J.T. Yao, V.V. Raghvan, G.Y. Wang (Eds.): WSS'04, pp. 155-162, 2004.

Enhancing Peer-to-Peer Network with RDF and Web

Yanwu Wang
Dept. of Computer Science
Xi'an Jiaotong University
Xi'an , China, 710049
blunt_hust@yahoo.com.cn

Huanzhao Wang
Dept. of Computer Science
Xi'an Jiaotong University
Xi'an, China, 710049
hzhwang@mail.xjtu.edu.cn

Abstract

As a new computing paradigm, peer-to-peer (P2P) has gained great popularity. However resource sharing, resource management and resource localization meets great challenge due to the distribute nature of P2P network. In this paper we introduce the using of Semantic Web recommendation RDF in P2P network. Through annotating resources with RDF metadata, the distributed resources link to each other in a semantic way. Thus application like resource discovery and searching, semantic routing and collaborative activities can become more efficient. For this purpose, we propose Web-based P2P metadata Network Infrastructure (WMNI) to support RDF metadata management, exchange and query. Another feature of WMNI is that it supports accessing P2P resources simply via a web browser i.e. User need not to install P2P application and bootstrap to join the P2P network. This feature is really important to capability- constraint devices like PDA and cell phone.

1. Introduction

P2P computing has proved to be a great success in distributed computing and instant messenger services. Applications such as Napster [1], Gnutella [2], Freenet [3] have been developed to share content over internet. In contrast to the conventional centralized way of managing resource, P2P network organizes these distributed resources in a decentralized manner. Furthermore the resources shared by peers are heterogonous. That means the resources are different in types (digital resource, education resource, medical resource), size etc. However most existing P2P applications characterize resource only with a few keywords. Keywords searching is of low preciseness and has "too many or none" problem: too few keywords lead to many searching results and too many keywords lead to no searching results. Another problem that P2P application faces is efficient resource query routing. Distributed Hash Table (DHT) method used in CAN [4], Chord [5] resolve keywords to location where resources are located. DHT maps each shared contents to string of number using hash function. Mapping resources into non-sense number induces the loss of relation between resources. More disadvantage of DHT is discussed in [6].

RDF [7] metadata itself is not something new, however in P2P network it really is to our best knowledge. Using rich metadata to describe the various heterogonous resources is a reasonable solution. But it incurs interoperability problem. For example someone would like to comment the writer of a book with "author", while another people would like to use "creator" instead. What's more the meaning of "author" in different context is also different. W3C consortium's recommendation Resource

Description Framework (RDF) is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information [7]. We adopt RDF/RDFS to describe resources and relation between resources, thus contents shared by peers are semantically related, e.g. Distributed resources of the same type are structured into a tree or list which can be efficiently navigated by user. The reasoning ability of RDF also enables the content-based search possible and improves the resource discovery ability of peer. Through the exchange of RDF metadata between different peer, resources are easily located. Resource query can route in a semantic way since each peer "know" which peer holds the requested resource.

In order to access resource in the P2P network, user has to install the P2P application in advance, starting the bootstrap to join the P2P network. However popular devices like PDA, cell phone have limited CPU cycles, finite memory and narrow link bandwidth. Our solution is to utilize web to present these resource to user. As shown in Figure 1, any device with a web browser can obtain desired resource via a P2P network Access Peer (AP). The expense of Web approach is protocol convention.

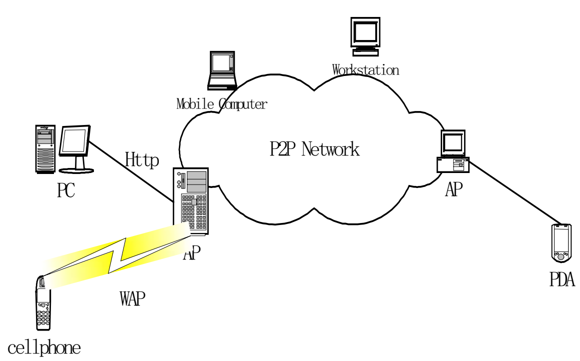

Figure 1. Access resource in P2P via web

For the above considerations, we propose our Web-based Metadata P2P Network Infrastructure (WMNI) upon JXTA. WMNI deals with RDF metadata management generated by different peer, metadata query and metadata exchange between peers. JXTA [8] is an open source computing platform designed for P2P computing, initiated by Sun Microsystems. Its goal is to develop basic building blocks and services to enable innovative applications for peer groups.

