第 35 卷 第 1 期 计 算 机 工 程 2009 年 1 月
Vol.35 No.1 Computer Engineering January 2009
·博士论文· 文章编号：1000—3428(2009)01—0001—03 文献标识码：A 中图分类号：TP311

基于 Skip Graph 的 P2P 覆盖网信任证链发现

徐 浩[1,2,3]，阎保平[1]

(1. 中国科学院计算机网络信息中心，北京 100080；2. 中国科学院计算技术研究所，北京 100080；3. 中国科学院研究生院，北京 100039)

摘 要：提出一种基于 Skip 图 P2P 覆盖网模型的分布式信任证存储和发现机制。为保证索引和发现效率，采用 RT_1^T 语言作为信任证描述语言，其语义属性作为 Skip 图中的索引关键字，构造了基于关键字前缀相似和支持范围查询的 P2P 覆盖网。通过试验评测和分析，该机制具有较高的查询效率和负载均衡机制。

关键词：Skip 图；信任证；信任证链；点对点

Skip Graph-based Credential Chain Discovery on P2P Overlay Network

XU Hao[1,2,3], YAN Bao-ping[1]

(1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100080; 2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080; 3. Graduate University of Chinese Academy of Sciences, Beijing 100039)

【Abstract】 This paper proposes a peer-to-peer framework, which adopts Skip graph named CredentialIndex as the overlay network for credential storage and credential chain discovery. To guarantee discovery efficiency, CredentialIndex schemes RT_1^T policy language and extracts its semantic attributes as indexing keys in Skip graph. The P2P overlay network is constructed to aggregate prefix similar keys to support range query and keep load balance on peer nodes. Evaluation shows that the CredentialIndex system performs considerable credential chain discovery efficiency and reasonable load balance.

【Key words】 Skip graph; credential; credential chain; peer-to-peer

随着分布式技术的发展，跨安全域的资源共享与访问的安全问题已逐步成为研究热点。一些研究者为此提出了信任管理机制(如RT[1]等)和自动信任协商机制(如TrustBuilder[2]等)，它们多是基于委托授权方法来实现信任关系建立与访问控制。其决定访问控制的过程就是在分布式网络中查找一条从资源拥有者到请求者的授权信任证链。其核心问题就是判断并找出这条链，即信任证链发现[3]。然而，上述系统并没有很好地解决分布式存储大量信任证和信任证链发现的问题。本文采用Skip图[4]作为信任证索引数据结构，构造P2P覆盖网，使得信任证较为均衡地分布在P2P节点上；利用Skip图分布式数据结构的特性设计出具有前缀匹配和范围查找的 RT_1^T 信任证搜索算法，结合信任证图机制给出分布式环境中信任证链发现的算法。

1 信任证劈分

P2P索引算法使用(key, vlaue)对来保存索引信息，达到高效路由和检索。本文使用的RTML[1]是一种结构化的XML语言，需要通过劈分机制形成键值，从而实现信任证的索引，同时，信任证劈分机制应尽量保留和准确表达原有信任证中的内容信息。本文将提取部分关键信息来劈分RTML中描述的 RT_1 的信任证。

RTML 是一种实现 RT_1^T 的基于 XML 的策略语言，描述了信任证和策略的定义。由于 RT_1 和 RT_1^T 的差别仅限于它们支持的信任证种类上有所不同，因此本文仅讨论 RT_1 信任证。

定义 1 一个信任证描述文档可以表示为一个五元组：C=(Issuer, Subject, Role, Validity, Signature)，其中，Issuer 为签发者公钥；Subject 为主体公钥；Role 为角色；Validity 为有效期；Signature 为签名。

本文的 P2P 索引系统利用关键字作为索引来完成资源的查找。为了支持复杂的信任证查询，按照查找信任证所需信息的重要程度，本文抽取 XML 文档中的签发者、主体和角色这 3 个元素作为索引关键字。

另外，RTML 文档中的元素和属性是以树的形式存储的，和基于索引的搜索算法类似。本文通过下列规则来生成关键词，即索引键值：

(1)抽取当前元素的第 1 个属性值，保留定义名称和 name 属性信息。

(2)使用分隔符“/@”将当前属性和元素分开，使用“/”将元素分开。

(3)如果子元素或者子属性没有被抽取，则使用当前元素来识别。即可以使用当前元素的劈分片断来识别其子元素或者子属性。这一规则利用了 Skip 图保持键值局部性特点。

