# EXPLOITING DATA DIVERSITY AND MULTIUSER DIVERSITY IN NONCOOPERATIVE MOBILE INFOSTATION NETWORKS

Wing Ho Yuen Roy D. Yates Siun-Chuon Mau
WINLAB, Rutgers University
andyyuen@winlab.rutgers.edu ryates@winlab.rutgers.edu siunmau@winlab.rutgers.edu

*Abstract* — **In wireless networks, it is often assumed that nodes cooperate to relay packets for one another. Although this is a plausible model for military or mission based networks, it is unrealistic for commercial networks and future pervasive computing environments. We address the issue of noncooperation between nodes in the context of content distribution in mobile infostation networks. We assume all nodes have common interest in all files cached in the fixed infostations. In addition to downloading files from the fixed infostations, nodes act as mobile infostations and exchange files when they are in proximity. We stipulate a social contract such that an exchange takes place only when each node can obtain something it wants from the exchange. Our social contract opportunistically aligns the individual node's interest with that of the whole distribution network and hence enables much higher system efficiency compared to downloading only from fixed infostations while not requiring true cooperation among nodes. We show by analysis and simulations that network performance depends on the node density, mobility and the number of files that are being disseminated. Our results point to the existence of data diversity for mobile infostation networks. The achievable throughput increases as the number of files of interest to all users increases. We have also extended the common interest model to the case where nodes have dissimilar interests. Our simulation results show that as mobile nodes change from having identical interests to mutually exclusive interests, the network performance degrades dramatically. We propose an alternate user strategy when nodes have partially overlapping interests and show that the network throughput can be significantly improved by exploiting multiuser diversity inherent in mobile infostation networks. We conclude that data diversity and multiuser diversity exist in noncooperative mobile infostation networks and can be exploited.**

# 1 Introduction

In generic mobile ad hoc networks, nodes communicate with each other through multihop routing. However, the achievable capacity in these networks is low as demonstrated by simulation studies [1, 2]. Although rate adaptation [10] or power control [20] techniques can improve network capacity, it is unlikely that these measures will increase capacity further by several orders of magnitude. Indeed, [8] showed that the asymptotic per-node capacity of a wireless multihop network goes to zero as the number of nodes tends to infinity, even under the optimistic assumption of perfect scheduling and power control.

Recently, a new ad hoc network paradigm known as mobile infostation networks [7] has been proposed. In a mobile infostation network, any pair of nodes communicates only when they are in proximity and have a very good radio channel. Under this transmission constraint, any pair of nodes is intermittently connected as mobility shuffles the node locations. The network capacity of mobile infostation networks compares favorably to conventional multihop ad hoc networks. Using a two-hop relay strategy, Grossglauser and Tse [7] showed that the per-node throughput of a mobile infostation network is $O(1)$, independent of the number of nodes. This capacity improvement comes from the exploitation of node mobility to physically carry the packets around the network, and is independent of the underlying mobility model, as long as the mobility process is ergodic.

Nevertheless, the order of magnitude improvement in network capacity comes at a cost. End-to-end transmissions incur a random delay that is at the same time scale of the mobility process. Thus, a mobile infostation network is applicable to delay tolerant applications with a heavy bandwidth requirement, say, in a content distribution application where all nodes are subscribers to a movie or news content provider. In this type of applications, a user is not concerned and aware of the movie download schedules. The application typically runs in the background for a few hours or even a few days as a user commutes to different places in his daily routine. This is consistent with the plethora of software applications in ubiquitous computing environments [27], where computing systems become invisible and fade into the background and work for the users. In this case, we can draw a parallel of *ubiquitous networking environments* since users are not aware of the background networking in the mobile infostation communication paradigm.

Motivated by the dramatic capacity improvement of mobile infostation networks, there is substantial literature that address exploitation of node mobility to improve data dissemination. While [3, 7, 16, 17, 25] provide theoretic analyses of capacity and delay, many other papers focus on performance evaluation of protocols and applications. The potential spectrum of applications ranges from biological information acquisition systems used in habitat monitoring of endangered wildlife species [12, 22, 23] on one hand, to mundane movie and news downloading in a content distribution network [30] and location specific information services [19] on the other hand. Node mobility can also be exploited in conventional ad hoc networks that use multihop forwarding [9, 15, 26, 32]. Instead of dropping packets when a network is partitioned, packets are buffered and handed over to another node when a node is reconnected to the network. This promotes robust network performance when network connectivity is intermittent. Finally, similar networking problems are also being considered under the banner of delay tolerant networking [4, 9].

Most of the above work relies on some sort of node cooperation in the underlying network model. For some applications such as habitat monitoring of wildlife species, sensor nodes are deployed from a single organization and the cooperation assumption is valid. On the other hand, in commercial applications each node in the network is autonomous and may act selfishly. A node usually has disincentives to relay other people's packets since it is expending its own bandwidth and energy resources in a transmission. The cooperation assumption is thus unrealistic.

In this paper we address the issue of noncooperation in the context of a mobile infostation network for movie downloading. All nodes are subscribers to a movie content distribution network. A movie is divided into $K$ files which are then cached in a network of fixed infostations, access points providing pockets of high-speed short-range coverage [5]. When a node comes close to an infostation, files can be downloaded. In an entirely noncooperative network, this would be the only mechanism for file dissemination. It only uses the high-speed channel between an infostation and a node near it, while wasting all the equally excellent channels between closely located nodes. A more efficient system would have any two nodes in proximity to act as mobile infostations to exchange copies of their files. With sufficient node density, a node obtains most of the files from node-to-node file exchanges. Data dissemination is thus distributed to all nodes and all locations in the network.

It is possible to allow file exchanges among mobile nodes while keeping the network essentially noncooperative by stipulating the following *social contract* for all nodes in the network. When two nodes meet, they inspect the file contents of each other. If each node identifies a file that it wants, a bilateral file exchange takes place. Conversely, if either node cannot find a file it wants, no file exchange takes place since that node has no immediate incentive to transmit a file to the other. This social contract opportunistically aligns the interests of individual nodes with the collective interest of the content distribution network.

We have shown by analysis and simulations that the networking performance of this file exchange mechanism depends on node mobility and density. More importantly, we find that both fairness and throughput of the network improve as the number of files in the network increases. We identify this phenomenon as a new form of diversity. Traditional communication diversity techniques exploit the variations of signal strength over temporal, spatial and frequency domains. *Data diversity*, on the other hand, arises when the number of files interested by an individual increases. It is a consequence of noncooperation among nodes.

We have also extended the common interest model to the case where each node has dissimilar interest. This is applicable to the contexts in which multiple movies or TV shows are cached in the infostations. When nodes have mutually exclusive or partially overlapping interests, network performance degrades drastically. We have identified two user strategies for the dissimilar interest model. Our simulation results show that network throughput can be significantly improved by exploiting multiuser diversity inherent in mobile infostation networks.

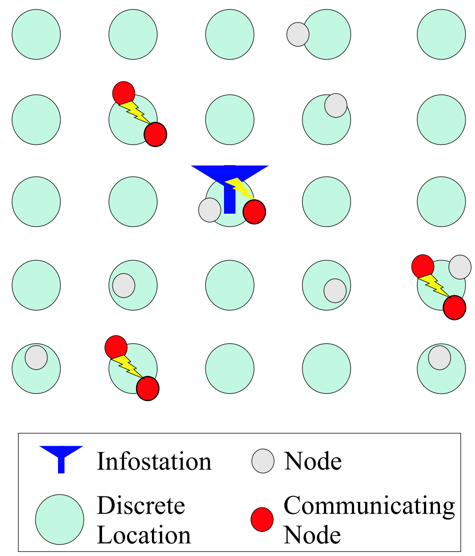



Figure 1: Illustration of the network model.

The rest of the paper is organized as follows. In section 2, we describe the system model. Section 3 is devoted to performance analysis, and the results are verified by simulations in section 4. We describe a new form of diversity - data diversity in section 5. In section 6, we extend our common interest model to the case where nodes have partially overlapping interests. Simulation results of two user strategies are discussed. The results are interpreted further as a form of multiuser diversity in section 7. Finally, conclusions are drawn in section 9.

## 2 System Model

This work is largely motivated by [7] which employed a signal to interference ratio (SIR) based link quality model to demonstrate the that $N$ nodes in a region could maintain $O(N)$ simultaneous transmissions with acceptable SIR. However, in this work, we look to employ a simpler communication model in order to demonstrate the effect of the social contract on content distribution. As shown in Figure 1, the geography consists of $L$ discrete locations in a square grid with an infostation at the center of the grid. The infostation cache holds the $K$ files of a movie. We assume the geography wraps around at each boundary, effectively creating a toroidal grid. We refer to this $L$ node wraparound grid with one infostation and $L-1$ *regular locations* as a *block*. A block is intended to mimic a typical multi-infostation network in which an infinite grid of infostations populate an infinite plane. The number of locations $L$ relative to the single infostation serves to characterize the density of fixed infostations over the terrain.

The $L$ location grid is populated with $N$ nodes with independent mobility processes. In our simulation experiments, we assume that time is discretized such that at each unit of time, each node randomly and independently moves in one of the four directions with equal probability $q = 0.25$. When two or more nodes are at the same location at the same time, we say those nodes are *neighbors*.

In our communication model, each node either downloads files from an infostation or exchanges files with a neighbor. At the infostation, only file downloading is allowed. At any other locations, file exchanges between mobile nodes are permitted. Given a particular radio bandwidth, the size of a file is chosen such that the time a node occupies a location allows for either a bilateral file exchange between neighbors at a regular location or for two files to be downloaded from the infostation.

