Optimal Cellular Traffic Offloading Through Opportunistic Mobile Networks by Data Partitioning

Ning Wang and Jie Wu
Center for Networked Computing, Temple University, USA
Email: \{ning.wang, jiewu\}@temple.edu

Abstract—In cellular traffic offloading through opportunistic mobile networks, existing schemes rely on the assumption that data can be entirely transmitted at each contact. However, transmission probability exponentially decreases as data size increases. That is, the contact duration in each contact might be insufficient for delivering large data. The objective of this paper is to find an optimal traffic offloading scheme through data partitioning so that the data delivery latency is minimized. There is a trade-off in data partitioning. Each small chunk in a path has a higher delivery probability than original data, and consequently, has a shorter delivery latency under the persistent transmission model with re-transmission. However, the destination needs to receive all the chunks in multiple paths to retrieve the data. A delay in any path will lead to a longer delivery latency. We formulate the optimal cellular traffic offloading problem and propose an approach to generate forwarding paths with possible heterogeneous data chunks. Specifically, we discuss the optimal solution for single-hop direct forwarding with multiple offloading helpers and optimal chunk sizes. Then, we propose using the node’s social-feature to generate multiple edge-disjoint multi-hop forwarding paths. Extensive experiments on realistic traces show that our scheme achieves a much better performance than those without partitioning.

Index Terms—Data offloading, data forwarding, delay tolerant networks, opportunistic mobile networks.

I. INTRODUCTION

The widespread availability of personal mobile devices has generated an opportunistic mobile network in which mobile users walk around and communicate with each other via Bluetooth or Wi-Fi in their carried short-distance wireless communication devices. The Cisco 2014-2019 White Paper [1] points out that, as of 2014, the number of mobile-connected devices has exceeded the world’s population and has led to many contact opportunities. New technologies, like Wi-Fi Aware [2], extend Wi-Fi’s capabilities in its discovery mechanism and provides better user experiences. As a result, the opportunistic mobile network has become a promising traffic offloading approach to relieve the overloaded 3G/4G cellular communication system currently in place.

Cellular offloading is that the base station sends data through cellular links to offloading helpers within convergence. These offloading helpers will forward data gradually to the destination through opportunistic links. However, previous data forwarding schemes [3–6] assume that content can always be entirely transmitted over a single opportunistic contact. With the rapidly increasing sizes of multimedia contents in recent years, such an assumption may no longer hold. We verified the INFOCOM06 and SIGCOMM09 traces [7, 8], two real-world human mobility traces in international conferences. The results show that the majority of contact durations are less than 2 minutes. Therefore, large data has few contact opportunities to be successfully forwarded.

This paper proposes a cellular traffic offloading scheme with a data partitioning to minimize data delivery latency. The original data is partitioned into multiple data chunks at the base station and sent through multiple, and preferably node-disjointed, paths. Each path is a sequence of contacts, and each chunk is sent along one distinct path. An illustration is shown in Fig. 1, where data is partitioned into two data chunks and offloaded to two offloading helpers to deliver the data opportunistically. We try to determine how much data should be forwarded to whom through data partitioning so we can fully utilize this abundance of relatively short contact opportunities. There is a trade-off in the optimal data partitioning. A small chunk size has a higher forwarding probability, and as a result, a shorter delivery latency. However, the destination needs to collect all the data chunks from multiple paths to retrieve the data. A long delay in any of the paths will lead to a long delivery latency. In addition, heterogeneous forwarding abilities of different paths make the proposed problem challenging.

For the proposed problem, we need to determine not only the number of routes (i.e., the number of data chunks) but also the size of each chunk (which can be different due to the heterogeneity of each route) to obtain the optimal performance in terms of minimum data delivery latency. We first prove that these two objectives can be unified into one objective in the optimal strategy, thus simplifying the problem. In addition, we observe the following: the number of contacts decreases exponentially as the contact duration increases. Based on the observation, we propose the optimal data partitioning in direct
forwarding routing (i.e., single-hop path, directly forwarding data chunk to the destination). In multi-hop path generation, we use the social features of nodes to build a social-hypercube to generate multi-hop node-disjointed chunk forwarding paths based on the literature [5]. Then, we apply the data partitioning scheme in the generated forwarding paths.

The contributions of this paper are summarized as follows:
- To our best knowledge, we are the first to discuss optimal heterogeneous data partitioning strategies in opportunistic mobile networks.
- We propose an optimal data partitioning strategy for direct forwarding routing, and we determine both the number of routes or data chunks and the optimal data chunk for each route.
- We propose using the node’s social-feature to generate multiple edge-disjoint multi-hop forwarding paths and conduct data partitioning.

