NONCOOPERATIVE CONTENT DISTRIBUTION IN 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 all nodes cooperate to relay packets for each other. 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. 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 occurs only when each node can obtain something it wants from the exchange. We show by analysis and simulations that network performance depends on 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. As the number of files of interest to all users increases, the achievable throughput increases. Moreover, each user has a more fair share of the total network throughput. In particular, when the number of files of shared interest is large, the transmission of each channel is only limited by contention, indicating the noncooperation strategy achieves near optimum resource utilization.

I. 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], [7]. Although rate adaptation [5] or power control [8] techniques have shown demonstrative improvement on network capacity, it is unlikely that capacity will be increased further by several orders of magnitude. In particular, [4] showed that the asymptotic throughput capacity of a wireless multihop network goes to zero as the number of nodes tends to infinity, even under the assumption of optimal scheduling and power control.

Recently, new mobile networking architectures have been proposed that exploit node mobility to achieve large network capacity. Instead of using multihop routing, networking is brought about by node mobility. In [3], nodes are connected intermittently when they are in proximity. It was shown in [3] that with a two hop relay model, the steady-state per-node network throughput scales with the number of nodes. Whereas [3] considered unicast communication, [6] considered content distribution using single hop multicast. In order to expedite data dissemination, a node also relays packets for other nodes if it has not done so for some time. The above works assume that nodes cooperate to relay packets for each other. Although typical in the wireless networking literature, this assumption is often unrealistic. When a node relays a packet for some other nodes, it expends its own bandwidth and power resources. A node therefore has no immediate incentive to forward packets for others.

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 [2]. 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 proximate nodes to act as mobile infostations to exchange copies of their files. When there are many nodes, 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 of the nodes 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.

We have shown by analysis and simulations that networking performance 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 due to the enlargement of individuals' preferences of data, and is a consequence of the assumption of noncooperation among the nodes. We conjecture that data diversity has important ramifications in the performance of other networking contexts such as multihop ad hoc networks.

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. Finally, we discuss the implications of our results in section 5.

II. SYSTEM MODEL

We look to employ a simple setting that demonstrates the characteristics of this peer to peer content distribution mechanism. 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

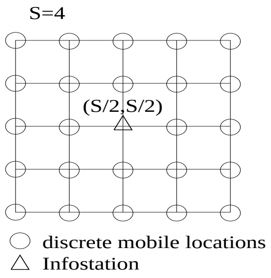


Fig. 1. Illustration of the network model.

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 system bandwidth, the size of a file is chosen such that during the time a node occupies a location at most 2 files are downloaded, or 1 file is exchanged between 2 of the collocating nodes.

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. Second, the probability of exchange is 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*. Specifically, a node may want to exchange for a file because it is genuinely interested in that file. 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 user strategy then must specify which file should be picked from the other node. In this paper, however, there is no distinction between the above models. Since 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.

After two neighbors agree to exchange files, each downloads one file from the other. In an encounter in which there are multiple files of interest, a node must decide which file to download. 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, a node randomly selects two files that he does not have for downloading at an infostation. 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. Under the social contract the two selected nodes may not perform file exchange. A practical node selection protocol should avoid this by scheduling transmissions only to the node pair with an exchange agreement. The random selection of nodes in this paper is used to facilitate performance analysis and provide a lower performance bound to an ideal node selection scheme. On the other hand, the social contract implicitly assumes there are no misbehaved 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 must be added to ensure the integrity of the files being exchanged.

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 [9].

III. PERFORMANCE ANALYSIS

When two or more mobile nodes are at the same location, a two-step process determines whether a file exchange occurs. 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 β, 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 β 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. We make the following key 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 II. 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 agrees closely with the analytical results, indicating that these assumptions work well in systems with moderately large number of files $K = 500$ 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. Based on the transmission concurrency constraint of two file transfers per location per unit time, the optimal N is about $2L$. A more careful optimization of $\beta(N)$ in (3) in the large N, L while keeping $\rho \triangleq N/L$ constant limit, i.e., equation (11), reveals that $\rho_{\text{opt}} \approx 1.8$. One can use this result to determine the optimal spatial density of fixed infostations based on the anticipated spatial density of mobile subscribers.

