(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2009 年6 月11 日 (11.06.2009)

PCT

(10) 国际公布号
WO 2009/071008 A1

(51) 国际专利分类号:
H04L 12/24 (2006.01)

(21) 国际申请号: PCT/CN2008/072836

(22) 国际申请日: 2008 年10 月27 日 (27.10.2008)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
200710188100.1
2007 年11 月22 日 (22.11.2007) CN

(71) 申请人 (对除美国外的所有指定国): **华为技术有限公司(HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为基地总部办公楼, Guangdong 518129 (CN)。

(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): **施广宇(SHI, Guangyu)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为基地总部办公楼, Guangdong 518129 (CN)。 **陈坚(CHEN, Jian)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为基地总部办公楼, Guangdong 518129 (CN)。 **龚皓(GONG, Hao)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为基地总部办公楼, Guangdong 518129 (CN)。

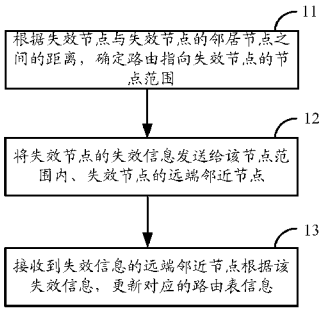
(74) 代理人: **北京中博世达专利商标代理有限公司(BEIJING ZBSD PATENT & TRADEMARK AGENT LTD.)**; 中国北京市海淀区大柳树路17号富海大厦B座501室, Beijing 100081 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

[见续页]

(54) Title: METHOD, EQUIPMENT AND SYSTEM FOR UPDATING ROUTING TABLE AFTER NODE FAILURE IN PEER-TO-PEER NETWORK

(54) 发明名称: P2P对等网络中节点失效后的路由更新方法、设备及系统



11 DETERMINING THE NODE RANGE WHICH ROUTING TO THE FAILED NODE ACCORDING TO THE DISTANCE BETWEEN THE FAILED NODE AND ITS NEIGHBOR
12 SENDINGTHE FAILURE INFORMATION OF THE FAILED NODE TO LDN OF THE FAILED NODE IN THE NODE RANGE
13 THE LDN WHICH RECEIVES THE FAILURE INFORMATION UPDATING THE INFORMATION OF ROUTING TABLE ACCORDING TO THE FAILURE INFORMATION

图 1 / Fig. 1

(57) Abstract: A method for updating routing table after node failure in peer-to-peer(P2P) network is disclosed. The method includes: determining the node range which routing to the failed node according to the distance between the failed node and its neighbor, sending the failure information of the failed node to Long Distance Neighbor(LDN) of the failed node in the node range, and the LDN updating the information of routing table according to the failure information. A network equipment and a P2P communication network are also disclosed. Then the perceptive faculty of the node churn in the whole P2P network is enhanced, and the routing lookup efficiency of the whole P2P network and the stability of the system are improved.

[见续页]

WO 2009/071008 A1

(84) 指定国（除另有指明，要求每一种可提供的地区保护）：ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布：
— 包括国际检索报告。

(57) 摘要：

本发明实施例公开了一种P2P对等网络中节点失效后路由表信息的更新方法，该方法包括：根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围；将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点；所述远端邻近节点根据所述失效信息更新对应的路由表信息。本发明同时公开一种网络设备和P2P对等通信网络。采用本发明实施例可以增强整个P2P对等网络对节点扰动行为的感知度，提高整个P2P对等网络的路由查找效率和系统的稳定性。

P2P 对等网络中节点失效后的路由更新方法、设备及系统

技术领域

本发明涉及通信技术领域，尤其涉及 P2P 对等网络中节点失效后的路由更新方法、设备及系统。

背景技术

P2P（Peer-to-Peer，表示对等体之间的一种对等关系）系统与传统的客户机/服务器模式系统不同，在对等体之间进行操作，每个对等体，即 P2P 系统中每个节点（Peer），既可以为其他节点提供服务，又可以接受其他节点提供的服务。P2P 系统可以按照其拓扑结构进行分类，一般分为：集中化拓扑（Centralized Topology）、全分布式非结构化拓扑（Decentralized Unstructured Topology）、全分布式结构化拓扑（Decentralized Structured Topology，也称作 DHT 网络）和混合型拓扑。

当前P2P系统的主要拓扑结构为结构化拓扑DHT(Distributed Hash Table，分布式哈希表）网络，P2P 系统中的很大部分应用都是基于 DHT 网络所构成的。在这些系统中，节点通过它的一些唯一属性，如 IP 地址，哈希得到唯一标识 NodeId，标识对应的数据项以键值对<key，value>的方式表示，其中键key是对于数据项的索引，而值value可以是数据项的定位地址如IP或者URL。通过哈希赋予数据索引键以唯一标识，并将此键对应的键值对存储到与此键标识最邻近的节点。查询时，可以通过查询键值对哈希得到唯一标识，并通过此唯一标识找到与之最邻近的节点（此节点存储了数据项所在的地址）。

另一方面，P2P 系统同时也是一种自组织形态的网络，在该网络中，节点可以随意加入或退出，这种随意性会造成资源定位的不准确和网络的扰动，网络扰动的程度大小直接影响路由发现方法的效率。网络扰动（Churn、fluctuation of network）包括节点的加入、退出、失败、迁移、并发加入过程、网络分割等。P2P 网络的 DHT 路由查找方法如何处理不同的网络扰动 churn，

将直接影响整个 P2P 网络的路由效率和负载开销。

现有技术中给出如下两种方式解决 P2P 网络扰动问题:

现有技术一

在 DHT 网络中解决 churn 问题时，考虑三个重要因素：快速的替代节点、超时检测和最近邻居节点选择。该技术对于网络扰动的直接对抗方法是：通过 P2P NodeId 空间中的顺序 K 个节点互相 PING，来保持节点之间不断连。

在 P2P 网络的运行过程中，每个节点每隔一段时间向它周边顺序 K 个节点发送 PING 维护信息，这顺序 K 个节点接收到源节点所发过来的 PING 消息后，立刻反馈一个信息告知源节点自己正常存活，源节点的路由表不用做任何修改。而当网络中某节点失效时，其周边的邻居节点能够通过这种主动探测 PING 的方式来发现这种节点扰动行为，找出失效节点，并广播该失效信息给顺序 K 个节点更新其路由表链接。

发明人在实现本发明的过程中，发现现有技术一存在如下不足:

现有技术一中，通过 P2P NodeId 空间中的顺序 K 个节点互相 PING 来保持节点之间不断连，即维护信息只在 K 个邻居节点间传递。当网络中存在节点失效行为时，由于路由表信息只能保持 K 段连续，因此在网络中查找一个节点时，很可能会出现不同程度的“爬行”现象，即由于原本指向失效节点的 peer 节点路由表信息均失效，远方的节点只能通过再次递归查找的方式才能找到失效节点的替代节点。两个节点之间的距离相隔越远，这个问题就越严重，失效后节点间递归查找的次数也就越多。当网络中 churn 问题很剧烈时，假如一条 P2P 路径中的多个关键路由节点都失效了，那么路由表信息在网络中的“爬行”现象将会十分明显，节点在路由时必须通过逐 K 跳逐 K 跳来查找下一步路由。

并且，在 P2P 网络中，churn 现象和“爬行”行为还存在累计效应，当网络中失效节点数目不断递增时，由于失效信息得不到有效广播，大部分节点仍维持原有的失效路由表，而导致“爬行”的次数将会不断增多，“爬行”的

