Implementation and Evaluation of Autonomous Collaborative Discovery of Neighboring Networks

Tao Zhang, Sunil Madhani, Ashutosh Dutta, Eric Van den Berg

Telcordia Technologies, Piscataway, NJ 08854, USA

Yoshihiro Ohba, Kenichi Taniuchi

Toshiba America Research Inc., NJ 08854, USA

Shantidev Mohanty

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract **Emerging mobile devices often want to discover neighboring networks in real time. For example, a mobile may want to know the capabilities of the surrounding WiFi networks so that it can intelligently determine which one to use next and avoid wasting scarce battery power to attempt to connect the wrong networks. Knowing the address of the authentication server in a neighboring network may allow a mobile device to authenticate with the neighboring network while the device is still using the current network to transport its user traffic, even before the device detects the radio signal from the neighboring network. The Internet Engineering Task Force (IETF) is developing a Candidate Access Router Discovery (CARD) protocol for a mobile device to discover the address and capabilities of access routers in a neighboring network. However, it requires all neighboring access routers to implement CARD and to dynamically exchange network information with each other and with the mobiles, which is difficult to implement especially when neighboring networks belong to different network providers. Furthermore, CARD can only allow a mobile to obtain the information a local network provider configures its CARD protocol to provide, making it difficult to provide consistent information from network to network and to support mobile devices that may need different information depending on their different networking capabilities and user applications. This paper proposes and evaluates a new approach for real time discovery of neighboring network information: Autonomous Collaborative Discovery (ACD). It is *autonomous* because regular mobile users and their devices act autonomously to collect information and make the information available to other mobile users and devices. It is *collaborative* as the autonomous actions of the mobile users and devices help each other to discover the information they want.**

I. Introduction

Emerging mobile devices (e.g., mobile phones, Personal Digital Assistants, notebook computers, communication devices in vehicles) need to discover a growing range of information about neighboring networks, i.e., *network neighborhood discovery*. Here, a neighboring network is a nearby network that a user can move into next from its current network; its radio coverage area may or may not overlap with the radio coverage area of the mobile device's current network. For example, if a mobile device knows the addresses of the authentication server and the IP address server in a neighboring network, it can authenticate with, and acquire an IP address from, this neighboring network while it is still using its current network even before detecting the radio signal from the neighboring network.

To further illustrate the needs for new network neighborhood discovery capabilities, we consider a user with a WiFi phone walking down a street that is covered by overlapping WiFi networks as shown in Figure 1. As she walks into areas covered by multiple WiFi networks, her WiFi phone seeks to connect to these networks. With today's methods, the mobile phone typically attempts to connect to the network with the best signal strength or signal to noise ratio. However, it may then fail to authenticate with this network if the user is not allowed to use the network. The mobile phone could lose network connection for a potentially long period of time while it attempts to find the right network. Furthermore, today's methods allow the phone to authenticate with a target network only when it is inside radio coverage area of the target network. Given the small sizes of WiFi networks, this may give the mobile phone little time to attempt the network connections when the overlapping areas of the networks are small and the user is moving fast such as when the user is in a moving vehicle. As illustrated in Figure 1, suppose that the mobile user is allowed to use network *A* but not networks *B* and *C*. Using today's methods, her mobile phone will not know whether it will fail to connect to, or authenticate with, networks *B* and *C* until after it attempts to connect to these networks.

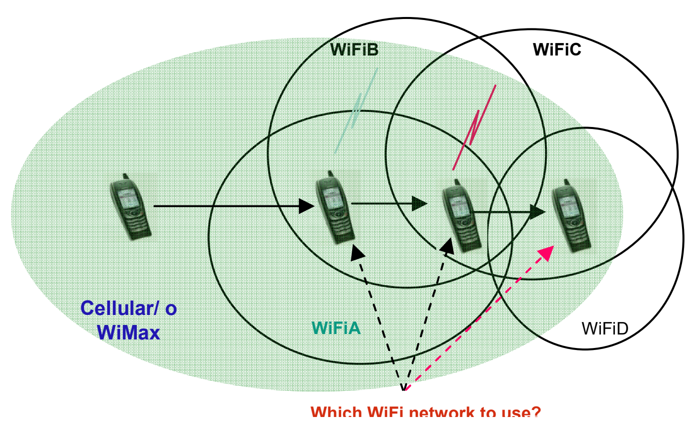

Figure 1: Benefit of network neighborhood discovery.

