

基于非结构化 P2P 的医疗物联网信息发现^{*}

邢 丹 姚俊明

(济宁医学院医学信息工程学院 日照 276826)

〔摘要〕 介绍基于 P2P 的物联网物品信息发现机制相关研究概况, 分析比较 3 种分布式 P2P 拓扑结构, 提出基于非结构化 P2P 进行物联网的信息发现, 即采用基于兴趣的层次化拓扑构建选择和基于兴趣簇的具有自适应能力的搜索算法, 并对该算法进行分析。

〔关键词〕 非结构化 P2P; 医疗物联网; 信息发现; 兴趣

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2015.10.012

Information Discovery for Medical Internet of Things Based on Unstructured P2P XING Dan, YAO Jun-ming, College of Medical Information Engineering, Jining Medical University, Rizhao 276826, China

〔Abstract〕 The paper introduces researches related to the discovery mechanism of things and information for the Internet of Things based on P2P, analyzes and compares three kinds of distributed P2P topological structures, and proposes information discovery for the Internet of Things based on unstructured P2P. This means to use interest-based selection of hierarchical topological construction and the self-adaptable search algorithm based on interest clusters and analyze this algorithm.

〔Keywords〕 Unstructured P2P; Medical Internet of Things; Discovery service; Interest

1 引言

物联网是未来智慧医疗的关键技术之一, 可将其应用在智慧健康管理系统中。而在典型的物联网——EPCglobal (国际物品编码协会和美国统一代码委员会的一个合资公司) 网络中, 物联网信息发现是通过对象命名服务 (Object Naming Service,

ONS) 来完成的。ONS 的实质是将一个物品的产品电子代码 (Electronic Product Code, EPC) 映射到 1 个或者多个网络资源地址统一资源标识符 (Uniform Resource Identifier, URI), 通过 URI 可以查找到关于物品的详细信息, 即对应着一个 EPC 信息服务。目前, 传统物联网中的分布式物品信息节点具有各节点储存的数据信息更新频繁、不同节点储存着相同物品名称的不同数据部分、节点数目巨大、节点相关的数据信息具有很高的安全敏感性等特点^[1], 这就决定了采用传统的查找方式必然会对根服务器造成很大的查询压力, 形成查询瓶颈, 降低查询效率。因此, 设计高效可行的医疗物联网信息发现机制是亟需解决的问题。而对等网 (Peer to Peer, P2P) 技术具有非中心化、可扩展性及健壮性能好、隐私保护及负载均衡等优点, 从而能够确保联网物品的相关信息能够被高效、准确和安全地寻址、定

〔修回日期〕 2015-05-06

〔作者简介〕 邢丹, 讲师, 发表论文 7 篇; 姚俊明, 讲师, 硕士, 发表论文 6 篇。

〔基金项目〕 济宁医学院青年基金项目“基于 P2P 的物联网资源发现机制的研究”; 山东省自然科学基金项目 (项目编号: ZR2011HL002); 济宁医学院校级科研计划项目 (项目编号: JY2013KJ036)。

位和查询。

2 基于 P2P 的物联网物品信息发现机制相关研究

BRIDGE 项目以 EPCglobal 相关规范为基础，提出了物联网中发现服务（Discovery Service，DS）的 8 种可行模式，从各方面进行了比较^[7]。这 8 种模式与 ONS 系统相比，都能够很好地完成任意给定的物品编码到商品流通链中与其相关联的多个 EPC 信息服务（EPC Information Services，EPCIS）的定位查找操作，在保证查全率的前提下实现动态查询。孔宁^[1]对这 8 种模式从安全性、响应速度等方面进行分析，得出 EPCIS 目录模式和请求传播模式是较好的两种模式，提出了采用基于半分布式 P2P 的拓扑结构设计物联网中的物品信息发现机制，通过改进并提高其发现效率及网络可扩展性，可使其满足物联网的物品信息发现需求，但未对此实现过程进行深入研究。黄宇等^[2]采用基于半分布式 P2P 网络的物联网信息发现服务，在超级节点间的路由算法中采用了分布式 P2P 中的 Chord 算法来实现，但未对其做详细分析。李占波等^[3]采用改进分布式哈希表（Distributed Hash Table，DHT）的结构化 P2P 的 ONS 解析机制，结果表明基于 DHT 的 ONS 解析系统能够有效地应对单点失效，相比于 Epcglobal ONS 系统具有更好的负载均衡性能。本文在请求传播模式基础上，采用非结构化 P2P 进行物联网的信息发现。

3 3 种分布式 P2P 拓扑结构分析

表 1 分布式 P2P 网络拓扑性能综合比较

比较标准	全分布式非结构化拓扑结构	全分布式结构化拓扑结构	半分布式拓扑结构
可扩展性	差	好	中
可靠性	好	好	中
可维护性	最好	好	中
发现算法效率	中	高	中
复杂查询	支持	不支持	支持

