基于投票的 P2P 文件真实性认证协议*)

李剑宇 胡 磊 鲁 力 汪维家
(信息安全国家重点实验室(中国科学院研究生院) 北京 100049)

摘 要 提出了一种基于电子投票的文件真实性认证协议。需要验证文件真实性的发起者将投票请求发送给多个代理节点,由代理节点泛洪(Flood)投票并收集结果返回给发起者。发起者根据各选区的非重复选票确定文件的真实性。交互过程中使用洋葱路由和路由交换表,实现了发起者匿名、投票者匿名和文件提供者匿名。理论分析和实验表明,用该协议进行文件真实性认证,成功率在95%以上。与已有文件真实性认证协议比较,交互次数大大减少,并提供了参与者匿名。

关键词 P2P 系统,文件真实性,匿名,电子投票,洋葱路由

A Voting-Based File Authentication Protocol

LI Jian-Yu HU Lei LU Li WANG Wei-Jia
(State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences, Beijing 100049)

Abstract An electronic voting-based file authentication protocol(EV-FAP) is presented. The initiator asks a number of proxies to flood an electronic voting and return the tickets collected. The initiator tallies returned tickets and determines that file is the authentic copy of the document. The protocol archives initiator anonymity, voter anonymity and provider anonymity by utilizing means of onion routing and switch route table. Theoretical analysis and experiential simulation show that the protocol can successfully authenticate the authenticity of the file searched with a probability exceeding 95%. Compared with existing file authentication protocols, it reduces communications and provides anonymity of participants.

Keywords P2P system, File authenticity, Anonymity, Electronic voting, Onion routing

1 引言

对等网(Peer-to-Peer Network,以下称为 P2P) 使用户可以自主地进行文件共享与交换。系统中的每个节点都是对等的,既充当服务器又充当客户机,并具有高度的自主性(Autonomy)。P2P 把网络中各节点的计算、存储、文件等各种资源集中起来为所有用户服务。相对于传统的 C/S 模型,P2P 具有灵活性、高动态性、高可扩展性和高健壮性等优点。目前,P2P 流量已占据 Internet 流量相当大的部分,例如在 2002 年,P2P 流量占据了校园网 43% 的带宽,而 WWW 只占 14%[1]。

随着 P2P 应用范围的不断扩展,P2P 服务的安全性与可靠性受到极大的关注[2]。其中文件真实性认证是 P2P 系统重要的安全服务之一。文件的真实性(Authenticity) 是指:用户获得的文件是未经篡改的副本。而文件真实性认证是判定文件真实性的过程。举例来说,如果一个节点发起一个"物种起源"的查询,结果收到 3 个应答:应答之一可能恰好是达尔文所著的《物种起源》;另外的一个应答可能是修改了几个关键段落的《物种起源》;而第三个应答可能是相信上帝造物论的基督徒的作品。判断这些应答中哪一个是真实文件的过程就是文件真实性认证[3]。

Damiani 提出了基于用户的信誉度(Reputation) 来确定 P2P 文件真实性的机制[4,5]。其具体过程是对文件提供者的信誉度发起投票,并根据投票结果来选择一个文件提供者,确认其身份后向其发出下载请求。该机制能有效地挫败假冒用户攻击。但该协议需要信誉机制的支持,因此要求发起者、投票者和文件提供者在交互中提供自己的身份信息,这不符合匿名的要求。该协议从发起查询到请求下载需要 8 次交互,这是一个很大的开销。文[6]给出了一种在各图书馆之间以投票的方式验证电子学术期刊真实性的协议。在该协议中,发起者通过让投票者提供某种与身份相关的证据和动态地改变投票者的方式来防止恶意用户的扰乱、假冒和磨损攻击。但该协议存在两个主要问题,首先投票者生成证据需要消耗较多的计算资源,这在有组织的各图书馆之间是可行的,但在无中心的 P2P 系统中,大多数用户并不愿意因为参加义务投票而消耗太多的计算资源。另外,该协议同样也不能实现参与者匿名。

本文提出一种基于电子投票的文件真实性认证协议,称为 EV-FAP(Electronic voting-based file authentication protocol)。其主要思想是发起者以洋葱路由(Onion routing)[7] 给多个代理发送投票请求,由代理泛洪投票并收集投票结果返回给发起者。发起者根据投票结果确定文件的真实性。与文[4~6]等文件真实性协议比较,本协议的显著优点是交互次数少,并且实现了发起者匿名、投票者匿名和文件提供者匿名。通过理论分析和仿真实验,本协议的文件真实性认证成功率在 95%以上。另外,本协议适合于两种情况下的文件真实性认证:(1)发起者已经有某个文件,通过投票验证此文件是否为真实副本;(2)发起者没有要求投票的文件,通过投票

