Handover Cost Optimization in Traffic Management for Multi-homed Mobile Networks

Shupeng Wang[1], Jianping Wang[2], Mei Yang[3]*Xiaochun Yun[1], and Yingtao Jiang[3]

[1] Intelligent Software Division, Institute of Computing Technology, Chinese Academy of Engineering, Beijing, 100080, China
[2] Department of Computer Science, City University of Hong Kong Kowloon, Hong Kong
[3] Department of Electrical and Computer Engineering University of Nevada, Las Vegas, NV 89154, U.S.A.
Emails:[1]{wshp, yxc}@hit.edu.cn, [2]jianwang@cityu.edu.hk, [3]{meiyang, yingtao}@egr.unlv.edu

Abstract. In a multi-homed mobile network, the traffic management problem, which is to redistribute the traffic to different available wireless access networks at different geographic areas on a journey, is important and challenging. When a wireless access network is not available in an area, the traffic carried on that wireless access network has to be switched to other wireless access network(s) available in that area, which incurs a handover process associated with certain cost. In this paper, we study the traffic management problem for multi-homed mobile networks with the objective of minimizing the overall handover cost. Two different traffic distribution modes are considered in our study: the Fractional Flow Assignment (FFA) mode which assumes the bandwidth demand from one connection is infinitely splittable and can be assigned to different wireless access networks, and the Integral Flow Assignment (IFA) mode which assumes that at any time the bandwidth demand from each connection can only be assigned to a single wireless network. We show that the general FFA problem and the IFA problem are both NP-hard and propose optimal solutions to the four special cases of the FFA problem and heuristic solutions to the general FFA problem and the IFA problem. Extensive simulations are conducted to verify the impact of the capacity-demand ratio on the performance of the proposed solutions.

1 Introduction

Providing Internet connectivity in a vehicle such as ships, aircrafts, and trains that contain mobile networks is desirable for passengers and also commercially attractive to the transportation systems [13][15]. In the next few years, mobile network communication is expected to grow rapidly [15]. In a typical vehicular communication scenario, the communication devices in large vehicles (e.g. buses or trains) form a local area network and connect to the mobile router (MR) which is installed on the vehicle and connected to the Internet. By employing an

MR as the gateway, all devices within the mobile network can achieve Internet connectivity [10][13].

The mobile network may be multi-homed, which means that the MR has several interfaces each connected to a different base station or there exist multiple MRs in the mobile network [13]. Multi-homing provides fault-tolerance and load sharing capabilities to the mobile networks. Potential multi-homed configurations for mobile networks and the solutions that support such configurations are discussed in [12].

When a wireless access network is not available in an area, the traffic carried on that wireless access network has to be switched to other wireless access network(s) available in that area, which incurs a handover process associated with certain cost. For a multi-homed mobile network moving on a predefined route, the wireless access networks available in different geographic areas are known to the mobile network. Such a pre-knowledge makes it feasible for the mobile router to seamlessly switch to the available wireless access networks when moving into a new area. It also provides information for traffic management which aims to assign the traffic demands from the mobile network to different available wireless access networks along the journey with minimum handover cost.
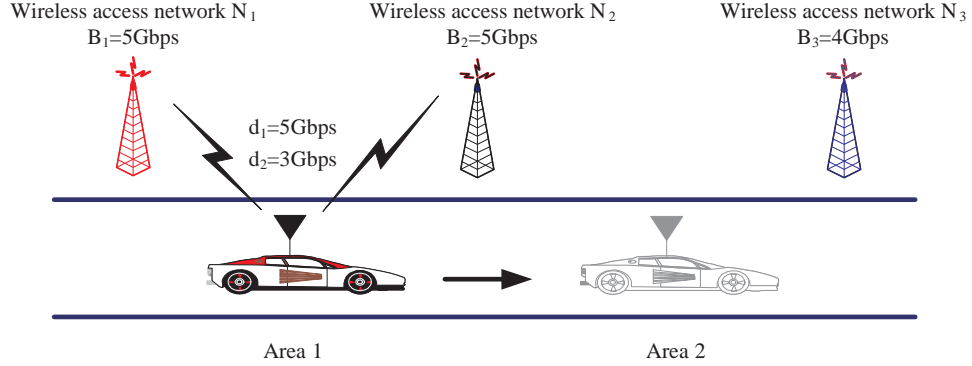
Fig. 1. Traffic distribution among different wireless access networks.

Traffic management in multi-homed mobile networks is important. Fig. 1 illustrates an example in which a vehicle is travelling from area 1 to area 2. In area 1, wireless access networks N_1 and N_2 are available, each with bandwidth capacity of 5Gbps. In area 2, wireless access networks N_2 and N_3 are available and N_3 has bandwidth capacity of 4Gbps. The mobile network residing in the vehicle has two connections with the bandwidth demand $d_1 = 5Gbps$ and $d_2 = 3Gbps$, respectively. A random assignment can be to distribute d_1 to N_1 and d_2 to N_2 in area 1. Under such an assignment, when the mobile network moves to area 2 where N_1 is no longer available, d_1 has to be redistributed as follows: 1Gbps to N_2 and other 4Gbps to N_3. Thus the handover cost is incurred by the traffic redistribution from N_1 to N_2 and from N_1 to N_3. However, with the pre-knowledge of the available wireless access networks at each area, a better

assignment is to assign d_1 to N_2 and d_2 to N_1. Then in area 2, the handover cost is only incurred by redistributing 3Gbps traffic from N_1 to N_3, which is less than that incurred in the random assignment.

