第 42 卷 第 6A 期
2015 年 6 月

计 算 机 科 学
Computer Science

Vol. 42 No. 6A
June 2015

基于混合双层模型的 DHT 网络路由表快照算法

余 杰[1] 李 强[2] 李莎莎[1,2] 马 俊[1] 李舟军[2]

(国防科学技术大学计算机学院 长沙 410073)[1] (北京航空航天大学计算机学院 北京 100000)[2]

摘 要 DHT 网络是目前应用最广泛的 P2P 协议,路由表是其进行自组织的关键组件。由于 DHT 网络的完全分布特点,对其全局路由表快照进行测量是一个研究难点和热点。提出了基于混合双层模型的 DHT 路由表快照算法:首先通过引入路由查询重复度这一重要概念来定义 DHT 网络快照和路由表快照采集的效率,然后提出了先宽度优先搜索后深度优先搜索的全局快照混合搜索策略;最后基于路由表的不均匀特性提出了路由表快照自适应搜索策略。在 Kad 网络上的真实实现表明,全局快照混合搜索策略的平均效率比 Blizzard 高 91.2%,比宽度优先搜索高 64.5%,比深度优先搜索高 27.4%;路由表快照自适应搜索策略在 $g=5$ 时具有最佳的路由表快照采集效率,比随机搜索策略高 187.4%,比 $g=7$ 时高 38.9%。

关键词 DHT,路由表,双层模型,混合策略,自适应策略

中图法分类号 TP393 文献标识码 A

Hybrid Method for Capturing Snapshot of Routing Table of DHT Networks

YU Jie[1] LI Qiang[2] LI Sha-sha[1,2] MA Jun[1] LI Zhou-jun[2]

(College of Computer, National University of Defense Technology, Changsha 410073, China)[1]

(School of Computer Science and Engineering, Beihang University, Beijing 100000, China)[2]

Abstract DHT network is the most widely used P2P protocol in the world and routing table is the key component for its self-organization. It's difficult to measure the global routing table of a DHT network due to its feature of totally distributed architecture. In this paper, we propose a hybrid method to capture the snapshot of the routing table of DHT network. We first introduce a repetition degree to define the efficiency when crawling the snapshot. Then, we propose a DHT snapshot method that uses breadth-first search first and then changes to depth-first search. Finnally, we propose a self-adaptive method to crawl the routing tables based on the un-even feature of routing table. The experiment on Kad network shows that, the DHT snapshot method is 91.2% better than Blizzard, 64.5% better than breadth-first and 27.4% better than depth-first. Routing table snapshot method reaches the best when $g=5$ and it is 187.4% better than the random method and 38.9% better than $g=7$.

Keywords DHT, Routing table, Two layer model, Hybrid strategy, Self-adaptive strategy

近十年来,对等网络(Peer to Peer,P2P)技术的相关研究在国际上获得了广泛关注。随着 P2P 协议设计[1]、路由算法[2]、搜索优化[3]等技术的日趋成熟,当前在 P2P 研究方向最活跃的领域之一是 P2P 网络测量。P2P 网络测量是进行 P2P 网络性能分析、用户行为分析、安全缺陷分析、安全攻击分析等研究的重要基础。由于 P2P 网络是建立在因特网上的大规模分布式协作网络,针对这类网络进行测量,不仅要考虑网络的运行状况和自身的性能指标,还需要关注用户行为与 P2P 系统性能之间的相互关系,以刻画真实的 P2P 应用网络与其协议的设计初衷是否一致。P2P 网络测量技术也是进行 P2P 网络拓扑研究的重要前提和基本保障。

从 P2P 网络的拓扑结构来看,其经历了集中式拓扑、全分布非结构化、全分布式结构化 3 个阶段。由于单点失效、搜索效率等因素的限制,目前应用最广泛的 P2P 网络是采用分布式哈希表(Distributed Hash Table,DHT)技术的完全分布式结构化拓扑,DHT 将文件中抽取的部分信息(通常是文件名)哈希成唯一的关键字(key),然后通过关键字来定位文件。最典型的 DHT 算法有 Chord、CAN、Pastry、Kademlia[4] 和 Tapestry。P2P 应用系统中获得广泛使用的 DHT 网络绝大部分都是基于 Kademlia 协议,如 eMule/aMule、Bittorrent 等,在全球有数百万的用户同时在线。本文的研究重点是基于 Kademlia 协议的 DHT 网络的路由表快照。

