Distributed dynamic mobile multicast

Yong Cui a, Shengling Wang b,∗, Sajal K. Das c

a Tsinghua University, Beijing, China
b Institute of Computing Technology, Chinese Academy of Sciences, China
c University of Texas at Arlington, USA

A R T I C L E   I N F O

Article history:
Available online 2 December 2010

Keywords:
Mobile multicast
Multicast tree configuration cost
Multicast packet delivery cost
Optimal service range

A B S T R A C T

Traditional mobile multicast schemes have either high multicast tree reconfiguration cost or high packet delivery cost. The former affects service disruption time while the latter affects packet delivery delay. Although existing region-based mobile multicast schemes offer a trade-off between two costs to some extent, most of them do not determine the size of the service range, which is critical to network performance. In this paper, we propose a novel approach, called Distributed Dynamic Mobile Multicast \((D^2M^2)\), to dynamically determine the optimal service range according to the mobility and service characteristics of a user. We derive an analytical model to formulate the costs of multicast tree reconfiguration and multicast packet delivery. The model is based on a Markov chain that analyzes a mobile node’s movement in a 2D mesh network. As the complexity of computing steady probability is high, we aggregate the Markov states by leveraging mobility symmetry. Simulation shows that the network performance is enhanced through \(D^2M^2\).

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The rapid progress of mobile networks has led to the tremendous demand for mobile services. One such service is mobile multicast because it outperforms the basic broadcast strategy by sharing resources among common links, while sending messages to a set of predefined destinations [22]. Due to its efficiency and flexibility, mobile multicast has gained a wide spectrum of applications, such as video-conferencing, video-on-demand, stock-quote distribution, and so on. Therefore, mobile multicast has received significant attention in recent years.

However, mobile multicast poses significant challenges because it must deal with not only dynamic group membership, but also dynamic locations of its members. To this end, Mobile IP proposes two basic schemes: the remote subscription (RS) and bi-directional tunneling (BT) [17,9].

In the RS, a mobile node (MN) is required to re-subscribe to its desired multicast groups every time it enters a new subnet. Thus, the multicast packets will be directly forwarded to the MN. However, the frequency of multicast tree reconfiguration is higher because it is directly proportional to users’ handoff frequency.

In the BT, on the other hand, the home agent (HA) joins in the desired multicast groups instead of the MNs. Hence, all multicast packets and signals are sent or received through the HA. Since this scheme hides a user’s mobility from other members of the group [12], no multicast tree needs to be updated after handoffs. However, the BT introduces the triangle routing problem, which makes the multicast packet delivery path far from optimal.

According to the above two mechanisms, RS has a high multicast tree reconfiguration cost while a low multicast packet delivery cost, which is exactly the opposite of the scenario in the BT. It is noteworthy that both multicast tree reconfiguration and multicast packet delivery costs are critical to the performance of mobile multicast. This is because a higher multicast tree reconfiguration cost incurs excessive signaling redundancy (from the viewpoint of network) and multicast service disruption time (from the viewpoint of users). This situation is worse in wireless mesh networks (WMNs) because the scarce and possibly asymmetrical wireless bandwidth require the amount of control signaling to be limited. In contrast, a higher packet delivery cost results from a longer packet delivery path and hence time. In multi-hop wireless networks, the longer the packet delivery path, the worse is the quality of service an MN receives.

Thus, a trade-off between these two costs is very important. This problem is similar to the trade-off between registration and packet delivery delays in mobile unicast. Hence, a natural idea to solve the trade-off problem in mobile multicast is to learn the idea of regional mobility management from mobile unicast schemes [20,4,13,24], thus leading to what are called the region-based mobile multicast schemes [12,21,23,29].
particular, we make the following contributions: according to the mobility and service characteristics of a user. In order to dynamically determine the optimal service range for an MMA, we propose a distributed dynamic mobile multicast (D2M2) scheme to dynamically determine the optimal service range for an MMA according to the mobility and service characteristics of a user. In particular, we make the following contributions:

(i) We propose an analytical model to formulate the cost of multicast tree reconfiguration and multicast packet delivery as functions of the service range. 

(ii) To establish the analytical model, we apply a Markov chain to analyze the mobility of MNs in a 2D mesh network. Because the complexity of computing steady state probability is high, we aggregate the Markov states by leveraging the mobility symmetry. In addition, a method is proposed to obtain the aggregate state set and the transition probability between any two aggregate states.

(iii) An iterative algorithm is derived to quickly find the optimal service range that balances the cost of multicast tree reconfiguration and multicast packet delivery. However, the computation of the optimal service range is an overhead of D2M2. To reduce this overhead, we present a mobile multicast architecture which groups the MMAs whose users have similar mobility and service characteristics into the same region range (SRR). Thus, only one MMA needs to compute the optimal service range for other MMAs in the same SRR.

D2M2 is also compatible with any multicast tree configuration algorithm. The MMA of an MN is decided by not only the MN’s location, but also the mobility and service characteristics of the MHA’s users. This makes each FA act as either a regular FA or an MMA. In other words, network traffic is feasibly allocated to each FA. Simulation experiments demonstrate that the optimal service range enhances network performance.

