Social-Aware Data Forwarding through Scattered Caching in Disruption Tolerant Networks

HyunAe Kim and HyungJune Lee[†]
Department of Computer Science and Engineering
Ewha Womans University, Seoul, South Korea
Email: nika1108@i.ewha.ac.kr, hyungjune.lee@ewha.ac.kr

***Abstract*—We present a social-aware data forwarding scheme from a mobile source to a mobile destination in disruption tolerant networks. By incorporating any scattered network devices as temporal data storage for forwarding into parts of forwarding path, we aim to find the most effective relay path consisting of mobile-to-stationary and stationary-to-mobile relays. We combine the "carry-and-forward" scheme for stationary-to-mobile transmission with the "store-and-forward" for mobile-to-stationary transmission for improving both delivery rate and packet delay performance. We formulate this relay selection problem into a mixed-integer linear program, considering two crucial QoS constraints of packet delivery rate and deadline. We find the optimal forwarding path of all relevant mobile relays as well as their corresponding stationary nodes as bridges between mobile relays. We validate our algorithm based on a real-world dataset in terms of routing cost and packet transmission time compared to a baseline counterpart.**

I. INTRODUCTION

In disruption tolerant networks (DTNs), routing from a user to another may be disrupted due to various reasons of limited radio range, sparsity of mobile users, separation of social groups, and battery drainout [13]. In this inherently connection-scarce networks, opportunistic data delivery is used to increase data accessibility. To make end-to-end delivery to a destination user successful, subsequent mobile users are needed as intermittent relays by following "carry-and-forward"-type transmissions.

To establish end-to-end delivery by exploiting a series of mobile relays, both node mobility and mobile contact are key driving forces for carrying and spreading information. However, contact events between two mobile users are transient, and guaranteeing data delivery is challenging and often infeasible. Instead, research works in DTNs tend to use steady-state statistical contact probability and contact interval to design some data forwarding schemes.

There have been numerous prior works on DTN routing. Some representative works [2], [6], [7], [9], [11] leverage the properties of centrality and/or similarity in contact patterns among mobile users. To deliver data to a dissimilar mobile user, mobile nodes with high centrality that have more popularity of connecting other nodes are preferred to be selected as relays. On the other hand, mobile users with high similarity in contact pattern can be used as effective relays in case of data delivery to a relatively similar mobile user.

However, these works rely too much on opportunistic mobile contact events, providing the only opportunity for data relay toward a destination. This suffers from low delivery chance with uncontrollably large packet transmission delay. Recently, data forwarding schemes have been proposed by using stationary nodes deployed in the network together with mobile relays [1], [5], [10], [12]. In particular, [5] presents a cooperative caching strategy that intentionally caches data at a set of network central locations, while optimizing the tradeoff between data accessibility and caching overhead.

In this paper, we focus on routing efficiency for exploiting surrounding stationary network devices as caches and also mobile users as mobile relays. We take into account some more practical constraints: packet delivery rate and packet delivery deadline at the same time based on an optimization approach that has not been well considered in previous DTN routing research. Given a desirable packet delivery rate and a packet delivery deadline, we aim to find the most efficient routing path consisting of mobile-to-stationary and stationary-to-mobile transmissions that minimizes total routing cost.

We combine the "carry-and-forward" scheme for stationary-to-mobile transmission with the "store-and-forward" for mobile-to-stationary transmission for improving both delivery rate and packet delay performance. Socially-closer users may meet more frequently with each other at some common places or paths [3], [4]. We exploit steady-state information of each mobile user's visit probability and inter-visit time over stationary nodes in the networks where they reflect implicit social behavioral pattern. By formulating the relay selection problem into a mixed-integer linear program (MILP), we optimize the relay selection of both mobile relays and stationary caches in terms of routing cost. At the same time, we also want to satisfy packet delivery constraints in delivery rate and deadline.

Our approach completes the procedure of finding all relevant mobile relays as well as their corresponding stationary caches for each mobile pair. This approach has an advantage over a traditional sequential sub-optimal solution that selects a series of suitable mobile relays, and then accordingly finds subsequent stationary caches as a bridge between mobile relays.