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

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

When a given mobile is at the infostation with $J = j$ neighbors, the probability β' 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,

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

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

$$P_E(l_i, l_j) = 1 - \frac{\binom{l_j}{l_i}}{\binom{K}{l_i}} \qquad 0 \leq l_i \leq l_j \leq K \tag{4}$$

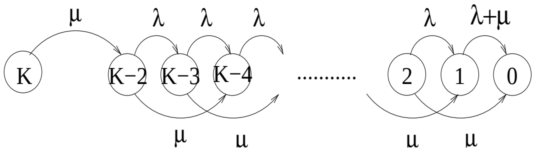


Fig. 2. Illustration of the Markov chain model.

From (4), we can derive a tight upper bound for the probability $P_{E^c} \triangleq 1 - P_E$ of no file exchange given two nodes with l_i and l_j files meet, for $\alpha K \leq l_i \leq l_j \leq (1-\alpha)K$ and $0 < \alpha < 1/2$. When K is large such that $\alpha K, (1-\alpha)K, (1-2\alpha)K \gg 1$, the asymptote of the upper bound for P_{E^c} coincides with the Stirling's approximation for P_{E^c} and is given by

$$\ln P_{E^c} \lesssim \Big[2\,(1-\alpha)\ln(1-\alpha) - (1-2\alpha)\ln(1-2\alpha)\Big] K \tag{5}$$

As the multiplier of K is negative for $0 < \alpha < \frac{1}{2}$, we deduce that when $0 < \alpha < \frac{1}{2}$,

$$\lim_{K\to\infty} P_E(l_i, l_j) = 1, \qquad \alpha K \leq l_i \leq l_j \leq (1-\alpha)\,K \tag{6}$$

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

To find an upper bound for P_{E^c} that is valid for most values of l_i and l_j, we set $0 < \alpha \ll 1/2$ in (5) and find

$$\ln P_{E^c} \lesssim -\,2\,\alpha^2 K\ , \tag{7}$$

implying that P_{E^c} can be neglected and $P_E \sim 1$ for $\alpha > O\Big(1/\sqrt{K}\Big)$. When the number of files in the system is large, file exchange almost always happens among collocating nodes 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 (6) 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 β' 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 β 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 yields the difference equation,

$$g_i = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} g_{i-1} + \frac{\mu}{\lambda+\mu} g_{i-2} \tag{8}$$

where the boundary conditions are given by $g_0 = 0$ and $g_1 = 1/(\lambda+\mu)$. Using z-transforms, we solve (8) to obtain

$$g_i = \frac{i(\lambda+2\mu) + \left(1 - \left(\frac{-\mu}{\lambda+\mu}\right)^i\right)\mu}{(\lambda+2\mu)^2} \tag{9}$$

It is obvious 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, $\lambda \sim \beta(\rho)$ and $\mu \sim \beta'(\rho)/L$ and the asymptote of the expected time for an arbitrary node to collect all K files is

$$E[T_4] \sim \frac{K}{\beta(\rho)} + \frac{L}{\beta'(\rho)} \tag{10}$$

where the asymptotes

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

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

are derived from (2) and (3). If we further allow K to grow large relative to both N and L, the corresponding throughput C_4 of a node is

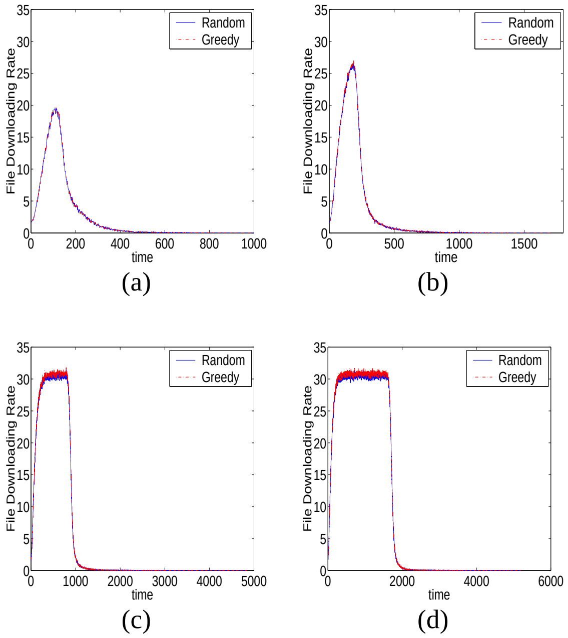
$$C_4 = \frac{K}{E[T_4]} \sim \beta(\rho), \qquad \frac{K}{N}, \frac{K}{L} \rightarrow \infty \tag{13}$$

We observe that the node density ρ that maximizes β 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 $E[T_4] = g_K = KL/2\beta'$ and

$$C_4 = \frac{2\beta'(\rho)}{L} \tag{14}$$

Hence, social contract provides an $O(L)$ or $O(N)$ (L and N are of the same order) improvement to the individual file collection rate. The social contract causes similar improvement to 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 β, file dissemination rate with 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 of social contract is of the order N.