*)国家自然科学基金(60373041)资助项目。李剑宇 硕士研究生,讲师,研究方向:网络安全、系统仿真等。

确定文件的真实副本，然后选择下载。

本文第 2 节给出关键概念和符号，第 3 节描述 EV-FAP，第 4 节分析协议的安全性和效率，第 5 节为仿真实验结果，最后总结全文。

2 相关概念和符号

在阐述协议之前，先介绍协议中涉及到的一些关键概念和符号。

2.1 相关概念

(1)洋葱(Onion)[7] 是一种递归的层次型数据结构，每一层依次指定路径上各对等节点的连接属性(如下一跳节点的 ID 等)，并依次用路径上各节点的公钥加密。洋葱路由(Onion Routing)是这种数据结构所描述的路径。路径上的节点称为 Onion Router，路径上每个节点收到洋葱后，用它的私钥解密最外层，得到下一跳节点的 ID 和内置的洋葱等。然后它对内置的洋葱进行填充以保持固定长度，并将其发送给下一个对等节点。经过路径上节点的层层解密，当到达最终节点(指定的接收者)时，接收者可以解密成明文的形式。洋葱路由可以确保发起者匿名，而发起者知道洋葱所经过的路径上所有节点的 ID 及公钥，当然也知道接收者的 ID 及公钥。图 1 描述的是一个三层洋葱的路由，发起者 $P1$ 构造洋葱

$Package_{12}=\{P_3,\{P_4\{PlainText\}K_{P4}\}K_{P3}\}K_{P2}$

并发送给 P_2(其中$\{\cdots\}K_{Pi}$表示用 P_i 的公钥 K_{Pi} 对$\{\cdots\}$中的内容进行公钥加密)，P_2 对收到的洋葱用自己的私钥解密得到下一跳的 ID P_3 和

$Package_{23}=\{P_4,\{PlainText\}K_{P4}\}K_{P3}$

P_2 填充 $Package_{23}$ 到固定长度并送给 P_3，P_3 收到洋葱后用自己的私钥解密，得到下一跳的 ID P_4 和 $Package_{34}=\{PlainText\}K_{P4}$，$P_3$ 填充 $Package_{34}$ 到固定长度并发送给 P_4，P_4 用自己的私钥解密得到明文。

路由上所有对等节点从上一级节点得到洋葱，解密并填充后按路径的指示发送到下一个节点。它们都无法判断真正的发起者与接收者。同样，接收者也不知道谁是真正的发起者。由于所有节点都在发送与接收洋葱，监听者无法判断某个洋葱的起始出发点与终点。如果路由节点对接收的多个洋葱搅拌后再发送，那么监听者就更无法跟踪洋葱的原始路由。

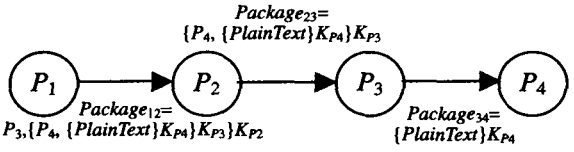

图 1 洋葱路由

为了叙述方便，我们称描述了从起点 P_1 到终点 P_n 完整路由的洋葱为从 P_1 到 P_n 的洋葱。

(2)真实文件分布率 即某一文件 F 的真实副本所有者与在线用户的比率。

2.2 符号定义

CCI 加密控制信息。包括加密算法和密钥；

EV-FAP 本协议的名称。Electronic voting-based file authentication protocol，基于电子投票的文件真实性认证协议；

F 当选文件。计票结束，获得票数最多的文件；

fakemix 伪装信息。用于填充在 *RO* 内部，使洋葱路由上的节点无法猜测下一节点是否为终点；

H_R 投票者 R 对自己拥有的候选文件所做的 Hash。

H_t $H_t=\mathrm{Hash}(sq\|H_R\|NetValue\|\mathrm{IP})$。即对 sq，HR，$NetValue$，IP 的级联做 Hash 函数运算。可选任意的安全 Hash 函数，如 SHA-1；

I 发起者。要求进行文件真实性认证的节点；

NetValue 投票者给出的自身的网络状况信息。如：网速 *(到达包的 TTL 值 + 1)；

P 代理节点，发起者选择的为他泛洪投票的节点；

query 指出被投票的文件。可以是文件名；

R 投票者。当空白选票经过时，对选票中的文件进行投票的节点；

RO[8] ReturnOnion 回馈洋葱。由洋葱路由的终点构造的一种洋葱；

sq 投票唯一标识符。发起者选择的随机数；

SRT 路由交换表(Switch route table)。保存在投票(查询)路由节点上的二维表格。有 5 个字段：sq，$from$，to，H_t，$time$。sq 记录了最近泛洪(Flood)过的投票的唯一标识符，$from$ 记录包的来源 IP，to 记录包的去向 IP，H_t 为返回 $Ticket$ 中包含的 H_t，$time$ 记录投票的到达时间，以便在超时后删除本记录；如表 1 所示。

表 1 路由交换表

Sequence	*From*	*To*	*Ht*	*time*
sq1	202. 38. 6. 59	210. 77. 22. 15	-	22:14′20″
sq2	61. 208. 74. 3	219. 56. 84. 24	Hti	22:15′45′
……	……	……	…	……

