Computing and Informatics, Vol. 30, 2011, 429–446

ON IMPROVING THE ROBUSTNESS OF PARTITIONABLE INTERNET-BASED MOBILE AD HOC NETWORKS

Sunho Lim

Department of Computer Science
Texas Tech University
Lubbock, TX 79409, U.S.A.
e-mail: sunho.lim@ttu.edu

Soo-Hoan Chae

School of Electronics, Telecommunication, and Computer Engineering
Korea Aerospace University
Geonggido, South Korea
e-mail: chae@kau.ac.kr

Abstract. Recent technological advances in portability, mobility support, and high speed wireless communications and users' insatiable interest in accessing the Internet have fueled to development of mobile wireless networks. Internet-based mobile ad hoc network (IMANET) is emerging as a ubiquitous communication infrastructure that combines a mobile ad hoc network (MANET) and the Internet to provide universal information accessibility. However, communication performance may be seriously degraded by network partitions resulted from frequent changes of the network topology. In this paper, we propose an enhanced least recently used replacement policy as a part of the aggregate cache mechanism to improve the information accessibility and reduce the access latency in the presence of network partitioning. The enhanced aggregate cache is analyzed and also evaluated by simulation. Extensive simulation experiments are conducted under various network topologies by using three different mobility models: random waypoint, Manhattan grid, and modified random waypoint. The simulation results indicate that the proposed policy significantly improves communication performance in varying network topologies, and relieves the network partition problem to a great extent.

Keywords: Aggregate cache, internet based mobile ad hoc network, network partition

1 INTRODUCTION

With the recent advent in wireless technologies and mobile devices, wireless networks have become a ubiquitous communication infrastructure. In addition, growing interest in accessing the wired network or Internet has fueled the development of mobile wireless networks, which can be used in many realistic applications. We envisage that in the near future, users will be able to access the Internet services and information anytime and anywhere.

To realize this vision, we consider an *Internet-based* MANET, called IMANET, which is an evolving communication infrastructure, that combines the wired Internet and wireless mobile ad hoc networks [1]. In an IMANET, a mobile terminal (MT) can not only connect to the Internet but also forward a message for communication with other MTs via wireless LANs (e.g. IEEE 802.11), as used in most prior studies [2, 3, 4, 5, 6, 7, 8, 9, 10]. An MT can cache a limited number of data items. The size of the cache space in an MT is relatively small compared to the total number of data items in a server/database, which is accessible via the access point (AP) to a wired network. The IMANET is gaining a lot of attention since it is quite relevant to realistic Internet applications in terms of flexible accessibility and information availability. For instance, in museums or shopping malls, users may need to access various types of information provided by local service provider through an electronic guide such as an info-station [11]. Users can form an ad hoc network and share information directly or indirectly by accessing the info-station through selected access points to the Internet.

However, network partitioning is one of the main challenges in IMANETs. An MT may incur frequent wireless link disconnections because of the limited communication range. As the MT moves around, the network topology can change dynamically. Thus, the network is likely to be partitioned into several isolated subnetworks. In an IMANET, network partitioning affects the information accessibility and access latency because it prevents the MTs from accessing the AP. For instance, an MT can be completely out of the communication range of an AP or other MTs. Also, a set of MTs can be isolated from an AP or other MTs, and thus their requests can not be processed by an AP or any other isolated MTs (see Figure 1). Although connectivity can be restored by adjusting the MT's transmission power [12] or by modifying the MT's mobile trajectory [13], we do not consider these techniques in this paper for the sake of simplicity.

In this paper, we investigate an aggregate caching technique to relieve the network partitioning. Since caching frequently accessed data items on the MTs is an effective technique to improve the communication performance, we exploit a performance tradeoff between data accessibility and access latency in data item duplica-

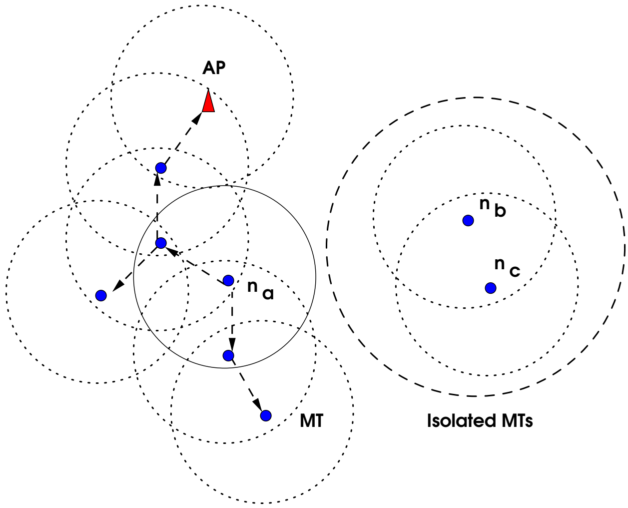


Fig. 1. An MT (n_a) broadcasts a request packet for information access which is forwarded to the AP and other connected MTs except isolated MTs (n_b and n_c) in the IMANET. Requests from the isolated MTs cannot be answered.

tion in the presence of network partition in an IMANET. The rationale behind this caching approach is that even isolated MTs can still access the data items from their neighbor. Although frequent change of network topology affects the communication performance, little work has been reported for analyzing the impact of network partitioning on the communication performance. To our knowledge, none of the prior work has considered the impact of various mobility patterns that may lead to network partitioning in an IMANET through caching. The main contributions of the paper are twofold:

- We propose an enhanced least recently used replacement policy, called AC_LRU, as part of the aggregate cache management to improve information accessibility in a dynamically changing environment, where the MTs could be isolated due to several reasons. Also the enhanced aggregate cache is analyzed and also evaluated by simulation.
- We analyze the performance of the proposed methodology using three different mobility models. These are random waypoint mobility, Manhattan grid mobility, and modified random waypoint mobility. These mobility patterns exhibit different network partitioning to evaluate the robustness of the proposed caching scheme.