This paper basically provides an optimal practical strategy for the following fundamental problem: Given any surrounding scattered network devices as temporal memories in DTN networks, how to exploit them selectively for effective ad-hoc

---

[†] HyungJune Lee is the corresponding author.

communication between mobile users with QoS.

The remainder of this paper is organized as follows: After presenting the system model in Sec. II, we propose our optimal relay selection scheme in Sec. III. We discuss the evaluation results of our approach in Sec. IV, and then conclude this paper in Sec. V.

II. SYSTEM MODEL

This paper considers the problem of data forwarding from a mobile source to a mobile destination where stationary nodes (e.g., 802.11 access point, Bluetooth beacon, or NFC tag) are deployed in the networks, and other mobile users move around the network. The objective of this paper is to find a set of mobile relays that can relay data from the source user to the destination user with high probability within the packet deadline constraint. To relay a mobile user to another, it is allowed to use stationary nodes as temporary data *caches* with the "store-and-forward"-type relaying. Our goal is to minimize total routing cost considering all of stationary and mobile relaying.

Stationary nodes are scattered over the network and not necessarily connected one another. The stationary nodes can communicate with mobile nodes within wireless radio range, and vice versa. We do not constrain ourselves to use only one determined wireless standard, but rather either 802.11x, Bluetooth, 802.15.4, or NFC can be used for the hop-by-hop "store-and-forward" transmission. We assume that mobile users actively roam around the network within structured environment such as streets and buildings. To guarantee a stable packet relaying from a mobile user to stationary nodes, it is allowed to multicast to multiple stationary nodes, increasing the probability to successfully relay to the next mobile relay user (or destination) within less amount of relay time. It is also assumed that each packet needs to be delivered from its source user to its destination user within packet deadline over the entire relay paths of mobile relays and stationary caches. It should be noted that stationary nodes are not necessarily connected among themselves.

The problem of data forwarding can then be described as the relay selection problem (as illustrated in Fig. 1) of selecting both stationary caches and mobile relays that minimizes total routing cost, while achieving high packet delivery rate, and satisfying packet deadline constraint.

III. OPTIMAL RELAY SELECTION

In DTN routing, the extraction of social network structure improves data forwarding efficiency by finding a series of mobile bridges to a destination user [2]. However, this traditional "carry-and-forward"-type opportunistic forwarding inevitably suffers from very long delay, which is usually not applicable in real-world applications.

In particular, two somewhat dissimilar mobile users in terms of social relationship and movement behavior over time and space tend to not share common characteristics with each other, and their direct contact probability would accordingly be very low. This means that it is very challenging to deliver data from a mobile user to a different type of user, causing a considerably large packet transmission time.

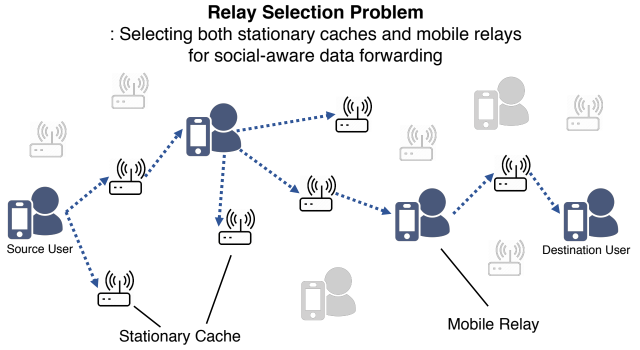


Fig. 1. Relay selection problem where stationary nodes can be used as temporary data caches, and mobile users connect to a part of them along their move.

As network devices with many different wireless standards such as 802.11x, Bluetooth, 802.15.4 Zigbee, and NFC have been ubiquitously deployed in cities, these devices can be utilized as temporary data storage beyond their original communication function. For example, even if the direct contact probability between dissimilar mobile users is low, they may have some common visiting places such as library, cafe, or restaurant, which are all static.

