Meta Service Discovery

John Buford
Panasonic Digital Networking Lab
Princeton, NJ, USA
buford@research.panasonic.com

Alan Brown
University of Stirling
Stirling, Scotland, FK9 4LA
abr@cs.stir.ac.uk

Mario Kolberg
University of Stirling
Stirling, Scotland, FK9 4LA
mko@cs.stir.ac.uk

Abstract

Meta service discovery is used to find and select a service discovery mechanism by context. As multiple service discovery mechanisms (SDM) proliferate across various administrative domains, mobile devices will require a way to locate and select the appropriate mechanism according to the context of the mobile device, such as network domain, location, protocol, and application. We define Meta Service Discovery and explain the motivation for it. We describe our results in building a Meta Service Discovery capability on three existing DHTs and integrated into a new broadcast-oriented SDM. Finally, we analyze the sizing and distribution of DHT entries, including hash distribution of SDM entries according to a geographic population-density scheme.

1. Introduction

Service discovery and advertisement is fundamental to service interoperability in pervasive computing, and many service discovery protocols have been developed [1][2][3].

Many portable devices are now equipped with TCP/IP networking function and applications on these devices need to describe, advertise, discover, and invoke services on the network. However there are numerous incompatible service description, advertisement, discovery, notification, and invocation mechanisms.

Further, there are several proposals for creating wide-area service overlays using structured peer-to-peer architectures. Due to the relatively low barrier of entry in creating and deploying peer-to-peer systems, it seems likely that multiple wide area peer-to-peer service overlays will emerge with different service discovery and advertisement methods.

In addition, the offering of services is frequently partitioned by administrative domain for AAA (authentication, authorization, accounting) purposes. Before a peer can discover a service, it must first determine 1) which administrative domain to operate in, and 2) which service discovery mechanism(s) (SDM) are used by that domain. A mobile device may be interested in discovering service domains by attributes such as location, operator, the number of available services, type of application domain, or other attributes of service discovery. We refer to the discovery of a service discovery mechanism as Meta Service Discovery.

2. Motivation and Related Work

There are many service discovery mechanisms (SDM). These include SDM as network services (such as SLP), SDM integrated into device network adapters (such as Bluetooth), SDM integrated in middleware platforms (such as Web services, Jini, and JXTA):

- Bluetooth Service Discovery Protocol (SDP) [4]
- Service Location Protocol (SLP) [5]
- Simple Service Discovery (SSDP) in UPnP [6]
- Web services and UDDI [7][8][9]
- Salutation [10]
- Jini [11]
- JXTA [12]
- Wide-area extensions to SLP: WASRV [13], wide-area SLP [14]
- Secure Service Discovery Service (SSDS) [15]

Further there are various proposals for creating service overlays using structured peer-to-peer systems [16][17][18].

We argue that multiple SDMs, both per domain and in the public Internet, will co-exist in the future for several reasons:

1. The low barrier of entry to create both application specific (e.g., file-sharing) and general purpose (e.g., Open DHT) peer-to-peer overlays is leading to the proliferation of P2P overlays, and service-oriented overlays will follow this trend.

2. Service discovery in pervasive computing is an extension of web search, and there will be economic incentives for multiple SDMs just as there are multiple web search sites today.
3. Application-specific SDMs will be designed and managed for the benefit of specific user communities, somewhat analogous to the multitude of web portals today.
4. Administrative control is argued in [19] as a basis for hierarchical and multi-ring overlay architectures. The same arguments apply to the use of multiple SDMs.
5. Experience has shown that standardization does not guarantee a single solution in the market place, the large number of incompatible instant messaging and presence protocols are one example of this.
6. From the various efforts to develop complex query languages and semantic lookup over DHTs, the tradeoff between query language power and efficiency and overhead could lead to a spectrum of service overlays with query languages matched to the needs of the application domain.

Consequently, a method is needed to identify and select service discovery mechanisms. For example, a mobile node may roam to a new location. Applications on the node wish to discover services available in the new context. The node must first determine which service discovery mechanisms are available to use in this context. The node uses meta service discovery to identify SDMs by context. In this step, it may obtain a client stub to download and use for accessing the SDM. Then it communicates directly with each SDM of interest, issuing discovery requests and obtaining service descriptions.