The remainder of the paper is organized as follows. The related works are in Section II. The problem statement is introduced in Section III. The optimal direct forwarding is shown in Section IV. The social-hypercube forwarding path routing algorithm is presented in Section V. Experiment results are shown in Section VI. We conclude the paper in Section VII.

II. RELATED WORKS

In this section, we capture some important data routing issues in opportunistic mobile networks [3, 4, 9–18].

A. Replication-based Routing

The earliest routing scheme in opportunistic mobile networks is blind flooding routing [3], which does not control data copy replication. To control the number of data copies, researchers proposed data replication routing based on a quality metric. In delegation forwarding [4], relay replicates data when an encountered node has the best quality metric that it has seen. Basically, we can categorize these routing schemes into two types: contact-capability-based schemes, including [9, 10], and social-concept-based schemes, including [5, 6]. They used physical distance or social distance from the destination to guide data replication. Many works [12–14] use different quality metrics and copy control methods. However, network resource consumption is still high. Our approach differs in that we partition data rather than replicating it, which consumes fewer network resources (e.g., buffer and energy).

B. Contact-duration-aware Routing

In [11], the authors optimized the data routing based on the contact duration between nodes. However, the assumption that there is a time-ordered contact graph may not be practical, since the mobile social network is an autonomous network with much uncertainty. In [15], the authors proposed data partitioning on all better routes except the direct route to the destination. However, selecting too many forwarding paths may lead to poor performance, as this paper will later explain. In addition, they failed to discuss the route dependency problem. In this paper, we discuss the optimal data partitioning strategy and apply the concept of hypercube routing [5]. Note that [5] is replication-based routing, which is different from the proposed single-copy partition-based approach. In [17], the authors analysis the influence of limited contact duration in routing decision.

III. PROBLEM STATEMENT

In this section, we introduce the contact-aware network model, describing the idea and the challenges it presents. We use a motivational example to illustrate it.

A. Contact-Aware Network Model

In this paper, we consider cellular offloading through an opportunistic mobile network. In this model, there are \( N \) mobile nodes called offloading helpers within the base station’s coverage. The original data can be partitioned and offloaded to offloading helpers through the cellular network. After that, offloading helpers can opportunistically contact the destination node or other nodes called relay nodes through short range interfaces (e.g., Bluetooth or Wi-Fi) to forward data. Note that two encountered nodes do not know the contact duration. If the data transmission is not finished when the encountered node leaves, the data forwarding fails and the encountered node will discard the incomplete data. The current node will transmit this data again during its next contact opportunity between them. This is called persistent transmission because the node will conduct repeated transmissions in next contacts until the first successful transmission.

Based on the contact-aware network model, the base station makes a data-forwarding decision, i.e., which data to send and how much data to send, data delivery probability increases as the data size decreases over the opportunistic path/route. When the data size is small enough, it can utilize all contact opportunities. Let us denote this delivery probability of the path \( i \) as \( p_i \) as the base delivery probability during one contact opportunity, i.e., contact probability without considering the real contact duration. Then, as the data size increases, the delivery probability of a path decreases called decay function, \( \beta(s) \), which is non-increasing with the data size \( s \). Therefore, the delivery probability of a path for data size \( s \) is,

\[
p_i(s) = p_i \cdot \beta(s).
\]

Therefore, the overall delivery probability, \( P \), of a data with \( m \) paths is \( P = \prod_{i=1}^{m} p_i(s_i) \).

B. Problem and Challenges

The proposed optimal data-partitioning problem is defined as follows, data of size \( S \) can be partitioned/fragmented into small data chunks, \( \{s_1, \ldots, s_m\} \), and each data chunk is forwarded along one forwarding path. Then, the main problem of data partitioning offloading is to find an optimal strategy so that the delivery delay of the data can be minimized. To do that, first, we need to generate a set of forwarding paths. After that, the problem becomes a joint optimization of the path number and the chunk size for each path.

