

# Network Planning for Multicast Using Partitioned Virtual User Domains<sup>\*</sup>

Xuezhou Ma<sup>1</sup>, Selcuk Cevher<sup>1</sup>, M. Umit Uyar<sup>1</sup>, Mariusz Fecko<sup>2</sup>, John Sucec<sup>2</sup>,  
and Sunil Samtani<sup>2</sup>

<sup>1</sup> The City College and Graduate Center of the City University of New York, NY,  
USA

{zxue00,scevher,uyar}@ccny.cuny.edu

<sup>2</sup> Applied Research, Telcordia Technologies Inc., Piscataway, NJ, USA  
{mfecko,jsucec,ssamtani}@research.telcordia.com

**Abstract.** We introduce a new network planning approach for management of multicast communications in large data dissemination networks. Our approach, addressing *Channelization Problem*, considers both the common user interests and geographical dependencies while evaluating multicast configurations. We first perform *Global Similarity Classification* to partition users into virtual domains based on their characteristics within the same geographical vicinity. Second, based on the resource availability of a given network, *Localized Multicast Update* mechanism is used to evaluate the benefits of forming a new multicast group and the cost of the management overhead. We evaluate the performance of the proposed approach using our simulation software. Preliminary results show that, using two-stage planning, the performance of our approach is at least 40% better than the existing greedy approaches.

**Keywords:** Multicast, network planning, channelization.

## 1 Introduction

Multicast is the most efficient mechanism for information delivery in large scale data dissemination network (LSDDN) applications which require delivering a large number of information flows to many users. In management of such networks, multicast planning prior to network deployment is of particular importance. Moreover, properly planned multicast communications can assist in network management itself by disseminating management information.

The nature of LSDDN applications, such as multi-player online games [4], and distributed event notification systems [6], is such that individual users are

---

<sup>\*</sup> This work has been prepared through collaborative participation in the Communications Networks Consortium sponsored by the U.S. Army Research Lab under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. Copyright © 2007 The City University of New York and Telcordia Technologies, Inc. All rights reserved.

not interested in all information flows. Instead, users in the same geographical vicinity show common characteristics by having high similarity in their information flow interests, while users not in the same neighborhood do not have such characteristics. Each user receives all of the information flows mapped to the multicast groups to which it subscribed [2]. *Channelization problem* in multicasting is defined as finding an optimal mapping of information flows to the multicast groups, and an optimal subscription of users to the multicast groups so as to minimize the total bandwidth consumed. Fig. 1 shows the two components of channelization problem.

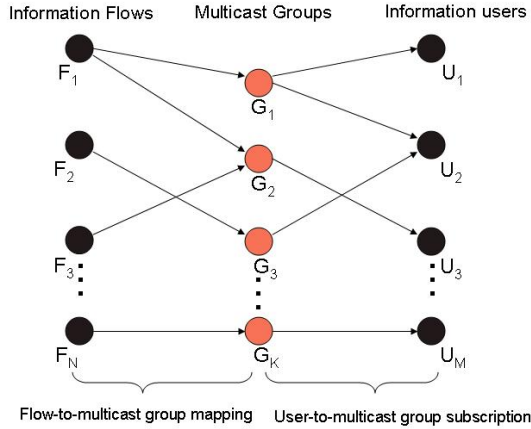


Fig. 1. Two components of channelization problem

Although it is desirable to use multicast for LSDDN to reduce the overall bandwidth demand, introduction of multicast groups requires configuration and management overhead [3] [1] (e.g., in terms of routing, determining the proper multicast route for each multicast group will consume additional bandwidth). Hence, there is a clear need for multicast group planning algorithms that provide a satisfactory balance between the unwanted traffic per user and multicast group management overhead, resulting in a near-optimal total bandwidth consumption. Specifically, the channelization problem addresses: (i) selection of the number of multicast groups to minimize total bandwidth consumption, (ii) update of multicast group planning in accordance with the network topology which dynamically changes due to user mobility, (iii) adaptation of the mapping algorithms to the networks with different characteristics (e.g., end-user tolerance to unwanted packets), and (iv) reduction of the computational complexity of the mapping algorithms.