This paper is organized as follows: Section 2 presents related work. The overview of D2M2 is given in Section 3, while Section 4 presents the problem formulation and modeling. Section 5 discusses how to obtain the optimal service range. Section 6 describes some implementation issues and Section 7 presents the performance evaluation. Finally, we conclude the paper in Section 8.

2. Related work

In RS, an MN needs to re-subscribe to the FA whenever a handoff occurs. In this scenario, the packet loss rate might vary between 1% and 30% [18]. To reduce the packet loss, Jiunn-Ru and Wanjun [8] proposes a mechanism permitting an old agent to deliver the packets received by an MN while roaming into a new one. A similar method is proposed in [11], which makes an MN receive the multicast packets soon after handoff through the tunnel between the new FA and the old one. Jianhong et al. [7] proposes a scheme which let an MN still maintain the connection to the previous foreign network until it finds an available multicast router in another foreign network. Thus, the multicast packet loss is reduced. The schemes [8, 11, 17] are based on RS, so their multicast tree maintenance cost is high. To solve this issue, hierarchical multicast management techniques are proposed [21, 30, 14, 15]. However, they require the hierarchical network topology, and hence not suited for flat topologies such as wireless mesh networks (WMNs).

As mentioned earlier, the bi-directional tunneling (BT) scheme has the problem of triangle routing. To optimize the packet delivery path, Romdhani et al. [19] deploys the multicast router proxies in the home network for redirecting multicast traffic to the mobile receivers. But when the users are far from the home network, packet delivery path is not optimal.
In addition, the BT scheme has the tunnel convergence problem. That is, multiple HAs all happen to have MNs belonging to the same multicast group at the same FA. Thus one copy of each multicast packet would be delivered to the FA by each HA, leading to a significant transmission redundancy. To solve the above problem, Harrison et al. [5] uses the concept of designated multicast service provider (DMSP). A DMSP is one of the HAs selected by an FA when it serves multiple MNs of the same multicast group and some of these MNs belong to different HAs. Only the DMSP forwards multicast packets to the FA, thus avoiding duplicate packet delivery.

In addition to the above schemes, several hybrid solutions exist that make use of the complementary advantages of BT and RS. In the dual subscription (DS) method [6], an MN always keeps the home subscription and subscribes to an FA once it visits a new subnet. When the handoff is over, the MN adopts RS instead of BT. Literatures [3,31] selectively use BT and RS based on the mobility pattern of MNs. The RS is used when the MN is immobile or with low mobility speed. Otherwise, the BT is employed.

For trade-off between the multicast packet delivery path and the multicast service disruption time, the region-based mobile multicast schemes [12,21,23,29] were proposed. The range-based mobile multicast (RBMoM) [12] scheme introduces two concepts: the multicast home agent (MHA) and the service range. An MHA is a router to tunnel multicast packets to an MN's current FA. The service range is a given distance (measured by hops) between the MHA and FAs. An MHA only serves the MNs roaming around the FAs within its service range. Once the tunnel distance between the MHA and the FA is larger than the service range, the MNA handoff takes place.

MobiCast [21] adopts domain FAs (DFAs) to deliver the multicast packets for MNs within the domains. When an MN is a multicast source, it sends the multicast packets to the DFA by encapsulation, through which the packets can be sent out. When an MN is a multicast receiver, the DFA forwards the multicast packets to the MN by multicasting.

Wang and Chen [23] introduces multicast agents to serve as the access points of MNs (via FAs) to the multicast backbone. The service area of a multicast agent covers multiple foreign networks. The packets from the multicast agent to the FA and those from the FA to the MNs are all forwarded in multicast mode. Ye et al. [29] employs the mobile multicast gateways (MMGs) to manage the multicast service. The multicast packets are delivered from the MMG to the FA through tunneling or multicasting.

The above region-based mobile multicast schemes [12,21,23,29] focus on how to deliver the packets from the MMA to the FA. Different delivery modes require different operations in the MMAs. However, these schemes do not propose an algorithm to quantitatively determine the optimal service range, which is crucial to enhance network performance in terms of the multicast packet delivery time and the multicast service disruption time. This motivates us to propose a scheme called distributed dynamic mobile multicast ($D^2M^2$) to derive an optimal service range.

3. Overview of $D^2M^2$

Like region-based mobile multicast schemes, $D^2M^2$ introduces MMAs as the entities for managing mobile multicast within the service ranges. The aim of $D^2M^2$ is to find the optimal service range for each MMA to realize the trade-off between the costs for multicast tree reconfiguration and multicast packet delivery.

Before proceeding further, let us define a few terminologies.