2003 年,R. Bhagwan 等[5]对 Overnet 的主机可用性(host availability)进行了测量,通过对 2400 个节点连续 14 天的测量,发现了如下结论:由于许多节点周期性地改变 IP 地址,Kad 网络存在 IP 别名问题;由于 Overnet 网络的动态性,节点的平均可用性随着观察周期的增长而减少。然而,Overnet 从 2006 年 9 月开始由于法律原因已经停止开发和运营。

本文受国家自然科学基金项目(61103015,61303190,61303191)资助。

余 杰(1982—),博士,副研究员,主要研究方向为分布式系统与操作系统;李 强(1982—),博士生,主要研究方向为 P2P 网络;李莎莎(1982—),博士,助理研究员,主要研究方向为自然语言处理;马 俊(1982—),博士,助理研究员,主要研究方向为网络安全;李舟军(1963—),博士,教授,博士生导师,主要研究方向为分布式系统与自然语言处理。

2006 年,R. Brunner[6]通过修改 aMule 客户端,对 Kad 的性能进行监控,分析了 Kad 网络的关键字 keyword、迭代过程 iteration process、主机可用性、发布过程 publishing 等。2007 年,M. Steiner 等[7]对 eDonkey2000 中的 Kad 网络进行了全局测量,对 Kad 网络的节点数目和分布、会话长度(session length)、会话间隔(inter-session time)、永久离开等特征进行了测量,并通过长达 6 个月的测量,指出 Kad 网络中存在 ID 别名问题。2005 年至 2009 年,研究人员针对 BitTorrent 系统中有 Tracker 和无 Tracker 的 DHT 网络都进行了比较深入的研究,对 BitTorrent 进行了测量、分析与建模[8],并分析了 BitTorrent 系统的可用性[9]。同时,S. Crosby 等[10]对 BitTorrent 系统中的 Mainline 和 Azureus 进行了分析、测量和连接性研究。本文作者也在 DHT 网络的测量、安全、效用等方面开展了较深入的研究,比如发现了 Kad 网络中大量的 ID 重复现象[11]、设计了 Kad 网络的 Cache 效用模型[12]、提出了 Kad 网络中 Sybil 攻击团体检测技术[13]等。

前期的 DHT 网络测量工作主要集中于 P2P 用户行为和流量特征的测量,而很少对 DHT 关键部件和过程的安全性进行测量。由于 DHT 网络的完全分布式和动态性特征,对其关键部件的全局测量需要大量的资源,同时也会花费较长的时间,从而会影响全局快照的准确性和实时性,因此,需要对全局快照的采集算法进行深入研究,提出高效、准确的路由表采集算法。本文在充分分析 DHT 网络结构、路由表组成以及路由过程的基础上,提出了基于混合双层模型的 DHT 路由表快照算法,首先采集 DHT 网络全局节点快照,然后逐个采集活跃节点的路由表快照。通过引入路由查询重复度这一重要概念来定义 DHT 网络快照和路由表快照采集的效率,通过分析宽度优先搜索和深度优先搜索的特点,提出了高效的全局快照混合搜索策略,通过分析路由表的不均匀特性,提出了路由表快照自适应搜索策略。在 Kad 网络上的真实实验表明,本文提出的全局快照混合搜索策略明显优于 M. Steiner 等[14]提出的 4-bit 子空间宽度优先搜索策略 Blizzard、宽度优先搜索和深度优先搜索,路由表快照自适应搜索策略在 $g=5$ 时具有最佳的路由表快照采集效率。

1 DHT 网络路由表

1.1 路由表的组成

路由表维护了节点运行过程中定位或接收到的联系人(contact)信息,并将这些信息按照合理的方式进行组织,以便进行快速地查找、更新和替换。

1.2 路由表的作用

路由表在 DHT 网络的运行与维护过程中都扮演了关键作用。

·启动过程(Bootstrapping Process):节点启动时,将上次运行后保存在本地的联系人信息加载入路由表,然后节点从路由表中随机选择联系人并向其发送 Bootstrapping 消息,活动的联系人会返回若干个新的联系人信息,节点成功加入 DHT 网络。

·路由过程(Routing Process):节点需要定位另外一个节点时,先从路由表中选择与之最近的联系人,向这些联系人迭代或递归地发送 Routing 请求,直到找不出更近的联系人。