t 选区号。每个代理泛洪投票所涉及的范围称为一个选区。T 是由发起者 I 选择并发送给代理的随机数，不同的 t 标识不同的选区；

Ticket 选票。$Ticket=\{sq,\{HR,NetValue\}K,Ht\}$；$X$，$Y$ 投票(查询)路由上的节点；

γ 当选文件必须达到的最低得票数，如取 $\gamma=2$；

λ 网络状况的阈值。

3 EV-FAP 方案

本节给出 EV-FAP 的具体方案。在描述方案之前，我们先提出文件真实性认证的基础假设。

3.1 文件真实性认证的基础假设

Daswani 等提出了文件真实性认证这一公开问题[3]，并列举了 4 种不同的判断文件真实性的标准：

(1)最古老的文件是真实的。这种定义认为最早提交到系统中的文件是真实副本；

(2)基于专家策略确认文件真实性。按照这种方式，一份文件如果经过了专家或权威节点的真实性确认，就被认为是真实的；

(3)基于投票策略确认文件真实性。为了解决单点失效问题，用投票的方法让更多的专家对文件的真实性进行确认，如果大部分专家认同，就认为文件是真实的；

(4)基于名誉确认文件真实性。一些专家可能比另一些专家更可信，可以考虑在投票中加大可信专家选票的权重。权重由名誉机制来提供，因此在这种方案中要求能保持、更新和传播名誉值。

上述标准中，(3)不需要(1)使用的时戳，不存在(2)的单

点失效问题，也避免了(4)需要确定投票者权重的名誉机制，因而可以用很小的代价实现文件真实性认证。因此本协议使用(3)作为文件真实性认证标准。现实生活中，当一个问题有多种选择而不能确定哪一个更明智时，常采用参与者投票的方法进行选择。需要投票确定一个文件的真实性时，投票的参与者就是这个文件副本的所有者。用户下载、保存并共享一个文件说明了用户对该文件有一定的了解和认可，因此可以参与投票。

本协议的设计基于以下3点合理性假设：

(1)绝大部分的P2P用户是善意的。与现实世界类似，在P2P环境中，恶意的用户只占很小的一部分；

(2)用户会删除他认为不真实的文件；用户发现下载了一个不真实的文件，不会保存起来占用宝贵的存储空间；

(3)在整个P2P系统中，被投票文件的真实副本的数量要比不真实副本占优势。

3.2 协议描述

下面以投票者没有要求投票的文件为例描述协议。投票过程实际上也是查询过程。

协议的基本思想是：当发起者需要某一文件时，他以洋葱路由发送请求给多个代理。代理以泛洪(Flooding)的方式发送投票，收集投票结果并通过回馈洋葱路由 RO 返回给发起者。发起者根据投票结果确定当选文件，并且可要求代理索取文件并中转给自己。协议如图2所示，具体过程由4个步骤组成。

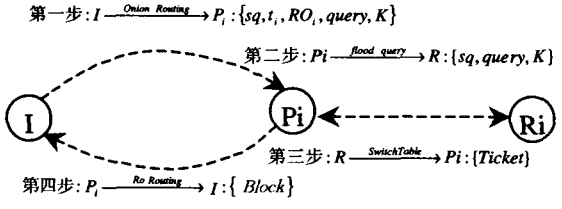

图2 投票过程示意图

(1)请求代理。当发起者 I 需要文件 F 时，请求代理泛洪投票。他将 m 个选区号 $t_1, t_2, \cdots, t_m$ 以洋葱路由分别发送给 m 个代理 $P_1, P_2, \cdots, P_m$，并确保发起者匿名。

具体操作如下：发起者 I 随机选择 m 个节点 $P_1, P_2 \cdots, P_m$ 作为代理节点，随机选择一个投票唯一标识符 sq 和 m 个选区号 $t_1, t_2, \cdots, t_m$。并构造从 $P_i(i=1, \cdots, m)$ 到 I 的 m 个回馈洋葱 RO_1，$RO_2, \cdots$，RO_m（RO 的结构将在(4)中详细描述）。I 将 sq、选区号 t_i 和洋葱 RO_i 等打包为从 I 到 P_i 的洋葱，然后发送出去。第 $i(i=1, \cdots, m)$ 颗洋葱中包括了{sq，t_i，RO_i，$query$，K}，其中 $query$ 指出要求投票的文件，可以是文件名或关键字；K 为 I 的公钥。该过程表示为

$$I \xrightarrow{Onion\ Routing} P_i:\{sq, t_i, RO_i, query, K, TTL\},$$

其中 $i=1, 2, \cdots, m$

(2)泛洪(Flood)投票。洋葱到达代理节点 P_i 后，P_i 打开洋葱的最后一层，得到{ sq，t_i，RO_i，$query$，K，TTL }。P_i 生成选票{ sq，$query$，K，TTL}，然后选择它的 n 个邻居泛洪出去。对等节点 Y 收到邻居节点 X 发送来的{ sq，$query$，K，TTL }后，Y 在自己的路由交换表中查找是否已经存在标识符为 sq 的记录：如果存在，则将包丢弃；如果不存在，则组织记录(sq、来源节点 X 的IP、查询的去向IP和时间)并存入路由交换表中，然后查找自身是否有 $query$ 指定的文件 F；如果没有，将 TTL 值减去1。若 TTL 不为零，则泛洪{ sq，$query$，K，TTL}。

(3)返回选票。如果 Y 有 $query$ 对应的文件 F，他将参与投票，我们称之为 R。R 对文件 F 的内容作Hash，得到 H_R，并计算 $H_t = Hash(sq \parallel H_R \parallel NetValue \parallel IP)$。然后构造 $Ticket=\{sq, \{H_R, NetValue\}K, H_t\}$，将 $Ticket$ 通过路由交换表沿原路径返回给 P_i。返回路径上的节点将 Ht 添加到路由交换表的对应记录中，以便发起者 I 以后能向 R 索取文件。