There are two factors that impact data dissemination. First there is a *transmission concurrency constraint* at each location. If there is more than one node at the infostation, contention is resolved by randomly picking one node for downloading. Similarly, when there are more than two neighbors at a location, two of the neighbors are randomly picked to perform a file exchange. The random picking of nodes for a wireless transmission is consistent to the node non-cooperation assumption. When nodes are non-cooperative, there is no co-ordination between transmissions of different nodes. Since each node wants to minimize its own downloading time, it attempts to seize the channel at every time slot. In actual implementation, each node may just simply broadcast a tone after a small random time at the beginning of a time slot to reserve it. The first two nodes that transmit a tone at a time slot seize the channel and are given the opportunity to negotiate for a file exchange. and may proceed to a file exchange eventually.

The second factor that affects data dissemination is captured by the probability of file exchange. This is in turn dictated by the *user strategy* which consists of two parts. The user strategy must determine first whether to exchange files according to a *social contract*. A social contract is observed by all nodes and governs whether a file exchange takes place or not. Specifically, a node may want to exchange for a file because it is genuinely interested in that file. In this case, a file exchange is warranted when both nodes find something that they are genuinely interested from each other. Alternatively, a node may want to exchange for a popular file, which is then used to facilitate future file exchanges. Thus even if a node cannot obtain a file of genuine interest, it may exchange for a file that it does not have. The above are two instantiations of social contract and will be discussed in this paper. In the first part of this paper, however, there is no distinction between the two social contracts. When all nodes have common interest in downloading the files of a popular movie, each node is genuinely interested in every file it does not have. In section 6, we extend the common interest model to the case where nodes have dissimilar interests that are partially overlapping. In that case, network performance is dependent on the choice of the social contract.

After two nodes have reached an agreement for a file exchange, a node must decide which file to download according to the user strategy. Two strategies are examined in this paper. For the random strategy, a node randomly selects a file it does not have from the neighbor node. Similarly, at the infostation, a node randomly selects to download two files that he does not have. For comparison, we also consider a greedy strategy which assumes that each node has full knowledge of the circulation of each file within the network. For an infostation download or a neighbor exchange, a node picks the file that is the least circulated among all files it does not have. This strategy is greedy since it maximizes the probability of exchange $P_E$ between two arbitrary nodes in a static snapshot.

We note that the selection of two arbitrary nodes for file exchange is suboptimal. Two nodes that seize the channel successfully may not perform a file exchange due to the social contract. This efficiency can be avoided by scheduling transmissions only to the node pair with an exchange agreement. However, scheduling solicits implicit co-operation between nodes. Nodes that are eager to transmit may be asked to refrain from transmitting. The decoupling of channel contention and user strategy for file exchange is thus consistent to the non-cooperative assumption. Each node would make every attempt to seize the channel at every time slot regardless of the likeliness of a file exchange. Incidentally, the modeling of node contention as random node selection greatly simplifies the performance analysis and provides a lower performance bound to an ideal scheduling scheme when nodes are cooperative.

On the other hand, the social contract implicitly assumes there are no misbehaving nodes. Each node makes no false claim on the files it possesses and ensures the integrity of all its disseminated files. The social contract provides a framework for studying non-cooperation between nodes. In a practical file exchange protocol, additional security mechanisms may be added to ensure the integrity of the files being exchanged. Apart from files authentication, reputation management algorithms [13] can also be used to dissuade nodes from misbehavior.

The proposed content distribution network admits a number of performance metrics to describe how quickly files are disseminated. We define $T_1$ as the time when 80% of the nodes get all of the files. A network operator is interested in this quantity, which is related to the networking efficiency and the revenue generated from the network. We define $T_2$ as the time when all nodes get 80% of the files. A network subscriber, on the other hand, will be interested in $T_2$, which is related to fairness and perhaps will influence his willingness to pay. We also define $T_3$ as the time for all nodes to get all the files. Finally $T_4$ is defined as the time for an arbitrary node to obtain all files. An analytical expression for $E[T_4]$ is obtained in the next section.

We also evaluate the network performance in terms of *throughput* $C_i$, which characterizes the average rate of file downloading per node. This is defined in terms of the networking time $T_i$ and is given by $C_i \triangleq K/E[T_i]$, for $i = 1, 2, 3, 4$. The units of $C_i$ are files per node per unit time. Note that we can view the distribution to a particular node of movies over time as a renewal process in which the renewal period equals $T_4$, the time required for the node to obtain one movie. Since the node obtains a reward of $K$ files in each renewal period, renewal-reward theory assures that the expected rate at which the node obtains files is precisely $C_4$ [21].

# 3 Performance Analysis

When two or more mobile nodes are at the same location, a two-step process determines whether a file exchange takes place. First, the nodes at that location follow a radio access protocol to determine which pair of nodes will attempt a file exchange. We use the term *access* to refer to the event that a node gets to be one of a pair of nodes that examines the files carried by the other. Under some simplifying assumptions, we will see that at a regular location the *access probability* is given by a constant $\beta$, that depends on the number of nodes $N$ and locations $L$ in the block. For a pair of nodes chosen in the access phase, the *exchange probability* $P_E$ denotes the probability that the two nodes can exchange files under the terms of the social contract. The exchange probability will depend on the file contents in each node, which in turn depends on the user strategy.

In this section we provide a simple approximate analysis of $\beta$ and $P_E$. We then develop a simple Markov chain model to obtain the expected networking time $E[T_4]$ and the corresponding throughput $C_4$ for each node. For the analysis, we make the following simplifying assumptions:

- **Memoryless Uniform Mobility** In each time unit, each node is randomly and independently at any of the $L$ locations with probability $p = 1/L$.
- **Independent Uniform Content Distribution** Given that node $i$ has obtained $l_i$ files, all combinations of $l_i$ out of $K$ files are equiprobable, independent of the files held by all other nodes.

It is not hard to see that these assumptions are inconsistent with the system model of section 2. In particular, when the number of locations is small and mobility is limited, nodes are likely to be neighbors frequently and have highly correlated content. Nevertheless, our simulation results agree closely with the analytical results, indicating that these assumptions work well in systems with moderately large number ($K = 500$) of files and reasonable mobility $q = 0.25$.

Due to the transmission concurrency constraint, the maximum number of simultaneous transmissions in the block equals $L$, the number of locations. For a given number of locations, it should be apparent that there is an optimum number of nodes $N$ such that the access probability is maximized. If the number of nodes in the network is small, the spatial transmission concurrency is not fully utilized. Similarly, if there are too many nodes in the block, only a fraction of nodes could schedule transmissions in the $L$ possible locations.

Given a particular node at a given location, memoryless mobility implies that the number of other neighbors at that location is a random variable $J$ with the binomial distribution

$$P[J = j] = \binom{N-1}{j} p^j (1-p)^{N-1-j} \quad j = 0, \ldots, N-1 \tag{1}$$

When a given mobile is at the infostation with $J = j$ neighbors, the probability $\beta'$ that the given node is chosen for the infostation download is $1/(j+1)$. Averaged over all $J$, the probability the given node is chosen for the download is

$$\beta' = \sum_{j=0}^{N-1} \frac{1}{j+1} P[J=j] = \frac{1-(1-p)^N}{Np} \tag{2}$$

Similarly, when a node is at a regular location with $J = j \geq 1$ other neighbors present, 2 out of $j+1$ nodes are randomly chosen. The conditional access probability that a given node is one of the two chosen nodes is $2/(j+1)$. Thus,

$$\beta = \sum_{j=1}^{N-1} \frac{2}{j+1} P[J=j] \tag{3}$$

$$= \frac{2[1-(1-p)^N - Np(1-p)^{N-1}]}{Np} \tag{4}$$

Based on (4), the optimal $N$ is around $2L$. Below, in equation (12), a more careful optimization of $\beta(N)$ in the limit of large $N$ and $L$ with fixed density $\rho \triangleq N/L$, reveals that $\rho_{\text{opt}} \simeq 1.8$. One can use this result to determine the optimal spatial density of fixed infostations based on the anticipated spatial density of mobile subscribers.

When nodes $i$ and $j$ have the opportunity to exchange files, the probability of exchange $P_E$ depends on the files each node is holding. Suppose nodes $i$ and $j$ have $l_i$ and $l_j$ files in their caches. An exchange between the nodes will occur *unless* one node has a collection of files that is subset of the other's collection. Assuming, without loss of generality, that $l_i \leq l_j$, an exchange failure occurs if node $i$ chooses its subset of $l_i$ files out of the $l_j$ files of node $j$. Since there are $\binom{K}{l_i}$ total ways for node $i$ to choose its files, the probability of exchange is

$$P_E(l_i, l_j) = 1 - \frac{\binom{l_j}{l_i}}{\binom{K}{l_i}} \qquad 0 \leq l_i \leq l_j \leq K \tag{5}$$

From (5), we can derive a tight upper bound for the probability $P_{E^c} \triangleq 1 - P_E$ of no file exchange between neighbor nodes with $l_i$ and $l_j$ files such that $aK \leq l_i \leq l_j \leq (1-a)K$ and $0 < a < 1/2$. When $K$ is large such that $aK$, $(1-a)K$, and $(1-2a)K$ are all much greater than 1, an asymptotic upper bound $\tilde{P}_{E^c}$ for $P_{E^c}$ coincides with Stirling's approximation for $P_{E^c}$ and is given by

$$\ln \tilde{P}_{E^c} = \Big[ \ln \frac{(1-a)^2}{1-2a} + 2a \ln \frac{1-2a}{1-a} \Big] K \tag{6}$$

As the multiplier of $K$ is negative for $0 < a < 1/2$, we deduce that when $0 < a < 1/2$,

$$\lim_{K \to \infty} P_E(l_i, l_j) = 1, \qquad aK \leq l_i \leq l_j \leq (1-a)\, K \tag{7}$$

That is, if each node has a non-vanishing fraction of all $K$ files, a file exchange almost certainly will occur when the number of files in the system is large.