·搜索过程(Searching Process):节点首先需要根据路由表在 DHT 网络中定位距离搜索目标最近的节点集合,然后向这些节点发出 Searching 消息,查找关键字或资源信息。

·发布过程(Publishing Process):节点首先需要根据路由表在 DHT 网络中的定位距离发布目标最近的节点集合,然后向这些节点发出 Publishing 消息,保存关键字或资源信息。

2 基于混合双层模型的 DHT 路由表快照算法

2.1 算法框架

由于 DHT 网络是完全分布式结构,针对其路由的快照采集包含两个步骤:

第一步 采集到当前 DHT 网络中的所有节点,称为全局快照搜索;

第二步 针对每个存活节点,采集其路由表快照信息,称为路由表快照搜索。

其算法框架的描述如表 1 所列。

表 1 DHT 网络全局路由表采集算法框架

```
1   /*全局数据结构*/
2   Vector<Node> nodes;// 当前 DHT 网络中的所有节点集合
3   Vector<Vector<Node>> routingTables ;//当前 DHT 网络的全局路由表
4   /*DHT 网络全局节点快照采集算法*/
5   void getDHTSnapshot()
6   {
7     foreach node∈nodes
8       k=getTargetID();//select target k
9       send routing request with target k to node;
10      waiting until receiving a response message;
11      foreach peer∈message
12        if nodes.contains(peer)=false
13          nodes.add(peer);
14  }
15  /*单个节点路由表快照采集算法*/
16  void getSingleRoutingTable(node)
17  {
18    Vector<Node> routingtable;//该节点的路由表
19    While(no more peers are found)
20    {
21      t=getTargetID();//select target t
22      send routing request with target t to node;
23      waiting until receiving a response message;
24      foreach peer∈message
25      if routingtable.contains(peer)=false
26        routingtable.add(peer);
27    }
28    routingtables.add(routingtable);
29  }
30  /*双层模型的 DHT 路由表快照算法*/
31  void getDHTRoutingTables()
32  {
33    nodes.add(initalnode);//设置初始节点
34    getDHTSnapshot(); //获取全局节点信息
35    foreach node∈nodes
36      getSingleRoutingTable(node);
37  }
```

2.2 基本搜索策略

上述算法均通过发送路由查询请求来采集接收节点的路由表信息或查找更多的存活节点。在发送路由查询请求时,需要在请求消息中设置一个目标 ID t(表 1 中第 8 行和第 21 行),接收节点在其路由表中寻距离 t 最近的 n 个联系人并返回给发送者。

定义 1(路由查询重复度 α) 假设每次路由查询响应消息中包含 n 个联系人信息,其中有 m 个联系人已经出现在之前采集的联系人集合中,则定义 $\alpha=m/n$。显然,$0\leqslant\alpha\leqslant1$。

要提高快照采集效率，必须要用尽可能少的查询消息获得尽可能多的联系人信息，也即 α 越小越好。理想情况下，最优的搜索策略是确保查询者每次获得的路由查询响应中包含的联系人信息都与之前得到的联系人信息完全不重复（即表 1 中第 12 行和第 25 行永远为 true，也即 α 永远为 0）。

不同的路由请求目标 ID 选择策略（对应于表 1 中的 *getTargetID()* 方法的不同实现），决定了路由响应消息中包含的联系人信息。总体而言，有两种基本的搜索策略。

定义 2（宽度优先搜索） 路由查询请求中的目标 ID 均匀分布在 ID 空间。这样在每个路由响应消息中包含的联系人会尽可能地分布在不同的区域中，采集到的节点信息就会均匀分布在整个 ID 空间。

定义 3（深度优先搜索） 路由查询请求中的目标 ID 与请求接收者的 ID 相同。这样在每个路由响应消息中包含的联系人会尽可能地分布在与接收者相同的区域中。

在 DHT 网络中，节点 P 的路由表中通常具有数十至数百个联系人，这些联系人并不是均匀分布在 ID 空间的。通常情况下，大部分联系人都是在 P 附近的 DHT 节点，这样才会确保每次路由查询迭代能够离目标节点更近。而上述的宽度优先搜索策略和深度优先搜索策略都没有考虑到路由表中的这种不均匀特性。

