

A Scalable Geographic Service Provision Framework for Mobile Ad Hoc Networks

Xiaojing Xiang
State University of New York at Buffalo
Buffalo, NY, USA
xxiang@cse.buffalo.edu

Xin Wang
State University of New York at Stony Brook
Stony Brook, NY, USA
xwang@ece.sunysb.edu

Abstract

Supporting scalable and efficient routing and service provision in Mobile Ad Hoc Networks (MANET) has been a big research challenge. Conventional topology-based unicast and multicast protocols are normally hard to scale due to the big overhead in their routing schemes. Supported by these routing protocols, conventional service discovery schemes also have limited scalability and efficiency. Basing their forwarding decisions only on the local topology, geographic-based unicast routing protocols have drawn a lot of attentions in recent years. However, current geographic unicast routing can not adapt to different traffic conditions in a service provision network and current geographic multicast protocols can hardly scale to a large network size and group size. We propose a geographic routing and service provision framework for MANET which possesses the features of scalability, efficiency, robustness and adaptability. In the framework, an efficient hierarchical structure is built and maintained, based on which a scalable membership management is deployed to efficiently track the resource and service states. An adaptive and reactive geographic unicast routing protocol and a scalable and robust multicast protocol are designed to meet the different routing requirements in a service provision network. With the support of all of these components, the service provision functions, such as service discovery, delivery and coordination, can be deployed in the framework.

1 Introduction

In recent years, we observe a proliferation of mobile wireless devices, which are widely used in transportation, military, security, health, education and so on, and play more and more important role in our daily lives and business. Besides the traditional resources such as storage, processor power and software, wireless devices are often equipped with GPS receivers, wireless interfaces (Bluetooth, Wi-Fi), camera, sensors, etc. The wide variety of new and convenient wireless resources motivate us to develop techniques to better exploit new service capabilities. However, service provision is still in an infancy stage in Mobile Ad Hoc Networks (MANET) and far behind that in the wired networks. The progress is mainly

hindered by the higher complexity and more constraints of MANET. Due to the dynamic topology, unreliable wireless connections, and the resource limitation in wireless devices, it is very challenging to develop reliable and efficient routing and service provision schemes in MANET.

Conventional approaches to service discovery ([6][26][23][27][8]) are mostly intended for wired networks and not suited to MANET, as wired networks have more stable infrastructures and hence support more reliable communications. Some schemes have also been proposed for MANET. The middleware-based solutions in [10] and [20] may generate unnecessary overhead due to the redundant flooding involved in each service discovery and delivery. Hence more attempts have been made (e.g., [19][7]) to integrate the service discovery and MANET routing to reduce the unnecessary overhead. Instead of using distributed resource searching approaches as above, the schemes in [12] [22] are centralized directory-based. The literature work indicates that service provision is closely related to routing. However, current service provision protocols are normally built on or simply extended from the existing topology-based MANET routing protocols and consequently inherit their limitations, e.g., limited scalability, relatively high control overhead, and unreliability. In MANET routing, lately there is a trend to develop position-based routing schemes which are more robust and efficient than the traditional topology-based schemes. In this work, we make use of the position information in our service provision mechanism to increase its scalability and efficiency. Instead of using complicated schemes to manage network topologies, only the positions of service nodes or zones need to be tracked.

The goal of our research is to develop a scalable and efficient geographic service provision framework to harness the resources available from mobile devices and wired networks to achieve truly ubiquitous resource-sharing. Our framework is designed to be *cross-layer* by integrating the application-layer service provision and the network-layer routing. Particularly, service provision and routing will share the same infrastructure to track the membership, positions and service states of service nodes. Our cross-layer design will significantly reduce the management redundancy and make different layers to work more closely. Our service provision framework will support the following functions: 1) *Service discovery*: lo-

ating the services based on user requests. 2) *Service delivery*: delivering relevant service data and control messages. Unicast routing is needed for the delivery between peer devices, while multicast routing would be required to support efficient group communications. 3) *Service coordination*: a service request may need to be satisfied by several service providers, or a service may have several candidate providers. To enable coordination, the service provision framework should be able to track a group of service providers and their services. Hence, efficient and scalable service and membership management is required to facilitate the selection of appropriate providers and the collaboration among multiple providers. All the above functions are very important for a complete service provision framework, while current work normally focuses on only a subset of the problems (e.g., service discovery and/or service delivery), without fully considering or efficiently supporting the other functions. Instead, our work is aimed to support all these functions. Our framework consists of the following components:

1. **A self-configuring, distributed and hierarchical structure**, which is formed through the self-organization of service nodes. The hierarchical structure can be constructed with a flexible number of layers, depending on the service nodes' capabilities and management requirements. The structure building process is efficient and scalable by making use of geographic information. With the support of the structure, we can implement scalable membership and service management as well as efficient service provision.
2. **Scalable membership management**, to manage the dynamic collection of service nodes and their services. Membership management is a key issue for service provision, based on which, many important functions such as service coordination, group communications, fault detection, service monitoring, economical model and security can be deployed. The membership management scheme must be able to handle the frequent joining and leaving of a possible large number of service nodes and their changing service states.
3. **Adaptive routing protocols**, to meet different routing requirements in service provision scenarios, specially to support both sporadic service control message transmissions and long-term service data deliveries.
4. **Service provision scheme**, to realize the service provision functions described above.