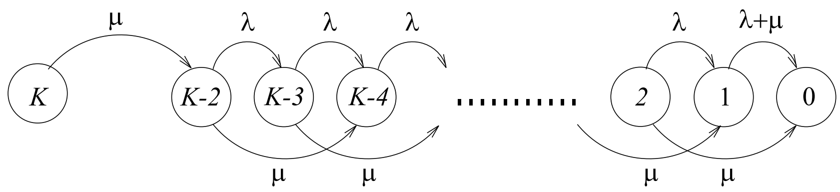



Figure 2: Illustration of the Markov chain model. The shown values denote the state transition rates. Note that the depiction of self transitions is omitted.

To find an upper bound for $P_{E^c}$ that is valid for most values of $l_i$ and $l_j$, we observe that the small $x$ approximation $\ln(1+x) \simeq x$ implies

$$\ln \tilde{P}_{E^c} \simeq -2a^2 K, \tag{8}$$

implying that $P_{E^c}$ can be made arbitrarily close to zero by choosing $a > O\Big(1/\sqrt{K}\Big)$. When the number of files in the system is large, file exchange almost always happens among neighbors during most of the file dissemination process. In practice, we can regard $P_E = 1$ when $K \geq 1000$. We will come back to this point when we discuss our simulation results in Figure 3.

In the following, we derive the expected networking time $E[T_4]$ for a node to obtain all files and the associated throughput $C_4$. We assume that $K$ is large such that (7) holds and we model the dynamics of movie downloading by the discrete time Markov chain illustrated in Figure 2. Denote the state as the number of files remaining to be downloaded to a node. Initially a node is at state $K$. Since the first two files must be obtained from an infostation, the next state is $K-2$. Subsequently, in states $k \in \{1, \ldots, K-2\}$, each unit of time allows the following possibilities:

- With probability $p$, the node encounters the infostation and then with probability $\beta'$ downloads two files. The state goes from $k$ to $k-2$ with probability $\mu = p\beta'$.
- With probability $1-p$, the node is at a regular location and then with probability $\beta$ participates in a file exchange. The state goes from $k$ to $k-1$ with probability $\lambda = (1-p)\beta$.
- With probability $1-\lambda-\mu$, no new files are obtained and the state stays the same.

Denote the expected first passage time from state $i$ to state 0 as $g_i$, where $(2 \leq i \leq K-2)$. Conditioning on the next state transition and rearranging yields the difference equation,

$$g_i = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} g_{i-1} + \frac{\mu}{\lambda+\mu} g_{i-2} \tag{9}$$

where the boundary conditions are given by $g_0 = 0$ and $g_1 = 1/(\lambda+\mu)$. Using z-transforms, we solve (9) to obtain

$$g_i = \frac{i(\lambda+2\mu) + \left(1 - \left(\frac{-\mu}{\lambda+\mu}\right)^i\right)\mu}{(\lambda+2\mu)^2} \tag{10}$$

It follows that $E[T_4] = 1/\mu + g_{K-2}$, where $1/\mu$ is the expected time until a node first encounters the infostation and obtains the first two files.

For a network with a single infostation supporting $N$ nodes over $L$ locations, we consider the large-system and many-files regime in which $N, L, K \gg 1$ while the spatial density of nodes $\rho \triangleq N/L$ is held constant. In this regime, $J$, the number of other nodes seen in a location, becomes a Poisson random variable with $E[J] = \rho$. From (2) and (4), the infostation download probability and the conditional access probability converge to

$$\beta'(\rho) = \frac{1 - e^{-\rho}}{\rho} \tag{11}$$

$$\beta(\rho) = \frac{2}{\rho}\Big(1 - (\rho + 1)e^{-\rho}\Big) \tag{12}$$

Coincidentally, [17] also provides a similar analysis on a grid network model, which agrees to (12) we obtained. Furthermore, $\lambda = \beta(\rho)$ and $\mu = \beta'(\rho)/L$ and the asymptote of the expected time for an arbitrary node to collect all $K$ files is

$$E[T_4] \approx \frac{K}{\beta(\rho)} + \frac{L}{\beta'(\rho)} \tag{13}$$

Here, the second term is equal to $1/\mu$ to account for the time for a node to fetch the first two files in an infostation encounter. The first term is an approximation to $g_{K-2}$ by assuming all remaining files are obtained from node to node file exchanges when infostation density is low, i.e. $L \gg 1$. If we further allow $K$ to grow large relative to both $N$ and $L$, the corresponding throughput $C_4$ of a node is

$$C_4 = \frac{K}{E[T_4]} \sim \beta(\rho), \qquad\qquad \frac{K}{N}, \frac{K}{L} \to \infty \tag{14}$$

We observe that the node density $\rho$ that maximizes $\beta$ also minimizes the expected networking time $E[T_4]$ and maximizes the throughput $C_4$.

To appreciate the extent to which social contract improves the rate of file dissemination of a completely noncooperative network, in which the only mechanism for file distribution is direct downloading from fixed infostations, we consider the Markov chain model for the latter. The corresponding difference equation for the first passage time from state $i$ to $0$ is $g_i = 1/\mu + g_{i-2}$ for $i \leq K - 2$, yielding $E[T_4]^{info} = g_K = KL/2\beta'$ and

$$C_4^{info} = \frac{2\beta'(\rho)}{L} \tag{15}$$

Hence, the social contract provides an $O(L)$, or equivalently $O(N)$ since $L$ and $N$ are of the same order, improvement to the individual file collection rate. The key ingredient in this improvement is the increase from $O(1)$ file deliveries per unit time made by an infostation to $O(N)$ peer-to-peer file exchanges per unit time. With more complex models for radio communication and user mobility, in particular those employed in [7], the ability to support $O(N)$ communication links in a population of $N$ mobile nodes will yield similar order-of-magnitude provements.

The social contract also leads to a similar improvement in the dissemination rate considered in our simulations, defined as the rate at which files are collected by nodes through either downloading from fixed infostations or file exchanges. Since the individual file collection rate $C_4$ is $\beta$, the file dissemination rate under the social contract is $N\beta$ during most of the dissemination process. On the other hand, the file downloading rate at an infostation is 2 if a node is present there, thus file dissemination rate without social contract is slightly less than 2. Therefore, the improvement offered by the social contract is of the order $N$.

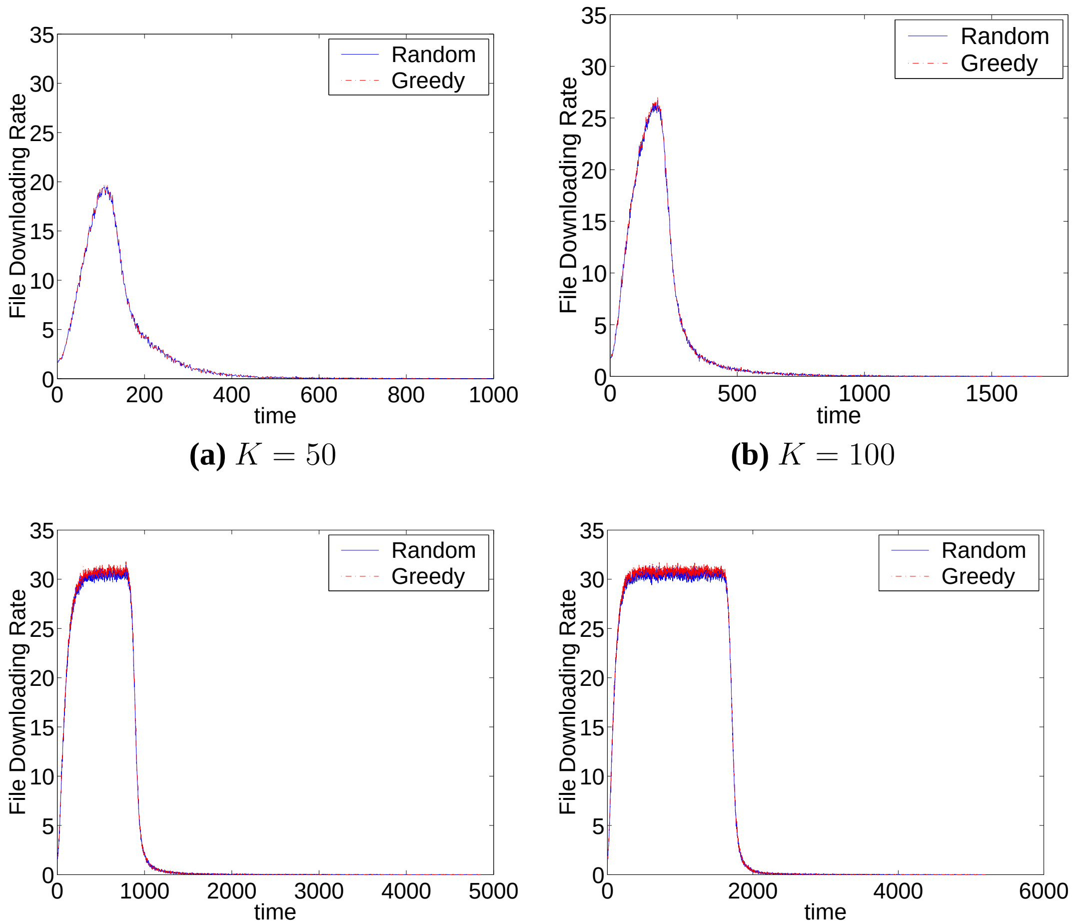
# 4 Simulation Results

In this section, we examine the impact of the number of nodes $N$ and number of files $K$ in the system on the network performance, evaluated in terms of the expected networking time $E[T_i]$ and throughput $C_i$. In our simulations, the network size is kept constant at $L = 25$ nodes. A node moves to one of the neighbor locations w.p. $q = 0.2$ at each unit time. The performance metrics are obtained from ensemble averaging over 100 simulations.