We conduct a simulation-based performance evaluation to observe the impact of caching, cache management, and mobility patterns. The overall results show that the proposed caching methodology can relieve network partitioning and improve system performance significantly in IMANETs.

The rest of this paper is organized as follows. The related work is carefully reviewed in Section 2. The aggregate cache management mechanism is presented in Section 3. Section 4 is devoted to performance evaluation and comparisons of various policies under different mobility models. Finally, we conclude the paper with future directions in Section 5.

2 BACKGROUND AND RELATED WORK

In this section, we describe the system model of partitionable IMANET and review the caching techniques applied in wired and wireless networks.

2.1 Partitionable Internet-Based Mobile Ad Hoc Network

IMANET is primarily aimed at increasing both connectivity and accessibility of MTs to the wired Internet. As shown in Figure 2, an IMANET consists of a set of MTs that can communicate with each other using an ad hoc communication protocol (illustrated by the dash-lines). As an MT moves based on a given mobility model, the network topology changes dynamically resulting in several isolated subnetworks. Thus, a set of MTs can be isolated from an AP or other MTs and their data access requests cannot be answered. In contrast to traditional MANET where MTs only can communicate among themselves, some of MTs can directly connect to the Internet and serve as APs for the rest of MTs in the IMANET.

An AP[1] is located in a communication area, is connected to a database server[2], plays a role as a gateway to the Internet, and is assumed to have access to any information. A database may be attached to an AP, a fixed router or a database server. We assume that the database consists of a set of data items in which a data item is the basic unit of an update or query operation. Also the database is only updated by the server, while a query is generated by the MTs for read-only requests. Due to limited storage space, an MT can cache only a small number of data items in the database.

2.2 Related Work

In most prior studies, replicating and caching schemes are deployed to increase the information availability, while reducing the access latency in wired and wireless networks.

Data replication and optimal relocation schemes have been widely researched in the content of distributed database systems. While replication is effective in

[1] An AP here is a logical notation. An AP equipped with appropriate antennas can directly communicate with the Internet through wireless infrastructures including cellular base stations (BSs), Low Earth Orbit (LEO), or geostationary (GEO) satellites.

[2] From an MT's point of view, since an AP is transparent to a database server, we use the terms AP and database server (later in short, server) interchangeably.

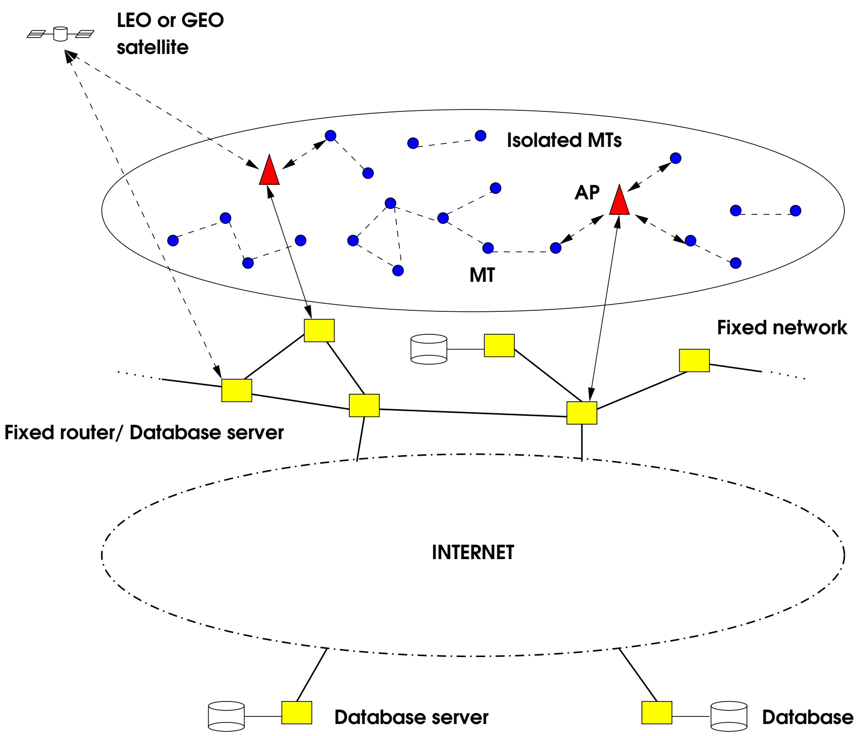


Fig. 2. A system model of partitionable IMANET

addressing the network partitioning problem that is caused by a link failure or system failure, maintaining consistency among replicated data items is a major problem. In other words, data availability and consistency are conflicting goals, which are difficult to optimize. Several techniques to relocate replicas of small data sets have been suggested [14, 15]. Also, a database may be dynamically relocated to adapt to system changes, called database migration [16], to improve overall system performance in broadband networks. However, all these works are carried out in fixed wired networks.