If this kind of place is surrounded by network devices, there is a higher chance that a mobile user can push data to a part of them, and then another user can pick up it later, based on the "store-and-forward" property. This is practically more feasible even if there is no direct contact between them, as long as indirect contact events can occur through *well-selected* stationary network caches.

We extend this problem into the one requiring multiple mobile relays with stationary caches available to use. We combine the "carry-and-forward" scheme for stationary-to-mobile transmission with the "store-and-forward" for mobile-to-stationary transmission. Our approach completes the procedure of finding all relevant mobile relays as well as their corresponding stationary caches for each mobile pair. It should be noted that although our approach focuses on sequential series of mobile-to-stationary and stationary-to-mobile relaying, our work does not exclude a possibility of combining with the opportunistic mobile-to-mobile relaying.

First, we define steady-state indirect contact probability of mobile user M_i to mobile user M_j through stationary node S_k as $P^{S_k}_{M_i->M_j}$. A high value of $P^{S_k}_{M_i->M_j}$ means that stationary node S_k is an effective bridge to relay data from mobile user M_i to mobile user M_j with high social accessibility, improving packet delivery rate. Also, the steady-state inter-visit time of mobile user M_i to stationary node S_k is defined as $\tau_{M_i->S_k}$. As mobile user M_i visits stationary node S_k more frequently in his/her own movement pattern, the value of $\tau_{M_i->S_k}$ decreases. This implies that it takes less amount of time to relay data from mobile user M_i to stationary node S_k, and vice versa. The steady-state indirect contact probability as well as inter-visit time can easily be calculated once empirical data consisting of contact events of mobile users to stationary nodes are collected, like Dartmouth dataset [8]. The contact probability is calculated for each mobile user to a visiting stationary node by measuring total connection time at that stationary node out of total connection time to any stationary node for the user. We calculate the inter-visit time by taking the average inter-connection time between the user's connections to the same stationary node.

We incorporate two packet constraints in delivery rate and delivery time, considered as important QoS metrics in practice. Our work requires delivery rate between mobile users M_i and M_j for all i and j to be larger than or equal to delivery threshold δ. To satisfy a given delivery threshold, multiple stationary caches need to be selected for data forwarding from mobile user M_i. However, multiple stationary cache selection can incur high routing cost in return. This causes a non-trivial trade-off problem between delivery rate and routing cost.

To employ the packet delivery time constraint, our work enforces delivery time less than a given packet deadline T_{max}. Under this constraint, a stationary cache S_k with very high contact probability for the mobile pair of M_i and M_j, but with very large inter-visit time $(\tau_{M_i->S_k}+\tau_{M_j->S_k})$ is prohibited for selection even though it is considered as a very good relay in terms of delivery rate. Another stationary cache with less contact probability, but with smaller inter-visit time can be selected instead, increasing the number of selected stationary caches and routing cost accordingly. This ends up with another non-trivial trade-off problem between packet deadline and routing cost.

To strengthen the packet delivery time constraint, our work further requires the worst case forwarding between mobile users M_i and M_j among various cases using multiple stationary caches. We define the longest inter-visit time from M_i to M_j through all possible intermediate stationary cache S_k as $C_{M_i \rightarrow M_j}$. Therefore, the accumulated value over the entire selected path should be less than packet deadline T_{max}. In this way, it guarantees that any selected data forwarding path other than the worst case forwarding path saves a relatively larger time margin from the deadline.

To find the optimal balance over two non-trivial trade-off problems, we formulate this problem of relay selection consisting of both stationary and mobile relays into a mixed-integer linear program (MILP). The proposed scheme simultaneously finds the optimal set of mobile-to-stationary and stationary-to-mobile paths that minimizes end-to-end routing cost, while satisfying packet delivery threshold δ and packet deadline T_{max}.

To formulate the optimization problem, we first introduce indicator functions $I^{S_k}_{M_i \rightarrow M_j}$ denoting whether data forwarding path from mobile user M_i to mobile user M_j through stationary cache S_k is selected as a part of entire routing path. Additionally, we introduce another indicator functions $J_{M_i \rightarrow M_j}$ denoting whether data forwarding from mobile user M_i to mobile user M_j is selected as a part of routing path.

