# A Community-Based Data Dissemination Scheme in Opportunistic Networks

Chun-Chih Lo, I-Hsun Chuang, Pey-Wen Lai, Bing-Jie Guo, and Yau-Hwang Kuo

***Abstract*—In opportunistic networks, data are often delivered by dissemination-based routing approaches using the store-carry-and-forward concept. However, the heavy overheads incurred by these approaches lead to inefficient data delivery. Generally, users in an area usually demand similar data, called "locality of demand", and this area is referred to as a community. Based on such property, this paper presents a community-based data dissemination (ComD) scheme to improve the data delivery efficiency in the opportunistic network. By delivering data to appropriate communities, users can obtain what they are actually interested in. Moreover, the ComD also uses the cooperative caching technique to reduce transmission redundancy. To optimize system utility, the proposed delivery model is formulated as a multiple knapsack problem to determine which data should be carried, and a greedy-based method is adopted to reduce the computational complexity. Simulation results show ComD can significantly improves the system utility and delivery rate under various situations.**

***Index Terms*—Opportunistic networks, community-based data dissemination, cooperative cache.**

## I. Introduction

Opportunistic networks are emerging network models evolved from Mobile Ad-Hoc Network (MANET). The communication in this type of network is intermittent and complete routes between sources and destinations rarely exist. The communication in this network is highly affected by human mobility [1], the link performance depends extensively on the characteristic of the mobility present in the network. Therefore, designing an efficient dissemination method is considered as one of the most compelling challenges in this type of network.

Several existing forwarding strategies [1] exploit human mobility and opportunistic contacts to deliver data via store-carry-and-forward principle. In the dissemination-based strategies, such as epidemic routing protocol [2] and Spray and Wait [3] scheme, data copies are diffused all over the network. As the data copies increases in the network, the better delay performance tends to be achieved. Although these methods tend to limit delivery delay, but suffer from heavy traffic overheads. In context-based strategies [4], contexts such as mobility patterns are exploited to determine suitable relay nodes. These approaches alleviate the data duplication problem but usually take many efforts to gather and analyze contexts.

Since it's difficult to establish steady connections in opportunistic network, the publish/subscribe [5], [6] service model appear to be more applicable in this challenging environment than client-server model. However, even by adopting the publish/subscribe model, existing data dissemination methods tends to be very costly in terms of resource usage to deliver data to a single user who has subscribed to receive specific data. The lack of considering the demands of other users results in inefficient data delivery.

Usually, users in the same area might have similar demands for specific information, and this property is called Locality of Demand (LoD) as in [7]. For example, people in a shopping district often tend to request shopping information such as which shop and items are on sale, and thus delivering this sales-related information to this district may increase the availability of information for those who are interested. In the paper, an area whose users have similar demands is called a community. Applying the LoD property, this paper presents a community-based data dissemination scheme (ComD) to deliver valuable data to the community which is interested in them. By this means, the potential benefit of collaborating within a community, the limited bandwidth can be used to satisfy more users that have similar demands. Therefore, the data value needs to be estimated to find valuable data. In many cases, users tend to access the latest and popular data, the data freshness and data popularity should be considered in the data value. The main objective of this paper is to maximize the system utility which indicates the utility of all data. However, the data utility that a data object can bring is related to both its data value and access times. The more requests users make for a data object, the higher data utility the data object can bring. Thus, the requirements of each community should be analyzed to estimate which data are more likely to be queried.

Since users might move from one community to another, they can help to carry various data in their cache space and deliver to all the passing communities. Therefore, this paper proposed an efficient cache management to decide which data should be carried during user's trip and which data should be dropped. To optimize the system utility, the delivery model is formulated and further mapped into a multiple knapsack problem (MKP). To reduce the computation overhead, the MKP is solved by a greedy-based algorithm. Moreover, the

Manuscript received March 10, 2015; revised September 21, 2015. The authors would like to thank the Ministry of Science and Technology for supporting this research, which is part of the project numbered 101-2221-E-006-259-MY3 and 101-2218-E-006-015-MY3.