(a) (b)

(c) (d)

Fig. 3. Average number of files obtained at each unit time over 100 simulations. (a) K=50, (b) K=100, (c) K=200, (d) K=500.

IV. 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 size of each *block* is kept constant at $L = 25$ nodes. A node moves to one of the neighbor locations w.p. $q = 0.25$ at each unit time. The performance metrics are obtained from ensemble averaging over 100 simulations.

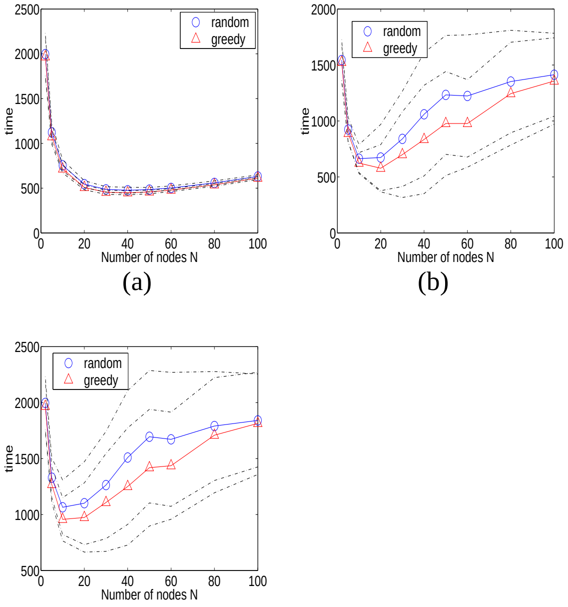
For performance evaluation, we define the *dissemination rate* as the total number of files obtained, either by download from the infostation or by file exchange, per unit time over all mobile nodes. Figure 3 shows the dissemination rate averaged over 100 simulations runs. The number of nodes is held constant at $N = 50$ and the number of files is varied ($K = 50, 200, 500, 1000$). In all cases, the differences between the random and the greedy strategies are 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. At first, 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 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 $t \to 0$, 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.

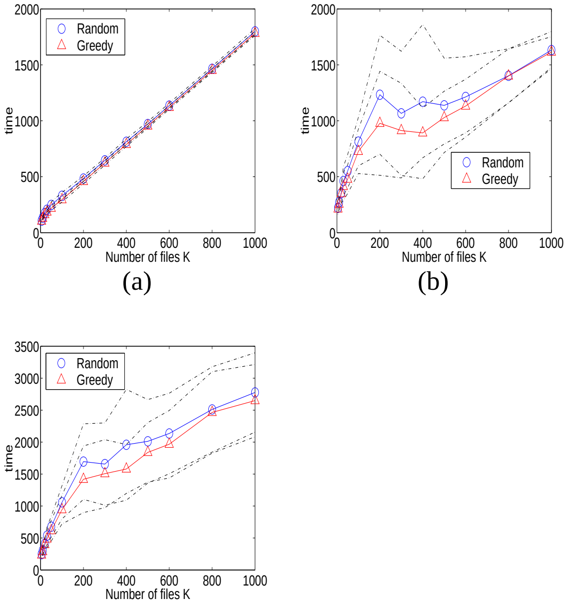
Lastly, for a finite population of nodes, we can mark the boundaries of the middle phase by the times about which all nodes have $O(\sqrt{K})$ and $O(K - \sqrt{K})$ files, based on the discussion of the upper bound of P_{E^c} after (7). We hence observe that the first and third phases require $O(L\sqrt{K})$ time roughly on the order of the time required for each node to acquire $\sqrt{K}$ file 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, 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 capacity 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 time $T_i, i = 1, 2, 3$ 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 $N^* = 45$ 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^*, $E[T_1]$ 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 $N \geq N^*$. 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.



