AN EXTENDED ANALYSIS OF AN INTEREST-BASED SERVICE DISTRIBUTION PROTOCOL FOR MOBILE AD HOC NETWORKS

Mohamed Hamdy and Birgitta König-Ries
Institute of Computer Science, Friedrich-Schiller University, Jena, Germany

Keywords: MANET, SOA, replication.

Abstract: In principle, service orientation is a suitable paradigm to allow for effective resource sharing in wireless ad hoc networks. However, special attention needs to be paid to ensure a high service availability since this is the basis for reliable service execution. Unfortunately, typical characteristics of ad hoc networks like ever-changing topology and limited resources have a dramatic negative effect on service availability. Therefore, special measures are needed to cope with this problem. In general, replication for services as well as data represents an efficient solution when the availability of some resource needs to be ensured. In this paper, we extend the service replication and distribution protocol presented in our prior work which is based on the interest of clients and providers of a specified service by taking into account not just one but all partitions of the network. By elaborating an extensive detailed simulation, the efficiency of replication and the allocation correctness are being examined. The results show that relying on our protocol is feasible. Since our protocol utilizes high level (application level) information about the available services in the ad hoc network only and does not rely on lower level information like network or protocol specifics, it is applicable in a wide range of settings.

1 INTRODUCTION

In many applications, wireless ad hoc networks represent the only feasible solution to achieve communication for each of the network participants. The mobile hosts are working in a collaborative way to achieve the core network functions such as routing. In addition, collaboration is needed to provide individual devices with information and/or functionality they do not possess themselves. Ad hoc characteristics like resource-restricted, wireless-enabled participants' devices, limitations of battery lifetimes, and the ever-changing network topology pose very tight constraints on all of the applications to be deployed on this type of network. In order to enable mobile hosts to share functionality as is needed to guarantee the functioning of the network, service orientation can be used: Provided functionality is offered as a service, required functionality is searched for via a service request.

However, provider mobile hosts can become temporarily or permanently unavailable at any point in time for some or all of the clients, the availability of the services offered by them cannot be guaranteed. Unavailability can be caused by service providers leaving the network, but also by the development of network partitions. Thus, mechanisms are needed to ensure service availability in such a setting.

The idea to replicate some resource in order to increase its availability is one of the classical concepts of computer science, applied in many applications and systems (e.g., DDBMS, RAIDs, DNS, ...). In fact, in many cases replication is the only feasible solution to ensure availability. Unfortunately, direct mapping of replication concepts into ad hoc networks is not possible as will be explained below.

The ad hoc network topology consists of a varying number of separate network partitions. Most of the related work is based on the prediction of the network partitioning behavior, host disjoins, etc. It assumes that all services offered are vital to the network, and thus aims at trying to deploy a copy (replica) of each of these vital services into developing new network partitions, and then finding solutions to manage the service and its replica concurrently in the different network partitions (Derhab et al., 2005; Derhab and Badache, 2005; Dustdar and Juszczyk, 2007). As mentioned in (Hamdy and könig Ries, 2008) this type

Hamdy M. and König-Ries B. (2008).
AN EXTENDED ANALYSIS OF AN INTEREST-BASED SERVICE DISTRIBUTION PROTOCOL FOR MOBILE AD HOC NETWORKS.
In *Proceedings of the International Conference on Wireless Information Networks and Systems*, pages 203-210
DOI: 10.5220/0002024202030210

of scheme couples the replication decision tightly to a specified lower layer component like the routing component as in (Derhab and Badache, 2005; Dustdar and Juszczyk, 2007). This is a tremendous disadvantage. Instead , our work in (Hamdy and könig Ries, 2008) proposes a protocol for service distribution for ad hoc networks based on the providers' and clients' interest in a specified service. This protocol avoids dealing with the lower layer component and can thus be flexibly used in wide range of settings. Also, it takes into account that not all deployed services will be equally important to all participants at all times. Therefore, we introduce a time-varying importance degree, the vitality of a service for a client. Moreover, how interested a provider is in hosting a service should be proportional to the overall interest in it by the client group of this server. This means that a service that clients and providers are interested in should be replicated, until interest decreases in which case it should be hibernated. Our protocol achieves this behavior by using two measurements (client-service-interest and provider-service-careless) and employing two opposite mechanisms, namely (a) replication mechanism: allows the service to be replicated to a specified client based on a certain client-service-interest level of it and (b) hibernation mechanism: allows a provider to hibernate a service after gaining some certain provider-service-careless or the gross client's interest of that service becomes low.