3. Description

Conceptually, a service discovery mechanism is a special type of service used to locate other services. However since existing SDMs are closed--that is, fixed with respect to protocol, advertisement and service description--using an SDM to find other SDMs corresponding to different protocols and formats has not been done. First, the number of services is generally far more than the number of SDMs, so SDMs have been designed with the assumption that a single SDM would be sufficient. Second, few SDMs have addressed wide-area, multi-domain service discovery. Now with the growing popularity of wide-area peer-to-peer applications, these assumptions can be re-considered as described in the previous section. Third, existing SDMs do not issue service advertisements for discovery by other SDMs.

Treating SDM discovery as a type of discovery problem, we can follow the approach of creating a searchable description and a method for SDMs. We are describe both a wide-area Meta Service Discovery method and a broadcast oriented method.

Figure 1 shows a document describing a SDM. Each tag contains an attribute of the SDM, such as domain, protocol, and location. Meta Service Discovery then becomes searching, advertising, and querying collections of such descriptions. In the wide-area approach, SDM descriptions are stored in a structured overlay. In the broadcast-oriented approach, SDM descriptions are broadcasted to other nodes on the network, such as a PAN or 802.11.

```
<SDM>
  <Domain>example.com</Domain>
  <Protocol>SLP</Protocol>
  <Location>New York</Location>
  <Operator>T. Smith</Operator>
  <Address>slp://services.example.com</Address>
  <Stub>http://services.example.com/slp.jar</Stub>
</SDM>
```

Figure 1 SDM description

An SDM description might contain other information about the service discovery protocol such as version number, protocol dependencies, or security considerations. Note that no changes are needed to existing service discovery protocols and that publishing an SDM description can be an administrative task.

3.1 Meta Service Discovery in Structured Overlay

In this scenario, SDM descriptions (Figure 1) are stored in a DHT implemented using a structured overlay. There are several techniques [17][20] for storing structured documents in a DHT so that the queries can be made against fields of the document. These techniques involve partitioning a structured document in to strands [17] and breaking up strings into substrings [20] which can be separately indexed. As discussed later, this leads to multiple index entries per SDM description.

Figure 2 Meta service discovery global index entries and lookup by various types, for discovery of various SDMs illustrates the meta service discovery on a structured overlay using a distributed hash table (DHT). For simplicity, we show the wide-area DHT as a single ring, but it could also be a hierarchical or multi-ring overlay. Entries in the DHT are depicted at the bottom of the diagram, and represent various SDM repositories, servers, and specialized service overlays.

Each SDM repository, server or service overlay is indexed by its SDM description. Queries for SDM entries may be by any combination of fields in the SDM description, such as domain, location, size, and operator.

Since SDM descriptions are significantly less complex than most service descriptions (for example, those found in WSDL, SLP, or UPnP), the mapping of SDM descriptions to the DHT hash indexing is straightforward, and queries are simpler than if full service descriptions were directly hashed in to the DHT index. We describe some example mappings that we have implemented in three existing structured overlays in section 4.

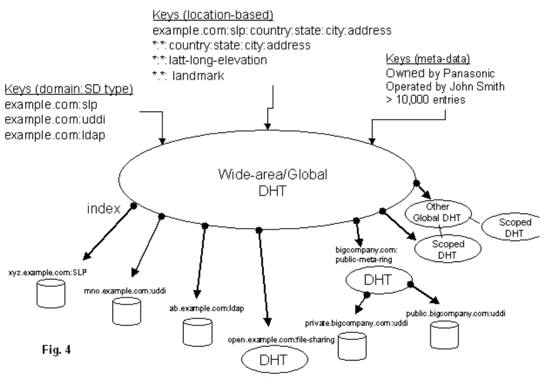

Figure 2 Meta service discovery global index entries and lookup by various types, for discovery of various SDMs

3.2 Meta Service Discovery in a Broadcast Service Discovery Protocol

We have separately developed a new service advertisement method called Data Dissemination Service Discovery (DDSD) [21]. In DDSD, a broadcast node broadcasts a sequence of advertisements, which represent the services of nodes connected to the broadcast network. Any node in the broadcast network can be the DDSD broadcaster, and the broadcast role may change from time to time. The selection of the DDSD broadcaster can be based on node resources, availability, and uptime. An important and novel property of DDSD is power conservation, in that nodes may go into power standby state while the broadcast node continues to advertise their services. Another node uses the broadcast to find a service of interest. If the service offering node is in power standby state, the node may signal the device to wakeup prior to invoking the service.

Figure 3 shows DDSD broadcasts containing various advertisements. In (a) the broadcast contains a sequence of advertisements collected from various nodes, an index for rapidly locating advertisements of interest, and timestamps to indicate how recent a given advertisement is. In (b) the broadcast contains advertisement groups, which is used to enforce group-based access control of service advertisements and services. In (c) the broadcast contains advertisements for other service discovery mechanisms, labeled SDM1 and SDM2. It also contains multiple instances of service1 advertisement, encoded according to SDM1 and SDM2 formats.