(4)当前仅抽取 principal, roleterm 和 parameter 3 个元素。由于 Skip 图能保持键值局部性和树形结构特点，因此很容易

基金项目：国家“863”计划基金资助项目(2002AA104240, 2004AA104240, 2006AA01A106, 2006AA01A120)；国家自然科学基金资助项目(90412011)；国家科技部基金资助项目(2003DKA5G015)

作者简介：徐 浩(1978－)，男，助理研究员、博士研究生，主研方向：网格安全与中间件，网格应用；阎保平，研究员、博士、博士生导师

收稿日期：2008-05-08 **E-mail：**morrise@cnic.ac.cn

进行扩展，以抽取更多元素。

根据上述规则，表 1 为一个信任证实例劈分成的片段。

表 1　RT_0信任证片段描述

简称	片段
Isu	/principal@IssuerKey=”/X9TgR11…”
Sub	/principal@SubjectKey=”ujN6AfA…”
Rol	/roleterm@name=”socketPerm”
Hst	rol/parameter@name=”host”
Prt	rol/parameter@name=”port”

2　P2P 覆盖网中的信任证查找

通过对信任证文档劈分得到(key, value)对后，需要将这些关键字(key)分发到 P2P 网络中，实现信任证的分布式存储和检索。在本文设计的 CredentialIndex(简称 CI)系统中，利用 Skip 图来构造 P2P 的覆盖网。

2.1　Skip 图

Skip图是一种分布式数据结构，是Skip List在分布式环境下的延续和发展，其搜索的主要思想是贪婪算法。在Level0的链表中，所有节点有序分布组成双向链表，相互间隔为 1。随着Level的增加，所有节点仍出现在当前层的不同链表中，节点间间隔增大的程度由P值决定。其中，$P=|\Sigma|^{-1}$，Σ表示字母表中字母数量，在一般的示例中，取$\Sigma=\{1, 0\}$。不同层上的不同链表通过关系向量来表示，关系向量由系统设定的字母表Σ随机生成，关系向量中标识长度就是Level的值，相同标识长度下的不同串表示当前层上的不同链表。

如图 1 所示为一个 3 层的简单 Skip 图。若从节点 20 需要查找节点 39，则从节点 20 在第 2 层开始查找邻居，发现没有节点 39，则下降一层，发现 39 为其邻居，从而发现节点 39。

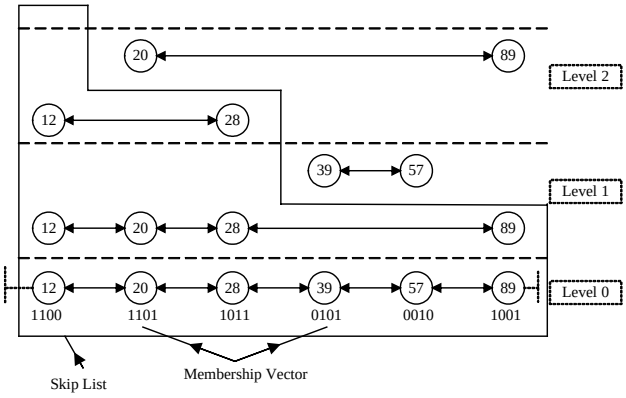

图 1　一个简单的 Skip 图

采用 Skip 图作为 P2P 覆盖网具有以下 3 个优势：(1)区别于 DHT 系统，Skip 图不使用哈希函数，使得信任证的语义信息被保持；(2)Skip 图支持范围查询、前缀查询等复杂查询；(3)Skip 图保持了键值之间的局部特性，即相似的键值将分布在邻居节点上。

2.2　P2P 覆盖网的构造

构造 P2P 覆盖网需实现 2 种类型的操作：关键字维护和节点维护。关键字维护包括查找、增加、更新和删除；节点维护包括节点加入和离开。

(1)关键字查找：该操作可以提高 Skip 图的查找效率。它允许不同信任证相同劈分的存在，同时还能够在查找不到关键字时给出离关键字最近的值。基于 Skip 图的关键字查找操作，其时间复杂度为 $O(\log N)$，消息传递开销为 $O(\log N)$，其中，N 为基于 Skip 图的覆盖网中节点个数。

(2)关键字插入和删除：插入操作和 Skip 图中的 insertOp 操作基本相同，仅仅在处理具有相同关键字时，需要改变 link 操作，将相同关键字及其值插入已有的节点中。关键字的删除操作同样需要使用类似的方法改变 link 操作。插入和删除操作的时间复杂度为 $O(\log N)$，消息传递开销为 $O(\log N)$，当签发和撤消信任证时才执行插入与删除。