The above problem can be addressed if

- The mobile phone can discover which neighboring networks the user is allowed to use before it even enters the radio coverage areas of these networks. This allows the phone to significantly narrow down the choices of the networks.
- The mobile phone can discover the addresses of the authentication servers in the neighboring networks, and then use its current network connection to authenticate with a target neighboring network, *before* it enters the radio coverage area of the target networks. This will give the mobile phone significantly more time to find the right networks and connect to them.

The Internet Engineering Task Force (IETF) has been developing a Candidate Access Router Discovery (CARD) protocol [4] that allows user devices to discover the IP address and capabilities of access routers in neighboring networks. CARD, however, has the following limitations:

- It allows a mobile to discover the access router of a neighboring network only after the mobile detects radio signal from the neighboring network.
- It requires neighboring access routers to use the CARD protocol to dynamically exchange network information, which is difficult when neighboring networks belong to different network providers.
- It requires all access routers to implement the CARD protocol to communicate with mobile users, which is a difficult proposition.
- A mobile device can only obtain the information a local network provider configures its CARD protocol to provide. Thus, the set of available information can vary significantly from network to network and hence is unsuitable to support mobile devices that may need different information depending on their different networking capabilities and user applications.

Several frameworks exist for discovering network services and devices and they are often referred to as service discovery approaches [1][2][3][5]. Here, network services include shared network devices such as printers, displays, and disks; software utilities; and information such as database and files. These service discovery mechanisms, however, do not support discovery of information regarding neighboring networks.

We advocate that future approaches for information discovery should allow mobile users and devices to

- Participate in the creation of information and share the information they accumulate with each other.
- Discover dynamic, location-based information, and time-sensitive information.
- Discover information they want, not only what network or service providers think they may want.
- Discover information in and about neighboring networks.
- Discover information in a timely manner to support fast-moving mobiles such as vehicles.

II. PROPOSED APPROACH

This paper describes a new approach for real-time discovery of neighboring networks and their capabilities: Autonomous Collaborative Discovery (ACD), which is based on the approach in [7].

ACD is *autonomous* in the sense that mobile devices act autonomously to collect information they want and make the information available to other mobile devices. It is *collaborative* as the autonomous actions made by the mobile devices help each other to discover the information they want.

Section II.A describes the concept, architecture, and operations of ACD. Section II.B describes a sample schema and its implementation for the storing and retrieval of network information.

A. Concept, Architecture, and Operation of CS-CDS

We use the network configuration shown in Figure 2 to illustrate the principles and operations of ACD. The networks shown in Figure 2 can be any type of wireless or wired networks including wireless Local Area Networks (WLANs) and cellular systems, and may belong to different providers.

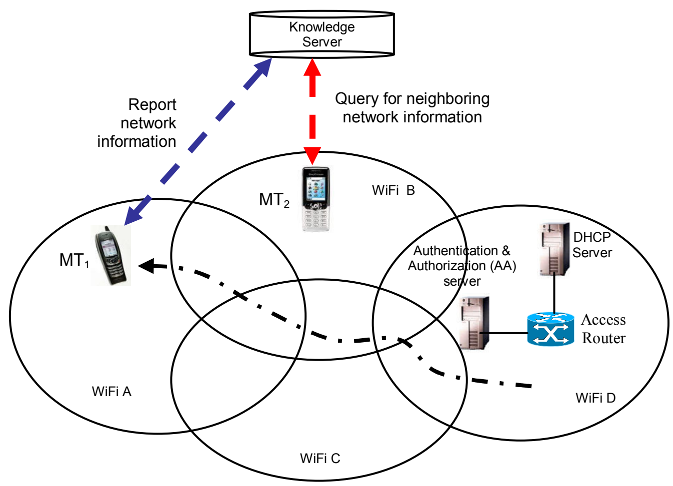

Figure 2: Autonomous Collaborative Discovery of network information.

