文章编号：2096-1618(2018)02-0155-05

# 一种基于 Merkel 树的 P2P 网络虚假下载片段检测方法

姜文婷， 林少锐
(广东电网有限责任公司电力调度控制中心，广州 广东，510600)

**摘要**：P2P 网络具有开放、匿名、自组织的特点，在为用户提供方便服务的同时，也为网络中的恶意节点提供虚假文件、为发动攻击行提供便利条件。当前的研究多采用信任模型构建的方法，通过以往的交易评估节点的可信程度为节点选择高质量服务、避免不安全交互提供选择依据，虽然可以在一定程度上提升网络安全性能，但计算过程依赖于反馈和推荐信息，对于大规模网络中反馈稀疏的情况评估性能较差，存在大量冗余信息，特别当恶意节点提供虚假反馈，发动共谋攻击、女巫攻击等针对信任模型的攻击时难于应对。为此，提出一种基于 Merkel 树的 P2P 网络虚假点检测方法，利用 hash 值检测文件片段的完整性。一旦检测到虚假节点之后，将其清除于网络，以增强 P2P 网络的安全性。

**关 键 词**：对等网络；虚假节点；Merkel tree；分布式证书
**中图分类号**：TP393 **文献标志码**：A
**doi**：10.16836/j.cnki.jcuit.2018.02.009

## 0 引言

在电力综合数据通信网中 P2P 网络有大量的应用。P2P 网络是依靠用户群交换信息的互联网体系、无中心服务器。网络中，相互连接的多台计算机共享自己的存储能力、处理能力、网络连接能力等以及各种硬件资源，如打印机、磁盘空间。整个网络不依赖专用的中心集中式服务器，可以不经过中间实体实现各个节点之间共享资源的直接访问。这些对等节点既可以向其他节点发出请求信息，获取网络资源、服务以及内容，又可以提供资源共享、信息交换等服务，以响应其他节点的请求。任何一个节点无法直接连接到其他节点，必须依靠用户群进行数据传递。

P2P 网络的基本结构包括如下几种：

(1)集中式对等网络。集中式对等网络为网络中各个节点提供目录查询服务，结构较为简单，基于中央目录服务器，由于数据传递无须再次经过中央服务器转发，因而显著降低中央服务器的负荷。但是，集中式对等网络仍保留中央节点，容易形成传输瓶颈，面临着 DOS 拒绝服务攻击以及过量存储负载问题，而且，网络带宽的限制导致了集中式对等网络的扩展性不强，不适合大型网络。

(2)无结构分布式网络。在无结构分布式网络中，不存在中央服务器，所有参与的结点接入整个网络的方式是通过邻近节点之间的数据传递，通过发送查询包到其他节点的机制在网络中搜索需要的资源。主要方法是，请求节点向相邻节点发送包含查询内容的查询包，相邻的节点收到后继续转发查询包给相邻的节点，以扩散方式在网络中传播，为防止消息泛滥，一般设置一个适当的生存时间(Time To Live)，生存时间随着查询时间的推移逐渐降低，当生存时间的值降为 0 时，停止发送查询包。在这种无结构分布式网络中，节点的进出较为自由，存在组织方式不够集中，稀疏性大的问题，可以迅速找到热门内容，对于冷门内容，小的生存时间情况下难以找到，如果设置较大的生存时间，查询将消耗大量流量。特别是，当网络发展到较大规模时，小的生存时间仍会导致网络流量的大幅增长。但是，持有丰富资源的类服务节点在网络中，可以大大减少查询时间。

(3)结构化分布式网络。结构化分布式网络基于分布式哈希表(distributed hash table)技术，实现整个 DHT 网络的寻址和存储。在不需要服务器的情况下，分布式哈希表采用 Hash 方式对数据分片，即将分片后的数据根据哈希函数映射到某台机器(节点)上存储，节点负责一个小范围的路由，各自管理自己的散列块，负责存储一小部分数据。由于，分布式系统中，经常存在节点的宕机、某个节点加入或者移出集群，因而使通常采用一致性哈希函数对机器和数据进行统一运算，避免使用普通哈希带来的扩容造成元素位置移动的问题。当用户发出查询请求时，节点首先检查本地是否有待查数据，如果不在本地，则从自己的路由表中，找