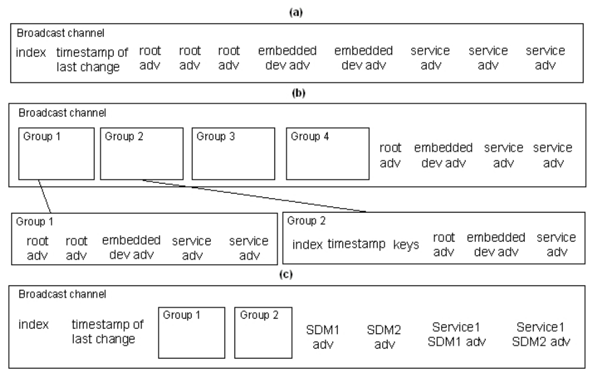

Figure 3 In DDSD, meta indices are included in the broadcast (a) sequence of advertisements, (b) groups of advertisements, (c) SDM advertisements and service advertisements with multiple encodings.

4. Mapping SDM Descriptions to Existing P2P Overlays

We created a large dataset of SDMs and used three available structured overlay indexing systems (FreePastry-an implementation of Pastry [22], OpenDHT [23], and INS/Twine [17] which uses Chord) to evaluate the indexing of SDMs in existing structured overlays. For FreePastry and OpenDHT we added an interface to decompose an SDM description into combinations of fields so that a wildcard lookup could be made by any combination of [domain, SDM protocol, location, and a free text field]. In these two systems, each SDM resulted in 15 index entries. INS/Twine provides an interface to map structured documents into index entries into Chord. For INS/Twine we encoded the SDMs in our dataset into the format used by INS/Twine.

For FreePastry, we used the FreePastry file storage mechanism and ran 220 virtual nodes and stored 20 SDMs, leading to a total of 300 entries. We then issued 300 queries to match the 300 entries, all of which returned a result. Due to a synchronization problem with the FreePastry implementation, we did not test larger data sets. One limitation of FreePastry is that duplicate keys are overwritten rather than appended. Since the SDM is decomposed into separate index entries for each field for wildcard purposes, collisions on protocol name, for example, were common.

At the time of writing, OpenDHT is deployed on about 200 nodes on PlanetLab. We inserted 72,019 SDMs into PlanetLab the network, which led to 1,072,014 total entries, for which we issued 1,008,140 queries. Unlike FreePastry, OpenDHT does permit multiple duplicate keys, and a lookup returns all matching keys. During the querying we received 7,334 errors (0.73%). OpenDHT errors are attributed to node or network conditions. OpenDHT successfully handled the SDM mapping and queries needed for these experiments.

For INS/Twine we ran 200 nodes and stored 200 SDMs leading to 2200 total entries. We then issued 2200 queries to validate the data entries. INS/Twine successfully handled the SDM mapping and queries needed for these experiments.

Table 1 Sizing estimates of number of SDM entries in wide-area index

SDM Attribute	Values	Sizing
Domain name	RFC 1034 syntax	DNS Resource Records type A: ~70M
Type	Protocol identifier	< 100
Location	Lat-Long Street Address Landmark	Number of grid points in top 40 cities world wide is 3.5M, based on sq area

5. Sizing and Key Distribution Analysis

Next we estimate the number of SDM entries needed if Meta Service Discovery were widely deployed. We assume that each domain that has type A DNS Resource Record advertises at least one SDM. The most recent measurement is 71,723,098 type A DNS entries [24].

Each SDM might also be referenced by location and we use the following approach to index locations. In general, a location could be according to street address, landmark, or latitude-longitude (LL). Street address and landmarks can be converted to a standard format and directly hashed, or they can be converted to a corresponding latitude-longitude. LL can be normalized to decimal format and aligned to the nearest grid point. The resulting grid point can be directly indexed. The grid alignment approach considerably simplifies lookup. For estimating sizing and measuring key distribution, we assumed that SDM locations would correspond to population densities. We obtained two data sets, one containing the LL position for the 2555 largest cities and another containing the square area for the 40 largest cities [28]. Both are worldwide datasets. For the 40 largest cities, assuming a grid spacing of 1 city block (about 200 meters), there are about 3.5M grid points in the largest 40 cities.

These estimates are summarized in Table 1.

