Relay Assisted Multicast with Markov Mobility

Xiaoying Gan, Chen Feng, Zhida Qin, Ge Zhang, Huaying Wu, Luoyi Fu, Xinbing Wang, Huadong Ma

Abstract—In this paper, we investigate the capacity and delay tradeoff under multicast scheme in the MANET, based on a general Markovian mobility model. To reduce the delay of the network, we propose a relay-assisted multicast scheme. Specifically, a two hop relay algorithm is developed, in which Lyapunov drift is utilized to derive the average packet delay. In addition, we utilize the cache in relay nodes and propose the two hop relay algorithm with redundancy. Theoretical analysis indicates that the network delay is significantly decreased while the capacity remains the same. To guarantee the fairness and efficiency of the network, a two hop relay selection algorithm with redundancy is proposed to decide which packet to serve in a queue. Moreover, the minimum energy function is applied to characterize the energy consumption of each node. We accordingly derive the accurate minimum energy function under the proposed relay-assisted multicast scheme. Furthermore, we design an efficient minimum energy algorithm, which pushes the actual energy consumption arbitrarily close to the minimum energy function at the cost of increasing delay. Theoretical results show that the optimal energy-delay tradeoff is achieved in our proposed algorithm. Numerous experiments are carried out to evaluate the performance of our proposed algorithms, where the experimental results well conform our theoretical findings.

Index Terms—Mobile ad hoc networks, multicast, capacity delay tradeoff.

1 Introduction

With the prevalence of smart phones, laptops and mobile sensors over the last years, the Mobile ad hoc network (MANET) has paved the way for numerous original and exciting applications. Typical examples include personal communications, earthquake rescue and sensor networks. These newly emerged applications typically require fast and reliable communications and thus have brought great challenges on the study of network delay in the MANET. MANETs, consisting of a set of fixed or mobile nodes, are characterized by intermittent connectivity and frequent network partitioning. The effect of MANETs on network performance is first considered by Gossglauser and Tse in [2]. A 2-hop relay algorithm is accordingly proposed under which the network capacity is a constant at the cost of very large delay. Although their communication scheme is not satisfactory in average delay, it pioneered the research for MANETs.

As an essential communication method for supporting information propagation in MANETs, multicast has attracted massive attention from researchers recently. Particularly, researchers have to deal with node mobility which leads to frequent topology changes. Multicast flows in wireless networks are envisioned to be predominant in numerous emerging practical situations, such as vehicular networks, online video sharing network and the edge computing. In order to meet these practical needs, it is imperative to study the performance of multicast in MANETs.

The tradeoff between capacity and delay in MANETs under store-carry-forward schemes has attracted extensive attention and is investigated by many researchers under numerous mobility models [3], [4], [5]. One of the significant and practical models is the Markovian Mobility model. In [4], Urgaonkar studied the network capacity of unicast under the Markovian Mobility model. The results imply the possibility to exploit node dependence to obtain a better performance of constant capacity and constant delay. Inspired by them, we intend to improve the network performance under multicast traffic pattern with the Markovian Mobility model.

We study capacity and delay tradeoffs in Markovian mobility with multicast traffic. To investigate the tradeoff from an exact and general perspective, we aim to obtain the exact formula of capacity and delay under relay-assisted multicast scheme with general Markovian Mobility. In the cell partitioned MANET, not only same cell communications is allowed, we also allow users to communicate between adjacent cells. In the MANET, users are statically partitioned into groups of size $k+1$, where each node sends packets to the other k nodes. The mobility of users is captured by the general Markovian mobility model with well-defined steady-state location distribution, which makes it possible to predict the distribution of all users. We investigate the relay-assisted multicast performance of both one duplicate and multiple duplicates scenarios. Our main contributions are summarized as follows.

- The *Two hop relay algorithm* is designed in which Lyapunov drift is exploited to obtain exact capacity and delay. Our algorithm is proven to achieve a better delay and capacity tradeoff than that of [6]. Our

This work was supported by National Key R&D Program of China 2018YFB1004705, NSF China (No. 61822206, 61602303, 61532012, 61672342, 61671478, 61829201), CCF Tencent RAGR 20180116, the Science and Technology Innovation Program of Shanghai under Grant (17511105103, 18510761200) and the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2018D06.

Xiaoying Gan is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China, and National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, P. R. China (email: ganxiaoying@sjtu.edu.cn).

Chen Feng, Zhida Qin, Ge Zhang, Huaying Wu, Luoyi Fu, Xinbing Wang are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China (email: {fengchen, zanderqin, sjtu.zg123, wuhuaying, yiluofu, xwang8}@sjtu.edu.cn).

Huadong Ma is with the School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, P. R. China (email: mhd@bupt.edu.cn).

Some preliminary results of this paper have been presented at IEEE INFOCOM Orlando, FL, 2012 [1].

analysis shows that network capacity only depends on the steady location distribution of users.

- To further improve the delay performance, we allow cache and packet redundancy in multiple relay nodes, proposing a *Two hop relay algorithm with redundancy*. Theoretical analysis shows that the delay in the MANET under Markovian mobility can be largely improved by providing more packet copies. Meanwhile, the capacity-delay tradeoff remains the same as the algorithm without redundancy. In case of high level redundancy in node's cache, we propose the *two hop relay selection algorithm with redundancy*. It is revealed that its delay is similar to the previous redundancy allowed algorithm without impairing the capacity.
- The minimum time average power is studied to maintain the network stability. Our analysis shows that the minimum energy consumption grows linearly with the stable capacity of the network. Furthermore, we present a *Minimum energy algorithm* which pushes the actual energy consumption arbitrarily close to the minimum energy by sacrificing the delay performance. It is shown that the optimal energy and delay tradeoff is achieved in our algorithm.

The reminder of the paper is organized as follows. Section 2 is devoted to the related works. In section 3 we present the system model and some definitions. The exact expression for multicast capacity is investigated in section 4 and we propose a *Two hop relay algorithm* to realize this capacity. In section 5 we study the cache-enabled network with redundancy and show the delay performance. The expression of the minimum energy function and the tradeoff between energy and delay are studied in section 6. We conduct the numerical simulation in section 7. Finally, we conclude this paper in section 8. For readability, some proofs are deferred to the Appendix.

2 Related Works

Capacity is one of the most critical metric for MANET performance, which has drawn extensive efforts on improving the network capacity. Mobility is shown to increase the capacity of wireless networks while incurs noticeable delay. This has spawned numerous studies on balancing the capacity and delay. Neely and Modiano [7] noticed the problem and showed that there is a tradeoff of $\Omega(n)$ between delay and capacity under i.i.d. mobility model. And recently, the capacity delay tradeoff is studied in the practical random way point model [8]. The broadcast capacity and delay scaling is studied in highly mobile wireless networks [9]. Ren *et al.* investigated the impact of the directional antenna on the capacity under delay constraints [10]. Luo *et al.* shows that the performance of capacity and delay is improved with the help of supportive infrastructures [11]. The secrecy constraint is further in [12] considered to study its impact on capacity and delay.

The relay-assisted technique with cache has been widely used in the wireless network. In recent works [13], [14], [15], the authors employ the relay-assisted transmission allowing cache in wireless networks to improve the network capacity throughput. The multicast capacity in relay-assisted network is also studied in [16]. A general theoretical framework is proposed for the relay assisted MANET [17]. Some works focus on the multicast capacity-delay performance in mobile wireless ad hoc networks. Zhang *et al.* in [3] investigated the tradeoff of capacity-delay in cognitive radio MANETs. The multicast network performance for i.i.d. mobility model is explored in [5]. The random walk mobility model is studied in [18]. A 2-D independently and identically distributed mobility model is investigated in [19]. Further, the impacts of user mobility correlations on the network capacity and delay were analyzed in [20]. Yang in [21] *et al* investigated the cooperative multicast performance with relay assisted transmissions in a MANET. However, none of them has investigated the impacts of Markovian Mobility model on the network multicast capacity and delay performance.

A flurry works are devoted to optimizing the energy consumption in wireless networks from various aspects, such as, routing protocol [22], MAC (Media Access Control) protocol [23] and network coding [24]. Neely further considered the delay and studied the energy delay tradeoff in the context of multiuser [25]. A game theory framework is established in [26] to satisfy both delay and energy constraints. The relation between the delay and maximal energy efficiency is explored in [27] and non-tradeoff relation is found under certain conditions. The stochastic power control problem is formulated in [28] and power control algorithms are accordingly designed satisfying delay constraints. Efficient transmission schemes with theoretical performance guarantee are proposed in [29] to optimize the energy consumption in delay-constrained MANETs.

3 System Model

Cell-Partitioned Network Model: The MANET is partitioned into C nonoverlapping small cells with arbitrary shapes and sizes, as shown in Fig. 1. Each cell $c \in \{1, 2, ..., C\}$ is adjacent to a constant number B_c of cells, and the maximal value of B_c is smaller than a finite constant J. N mobile nodes move from cells to cells independently under a mobility model, and each node is allowed to visit any cell in the network. We define the node-per-cell density as $\theta = N/C$. We assume each node is equipped with a cache and its capacity is limited.