A number of studies have been conducted to reduce the Web traffic and overall network congestion by deploying various caching schemes [17, 18, 19] in the Internet. A cooperative caching [17] is suggested, in which a couple of individual caches are treated as a unified cache and they interact among themselves to eliminate duplicate copies and increase cache utilization. Fan et al. [18] proposed a summary cache, where the proxies share their summary of cache contents represented by bloom filters. When a proxy has a cache miss for a request, it sends the request to other proxies based on the periodically updated summaries of cache contents in other proxies. In [20], a proxy cache relocation scheme based on the prediction of user's mobility is proposed to reduce delay during a handoff in cellular networks. Cao and

Das proposed a counter based cache invalidation scheme [21] using the updated invalidation report (UIR) mechanism [22] in the presence of server failure, MTs failure, and disconnection in cellular networks. In this scheme, a server maintains a counter to identify hot data items and adjusts updated data items' broadcast. It efficiently utilizes the broadcast bandwidth and reduces the query latency. IR broadcast error due to time-varying link quality was further considered in [23]. However, none of these techniques can be directly applied to a mobile environment, where an MT is multi-hops away from an access point.

In particular in MANETs, it is important to cache frequently accessed data items not only to reduce the average latency, but also to save wireless bandwidth. Hara [24] proposed a replica allocation method to increase information availability in MANETs. In this scheme, an MT maintains a limited number of duplicated data items if they are frequently requested. Replicated data items are relocated periodically at every relocation period based on the following: each MT's access frequency, the neighbor MTs' access frequency, or overall network topology. Occurrence of updates on a data item is further considered in [25]. Since an MT cannot access anything when it is isolated from others, replication is an effective means to improve information availability. Due to limited size of information that an MT can hold, simply replicating data items and accessing among them in a MANET cannot fulfill users' requirements to access a wide variety of information database, which is usually available in the Internet and not in MT itself. Unlike our approach, these schemes focus on data dissemination and its reallocation, and thus a network partition is not considered.

Li and Rus [13] proposed an algorithm in which an MT actively modifies its trajectory to transmit a message instead of waiting until the network topology is changed. This guarantees minimal message transmission time. In [26], a data dissemination strategy has been suggested in the presence of network partitions in MANETs, where a subset of MTs (quorum) is selected as a server. Since not all the servers in the quorum are reachable from the MTs, several heuristics are suggested to make update/query operations efficient. However, they do not consider cache management including cache admission control and cache replacement. A limited research effort [6, 7, 8, 10] has been made in cache invalidation and its consistency issues in IMANET.

To summarize, to the best of our knowledge, none of previous work has touched an aggregated caching mechanism for partitionable networks in the realm of IMANETs.

3 A CACHE MANAGEMENT FOR PARTITIONABLE NETWORKS

In IMANET, caching data items in the local cache helps in reducing latency and increasing accessibility. If an MT is located along the path in which the request packet travels to an AP, and has the requested data item in its cache, then it can serve the request without forwarding it to the AP. In the absence of caching, all the

MTs' request should be forwarded to the appropriate APs. Since the local caches of the MTs virtually form an aggregate cache, a decision as to whether to cache the data item depends not only on the MT itself, but also on the neighboring MTs.

In addition, a time-varying network topology may lead to several isolated subnetworks from the APs. Network partitioning affects both accessibility and latency, because it prevents the MTs from accessing the APs. Although caching helps in improving performance, too much duplication of the same data items due to aggressive caching is a potential problem in the caching scheme. Since duplication reduces data availability, a controlled duplication is essential to optimize data availability in partitionable networks.

To address this network partitioning constraint, we proposed an *aggregate caching* approach in [6] for IMANETs. The basic idea is to store data items in the local caches of the MTs , such that members of the IMANET can still obtain the requested information even if the network is partitioned. Thus, the aggregated local caches of the MTs can be considered as a unified large cache for the IMANET. If an MT is isolated from the Internet, it can search other reachable MTs for the requested data item. Also, if an MT is located far away from the Internet, it may request the data items from other closely located MTs to reduce latency. In addition, we use a broadcast-based Simple Search (SS) algorithm [6] to locate the requested data item. A four-way handshake is implemented to reduce network congestion and bandwidth consumption by multiple replied data items for a single query. In IMANET, requested data items can be received from local cache of the MTs as well as via an AP connected to the Internet.

In the following, we describe our cache management and its analysis for managing the aggregate cache.

3.1 Cache Management

In this paper, we propose an enhanced least recently used (LRU) replacement policy by combining the cache admission control and the traditional LRU policy, which is called *LRU with admission control (AC_LRU)*. When an MT receives the requested data, it triggers a cache admission control mechanism to decide whether the data item should be admitted to the cache. The admission control allows an MT to cache a data item based on the distance (measured by hop) to other APs or MTs, which have the requested data. If the MT is located within Γ hops from them, then it does not cache the data item; otherwise it caches the data item. Therefore, the same data items are cached at least Γ hops apart. Here, Γ is a system parameter. Also the MT victimizes the least accessed data item, when the cache is full.