For performance evaluation, we define the *dissemination rate* as the total number of files obtained, either by download from the infostation or by file exchange, per unit time over all mobile nodes. Figure 3 shows the dissemination rate averaged over 100 simulations runs. The number of nodes is held constant at $N = 50$ and the number of files is varied ($K = 50, 100, 500, 1000$). In all cases, the differences between the random and the greedy strategies were found to be very small. Thus, the random strategy is a good alternative to the greedy strategy for practical implementation.

From Figure 3, the $y$-intercept is slightly less than 2. Since the node density is high, it is probable to find at least a node at an infostation location and download 2 files at $t = 0$. The file dissemination process has three distinct phases. In the first phase, the infostation seeds the mobile nodes with files and the dissemination rate increases rapidly as nodes obtain the ability to exchange files. Once most nodes have visited the infostation, $P_E \simeq 1$ and the dissemination rate remains steady at a peak rate that is a function of the access probability $\beta(\rho)$. In particular, each node will exchange one file with probability $P_E\beta(\rho) \simeq \beta(\rho)$. Over all $N$ nodes, the dissemination rate is $N\beta(\rho)$. Once a node has acquired all $K$ files, the social contract dictates that the node refrain from file exchanges. As the number of nodes with all $K$ files becomes significant, we enter the third phase in which the dissemination rate declines to zero as time evolves. The remaining nodes must download their files directly from an infostation, prolonging the time to download the entire movie. For all values of $K$, our simulations exhibit a significant tail associated with this final phase of dissemination.

As mentioned in the last section, in the absence of node-to-node file exchanges, the rate of file downloading shown in Figure 3 would have been constantly the y-intercept value of about 2, as opposed to $N\beta(\rho)$ most of the time. The simulation results are consistent with the analysis in the last section. As $P_E \simeq 1$ for large $K$, in each unit of time, each node will obtain one file with probability $\beta(\rho)$. With $N$ nodes in total, the average dissemination rate




**(a)** $K = 50$ **(b)** $K = 100$

**(c)** $K = 500$ **(d)** $K = 1000$

Figure 3: Average number of files obtained at each unit time over 100 simulations. (a) K=50, (b) K=100, (c) K=500, (d) K=1000.

in the middle phase is $N\beta(\rho)$. In Figure 3, $N = 50$, $L = 25$, yields $\rho = N/L = 2$ and the middle phase dissemination rate is very close to $N\beta(2) \simeq 30$ files per unit time. The ratio of this rate to that of the completely noncooperative network is about 15—a dramatic improvement. Incidentally, we can interpret Figure 3 as a scaled version of $P_E$ as a function of $t$. When $t \rightarrow 0$, most nodes have nothing in their caches, thus $P_E(t) \simeq 0$. Similarly, $P_E(t) \simeq 0$ when $t$ is large since most of the nodes have finished downloading everything.

Lastly, for a finite population of nodes, we can mark the boundaries of the middle phase by the times about which all nodes have $m$ files, $\sqrt{K} \leq m \leq K - \sqrt{K}$ , based on the discussion of the upper bound of $P_{E^c}$ after (8). We hence observe that the first and third phases require $O(L\sqrt{K})$ time, roughly on the order of the time required for each node to acquire $\sqrt{K}$ files solely by visiting the infostation. On the other hand, in the middle phase, the system must deliver $O(NK)$ files in total at a dissemination rate of $N\beta(\rho)$ files per unit time, and this requires $O(K)$ time. As $K$ increases (with $N, L$ fixed although not small), this middle phase comes to dominate the total dissemination time. Hence, for large $K$,

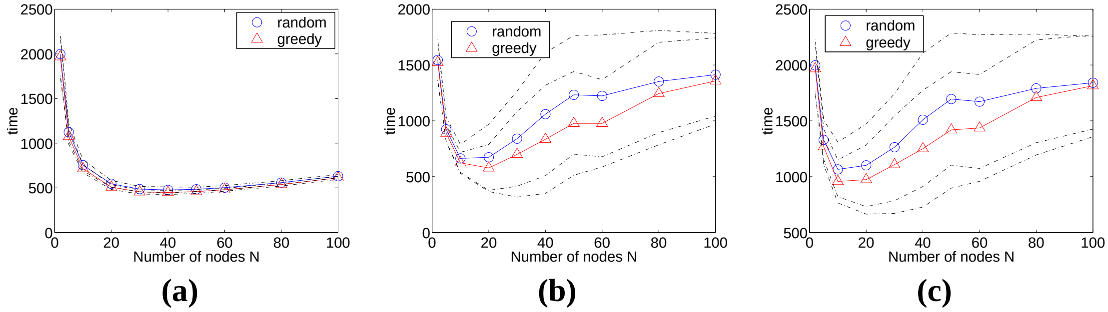



Figure 4: Average networking time vs. the number of nodes $N$. (a) $E[T_1]$ when 80% of all nodes obtain all files, (b) $E[T_2]$ when all nodes obtain 80% of all files, (c) $E[T_3]$ when all nodes obtain all files.

the average dissemination rate is effectively the same as the peak dissemination rate of the middle phase. In short, as $K \to \infty$, the curve of Figure 3 converges to a rectangle with a constant file dissemination rate of $N\beta(\rho)$ files per unit time for a duration of $K/\beta(\rho)$ time units. This conclusion is consistent with the observation that the peak dissemination rate $N\beta(\rho)$ is simply $N$ times the average per node throughput $C_4$. We note that as $K \to \infty$, the transmission of each channel is only limited by contention, indicating the noncooperation strategy achieves almost optimum resource utilization.

In Figure 4, the networking times $T_i$, $i = 1, 2, 3$, are plotted against the number of nodes $N$. The number of files is kept constant at $K = 200$. From (2), it is easily verified that $\beta(\rho)$ is maximized at $\beta = 1.7933$ users/location, or $N_{\text{opt}} = 45$ users over $L = 25$ locations. This agrees with our observation in Figure 4(a), confirming that $N \simeq 45$ also minimizes $E[T_1]$. When $N$ increases past $N_{\text{opt}}$, $E[T_1]$ increases due to the increased contention at each location; however, the increase is partially offset by the increased opportunity for exchanges; hence, $E[T_1]$ is fairly insensitive to $N$ when $N \geq N_{\text{opt}}$. When $N < N_{opt}$, $E[T_1]$ increases quickly for decreasing $N$. When $N$ is small and node density is low, the system performance is hampered by the limited availability of file exchanges. In this case, $E[T_1]$ is very sensitive to $N$ since a small increase in $N$ significantly increases the rate of file exchange.

In Figure 4(b), and 4(c), the optimum number of nodes that minimizes the networking time $T_2$ and $T_3$ are respectively $N_{\text{opt}} = 20$ and $N_{\text{opt}} = 10$ nodes, rather than $N = 45$ nodes. This disparity arises from the observation in Figure 3(a),(b) that when $K$ is not large, the total download time depends strongly on the duration of phase three which has a long tail. The tail length depends largely on the rate at which mobile nodes can download from the infostation. The tail decreases as $N$ decreases because fewer nodes results in each node having better access to the infostation. On the other hand, $T_1$ is unaffected by the long tail. A plausible reason is that networking is unfair to the last few nodes who have yet to complete their downloading; 80% of the nodes finish downloading all files well before hitting the long tail regime.

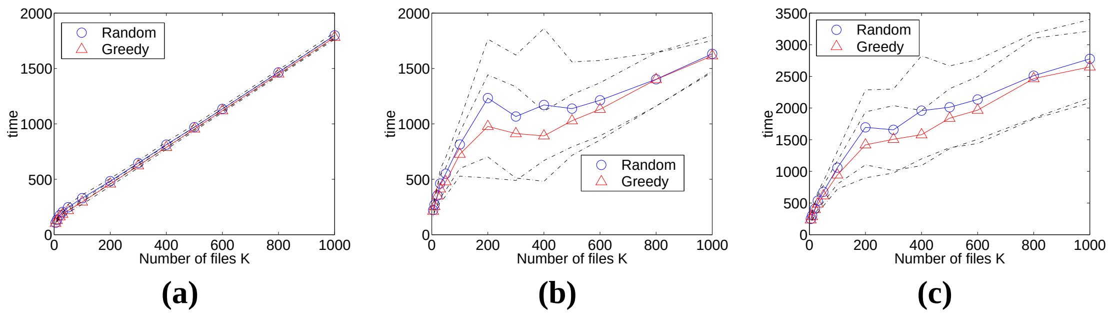


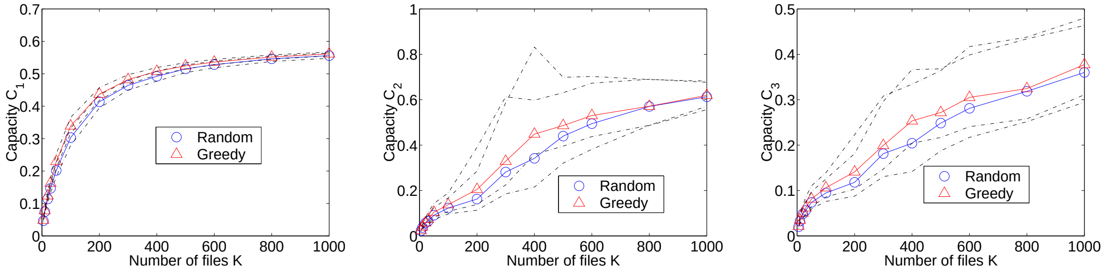

Figure 5: Average networking time vs. the number of cached files $K$. (a) $E[T_1]$ when 80% of all nodes obtain all files, (b) $E[T_2]$ when all nodes obtain 80% of all files, (c) $E[T_3]$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.