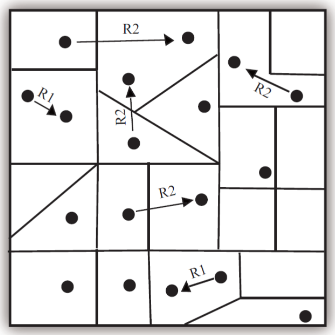


Fig. 1. A cell partitioned network: two nodes communicate with each other in the same cell at rate R_1 or between adjacent cells at rate R_2 ($R_1 \geq R_2$).

Markovian Mobility Model: Dividing time into constant length slots, we model the mobility process of users as the

TABLE 1
Frequently Used Notations

C	The number of non-overlapping cells in the MANET
N	The number of users in the network
θ	The node-per-cell density of the network
$\mathbf{P}$	The transition probability matrix of the Markov process
π	$\pi = \{\pi_c\}_{1\times C}$, The steady state distribution of the Markov process
k	The number of destination a packet must be transmitted to
R_1	The rate of the same cell transmission
R_2	The rate of the adjacent cell transmission
λ	The arrival rate of packets at each source node
μ	The multicast capacity
$\overrightarrow{U}(t)$	The vector of the unserved number of packets queued in each node at slot t
$\overline{D}$	The average delay of each source destination pair
T_i	The time needed for the ith destination to obtain the packet
$\Phi(\lambda)$	The minimum energy function
$\overline{e}$	The average energy consumption of each node
$U_s(t)$	The number of packets waiting in the buffer of the source at slot t
$\overline{U}_s$	The number of packets waiting in the buffer of the source in the steady state

finite state ergodic Markov Chain. We define P_{ij} as the conditional probability that a node moves to cell j in the current slot under the condition that it's in cell i in the last slot. In this model, nodes stay in their current cell in a time slot, and possibly move to an adjacent cell in the next slot. This implies that any entry P_{ij} of the transition probability matrix $\mathbf{P}$ of the Markov Chain is positive only when j is adjacent to i, i.e., $j \in B_i$. Otherwise, its value is 0. Note that such mobility process results in a well-defined steady-state location distribution $\pi = \{\pi_c\}_{1\times C}$ over all the cells which satisfies $\pi\mathbf{P}= \pi$ and the same is to all the nodes. Therefore, the delay and relay algorithm will be influenced by the distribution. Furthermore, the transition matrix $\mathbf{P}$ doesn't need to be known.

Multicast Traffic Model: We assume the number of users N can be divided by $k + 1$. The network is uniformly and randomly divided into numerous groups, each of which contains $k + 1$ nodes. We assume that packets from a node in some group must be transmitted to all the other k nodes in the group. Nodes not in the group can act as relays. Thus, each node i can be a source node involving k randomly and independently chosen destinations among all the other nodes in the network. Packets arriving at each source node are assumed to follow an i.i.d. process $A_i(t)$ of rate λ_i, and the maximal number of arrivals is bounded, i.e., $\max\{A_i(t)\} < A_{max}$. The relationship will not change when the nodes roam around. Thus, the source could transmit packets to its k destinations respectively under the same MANET environment.

Relay-assisted Communication Model: In our model, we assume each node has two states *Silent* and *Active*, where *Silent* indicates that the node stays idle without transmitting any packets and *Active* indicates the node is transmitting packets. It is assumed that packet transmission only happens between nodes in the same cell or adjacent cells. The rate of the same cell transmission is denoted as R_1 and the rate of the adjacent cell transmission is denoted as R_2, thus $R_1 \geq R_2$. Moreover, to relieve the interference, at any given slot, at most one transmitter is allowed to transmit in a cell and we apply different orthogonal communication channels among adjacent cells. Nodes can easily judge whether they are inside an active zone or not at any time slot. As to relay transmission, a packet can be transmitted to the destination by either a direct S-D transmission or indirect two hop transmissions.

Definition of Capacity: Packets are transmitted through the MANET according to certain scheduling scheme. The network is said to be stable if the arrival rate λ_i is achievable for all the nodes such that the queue of each node will not be infinite as the time tends to infinity. For simplicity, the arrival rate for all the nodes is assumed to be the same, denoted as λ. The capacity of the network is defined as the maximal arrival rate λ that the network can stably support over all possible scheduling and routing algorithms.

Definition of Delay: The delay of a packet is the time it takes for the packet to reach all its k destinations after it leaves the source. The total network delay is obtained by averaging the delay over all packets, all source-destination pairs and all random network configurations in the long term.

Frequently used notations are summarized in Table 1 for the reader's convenience.

4 Multicast Capacity with Two Hop Relay

In this section, we begin with computing the exact multicast capacity μ, which is the maximum rate that the MANET can stably support. Then, we propose the *Two hop relay algorithm* for any $\lambda < \mu$ to hold the MANET stable and derive the corresponding delay. We further show that the tradeoff between delay and capacity under this algorithm is superior than that of [6].

4.1 Exact Multicast Capacity

In a mobile ad hoc wireless network, the source can transmit packets to the destination by either a direct S-D transmission or indirect two hop transmissions via relays. To be specific, in our MANET model, there are four kinds of transmission pattern, i.e., 1) a direct S-D transmission in the same cell; 2) a source to relay (or relay to destination) transmission in the same cell; 3) a direct S-D transmission between adjacent cells; 4) a source to relay (or relay to destination) transmission between adjacent cells. However, for a specific packet, only one of the four kinds of transmissions takes place in one slot.

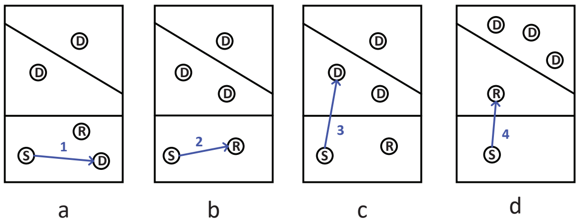


Fig. 2. An example with $k = 3$ to illustrate the four kinds of transmission patterns. a) The first kind; b) the second kind; c) the third kind; d) the fourth kind.

Since the two hop relay transmission consumes more network resources, the direct S-D transmission has priority when both the direct or two-hop transmissions happen simultaneously. Moreover, note that the same cell transmission rate R_1 is no less than the adjacent cell transmission rate R_2, thus the first kind of transmission has the highest priority and the fourth kind of transmission has the lowest priority. The second and third kinds of transmissions need to be discussed at two different scenarios. For a specific packet to be transmitted to its destination, these two kinds of transmissions may happen at the same time. The second kind of transmission takes two hops and involves the same cell transmission, while the third kind of transmission takes only one hop and involves the adjacent cell transmission. Thus, in order to maximize the network capacity, when $\frac{R_1}{2} \geq R_2$ (i.e., $R_1 \geq 2R_2$), the second kind of transmission should have a higher priority than the third kind of transmission. And when $\frac{R_1}{2} < R_2$ (i.e., $2R_2 > R_1 \geq R_2$), the third kind of transmission should have priority over the second kind of transmission.

We use an example in Fig. 2 to further illustrate and compare these four kinds of transmission pattern. The multicast group in Fig. 2 is distributed in three cells with $k = 3$. The node with symbol S denotes the source node; the node with symbol R denotes the relay node; the nodes with symbol D denote the destination node. As we can see, there is one source node, three destination nodes and one relay node. Note that only one of these four kinds transmission patterns can take place at one slot. We will discuss these four figures as follows.

- In Fig. 2a, the source node, the relay node, and the destination node are in the same cell. Since the direct S-D transmission has the highest priority, the source node will transmit the packet directly to the destination node in the same cell.
- In Fig. 2b and 2c, the source node and the relay node are within the same cell, while two destination nodes are in adjacent cells. In this way, a relay transmission in the same cell or a direct S-D transmission may take place at the same time. According to the discussions above, when $\frac{R_1}{2} \geq R_2$, the second kind of transmission has priority over the third, thus the source node will transmit the packet via the relay node, as shown in Fig. 2b. When $\frac{R_1}{2} < R_2$, the third kind of transmission has priority over the second kind, the source node will directly transmit the packet to the destination node in the adjacent cell, as shown in Fig. 2c.
- In Fig. 2d, there is no destination node in the same cell and adjacent cell of the source node. In this way, only the fourth kind of transmission can happen. Thus, the source will transmit its packet to the relay node in the adjacent cell.

