

Semantic Driven Hashing (SDH): An Ontology-based Search Scheme for the Semantic Aware Network (SA Net)

Chatree Sangpachatanaruk[†]

Taieb Znati^{*,†}

[†]Department of Information Science and Telecommunications ^{*}Department of Computer Science
University of Pittsburgh, Pittsburgh, PA, 15260

Abstract

The success of personalized resource discovery depends on its ability to allow users to discover, extract and integrate information of interest from heterogeneous sources, and its ability to provide these users with efficient tools to manipulate and convert the discovered information into knowledge. We propose Semantic Aware Network (SA Net), a structured peer-to-peer (P2P) overlay architecture to support the basic functionalities of personalized resource discovery. The SA Net uses an ontology-based representation of the resources to enable a semantic resource discovery and access that reflects the interest of the user. In this paper, we describe the the SA Net search scheme “Semantic-driven Hashing” (SDH), which uses lexical-based ontology to provide a foundation for indexing and searching in structured P2P overlay infrastructure.

1. Introduction

Structured P2P overlay networks have gained attention as a viable solution to effectively facilitate resource discovery and sharing in large scale networks. The basic tenet of this technology is a distributed hash table (DHT), which is used to infer a structure that regulates location and access to a distributed set of resources. Using the DHT information, overlay nodes that currently manage the resource can be identified deterministically and in a finite number of steps, thereby reducing considerably the searching overhead due to communication and routing. The properties of structured overlay networks make them a viable solution to support the design of a resource discovery for and enable an effective management of a distributed resource indexes of interests to different communities. The basic DHT-based overlay structure, however, exhibits several shortcomings that need to be addressed in order for such a structure to meet the performance requirements of personalized resource discovery.

One critical limitation of DHT-based structure, with respect to resource discovery, is its lack of support for location-based queries that go beyond perfect matching. Resource distribution in a structured P2P overlay

network is based on a hash table, which is used for looking up a specific item based on a single key. This key is mapped to only one location in the overlay network. As a result, each resource is uniquely handled by a specific node in the network. Resource indexing, on the other hand, requires a flexible scheme to handle multiple-key mapping. The scheme must support resource distribution, whereby resources are spread among overlay nodes based on their multiple semantic keys.

To address the limitation of DHT-based structures and meet the design requirement of multiple key-mapping resource discovery scheme, we propose a novel Semantic Aware Network architecture, referred to as SA Net, for personalized Internet resource discovery. Within this structure, we also propose an ontology-based search scheme for a DHT overlay architecture, referred to as “Semantic Driven Hashing (SDH)”. In the following section, the concept of the SDH, its components, and main algorithms, will be described.

2. SA Net Search Scheme

SA Net is an agent-based system which achieves its semantic richness through the use of explicit ontologies to represent resources. SA Net further enhances the DHT-based resource distribution scheme by using the unique identifier assigned to each ontology as a key to locate the overlay node responsible for maintaining the resource index associated with the underlying ontology. In other words, the ontology-based hashing scheme, thereof referred to as “*Semantic-Driven Hashing (SDH)*”, utilizes ontologies, instead of resource names, as the hash input to generate the key necessary to distribute the resource among overlay nodes.

To further explain the SDH-based resource discovery scheme, we consider an ontology that describes the concept of “*telecommunication*”, as depicted in Figure 1. Based on the above ontology, a given Internet user can express personal interest in communication by using the keyword “*telecommunication*”. When this keyword is issued as part of a search, the resulting query returns information of all

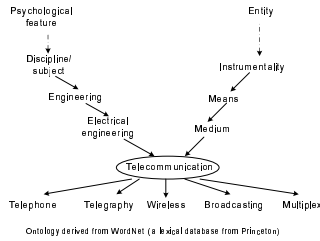


Figure 1. Ontology example.

resources which are directly related to the associated concept, namely *telephone*, *telegraphy*, *wireless*, *broadcasting*, and *multiplex*. When used in conjunction with the keyword of “*wireless*”, however, the ontology can be scoped down to the resources that are associated with the new concepts (i.e. telecommunication that is wireless), thereby capturing more specifically the user’s interests. In addition, when semantic relations are stored with the concepts among the overlay nodes, networks of resources for a user can be proactively formed to reflect the user interest.