The rest of the paper is organized as follows. Section 2 gives a brief introduction to JXTA. In section 3 we present our proposed framework WMNI and elaborate it. In section 4 we have a detailed discussion of accessing P2P resource via web. Section 5 gives example applications built on WMNI. Section 6 is related works and Section 7 is conclusion.

2. JXTA architecture

Almost every P2P application introduces a different protocol, replicating already done work and causes unnecessary incompatibilities. JXTA provides protocols [8] that standardize the core P2P functionality, which includes peer discovery, self-organization into peer groups, advertisement and discovery of resources, communication and monitoring. Figure 2 shows JXTA's software architecture.

JXTA Core Layer deals with peer establishment, communication management such as routing, and other low-level plumbing. JXTA Services Layer is a service layer that deals with higher-level concepts, such as indexing, searching, and file sharing. These services, which make heavy use of the plumbing features provided by the core, are useful by themselves but also are commonly included as components in an overall P2P system. In JXTA all components such as Peers, Peer Groups, Pipes, peer services, Peer Group Services, are represented with corresponding advertisement. An advertisement is an XML document denoting these components. Our framework is built on the JXTA Services layer. WMNI makes use of JXTA core to provide basic service like peer organization and communication, but the methods that we proposed is not limited to JXTA.

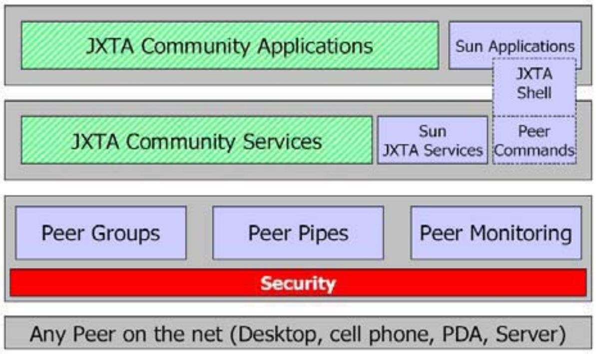

Figure 2. Software architecture of JXTA

3. WMNI architecture

The WMNI architecture aims to provide an underlying RDF metadata infrastructure to facilitate the development of various P2P applications. Our basic idea is to utilize RDF metadata to annotate resource, making distributed resources link to each other in a semantic way. WMNI consists of the following components: information space to store RDF metadata, management and exchange of metadata between peers, interfaces to query and reason on metadata. Figure 3 shows our proposed architecture WMNI.

3.1 RDF and data storage

Relations between resources are expressed with RDF statements, so back-end databases are employed to support this feature. The graphic representation of RDF is very similar to ER diagram, so RDF metadata can be easily stored in relation database. Another alternative method is to store these statements in XML database since RDF data is serialized with XML syntax. Various tools are already available for handing RDF-based metadata. Jena [9] is an open source project hold by HP. It provides java-programming interface for reading, writing, querying RDF in XML. Redland [10] is a library that provides a high-level interface for RDF allowing the RDF graph to be parsed from XML, stored, queried and manipulated.

Shared resources can be stored either in database or file system according to the characteristic of content. Relation Database is suitable for storing highly structured data that are generated dynamically such as report forms. File systems can be used to store video files, web document.

To associate each resource with its corresponding annotation (RDF statements), an additional Object Oriented Database storage is required to record the mapping between resources and RDF statements. Each shared contents is assigned with a Unique Resource Identifier (URI). This identifier can be acquired through hash digest, for example SHA1, on the shared content.

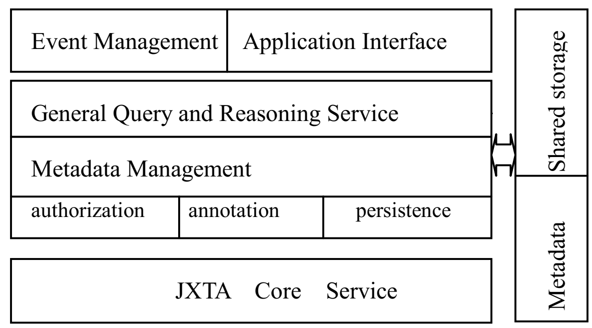

Figure 3 Architecture of WMNI

3.2 Metadata management

Metadata management includes annotation, persistence and authorization operation on metadata. When a user decides to share a resource (video file, web document etc) within the P2P network, several steps are taken in the following order. First the sharing contents are annotated according to some RDF schemas, and then the annotation is saved in metadata storage. To announce the existing of content to other JXTA peers, the information about the metadata is encapsulated in an advertisement, and then propagates to other peers. Here we have two problems: what granularity of metadata information to send and which peers to receive the information. Granularity means the abstract levels of information to be published, e.g. in schema level (DC[11], LOM[12]) or in entries level (DC, DC:Creator, DC:language, LOM). The second question concerns about clustering of information. Sending the information to all peers blindly may incur unnecessary traffic. In [13] various cluster strategies are covered based on the consideration of efficient query message routing.

