



SSDHT: 基于社交网络的 DHT 安全增强机制

李攀攀¹, 谢正霞², 乐光学¹, 马柏林¹, 陈丽¹, 刘建华^{3,4}

(1. 嘉兴学院数理与信息工程学院, 浙江 嘉兴 314001;

2. 嘉兴学院建筑工程学院, 浙江 嘉兴 314001;

3. 上海师范大学信息与机电工程学院, 上海 200234;

4. 绍兴文理学院计算机科学与工程系, 浙江 绍兴 312000)

摘要: P2P 系统在文件共享等领域中得到了广泛的应用, 但 DHT (distributed hash table) 网络无中心、无认证、缺乏节点身份验证机制, 使得现有的基于 DHT 的 P2P 系统易受到 Sybil 攻击等外部攻击。提出一种基于社交网络的 DHT 安全增强机制, 将社交网络中节点的信任关系引入 DHT 网络中提高对 Sybil 节点的识别能力。以 KAD (Kademlia) 算法为例进行了实验验证, 基于 Facebook 和 Twitter 数据集的实验结果表明本文提出的安全机制适用于大规模动态的网络, 能够有效防御 Sybil 攻击。

关键词: P2P; DHT; 社交网络; Sybil 攻击

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018276

SSDHT: social network-based secure DHT mechanism

LI Panpan¹, XIE Zhengxia², YUE Guangxue¹, MA Bolin¹, CHEN Li¹, LIU Jianhua^{3,4}

1. College of Mathematics and Information and Engineering, Jiaxing University, Jiaxing 314001, China

2. College of CML Engineering and Architecture, Jiaxing University, Jiaxing 314001, China

3. College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China

4. Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China

Abstract: P2P systems have been widely used in the field of file sharing, etc. However, DHT (distributed hash table) networks have no center, no authentication, and lack of node authentication mechanism, making existing DHT-based P2P systems vulnerable to external attacks such as Sybil attacks. A DHT security enhancement mechanism based on social network was proposed, which introduced the trust relationship of nodes in the social network into the DHT network to improve the recognition ability of Sybil nodes. The KAD (Kademlia) algorithm was used as an example to verify the experiment. Experimental results based on Facebook and Twitter datasets show that the proposed security mechanism is suitable for large-scale dynamic networks and can effectively defend against Sybil attacks.

Key words: P2P, distributed hash table, social network, Sybil attack

收稿日期: 2018-04-25; 修回日期: 2018-10-31

基金项目: 国家自然科学基金资助项目 (No.61572014); 浙江省自然科学基金资助项目 (No.LY16F020028, No.LY15F020040)

Foundation Items: The National Natural Science Foundation of China (No.61572014), Zhejiang Provincial Natural Science Foundation of China (No.LY16F020028, No.LY15F020040)

1 引言

目前 P2P 技术已成为文件共享、即时通信、视频流媒体的主流实现技术。2018 年 1 月, CNNIC (China Internet Network Information Center, 中国互联网络信息中心) 发布的第 41 次《中国互联网络发展状况统计报告》中指出, 75% 的网民使用和收看基于 P2P 技术的网络视频应用^[1]。现有主流 P2P 系统 (如 BitTorrent、eMule) 均基于 DHT 网络, DHT 网络的拓扑结构相对稳定、规则, 具有可扩展性良好、可靠性高和资源定位准确等优点^[2]。但由于 DHT 网络无中心、无认证、缺乏节点身份验证机制, 现有的基于 DHT 网络的 P2P 系统非常容易受到外部攻击, 如 Sybil 攻击、Eclipse 攻击及路由和存储攻击等^[3]。目前研究工作都致力于防止或减轻攻击危害, 尤其是降低 Sybil 攻击的威胁, 因为 Sybil 攻击往往是部署其他大规模网络攻击的先决条件, 因此如何防御 Sybil 攻击, 提高 DHT 网络的安全性, 尤其是 DHT 路由的安全性一直是学术界的热点问题。

当前针对 Sybil 攻击对抗方法的研究分为中心化结构和分布式结构两大类, 前者相关工作主要包括可信证书^[4-7]和虚拟货币机制^[8-10]等, 其方法依赖于中心组件, 这也是系统瓶颈, 影响 P2P 系统的可靠性和扩展性。分布式结构的方法通过检查节点是否有足够的资源, 判断节点是否为 Sybil 节点, 该方法的事实基础是用户在网络中拥有多个节点身份时, 每个节点身份的资源少于普通节点的概率很大, 代表性工作包括资源测试^[11]、计算挑战^[12]、IP 地址过滤^[13]、数据流检测^[14]等, 但这些方法在节点能力异构性突出的环境中存在较高的误报率。

近年来, 社交网络的发展引人注目, 约有一半以上的网民通过社交网络沟通交流、分享信息。社交网络已成为覆盖用户最广、传播影响最大、商业价值最高的 Web 2.0 业务^[1]。一些研究者尝试

利用社交网络中需要人为参与和判断的信任关系来识别 Sybil 节点^[10-19], 为增强 DHT 的安全提供了新的思路。但已有机制对网络带宽的占用大, 并由此带来可扩展性差、不适应动态网络等问题。

本文提出一种基于社交网络的 DHT 安全增强机制, 将社交网络中节点的信任关系引入 DHT 网络中以提高对 Sybil 节点的识别能力。本文以 KAD 算法为例进行了实验验证, 基于 Facebook 和 Twitter 数据集的实验结果表明, 本文提出的安全机制适用于大规模动态的网络, 能够有效地防御 Sybil 攻击。