时间也越来越长，最恶劣的情况下查找一个节点时间复杂度会达到 O(N)次（N 为 P2P Overlay 中节点数目），严重影响了 P2P 网络的路由查找效率。

现有技术二

通过整个网络中节点的历史生命周期信息，来预测某节点当前的存活概率。具体的，采集整个 P2P 网络中节点历史生命周期分布概率，推算出某节点在下一时间段内的存活概率，以此对整个网络的节点行为产生一种预判信息来决定网络下一步的行为。当预测出某下一时段内该节点较为稳定时，其邻居节点会减少对该节点的 PING 维护操作，从而降低了维护带宽的开销。

发明人在实现本发明的过程中，发现现有技术二存在如下不足:

P2P 网络越庞大越复杂，节点的生命周期就越难预测准确。在预测的过程中需要设定很多系统参数，参数的数值选定存在很大的特殊性，不同的参数设定可能会给预测结果带来严重的不准确性。

另外，现有技术二虽然降低了系统维护开销，但是对于 P2P 网络的路由查找方法的效率并没有任何改进。当网络中出现严重的扰动行为时，节点的查找复杂度仍可能会达到 O（N）次，“爬行”现象在路由查找过程中依然存在，路由查找效率较低。

发明内容

本发明实施例提供一种 P2P 对等网络中节点失效后路由表信息的更新方法、设备及系统，用以增强整个网络对节点扰动行为的感知度，提高整个网络的路由查找效率和系统的稳定性。

本发明实施例采取以下技术方案:

本发明实施例提供一种 P2P 对等网络中节点失效后路由表信息的更新方法，该方法包括:

根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围;

将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远

端邻近节点;

所述远端邻近节点根据所述失效信息更新对应的路由表信息。

本发明实施例还提供一种网络设备，包括:

第一确定模块，用于根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围;

第一发送模块，用于将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点。

本发明实施例还提供一种网络设备，所述网络设备为失效节点的远端邻近节点，包括:

接收模块，用于接收失效节点的失效信息;

处理模块，用于根据所述失效信息更新对应的路由表信息。

本发明实施例还提供一种通信网络，包括失效节点，还包括失效节点的邻居节点、远端邻近节点，其中:

失效节点的邻居节点，用于根据本节点与所述失效节点之间的距离，确定路由指向所述失效节点的节点范围；将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点;

远端邻近节点，用于接收所述失效信息，根据所述失效信息更新对应的路由表信息。

本发明实施例中，根据失效节点与邻居节点之间的距离，确定路由指向所述失效节点的节点范围；将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点；所述远端邻近节点根据所述失效信息更新对应的路由表信息，与现有技术中节点失效时仅将失效信息通知K个邻居节点的技术方案相比，增强了整个P2P对等网络对节点扰动行为的感知度，提高了整个P2P对等网络的路由查找效率和系统的稳定性。

附图说明

图 1 为本发明实施例中 P2P 对等网络中节点失效后更新路由表信息的处

理流程图;

图 2 为本发明实施例中 P2P 对等网络中节点失效后更新路由表信息的一个具体实例的处理流程图;

图 3 为本发明实施例中 P2P 对等网络中节点失效后更新路由表信息的一个具体实例在网络结构中的处理示意图;

图 4 为本发明实施例中 Kademlia DHT 网络中节点失效的示意图;

图 5 为本发明实施例中 Kademlia DHT 网络中节点失效时的处理示意图;

图 6 为本发明实施例中 Chord 网络的结构示意图;

图 7A、图 7B、图 7C、图 8A、图 8B 为本发明实施例中网络设备的结构示意图;

图 9 为本发明实施例中通信网络的结构示意图。

具体实施方式

下面结合说明书附图对本发明实施例进行详细说明。

如图 1 所示，本发明实施例中，P2P 对等网络中节点失效后，更新路由表信息的处理流程如下:

步骤 11、根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围。

步骤 12、将所述失效节点的失效信息发送给该节点范围内失效节点的远端邻近节点。远端邻近节点（Long Distance Neighbor，LDN）可以有 i 个，若节点 B 的 NodeId 和节点 A 的 NodeId 仅第 i 位不同，其他位都相同或相隔距离最小，则节点 B 称为节点 A 的 LDN[i]（远端邻近节点[i]）。

步骤 13、接收到所述失效信息的远端邻近节点根据该失效信息，更新对应的路由表信息，将指向失效节点的路由表信息指向失效节点的替代节点。

图 1 所示流程中，在确定路由指向失效节点的节点范围之前，可以按设定周期对网络中邻居节点的状态进行探测，确定失效节点，按设定周期的不同，失效节点可以是当前网络中突然失效的节点，也可以是一段时期前网络

中失效的节点。不同网络中，根据失效节点与失效节点的邻居节点之间的距离确定的路由指向失效节点的节点范围不同，一个实施例中，可以根据失效节点与失效节点的邻居节点之间的距离、网络的路由特性，确定路由指向失效节点的节点范围。

在步骤 12 中，可以根据网络存储的键值特性，在失效节点的邻居节点中确定发送失效信息的替代节点，由替代节点将失效节点的失效信息发送给路由指向失效节点的节点范围内、失效节点的远端邻近节点。

失效信息包括替代节点的地址；另外，失效信息还可以包括失效节点的标识、或失效节点的地址、或失效节点的键值、或失效节点的标识、地址、键值的两两组合或三者的组合。一个实施例中，失效信息还可以包括失效节点对其远端邻近节点的影响范围；远端邻近节点在接收到失效信息后，可以将失效信息转发给该影响范围内的节点。

远端邻近节点确定该影响范围未超过本节点的邻居节点维护范围时，可以直接将失效信息转发给该影响范围内的节点；或，确定该影响范围超过本节点的邻居节点维护范围时，较佳的，将失效信息转发给离该影响范围最近的本节点的邻居节点，由该邻居节点继续将失效信息转发给该影响范围内其它节点，当然，此时远端邻近节点也可以将失效信息转发给另外的节点，由接收到转发的失效信息的节点继续将失效信息转发给该影响范围内其它节点。该影响范围内的节点在接收到失效信息后，更新对应的路由表信息，即将指向失效节点的路由表信息指向失效节点的替代节点。

图 2 为本发明实施例中 P2P 对等网络中节点失效后更新路由表信息的一个具体实例，该流程在 P2P 对等网络中的实施可参见图 3 所示的路由表信息更新示意图。其处理流程如下：

步骤 21、节点 peer 主动探测邻居节点存活状态。例如，节点 B 每隔周期时间对周边邻居节点进行一次主动探测，判断邻居节点 F 的存活状态。

步骤 22、当某一节点（节点 F）失效时，邻居节点（节点 B）通过步骤

21 发现节点 F 失效后，计算失效信息通知范围。邻居节点（节点 B）根据自身 NodeId 和失效节点（节点 F）NodeId 之间的相隔距离，可以基于不同的 DHT 网络路由方法特性，推算出整个网络中哪些路由表指向失效节点的节点范围区间。

步骤 23、邻居节点（节点 B）确定发送失效信息的替代节点（节点 A）。该步骤与步骤 22 的执行顺序并无要求，可以同时进行，也可以先执行步骤 22，再执行步骤 23，或者可以先执行步骤 23 再执行步骤 22。

邻居节点在计算出失效节点所影响的范围区间后，可以根据不同 DHT 网络存储 Key 值的特性，选出最终由哪个邻居节点主动发送节点失效信息并承担替代节点的责任。若以节点 NodeId 远近来选择存储 Key 值的 Kademlia 网络或 Pastry 网络，那么作为替代节点的则是离失效节点异或距离最小的邻居节点。若以最近的后续节点来存储 Key 值的 Chord 网络或 Koorde 网络，作为替代节点的则是失效节点的后续节点。