The proposed problem is challenging even if the forwarding paths are known. An illustration of the challenge is shown...
in Fig. 2. The source node has a data item with size $S$ to the destination, and there are three disjointed opportunistic paths labeled. The base delivery probabilities of all three paths are 0.9, 0.8 and 0.7, respectively. Specifically, the probability decays with data size $S$, $2S/2$, $S/2$, and $S/3$, are 0.10, 0.36, 0.47 and 0.64, respectively. Without data partitioning, the best forwarding probability of a single copy is only 0.09. However, we can further partition the data into three even chunks and use all three paths, as shown in Fig. 2(a), delivery probability then becomes 0.132, which is better than without data partitioning. If we use the best two paths and distribute data evenly in two paths as shown in Fig. 2(b), delivery probability becomes 0.159. But this is still not optimal. If $\frac{3}{2}S$ is delivered via the first path and the remaining data is delivered by the second path, the overall delivery probability is 0.166.

The proposed problem is challenging due to the trade-off. If we partition the data into smaller chunks, each small chunk has a higher delivery probability to the destination. However, the destination needs to receive chunks from more routes to recover the original data. In addition, considering heterogeneous path delivery abilities, we need to determine not only the number of data chunks but also the size of each data chunk to optimize the performance.

IV. DIRECT FORWARDING ROUTING

In this section, we will discuss the optimal data offloading with data partitioning in single-hop direct forwarding, i.e., the offloading helper directly forwards data to the destination.

A. Chunk Assignment and Path Selection

Each offloading helper’s forwarding ability to destination is different. Therefore, we need to determine not only the number of offloading helpers but also the size of each data chunk carried by each offloading helper. These two correlated objectives complicate our objective.

In data partitioning routing, the bottleneck of the data delivery delay is determined by the route with the largest delivery delay among all the routes. Therefore, if we can minimize the maximal delivery delay of the data, the overall delivery delay will be minimized. We prove this according to the following theorem.

Theorem 1. If $\frac{d\ln\beta(s)}{ds}$ is non-increasing for $s$, to maximize the overall delivery probability, $\beta(s_1) = \cdots = \beta(s_m)$.

Proof. To prove the Theorem 1, let us assume the sizes of data chunks in two different paths are $s_1$ and $s_2$ and $\Delta$ is a small value. Without loss of generality, we assume that $s_1 < s_2$. It is equivalent to prove that the following inequality is true,

$$p_1\beta(s_1) \times p_2\beta(s_2) \leq p_1\beta(s_1 + \Delta) \times p_2\beta(s_2 - \Delta). \tag{2}$$

The idea of Eq. 2 is to balance the data size in two paths. We can re-write Eq. 2 as the following equation since the $\beta(s)$ function is positive.

$$\frac{\beta(s_2)}{\beta(s_2 - \Delta)} \leq \frac{\beta(s_1 + \Delta)}{\beta(s_1)}. \tag{3}$$

Since $\beta(s)$ is a non-increasing function, and therefore,

$$1 - \frac{\beta(s_2)}{\beta(s_2 - \Delta)} \geq 1 - \frac{\beta(s_1 + \Delta)}{\beta(s_1)}. \tag{4}$$

This equals

$$\frac{\beta(s_2) - \beta(s_2 - \Delta)}{\beta(s_1) - \beta(s_1 + \Delta)} \geq \frac{\beta(s_1) - \beta(s_1 + \Delta)}{\beta(s_1) - \beta(s_1 + \Delta)},$$

which can be approximated as

$$\frac{1}{\beta(s_2)} \frac{d\beta(s_2)}{ds_2} \leq \frac{1}{\beta(s_1)} \frac{d\beta(s_1)}{ds_1}. \tag{6}$$

Therefore, when $\frac{d\beta(s)}{ds}$ is non-increasing, Eq. 2 is true. This means that evenly partitioning data is the optimal strategy.

The insight behind Theorem 1 is that if the condition in Theorem 1 is satisfied, then the decay function dominates the result, i.e., a path with a large data chunk can decrease the overall performance significantly. Therefore, a good strategy is to ensure that the maximum data chunk size is minimized. Therefore, we partition data so that each path has the same chunk size. We verified the INFOCOM06 and SIGCOMM09 traces [7, 8], two real-world human mobility traces in international conferences, and observe the following:

$$\beta(s) = \lambda e^{-\lambda s}. \tag{7}$$

This fits Theorem 1. Fig. 3 shows the fitting result: evenly partitioning data is optimal in mobile opportunistic networks.

In direct forwarding routing, relay nodes directly deliver the data chunks to the destination. Based on the analysis in [9] and [7], the contact frequency between nodes follows the exponential distribution with a rate parameter $\mu$. The insight of the distribution is that majority of contacts are among friends rather than strangers.