The design goals of our service provision mechanism are: 1) *Scalability*: the protocol should be scalable to large network terrain and large number of service nodes. 2) *Efficiency*: the protocol should have relatively low control overhead. 3) *Robustness*: the service provision should be robust to possible failures. 4) *Adaptability*: the framework should be able to adapt its behavior based on the network environments to maximize its performance. All of the above properties are critical.

In our framework, the scalable hierarchical structure and efficient membership management will facilitate the design of the scalable and efficient service discovery and coordination. And the proposed adaptive geographic unicast protocol and efficient multicast scheme will help provide robust and adaptive service delivery.

The rest of the paper is organized as follows. In Section 2, we discuss some related work on MANET service provision and geographic routing protocols. A detailed description of our service framework is given in Section 3. Section 4 presents the simulation studies on the framework. Section 5 concludes this paper.

2 Related Work

2.1 Service Provision in MANET

Service related protocols proposed in MANET are mainly for service discovery and these discovery schemes are normally integrated with different kinds of routing protocols including the reactive ([19] [7]), proactive ([14]), hybrid ([28][17]) unicast protocols, multicast protocol ([5]) and any-cast routing protocol ([18]). These approaches are based on conventional routing protocols, and hence inherit their limitations including limited scalability, high overhead and unreliability. Also it is hard to support service coordination through these distributed searching schemes in a dynamic network. In centralized directory-based schemes [12] [22], some mobile nodes hold the service directory to assist the communications between service providers and clients. Although service coordination is easier, such centralized management is hard to scale and the centralized directories lead to bottlenecks. In [21], local directories are constructed and form the backbone of the network. A request will be searched in a local directory or multiple directories by distributed searching. Its topology-based scheme is still hard to scale to a larger network (e.g., with several hundreds of nodes).

2.2 Geographic Routing in MANET

Geographic unicast routing [11][2] assume mobile nodes are aware of their own positions through certain positioning system (e.g., GPS), and a source can obtain the destination's position through some location service [13]. An intermediate node makes packet forwarding decisions based on its knowledge of the neighbors' positions and the destination's position inserted in the packet header by the source. As the forwarding decisions are only based on the local topology, geographic routing is more scalable and robust in a dynamic environment. However, these protocols adopt a proactive beaconing scheme to maintain the local topology no matter if there are data traffics. To reduce control overhead under light traffic, in [32] [9], neighbors of the forwarding node contend for packet forwarding according to their distances to the destination. These simple contention-based schemes may result in

redundant forwarding, and lead to large collision probability under high traffic load.

Group communications are important in service provision. With a one-to-many or many-to-many transmission pattern, multicast is an efficient method to realize group communications. The high dynamics of MANET, however, makes the design of a scalable, robust and efficient multicast protocol much more challenging due to the difficulty in group membership management, multicast packet forwarding and the maintenance of a tree- or mesh-based multicast structure. Similar to unicast routing, lately some geographic multicast schemes have been proposed to increase the routing performance in MANET, in which the protocols presented in [1], [4] and [16] need to put the information of all the group members into the packet header and can hardly scale to large group. And the scheme in [24] is hard to scale to large network size due to its periodic local-range and network-range membership flooding. To support the scalable service provision, we need a multicast routing protocol which is robust and scalable to both group size and network size.

3 The Cross-layer Framework

In this section, we first introduce the scenarios and notations used in the framework description, and then we present the four components of our framework.

3.1 Scenarios and Notations

The service framework we consider has two basic network structures: *stand-alone* and *hybrid*. In the stand-alone case, individual wireless devices connect with each other through infrastructure-free ad hoc networks. In a hybrid structure, wireless devices are connected through one or more gateways to the wired networks. The gateways can be static (e.g., on an access point or base station) or mobile (e.g., when located in a mobile device). Each mobile node is assumed to be aware of its own position.

We define the following notations which will be used in our presentation.

SP: Service provider.

SR: Service requestor.

SN: Service node. Besides *SP*, *SNs* also include service servers and clients. After the requested service has been supported, the provider and requestor become service server and client. *SNs* will be managed by our membership management mechanism.