In this paper, we provide a formal definition of channelization problem with respect to *no false exclusion* and *minimum false inclusion* requirements [1]. We consider the issues of homogeneous characteristics of the users within the same geographical boundary, routing overhead, total bandwidth consumption, and introduce a new multicast group planning approach which consists of two

main components: *Global Similarity Classification* and *Localized Multicast Group Update*. We provide simulation results for comparing our approach with existing greedy algorithms with respect to the number of multicast groups and total bandwidth consumption.

## 2 Background

Based on the channelization problem structure shown in Fig. 1, a data dissemination system is defined as a triple  $(F, G, U)$  where  $F = \{F_1, F_2, \dots, F_N\}$  is the set of information flows with each flow  $F_i$  assigned a transmission rate  $\lambda_i$ ,  $G = \{G_1, G_2, \dots, G_K\}$  is the set of multicast groups, and  $U = \{U_1, U_2, \dots, U_M\}$  is the set of users. In generalized channelization problem, the information flows (or users) might be mapped to more than one multicast groups. The user-flow interest matrix  $W$ , flow-to-multicast group mapping matrix  $X$ , and the subscription matrix  $Y$  are defined as follows [1]:

$$W = (w_{jm})_{M \times N}, \text{ where } j \in U, m \in F, M = |U|, N = |F|,$$

$$w_{jm} = \begin{cases} 1 & \text{user } j \text{ is interested in information flow } m \\ 0 & \text{otherwise} \end{cases}$$

$$X = (x_{im})_{N \times K}, \text{ where } i \in F, m \in G, N = |F|, K = |G|,$$

$$x_{im} = \begin{cases} 1 & \text{flow } i \text{ is assigned to multicast group } m \\ 0 & \text{otherwise} \end{cases}$$

$$Y = (y_{jm})_{M \times K}, \text{ where } j \in U, m \in G, M = |U|, K = |G|,$$

$$y_{jm} = \begin{cases} 1 & \text{user } j \text{ subscribes to multicast group } m \\ 0 & \text{otherwise} \end{cases}$$

Finding an optimal solution to the generalized channelization problem is NP-Complete [1]. However, some constrained polynomial-time instances can be considered since practical networking situations might not require the solution of the generalized form of channelization problem.

The existing solutions for channelization problem are either approximations using greedy approaches, such as User Based Merge (UBM) and Flow Based Merge (FBM) [1], or centralized control mechanisms such as Optimization-Based Congestion Control [3]. Since greedy approaches are based on a fixed number of multicast groups, they may not be appropriate for mobile networks. Also, such greedy approaches can be computationally inefficient since the flow-to-multicast group mapping and user-to-multicast group subscription have to be recalculated even for small amount of changes in network topology. On the other hand, centralized control mechanisms can be tuned to provide an acceptable level of total bandwidth consumption. However, they have to process feedback from all users leading to an unscalable multicast protocol overhead.

In general, the existing solutions need two improvements: (i) the benefits to the algorithms must be evaluated due to possible common interests of users residing in the same vicinity (compared to the users in different geographical

locations), (ii) a more efficient solution may emerge if the number of multicast groups is a result of diversity of user flow interests as opposed to a fixed number.

The first direction mentioned above can reduce the hardness of the channelization problem to a polynomial solution since the similarity consideration. Similarly, the second improvement will remedy the typical cases where the tolerance to routing overhead in the network is limited.

### 3 Our Approach

In channelization problem, the flow-to-multicast group and user-to-multicast group mappings must be such that: (i) all data needed by a user is mapped to one or more multicast groups to which the user either subscribes or is designated as the data recipient (no false exclusion), (ii) the amount of unneeded data received by users belonging to various multicast groups carrying the needed data is minimized (minimum false inclusion).

Additionally, the total bandwidth consumption should be taken into consideration while determining the mappings. Furthermore, any increase in the number of multicast groups requires additional routing overhead to determine the proper multicast routes. Since multicast groups is a limited resource, the number of multicast groups should be determined based on a design tradeoff involving all of the issues just described (i.e., minimum false inclusion, no false exclusion, total bandwidth consumed, and routing overhead).

#### 3.1 Global Similarity Classification

Although both constrained and generalized channelization problems are NP-Complete, in real LSDDN applications (e.g., hierarchical army networks, large wireless ad-hoc networks [5], online soccer games [4]), if the users can be partitioned into virtual domains at the planning stage based on their common characteristics and the geographical vicinity, the problems' computational complexity can be reduced to polynomial-time as described below.