To sum up, when $R_1 \geq 2R_2$, the priority order of the four kinds of transmissions is: $1) \succ 2) \succ 3) \succ 4)$; and when $2R_2 > R_1 \geq R_2$, the priority order of the four kinds of transmissions is: $1) \succ 3) \succ 2) \succ 4)$. Following this priority order of transmissions in different cases, we can compute the capacity μ. Thus, we show the multicast capacity μ in the following theorem:

Theorem 1. *In a cell partitioned MANET with N nodes and C cells, nodes move around according to the Markovian mobility model described in Sec. 3. The mulicast capacity of the MANET is given by:*

$$\mu = \begin{cases} \frac{R_1 q + R_1 p + R_2 q' + R_2 p'}{(k+1)\theta} & \textit{if} \quad R_1 \geq 2R_2, \\ \frac{2R_1 q + 2R_2 q'' + R_1 p'' + R_2 (p' - q')}{(k+1)\theta} & \textit{if} \quad 2R_2 > R_1 \geq R_2, \end{cases}$$

where $q = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding a source-destination pair in cell c in a slot];

$p = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding at least 2 nodes in cell c in a slot];

$q' = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding exactly 1 node in cell c and its destination in an adjacent cell in a slot];

$p' = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding exactly 1 node in cell c and at least 1 node in an adjacent cell in a slot];

$q'' = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding no source-destination pair in cell c but at least 1 source-destination pair in an adjacent cell in a slot];

$p'' = \frac{1}{C}\sum_{c=1}^{C} Pr$[finding no source-destination pair in cell c and any adjacent cell but at least 2 nodes in cell c in a slot].

Based on the system model and assumptions, we can calculate these probabilities as follows. The detailed derivation of the probabilities is provided in Appendix A.

$$q = \frac{1}{C}\sum_{c=1}^{C} \left(1 - \left[(1-\pi_c)^{k+1} + \binom{k+1}{1}\pi_c(1-\pi_c)^k\right]^{\frac{N}{k+1}}\right),$$

$$p = \frac{1}{C}\sum_{c=1}^{C} \left(1 - (1-\pi_c)^N - \binom{N}{1}\pi_c(1-\pi_c)^{N-1}\right),$$

$$q' = \frac{1}{C}\sum_{c=1}^{C} \left(k\binom{N}{1}\pi_c(1-\pi_c)^{N-1}\Pi_{adj}(c)\right),$$

$$p' = \frac{1}{C}\sum_{c=1}^{C} \left((1 - (1-\Pi_{adj}(c))^{N-1})\binom{N}{1}\pi_c(1-\pi_c)^{N-1}\right),$$

$$q'' = \frac{1}{C}\sum_{c=1}^{C}\sum_{i=1}^{\frac{N}{k+1}} \frac{(k+1)^i\binom{\frac{N}{k+1}}{i}}{\binom{N}{i}}\binom{N}{i}\pi_c^i(1-\pi_c)^{N-i}(1-(1-\Pi_{adj}(c))^{ki}),$$

$$p'' = \frac{1}{C}\sum_{c=1}^{C}\sum_{i=2}^{\frac{N}{k+1}} \frac{(k+1)^i\binom{\frac{N}{k+1}}{i}}{\binom{N}{i}}\binom{N}{i}\pi_c^i(1-\pi_c)^{N-i}(1-\Pi_{adj}(c))^{ki},$$

where $\Pi_{adj}(c)$ is the sum of the conditional steady-state probability of a node being in any adjacent cell of cell c given that this node is not in cell c, i.e., $\Pi_{adj}(c) = \frac{1}{1-\pi_c}\sum_{i \in B_c} \pi_i$.

Proof. The proof of the multicast capacity is provided in Appendix B. □

4.2 Two Hop Relay Algorithm

In this subsection, we propose a *Two hop relay algorithm* which keeps the network stable for any arrival rate λ. We only present the scheduling algorithm under the first case $R_1 \geq 2R_2$. Similarly, the algorithm in the other case can be obtained.

Two hop relay algorithm: In the algorithm, each slot is further equally divided into two subslots. With probability $\frac{1-\delta}{2}$, the transmission follows the first procedure. Otherwise, the transmission follows the second procedure.

1) In an odd subslot, nodes transmit in the source-to-relay mode:

- If a cell has at least two nodes, one node is randomly chosen as a sender and another node as the receiver. If the sender needs to transmit a new extraneous packet, the packet is relayed to the receiver at rate R_1 and then deleted from the buffer. Else, stay idle.
- If a cell only has one node and its adjacent cells have at least one node, the only node is chosen as a sender and another random node in adjacent cells is chosen as a receiver. If the sender needs to transmit a new extraneous packet, the packet is relayed to the receiver at rate R_2 and then deleted from the buffer. Else, stay idle.

2) In an even subslot, nodes transmit in the relay-to-destination mode:

- If a cell has at least two nodes, one node is randomly chosen as a sender and another one as a receiver. If the sender received packets from other nodes scheduled for the receiver but the receiver has not obtained the packet yet, then the latest packet is chosen and transmitted at rate R_1. If the packet has been received by all the destinations, it is deleted from the buffer of the sender. Else, stay idle.
- If a cell only has one node and its adjacent cells have at least one node, the only node is chosen as a sender and another random one as the receiver. If the sender received packets from other nodes scheduled for the receiver but the receiver has not obtained the packet yet, then the latest packet is chosen and transmitted at rate R_2. If the packet has been received by all the destinations, it is deleted from the buffer of the sender. Else, stay idle.

In the second procedure, the source node regards all the nodes it meets as relay nodes. Thus, according to the scheduling scheme, all the packets will be delivered along a 2-hop path: source-relay-destinations. The performance of the algorithm can be evaluated by applying the Lyapunov Drift analysis [30]. Let $\overrightarrow{U}(t) = \big(U_1(t), U_2(t), ..., U_n(t)\big)$ be the vector of the unserved number of packets queued in each node at slot t. The Lyapunov function $L\big(\overrightarrow{U}(t)\big)$ is defined as a non-negative function of $\overrightarrow{U}(t)$. Then we present the following lemma:

Lemma 1. (Network Stability-Sufficient Condition using Lyapunov Drift)*: If there exists a positive integer d such that for all slots t, the Lyapunov function evaluated d steps into the future satisfies:*

$$\mathbb{E}\{L(\overrightarrow{U}(t+d)) - L(\overrightarrow{U}(t)) | \overrightarrow{U}(t)\} \le B - \sum_i \theta_i U_i(t), \quad (1)$$

for some positive constants B, $\{\theta_i\}$, and if $\mathbb{E}\{L(\overrightarrow{U}(t))\} < \infty$ for $t \in \{0, 1, ..., d-1\}$, then the network is stable, and:

$$\limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \big[\sum_i \theta_i \mathbb{E}\{U_i(\tau)\}\big] \le B. \quad (2)$$

This lemma has been proved by using a telescoping series argument in [30]. On this basis, we obtain Theorem 2.

Theorem 2. *Consider a cell partitioned MANET with N users and C cells, under the* two hop relay algorithm, *if the input rate λ for each node such that $\lambda = \rho\mu$ for some $0 \le \rho < 1$, and nodes move around following the Markovian mobility model as described in Sec. 3, the average packet delay $\overline{D}$ satisfies:*

$$\overline{D} \le \frac{2dBN\log k}{(k+1)\lambda\mu\varphi(1-\rho)},$$

where $\delta = \frac{1-\rho}{2}$, B is a constant given by (10), φ is a positive constant given by $\varphi \triangleq \frac{R_1 p + R_2 p'}{R_1 p + R_2 p' + R_1 q + R_2 q'}$.

Proof. The proof is provided in Appendix C. □

5 Two Hop Relay with Redundancy

As mentioned in the system model, each node is equipped with a cache. Therefore, in this section, we utilize redundancy (i.e., more than one relays to hold the same packet) to improve the delay performance. However, redundancy poses new challenges on network design. The first one is how to delete the duplicated packet. The scheduling scheme applied to the network should ensure that all the redundancies of each packet should be deleted when all of the k destinations have received the packet. The second challenge lies in the minimum delay. The purpose of this section is designing an efficient algorithm with a smaller delay.

In general, the cache capacity for each node is limited. When the number of packet replications waiting at the relays' cache is large, the service time for packets queue cannot be ignored. In this way, we will talk about two cases: 1) the length of packets queue in relay nodes is small; 2) the queue length in relay nodes is large. We will study the network performance for these two cases in two subsections, respectively.

5.1 Two Hop Relay with Redundancy

Assume the new packet arriving at the source node is marked with a sender number SN. To indicate which packet is needed, the destination will send a request number RN. We next present the proposed relay algorithm with redundancy, where each node is allowed to hold at most m redundancies. That is to say, in the network, at most m nodes (including the source node) hold the same packet.

Two hop relay algorithm with redundancy: Each slot is further equally divided into two subslots. With probability $\frac{1-\delta}{2}$, the transmission follows the first procedure. Otherwise, the transmission follows the second procedure.