In the literature, Montavont *et al.* investigated the multi-homing issue in nested mobile networks and proposed an optimal method for the hosts to discover the hierarchy of mobile routers [10]. However, it does not consider the problem of optimizing handover cost between different wireless access networks. In [6], Goldenberg *et al.* studied the problem of how to assign traffic dynamically such that cost and performance are optimized assuming that the user is already multi-homed to a set of ISPs. Nevertheless, the smart routing algorithm proposed in [6] is not applicable to mobile networks where the available wireless access networks are dynamically changing along the journey. In [4], the authors studied the traffic distribution scheme to maximize the total profit for the mobile hotspot given the service classes provided in a mobile vehicular network. Two traffic distribution schemes are considered in their study, *in-transit* which computes the optimum traffic distribution on-the-fly and *pre-transit* which computes the traffic distribution before the journey starts. However, no handover cost has been considered in their study.

In this paper, we investigate the traffic management problem for multi-homed mobile networks which is to distribute the traffic to different available wireless access networks with the objective of minimizing the total handover cost. In our study, we do not consider the cost charged for the traffic volume at each wireless access network since all wireless access service provider intends to apply the same charging policy due to competition. Our study is focused on the pre-transit traffic distribution [4]. We assume that the connections can be originated at any stage and terminated at any stage on the journey. Two different traffic distribution modes are studied: the Fractional Flow Assignment (FFA) mode and the Integral Flow Assignment (IFA) mode. We show that the general FFA problem and the IFA problem are both NP-hard and propose optimal solutions to four special cases of the FFA problem and heuristic algorithms to the general FFA problem and the IFA problem.

In the rest of the paper, we first present the problem description in Section II. In Section III, we study the fractional flow assignment problem. The study of the integral flow assignment problem is presented in Section IV. Section V presents the simulation results of the proposed algorithms for the two problems. Section VI concludes the paper.

2   Problem Description

Without loss of generality, we assume that the entire journey contains S geographic stages and N different wireless networks. For each stage s, $1 \leq s \leq S$, the availability of each wireless access network i is known in advance, where $1 \leq i \leq N$. Specifically, we use $a_{si} = 1$ to denote that wireless access network i is available for stage s, and $a_{si} = 0$ for otherwise. Assume that the bandwidth capacity of wireless access network i is B_i, then $a_{si}B_i$ gives the available bandwidth from wireless access network i at stage s, and the total available bandwidth at

stage s is given by $\sum_{i=1}^{N} a_{si}B_i$. We also define the duration of a wireless network i at stage s, denoted by L_{si} and $L_{si} = min\{s'|s' \geq s | a_{s'i} = 0 \text{ or } s' = S\} - s + 1$, as the number of stages for which the wireless access network will be continuously available from stage s.

Assume there are M different connections. Each connection $m, (1 \leq m \leq M)$ has a bandwidth demand d_m that has to be guaranteed during the lifetime of the connection. We use $D = \sum_{i=1}^{M} d_m$ to denote the total bandwidth demand from all connections. Let S_m^S and S_m^E denote the starting stage and the ending stage of connection m during the journey, respectively. The duration of a connection m (in the number of stages) can be represented as: $I_m = S_m^E - S_m^S + 1$.

At the beginning of a stage s, because the network availability changes, some connections may need to be switched from one wireless network i to another wireless network j, which is referred to as a *handover*. In [11], it defines the handover cost as a function of the bandwidth demand, charging rate, and power consumption. For ease of understanding, we assume that the handover cost from wireless access network (shortened as network) i to network j is determined by $C_{ij}x$ for a connection with bandwidth demand x, where C_{ij} is referred to as the handover cost coefficient between network i to network j. C_{ij} can be determined according to the detailed requirements of a specific application. The *traffic management problem* is to assign the connections to different wireless networks with the objective to minimize the total handover cost for all connections during the entire journey subject to that the bandwidth requirement for each connection is satisfied in any stage.

Similar to [6], we consider two traffic distribution modes: the fractional flow assignment mode and the integral flow assignment mode. Under the FFA mode, it is assumed that flows are infinitely splittable and the traffic demand from one connection can be assigned to multiple wireless networks simultaneously. For the FFA mode, the resource can be utilized efficiently with the cost of extra management complexity due to synchronization overhead. On the other hand, the IFA mode assumes that at any time the bandwidth demand from each connection can only be assigned to a single wireless network, though multiple connections can share one wireless network subject to the bandwidth constraint. For ease of presentation, we refer the traffic management problem under the FFA mode as the FFA problem, and the traffic management problem under the IFA mode as the IFA problem.

3   Fractional Flow Assignment Problem

In real applications, the starting and ending stages of each connection may be varied. According to the situations of the staring and ending stages, we classify the FFA problem into five cases, as shown in Tab. 1. Case I is the general case, in which each connection starts from a random stage and ends at a random stage. In Case II, all connections start from the same stage and end at the same stage. In Case III, all connections start from random stages and end at the same stage. In Case IV, all connections start from the same stage and end at random stages. We also consider a special case, Case V, which assumes the handover

cost coefficient is identical between any two networks and there is no constraint for the starting/ending stage of the connections. In Cases I, II, III and IV, the handover coefficient is assumed to be different for different pairs of networks.

Table 1. Five cases of the FFA problem.

	Starting stages	Ending stages	Handover coefficient
Case I	Random	Random	Different
Case II	Identical	Identical	Different
Case III	Random	Identical	Different
Case IV	Identical	Random	Different
Case V	Random	Random	Identical

In this section, we study the five cases of the FFA problem as follows. We first show that Case I of the FFA problem (i.e., the general FFA problem) is NP-hard. Then we provide optimal solutions to Cases II, III, and IV of the FFA problem based on the network flow model. Next we show the optimal solution to Case V. Finally, we propose two heuristic algorithms to solve Case I of the FFA problem.