Annotation service: To make the distributed resource in P2P network semantically link to each other. We propose to comment resource with RDF metadata according to RDF Schemas. In Semantic Web context, RDF Schemas are vocabulary of domain. The design of domain specific vocabulary is so-called ontology engineering. It is really a sophisticated process and only can be done by domain expert. Several standard RDF Schemas have been developed in some domain. For example Dublin Core (DC) is designed for category and retrieve learning material, IEEE-LOM/IMS (LOM) is designed for exchanging educational resources.

Persistence and monitoring: persistence includes add, remove, update of metadata in RDF database. Monitoring keeps track of modification of RDF metadata and triggers modification events. The interfaces provided by database system are primitive. These interfaces should be published as JXTA service so that others peers can consume them. As we have mentioned early in this section, metadata changes in one peer should be publish to others peers. So each persistence operation will be monitored and trigger an event. Other peers are notified with the event in a JXTA message. We introduce two modes to dispatch the message to other peers: *active mode* and *lazy mode*. In active mode, when a persistence event occurs, a message is sent to remote peers immediately. In lazy mode, only when the peer receives a query, would the peers send such a message. Different model can be adopted in different application environment. The active model is suitable for collaborative application like blackboard writing/reading, while lazy mode suitable for common content sharing application.

Replication, leasing, and authorization: The objective of metadata replication is to facilitate resource localization, while the objective of data replication is to achieve reliable storage in case of network failure. Here we mainly deal with replication of RDF metadata. The redundancy of metadata can utilize the JXTA replication service. Leasing is lifetime management of replicas. To avoid RDF storage from being flooded with metadata, outdated information should be removed according to a certain strategy. Some simple strategies like LRU (least recently used) or LFU (least frequent used) can be adopted. Authorization deals with security issue of RDF metadata. Metadata authorization prevents malicious user from fabricating metadata.

3.3 Query and reasoning

The common query and reasoning service is designed to shield the difference between underlying metadata query language, presenting user an easy to use interface. As discussed before, the resource description (i.e. metadata) in our system can be stored in XML database or traditional relation database. Since the metadata is expressed with XML, some XML query languages like XPath, XQuery can be used to query metadata. Also some SQL like languages, for example RQL [14], have been developed to query RDF metadata in database. These query languages have different syntax. Thus we need a general-purpose query language that acts as bridge between them. RDF-QEL [15], an XML-based Query Exchange Language, provides the syntax for an overall standard query interface across heterogeneous peer for any kind of RDF metadata. The General query service is application interfaces for this uniform language. Peer who wants to perform some query task can simply register query request to peers that provide query service. When query service is invoked, the query message is distributed to P2P network transparently. The requesting peer is notified with advertisement when the query process is finished.

4.4 Application interface

The peers in our system are highly heterogeneous peers. That means they may apply different service implementation .The service implementation is service provider of application interface. This idea is very similar to JNDI. The service interfaces are highly abstract, independent of its implementation. Applications developed in one peer can easily migrate to another peer after only modifying the service binding statement. With these application interfaces powerful and efficient application can be built.

3.5 Event management

The P2P network is dynamic and communication often takes long delay, asynchronous communication is preferred. Event subscribing and notification model satisfies this requirement. As discussed before metadata persistence operation will trigger events. The "interested peers" are notified with the changes of metadata information. In addition, the metadata query service also works in asynchronous mode. When a user issue a query, the query request is submitted to P2P network, registering with a callback function to handle the notification. Event management is especially important for those collaborative P2P applications.

4. Accessing resource via Web

4.1 Notion definition

The following is notions that are used later:

Foreigner: User who consumes the P2P network resource without having to install P2P application in advance. In this paper the foreigner acquires resources through a web browser.

Service Peer (SP): peer that provides resources consumed by foreigner.

Access Peer (AP): peer that mediates the communication between foreigner and SP.

4.2 web-base solution

The web-base approach enables those capability-constraint devices, i.e. foreigner, to access resources in P2P network. Peers communicate with each other using specific platform protocols, which are not compatible with current Internet protocols such as Http, RMI, and IIOP. Foreigners cannot interact with peers if they don't participate the P2P network in advance. One solution is to re-implement P2P service as web service. JXTA SOAP binding [16] is designed to allow SOAP communication over the JXTA P2P network. [17] discusses how to expose existing P2P service as web service. The JXTA-RMI project [18] allows to develop applications with familiar Remote Method Procedure. Through running a java applet in the client's web browser, foreigners can communicate with peer directly. But [16], [18] are ongoing, not mature and restricted to JXTA platform only. Exposing JXTA service as web service seems attractive and promising, however, the protocol conversion needs much technique work due to considerable protocol incompatibility. [19] built a web system over its SIONET P2P system. It employed a P2P application control in every peer to eliminate protocol difference.