By maintaining an appropriate number of replicas for the current service interest, the following main advantages are achieved:(a) increasing the availability, (b) avoiding the time and computation intensive operation of network partition prediction and detection and resolving the coupling to the lower network layers , (c) introducing the service interest as a realistic measurement to be used in the replication process besides the hibernation mechanism which represents a realistic behavior of service providers in many cases, (d) introducing the ability of tuning the degree of replication by utilizing the information of the application layer.

The work of (Hamdy and könig Ries, 2008) estimated the performance of the service distribution protocol on just one partition of the network. In this paper we are extending the protocol to be used on the whole set of ad hoc network partitions. The concepts to be evaluated by the current work are: Is the generated number of replicas based on the interest enough to satisfy the whole number of formed partitions? How is the efficiency of the replica placement based on the service interest? In order to evaluate our concepts, we developed our own network model and its related performance measurement. A detailed simulation and further analysis of the results show a good and promising performance.

The rest of this paper is organized as follows: In Section 3, the network model that we use to evaluate our concepts and the proposed protocol is presented. The proposed protocol is presented in Section 4 with its two mechanisms of replication and hibernation. In Section 5, a primary and extended results of an elaborated simulation for our proposed protocol are analyzed. Finally, the related work and conclusions are discussed in Sections 2, and 6 respectively.

2 RELATED WORK

J. Böse et al. introduce in (Böse et al., 2005) an adaptive pull protocol for data dissemination over ad hoc networks. That work estimates the data freshness considering the data load, by comparing their proposed optimistic protocol to others techniques like flooding and combinations of the proposed protocol with flooding feature, they can nearly save 13% of the network load and achieve high freshness rates. S. Moussaoui et al proposed in (Moussaoui et al., 2006) a method for data replication in ad hoc networks with after building the required replicas of available data, starts a recovery stage to overcome the effects of the mobility and ever-changing topology. The work is based on the frequency of accessing based on a moving averages equation (like in (Fu and Cheung, 1994)). Also, Work of T. Hara in (Hara, 2003) is considering the data accessibility frequency to introduce data replicas in many approaches of replicating a specified data item on the whole mobile hosts.

In (Hähner, 2007) J. Hähner introduces a survey over many data consistency models, then introducse full and partial replication algorithms taking into consideration the data consistency based on ordering the observer's graphs.

In (Derhab et al., 2005; Derhab and Badache, 2005) A. Derhab et al., by estimating the link quality and employing a partition prediction mechanisms based on TORA (Park and Corson, 1997) supplies two mechanisms for pull based replication; (a) replication (pre-partition formation) and (b) merging (after two partitions merged) mechanisms. In (Dustdar and Juszczyk, 2007; Juszczyk, 2005) S.Dustdar et al. introduce algorithms that take care of replication and synchronization of services in ad hoc mobile networks. Based on a global view of all network nodes, a replication mechanism (component) by the original service node moderates the replication process per predicted partition, replicas in the new formed partitions are supposed to be hosted by a powerful elected node. The used service model assumes presence of master nodes in order to keep services synchronized. (Hauspie et al., 2001) M.Hauspie et al. (2001), other research like (Wang and Li, 2002) goes also on the same fashion and concepts of link evaluation and availability of global view about the ad hoc network, (Derhab et al., 2005) presented a comparison between these approaches.

3 NETWORK MODEL

The goal of this research is to evaluate the performance of the proposed replication/hibernation mechanism across all network partitions. We use the network model described in (Hamdy and könig Ries, 2008), which models the network at a certain time as an undirected, unweighted graph G(N,E) where N represents the set of uniquely identified nodes, and E is the set of edges representing network connections between nodes. $G_x(N_x,E_x)$ represents one of the network partitions, where: $G(N,E) = G_1(N_1,E_1)\cdots\bigcup G_x(N_x,E_x)\cdots\bigcup G_k(N_k,E_k)$, $N = N_1\cdots\bigcup N_x\cdots\bigcup N_k$, and $E = E_1\cdots\bigcup E_x\cdots\bigcup E_k$. Each of the mobile hosts can cover a fixed range with radius R, a connection is established between two nodes if the distance between them is less than or equal to R. All mobile hosts are placed in a square area. The other components of the network model, namely mobility , service, and calling models are described in the following paragraphs.