(3)关键字更新：利用查找操作，找到该关键字，并执行 update 操作更新信任证(当信任证过期后 renew 时才使用该功能)。

(4)节点加入和离开：当一个新的关键字无法插入到任意当前节点时，需要调用节点加入操作；或者当前节点为热点(关键字存储大于某个临界值)时，需要调用节点加入操作来实现该节点的负载均衡。该操作的时间复杂度为 $O(\log N)$，消息传递开销为 $O(\log N)$。当节点离开时，需要利用修复机制恢复 Skip 图的连接，并且检查信任证的丢失分片。然后将无法恢复的信任证关键字重新分布到其他 Skip 图的节点上。

3　信任证链发现

采用 CI 系统，信任证链发现过程就是通过不断发起包含部分信任证关键字的查询来获得满足要求的信任证，从而逐步构成信任证图的过程。对于同一个信任证集合来说，查找过程中构造的信任证图越小，表明通过网络获取的信任证数目越少，信任证链发现的效率就越高。

CI 系统通过 3 个步骤完成信任证链的发现：(1)按照信任证图机制判断当前需要查询的信任证所包含的关键字，即签发者、主体和角色集合。(2)在 Skip 图中查找出所有与请求信任证所包含关键字匹配的信任证。(3)在匹配引擎中获得准确的信任证。

3.1　信任证查找

信任证查找是在 P2P 覆盖网中获取语义关键字并且获得扩展的信任证描述。当在某个 Skip 图节点上发起给定关键字的查询请求时，CI 系统首先将当前节点中的关键字和查询关键字进行比较，然后查询它的邻居节点，找到拥有与查询关键字最相近前缀的节点，然后将查询消息转发给这个邻居节点。通过此种方法不断查找，直到给定的查询关键字找到为止，或者查询到最下一层 Skip List 并且找不到所查询关键字而以失败告终。

成功完成信任证查询后，需要执行匹配过程。信任证文档中元素和属性都是以它们的名字、类型和结构作为其表示的形式和语义，信任证文档的相似度可以理解为这些形式和语义的相似程度。利用关键字最大前缀相似机制来比较信任证索引关键字的相似程度。

定义 1　键值惟一数值表示 $Uni(key)=\sum_{i=0}^{\|key\|-1} d^{Max_s-i-1}\times Num(key[i])$，其中，$\|key\|=length(key)$；$Num(key[i])$ 表示 key 第 i 个字符所表示的数字。规定 key 的字符串最大长度 Max_s，即有 $1\le\|key\|\le Max_s$；d 表示当前数值为 d 进制数值；$Num(0)\prec\cdots\prec Num(a)\prec\cdots\prec Num(z)$。简单地，令 $Num(0)=0$，$Num(z)=35$，且不考虑字母大小写，其他字符归 0；容易得到 $d=36$，键值唯一数值是一个 36 进制的数值。

定义 2　相似度函数 $Sim(key_1,key_2)=\dfrac{|Uni(key_1)-Uni(key_2)|}{Uni(Z)-Uni(O)}$，其中，$Z$ 表示长度为 Max_s 的字符串“$z\cdots z$”；O 表示长度为

Max_s 的字符串“ $o\cdots o$ ”，显然 $0 \le Sim(key_1, key_2) \le 1$ ，并且前缀越相近，2 个键值越相似。

利用定义 1 和定义 2，可以找到信任证查询请求的确切结果或者相似结果。给定一个查询，所有信任证都以查询得到的相似度进行排序，从而可以根据相似度的高低选择符合条件的信任证构造信任证图，筛选掉无关的信任证，减少信任证构造的大小，提高信任证发现效率。

3.2 信任证链发现

当查找到所有满足要求的信任证后，利用当前查询到的信任证集合构造信任证图，从而发现信任证链。由于 CI 系统中包含的索引关键字不再是签发者和主体二维信息，而是签发者、主体和角色集合作为索引关键字，使得查找所得到的信任证范围更小、更精确，构造的信任证图也构造较小。

本文提出的基于 Skip 图的信任证链发现算法(SkipGraph-based Credential Chain Discovery, SGCCD)主要思想是：根据用户所发出的对某个角色的请求，从用户的主体和角色 2 个方向，分别构造 2 个独立的信任证图。通过队列方式循环，不断扩展这 2 个信任证图，直到连通或找不到相关信任证，算法结束。如果 2 个图连通，说明存在 1 个信任证链，否则不存在信任证链。当算法扩展信任证图时，首先需要构造查找所需的信任证关键字，然后调用 CI 的搜索功能，得出符合要求的信任证集合，并且比较正反 2 个方向所构成的信任证集合大小，选择小的一方作为查找方向，从而构造出尽量小的信任证图。