In contrast to [19], our solution is a light-weight one. Figure 4 illustrates the suggested solution. Both AP and SP contain a Web server, running a Java servlet engine. The foreigner connects to an AP querying for some resources. The AP sends the query to P2P network through query interface provided by WMNI and locates the SP. Then AP relays the communication between foreigner and SP. The Http packet is load of JXTA pipe. First AP encapsulates the request http packet sent from foreigner into JXTA message, and then AP transmits the JXTA message to SP in JXTA pipe. SP parses the http message from JXTA message and sends web service page to foreigner in a reverse direction. The service page may

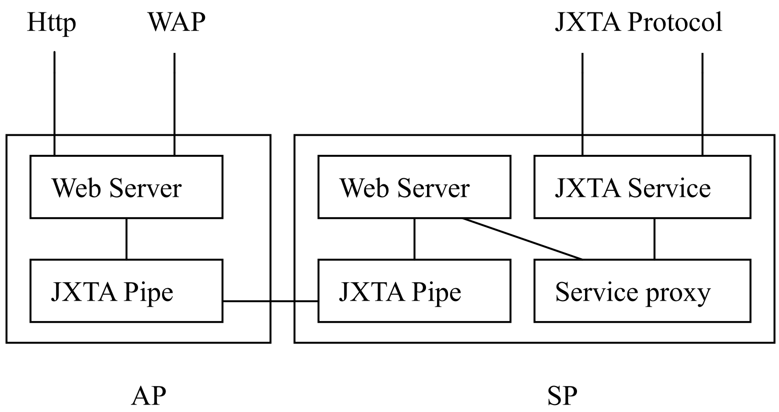

Figure 4. Skeleton of Web approach

contain JXTA Service proxy (proxy may reside in Java component e.g. Applet). A proxy acts as broker between Web client and JXTA service: foreigner invokes the interfaces supplied by proxy to access shared resource, in turn the proxy calls the JXTA service.

4.3 Functionality of proxy

The design of service proxy should satisfy at least the following requirement: 1) The proxy provides interfaces that can be invoked through a Java component. Each proxy is corresponding to a JXTA service.2) message wrapping. Input information submitted by foreigner through web page is a series of Java objects (String, Integer etc). Proxy should serialize these Java objects in form of JXTA message, and vice versa. 3) Proxy should also be responsible for mediation between socket and JXTA pipe. Peers in JXTA network communicate with each other using pipes. Furthermore, most communication in P2P network is asynchronous due to relative long delay. The result from SP can be send back to foreigners in a *push-like* notification. Event subscribing and notifying module in WMNI supports this feature.

5. Application

To further illustrate our proposed WMNI, some potential applications that can be built on WMNI are discussed.

5.1 Collaborative activity

Information exchange and sharing is common in collaborative activities. One user interacts with another user to acquire some information. The user may extract some part of information item that he is interested in from message he received. After processing, some new information is added and forwarded to other collaborative participants. In this context, two kinds of interactions exist: interaction between users and interaction between user and information. It is easy to describe the interaction with RDF statements. The activity participant and information item are mapped to *resource* in RDF statement; interactions are mapped into *predicate*. Through the annotation of interaction, the information can be reused. We believe that WMNI can facilitate the development of such collaborative application. Firstly, activity participants communicate and interact with each other in a P2P manner. While in traditional C/S collaborative application, the server is responsible for session management and need to keep track the state of participant. So the server often becomes complicated and is not easy to be maintained. Secondly, annotating interaction can solve the evolvement and interoperability problems of collaborative application.

5.2 Metadata clustering and semantic routing

P2P network is a kind of application overlay. The peers are not aware of the physical network structure, so message routing in P2P network cannot apply the mechanism used by Internet. Routing strategies like flooding, random walk are not scalable. In [6] disadvantages of DHT are also discussed. Based on our metadata infrastructure, we can set up semantic routing table for message routing. Table 1 shows a sample schema level routing table.