Partitioning the users into virtual domains has three key benefits:

- It reduces the network layer protocol overhead. In most routing protocols, for example, the aggregate route update overhead grows as  $O(n^2)$ , where  $n$  is the number of routers in a routing domain [5]. Therefore, using the concept of virtual domains instead of considering all users as one single group reduces overall maintenance complexity for routing.
- If the virtual domains are well chosen, homogeneous characteristics among the users in the same domain will make channelization problem much easier. For example, in the context of Optimal Configuration Approach [1], when the number of virtual domains equals to the expected number of multicast groups, the partitioned result will independently work as a near-optimal configuration without any further consideration.
- Virtual user domain concept can fit to mobile networks characterized by highly dynamic arrival and departure of users.

Typically, the users in the same vicinity have common information flow interests with a higher probability than the ones in different geographical regions. For global similarity classification, we propose an interest-aware algorithm inspired by *overlap matrix technique* [2] to partition the users into virtual domains based on their common interests and geographical dependencies.

Let us first define similarity indicator  $\delta_i(j_1, j_2)$  of users  $U_{j_1}$  and  $U_{j_2}$  as:

$$\delta_i(j_1, j_2) = \begin{cases} 1 & \text{if } w_{j_1, i} \oplus w_{j_2, i} = 0 \\ 0 & \text{if } w_{j_1, i} \oplus w_{j_2, i} = 1 \end{cases}$$

where  $w_{j_1, i}$ ,  $w_{j_2, i}$  represent the flow interests of the users  $U_{j_1}$  and  $U_{j_2}$  regarding information flow  $F_i$ , respectively. We maintain an  $M \times M$  ( $M=|U|$ ) *similarity table* to store the similarity rates. Each entry  $k_{j_1, j_2}$  of the similarity table represents the similarity rate for the user pair of  $(U_{j_1}, U_{j_2})$ :

$$k_{j_1, j_2} = c_{j_1, j_2} \left( \frac{\sum_{i \in F} \delta_i(j_1, j_2)}{\sum_{i \in F} (1 - \delta_i(j_1, j_2))} \right) \quad (1)$$

where  $c_{j_1, j_2}$  is a proximity dependent coefficient ( $0 < c_{j_1, j_2} < 1$ ) where  $c_{j_1, j_2} = 0$  if the users are far from each other,  $c_{j_1, j_2} = 1$  if they are in the same vicinity, and between 0 and 1 depending on their distance. When  $\sum_{i \in F} (1 - \delta_i(j_1, j_2)) = 0$ ,

we set  $k_{j_1, j_2}$  as  $\infty$ , which indicates the highest value.

We introduce Algorithm 1 as shown in Fig. 2 which uses global similarity classification to find a set of virtual domains based on the similarity rates in similarity table. In the *for* loop in Fig. 2 (lines 1 through 10), the exclusive-or operator shown as  $\oplus$  in line 3 checks if the users  $U_{j_1}$  and  $U_{j_2}$  have common interest on information flow  $F_i$  ( $\oplus$  operator results in 0 if  $w_{j_1, i}$  and  $w_{j_2, i}$  are the same, and 1 otherwise). The counter called *overlap* is incremented if there is a common interest (line 5), otherwise, the counter *non-overlap* is incremented (line 7). In each iteration of the *while* loop (lines 11 through 13), the user (or group) pair of  $(j_1, j_2)$  with the highest similarity rate  $k_{j_1, j_2}$  is selected among the unselected pairs. If neither of the users  $U_{j_1}$  and  $U_{j_2}$  belongs to any virtual domains created before, a new virtual domain is created for them. If one of the users  $U_{j_1}$  or  $U_{j_2}$  is already a member of a virtual domain, the user currently not in this domain will be added to the domain.

As can be seen in Fig. 2, global similarity classification reduces the computational complexity of the channelization problem. Moreover, if used after network deployment, Algorithm 1 does not have to run for all user pairs each time a user arrives in the network since the only decision to make in Algorithm 1 is either to assign a new virtual domain to the new user or include it in one of the existing domains, while greedy approaches such as FBM may require a computation with  $O(N^2M + (N - K)^2M)$  running time to recover from such changes in the network topology.

