December 2003

Supporting Group-Oriented Mobile Services

Upkar Varshney
Georgia State University

SUPPORTING GROUP-ORIENTED MOBILE SERVICES

Upkar Varshney
Department of Computer Information Systems
Georgia State University
uvarshney@gsu.edu

Abstract

The emerging mobile services, including mobile commerce, have received considerable interest among researchers, developers, service providers, and users. Some of these mobile services will require group communications among mobile users. An efficient way to support this requirement is by using wireless multicast, where group connectivity among users is maintained and the information exchange is performed by the underlying networking infrastructure. In this paper, multicast requirements of mobile services under varying levels of connectivity are addressed. More specifically, the design and operation of three protocols for supporting multicast transactions is presented. Since these protocols could lead to different response time and transaction completion probabilities, performance evaluation under different connectivity and protocol conditions has been performed. These results can be used to select one or more multicast protocol(s) for group-oriented mobile applications.

Keywords: Mobile services, mobile applications, group communications, wireless multicast, multicast protocols, performance evaluation, analytical modeling, transaction, and connectivity

Introduction

Several group-oriented mobile applications and services such as mobile financial applications, mobile advertising, mobile auction, mobile office, mobile entertainment, and multiparty games have been proposed (Varshney and Vetter 2002). These applications require the support for group communications among mobile users. This can be efficiently supported by using wireless multicast, where the group users are sent messages simultaneously using one to many or many to many connections (Deering 1989). As mobile users move to different locations or leave the group or new users join the group, the membership information is updated for network traffic routing. The work on wireless multicast has been primarily on the routing aspects such as AMRIS (Wu et al. 1998), AMRoute (Liu et al. 1999), and ODMRP (Lee et al. 1999). Performance issues in routing protocols are addressed by Carson and Macker (1999). Although routing is an important aspect of multicast, mobile services will be benefited significantly if a multi-layer approach is employed where application, middleware and networking issues are studied together. Group-oriented mobile services are different from peer-to-peer services, which focus on how to enable sharing of information among users on widely different networking and computing platforms.

The group-oriented mobile services would require continued group communications among mobile users. To support group-oriented mobile services, multicast requirements must be identified and high performance network protocols must be designed and evaluated. In this paper, the following work has been done: (a) identification of the requirements of group-oriented mobile services (b) design of three multicast protocols to support group communications under brief disconnectivity and intermittent connectivity (c) design of an analytical model for evaluating the performance of the proposed multicast protocols. In addition to these important contributions, the paper also presents several new metrics and parameters for measuring the performance. Using the analytical model, performance of the proposed multicast protocols has been evaluated under different connectivity patterns and application requirements. These results show that high transaction completion probability could be achieved by all of the three protocols under some cases. It has also been shown that brief and intermittent connectivity of mobile users can affect transaction completion probability significantly. One interesting result of our research also shows that flexible and adaptable group-oriented mobile services can achieve a higher percentage of transaction completion.

The paper is organized as follows. In section 2, several group-oriented mobile services and their specific requirements are presented. Then in section 3, the group connectivity in mobile and wireless environment is discussed. The design of three multicast protocols is presented in section 4 and the performance evaluation using analytical model is included in section 5. Finally, some concluding remarks are made in section 6.

Group-Oriented Mobile Services

The group-oriented mobile services involve simultaneous communications among mobile users. The mobile services could fit under this category are Mobile Auction, Mobile Games, Mobile Financial Services, Mobile and Locational Advertising, Mobile Entertainment Services, Mobile Distance Education, Proactive Service Management, Product Recommendation Systems, Mobile Inventory Management, and Product Location and Search. Some of these services are illustrated in Figure 1 and more details can be found in (Varshney and Vetter 2002). The focus here is to derive the requirements of these mobile services. These services have been classified into multiple groups based on similar multicast requirements. This classification would reduce the complexity of requirement analysis and will also help in modeling and performance evaluation. The criteria used were the type of communications, number of entities, response time, and connectivity for deriving the requirements of these services as shown in Table 1.