2.3 全局快照混合搜索策略

本节分析 DHT 网络全局节点快照采集算法中应采取的最佳搜索策略。下面首先分析在进行 DHT 网络全局快照采集时采用不同策略的特点。

一方面，使用宽度优先搜索策略是在整个 ID 空间进行联系人采集，在搜索开始时会具有较小的路由查询重合度，但在若干次迭代后会获得更多的重复联系人，从而导致搜索收敛缓慢。因此，可以认为宽度优先搜素策略的特点是：启动快，降速也快。

结论 1 使用宽度优先搜索策略进行 DHT 全局快照采集时，路由查询重复度 α 从 0 逐渐增加到 1。

另一方面，由表 1 可知，由于每次的路由请求接收者都不同，使用深度优先搜素策略每次所发现的 ID 空间基本上是相同的，也即它所得到联系人信息与已知信息的重复数目基本上是稳定的。因此，可以认为深度优先搜素策略的特点是：启动慢，降速也慢。

结论 2 使用深度优先搜索策略进行 DHT 全局快照采集时，路由查询重复度 α 基本稳定在某个固定值 b。

要高效地获取全局节点信息，必须让采集算法具有启动快、降速慢的特点。由于路由查询重复度 α 值在每次查询迭代时都可以实时计算出来。因此，本文提出一种全局快照混合搜索策略。

算法 1 全局快照混合搜索策略

在进行 DHT 网络全局快照采集时，首先采用宽度优先搜索，直至 α 值从 0 增长到 b，然后切换至深度优先搜索，直至获得所有节点信息。

需要说明的是，算法 1 不仅分别利用了宽度优先搜索策略和深度优先搜索策略的优势，还通过宽度优先搜索策略为优先搜索策略获得了在全局 ID 空间更加均匀的启动节点集合，从而使得深度优先搜索策略可以在不同的区域并行搜索，在一定程度上降低搜索后期 α 增加的速度，从而进一步提高全局快照采集效率。

2.4 路由表快照自适应搜索策略

本节分析了 DHT 网络单个路由表快照采集算法中应采取的最佳搜索策略。

路由表中的联系人信息通常按照不规则多叉树的方式组织，距离节点越远，联系人越少。由于每个节点的路由表联系人总数不超过 $O(10^2)$ 数量级，并且是实时动态更新的，路由表随着节点加入 DHT 网络的时间和所处 ID 空间的不同而具有较大差别，不宜用固定的算法来进行快照采集。本文提出一种自适应的路由表快照搜索策略。

算法 2 路由表快照自适应搜索策略

在进行 DHT 网络节点的路由表快照采集时，采用基于 α 值的动态宽度优先搜索算法：首先将整个 ID 空间划分为 g 个子空间 zone，对子空间按照距离节点由远至近进行编号，设为 zone_i，其中 $1 \leqslant i \leqslant g$；然后在每个子空间 zone_i 中采用剩余迭代次数为 $(1-\alpha)2^i$ 的宽度优先搜索策略。

上述算法中，距离节点越近的子空间，剩余迭代次数越多，直到某次迭代后计算出的路由查询重复度 α 为 0。

3 实验验证

本文在 Kad 网络上进行算法的可行性验证。Kad 是基于 Kademlia 协议的一种 DHT 实现，并广泛应用于电驴（eMule/aMule）等系统中，目前拥有数百万的在线用户。

3.1 实验环境

本文使用 Java 语言编程实现了上述算法，并将其部署在一个通用 PC 电脑上。系统的 CPU 为 1.8 GHz、内存为 4GB，安装 Ubuntu Kylin 14.04 系统，Java 版本为 0penJDK 1.7，网络带宽为 10Mbps。为了减小实验对 Kad 网络正常运行的影响，本文在实验程序中控制了采集报文的发送频率，并尽量缩短每次实验采集时间。图 1 是本文实现环境的截图。