3.1. Basic terms

**Definition 1.** Optimal service range, denoted as $K_{opt}$, is defined as the distance $K$ (measured by hops) between an MMA and an FA that minimizes the sum of average multicast tree reconfiguration cost ($C_2$) and average multicast packet delivery cost ($C_3$). Thus, $K_{opt}$ is given by:

$$K_{opt} = \{K | \min(C_2(K))\}$$  \hspace{1cm} (1)

where the overall cost function $C_T$ is defined as:

$$C_T(K) = C_2(K) + C_3(K).$$  \hspace{1cm} (2)

The term “Optimal service range” brings another definition:

**Definition 2.** MMA region, denoted as $R_M$, is a set of FAs whose hop distance $D(FA, MMA)$ to the MMA is within the optimal service range. Thus, $R_M$ is given by:

$$R_M = \{FA | D(FA, MMA) \leq K_{opt}\}.$$  \hspace{1cm} (3)

3.2. Tunnel convergence problem

In $D^2M^2$, an MMA may deliver packets to an FA by tunneling or multicasting. When employing tunneling, similar to RBMoM, $D^2M^2$ faces the tunnel convergence problem. We illustrate this problem with the help of Fig. 2, where $\text{MN}_{a1}, (\phi = a, b, \ldots, \omega = 1, 2, \ldots)$ denotes that the MN is the $\phi$th user of MMA$_a$, Assume all of $\text{MN}_{a1}, \text{MN}_{a2}, \ldots, \text{MN}_{a1}$ belong to the same multicast group. Because these MNs locate at the same FA, the common FA will receive $z$ copies of each multicast packet. In the worst case, the number of duplicate multicast packets increases with the number of MNs, leading to a serious bandwidth wastage in the network. To solve the above problem, we use the concept of DMSP in MoM [5]. In $D^2M^2$, the DMSP is one of the MMAs selected by an FA when it serves multiple MNs within the same multicast group and some of these MNs belong to different MMAs. Only the DMSP forwards multicast packets to the FA, thus avoiding duplicate packets received at the FA. For the case described in Fig. 2, only one copy (rather than $z$ copies) of each multicast packet is received by the FA, thus enhancing the network’s bandwidth utilization.

However, DMSP introduces another, i.e., which MMA should be chosen as the DMSP? To resolve this issue, $D^2M^2$ adopts the near-to-FA policy [5] for reducing multicast packet delivery time. This policy appoints the MMA with the shortest distance to the FA. According to the near-to-FA policy, if a new MN enters an FA and its MMA is nearer to the FA than the current DMSP, the MMA should be selected as the new DMSP. In other words, DMSP handoff occurs. Otherwise, if all MNs of a DMSP move away from the FA, a new DMSP should be re-selected.

3.3. Detailed operations

In $D^2M^2$, time is divided into equal length intervals. At the beginning of a time interval, $K_{opt}$ should be updated. Moreover, each MN is required to record its MMA and the corresponding $K_{opt}$. The initial MMA of an MN is the HA. When an MN arrives at a
foreign subnet, it will contact the FA with the MMA information. In this paper, we assume the packets are delivered from the MMA to the FA by tunneling. Hence, the DMSP should be selected to avoid the tunnel convergence problem. In addition, the FAs and MMAs are routers. To maintain connectivity, the routers must exchange some routing information, such as hop counts. Therefore, the FAs can obtain the hops to the MMA in an actual scenario and accordingly execute the following the procedure as shown in Fig. 3.

(1) When the distance between the FA and MMA is greater than $K_{opt}$, the FA must join this multicast group if it not already. The FA should act as the new MMA for the MN and the DMSP for the multicast group. Finally, the FA should inform the old MMA of deleting all the information on MN.

(2) When the distance is smaller than or equal to $K_{opt}$, the MMA is the only informed of the current FA. Similar to RBMOM, in order to get a shorter delivery path, if the FA at which an MN resides is only informed of the current FA. Similar to RBMOM, in order to get a shorter delivery path, if the FA at which an MN resides

Fig. 3. Operations of FA in $D^2M^2$.

4. Problem formulation and modeling

As described above, $K_{opt}$ is the key to $D^2M^2$. To compute $K_{opt}$, we must derive the overall cost $C_T$ beforehand. In light of Eq. (2), $C_T$ is a function of $K$. In the following, we first assume that the MMA region is confined by $K_{opt}$, and then we find the value of $K$ that minimizes $C_T$ to obtain $K_{opt}$.

By Definition 1, $C_T$ includes $C_P$ and $C_M$, where $C_P$ is the average cost for delivering multicast packets from the MMA to the FA and $C_M$ is the average cost for multicast tree configuration. Assuming that the multicast packet delivery cost is proportional to the distance, the distance between the MMA and the FAs follows a uniform distribution, $D_{MF}$ can be calculated as:

$$D_{MF} = \frac{4K^2 + 4(K - 1)^2 + \cdots + 4}{4K + 4(K - 1) + \cdots + 4} = \frac{2K + 1}{3}. \quad (4)$$

According to Fig. 3, when an MN roams out of an MMA region, the current FA should act as the new MMA. If the new MMA has not joined in the multicast group, multicast tree reconfiguration occurs. Hence $C_S$, the average multicast tree reconfiguration cost of an MMA, can be calculated as:

$$C_S = (1 - P_M)(C_M/(m_k \cdot T)) \quad (5)$$

where $P_M$ is the probability that the MMA has joined in the multicast group; $C_M$ is the average latency that an MMA joins in the multicast group; $T$ is the average time that an MN resides in the MMA; $m_k$ is the average number of times the FA is visited before an MN goes out of an MMA region, i.e., the average number of intra-MMA region handoffs, when the tunnel distance between the MMA and the FA is $K$. As a result, $C_M/(m_k \cdot T)$ represents the multicast tree reconfiguration cost within a unit time.