LC, *RC*, *GC*: Local coordinator, regional coordinator, global coordinator. They will be introduced in the next subsection.

3.2 Hierarchical Service Provision Infrastructure

In this subsection, we present the construction and maintenance of our geographic-based hierarchical system architec-

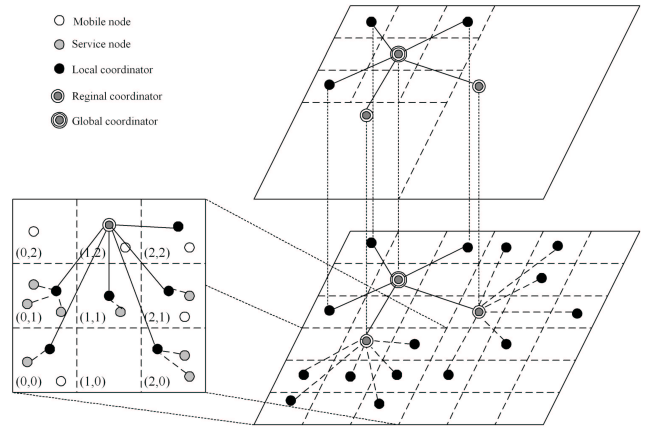


Figure 1. Overview of the hierarchical structure.

ture. Our system architecture follows a hybrid model which takes advantage of the strength of both centralized and distributed structures. The hierarchical structure is designed to be flexible. Specifically, the number of management layers in the structure is adjusted according to the capability of the coordinators, the density of service nodes and the service requirements. We define *service coordinator* at each layer to facilitate service management. According to the management range and layer, a coordinator can be classified as local, regional or global coordinator. The functions supported by a coordinator are also kind of services provided to other wireless nodes. An example of the hierarchical structure is shown in Fig. 1. The whole network is divided into size-manageable square zones. The zone-structure is virtual and constructed and maintained only on management need.

Based on the local zone structure, an upper layer is built in the network range and a *GC* is elected for global management. The *GC* only needs to track the aggregated information of those zones having *SNs* instead of every *SN*. Although a two-tier structure is enough for a normal MANET, in order to scale to a larger network or manage a larger number of *SNs*, more layers can be introduced between *GC* and *LCs*. Whenever the management load at *GC* is high, it can assign some *RCs* to manage regions consisting of multiple zones. A *RC* will aggregate regional service information and report to *GC* to mitigate *GC's* management load.

3.2.1 Virtual zone construction

As mentioned above, the network terrain is divided into virtual square zones. The length of a side of the zone square is defined as *zone size*. Each virtual zone has a zone ID (*zID*) to help identify and locate a zone. A node can calculate its *zID* (*a*, *b*) from its position coordinates (*x*, *y*) as: $a = \lfloor \frac{x-x_0}{zone_size} \rfloor$ and $b = \lfloor \frac{y-y_0}{zone_size} \rfloor$, where (*x*₀, *y*₀) is the position of the virtual origin. For simplicity, we assume all the zone IDs are positive. For a zone with *zID* (*a*,*b*), the position of its center (*x*_{*c*}, *y*_{*c*}) is: $x_c = x_0 + (a + 0.5) \times zone_size$, $y_c = y_0 + (b + 0.5) \times zone_size$. A packet destined to a zone

will be forwarded towards its center.

3.2.2 Adaptive service coordinator election

In this subsection, we introduce how different service coordinators are elected and maintained to efficiently address the service need.

a) Local service coordinator election

A LC is in charge of a local zone. Only when a zone has SN in it, a LC is elected and maintained to avoid unnecessary management overhead. When a SN just moves into a new zone, it queries its neighbors in the zone for LC. If failing to get LC information, it will announce itself as LC by flooding into the zone a LEADER message with its position. A LC needs to announce its role in its zone periodically at a time interval $Intval_{LC}$, and a node in the zone will cache the LC information for twice of $Intval_{LC}$. If not receiving any further announcement for this caching period, a SN will announce itself as LC after waiting for a random period. In case there exist multiple LCs, the one with the largest address wins. The SNs in a zone need to periodically refresh their positions and service states to LC as to be described in Section 3.3. Before a LC leaves the zone it is in charge, if there are no other SNs in the zone, it will notify GC instantly, or the zone's information will finally be timed out at GC without updating; otherwise, it will handover its leadership and service records to one of the SNs selected according to certain rule (e.g., the one closest to the zone center, the one with more energy left, or the one with lowest velocity).

b) Global service coordinator maintenance

A higher level service aggregation and management can be used to further facilitate service coordination and management. In a MANET, a GC can be elected on demand or pre-designated and located at some powerful node such as wireless access point or base station at network formation time. The GC will have a global view of the resource and service states of the whole network through periodic reporting from local service zones as described in Section 3.3.