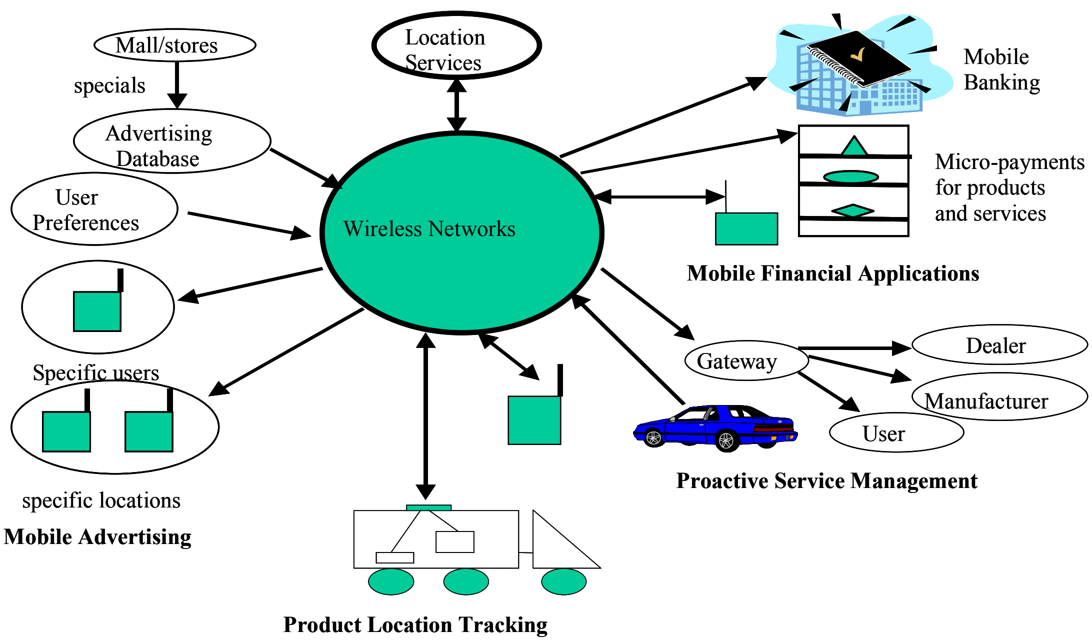

Figure 1. Group-Oriented Mobile Services

Group Membership Management

One major issue in multicast support is membership management. When a group application starts, different participants will have to be located and joined in a group. The information on these participants will have to be maintained in a group list. As some participants leave and others join, the list will be updated. Also the users who move to another location while being a member should be kept in the group list. Due to the user mobility involving leaving or joining a multicast group, or leaving at one point and joining at the other, the amount of membership overhead could become significant. These requirements are difficult to support in wireless networks due to a mixture of point-to-point link as in cellular and PCS networks and broadcast links as in satellite and wireless LANs. Combining this with the limited and variable bandwidth in wireless networks will make the support for multicast even more difficult. The wireless links could also be asymmetric where the bandwidth available in two directions is not equal (or the links could even be unidirectional). Additionally, user mobility and possibly dynamic topology of ad hoc wireless networks would make wireless multicast even more difficult for mobile services.

Table 1. Group-Oriented Mobile Services and Requirements

Group-Oriented Mobile Services	Description	Type of Communication and Number of Entities	Multicast Requirements and Response Time
Mobile Auction, Interactive Games, Financial Services	Users to buy or sell certain items, play multiparty games, conduct multi-party financial transactions	Real-time multicast with active participation from multiple users	Very low delays Continued connectivity required
Mobile and Locational Advertising	Applications turning the wireless devices into a powerful marketing medium	Asymmetric non-real-time multicast involving hundreds or more devices	Higher delays can be tolerated Intermittent connectivity or brief disconnectivity may be tolerated
Mobile Entertain-ment Services/ Mobile Distance Education	entertainment or distance education services to users on per event or subscription basis	Asymmetric Real-time Multicast involving multiple users	Large bandwidth and low delays Intermittent connectivity or brief disconnectivity significantly affects the overall experience.
Proactive Service Management	Providing users the information they will need in very-near-future	Asymmetric non-real-time multicast involving a few entities	Higher delays can be tolerated Intermittent connectivity or brief disconnectivity can be tolerated
Product Recommendation Systems	To receive recommendation on products & services from a 3rd party or customers	Asymmetric non-real-time multicast involving large number of entities	Higher delays can be tolerated Intermittent connectivity or brief dis-connectivity can be tolerated
Mobile Inventory Management/ Product Location	Reducing the inventory by managing in-house and inventory-on-move/to find the location of product and services	Unicast/Multicast involving few entities	Response time of few seconds Intermittent connectivity or brief dis-connectivity increases the delay