**Algorithm 1. Global Similarity Classification**

**Input :**  $U$  (set of users in the network),  $F$  (set of information flows),  $W$  (user-flow interest matrix), and  $C$  (boundary dependent coefficients)

**Output :** Partitioned virtual user domains

PARTITION ( $W, F, U, C$ )

```

1  for all possible user pairs ( $U_{j_1}, U_{j_2}$ )
2      do for  $m \leftarrow 1$  to  $N$           /*  $N$  is the number of flows */
3          if ( $w_{j_1 m} \oplus w_{j_2 m}$ ) = 0
4              then
5                   $overlap \leftarrow overlap + 1$ 
6              else
7                   $non-overlap \leftarrow non-overlap + 1$ 
8
9           $k_{j_1, j_2} \leftarrow c_{j_1, j_2} (overlap \div non-overlap)$ 
10
11 while there is any user not partitioned to a virtual domain
12     do select the maximum similarity rate  $k_{j_1, j_2}$  and put
13     corresponding ( $U_{j_1}, U_{j_2}$ ) into the same virtual domain.
```

**Fig. 2.** Algorithm for global similarity classification

Virtual domains can be represented as an undirected graph  $G_N = (V, E)$  named *User Neighborhood Graph*, where  $V$  is the set of vertices (i.e., network users), and  $E$  is the set of edges among the vertices. There is an edge between any user pair ( $U_i, U_j$ ) as long as the users  $U_i$  and  $U_j$  are in the same neighborhood. Since each user can belong to only one neighborhood,  $G_N = (V, E)$  is naturally *disconnected*, and consists of *complete* subgraphs. Hence:

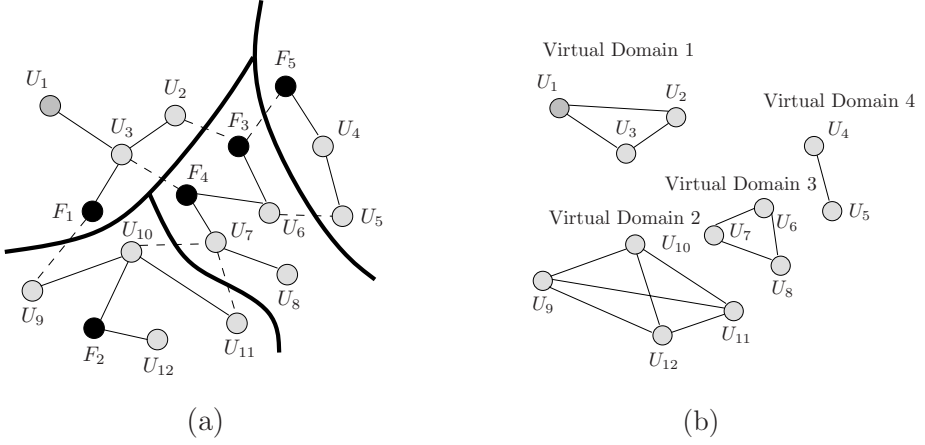
$$G_N = (V, E) \begin{cases} V = \{u : u \text{ is a network user}\} \\ E = \{(u, v) : u \text{ and } v \text{ are in the same neighborhood}\} \end{cases}$$

Fig. 3a depicts an example network in which black vertices represent information flows while gray vertices are for network users, and lines between vertices represent the communication links. Vertices in the same neighborhood are connected to each other with solid edges whereas dotted edges connect vertices in different neighborhoods. Fig. 3b shows  $G_N$  for the example network in Fig. 3a.

### 3.2 Constrained Channelization Problem

The cost  $C_{sub}$  of user-to-multicast group subscription of channelization problem is computed based on  $X$  and  $Y$  matrices, as defined in Section II, as follows:

$$C_{sub} = \sum_{i \in F} \sum_{j \in U} \sum_{m \in G} x_{im} y_{jm} \lambda_i \quad (2)$$



**Fig. 3.** Example of global similarity classification

where  $x_{im}$ ,  $y_{jm}$  are the members of  $X$  and  $Y$  matrices,  $F$ ,  $U$ ,  $G$  are the sets of information flows, multicast groups, and users, respectively, and  $\lambda_i$  is the transmission rate of the corresponding information flow  $F_i$ . The cost  $C_{map}$  of flow-to-multicast group mapping is computed based on  $X$  matrix as follows:

$$C_{map} = \sum_{i \in F} \sum_{m \in G} x_{im} \lambda_i \quad (3)$$

Hence, the total cost  $C_c$  of channelization problem is the sum of Eqs. (2) and (3):

