Approach of Association Rules Mining for Service Discovery in Mobile Adhoc Network

Ashish T. Bhole
Assistant Professor
Department of Computer Engineering,
SSBT's College Of Engineering and Technology,
Bambhori, Jalgaon, INDIA

Shital Ranade
Research Scholar
Department of Computer Engineering,
SSBT's College Of Engineering and Technology,
Bambhori, Jalgaon, INDIA

ABSTRACT
The Service Discovery in Mobile Ad hoc Networks (MANET) is a difficult task because of the changing nature of such networks. Different services are available in MANET considering time and location. Many varying traditional solutions to service discovery of Internet are not well suited for MANET because of their Ad hoc nature. Consequently, Service Discovery (SD) in the network a perquisite for efficient usage of network resources is a complex problem. In this work Association Rules mining algorithm are used to get service discovery. Using the correlation among the services and piggybacking future service request answers along with current service requests, there is significant gain in performance. Two algorithms of association rules mining are used for discovering the services in MANET. FPTree algorithm is already been used for service discovery. We have approached towards Apriori algorithm for service discovery in MANET with the significant gain in performance.

General Terms
MANET, Data Mining.

Keywords
Service Discovery, Association Rules Mining, FPTree Algorithm, Apriori Algorithm.

1. INTRODUCTION
Mobile ad hoc network is an autonomous system of mobile routers connected by wireless links, the union of which forms an arbitrary graph. The routers are free to move randomly and organize themselves arbitrarily making the network's wireless topology change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the larger Internet. The strength of the connection can change rapidly in time or even disappear completely. Nodes can appear, disappear and reappear as the time goes on and all the time the network connections work between the nodes that are part of it. Ad hoc networks are consisting of nodes not connected to any static infrastructure. An ad-hoc network is a LAN or other small network, especially one with wireless connections, in which some of the network devices are part of the network only for the duration of a communications session or, in the case of mobile or portable devices, while in some close proximity to the rest of the network.

1.1 Service Discovery in MANET
Service discovery allows end systems to discover desired services on networks automatically, eliminating configuration by system administrators. Service discovery is a technology that can dynamically map service descriptions into service access points. Service discovery involves advertisements, exploration and indexing of available services in a dynamic way. It provides a powerful and flexible way for service users to locate available desired services on networks. Services are of various types, such as printing services, computational services, and storage services. Service discovery differs from other resource discovery such as information retrieval by focusing on where desired services are provided. Traditionally, service users depend on a priori knowledge or manual configuration to learn about desired services on networks, involving non-trivial administrative overhead as more devices are network enabled and more services are available on networks. Moreover, administrative configuration becomes difficult or even impossible in certain situations such as ad-hoc networks. By using service discovery technology, service users no longer need to know the access points of desired services via a priori knowledge. Instead, they just need to specify the characteristics of desired services, which will be mapped into available service access points automatically in any network that supports service discovery. Service discovery is a technology for building autonomic networking and distributed systems. Thus, service discovery is the ability to discover and form an ad-hoc network without explicit user direction. It facilitates devices and services to properly discover, configure, and communicate with each other. Service discovery minimizes administrative overhead and increases usability [1]. So the main purpose of service discovery is to detect services and devices offered by devices and computers in a network and to announce offered services to devices and computers.

2. LITERATURE SURVEY
Various service discovery techniques are currently being developed and standardized by the various industry consortiums and organizations. Service discovery is the process of mapping a service description to a so called service location. Service discovery in wireless networks has been classified into two types of approaches, centralized and decentralized. Other approaches are also discussed. Centralized and Decentralised approaches are discussed further.

2.1 Centralized Approaches

Centralized service discovery is the one in which one dedicated server; named *directory agent* is responsible for registering and maintaining a database of all the services present within the network nodes. These service providing network nodes are termed as *service agents.* Every service request and response is routed through *directory agent.* Some conventionally adopted protocols, including Sun's Jini, the service directory service Universal Description, Discovery, and Integration, DReggie, Light Weight, rely on a central directory to register and discover services.