1) In an odd subslot, nodes transmit in the source-to-relay mode:

- If a cell has at least two nodes, one node is randomly chosen as a sender and another node as the receiver. If the sender needs to transmit a new extraneous packet with a sender number SN, the packet is relayed to the receiver at rate R_1 until m duplicates are delivered to different relay nodes (maybe some are destinations), or until the SN packets are received by all the k destinations. After that, the sender number is added to $SN+1$. Else, stay idle.
- If a cell only has one node and its adjacent cells have at least one node, the only node is chosen as the sender and another random node in adjacent

cells is chosen as a receiver. If the source needs to transmit a new extraneous packet marked with SN, the packet is relayed to the receiver at rate R_2 until m duplicates are delivered to different relay nodes (maybe some are destinations), or until the SN packets are received by all the k destinations. After that, the sender number is added to $SN+1$. Else, stay idle.

2) In an even subslot, nodes transmit in a relay-to-destination mode: The relay-to-destination transmission happens with a handshake:

- The receiver delivers its current RN number for the packet it needs.
- The transmitter should transmit the packet with $SN = RN$ following the next two sequences:
    - If a cell has at least two nodes, one node is randomly chosen as a sender and another node as the receiver. If the sender received a packet with number SN from other nodes that is destined for the receiver and has not been received by the receiver yet, then transmit the packet at rate R_1. If the packets with $SN = RN$ are received by all the k destinations, delete these packets from the buffer of all the senders. Else, stay idle.
    - If a cell only has one node and its adjacent cells have at least one node, the only node is chosen as the sender and another random node in adjacent cells is chosen as a receiver. If the sender received a packet with number SN from other nodes that is destined for the receiver and has not been received by the receiver yet, then transmit the packet at rate R_2. If the packets with $SN = RN$ are received by all the k destinations, delete these packets from the buffer of all the senders. Else, stay idle.

To evaluate the algorithm's performance, we study an extreme case which leads to the largest delay. First, the packet will be delivered to m relay nodes with time T_f when none of its k destinations receive the packet; second, under the condition that m relay nodes hold the packet, it takes time T_s for all its k destinations receive the packet. Hence, the delay is $T = T_f + T_s$. We next begin to calculate $\mathbb{E}(T_f)$ and $\mathbb{E}(T_s)$ respectively.

$\mathbb{E}(T_f)$: Similar to the calculation of D in the previous section, we can derive T_f by converting it to the cover time of a Markov Chain with m states. Recall that the first procedure of the algorithm only happens in odd subslots (i.e., *source-to-relay* transmission), thus the average hitting time of the state i (i.e., the ith relay node receives a copy of the packet) is in the order of $\Theta(\frac{1}{k})$ by applying *Lemma 1* and Little's law. Similarly, by *Lemma 2*, we have $\mathbb{E}(T_f) = \Theta(\frac{\log m}{k})$.

$\mathbb{E}(T_s)$: This problem is set in a scenario where m relays already hold the same replications of a packet. We try to find the time needed for all the k destinations receive the packet. This problem is converted to solving a m to k cover time problem for a Markov chain. First, let T' be the time needed for one of the k destinations to receive the packet when there is only one relay node have the packet. In this case, we then present the following important lemma.

Lemma 2. *Let $R_1, R_2, ..., R_m$ denote the m relays holding the same duplicate of a packet, and $D_1, D_2, ..., D_k$ denote its k destinations requiring the packet. Then the m to k cover time is:*

$$\mathbb{E}[T^{m\to k}] = \frac{\log k}{m}\mathbb{E}[T']. \tag{3}$$

Proof. First we derive the expected time for the m relays to cover any one destination node, which is the m to 1 cover time $T^{m\to 1}$. Let $I(R_i, D_j, t)$ be the indicator function that any relay R_i transmits a duplicate of the packet successfully to any destination D_j at time slot t. Then we have:

$$\begin{aligned}
\mathbb{E}[T^{m\to 1}] &= \mathbb{E}[min\sum_{i=1}^{m} I(R_i, D_j, t) \geq 1] \\
&= \mathbb{E}[max\sum_{i=1}^{m} I(R_i, D_j, t) = 0] \\
&= \frac{1}{1-(1-Pr(I(R_i, D_j, t)=1))^m} = \frac{\mathbb{E}[T']}{m}.
\end{aligned}$$

By using the same method as lemma 2, we have that the m to k cover time $\Theta(\frac{\log k \mathbb{E}(T')}{m})$. □

Recall that the second procedure only happens in even subslots (i.e., *relay-to-destination* transmission), then $\mathbb{E}[T']$ is in the order of $\Theta(\frac{N}{k})$ by applying *Lemma 2* and Little's law. Thus from *Lemma 3*, we have $\mathbb{E}(T_s) = \Theta(\frac{N\log k}{mk})$.

Combining the two procedures we have $\mathbb{E}(T) = \mathbb{E}(T_f) + \mathbb{E}(T_s) = \Theta(\frac{\log m}{k}) + \Theta(\frac{N\log k}{mk})$. It is clear that when the value of m changes, the value of $\Theta(\frac{\log m}{k})$ and $\Theta(\frac{N\log k}{mk})$ varies. Thus we have theorem 3:

Theorem 3. *In the cell partitioned network under the* two hop relay algorithm with redundancy, *the network capacity can achieve $\Theta(\frac{1}{km})$ with corresponding delay of* $\begin{cases}\Theta(\frac{N\log k}{km}) & m = o(N),\\ \Theta(\frac{\log m}{k}) & m = \Theta(N).\end{cases}$

Thus when $m = \Theta(N)$, the network delay is $\mathbb{E}(T) = \Theta(\frac{\log N}{k})$; when $m = o(N)$, the network delay is $\Theta(\frac{N\log k}{mk})$. Hence, when m takes different values, the network delay is in the interval $[\Theta(\frac{\log N}{k}), \Theta(\frac{N\log k}{k})]$. Also, since on average the redundancy of each packet is m, the network capacity is $\Theta(\frac{1}{km})$. The tradeoff between delay and capacity remains the same as the tradeoff without redundancy.

5.2 Two Hop Relay Selection with Redundancy

In the previous subsection, we studied the upper bound of delay when the number of packets in the buffer is small enough to regard the service time as zero. But when many packets are waiting in the buffer, the queue of packets cannot be ignored. That is, the packet arrived earlier should be served with priority. In this way, we propose an algorithm to guarantee the fair and efficiency of the network.

Similar to the two hop relay algorithm with redundancy in Section 5.1, a new packet at the source node is marked with number SN, and the request at the destination is marked with number RN. Each packet is allowed to have at most m replications. We assume each node maintains

3 individual queues in the buffer: The first queue stores packets locally generated; the second queue stores packets whose m replications have already been sent out but the destination has not received yet; the third queue stores packets from other $S-D$ flows. The transmission sequence for each packet is determined by the three queues. The new packet at the first queue will wait until the replications of packets with higher priority have been sent out, which results in a period of service time at the source node S. Similarly, the destination node D will send out requests of the packet i just after it has received packets with higher priority, which results in a period of service time at D. Then we have the following algorithm.

Two Hop Relay Selection Algorithm With Redundancy: All packets have been stamped with its arrival time in the send number SN. Each slot is further divided into two equal subslots. With probability $\frac{1-\delta}{2}$, the transmission follows the first procedure. Otherwise, the transmission follows the second procedure.

1) In an odd subslot, nodes transmit in the source-to-relay mode:

- A node is randomly selected as a transmitter S and another node as a receiver D, then the S initiates a handshake to obtain the number RN from the destination D, and the source S executes the same actions as the two hop relay algorithm .
- If a cell only has one node and its adjacent cells have at least one node, the only node is selected as the transmitter S and a random node in the adjacent cells is chosen as the receiver D, then the S initiates a handshake to obtain the number RN from the destination D, and S executes the same actions as the two hop relay algorithm. Else, stay idle.

2) In an even subslot, nodes transmit in the relay-to-destination mode: the relay-to-destination transmission happens with a handshake to obtain the RN from the destination D:

- Firstly, get the number SN from the first queue. If $SN > RN$, then S retrieves the packet i with $SN = RN$ from its second queue and sends packet i to the destination D. Else, if $SN = RN$, then S sends packet i directly to D.
- Secondly, get SN from the third queue. If S has a packet i in the third queue dedicated to D with $SN = RN$, then S sends packet i to node D. Else, if D has one copy of i, S remains idle. Moreover, if the D has not stored the copy of i, S sends a copy of packet i to D. If all copies for packet i have been distributed, S puts packet i to the end of the second queue and moves ahead the remaining packets in the first queue.

To evaluate the performance of this algorithm, we also study the extreme case which leads to the largest delay. Firstly, the packet has to wait until packets with higher priority are transmitted (suppose the number of such packets is z). Secondly, the packet should be delivered to m relays when none of its k destinations receive the packet. Thirdly, under the condition that m relay nodes hold the copies of the packet, it takes a time period for the packet to reach all its k destinations. Now we give a proof that the largest delay of the first step should be taken from the service time at the destination node D instead of the service time at source node S.