$$C_c = C_{sub} + C_{map} = w_1 \sum_{i \in F} \sum_{j \in U} \sum_{m \in G} x_{im} y_{jm} \lambda_i + w_2 \sum_{i \in F} \sum_{m \in G} x_{im} \lambda_i \quad (4)$$

where  $w_1$  and  $w_2$  are introduced to provide relative weights to the costs of subscription and mapping, respectively. Considering the costs of both channelization problem  $C_c$  and routing management  $C_r$  (i.e., setting up and maintaining the multicast routes), the total cost  $C_{tot}$  of multicasting is defined as follows :

$$C_{tot} = C_c + C_r = w_1 \sum_{i \in F} \sum_{j \in U} \sum_{m \in G} x_{im} y_{jm} \lambda_i + w_2 \sum_{i \in F} \sum_{m \in G} x_{im} \lambda_i + w_3 C_r \quad (5)$$

where  $w_3$  is the corresponding relative weight for  $C_r$ . For simplicity, without loss of generality, we evaluate  $C_{tot}$  with the following approximations: (i) the cost of routing management increases linearly with the number of multicast groups, (ii)  $w_1 = w_2 = w_3$  so that they can be ignored in the computation, (iii) one user subscribes to only one multicast group. Now,  $C_{tot}$  becomes:

$$C_{tot} = \sum_{i \in F} \sum_{j \in U} \sum_{m \in G} x_{im} y_{jm} \lambda_i + \sum_{i \in F} \sum_{m \in G} x_{im} \lambda_i + \sum_{m \in G} \gamma_m \quad (6)$$

where  $\gamma_m$  represents routing overhead for the multicast group  $G_m$ . We now consider cost of merging two multicast groups  $G_k$  and  $G_{k+1}$ . Let the total cost function  $C'_{tot}$  after merging be:

$$C'_{tot} = \sum_{i \in F'} \sum_{j \in U'} \sum_{m \in G'} x'_{im} y'_{jm} \lambda_i + \sum_{i \in F'} \sum_{m \in G'} x'_{im} \lambda_i + \sum_{m \in G'} \gamma'_m \quad (7)$$

where  $F'=F$ ,  $U'=U$ ,  $|G'|=|G|-1$ . Under the constrained subscription framework, we define the cost changes in both flow-to-multicast group mapping  $C_{map}$  and user-to-multicast group subscription  $C_{sub}$  after the multicast group merging as  $\Delta C_{map}$  and  $\Delta C_{sub}$ , respectively. Thus, the difference  $\Delta C_{tot}$  of  $C_{tot}$  and  $C'_{tot}$  is:

$$\begin{aligned} \Delta C_{tot} = C_{tot} - C'_{tot} = & \lambda_i \left( \sum_{i \in F} \sum_{j \in U} \sum_{m \in G} x_{im} y_{jm} - \sum_{i \in F'} \sum_{j \in U'} \sum_{m \in G'} x'_{im} y'_{jm} \right) + \\ & \lambda_i \left( \sum_{i \in F} \sum_{m \in G} x_{im} - \sum_{i \in F'} \sum_{m \in G'} x'_{im} \right) + \left( \sum_{m \in G} \gamma_m - \sum_{m \in G'} \gamma'_m \right) = \Delta C_{map} + \Delta C_{sub} + \Delta C_r \end{aligned} \quad (8)$$

where  $\Delta C_{map}$  is determined based on three different cases: (i) if the flow is not sent to any multicast groups which have been merged,  $\Delta C_{map} = 0$ , (ii) if the flow is sent to only one of the multicast groups which have been merged,  $\Delta C_{map} = 0$  (iii) if the flow is sent to both multicast groups which have been merged,  $\Delta C_{map} = \sum_{i \in F} \delta_i \lambda_i$ , where  $\delta_i = 1$  if  $w_{ji} \times w_{j'i} \neq 0$ , and 0 otherwise. In the

same way,  $\Delta C_{sub}$  is determined as: (i) if the user is not subscribed to any multicast groups which have been merged,  $\Delta C_{sub} = 0$ , (ii) if the user is subscribed to only one of the multicast groups which have been merged,  $\Delta C_{sub} = - \sum_{i \in F} \delta'_i \lambda_i$

where  $\delta'_i = 1$  if  $w_{ji} \oplus w_{j'i} = 1$ , and 0 otherwise. Therefore, the total difference of cost functions before and after merging is:

$$\Delta C_{tot} = \left( \sum_{i \in F} \delta_i \lambda_i - \sum_{i \in F} \delta'_i \lambda_i \right) + \left( \sum_{m \in G} \gamma_m - \sum_{m \in G'} \gamma'_m \right) \quad (9)$$