(4)代理反馈。P_i 将收到的所有 $Ticket$ 封装为 $Block_i = \{t_i, Ticket_1, Ticket_2, \cdots, Ticket_{n'}\}K$，按照 RO_i 所指示的路由返回给 I。

$$RO = \{V_1, CCI_1, \{V_2, CCI_2, \{V_3, \cdots \{I, \{sq, fakemix\}K\}K_{Vs}\cdots\}K_{V2}\}K_{V1}\}K_P,$$

其中 CCI 包括加密算法和密钥，指出当前节点对外部的数据包 $Block$ 使用的加密算法和密钥。注意到 RO 也是一个洋葱，不过这个洋葱不是由洋葱路由的起始节点构造的，而是由终点构造的，因此称之为回馈洋葱。这使代理节点无法知道发起者 I 的任何信息。与普通洋葱的另一个区别是在 RO 的内部放入的是 $fakemix$ 而不是要传送的信息；要传送的信息是 $Block$，不包含在 RO 中。

回馈洋葱路由上的节点 V_k 收到 $RO_k = \{V_{k+1}, CCI_k, RO_{k+1}\}K_{Vk}$ 和 $Block_{k-1} = E_{k-1}\{\cdots E_1\{E_0\{Block\}\}\}$ 后，V_k 用私钥解密 RO_k，得到 V_{k+1}、CCI_k 和 RO_{k+1}，根据 CCI_k 加密 $Block_{k-1}$，得到 $Block_k = E_k\{E_{k-1}\{\cdots E_1\{E_0\{Block\}\}\}\}$，对 RO_k 进行填充后与 $Block_k$ 一起送给节点 V_{k+1}。如此传递下去，最后节点 I 收到{sq，$fakemix$}K 和 $Block_s = E_s\{E_{s-1}\{\cdots E_1\{E_0\{Block\}\}\}\}$。

如果 P_i 没有收到任何 $Ticket$，它将不对 I 做出回应。

(5)发起者计票。I 收到 P 的返回包，分离出 $Block_s$，并层层解密得到{ t'_i，$Ticket_1$，$Ticket_2$，$\cdots$，$Ticket_{n'}$ }。I 首先验证 t'_i 是否等于 t_i：如果不是，则视为无效包，包中所有选票视为废票。然后检查选票的格式是否正确，格式不正确或不能识别的选票为废票。对于Hash值相同的选票，只留下 $NetValue > \lambda$ 的任意一张。如果都小于 λ，则留下 $NetValue$ 最大的一张选票，其余删除。

I 对所有返回包中的 $Block$ 做相似处理后，开始计票。设返回的所有不同内容文件的Hash值的集合为 $H=\{h_1, h_2, \cdots, h_r\}$，$h_1, h_2, \cdots, h_r$ 出现的次数分别为 $c_1, c_2, \cdots, c_r$。令

$$c_s = \max(c_1, c_2, \cdots, c_r)。$$

若 $c_s \geqslant \gamma$ 且序列 $c_1, c_2, \cdots, c_r$ 中等于 c_s 的值唯一，则认为本次投票真实有效。设 h 是对应 c_s 的文件的Hash值，I 抽取投票为 h 的回应者 R_j。以洋葱路由向 P_i 发送确认信息 $Confirmation=\{H_t, RO\}$，要求 P_i 向提供者 R_i 索取文件，并以 RO 指示的路径逐层加密后送回 I。

4 分析

上述协议在真实文件占多数的条件下，实现了文件的真实性认证。相对于提供文件真实性认证的其它协议，该协议从发起投票到下载文件只需3次交互，与Gnutella的文件查询协议的交互次数相同，因此通信交互是高效率的。协议中代理节点泛洪投票的过程与Gnutella的文件查询协议相似。但本协议对路由交换表进行了改进以实现投票者匿名，通过代理分片投票和在返回中加入文件的Hash值实现了文件的

真实性认证，通过发起者与代理之间的洋葱路由和回馈洋葱路由实现了发起者匿名。为了控制每次投票的通信耗费，发起者可以将 TTL 值设置为比 Gnutella 协议的 *TTL* 小 1。

4.1 匿名性分析

投票者是否要保持匿名有两种不同的观点：不保持投票者匿名和保持投票者匿名。在网络环境中，用户要求保持自己的私有信息不被泄露，如果要求投票者提供真实身份，投票者就暴露了自己拥有被投票的文件，这就会使部分用户不愿投票，甚至不愿共享拥有的文件，这与 P2P 共享资源的发展初衷相背离。我们的协议提供了投票者匿名。在协议中，从发起者 I 到代理节点使用了洋葱路由，从代理节点到发起者 I 使用了回馈洋葱路由，因此代理节点及洋葱路由上的节点无法得知真正的发起者，当然投票者和文件的提供者也不知道谁是真正的发起者。投票过程中的路由交换表，使代理节点、发起者无法知道真正的投票者和文件的提供者。因此整个协议确保了发起者匿名、投票者匿名、文件提供者匿名。