Chun-Chih Lo, I-Hsun Chuang, Pey-Wen Lai, Bing-Jie Guo, and Yau-Hwang Kuo are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan (e-mail: cobrageo@cad.csie.ncku.edu.tw, jacky@cad.csie.ncku.edu.tw, pwlai0802@cad.csie.ncku.edu.tw, leewon@cad.csie.ncku.edu.tw, kuoyh@cad.csie.ncku.edu.tw).

Yau-Hwang Kuo is with the Department of Computer Science, National Chengchi University, Taipei, Taiwan.

doi: 10.18178/ijfcc.2015.4.5.406

proposed cache management also supports cooperation between nodes to alleviate the data duplication problem. That is, users who have similar moving paths can carry data cooperatively to avoid redundancy. The simulation results show that the ComD works well in system utility provision, cache efficiency and delivery rate.

The rest of the paper is organized as follows. In the Section II, existing data dissemination schemes in opportunistic networks are reviewed. Then, details of ComD are described in the Section III; simulation results are presented in the Section IV. Finally, the paper is concluded in the Section V.

## II. Related Work

In opportunistic networks, forwarding data efficiently is one of the most challenged issues. Traditional routing methods are inapplicable for opportunistic networks due to the absence of network topology knowledge. Dissemination-based routing methods work well by exploiting node contacts to deliver data [1]. However, they diffuse data into the whole network which incurs heavy traffic overheads. Thus, context-based routing strategies are proposed.

### A. Context-Based Routing Strategy

The context-based routing strategy takes node contexts into account for selecting relay nodes. Community-based schemes, such as ContentPlace (CP) [8] that aims at making data available in those interested areas, belong to this strategy. CP learns contexts of both nodes and their social communities, and then data utility can be derived based on the community requirements. It improves data availability and minimize delay ratio for a connection request but only optimizes the utility of a single destination rather than the whole system. In addition, the performance of CP relies on the prediction of social behavior. The lack of cooperation between nodes also leads to data redundancy. In CDTS [9], the community is composed of nodes which often contact with each other, and the node with higher leaving probability will be chosen as the relay node. However, the actual demands of nodes are not considered which may result in blindly carrying data that are not needed by others. Also, the cache management in CDTS only considers the Time-to-live (TTL) of data which is inadequate. Some other community-based schemes such as [9], [10] also exploit social behavior to estimate attributes of nodes. However, the more accurate estimations are needed, the more effort has to be made. Contrarily, ComD exploits the information of moving paths provided by nodes which is more accurate and easily-accessible than estimating node behavior by the social relation.

### B. Cooperative Caching Techniques

Cache policy is another attractive issue in opportunistic networks since the cache space of mobile device is limited. GroupCaching (GC) [11] allows mobile nodes to form a group with their one-hop neighbors. Within a group, group members exchange their cache status periodically, so it could diminish data redundancy and improve cache efficiency. However, GC only considers the Time-to-live (TTL) parameter of data which is insufficient for real network environments.

## III. Community-Based Data Dissemination

### A. System Model

As depicted in the Fig. 1, the network is composed of several communities. In a community, users equipped with mobile devices, called nodes, are moving around and communicating with each other. In addition, a node might leave its home community, passing through some intermediate communities, and then arrive at its destination community. In this paper, various data are classified into different channels and stored in the cache space of mobile devices. Nodes will deliver these data to not only the destination but also all passing communities during its trip time. According to the LoD, nodes in a community might subscribe to similar channels. Therefore, delivering data of these channels to the community may increase the availability of information and obtain better system utility. Consequently, the main objective of this paper is delivering appropriate data to the communities which are interested in them. Moreover, in a community, several nodes might have similar destination and passing communities. These nodes can corroborate to carry data for reducing data redundancy.

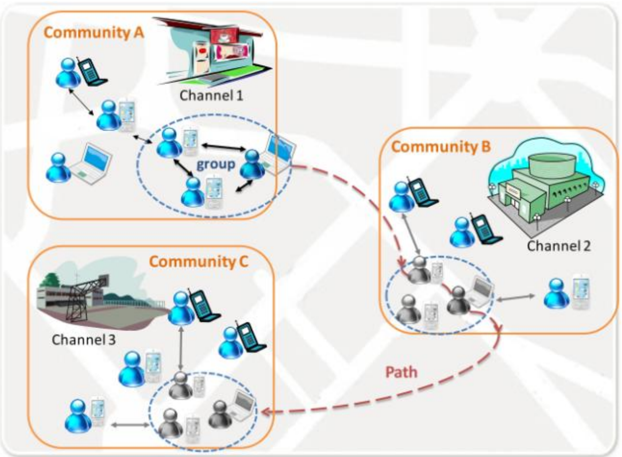


Fig. 1. Scenario of community-based data dissemination.