An example exhibiting flow-to-multicast group mapping and user-to-multicast group subscription before and after multicast group merging is shown in Fig. 4a and Fig. 4b, respectively. Both users  $U_j$  and  $U_{j+1}$  are interested in the information flows  $F_i$  and  $F_{i+1}$ . The user  $U_{j+2}$  is interested in the information flows  $F_{i+1}$  and  $F_{i+2}$ . In Fig. 4a, which shows the multicast group configuration before merging, information flow  $F_{i+1}$  is sent to both multicast groups  $MG_k$  and  $MG_{k+1}$ , which makes the cost  $C_{map}$  of the flow-to-multicast group mapping equal to 235 and the cost  $C_{sub}$  of the user-to-multicast group subscription equal to 360. In Fig. 4b, the two multicast groups are merged into the new multicast group  $MG_{new}$ . The benefit after the merging is that  $F_{i+1}$  need to be sent only to one multicast group now which makes the cost  $C'_{map}$  of the flow-to-multicast group mapping reduce to 160. On the other hand, the cost  $C'_{sub}$  of the user-to-multicast group subscription increases to 480 due to the unwanted traffic generated since all of the three users are the members of a single multicast group now.



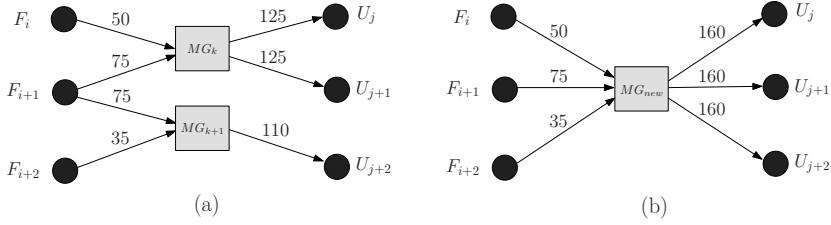


Fig. 4. An example for multicast group merging

### 3.3 Localized Multicast Group Update

In **resource-rich** networks (e.g., multi-player online games) are defined as networks whose users assign a greater weight to the cost of flow delivery than multicast routing overhead. In resource-rich networks, the users may tolerate very small amount of unwanted traffic, which requires the assignment of a separate multicast group for each user group even if it may not be a good choice for routing management. In some other networks, called **resource-constrained** networks, users may quickly recover from large amounts of unwanted traffic while they wait for a long time for the multicast tree maintenance. In resource-constrained networks, groups should be merged to reduce routing management as much as possible. In resource-constrained networks, if the pairwise merging cost  $\Delta C_{tot}$  for any possible virtual domain pair  $(V_i, V_j)$  is positive, we may need to further reduce the number of multicast groups by merging this virtual domain pair in our localized multicast group update.

In resource-rich networks, starting with  $M$  multicast groups, we create one multicast group for each user, and apply our localized multicast group update inside of each virtual domain simultaneously. The user pairs with minimum merging costs inside of each virtual domain will be merged during each iteration of update until the amount of end-user unwanted traffic for one user is over the expected value. Fig. 5 shows the localized multicast group update algorithm, which first the merges the multicast group pairs that have the least merging cost. It is applied among all virtual user domains for resource-constrained networks, and inside each virtual domain for resource-rich networks. The output of the algorithm is the final multicast group configuration which is both suitable to user groups' homogeneous characteristics and to each end user's unique characteristic.

An example for user-flow-interest matrix is shown in Table 1. Without loss of generality, we set the transmission rate  $\lambda_i$  for  $F_i$  as one of two values 1 and 10. To show the implementation of  $\Delta C_{tot}$  from Eq. (9) into a user-flow-interest matrix, we define  $\Delta C_c$  as:

$$\Delta C_c(U_i, U_j) = \Delta C_{map}(U_i, U_j) + \Delta C_{sub}(U_i, U_j) \quad (10)$$

Because  $\Delta C_c(U_1, U_2)$  and  $\Delta C_c(U_3, U_4)$  have the same maximum positive cost changes among all user pairs, we decide to merge the user pairs of  $(U_1, U_2)$  and  $(U_3, U_4)$ , as shown in Table 2. Similarly, we will merge user pairs  $(U_{1,2}, U_5)$