Remark 1. *For a given time slot and a source-to-destination pair, we use p_1, p_2 and p_3 to denote the probability that the source node S conducts a successful transmission, the probability that the S conducts a source-to-destination transmission and the probability that the S conducts other type of transmission, respectively. Then we have*

$$p_1 = p+p', \quad p_2 = q+q', \quad p_3 = p_1-p_2 = p+p'-q-q'. \tag{4}$$

The definitions of p, q, p' and q' could be found in Theorem 1.

Lemma 3. *For a source-to-destination pair, suppose that there are m copies of packet i in the network when the destination node D starts to request for the i, $1 \leq z \leq f$. We use T_r and T_d to denote the corresponding service time of packet at the S and the D, respectively. Then we have*

$$\mathbb{E}\{T_r\} < \mathbb{E}\{T_d\}. \tag{5}$$

Proof. For a given time slot, Let P_4 denote the probability that the destination node D will receive packet i at the next time slot, and P_5 denote the probability that the S will successfully deliver a new copy of i, respectively.

In the next time slot, the destination node may receive packet i either from S or from one of the $m-1$ relays. Then we have

$$\begin{aligned} P_4 =& p_2 + \frac{m-1}{2}(1-\frac{1+J}{C})\sum_{t=0}^{N-3}\binom{N-3}{t}\Big[\sum_{k=0}^{t}\binom{t}{k}(\frac{1}{C})^{k+1}\cdot \\ & (\frac{J}{C})^{t-k}(1-\frac{1+J}{C})^{N-3-t}\frac{1}{t+1}(\frac{1}{k+2}+\frac{J}{k+1})\Big] \\ =& p_2 + \frac{p_3(m-1)}{2N-4}. \end{aligned} \tag{6}$$

According to the source-to-relay transmission, a relay node is randomly selected from the one-hop neighbors of node S. Therefore, the S can successfully deliver packet i if a node instead of these $m-1$ relay nodes that have already received copies is selected as receiver in the source-to-relay transmission. We can obtain that

$$\begin{aligned} P_5 =& \frac{1}{2}(1-\frac{1+J}{C})\sum_{k=1}^{N-2}\binom{N-2}{k}\Big[\sum_{i=0}^{k}\binom{k}{i}(\frac{1}{C})^{i}(\frac{J}{C})^{k-i}\cdot \\ & (1-\frac{1+J}{C})^{N-2-k}\frac{1}{i+1}\frac{N-m-1}{N-2}\Big] \\ =& \frac{p_3(N-m-1)}{2N-4}. \end{aligned} \tag{7}$$

It is easy to understand that:

$$\begin{aligned} \mathbb{E}\{T_d\} &= \frac{1}{P_4} = \frac{1}{p_2+\frac{p_3(m-1)}{2N-4}} \geq \frac{2N-4}{p_3(m-1)}, \\ \mathbb{E}\{T_r\} &= \frac{1}{P_5} = \frac{1}{\frac{p_3(N-m-1)}{2N-4}} = \frac{2N-4}{p_3(N-m-1)}. \end{aligned} \tag{8}$$

For $m = o(N)$, we have $\mathbb{E}\{T_r\} < \mathbb{E}\{T_d\}$. □

Hence, in this case the delay is $T = T_d + T_f + T_s$, where T_d, T_f and T_s are the time for the three steps of the above extreme case respectively. It is easy to derive that $T_d = zT_s$, where $z = \frac{2k}{p_3 \log k}$, thus from Lemma 3 we have $\mathbb{E}(T_s) = \Theta(\frac{N \log k}{mk})$, $\mathbb{E}(T_d) = \Theta(\frac{zN \log k}{mk})$ and the previous section proves that $\mathbb{E}(T_f) = \Theta(\frac{\log m}{k})$. Thus we have *Theorem 4:*

Theorem 4. *In the cell partitioned MANET under the* Two Hop Relay Selection Algorithm With Redundancy, *the network capacity can achieve $\Theta(\frac{1}{km})$ with corresponding delay of $\Theta(\frac{zN \log k}{km})$.*

From *Theorem 4*, we know that when $m = o(N)$, the network delay is $\Theta(\frac{zN \log k}{mk})$. Hence, when z takes different values from $o(N)$ to $\Theta(N)$, the upper bound of delay is in the interval $[\Theta(\frac{N \log k}{k}), \Theta(\frac{N^2 \log k}{k})]$. The delay and capacity tradeoff remains the same as the tradeoff of the two hop algorithm when $z = o(N)$.

The main differences between the algorithm with and without relay selection lies in the network scenarios. When applying the two hop relay algorithm with redundancy, the length of packets in the node cache is assumed to be negligible. The two hop relay selection algorithm with redundancy considers much heavy traffic scenarios. In this way, the queue of packets at each node cannot be ignored. Such waiting time need to be considered in the network delay, since many packets are waiting at the nodes' cache. The two hop relay selection algorithm with redundancy is proposed to optimize transmission schedule for multiple packets waiting at the nodes' buffer. We further prove that the algorithm with relay selection can achieve the same capacity-delay tradeoff compared with the scenario without packet duplication.

6 Minimum Energy Consumption

In this section, we investigate the energy needed to transmit the packets across the MANET. We introduce a performance metric to evaluate the energy consumption named minimum energy function $\Phi(\lambda)$ (first defined in [4]): the minimum time average energy per node needs to keep the network of an input rate λ stable, over all scheduling and routing schemes.

Note that each node has two states: transmitting packets using full power and staying idle without consuming any power. For brevity, we only study the minimum energy function in the first case $R_1 \geq 2R_2$ in Theorem 5. The other case $R_2 \leq R_1 < 2R_2$ can be derived in a similar way.

Theorem 5. *In the case $R_1 \geq 2R_2$, the minimum energy function $\Phi(\lambda)$ can be exactly expressed in a piecewise linear function of the input rate λ:*

$$\Phi(\lambda) = \begin{cases} \frac{k\lambda}{R_1} & \textit{if} \quad C_1, \\ \frac{q}{\theta} + \frac{k+1}{R_1}\big[\lambda - \frac{R_1 q}{k\theta}\big] & \textit{if} \quad C_2, \\ \frac{p}{\theta} + \frac{k}{R_2}\big[\lambda - \frac{R_1 p}{(k+1)\theta} - \frac{R_1 q}{(k+1)k\theta}\big] & \textit{if} \quad C_3, \\ \frac{p+q'}{\theta} + \frac{k+1}{R_2}\big[\lambda - \frac{R_2 q' + R_1(kp+q)}{(k+1)k\theta}\big] & \textit{if} \quad C_4, \end{cases} \tag{9}$$

where C_1: $0 \leq \lambda < \frac{R_1 q}{k\theta}$, C_2: $\frac{R_1 q}{k\theta} \leq \lambda < \frac{R_1 p}{(k+1)\theta} + \frac{R_1 q}{(k+1)k\theta}$, C_3: $\frac{R_1 p}{(k+1)\theta} + \frac{R_1 q}{(k+1)k\theta} \leq \lambda < \frac{R_2 q' + R_1(kp+q)}{(k+1)k\theta}$ and C_4: $\frac{R_2 q' + R_1(kp+q)}{(k+1)k\theta} \leq \lambda < \mu$. Note that when $k = 1$, it becomes a unicast network, whose minimum energy function is the same as that of [4].

Next, we will prove the necessity and sufficiency of this function.

6.1 Proof of Necessity

Let $X_{ab}(T)$ denote the number of packets transmitted to all the k destinations during $(0, T)$ by exactly a same cell transmissions and b adjacent cell transmissions. Because of the stability and ergodicity of the Markov Chain, the time average energy consumption $\overline{e}$ satisfies:

$$\overline{e} \geq \sum_{a,b|a+b \geq k} (\frac{a}{R_1} + \frac{b}{R_2}) \frac{x_{ab}}{N}, \tag{10}$$

where $x_{ab} \triangleq \lim\limits_{n \to \infty} \frac{X_{ab}(T)}{T}$.

Note that x_{ab} can not be any value due to the system model and assumptions presented in Sec. 3. Specifically, x_{ab} must be in the constraint set $\Omega = \Omega_0 \cap \Omega_1 \cap \Omega_2 \cap \Omega_3$, whose definitions are follows:

$$\Omega_0 \triangleq \{x| \sum_{a,b|a+b \geq k} x_{ab} = N\lambda\}, \quad \Omega_1 \triangleq \{x| \frac{kX_{k0}}{R_1} \leq c_1\},$$

$$\Omega_2 \triangleq \{x| \frac{1}{R_1} \sum_{a \geq k} a x_{a0} \leq c_1 + c_2\},$$

$$\Omega_3 \triangleq \{x| \frac{1}{R_1} \sum_{a \geq k} a x_{a0} + \frac{k x_{0k}}{R_2} \leq c_1 + c_2 + c_3\},$$

where c_1 is the maximum rate of direct source to destination transmission opportunities within the same cell, c_2 is the maximum rate of source to relay (or relay to destination) transmission opportunities within the same cell and c_3 is the maximum rate of direct source to destination transmission opportunities between adjacent cells. After some calculation, we have $c_1 = Cq$, $c_2 = C(p - q)$ and $c_3 = Cq'$.