With ACD, regular mobile devices act as *scouts* to collect information about the networks they visit during their routing use of the networks, and make the information they accumulate available to other mobile devices. In essence, all mobile devices may participate in the creation and sharing of knowledge about the networks.

A mobile device can collect a broad range of information about each network it detects or uses, without any special assistance from the network providers, as described in [7].

Today's mobile devices typically keep the information they accumulate to themselves ("selfish") and discard the information after they no longer need it ("memory-less"). With ACD, *the information a mobile device accumulates will be made available to other authorized users and devices*.

To share the information, each mobile device reports its collected information to a functional entity referred to as *Knowledge Server*. Other approaches to information sharing may also be used. For example, a peer-to-peer information discovery and sharing method was recently proposed [6].

The knowledge server maintains all the information reported by the mobile devices that subscribe to its services, and makes the information available to all subscriber mobiles. When a mobile device needs to discover any information, it queries the knowledge server. For example, as shown in Figure 2, the first mobile, MT_1, traveled through networks D*, C, B, and A* and served as a scout to discover the existence of these networks, some network information (e.g., addresses of the access routers, IP address servers, and authentication servers), and the relationship among the networks (e.g., networks *A, C, and D* are neighbors of network B). It reports such information to the knowledge server. When mobile MT_2 later moves into network *B*, it may query the knowledge server to find the IP address of the authentication server in each neighboring network.

The knowledge server can be distributed or centralized. They can be operated by a Knowledge Provider that is independent of network providers, as illustrated in Figure 2. The knowledge server only receives, and hence only needs to maintain, information that its subscribers are interested in. It only needs to maintain information regarding the locations and the networks its subscribers traveled.

Each mobile user or device does not need to know what information other users or devices want. Instead, it collects the information it wants for itself. For example, when mobile MT_1 in Figure 2 traveled through networks *D* and observed a traffic accident, it will report the accident and its location to the knowledge server. Then, when mobile MT_2 later enters network *B or C,* it can find out, by querying the knowledge server, not only the IP address of the authentication server in network *D* but also about the traffic accident.

Information collected by all the subscriber mobiles of the knowledge server will collectively satisfy the needs of all these subscriber mobiles even when the information collected by each mobile will never be useful to any other mobile. For example, mobile MT_1 in Figure 2 has visited networks *A, B*, *C*, and *D* and reported the information it collected about these networks to the knowledge server. When it returns to any of these networks later, it will be able to find the information regarding the neighboring networks by querying the knowledge server.

Multiple mobiles may report the same information repeatedly if no effort is taken to reduce the unnecessary reporting. On the other hand, the mobiles need to update the knowledge server frequently enough to keep the dynamic information stored on the knowledge server up to date. To balance the need for keeping information up to date and the need for reducing unnecessary reporting of duplicated information, we use a method described in [7].

To make ACD independent of specific networking technologies and network providers, it uses an application–layer standard protocol between the mobile and knowledge server for the mobile to report information to, and retrieve information from, the knowledge server. ACD does not impose any special requirements on such an application-layer protocol. We are using the Simple Object Access Protocol (SOAP) [9] in our testbed.

B. Database Schema and Implementation

The information on the knowledge server should be stored in a standard and easy to access manner. We are using RDF (Resource Description Framework) [6] based schema to store the information elements along with its characteristics in the mobility server. RDF-based query and response mechanism using XML provides the ability for a mobile to query specific information element by given characteristics of the information elements in a granular manner. For example, network information elements could be Access Point, DHCP server or AA server. The characteristics of these network information elements could be SSID, location-info (geo-coordinate), layer-2 (L2) security information.