When the GC is mobile, to track the location of the GC, the concept of *global coordinator zone* (GCZone) is used. When a node wants to start a GC service, it will announce its current zone as GCZone into the network. Each node in the network will cache the GCZone information. A new entering node can get GCZone information by querying its neighbors. The GC can also make periodic announcement to the whole network at a long interval (e.g., in minutes). When GC moves to a new zone, it needs to register its current zone to GCZone, and the registration message will be flooded in GCZone so that all the GCZone nodes know GC's current zone. GC will make its position known to other nodes in its zone through LEADER message. When a node has a message for GC and doesn't know its position, it will send the message to GCZone and the first node in GCZone receiving the message will forward it to GC's current zone.

Normally GCZone is kept unchanged except in three cases:

1) GC will announce its current zone as GCZone when its dis-

tance to GCZone is larger than D for route optimization, or when it cannot reach GCZone due to network partition or failures of GCZone nodes; 2) When GCZone is to be empty, the last moving out node will announce its entering zone as new GCZone; 3) GC can handover its role to another GC candidate, and the new GC can announce its zone as GCZone. For the consistency of GCZone information, each GCZone has a seqNo. The seqNo is increased by one each time GCZone changes. When multiple GCZones exist, the one with larger seqNo wins or the one with larger zID wins when having the same seqNo.

c) Regional service coordinator assignment

When the network size or the number of service zones increases and the load of GC is high, GC can distribute its responsibility by assigning some RCs in certain regions according to the service and management need. GC will then inform the service zones in a region about their corresponding RC, so that these zones will send their information to the RC instead of GC. A zone that is newly formed within a region and doesn't know RC will send report to GC, which will notify the zone the existence of RC. A RC will announce its liveness and position periodically to the zones it is in charge through multicasting. If a zone fails to receive such announcement within a predefined timeout period, it will try to contact the RC, and if this also fails, it will send its future reports to GC directly again. When leaving its region or not having enough resource to provide RC service, a RC may handover its role to another node in its region based on its own knowledge of the nodes' capability to provide RC services, or handover the management role of its region back to GC when no candidate is available. Similar to GC, to distribute its load, a RC can further sub-divide its region and introduce more service management layers.

3.3 Scalable Resource and Service Membership Management

As discussed earlier, to facilitate service provision, the SNs will join the service framework and their resource and service information will be managed through the hierarchical infrastructure. The membership management not only needs to track the service states and positions of SNs, but also needs to handle the frequent joining and leaving of SNs from the service management system.

In a local zone, a SN will periodically report to LC through a REGISTER message its position and service states (e.g., the available resources and the average service load) at a time interval $Intval_{local}$. And LC will remove a SN's information if not refreshed within $2 \times Intval_{local}$. $Intval_{local}$ can be decided by each SN adaptively (e.g., based on the SN's moving speed), in which case, the SN must inform its $Intval_{local}$ to LC during the reporting. A LC (or RC) will periodically report the aggregated information to its higher level service coordinator at a longer interval $Intval_{regional}$ (or $Intval_{global}$) or when the service state changes, and the reporting interval can be also adaptive. A service record will be removed by

a coordinator upon timeout or when requested by the corresponding SN or zone. For example, when a SN moves to a new zone, it will ask its old LC to remove its record. Otherwise, its record will be removed when timeout.

3.4 Efficient and Robust Routing

3.4.1 Reactive geographic unicasting

In a service provision scenario, besides long-term data transfers, many short-term connections may be needed to transmit control information, while no communication is needed for a long period when the providers are busy with certain services. Current geographic routing protocols either proactively distribute geographic topology information [11] [2] which will generate unnecessary control overhead when there is no traffic or use simple contention-based schemes to forward packets [32] [9] which result in high packet loss under heavy traffic load. Therefore, the literature routing protocols can not work efficiently and adapt to different traffic demands in a service provision network.

In the framework, we use our Self-Adaptive On Demand Geographic Routing Protocol with Geographic Reactive Mechanism (SOGR-GR) proposed in [30] to provide efficient and robust information delivery. To be brief, in SOGR-GR, the local geographic topology information is distributed adaptively based on the traffic demand and network environment. The one-hop neighbors' positions, are obtained through reactive beaconing mechanism. And to make the routing protocol more adaptive, the protocol parameters, e.g., the beaconing interval, are set by each node according to network environments, data traffic and its own requirements.