3.1   NP-Completeness

Theorem 1: The decision version of the general *FFA* problem is NP-complete.

   Proof. We prove the theorem by the reduction from the set-packing problem, a known strongly NP-complete problem, to the decision version of the general FFA problem.

   The *Set-Packing* problem: Given a collection $\mathcal{A}$ of finite sets and a positive integer $K \leq |\mathcal{A}|$. The problem is to ask if $\mathcal{A}$ contains at least K mutually disjoint sets.

   Here we use a special case of the set-packing problem, in which each set has three elements and the elements in each set are consecutive positive integer in the range $[1, \cdots, 3g]$. This special case is also an NP-complete problem [5].

   Given an instance of this case of the set-packing problem, we can construct an instance of the general FFA problem as follows. Let there be $n = |\mathcal{A}|$ connections, where $|\mathcal{A}|$ denotes the number of sets in $\mathcal{A}$. Each connection m corresponds to a set in $\mathcal{A}$ and the stage id spanned by the connection corresponds to an element in the set. Assume that the bandwidth demand of each connection is 1, and there are $3g$ stages and $3g + 1$ networks. The network 1 is available from stage 1 to stage $3g$ and has bandwidth capacity 1. The network $s(s = 2, \cdots, 3g + 1)$ is available in stage $s - 1$ and has the bandwidth capacity $B_s(B_s > n)$. The handover cost from network i to j $(i = 2, \cdots, 3g+1, j = 2, \cdots, 3g+1, i \neq j)$ is 1, and the handover cost from network 1 to i $(i = 2, \cdots, 3g + 1)$ is $h_{1,i}$ $(h_{1,i} > 1)$. The decision version of the general FFA problem is to decide whether there is an assignment of the bandwidth demand of the n connections to the $3g+1$ networks with the total handover cost at most H, where $H = 2 * (n - K)$, $K \leq g$ and $n > g$.

   (1) Suppose that the set-packing problem has a feasible solution $U_{t_1}, \cdots, U_{t_i}, \cdots,$ $U_{t_K}(1 \leq t_i \leq n)$. Then we can allocate at least K connections $m_1, \cdots, m_K$ to

network 1, and the handover cost of the K connections is 0. Other connections can be allocated to networks 2 to $3g + 1$. It is evident that the handover cost is at most $2 * (n - K)$. This gives a feasible solution to the general FFA problem.

(2) Suppose that the general FFA problem has a feasible solution. Then we know that the handover cost is at most $2*(n-K)$. One possible allocation scheme for the general FFA problem is to assign at least K connections to network 1 and assign other connections to networks 2 to $3g+1$, each contributing handover cost 2. According to the bandwidth assumption of the connections and networks, the K connections should be disjoint. This implies a feasible solution of the set-packing problem.   □

For the general FFA problem, we will propose two heuristic algorithms to solve it. As these two algorithms will use part of the solutions to Cases II-IV, we will first present the solutions to these three cases.

3.2   Solutions to Cases II, III, and IV of the FFA Problem

The solutions to Cases II, III, and IV of the FFA problem all base on the minimum cost network flow algorithm. But the network flow models constructed for each case are slightly different. Hence, we describe the construction of the network flow model for each case first.
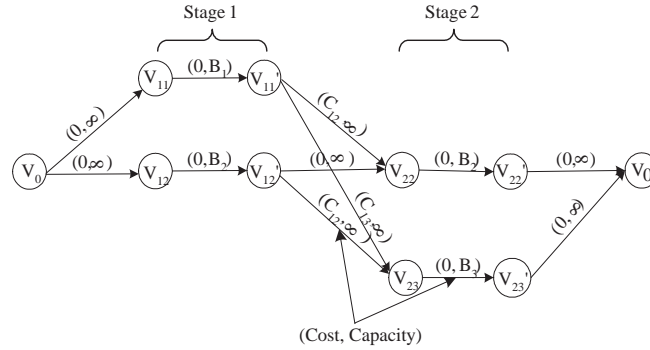
Fig. 2. Network flow model of Case II of the FFA problem.

The network flow model for Case II is constructed as follows.

Step 1.   For each stage s and network i, if $a_{si} = 1$, we define two nodes, denoted by V_{si} and V'_{si}, and add an edge from V_{si} to V'_{si} with its capacity set as B_i and its cost set as 0.

Step 2.   For any two nodes at adjacent stages s and $s + 1$, i.e., V'_{si} and $V_{s+1,j}$, we add an edge from V'_{si} to $V_{s+1,j}$ with its capacity set as infinity (∞) and its cost set as C_{ij} if $i \neq j$ or 0 if $i = j$.

Step 3.   Define a single sink node V'_0, and add an edge from each node V'_{Si} to V'_0 for $i = 1, \cdots, N$ with its capacity set as ∞ and its cost set as 0. Here S represents the ending stage.

Step 4.   Define a single source node V_0, and add an edge from V_0 to each node V_{1i} for $i = 1, \cdots, N$ with its capacity set as ∞ and its cost set as 0.

Fig. 2 shows the network flow model of Case II for an example with two stages and three networks where networks 1 and 2 are available in stage 1 and networks 2 and 3 are available in stage 2.

We can construct the network flow model for Case III by modifying the model for Case II as follows. All connections can be partitioned into sets, each including connections starting from the same stage. We denote these sets as $E_1, E_2, \cdots, E_S$, where E_s represents the set with connections starting from stage s $(1 \leq s \leq S)$. Then we construct the network flow model following the first three steps for case II and adding the following steps.