Figure 3 shows a partial view of the RDF-based tree that we have built for the network elements in the neighborhood networks and their dependency.

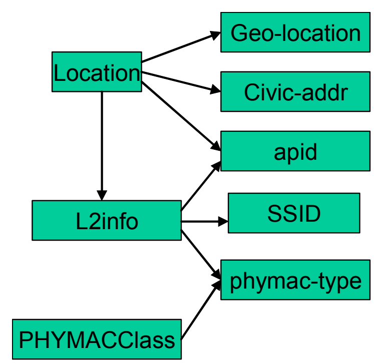

Figure 3: Partial graphical view of a sample schema.

The following is the schema representing the above graphical view:

```
<rdf:RDF xml:lang="en"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:nd="http://www.networkdiscovery.org/2004/10/rdf-schema/">
<rdfs:Class rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/Location">
 <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/geo-location">
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/Location"/>
 <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
<rdf:Property rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/civic-addr">
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/Location"/>
 <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
<rdfs:Class rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/L2info">
 <rdfs:subClassOf
rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/Location"/>
```

```
</rdfs:Class>
<rdf:Property rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/apid">
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/L2info"/>
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/Location"/>
 <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
<rdf:Property rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/ssid">
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/L2info"/>
 <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
</rdf:Property>
<rdf:Property rdf:about="http://www.networkdiscovery.org/2004/10/rdf-schema/phymac-type">
 <rdfs:domain rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/L2info"/>
 <rdfs:range rdf:resource="http://www.networkdiscovery.org/2004/10/rdf-schema/PHYMACClass"/>
</rdf:Property>
... (SNIPPED...)
```


III. EVALUATIONS

A critical question is how long the *initial learning period* (i.e., the time it takes for scouts to discover all the networks of interest) will be and how the number of scouts impacts the length of this initial learning period? This section describes an alnalytical model and uses it to answer these questions.

We consider a scenario where the network a scout moves into next depends on where the scout is before the move. For example, a scout always moves to one of the closest networks or will move out of the region of interest. The case where the network the scout moves into next is independent of the netowrks the scout has been is given in [7].

We first consider a single scout in our analysis and then extend the analysis to multiple scouts. We assume that the time the scout spends inside each network is a random variable and is independent of which network it moves into next. If the average time a scout spends in each network is the same and is denoted by μ, the initial learning period will be the μ times the average number of movements the scouts made before discovering all the networks. Here, we say a scout made a movement each time it moves out of a network. In the rest of this section, we focus on evaluating the number of movements a scout needs to made before discovering the networks of interest.

We denote by $S=\{1,2,3,\ldots N\}$ the set of the networks to be discovered, denote by T_n^N the number of networks discovered by n movements, and denote by T_N the number of movements it takes to discover all N networks..

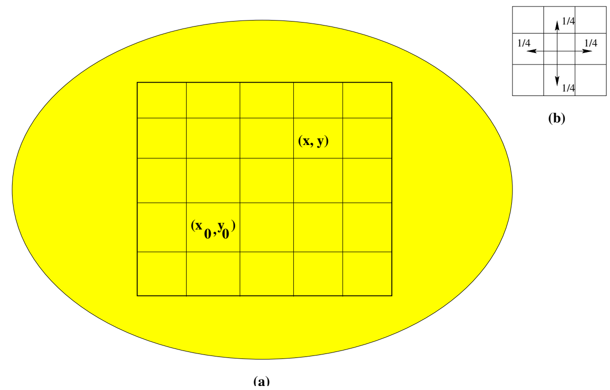

Figure 4: Networks of a particular geographic region.

For analysis purpose only, we assume that the N networks form a geographical grid as shown in Figure 4 (a).