3.4.2 Scalable and robust geographic multicasting

Examples of group communications in our framework include the communications between the GC and RCs, RC and its managed zones, LC and its local SNs, and communications among multiple servers that collaborate to provide a requested service. Multicast is an efficient method to realize group communications. In our framework, based on the hierarchical membership management, service coordinators at different layers naturally maintain the information and positions of their managed zones or SNs. When a node intends to send a packet to a group of receivers (zones or nodes), the packet delivery follows the scheme in our Robust and Scalable Geographic Multicast Protocol (RSGM) proposed in [29]. To summarize, with the guidance of receivers' positions, the packets are forwarded along an efficient *virtual-tree* based structure, but there is no need to actively maintain a tree structure. This not only reduces maintenance overhead but also makes the multicast transmissions more robust to network dynamics.

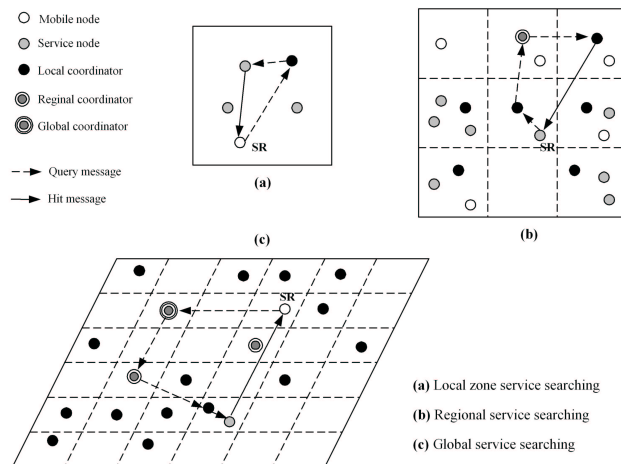


Figure 2. Service discoveries in different layers.

3.5 Service Provision

In our framework, the sending of control and data information in the service delivery can be implemented through our unicast and multicast protocols. The other two functions required in service provision can be supported by our hierarchical service and resource management architecture as discussed in this subsection.

3.5.1 Service discovery

The service discovery follows our hierarchical structure from the bottom layer to the top layer, i.e., first searching for the services and resources locally, then regionally, and at last globally. Examples are shown in Fig. 2. As our service domains are formed based on geographic information, this searching procedure naturally follows *proximity principle*. Specially, when a service requestor (SR) wants to request one or multiple services, it will send a Query message with service descriptions (including service IDs and parameters) to an appropriate service coordinator. When there is a LC in its zone, it issues the Query to its LC; otherwise, if the SR knows RC, it will issue Query to RC. When neither LC nor RC is known, SR will send Query to GC directly or through GC-Zone. If having no information on GC, GC-Zone, RC or LC, SR can start an expanded ring search, which is actually a fully distributed searching as in some service discovery protocols.

When a LC receives a Query or it itself has service requests, if any of the requested services can be satisfied by some local SPs according to its record, the Query will be forwarded to one or more candidate SPs. On receiving the Query, if it can provide the service, the SP will send back a Hit message describing what it can provide to the LC or initial SR (which one to respond to is based on the policy and the service request). If the requested services could not be satisfied locally, the LC will follow the same searching procedure as above and resort to its upper layer service coordinators, and

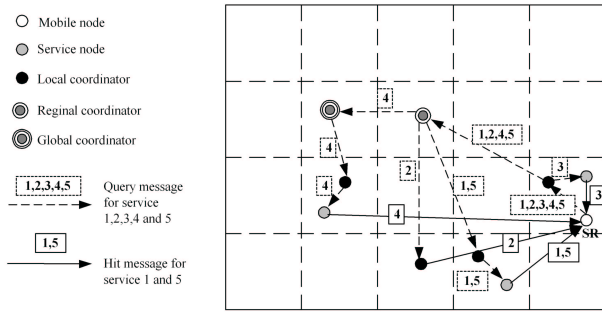


Figure 3. Service coordination in discovery.

the last option is to look for SPs through expanded ring search. When a RC or GC receives a Query, it will process the message similarly. If any of the requested services can be satisfied by the service zones recorded, the Query will be further forwarded to the selected candidate zones. Selection criteria such as QoS requirements, geographic closeness will be used when multiple candidates are available. When reaching a destined service zone, the Query will be forwarded to its LC, which will forward the Query to appropriate SPs to check if the services can be supported. Without receiving any Hit message, the initial SR or corresponding coordinator can retry by reissuing the Query or resorting to another SP candidate or upper coordinator.

We also use service cache to optimize the service discovery. Each node keeps a service cache. Whenever a node receives a Hit or REGISTER message from a SP, it will cache the SP's position and service information. The information will be removed after caching for a period $Intval_{cache}$. During the service discovery, when a non-coordinator node receives a Query, it will look up its service cache for the requested services, and forward the Query to qualified SPs if available, instead of forwarding the Query to destined coordinator. For a coordinator, if the requested services can not be satisfied by its service records, it will try to search for non-local SPs in its service cache, and if this also fails, it will forward the Query to its upper coordinator. Other caching techniques ([31]) can also be applied to further improve the data access performance.