The primary idea is that, in order to increase accessibility, we try to cache as many data items as possible, while trying to avoid too much duplication. There is a tradeoff between access latency and data accessibility in data item duplication. If the popular data items are duplicated a lot (less), then the latency to average access is reduced (increased) because there is a high (less) probability of finding those data items in another closer MT. With high duplication, however, the number of distinct data items in the aggregate cache is less. Thus, the probability of finding less popular data items from other MTs becomes low. Even though the number of copies of popular data items reduces due to limited duplication, they still are accessible from less closer MTs with a longer delay.

Although caching popular data items aggressively in closer MTs helps in reducing the latency, in this work we give more weight to the data accessibility than to access latency. A rationale behind this is that it is meaningless to reduce access latency when a set of MTs is isolated from other MTs or the AP, and they can not access any interested data items. Instead of waiting until the network topology changes, it is better for the MTs to have a high probability of finding the requested data items. Since the Γ parameter enables more distinct data items to be distributed over the entire cache, more data items can be accessed and thus the overall data accessibility is increased.

3.2 Analysis

We analyze the aggregate cache (AC) and compare its performance with no cache (NC) case in the IMANET. Table 1 lists the notations used in the analysis. First, in no cache case, P_{NC} is a probability of a *request* packet to be delivered to an AP. Thus, we have

$$\begin{aligned} P_{NC} &= \overbrace{P_e \cdot \prod_{i=1}^{n}(1-P_{ap}) \cdot \ldots \cdot P_e \cdot \prod_{i=1}^{n}(1-P_{ap})}^{d-1} \cdot P_e \cdot P_{ap} \cdot \prod_{i=1}^{n-1}(1-P_{ap}) \\ &= (P_e)^d \cdot \left\{ \prod_{j=1}^{d-1} \prod_{i=1}^{n} (1-P_{ap}) \right\} \cdot P_{ap} \cdot \prod_{i=1}^{n-1}(1-P_{ap}) \qquad (1) \end{aligned}$$

where $P_{ap} = \frac{N_{ap}}{N_{mt}+N_{ap}}$.

When an MT is isolated from other MTs or AP, which means P_e is zero, it fails to receive a requested data item and P_{NC} is zero. In IMANET, N_{ap} is relatively smaller than N_{mt}. In this paper, since we have deployed a single AP, N_{ap} is one. However, in the aggregate cache case, P_{AC} is a probability of a *request* packet to be delivered to the MTs, which are located along the path to which the packet is relayed to an AP, and cached requested data item. Thus, we have

$$\begin{aligned} P_{AC} &= \overbrace{P_e \cdot \prod_{i=1}^{n}(1-Q_c) \cdot \ldots \cdot P_e \cdot \prod_{i=1}^{n}(1-Q_c)}^{d'-1} \cdot P_e \cdot \left\{ Q_c + (1 - \prod_{i=1}^{n}(1-Q_c)) \right\} \\ &= (P_e)^{d'} \cdot \left\{ \prod_{j=1}^{d'-1} \prod_{i=1}^{n} (1-Q_c) \right\} \cdot \left\{ Q_c + (1 - \prod_{i=1}^{n}(1-Q_c)) \right\} \qquad (2) \end{aligned}$$

where $Q_c = \begin{cases} P_{ap} & \text{if } P_c = 0 \\ P_c & \text{otherwise.} \end{cases}$

When a requested data item is not found in any MTs but AP, which means P_c is zero, P_{AC} is the same as P_{NC}. When a network is partitioned, an MT may not access to AP but a set of adjacent MTs, which are out of communication range of AP either. Unlike to no cache case in which P_{NC} is zero, an MT can access the local caches of adjacent MTs, and P_c is greater than or equal to zero. Thus,

$$P_{NC} \leq P_{AC}. \tag{3}$$

Notation	Meaning
P_{NC}	The probability that a requested data item is received in no cache case
P_{AC}	The probability that a requested data item is received in aggregate cache case
P_e	The probability that there is a node within a communication range
P_c	The probability that a requested data item is in local cache
P_{ap}	The probability that a node is an AP
L_{NC}	The average number of hops in no cache case
L_{AC}	The average number of hops in aggregate cache case
N_D	The total number of data item
N_{mt}	Total number of MTs in the area
N_{ap}	Total number of APs in the area
n	The average number of MTs which are located within a communication range
d	The average number of hops to an AP
d'	The average number of hops to the MTs that cache requested data item

Table 1. Notations

Next, L_{NC} is a number of hops for a *request* packet to be delivered to an AP, and we have

$$\begin{aligned} L_{NC} &= P_e \cdot (1 - P_{ap}) \cdot P_{ap} \cdot 1 + (P_e)^2 \cdot (1 - P_{ap})^2 \cdot P_{ap} \cdot 2 + \dots \\ &\quad + (P_e)^{d-1} \cdot (1 - P_{ap})^{d-1} \cdot P_{ap} \cdot (d-1) + (P_e)^d \cdot (1 - P_{ap})^d \cdot P_{ap} \cdot d \\ &= P_{ap} \cdot \sum_{i=1}^{d} (1 - P_{ap})^i \cdot (P_e)^i \cdot i. \end{aligned} \tag{4}$$