Each time, a scout moves out of a visited network, it goes into one of the four neighboring networks with equal probability as shown in Figure 4(b). Then, T_n^N is given by

$$T_n^N = \sum_{i=1}^{N} 1(i_n) \quad (1)$$

where $1(i_n)$ is the identity function and i_n is the event that network i is discovered by n movements. The identity function is given by

$$1(x) = \begin{cases} 1, & if\ \ x\ is\ true \\ 0, & if\ \ x\ is\ \ false \end{cases}$$

The expected number of networks discovered in n movements is therefore given by

$$E[T_n^N] = \sum_{i=1}^{N} E[1(i_n)] = \sum_{i=1}^{N} p_n(i) \quad (2)$$

where $p_n(i)$ is the probability that the network i is discovered in the first n movements. To calculate the value of $p_n(i)$, we consider that the scout starts from the network located at *(x_0, y_0)* in the grid. Let $\Delta x_i=x-x_0$, *and* $\Delta y_i=y-y_0$. be, respectively, the differences between the x and y co-ordinates of the starting network and network i.. Now the probability that the scout will discover network i during its n[th] movement is given by

$$p_{(\Delta x_i,\Delta y_i)}(n) = \alpha_{(\Delta x_i,\Delta y_i)}(n) - \sum_{j=0}^{n-1} p_{(\Delta x_i,\Delta y_i)}(j)\,\alpha_{(0,0)}(n-j) \quad (3)$$

where $\alpha_{(\Delta x_i,\Delta y_i)}(j)$ is the probability that the scout is in network i during its j^{th} movements The second term on the right hand side of (3) is the probability that the scout visited the network i during its earlier visit and again during the n^{th} movement. The expression for $\alpha_{(\Delta x_i,\Delta y_i)}(j)$ is given by :

$$a_{(\Delta x_i, \Delta y_i)}(j) = \frac{f(j, \Delta x_i, \Delta y_i)}{4^j}$$

where $f(j, \Delta x_i, \Delta y_i)$ is the number of possible paths from the starting netowrk to network i when exactly j movements are performed by the scout. For a two-dimensional grid as shown in Figure 4(a), a scout can move in four different directions: LEFT, RIGHT, UP, and DOWN. We assume r ($0 \leq r \leq j$) to be the number of RIGHT movements performed by the scout. Given that the total number of movements performed by the scout is j, the number of LEFT, UP, and DOWN movements by l, u, and d, respectively, and their expressions are given by:

$$l = r + \Delta x_i$$

$$u = \frac{1}{2}(j - 2r - \Delta x_i - \Delta y_i)$$

$$d = \frac{1}{2}(j - 2r - \Delta x_i + \Delta y_i)$$

Using the above expressions, $f(j, \Delta x_i, \Delta y_i)$ is given by: $f(j, \Delta x_i, \Delta y_i) = \sum_{r=0}^{j} g(r, \Delta x_i, \Delta y_i)$

where $g(r, \Delta x_i, \Delta y_i) = \frac{j!}{r!l!u!d!}$.

Now using (3), the probability of discovering the network i when the scout starts starts from (x_0, y_0) within n movements is given by

$$p_n(i) = \sum_{k=0}^{n} p_{(\Delta x_i, \Delta y_i)}(k) \qquad (4)$$

Using (4), $E[T_n^N]$ can be calculated numerically, for relatively small N.

The problem of determining the number of sites visited by a two-dimensional random walk has received considerable attention in the mathematics and theoretical physics community [10]. Using generating functions, many results for T_n^N can be derived. The average of T_N for large N is:

$$E[T_N] \approx N \log^2 N \qquad (5)$$

IV. PERFORMANCE EVALUATION

We analyze the network discovery process for a given grid of networks as shown in Figure 4 (a). First we analyze the network discovery process using one scout. Then we carry out the simulation studies for network discovery when multiple scouts are present.