Theorem 2. In terms of maximizing the data delivery probability in direct forwarding routing, there is a percolation point, i.e., if $\lambda\mu > 1$, we should select as many offloading forwarders as possible. Otherwise, only selecting one offloading forwarder is optimal, where $\mu$ and $\lambda$ are distribution parameters of $p_i$ and $\beta(s)$, respectively.
Routing. Figs. 4 and 5 are trace analysis results from real traces. Therefore, we can use social information to guide nodes with similar social features have a large contact. A common assumption of opportunistic mobile networks is that nodes with similar social features have a large contact frequency. Therefore, we can use social information to guide routing.

A. Forwarding Path Generation

1) Social-Feature-Space: Since mobile devices are carried by humans, opportunistic mobile networks have social features [5]. A common assumption of opportunistic mobile networks is that nodes with similar social features have a large contact frequency. Therefore, we can use social information to guide routing. Figs. 4 and 5 are trace analysis results from real traces.

![Fig. 3. Probability density function of contact duration in INFOCOM and SIGCOMM datasets.](image)

![Fig. 4. Contact summary in different feature distances in the INFOCOM trace.](image)

![Fig. 5. Contact summary in different feature distances in the SIGCOMM trace.](image)

Proof. Assume that the original data is partitioned into \( m \) paths to its destination, \( s_1, s_2, \ldots, s_m \). \( \sum_{i=1}^{m} s_i = S \). Therefore, the forwarding ability of path \( i \) is \( p_i(s_i) = p_i \cdot \beta(s_i) = \mu \lambda e^{-(s_i+\lambda s_i)} \), and the overall delivery probability with \( m \) paths is proportional to

\[
P = \prod_{i=1}^{m} p_i \beta(s_i) = \prod_{i=1}^{m} \mu \lambda e^{-(s_i+\lambda s_i)}.
\]

(8)

Based on Theorem 1, the delivery probability of any two offloading helpers should be the same in the optimal solution. That is, \( e^{-\lambda s_1} = e^{-\lambda s_2} \). Then, we get the following equation.

\[
s_i = S/m
\]

(9)

The problem is reduced to find the optimal \( m \) as in Eq. 8.

\[
P = \prod_{i=1}^{m} (\mu \lambda e^{-\lambda s_i})^m = (\mu \lambda)^m e^{-\mu \lambda S}
\]

(10)

Since the function \( P \) is a concave function in terms of \( m \), when its first differential reaches 0, the function is maximum.

\[
\frac{dP(m)}{dm} = \frac{d(\mu \lambda)^m e^{-\mu \lambda S}}{dm} + (\mu \lambda)^m \frac{de^{-\mu \lambda S}}{dm} = (\mu \lambda)^m \ln(\mu \lambda) e^{-\mu \lambda S}
\]

(11)

Therefore, there is a percolation point, i.e., if \( \lambda \mu > 1 \), we should select as many offloading forwards as possible. Otherwise, only selecting one offloading forwarder is optimal.

The insight is that when \( \mu \lambda < 1 \), we prefer not to partition data since the base forwarding ability decays very quickly as the path number increases and the latter path will be the bottleneck. If \( \mu \lambda > 1 \), we prefer to partition data. It is because that the base forwarding abilities are similar and therefore if we partition data into small chunks, the each path’s forwarding ability increases significantly.

V. SOCIAL PATH ROUTING

In this section, we propose to use social-feature-space and social-hypercube routing scheme to generate multi-hop paths.

A. Forwarding Path Generation

1) Social-Feature-Space: Since mobile devices are carried by humans, opportunistic mobile networks have social features [5]. A common assumption of opportunistic mobile networks is that nodes with similar social features have a large contact frequency. Therefore, we can use social information to guide routing. Figs. 4 and 5 are trace analysis results from real traces.
Algorithm 1 Forwarding path: source node

Input: Social feature of nodes and the base forwarding ability of each path.

Output: The m disjoint paths and the corresponding chunks.

1: for \( i = 1 : n \) do
2: \[ \text{if } i \in \{1, 2, \ldots, k\} \text{ then} \]
3: Use the the circular left shifts \( C' \) to generate a path;
4: \[ \text{if } i \in \{k + 1, \ldots, n\} \text{ then} \]
5: Reach the dimension \( i \) first, then use any permutation of \( C \) to generate a path;
6: Sort all paths based on the base delivery probability.
7: for \( m = 1 : n \) do
8: Assign data chunk evenly according to Theorem 1.
9: Find the \( m \) with largest forwarding ability.