**收稿日期**：2018-02-27
**基金项目**：广东电网科技资助项目(036000KK52170002)；国家重点研发计划资助项目(2016YFB0901200)

一个和数据 id 距离最近、并且存活在网络中的节点 next。如果该节点的 id 和数据 id 相等,那么为目的节点;如果不相等,则到 next 进行递归查找。找到哈希表中对应的存储节点后,获得所需资源的地址信息,然后与对应节点建立连接,进行通信。一致性哈希有多种实现算法,经典算法有 Chord 算法、KAD 算法(kademlia)。这种结构化分布式网络,可扩展性高,结构性强,适合规模较大的网络。

由于 P2P 网络中,每个节点自愿的、随机的,动态的,以匿名的方式进行通信或文件资源的共享,因此可能会受到恶意入侵和破坏。通常 P2P 网络的安全攻击包括:

(1)中毒攻击。节点提供内容与描述不同的文件,共享夹杂计算机病毒的文件,肆意破坏网络。

(2)拒绝服务攻击。节点发送大量伪造的 TCP 连接请求,使被攻击方耗尽资源,使网络运行缓慢,甚至彻底崩溃。

(3)背叛攻击。节点使用网络却没有提供自己的资源。

(4)共谋攻击。多个恶意节点团结起来,向其他诚实节点提供虚假信息,进行不正当交易,以获取其他诚实节点的信任。

(5)女巫攻击。恶意节点声称自己有多个身份,并欺骗信誉系统,使其相信它拥有不成比例的巨大影响力。攻击者通过创建大量的假名标识破坏对等网络的信誉系统,使用它们获得不成比例的大的影响。

(6)节点故障。当某个正常节点路由故障或路由表错误时,向这个节点发送的请求资源长期得不到相应,被网络判定为恶意节点,被隔离。如何区分正常故障节点和恶意节点是一个难题。

## 1 相关工作与存在的问题

如何使诚实的节点在未知网络环境中,有效地甄别恶意节点,减少 P2P 网络中恶意攻击的影响,规避风险,提供一个安全、公平的文件共享下载环境,提升 P2P 网络的安全性是 P2P 系统面临的难题之一。

### 1.1 信任模型的构建

当前的研究多采用节点间建立信任模型的方法,通过评估节点的可信程度为节点选择高质量服务,避免不安全交互。通过节点以往的交易信息及参与交易的节点的反馈信息,其他节点通过一定的计算、评估,进而对此节点进行信任值判定,网络中的节点都可以获得这个信任值。在多个节点提供相同服务的情况下,首先选择信任值高的节点,设有激励机制,激励节点采取有利于网络通信的行为,以增强 P2P 网络的安全性、稳定性。节点在 P2P 网络中表现不合法行为也作为影响信任值的因素,此外,网络、存储、计算、带宽等特性也作为信任值的考量。

### 1.2 现有信任模型

基于证书和策略的信任管理:节点之间的信任关系通过证书建立,根据定义的策略限制对资源的访问。典型的信任模型有 REFEREEt、PolicyMaker、SPKI/SDSI、KeyNote。

基于声誉的信任管理:节点与其他节点建立信任关系,通过与其他节点之间的交互行为,赋予一定的信任值,信任值包括节点的全局声誉以及局部声誉。典型模型有 Jiminyt、HISTOS、NICE、DMRep、TrustMel、GossipTrust、PACEt、REBCON、RunTrus、PowerTrust、PeerTrust、SuperT′mst、EigenRep、P2Prep 和 RETM 等。

基于社会网络的信任管理:利用节点之间的社会管理机制,通过分析团体关系形成对节点的评价。

目前上述研究工作存在的问题包括:

(1)节点身份的认证。节点以匿名加入网络,进行通信,判断出虚假节点后,没有有效的机制将其清除于网络,没有得到惩罚。

(2)信任值的有效度量。现有的信任模型仅仅通过节点的交易次数和历史交易情况计算信任值是不可靠的,由于评价是用户对交互过程的主观评测,具有不可控性,信任值是一个复杂、主观的概念,受许多因素影响。

(3)信任值的可靠存储。信任值在信任管理技术中充当关键因素,决定了信任管理技术是否有效,因此安全存储信任值是重要的环节,并且需要防止它被泄露、篡改和利用,而且信任值是动态的。现有的方案有存储在硬件设备中,对设备的性能要求较高,依赖设备的安全性。目前对于信任值的分布式存储还没有公认的解决方案。

(4)难以抵抗共谋攻击。恶意节点,互相夸大,利用多次无关紧要的交易中表现诚实,从而赢得很高的信任值,利用这种高信任值,在与其他诚实节点的交易中进行欺骗,对其他节点提供虚假推荐达到抬高或诋毁其他节点的目的,从而颠覆该信任模型。

(5)存在数据冗余。信任值的计算过程依赖于反馈和推荐信息,对于大规模网络中反馈稀疏的情况评估性能较差,存在冗余信息。

为增强网络提供服务的高效性和可靠性,在下载机制中需要考虑数据完整性的验证、节点恶意行为的识别、恶意节点的加入控制,节点提供可靠下载服务的可行度评估。

## 2　基于 Merkel Tree 的虚假片段检测

### 2.1　P2P 网络虚假片段检测方法

P2P 对等网络中每一个节点,关系对等,每个节点既充当资源的发布者,也充当资源的接收者,节点间可以直接进行连接和数据通信,可以自由地加入或者离开网络,并且网络的特性不因为单个节点的加入、离开而受到改变。P2P 网络无中心服务器的参与,任意两个节点之间可以直接进行数据传递。随着节点数量的急剧增长,P2P 网络可以提高资源分享的速率,而不带来网络的拥塞。P2P 网络目前普遍应用于文件下载中,例如:Napster, Gnutella, BitTorrent 和 eMule 等系统。

P2P 网络中,文件被分解成片段到不同点节点,这使网络服务器带宽不会限制节点的最大下载速度,使服务端不用担心用户数量过多带来带宽限制的问题,网络中参与下载的用户越多,下载速度越快。要下载必须从中心索引服务器获取一个种子文件,种子文件具有索引的功能,文件名的长度和文件的哈希值存储其中,以及一个指向中心索引服务器的 URL。种子文件中的哈希值是每一块要下载的数据的消息摘要,摘要里面包含了时间戳,数字签名,节点证书等信息,在下载的时候进行验证。

P2P 网络下载片段的步骤为:

(1)在线的节点 A 想要通过 P2P 网络下载文件,在线节点 A 向 tracker 服务器发送请求需要下载的文件,tracker 服务器随即发送给 A 种子文件。A 从中知道哪些节点拥有需要的文件片段,这些节点称作在线共享方,从而建立连接。

(2)Tracker 服务器向在线共享方 $B_1$,$B_2$,$B_3$, …,$B_n$ 发出询问,表明想要下载共享片段的起始位置 S,结束位置 E。

(3)在线的共享方根据收到的询问 REQ,计算起始位置到结束为止片段的 *hash* 值,并返回该值给 A。

(4)节点 A 收到来自不同在线共享方的哈希值,如果共享方 B 是可信的,那么计算出的 *hash* 值应当是相同的(拜占庭将军问题,忠诚的将军做出的选择是一样的),因此 A 从收到的 *hash* 值中,选择出项次数最多的 *hash* 值对应的一组共享方 B。

(5)在线的节点 A 从已经选出的在线共享方 Set B 中确定一个共享片段下载的节点。从 Set B 中选择一个距离最近的一个,传输速度最大的一个,带宽最大的一个,即最先收到 *hash* 答复的那个 $B_i$,进行片段的下载。