Mobility Model. We use the random waypoint mobility model (Lin et al., 2004), in which each mobile node picks a random constant speed uniformly between some preset interval (in our model [1..12]m/s), then generates a random destination location to visit after waiting for a pause time uniformly selected (in our model) between [0..30] minutes. A slight modification was introduced. By introducing the ”mobility index”, which is a percentage [0..100]%, we can change the mobility status of the network. The higher the mobility index, the higher speed value selection is allowed, and lower pause times are generated, and vice versa.

Service Model. To simplify the analysis, the network is maintaining just one service. This service is placed on the first created node in the network. Three assumptions are made in the service model; (a) all mobile nodes can participate in the replication mechanism; (b) the original service is replicable; (c) all participants do not mind to cache the replicas in case of service hibernation. Each replica is described by a requirement index which quantifies the requirements needed to run this service. These values are generated as a normal distribution of about 20% of a general requirement index. The requirement index is a mimic of the reality; normally and even if two providers provide the same service, requirements by each of them to use its service (or get a replica) will differ. Clients are supposed to find the minimum requirement index from the neighboring services to communicate with. This varying of the offered requirement index is one of the responsible components of distributing the interest of the clients among the offered services/replicas in the network. By the work in this paper, since we are relying on a optimistic replication model, the synchronization management of the service/replicas is not considered. Another current ongoing work is addressing the concurrent service synchronization and states.

Calling Model. Initially, all of the created nodes seek for the initial (original) service provider node; only those nodes with at least one feasible path to the provider node are supposed to start evaluating the service calling and be involved in the related replication/hibernation processes. After a while, service/replicas prevalence through the network is supposed to cover as much as possible of the ad hoc formed partitions. Variant calling rates are maintained by each node; the calling rate is generated between [0..4] calls per minute, the calling rate is supposed to be constant during a calling period of {5,10,15} minutes, and after a pause time of {0,5,10,15} minutes, the node is supposed to select another calling rate and so on. Calling rate, calling period, and pause period are uniformly randomly generated.

4 THE APPLIED PROTOCOL

The main players in the protocol components are the two measurements of the client-service-interest and the provider-service-carelessness.For simplicity, currently, both of these measurements are based just on the client calling frequency and the service requirement index. For a client, it will be considered to be ”a replica-interested client” if it achieves a certain number of calls within a specified time interval; then the replication mechanism should start. On the other hand, a provider will be considered as a ”service-careless provider” if it receives no more than one call in a specified time interval; then the hibernation mechanism should start. Our motivation for this research was more to investigate the concepts of replicating the services based on the interest and watching the service prevalence than on establishing a sophisticated computation of the interest itself. Finding more expressive definitions of interest is, however, part of our ongoing work. As in (Hamdy and könig Ries, 2008), the core component actions of the replication and hibernation mechanisms are described below.

Replication Mechanism. The core actions of the replication mechanism are as follows:

- *Restore from Cache.* If a client is interested enough to host a replica, it should search first if it had a replica before, if yes it restores it.
- *Find Least Requirement Service.* If a client is interested enough to host a replica, then, it should discover the replica with the least requirement index.
- *Pass a Replica.* If a client is interested enough to host a replica, it receives a replica from its provider.
- *Switching to the Local Service.* When a replica is received by a client node (new provider), then the node switches its calling to the local replica.
- *Publish.* Allows publishing the new service/replica status.
- *Check the Correctness.* Enforces the interested client to check if it can achieve a certain correctness of replica placement if it receives its own replica.

Hibernation Mechanism. The core actions of the hibernation mechanism are as follows:

- *Shutdown.* Hibernates a local replica.
- *Publish.* Allows publishing the new replica status.
- *Find another Service.* Finds another replica of the called service, if that service is not found.

5 SIMULATION AND DISCUSSIONS

A detailed simulation for the extended application of the proposed protocol of (Hamdy and könig Ries, 2008) has been elaborated. The results are divided mainly into two groups, the first group comes from applying just the replication mechanism (R group), and the other one comes from applying both replication and hibernation mechanisms(R-H group). In our performance analysis, four main performance measurements have been introduced:

- *Service Availability* is the ratio between the time during which at least one replica was available in any of the network partitions to the total running time of the network.
- *Success Ratio*is the ratio between the number of successful service calls to the overall number of calls in the entire network.
- *Service Prevalence*is the ratio between the number of mobile hosts that hosted a replica to the whole number of network participants.
- *Residence Time*is the average time that the replica remained running (not hibernated) on some mobile host.