### B. Definitions and Assumptions

#### 1) Channel

As in Podcasting [12], data are classified into various channels according to their characteristics. For example, music channel, movie channel, shopping channel and so on.

#### 2) Mobile node

Each node will subscribe to some channels and demand the data belonging to these channels. Cache space of mobile device is divided into private and public cache and cache space is limited. The private cache is used to store node's own data, and the public cache is used store and carries data to share with other nodes. For simplicity, only the public cache will be discussed. Furthermore, nodes have capability to decide their destinations and passing communities by predefining their moving path [13]. By exchanging moving information, nodes can cooperate and carry data for other nodes that have the same interests.

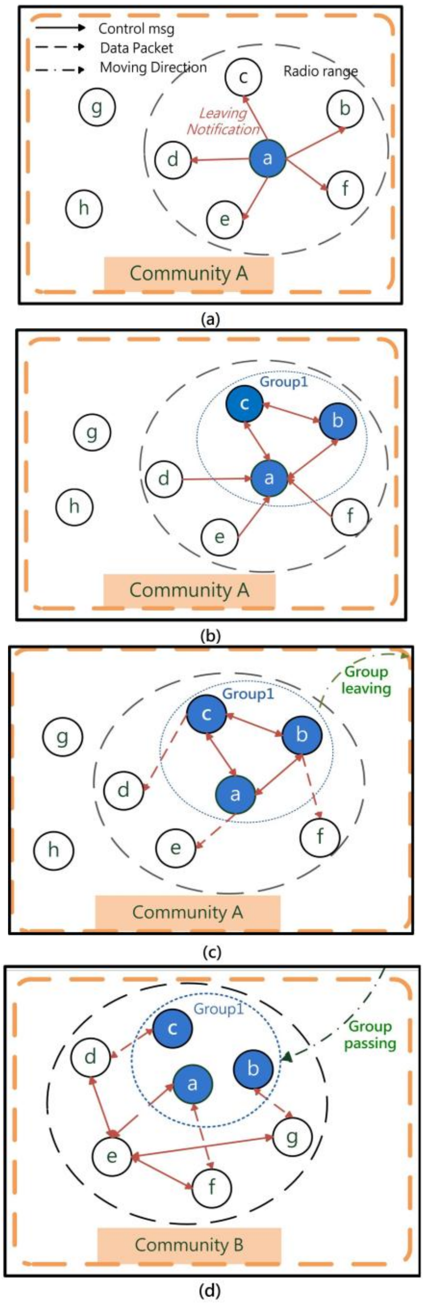


Fig. 2. System flow community.

Based on the LoD property, a community is defined as an area whose users have similar demanded channels. Nodes in a community are more likely to query the data belonging to the demanded channels. These demanded channels might be related to the local characteristics and seldom change. For example, the shopping channel is a demanded channel of shopping district, and thus delivering sales information to this district may increase the availability of information for those who are interest in it and attract more users to query for such information. Therefore, the value of LoD indicates the ratio of data access times of all channels. Generally, the value of LoD can be decided according to the actual data requirements which are obtained by analyzing the data requests in the community.

*3) Data*

In the paper, data value is decided by two factors, Data Popularity and Data Freshness.

**Data Popularity**: Data Popularity indicates how popular the data are. The number of data access, called data access times, is the most significant index of Data Popularity. Thus, nodes will record data access times when data are requested and used the data access times to represent the Data Popularity.

**Data Freshness:** Data Freshness is the other factor should be considered because the latest data are more attractive. As we believe the freshness of data decreases with their life time by a nonlinear decreasing rate, we apply an exponential decay function of data life time to represent Data Freshness. Therefore, the data value is defined as follows:

$$V_k = P_k \times F\left(age(k)\right) = P_k \times e^{-\lambda \cdot age(k)} \qquad (1)$$

where $V_k$ indicates the data value of $k$-th data. $P_k$ is the value of access times which represents the Data Popularity. $F(x)$ is an exponential decay function of data life time which represents the non-linear decreasing Data Freshness, and $age(k)$ indicates the data life time. Furthermore, $\lambda$ is used to control how fast the factor decays with data life time.