(6)最后检索数据块,下载完所有共享片段后,检测是否下载了虚假片段。

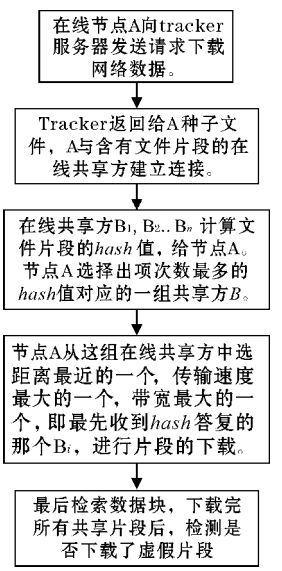


图 1　P2P 网络虚假节点检测流程图

### 2.2　Merkel 树的构建过程

首先,P2P 网络 tracker 服务器需要构建每个文件的 Merkel 树。构建步骤如下:

(1)将一个文件等分成 $n$ 个片段,每个片段长度为 $x$,$x$=FileLength/$n$,分别计算每个片段的 *hash* 值,作为初级子 *hash* 值。*hash* 值计算函数可以是 MD5,SHA1 等。

(2)将得到的初级子 *hash* 值与相邻的 *hash* 值串联合并成一个字符串,然后运算这个字符串的哈希,得到了一个“子哈希”。

如果最底层的哈希总数是单数,直接对它进行哈希运算,最后一个初级子 *hash* 值单独形成新的网络数据片段。

(4)相同的方法向上推导,获得数目更少一级的哈希,最终自下而上构建出一颗哈希树,树的底层是 $hash=(\mathrm{File}[aX,(a+1)X],a=0,1,2,\cdots,n-1$。到了树根的这个位置,这一代就剩下一个根哈希 Merkel Root。Merkel 树构建成功。

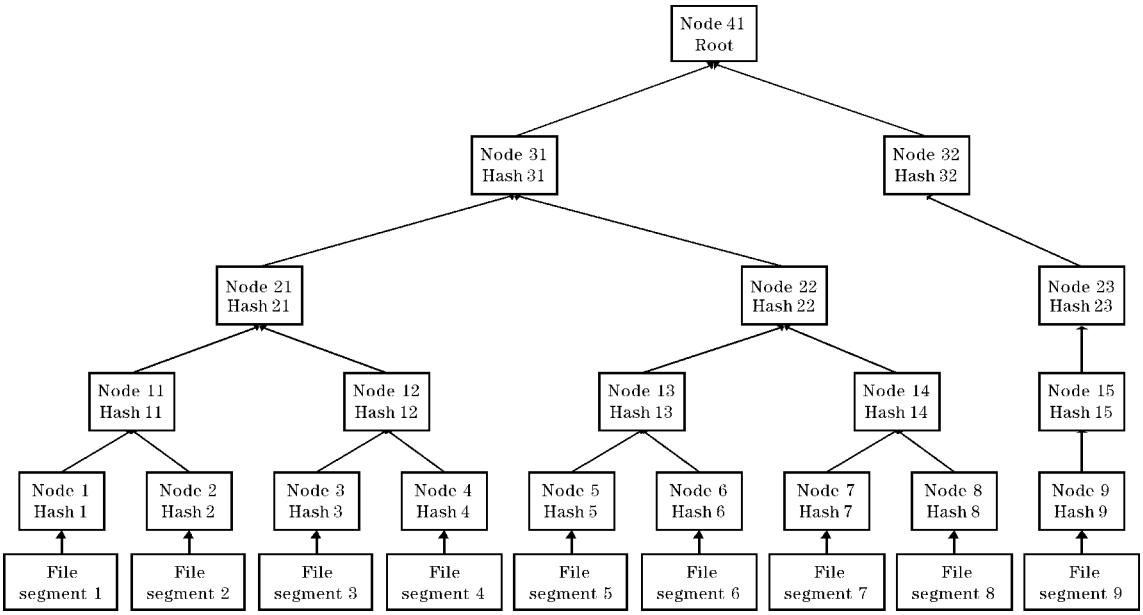


图 2 文件片段 hash 值构建 Merkel 树的过程

## 2.3 检测虚假片段的过程

请求节点 A 下载完所有文件片段后,可以构建文件的 Merkel 树,将 Merkel 树与来自可信源 tracker 服务器的该文件 Merkel 树进行对比,检索是否有虚假数据块。