The average latency, $C_M$, that an MMA joins in the multicast group can be obtained from some empirical data. For this purpose, an MMA needs to record the latency each time it joins a multicast group. Then the MMA uses the statistical data to estimate the value of $C_M$. At the beginning of each time interval, $C_M$ is employed as a given input to compute $K_{opt}$. Besides $C_M$, the values of $m_k$ and $P_M$ should be obtained. How to derive them is explained in the following subsections.

4.1. Average number of intra-MMA region handoffs

For simplicity, we model the network as an $n \times n$ mesh network as in [16]. As shown in Fig. 4, each vertex is regarded as an FA with MMA function. We mark the FAs according to their positions relative to X- and Y-axis. In our model, an MN moves randomly towards one of four directions, and thus the movement probability of each direction is equal. Because the probability that an MN visits an AR is only related to the neighbors of the AR and the movement direction probability of the MN, we study the mobility of MNs using the Markov chain, where a state represents the FA which an MN accesses, and the transition probability is the MN’s movement direction probability.

It can be observed that MNs’ movement patterns at some FAs are identical due to the same transition probabilities. For example, when $K = 1$ and the MMA is $(0, 0)$, the MNs accessing one of $(1, 0), (0, 1), (0, -1)$ and $(-1, 0)$ could either move to $(0, 0)$ or go out of the MMA region; When $K = 2$, these MNs moves to $(0, 0)$, the corner of the region and the middle of the region edge with the probabilities of 0.25, 0.25 and 0.5 respectively. Although the movement pattern rotates by some angles, this rotation does not affect the mathematical analysis. To reduce the total number of the computational states, we aggregate the states using the movement symmetry (see Fig. 5).

Definition 3. Several FAs belong to an aggregate state only if with reference to the MMA region, the MNs that access one of these FAs have the identical patterns of movement to other FAs within and outside of the MMA region.

Fig. 4. Node numbering in a 2D mesh network topology.
From Fig. 5, the states that belong to one aggregate state are symmetrical about X-axis, Y-axis and the origin. In detail, the states $(i, j), (-i, -j), (-i, j), (i, -j), (j, i), (-j, i), (-j, -i)$ for $i, j = 1, 2, \ldots$, are all in one aggregate state. In our analyses, if $i \geq j$, the aggregate state is marked as $i_\_j$, otherwise $j_\_i$ as described in formula (6).

$$\{(i, j), (-i, j), (i, -j), (-i, -j), (j, i), (-j, i), (j, -i), (-j, -i)\} = \begin{cases} i \geq j \\ i < j \end{cases}$$  \hspace{1cm} (6)

In terms of Eq. (6), the FAs after being aggregated are shown in Fig. 6, where the rhombuses surrounding different FAs represent different MMA regions confined by different service ranges. In light of this figure, the same aggregate states have the same neighbors and transition probabilities to its neighbors. For example, each aggregate state $(1, 0)$ has the same neighbors, i.e., $(2, 0), (0, 0), (1, 1)$, and the transition probabilities to the neighbors are 0.25 and 0.5 respectively.

In our scheme, region is a logical concept. It actually works as a counter. Hence, once an MN moves out of a region, it never comes back. Due to this trait, we introduce the absorbing state “O” which represents the FAs outside the MMA region. Fig. 7 shows the aggregate states in an MMA region after adding state “O” when $K = 3$. Observing the movement direction probabilities as in Fig. 7, the aggregate state transition diagram for the regular Markov chain when $K = 3$ can be derived as shown in Fig. 8.

As shown in Fig. 7, since all the neighbors of $(0, 0)$ are $(1, 0)$, the probability that the aggregate state transfers from $(0, 0)$ to $(1, 0)$ is 1. In addition, because the aggregate state $(3, 0)$ has three “O” neighbors and one $(2, 0)$ neighbor, the transition probabilities from $(3, 0)$ to O and to $(2, 0)$ are 0.75 and 0.25, respectively. According to the property of an absorbing state, once it is entered, it cannot be left. Hence, there is self-transition to this state with a probability of 1.

Similarly, when $K = 1$ and $K = 2$, the aggregate state transition diagrams for the regular Markov chain are shown in Fig. 9(a) and (b) respectively.

Let $S_k$ denote the set of aggregate states when the tunnel distance between the MMA and the FA is $K$. From Fig. 7, $S_k$ is given by:

$$S_k = \{0 \cup \{i \_j \mid i + j \leq K \} \cap \{j \leq i \} \mid i, j \in \{1, 2, \ldots\} \}$$  \hspace{1cm} (7)

For example, when $K = 2$, $S_2 = \{0, 1_0, 1_1, 2_0, O\}$, and when $K = 3$, $S_3 = \{0, 1_0, 1_1, 1_2, 2_0, 2_1, 3_0, O\}$. Let the
state of the neighbors of \(i,j\) be \(N_{i,j}\), which is given by:

\[
N_{i,j} = \begin{cases} 
(i + 1,j, i + 1,j, i + 1, j - 1) \cap S_k \\
(i + 1,j, i - 1,j, i + 1, j - 1, O) \cap S_k \\
(1 < j < K) \\
(1 + j = K).
\end{cases}
\]

