Computer Science and Application 计算机科学与应用, 2012, 2, 47-50
http://dx.doi.org/10.12677/csa.2012.21009 Published Online March 2012 (http://www.hanspub.org/journal/csa)

Fault-Tolerant Method in P2P Information Management Systems*

Lu Cai[1], Jian Zhao[2]

[1]School of Information Management and Information Systems, Chang'an University, Xi'an
[2]School of Basic Education, National University of Defense Technology, Changsha
Email: Centaur_Laurel@hotmail.com, zhaojian@nudt.edu.cn

Received: Jan. 2nd, 2012; revised: Jan. 19th, 2012; accepted: Feb. 3rd, 2012

Abstract: FissionE is a Kautz graph based infrastructure of P2P information management systems. It has the optimal network diameter given node degree $d = 2$. In order to address the problem of degraded routing performance caused by node failures, in this paper we propose a fault-tolerant routing algorithm for the FissionE system. The basic idea is to bypass failed node or link with some certain mechanism, so that FissionE can achieve better routing performance.

Keywords: P2P Information Management System; Kautz Graphs; Fault-Tolerance

P2P 信息管理系统中的容错方法*

蔡　璐[1]，赵　舰[2]

[1]长安大学经济管理系，西安
[2]国防科技大学基础教育学院，长沙
Email: Centaur_Laurel@hotmail.com, zhaojian@nudt.edu.cn

收稿日期：2012 年 1 月 2 日；修回日期：2012 年 1 月 19 日；录用日期：2012 年 2 月 3 日

摘　要：FissionE 是一种基于 Kautz 图的 P2P 信息管理系统网络架构，在给定节点度数($d = 2$)下具有最优的网络直径。针对结点失效导致的 FissionE 路由性能较差的问题，本文对 FissionE 的容错路由算法进行研究，其基本思想是：如果下一跳结点失效或网络连接失效，那么将采用某种方法“绕过”失效的结点或连接，从而获得较好的路由性能。

关键词：P2P 信息管理系统；Kautz 图；容错

1. 引言

近年来，P2P 技术作为一种新型的网络计算技术蓬勃兴起，已经被广泛地应用于信息管理系统领域，具有广阔的发展前景[1-3]。P2P 信息管理系统中各结点的地位平等，每个结点既是客户机又是服务器，不依赖于任何中央服务器，可以充分利用各个结点上的资源，极大地丰富了网络中的资源，各个结点可以直接进行交互和协作，因而有效地提高了资源的利用率[4-6]。

然而，现有的 P2P 信息管理系统难以应用于结点状态不稳定的环境[7]。这是由于一方面网络拓扑变化频繁，另一方面 P2P 信息管理系统中没有中央控制，各结点的状态信息无法进行汇聚，进而导致路由性能较差。

FissionE 是一种基于 Kautz 图的信息管理系统网络架构。FissionE 创造性地提出了许多新型的规则和算法，例如 Kautz_hash 命名算法、邻居关系不变量拓扑构造规则和局部优化动态维护机制，并且详细地分析和证明这些机制的正确性。在给定节点度数($d = 2$)

*资助信息：本文研究成果受到国家自然科学基金(批准号：60903205)和博士点基金(批准号：20094307110008)的资助。

下 FissionE 具有最优的网络直径。

针对节点失效导致的 P2P 信息管理系统性能较差的问题，本文对 FissionE 的容错方法进行研究，其基本思想是：如果下一跳结点失效或网络连接失效，那么将采用某种方法“绕过”失效的结点或连接，从而获得较好的性能。模拟实验表明，本文提出的方法能够在存在节点失效的情况下实现高效的 FissionE 信息管理系统网络架构。

2. 相关研究

本节将对 FissionE 的基本设计(包括 Kautz 图拓扑结构、FissionE 路由算法等)进行回顾。

Kautz 串[2]是指在由 k 个字符组成的字符串中，如果每个字符只能取从 0 到 d 的整数且相邻的两个字符不同，那么就称这个字符串是基底为 d、长度为 k 的 Kautz 串。Kautz 空间 KautzSpace(d, k)是指所有基底为 d、长度为 k 的 Kautz 串的集合。Kautz 图[2]K(d, k)是每个结点的出度和入度都是 d 且网络直径是 k 的有向图，Kautz 图 K(d, k)中每个结点的标识都是 Kautz 空间 KautzSpace(d, k)中的一个 Kautz 串，Kautz 图 K(d, k)中任意一个结点 $U\left(U=u_1u_2,\cdots,u_k\right)$ 都有 d 条到结点 $V\left(V=u_2u_3,\cdots,u_ka\left(a\neq u_k\right)\right)$ 的出边(记作 $U\to V$)。图 1 给出了 Kautz 图 K(2, 3)的示例。