In our experiments, we use the number of transmissions from mobile-to-stationary multicast relaying in case of using multiple intermediate caches and stationary-to-mobile unicast relaying as the end-to-end routing cost.

We set the objective function of end-to-end routing cost to minimize as

$$\sum_{i,j,k} I^{S_k}_{M_i->M_j} + \sum_{i,j} J_{M_i->M_j}$$

where $\sum_{i,j,k} I^{S_k}_{M_i->M_j}$ is the incurred routing cost of total number of mobile-to-stationary relay transmissions considering multicast to stationary caches, and $\sum_{i,j} J_{M_i->M_j}$ is the incurred routing cost of total number of stationary-to-mobile relay transmissions with a single unicast from one stationary cache to the next mobile relay.

To this end, the relay selection problem can be formulated into a mixed-integer linear program consisting of binary integer variables of $I^{S_k}_{M_i->M_j}$ and $J_{M_i->M_j}$ and non-negative non-integer variables $C_{M_i \rightarrow M_j}$ as follows:

$$\begin{aligned}
\text{minimize} \quad & \sum_{i,j,k} I^{S_k}_{M_i \rightarrow M_j} + \sum_{i,j} J_{M_i \rightarrow M_j} && (1)\\
\text{subject to} \quad & \sum_{j} J_{M_S \rightarrow M_j} = 1 && (2)\\
& \sum_{i} J_{M_i \rightarrow M_D} = 1 && (3)\\
& \sum_{i} J_{M_i \rightarrow M_S} = 0 && (4)\\
& \sum_{j} J_{M_D \rightarrow M_j} = 0 && (5)\\
& \sum_{j,k} I^{S_k}_{M_S \rightarrow M_j} \geq 1 && (6)\\
& \sum_{i,k} I^{S_k}_{M_i \rightarrow M_D} \geq 1 && (7)\\
& \sum_{i} J_{M_i \rightarrow M_j} = \sum_{k} J_{M_j \rightarrow M_k} \leq 1 \quad \forall j && (8)\\
& I^{S_k}_{M_i \rightarrow M_j} (\forall k) \leq J_{M_i \rightarrow M_j} \leq \sum_{k} I^{S_k}_{M_i \rightarrow M_j} && (9)\\
& \sum_{k} P^{S_k}_{M_i \rightarrow M_j} \cdot I^{S_k}_{M_i \rightarrow M_j} \geq \delta \cdot J_{M_i \rightarrow M_j} && (10)\\
& (\tau_{M_i \rightarrow S_k} + \tau_{M_j \rightarrow S_k}) \cdot I^{S_k}_{M_i \rightarrow M_j} \leq C_{M_i \rightarrow M_j} \ \forall k && (11)\\
& \sum_{i,j} C_{M_i \rightarrow M_j} \leq T_{max}. && (12)
\end{aligned}$$

To embed the source and destination conditions, we enforce constraints (2) – (7). Constraints (2) – (3) ensure that the source user should start forwarding data toward the network, and it should be delivered eventually to the destination user. Constraints (4) – (5) avoid any routing loop case in the entire forwarding path. Constraints (6) – (7) make sure that the source user should start forwarding data toward the network through at least one stationary cache, while it should be delivered eventually to the destination user through at least one stationary cache.

To make the flow conservation for packet ingress and egress over the network, we constrain ourselves with constraint (8).

Constraint (9) describes intertwined relationships between $I^{S_k}_{M_i \to M_j}$ and $J_{M_i \to M_j}$. This enforces a condition that if the data forwarding path from M_i to M_j is not selected, i.e., $J_{M_i \to M_j} = 0$, all the relevant stationary paths should not be selected, i.e., $I^{S_k}_{M_i \to M_j} = 0 \; \forall k$. On the other hand, once the data forwarding path from M_i to M_j is selected, i.e., $J_{M_i \to M_j} = 1$, at least one stationary cache in the middle between M_i and M_j requires to be selected, i.e., $1 \leq \sum_k I^{S_k}_{M_i \to M_j}$.