Next we define $m(x) \triangleq \sum\limits_{a,b|a+b \geq k} (\frac{a}{R_1} + \frac{b}{R_2}) \frac{x_{ab}}{N}$. Thus, the minimum energy problem can be converted to solving an optimization problem:

$$\overline{e} \geq \inf_{x \in \Omega} f(x). \tag{11}$$

Applying the similar method in [4], we can convert this optimization problem to a less strict one as follows:

$$\overline{e} \geq \inf_{x \in \tilde{\Omega}} g(x), \tag{12}$$

where $\Omega \subset \tilde{\Omega} = \tilde{\Omega}_0 \cap \tilde{\Omega}_1 \cap \tilde{\Omega}_2 \cap \tilde{\Omega}_3$ and $g(x) \leq f(x)$. The new constraint sets $\tilde{\Omega}_0$, $\tilde{\Omega}_1$, $\tilde{\Omega}_2$, $\tilde{\Omega}_3$ are defined as follows:

$$\tilde{\Omega}_0 \triangleq \Omega_0, \quad \tilde{\Omega}_1 \triangleq \Omega_1,$$

$$\tilde{\Omega}_2 \triangleq \{x| \frac{k x_{k0}}{R_1} + \frac{(k+1)}{R_1} \sum_{a \geq k+1} x_{a0} \leq c_1 + c_2\},$$

$$\tilde{\Omega}_3 \triangleq \{x| \frac{k x_{k0}}{R_1} + \frac{(k+1)}{R_1} \sum_{a \geq k+1} x_{a0} + \frac{k x_{0k}}{R_2} \leq c_1 + c_2 + c_3\}.$$

Following the similar method in [4], we can prove *Theorem 5*. The details of the derivation of the minimum energy function under four different conditions are provided in Appendix D.

6.2 Proof of Sufficiency

Note that the minimum energy function $\Phi(\lambda)$ derived in *Theorem 5* is a piecewise linear function about λ. In this subsection, we aim at designing a scheduling algorithm which stabilizes the network under different values of λ. Our scheduling schemes also yield the average energy consumption arbitrarily close to the minimum energy function $\Phi(\lambda)$. However, when the average energy cost becomes closer to $\Phi(\lambda)$, the network delay increases. It is still assumed that packets can only be delivered via at most two hops. The difference lies in that we prefer a smaller energy consumption even at the cost of delay. Thus, according to the value of the input rate λ, the algorithms greedily exploit part of the transmission opportunities similar to that of [4].

6.3 Minimum Energy Algorithm

In this part, we present the algorithm for the first case $0 \leq \lambda < \frac{R_1 q}{k\theta}$ and discuss its performance on average energy consumption and delay. The other three cases can be analyzed with similar method. We describe the minimum energy algorithm in the first case $0 \leq \lambda < \frac{R_1 q}{k\theta}$ as follows:

Minimum energy algorithm: If there is an S-D pair in the same cell, we randomly choose a pair. If a new packet arrives, then with probability $\beta\rho$ $(1 < \beta < \frac{1}{\rho})$ the packet is transmitted to the destination with rate R_1. If this packet has been received by all the destinations, delete it from the cache; else, stay idle.

We now begin to analyze the performance of the algorithm. When the input rate $\lambda = \frac{\rho R_1 q}{k\theta}$ $(0 < \rho < 1)$, packets can only be transmitted through direct S-D transmission within the same cell. Since such transmission costs one unit power, the expression of the average energy consumption $\overline{e}$ can be written in terms of the probability of it $\beta\rho C q$. Hence, we have

$$\bar{e} = \frac{\beta\rho C q}{N} = \frac{\rho q}{\theta} + \frac{(\beta-1)\rho q}{\theta} = \Phi(\lambda) + \frac{(\beta-1)\rho q}{\theta}. \quad (13)$$

As for the corresponding delay, we only analyze it in order sense. For simplicity, we assume that at a time slot, with probability ρ, $\lambda = \frac{R_1 q}{k\theta}$ packets arrive at a node. Otherwise, no packet arrives. This assumption ensures that the average input rate λ is still $\frac{\rho R_1 q}{k\theta}$. Recall the probability that an S-D pair in the same cell at a slot is q. Therefore during a time slot, if no new packet arrives, packets queued in the cache of the source can decrease by R_1.

Let $U_s(t)$ denote the number of packets waiting in the buffer of the source at slot t, and $\overline{U}_s$ be the steady state average. Hence, $Pr(U_s(t) \leq R_1\overline{U}_s) \geq 1 - \frac{1}{R_1}$ (This inequality is true for any nonnegative variables). We further assume that an S-D transmission only happens in the same cell at a time slot when at least R_1 packets are in the buffer of the source, which is reasonable for parameter β. In each time slot where at least R_1 packets queued in the cache, with probability $(1-\rho)q$ the number of the packets decreases by R_1 packets. Let T denote the smallest integer bigger than $\overline{U}_s$. Then, the probability that at time slot t less than R_1 packets in the buffer is greater than or equal to the probability that $U_s(t-T) \leq R_1\overline{U}_s$, which implies that no packet arrives during T successive slots but there is always an S-D pair within the same cell. Therefore, the probability of the event Π that less than R_1 packets queued in the cache of the source is

$$Pr(\Pi) \geq q\big((1-\rho)q\big)^{\overline{U}_s+1} \triangleq \delta.$$

Hence we have

$$\overline{U}_s = \frac{\log(q/\delta)}{\log(1/((1-\rho)q))} - 1.$$

In a word, the discussion above can be summarized as the following theorem.

Theorem 6. *For the cell partitioned network and mobility models as described in Sec. 3, the average energy consumption $\overline{e}$ of the* Minimum energy algorithm *with input rate $\lambda = \frac{\rho R_1 q}{k\theta}$ $(0 < \rho < 1)$, a control parameter β $(1 < \beta < \frac{1}{\rho})$, and δ which is defined later, is as follows:*

$$\bar{e} = \Phi(\lambda) + \frac{(\beta-1)\rho q}{\theta}.$$

And the energy and delay tradeoff under this algorithm is optimal, which satisfies

$$\bar{e} = \Theta(\delta), \quad D = \Theta(\log\frac{1}{\delta}).$$

Applying Little's law, we know that the delay decreases logarithmically with δ. Recall that δ is the lower bound of the probability that the buffer has less than R_1 packets. Thus, $\beta\rho \leq 1-\delta$. Hence applying it to Equation (13), we have $\overline{e} \geq \Phi(\lambda) + \frac{(1-\delta)q - \rho q}{\theta}$, which grows linearly with δ. By [25], we derive that the tradeoff between energy and delay is optimal.

6.4 The summary of theorems

For the reader's comprehension, we summarize the theorems and their corresponding achievements in Table 2.

TABLE 2
Theorems and the corresponding achievements

Thm.s	Metrics	Achievements
Thm.1	Exact capacity	$\mu = \begin{cases} \frac{R_1 q + R_1 p + R_2 q' + R_2 p'}{(k+1)\theta}, \\ \frac{2R_1 q + 2R_2 q'' + R_1 p'' + R_2(p'-q')}{(k+1)\theta}. \end{cases}$
Thm.2	Exact delay	$\overline{D} \leq \frac{2dBN\log k}{(k+1)\lambda\mu\varphi(1-\rho)}$
Thm.3	Capacity and delay for two hop relay algorithm with redundancy	$\lambda = \theta(\frac{1}{km}),$ $D = \begin{cases} \Theta(\frac{N\log k}{km}) & m = o(N), \\ \Theta(\frac{\log m}{k}) & m = \Theta(N). \end{cases}$
Thm.4	Capacity and delay for two hop relay selection algorithm with Redundancy	$\lambda = \theta(\frac{1}{km}), D = \Theta(\frac{zN\log k}{km})$
Thm.5	The expression for minimum energy function	$\Phi(\lambda) = \begin{cases} \frac{k\lambda}{R_1} \\ \frac{q}{\theta} + \frac{k+1}{R_1}[\lambda - \frac{R_1 q}{k\theta}] \\ \frac{p}{\theta} + \frac{k}{R_2}[\lambda - \frac{R_1 p}{(k+1)\theta} - \frac{R_1 q}{(k+1)k\theta}] \\ \frac{p+q'}{\theta} + \frac{k+1}{R_2}[\lambda - \frac{R_2 q' + R_1(kp+q)}{(k+1)k\theta}] \end{cases}$
Thm.6	Average energy consumption & Optimal energy and delay tradeoff	$\bar{e} = \Phi(\lambda) + \frac{(\beta-1)\rho q}{\theta} = \Theta(\delta), D = \Theta(\log\frac{1}{\delta})$

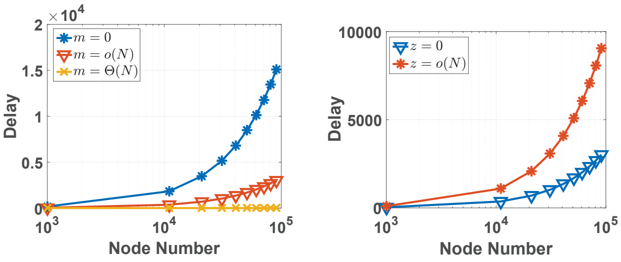
7 Simulations

In this section, we attempt to evaluate the performance of the proposed algorithms. The purpose is three-fold. (1) We compare two hop relay algorithms with and without redundancy, with and without relay selection to verify the improvement of delay performance brought by redundancy and relay selection. (2) We demonstrate the advantage of our algorithms by comparing with existing algorithms. (3) The minimum energy algorithm is carried out for validation. All the algorithms are implemented in C++ and ran on a Windows x64 PC (Intel Core i-5 @2.3Ghz, 4GB RAM).