Algorithm 2 Forwarding path: relay node

Input: The forwarding path, \( l_i \), of the current, chunk holder, the group \( g_j \) of the encountered node.

if The encountered node is in the destination’s group then
1: Forward data to the encountered node.
2: if \( g_j \in l_i \) then
3: Delete the groups before \( g_j \) in \( l_i \).
4: Forward the data and \( l_i \) to the encountered node.

Once the paths are generated, the chunks in each path will be forwarded to the destination along the forwarding path group by group. Note that in Algorithm 2, some chances to shorten the distance to the destination with more than one group will not be used. Therefore, the short-cut to the destination is allowed and can improve the performance significantly. For example, \( g_0 \) can forward the chunk directly to \( g_6 \) as a shortcut for path \((g_4, g_0)\).

VI. EXPERIMENTS

A. Experiment Setting

Detailed dataset information is shown in Table I. Particularly, In SIGCOMM trace, there is no contact duration information. In the experiments, if we detect two contacts for two nodes within 120 seconds, we regard these two nodes are contacted. In addition, nodes without social features are removed from experiments. the data size varies from 4MB to 50MB in the INFOCOM trace and from 100MB to 600MB in the SIGCOMM trace. For both traces, the data deadline is one day. The contact bandwidth between nodes is assumed to be constant, 100kbps. In the experiments, both the source and the destination nodes are randomly selected in each round.

B. Algorithm Comparison

In our experiments, we propose 4 algorithms for performance comparison. They are divided into 2 categories. (1) comparison algorithms: the social feature-space routing algorithm without data partitioning, denoted as the FN algorithm uses the same number of path as the FY algorithm; the single-copy probability-based algorithm without data partitioning, denoted as the SN algorithm. (2) The partition versions of the aforementioned two algorithms denoted as the FY algorithm and the SY algorithm, respectively.

C. Experimental Results

1) With/Without data partitioning: Figs. 7(a) and 7(b) show the delivery delay and the delivery ratio for different data sizes in the INFOCOM trace. In Fig. 7(a), we find that delivery delay can be greatly reduced using data partitioning. Note that the FY algorithm reduces 20%, 60% delivery delay of the SY and SN algorithm when the data size is large. The proposed data partitioning algorithm achieves a similar performance as for the data delivery ratio when the data size is small. As the data size increases, the proposed data partitioning algorithm achieves a better performance. In summary, the performances of these algorithms from best to worst follow the order: FY, SY, FN, and SN. The delivery ratio improves by 2%, 15%, and 45%, respectively. Figs. 8(a) and 8(b) show the results of the performance in the SIGCOMM trace, which are similar to those in the INFOCOM trace. With proper data partitioning, the FY further reduces the delay by 40%.

2) Different data partitioning strategies: Figs. 7(c) and 7(d) show the results of different data partitioning algorithms in the INFOCOM trace. The data size is 50MB in this set of experiments. Results show that without data partitioning, the performance of the proposed algorithms is really poor. In fact, in such a setting, the data cannot be delivered to the destination with limited contact opportunities. However, if we partition the data into small chunks increase contact opportunities can achieve a relatively good performance. In such as setting, FY and SY algorithms achieve better performance with smaller chunk size. In the SIGCOMM dataset, we find similar results when the data size is 600MB. In addition, we further verify the necessary to find the optimal data partitioning strategy. In Figs. 8(c) and 8(d), if we partition the data into many

<table>
<thead>
<tr>
<th>Table I: Dataset Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>INFOCOM06</td>
</tr>
<tr>
<td>SIGCOMM09</td>
</tr>
</tbody>
</table>
chunks, the performance of the FY and SY algorithms is even worse than algorithms without partitioning. The FY and SY algorithms show the best performance when data is partitioned into 3 chunks. This verifies the disadvantage of the algorithm proposed in [15], which always uses all the paths.

VII. CONCLUSION

In this paper, we observe limited contact duration in opportunistic mobile networks. Unlike existing works that use data replication schemes to improve performance, we propose a data partitioning forwarding algorithm to utilize short-contact duration in networks to minimize the data delivery delay. Specifically, we discuss optimal solution for direct forwarding with multiple offloading helpers and optimal chunk size on each helper based on the observation that the number of contacts decreases exponentially as the contact duration increases. Then, we propose using the node’s social-feature to generate multiple edge-disjoint multi-hop forwarding paths. Extensive experiments on realistic traces show that our scheme achieves a better performance than those without partitioning.

VIII. ACKNOWLEDGMENT

This research was supported in part by NSF grants CNS1629746, CNS 1564128, CNS 1449860, CNS 1461932, CNS 1460971, CNS 1439672, CNS 1301774, and ECCS 1231461.

REFERENCES