To make total contact probability from M_i to M_j through all selected stationary caches above the delivery threshold δ, constraint (10) is employed. Only if the path from a specific M_i to M_j is selected, i.e., $J_{M_i \to M_j} = 1$, we force this constraint.

Lastly, we convert a constraint of the worst case packet delay related to $C_{M_i \to M_j}$ into a series of inequalities. Constraints (11) – (12) ensure that the worst case delay is less than or equal to packet deadline T_{max}.

Mobile source devices or a centralized server upon request run the optimization algorithm, and obtain the optimal set of stationary and mobile relays by using MATLAB `intlinprog` or AMPL/CPLEX MILP solvers.

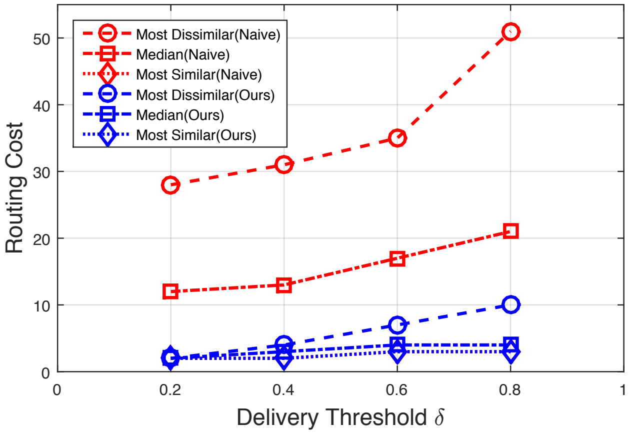
IV. EVALUATION

We validate our proposed relay selection scheme using Dartmouth dataset [8] where wireless access points (APs) are used as stationary caches for validation purpose. We picked up the most active 15 mobile users in the sense of having connected to the largest number of distinctive APs on association during the fall semester of 2003 where total 475 associated APs have appeared in the dataset.

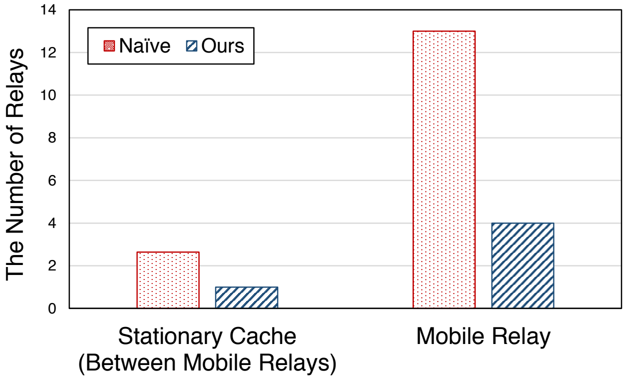
From this dataset, we calculate steady-state indirect contact probability between two mobile users and inter-visit time of each mobile user to a stationary AP out of its associated APs. We employ these steady-state empirical information into our relay selection algorithm as presented in Sec. III.

We evaluate network performance of our proposed scheme in terms of routing cost and packet transmission time. We define routing cost as totally incurred hop count from a mobile source to a mobile destination considering multicast hop count from a mobile node to multiple selected stationary caches and also single hop count from one of the caches to a mobile node. Also, we measure packet transmission time over entire paths of mobile-to-stationary and stationary-to-mobile relaying and check whether the incurred packet transmission time satisfies the given packet deadline constraint. We observe how these performance metrics are affected by varying the constraints of delivery threshold δ and packet deadline T_{max}.

To compare our algorithm in terms of routing efficiency with a baseline counterpart, we devise a *Naive* algorithm based on [2]. In this *Naive* algorithm, a mobile node first finds the most similar mobile node as the next mobile relay in terms of steady-state visit probability to stationary nodes so as to increase the contact chance (motivated by [2]). Then, it sequentially chooses a set of stationary caches with the highest visit probabilities, and the next mobile node is able to receive data from one of the caches. For fair comparison, we make the sum of the visit probabilities for the selected stationary caches larger than or equal to the delivery threshold δ as done in our algorithm. This procedure continues until the data is finally delivered to the destination user.