7.1 Simulation Setup

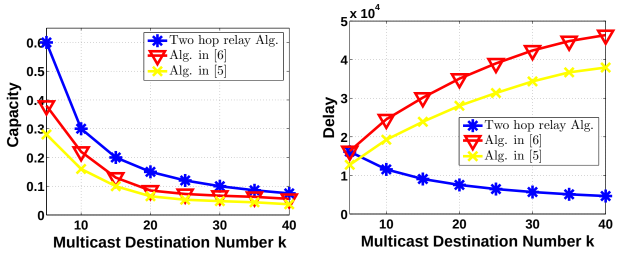
The network area is assumed to be a large square *w.l.o.g.* and further partitioned into 20×20 non-overlapping small squares (i.e., cells) of the same size. To eliminate the edge effect, we assume the two pairs of parallel edges of the area are adjacent. Then, each cell is adjacent to four cells. Nodes are initially uniformly and randomly distributed on the area. At the end of each time slot, a node stay in the current cell or move to either north, south, east or west directions with the same probability 0.2. The rate of same cell transmission R_1 is set to be 2 packets/slot and the rate of adjacent transmission $R_2 = 1$ packet/slot.

To measure the capacity of the network, according to the definition presented in Section 3, we gradually increase the arrival rate λ at each node until the queue of each node tends to keep growing over time. The maximal arrival rate is treated as the capacity. In terms of the delay, we calculate the delay of each packet by subtracting its arrival time from the time when it is received by all its destinations. Then, we average the delay of all the packet as the network delay. In each time slot, we average the total energy consumption of the network over all the users. Then, the average energy consumption $\bar{e}$ is calculated as the time average energy consumption over a long period.



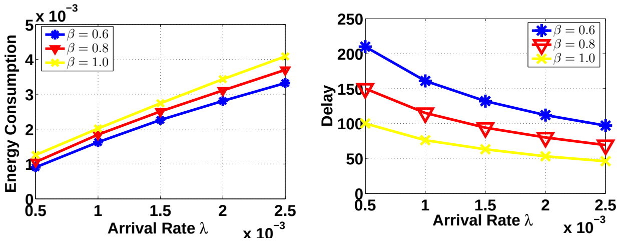
(a) With and without redundancy (b) With and without selection

Fig. 3. Network delay of proposed algorithms.



(a) Network capacity (b) Network delay

Fig. 4. Performance comparison with existing algorithms.



(a) Energy Consumption (b) Network Delay

Fig. 5. Performance of the minimum energy algorithm.

7.2 Simulation Results

We first study the impact of packet redundancy and relay selection on delay performance. We obtain the network delay under network size $N \in \{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9\} \times 10^4$. Fig. 3(a) illustrates the delay of two-hop relay algorithm with redundancy and without redundancy, i.e., $m > 0$ and $m = 0$, where m is the number of content redundancies. We can see that the packet redundancy could significantly reduce the network delay. With a larger redundancy, the network delay becomes smaller. Fig. 3(b) evaluates the delay performance for the two-hop relay selection algorithm with and without redundancy, i.e., $z = o(N)$ and $z = 0$, where z is the average number of packets waiting at the node buffer. When $z = o(N)$, packet congestion happens at the node and the delay increases.

To verify the performance of our algorithm, we compare our two hop relay algorithm with similar works. The algorithms in [5] and [6] are introduced for comparison. Particularly, the two comparative algorithms are carried in our settings where the movement of users is depicted by the Markovian mobility model. Fix network size $N = 10^3$, we compare the capacity and delay under different group sizes $k = \{5, 10, 15, 20, 25, 30, 35, 40\}$. The results are reported in Fig. 4. From Fig. 4(a), we can see that our algorithm achieves nearly 30% to 50% larger capacity than other two counterparts. In terms of the network delay, the delay of our algorithm gradually decreases with the group size, while the delay in [5] and [6] keep increasing. The simulation result well conforms with our theoretical findings.

The minimum energy algorithm is further implemented for validation. Note that only same cell transmissions happen in the algorithm. We assume each transmission costs one unit power. We consider a network of $N = 10^3$ nodes. Fix $R_1 = 2$ packets/slot and group size $k = 5$, we first measure the average energy consumption under different β with respect to arrival rate $\lambda = \{0.5, 1, 1.5, 2, 2.5\} \times 10^{-3}$. The results are presented in Fig. 5(a). We find that the energy consumption grows almost linearly with the arrival rate and monotonically increases with β. This result verifies our theoretical result that the average energy consumption $\bar{e} = \frac{k\lambda}{R_1} + \frac{(\beta-1)\rho q}{\theta}$ is a linear function of λ. We further evaluate the delay of the algorithm under the same setting by Fig. 5(b). We see that the delay decreases with the arrival rate λ almost logarithmically. Moreover, the delay decreases as β grows. Combining the two figures, we can infer that the delay decreases logarithmically with the average energy consumption. This result accords with the theoretical conclusion in Theorem 6 that there is a logarithmic factor between $\bar{e}$ and D.

8 Conclusion

In this paper, we have studied capacity, delay and energy consumption of relay-assisted multicast scheme in MANETs under a general Markovian mobility model. A *Two hop relay algorithm* is proposed and the exact capacity and delay are derived by the Lyapunov drift analysis. Moreover, we allow redundancy in the cache-enabled network to further improve the network performance. By studying the queue delay in node-side cache, we show that multicast delay can be significantly improved without impairing the capacity. Moreover, we apply the minimum energy function to depict the energy consumption for our multicast network. A *Minimum energy algorithm* is designed and proven to push the actual energy consumption arbitrarily close to the minimum energy function by sacrificing the delay. Theoretical results show that we achieve the optimal energy and delay tradeoff.

References

[1] Shangxing Wang, Youyun Xu, and Xinbing Wang. Motioncast with general markovian mobility. In *INFOCOM, 2012 Proceedings IEEE*, pages 756–764. IEEE, 2012.

[2] Matthias Grossglauser and David NC Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM transactions on networking*, 10(4):477–486, 2002.

[3] Jinbei Zhang, Yixuan Li, Zhuotao Liu, Fan Wu, Feng Yang, and Xinbing Wang. On multicast capacity and delay in cognitive radio mobile ad hoc networks. *IEEE Transactions on Wireless Communications*, 14(10):5274–5286, 2015.

[4] Rahul Urgaonkar and Michael J Neely. Network capacity region and minimum energy function for a delay-tolerant mobile ad hoc network. *IEEE/ACM Transactions on Networking (TON)*, 19(4):1137–1150, 2011.

[5] Shan Zhou and Lei Ying. On delay constrained multicast capacity of large-scale mobile ad hoc networks. *IEEE Transactions on Information Theory*, 61(10):5643–5655, 2015.

[6] Xinbing Wang, Wentao Huang, Shangxing Wang, Jinbei Zhang, and Chenhui Hu. Delay and capacity tradeoff analysis for motioncast. *IEEE/ACM transactions on networking*, 19(5):1354–1367, 2011.

[7] Michael J Neely and Eytan Modiano. Capacity and delay tradeoffs for ad hoc mobile networks. *IEEE Transactions on Information Theory*, 51(6):1917–1937, 2005.

[8] Jiajia Liu, Nei Kato, Jianfeng Ma, and Toshikazu Sakano. Throughput and delay tradeoffs for mobile ad hoc networks with reference point group mobility. *IEEE Trans. Wireless Communications*, 14(3):1266–1279, 2015.

[9] Rajat Talak, Sertac Karaman, and Eytan Modiano. Capacity and delay scaling for broadcast transmission in highly mobile wireless networks. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 8. ACM, 2017.