Transaction is a unit of work where all steps are necessary in an ordered sequence. A group transaction could involve multiple components as shown in Figure 2. For example, a user wants to initiate a mobile transaction. Such request will be forwarded to a transaction server. Then a multicast session will be initiated with the help of a multicast server. The transaction could also require accessing data center and the services of third party (not shown here). Also the duration of multicast session is dependent on the mobile application. For example, mobile advertising only needs to send one or more messages to users and a group response is not necessary. If any of the users need to purchase something in response to such user-specific or location specific advertising, such transactions could take place in a unicast fashion.

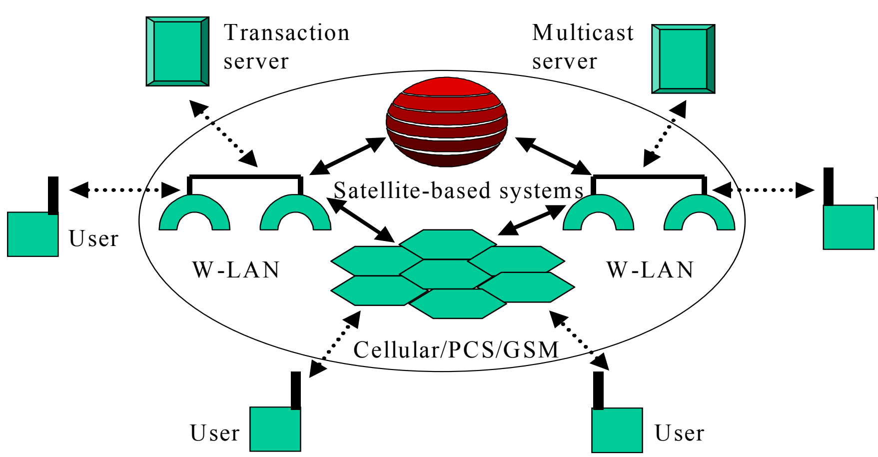

Figure 2. Multicast Architecture for Mobile Services

Although the wireless infrastructure shown in Figure 2 corresponds to a fixed-topology wireless network such as cellular, PCS or GSM network, mobile services could also run in dynamic topology or ad hoc wireless networks where users and servers will be mobile or portable. Such environment is characterized by mobile and/or portable nodes, the lack of the fixed infrastructure support, dynamic topologies, bandwidth-constrained variable capacity links, energy-constrained operation, and limited physical security. In such networks, some (or all) nodes function as routers. As the nodes move, the interconnection pattern among nodes change, so new routes have to be discovered and maintained using routing protocols with substantial overhead/bandwidth cost. The multicast routing protocols designed for infrastructure-based networks can not be used in ad hoc networks for three reasons: (a) the routes in ad hoc networks change frequently, (b) the fixed infrastructure does not exist for such route (re) computation, and, (c) there are power, coverage and bandwidth restrictions (possible unidirectional links or links with unequal bandwidth in two directions). If such protocols were to be used in ad hoc wireless networks the amount of processing and storage overhead could exceed limited capabilities of nodes. IETF's Working Group on Mobile Adhoc Networks (MANET) suggests desirable properties of multicast routing in ad hoc networks including distributed operation, loop free routing, security to support sensitive applications, possible support for unidirectional links, on-demand (reactive) route discovery, and proactive operation if resources allow. More details of multicast routing are discussed by Carson and Macker (1999).

In such an environment, one of more nodes (devices or servers) could be dynamically selected to act as group leaders based on their ability to locate one or more required servers. Since users, devices, networks, and servers could all be mobile, completing a series of transactions using a multicast session before one node moves out of the range will become a challenging task. Fortunately, many of the mobile commerce transactions are of short duration, thus could be completed before a user moves to another location. It is possible to use (slow) mobility level in selecting group leaders and servers for mobile transactions. Such nodes could perform various m-commerce functions such as support for transaction, user preferences and pricing. For additional functions, one or more nodes would require connectivity to other servers using one or more additional wireless networks such as satellite or overlapped cellular networks. The users should also have the ability to discover server and available services that could be used directly or with the help of other users/networks. It is also possible that mobile commerce applications would run in a combination of infrastructure and ad hoc topology wireless networks.