3.5.2 Service coordination

A service request may need to be satisfied by several SPs. Our membership and service management framework can efficiently support the search of multiple SPs. During service discovery, when a request can not be fully satisfied by the resources tracked by a LC, it will be further issued to the RC and then GC. Based on the knowledge of larger-range resource and service information, RC and GC can allocate necessary remaining resources to the SR. An example of coordination in service discovery is shown in Fig. 3, in which the SR issues a Query for five services, one of which is satisfied in local zone, another three are satisfied regionally, and the last one is satisfied with the help of GC.

4 Performance Evaluation

4.1 Simulation Overview

We implemented the proposed scheme within the Global Mobile Simulation (GloMoSim) [25] library. We represent our scheme as SGSP (Scalable Geographic Service Provision). For performance reference, we also implemented the distributed service discovery protocol AODV-SD [7]. In our hierarchical structure, the zone size is set as 400m, $Intval_{LC} = 4s$ and $D = 1000m$. Since a MANET is normally manageable with a two-layer structure, in our simulation, our hierarchical structure has two layers. For membership and service management, $Intval_{local} = 4s$ and $Intval_{global} = 6s$. In service discovery, for fair comparison, caching has been used for intermediate nodes in both AODV-SD and SGSP with $Intval_{cache} = 8s$. The simulations were run with 300 nodes randomly distributed in the area of $2000m \times 2000m$. The nodes movement follows the *random waypoint* model [3]. The moving pause time was set as 0 second, minimum speed was 0 m/s and default maximum speed was 20 m/s except in the performance evaluations by varying moving speed.

We set the service distribution based on the models in [15] and [21]. Each SP offered 5 different services. The total number of service types offered by the network followed the square root distribution in [15] as $40 + 5 * \sqrt{N}$, where N is the number of SPs, e.g., with 75 SPs, there are 375 services in 137 types in the network. 25% of the mobile nodes were SPs. Each mobile node in the network sent a service query every 100s. A query requests one type of service not offered by the requestor itself. When multiple SPs available for one service, we used a simple selection criteria according to the geographic closeness. Each simulation lasted 500 simulation seconds. A simulation result was gained by averaging over 20 runs with different seeds. The following metrics were studied:

1. *Service discovery success ratio*: The ratio that issued Query messages received corresponding Hit messages. In the simulation, the Hit message was sent to the SR, and a SR wouldn't retry if not receiving Hit.
2. *Number of message transmissions*: The total number of control message transmissions. Each forwarding of the message was counted as one transmission. To provide more insights on the performance of different components of the framework, we also measured separately the message transmissions generated by the hierarchical structure building and maintenance, membership management, adaptive routing and service discovery.
3. *Average discovery delay*: The average time interval between a node issuing a service query and its receiving of the corresponding Hit message from the SP.

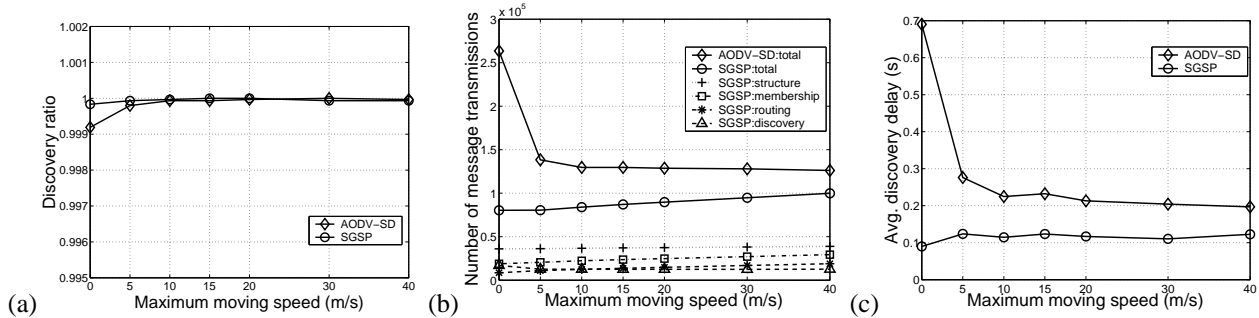


Figure 4. Performance vs. maximum moving speed: (a) discovery success ratio; (b) control overhead; (c) average discovery delay.