步骤 24、替代节点（节点 A）查找失效影响范围区间内的远端邻近节点 LDN。即，作为替代节点的邻居节点使用路由查找方法找出各个区间内失效节点所对应的 LDN[i]节点，即各个区间内和失效节点相隔距离最近的节点。

步骤 25、替代节点（节点 A）依次向失效节点的远端邻近节点 LDN 发送失效信息。在找出失效节点的各个影响区间的 LDN[i]后，由替代节点依次向 LDN 发送节点的失效信息，失效信息包括包括替代节点的地址；另外，失效信息还可以包括失效节点的标识、或失效节点的地址、或失效节点的键值、或失效节点的标识、地址、键值的两两组合或三者的组合。

邻居节点还可以向各个 LDN 通知失效节点对其影响的范围，例如在失效信息中包括失效节点对 LDN 的影响范围，通知这些 LDN 更新其对应的路由表信息并依据失效信息影响的范围在 LDN 的邻居内转发这个失效信息。

步骤 26、远端邻近节点 LDN 接收到失效信息后，更新对应的路由表信息，将指向失效节点的路由表信息指向替代节点（节点 A）。进一步地，LDN 节点

根据接收到的失效节点对其影响的范围后，根据影响范围向自己的邻居节点转发失效信息。

各个 LDN 节点在收到失效节点信息对其影响的范围后，可以根据范围计算出自己周边有多少邻居节点需要知道这一节点失效信息，并向这些节点转发这一信息。如果需要通知的范围超出了自己邻居节点的范围，LDN 则通知离该范围最近的邻居继续向范围内其他节点通知（例如：组播）这一信息，通过多跳的形式来达到全范围内的节点通知，当然，此时远端邻近节点也可以将失效信息转发给另外的节点，由接收到转发的失效信息的节点继续将失效信息转发给该影响范围内其它节点。

影响范围内的节点更新对应的路由表信息。在影响范围内的节点收到节点失效信息后，将路由表中原来指向失效节点的指针指向失效节点的替代节点，完成路由表信息的更新操作。

下面分别以不同的 P2P 对等网络为例，说明本发明实施例方法。

一、以 Kademlia DHT 网络为例

以图 4 所示 Kademlia 网络中的节点 10110 失效为例，根据本发明实施例方法，当网络中节点 10110 失效时，它的邻居节点 1010 和 10111 通过主动探测得知节点 10110 失效后，可以根据自身 NodeId 和失效节点 NodeId 的相隔距离推算出失效信息所影响的范围区间。根据 Kademlia 网络的特性，当节点 F 失效时，那些所有路由表指针都指向过 F 的节点包括在以下范围内：

$$\begin{cases} 2^{m-i} < XOR(N,F) \le 2^{m-i+1} & (1) \\ XOR(N,F) < XOR(N,X) & (2) \\ XOR(N,F) < XOR(N,Y) & (3) \end{cases}$$

其中，节点 F 失效产生扰动，节点 X 和 Y 为节点 F 的前邻居节点和后邻居节点，节点 N 为需要通知的节点。m 为 NodeId 的位数，Kademlia 网络中 m 为 160，i 的范围从 1 到 m。

如上述公式所述，影响节点 N 的范围必须满足：(1) 节点 N 和失效节点 F

的异或距离大于 2^{m-i} 且小于等于 2^{m-i+1}，(2) 节点 N 和 F 的异或距离小于节点 N 和 X 的异或距离，(3) 节点 N 和 F 的异或距离小于节点 N 和 Y 的异或距离。当节点 N 都满足上述三式时，N 即为方法所需要通知的节点。

节点 1010 和 10111 可以根据自己 NodeId 和失效节点 10110 的 NodeId 之间的差值，确认失效节点的替代节点，由于节点 10111 与失效节点 10110 之间 NodeId 的差值小于节点 1010 与失效节点 10110 之间 NodeId 的差值，因此确认替代节点为邻居节点 10111。邻居节点 10111 在根据上述公式确定的各影响范围区间内各找出一个 LDN 节点并把这一失效信息通知到各个 LDN 节点。如图 5 所示，本实例中，根据 Kademlia 路由表的对称性，节点 10110 的 LDN[1] 就是 00110，这是由于 00110 除了首位和 10110 不同外，其他位都相同。而根据 Kademlia K 桶路由表的构成方式，节点 00110 和 10110 均会在对方的 K 桶路由表中。另外，根据上述公式可以得知，节点 10110 的 LDN 还包括 LDN[1]=00110，LDN[2]=111，LDN[3]=10010，LDN[4]=1010，LDN[5]=10111。

在计算出需要通知的节点范围且找到各自范围内的 LDN 节点后，LDN 节点会分别查找各自范围内的节点 N，并将该扰动信息传递给这些节点 N，使其更新自身的路由表信息。

通过这样主动通知，当远端节点需要查找失效节点时，能够直接知道该节点的信息状态和邻居替代节点，避免了传统 P2P 网络中远方的节点通过缓慢“爬行”的方式来查找失效节点。

二、以 Pastry DHT 为例

Pastry 网络的路由表结构和 Kademlia 方法比较类似，二者的区别在于，Pastry 网络在 Kademlia 网络的基础上，将二进制 NodeId 变成多进制 NodeId。实施中，可以先将节点 NodeId 由多进制转化为二进制，再依照 Kademlia 的方法来在 Pastry 网络上实现。具体处理过程如下：

节点 F 扰动时，它的最近邻居节点 X 检测到这一扰动后，可以结合失效节点的另一个邻居将这一扰动信息根据下式的范围找出首位和节点 F 不同，

其他位和节点 F 异或差值最小的节点 LDN[1]，通知它们节点 F 已经失效，以后指向节点 F 的路由都以指向节点 X 代替，并通知它们在范围内转发该信息。依次类推，邻居节点 X 分别通知其他那些 i 位与节点 F 不同，其他位差值最小的 LDN[i]。告诉它们节点 F 的失效信息，并让它们代理通知各自范围内的其他节点更新其路由表。

$$\begin{cases} 2^{m-i} < XOR(N,F) \le 2^{m-i+1} & (1) \\ XOR(N,F) < XOR(N,X) & (2) \\ XOR(N,F) < XOR(N,Y) & (3) \end{cases}$$

其中，k 为 k 进制基数，i 的范围从 1 到 m。X、Y 为失效节点 F 的前后邻居，N 为需要通知的节点。

如上述公式所述，影响节点 N 的范围必须满足: (1) 节点 N 和失效节点 F 的异或距离大于 2^{m-i} 且小于等于 2^{m-i+1}，(2) 节点 N 和 F 的异或距离小于节点 N 和 X 的异或距离，(3) 节点 N 和 F 的异或距离小于节点 N 和 Y 的异或距离。当节点 N 都满足上述三式时，N 即为方法所需要通知的节点。

Pastry 和 Kademlia 在实施中不同之处在于，Kademlia 采用的是二进制 NodeId，所以 Kademlia 在计算两个节点间距离的是采用的异或 XOR 的方式。而 Pastry 由于采用的是多进制 NodeId 的方式，所以在实现上需要先将 Pastry 网络里的多进制 NodeId 转化为二进制 NodeId 再执行发明步骤。

三、以 Chord DHT 为例

如图 6 所示，Chord 网络中，将 n 个节点哈希到具有 logn 位的标识环上。每个节点 x 指向它的直接后续 successor 节点（沿环顺时针方向的最近节点）。它将维护一个有 m=logn 个表项的指针表（finger table）。第 i 表项存储 $x+2^{i-1}$ 的后续标识。