4 实验结果与分析

基于 SGCCD 算法开发了 CI 系统，和其他信任证链发现系统不同，信任证是分布存储在对等节点中，实现负载均衡。本文模拟在 Internet 环境中 1 000 个用户作为签发者和主体产生的 100 000 个 RT_1 的信任证，构建一个 Skip 图仿真信任证劈分后的索引片段。首先给出平均查询跳数(查找过程中访问的节点数)和查找时间。然后根据信任证发现算法与现有几种发现方法进行比较。

4.1 信任证搜索跳数

将上述信任证劈分后所产生的关键字分布到 CI 系统中，得出如图 2 所示的平均查询跳数。关键字平均查找跳数为 $\log(N)$，其中，N 为系统中 P2P 节点的数目。

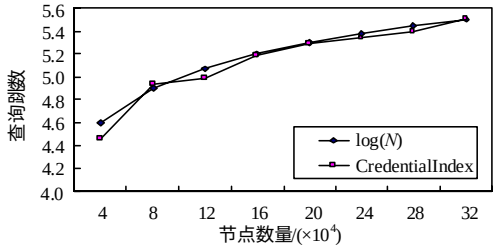

图 2 CredentialIndex 平均查询跳数

4.2 信任证搜索时间

如图 3 所示，给出了基于 Skip 图的信任证搜索算法试验结果。从图中可以看出，对 10 万个信任证规模的 P2P 覆盖网来说，随着信任证的不断加入，查询时间缓慢增加，其平均查询时间在 2.5 s~3.4 s 之间，能够被实际系统接受。因此，此种算法效率较高。

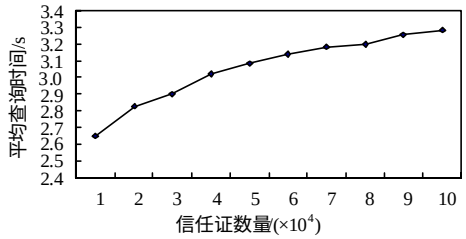

图 3 CredentialIndex 平均查询时间

4.3 信任证链发现

根据SGCCD算法，统计出了 3 种情况下信任证图的大小：信任证全部存放在签发者一端，采用逆向查找算法[3]；信任证全部存放在主体一端，采用正向查找算法；信任证利用本文提出的基于Skip图的P2P覆盖网存储，采用SGCCD算法。

实验结果如图 4 所示，从图中可以看出，对于随机生成的相同信任证集合和相同的信任证链发现，SGCCD算法生成的信任证图大小只有算法[3]生成的信任证图大小的 20%左右，这完全得益于支持基于更多关键字的前缀匹配查询方法，使得每次查找所得到的信任证集合更少。

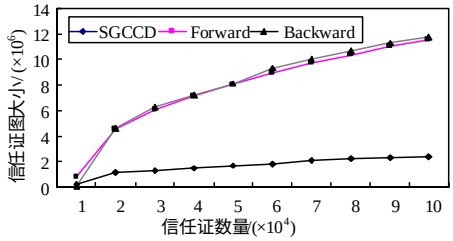

图 4 不同信任证链发现算法比较试验

5 结束语

本文开发了一个在 P2P 覆盖网中基于部分关键字前缀和相似度的信任证搜索系统，并在此基础上提出了基于 RT_1^T 格式信任证的 SGCCD 算法，该算法在发现过程中能够查找较少的信任证，从而极大地减少了网络中信任证的传输开销，优化信任证链发现的性能。下一步的工作包括如何动态劈分信任证文档和如何提高多层 Skip 图的关键字分发效率等。

参考文献

[1] Li Ninghui, Mitchell J C, Qiu Yu, et al. RTML: A Role-based Trust-management Markup Language[EB/OL]. (2002-08-31). http://crypto.stanford.edu/~ninghui/papers/ rtml.pdf.

[2] Seamons K E, Chan K E, Child T, et al. TrustBuilder: Negotiating Trust in Dynamic Coalitions[C]//Proc. of DARPA Information Survivability Conference and Exposition. Washington D. C., USA: [s. n.], 2003: 49-51.

[3] Li Ninghui, Winsborough W H, Mitchell J C. Distributed Credential Chain Discovery in Trust Management[C]//Proc. of the 8th ACM Conf. on Computer and Communications Security. New York, USA: ACM Press, 2001: 156-165.

[4] Aspnes J, Shah G. Skip Graphs[C]//Proc. of SODA'03. New York, USA: ACM Press, 2003.