Protocols for Supporting Group Communications

The user mobility combined with the link characteristics makes the continuous communications among group users difficult. This would affect the performance and outcome of group-oriented mobile services such as mobile auction, games, and financial services where user participation is important. No response from a user for some time could lead to some challenges. One is to determine if the user has left the group or is experiencing a brief dis-connectivity. If there is a reason to believe that user is experiencing connectivity problems, then the state of the application could be maintained until a time-out occurs or the user sends some information again. On the other hand, if the user response could not have affected the outcome for the group application, then it would be wasteful for everyone in the group to wait for this user. It is possible that such determination could be difficult for all possible states of mobile applications.

One major issue is what is a reasonable time to wait for response from mobile users. As time-out is a possible technique for membership management (remove a user from the group application if do not hear anything for a given time) in such applications, it is not clear what should be the value of time-outs. The amount of wait could affect the performance of group-oriented mobile services. If a small value is chosen, it will lead to frequent removal of mobile users from the group when experiencing a brief dis-connectivity. Later on, these users could attempt to connect to the group but since the state of application is likely to have changed significantly (such as bidding in a mobile auction), these users have lost the chance to affect the outcome of the group application. On the other side, if a higher value of time-out is chosen, then all the connected users are forced to wait for these users who could have been permanently disconnected. This will also reduce the performance (speed) of mobile services for most users. To avoid these extremes, finding an optimal or near optimal time-out is necessary for achieving a high level of performance from mobile applications for most users.