Step 4. Define a single source node V_0.
Step 5. For each nonempty connection set E_s, $1 \leq s \leq S$, do the following steps:
  – If there are more than one node in stage s, then define one node V_s and add an edge from V_s to each node in stage s with its capacity set as ∞ and its cost set as 0.
  – Add an edge from V_0 to V_s (or to the single node in the stage) with its capacity set as $\sum_{m \in E_s} d_m$ and its cost set as 0.

Similarly we can construct the network flow model for Case IV by modifying the model for Case II. Firstly, we partition the connections into different sets, each including connections ending at the same stage. We denote these sets as $E_1, E_2, \cdots, E_S$, where E_s represents the set with connections ending at stage s. Then we construct the network flow model following the first two steps for case II and adding the following steps.

Step 3. Define a single source node V_0, and add an edge from V_0 to each node V_{1i} for $i = 1, \cdots, N$ with its capacity set as ∞ and its cost set as 0.
Step 4. Define a single sink node V_0'.
Step 5. For each nonempty connection set E_s, $1 \leq s \leq S$, do the following steps:
  – If there are more than one node in stage s, then define one node V_s' and add an edge from each node in stage s to V_s' with its capacity set as ∞ and its cost set as 0.
  – Add an edge from V_s' (or from the single node in the stage) to V_0' with its capacity set as $\sum_{m \in E_s} d_m$ and its cost set as 0.

Then on the network flow model constructed for each case of the FFA problem, we need solve the Minimum Cost Network Flow (MCNF) problem which sends D units of flow from V_0 to V_0'. The MCNF problem is a classic optimization problem and can be solved readily using algorithms presented in [1]. From the network construction, we see that there is a one to one mapping between a feasible solution to the MCNF problem and a feasible solution to the corresponding case of the FFA problem, and the flow cost of a solution to the MCNF problem exactly reflects the handover cost of a solution to the FFA problem. Therefore, we have the following theorem.

Theorem 2: Cases II, III, and IV of the FFA problem can be solved in polynomial time by a minimum cost network flow algorithm.

3.3 Solution to Case V of the FFA Problem

We then consider Case V of the FFA problem which assumes identical handover cost coefficient between any two networks, i.e., $C_{ij} = C$ for all $i \neq j$. We propose a more efficient greedy algorithm for this special case. When all handover cost coefficients are equal, the key factor to the total handover cost is the number of handovers. The ideas of the greedy algorithm are 1) to postpone a handover as late as possible, and 2) to choose a network that has the longest duration when a handover is needed. We first formally state these two properties as follows.

Theorem 3: For the FFA problem with identical handover cost coefficient, there exists an optimal solution in which 1) no traffic will be switched if its current network is still available at the next stage; and 2) when some traffic is switched, it must be switched to an available network with the longest duration.

Proof. We need consider the following two cases. 1) Suppose that in an optimal solution there is some bandwidth x being switched from network i to j at the beginning of stage s while network i is still available. Then we can keep this part of bandwidth x within network i for stage s, and postpone the switch to stage $s+1$. We consider two cases. i) If at the beginning of stage $s+1$ there is no bandwidth being switched from network j to other networks, then we switch bandwidth x from network i to network j at stage $s+1$. In doing this, there is no increase to the overall handover cost. ii) If at beginning of stage $s+1$, some bandwidth is switched from j to other network k, then we can switch part of x to network j and part of x to k if network j does not have enough bandwidth. This will not increase the overall handover cost either.

2) Suppose that in an optimal solution there is some bandwidth x being switched from network i to network j while another network k with longer duration has unused bandwidth. Then we can reassign (part of) the bandwidth x to k, which will not increase the handover cost at stage s. Similar to (1), the new bandwidth on network k cannot cause more handover in future stages.   □

Theorem 3 leads to the following greedy algorithm. In the algorithm description, we use x_{si} to denote the bandwidth usage of network i at stage s.

Greedy Algorithm
Begin
  Initialization: Let $x_{in} = 0$ for $i = 1, \cdots, N$
  For stage $s = 1, \cdots, S$, **do**
    1. Keep the bandwidth assignment in the previous stage if possible:
      For $i = 1, \cdots, N$, **do**
        If $x_{s-1,i} > 0$ **and** $a_{si} = 1$, **then**
          let $x_{si} = x_{s-1,i}$.
          let $B'_{si} = a_{si}B_{si} - x_{si}$ be the left available bandwidth.
    2. Calculate the total bandwidth that needs to be switched as
    $D_s = D - \sum_{i=1}^{N} x_{si}$.
    Sort available networks in the nonincreasing order of their
    current available duration.
    3. Let the networks after sorting be $\pi(1), \pi(2), \ldots, \pi(N)$, i.e.,
    $L_{s,\pi(1)} \geq L_{s,\pi(2)} \geq \cdots \geq L_{s,\pi(N)}$.
    Find a minimum k such that $\sum_{i=1}^{k} L_{s,\pi(i)} \geq D_s$
    Assign D_s to the first $k - 1$ networks,
    each with the remaining bandwidth capacity of the network and
    assign D_s to $\pi(k)$ with $D_s - \sum_{i=1}^{k-1} L_{s,\pi(i)}$.
End

3.4   Two Heuristic Algorithms for the General FFA Problem

As shown in Section 3.1, the general case (Case I) of the FFA problem is NP-hard. In the following, we propose two heuristic algorithms for this case.