2 相关工作

随着社交网络的兴起, 研究者开始利用社交网络识别 Sybil 节点, 缓解针对 DHT 的攻击。Marti^[15]早在 2005 年就提出 SPROUT DHT 路由算法, 基于社交网络中节点间跳数来判断节点间的信誉度, 尝试利用社交网络节点间的信誉度进行路由选择。同年 Danezis 等人^[16]提出的引导图模型给出了利用社会网络实现安全路由的初步建议。Yu 等人^[17]在 2006 年和 2010 年先后提出的 SybilGuard 和 SybilLimit 系统中采用了攻击边的概念, 每一条由良性节点社团到恶意节点社团的逻辑连接称为一条攻击边, 并基于 Boyd 和 Flaxman^[19-20]提出的社会网络快速混合 (fast mixing) 的观点, 采取一种特殊的随机路由的方式保证良性节点以很大的概率同另一个良性节点进行交互。假定 n 为网络规模, g 为攻击边数, m 为网络中边的数量 (后同)。在系统中攻击边条数复杂度为 $O(n/\ln n)$ 的情况下, SybilLimit 可以保证良性节点团体能接受的 Sybil 节点数的复杂度为 $O(g\sqrt{n})$ 。但是 SybilLimit 中每次需要验证复杂度为 $O(\sqrt{m})$ 的随机路由, 如果直接采用 SybilLimit 提出的算法验证另外一个节点是否为良性节点, 需要消耗网络带宽的复杂度为 $O(n^2\sqrt{m})$, 效率低于泛洪查找。Tran 等人^[21]提出



的 SumUP 系统通过社会网络实现了客体的评价,但 SumUp 要依赖于一个可信服务器来存储节点交易的历史信息,不适用于完全分布式的 P2P 网络。

Lesniewski-Laas^[22]基于上述工作提出了一种基于社会网络的 DHT 方法,并提出了改进的 Whanau 方法^[23]。Whanau 方法中每个节点需要 $O(\sqrt{m})$ 次随机路由构建路由表。Whanau 方法实现了单跳的路由查询,路由查询速度快但 Whanau 方法中每个节点需要维护 $O(\sqrt{nlbn})$ 大小的路由表,这使得 Whanau 方法在网络抖动时难以保证系统处于正确状态。Mittal 等人^[24]在 Whanau 的基础上提出了 X-vince 方法,利用社交网络实现了复杂度为 $O(lbn)$ 跳的路由查询,但 X-vince 方法中节点只能与社交网络中的邻居节点通信,而不能与自身路由表中的节点直接通信,导致 DHT 查询操作的时延明显增加,且 X-vince 节点间的消息传递路径中有一个节点离线时,就需重新构造一条消息传递路径,因而 X-vince 方法不适合动态网络。现有基于社交网络的 DHT 系统见表 1。

表 1 现有基于社交网络的 DHT 系统

方法	路由表	通信方式	路由查询	动态网络
Whanau	$O(\sqrt{nlbn})$	直接通信	快	不适合
X-vince	$O(lbn)$	间接通信	慢	不适合

如表 1 所述,现有基于社交网络的 DHT 系统存在一定的局限性。针对上述问题,本文将社交网络中的信任关系引入 DHT 网络中识别 Sybil 节点,在节点信息发布阶段增强 DHT 的安全性。节点只需维护复杂度为 $O(lbn)$ 的路由表,且节点可与路由表中的节点直接通信,缩短查询时延,避免不必要的带宽消耗。

3 基于社交网络的 DHT 安全增强机制

DHT 网络具有无中心、无认证、缺乏节点身份验证机制等特点,使得现有的基于 DHT 的 P2P

系统易受到 Sybil 攻击等外部攻击。针对上述问题,本文提出 SSDHT (social network based secure DHT mechanism, 基于社交网络的 DHT 安全增强机制),将社交网络中节点间的信任关系引入 DHT 网络中识别 Sybil 节点,增强 DHT 的安全性。

3.1 攻击模型及假设

本文研究所针对的攻击模型如图 1 所示,网络中存在 k' 个良性节点 $(x_1, x_2, \dots, x_{k'-1}, x_{k'})$ 和 j' 个攻击节点 $(y_1, y_2, \dots, y_{j'-1}, y_{j'})$, 节点间通过社交网络的逻辑连接进行连接。在攻击模型中,良性节点遵守系统规则,而攻击节点伪造并冒充合法节点 ID,注册大量的 Sybil 节点 $(z_1, z_2, \dots, z_{l'-1}, z_{l'})$, 进而在 Sybil 攻击的基础上部署其他大规模网络攻击,如路由污染、索引毒害等。本文中定义良性节点社团 $V(L) = \{x_1, x_2, \dots, x_{k'-1}, x_{k'}\}$; Sybil 节点社团 $V(S) = (y_1, y_2, \dots, y_{j'-1}, y_{j'}) \cup (z_1, z_2, \dots, z_{l'-1}, z_{l'})$ 。攻击节点社团中为数众多的攻击节点和 Sybil 节点往往被一个攻击者控制。

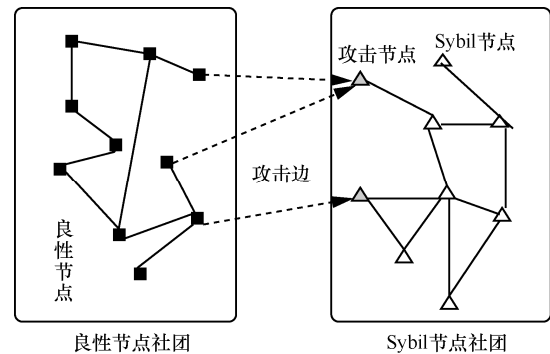


图 1 攻击模型

本文中节点间的信任关系具有不对称性,节点 x_i 信任节点 x_j , 不代表着节点 x_j 信任节点 x_i 。如果在社交网络中存在 x_i 到 x_j 的逻辑连接,则称 x_i 信任 x_j , 表示为 $x_i \rightarrow x_j$ 。本文中引用参考文献[5-6]中攻击边的概念,若 $x_i \in V(L)$, $y_i \in V(S)$, 且 $x_i \rightarrow y_i$, 则称 $x_i \rightarrow y_i$ 为一条攻击边, $V(G) = \{x_i \rightarrow y_i | x_i \in V(L), y_i \in V(S)\}$ 为攻击边集合。Sybil 节点社团可在社团 $V(S)$ 内建立非常多的信任关系, 令 $V(G') = \{x_i \rightarrow y_i | x_i \in V(S), y_i \in V(S)\}$, 则有 $0 \leq |V(G')| < +\infty$ 。但社

