. PhD and lecturer of Sichuan University, China. His main research interests include peer-to-peer networks, information security and trusted computing.