(a) (b) (c)

Fig. 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.



(a) (b) (c)

Fig. 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.

In Figure 4(b),(c), the optimum number of nodes that minimizes the networking times T_2 and T_3 are respectively $N^* = 20$ and $N^* = 10$ 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; 80% of the nodes finish downloading all files well before hitting the long tail regime.

With reference to Figure 5, the networking times are plotted against the number of files K cached in an infostation. It is obvious that the networking time $T_i, i = 1, 2, 3$ could be fitted to an asymptote as $K \to \infty$. The variance for $E[T_1]$ is small,

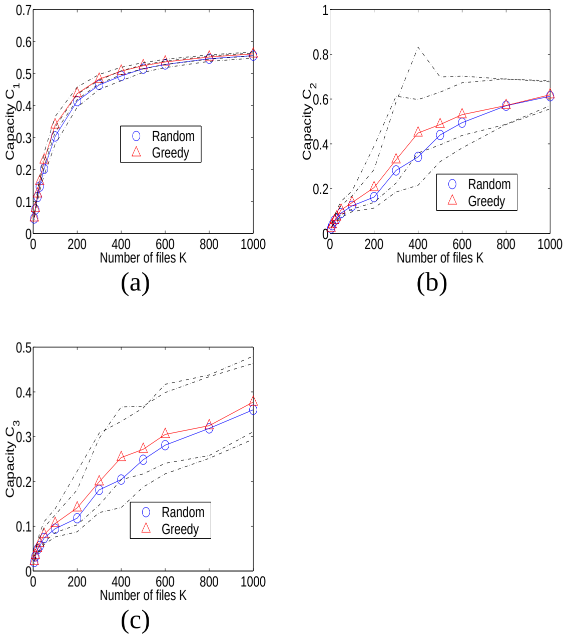


Fig. 6. Throughput capacity vs. the number of cached files K. (a) C_1 when 80% of all nodes obtain all files, (b) C_2 when all nodes obtain 80% of all files, (c) C_3 when all nodes obtain all files.

indicating that the networking effect due to node mobility is deterministic. The slope of the asymptote is found to be around 1.63, which is equal to $1/\beta(N)$. $E[T_2]$ and $E[T_3]$, on the other hand, exhibit larger variances. The slope of the asymptotes for $E[T_2]$ and $E[T_3]$ 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 $P_E \to 1$ 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 $E[T_1] > E[T_2]$.

The corresponding throughputs are plotted versus the number of files K in Figure 6. It is instructive to find the asymptotic value of throughput C_i^∞ as $K \to \infty$. To do this, we use the intuition captured in (10) and approximate the asymptote of T_i by

$$T_i^\infty = m_i K + c_i \tag{15}$$

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

$$C_i^\infty = \lim_{k\to\infty} \frac{K}{T_i} = \lim_{k\to\infty} \frac{K}{T_i^\infty} = \frac{1}{m_i} \tag{16}$$

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

$$\lim_{K\to\infty} C_3 = \lim_{K\to\infty} C_4 \tag{17}$$

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

V. DISCUSSION

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 over spatial, time and frequency domains due to multipath fading. Diversity arises when multiple signals are received, in which the strongest signal component is selected for decoding, for instance. In contrast to receiver diversity, we argue that *data diversity* is exhibited in noncooperative content distribution. Under a social contract, each node 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. (4) and (6) dictate that file exchange is more efficient as the number of file selections K increases. There are, however, some differences between receiver diversity and data diversity. We note that receiver diversity is the result of a passive environment and we can exert no influence to the outcome. Data diversity, on the other hand, is the consequence of our social contract, over which we have complete control. Nevertheless, the social contract provides a general framework to study non-cooperation content distribution in mobile infostation networks. We have shown that data diversity is relevant to noncooperative data disseminations, 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 this infostation infrastructure to disseminate their content (that are not highly overlapping) to a common group of subscribers. If a subscriber has files from provider A and he encounters another subscriber with file from provider B, these two 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.

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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] C.E. Perkins, E.M. Royer, S.R. Das, and M.K. Marina. Performance comparison of two on-demand routing protocols for ad hoc networks. *IEEE Personal Communications*, 8(1):16–28, Feb. 1998.

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

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