Table 1. Sample schema level routing table

Schema	Destination Super Peer
Dublin Core	Super Peer A
Mp3	Super Peer B
Dublin Core	Super peer B
….	…

In WMNI, all resources are annotated according to RDF schemas. Schema is a collection of domain vocabularies. Initially the peers are divided in small groups (the network diameter of group is several hops). The routing table is held in a *super peer*. The notion of *ordinary peer* and *super peer* is corresponding to host and router in Internet respectively. First the metadata information in a group is clustered to super peer in different level (schema level, property level etc.), then the super peer exchange with each other to setting up the semantic routing table. In three situations the routing table need to be updated: peer join, peer departure and metadata information modification. The event management module in WMNI supports "event registry and notification " and can fulfill this requirement.

6. Related work

In [20] Xin implemented a metadata search layer as an enhancement of JXTA content manager service. They used Dublin core Schema to annotate shared contents. Instead of storing the metadata in separately RDF storage as we have proposed, they stored annotation together with resource advertisement. Edutella [8] aims to build an metadata infrastructure to connect peers in P2P network based on exchange of RDF metadata. It concentrates on providing a common data model and Edutella query exchange language.

7. Conclusion

In this paper, we introduce the use of RDF metadata in P2P network to link distributed resources in a semantic way. To make capability-constraint devices access P2P network in a convenient way, we propose to present the resources with Web approach. The proposed WMNI provides an underlying metadata infrastructure that support the development of efficient P2P application. WMNI also enables capability-constraint devices(PDA, cell phone) to consume the resources without installing the P2P application in advance. Although WMNI is based on JXTA, but it is not limited to JXTA. To get concrete understanding the advantage of WMNI, two applications are discussed. 1)collaborative application: we use RDF statements to annotate the interaction between activities participant. 2)metadata clustering and semantic routing: Peers exchange metadata with each other to built up semantic routing table for resource query.

References

[1] Napster LLC. http://www.napster.com

[2] The Gnutella Project , http://www.gnutella.com

[3] Freenet Home Page, http://freenet.sourceforge.com/.

[4] S. Ratnasamy, R. Francis, M.Handley, R. Krap, J. Padye, and S.Shenker. A scalable content-addressable network. In ACM SIGCOMM, 2001.

[5] I. Stocia, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In ACM SIGCOMM, 2001.
[6] Zhou J., Vijay Dialani, De Roure D. and Hall W, A Distance Based Semantic Search Algorithm for Peer-to-Peer Open Hypermedia Systems, Proceedings of the Fourth International Conference on Parallel and Distributed Computing Applications and Technologies, Pages:7-11 Aug. 2003
[7] Resource Description Framework, W3C consortium http://www.w3.org/TR/1999/PR-rdf-syntax-19990105
[8] The JXTA Project ,www.jxta.org
[9] Jena: A Semantic Web Framework for Java. http://jena.sourceforge.net/index.html
[10] Redland RDF Application Framework . http://www.redland.opensource.ac.uk/
[11] Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1", http://dublincore.org/usage/terms/dc/current-elements/, October 2002.
[12] IEEE Learning Technology Standard Committee Working Group 12, "Final 1484.12.1 LOM draft standard", http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf, September 2002
[13] Alexander L¨oser1, Wolf Siberski, MartinWolpers, and Wolfgang Nejdl. Information Integration in Schema-based Peer-To-Peer Networks. Lecture Notes in Computer Science, Springer-Verlag Heidelberg, Volume 2681 / 2003, 258 - 272 ,January 2003
[14] G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis and Michel Scholl. RQL, A Declarative Query Language for RDF. The Eleventh International World Wide Web Conference (WWW'02), Honolulu, Hawaii, USA, May 7-11, 2002
[15] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjorn Naeve, Mikael Nilsson, Mathias Palmer and Tore Risch. EDUTELLA: A P2P Networking Infrastructure Based on RDF. WWW2002, May 7–11, 2002, Honolulu, Hawaii, USA.ACM 1581134495/ 02/0005.
[16] JXTA SOAP Project , http://soap.jxta.org/servlets/ProjectHome
[17] Changtao Qu and wolfgang Nejdl. Interacting the Edutella/JXTA Peer-to-Peer Network with Web Services. International Symposium on Applications and the Internet, IEEE,Proceedings, 2004
[18] jxta-rmi , http://jxta-rmi.jxta.org/servlets/ProjectHome
[19] Shoji Kimura,Koji Sakai, Akira Kurokawa, and Hiroshi Sunaga. Implementation and Evaluation of the P2P Web System, The 9th Asia-Pacific Conference on Communications, APCC 2003 , IEEE ,Volume: 1 ,21-24
[20] Xin Xiang, Yuanchun Shi and Ling Guo. Rich Metadata Searches Using the JXTA Content Manager Service. http://media.cs.tsinghua.edu.cn/~xiangx/paper/cms.pdf