

The Bio-Networking Architecture
Bi-weekly report #10 (October 14, 2002): Distributed Discovery
PI: Tatsuya Suda (suda@ics.uci.edu)
University of California, Irvine
<http://netresearch.ics.uci.edu/bionet/>

Introduction

In the Bio-Networking Architecture, a network application is implemented as a decentralized collection of autonomous objects called *cyber-entities*. One of the challenges with such a distributed framework is discovery of cyber-entities. Discovery provides the ability for applications to locate specific cyber-entities that may represent available services, information, or users.

In the bi-weekly report submitted on September 16, 2002, the PI evaluated dynamic properties of a discovery algorithm developed with support from DARPA. In this report, the PI focuses on further evaluation of the dynamics of the discovery algorithm described in the report submitted on September 16, 2002. In the discovery algorithm that the PI focuses in this report, each cyber-entity contains a set of keywords, and discovery involves locating cyber-entities that match some particular keywords. Cyber-entities also contain a limited set of relationships to other cyber-entities (i.e., links that include information about other cyber-entities). These relationships between cyber-entities together form a network on which discovery queries are forwarded. The discovery algorithm in this report forwards discovery requests multiple hops and returns discovery results along the same path that the discovery request traversed.

Dynamics in the environment the PI considers includes cyber-entity availability, failures of discovery request processing, changes in discovery demand and changes in content provided by the cyber-entities. In this report, the PI focuses on the behavior of the discovery algorithm relative to dynamic discovery demand and also compares an algorithm variation that uses relationships in a probabilistic manner. Discovery demand is characterized by the rate at which requests are issued and the distribution of keywords used in those requests. The distribution of keywords in discovery demand may be biased towards a subset of keywords available within the network, and over time, user demand may change to a different subset of keywords. In this report we evaluate the discovery algorithm with respect to the rate at which the distribution of keywords of discovery demand changes over time and we compare the probabilistic variation to the discovery algorithm described in the report submitted on September 16, 2002.

Discovery Algorithm

The discovery algorithm in this report addresses the dynamic discovery demand through maintaining a history value for each relationship. Relationship history summarizes information about how a relationship partner has performed in past discoveries and is defined as the ratio of successful discovery requests on a relationship relative to all the discovery queries forwarded on that relationship. As more discovery requests are processed, relationship history is updated, providing a method to optimize relationships towards the current discovery demand. Discovery demand may be biased towards a subset of keywords available within the network, and based on this biased discovery demand, history enhances the performance of the network by adapting relationships to better support frequently searched-for keywords. As discovery demand changes, history is continuously updated, causing relationships to reflect the changes in discovery demand, better supporting the new discovery demand.

In addition, this report includes evaluation of a variation of the discovery algorithm described in the bi-weekly report submitted on September 16, 2002. The discovery algorithm in this report applies a probabilistic approach to selecting which relationship to use for forwarding. The probability of selecting relationship A for discovery forwarding = $SIM_A * HST_A$. SIM_A is the normalized similarity value of relationship A over all relationships and HST_A is the normalized history value of relationship A over relationships with similarity equal to SIM_A .

This probabilistic approach to discovery may be beneficial towards handling dynamic discovery demand. The discovery algorithm described in the bi-weekly report submitted on September 16, 2002 presents a greedy approach to discovery forwarding: forward to the most similar relationship first and break ties by strongest history. A greedy approach may consistently reinforce a particular relationship to locate certain cyber-entities, and is unlikely to use other less-reinforced relationships for forwarding. A probabilistic approach for discovery request forwarding probabilistically uses both strongly reinforced and weakly reinforced relationships for locating cyber-entities. By using both types of relationships for discovery, the probabilistic approach compares multiple relationships for discovery and potentially identifies a better relationship for discovery in dynamic environments. Also, the greedy approach may fail to compare relationships that arise from the other dynamic considerations not discussed in this report (such as newly introduced cyber-entities), whereas, the probabilistic approach may compare relationships and identify the better performing or reliable relationships. However, there is overhead associated with using weakly reinforced relationships, and thus, probabilistic forwarding may perform worse depending on how frequently alternate relationships are used and the quality of those alternate relationships.