根据 Kautz 串的定义，Kautz 图 K(d, k)中共有 d^k+d^{k-1} 个结点，在目前所有的拓扑图中最接近 Moore 界 $\left(1+d+d^2+\cdots+d^k\right)$，并且在 $k=2$ 时，Kautz 图中共有 $d+d^2$ 个结点，只比 Moore 界 $\left(1+d+d^2\right)$ 少一个，可见 Kautz 图是结点数目最多的拓扑图。Kautz 图 K(d, k)的网络直径达到了由 Moore 界推出的网络直径下界 $\left[\log_d\left(N\times\left(d-1\right)+1\right)\right]-1$，可见 Kautz 图具有最优网络直径。Kautz 图的连接度为 d，即在 Kautz 图的任意两个结点之间存在 d 条互不相交的路径，可见 Kautz 图的容错性非常好。

在静态 Kautz 图中采用长路径路由算法，如果 $u_k\neq v_1$，那么从结点 U 到结点 V 的路由路径为：$U=u_1u_2\cdots u_k\to u_2u_3\cdots u_kv_1\to u_3u_4\cdots u_kv_1v_2\to\cdots\to u_kv_1v_2\cdots v_{k-1}\to v_1v_2\cdots v_k=V$；如果 $u_k=v_1$，那么从结点 U 到结点 V 的路由路径为：$U=u_1u_2\cdots u_k\to u_2u_3\cdots u_kv_2\to u_3u_4\cdots u_kv_2v_3\to\cdots\to u_kv_2\cdots v_{k-1}v_k=V$。

由此可见，Kautz 图 K(d, k)中任意两个结点之间的路由路径长度为 k 或 $k-1$，平均路由路径长度为 $d/\left(d+1\right)\times k+1/\left(d+1\right)\left(k-1\right)=k-1/\left(d+1\right)$。例如在 Kautz 图 K(2, 3)中从结点 101 到结点 212 的路由路径为 101→012→121→212，如图 1 所示。

FissionE 中每个结点的标识都是一个基底为 2 的 Kautz 串。最初 FissionE 与静态 Kautz 图相同，但是随着结点的动态加入和退出，结点标识的长度也在动态变化，FissionE 针对此提出了邻居关系不变量拓扑构造规则，使 FissionE 中邻居结点的标识长度相差始终不超过 1，维护了良好的拓扑结构。FissionE 中每个结点的路由表中都记录了两种邻居的信息，即出边邻居和入边邻居，结点 $U=u_1u_2,\cdots,u_k$ 的出边邻居是结点 $V=u_2u_3\cdots u_kq_1\cdots q_m$（$u_2u_3\cdots u_kq_1\cdots q_m$ 是基底为 2 的 Kautz 串且 $0\le m\le 2$）；结点 $U=u_1u_2,\cdots,u_k$ 的入边邻居是结点 $W=au_1u_2\cdots u_i$（$au_1u_2\cdots u_i$ 是基底为 2 的 Kautz 串且 $k-2\le i\le k$）。图 2 给出了一个 FissionE 拓扑示例。

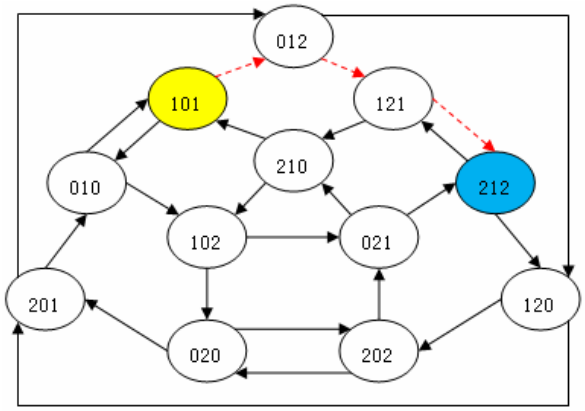

Figure 1. Routing path from source node 101 to destination node 212 in K(2, 3)
图 1. K(2, 3)中从源结点 101 到目的结点 212 的路由路径

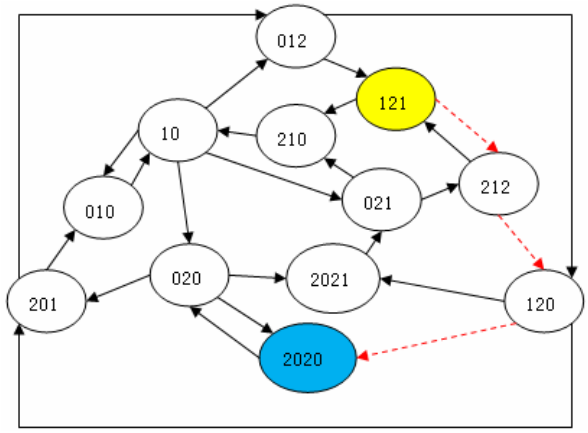

Figure 2. Routing path from source node 121 to destination node 2020
图 2. 从源结点 121 到目的结点 2020 的路由路径