(a) Routing cost with respect to delivery threshold δ (where T_{max} = 8000 sec).



(b) The number of selected relays in terms of stationary cache and mobile relay (where δ = 0.8, T_{max} = 8000 sec for the most dissimilar src-dst pair).

Fig. 2. Network performance of routing cost and the number of selected relays compared to a $Naive$ algorithm.

First, we evaluate network performance of routing cost and the number of selected relays compared to *Naive* in Fig. 2. We measure routing cost with respect to delivery threshold δ for three different source-destination pairs. By quantifying the L^2 norm distance between visit probability vectors of two mobile users (referring to [1]), we choose three sets of the most similar, median, and the most dissimilar mobile user pairs. As in Fig. 2(a), in case of the most similar source-to-destination user pair, both our algorithm and *Naive* consumes the same routing cost over the range of delivery threshold from 0.2 to 0.8. As a source user and a destination user becomes

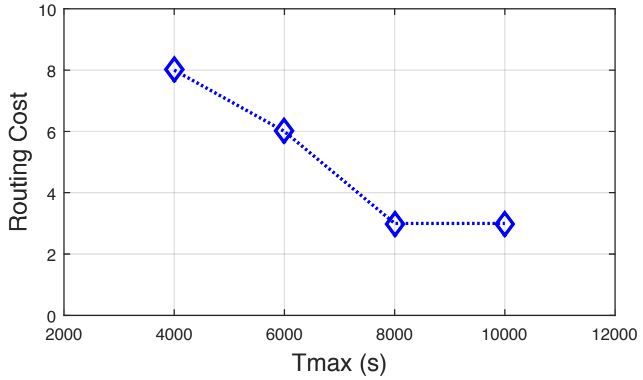


Fig. 3. Routing cost with respect to packet deadline T_{max} (where δ = 0.8 for the most similar src-dst pair).

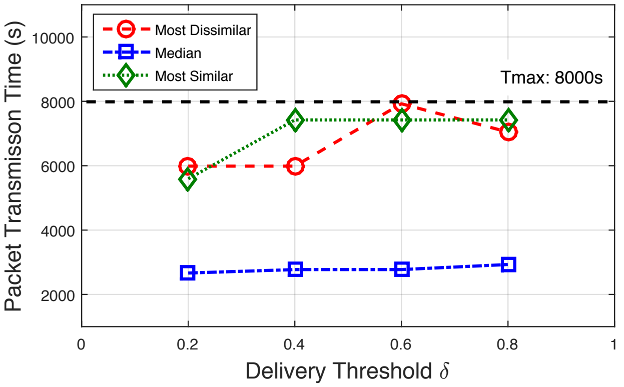


Fig. 4. Incurred packet transmission time with respect to delivery threshold δ for a given packet deadline of T_{max} = 8000.

dissimilar, our algorithm starts outperforming *Naive* with a factor of up to 5. As delivery threshold δ increases, routing cost of our algorithm slightly increases, whereas *Naive* consumes routing cost more steeply. This implies that our algorithm finds effective mobile relays in the level of mobile user relationship through stationary caches as an effective bridge to the next mobile relay, significantly reducing the end-to-end routing cost for data forwarding.

Regarding the number of selected relays, both our algorithm and *Naive* exploit a relatively small number of stationary caches by selecting few crucial caches with high visit probability that can satisfy the given delivery threshold. On the other hand, our algorithm uses much fewer mobile relays up to the destination user as opposed to *Naive*, with a factor of around 3. This demonstrates that our algorithm successfully selects fewer stationary caches and mobile relays that can keep subsequent contacts stable up to the destination user.

We investigate how a given packet deadline affects our algorithm in terms of routing cost in Fig. 3. As packet deadline T_{max} increases, our algorithm finds a more efficient path consisting of mobile-to-stationary and stationary-to-mobile relays. This is due to the fact that given a more relaxed time constraint, the degree of optimization in routing efficiency increases. This shows an interesting trade-off relationship between delay and routing efficiency.