Let \(p_{i,j}^{(i,j)}\) represent the transition probability from the aggregate state \(i,j\) to \(i',j'\). From Figs. 8 and 9, the transition probabilities among some special neighbors can be summarized as follows:

\[
\begin{align*}
p_{0,0,1},0 &= 1 \\
p_{K,0,0} &= 1/4 \\
p_{i,0,j} &= 1/2, \quad i + j = K, j \neq 0 \\
p_{i,0,j-1,0} &= 1/4, 1 \leq i < K - 1 \\
p_{i,0,i+1,0} &= 1/4, 1 \leq i < K - 1 \\
p_{0,0} &= 1.
\end{align*}
\]

**Definition 4.** The aggregate state \(i_2,j_2\) is a special neighbor of \(i_1,j_1\) if their relationship satisfies the following:

\[
(i_1,j_1, i_2,j_2) \in \{(0,0), (1,0), (2,0), (0,0)\} \\
\cup \{(i_1,j_1, 0)|i_1 + j_1 = K, j_1 \neq 0\} \\
\cup \{(i_1,0, i_1 - 1, 0), (i_1,0, i_1 + 1, 0)|1 \leq i_1 \leq K - 1\}.
\]

If \(i_2,j_2\) is a special neighbor of \(i_1,j_1\), the aggregate state transition probability from \(i_1,j_1\) to \(i_2,j_2\) can be calculated by Eq. (9), and otherwise by formula (11):

\[
p_{i_1,j_1,i_2,j_2} = 1/|N_{i_1,j_1} - \{\text{special neighbors of } i_1,j_1\}| \
\text{where } |N_{i_1,j_1} - \{\text{special neighbors of } i_1,j_1\}| \text{ is the cardinality of the set } N_{i_1,j_1} - \{\text{special neighbors of } i_1,j_1\}.
\]

Let \(A_k\) be the number of all aggregate states excluding State \(O\) when the tunnel distance between the MMA and the FA is \(K\). In light of Eq. (7), \(A_k\) can be calculated as:

\[
A_k = \left\{(K + 3)(K + 1)/4 \quad \text{if } K \text{ is odd} \right. \\
\left. (2K^2)/4 \quad \text{if } K \text{ is even.} \right.
\]

We describe the transition relationship among the aggregate states except state \(O\) by the transition probability matrix \(P_{A_k \times A_k}\).

\[
P_{A_k \times A_k} = \begin{bmatrix}
p_{0,0,0,0} & p_{0,0,1,0} & \cdots & p_{0,0,k,0} \\
p_{1,0,0,0} & p_{1,0,1,0} & \cdots & p_{1,0,k,0} \\
\vdots & \vdots & \ddots & \vdots \\
p_{K,0,0,0} & p_{K,0,1,0} & \cdots & p_{K,0,k,0}
\end{bmatrix}.
\]

Let \(B\) be the fundamental matrix for an absorbing Markov chain. Then, \(B\) is given by formula (14), where \(I\) is the identity matrix with \(A_k \times A_k\) elements. Thus,

\[
B = (I - P_{A_k \times A_k})^{-1}.
\]

According to the properties of the absorbing Markov chain, the average handoff times that an MN needs to go out of an MMA region can be derived as the mean times from the initial state to the absorbing state. According to [11], \(m_k\) can be calculated as:

\[
m_k = \sum_{j=1}^{m_{A_k}} p_{i,j}
\]

where \(p_{i,j}(j = 1, 2, \ldots, A_k)\) are the elements of the first row in matrix \(B\).

4.2. Probability that an MMA joins in the multicast group

For a given multicast group, an MMA certainly knows whether it has joined in the multicast group or not. However, the MMA does not know which multicast group that an MN arriving at the next time interval belongs to and hence does not know whether it has joined in this multicast group. As a result, the probability calculation becomes challenging. In an actual scenario, there are many methods to compute the probability, such as using the history of statistical data. Here we introduce one of these methods.

Assume that there are \(E\) multicast groups in the network and the average number of members per multicast group is given by \(G\). Fig. 10 shows a state diagram of a Markov chain describing the behavior of an MMA [25]. In this figure, each state is represented by \((i, \eta)\), where \(\eta\) indicates the number of MNs serviced by the MMA. \(\eta \in \{0, 1, 2, \ldots, E\}\) and \(\eta \in \{0, 1, 2, \ldots, E\}; \) Here \(\eta = 0\) means that the MMA has not joined in any multicast group whereas \(\eta = 1\) means that the MMA has joined in the \(i\)th \((i = 1, 2, \ldots, E)\) multicast group; \(\mu\) is the rate that an MN leaves the MMA, which can be easily obtained by \(\mu = 1/T\); and \(\lambda\) is the average user arrival rate at the MMA.