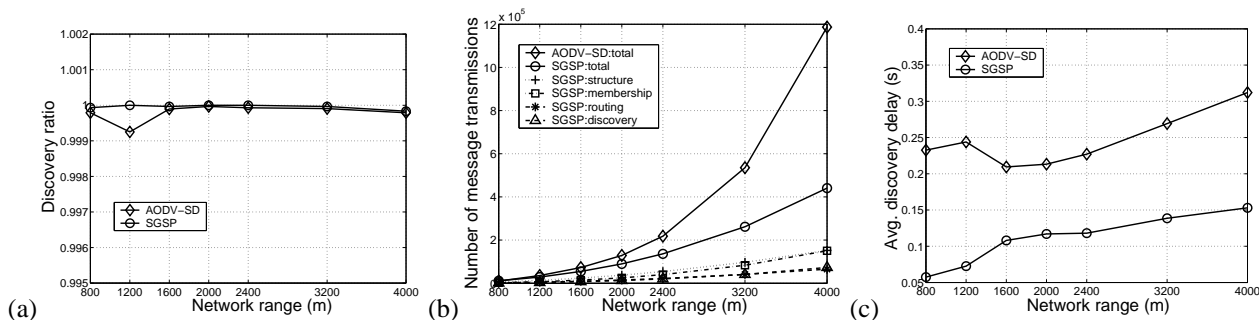


Figure 5. Performance vs. network range: (a) discovery success ratio; (b) control overhead; (c) average discovery delay.

4.2 Simulation Results

We present the simulation results with various moving speeds and network sizes.

4.2.1 Effect of moving speed

We first study the protocol performance by varying the maximum moving speed from 0 m/s to 40 m/s. From Fig. 4, SGSP keeps a more stable performance and performs much better than AODV-SD under almost all the mobility cases due to its robustness to dynamics. AODV-SD's performance degrades in a low-mobility or static network since mobility will help the network connection. AODV-SD is based on expanded ring search, in which each time the discovery fails, the requestor will increase the flooding range of its query message. Both protocols can achieve a nearly 100% discovery ratio. The high discovery ratio of AODV-SD is at the expense of relatively high searching overhead as shown in Fig. 4 (b). In SGSP, with the help of the hierarchical structure and membership and service management, no flooding is involved in the discovery. The efficiency is at the cost of the overhead in hierarchical structure maintenance and membership management. From Fig. 4 (b), the overheads for hierarchical structure, membership management and routing increase slightly with the increase of mobility. For the hierarchical structure maintaining, as mobility increases, there are more LC han-

doers and elections due to more zone crossings. In membership management, the increasing zone crossings result in more frequent changes of service zones and hence more reporting to GC for the changes. When the mobility increases, our routing protocol can adapt its parameter values accordingly and generate more beacons to timely catch the topology changes. By following our service management architecture, a SP can be located more efficiently and quickly as shown in Fig. 4 (c). With expanded ring searching, it may require several rounds for AODV-SD to discover a SP, which leads to a longer discovery delay.

4.2.2 Effect of network size

To study the protocol's scalability to large network size, we varied the network range from $800m \times 800m$ to $4000m \times 4000m$. The node density is kept as before, so the total number of nodes is varied from 48 to 1200. Confirmed by the results in Fig. 5, SGSP has a much better scalability to network size than AODV-SD. Both protocols can keep a nearly 100% discovery ratio with different network sizes. The overheads of both protocols increase when the network size increases with AODV-SD's overhead increasing much more sharply. When the network size reaches $4000m \times 4000m$ with 1200 nodes, AODV-SD's overhead is almost three times of SGSP's overhead. With a larger network size, AODV-SD's expanded ring search may extend to a larger range, while the network-layer

messages in SGSP may traverse longer paths, both of which contribute to the larger message overheads. Since the SP density is kept as 25%, there are more SPs when the network size increases, and hence there are more overheads in hierarchical structure maintenance and membership management for SGSP. As expected, the discovery delays of both protocols increase with the increase of network size, as the candidate SP may be farther away from the SR in a larger network terrain.

5 Conclusions

We have designed a scalable and efficient geographic routing and service provision framework for MANET. In the framework, we build a flexible and self-organizing hierarchical structure which can adapt to the service and management requirements in the network. Based on the structure, we propose a scalable resource tracking and service membership management mechanism to support timely and coordinative service discovery. We also integrate two geographic routing protocols proposed in our earlier work into the service provision framework: an adaptive and reactive unicast protocol to support both short-term control message transmissions and long-term data transfers required by a service provision system, and a scalable and robust multicast protocol to efficiently support group communications. Based on all of the above, instead of addressing only partial service functions, our service provision framework can comprehensively support efficient service discovery and delivery and robust service coordination.