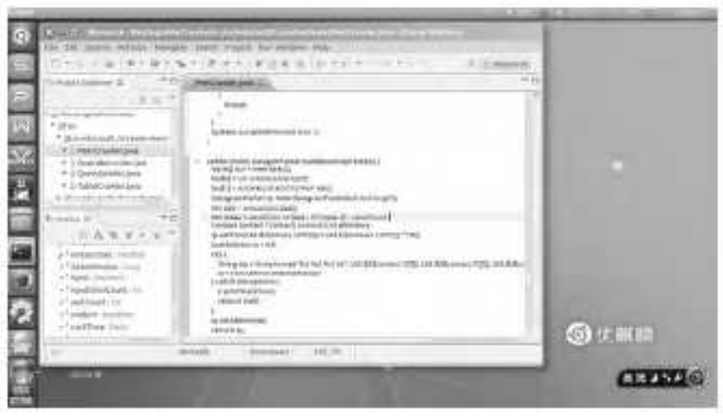

图 1 实现桌面环境截图

3.2 DHT 网络全局快照搜索策略对比

M. Steiner 等[14]提出了一种基于 4-bit 子空间宽度优先搜索的 Blizzard 策略，该方法通过对每个节点连续发送 16 个目标 ID 均匀分布在 4-bit 空间的路由请求消息来获取 DHT 全局快照。为了对比，本文在实验中也实现了该算法。因此，本实验主要对比以下 4 种搜索策略：Blizzard、宽度优先搜索、深度优先搜索和混合搜索。实验过程中，取 $b=0.5$。

为了减小 DHT 动态变化对实验的影响，本文在每天不同时间共进行 10 组实验，每组实验运行 10 分钟。每种策略所采集到的全局快照节点数目随时间的变化情况如图 2 所示。

由图 2 可知：

1）Blizzard 的效率最低。这是由于通常一个路由表中只有 40～120 个联系人，按照 4-bit 空间发送请求获得的 $16 \times 8 = 128$ 个联系中有大量的重复。

2）宽度优先搜索和深度优先搜索两种策略的增长趋势与

（下转第 270 页）

间进行分析，得出了理论上限值，对于 iBGP 定时器、WRATE 设置以及 iBGP 收敛的研究具有很好的借鉴意义。本文还对基于 RCP 的 MP-RCP 协议进行对比，验证了 MP-RCP 具有较好的性能。

参 考 文 献

[1] Mühlbauer W, Maennel O, Uhlig S. Building an as-topology model that captures route diversity[C]// Proc. of ACM Sigcomm'06. Pisa, Italy: ACM Press, 2006: 195-206

[2] Walton D, Retana A, Chen E, et al. Advertisement of Multiple Paths in BGP[EB/OL]. http://www. draft-walton-bgp-add-paths-06. txt, 2008

[3] van den Schrieck V, Francois P, Pelsser C, et al. Preventing the Unnecessary Propagation of BGP Withdraws[C]// Processing of Information IIFIP International Federation (NETWORKING 2009). LNCS, 2009: 495-508

[4] Caesar M, Caldwell D, Feamster N, et al. Design and Implementation of a Routing Control Platform[C]// Proc. of NSDI '05. Boston, MA Berkeley, CA, USA: USENIX Association, 2005: 15-28

[5] 赵丹. 基于逻辑集中控制的网络路由关键技术研究[D]. 长沙：国防科技大学, 2013

[6] 程柏林，胡乔林，陈新，等. MP-RCP：基于 RCP 的快速恢复 iBGP 协议[J]. 计算机应用与软件, 2014(1): 127-131, 147

[7] Pei D, Zhang Bei-chuan, et al. An analysis of convergence delay in path vector routing protocols[J]. Computer Networks, 2006, 50(3): 398-421

[8] Qiu J. simBGP: a lightweight event-driven BGP simulator[EB/OL]. http://www. bgpvista. com/simbgp. php, 2009

(上接第 265 页)

本文的结论一和结论二基本一致。

3) 混合搜索策略综合利用了宽度优先搜索和深度优先搜索的优势，10 分钟内其平均效率比 Blizzard 高 91.2%，比宽度优先搜索高 64.5%，比深度优先搜索高 27.4%。