交网络中信任关系需要人为的参与和判断。研究成果表明^[13-14], 攻击者很难在大量的 Sybil 节点社团 $V(S)$ 和良性节点社团 $V(L)$ 间建立起足够多的信任关系, 即限定 $|V(G)|=w(w \in \mathbb{N})$ 。与参考文献[24]中相同, 本文中也假设 $|V(G)|$ 中的攻击边遵循随机分布。

3.2 系统的基本框架和原理

针对上述攻击模型, 为评估 DHT 网络节点间的信任关系、识别 Sybil 节点, SSDHT 将社交网络中的信任关系引入 DHT 网络中。系统由 DHT 网络与社交网络构成, 基本框架如图 2 所示。DHT 网络实现准确可靠的资源定位, 每个节点负责维护一段路由信息和索引列表, 从而实现整个 DHT 网络的寻址和存储。社交网络解决 DHT 网络中节点互不信任的问题, 增强 DHT 的安全性。

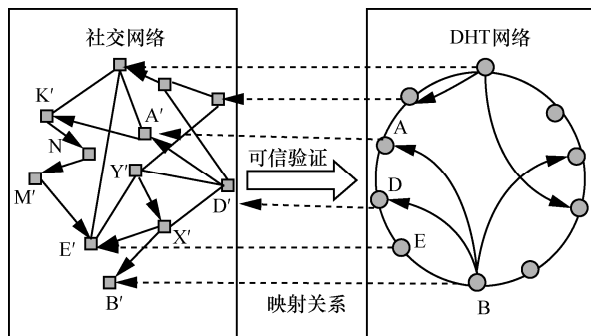


图2 系统的基本框架

设 DHT 网络节点集合 $V(\text{DHT}) = \{x | x \text{ 是 DHT 网络中的节点}\}$, 社交网络节点集合 $V(\text{SN}) = \{x' | x' \text{ 是社交网络中的节点}\}$ 。 $V(\text{DHT})$ 与 $V(\text{SN})$ 均为非空集合, DHT 网络中节点注册阶段用社交网络进行身份验证, 对于 $V(\text{DHT})$ 中每个节点 x , 在 $V(\text{SN})$ 中有唯一确定的一个节点 x' 与之对应, 即存在从 DHT 网络到社交网络的映射。记作 $f: V(\text{DHT}) \rightarrow V(\text{SN})$ 。

本节中用到的符号定义如下。

信任路径 $p[x, y]$: 如果节点 x 与节点 y 在社交网络中存在直接或间接的信任关系, 则存在 x 到 y 的信任路径, 记作信任路径 $p[x, y]$ 。如图 2

所示, 节点 A' 与节点 B' 间存在间接信任关系, 则 A' 存在到 B' 的信任路径。信任路径 $p[A', B']$ 依赖信任关系 $\{A' \rightarrow Y', Y' \rightarrow X', X' \rightarrow B'\}$ 。

Route (x): 节点 x 的路由表。

$x \leftarrow y$: 节点 x 收到节点 y 发送的消息。

Route (x) $\leftarrow y$: 节点 x 利用节点 y 的信息更新路由表。

Max ($x \rightarrow y$): 依赖信任关系 $x \rightarrow y$ 的信任路径的最大数量。

基于上述系统框架, SSDHT 利用社交网络中节点间的信任关系构建信任路径识别 DHT 网络中的 Sybil 节点, 节点路由表更新时, 只将验证通过的非 Sybil 节点加入路由表中。如算法 1 所示, $V(\text{DHT})$ 中的节点 x 在将另一节点 y 插入其路由表之前, 先利用对应的 $V(\text{SN})$ 中节点 x' 的信任关系构造到 y' 的信任路径 $p[x', y']$, 当且仅当信任路径 $p[x', y']$ 存在时, x 认为 y 为良性节点, x 将 y 插入路由表中。

算法 1 n-Identify (Node x , Node y)

输入: Node x , Node y

输出: Route (x)

初始化系统参数, 包括 Node x , Node y

$x \leftarrow y$

$x' = f(x)$, $y' = f(y)$

if $p[x', y'] \neq \text{NULL}$ then //存在信任路径 $p[x', y']$

Route (x) $\leftarrow y$

else

x refuse to insert y into Route (x)

end if

return Route (x)

当 $|V(G)|=1$ 且 $\text{Max} (x \rightarrow y) = \delta (x \in V(L), y \in V(S))$ 时, 即使攻击节点伪造足够多的 Sybil 节点, 并以不同的节点 ID 反复加入 DHT 网络中。根据算法 1, 至多有 δ 个 Sybil 节点进入良性节点的路由表中, 即单个 Sybil 节点社团的大小被限定为 δ 。



当 $|V(G)|=w$ 时, SSDHT 可以保证良性节点社团路由表中 Sybil 节点不超过 $O(\delta \cdot g)$ 。

3.3 关键算法与数据结构

社交网络的拓扑结构如图 3 所示, 其中节点 E、F 是节点 B 在社交网络中的邻居节点。节点 K、R 是节点 B 路由表中的节点。虚线表示节点间的信任路径。

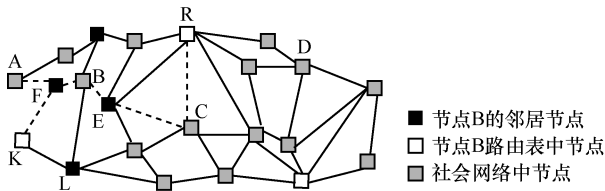


图 3 社交网络的拓扑结构

3.3.1 路由表的结构

与 KAD 算法^[24]相同, SSDHT 采用异或 (XOR) 衡量节点间的距离, 对于 ID 分别为 x 和 y 的两个节点, 它们的距离 $d(x, y) = x \text{ XOR } y$ 。如图 4 所示, 每个节点需维护 $\text{lb}n$ 个表 (n 为网络规模, 默认 $\text{lb}n = m$, m 是节点 ID 的位数), 对于第 i ($0 < i < m$) 个表, 该表内保存着和该点距离为 $2^{i-1} - 2^i$ 的 k 个节点信息 (节点 IP 地址、UDP 端口号、节点 ID)。这些表称为 k 桶, 其中 k 是系统可变参数, k 桶覆盖距离的范围呈指数关系增长。