The first heuristic algorithm is based on the idea of grouping the connections with the same ending stage in one group and assigning the connections in such a group at one time. Similarly, we can group the connections with the same starting stage in one group and assign the connections following the group order. Here, we just use the first grouping method to elaborate the algorithm as the grouping method has not much impact to the algorithm performance. As one may see, the solution to Case III of the FFA problem can be directly applied to solve the assignment of the connections in each group. We hence name this heuristic algorithm as the Minimizing Group Connection Cost (MGCC) algorithm. The assignment order of these groups follows the order of the ending stage of these groups in such a way that the later the ending stage, the earlier the group is assigned. The MGCC algorithm works as follows.

MGCC Algorithm
Begin

    1. Divide the connections into different groups such that
the connections in a group have the same ending stage.
Sort them in the decreasing order of their connections'
ending stage. Let the sorted groups be $g(1), g(2), \cdots, g(S)$, i.e.,
$e_{g(1)} > e_{g(2)} > \cdots > e_{g(S)}$, where $e_{g(i)}$ denotes
the ending stage of the $g(i)$ for $i = 1, \cdots, S$.

    2. Select one unassigned group $g(i)$ following the sorted order.
Construct the network flow model for the group following the
steps for Case III (refer to Section 3.2), except only defining
nodes of connections from the first stage to the ending stage.

    3. Solve the FFA problem on the constructed network flow
model using the MCNF algorithm.

    4. Update the available bandwidth capacity of each network
in each stage by subtracting it by the bandwidth allocated to all
connections in group $g(i)$.

    5. Repeat Steps 2 to 4 until all groups are assigned.

End

The second heuristic algorithm is based on the idea of assigning one connection at one time. Hence, we name the second heuristic algorithm as the Minimizing Single Connection Cost algorithm of FFA (MSCC-FFA). The assignment order of the connections follows the order of the duration of these connections, i.e., the longer the duration, the earlier the connection is allocated. The MSCC-FFA algorithm works as follows.

MSCC-FFA Algorithm
Begin

    1. Sort the connections in the nonincreasing order of their
duration (in the number of stages), and sort the connections
with identical duration in the nonincreasing order of
their bandwidth demand. Let the sorted connections be
$\pi(1), \pi(2), \cdots, \pi(M)$, i.e., $I_{\pi(1)} \geq I_{\pi(2)} \geq \cdots \geq I_{\pi(M)}$,
where $I_{\pi(m)}$ denotes the duration of connection m.

    2. Select an unassigned connection $\pi(m)$ following the sorted order.
Construct the network flow model for the connection according to
the steps used for Case II (refer to Section 3.2).

    3. Solve the FFA problem on the constructed network flow
model using the MCNF algorithm.

    4. Update the available bandwidth capacity of each network in every
stage by subtracting it by the bandwidth allocated to connection
$\pi(m)$.

    5. Repeat Steps 2 to 4 until that all connections are assigned.

End

For each connection, the MSCC-FFA algorithm allocates its bandwidth demand using the MCNF algorithm. Hence, it will run the MCNF algorithm M

times to allocate all the M connections. The complexity is larger than that of the MGCC algorithm, which only needs to run the number of groups times of the MNCF algorithm.

4 Integral Flow Assignment Problem

The IFA problem can be described by an integer linear programming. The following variables are defined.

y_{sim}: binary variable, $y_{sim} = 1$ if and only if connection m is assigned to network i at stage s.

Z_{sm}: handover cost for connection m at stage s.

Then the objective of the IFA problem is to minimize the total handover cost:

$$\sum_{m=1}^{M} \sum_{s=2}^{S} Z_{sm}$$

subject to

$$\sum_{i=1}^{N} y_{sim} = 1, \text{ for } s = 1, \cdots, S; m = 1, \cdots, M. \tag{1}$$

$$\sum_{m=1}^{M} y_{sim} d_m \leq B_i, \text{ for } s = 1, \cdots, S; i = 1, \cdots, N. \tag{2}$$

$$Z_{sm} \geq (y_{s-1,im} + y_{sjm} - 1)C_{ij}, \text{ for } s = 1, \cdots, S;$$
$$i, j = 1, \cdots, N; m = 1, \cdots, M. \tag{3}$$

$$y_{sim} \in (0, 1). \tag{4}$$

In the formulation, constraint (1) guarantees that any connection will be assigned to a single network at any time; constraint (2) states that all bandwidth allocated is bounded by the bandwidth capacity of each network; and constraint (3) is used to measure the handover cost of each connection at each stage.

4.1 NP-Completeness

The following theorem shows that the IFA problem is computational intractable [5].

Theorem 4: The decision version of the IFA problem is NP-complete. **Proof.** We prove the theorem by the reduction from the 3-Partition problem, a known strongly NP-complete problem [5], to the decision version of the IFA problem.

The *3-Partition* problem: Given a set of non-negative integers $A = a_1, a_2, \cdots, a_{3m}$ and a non-zero integer B, where $\sum a_i = mB$, and $B/4 < a_i < B/2$ for $i = 1, \cdots, 3m$. The problem is to ask if A can be partitioned into m subsets $A_1, \cdots, A_m$ such that each A_j has three elements of which the sum is equal to B.

Given an instance of the *3-Partition* problem, we can construct an instance of the IFA problem as follows. Let there be $3m$ connections, each connection i corresponding to an element in A, and thus having a bandwidth demand a_i. Let there be m networks, each with an available bandwidth B. The decision version of the IFA problem is to feasibly assign each connection to a network.

(1) Suppose that the *3-Partition* problem has a feasible solution $A_1, \cdots, A_m$. Then we have a feasible solution to the corresponding IFA problem by letting each network j serve the three connections that correspond to the three elements in A_j. Obviously the total bandwidth requirement for each network is exactly B, and thus a feasible assignment is obtained.