(1)请求下载文件的节点 A 将获得的网络数据 hash 值列表构建 Merkel 树。

(2)将构建好的 Merkel 树的根信息与来自 tracker 服务器的根做比较,如果比较结果相同,则判定此次下载过程中,无虚假节点

(3)若比较结果不同,则继续向下比较子节点 hash 值,直到找到相异的初级子 hash 值,根据数字签名,找到提供该片段的节点,判定为虚假节点。

(4)检索比较完毕。

以上过程的理论复杂度是 Log($N$)。$n$ 为 Merkel 树的叶子结点个数,即文件片段个数。

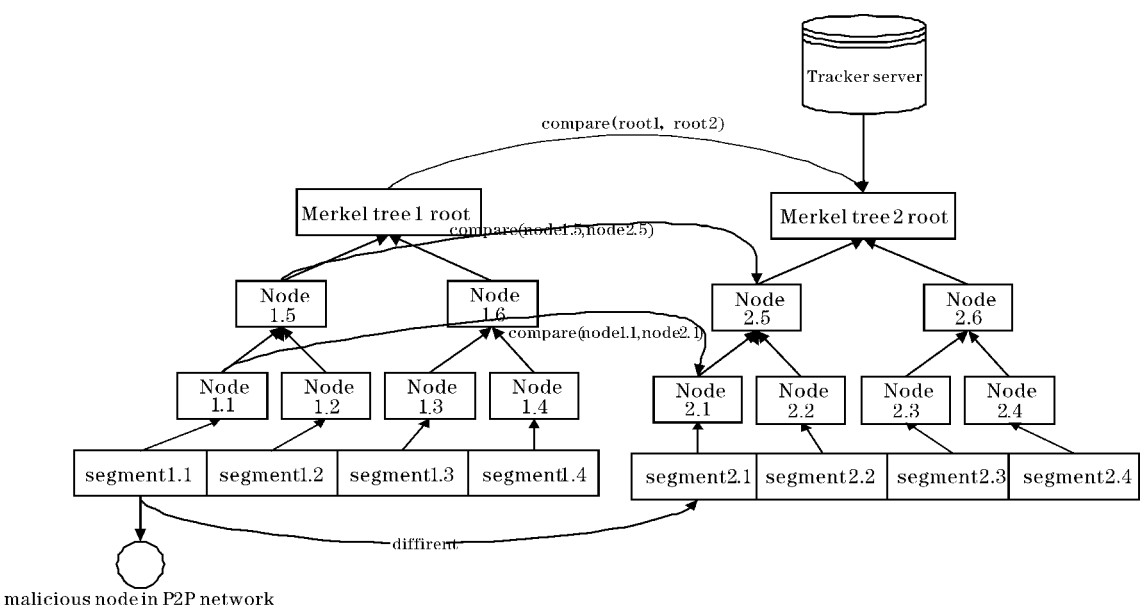


图 3 虚假片段的检测过程

## 2.4 分析

利用 Merkel 树来进行文件完整性校验,虚假节点检测会带来时延。首先是时间开销,MD5,SHA1 哈希算法的时间复杂度通常为 $O(n)$,时间开销随着文件块字节数的增大而增大。其次,是传输时间开销。在