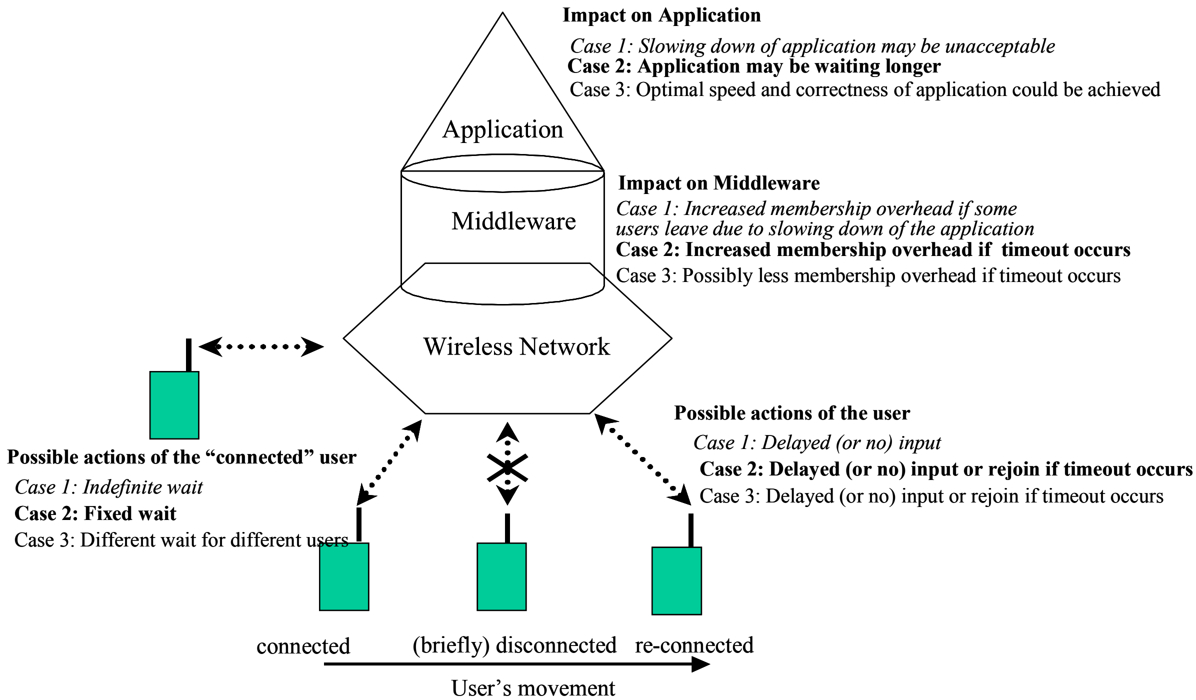

Figure 3. The Intermittent Connectivity and Mobile Services

A possible scenario is shown in Figure 3, where impact of intermittent and brief dis-connectivity on different layers of a wireless network is illustrated (Varshney, 2002). In such a scenario, three different solutions are possible. In the first case, the group is forced to wait until the user sends an input. The second case involves waiting only for some time as determined by previous values and/or using information on brief dis-connectivity. In the third case, the group waits for different time based on the type of user: passive listener, active listener, and active participant. For the first type, the user can be removed from the group and can join later without affecting the status of the group. For the second type, the group can wait for some time, however for the third type, the group should wait longer as the input from such users can affect the outcome of the group application. Then a trade-off can be done between overhead due to the continued information transmission to users who may no longer be members (larger time-outs are used) to the overhead of join operation later (if smaller time outs are used). These three different schemes should be compared under different mobility patterns for various mobile applications. It is possible that applications that require larger response time would work better with first scheme due to its simplicity and less membership overhead, although in its worst case it has no upper bound on delays caused by certain users. More sophisticated mobile applications with low delay requirements could benefit from second or third schemes. The choice between these two could be made by the tolerable amount of processing overhead. We have designed three protocols for group communications in mobile services (Table 2). These will be compared under different users, transactions, and network conditions. The mobile services that could tolerate larger response time would work better with first protocol due to its simplicity and less overhead, although it has no upper bound on delays. The mobile services requiring lower delays will benefit from second or third protocol. The overall multicast process is shown in Figure 4.

Table 2. Steps of Three Protocols

Protocol	Description	Comments
1: all users equal & unlimited waiting	If no input from a user, wait until the user responds or sends a leave message	Simplest to implement Likely to work well for transactions with tolerance for higher delays Likely to have the least affect of brief-disconnectivity or intermittent connectivity (highest probability of transaction completion)
2: all users equal & limited waiting	Step 1: If no input from a user, wait only for the predetermined time Step 2: If transaction needs the user, block the group transaction	Simple to implement Likely to work well for transactions with lower delay requirement Likely to have the effect of brief-disconnectivity or intermittent connectivity (reasonable probability of transaction completion)
3: Status-based waiting	Step 1: If no input from a user, wait for certain time based on the type of user Step 2: If transaction needs the user, block the group transaction	More complex to implement Likely to work well for transactions with lower delay requirement Could lead to a more optimal combination of response time and transaction completion probability under intermittent connectivity