[10] Jiajie Ren, Guanglin Zhang, and Demin Li. Multicast capacity for vanets with directional antenna and delay constraint under random walk mobility model. *IEEE Access*, 5:3958–3970, 2017.

[11] Zhe Luo, Xiaoying Gan, Xinbing Wang, and Hanwen Luo. Optimal throughput-delay tradeoff in manets with supportive infrastructure using random linear coding. *IEEE Trans. Vehicular Technology*, 65(9):7543–7558, 2016.

[12] Xuanyu Cao, Jinbei Zhang, Luoyi Fu, Weijie Wu, and Xinbing Wang. Optimal secrecy capacity-delay tradeoff in large-scale mobile ad hoc networks. *IEEE/ACM Transactions on Networking (TON)*, 24(2):1139–1152, 2016.

[13] Wayan Wicke, Nikola Zlatanov, Vahid Jamali, and Robert Schober. Buffer-aided relaying with discrete transmission rates for the two-hop half-duplex relay network. *IEEE Transactions on Wireless Communications*, 16(2):967–981, 2017.

[14] Mohsen Mohammadkhani Razlighi and Nikola Zlatanov. Buffer-aided relaying for the two-hop full-duplex relay channel with self-interference. *IEEE Transactions on Wireless Communications*, 17(1):477–491, 2018.

[15] Ke Xiong, Pingyi Fan, Chuang Zhang, and Khaled Ben Letaief. Wireless information and energy transfer for two-hop non-regenerative mimo-ofdm relay networks. *IEEE Journal on Selected Areas in Communications*, 33(8):1595–1611, 2015.

[16] Jing Li and Young-Han Kim. Partial decode-forward relaying for the gaussian two-hop relay network. *IEEE Transactions on Information Theory*, 62(12):7078–7085, 2016.

[17] Jia Liu, Min Sheng, Yang Xu, Jiandong Li, and Xiaohong Jiang. End-to-end delay modeling in buffer-limited manets: A general theoretical framework. *IEEE Transactions on Wireless Communications*, 15(1):498–511, 2016.

[18] Jinbei Zhang, Xinbing Wang, Xiaohua Tian, Yun Wang, Xiaoyu Chu, and Yu Cheng. Optimal multicast capacity and delaytradeoffs in manets. *IEEE Transactions on Mobile Computing*, 13(5):1104–1117, 2014.

[19] Shan Zhou and Lei Ying. On delay constrained multicast capacity of large-scale mobile ad hoc networks. *IEEE Transactions on Information Theory*, 61(10):5643–5655, 2015.

[20] Riheng Jia, Feng Yang, Shuochao Yao, Xiaohua Tian, Xinbing Wang, Wenjun Zhang, and Jun Xu. Optimal capacity–delay tradeoff in manets with correlation of node mobility. *IEEE Transactions on Vehicular Technology*, 66(2):1772–1785, 2017.

[21] Bin Yang, Yulong Shen, Xiaohong Jiang, and Tarik Taleb. Generalized cooperative multicast in mobile ad hoc networks. *IEEE Transactions on Vehicular Technology*, 67(3):2631–2643, 2018.

[22] Tejpreet Singh, Jaswinder Singh, and Sandeep Sharma. Energy efficient secured routing protocol for manets. *Wireless Networks*, 23(4):1001–1009, 2017.

[23] Deyu Zhang, Zhigang Chen, Lin X Cai, Haibo Zhou, Sijing Duan, Ju Ren, Xuemin Shen, and Yaoxue Zhang. Resource allocation for green cloud radio access networks with hybrid energy supplies. *IEEE Transactions on Vehicular Technology*, 67(2):1684–1697, 2018.

[24] Huaying Wu, Xiao-Yang Liu, Luoyi Fu, and Xinbing Wang. Energy-efficient and robust mask-encoder for wireless camera networks in internet of things. *IEEE Transactions on Network Science and Engineering*, 2018.

[25] Michael J Neely. Optimal energy and delay tradeoffs for multiuser wireless downlinks. *IEEE Transactions on Information Theory*, 53(9):3095–3113, 2007.

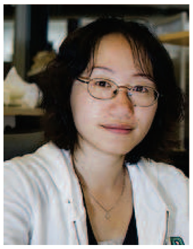
[26] Messaoud Doudou, Jose M Barcelo-Ordinas, Djamel Djenouri, Jorge Garcia-Vidal, Abdelmadjid Bouabdallah, and Nadjib Badache. Game theory framework for mac parameter optimization in energy-delay constrained sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 12(2):10, 2016.

[27] Changyang She and Chenyang Yang. Energy efficiency and delay in wireless systems: Is their relation always a tradeoff? *IEEE Transactions on Wireless Communications*, 15(11):7215–7228, 2016.

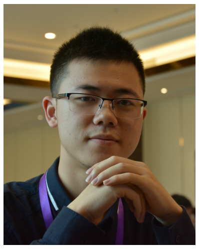
[28] Imtiaz Ahmed, Khoa Tran Phan, and Tho Le-Ngoc. Optimal stochastic power control for energy harvesting systems with delay constraints. *IEEE Journal on Selected Areas in Communications*, 34(12):3512–3527, 2016.

[29] Luoyi Fu, Xinzhe Fu, Zesen Zhang, Zhiying Xu, Xudong Wu, Xinbing Wang, and Songwu Lu. Joint optimization of multicast energy in delay-constrained mobile wireless networks. *IEEE/ACM Transactions on Networking*, 26(1):633–646, 2018.

[30] Michael James Neely. *Dynamic power allocation and routing for satellite and wireless networks with time varying channels*. PhD thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2003.



Xiaoying Gan received the Ph.D. degree in Electronic Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006. From 2009 to 2010, she was a Visiting Researcher with the California Institute for Telecommunications and Information, University of California at San Diego, San Diego, CA, USA. She is currently an Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. Her current research interests include network economics, social aware networks, heterogeneous cellular networks, multiuser multi-channel access, and dynamic resource management.

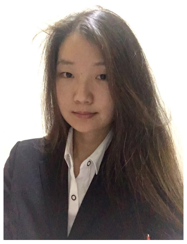


Chen Feng received his B. E. degree in Communication Engineering from Tianjin University, China, in 2016. He is currently pursuing the Ph.D. degree in Electronic Engineering at Shanghai Jiao Tong University, Shanghai, China. His current research interests are in the area of wireless social networks and information diffusion.

Zhida Qin received the B.S. degree in Huazhong University of Science & Technology, Wuhan, China, in 2014, and is currently pursuing the Ph.D. degree in Electronic Engineering at Shanghai Jiao Tong University, Shanghai, China. His current research interests include wireless network capacity, social aware networks, information diffusion.

Ge Zhang is currently pursuing his B.E. degree in Information Engineering at Shanghai Jiao Tong University, Shanghai, China. His current research interests include social aware networks and information diffusion.

Huaying Wu received her B. E. degree in Information Engineering from Xi'an Jiao Tong University, China, in 2015. She is currently pursuing the Ph.D. degree in Electronic Engineering at Shanghai Jiao Tong University. Her current research of interests are in the area of wireless networks and social networks.



Luoyi Fu received her B. E. degree in Electronic Engineering from Shanghai Jiao Tong University, China, in 2009 and Ph.D. degree in Computer Science and Engineering in the same university in 2015. She is currently an Assistant Professor in Department of Computer Science and Engineering in Shanghai Jiao Tong University. Her research of interests are in the area of social networking and big data, scaling laws analysis in wireless networks, connectivity analysis and random graphs.

Xinbing Wang received the B.S. degree (with hons.) in automation from Shanghai Jiao Tong University, Shanghai, China, in 1998, the M.S. degree in computer science and technology from Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree with a major in electrical and computer engineering and minor in mathematics from North Carolina State University, Raleigh, in 2006. Currently, he is a Professor in the Department of Electronic Engineering, and Department of Computer Science, Shanghai Jiao Tong University, Shanghai, China. Dr. Wang has been an Associate Editor for IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON MOBILE COMPUTING, and ACM Transactions on Sensor Networks. He has also been the Technical Program Committees of several conferences including ACM MobiCom 2012,2014, ACM MobiHoc 2012-2017, IEEE INFOCOM 2009-2017.

Huadong Ma received the BS degree in mathematics from Henan Normal University in 1984, the MS degree in computer science from the Shenyang Institute of Computing Technology, Chinese Academy of Science in 1990, and the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Science in 1995. He is currently a professor and the director of the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, dean of the School of Computer Science, Beijing University of Posts and Telecommunications, China. He visited UNU/IIST as a research fellow in 1998 and 1999, respectively. From 1999 to 2000, he held a visiting position in the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan. He was a visiting professor at The University of Texas at Arlington from July to September 2004, and a visiting professor at the Hong Kong University of Science and Technology from December 2006 to February 2007. His current research focuses on multimedia system and networking, Internet of things and sensor networks, and he has published more than 100 papers and four books on these fields. He is a member of the IEEE and the ACM.