Jini [2] is a distributed service-oriented architecture developed by Sun Microsystems. JINI system provides mechanism for service construction, lookup, communication and use in a distributed system. Services in a Jini system communicate with each other by using a service protocol, which is a set of interfaces written in the Java programming language. Two examples of services are printing a document and translating from one word processor format to some other. Jini Lookup Service (JLS) maintains dynamic information about the available services in a Jini federation (a collection of Jini services). When a Jini service wants to join a Jini federation, it first discovers one or many Jini Lookup Service from the local or remote networks. The service then uploads its service proxy (i.e. a set of Java classes) to the Jini Lookup Service. The service clients can use this proxy to contact the original service and invoke methods on the service.

UDDI [3] Universal Description, Discovery and Integration define a universal method for enterprises to dynamically discover and invoke Web services. The information of available services is concentrated in directories. The service advertisement is based on multicasting, and the service register is based on unicasting.

The project DReggie [4] is an attempt to enhance the matching mechanisms in Jini and other service discovery systems. The key idea in DReggie is to enable these service discovery systems to perform matching based on *semantic information* associated with the services. As is well-known, service matching in existing systems is strictly syntactic (i.e., string matching). Semantic information of services consists of their extensive descriptions including, but not limited to, capabilities, functionality, portability and system requirements. Semantic service matching introduces the possibilities of fuzziness and inexactness of the response to a service discovery request. In the DReggie system, a service discovery request contains the description of an "ideal" service one whose capabilities match exactly with the requirements. Thus, matching now involves comparison of requirements specified with the capabilities of existing services. Depending on the requirements, a match may occur even if one or more capabilities do not match exactly.

Light Weight [5] is very light-weight, robust and reliable model for service discovery in wireless and mobile networks by taking into account the limited resources to which are subjected the mobile units. Three processes are involved in service discovery protocols using virtual dynamic backbone for mobile ad hoc networks: registration, discovery and consistency maintenance. More specifically, the model analytically and realistically differentiates stable from unstable nodes in the network in order to form a subset of nodes constituting a relatively stable virtual Backbone.

2.2 Distributed Approaches

In decentralized approach, also known as distributed service discovery, there is no dedicated *directory agent*. Some distributed approaches are Allia, Konark, GSD, IBM DEAPspace and UPnP.

Allia [6] is a peer-to-peer caching based and policy-driven agent service discovery framework to facilitate cross-platform service discovery in ad-hoc environments for mobile electronic commerce applications. This approach removes the problems associated with structured compound formation of agent communities in mobile commerce environment and achieves high degree of flexibility in adapting itself to the changes of the ad-hoc environment. This framework takes into consideration device capabilities and limitations, user preferences regarding usage of the devices, application specifics with respect to mobile commerce and adapts accordingly.

Konark [7] is a service discovery and delivery protocol designed specifically for ad-hoc, peer-to-peer networks, and targeted towards device independent services in general and m-commerce oriented software services in particular. It has two major aspects service discovery and service delivery. For discovery, Konark uses a completely distributed, peer-to-peer mechanism that provides each device the ability to advertise and discover services in the network. The approach towards service description is XML based. It includes a description template that allows services to be described in a human and software understandable forms. A micro-HTTP server present on each device handles service delivery, which is based on SOAP. Konark provides a framework for connecting isolated services offered by proximal pervasive devices over a wireless medium.

Group based SD (GSD) [8] protocol is based on the concept of peer-to-peer caching of service advertisements and group-based intelligent forwarding of service requests. It does not require a service to register to a registry or lookup server. Services are described using an ontology based on the DARPA Agent Mark up Language (DAML+OIL). It exploit the semantic class/subclass hierarchy of DAML to describe service groups and use this semantic information to selectively forward service requests to respective nodes. DAML-based service description helps us in achieving increased flexibility in service matching.