(2) Suppose that the IFA problem has a feasible solution. Then we know that each network cannot have more than three users, for otherwise the total bandwidth will be larger than its capacity B. Consequently, the total bandwidth requirement for each network must be exactly mB because the overall requirement is mB. This implies a feasible solution to the *3-Partition* problem.   □

Although the IFA problem can also be classified into different cases as for the FFA problem, any case of the IFA problem is NP-hard. Hence, we do not differentiate the different cases of the IFA problem.

4.2   Heuristic Algorithm

We instead propose a heuristic algorithm for the IFA problem. Similar to the MSCC-FFA, the heuristic algorithm is based on the idea of assigning one connection at one time. The IFA problem with one connection can be solved using the MCNF algorithm.

For the IFA problem with one connection starting from the first stage and ending at the last stage, the network flow model can be constructed as follows, which is different from the steps described in Section 3.2.

Step 1.   For each stage s and network i, if $a_{si} = 1$, we define one node V_{si} with bandwidth capacity B_i.

Step 2.   For any two nodes at adjacent stages s and $s + 1$, i.e., V_{si} and $V_{s+1,j}$, we define an edge from V_{si} to $V_{s+1,j}$. The capacity of the edge is set as 1 if V_{si} has enough bandwidth to satisfy the bandwidth requirement of the connection or 0 otherwise. And the cost of the edge is set as C_{ij} if $i \neq j$, or 0 if $i = j$.

Step 3.   Define a single source node V_0 and an edge from V_0 to each node V_{1i} for $i = 1, \cdots, N$ with its capacity set as 1 and its cost set as 0.

Step 4.   Define a single sink node $V_0^{'}$ and an edge from each node V_{si} to $V_0^{'}$ with its capacity set as 1 and its cost set as 0.

For the IFA problem with one connection starting from a random stage and ending at a random stage, we can apply the above steps to construct the network flow model except only defining nodes from the starting stage to the ending stage.

On the constructed flow network, we then solve the MCNF problem which sends 1 unit of flow from V_0 to $V_0^{'}$ with the minimum cost using the MCNF algorithms [1]. The solution to the IFA problem with one connection can be obtained from the solution to the MCNF problem by taking the nodes (networks) with its incoming or its outgoing edge having 1 unit of flow.

We name the heuristic algorithm as the Minimizing Single Connection Cost algorithm of IFA (MSCC-IFA). As shown below, the steps in the MSCC-IFA algorithm are similar to the steps in the MSCC-FFA algorithm.

MSCC-IFA Algorithm
Begin

1. Sort the connections in the nonincreasing order of their duration, and sort the connections with identical duration in the nonincreasing order of their bandwidth demand. Suppose that the connections after sorting are $\pi(1), \pi(2), \cdots, \pi(M)$, i.e., $I_{\pi(1)} \geq I_{\pi(2)} \geq \cdots \geq I_{\pi(M)}$.

2. Select an unassigned connection $\pi(m)$ following the above order. Construct the network flow model for the connection according to the steps described above.

3. Solve the IFA problem on the constructed network flow model using the MCNF algorithm.

4. Update the available bandwidth capacity of each network in every stage by subtracting it by the bandwidth allocated to connection $\pi(m)$.

**5.** Repeat Steps 2 to 4 until that all connections are assigned.

End

The complexity of the MSCC-IFA algorithm is similar to the complexity of the MSCC-FFA algorithm.

Next we give an example to show how the MSCC-IFA algorithm works. In the example, assume that there are two stages, three networks with bandwidth capacity 20, 15 and 18, respectively and two connections with bandwidth demands 10 and 14, respectively. In stage 1, networks 1 and 2 are available, and in stage 2, networks 1 and 3 are available. The handover cost coefficients are $C_{1,1} = 0, C_{1,2} = 2, C_{1,3} = 1, C_{2,1} = 3, C_{2,2} = 0, C_{2,3} = 2, C_{3,1} = 2, C_{3,2} = 1, C_{3,3} = 0$.

According to MSCC-IFA algorithm, we first assign the bandwidth demand of connection 2. Then we construct the network flow model as shown in Fig. 3(a). According to the MCNF algorithm, the flow path with the minimum cost which can meet the bandwidth demand 14 is $V_0 \rightarrow V_{11} \rightarrow V_{21} \rightarrow V_0'$. So connection 2 is assigned to N_1 in both stages, and the handover cost is 0. The bandwidth capacity of N_1 is updated by subtracting the bandwidth demand of connection 2, 14. Next, we assign the bandwidth demand of connection 1. The network flow model constructed for connection 1 is shown in Fig. 3(b). The flow path obtained by the MCNF algorithm is $V_0 \rightarrow V_{12} \rightarrow V_{23} \rightarrow V_0'$, and the handover cost is 20. So connection 1 is assigned to N_2 and N_3 in stage 1 and stage 2, respectively. The total handover cost for connections 1 and 2 is 20.

5   Performance Evaluation

To evaluate the performance of the proposed solutions to the FFA and IFA problems in Sections III and IV, simulations have been conducted for different scenarios. For the FFA problem, the MGCC and MSCC-FFA for the general case (Case I) are simulated as the solutions to other cases are optimal. For the IFA problem, the MSCC-IFA for Case I is simulated.