FissionE 中消息路由的过程和 Kautz 图的长路径路由算法相似，设 X 为当前结点的标识，T 为目标 Kautz 串，M 为 X 和 T 的前后匹配位数，$|X|$ 为 X 的长度，消息路由算法为重复进行 $L=|X|-M(X,T)$ 和 $L'=L-1$ 直至 $L'=0$。例如在上述拓扑示例中从结点 121 到结点 2020 的消息路由路径为 121→212→120→2020，如图 2 所示。

3. FissionE 的容错路由算法设计

3.1. Kautz 图容错路由

为降低失效结点对消息路由的影响，本文提出在 Kautz 图中采用如下兄弟结点容错机制。

结点 $X=x_1x_2,\cdots,x_k$ 的两个出边邻居 $B=x_2x_3\cdots x_k y\ (y=0,1,2$ 且 $b\neq x_k)$互为兄弟结点，当结点 X 处理路由消息 Routing(T, L, P)时，如果路由消息要转发到的出边邻居 $W=x_2x_3\cdots x_k a$ 已经失效，那么将路由消息转发给 W 的兄弟结点 $B=x_2x_3\cdots x_k b$，之后按照长路径路由算法，路由至目的结点。例如在 Kautz 图 K(2, 3)中，理想情况下从源结点 102 到目的结点 120 的路由路径为 102→021→212→120。当结点 212 失效后，路由路径变为 102→021→210→101→012→120，如图 3 所示。

3.2. FissionE 容错路由

在现有 FissionE 设计中，结点被动退出后路由消息很有可能还会被转发给已经失效的结点，为增加消息路由的鲁棒性，一种选择是仿照 Kautz 图容错路由算法进行路由。但是，由于在 FissionE 中结点的出边邻居个数可能为 1，因此我们提出如下改进的 FissionE 容错方法。

标识为$W=u_2\cdots u_k x_1\cdots x_j\ (0\le j\le 2)$ 和$V=\tilde{u}_2u_3\cdots u_k q_1\cdots q_m$ ($\tilde{u}_2$ 与 u_2 相对于 u_1 互补，$0\le m\le 2$)的结点互为容错结点。当路由消息路由至$U=u_1u_2,\cdots,u_k$ 时，假设下一跳结点 $W=u_2\cdots u_k x_1\cdots x_j$ ($0\le j\le 2$) 已经失效，那么 U 将把消息转发至 W 的容错结点(即 $V=\tilde{u}_2u_3\cdots u_k q_1\cdots q_m$)，然后从 V 再路由至目的结点。

例如在图 2 所示的 FissionE 拓扑中，212 和 012 互为容错结点，理想情况下从源结点 121 到目标 Kautz 串 120 的路由路径为 121→212→120，如图 4 所示。当结点 212 失效后，路由路径变为 121→210→10→012→120，如图 5 所示。图 6 给出了 FissionE 容错机制。

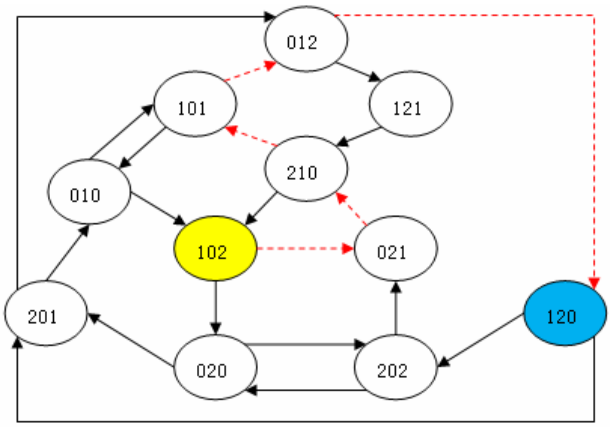

Figure 3. Routing path after node 212 fails in K(2, 3)
图 3. K(2, 3)中结点 212 失效后的路由路径

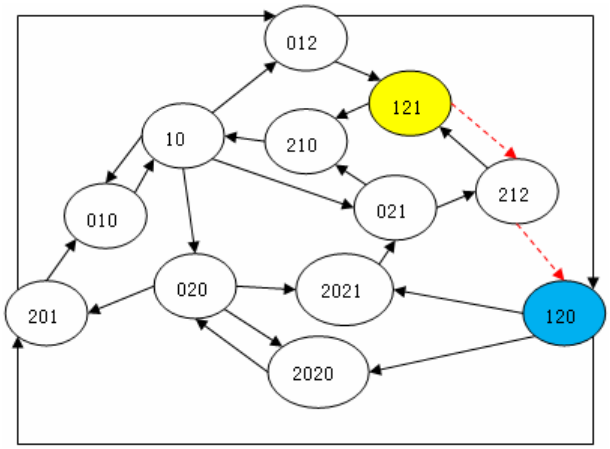

Figure 4. Ideal routing path from source node 121 to destination node 120
图 4. 从源结点 121 到目的结点 120 的理想路由路径