*C. Community-Based Data Dissemination Design*

ComD is proposed to decide which data should be carried during the trip of a node and which data should be discarded due to the limited cache space of a node. The ComD is composed of three parts, Temporal Group Formation, Cooperative Community-based data Dissemination Procedure (CCD Procedure), and Cooperative Community-based Cache Management (CCCM).

*1) Temporal group formation*

Since nodes in a community might have similar destination and passing communities during their trip, these nodes can cooperate to deliver data. First, when a node Na would like to leave its home community, the destination and passing communities during its trip is notified to its neighbors as depicted in Fig. 2(a). If a temporal group with similar destination and passing communities exists, Na can join this group and become the group leader. Otherwise, Na and its neighbors with similar destination and passing communities can form a new group whereas Na is the group leader as in Fig. 2(b). Else, Na will leave alone.

If a node has been involved in a temporal group, it will not form or participate in other temporal group before that node is arrived at its destination. Also, because group members might leave the home community separately, no new node can join this temporal group after any group member has left the home community.

*2) CCD procedure*

After a temporal group is formed, Na queries all nodes in the community about their cache statuses, including cached data, data channel, data size, data access times and data lifetime. Since the group leader can manage public cache of its group members, Na further employs the CCCM which is described in the following section to derive an optimal cache schedule. Because the main objective of the paper is to optimize the system utility, the group leader should consider all passing communities during its trip when employing CCCM. Then, Na assigns its group members to carry expected data based on the derived cache schedule as in the Fig. 2(c).

Because once nodes are leaving a community, the cache space will decrease and so will the system utility. To sustain high system utility, Na also has to employ the CCCM for the remaining nodes in the home community to decide which data should be kept, and then nodes can start to exchange their data. After exchanging expected data, nodes in the temporal group can leave the home community at any time.

Since nodes in the temporal group might leave their home community separately, the arrival time at passing communities might be different. When nodes of the temporal group arrived at a passing community, a node in the passing community will employ CCCM to decide which data should be downloaded from these arriving nodes as in Fig. 2(d).

*3) CCCM*

The CCCM is applied to derive the optimal cache schedule for deciding which data will be carried. As communities might demand different channels, the data utility of a data object might be different in diverse communities. The data utility for a community is defined as follows:

$$U_{A,k} = V_k \times LoD_{A,k} \tag{2}$$

where $U_{A,k}$ denotes the data utility of the $k$-th data in the community $A$, and $V_k$ is the data value of $k$-th data. Moreover, if the channel of the $k$-th data is $CH_k$, $LoD_{A,k}$ means the LoD value of $CH_k$ in the community $A$. The demanded channels of communities are assumed well-known and barely changed. Then, the total utility can be defined as follows:

$$\sum_{\forall i}\sum_{\forall A}\sum_{\forall k} U_{A,k} \times I_{i,k} \qquad A \in Comm, k \in DS, I_{i,k} \in \{0,1\} \tag{3}$$

where $DS$ is the available data set and $I_{i,k}$ indicates whether the node $i$ carry the $k$-th data. *Comm* is a community set with diverse meanings in different conditions. When the temporal group leaving, *Comm* involves all passing communities it will go through. Else, when CCCM is performed for the nodes in the home community or passing communities, *Comm* indicates the community itself. Then, (3) can be simplified as (4).

$$\sum_{\forall i}\sum_{\forall k} C_k \times I_{i,k} \qquad where\ C_k = \sum_{\forall A} U_{A,k} \tag{4}$$

Thus, the optimal cache schedule can be derived as follows:

$$\max \sum_{\forall i}\sum_{\forall k} C_k \times I_{i,k} \qquad i = 1,...,m\ and\ k = 1,...,n$$
$$s.t. \sum_{k=1}^{n} I_{i,k} \times S_k \le W_i$$
$$\sum_{i=1}^{m} I_{i,k} \le 1 \tag{5}$$

In (5), $m$ indicates the number of nodes; $n$ is the number of data. For a temporal group, the group leader considers $m$ as the number of group members; for a community, $m$ indicates the number of nodes in the community. $S_k$ indicates the size of the $k$-th data; $W_i$ is the size of public cache of node $i$.