IBM has developed DEAPspace [9] that addresses the service discovery problem in wireless single-hop ad-hoc networks. Each device in DEAPspace maintains a view of all the services present in the network, and periodically exchanges its view of the world, i.e., the full list of its service map, with its neighbours. The periodic broadcast is scheduled in a proactive way in that when a device finds its local services absent in messages broadcast by its neighbours or about to expire, it schedules its broadcast sooner than usual. They showed the timely convergence of their approach compared to regular periodic broadcast scheme.

Universal Plug and Play (UPnP) [10] is supported by Microsoft. This architecture is designed to extend the original Microsoft Plug and Play peripheral model. UPnP works primarily at lower layer network protocols suite (i.e. TCP/IP). UPnP uses the Simple Service Discovery Protocol (SSDP) for discovery of services on Internet Protocol based networks. When a service joins a network, it sends out an advertisement

message, notifying the world about its presence. The advertisement message contains a Universal Resource Locator (URL) that identifies the advertising service and a URL to a file that provides a description of the advertising service. When a service client wants to discover a service, it can either contact the service directly through the URL that is provided in the service advertisement, or it can send out a multicast query request.

2.3 Other Approaches

The IETF's Service Location Protocol (SLP) [11] represents the services in the form of URL and attributes and can work in directory-based or directory-less fashion. SLP has been developed to reduce the overload of the centralized directory. The SLP provides a flexible and scalable framework for providing hosts with access to information about the existence, location, and configuration of networked services. Traditionally, users have had to find services by knowing the name of a network host (a human readable text string) which is an alias for a network address. SLP eliminates the need for a user to know the name of a network host supporting a service. Rather, the user supplies the desired type of service and a set of attributes which describe the service. Based on that description, the Service Location Protocol resolves the network address of the service for the user. SLP provides a dynamic configuration mechanism for applications in local area networks. Applications are modelled as clients that need to find servers attached to any of the available networks within an enterprise. For cases where there are many different clients and/or services available, the protocol is adapted to make use of nearby Directory Agents that offer a centralized repository for advertised services. Such approaches are typically based on either pull- or push based broadcasting.

A novel resource and service discovery mechanism for MANET's using Routing Intelligent Mobile Agents (RIMA) [12] is been proposed. RIMA's periodically collect routing, resource and service availability information and index the same in appropriate RIMA nodes. Every mobile node is close to at least one RIMA node. The mobile node running an application in need of a resource or a service dispatches a Discovery agent to its nearest RIMA node. The index available at the RIMA node will guide the Discoverer agent to the resource or service node.

3. RELATED WORK

Using data mining for MANET service discovery is a relatively new concept and only a small amount of work is done this direction. Hu [13] applied the mining technique for composite web service discovery but his work was not applied to Mobile Ad hoc Network. Jabas [14] applies the mining technique to MANET traffic to find interesting relations ships among MANET nodes. Their research was motivated by the fact that these relationships can assist routing, MAC and other protocols of MANET.

Service Mining [13] for Composite Service Discovery approach focuses on how to discover composite services through service mining. Manually searching for required services takes lot of time and energy for users. Provide web service users with the suggestions of composite service patterns. Thus, analyzing the relationship between web services becomes critical for us. Normally there are two kinds of service patterns: non-sequential patterns and sequential patterns. Approach is based on two data mining techniques, which are Multilevel Association Rule and Sequential Pattern Mining. There are three requirements can be match illustrated below.

1. Discover related services in more efficient way. Web service user can reduce time of manually searching for required services on UDDI since they can adopt discovered patterns which represents the frequently-used patterns and highly-related web services.
2. Discover related services to reach user's requirement. The more complete web usage data we can collect, the more different patterns we can extract. Different patterns represent different user behavior, and thus generation of enormously different patterns better satisfy users' requirement.
3. Discover syntactic and semantic relationship among related services. According to our approach, we can obtain service patterns in syntactic relationship and in semantic relationship. The web usage data mainly depends on users' behavior, and users may choose one of them or both of them to complete a specific task.