5.1 Configurations

The mobile hosts are placed in a 500 $meter^2$ area. The transmission range of each node is fixed to 120 meters. The network operation time has been set to be 2 hours per sample run; results are obtained from the average of 20 runs. The replication threshold is set to be 4 calls per minute; the hibernation threshold is set to be 1 call in 5 minutes.
The network size is varying from 10 to 140 nodes. In case of varying the network size, the mobility index is fixed to be 50% and the maximum allowed prevalence is set to be 100%. In case of observing the effects of varying the mobility index, the network size is set to be 50 nodes, and the maximum allowed prevalence is set to be 100%. Finally, in case of varying the maximum allowed service prevalence, the network size is fixed to be 50 nodes, and the mobility index is 50%.

5.2 Basic Performance Analysis

In both of the proposed groups of experiments (R and R-H) the network size, mobility index, and maximum allowed prevalence are varying and the service availability, success ratio, and residence time are observed.

Service Availability. In the (R) group of experiments, by definition, the service is always available in some partition, it might, however, be inaccessible by nodes from other partitions as depicted in Figure 1(a,b,c). For the (R-H) group of experiments, allowing the hibernation process enables the provider nodes to evaluate their carelessness about their offered service, this leads to shutting down replicas by some providers. Figure 1(a) shows, that the higher network sizes the higher service availability. Starting from a moderate network size of 30 nodes, about 71% of service availability is observed.
The effect of the mobility is presented in Figure 1(b): the higher the mobility index, the higher service availability, the reason is that higher speeds and lower pause times result in the service host nodes to traverse

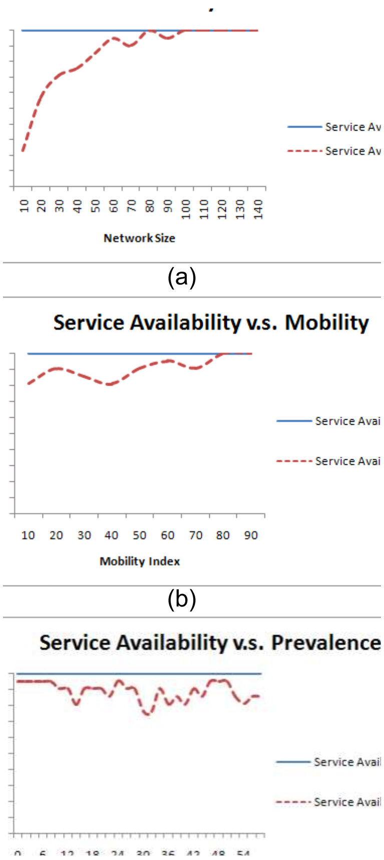

Figure 1: Service Availability.

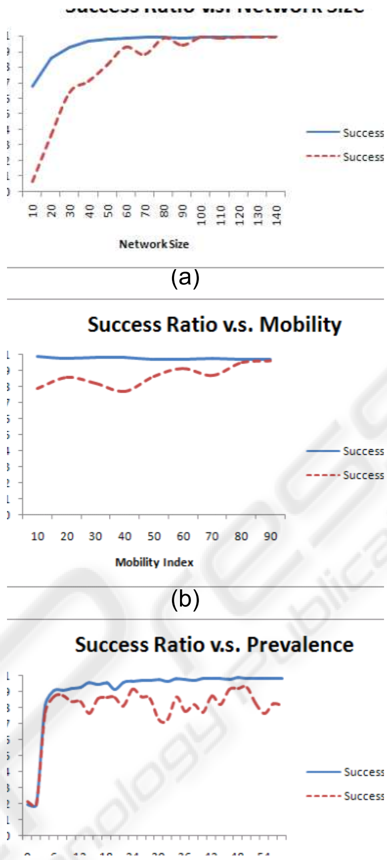

Figure 2: Success Ratio.

more of the network partitions in a shorter time; this traversing enables more pervasive replication in the whole network. The proposed protocol shows achieving high service availability for the lower mobility indices as well, about 80% as a minimum service availability for the minimum mobility index 10%.
In Figure 1(c), The smaller values of the maximum allowed service prevalence lead to achieve higher and steady service availability, these small values of the maximum allowed prevalence concentrate the overall client interest to a few number of provider nodes on the whole network, this interest concentration pushes those provider to keep their services one and of course minimize their provider-service-careless measurements. The higher values of maximum allowed prevalence increase the number of the service providers and of course distribute the gross interest of the client among them, this leads to varying achieved service availability, Despite of that, the minimum achieved service availability lies above 75%.