4.2 协议所满足的电子投票的特性

电子投票协议(Electronical voting protocol)[9]是以 Internet 为媒介，由选举委员会主持，让合法的选举者通过投票对某一事物表达自己的态度。电子投票协议要满足：(1)完全性(Completeness)：所有合法选票都能被正确统计；(2)合理性(Soundness)：不诚实的选举者不能扰乱选举；(3)秘密性(Privacy)：所有选票都是秘密的，即能保护个人隐私；(4)不可重复性(Unresuability)：任何选举者不能投两次票；(5)公平性(Fairness)：任何事情都不能影响选举；(6)可验证性(Verifiability)：所有选举者都能检验他们的选票最后是否被统计上。

在 EV-FAP 中，因为选举管理者与计票者为同一实体，即发起者 I，为了得到正确的结果，发起者没有作弊的动机，所以不需用特殊机制约束计票者(发起者 I)以实现完全性与可验证性。使用路由交换表并对选票加密，使得返回路由节点以及代理节点无法知道选票的内容，这确保了秘密性。查询路由上的节点不会重复泛洪投票，因此投票者不能对同一 *sq* 投票多次，确保了不可重复性。由于被投票者是文件，与投票者不是同一类实体，因此能满足公平性。代理、(回馈)洋葱路由节点丢弃选票的行为和投票的合理性将在 4.4 小节中进行分析。

4.3 认证成功概率分析

依据得票多少，假冒文件与真实文件公平竞争成为当选文件。

在 3.1 小节的假设(2)“用户会删除他认为不真实的文件”和假设(3)“被投票文件的真实副本在整个 P2P 系统中占多数”下，理论上有较多的代理返回真实副本回应。第 5 节的仿真实验结果显示，在真实副本占优势的典型参数设置下，真实文件能以 95%以上的概率当选。

4.4 攻击分析

我们拟参照密码体制用两个维度(攻击与目标)对协议进行分析。

按敌手拥有的能力不同，攻击的方法有 3 种：(1)网络监听(Sniffer)；(2)弃包(Drop)；(3)恶意投票(Maliciously Vote)：即在选票中放入假冒文件的 Hash 值或(和)假冒用户投票。洋葱路由上的节点、代理节点、查询路由上的节点具有弃包的能力。代理节点、查询路由上的节点具有恶意投票的能力。因此 3 种攻击能力是由弱到强的包含关系，有恶意投票的能力必定有弃包与网络监听的能力。

协议的目标为：(1)投票有效(Effective Voting)，能根据投票结果做出判断，而不会出现返回的选票数太少或获票最多的两个 Hash 值的选票数相同等情况；(2)投票结果正确(Correct Result)，整个投票有效，且真实文件当选；(3)投票者匿名(Voter anonymity)；(4)发起者匿名(Initiator anonymity)。能达到目标(2)则一定能达到目标(1)，但其它目标之间没有隶属关系。

本协议能以较大概率挫败恶意投票攻击，达到目标投票结果正确、实现投票者匿名和发起者匿名。下面按照协议抵抗攻击的能力对其达到的目标进行分析。

协议中共有 7 类不同的角色：发起者、(回馈)洋葱路由上的节点、代理节点、查询路由上的节点、投票节点、文件提供节点、其它节点。

(1)网络监听与投票者匿名和发起者匿名。由于协议使用了洋葱路由、回馈洋葱路由、路由交换表以及对传输内容加密，监听数据包得不到包的传输内容。代理节点如果把 I 的公钥替换成自己的公钥进行泛洪，可以解密得到投票的内容，但其中只有文件的 Hash 值和 *NetValue*，却无法得知投票者的任何身份信息。因此协议总可以满足投票者匿名和发起者匿名。

(2)弃包。洋葱路由上的节点和查询路由上的节点弃包引起的后果是不同的。代理是以泛洪的方式进行查询的，设查询路由上的每个节点都选择 n 个邻居进行泛洪，设某一节点 A 与代理 P 的距离为 e，当 P 收到一张投票时，A 有机会接触该选票的概率为 n^{-e}。如果对应代理原本能收到 l 张选票，那么查询路由上各节点能阻止代理收到任何选票的概率为 $(n^{-e})^l=n^{-l\cdot e}$。取 $l=3$，$e=2$，$n=4$，则 A 能阻止 P 收到任何选票的概率只有 $n^{-l\cdot e}\approx 0.0002441$。

洋葱路由和回馈洋葱路由上的节点弃包会使发起者少收到一个回应。设洋葱路由上的节点数为 f，回馈洋葱路由上的节点数为 g，那么只有当洋葱路由及回馈洋葱路由上的所有节点都不是恶意节点时 I 才能收到对应代理的 *Block*。设洋葱路由和回馈洋葱路由上的节点为恶意节点的概率为 p，那么 I 收到对应代理返回的 *Block* 的概率为 $(1-p)^{f+g-1}$。如果 I 向 m 个代理发出投票请求，那么 I 收到的返回 *Block* 的个数的数学期望为 $m\cdot(1-p)^{f+g-1}$。I 可以精心选择可靠的代理、洋葱路由和回馈洋葱路由上的节点，使 $(1-p)^{f+g-1}$ 较大。例如，取 $f=g=4$，$p=0.04$ 时，$(1+p)=^{f+g-1}\approx 0.75$。取 $m=10$，则 $m\cdot(1+p)^{f+g-1}\approx 7.5$。投票有效性可以得到保障。