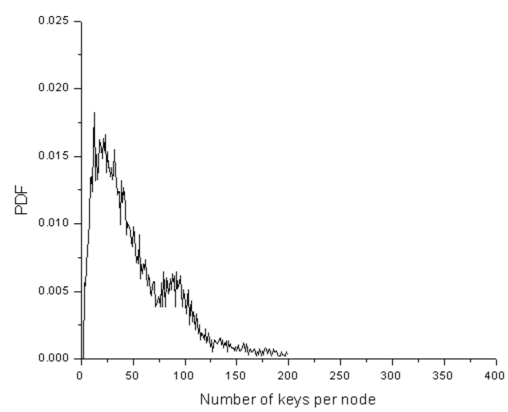

Figure 4 PDF of key distribution of 405,097 real domain names in a 10,000 node network

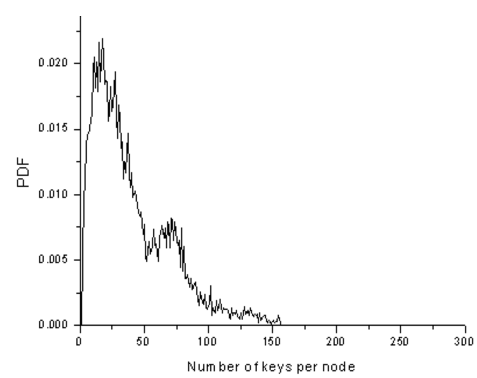

Figure 5 PDF for location based key distribution of 500,000 keys over 10,000 nodes

5.1 Key Distribution

As described previously, each SDM is decomposed into separate fields and indexed into combinations of fields. We evaluated the key distribution of the domain name and location entities separately.

For the distribution of domains, we obtained a list of 405,097 domains registered in 1997 from [29]. After hashing each domain using SHA1, we inserted the keys into a network consisting of 10,000 nodes. Figure 4 shows the Probability Density Function (PDF) of the key distribution using real domain names. These results are similar to those measured in [25] and demonstrate that by indexing domain names we incur no penalty compared to randomly generating keys.

For location, we used a normalized LL format as described previously. We randomly generated positions in 2555 cities on a 200 meter grid spacing. We then hashed these positions using SHA1. Figure 5 shows the PDF for 0.5 million location entries over 10,000 nodes. Figure 6 shows the 1st and 99th percentile and mean keys per node for a range of keys. These results, like those collated for the distribution of domain names, are at least as uniform as those measured in [25] for randomly generated keys.

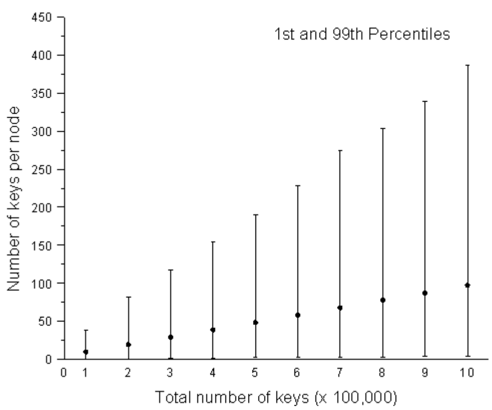

Figure 6 Number of keys per node vs total numbers of keys for 10,000 nodes

6. Conclusion

Meta service discovery is used to find and select a service discovery mechanism by context. As multiple service discovery mechanisms (SDM) proliferate across various administrative domains, mobile devices will require a way to locate and select the appropriate mechanism according to the context of the mobile device, such as network domain, location, protocol, and application.

In this paper, we introduced Meta Service Discovery and described mechanisms for achieving Meta Service Discovery in both wide area structured overlay and broadcast network contexts. We used three structured overlay-based index systems (FreePastry, OpenDHT, INS/Twine) to index an SDM data set and issued wild card queries in each system.

We estimated the sizing of the key set if indexing of SDM were widely deployed. We further analyzed the key distribution of SDM index entries, focusing on location and domain attributes. Assuming SHA1 hash function, we obtained key distribution results that are comparable to measurements from hashing random strings.

7. References

[1] F. Zhu, M. Mutka, L. Ni. Classification of Service Discovery in Pervasive Computing Environments. MSU-CSE-02-24, Michigan State University, EastLansing, 2002.

[2] C. Lee and S. Helal. Protocols for Service Discovery in Dynamic and Mobile Networks. Intl. J. of Computer Research. v. 11 n 1 pp. 1-12. 2002.

[3] Bettstetter, C. and C. Renner, A Comparison of Service Discovery Protocols and Implementation of the Service Location Protocol, Proc. EUNICE 2000, September 2000.