参见图 5 和图 6，当节点 F 出现扰动时，根据本发明实施例方法，其邻居节点可以通过主动探测发现节点 F 的扰动行为，并根据自身 NodeId 和失效节点 F NodeId 的距离推算出整个 Chord 环中那些路由表指针都曾指向过 F 的节

点 N 范围:

$$predecessor(F) < (N + 2^i) \bmod 2^m \le F \quad 0 \le i < m$$

其中，predecessor(F)为节点 F 的前邻居节点。如上述公式所示，节点 N 的 NodeId 加上 2^i 且对 2^m 取模后将大于失效节点 F 的前邻居节点 NodeId 且小于等于 F 的 NodeId，在上述不等式范围内的节点 N 即为需要通知的节点。

在获得扰动影响范围后，根据 Chord 网络特性，确认失效节点的替代节点为其后续节点。替代节点可以根据范围找出各个范围内的 LDN 节点。如节点 N8 的 LDN[m-1]应当是路由表的最后一项指向 N8 的那个节点，即(NodeId+2^{m-1})所对应 successor 后续节点为 N8 的那个节点 NodeId。而它的 LDN[i]则为(NodeId+2^i)所对应 successor 指向 N8 的那些节点，其中 i 的范围从 0 到 m-1。由 F 的替代节点通知 Chord 环中各个影响区间内的 LDN，并告之节点失效信息和各自扰动范围。这些 LDN 在各自的范围内组播此信息，通知那些路由表中曾指向过 F 的节点，告知它们节点 F 已经失效，替代节点为 F 后续邻居节点。范围区间内的节点根据此信息更新自己的路由表。

四、以 Koorde DHT 为例

Koorde 是在 Chord 的基础上引入了 de Bruijn 图的思想，来对 Chord 路由改进的一种 DHT（分布式哈希表算法）。

在一个 Koorde 网络中，当节点 F 失效时，它的前邻居节点 X 和后邻居节点 Y 探测出该节点失效信息。根据 Koorde 网络的结构特性，Koorde 网络中节点 Key 值是存储在其最接近的后续邻居节点中的。所以当节点 F 失效，那些需要得知改信息的节点 N 范围如下所示，并且指定失效节点的替代节点为其后续邻居节点 Y。

$$\begin{cases} F \in (X, Y] & (1) \\ F \notin (N, successor(N)] & (2) \\ i \circ topBit(kshift) \notin (N, successor(N)] & (3) \end{cases}$$

如上式所示，当节点 N 满足：(1) 失效节点 F 的前后邻居节点分布为 X

和 Y，(2) F 不属于节点 N 和 N 的后续节点区间（区间包括 N 的后续节点 successor(N)但不包括 N），(3) F 的 de Bruijn 虚拟节点 i 经过 topBit(kshift)移位操作后也不属于节点 N 和 N 的后续节点区间（区间包括 N 的后续节点 successor(N)但不包括 N）。当节点 N 同时满足上述三式时，N 就为所要通知的节点。

节点 Y 通过向范围内的 LDN 节点组播该失效信息，让其在范围内转发节点失效信息。

综上所述，本发明实施例利用在现有的 DHT 路由表上实现对称机制，当网络中节点出现扰动时，迅速根据扰动节点和周围邻居节点 NodeId 之间的距离推算出扰动信息可能影响的范围，并由邻居节点将其扰动信息主动通知到范围内迫切需要知道这一信息的远方节点，继续维护其原两节点间的一跳路由，避免远方节点通过“爬行”的方式来查找扰动节点。当节点失效时，通知路由表中远端的节点比仅通知邻居 K 个节点更能提高整个网络对节点扰动行为的感知度，通知信息数量级为 O(logN)，也有可能为 O(1)，其值和 DHT 中每个节点的路由表大小相关，从而能够直接提高整个 P2P 网络在 churn 环境下的路由查找效率，降低系统维护开销，提高系统稳定性。

基于同一发明构思，本发明实施还提供一种网络设备，如图 7A 所示，包括：第一确定模块 71、第一发送模块 72；其中，第一确定模块 71，用于根据失效节点与失效节点的邻居节点之间的距离，确定路由指向失效节点的节点范围；第一发送模块 72，用于将失效节点的失效信息发送给节点范围内、失效节点的远端邻近节点。

如图 7B 所示，一个实施例中，图 7A 所示的网络设备还可以包括探测模块 73，用于按设定周期对网络中邻居节点的状态进行探测，确定失效节点。

一个实施例中，第一确定模块 71 还可以用于根据失效节点与失效节点的邻居节点之间的距离、网络的路由特性，确定路由指向失效节点的节点范围。

如图 7C 所示，一个实施例中，图 7A 所示的网络设备还可以包括第二确

定模块 74，用于根据网络存储的键值特性，在失效节点的邻居节点中确定发送失效信息的替代节点。

若网络为 Kademlia DHT 网络或 Pastry DHT 网络，第二确定模块 74 还可以用于确定替代节点为离失效节点异或距离最小的邻居节点；或，若网络为 Chord DHT 网络或 Koorde DHT 网络，第二确定模块 74 还可以用于确定替代节点为失效节点的后续节点。

失效信息包括替代节点的地址。失效信息还可以包括失效节点的标识、或失效节点的地址、或失效节点的键值的组合、或失效节点的标识、地址、键值的两两组合或三者的组合。失效信息还可以包括失效节点对所述远端邻近节点的影响范围。

基于同一发明构思，本发明实施例还提供一种网络设备，该网络设备为失效节点的远端邻近节点，其结构如图 8A 所示，包括：接收模块 81、处理模块 82；其中，接收模块 81，用于接收失效节点的失效信息，失效信息包括失效节点的替代节点的地址；处理模块 82，用于根据失效信息更新对应的路由表信息。

一个实施例中，失效信息还包括失效节点的标识、或失效节点的地址、或失效节点的键值的组合、或失效节点的标识、地址、键值的两两组合或三者的组合。

失效信息还可以包括失效节点对本节点的影响范围；此时，如图 8B 所示，图 8A 所示的网络设备还可以包括：第二发送模块 83，用于在接收模块 81 接收到失效信息后，将失效信息转发给影响范围内的节点。

一个实施例中，失效信息包括失效节点对本节点的影响范围，此时，第二发送模块 83 还可以用于在接收模块 81 接收到失效信息后，在影响范围未超过本节点的邻居节点维护范围时，直接将失效信息转发给影响范围内的节点；在影响范围超过本节点的邻居节点维护范围时，将失效信息转发给离影响范围最近的本节点的邻居节点，通知该邻居节点继续将失效信息转发给影

响范围内其它节点。此处影响范围的确定与前述在不同的 P2P 对等网络中的影响范围的确定方法类似，包括在 Kademlia 网络、Pastry 网络、Chord 网络、Koorde 网络中影响范围的确定方法。

基于同一发明构思，本发明实施例还提供一种 P2P 对等通信网络，其结构如图 9 所示，包括失效节点 91，还包括失效节点的邻居节点 92、远端邻近节点 93，其中：失效节点的邻居节点 92，用于根据本节点与失效节点之间的距离，确定路由指向失效节点的节点范围；将失效节点的失效信息发送给节点范围内、失效节点的远端邻近节点；远端邻近节点 93，用于接收失效信息，根据失效信息更新对应的路由表信息。

本领域普通技术人员可以理解上述实施例方法中的全部或部分步骤是可以通过程序来指令相关的硬件完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：ROM、RAM、磁盘或光盘等。