L_{NC} is greater than or equal to one, because an MT is located at least one hop apart from an AP. Similarly, we have

$$\begin{aligned} L_{AC} &= P_e \cdot (1 - P_c) \cdot P_c \cdot 1 + (P_e)^2 \cdot (1 - P_c)^2 \cdot P_c \cdot 2 + \dots \\ &\quad + (P_e)^{d'-1} \cdot (1 - P_c)^{d'-1} \cdot P_c \cdot (d'-1) + (P_e)^{d'} \cdot (1 - P_c)^{d'} \cdot P_c \cdot d' \\ &= P_c \cdot \sum_{i=1}^{d'} (1 - P_c)^i \cdot (P_e)^i \cdot i. \end{aligned} \tag{5}$$

In the aggregate cache case, when an MT caches requested data item, P_c and L_{AC} is 1 and 0, respectively. When no MTs caches requested data item but AP, which is the worst case, L_{AC} is the same as L_{NC}. Since a request data item can be obtained in any near located MTs or AP, Equation (5) can be expressed as follows:

$$L_{AC} = P_c \cdot \sum_{i=1}^{D} (1 - P_c)^i \cdot (P_e)^i \cdot i. \tag{6}$$

Here, D is the minimum number of hops between d and d', $D = \min(d, d')$. Thus,

$$L_{AC} \leq L_{NC}. \tag{7}$$

In summary, an aggregate cache enhances the communication performance in terms of increasing the accessibility by utilizing adjacent MTs' caches, and reducing a delay by minimizing the number of hops. Since the proposed model is simple and intended only to estimate the performance trend, we conduct detailed performance evaluation through extensive simulation in the following section.

4 PERFORMANCE EVALUATION

4.1 Simulation Testbed

In order to evaluate the efficiency of the proposed scheme, we have developed a event-driven simulator using CSIM [27]. We use a $3\,000 \times 1\,500$ (m^2) rectangle wrap around network to examine the proposed idea. We assume that an AP is located in the center of an area. The 50 MTs are randomly located in the network. An MT has 30 caching slots. The request arrival pattern follows the Poisson distribution with a rate of λ. The inter request time is 30 seconds. The speed (s) of the MTs is uniformly distributed in the range ($0.0 < s \leq 2.0$ m/sec), and the pause time is 200 seconds. The total number of data items is 10 000, and each data item size is 10 KB. We use Zipf distribution [28] to model the data item access pattern. The Zipf distribution is often used to model a skewed access pattern, where θ is the access skewness coefficient that varies from 0 to 1.0. Setting $\theta = 0$ corresponds to the uniform distribution. Here, we set θ to 0.8 according to studies on real web traces [29].

4.2 Mobility Models

The random waypoint mobility [30] is primarily used in most of the MANET studies. Since a single mobility model may not capture the different scenarios, we use two other mobility models [31] in our simulation.

Random waypoint mobility (M_r): With this approach, the MTs travel toward a randomly selected destination in the network. After they arrive at the destination, they choose a rest period (pause time) from a uniform distribution. After the rest period, the MTs travel towards another randomly selected destination, repetitively. If the pause time is 0, then the MTs always move.

Manhattan grid mobility (M_g): It models a Manhattan-like street, where an MT moves only along the predefined paths and changes its direction at an intersection with equal probabilities. In this paper, we use three horizontal and vertical paths. Two vertical and horizontal paths are placed in the left- and rightmost edges, and bottom and top edges, respectively. The third vertical and horizontal paths are in the middle of the area. The AP is located at the center of intersection. In this model, the connectivity among MTs is much lower than that of M_r.

Modified random waypoint mobility (M_a): It is a variation of M_r and models a downtown or city, where people are crowded around a place or event, called an *attraction point*, such as a theater, shopping mall, or building. An MT chooses the direction toward the attraction point rather than selecting a randomly selected destination. We use three attraction points at $(500, 1\,250)$, $(1\,500, 750)$, and $(2\,500, 250)$ in the given area, respectively. Since it is reasonable to have an AP in a place where people are crowded, one of the attraction points is co-located with the AP. In this model, the connectivity is higher than that of the previous two mobility models.

4.3 Simulation Results

We investigate the impact of caching, cache management, and network topology on the IMANET performance with the three different mobility models.

4.3.1 Impact of Caching

We first examine the impact of caching on the IMANET performance. Here, we define a performance metric, called *accessibility*, as the fraction of successful requests over all of requests. In Table 2, accessibility is greatly improved over the entire mobility patterns when we use the aggregate cache compared to the no cache case. With caching, there is a high probability of the requested data being cached in the MT's local cache or at other MTs'. Even though a set of MTs is isolated from the AP, in contrast to the no cache case, they still try to access the cached data items among themselves.

	M_r	M_g	M_a
No caching	30.3851	11.2200	43.0580
Our approach	42.9534	24.5801	54.9855

Table 2. Accessibility comparison with three different mobilities

4.3.2 Impact of Cache Management

Next, we examine the impact of cache management. The proposed AC_LRU policy is compared to the three variations of schemes [6], that uses our admission control and chooses a victim data item based on a given weight on distance (TDS_D), access frequency (TDS_T), or both (TDS_N). In Figure 3, AC_LRU scheme exhibits better performance than that of all the variations of TDS over the entire mobility patterns. In M_r, AC_LRU shows the best performance because there is a high probability that both a less popular data item will be selected as a victim and a popular data item to be found in other MTs. The performance becomes worse when the mobility pattern is M_g. It implies that the low connectivity prevents a set of MTs from accessing other MTs or the AP. Thus, the isolation period of the MTs is relatively longer in M_g than that of other mobility models. Since the MTs have a high probability to access closer MTs, our admission control that helps in caching large number of distinct data items improves performance. In M_a, however, the AP is not only co-located with one of the attraction points but also between the other two attraction points. Since the MTs move between attraction points, there are MTs which are located between the AP and an attraction point. Thus, an MT has a high probability of being connected to the AP through multi-hop relay even if it is located far away from the AP. The overall accessibility is increased due to high connectivity among the MTs.