Success Ratio. The success ratio is much important than the service availability because it indicates the service accessibility form all of the network partitions over the operation time. In Figure 2(a), even applying just the replication mechanism -in (R) group- does not insure 100% success ratio for moderate and low network sizes, it achieves 68% for a very low network density (10 nodes). The difference between success ratios of (R) and (R-H) groups of results is due to the reduced service availability in the (R-H) group. Starting from a network size of 30 nodes, the success ratio is above 70%, and the average difference between the two groups is less than 15%. The effect of varying the mobility index in Figure 2(b) is on the same fashion of the service availability, for the (R) group, the success ratio is constant about 100% because the network size is sufficient to achieve enough interest by the different partitions to maintain at least one replica inside each of them for all values of the mobility index. For the (R-H) group, because the higher speeds and lower pause times enable wider service dissemination on more network partitions, the success ratio increases as the mobility index increases. Figure 2(c) shows the effects of varying the maximum allowed prevalence ratio, both curves are dramatically increasing by slight increments of the maximum allowed prevalence, starting from 6% allowed prevalence the (R)

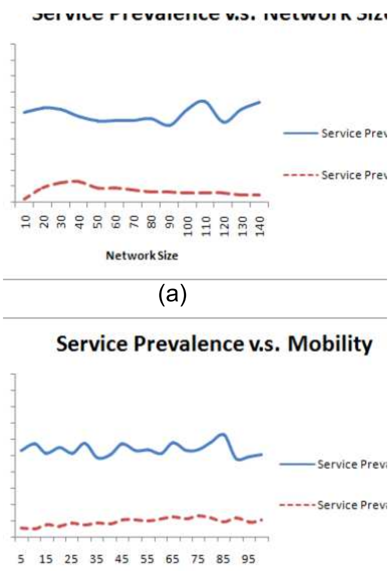

Figure 3: Service Prevalence.

curve is increasing very slowly to close to 100%, on the other hand, the (R-H) curve is hesitating about average value about 83%. The valuable notice here is that there is no need to have high prevalence ratios to achieve high success ratios.

Service Prevalence. Figures 2(a,b) show the effects of both the varying network size and mobility index on the service prevalence. In the (R) group, in Figure 2(a) we can easily deduce that the prevalence ratio is to be about half (55%) of the number of participants which represents a high number of simultaneously running replicas over all network partitions. Also, varying the mobility index (Figure 2(b)) has the same effect, it produces about (55%) service prevalence on the network. On the other hand, in the (R-H) group, by applying the hibernation mechanism, a significant reduction of the service prevalence ratio could be achieved. The reduction value is about(48%), 2(a), in case of varying the network size and about(44%), 2(b), in the case of varying the mobility index. The notable result here, is the effect of the hibernation mechanism in enhancing the many criteria related to the number of the running replicas like minimize the required effort of service/replicas synchronization and of course link's utilization.

Residence Time. Figures 4(a,b) are showing that applying just the replication mechanism of the (R) curves makes the average residence time of the service by the hosting node seem to be constant, the reason is that, the gross interest of the network participants is divided on the same set of providers, since

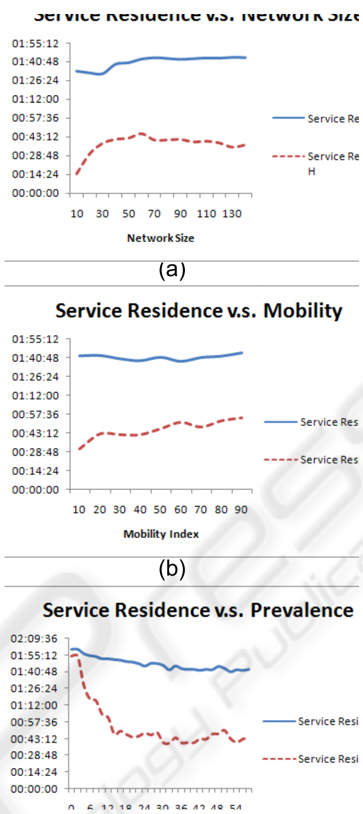

Figure 4: Service Residence Time.

that gross interest comes from the uniform distribution of the calling rates, it is always supposed to have a constant average. On the other hand, in case of applying replication/hibernation mechanisms together, by the (R-H) curves, the sets of the hosting nodes are supposed to be increased, not all of that participants can receive the same client interest portion, so many providers trigger the hibernation mechanism and shut down their service, this behavior makes the average residence time decrease affected by both increasing of the mobility index and the maximum allowed prevalence.