从表 1 可以看出，全分布式结构化拓扑结构的整体性能最好，采用此拓扑结构，物品信息发现机制能够很好地满足网络可扩展性、容错性以及查询的效率、查全率等要求；但是由于 DHT 算法本身的限制，其不支持复杂查询，因此无法满足物联网中对于物品内容的复杂查询。而半分布式拓扑结构的优点是性能、可扩展性较好，较容易管理；但对超级节点的依赖性大，易于受到攻击，容错性也受到影响。而全分布式非结构化拓扑结构由于采用了随机图的组织方式，结点度数服从幂律（Power-law）规律，从而能够较快发现目的结点，面对网络的动态变化体现了较好的容错能力，因此具有较好的可用性；支持复杂查询，如带有规则表达式的多关键词查询、模糊查询等。目前对此类结构的研究主要集中于改进发现算法和复制策略以提高发现的准确率和性能。根据物联网的特殊性，物品的信息发现机制需要具有各节点高度独立的自治性；网络具有良好的可扩展性和容错性；查询的效率和查全率高；支持内容、语义等复杂查询的特点。综合以上分析，本文采用非结构化拓扑结构来组织节点，但由于没有确定拓扑结构的支持，非结构化网络无法保证资源发现的效率，即使需要查找的目的结点存在发现也有可能失败，因此发现的准确性和可扩展性是非结构化网络面临的两个重要问题，需要采用优化的改进算法来发现资源。

4 基于非结构化 P2P 的医疗物联网物品信息发现机制

4.1 非结构化 P2P 的拓扑构建选择

传统的 P2P 网络结构均没有考虑语义和节点能力，因此其节点在网络中的位置存在巨大的随意性，忽视节点资源和网络结构的对应，使得其网络对于资源的搜索查询具有较大的盲目性，搜索的开销巨大，且效率较低。目前研究热点之一是引入节点兴趣^[4]，将兴趣相似节点聚集在一起，以提高资源搜索的成功率；同时，在资源搜索过程中，搜索查询消息只需要访问网络中少量节点，产生较高的搜索成功率和较低的资源搜索查询开销。周晓波

等^[5]采用了一种基于层次化兴趣表达的拓扑生成模型，通过兴趣的定义及维护从统计上在资源和网络拓扑之间建立了联系。在物联网的应用中，尤其是物品流通领域，通常所关心的商品是按照人群来进行划分的，如孕妇经常关心孕妇服装、待产包等，且其所感兴趣的商品通常存储在相对固定的位置。基于此，选取具有兴趣相似的节点来组织网络拓扑。这其中需要解决两个关键问题，其一是兴趣的表示问题。目前常用节点兴趣的描述方法主要有主题分类、向量空间模型（Vector Space Model, VSM）和资源描述框架（Resource Description Framework, RDF）^[1]。黄宇等^[2]采用兴趣树来表示资源，本文采用基于向量空间模型的方法进行兴趣树的组织，对兴趣树进行广度优先遍历得到资源所处的社区和兴趣的编号。另一个是社区维护问题，其过程描述为：当一个节点通过一个启动（Bootstrap）节点加入到网络中，将继承 Bootstrap 的邻居

表，且被加入到 Bootstrap 的邻居表中。随后，将不断地发起查询，对于那些成功的查询，把对应节点作为自己的邻居候选，然后对邻居表及候选表进行排序，删除那些与自己的社区编号距离太远的节点。如此不断反复，节点将逐渐地从初始的社区跨越到其社区编号所标志的社区中去。

4.2 非结构化 P2P 搜索算法的选取

采用非结构化 P2P 拓扑结构来进行资源发现，其中一个关键问题是采用何种搜索算法。非结构化的搜索策略可分为盲目搜索和信息搜索两大类。盲目搜索通过洪泛方式来搜索想要的资源，虽然查询效果较好，但需要耗费较多；信息搜索存储了资源已有的知识，所以能够比较快地找到资源。钱宁等^[6]将其更细致地划分为基于本地索引的方法、基于层次的方法和基于聚集的方法，表 2 归纳总结了各个方法的优缺点。

表 2 非结构化 P2P 搜索算法分析

方法	搜索成功率	产生消息量	网络动态变化	可维护性
基于本地索引	较高	被动方式：不增加主动方式：额外开销	一般	简单
基于层次	较高	多	不好	复杂
基于聚集	较高	少	好	一般

鉴于拓扑结构的构建选择了基于节点兴趣，即将资源和网络的拓扑结构建立了联系，采用簇的方式为聚集节点；同时为了符合物联网的自身特点，综合以上，采用基于聚集方法中的基于兴趣的方法进行搜索。其次，为了使节点具有高度独立的自治特性，采用具有自适应能力的基于簇搜索算法，即 SAICA 算法。

4.3 SAICA 算法具体工作过程

对于一个给定的查询，根据“六尺度”原则可知它的结果集很大程度上就在对应的簇集，而不是散落在整个系统，这样资源搜索问题就转化成如何快速定位对应簇集的问题。下面给出搜索的基本过程：（1）确定资源所在的簇集。当一个节点发起查询时，如果该节点是服务节点，跳（2）；否则，节点首先计算查询项 q 与该节点维护的资源 r 的相