(3)恶意投票。查询路由上的节点和代理节点进行恶意投票攻击时，会使 I 得到的选票集 $H=\{h_1, h_2, \cdots, h_r\}$ 中错误 Hash 值的个数增加。但由于统计选票之前删除了每个 *Block* 中的重复 Hash 值，因此恶意节点即使多次投票在同一个 *Block* 中也只能留下一票。只要 $c_s>1$(c_s 的定义见 3.2 小节的第(5)步)，恶意投票攻击就会失效。恶意节点要破坏投票有效和投票结果正确的目标，就必须在尽可能多个选区内投票，使自己在选票集 H 中的恶意选票数大于等于真实文件的选票数 k。

定理 设系统中的节点总数为 τ，每个节点的邻居数为 n，I 选定了 m 个代理，那么恶意节点能在 k 个选区内投票的概率为

$$\frac{m!}{(m-k)!}\sum_{0\leqslant\xi_1,\xi_2,\cdots,\xi_k\leqslant TTL}\theta(\xi_1,\xi_2,\cdots,\xi_k)\cdot\omega(\xi_1,\xi_2,\cdots,\xi_k)$$

其中

$$\theta(\xi_1,\xi_2,\cdots,\xi_k)=\frac{\binom{\tau}{\varepsilon_k}P\prod_{i=1}^{k}\binom{\varepsilon_i}{\xi_i}}{\prod_{i=1}^{k}\binom{\tau}{\xi_i}},$$

$$\omega(n,\xi_1,\xi_2,\cdots,\xi_k)=\frac{n^{\varepsilon_k}}{\tau^k},\varepsilon_i=\sum_{j=1}^{i}\xi_j$$

证明：设节点 A 到 k 个代理的路由长度分别为 $\xi_1,\xi_2,\cdots,\xi_k$。令 $\varepsilon_i=\sum_{j=1}^{i}\xi_j$，则节点 A 到 k 个代理的 k 个路由节点集两两没有交集的概率为

$$\theta(\xi_1,\xi_2,\cdots,\xi_k)=\frac{\binom{\tau}{\varepsilon_k}P\prod_{i=1}^{k}\binom{\varepsilon_i}{\xi_i}}{\prod_{i=1}^{k}\binom{\tau}{\xi_i}},$$

又因为节点 A 与某一代理 P 路由长度是 ξ 的概率为 $\sigma(n,\xi)=\frac{n\xi}{\tau}$，则节点 A 到 k 个代理路由长度分别为 $\xi_1,\xi_2,\cdots,\xi_k$ 的概率为

$$\omega(n,\xi_1,\xi_2,\cdots,\xi_k)=\prod_{i=1}^{k}\sigma(n,\xi_i)=\frac{n^{\sum_{i=1}^{k}\xi_1}}{\tau^k}$$

因此恶意节点能在 k 个选区投票的概率为

$$\frac{m!}{(m-k)!}\sum_{0\leqslant\xi_1,\xi_2,\cdots,\xi_k\leqslant TTL}\theta(\xi_1,\xi_2,\cdots,\xi_k)\cdot\omega(\xi_1,\xi_2,\cdots,\xi_k)$$

取 $\tau=20000$，$m=10$，$n=4$，按照上述定理计算可以得到，一个恶意节点能在 5 个和 6 个选区内投票的概率分别为 0.00004327885和 $3.6771*10^{-6}$。如果恶意节点数达到 6%，存在能在 5 个选区内投票的恶意节点的概率为

$1-(1-0.00004327885)^{20000*0.06}=0.05061$

设发起者 I 能得到 k 个代理提供的真实文件的 Hash 值，那么恶意节点在 k 个选区内投票就能破坏目标投票的有效性，如果在 k 个以上的选区内投票就能破坏目标投票结果正确。从以上理论分析和示例数据可以看出，在“绝大部分的 P2P 用户是善意的”的假设下，投票的有效性和投票结果的正确性目标遭到破坏的概率是较小的。这点同样可以通过第 5 节的实验仿真看出来。

5 实验仿真

我们构造实验从 4 个方面检测我们协议的有效性：(1)回应代理数量：是否有足够的代理回应，使发起者 I 能够做出判断；(2)假冒文件所占比率对认证成功率的影响：随着假冒文件所占比率的增加，本协议成功进行文件真实性认证的概率变化情况；(3)恶意节点数量对认证成功率的影响：不同比例的恶意节点对本协议成功进行文件真实性认证的概率的影响；(4)协议的通信耗费：当查询涉及相同数量的节点时，本协议与著名的 Gnutella 协议在网络上泛洪包的数量比率。

实验设置：我们使用的 P2P 层的拓扑结构来自于文[10]的 DSS Clip2 trace，选择了 2001 年 3 月 18 日和 2001 年 4 月 5 日监测到的 Gnutella 的拓扑结构。以邻接表的形式表示 P2P 节点的邻居关系，模拟对文件投票。选择了 0.001，0.002，0.003，0.004，0.005，0.006，0.007，0.008，0.009，0.01，0.015，0.02，0.025，0.03，0.035，0.04，0.045，0.05，0.06，0.07，0.08，0.09，和 0.1 的 23 种“真实文件分布率”(见 2.1 小节)，对每种“真实文件分布率”分别选择 3000 个不同发起者，每个发起者选择 10 个代理进行仿真，然后对响应结果求平均值。我们的实验在两个不同拓扑的网络上的实验结果是类似的，下面给出在 2001 年 4 月 5 日拓扑结构下的仿真结果。