Simulation Results

Discovery demand may be biased towards particular keywords within the network. In this simulation, the PI compares the effects of biased discovery demand and how continuously changing discovery demand impacts discovery performance. For the simulation results presented below, discoveries are biased towards 30 particular keywords in the network (out of 3000 keywords that exist in the network). Figure 1 shows the probability of keywords being in a request. As shown, 30 keywords have much higher probability than all others. To generate a discovery request, one of the 30 keywords is selected, and the other 3-4 keywords of the request are chosen such that the request will match some cyber-entity in the network.

Discovery requests are issued every interval, and also, at every interval the discovery demand, with probability α , slightly changes which keyword set discovery requests are biased towards. The discovery algorithm is evaluated with $\alpha = 0$ (no dynamics), 0.0002, 0.002, 0.02, and 1. (Note for $\alpha = 1$, the keyword set completely changes).

Figure 2 compares search times for three methods to performing discovery: forwarding using similarity and no history (“Sim only”), forwarding using both similarity and history (“Sim & Hist”), and the probabilistic approach to forwarding described above (“Prob Sim & Hist”). From Figure 2, it appears that the approach of “Sim & Hist” performs

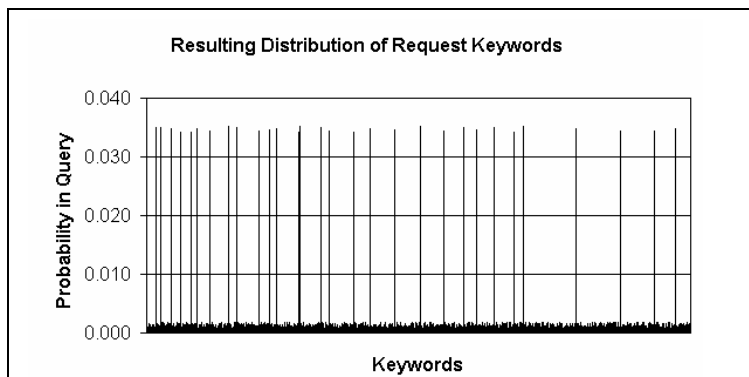


Figure 1

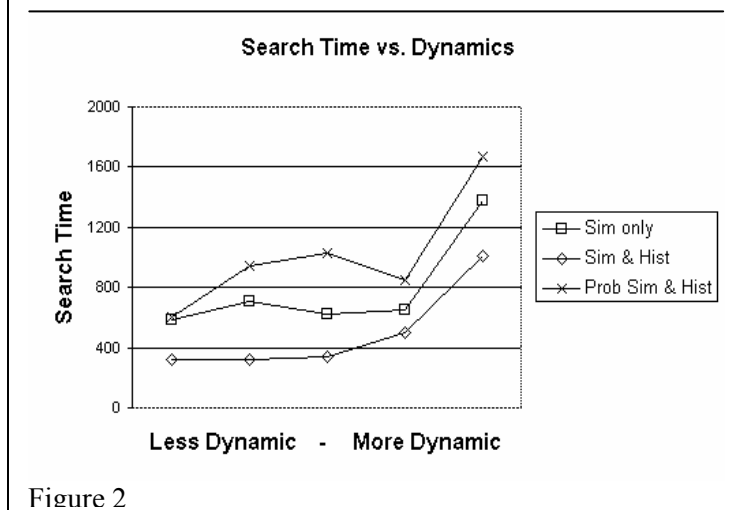


Figure 2

best when there is little or no dynamics in discovery demand and that “Sim & Hist” performs worse as the discovery demand becomes more dynamic. The probabilistic approach to forwarding does not appear to show its benefits in these simulations. However, the probabilistic approach is expected to be robust to dynamic environments, and the PI intends to continue simulation study to identify areas/conditions under which the probabilistic approach provides good performance.