本发明实施例中，根据失效节点与邻居节点之间的距离，确定路由指向所述失效节点的节点范围；将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点；所述远端邻近节点根据所述失效信息更新对应的路由表信息，并根据节点范围通知其周边的邻居节点更新各自对应的路由表信息，与现有技术中节点失效时仅将失效信息通知 K 个邻居节点的技术方案相比，增强了整个 P2P 对等网络对节点扰动行为的感知度，提高了整个 P2P 对等网络的路由查找效率和系统的稳定性。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若对本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

权 利 要 求 书

1、一种 P2P 对等网络中节点失效后路由表信息的更新方法，其特征在于，该方法包括:

根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围;

将所述失效节点的失效信息发送给所述节点范围内所述失效节点的远端邻近节点;

所述远端邻近节点根据所述失效信息更新对应的路由表信息。

2、如权利要求 1 所述的方法，其特征在于，在确定路由指向所述失效节点的节点范围的步骤之前，该方法还包括:

按设定周期对网络中邻居节点的状态进行探测，确定失效节点。

3、如权利要求 1 所述的方法，其特征在于，所述根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围的步骤，进一步包括:

根据失效节点与所述失效节点的邻居节点之间的距离、网络的路由特性，确定路由指向所述失效节点的节点范围。

4、如权利要求 1 所述的方法，其特征在于，所述根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围步骤之后，该方法还包括:

根据网络存储的键值特性，在所述失效节点的邻居节点中确定发送所述失效信息的替代节点。

5、如权利要求 4 所述的方法，其特征在于，

所述网络为 Kademlia DHT 网络或 Pastry DHT 网络，所述替代节点为离失效节点异或距离最小的邻居节点;

或，所述网络为 Chord DHT 网络或 Koorde DHT 网络，所述替代节点为所

述失效节点的后续节点。

6、如权利要求 4 所述的方法，其特征在于，所述失效信息包括所述替代节点的地址。

7、如权利要求 6 所述的方法，其特征在于，所述失效信息还包括失效节点的标识、或失效节点的地址、或失效节点的键值的组合、或失效节点的标识、地址、键值的两两组合或三者的组合。

8、如权利要求 6 所述的方法，其特征在于，所述失效信息还包括所述失效节点对所述远端邻近节点的影响范围;

该方法进一步包括:

所述远端邻近节点接收到所述失效信息后，将所述失效信息转发给所述影响范围内的节点。

9、如权利要求 8 所述的方法，其特征在于，所述远端邻近节点将所述失效信息转发给所述影响范围内的节点包括:

确定所述影响范围未超过本节点的邻居节点维护范围时，直接将所述失效信息转发给所述影响范围内的节点;

确定所述影响范围超过本节点的邻居节点维护范围时，将所述失效信息转发给离所述影响范围最近的本节点的邻居节点，通知该邻居节点继续将所述失效信息转发给所述影响范围内其它节点。

10、如权利要求 8 或 9 所述的方法，其特征在于，该方法进一步包括:

所述影响范围内的节点接收到所述失效信息后更新对应的路由表信息。

11、一种网络设备，其特征在于，包括:

第一确定模块（71），用于根据失效节点与所述失效节点的邻居节点之间的距离，确定路由指向所述失效节点的节点范围;

第一发送模块（72），用于将所述失效节点的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点。

12、如权利要求 11 所述的设备，其特征在于，还包括:

探测模块（73），用于按设定周期对网络中邻居节点的状态进行探测，确定失效节点。

13、如权利要求 11 所述的设备，其特征在于，所述第一确定模块（71）进一步用于根据失效节点与所述失效节点的邻居节点之间的距离、网络的路由特性，确定路由指向所述失效节点的节点范围。

14、如权利要求 11 所述的设备，其特征在于，所述设备还包括:

第二确定模块（74），用于根据网络存储的键值特性，在所述失效节点的邻居节点中确定发送所述失效信息的替代节点。

15、如权利要求 14 所述的设备，其特征在于，所述网络为 Kademlia DHT 网络或 Pastry DHT 网络，所述第二确定模块（74）进一步用于确定所述替代节点为离失效节点异或距离最小的邻居节点;

或，所述网络为 Chord DHT 网络或 Koorde DHT 网络，所述第二确定模块进一步用于确定所述替代节点为所述失效节点的后续节点。

16、如权利要求 11 所述的设备，其特征在于，所述第一发送模块（72）所发送的失效信息包括替代节点的地址。

17、一种网络设备，其特征在于，包括:

接收模块（81），用于接收失效节点的失效信息，所述失效信息包括所述失效节点的替代节点的地址;

处理模块（82），用于根据所述失效信息更新对应的路由表信息。

18、如权利要求 17 所述的设备，其特征在于，所述接收模块（81）所接收的失效节点的失效信息还包括失效节点的标识、或失效节点的地址、或失效节点的键值的组合、或失效节点的标识、地址、键值的两两组合或三者的组合。

19、如权利要求 17 所述的设备，其特征在于，所述接收模块（81）还用于接收所述失效节点对本节点的影响范围的失效信息;

所述设备还包括:

第二发送模块（83），用于在所述接收模块（81）接收到所述失效信息后，将所述失效信息转发给所述影响范围内的节点。

20、如权利要求 19 所述的设备，其特征在于，所述第二发送模块（83）进一步用于在所述接收模块（81）接收到所述失效信息后，在所述影响范围未超过本节点的邻居节点维护范围时，直接将所述失效信息转发给所述影响范围内的节点；在所述影响范围超过本节点的邻居节点维护范围时，将所述失效信息转发给离所述影响范围最近的本节点的邻居节点，通知该邻居节点继续将所述失效信息转发给所述影响范围内其它节点。

21、一种 P2P 对等通信网络，包括失效节点（91），其特征在于，还包括失效节点的邻居节点（92）、远端邻近节点（93），其中：

失效节点的邻居节点（92），用于根据本节点与所述失效节点之间的距离，确定路由指向所述失效节点的节点范围；将所述失效节点（91）的失效信息发送给所述节点范围内、所述失效节点的远端邻近节点；

远端邻近节点（93），用于接收所述失效信息，根据所述失效信息更新对应的路由表信息。

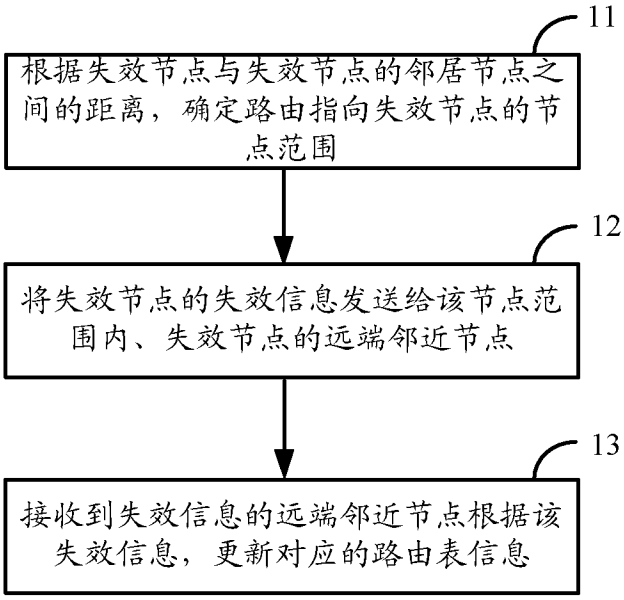

11
根据失效节点与失效节点的邻居节点之间的距离，确定路由指向失效节点的节点范围
12
将失效节点的失效信息发送给该节点范围内、失效节点的远端邻近节点
13
接收到失效信息的远端邻近节点根据该失效信息，更新对应的路由表信息