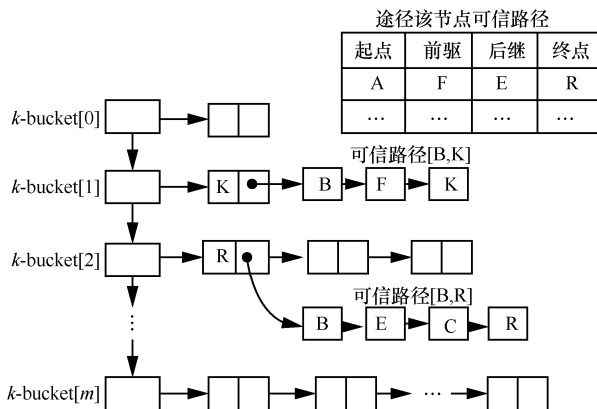


图 4 路由表结构

为了应对 DHT 网络的抖动和社交网络的拓扑变化, 除上述节点信息, 路由表中还需保存以

下信息。

- 节点到其路由表中节点的信任路径。以图 3 中节点 B 的路由表为例, R 是 B 的路由表中节点, 如图 4 所示, B 的路由表中除保存着节点 R 的 IP 地址、UDP 端口号、节点 ID 外, 还需保存信任路径 $p[B, R]$ 依赖的信任关系 $\{B \rightarrow E, E \rightarrow C, C \rightarrow R\}$ 。
- 经过该节点的信任路径的起点、前驱节点、后继节点及终点。信任路径 $p[A, R]$ 经过节点 B, B 路由表中需保存信任路径 $p[A, R]$ 的起点 A、前驱节点 F、后继节点 E 及终点 R。

3.3.2 基于 DHT 网络节点 ID 的信任路径构造算法

如第 3.3.1 节中所述, k 桶覆盖距离的范围呈指数关系增长, 因此 DHT 网络节点路由表中离自己距离近的节点的信息多, 离自己距离远的节点的信息少。基于上述特点, 本文提出一种基于 DHT 网络节点 ID 的信任路径的构造算法。

对于 DHT 网络中 ID 分别为 x, y 的两个节点, 在社交网络中都存在唯一确定的节点 $x' = f(x), y' = f(y)$ 与之对应, SSDHT 将 DHT 网络中节点 ID 映射到社交网络中。如算法 2 所示, 当社交网络中节点 x' 构造信任路径 $p[x', y']$ 时, 向距目标节点 y' 对应的节点 ID y 距离近的方向收敛。可信路径构造时, 每次查询都能从距离更接近目标 ID y 的节点的 k 桶中获取节点信息, 从而保证信任路径构造过程收敛。算法中用到的符号定义如下。

- $\text{ND}(x)$: 节点 x 在社交网络中直接信任的节点。
- $\text{RL}(x)$: 节点 x 的路由表中保存的信任路径。
- $M[x, y]$: 信任路径 $p[x, y]$ 的下一跳节点。
- $\text{FL}(x)$: 查找过的节点集合。
- $d(x, y) = x \text{ XOR } y$ 。
- T : 查找次数。
- t' : 超时时间。

算法 2 p-Search(Node x , Node z)

输入: Node x 为起始节点, Node z 为目标节点。

输出: $p[x,z]$ -信任路径 $p[x,z]$ 。

初始化系统参数, 包括 Node x , Node z , $FL(x)$, $T, t', i=0$;

```

while( $i < T$ )//最大尝试次数为  $T$ 
     $i++$ ;
     $V(x) = \{Route(x) - FL(x)\}$ ;
    if  $V(x) == \emptyset$  //随机选择一个邻居节点
        作为中转节点
         $y \leftarrow \text{random}\{ND(x)\}$ 
        //选择路由表中距  $z$  最近的节点  $y$  作为中转节点, 即  $x$  经  $y$  到  $z$ 
        else if  $r \in V(x) \&\& [\exists r_i \in V(x), d(r_i, z) < d(r, z)]$  then
             $y \leftarrow r$ ;
        end if
         $FL(x) \leftarrow y$  //加入查找过的节点集合
        while( $t < t'$ )//超时时间
             $n = N(x, y)$  //下一跳节点  $n$ 
            while ( $n! = y$ )
                //考察  $n$  的邻居节点或保存信任路径中是否包含  $z$ 
                if  $z \in ND(n) || z \in RL(n)$  then
                    //回溯的方式返回信任路径  $p[x,z]$ 
                    return  $p[x,z]$ 
                end if
                //考察  $n$  的邻居节点或路由表中是否存在比  $y$  距  $z$  估计距离更小的点  $m$ 
                if  $m \in [ND(n) \cup RL(n)] \&\& F(x, m, z) < F(x, y, z)$  then
                     $y \leftarrow m$ ; //  $m$  作中转节点
                end if
                 $n \leftarrow N[x, y]$ ;
            end while
        //路由表中存在估价距离更小的点  $m$ 
        if  $\exists m \in Route(y), F(x, m, z) < F(x, y, z)$ 
             $y \leftarrow m$ ;

```

else

break;

end if

end while

end while

$F(x,y,z)$ 为节点 x 经节点 y 到节点 z 的估价距离。

$$F(x, y, z) = G(x, y) + \sqrt{\frac{\ln n}{m \cdot \text{lb}d}} \text{lb}D(y, z) \quad (1)$$

其中, $G(x, y)$ 表示从节点 x 到节点 y 的实际距离, 在无权网络中为节点 x 到节点 y 的跳数。式 (1) 后半部分表示节点 y 到节点 z 的估价距离, 其中 n 为网络规模, d 为节点的平均度, m 为节点 ID 的位数。

考虑如下情况, $x_1, x_2 \in$ 良性社团集合 $V(L)$, $y \in$ Sybil 社团集合 $V(S)$, 节点 x_2 请求添加到节点 x_1 的路由表。若信任路径 $p[x_1, x_2]$ 的构造请求转发到 y 时, y 可发动两种攻击: 丢弃该消息、篡改或伪造该消息。

针对上述攻击, 本文提出以下策略保证信任路径构造的安全性。

节点 x_1 在构造信任路径 $p[x_1, x_2]$ 时, 选择 r 个点作为初始节点, 同时构造 r 条信任路径 $p[x_1, x_2]$, 只要其中一条路径中不存在攻击节点 y , 就可完成信任路径 $p[x_1, x_2]$ 的构造。