We measure packet transmission time from a source user to a destination user to check whether our algorithm meets a given packet deadline constraint in Fig. 4. Irrespective of the similarity between the source user and the destination user, our algorithm successfully delivers data to the destination within a given deadline.

Lastly, we quantify computation complexity in terms of algorithm running time. As varying the number of mobile users used in the experiments, the numbers of variables and constraints vary, and accordingly affect the computation time. We measure the running time for solving our formulated MILP program on Intel(R) Core(TM) i7-4790 CPU with 3.60 GHz with MATLAB `intlinprog` solver as in Table I. As the number of mobile nodes used as relays increases, the required computation time increases up to several minutes on real-time. In practice, a central server can pre-compute the optimal forwarding path for each pair of mobile source and mobile destination for various packet deadline, and record them in a table. When a mobile source needs to send data to a specific mobile destination, it can obtain the optimal forwarding path upon request by referring to the table, improving practical feasibility.

Mobile nodes	Variables	Constraints	Running time(s)
5	2798	5572	0.23
10	11786	23418	1.83
15	25630	50876	70.87

TABLE I
COMPUTATION COMPLEXITY FOR SOLVING OUR PROPOSED FORWARDING ALGORITHM.

V. CONCLUSION

We have presented a novel data forwarding scheme by extending the traditional data forwarding problem relying on only mobile contacts into a relay selection problem with both mobile and stationary relays. Our approach exploits any scattered network devices as temporal data cache for delivering to the next mobile user to reduce routing cost, while satisfying packet delivery rate and packet delay constraints.

We have formulated the extended relay selection problem into a mixed-integer linear program. We have obtained the optimal forwarding strategy of mobile-to-stationary and stationary-to-mobile relay paths at the same time. Our work has been validated with real-world dataset and has outperformed a baseline counterpart in routing efficiency, achieving QoS constraints.

For future work, we would devise a distributed data forwarding that can allow to adjust forwarding path depending on network status by lowering computation complexity. Also, by considering both direct contact events for mobile-to-mobile and indirect contact events for mobile-to-stationary-to-mobile in the relay selection problem, we may find a more efficient forwarding strategy depending on different QoS requirements.

ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2015R1D1A1A01057902), and by the Ministry of Science, ICT, and Future Planning(NRF-2013R1A1A1009854, NRF-2013R1A2A2A04014772).

REFERENCES

[1] H. Cho, Y. Park, and H. Lee. Routing relevant data to group mobile users by mining social trajectory pattern. *The Journal of Korean Institute of Communications and Information Sciences*, 38(11):934–936, 2013.

[2] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pages 32–40. ACM, 2007.

[3] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[4] H. Gao and H. Liu. Data analysis on location-based social networks. In *Mobile social networking*, pages 165–194. Springer, 2014.

[5] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa. Cooperative caching for efficient data access in disruption tolerant networks. *IEEE Transactions on Mobile Computing*, 13(3):611–625, 2014.

[6] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, pages 299–308. ACM, 2009.

[7] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11):1576–1589, 2011.

[8] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo. Crawdad trace dartmouth campus syslog (v. 2004-12-18). *CRAWDAD wireless network data archive*, Dec 2004.

[9] Z. Li, C. Wang, S. Yang, C. Jiang, and X. Li. Lass: Local-activity and social-similarity based data forwarding in mobile social networks. *IEEE Transactions on Parallel and Distributed Systems*, 26(1):174–184, 2015.

[10] M. J. Pitkänen and J. Ott. Redundancy and distributed caching in mobile DTNs. In *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*, page 8. ACM, 2007.

[11] K. Wei, X. Liang, and K. Xu. A survey of social-aware routing protocols in delay tolerant networks: Applications, taxonomy and design-related issues. *IEEE Communications Surveys & Tutorials*, 16(1):556–578, 2014.

[12] L. Yin and G. Cao. Supporting cooperative caching in ad hoc networks. *IEEE Transactions on Mobile Computing*, 5(1):77–89, 2006.

[13] Z. Zhang. Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges. *IEEE Communications Surveys & Tutorials*, 1(8):24–37, 2006.