With reference to Figure 5, the networking times $t_i$ are plotted against the number of files $K$ cached in an infostation. It is obvious that the networking time $T_1$ can be fitted by a straight line for large $K$. The variance for $E[T_1]$ is also small, indicating that the networking effect due to node mobility is largely deterministic. The slope of the asymptote is found to be around 1.63, which is equal to $1/\beta(N)$. On the other hand, $T_2$ and $T_3$ exhibit larger variances. The slope of the asymptotes for $E\left[T_2\right]$ and $E\left[T_3\right]$ are 1.1 and 1.6. When $K \leq 500$, we observe that $E[T_2]$ is larger than $E[T_1]$. Beyond $K = 500$, $E[T_2]$ is smaller than $E[T_1]$. This demonstrates that as $K$ increases, the networking between the nodes is more fair. That is, all nodes have approximately the same file downloading time. A plausible reason is that $P_E \rightarrow 1$ as $K$ increases. The downloading rate is no longer influenced by individual file content, but depends primarily on mobility and contention. For large $K \geq 500$, the downloading time is long compared with the time scale of mobility ergodicity. Each node therefore has a downloading time that is almost the same, such that $E[T_1] > E[T_2]$.

# 5 Data Diversity

In Figure 5, we showed that the networking time $E[T_i],\ i = 1, 2, 3$ can be fitted nicely to an asymptote as $K$ increases. The corresponding throughputs are plotted in Figure 6 versus $K$. We observe that the throughput is an increasing function of $K$. It is instructive to find the asymptotic value of throughput $C_i^{\infty}$ as $K \rightarrow \infty$. To do this, we use the intuition captured in (13) and approximate the asymptote of $T_i$ by

$$T_i^{\infty} = m_i K + c_i \tag{16}$$




**(a)** **(b)** **(c)**

Figure 6: Throughput vs. the number of cached files $K$. (a) $C_1$ when 80% of all nodes obtain all files, (b) $C_2$ when all nodes obtain 80% of all files, (c) $C_3$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.

where $m_i$ is the slope and $c_i$ is the vertical intercept. Since the asymptote $T_i^\infty$ approaches $E[T_i]$ arbitrarily close when $K \to \infty$, we compute the asymptotic throughput as

$$C_i^\infty = \lim_{k\to\infty} \frac{K}{T_i} = \lim_{k\to\infty} \frac{K}{T_i^\infty} = \frac{1}{m_i} \tag{17}$$

Recall that $m_3 = 1.63$ as read from Figure 5(c). Thus $C_3 = 0.613$ files per node per unit time, or $30.65$ files per unit time in our network where $N = 50$. This agrees with our result in Figure 3(d). When $P_E \simeq 1$, the rate for data dissemination is around $30$ files per unit time. Incidentally, we observe that

$$\lim_{K\to\infty} C_3 = \lim_{K\to\infty} C_4 \tag{18}$$

When $K \to \infty$, networking is fair and each node has the same asymptotic throughput. Thus, our simulation results are consistent with our simplified analysis.

The apparent increase in throughput can be understood using the concept of *data diversity*. In wireless communications, diversity refers to the exploitation of variations in signal strength due to multipath fading. Since multipath fading exhibits signal variations over spatial, time and frequency domains, diversity techniques can be applied to select the strongest signal component over the respective domains. Diversity can also be exploited in a more general sense. In multiuser diversity, for instance, a receiver exploits the variability of received signal strength over different mobile nodes, and selects the node with the best channel for transmission.

Whereas the above techniques belong to the category of communication diversity, we argue that a new form of diversity, coined *data diversity*, is exhibited in noncooperative content distribution. When nodes are not cooperating, each node effectively has a preference list of files that evolves with time. If the number of disseminated files is large, there are more selections from a node's perspective. Here, a node opportunistically chooses another node with large selection to exchange files with. Equations (5), (7) and (8) dictate that file dissemination under the social contract is more efficient when there are more selections available for each node. We have shown that data diversity is relevant to noncooperative data dissemination, which is gaining more attention in the networking community. Data diversity may also have implications to other peer to peer networks other than mobile infostation networks such as content distribution on the wired Internet.

Consider the possibility that several content providers use the mobile infostation infrastructure to disseminate their content (that are not highly overlapping) to a common group of subscribers. If a subscriber has files from content provider A and it encounters another subscriber with files from content provider B, these files generally would not be inter-exchangeable since they originated from different content providers. However, our results point out that content distribution for each provider would be more efficient, in terms of both throughput and fairness, if there were mutual agreements between content providers such that all files are inter-exchangeable, effectively increasing the content size $K$.

On the other hand, even if content providers do not collaborate in data dissemination, data diversity can still be useful, say, in the dissemination of a single movie of a movie distribution network. Consider the scenario when a DVD quality movie is disseminated in a highway infostation network populated with fast vehicular subscribers. A typical drive-through infostation has a coverage radius of 20m [6]. A vehicle at a speed $20$ m/s ($45$ mi/hr) therefore has a connection time of 2 seconds when it is in the coverage area of an infostation. Similarly, for two vehicles moving in opposite direction, the connection time is only 1 second. Suppose the infostation radios operate at a modest data rate of 160Mbit/s (which still substantially outperform the state of the art 54Mbit/s 802.11a access points available today). In order to facilitate the file exchange of two data files in the worst case of a head-on mobile to mobile encounter, the file size should be no more than 10MByte. On the other hand, the typical size of a DVD quality movie is roughly 5GByte. Thus, a movie should be split into $K = 500$ files and cached in fixed infostations for dissemination. Our simulation results in Figure 6(c) have shown that with a modest content size of $K = 500$ files, the achievable per node throughput $C_3$ is 80% of the theoretical per node throughput $\lim_{K\to\infty} C_4$ for asymptotically large $K$. Thus, without even relying on the cooperation between the content providers, we can enjoy the benefits of data diversity in the dissemination of a single movie.

# 6 Dissimilar Interests

In our basic model, we assume all nodes have a common interest in $K$ files. In this section, we extend the common interest model to the case where each node has interest in only a subset of the $K$ files cached in the infostation. Depending on the type of content, the interests of the nodes can be *mutually exclusive* or *partially overlapping*. For instance, suppose multiple movies, say $1/\alpha$ movies are cached in the infostations, where $0 < \alpha \leq 1$. Each movie has the same length and is divided into $\alpha K$ files. If each node is interested in one movie only, then any two nodes will have interests that are either exactly the same or mutually exclusive. More generally, the interests of all nodes are partially overlapping.

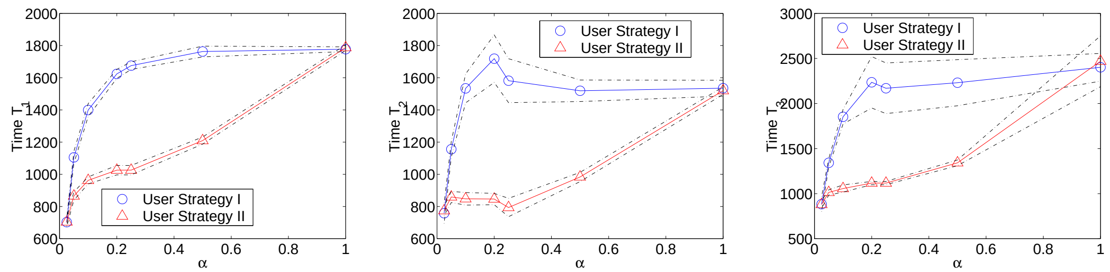
Consider the case where multiple TV shows are cached in the infostations. Without loss of generality we assume each TV show is stored as one file. Each node is interested in $\alpha K$ TV shows or files that is randomly selected from all $K$ cached files. We redefine $T_i$ as the downloading time for all files that a node is interested, i.e. $\alpha K$ files. The corresponding throughput is redefined as $C_i \triangleq \alpha K / E[T_i]$.

Recall in section 2 that a user strategy consists of two parts. Suppose two nodes seize the local channel successfully. First the two nodes must determine whether to exchange files. Second, upon an agreement of performing a file exchange, each node determines what to exchange as specified by the *random* or *greedy* strategy. In the common interest model, each node is interested in every file cached in the infostations. A node therefore is genuinely interested in every file that it does not have. In the dissimilar interest model, however, the above assumption is no longer valid. We can differentiate two user strategies in which neighbor nodes determine whether to exchange files. In **user strategy I**, neighbor nodes $A$ and $B$ perform a file exchange only if both nodes discover a file of genuine interest on inspection of each other's caches. In **user strategy II**, nodes $A$ and $B$ are obliged to exchange files if each node has a file that the other node does not have, whether or not those files are of genuine interest.

Once the nodes agree on a file exchange, either the *random* or *greedy* downloading strategy can be used in both user strategies. Nevertheless, we have demonstrated through analysis and simulations in earlier sections that the random and greedy downloading algorithms have almost identical performance. Hereafter, we consider only the random downloading strategy when we compare the performance of user strategy I and II in the simulation studies.