图 1

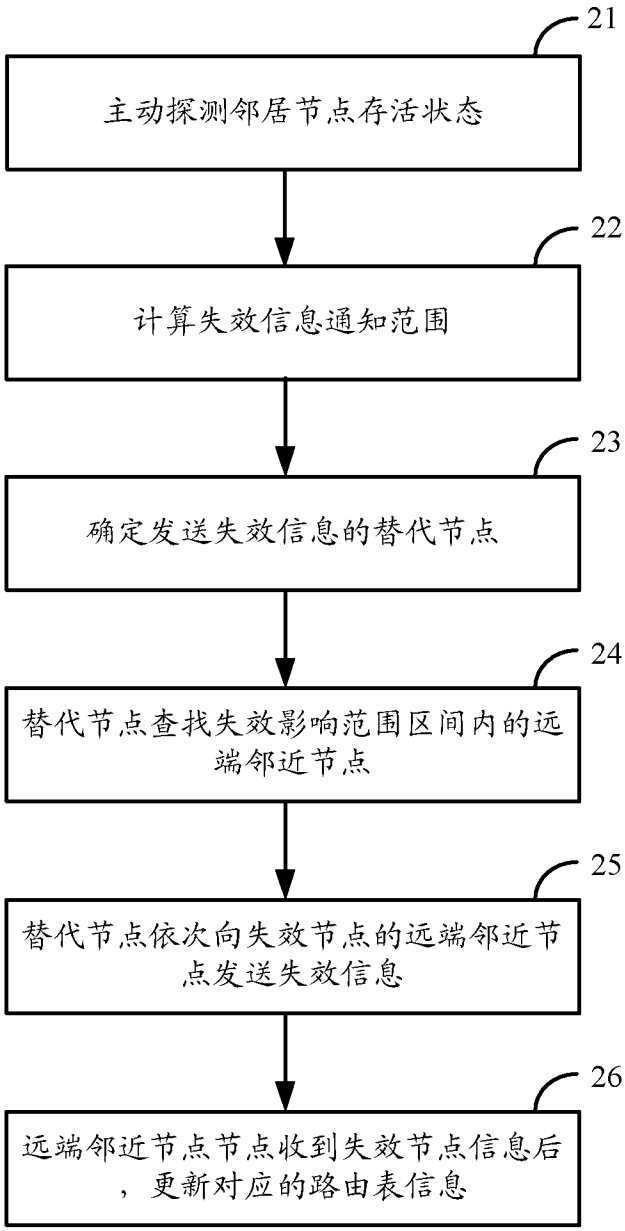

21
主动探测邻居节点存活状态
22
计算失效信息通知范围
23
确定发送失效信息的替代节点
24
替代节点查找失效影响范围区间内的远端邻近节点
25
替代节点依次向失效节点的远端邻近节点发送失效信息
26
远端邻近节点节点收到失效节点信息后，更新对应的路由表信息

图 2

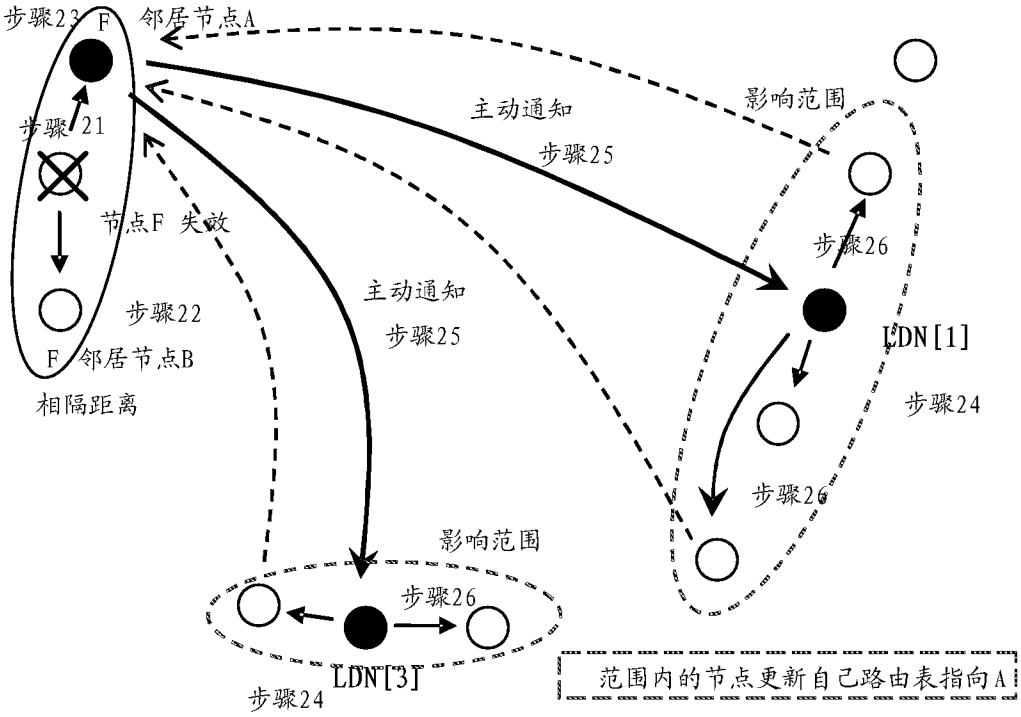

步骤23
F
邻居节点A
步骤21
节点F 失效
步骤22
F 邻居节点B
相隔距离
主动通知
步骤25
影响范围
步骤26
LDN[1]
步骤24
主动通知
步骤25
步骤26
影响范围
步骤26
LDN[3]
步骤24
范围内的节点更新自己路由表指向A