References

- [1] S. Basagni, I. Chlamtac, and V. R. Syrotiuk. Location aware, dependable multicast for mobile ad hoc networks. *Computer Networks*, 36(5-6):659–670, August 2001.
- [2] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. In *Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DialM 99)*, August 1999.
- [3] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. G. Jetcheva. A performance comparison of multihop wireless ad hoc network routing protocols. In *ACM/IEEE MOBICOM*, 1998.
- [4] K. Chen and K. Nahrstedt. Effective location-guided tree construction algorithms for small group multicast in MANET. In *IEEE INFOCOM*, 2002.
- [5] L. Cheng. Service advertisement and discovery in mobile ad hoc networks. In *Proc. of CSCW*, November 2002.
- [6] K. Edwards and T. Rodden. *Jini example by example*. Prentice Hall PTR, June 2001.
- [7] J. Garcia-Macias and D. Torres. Service discovery in mobile ad-hoc networks: better at the network layer? In *International Conference Workshops on Parallel Processing*, 2005.
- [8] E. Guttman, J. Veizades, C. Perkins, and M. Day. Service location protocol, version 2. *IETF RFC 2608*, June 1999.
- [9] M. Heissenbuttel and T. Braun. BLR: beacon-less routing algorithm for mobile ad-hoc networks. *Elsevier's Comp. Commun. Journal*, 27(11):1076–1086, July 2004.
- [10] S. Helal. Konark - a service discovery and delivery protocol for ad-hoc networks. In *IEEE WCNC*, 2003.
- [11] B. Karp and H. T. Kung. Greedy perimeter stateless routing for wireless networks. In *ACM/IEEE MOBICOM*, 2000.
- [12] U. Kozat and L. Tassiulas. Network layer support for service discovery in mobile ad hoc networks. In *IEEE INFOCOM*, 2003.
- [13] J. Li and et al. A scalable location service for geographic ad hoc routing. In *ACM/IEEE MOBICOM*, 2000.
- [14] L. Li and L. Lamont. A lightweight service discovery mechanism for mobile ad hoc pervasive environment using cross-layer design. In *IEEE PerCom*, 2005.
- [15] C. Mastroianni, D. Talia, and O. Verta. A P2P approach for membership management and resource discovery in grids. In *International Conference on Information Technology: Coding and Computing (ITCC)*, volume 2, pages 168–174, 2005.
- [16] M. Mauve, H. Fubler, J. Widmer, and T. Lang. Position-based multicast routing for mobile ad-hoc networks. In *Poster section in ACM MOBIHOC*, June 2003.
- [17] N. Nahata, P. Pamu, S. Garg, and A. Helmy. Efficient resource discovery for large scale ad hoc networks using contacts. *ACM SIGCOMM Computer Communications Review*, 32(3), July 2002.
- [18] V. Park and J. Macker. Anycast routing for mobile services. In *Proc. of Conference on Information Sciences and Systems (CISS)*, March 1999.
- [19] C. Perkins and R. Koodli. Service discovery in on-demand ad hoc networks. *IETF Internet Draft. Work in progress*, October 2002.
- [20] O. Ratsimor. Allia: alliance-based service discovery for ad-hoc environments. In *Proc. of the ACM Mobile Commerce Workshop*, 2002.
- [21] F. Sailhan and V. Issarny. Scalable service discovery for MANET. In *IEEE PerCom*, 2005.
- [22] M. Storey, G. Blair, and A. Friday. MARE: resource discovery and configuration in ad hoc networks. *Mobile Networks and Applications*, 7(5), October 2002.
- [23] The Salutation Consortium. Salutation architecture specification v2.0c (part-1), June 1999.
- [24] M. Transier, H. Fubler, J. Widmer, M. Mauve, and W. Effelsberg. Scalable position-based multicast for mobile ad-hoc networks. In *Proc. of the First International Workshop on Broadband Wireless Multimedia: Algorithms, Architectures and Applications (BroadWim 2004)*, October 2004.
- [25] UCLA Parallel Computing Laboratory. Glomosim. <http://pcl.cs.ucla.edu/projects/glomosim/>.
- [26] Uddi version 3.0 published specification, July 2002.
- [27] UPnP Forum. Universal plug and play device architecture, version 1.0, June 2000.
- [28] C. Ververidis and G. Polyzos. Routing layer support for service discovery in mobile ad hoc networks. In *IEEE PerCom*, 2005.
- [29] X. Xiang, Z. Zhou, and X. Wang. Robust and scalable geographic multicast protocol for mobile ad hoc networks. In *IEEE INFOCOM Minisymposium*, 2007.
- [30] X. Xiang, Z. Zhou, and X. Wang. Self-adaptive on demand geographic routing protocols for mobile ad hoc networks. In *IEEE INFOCOM Minisymposium*, 2007.
- [31] L. Yin and G. Cao. Supporting cooperative caching in ad hoc networks. *IEEE Trans. Mobile Computing*, 5(1), January 2006.
- [32] M. Zorzi and R. R. Rao. Geographic random forwarding (GeRaF) for ad hoc and sensor networks: energy and latency performance. *IEEE Trans. Mobile Computing*, 2(4), Oct.-Dec 2003.