One way to enhance MANET performance is to disclose the hidden characteristics (patterns) of MANET and to utilize them in the distributed algorithms. This can be achieved by analyzing the MANET traffic in a tempo-spatial domain. In MANET mining, a new framework is proposed i.e. to apply the association rules to the packets traffic in MANET with respect to time. This technique paves the way for the other MANET algorithms.

Association Rules mining algorithms find out correlation among the services of one session and can predict future discovery requests based on current service discovery request. Noman Islam has used the FPTree algorithm of Association Rules mining algorithm for service discovery in MANET. Using the correlation among the services and piggybacking future service request answers along with current service requests authors have proved significant gain in performance [15].

The Association Rules mining component uses the records stored in log database and applies FPTree mining algorithm on the these records to determine the correlation among the service requests. The algorithm determines what service requests are issued together i.e. frequent item sets. The frequent item sets thus discovered can be used to predict future service discovery requests based on the current service discovery requests. The FPTree approach has been applied to broadcast approach and found to give significant improvement in results. The approach has been implemented in JIST/SWANS network simulator [16, 17]. The simulation has been done in a field of 500 × 500 with random placement of nodes. They used the default configuration parameters of JIST/SWANS for simulation. Service discovery requests are generated in a session such that they are correlated. Authors developed a Correlated Data Generator (CDG) module that issues service discovery requests based on a correlation matrix **CM** of **n** × **n**, where **n** is the number of services in the network. The correlation matrix is built programmatically using Java Random Number Generator that follows a Uniform Distribution. An **n** × **n** matrix of random numbers **R** is generated and based on this matrix, the correlation matrix **CM.** It was the novel approach to improve service discovery in MANET using Association Rules mining. The approach had been applied to broadcast approach and found to give significant improvement in results [15].

4. PROPOSED SYSTEM

Figure 1 shows the proposed approach of service discovery in MANET using Association Rules Mining. Mainly there are three main components.

1. Log Database: - It consists of various session records. Whenever the service request is send by the consumer the record will be logged in log database. A session consists of number of service discovery request by consumer and reply and request associated with that request. The Fields in service request and service reply is shown in Table 1 and Table 2 respectively.

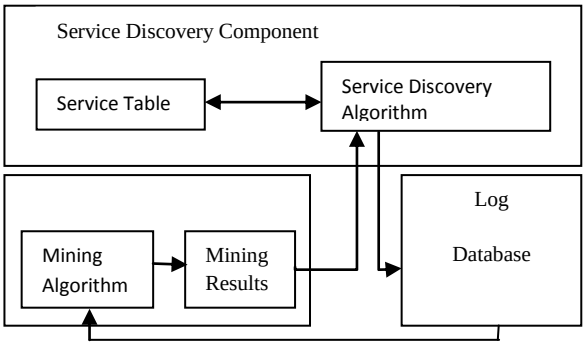

Figure 1: Proposed Approach Modules

2. Association Rules Mining component: - It uses the records stored in log database. It uses the mining algorithms on these algorithms to determine correlation among service requests. It can use either Apriori or FPTree.

Table 1: Service Request format

Field	Description
Service ID	The unique identifier of the service to be requested
Originator	The node who has initiated the service discovery request
TTL	Time to Live value to minimize network overhead
Sender	The address of the previous hop who has forwarded the SREQ
Request ID	The unique identifier for the service request
Session ID	The unique identifier of the session

3. Service Discovery Component: - If the incoming message M is a service discovery request i.e. SREQ, the method extracts the requested service from the packet and looks for this service in its local service table. If the service table contains the desired service, then all the services related to this service are also picked and they are also returned along with the service in SREP message. If no service is found in service table, then SREQ is propagated ahead. If the incoming message is SREP, the node sees reply of its own SREQ, then it inserts this information in its service table along with other related service information. If SREP doesn't belong to node, it propagates the SREP message along with recording service and related information in its local service table.