(a) Connection 2   (b) Connection 1
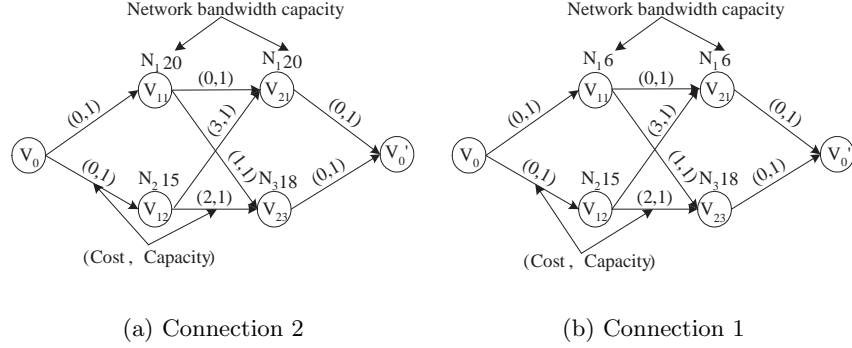
Fig. 3. Network flow models.

5.1   Simulation Settings

For all the simulations, we assume that the bandwidth capacity of each network and the bandwidth demand of each connection follow uniform distribution. The following notations are used in describing simulation settings.

B_u: the upper boundary of the bandwidth capacity of the networks.

B_l: the lower boundary of the bandwidth capacity of the networks.

B_A: the average bandwidth capacity, $B_A = (B_u + B_l)/2$.

B_T: the expectation of the total bandwidth capacity of the networks.

P: the probability of the networks being available.

D_u: the upper boundary of the bandwidth demand of the connections.

D_l: the lower boundary of the bandwidth demand of the connections.

D_T: the expectation of the total bandwidth demand of connections.

D_A: the average bandwidth demand, $D_A = (D_u + D_l)/2$.

M_A: the average number of connections in one stage.

Hence the expectation of the total bandwidth capacity provided by the N networks can be calculated as: $B_T = B_A * P * N$. The expectation of the total bandwidth demand required by the M connections can be calculated as: $D_T = D_A * M_A$. In addition, we define the ratio of the total bandwidth capacity to the total bandwidth demand (referred as the *capacity-demand ratio*) as:

$$R_T = \frac{B_T}{D_T} \tag{5}$$

In order to satisfy all the bandwidth demand, we should have $R_T \geq 1$.

In the following, we will present the simulation results of the total handover cost with three varying parameters: the number of connections (M), the number of networks (N), and the number of stages (S). For all experiments, 1000 instances are simulated and the results shown are the average of successful cases

(i.e., the cases that have all bandwidth demand satisfied). By default, the number of successful cases is over 50% of the 1000 instances.

The basic simulation settings for the experiments are shown in Tab. 2. Then we have $B_A = 7.5Mbps$. Note that for different cases, the calculation of M_A is different, which leads to the different R_T values even with the same setting.

Table 2. Basic simulation settings.

Parameter	Value
Range of bandwidth demand of connections ($[D_l, D_u]$)	Uniform distribution in $[50, 512]Kbps$
Range of bandwidth capability of networks ($[B_l, B_u]$)	Uniform distribution in $[5, 10]Kbps$
Handover cost coefficient	Uniform distribution in [5,10]
Network availability probability (P)	80%

5.2   Simulation Results

In the following, we present the results of MGCC, MSCC-FFA, and MSCC-IFA for the general case. In the simulations, we assume that each connection may start at a random stage and end at a random stage and the starting and ending stages of the connections follow uniform distribution. Three experiments are conducted. For all experiments, we use the R_T value to analyze the results.

In order to calculate R_T, we need calculate D_T. Hence we need find out M_A, which can be calculated in the following way. According to the assumption, the probability that one connection starts from the k_{th} stage is $1/S$ and the probability that the connection ends at the $l_{th}(l \geq k)$ stage is $1/(S - k + 1)$. The expectation of the duration of a connection m can be computed as: $I_m = \sum_{k=1}^{S}(\frac{1}{S} * \sum_{l=k}^{S} \frac{l-k+1}{S-k+1})$, and the total duration of all the connections is $I_{sum} = M * I_m$. Then we can get $M_A = I_{sum}/S$. Then, we have $D_T = D_A * M_A$.

Experiment 1 In the first experiment, we vary the number of connections and fix the number of stages and the number of networks at 5 and 5, respectively. The number of connections is chosen from the range of 20 to 100 in the step of 20. The upper bound of 100 connections is chosen because the minimum value of the maximum number of connections that can be supported by the networks is set by $B_T/D_A = 109.3$.

The average total handover costs of MSCC-FFA, MGCC, and MSCC-IFA are shown in Tab. 3. As shown in the table, the total handover costs of MSCC-IFA are larger than the results of MSCC-FFA and MGCC except for M=20, which is due to the natures of the IFA problem and the FFA problem. And the results of all three algorithms increase with the number of connections increasing while R_T's value decreasing. This is consistent with our intuition that when R_T's value is decreased, limited bandwidth is shared by more number of connections, which causes more number of handovers.

Table 3. The total handover cost results of MSCC-FFA, MGCC, and MSCC-IFA vs. number of connections for Case I.

M	20	40	60	80	100
MSCC-FFA	642.327	915.08	1646.02	2949.96	3513.59
MGCC (FFA)	641.683	912.173	1569.94	2695.47	3130.92
MSCC-IFA	641.683	919.268	1659.46	3041.22	3490.19
R_T	13.64	6.82	4.55	3.41	2.73

Experiment 2 In the second experiment, we vary the number of networks and fix the number of stages and the number of connections at 5 and 50, respectively. The maximum value of the minimum number of networks that are needed to support all the connections is given by $D_T/(P*B_A) = 1.8$. Hence we choose the number of networks from 2 to 10 in step of 2.