5.3 Extended Performance Analysis and Discussion

In this extended analysis both of the replication/hibernation mechanisms are applied together. With these experiments we measure the correctness of replica placement, and the suitable degree of replication.

Replication Allocation. Obviously, if we can compute the optimum number of ad hoc partitions or clusters for a certain point in time, placing a replica inside each partition will be the minimum required number of replicas to keep service/data available for all participants. Our proposed measure of the correctness of replica placement is linearly based on the partition size and the number of already available replicas in it.

Correctness Ratio *CR***.** In figure 5, the correctness ratio of the placement process is bounded between 0% and 100%, if there is no replicas in the partition the ratio should be 0%, else if there are one or two replicas in the partition the ratio will be 100%, otherwise ,for simplicity, the ratio is linearly inversely proportional to the number of replicas, the ratio becomes zero at a number of replicas equals to the partition size (Pz). Normally, at least one replica per partition is an optimal case. Finding two replicas in the same partition is very healthy from replication point of view.The following equation describes the value of the correctness ratio $CR_t(P_i)$ in an ad hoc formed partition P_i at a certain moment of time t:

$$CR_t(P_i) = \begin{cases} 0 & no(replicas) = 0 \\ 1 & no(replicas) in \{\ 1,2\ \} \\ \frac{Pz - no(replicas)}{Pz-2} & no(replicas) > 2 \end{cases}$$

where $no(replicas)$: the available number of replicas in the partition.

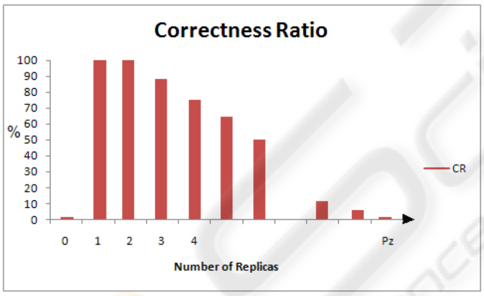

Figure 5: A Linear Ratio of the Allocation Correctness.

In Figure 6(a,b,c), both of the correctness ratio of the replica placement process and the replication degree are investigated against the network size, mobility index, and the maximum allowed service prevalence. Allocation correctness is directly proportional to bnetwork size as in Figure 6(a). By our approach of replica placement (based on the client interest), we can achieve a moderate correctness ratio for moderate network sizes (about 50 nodes). The higher mobility increases the correctness ratio, starting from the value of 50%, for 50 nodes, the correctness ratio is about 35%, as in Figure 6(b). Starting from maximum allowed prevalence ratio equals to 12%, (Figure 6(c)), the achieved allocation correctness is about 35%. So, the allocation correctness is increased as the network

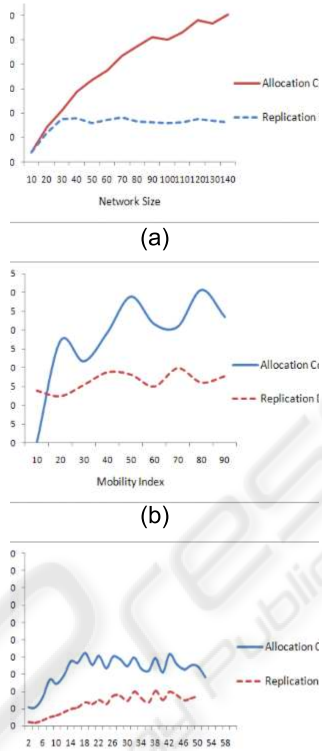

Figure 6: Allocation Correctness and Replication Degree.

size increases, the same holds for both higher mobility and maximum allowed prevalence ratio until certain values are reached, then it becomes steady varying about a certain average.