如果一定的时间 t' 内, 节点 x_2 没有返回的信任路径 $p[x_1, x_2]$, 节点 x_1 将选择其他节点作为初始节点, 重新构造可信路径 $p[x_1, x_2]$ 。

x_2 将发给 x_1 消息的摘要用 x_2 的私钥加密, 与消息一起传送给节点 x_1 。节点 x_1 在收到 x_2 的消息后, 用 x_2 的公钥解密被加密的摘要信息, 并同原文产生的摘要信息做比较, 判断信息是否被篡改。

3.4 DHT 网络的抖动与社交网络信任关系的变化

3.4.1 DHT 网络的抖动

节点首次加入 DHT 网络的算法如算法 3 所述, 设节点 x 与已在 DHT 网络中的节点 y 取得联系, 则节点 y 为节点 x 的引导节点, 负责新加入节点 x 的初始化工作。在节点在它的某个 k 桶中



插入引导节点 y 之前, x 需基于社交网络验证 y 是否为良性节点, 只有当 x 认为 y 为良性节点, x 才将 y 添加到路由表中, 记录信任路径 $p[x, y]$, 并向引导节点 y 发起查询操作请求来定位自己; 反之节点 x 拒绝将节点 y 添加到自己的路由表中, x 尝试与网络中的其他节点取得联系。DHT 网络中节点的退出不需要发布任何消息。

3.4.2 社交网络信任关系的变化

DHT 网络中利用社交网络节点间的信任关系识别 Sybil 节点, 社交网络信任关系的变化会导致 DHT 网络中节点间信任关系的变化。如图 4 所示, 信任路径 $p[A, C]$ 所依赖的信任关系是 $B \rightarrow E$, 如果节点 B 在社交网络中取消了对节点 E 的信任, 此时 A 需重新构造信任路径 $p[A, C]$ 。针对上述问题, 本文提出以下策略保持 DHT 网络和社交网络中信任关系的一致性。

算法 3 n -Start (Node x , Node y)

输入: Node x 为新加入的节点, Node y 为已在网络中的节点。

输出: Route (x) 为节点 x 的路由表。

初始化系统参数, 包括 Node x , Node y ,
Route (x)
//新加入网络的节点, 被分配一个随机生成的节点 ID
 $x \leftarrow \text{random ID}$
//基于社交网络验证节点 y 是否是一个良性节点
 $R \leftarrow \text{p-Search}(x, y);$
if $R \neq \text{NULL}$ then
// x 认为 y 是良性节点, x 将 y 加入自由路由表中

Route(x) $\leftarrow \{\text{Node } y, p[x, y]\}$

look_up(x); //查询请求定位 x

else

连接 DHT 网络中的其他节点;

end if

return Route (x)

(1) 信任关系 $x_1 \rightarrow x_2$ 失效时, 如果节点 x_1 处于在线状态, x_1 将通知依赖信任关系 $x_1 \rightarrow x_2$ 的信任路径 $p[x, y]$ 的起点 x 信任关系 $x_1 \rightarrow x_2$ 失效。 $p[x, y]$ 的起点 x 收到 x_1 的通知后, 查询其路由表中的信任路径 $p[x, y]$ 是否为依赖信任关系 $x_1 \rightarrow x_2$, 如果是, x 将节点 y 从路由表中移除, 并通知信任路径 $p[x, y]$ 中的节点信任路径已失效。

(2) 信任路径 $p[x, y]$ 的起点 x 定时轮询信任路径中的节点 $\{x_1, x_2, \dots, x_{n-1}, x_n\}$ 是否在线。由于社交网络中信任关系的变化相对 DHT 网络的抖动并不频繁, 因此 SSDHT 中轮询时间间隔设为 300 s, 且只有当信任路径中 $p[x, y]$ 某个节点 x_i 长时间离线 (SSDHT 中设为 3 600 s) 时, 节点 x 才重新构建信任路径 $p[x, y]$ 。

(3) 当节点 x_i 重新加入 DHT 网络时, 节点 x_i 向信任路径 $p[x, y]$ 的起点 x 查询信任路径是否存在, 并删除已失效的信任路径。

3.4.3 节点的负载均衡

假设节点 x 在社交网络中存在 n 个信任关系 $\{x \rightarrow x_1, x \rightarrow x_2, \dots, x \rightarrow x_{n-1}, x \rightarrow x_n\}$, 其中节点 $\{x_1, x_2, \dots, x_{n-1}\}$ 已加入 DHT 网络中, x_n 为新加入的节点。考虑如下情况, 当某个信任路径构造请求被转发到 x 点时, 理论状态下, $\{x_1, x_2, \dots, x_{n-1}, x_n\}$ 被选为下一跳节点的概率是相同的, 均为 $1/n$, 但节点 x_n 加入 DHT 网络之前, 节点 $\{x_1, x_2, \dots, x_{n-1}, x_n\}$ 中已经有若干信任路径的存在, 这种情况下节点加入的时间越早, 负荷越高, 越容易导致 DHT 网络出现负载不均衡的问题。

针对上述问题, 本文提出的负载均衡算法如算法 4 所示, 其中 $|r|$ 表示信任路径 r 的长度。当节点 x_n 新加入 DHT 网络时, 如果节点 x 的路由表中存在长度超过一定阈值 l 的信任路径, 节点将尝试利用新加入的节点 x_n , 重新构建该信任路径。

算法 4 load-balance (RL(x), ND(x))

输入: RL(x) 为 x 的路由表中保存的信任路径, ND(x) 为 x 在社交网络中的邻居节点。

输出: $OL(x)$ 为 x 在社交网络中已经加入 DHT 网络中的邻居节点。

初始化系统参数, 包括 $RL(x)$, $ND(x)$, $OL(x)$

$OL(x) \leftarrow x_n$; // x_n 为新加入的节点

// 是否存在长度超过一定阈值 l 的信任路径

for $p[x,y] \in RL(x)$ do

if $|p[x,y]| > l$ then

$p = p\text{-Search}(x, y)$ // 重新构建信任路径

if $|p| < |p[x,y]|$ then

$p[x,y] = p$

end if

end if

end for

return $OL(x)$

3.4.4 节点信任度更新

SSDHT 中信任关系具有不对称性、信任度动态性、信任度差异化的特点。因此, 节点构建信任路径时, 可将信任度作为一个加权系数, 一定范围内优先选择信任度高的节点作为下一跳节点。