Let \(O_{ij}(1 \leq i \leq E \cdot G; 1 \leq j \leq E)\) be the probability that a member of the \(j\)th multicast group roams out of its old MMA region and currently becomes the \(i\)th user of the current MMA who has not yet joined in the multicast group. Without loss of generality, we assume the MN has an equal probability to join in any multicast group. In other words, the probability that a user belongs to the \(j\)th \((i \leq j \leq E)\) multicast group is \(1/E\). In addition, we assume that the MMA’s user arrival rate follows the Poisson distribution. Hence, \(O_{ij}(1 \leq i \leq E \cdot G; 1 \leq j \leq E)\) can be calculated as:

\[
O_{ij} = \frac{\lambda e^{-\lambda}}{i!} \left(1 - \frac{1}{E}\right)^{i-1} \left(\frac{1}{E}\right) \quad (1 \leq i \leq E \cdot G; 1 \leq j \leq E).
\]

The probability that the MMA joins in the multicast group is given by:

\[
P_M = (1/E) \cdot \sum_{i=1}^{E} \sum_{j=1}^{E} P_{ij} = (1/E) \cdot \left(1 - \sum_{i=0}^{E} P_{i,0}\right)
\]

where \(P_{ij}\) is the steady state probability of \((i,j)\) \((0 \leq i \leq E \cdot G; 1 \leq j \leq E)\). To solve \(P_M\), let \(\lambda_i = (E \cdot G - i)\lambda(0 \leq i \leq E \cdot G - 1)\) and \(\mu_i = i/T\) for \(0 \leq i \leq E \cdot G\), and then we can obtain the following balance equations from Fig. 10.
\[
\lambda_0 P_{0,0} = \mu_1 \sum_{k=0}^{E} P_{1,k} \left( \lambda_i + \mu_i + \sum_{j=0}^{E} O_{i,j} \right) P_{i,j} = \lambda_{i-1} P_{i-1,0} + \mu_{i+1} P_{i+1,0} \\
\quad \quad \quad \quad \quad (1 \leq i \leq E \cdot G - 1)
\]

\[
(\lambda_i + \mu_i) P_{i,j} = O_i P_{i,j} + \mu_{i+1} P_{i+1,j} (1 \leq j \leq E) \\
(\lambda_i + \mu_i) P_{i,j} = 0 \quad (2 \leq h \leq E \cdot G - 1, 1 \leq i \leq E) \\
(\mu_{E+C} + \sum_{j=0}^{E} O_{E+C,j} P_{E+C,0} = \lambda_{E+C-1} P_{E+C-1,0} \\
\mu_{E+C} P_{E+C,0} = O_{E+C,0} P_{E+C,0} + \lambda_{E+C-1} P_{E+C-1,0} (0 \leq t \leq E).
\]

To solve Eq. (18), we use the method in [2]. The probability \( P_q \) that the FA serves \( q \) number of MNs is given by:

\[
\begin{align*}
P_q &= P_{0,0} \\
&= \sum_{j=0}^{E} P_{j,0} \quad 1 \leq q \leq E \cdot G.
\end{align*}
\]

Applying Eq. (18) to (17), we obtain:

\[
\begin{align*}
(\lambda_i + \mu_i) P_{i,j} &= \lambda_{i-1} P_{i-1,0} + \mu_{i+1} P_{i+1,0} (1 \leq i \leq E \cdot G - 1) \\
(\mu_{E+C} + \sum_{j=0}^{E} O_{E,C,j} P_{E,C,0} &= \lambda_{E+C-1} P_{E+C-1,0} \\
\mu_{E+C} P_{E,C,0} &= O_{E+C,0} P_{E,C,0} + \lambda_{E+C-1} P_{E+C-1,0} (0 \leq t \leq E).
\end{align*}
\]

According to Eq. (20), \( P_0 (0 \leq n \leq E \cdot G) \) is given by:

\[
P_n = (\alpha (E/\mu + 1 - \alpha(E/\mu))^{n/(1 + \alpha(E/\mu))})^{E/\mu} \quad 0 \leq n \leq E \cdot G.
\]

From the above formulas, the steady state probabilities can be calculated by Eq. (22), where \( P_{E,0} \) and \( B_k \) can be calculated with the help of Eq. (23).

\[
\begin{align*}
P_{0,0} &= P_{E,0} \sum_{k=1}^{E} B_k \\
&= \sum_{j=1}^{E} P_{j,0} - P_{i,0} \quad 1 \leq i \leq E \cdot G - 1.
\end{align*}
\]

\[
P_{E,C,0} = P_{0,0} \sum_{k=1}^{E} B_k \\
B_k = \lambda_{k-1} / \left( \lambda_k + \mu_k + \sum_{i=1}^{E} O_{E,C,i} - \mu_{k+1} \cdot B_{k+1} \right) \\
B_{E,C} = \lambda_{E,C-1} / \left( \mu_{E+C} + \sum_{i=1}^{E} O_{E,C,i} \right) \\
1 \leq k < E \cdot G.
\]

5. Solution of optimal service range

According to Definition (1), \( K_{opt} \) is the value of \( K \) that minimizes the cost function. Since \( K_{opt} \) can only be an integer and the cost is not a continuous function of \( K \), the following method is adopted to derive \( K_{opt} \) [25]. First, we define the following functions:

\[
\Delta(K) = \begin{cases} 
1, & \text{if } C_T(K) > C_T(K - 1) \\
0, & \text{otherwise} 
\end{cases}
\]

\[
\varphi(K) = \begin{cases} 
0, & \text{if } K \neq 0 \\
1, & \text{otherwise} 
\end{cases}
\]

Then formulas (24) and (25) lead to the following equation:

\[
\varphi(\Delta(K)) = \begin{cases} 
0, & \text{if } C_T(K) > C_T(K - 1) \\
1, & \text{otherwise} 
\end{cases}
\]

6. Implementation issues

For adapting well to the dynamics of mobile networks, \( K_{opt} \) is calculated based on the user mobility and service characteristics, where the mobility characteristic is captured by the average FA residence time (i.e., \( T \)), while the service characteristic is captured by the average multicast packet arrival rate (i.e., \( \alpha \)). These parameters may change dynamically, implying that \( K_{opt} \) can be different and adjusted from time to time. Now \( T \) can be calculated by the method introduced in [1], while the algorithms for estimating \( \alpha \) can be found in [27,28].