[4] Bluetooth Specification version 2. Vol. 3 Core System Package. Part B Service Discovery Protocol (SDP).

[5] E. Guttman, C. Perkins, J. Veizades, M. Day. Service Location Protocol, Version 2. IETF RFC 2608. June 1999.

[6] UPnP Forum. UPnP Device Architecture. Version 1. June 8, 2000.

[7] K. Gottschalk, S. Graham, H. Kreger, J. Snell. Introduction to Web Service Architecture. IBM Systems J. v 41 n 2. 2002.

[8] W3C. Web Services Architecture. http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/. Feb 11, 2004.

[9] Microsoft. Web Services Dynamic Discovery (WS-Discovery). Oct 2004.

[10] Salutation Architecture Specification (part 1) Version 2.0c. www.salutation.org

[11] Sun Microsystems. Jini Architecture Specification v 2. June 2003.

[12] Bernard Traversat, Ahkil Arora, Mohamed Abdelaziz, Mike Duigou, Carl Haywood, Jean-Christophe Hugly, Eric Pouyoul, Bill Yeager. Project JXTA 2.0 Super-Peer Virtual Network. www.jxta.org

[13] J.Rosenberg, H.Schulzrinne, B.Suter. Wide Area Network Service Location. IETF draft-ietf-svrloc-wasrv-01.txt. Work in progress. Nov 14, 1997.

[14] An-Cheng Huang and Peter Steenkiste. A Flexible Architecture for Wide-Area Service Discovery,. The Third IEEE Conference on Open Architectures and Network Programming (OPENARCH 2000), March 26-27, 2000.

[15] Czerwinski, S. E., Zhao, B. Y., Hodes, T. D., Joseph, A. D., and Katz, R. H. 1999. An architecture for a secure service discovery service. In *Proceedings of the 5th Annual ACM/IEEE international Conference on Mobile Computing and Networking* (Seattle, Washington, United States, August 15 - 19, 1999). MobiCom '99. ACM Press, New York, NY, 24-35

[16] Gu, X., Nahrstedt, K., and Yu, B. 2004. SpiderNet: An Integrated Peer-to-Peer Service Composition Framework. In *Proceedings of the 13th IEEE international Symposium on High Performance Distributed Computing (Hpdc'04) - Volume 00* (June 04 - 06, 2004). HPDC. IEEE Computer Society, Washington, DC, 110-119.

[17] M. Balazinska, H. Balakrishnan, D. Karger. INS/Twine: A Scalable Peer-to-Peer Architecture for Intentional Resource Discovery. Pervasic Computing 2002.

[18] C. Schmidt and M. Parashar, *A Peer-to-Peer Approach to Web Service Discovery*, World Wide Web Journal, Vol. 7, Issue 2, June 2004

[19] A. Mislove, P. Druschel. Providing Administrative Control and Autonomy in Structured Peer-to-Peer Overlays. IPTS 2004.

[20] M. Harren, J. Hellerstein, R. Huebsch, B. T. Loo, S. Shenker, I. Stoica. Complex Queries in DHT-based Peer-to-Peer Networks. IPTPS 02.

[21] E. Celebri, J. Buford. Service Discovery Using Data Dissemination in Broadcast Networks. Submitted.

[22] A. Rowstron and P. Druschel, Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), Heidelberg, Germany, pages 329-350, November, 2001.

[23] Rhea, S., Godfrey, B., Karp, B., Kubiatowicz, J., Ratnasamy, S., Shenker, S., Stoica, I., and Yu, H. 2005. OpenDHT: a public DHT service and its uses. In *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications* (Philadelphia, Pennsylvania, USA, August 22 - 26, 2005).

[24] Netcraft (26th September 2005) http://news.netcraft.com/archives/web_server_survey.html

[25] Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., and Balakrishnan, H. 2003. Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Trans. Netw.* 11, 1 (Feb. 2003), 17-32.

[26] A. Gulbrandsen, P. Vixie, L. Esibov. A DNS RR for specifying the location of services (DNS SRV). IETF RFC 2782. Nov. 2000.

[27] R. Cox, A. Muthitacharoen, and R. T. Morris, Serving DNS using a Peer-to-Peer Lookup Service, in IPTPS, Mar. 2002.

[28] Grid-Arendal, Online GIS and Maps Database (29th September 2005) http://www.grida.no/db/gis/prod/html/glohuma1.htm

[29] Internet Systems Consortium (29th September 2005), http://www.isc.org/index.pl?/ops/ds/