5. PROPOSED WORK

Proposed system is being simulated using JiST packages along with JAVA language on Net beans platform. This section discusses in detail the experimental results of the proposed approach.

Table 2: Service Reply format

Field	Description
Service ID	The unique identifier of the service to be requested
Provider	The node who has requested service
Requestor	The node whose service request is been responded
TTL	Time to Live value to minimize network overhead
Request ID	The unique identifier for the service request
Session ID	The unique identifier of the session

5.1 Simulation Parameters

Various parameters are been considered for comparing the performance of these two algorithms. Different results are compared by varying the number of nodes during the simulation. As the number of nodes increases, the number of request going outside drastically increases without using association rules algorithm. For FPTree and Apriori algorithm they are slightly changing. Number of services available within the network also affects the performance. For the proposed approach we have limited the number of services to 3 and 5. Simulation is carried out for three conditions as follows:

a. Without using mining: Here mining component is not used. Simply logging component and service discovery component is present.

b. Using FPTree Algorithm: It uses the FPTree algorithm for service discovery. This algorithm is already been implemented [15].

c. Using Apriori Algorithm: This algorithm is my approach for service discovery in MANET. Here item sets are generated using Apriori algorithm to predict the future request.

5.2 Simulation

Number of nodes makes the significant affect on the service discovery. Assume the number of services 5, TTL 3 and range of nodes from 40 to 55 and its results are shown in Figure 2. Figure 2 shows the comparative results for without using mining, FPTree and Apriori. From the above it is observed as follows:

1. Without using association rules mining algorithm the request going outside is much more.
2. Using FPTree algorithm request going outside for nodes 40 to 45 it is decreasing and increasing for 46 to 50. Than again decreasing for nodes 51 to 55.Thus results obtain in FPTree are in wavelet form.
3. Using Apriori algorithm at 40 it is more and than goes on decreasing. At 50 it is 1 and then remains static.

5.3 Results and Discussion

Without using any of association rules algorithm request went outside is significantly more. Thus it consumes more time and more request response is generated. This leads to the generation of more traffic as shown in Figure 3.

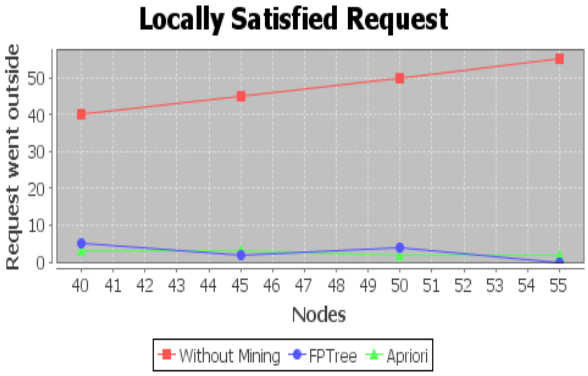

Figure 2: Simulation for 40 to 55 Nodes

From the FPTree algorithm implementation there is significant gain in the performance. But the point to be noted is for the same range of nodes request went outside in not constant. The results obtained are in fluctuating manner as wavelet form. For the slightly variance in number of nodes not the same results will be obtained as that of previous nodes. Thus future request generation are less obtained.

service discovery in Mobile Ad hoc Network. Both Apriori and FPTree algorithm are suitable for finding association between service request and its response.

Without using Association rules mining for service discovery are having poor results. FPTree algorithms results are fluctuating. Apriori is better than FPTree algorithm as results have proved. As the number of nodes increases static ratio is gained in Apriori algorithm simulation.