Figure 4 plots local cache hit (h_{local}) and remote cache hit (h_{remote}) against mobility patterns. h_{remote} implies that the requested data is available in other MTs' local cache. Even if a data item is popular and can be received from a nearby MT, it is not cached due to the cache admission control. Since M_g has a low connectivity, its accessibility is mainly contributed by accessing local cache of other MTs rather than AP. Thus, the cache hit is higher than that of other mobility patterns.

4.3.3 Impact of Network Topology

Finally, we observe the communication overhead in terms of communication message rate of the MTs after the simulation ends. In Figure 5, the message rates of the M_a are higher than that of M_r and M_g. Since the connectivity of M_a is relatively high, the MTs have a high probability to communicate among themselves. Thus, the message rate is increased. Also TDS_D shows slightly higher message rate than that of other policies with the M_a pattern, because data items are treated more fairly in the sense that the number of duplications for the most popular data items is restricted due to the cache admission control.

In summary, AC_LRU has the best accessibility and its further improvement is made possible by changing the Γ value, which controls the degree of duplication for the popular data items. Therefore, the aggregate cache improves IMANET performance and the proposed AC_LRU scheme is a viable cache replacement policy to provide better performance in varying topologies of IMANETs.

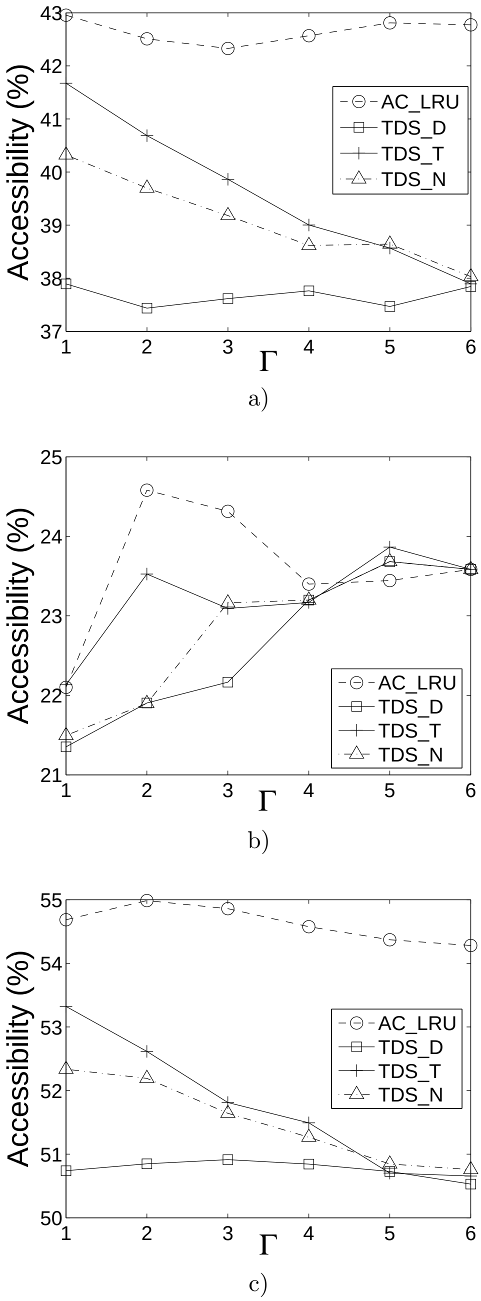


Fig. 3. Accessibility as a function of Γ with different mobility patterns: a) M_r, b) M_g, c) M_a

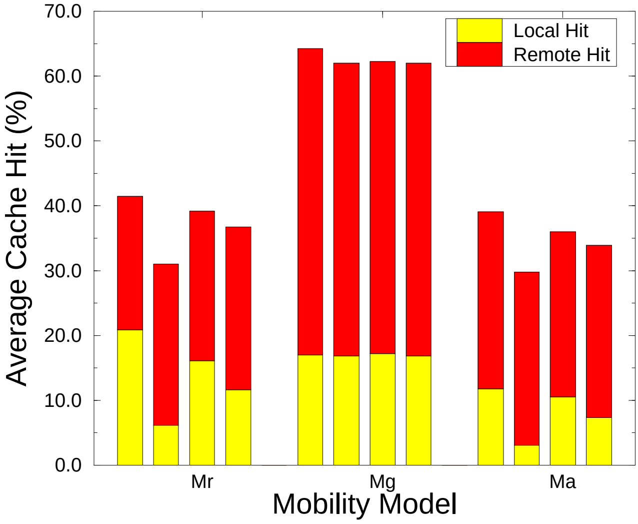


Fig. 4. Average cache hit ratio (h) comparison. (Four stack bars for different replacement policies are shown for three mobility models. The AC_LRU, TDS_D, TDS_T, and TDS_N are plotted from left to right.)