To ensure global consistency in semantic classification, however, SA Net must support a globally defined ontology map which represents the well-known concepts and which can be used to guide an SDH-based search query. As such, given a keyword, SA Net can consistently identify which concept it is related to. To achieve this goal, SA Net uses Wordnet lexical network, referred to as *Ontology Semantic Network (OSN)*, where a node represents a concept as described by a set of similar words, and where an edge between two nodes represents the relation between these concepts[1]. Each node in OSN is assigned a numerical ID that is used as the hash input for an SDH-based search.

When a query to locate or advertise a resource is issued, the SDH scheme consults the OSN to obtain the ID of the concept associated with the resource specified in the query. This ID is then hashed to obtain the key in the DHT space, of the overlay node where the metadata resource index is maintained. This key is then used to route the request to the identified overlay node. The pseudo code of the SDH algorithm is shown in Algorithm 1.

Algorithm 1 SDH algorithm: Retrieve DHT keys for keyword K .

```

SDH(Keyword  $K$ ){
   $OSN\_id[] \leftarrow Ontology\_discovery(K)$ 
  {retrieve ontology IDs from the keyword using OSN}
   $OSN\_keys[][]$ 
  for  $i = 0$  to  $OSN\_id.length$  do
    {obtain DHT key for each OSN ID}
    for  $j = 0$  to  $NUMBER\_REPLICAS$  do
      {one concept key has several replicas for fault-tolerance}
       $OSN\_keys[i][j] \leftarrow Hash(OSN[i] + j)$ 
    end for
  end for
  return  $OSN\_keys$ ;

```

3. Related Works

The recent trend to improve resource discovery for the structured P2P networks is to enable semantic-based data location and search by associating semantics into the structure to regulate resource distribution. Recent frameworks propose to use information retrieval techniques, namely vector space model (VSM), and latent semantic indexing (LSI), to obtain semantic representation from the textual information appearing in the resource and queries [4]. Another approach uses XML techniques to generate semantic representation from metadata attributes [2]. The most similar work to the SDH proposed scheme is the approach that uses a concept of global catalog to provide a standard for semantic classification [3]. While the first approach does not work well with non-textual resources, the second approach is likely to provide only perfect-matched result for a search, because of the indexing and searches that depend on keyword attributes. The catalog approach, even though, provides a better scheme for semantic-based resource classification, its naming still depends on the keywords. SA Net, on the other hand, uses global classification, but contrary to the catalog-based approach, the classification does not depend on word appearances. Instead, it relies on the ontologies that capture the keywords’ meanings and semantic relations, which can be used to infer related resources more intuitively.

4. Conclusion

This paper presents the search scheme of SA Net, a semantic aware network to enable personalized resource discovery on the Internet. The SA Net search scheme enhances DHT search by using ontology, a semantic representation, as the hash input instead of filename or simple keywords. The meanings and relations derived from ontologies will allow personalization to evolve effectively. As such, a personalized view of the resources can reflect user’s interest in a transparent manner. Several parts of the SA Net are currently being developed. These include personalization built upon this search scheme and mechanisms to provide security and fault tolerance for the SA Net architecture.

References

- [1] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, May 1998.
- [2] L. Garces-Erice, P. Felber, E. Biersack, K. Ross, and G. Urvoy-Keller. Data indexing in dht peer-to-peer network. In *Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS-04)*, March 2004.
- [3] A. Loser. Towards taxonomy-based routing in p2p networks. In *Proceedings of Workshop on Semantics in Peer-to-Peer and Grid Computing*, pages 35–50, Newyork, USA, May 2004.
- [4] C. Tang, Z. Xu, and S. Dwarkadas. Peer-to-peer information retrieval using self-organizing semantic overlay networks. In *ACM SIGCOMM 2003*, pages 175–186, August 2003.