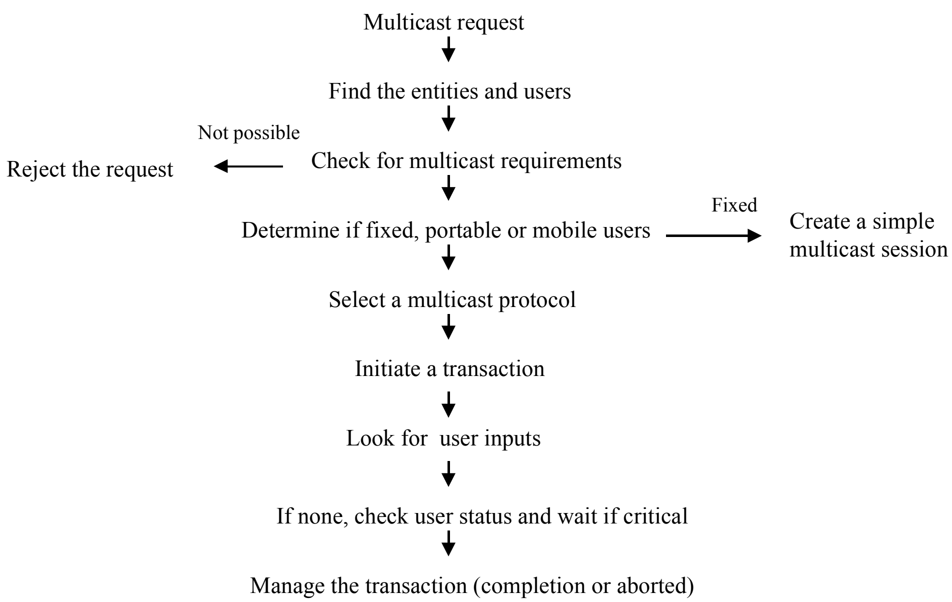

Figure 4. A Step-by-Step Description of Multicast Process

Performance Evaluation of Protocols

To evaluate the performance of three multicast protocols, several new parameters for group-oriented mobile services can be defined. The first is Transaction to Mobility Product (TMP), the product of average transaction length and mobility (=avg. length of dis-connection X frequency). Some transactions could be very long, leading to a higher TMP. If users experience longer or more frequent connectivity interruptions, then TMP would also be high. From a performance point-of-view, lower the TMP higher is the chance of transaction completion. Mobile Member Ratio (MMR) defines the fraction of users who are highly mobile or experience brief dis-connectivity. This could vary over a range and is affected by both user mobility and characteristics of wireless links. The effect of high MMR would be more on services that require real-time multicast (low delays). Transaction to Session-length Ratio (TSR) defines the number of transactions in a unit of time during a multicast session. Source to Members Ratio (SMR) defines the fraction of users in a session that could act as source of multicast transaction. This would be dependent on the type of mobile services and a high value will affect the performance negatively, as more sources will increase the frequency of larger wait under connectivity problems. Entity to Transaction Ratio (ETR) defines the number of entities involved in a transaction. A higher value of ETR is likely to increase the wait and thus would result in an increased response time or lower completion probability for transactions. The output parameters to measure are transaction completion probability, transaction response time, and the probability of transaction completion under brief-disconnectivity and intermittent connectivity. The variable parameters are number of users, number of multicast sessions, number of users in a session, number of transactions in a session, the frequency and duration of disconnectivity periods. To compare the performance of three protocols, we have developed an analytical model. The model currently computes transaction completion probabilities for each of the protocol under different number of users, connectivity and dis-connectivity duration, and number of users that must remain connected for a transaction to complete. The model is currently being extended to include transaction response time for three protocols.

A transaction could be completed if M out of N users are connected (for some transactions M has to be equal to N). The probability of transaction completion can be given as

$$P_T = \sum_{i \geq M}^{N} {}^{N}C_i * (P_c)^i * (1- P_c)^{N-i}$$

here P_c is the probability that a user remains connected

$_c$ is given by

$$(D_{on}/(D_{on}+D_{off})) + (1/ N_{off}) \sum_{i=1}^{Noff} Prob(T_o \geq D_{off\text{-}i})(D_{off}/(D_{on}+D_{off}))$$

where Don and Doff are the average time spent in the connected and disconnected stages, respectively. Non and Noff are the number of times users were in connected and disconnected states. Doff-i is the time spent in the ith disconnection and To is the time-out before a transaction is blocked. The value of To is dependent on the protocol used.

The active state of group can be given by

$$G = 1- (\sum_{i=1}^{Noff} \min(To, Doff\text{-}i)/(\sum_{i=1}^{Noff} Doff\text{-}i + \sum_{j=1}^{Non} Don\text{-}j))$$

Using these equations, the performance of three proposed protocols is derived. The results are expressed in terms of transaction completion probabilities. Figure 5 shows that connectivity and dis-connectivity pattern does not affect protocol 1 as it allows for unlimited waiting time for all users. However, the impact is seen for protocol 2 and 3, which perform better as connectivity is improved. As protocol 3 reduces waiting time further, it leads to lower transaction completion probability. However, this results in a better transaction response time. The impact of minimum number of users that must remain connected for a transaction to get completed is shown in Figure 6. As the number is reduced, the performance is very high for all protocols. When very strict connectivity is required by application for completing a transaction, protocol 3 performs poorly as it would timeout faster, thus leading to a reduced transaction completion probability.