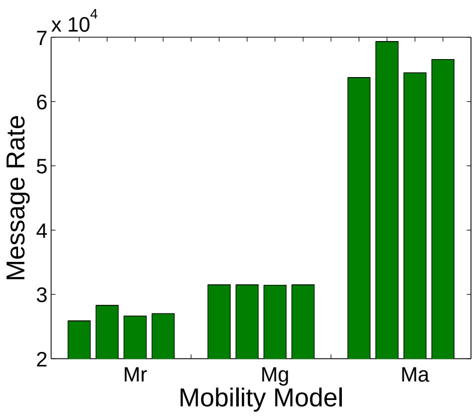


Fig. 5. Message rate comparison. (Four stack bars for different replacement policies are shown for three mobility modeles. The AC_LRU, TDS_D, TDS_T, and TDS_N are plotted from left to right.)

5 CONCLUDING REMARKS AND FUTURE WORK

In this paper, we proposed a novel cache replacement policy as a part of the aggregate cache mechanism not only to enhance the communication performance, but also to relieve the impact of changing network topology. The aggregate caching concept combines the local cache of each MT in forming a unified cache that can soothe the limited accessibility in case of network partition. The proposed policy combines the cache admission control to prevent high data duplication by enforcing a minimum distance between the same data items, and the cache replacement policy to improve the cache hit ratio and accessibility.

We first provided a simple analysis of the aggregate cache but it intended only to estimate the performance trend. Thus, a simulation based performance study was conducted to examine the proposed scheme from the three different perspectives: impact of caching, impact of cache management, and impact of network topology. For detail performance evaluation under various network topologies, we deployed three different mobility models: random waypoint, Manhattan grid, and modified random waypoint. The simulation results indicated that the proposed policy showed better accessibility and higher cache hit in varying network topologies, and relieved the network partition problem to a great extent.

There are many issues that need further investigation to exploit the full potential of IMANET. First, in this paper, we implicitly assumed that data items are never updated and a request query is answered by the latest updated data item from either a local cache or the server. We would relax this assumption to incorporate data modification capability. This brings in the cache update and consistency issues. Thus, we plan to investigate an adaptive cache consistency scheme to support growing diversity in applications' and users' demand requiring a certain consistency level with the server. Second, we plan to investigate several design aspects that often have conflict requirements in designing a caching scheme: consistency level (strong or weak), consistency control (push or pull), and consistency maintenance (stateful or stateless). For example, to ensure a certain level of consistency, either the server or an MT can initiate a procedure to control the consistency through push, pull, or combination of push and pull operations. The server may use a push operation (e.g. broadcast an IR periodically), while an MT may use a pull operation (e.g. polling). To keep track of status of each cached data item, the server may be a stateful server or stateless server. Third, we plan to expand our simple analysis of aggregate cache and build an extensive analytical model of partitionable IMANET to gain a performance insight.

Acknowledgements

This research was supported in part by Startup Grant in Dept. of Computer Science at Texas Tech University and National Science Foundation (NSF) under the grants CNS-0831673 and CNS-1004210.

REFERENCES

[1] Corson, M. S.—Macker, J. P.—Cirincione, G. H.: Internet-Based Mobile Ad Hoc Networking. IEEE Internet Computing, 1999, July-August, pp. 63–70.

[2] Sailhan, F.—Issarny, V.: Cooperative Caching in Ad Hoc Networks. Proc. MDM 2003, pp. 13–28.

[3] Luo, H.—Ramjee, R.—Sinha, P.—Li, L.—Lu, S.: UCAN:A Unified Cellular and Ad-Hoc Network Architecture. Proc. ACM MOBICOM, 2003, pp. 353–367.

[4] Wei, H.—Gitlin, R.: Two-Hop-Relay Architecture for Next-Generation WWAN/WLAN Integration. IEEE Wireless Communications, 2004, April, pp. 2–9.

[5] Cavalcanti, D.—Agrawal, D.—Cordeiro, C.—Xie, B.—Kumar, A.: Issues in Integrating Cellular Networks, WLANs, and MANETs: A Futuristic Heterogeneous Wireless Network. IEEE Wireless Communications, 2005, April, pp. 2–9.

[6] Lim, S.—Lee, W.—Cao, G.—Das, C. R.: A Novel Caching Scheme for Improving Internet-based Mobile Ad Hoc Networks Performance. Ad Hoc Networks Journal, Vol. 4, 2006, No. 2, pp. 225–239.

[7] Lim, S.—Lee, W.—Cao, G.—Das, C. R.: Cache Invalidation Strategies for Internet-based Mobile Ad Hoc Networks. Computer Communications Journal, Vol. 30, 2007, No. 8, pp. 1854–1869.

[8] Li, W.—Chan, E.—Wang, Y.—Chen, D.: Cache Invalidation Strategies for Mobile Ad Hoc Networks. Proc. ICpp. 2007, pp. 57–64.

[9] Zhao, J.—Zhang, P.—Cao, G.: On Cooperative Caching inWireless P2P Networks. Proc. IEEE ICDCS, 2008, pp. 731–739.

[10] Li, W.—Chean, E.—Chen, D.—Lu, S.: Maintaining Probabilistic Consistency for Frequently Offline Devices in Mobile Ad Hoc Networks. Proc. IEEE ICDCS, 2009, pp. 215–222.