We have performed simulations to study the network performance for both the multiple movies model and the TV show model. For the multiple movies model, each node is interested in exactly one movie consisting of $\alpha K$ files. The interest of each node is fixed in all simulations. For the TV show model, a node is interested in each file with probability $\alpha$. Thus each node is interested in $\alpha K$ files on average. Individual node interests are varied across simulations. The network performance is evaluated in terms of $\alpha$, which characterizes the extent of overlapping interest with other nodes. When $\alpha$ is very small, each node is interested in a small fraction of all files. The interests of any two nodes are likely to be mutually exclusive. As $\alpha$ increases, more nodes are interested in the same files. It is therefore more probable for a node to run into another node that has the same interest. When $\alpha = 1$, all nodes are interested in all $K$ files and our model reduces to the common interest model.

We assume a system with $N = 40$ nodes in each infostation block and $K = 1000$ files. We consider the multiple movies model with $1, 2, 4, 5, 10, 20, 40$ movies distributed at the infostations, corresponding to $\alpha = 1, 0.5, 0.25, 0.2, 0.1, 0.05, 0.025$. In the case of 40 movies, each node is interested in different movies and have mutually exclusive interest. The number of nodes having the same interest increases with $\alpha$. When $\alpha = 1$, all nodes have a common interest for the same movie. Denote $E[T_i^{\alpha,j}],\ i = 1, 2, 3$ as the expected networking time of user strategy $j$, where $j = 1, 2$. We are interested in finding the ex-




**(a)** **(b)** **(c)**

Figure 7: Average networking time vs. the fraction of interested files $\alpha$ for multiple movies model. (a) $E[T_1]$ when 80% of all nodes obtain all files, (b) $E[T_2]$ when all nodes obtain 80% of all files, (c) $E[T_3]$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.

pected networking time for both user strategies. In the TV show model, we consider 1000 TV shows are being distributed on the infostation networks. Each node is interested in a show with probability $\alpha$, where $\alpha$ takes the values of $1, 0.75, 0.5, 0.2, 0.1, 0.05, 0.025$. In each simulation trial, the interest of each node is changed. Both the multiple movies and the TV show model have similar performance, which is not surprising. For illustration purposes, we focus on the multiple movies model below and will revisit the results for the TV show model in the end of the next section.

Referring to Figure 7, the networking time of both user strategies is plotted versus $\alpha$. We observe that even when $\alpha$ is very small, the downloading time of user strategy I is quite large. In particular, when $\alpha = 0.025$, the number of files wanted by each node is only $\alpha K = 25$. The corresponding expected networking time $E[T_i^\alpha]$, $i = 1, 2, 3$ for both user strategies is approximately 700, 750, and 850 units. At $\alpha = 0.025$, each file is desired by one node. This is easily seen since by symmetry, each file is desired by $\alpha N = (0.025)(40) = 1$ node. Suppose all nodes observe user strategy I. It is obvious there is no file exchange between nodes since each node keeps only files that is wanted by that particular node only. On the other hand, when user strategy II is used, file exchanges between nodes are allowed. Nevertheless, a node never fetches a file and benefits from a file exchange since all nodes have mutually exclusive interest. For both user strategies, each node has to download every desired file directly from an infostation. The absence of concurrent file exchanges in conjunction to infostation downloading explains the long and identical networking time.

Referring to Figure 7 again, it is obvious that $E[T_i^{\alpha,1}]$ and $E[T_i^{\alpha,2}]$ are increasing with $\alpha$ for $i = 1, 3$. This is plausible since in general, more time is needed for a fraction of nodes to finish file downloading as the number of desired files increases. An interesting (although not statistically significant) exception is observed for $E[T_2^{\alpha,1}]$, and might be explained by the following. When the number of files $\alpha K$ to be downloaded is small, a node usually runs into other nodes that have mutually exclusive interests. The node therefore has to download most of the files directly from the infostations, unable to enjoy the benefit of spatially concurrent file exchanges. As a result, these nodes have a large networking time. As $\alpha$ increases further, most, if not all, of the nodes participate in beneficial file exchanges due to the presence of nodes with the same interests. Since $E[T_2^{\alpha,1}]$ is dominated by the nodes without file exchanges when $\alpha$ is small, this explains the peak at $\alpha = 0.2$.

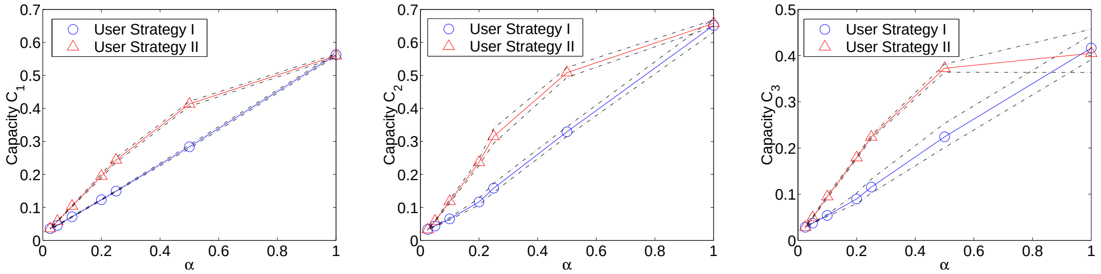
In order to explain the increasing trend of networking time with $\alpha$, and to characterize the performance difference for both user strategies, we examine the mechanism of the data dissemination in the following. As $\alpha$ increases from $\alpha = 0.025$, there are more nodes with the same interests. Each file is desired by $\alpha N$ users on average. Consider user strategy I. Approximately $\alpha N$ nodes are willing to act as the *networking agents* for each file and possibly carry the file in their cache as these nodes roams around the network. When $\alpha$ gets larger, the number of networking agents for each file increases. Since the circulation of a particular file is constrained by the number of networking agents for that file, increasing $\alpha$ effectively promotes the circulation of each file. This impacts the number of node-to-node file exchanges favorably, allowing more simultaneous file exchanges to take place. Consequently, the networking time $E[T_1^{\alpha,1}]$ and $E[T_3^{\alpha,1}]$ flatten quickly as $\alpha$ is increased.

For user strategy II, the networking time is consistently smaller than that of user strategy I as $\alpha$ increases from 0.025. Although nodes have little overlap of common interests when $\alpha$ is small, user strategy II dictates that a file exchange ensues whenever each node can retrieve a file that it does not have on inspection of the cache of the other node. Thus, all $N$ nodes are willing to act as the networking agents for all files. The circulation of each file is not constrained by the particular interests of each node. Since nodes are more admissible and willing to carry files in user strategy II, the networking time is consistently smaller.

In the case $\alpha = 1$, our dissimilar interest model reduces back to the common interest model. Both user strategies I and II have identical networking time $E[T_i^{\alpha}],\ i = 1, 2, 3$, that agrees to the corresponding values $E[T_i],\ i = 1, 2, 3$ for the common interest network model. When $K$ is reasonably large (in our case $K = 1000$), data diversity dictates that $P_E \to 1$ and the networking time is then only constrained by the contention probability $\beta$ given by (13).

# 7 Multiuser Diversity

In Figure 7, we showed that the networking time $E[T_i^{\alpha}],\ i = 1, 2, 3$ for user strategy II is always less than that of user strategy I. The corresponding network throughput is plotted versus $\alpha$ in Figure 8. Again, $x$-axis denotes the fraction $\alpha$ of files that each node is interested in, where $\alpha$ takes the values of 0.025, 0.05, 0.1, 0.2, 0.25, 0.5,1. We observe that for both user strategies, the network throughput $C_i^{\alpha},\ i = 1, 2, 3$ is strictly increasing with $\alpha$. The throughput of user strategy II is consistently larger than that of user strategy I when nodes have dissimilar interests ($\frac{1}{N} < \alpha < 1$). The throughput of both strategies coincide when $\alpha \leq \frac{1}{N}$ and $\alpha = 1$. When $\alpha \leq \frac{1}{N}$, all nodes have mutually exclusive interests. Even though user strategy II allows node-to-node file exchanges, there is no corresponding gain




**(a)** **(b)** **(c)**

Figure 8: Throughput vs. the fraction of interested files $\alpha$ for multiple movies model. (a) $C_1$ when 80% of all nodes obtain all files, (b) $C_2$ when all nodes obtain 80% of all files, (c) $C_3$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.

in network throughput. Similarly, when $\alpha = 1$, our model reduces back to the common interest model. Thus both user strategies I and II have almost identical capacities.

The increasing trend of network throughput with $\alpha$ can be understood using the concept of *multiuser diversity* inherent to mobile infostation networks. The efficiency of dissemination of this file is dependent on the willingness of the mobile nodes to carry it across the network. If a node is willing to carry a particular file, then the node is effectively acting as a *networking agent* for that file. For user strategy I, each file is wanted by approximately $\alpha N$ nodes, who are willing to act as the networking agents for the file. For strategy II, each node is obliged to carry every file even if the file is not wanted by the node. The number of networking agents is then equal to the number of nodes $N$ irrespective of $\alpha$. We argue that the performance improvement of user strategy II is an exploitation of multiuser diversity, where the number of nodes willing to act as networking agents for each file is increased. Since the circulation of a particular file is equal or less than the number of networking agents for that file, the actual circulation of each file improves as the number of networking agents increases. As a consequence of improved file circulation, the efficiency of file exchanges improves as stipulated by data diversity, allowing multiple spatially concurrent file exchanges to take place.