Computing \( K_{opt} \) is the overhead of \( D^2M^2 \) for enhancing system performance. To reduce this overhead, we propose a mobile multicast architecture, which employs the following concepts:

Definition 5. Same range region (SRR) is a set of MMAs whose users have similar mobility and service characteristics, i.e., \( \alpha \) and \( T \), and thus have a similar optimal service range.

Definition 6. Computation agent (CA) is an MMA in an SRR that takes the responsibility of periodically computing \( K_{opt} \) for each MMA in the same SRR.

The proposed architecture is shown in Fig. 11. At the beginning of a time interval, the CA computes \( K_{opt} \) according to the aggregate users' \( \alpha \) and \( T \), and then transfers the optimal service range value to each MMA in the same SRR. Since \( K_{opt} \) is only computed by the CA instead of each MMA, the computation overhead is reduced significantly.

In fact, the user mobility and service characteristics are closely related to the location of the FAs. For example, the users have higher mobility in the streets or highways, while lower mobility in the residential areas; the user traffic is higher in the business districts while lower in rural areas. Therefore, in actual scenarios, the
SRRs can be set by the internet service providers according to the location of FAs, the user density of region, the type of users, and so on. After setting the SRRs, some global parameters such as the multicast group number in the network (E) and the average number of members per multicast group (G) can be statistically analyzed and transferred only among the CAs. And these parameters are finally used for computing the optimal service range.

7. Performance analysis

This section compares the performance of $D^2M^2$ with RS, BT and RBMoM. Among these schemes, RBMoM and $D^2M^2$ have limited service ranges, so they can be grouped as the service-range-based mobile multicast (SRBMM) scheme. In fact, both RS and BT are the extremes of SRBMM [12]. If the service range $= \infty$, SRBMM is the same as BT; If the service range $= 0$, SRBMM is the same as RS. In this section, we first perform simulation experiments to observe the change of tunnel distance as well as the number of multicast tree reconfigurations in SRBMM, RS and BT at different service ranges and mobility rates. Finally, we compare the performance of $D^2M^2$ and RBMoM in terms of cost values using both simulation and numerical analyses.

The simulation was carried out using a custom C++ based simulator. The simulation topology is a 2D mesh network with a size of 20 × 20 which is similar to the one in Fig. 5 (a 7 × 7 mesh network). Each vertex in the mesh network is regarded as an FA with MMA function. These vertexes are numbered according to their positions relative to the X- and Y-axis.

At the beginning of each experiment, the MN accesses (0, 0); after residing in an FA for $T$ time units, it either stays in the same subnet or moves into one of four neighboring FAs’ coverage range with equal probabilities. The average FA residence time ($T$) and the average multicast packet arrival rate ($\alpha$) are different in different experiments, but in each experiment, they are static. After each handoff of the user, we need to determine whether it roams out of the MMA region. If yes, the average multicast tree configuration cost ($C_t$) and the average multicast packet delivery cost ($C_p$) can be calculated. Hence, the cost value can be obtained. Each experiment lasted 2000 time units and was repeated 20 times, following which we collected the simulation statistics to analyze the performance of $D^2M^2$ and other schemes. The main parameters used in the simulation experiments and numerical analyses are shown in Table 1.

7.1. Comparison among RS, BT and SRBMM

Fig. 12 shows how the tunnel distance in RS, BT and SRBMM changes with the service range, $K$. We observed that the tunnel distance in RS is equal to zero, which accords with the mechanism of RS. Moreover, the tunnel distance in BT is the longest. This is because BT has no limit in tunnel length. Once the MN accesses a new FA that has not joined in the multicast group, the tunnel is rebuilt from the HA. In addition, the tunnel distance in SRBMM increases as the service range increases. The reason is that the tunnel distance is confined by the service range in SRBMM. The service range is the maximum tunnel distance between the MMA and the FA. If the service range increases, the tunnel distance naturally increases as well. Fig. 12 verifies the phenomenon described in the introduction. That is, the packet delivery path is optimal in RS, while near optimal in SBMM, and longest in BT.

Fig. 13 shows how the frequency of multicast tree reconfiguration of RS, BT and SRBMM ($K = 2, 4$ and 6) change with $T$. From Fig. 13, we can see that when $T$ increases, the frequency of multicast tree reconfiguration in RS and SRBMM decrease while keeps zero in BT. Moreover, the frequency of multicast tree reconfiguration is the highest in RS and lower in SRBMM. This is because $T$ reflects the movement rate of the user. The shorter $T$ is, the faster is the movement rate, which leads to a higher handoff probability. In RS, once the MN handoff happens, the multicast tree should be reconfigured. A higher handoff probability leads to higher frequency of multicast tree reconfiguration. While in SRBMM, the multicast tree reconfiguration only happens when the MN roams out of the service range $K$, which greatly confines the frequency of multicast tree reconfiguration. In addition, Fig. 13 further shows the frequency of multicast tree reconfiguration decreases when the service range $K$ increases. The reason is that the MMA handoff frequency decreases as $K$ increases, which leads to the decrease in the frequency of multicast tree reconfiguration. Fig. 13 verifies the phenomenon described in the introduction. That is, the performance of multicast tree reconfiguration is the worst in RS and the best in BT.