A. Probability of nework discovered vs. time

Figure 4 shows that a scout starts to move from a randomly selected network in a grid of 25 networks. The x and y axes show the grid of networks where networks are identified by locations on the grid and the z axis shows the probability that each network is discovered. Different networks are identified by their locations on the grid. Initially, only the network where the scout started is discovered. Hence, it has a 100% probability of discovery, and all other networks have zero probability of discovery initially.

Figure 5 shows the discovery probabilities of the networks after 150 movements performed by the scouts, which shows that as time passes the networks that are farther from the starting locations of the scouts start to be discovered. Eventually all the networks have probability of discovery as 1, i.e., the all are discovered.

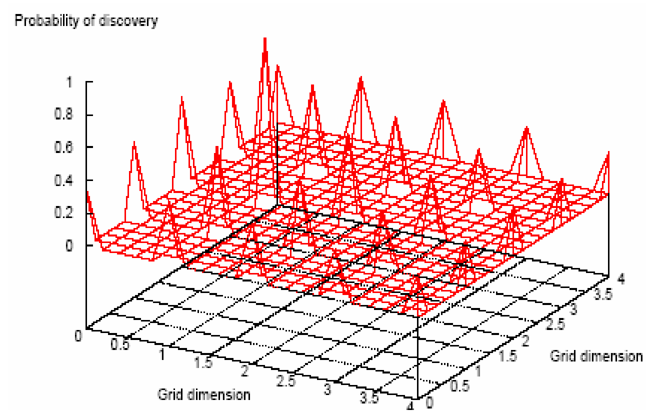

Figure 5: Network discovery after 150 movements.

B. Number of networks discovered vs. time

Figure 6 shows the number of networks discovered as time passes using both analytical and simulation studies by one scout. It shows that the rate of discovery of networks decreases as the time increases. This is because initially the probability that the scouts will enter a new network upon a movement is higher as there are more undiscovered networks. However, as more networks are discovered, the probability that a scout will enter an undiscovered network decreases. Therefore, the rate of network discovery decreases as time passes. Figure 6 also shows that the number of networks discovered *vs.* time match closely for both analytical model and simulation studies.

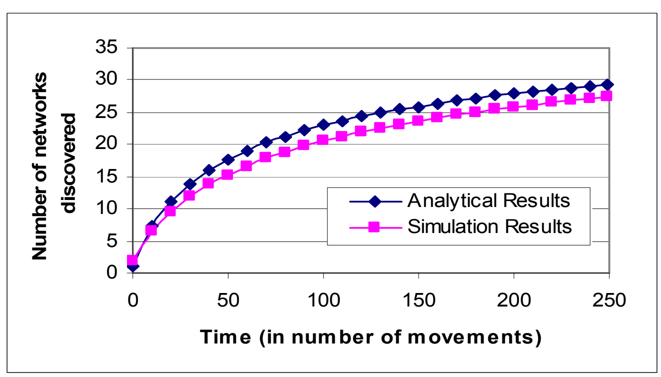

Figure 6: Comparison of analytical and simulation results.

C. *Effects of the number of scouts on network discovery*

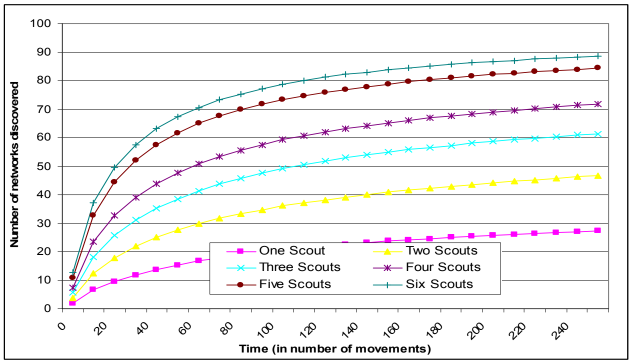

Figure 7: Effect of number of scouts on network discovery.

We carried out simulations using different number of scouts to discover the networks in a particular geographic region of inerest, such as a city, as shown in Figure 4 (a). The simulations assumed the same mobility pattern as the one used in our analytical models.