文件完整性校验的每一个分块校验过程中,如果应用程序暂停接收数据时间过长,发送节点将接受节点从发送服务队列中移除,甚至断开节点的传输连接,最终使节点必须再等一段时间后,才能够继续获取发送节点的服务。因此,如果平台计算能力较差、文件太大、哈希计算的时间与文件下载的时间较长时,文件完整性校验、虚假节点检测将大大影响数据传输的效率。

可信中心索引服务器存储了每个文件片段的 Merkeltree,一旦遭受攻击,整个网络将瘫痪。信任中心难以部署,交易签名和节点认证需要耗费大量计算机资源,节点要求具有强大的计算能力,存在一定的局限性。

方案还可以如下增强:给系统中的每个节点一个证书标识节点,使其具有合法的网络身份,并且可以被系统中的其他网络节点识别,保证终端的可信度;节点之间的交易信息都有一个时间戳和唯一的密码签名,使 P2P 网络中的交易具有可审计性;检测机制可以成功识别、防备恶意节点传播的虚假信息或错误的推荐信任消息,并且立即采取相应的有效措施将其清除,提高了系统的健壮性与容错性。

## 3　结束语

对 P2P 网络中现有的节点信任模型存在大量冗余信息,且难以应对共谋攻击,无法检测虚假片段文件的问题,提出一种基于 Merkeltree 的网络虚假片段检测方法增强 P2P 网络的安全性。该方法基于 Merkel tree,具有很好的可靠性和稳定性。

## 参考文献:


[1]　韩磊磊. P2P 网络的恶意节点检测模型[J]. 计算机工程与设计,2001(2):484-485.

[2]　刘凤鸣. 基于对等网络的带虚假反馈检测的声誉系统[D]. 湘潭:湘潭大学,2006(5).

[3]　宁晓莉,黄遵国. DHT 网络中并发下载及安全防御机制的实现[J]. 计算机工程与科学,2008,30(1).

[4]　贺鹏程,王劲林,邓浩扛,等. P2P 文件完整性校验延迟隐藏算法[J]. 计算机工程,2010,36(15):29-31.

[5]　胡玲. 可信计算技术在 P2P 网络信任模型中的应用研究[D]. 南京:南京邮电大学, 2011.

[6]　冯景瑜,张玉清,陈深龙,等. P2P 声誉系统中 Good Rep 攻击及其防御机制[J]. 计算机研究与发展,2011,48(8):1473-1480.

[7]　王勇,侯洁,白杨,等. 基于反馈相关性的 P2P 网络信任模型[J]. 计算机科学,2013,40(2):103-107.

[8]　廖歹法,文朝阳. P2P 网络中基于模糊评判的推荐信任模型[J]. 计算机工程与设计,2014,35(4):1183-1187.

[9]　任爽. 基于 FP-Outlier 挖掘的 P2P 网络恶意节点检测模型[D]. 大连:大连理工大学,2015.

[10]　唐伯浩. P2P 文件共享系统中信任管理机制研究[D]. 长春:吉林大学,2016.


# A Malicious Download Fragment Detection Method for P2P Network based on Merkel Tree


JIANG Wen-ting,　LIN Shao-rui

(Guangdong Power Grid Co., Ltd. Power Dispatching Control Center, Guangzhou 510600, China)



**Abstract**: P2P network provides the conditions for the malicious nodes to provide fake documents and start its attack while provides a convenient and efficient service for users since it has the features of open, anonymous, self-organization. Current most researches use the method of trust model, through the credibility of the transaction evaluation node to node selection of high quality service, avoid unsafe interaction and provide basic data, although can improve the network security performance to a certain extent, but its calculation depends on the feedback and recommendation information for poor performance evaluation of sparse feedback in large-scale network, there are a lot of redundant information, especially when malicious nodes provide false feedback, to launch attacks against collusion attack, witch attacks are difficult to deal with the trust model. To solve the above problem, this paper proposes a solution that is detect and verification the malicious nodes based on Merkel tree. Once malicious node is detected, exclude the node out of the P2P networks in order to enhance the security of P2P network.



**Keywords**: P2P network; malicious nodes; Merkel tree; distributed authentication