5.1 回应代理数量

要使本协议成功地进行文件真实性认证，就必须确保发起者能收到足够数量的真实文件所有者的投票，而且是通过不同代理返回的，即达到“投票有效”的目标。图 3 对不同的 TTL 值，给出了真实文件分布率与响应代理的数量变化关系。可以注意到，如果设定 0.002 的节点拥有真实文件，$\gamma=2$，$TTL=3$，则达到目标“投票有效”的概率为 99.5%。

5.2 假冒文件所占比率对认证成功率的影响

当假冒文件存在时，达到“投票结果正确”的目标是协议有效性的一个重要指标。这必须保证返回真实文件 Hash 值的代理数量超过返回假冒文件 Hash 值的代理数量。本文中，假冒文件特指某一版本的假冒文件，而不是被投票文件 F 的各种假冒文件的总和。只要每种版本的假冒文件数量(而不是所有假冒版本的总数量)都小于真实文件的数量就符合 3.1 的假设(3)。取 $TTL=4$，假冒文件与真实文件的比例 ρ 分别为 0.5、0.7、0.9 时，图 4 给出协议达到“投票结果正确”目标的概率。即使真实文件的分布率为 0.003，$\rho=0.7$，达到“投票结果正确”的目标的概率仍有 94.2%；如果真实文件分布率大于等于 0.001，且 $\rho=0.5$，则协议达到“投票结果正确”目标的概率大于等于 94.6%。随着投票和下载次数的增多，假冒文件与真实文件的比率会越来越小。

5.3 恶意节点数量对认证成功率的影响

当恶意节点的比例增加时，能否达到投票结果正确的目标反映了协议抵抗恶意节点攻击的能力。图 5 反映了当 $TTL=4$，恶意节点比例从 2%增加到 10%时，在不同的文件分布率下，协议达到目标“投票结果正确”的概率。当恶意节点比例小于等于 6%(假设“绝大部分的 P2P 用户是善意的”)时，如果真实文件的分布率超过 0.006，则有大于 88.2%的投票能达到投票结果正确的目标。随着真实文件的分布率的提高，达到投票结果正确的目标的概率稳定在 95%的水平。当恶意节点数为 10%时，随着真实文件的分布率的提高，达到目标投票结果正确的概率稳定在 90%的水平。

5.4 协议的通信耗费

通信耗费关系到协议的可扩展性。本协议投票过程与 Gnutella 的查询过程相似。模拟查询的结果表明，要让两种协议的投票包(查询包)涉及相同数量的节点，则在网络上需要泛洪数量相当的投票包(查询包)。从图 6 可以看到在涉及相同数量的节点时，本协议与 Gnutella 协议泛洪包的数量之比在 1±5%的范围内波动。

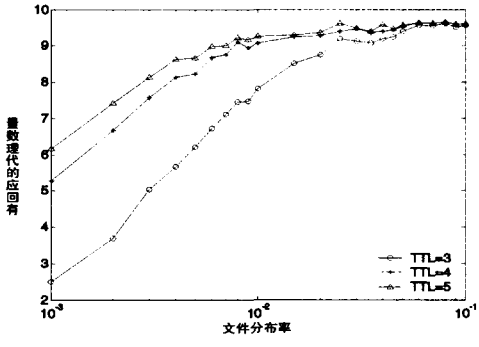

图 3 回应代理数量

(下转第 137 页)

传播特点进行了分析，建立了基于 TCP/IP 简单信息的访问兴趣度一般关系模型，在此基础上进一步提出了一种新的基于源地址活跃度、目标地址离散度和响应度准则等多个测度联合约束的快速蠕虫定位算法。其中针对地址离散度评价，又提出基于位图(Bitmap)重量的评价方法。实验表明，基于该约束算法能实现目标网络的蠕虫传播源快速定位。为提高算法在骨干网上的处理速度，研究中提出了基于二级 HASH 页表结构的快速 TIR 树构建算法，实验结果表明该算法极大地提高了信息的处理效率。

参 考 文 献

1 Spafford E H. The Internet Worm: Crisis and Aftermath. Communications of the ACM, 1989, 32(6)678～687
2 The Morris Internet worm. http://snowplow. org/tom/worm/worm. html
3 CERT Advisory CA-2001-19 "Code Red" Worm Exploiting Buffer Overflow In IIS Indexing Service DLL. http://www. cert. org/advisories/CA-2001-19. html
4 Computer worm grounds flights, blocks ATMs - Jan. 26, 2003, CNN. com. http://www. cnn. com/2003/TECH/internet/01/25/internet. attack/
5 Zou C C, Towsley D, Gong W. On the Performance of Internet Worm Scanning Strategies: [Umass ECE Technical Report. TR-03-CSE-07]. Nov. 2003
6 Bakos G, Berk V. Early Detection of Internet Worm Activity by Metering ICMP Destination Unreachable Messages. In: Proc. of the SPIE Aerosense, 2002
7 Maxion R A. Anomaly detection for diagnosis. In: Proc. of the 20th Intl. Symposium Fault-Tolerant Computing(FTCS-20), 1990. 20～27
8 Maxion R A, Feather F E. A case study of Ethernet anomalies in a distributed computing environment. IEEE Transactions on Reliability, 1990, 39(4): 433
9 Huang P, Feldmann A, Willinger W. A non-intrusive, wavelet-based approach to detecting network performance problems. In: Proc. of ACM SIGCOMM Internet Measurement Workshop, San Francisco Bay Area, 2001
10 Lakhina A, Crovella M, Diot C. Diagnosing Network-wide Traffic Anomalies. In: Proc. ACM SIGCOMM, August 2004
11 Gu RongJie, Xia DeLin, Yan PuLiu Yan. An Adaptive Internet Backbone Malicious Activies Detection System Based on Frequent Pattern Mining. [J]GESTS International Transactions on Computer Science and Engineering, 2005, 12: 141～148