7.2. Comparison between $D^2M^2$ and RBMoM

Figs. 12 and 13 demonstrate that $K$ is the key to balance the trade-off between the tunnel distance and the multicast tree reconfiguration frequency. This is why we proposed $D^2M^2$ to adaptively adjust $K$ for enhancing the system performance. In this subsection, we compare the costs of $D^2M^2$ and RBMoM using both simulation results and numerical analysis.

Fig. 14(a) and (b) show how the cost value of $D^2M^2$ and RBMoM ($K = 3$ and 5) changes with the packet arrival rate $\alpha$ when $T = 20$, $C_M = 5$ and 10, respectively. Fig. 15(a) and (b) demonstrate how the cost value of $D^2M^2$ and RBMoM ($K = 3$ and 5) changes with $T$ when $\alpha = 0.5$, $C_M = 5$ and 10, respectively. It can be seen
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>Tunnel distance between MMA and FA</td>
<td>1–20 hops</td>
</tr>
<tr>
<td>$T$</td>
<td>Average FA residence time</td>
<td>10–100 time units</td>
</tr>
<tr>
<td>$C_M$</td>
<td>Multicast tree reconfiguration cost of FA</td>
<td>5 and 10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Average packet arrival rate</td>
<td>0.15–0.95 packets/time unit</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Average user arrival rate</td>
<td>0.5</td>
</tr>
<tr>
<td>$E$</td>
<td>Multicast groups in the network</td>
<td>3</td>
</tr>
<tr>
<td>$G$</td>
<td>Average number of multicast members per multicast group</td>
<td>30</td>
</tr>
</tbody>
</table>

that the cost value of $D^2M^2$ is smaller than that of RBMoM, thus demonstrating that $D^2M^2$ improves the performance of RBMoM through finding the optimal service range adaptively for each MMA.

8. Conclusion

In $D^2M^2$ an MMA only serves users, whose FAs are within the optimal service range. The optimal service range can realize the trade-off between multicast service disruption time and multicast packet delivery time. This trait makes $D^2M^2$ effectively adapt to the dynamics of mobile networks. For finding the optimal service range, we model the MN's movement in a 2D mesh network using Markov chain. To reduce the computational complexity, we aggregate the states by using the mobile movement symmetry. In addition, we propose a method to compute the aggregate state set and the transition probability between any two aggregate states once the service range is given. Computing the optimal service range incurs the overhead of $D^2M^2$ while enhancing system performance. To reduce this cost, we propose a mobile multicast architecture which introduces the concepts of SRR and CA. Only the CA instead of each MMA in the network needs to compute the optimal service range. Finally, the performance comparison demonstrates that $D^2M^2$ outperforms RBMoM in terms of the cost value. For example, the simulation shows that $D^2M^2$ enhances 27% and 39% cost value compared with RBMoM ($K = 3$, $T = 20$) when the average latency that the MMA joins in the multicast group ($C_M$) is 10 and 5 time units.

Acknowledgments

The authors are grateful to the anonymous reviewers for constructive comments which helped to improve the quality of the manuscript.

References


Yong Cui, born in 1976, received the Ph.D. degree in 2004 from Tsinghua University, PR China. He is now an associate professor at the Department of Computer Science in Tsinghua University. He directed several national R&D projects, including some projects funded by the Chinese Next Generation Internet Plan (CNGI), a project funded by the National Natural Science Foundation of China (NSFC), and a project funded by the Development Plan of the State High Technology Research of China (863). He also participated as a major player in a project funded by the National Basic Research Plan of China (973 Program) and a Key Project of the National Natural Science Foundation of China. Having published more than 60 papers in international journals and conferences, he also holds more than 20 Chinese patents. His major research interests include wireless mesh networks, mobility management, the next generation Internet architecture, quality of service and distributed routing protocols.

Shengling Wang, born in 1978, received the Ph.D. degree in Computer Science in 2008 from Xi’an Jiaotong University. After that, she worked at Tsinghua University as a postdoctoral fellow. Now, she is an assistant research fellow at the Institute of Computing Technology, Chinese Academy of Sciences. Her major research interests include QoS, wireless/mobile networks, internet of things.

Sajad K. Das, born in 1976, received the Ph.D. degree in Computer Science in 1988 from University of Central Florida, Orlando. He is now a Professor of the Department of Computer Science and Engineering in University Texas at Arlington. He is director of Center for Research in Wireless Mobility and Networking (CREW-MaN). His major research interests include cellular mobile communication and computing, wireless multimedia, sensor networks/computer security, pervasive computing, broadband networking, parallel and distributed processing, parallel algorithms and data structures, parallel discrete-event simulation and multiprocessor interconnection networks.