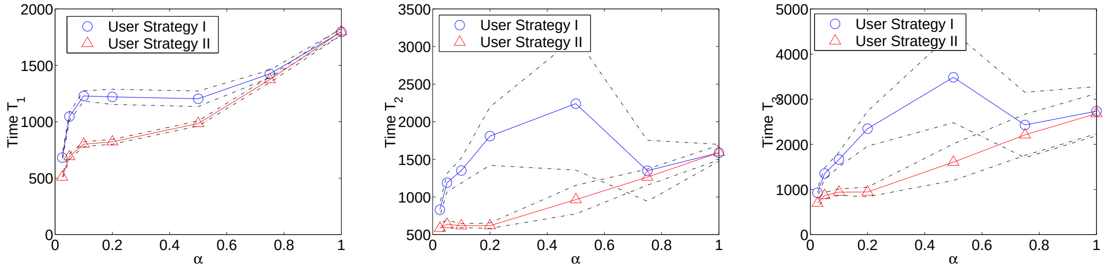
From the above argument, we expect the two user strategies have the greatest performance disparity when $\alpha$ is small. Figure 8, however, shows that the percentage performance disparity is maximum when $\alpha$ is about 0.5. We note that the increase of the number of networking agents indeed leads to a proportional increase in the number of files in circulation. However, when $\alpha$ is small, each file is of genuine interest to only a few nodes and most file exchanges involve files that are of no interest to either node. Thus even if the circulation of all files is increased significantly, the corresponding increase in the number of file exchanges is not beneficial.

There are two opposing factors that impact the performance of user strategy II. For small $\alpha$, the number of networking agents for user strategy II is increased dramatically by a factor of $1/\alpha$. However, most of the file exchanges are not beneficial since node interests are largely non-overlapping. For large $\alpha$, there is only a nominal increase in the number of networking agents. However, since most nodes have very similar interests, each node gets many desired files and benefits from file exchanges. Our simulation results show that for $\alpha = 0.5$, we achieve an attractive, and perhaps optimum, tradeoff in terms of throughput gain. The corresponding throughput $C_i^{\alpha,2},\ i = 1, 2, 3$ improvement of user strategy II over user strategy I is above 66% for all three cases.

Consider a movie distribution network in which 20 movies are cached in the infostations, making a total of $K = 1000$ cached files. Suppose each node is interested in only one movie of 50 files. This is equivalent to our multiple movies model with $\alpha = 0.05$. If all nodes observe user strategy I, the networking time $E[T_i^{\alpha,1}]$ is respectively $1100, 1200$ and $1300$ units. On the other hand, if all nodes observe user strategy II, the networking time $E[T_i^{\alpha,2}]$ is $825, 825$ and $1000$ units, roughly 70% of the original time. In content distribution, usually each node wants to minimize the networking time for files of genuine interest. Our simulation results point out that if a node acts as a networking agent for files he is not interested in, it actually expedites the file downloading process, reducing the networking time while enjoying a throughput gain as warranted by multiuser diversity. This is an interesting result because it demonstrates that each node has an incentive to act as a networking agent and assist in data dissemination without having an explicit node cooperation model.

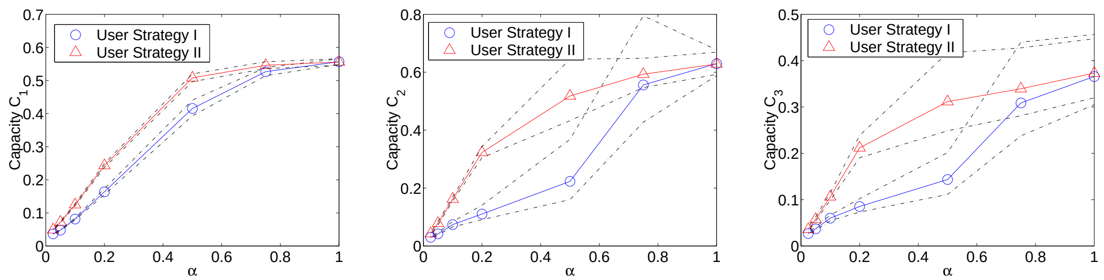
Although the exploitation of multiuser diversity in user strategy II yields better network throughput, it comes at a cost of increased energy consumption due to more frequent file exchanges. Thus there is a tradeoff between energy consumption and network throughput. We note that in mobile infostation networks, the transmit range is typically much smaller than that in multihop networks. In fact, [29] shows that a node should see around one node on average at the optimum transmit range. Classical results on multihop networks [14, 18], however, point out that a transmit range that sees 6 to 8 neighbors is optimum. If the network nodes have plentiful energy reserves, say infostations on vehicles, they should adopt user strategy II to tradeoff energy consumption for better throughput. On the other hand, nodes with very limited energy supplies can reduce their energy consumption by sacrificing some throughput. Moreover, nodes do not need to adopt the same user strategy in a network. Each node can independently decide what user strategy to adopt based on its current level of residual energy.

We note that in user strategy II, there is implicit cooperation between nodes. Each node is obliged to act as the networking agent for files that it is not interested in, That is, each node caches and disseminates personally uninterested files for other nodes as it roams the network. The performance gain of user strategy II over strategy I agrees with the intuition that more cooperation usually leads to better system performance. Although user strategy II requires implicit cooperation between nodes, there is no corresponding control overhead due to user cooperation. We do not assume the exchange of files of genuine interest to neighbor nodes takes priority over other types of file exchanges. In our implementation, when there are multiple neighbor nodes at the same location, the first two nodes that broadcast control messages to request a file exchange seize the channel. This rule is equivalent to randomly picking two nodes from all neighbor nodes with no signaling overhead and is completely determined by contention. Note that giving priority to exchanges of files of genuine interest may improve overall system performance if one




**(a)** **(b)** **(c)**

Figure 9: Average networking time vs. the fraction of interested files $\alpha$ for the TV show model. (a) $E[T_1]$ when 80% of all nodes obtain all files, (b) $E[T_2]$ when all nodes obtain 80% of all files, (c) $E[T_3]$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.




**(a)** **(b)** **(c)**

Figure 10: Throughput vs. the fraction of interested files $\alpha$ for the TV show model. (a) $C_1$ when 80% of all nodes obtain all files, (b) $C_2$ when all nodes obtain 80% of all files, (c) $C_3$ when all nodes obtain all files. The dashed lines denote the 1 standard deviation upper and lower bounds from the mean value.

can develop an efficient protocol between multiple neighbor nodes to determine the optimal node pair to exchange files.

For the TV show model, the networking times and capacities of user strategies I and II are plotted in Figure 9 and 10. The TV show model shares a lot of similar characteristics with the multiple movies model. Instead of having the same interest on $\alpha K$ files in every simulation, a node has different interests in each simulation, and interested in $\alpha K$ on average in each realization. Refer to Figure 9 and 10, user strategy II is consistently superior to user strategy I in terms of downloading time and throughput. The multiuser diversity argument continues to hold in the TV model. Under user strategy II, every node is willing to be the networking agent of any file and carry it for further dissemination inside the network. The diversity of multiple copies of the same file at multiple locations dramatically shortens the downloading time relative to that of user strategy I. Refer to Figure 9, the downloading time curves for both user strategies meet at both ends. Again, when $\alpha = 1$, the problem degenerates to the common interest problem. There is no differentiation between User Strategy I and II. On the other hand, it is likely that each file will be of interest to only one node when $\alpha <= 1/N$. Thus, an exploitation of multiuser diversity does not yield performance improvement. We note, however, that there is a small discrepancy of downloading time for User Strategy I and II. Due to the random interest of individuals, some files are desired by more than one node. In this case, User Strategy II can leverage on multiuser diversity to slightly decrease the downloading time. We also observe that the downloading time for User Strategy I $E[T_2^{\alpha,1}]$ and $E[T_3^{\alpha 12}]$ are not increasing with $\alpha$. A a similar observation was noted in the multiple movie model in the previous section for the case of downloading time $T_2$. Again, when $\alpha$ is small, a node has to download most of the files from the infostations. As $\alpha$ increases further, most nodes can participate in beneficial file exchanges due to the presence of nodes with the same interests. Thus, even when the total number of downloaded files are larger, the downloading time actually goes down. The above effect is more remarkable for the TV show model. Since the number of files a node is interested is random rather than deterministic, there may exists instances when few nodes are interested in a file even for reasonably large value of $\alpha$. This leads to exceptionally large downloading times $E[T_2^{\alpha,1}]$ and $E[T_3^{\alpha,1}]$ even when $\alpha$ is non-negligible. Refer to Figure 10, we observe that the discrepancy of the two user strategies is maximum when $\alpha = 0.5$. Again, similar observation is made for the multiple movies model. Although the ratio of networking agents diverges when $\alpha$ decreases, multiuser diversity does not improve network performance a lot. When $\alpha$ is small, most of the extraneous file exchange opportunities are futile due to incompatible node interests.

# 8 Discussions

It is apparent that the social contract defined in this paper describes one out of many possibilities to describe non-cooperation between nodes. We can modify the social contract in many different ways, say, allowing different number of file exchanges in a single transaction. If the circulation of a particular file is very small, a node $i$ is more eager to get this particular file in a rendezvous with node $j$. Knowing that the file is not widely circulated in the network, node $j$ infers node $i$ is willing to trade in the file for several files in return. The characterization of the number of files a node is willing to trade for is somewhat arbitrary and is related to the behavioral patterns of network users, and the knowledge of the file's circulation status and demand. This information will most likely not be available in individual nodes. It is not very meaningful to derive strategies to leverage on the number of files exchanges in a single transaction based on incomplete information on node behavioral patterns and the demand and supply of files. On the other hand, one-to-one file exchange is the appropriate exchange strategy when the circulation status of each file is unavailable to individual nodes. It is also intuitive that it is fair and most nodes would be obliged to follow the strategy.