[11] Ye, T.—Jacobsen, H. A.—Katz, R.: Mobile Awareness in a Wide Area Wireless Network of Info-station. Proc. ACM MOBICOM, 1998, pp. 109–120.

[12] Jiang, S.—He, D.—Rao, J.: A Prediction-based Link Availability Estimation for Mobile Ad Hoc Networks. Proc. IEEE INFOCOM, 2001, pp. 1745–1752.

[13] Li, Q.—Rus, D.: Sending Messages to Mobile Users in Disconnected Ad-hoc Wireless Networks. Proc. ACM MOBICOM, 2000, pp. 44–55.

[14] Sidell, J.—Aoki, P. M.—Devine, R.—Litwin, W.: Data Replication in Mariposa. Proc. ICDE, 1996, pp. 485–494.

[15] Wolfson, O.—Jajodia, S.: Distributed Algorithms for Dynamic Replication of Data. Proc. ACM PODS, 1992, pp. 149–163.

[16] Hara, T.—Harumoto, K.—Tsukamoto, M.—Nishio, S.: Dynamic Replica Allocation Using Database Migration in Broadband Networks. Proc. IEEE ICDCS, 2000, pp. 376–384.

[17] Dykes, S. G.—Robbins, K. A.: A Viability Analysis of Cooperative Proxy Caching. Proc. IEEE INFOCOM, 2001, pp. 1205–1214.

[18] Fan, L.—Cao, P.—Almeida, J.—Broder, A. Z.: Summary Cache: A Scalable Wide-Area Web Cache Sharing Protocol. Proc. ACM SIGCOMM, 1998, pp. 254–265.

[19] Wu, K.—Yu, P. S.—Wolf, J. L.: Segment-Based Proxy Caching of Multimedia Streams. Proc. WWW, 2001, pp. 36–44.

[20] Hadjiefthymiades, S.—Merakos, L.: Using Proxy Cache Relocation to Accelerate Web Browsing in Wireless/Mobile Communications. Proc. WWW, 2001, pp. 26–35.

[21] Cao, G.—Das, C. R.: On the Effectiveness of a Counter-Based Cache Invalidation Scheme and its Resiliency to Failures in Mobile Environments. Proc. IEEE SRDS, 2001, pp. 247–256.

[22] Cao, G.: A Scalable Low-Latency Cache Invalidation Strategy for Mobile Environments. Proc. ACM MOBICOM, 2000, pp. 200–209.

[23] Yeung, M. K. H.—Kwok, Y.: Wireless Cache Invalidation Schemes with Link Adaptation and Downlink Traffic. IEEE Trans. on Mobile Computing, Vol. 4, 2005, No. 1, pp. 68–83.

[24] Hara, T.: Effective Replica Allocation in Ad Hoc Networks for Improving Data Accessibility. Proc. IEEE INFOCOM, 2001, pp. 1568–1576.

[25] Hara, T.: Replica Allocation in Ad Hoc Networks with Period Data Update. Proc. MDM, 2002, pp. 79–86.

[26] Karumanchi, G.—Muralidharan, S.—Prakash, R.: Information Dissemination in Partitionable Mobile Ad Hoc Networks. Proc. IEEE SRDS, 1999, pp. 4–13.

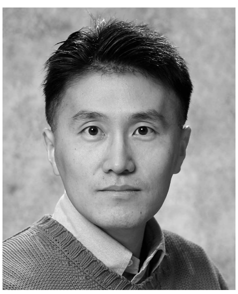
[27] The CSIM User Guides. Available on: `http://www.mesquite.com/documentation/`.

[28] Zipf, G. K.: Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge, MA 1949.

[29] Breslau, L.—Cao, P.—Fan, L.—Phillips, G.—Shenker, S.: Web Caching and Zipf-like Distributions: Evidence and Implications. Proc. IEEE INFOCOM, 1999, pp. 17126–17134.

[30] Johnson, D. B.—Maltz, D. A.: Dynamic Source Routing in Ad Hoc Wireless Networks. Mobile Computing, 1996, Kluwer, pp. 153–181.

[31] Gerharz, M.—Waal, C. D.—Frank, M.—Martini, P.: Link Stability in Mobile Wireless Ad Hoc Networks. Proc. IEEE LCN, 2002, pp. 30–42.



Sunho Lim received the B. Sc. degree (summa cum laude) in Computer Science and the M. Sc. degree in Computer Engineering from Hankuk Aviation University (a.k.a. Korea Aerospace University), Korea, in 1996 and 1998, respectively. He received the Ph. D. degree in Computer Science and Engineering from the Pennsylvania State University, University Park, in 2005. He is currently an Assistant Professor in the Department of Computer Science, Texas Tech University. He has authored/co-authored more than 30 technical papers in the areas of mobile P2P networks, mobile data management, and network security. He is a member of the IEEE.

Soo-Hoan Chae received the B. Eng. degree in Avionics from Korea Aerospace University in Korea in 1973. He received the M. Sc. degree in Computer Science and the Ph. D. degree in Computer Engineering from University of Alabama in USA in 1985 and 1988, respectively. He has been a Professor at the School of Electronic, Telecommunication, and Computer Engineering at Korea Aerospace University since 1989. His current interests include parallel/distributed computer architectures, real-time scheduling, communication software, and information retrieval.