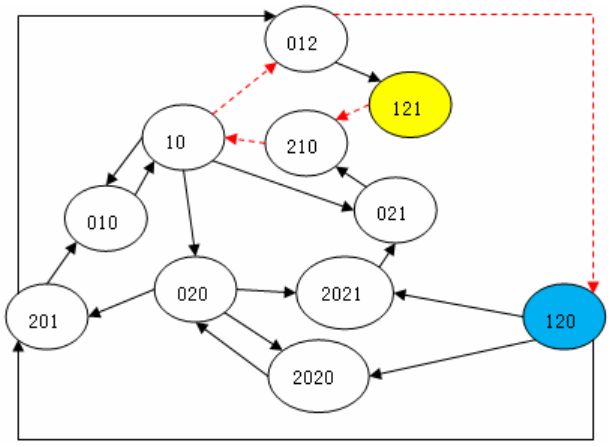

Figure 5. Routing path after node 212 fails
图 5. 结点 212 失效后的路由路径

```
Program FaultTolerantRouting (Target T, Current C)
//T = t1t2,⋯,tk 为目的结点，C 为当前结点
next_hop = FissionE_Routing (T, C)
if next_hop 有效
   then 正常路由至下 跳结点
else
      B = next_hop 的容错结点
      next_hop = FissionE_Routing (B, C)
return
```

Figure 6. Fault tolerant rouging in FissionE
图 6. FissionE 的容错路由机制

4. 模拟评估

我们通过修改 FissionE 模拟器对容错路由的性能(平均路径长度)进行评估。在模拟过程中，首先通过 Kautz_hash 算法随机产生一个目标 Kautz 串，并随机选择一个结点作为消息发起结点。消息采用标准 FissionE 路由算法进行路由。在每个消息的路由过程中，随机选择一个结点作为失效结点，当消息遇到失效结点时将启用如图 6 所示的容错路由算法。结点数从 256 变化到 32 K。模拟结果如图 7 所示。

图中也包括了没有失效结点时标准 FissionE 的平均路由路径长度。从图中可以看出，发生结点失效时的容错路由延迟仅比正常情况下的 FissionE 路由延迟高 50%左右，从而说明我们的容错路由算法是有效的。

5. 结论

本文对结构化 P2P 信息管理系统 FissionE 的容错方法进行研究。模拟实验表明，本文提出的方法能够

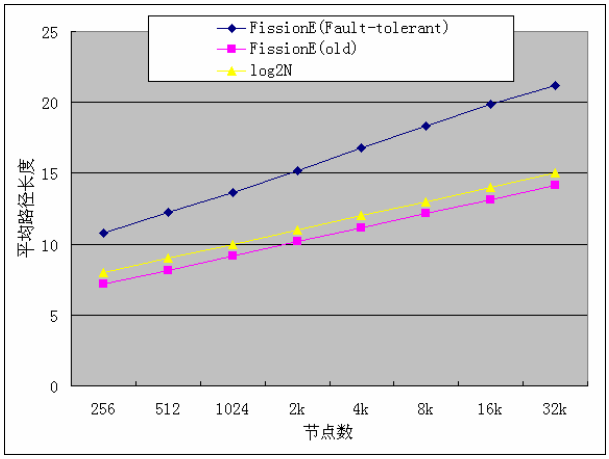

Figure 7. Fault tolerant performance
图 7. 容错性能

有效地提高 FissionE 的容错性能。

参考文献 (References)

[1] 李东升. 基于对等模式的资源定位技术研究[D]. 国防科技大学, 2005.
[2] 彭兰. P2P 技术与网络传播的未来[URL], 2004. http://news.xinhuanet.com/newmedia/2004-12/01/content_2282285.htm
[3] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, et al. Chord: A scalable peer-to-peer lookup protocol for internet applications. IEEE/ACM Transactions on Networking, 2003, 11(1): 17-32.
[4] M. Bawa, F. B. Cooper, A. Crespo, N. Daswani, et al. Peer-to-Peer research at Stanford. SIGMOD Record, 2003, 32(3): 23-28.
[5] Napster Website, 2001. http://www.napster.com
[6] S. Rhea, D. Geels, T. Roscoe and J. Kubiatowicz. Handling churn in a DHT. USENIX Annual Technical Conference, General Track, 2004.
[7] 刘敏. 结构化 P2P 系统容错机制研究[D]. 国防科技大学, 2008.