First, we consider the case where, when a scout moves out of a network, it may leave the region of interest. We refer to this scenario as the *unbounded scenario*. The results on the number of networks discovered *vs*. the time it takes to discover these networks for the unbounded scenario is shown in Figure 7. It shows that when the number of scouts increases, the number of networks discovered within a given amount of time increases rapidly.

Next, we consider the case in which the scouts never leave the region of interest. This implies that when a scout reaches the boundary of the geographic region, it comes back into the region. This is typical of the mobile users that stay in one particular city most of the time. We refer to this scenario as the *bounded scenario* because the scouts are bounded to the geographical region of interest.

Figure 8 shows the number of networks discovered *vs.* time for different number of scouts in the bounded scenario. It shows that when the number of scouts increases, the number of networks discovered within a given amount of time increases rapidly.

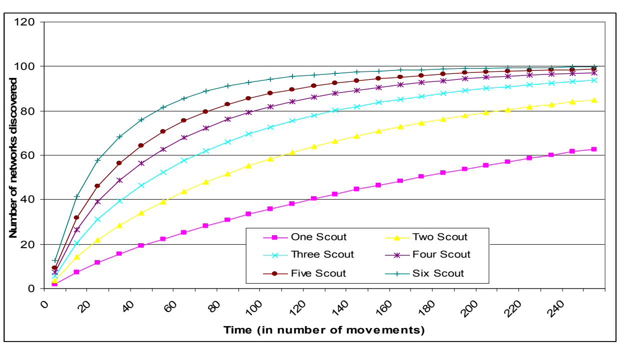

Figure 8: Effect of number of scouts on network discovery.

A comparison of Figure 7 and Figure 8 shows that the number of networks discovered in a given amount of time is less for the unbounded scenarios compared to the bounded scenarios. In the unbounded scenarios, the scouts spend part of their time in empty areas, thereby not contributing to network discovery.

V. CONCLUSIONS

This paper proposes a new framework that allows mobile devices to discover information about neighboring networks in real time. It shows how inform collected by mobile devices may be implemented using RDF-based schemas. It also provides analytical models and simulation results to evaluate the time it takes to collect information about networks in a geographical region and show that that a small number of scouts are typically sufficient to discover a large number of networks quickly.

VI. ACKNOWLEDGMENT

Authors would like to thank Henning Schulzrinne from Columbia University for many meaningful discussions related to network and service discovery.

REFERENCES

[1] Sun Microsystems Inc., http://www.sun.com/jini/

[2] Microsoft Corporation, "Universal Plug and Play Device Architecture", http://www.upnp.org/

[3] Salutation Consortium, "Salutation Architecture Specification", June 1, 1999, http://www.salutation.org/

[4] Marco Liebsch and Ajoy Singh (editors), "Candidate Access Router Discovery", Internet Draft draft-ietf-seamoby-card-protocol-06.txt, Internet Engineering Task Force, June 2004.

[5] K. Arabshian, H. Schulzrinne, "GloServ: Global Service Discovery Architecture", *Mobiquitous 2004*, Boston, USA.

[6] Tao Zhang, Eric van den Berg, Sunil Madhani, "Peer-to-Peer Information Discovery and Sharing among Mobile Users and Devices", accepted to appear in the *61st IEEE Vehicular Technology Conference (VTC), 2005.*

[7] Tao Zhang, Sunil Madhani, Eric van den Berg, Yoshihiro Ohba, "AC-CDS Autonomous Collaborative Collection, Discovery, and Sharing of User and Network Information", *The 3rd International Conference on Information Technology: Research and Education* (ITRE 2005).

[8] Resource Description Framework, http://www.w3.org/TR/rdf-primer

[9] SOAP, Simple Object Access Protocol, http://www.w3.org/TR/soap12

[10] F. van Wijland, S. Caser, H.J. Hilhorst, "Topology of the support of the two-dimensional random walk", *Journal of Physics*, A, Vol. 30, .p.507, 1997.