似度 Similumar (q, r)，计算结果和预先设定的阈值（Threshold）进行比较，如果计算结果大于或等于设定的阈值，则认为成功找到了匹配的资源，将与查询项 q 相似度大于等于阈值的资源作为查询结果返回，同时该节点将查询消息发送给资源连接的服务节点；如果查询项与该节点维护的资源的相似度小于设定的阈值，查询消息则传递给需求连接对应的服务节点。（2）确定查询内容所在的具体位置。当查询消息到达后，接收到查询消息的服务节点先查看查询内容是否在本机上，然后根据它维护的资源索引确定内容是否在簇集内的叶节点上，同时将查询按照一定策略发送给她邻居服务节点。

4.4 算法评价

直观上看，SAICA 算法利用判断查询和当前节
(下转第 63 页)

- 究 [D]. 北京: 北京中医药大学, 2007.
- 13 晏峻峰, 朱文锋. 粗糙集理论在中医证素辨证研究中的应用 [J]. 中国中医基础医学杂志, 2006, 12 (2): 90-93.
 - 14 徐蕾, 贺佳, 孟虹, 等. 决策树技术及其在医学中的应用 [J]. 数理医药学杂志, 2004, 17 (2): 161-164.
 - 15 卢延鑫, 姚旭峰, 王松旺. 利用自然语言处理技术提取致病因素信息研究 [J]. 医学信息学杂志, 2013, 34 (3), 55-58.
 - 16 贾李蓉, 于彤, 李海燕, 等. 中医药学语言系统的语义网络框架概述 [C]. 北京: 中国中医药信息大会, 2014.
 - 17 于彤. 中医药学语言系统的语义网络框架 [EB/OL]. [2013-07-15]. <http://www.tcmkd.com/ontologies/tcm-mls/>.
 - 18 Brown P F, Desouza P V, Mercer R L, et al. Class-based N-gram Models of Natural Language [J]. Computational Linguistics, 1992, 18 (4): 467-479.
 - 19 Berger A L, Pietra V J D, Pietra S A D. A Maximum Entropy Approach to Natural Language Processing [J]. Computational Linguistics, 1996, 22 (1): 39-71.
 - 20 Bengio Y, Ducharme R, Vincent P, et al. A Neural Probabilistic Language Model [J]. The Journal of Machine Learning Research, 2003, (3): 1137-1155.
 - 21 Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space [J]. arXiv preprint arXiv: 1301.3781, 2013.
 - 22 Miñarro - Giménez J A, Marín - Alonso O, Samwald M. Applying Deep Learning Techniques on Medical Corpora from the World Wide Web: a prototypical system and evaluation [J]. arXiv preprint arXiv: 1502.03682, 2015.
 - 23 Miñarro - Giménez J A, Marín - Alonso O, Samwald M. Exploring the Application of Deep Learning Techniques on Medical Text Corpora [J]. Studies in Health Technology and Informatics, 2013, (205): 584-588.
 - 24 Van der Maaten L, Hinton G. Visualizing Data Using t-SNE [J]. Journal of Machine Learning Research, 2008, (9): 2579-2605.

(上接第 57 页)

点所含资源的相似度决定查询的转发路径。如果当前节点所含的资源 and 查询的相似度小于设定的阈值, 那么该节点所属簇集内的节点拥有和查询相关资源的可能性也较小——因为根据节点面向兴趣转移的拓扑连接调整, 同一簇集内维护相似资源的节点。因此, 借助历史反馈信息, 将查询路由给需求连接的服务节点。反之, 查询可能已经被发送到一个由一组查询主题相关的资源所在节点构成的社区中, 当前簇集内包含与查询相关的大部分资源对象。因此根据 SAICA 算法的策略, 对于给定的查询, 目标社区的定位基本上可以在一个跳数内完成, 有效控制了消息数量和搜索路径长度, 从而提升了系统的整体搜索性能。

5 结语

本文通过分析提出了基于兴趣的层次化拓扑构建方法和基于兴趣簇的具有自适应能力的搜索算

法, 解决了医疗物联网中资源发现服务存在的弊端。下一步的工作是对其进行模拟实验, 分析其同 ONS 的查询效率问题, 以及不在兴趣范围内的节点的解决方案。

参考文献

- 1 孔宁. 物联网资源寻址关键技术研究 [D]. 北京: 中国科学院计算机网络信息中心, 2008.
- 2 黄宇, 金蓓弘. 非结构化 P2P 系统 Overlay 优化技术综述 [J]. 小型微型计算机系统, 2008, 29 (2): 238-243.
- 3 李占波, 张哲. 基于 DHT-P2P 新型的 ONS 解析机制 [J]. 计算机工程与应用, 2013, 49 (3): 91-94.
- 4 苏森. 无结构 P2P 网络中基于语义和节点存储能力的搜索关键技术研究. [D]. 北京: 北京邮电大学, 2011.
- 5 周晓波, 周健, 卢汉成, 等. 一种基于层次化兴趣的非结构化 p2p 拓扑形成模型 [J]. 软件学报, 2007, 18 (12): 3131-3138.
- 6 钱宁, 吴国新. 无结构化 P2P 网络资源搜索机制研究综述 [J]. 计算机科学, 2010, 37 (4): 10-11.