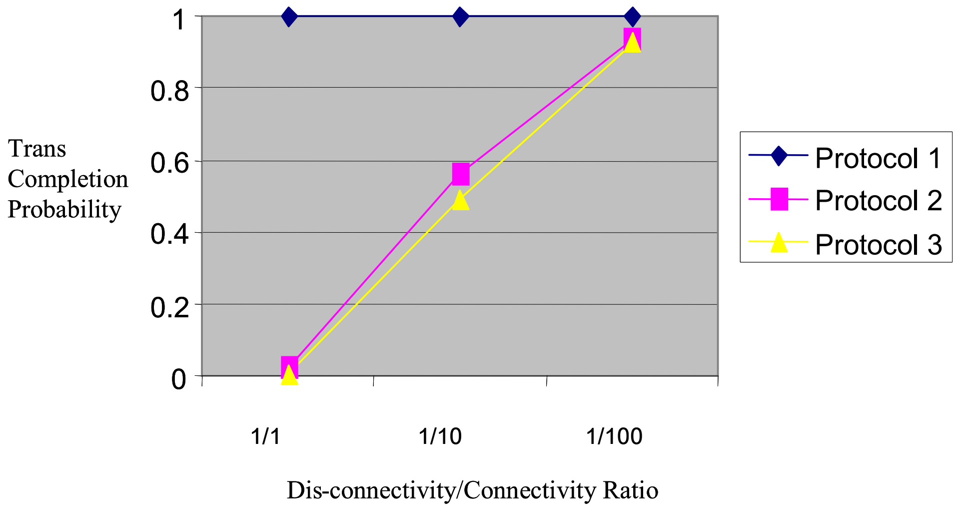

Figure 5. Transaction Completion Probability Under Vvarying Levels of Connectivity

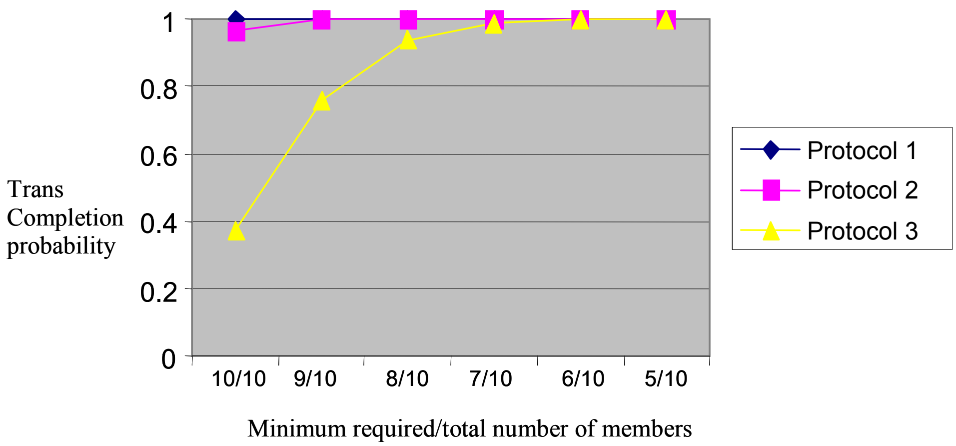

Figure 6. Transaction Completion Probability under Different Application Requirement

Conclusions and Future Scope

The group communications requirement of the emerging group-oriented mobile services can be supported by using wireless multicast. In this paper, three multicast protocols are presented and discussed. The protocols are designed to support diverse requirements of group-oriented mobile services. One of the major requirements is a high completion probability for multicast transactions, especially when mobile users experience brief dis-connectivity and/or intermittent connectivity. These protocols allow the applications to handle such events gracefully. The results presented here showed that performance of these three protocols depends on the connectivity pattern and the application requirements. It is shown that high completion probability could be achieved by all three protocols under different conditions. It is also found that group applications with higher flexibility would be awarded with significantly higher completion probability by all the three protocols. The further research in this area could include evaluating transaction response time and how to achieve an optimal combination of response time and transaction completion probability for different mobile services.

References

Carson, S. and Macker, J. (1999) "Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", IETF RFC 2501, Work in Progress.

Deering, S. (1989) "Host Extension for IP Multicasting", Internet RFC 1112, August

Lee, S., Su, W., and Gerla, M. (1999) "On-Demand Multicast Routing Protocol for Ad Hoc Networks", Internet-Draft, MANET Working Group, Work in Progress

Liu, M., Talpade, R., McAuley, A., and Bommaiah, E. (1999) "AMRoute: Adhoc Multicast Routing Protocol", Technical Research Report (CSHCN T.R. 99-1), Center for Satellite and Hybrid Communication Networks, University of Maryland (www.isr.umd.edu/CSHCN)

Varshney, U. and Vetter R. (2002) "Framework, Applications, and Networking Support for M-commerce", ACM/Kluwer Journal on Mobile Network and Applications (MONET) (7:3), June, pp. 185-198.

Varshney, U. (2002) "Multicast over Wireless Networks", Communications of the ACM, (45:12), December, pp.

Wu, C. Tay, Y. and Toh, C.-K. (1998) "Ad Hoc Multicast Routing Protocol Utilizing Increasing Id-numberS (AMRIS)", Internet-Draft, MANET Working Group, Work in Progress.