节点间信任度用离散变量标示, 首先由用户根据实际社会网络中的信任度标示社交网络中存在直接信任关系的节点的初始信任度。路由表中的节点的初始信任度由节点的推荐节点 (如图 3 所示, 节点 R 是节点 A 路由表中的节点, 信任路径 $[A, R]$ 经过节点 F, 则节点 F 为节点 R 的推荐节点) 的信任度和信任路径的长度决定。推荐节点的信任度越高, 信任路径的长度越短, 信任路径中存在攻击节点的概率越低, 节点的初始信任度越高。节点信任度的更新依赖于直接推荐机制, 即节点依据直接交易记录更新路由表中的节点的信任度, 同时更新该节点的推荐节点的信任度。

4 实验及分析

本文基于 Facebook 和 Twitter 的数据集进行了实验验证, 对 SSDHT 的功能和性能进行了评

测, 包括路由表中良性节点的比例、信任路径平均长度, 并讨论了 SSDHT 带宽开销及应对网络抖动的能力。

4.1 实验数据及步骤

本文中使用的实验数据集来源于参考文献[26-27]公开的社交网络数据集, 其中 Facebook 为信任关系对等的社交网络^[26], Twitter 为信任关系不对等的社交网络^[27]。实际社会网络中一个人的社交关系系数被限定在一定范围之内 (详见参考文献[21]), 为更接近真实的社会网络, 实验中去除了数据集中节点社交关系系数小于 5 或者大于 500 的节点, 实验数据集见表 2。

表 2 实验数据集

数据集	节点数	边数	节点的平均度
Facebook 数据集	9 495	219 334	23.1
Twitter 数据集	13 440	768 670	57.1

实验中首先从实验数据集中随机选取一个节点作为初始节点, 每个节点加入 DHT 网络时随机生成一个 24 bit 的 ID 值, 之后在该节点社交网络中的邻居节点中随机选择一个加入 DHT 网络中, 不断地递归重复这个过程^[24]。与 KAD 算法相同, 新加入的节点在它的某个 k 桶中插入引导节点, 然后向它的引导节点发起查询操作请求来定位自己, 直至路由表处于稳定状态。上述过程中, 单个 Sybil 节点社团大小被限定为 δ 时, 即依赖某个信任关系 $x \rightarrow y$ 的信任路径的最大数量为 δ 时 ($x, y \in V(L) \cup V(S)$), 会有少量良性节点被其他良性节点拒绝添加到路由表中, 但由于社交网络快速收敛的特性, 当 $\delta \in O(\ln n)$ 时, 绝大部分良性节点都会被其他良性节点接受。实验中当 $\delta = 2 \ln n$ 、 $T = 10$ 时, 良性节点被其他良性节点拒绝的概率小于 1%, 其中 n 为网络规模, T 为查找次数。因此后续实验中的参数设置为 $\delta = 2 \ln n$ 、 $T = 10$ 。

本文第 4.2 节中的实验 1 中通过考查路由表中良性节点的比例, 评价 SSDHT 抵御 Sybil 攻击的



能力。第 4.3 节中的实验 2 通过考查信任路径的平均长度评价 SSDHT 的带宽消耗及应对网络抖动的能力。

4.2 路由表中良性节点的比例

实验 1: 在实验数据集中随机选择 β 个节点, 直到系统中存在 g 条攻击边, 攻击节点可生成任意数量的 Sybil 节点 ($0 \leq |V(S)| < +\infty$)。良性社团集合中的节点周期性地加入和离开, 而 Sybil 节点社团中的节点长时间保持在线, 并利用 DHT 网络不断进行节点路由表更新, 保证抖动环境下进行正常的路由过程, 使尽可能多的 Sybil 节点进入良性节点的路由表中。

实验结果如图 5、图 6 所示, 系统处于稳定状态时, 路由表中良性节点的比例随着攻击边 g 的增加而下降。如 Twitter 数据集中, 当攻击边数为 1 000 时, 良性节点比例为 98.7%; 当攻击边数为 2 000 时 ($k=5$), 良性节点的比例为 97.4%。良性节点社团的路由表中 Sybil 节点数量 τ 与社交网络中攻击边数 g 正相关。

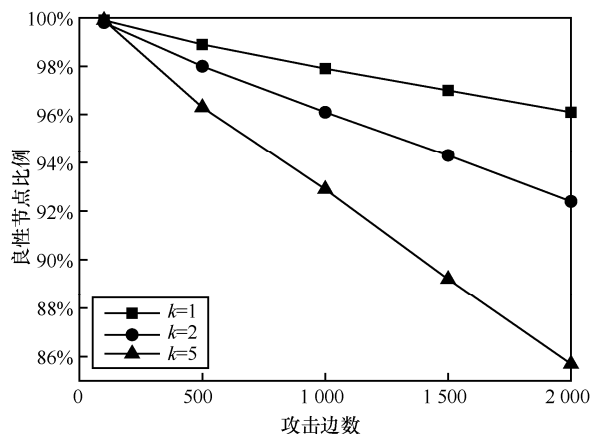


图 5 Facebook 数据集中路由表中良性节点的比例

提高 k 桶的值, 即增大路由表容量时, 良性节点社团的路由表中良性节点的比例随之提高, Sybil 节点的比例明显下降。如 Facebook 数据集中 $k=1$ ($g=1\ 000$), 良性节点比例为 92.9%; 当 $k=5$ 时, 良性节点比例为 97.9%。路由表中 Sybil 节点数 τ 没有随着路由表容量的增加而明显增加, 这说