**Localized multicast group update**

**Input :**  $W$  (user-flow interest matrix) and  $V$  (set of virtual user domains)

**Output :** Multicast group configuration

RESOURCE-CONSTRAINED NETWORK ( $W, V$ )

```

1  for all possible virtual user domain pairs ( $V_i, V_j$ )
2      do for  $h \leftarrow 1$  to  $N$                                 /* $N$  is the number of flows*/
3          if  $w_{ih} \times w_{jh} = 1$ 
4              then
5                   $overlap \leftarrow overlap + \lambda_h$ 
6              else if  $w_{ih} \oplus w_{jh} = 1$ 
7                   $non-overlap \leftarrow non-overlap + \lambda_h$ 
8
9           $\Delta_{C_c}(V_i, V_j) \leftarrow overlap - non-overlap$ 
10
11  $\Delta_{tot}(V_i, V_j) \leftarrow \Delta_{C_c}(V_i, V_j) + \Delta_{C_r}(V_i, V_j)$ 
12 Create multicast groups by merging the virtual user domains
     $v_i$  and  $v_j$  with the maximum  $\Delta_{tot}(V_i, V_j)$  value into the same group

```

RESOURCE-RICH NETWORK ( $W, V$ )

```

1  for all possible user pairs ( $U_i, U_j$ ) in each virtual domain  $V_k$ 
2      do for  $h \leftarrow 1$  to  $N$ 
3          if  $w_{ih} \times w_{jh} = 1$ 
4              then
5                   $overlap \leftarrow overlap + \lambda_h$ 
6              else if  $w_{ih} \oplus w_{jh} = 1$ 
7                   $non-overlap \leftarrow non-overlap + \lambda_h$ 
8
9           $\Delta_{C_c}(U_i, U_j) \leftarrow overlap - non-overlap$ 
10
11  $\Delta_{tot}(U_i, U_j) \leftarrow \Delta_{C_c}(U_i, U_j) + \Delta_{C_r}(U_i, U_j)$ 
12 Create multicast groups by merging the user pair with the maximum
     $\Delta_{tot}(U_i, U_j)$  value into the same group inside of virtual domain  $V_k$ 

```

**Fig. 5.** Algorithm for multicast group update

**Table 1.** Example for 5 users and 5 flows user-flow-interest matrix

	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
	$(\lambda_1 = 1)$	$(\lambda_2 = 1)$	$(\lambda_3 = 10)$	$(\lambda_4 = 10)$	$(\lambda_5 = 10)$
$U_1$	1	0	0	10	10
$U_2$	0	1	0	10	10
$U_3$	1	0	10	0	10
$U_4$	0	1	10	0	10
$U_5$	1	1	0	0	10

$$\Delta_{C_c}(U_1, U_2) = 10 + 10 - 1 - 1 = 18$$

$$\Delta_{C_c}(U_3, U_4) = 10 + 10 - 1 - 1 = 18$$

**Table 2.** User-flow-interest Matrix after merging the user pairs  $(U_1, U_2)$  and  $(U_3, U_4)$ 

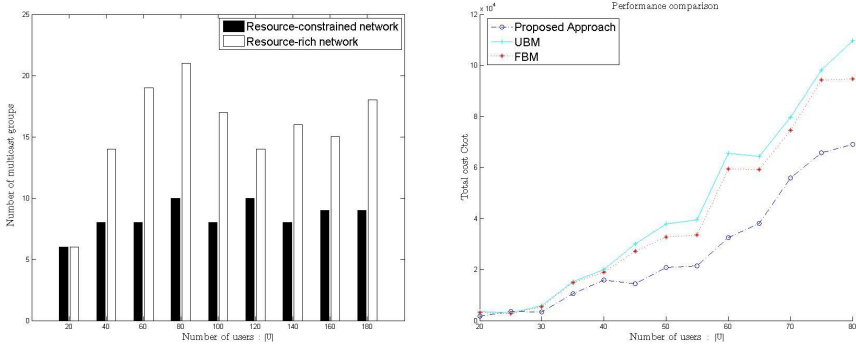
	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
	$(\lambda_1 = 1)$	$(\lambda_2 = 1)$	$(\lambda_3 = 10)$	$(\lambda_4 = 10)$	$(\lambda_5 = 10)$
$U_{1,2}$	1	1	0	10	10
$U_{3,4}$	1	1	10	0	10
$U_5$	1	1	0	0	10