7. REFERENCES

[1] Feng Zhu, Matt Mutka and Lionel Ni, "Service Discovery in Pervasive Computing Environments", IEEE Pervasive Computing, Volume 4, Issue 4, pp. 81-90, October 2005.

[2] Jini Technology Core Platform Specification, v. 2. 0, Sun Microsystems, 2003.

[3] UDDI Version 3.0 Published Specification, [Online] Available: http://www.uddi.org, Jul. 2002.

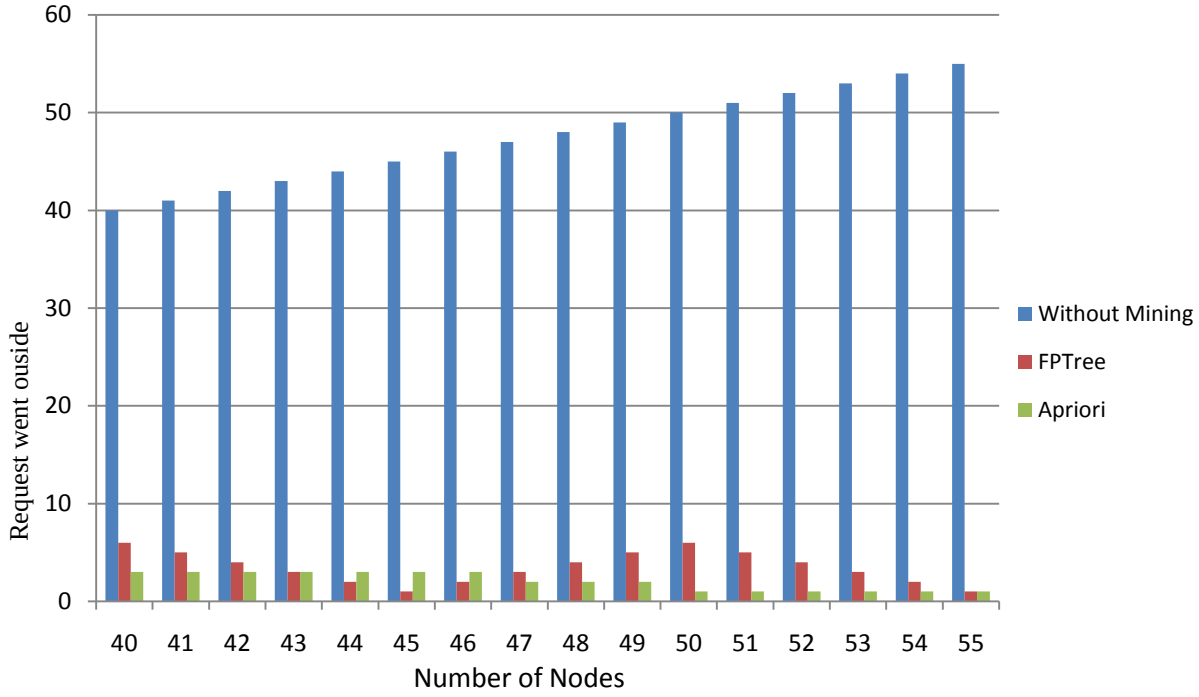

Figure 3: Simulation Comparisons for 40 to 55 Nodes

In the Apriori algorithm results observed are the static one. When number of nodes is certain ranges it gives the same results. Future request for the services generated here will be more effective as compared to the FPTree algorithm. From the above, Apriori is better than the FPTree algorithm in terms of varying number of nodes.

6. CONCLUSION

A simple approach for service discovery is been proposed over here. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and, unlike the if-then rules of logic, association rules are probabilistic in nature. The overall concepts required for the proposed system for the implementation of Association Rules. Association rule approach is some new aspect for the

[4] Dipanjan Chakraborty, F.P., Sasikanth Avancha and Anupam Joshi, "DReggie : Semantic Service Discovery for M-Commerce Applications", in Reliable and Secure Applications in Mobile Environment in Conjunction with 20th Symposium on Reliable Distributed Systems (SRDS), 2001.