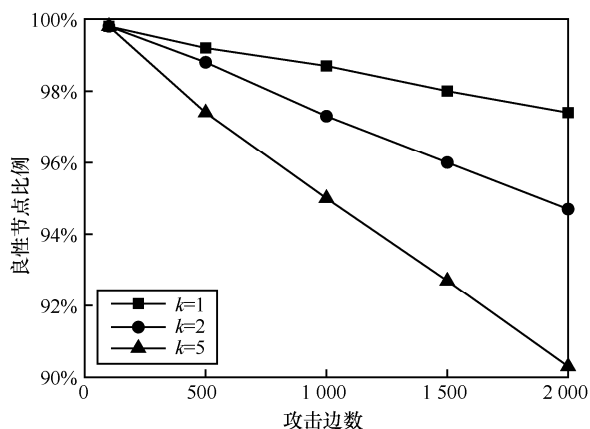


图 6 Twitter 数据集中路由表中良性节点的比例

明良性节点社团的路由表中 Sybil 节点的数量受单个 Sybil 节点社团的大小 δ 的限制。

4.3 信任路径的平均长度

实验 2: 统计良性节点社团路由表中信任路径的平均长度 γ , 考察并行查找次数 r 和 k 桶大小 (路由表容量) 对信任路径平均长度 γ 的影响。

实验结果见表 3, 信任路径的平均长度 γ 随着并行查找次数 r 的提高而降低, 如 Facebook 数据集中并行查找次数 r 从 1 提高到 10 时 ($r=5$), 信任路径的平均长度 r 从 34.3 下降到 12.3。信任路径的平均长度 γ 随路由表容量的增加而降低, 如在 Facebook 数据集中, k 桶的值从 1 增加 5 时 ($r=10$), 信任路径的平均长度 γ 从 17.4 降到 12.3。

表 3 信任路径的平均长度

k 桶大小	Facebook 数据集			Twitter 数据集		
	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
1	56.4	25.4	17.4	43.3	21.4	14.0
2	39.8	19.1	14.9	32.6	17.8	11.7
5	34.3	17.2	12.3	27.2	13.1	10.3

4.4 实验结果分析

良性社团节点的路由表中良性节点的比例反映了 SSDHT 对于 Sybil 攻击和路由攻击的防抵御能力。通过实验 1 可以看出, 良性社团节点路由表中 Sybil 节点的数量 τ 受攻击边的数量 g 和单个 Sybil 节点社团的大小 δ 的限制。当 $k=5$ 时, $g \approx n/5$

时, Facebook 数据集中良性社团节点路由表中良性节点比例为 96.1%, Twitter 数据集中路由表中良性节点为 97.4%。即使攻击边 g 的数量进一步增加, τ 的数量依然会受 g 的数量的限制。上述结果说明 SSDHT 能够有效地抵御 Sybil 攻击。

如本文第 3.2 节所述, 为保持社交网络与 DHT 网络信任关系的一致性, 信任路径 $p[x, y]$ 的起点 x 需定时轮询信任路径中的节点 $\{x_1, x_2, \dots, x_{n-1}, x_n\}$ 是否在线, 轮询时间间隔为 300 s。通过实验 2 可以看出, Facebook 数据集中信任路径的平均长度为 12.3, Twitter 数据集中信任路径的平均长度为 10.3 ($r=10, k=5$)。当节点的路由表中大小为 120 ($k=5$ 时, 理论上的饱和状态), UDP 报文的大小为 40 byte 时, Facebook 数据集中节点的平均带宽消耗为 0.19 kbit/s, Twitter 数据集中节点的平均带宽消耗为 0.16 kbit/s。

与 Whanau 方法^[23]的比较: Whanau 方法中节点路由表大小的复杂度为 $O(\lg n \cdot \sqrt{n})$, SSDHT 中为 $O(\lg n)$, 其中 n 为网络规模。以 Twitter 数据集为例, SSDHT 中节点的路由表饱和状态时为 120 ($k=5$), 而 Whanau 方法中节点路由表的大小大约为 500, 远高于 SSDHT, 且 Whanau 方法中节点路由表的大小随网络规模的增加而明显增加, 这使得 Whanau 方法不适合大规模网络。

与 X-vince 方法^[24]的比较: X-vince 方法中节点不能与其路由表中的节点直接通信, 只能通过社交网络中存在直接信任关系的节点进行消息的传递。X-vince 方法需实时轮询消息传递路径中的节点是否在线, 节点的平均带宽消耗为 4.8 kbit/s。当 X-vince 方法中消息传递路径存在节点离线时, 需立刻重新构造该消息传递路径。考虑 DHT 网络抖动, 在一段时间内, 10% 的节点处于离线状态时, X-vince 方法需重新构建 60% 的消息传递路径, 这使得 X-vince 方法不适合动态的网络。SSDHT 中节点平均的带宽消耗小于 0.2 kbit/s, 且信任路径中存在节点长时间离线时才需重新构建

该信任路径, 从而使 SSDHT 更适用于动态网络。

5 结束语

本文提出一种基于社交网络的 DHT 安全增强机制, 利用社交网络中的信任关系构造信任路径识别 Sybil 节点, 节点路由表更新时, 只将验证通过的非 Sybil 节点加入路由表中, 限制路由表中 Sybil 节点数的复杂度为 $O(g\delta)$ 。基于 Facebook 和 Twitter 数据集的实验中 SSDHT 中路由表良性节点比例在 90% 以上, 节点只需维护复杂度为 $O(\lg n)$ 的路由表, 平均带宽消耗小于 0.2 kbit/s。实验结果表明, SSDHT 适用于大规模的动态网络, 能够有效地防御 Sybil 攻击。

参考文献:

- [1] 中国互联网络信息中心. 第 41 次《中国互联网络发展状况统计报告》[R]. 2018.
CNNIC. The 41st "Statistical report on the development of China's internet"[R]. 2018.
- [2] 史建焘, 张宏莉. KAD 网络负载均衡技术研究[J]. 电信科学, 2012, 28(6): 68-72.
SHI J T, ZHANG H L. Study on load balancing of KAD network[J]. Telecommunications Science, 2012, 28(6): 68-72.
- [3] CHAN C, CHAN S. Distributed hash tables: design and applications[M]//Handbook of Peer-to-Peer Networking. Berlin: Springer Science, 2010: 257-280.
- [4] DOUCEUR J R. The Sybil attack[J]. Lecture Notes in Computer Science, 2002, 65(1): 251-260.
- [5] 肖禄, 黄韬, 刘江, 等. 一种基于物理拓扑的 DHT 物联网解析机制[J]. 电信科学, 2012, 28(6): 25-32.
XIAO L, HUANG T, LIU J, et al. A novel resolution service for internet of things based on physically topological DHT[J]. Telecommunications Science, 2012, 28(6): 25-32.
- [6] JAN M A, NANDA P, HE X, et al. A Sybil attack detection scheme for a forest wildfire monitoring application[J]. Future Generation Computer Systems, 2018(80): 613-626.
- [7] WANG C, ZHU L, GONG L, et al. Accurate sybil attack detection based on fine-grained physical channel information[J]. Sensors, 2018, 18(3): 878.
- [8] BOTT J, MILKAU U. Towards a framework for the evaluation and design of distributed ledger technologies in banking and payments[J]. Journal of Payments Strategy & Systems, 2016, 10(2): 153-171.
- [9] WANG G, MUSAU F, GUO S, et al. Neighbor similarity trust



- against sybil attack in P2P E-commerce[J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(3): 824-833.
- [10] ZHENG Y, LIN F, YANG Y, et al. Adaptive resource scheduling mechanism in P2P file sharing system[J]. Peer-to-Peer Networking and Applications, 2016, 9(6): 1089-1100.
- [11] JAN M A, NANDA P, HE X, et al. A Sybil attack detection scheme for a forest wildfire monitoring application[J]. Future Generation Computer Systems, 2018(80): 613-626.
- [12] RAJA K N, BENO M M. Secure data aggregation in wireless sensor network-Fujisaki Okamoto (FO) authentication scheme against sybil attack[J]. Journal of medical systems, 2017, 41(7): 107.
- [13] FREEDMAN M J, MORRIS R. Tarzan: a peer-to-peer anonymizing network layer[C]//The 9th ACM Conference on Computer and Communications Security, November 17-21, 2002, Washington, DC, USA. New York: ACM Press, 2002: 193-206.
- [14] RAJA K N, BENO M M. Secure data aggregation in wireless sensor network-Fujisaki Okamoto (FO) authentication scheme against sybil attack[J]. Journal of Medical Systems, 2017, 41(7): 107.
- [15] MARTI S, GANESAN P, GARCIA-MOLINA H. DHT routing using social links[M]. Berlin: Springer Berlin Heidelberg, 2005: 100-111.
- [16] DANEZIS G, LESNIEWSKI-LAAS C, KAASHOEK M, et al. Sybil-resistant DHT routing[C]//The 10th European Symposium on Research in Computer Security, September 12-14, 2005, Milan, Italy. Berlin: Springer, 2005: 305-318.
- [17] YU H, KAMINSKY M, GIBBONS P B, et al. Sybilguard: defending against sybil attacks via social networks[J]. IEEE/ACM Transactions on Networking, 2008, 16(3): 576-589.
- [18] YU H, GIBBONS P B, KAMINSKY M, et al. Sybillimit: a near-optimal social network defense against Sybil attacks[J]. IEEE/ACM Transactions on Networking, 2010, 18(3): 885-898.
- [19] BOYD S, GHOSH A, PRABHAKAR B, et al. Gossip algorithms: design analysis and applications[C]//IEEE INFOCOM 2005, March 13-17, 2005, Miami, FL, USA. Piscataway: IEEE Press, 2005.
- [20] FLAXMAN A D. Expansion and lack thereof in randomly perturbed graphs[J]. Internet Mathematics, 2007, 4(2-3): 131-147.
- [21] TRAN N, MIN B, LI J, et al. Sybil-resilient online content voting[C]//The 6th USENIX Symposium on Networked Systems Design and Implementation (NSDI'09), April 22-24, 2009, Boston, Massachusetts, USA. New York: ACM Press, 2009: 15-28.
- [22] LESNIEWSKI-LAAS C. A Sybil-proof one-hop DHT[C]//The First Workshop on Social Network Systems (SocialNets'08), April 1, 2008, Glasgow, Scotland, UK. New York: ACM Press, 2008: 19-24.
- [23] LESNIEWSKI-LASS C, KAASHOEK M F. Whanau: a Sybil-proof distributed hash table[C]//7th USENIX Symposium on Network Design and Implementation, April 28-30, 2010, San Jose, California, USA. New York: ACM Press, 2010: 3-17.
- [24] MITTAL P, CAESAR M. X-Vine: secure and pseudonymous routing using social networks[J]. Corr, 2012.
- [25] LIU B, WEI T, ZHANG C, et al. Improving lookup reliability in KAD[J]. Peer-to-Peer Networking and Applications, 2015, 8(1): 156-170.
- [26] REZVANIAN A, MEYBODI M R. Sampling algorithms for weighted networks[J]. Social Network Analysis and Mining, 2016, 6(1): 60.
- [27] KWAK H, LEE C, PARK H, et al. What is Twitter, a social network or a news media?[C]//The 19th International Conference on World Wide Web, April 26-30, 2010, Raleigh, NC, USA. New York: ACM Press, 2010: 591-600.

[作者简介]



李攀攀 (1983-), 男, 博士, 嘉兴学院数理与信息工程学院讲师, 主要研究方向为社交网络、网络安全、云计算等。



谢正霞 (1982-), 女, 嘉兴学院建筑工程学院工程师, 主要研究方向为社交网络、智能算法、云计算等。

乐光学 (1963-), 男, 博士, 嘉兴学院数理与信息工程学院教授, 主要研究方向为多云融合与协同服务、无线 mesh 网络与移动云计算、混成与嵌入式系统等。

马柏林 (1961-), 男, 博士, 嘉兴学院数理与信息工程学院教授, 主要研究方向为调和分析与小波分析、信号分析等。

陈丽 (1975-), 女, 博士, 嘉兴学院数理与信息工程学院讲师, 主要研究方向为移动通信、移动感知网络及机器学习等。

刘建华 (1978-), 男, 博士, 上海师范大学信息与机电工程学院副教授, 绍兴文理学院计算机科学与工程系副教授, 主要研究方向为无线传感器网络、云计算、网络与信息安全等。