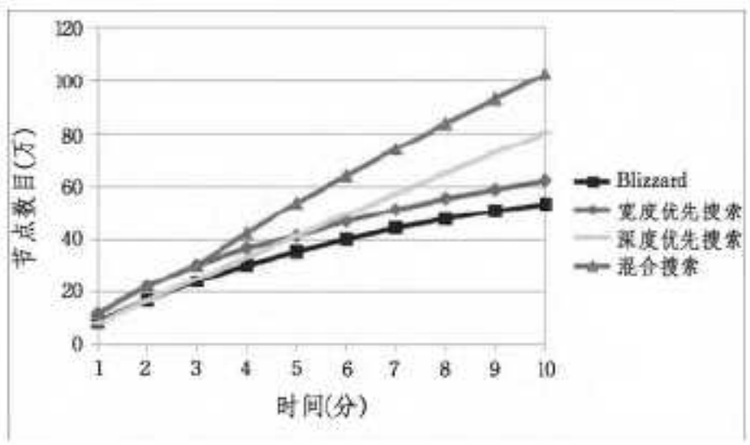

图 2 DHT 网络全局快照采集算法对比

3.3 路由表快照搜索策略对比

通过前期测量可知，Kad 节点的路由表中通常只有 40～120 个联系人。因此我们设置 g 分别为 3、5、7、9(将 g 设置为奇数，是为了将节点 ID 放在 $zone_g$ 的中间，提高检索效率)。为了对比，本文还实现一个随机算法，每次随机选择目标 ID 发送给接收者。

为了减小 DHT 动态变化对实验的影响，本文在每天不同时间共进行 10 组实验，每组实验随机选择 1000 个节点对其进行路由表快照获取。每种策略采集完每个节点路由表所用的路由请求报文平均数如表 2 所列。

表 2 不同搜索策略所使用的报文平均数

算法	随机	g=3	g=5	g=7	g=9
报文数	27.3	13.6	9.5	13.2	16.4

需要注明的是，$g=3$ 时平均使用了 13.6 个报文(上限是 2+4+8=14 个)，对于部分联系人较多的节点，该策略不能完全采集其路由表快照信息。

由表 2 可知：

1) 随机搜索策略的效率最低。这是由于它没有考虑到节点路由表的不均匀特性。

2) 对于 Kad 网络，选择 $g=5$ 具有最佳的路由表快照采集效率，比随机搜索策略高 187.4%，比 $g=7$ 时高 38.9%。

结束语 本文通过分析 DHT 网络的节点分布特性和路由表分布特性，提出了一个采集 DHT 网络全局路由表快照的混合双层模型。在 Kad 网络上的真实实验验证了本文提出算法的有效性。

下一步将使用现有算法测量更多 DHT 网络的全局路由表，分析各个路由表之间的关系、路由表存在的潜在风险，并深入研究各种路由表攻击的检测技术。

参 考 文 献

[1] Stoica I, Morris R, Karger D, et al. Chord: A scalable peer-to-peer lookup service for Internet applications[C]// Proceedings of SIGCOMM'01. 2001: 149-160

[2] Han J, Liu Y. Rumor Riding: Anonymizing Unstructured Peer-to-Peer Systems[M]. IEEE ICNP, Santa Barbara, California, USA, November, 2006

[3] 刘琼，徐鹏，杨海涛，等. Peer-to-Peer 文件共享系统的测量研究[J]. 软件学报, 2006, 17(10): 2131-2140

[4] Maymounkov P, Mazières D. Kademlia: A Peer-to-Peer Information System Based on the XOR Metric[C]// Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). 2002: 53-65

[5] Bhagwan R, Savage S, Voelker G. Understanding availability[C]// Proceedings of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS'03). 2003: 256-267

[6] Brunner R. A performance evaluation of the Kad-protocol[M]. Master Thesis, 2006

[7] Steiner M, Biersack E W, Ennajjary T. Actively monitoring peers in KAD[C]// Proceedings of the 6th International Workshop on Peer-to-Peer Systems (IPTPS'07). 2007

[8] Guo L, Chen S, Xiao Z, et al. Measurement, analysis, and modeling of BitTorrent-like systems[C]// Proceedings of IMC'05. 2005

[9] Neglia G, Reina G, Zhang H. Availability in BitTorrent Systems [C]// Proceedings of INFOCOM'07. 2007

[10] Jimenez R, Osmani F, Knutsson B. Connectivity Properties of Mainline BitTorrent DHT Nodes[C]// Proceedings of P2P'09. 2009

[11] Yu J, Fang C, Xu J, et al. ID repetition in Kad[C]// Proceedings of IEEE P2P'09. 2009

[12] Yu J, Lu L, Li Z, et al. A simple effective scheme to enhance the capability of web servers using P2P networks[C]// Proceedings of ICPP'10. 2010

[13] 李强，李舟军，周长斌，等. Kad 网络中 Sybil 攻击团体检测技术研究[J]. 计算机研究与发展, 2014, 51(7): 1614-1624

[14] Steiner M, Carra D, Biersack E W. Long term study of peer behavior in the KAD DHT[C]// IEEE/ACM Trans. Networking. 2009