The results of the second experiment are shown in Tab. 4. According to Tab. 4, the handover costs of MSCC-FFA, MGCC, and MSCC-IFA all decrease with the number of networks increasing while R_T's value increasing. This is also consistent with our intuition that when R_T's value is increased, more bandwidth is available for fixed number of connections, which leads to less number of handovers.

Table 4. The total handover cost results of MSCC-FFA, MGCC, and MSCC-IFA vs. number of networks for Case I.

N	2	4	6	8	10
MSCC-FFA	3745.86	2427.79	752.405	246.314	125.015
MGCC (FFA)	4331.13	2354.01	747.493	244.584	124.75
MSCC-IFA	5124.92	2441.07	763.149	250.933	125.63
R_T	2.17	4.35	6.52	8.7	10.9

Experiment 3 In the third experiment, we vary the number of stages from 5 to 25 and fix the number of networks and the number of connections at 10 and 50, respectively. The results of this experiment are shown in Tab. 5. According to Tab. 5, the handover costs of the MSCC-FFA, MGCC and MSCC-FFA algorithm increase with the number of stages increasing while R_T's value increasing. This is due to the fact that the increasing of the number of stages causes dramatic increase of the total handover cost, which overweighs the decrease of the handover cost due to the increasing of R_T's value.

In summary, as shown in Tabs. 3-5, both the capacity-bandwidth ratio R_T and the average duration per connection have significant impacts on the total handover costs of all algorithms for all cases. In general, the total handover cost decreases with R_T's value increasing. And longer average duration per connection leads to larger total handover cost.

Table 5. The total handover cost results of MSCC-FFA, MGCC, and MSCC-IFA vs. number of stages for Case I.

S	5	10	15	20	25
MSCC-FFA	68.235	1776.91	5468.02	11326.9	17887.3
MGCC (FFA)	68.97	1786.13	5478.71	11330	17894.1
MSCC-IFA	67.22	1775.19	5460.08	11314.8	17868.7
R_T	10.9	13.4	14.6	15.2	15.6

6   Conclusion

In this paper, we investigated the traffic management problem with the objective of minimizing the total handover cost for multi-homed mobile networks on a journey base. We studied the problem under two different modes, the FFA mode and the IFA mode. We showed that the general FFA problem is NP-hard and proposed two heuristic algorithms for the FFA problem. For the four special cases of the FFA problem, we proposed optimal solutions based on the MCNF algorithm on the network flow models. We proved the IFA problem is NP-hard and proposed the MSCC-IFA algorithm to solve it. Extensive simulations have been conducted to study the impacts of capacity-demand ratio and other factors on the proposed algorithms.

Our results will make the mobility transparency in multi-homed mobile networks a reality and help in maximizing the revenue of the transportation system by providing seamless Internet connectivity to different users in the mobile environment.

References

1. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1993.
2. M. Bechler, O. Storz, W. Franz, and L. Wolf, "Efficient discovery of Internet gateways in future vehicular communication systems," *Proc. 57th IEEE Semiannual Vehicular Tech. Conf. (VTC)*, pp. 965-969, 2003.
3. J. Ben-Othman and F. Valois, "Multiservice allocation in hierarchical cellular networks," *Proc. 4th IEEE Symp. Computers and Communications*, pp. 80-86, 1999.
4. A. Chung and M. Hassan, "Traffic distribution schemes for multihomed mobile hotspots," *Proc. IEEE Vehicular Technology Conf.*, pp. 2127-2131, 2005.
5. M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, San Francisco, CA: Freeman, 1979.
6. D.K. Goldenberg, L. Qiu, H. Xie, Y.R. Yang, and Y. Zhang, "Optimizing cost and performance for multihoming," *Proc. SIGCOMM*, pp. 79-92, 2004.
7. W. Kellerer, C. Bettstetter, C. Schwingenschlögl, *et al.*, "(Auto) mobile communication in a heterogeneous and converged world," *IEEE Personal Communications*, vol. 8, no. 6, pp. 41-47, 2001.
8. L. Li, B. Li, B. Li, and X. Cao, "Performance analysis of bandwidth allocations for multi-services mobile wireless cellular networks," *Proc. IEEE Wireless Communications and Networking Conf.*, pp. 1072-1077, 2003.

9. M. Mani and N. Crespi, "Handover criteria considerations in future convergent networks," *to appear in Proc. Globecom 2006*, San Francisco, CA.

10. J. Manner and M. Kojo, "Mobility related terminology," *(draft-ietf-seamoby-mobility-terminology-06.txt), Internet Draft, IETF,* Feb. 2004.

11. N. Montavont, T. Noel, and T. Ernst, "Multihoming in nested mobile networking," *Proc. Int'l Symp. Applications and the Internet-Workshops (SAINT)*, pp. 184-189, 2004.

12. C. Ng, E. Paik, T. Ernst, and M. Bagnulo, "Analysis of multihoming in network mobility support," *(draft-ietf-nemo-multihoming-issues-06)*, *Internet Draft*, *IETF*, Jun. 2006.

13. E. Perera, V. Sivaraman, and A. Seneviratne, "Survey on network mobility support," *Mobile Computing and Communications Review*, vol. 8, no. 2, pp. 7-19, 2004.

14. T.S. Randhawa and R.H.S. Hardy, "Performance analysis of multi-service cellular networks with mobile users," *Proc. IEEE Wireless Communications and Networking Conf.*, pp. 1390-1397, 2000.

15. M. Ronai, A. Petrescu, R. Tnjes, and M. Wolf, "Mobility issues in OverDRiVE mobile networks," *Proc. IST Mobile Summit*, pp. 16-18, 2003.