$$\begin{aligned}\Delta_{C_c}(U_{1,2}, U_5) &= 12 - 10 = 2 \\ \Delta_{C_c}(U_{3,4}, U_5) &= 12 - 10 = 2 \\ \Delta_{C_c}(U_{1,2,5}, U_{3,4}) &= 12 - 30 - 20 = -38\end{aligned}$$

based on  $\Delta_{C_c}(U_{1,2}, U_5)$ . Further, if  $\Delta_{C_r}$  is greater than pairwise merging cost  $\Delta_{C_c}(U_{1,2,5}, U_{3,4})$ , we may merge  $U_{1,2,5}$  and  $U_{3,4}$  in next iteration to reduce  $C_{tot}$ .

## 4 Simulation Results

To study the effectiveness of our approach, we implemented Algorithms 1 and 2 in Figs. 2 and 5, respectively, in C++. The simulation parameters are set as follows: (i) each information flow  $F_i$ , once created, is assigned a transmission rate  $\lambda_i$ , which takes one of the following values  $\lambda_h=100$ ,  $\lambda_l=10$ , with the equal probability; (ii)  $F_i$  is either popular or unpopular for users in its vicinity, with the probabilities of  $\mu_p$  and  $\mu_{up}$ , respectively; (iii) each user, once created, is assigned to one of the user groups with equal probability; (iv) resource-rich networks with more multicast groups are available to reduce unwanted traffic, and resource-constrained networks where more unwanted traffic is tolerated to reduce the expense for multicast group routing and management overhead.



**Fig. 6.** (a) Effect of number of multicast groups for different networks for  $|F| = 100$ ,  $|U|=100$ ,  $\lambda_h=100$ ,  $\lambda_l=10$ ,  $\mu_p=0.9$ ,  $\mu_{up}=0.05$  (b)  $C_{tot}$  for our algorithm, FBM, and UBM

Our approach yields 10.3 and 5.6 as the average number of multicast groups for resource-rich networks and the resource-constrained networks, respectively (Fig. 6.(a)). This result implies that our approach is flexible enough to handle different network conditions for multicast group assignments as demonstrated by the smaller number of average multicast groups for constrained networks compared to the resource-rich networks. We also evaluated the cost of  $C_{tot}$  for different approaches when number of multicast groups  $k$  is pre-determined for the network of Fig. 6.(a) (except for the different (un)popularity rates of  $\mu_{up} = 0.1, \mu_p = 0.85$ ). Fig. 6.(b) shows that our proposed approach achieves at least 40% better performance than UBM and FBM in spite of the fact that a  $k$  is available (which is supposed to enhance the performance of UBM and FBM).

## 5 Conclusions

<sup>1</sup>We introduced a new multicast group planning approach for channelization problem in management of large data dissemination networks. First, we propose a classification algorithm, which forms virtual user domains by considering both common interests and geographical dependency characteristics among users. Then we apply a greedy merging algorithm to reduce total consumed bandwidth, unwanted traffic and multicast group management overhead depending on the availability of network resources. Simulation experiments have validated that the performance of our algorithm is better than existing algorithms.

## References

1. Adler, M., Ge, Z., Kurose, J.F., Towsley, D., Zabele, S.: Channelization Problem. In: Large Scale Data Dissemination, ICNP (2001)
2. Papaemmanouil, O., Cetintemel, U.: SemCast: Semantic Multicast for Content-Based Data Dissemination. In: ICDE (2005)
3. Shapiro, J.K., Towsley, D., Kurose, J.: Optimization-Based Congestion Control for Multicast Communication. In: ICDE (2002)
4. Levine, B.N., Crowcroft, J., Diot, C., Garcia-Luna-Aceves, J.J., Kurose, J.F.: Consideration of Receiver Interest for IP Multicast Delivery. In: Infocom (2000)
5. Galli, S., Luss, H., Sucec, J., McAuley, A., Samtani, S.: A Novel Approach to OSPF-Area Design for Large Wireless Ad-Hoc Networks, IEEE ICC (2005)
6. Carzaniga, A., Rosenblum, D.S., Wolf, A.L.: Achieving Scalability and Expressiveness in an Internet-scale Event Notification Service. In: ACM PODC (2000)

---

<sup>1</sup> The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Lab or the U.S. Government.