In section 5, we observe that the attainable throughput of an individual node improves as the number of files being disseminated increases. In order to further improve the network performance of our system, a naive approach will be just simply dividing up a movie files into a large number of small files to leverage on the improved efficiency for a file exchange at large $K$. That is, instead of transmitting one file in each direction at the duration of one node encounter, many small files are being exchanged in each direction to fully utilize the available bandwidth. Nevertheless, the amount of overhead also increases with the use of small files. A fixed amount of bits must be set aside in each file for the purposed for synchronization and error detection. As illustrated in in section 5, for state of the art infostation structure, the file size should not be no more than 10MByte in a file transmission. This corresponds to the dissemination of $K = 500$ files for a DVD quality movie, a reasonably large number of files. Thus, we can reap the benefits of data diversity without further dividing up a 10MByte file into smaller chunks. This result is consistent to our initial assumption on the bandwidth constraint. Nodes can exchange one file in each direction in a node to node encounter.

It is plausible that downloading time of each node will decrease when nodes have some kind of implicit cooperation. For instance, a node $k$ that happens to be in the proximity of nodes $i$ and $j$ having a file exchange may simply eavesdrop the transmission and get two files from $i$ and $j$. Since file eavesdropping does not incur extraneous overhead to node $i$ and $j$, all network users will benefit if they implicitly approve of files eavesdropping. The legitimization of file eavesdropping, however, will dissuade nodes from actively participating in file exchanges. Instead of participating in file exchanges, a node may simply wait for opportunities to eavesdrop files. In the extreme case when all nodes revert to eavesdropping at all time, no data dissemination is possible. The determination of an optimal eavesdrop probability in a non-cooperative environment is thus an interesting question and should be studied further under a game-theoretic framework.

Throughout this paper, we have used the term ”non-cooperation” loosely. We go straight ahead to describe our social contract in the beginning, and abstain from providing definitions of node non-cooperation and cooperation. The line between cooperation and non-cooperation is vague. In particular, User Strategy II may be viewed as a relaxation of User Strategy I, and therefore misconstrued as a cooperative gesture between network nodes. Cooperation comes into the picture when nodes implicitly agree to carry files without intrinsic interest. Nevertheless, we note that in true non-cooperation, all network nodes are selfish and they only care about their performance. If a non-cooperative network node can reap performance gain through a unilateral change of his strategy, it will. As such, we claim there is no cooperation involved for user to adopt User Strategy II. There is nothing to gain for a node to unilaterally revoke to User Strategy I. Each node has incentive to employ user strategy II, to expedite file downloading of its own interest. There is no need for policing to ensure nodes enforcing User Strategy II.

# 9 Conclusion and Related Work

We have addressed the issue of noncooperation among nodes in the context of content distribution in mobile infostation networks. In the first part, we assume all nodes have a common interest of $K$ files cached in the infostations. We have shown that it is possible to drastically increase the rate of file dissemination of a completely noncooperative network by requiring the absolute minimal cooperation among users in the form of a social contract. A random and a greedy file downloading algorithms are examined and shown to have similar performance. We show that there exists some optimal node density in these networks such that the access probability of a node is maximized and the networking time is minimized. More importantly, we show that the total number of files cached in the infostations impacts the networking fairness and throughput. We identify this phenomenon as data diversity that is distinct from conventional communication diversity. When nodes are noncooperative and have individual preferences for data, the network exhibits data diversity and the throughput of each node increases with increasing content variety. In the second part, we extend the common interest model to the case where nodes have partially overlapping but dissimilar interests. Two user strategies are considered for this model. We show in our simulations that a file exchange strategy that takes better advantage of the multiuser diversity inherent in mobile infostations results in enhanced network performance. We conclude that both data diversity and multiuser diversity can be exploited in the mobile infostation architecture even if nodes are noncooperative.

In the present work, simple mobility and interference models are used to facilitate analysis. This approach has been fruitful, leading to the observations of two diversity phenomena in noncooperative content distribution. Nevertheless, a thorough examination of the implications of mobility and interference to the network performance of mobile infostations is called for. As a first step, the issue of interference modeling is addressed in a recent paper [28]. The effect of transmit range on network throughput is examined. We found out a stipulated transmit range improves the throughput of a mobile infostation network further. An optimal number of neighbors exists for mobile infostation networks that is distinct from the well known 6-8 magic number [11, 14, 24] for multihop ad hoc networks. Moreover, network throughput is linearly increasing with node density. Thus mobile infostation is an attractive alternative to multihop networking in future pervasive computing environments, where high node density dooms the throughput of multihop networks. On the other hand, the effect of mobility on mobile infostations is reported in [31]. Previous research assumes that the connection time in each node encounter is constant and is independent of node mobility. However, the connection time and thus the data rate of an observer node depends on node mobility and needs to be quantified. To this end a sophisticated mobility model has been proposed for highway networks that allows for performance analyses based on renewal theory.

# References


[1] J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, and J. Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proc. Mobicom*, pages 85–97, 1998.

[2] S.R. Das, C.E. Perkins, E.M. Royer, and M.K. Marina. Performance comparison of two on-demand routing protocols for ad hoc networks. *IEEE Personal Communications*, 8(1):16–28, Feb. 2001.

[3] A. El Gammal, J. Mammen, B. Prabhakar, and D. Shah. Throughput-delay trade-off in wireless networks. In *Proc. IEEE INFOCOM*, April 2004.

[4] Kevin Fall. A delay tolerant networking architecture for challenged internets. In *Proc. SIGCOMM 2003*, August 2003.

[5] R. H. Frenkiel, B. R. Badrinath, J. Borras, and R. Yates. The infostations challenge: Balancing cost and ubiquity in delivering wireless data. *IEEE Personal Communications*, 7(2):66–71, April 2000.

[6] R. H. Frenkiel, B. R. Badrinath, J. Borras, and R. Yates. The infostations challenge: Balancing cost and ubiquity in delivering wireless data. *IEEE Personal Communications*, 7(2):66–71, April 2000.

[7] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *Proceedings of IEEE INFOCOM '01*, volume 3, pages 1360–1369, 2001.

[8] P. Gupta and P.R. Kumar. Critical power for asymptotic connectivity in wireless networks. *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, pages 547–566, 1998.

[9] K. Harras, K. Almeroth and E. Belding-Royer. Delay tolerant mobile networks (dtmns): Controlled flooding schemes in sparse mobile networks. In *Proc. IFIP Networking*, May 2005.

[10] G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive mac protocol for multi-hop wireless networks. In *The seventh annual international conference on Mobile computing and networking*, July 2001.

[11] T.C. Hou and V.O.K. Li. Transmission range control in multihop packet radio networks. *IEEE Trans. Commun.*, 34(1):38–44, 1986.

[12] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.-S. Peh, and D. Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. In *ACM SIGARCH Computer Architecture News*, volume 30,5, December 2002.

[13] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Monlina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of twelfth International World Wide Web Conference*, May 2003.

[14] L. Kleinrock and J. Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. In *Proceedings of the IEEE National Telecommunications Conference*, pages 4.3.1–4.3.5, December 1978.

[15] Q. Li and D. Rus. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *ACM MOBICOM*, pages 66–75, 2000.

[16] X. Lin and N.B. Shroff. The fundamental capacity-delay tradeoff in large mobile ad hoc networks. In *Proc. Third Annual Mediterranean Ad Hoc Networking Workshop*, June 2004.

[17] M.J. Neely, E. Modiano, and C.E. Rohrs. Improving delay in ad-hoc mobile networks via redundant packet transfers. In *Proceedings of Conference on Information Sciences and Systems*, March 2003.

[18] R. Nelson and L. Kleinrock. The spatial capacity of a slotted aloha multihop packet radio network with capture. *IEEE Trans. Commun.*, 32(6):684–694, June 1984.

[19] M. Papadopouli and H. Schulzrinne. Effects of power conservation, wireless coverage and cooperation on data dissemination among mobile sevices. In *Proc. IEEE MobiHoc '01*, 2001.

[20] R. Ramanathan and Rosales-Hail. Topology control of multihop wireless networks using transmit power adjustment. In *Proc. IEEE INFOCOM '00*, pages 404–413, 2000.

[21] S.M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 1983.

[22] R. Shah, S. Roy, S. Jain, and W. Brunette. Data mules: Modeling a three-tier architecture for sparse sensor networks. In *IEEE SNPA Workshop*, 2003.

[23] T. Small and Z.J. Haas. The shared wireless infostation model - a new ad hoc networking paradigm (or where there is a whale, there is a way). In *Proc. IEEE MobiHoc '03*, 2003.

[24] H. Takagi and L. Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Trans. Commun.*, 32(3):246–257, March 1984.

[25] S. Toumpis and Goldsmith A.J. Large wireless networks under fading, mobility and delay constraints. In *Proc. IEEE INFOCOM 2004*, 2004.

[26] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. In *Technical Report CS-200006, Duke University*, 2000.

[27] M. Weiser. "the computer for the twenty-first century". In *Scientific American*, 1991.

[28] W.H. Yuen and C.W. Sung. On energy efficiency and network connectivity of mobile ad hoc networks. In *to appear, Procceedings of IEEE International Conference in Distributed Computing Systems*, 2003.

[29] W.H. Yuen and R.D. Yates. Optimum transmit range and capacity of mobile infostation networks. In *Procceedings of IEEE GLOBECOM 2003*, 2003.

[30] W.H. Yuen, R.D. Yates, and S.-C. Mau. Noncooperative content distribution in mobile infostation networks. In *Proceedings of IEEE WCNC*, 2003.

[31] W.H. Yuen, R.D. Yates, and C.W. Sung. Effect of node mobility on highway mobile infostation networks. In *Proceedings of ACM International Workshop on Modeling, Analysis and Simulations of Wireless and Mobile Systems*, Sep. 2003.

[32] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *ACM MOBIHOC 2004*, May 2004.