图 3

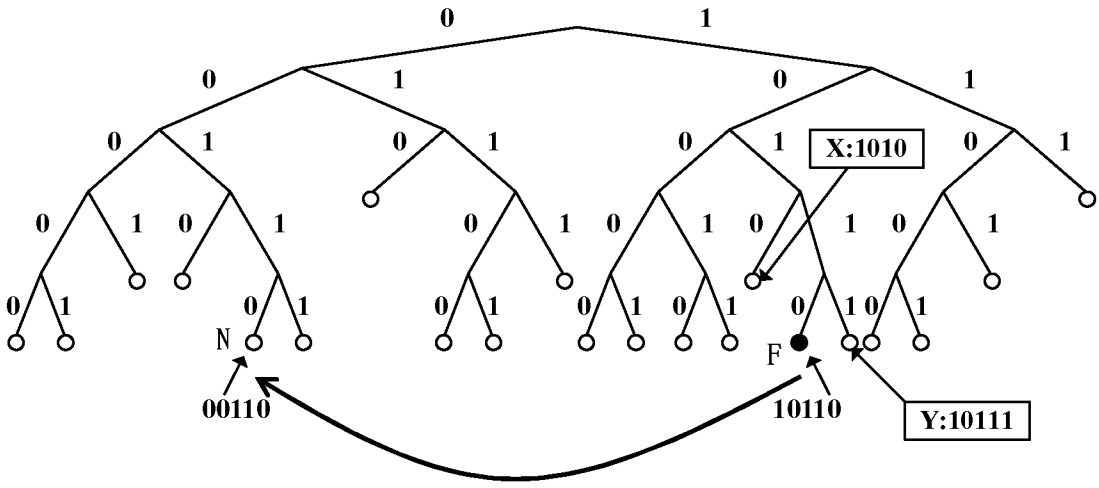

X:1010
N
00110
F
10110
Y:10111

图 4

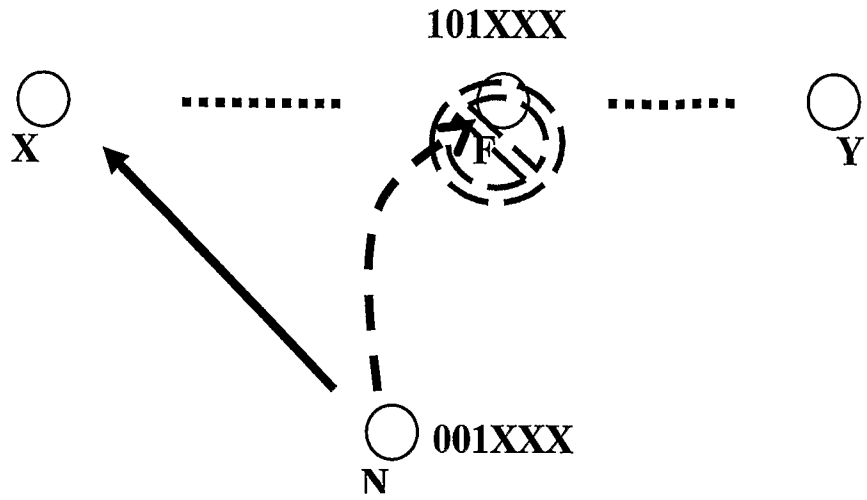

101XXX
X
F
Y
001XXX
N

图 5

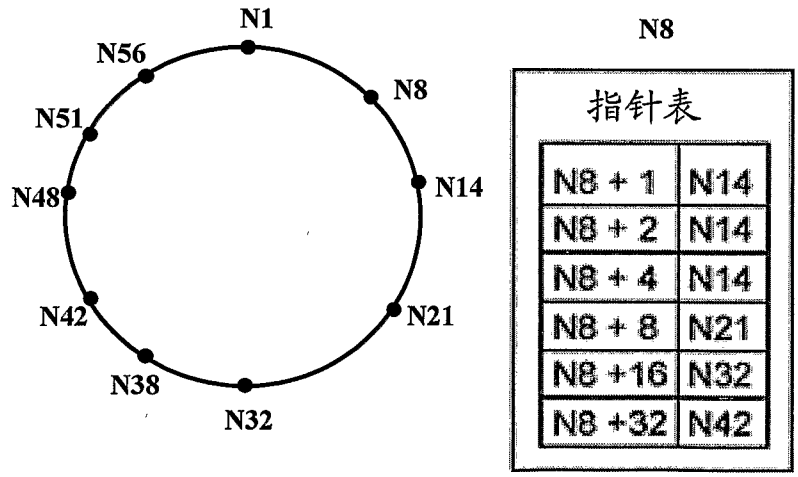

N1
N56
N8
N51
N14
N48
N42
N21
N38
N32
N8
指针表
N8 + 1 N14
N8 + 2 N14
N8 + 4 N14
N8 + 8 N21
N8 +16 N32
N8 +32 N42

图 6

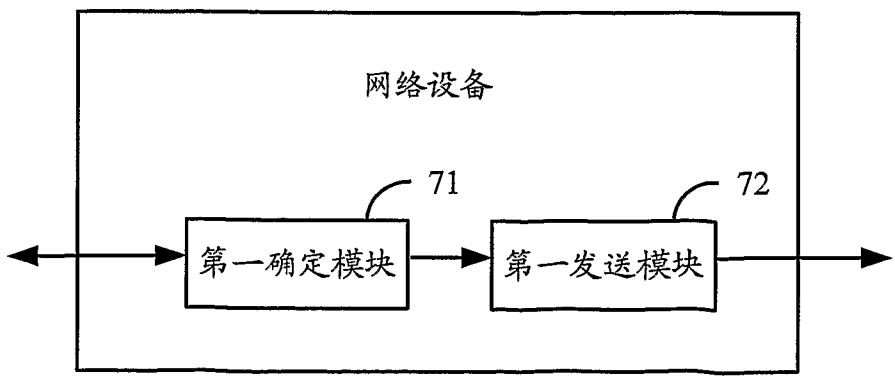

网络设备
71
72
第一确定模块
第一发送模块

图 7A

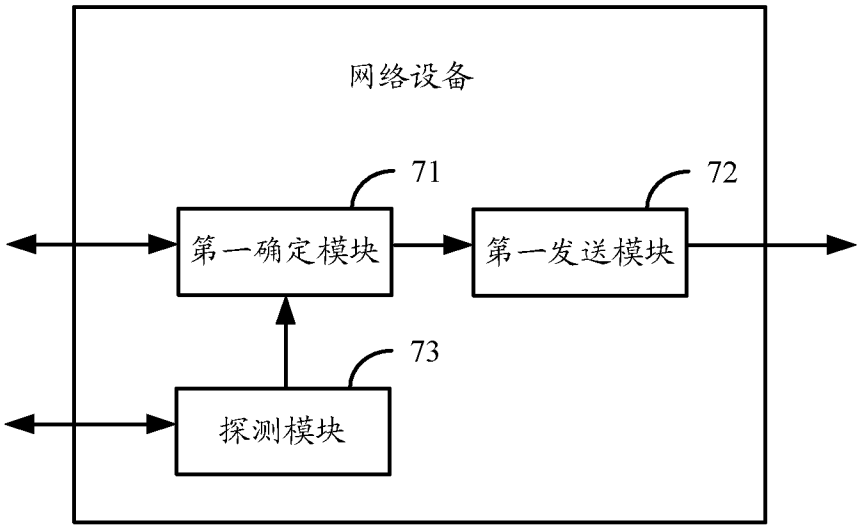

网络设备
71
72
第一确定模块
第一发送模块
73
探测模块

图 7B

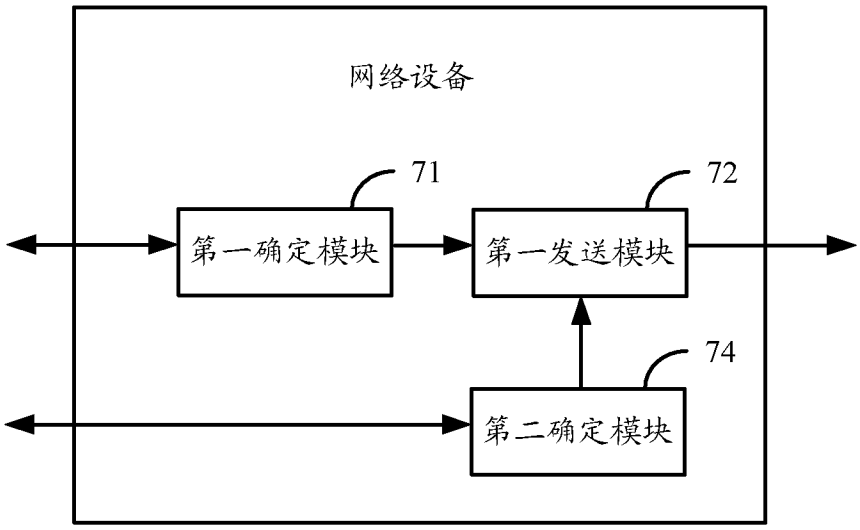

网络设备
71
72
第一确定模块
第一发送模块
74
第二确定模块

图 7C

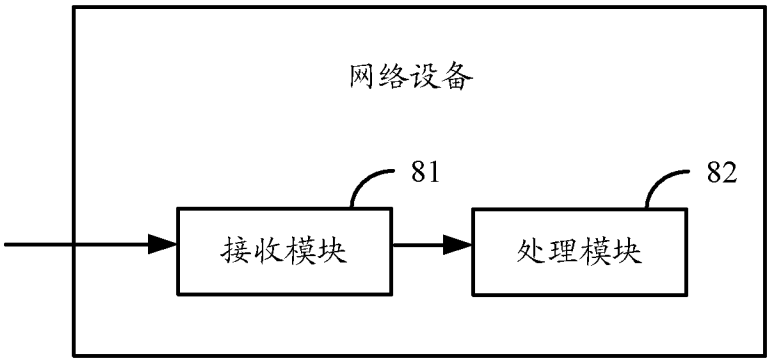

网络设备
81
82
接收模块
处理模块

图 8A

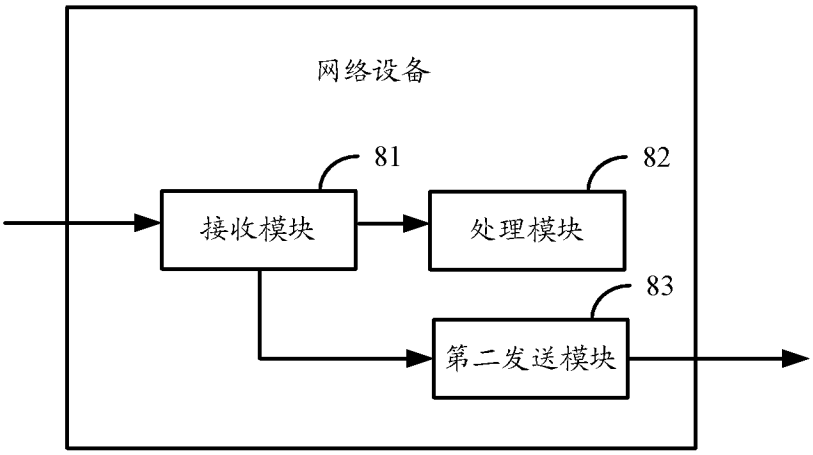

网络设备
81
82
接收模块
处理模块
83
第二发送模块

图 8B

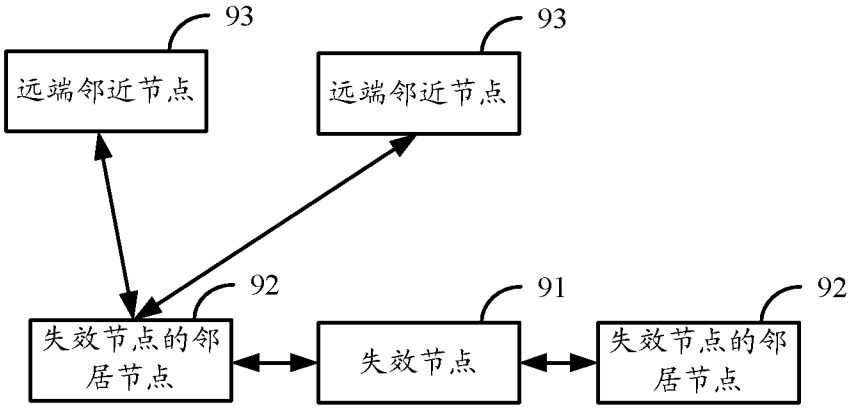

93
93
远端邻近节点
远端邻近节点
92
91
92
失效节点的邻居节点
失效节点
失效节点的邻居节点

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2008/072836

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/24(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04L, H04Q, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, PAJ, CNKI, CPRS: P2P, peer to peer, DHT, route, fail, fault, bound, neighbor, proximity, distance, range, region, churn, fluctuation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US2005/0063318A1(Xu et al.) ,24 Mar. 2005(24.03.2005) , paragraphs 0022-0063, figures 1-8	1-21
Y	CN1283076C(UNIV TSINGHUA), 01 Nov. 2006(01.11.2006), page 2 line15 – page 8 line 6 of description, figure 1	1-21
A	CN1681257A(MICROSOFT CORP), 12 Oct. 2005(12.10.2005), the whole document	1-21
A	EP1802070A1(NTT DoCoMo, Inc.), 27 Jun. 2007(27.06.2007), the whole document	1-21

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 Jan. 2009(12.01.2009)	**05 Feb. 2009 (05.02.2009)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer WEI, Feng Telephone No. (86-10)62413556

Form PCT/ISA/210 (second sheet) (April 2007)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2008/072836

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US2005/0063318A1	24.03.2005	None	
CN1283076C	01.11.2006	CN1564543A	12.01.2005
CN1681257A	12.10.2005	EP1583326A2	05.10.2005