(上接第 117 页)

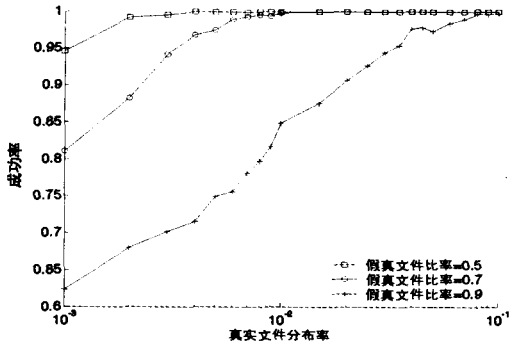

图 4 假真文件比率与认证成功率的关系

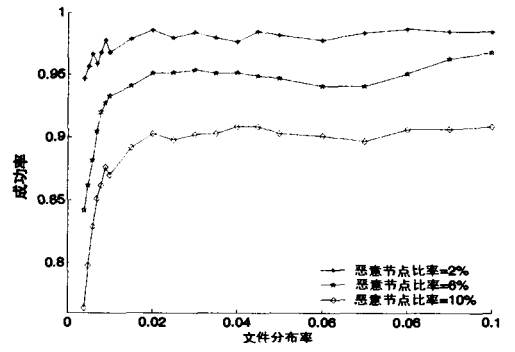

图 5 恶意节点数量与认证成功率的关系

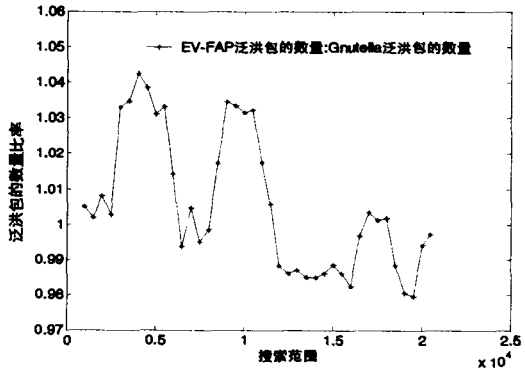

图 6 泛洪包的数量比率

总结 由于匿名性需求，已有的文件真实性认证协议并不能很好地适用于分布式 P2P 环境。本文提出了一种基于电子投票的文件真实性认证协议 EV-FAP，在确定文件真实性的同时实现了协议参与者匿名。我们对协议进行的理论分析表明，EV-FAP 能抵抗网络监听、弃包、恶意投票等攻击。相比已有的文件真实性认证协议，通过实验仿真，我们观察到 EV-FAP 具有协议交互次数少、效率高、认证成功率高等特点。我们相信 EV-FAP 是适合无中心 P2P 网络的文件真实性认证协议。

参 考 文 献

1 Saroiu S, Gummadi K P, Dunn R J, et al. An Analysis of Internet Content Delivery Systems. OSDI '02, 2002
2 Suryanarayana G, Taylor R N. A Survey of Trust Management and Resource Discovery Technologies in Peer-to-Peer Applications: [UCI-ISR Technical Report 04-7]. 2004
3 Daswani N, Garcia-Molina H, Yang B. Open Problems in Data-sharing Peer-to-peer Systems. ICDT2003, 2003
4 Cornelli F, Damiani E, Vimercati S D C D, et al. Choosing Reputable Servents in a P2P Network. In: Proceedings of the 11th World Wide Web Conference, 2002. 441～449
5 Damiani E, Vimercati D C D, Paraboschi S. A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks. In: ACM Conference on Computer and Communications Security 2002, 2002. 207～216
6 Maniatis P, Roussopoulos M, Giuli T, et al. Preserving Peer Replicas By Rate-Limited Sampled Voting. In: Proceedings of the 19th ACM Symposium on Operating Systems Principles, 2003
7 Goldschlag D M, Reed M G F P. Onion Routing for Anonymous and Private Internet Connections. Communications of the ACM, 1999, 42(2): 39～41
8 Xiao L, Xu Z, Zhang X. Low-cost and Reliable Mutual Anonymity Protocols in Peer-to-Peer Networks. IEEE Transactions on Parallel and Distributed Systems, 2003, 14: 829～840
9 冯登国，裴定一. 密码学导引. 北京：科学出版社，1999
10 Gnutella Protocol Specification. www. gnutella. co. uk/library/pdf/gnutella_protocol_0. 4, 2000-07