Degree of Replication. The meaning of the degree of replication here is the average number of replicas in each network partition per that partition size over the network operation time. Generally, the optimum number of replicas is depending on the network connectivity probability. The global connectivity probability is based on the network density and could not be precisely computed (Gianuzzi, 2004). Rather than creating probabilistic models for estimating the connectivity probability (Madsen et al., 2004).

Figures 6(a,b,c) show that the proposed protocol achieves a very low degree of replication, this degree seems to be varying about some average about 15% against each of the varying network size, mobility index, and maximum allowed prevalence.

6 CONCLUSIONS

In this paper we have shown that the proposed protocol of mobile service distribution in (Hamdy and könig Ries, 2008) is applicable on a real ad hoc mobile network simulation considering the formation of several network partitions. Simulation showed that by using interest measurements of a categorized group of clients, a certain number of running replicas could be generated which traversed through the network partitions achieving a high service availability and success ratio, while showing a low service prevalence on the network participants. Moreover, we propose definitions for both correctness of replica placement regarding the ad hoc formed network partitions and the replication degree. The proposed protocol shows promising results for both of these criteria.

REFERENCES

Böse, J.-H., Bregulla, F., Hahn, K., and Scholz, M. (2005). Adaptive data dissemination in mobile ad-hoc networks. *INFORMATIK 2005 - Informatik LIVE!, Band 2, Beitrge der 35. Jahrestagung der Gesellschaft fr Informatik e.V. (GI)*, pages 528–532.

Derhab, A. and Badache, N. (2005). A pull-based service replication protocol in mobile ad hoc networks. *European Transactions on Telecommunications*, 18:1–11.

Derhab, A., Badache, N., and Bouabdallah, A. (2005). A partition prediction algorithm for service replication in mobile ad hoc networks. In *Second Annual Conference on Wireless On-demand Network Systems and Services (WONS 2005)*, Swi.

Dustdar, S. and Juszczyk, L. (2007). Dynamic replication and synchronization of web services for high availability in mobile ad-hoc networks. *Service Oriented Computing and Applications*, 1:19–33.

Fu, A. W.-C. and Cheung, D. W.-L. (1994). A transaction replication scheme for a replicated database with node autonomy. In Bocca, J. B., Jarke, M., and Zaniolo, C., editors, *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, pages 214–225. Morgan Kaufmann.

Gianuzzi, V. (2004). Data replication effectiveness in mobile ad-hoc networks. In *PE-WASUN '04: Proceedings of the 1st ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, pages 17–22, Venezia, Italy. ACM.

Hähner, J. (2007). *Consistent Data Replication in Mobile Ad Hoc Networks*. PhD thesis, Fakult¨at Informatik, Elektrotechnik und Informationstechnik der Universität Stuttgart, Stutgart.

Hamdy, M. and könig Ries, B. (2008). A service distribution protocol for mobile ad hoc networks. In *International Conference on Pervasive Services (ICPS 08)*, Sorrento, Italy.

Hara, T. (2003). Replica allocation methods in ad hoc networks with data update. *Mobile Networks and Applications*, 8:2003.

Hauspie, M., Simplot, D., and Carle, J. (2001). Replication decision algorithm based on link evaluation for services in manet. Technical report, University of Lille, France.

Juszczyk, L. (2005). Replication and synchronization of web services in ad-hoc networks. Master's thesis, Technischen Universität Wien.

Lin, G., Noubir, G., and Rajaraman, R. (2004). Mobility models for ad hoc network simulation. In *The 23rd Conference of the IEEE Communications Society (INFOCOM 2004)*. College of Computer & Information Science Northeastern University Boston.

Madsen, T., Fitzek, F., and Prasad, R. (2004). Simulating mobile ad hoc networks: Estimation of connectivity probability. In *The Seventh International Symposium on Wireless Personal Multimedia Communications (WPMC'04)*, Abano Terme, Italy.

Moussaoui, S., Guerroumi, M., and Badache, N. (2006). *Data Replication In Mobile Ad Hoc Networks*, chapter Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Park, V. D. and Corson, M. S. (1997). A highly adaptive distributed routing algorithm for mobile wireless networks. In *INFOCOM (3)*, pages 1405–1413.

Wang, K. and Li, B. (2002). Efficient and guaranteed service coverage in partitionable mobile ad-hoc networks. In *IEEE INFOCOM02.*, New York,USA.