[5] Mallah R.A. and A. Quintero, "A Light-Weight Service Discovery Protocol for Ad Hoc Networks", Journal of Computer Science, pp. 330-337, 2009.

[6] Olga Ratsimor, D.C., Anupam Joshi and T. Finin, "Allia: Alliance-based Service Discovery for Adhoc

Environments", in WMC '02: Proceedings of the 2nd International Workshop on Mobile commerce, ACM

Service Discovery and Delivery Protocol for Ad hoc Networks", in 3rd IEEE Conference on Wireless Communication Networks (WCNC), 2003.

[8] Chakraborty, A.J. "GSD : A Novel Group-based Service Discovery Protocol for MANETS", in 4th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002), Stockholm, Sweden, 2002.

[9] M. Nidd, "Service Discovery in DEAPspace", in IEEE Personal Communications, pp. 39-45, 2001.

[10] Understanding Universal Plug and Play, 2000, Microsoft UPnP forum [Online] Available: http://www.upnp.org.

[11] E. Guttman, C.P., J. Veizades and M. Day, "Service Location Protocol, Version 2", IETF RFC 2608, 1999.

[12] Basavaraj Talwar, P.V.a.L.M.P., "A Method for Resource and Service Discovery in MANET", Wireless Personal Communications, pp. 301-323, 2007.

[13] Hu Jia Ying, "Service Mining for Composite Service Discovery", in Computer Science and Information Engineering, National Central University, 2006

[14] Jabas A., R.M. Garimella and S. Ramachandram, "MANET Mining: Mining Temporal Association Rules", in International Symposium on Parallel and Distributed Processing with Applications, Sydney, NSW, Australia, 2008.

[15] Noman Islam, Zubair A. Shaikh**,** "Service Discovery in Mobile Ad hoc Networks Using Association Rules Mining", IEEE 13th International Multitopic Conference, 2009.

[16] Barr R., Z.J. Haas and R.V. Renesse, "JiST: An Efficient Approach to Simulation using Virtual Machines", in Press, 2002.

[7] S. Helal, N.D, V. Verma and C. Lee, "Konark - A Software Practice & Experience, Vol. 35, No. 6, pp. 539-576, 2005.

[17] Barr R., Z.J. Haas and R.V. Renesse, "Scalable Wireless Ad hoc Network Simulation", in Handbook on Theoretical and Algorithmic Aspects of Sensor, Adhoc Wireless and Peer-to-Peer Networks, pp. 297-311, 2005.

8. AUTHORS PROFILE

Ashish T. Bhole is working as Assistant Professor in the Department of Computer Engineering, SSBT's COE & T, Bambhori, Jalgaon, India. He received B.E. Degree in Computer Engineering in 1999 from SSBT's COE & T, Bambhori, Jalgaon affiliated to North Maharashtra University, India, M. Tech. in Computer Science & Engineering in 2008 from S. A. T. I., Vidisha affiliated to Rajiv Gandhi Technological University, India and currently pursuing Ph.D. in Computer Science & Engineering. He has 12 years of teaching & research experience. He has published 16 papers in various International & National Journals, Conference Proceedings. He is a Member of Institution of Engineers (India) & Life Member of Indian Society for Technical Education. His research area includes Computer and Wireless Networks, Network Security, Routing Protocols and Traffic Engineering.

Shital Ranade is a Research Scholar in the Department of Computer Engineering, SSBT's COE & T, Bambhori, Jalgaon, India. She received B.E. Degree in 2008 from SSVPS's College of Engineering, Dhule affiliated to North Maharashtra University, India. She is presently working as lecturer in Department of Computer Engineeiring, KCES's COEIT, Jalgaon, India. She has 3.5 years of teaching experience. She has published two National and one International Papers in conference proceedings.