		US2005/0223102A1	06.10.2005
		JP2005323346A	17.11.2005
		AU2005201191A1	20.10.2005
		BR0501178A	01.11.2005
		CA2503360A1	30.09.2005
		KR20060045065A	16.05.2006
		MXPA05003462A	23.11.2005
		RU2005109223A	10.10.2006
EP1802070A1	27.06.2007	EP1802070B1	18.06.2008
		JP2007174672A	05.07.2007

国际检索报告

国际申请号
PCT/CN2008/072836

A. 主题的分类

H04L 12/24(2006.01) i

按照国际专利分类表(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: H04L, H04Q, G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

WPI,EPODOC,PAJ,CNKI,CPRS: 点对点，对等，网络，路由，失败，失效，结构化，范围，距离，邻，扰动，P2P, peer to peer, DHT, route, fail, fault, bound, neighbor, proximity, distance, range, region, churn, fluctuation

C. 相关文件

类　型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	US2005/0063318A1(Xu et al.) ,24.3 月 2005(24.03.2005) ,第 0022 段至第 0063 段、图 1-8	1-21
Y	CN1283076C(清华大学), 01.11 月 2006(01.11.2006), 说明书第 2 页第 15 行至第 8 页第 6 行、图 1	1-21
A	CN1681257A(微软公司), 12.10 月 2005(12.10.2005), 全文	1-21
A	EP1802070A1(NTT DoCoMo, Inc.), 27.6 月 2007(27.06.2007), 全文	1-21

☐ 其余文件在 C 栏的续页中列出。　　☒ 见同族专利附件。

* 引用文件的具体类型：
"A" 认为不特别相关的表示了现有技术一般状态的文件
"E" 在国际申请日的当天或之后公布的在先申请或专利
"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件
"O" 涉及口头公开、使用、展览或其他方式公开的文件
"P" 公布日先于国际申请日但迟于所要求的优先权日的文件
"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件
"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性
"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性
"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
12.1 月 2009(12.01.2009)	**05.2 月 2009 (05.02.2009)**
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号：　(86-10)62019451	受权官员 **魏峰** 电话号码：(86-10) **62413556**

国际检索报告

关于同族专利的信息

国际申请号
PCT/CN2008/072836

检索报告中引用的专利文件	公布日期	同族专利	公布日期
US2005/0063318A1	24.03.2005	无	
CN1283076C	01.11.2006	CN1564543A	12.01.2005
CN1681257A	12.10.2005	EP1583326A2	05.10.2005
		US2005/0223102A1	06.10.2005
		JP2005323346A	17.11.2005
		AU2005201191A1	20.10.2005
		BR0501178A	01.11.2005
		CA2503360A1	30.09.2005
		KR20060045065A	16.05.2006
		MXPA05003462A	23.11.2005
		RU2005109223A	10.10.2006
EP1802070A1	27.06.2007	EP1802070B1	18.06.2008
		JP2007174672A	05.07.2007