These formulas can be mapped into the Multiple Knapsack Problem (MKP) [14]. In the opportunistic network, it is important to solve the MKP problem efficiently. Thus, we adopt a greedy-based solution introduced in [15] to obtain an approximate solution rather than using heuristic algorithms.

## IV. Simulation

### A. Simulation Setting

In this section, the ComD are compared with epidemic, CP [8], and GC [11]. Since the cache space is limited, without loss of generality, the epidemic scheme tends to deliver the data with higher data value. As in the Table I, there are 4 channels and 6 communities with total 100 nodes in the system. For all communities, favorite channels are randomly assigned. For every communities, the values of LoD of all channels are assumed as 7/12, 3/12, 1/12, and 1/12. Initially, 1000 distinct data generated by nodes are distributed over all communities and classified into channels. The initial data lifetime is randomly assigned and increases over time. Moreover, Zipf's distribution [16] with parameter 1 is exploited to simulate the data access times. Also, nodes decide their destinations and passing communities randomly. For simplicity, system randomly selects one node in a random community as group leader to leave its community every 3 seconds and nodes in the same community with similar destination and passing communities will leave together. Nodes in a community will request data periodically and the data with higher data utility have higher probability to be requested.

In the simulation, gaining utility, delivery rate and cache efficiency are used to evaluate the system performance. The gaining utility indicates the data utility obtained by data delivery. The deliver rate indicates whether the requested data are successfully delivered is given by the number of received data divided by the total number of data requests. The cache efficiency indicates whether the data stored in the cache have been requested, and thus it is defined as the number of data which have been requested divided by the number of stored data.

TABLE I: Simulation Parameter

| | |
|---|---|
| **Mobile Nodes** | 100 |
| **Number of Community** | 6 |
| **Number of Channel** | 4 |
| **Number of Data** | 1000 |
| **Data Size** | 10~50 MB randomly |
| **Public Cache Space Size** | 300 MB |
| **Max Data Lifetime** | 200 sec |

### B. Simulation Results

In the simulations, we will analyze the results of all metrics under various situations including different numbers of group members and different numbers of passing communities in a trip. By default, the number of nodes in a temporal group is 12; the number of passing communities is 4.

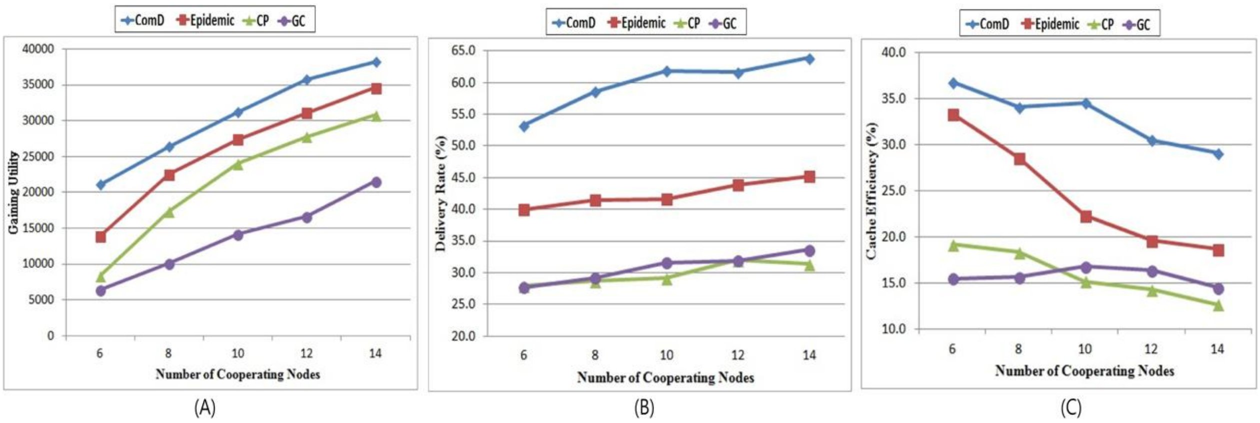


Fig. 3. Simulation results under various numbers of group members.

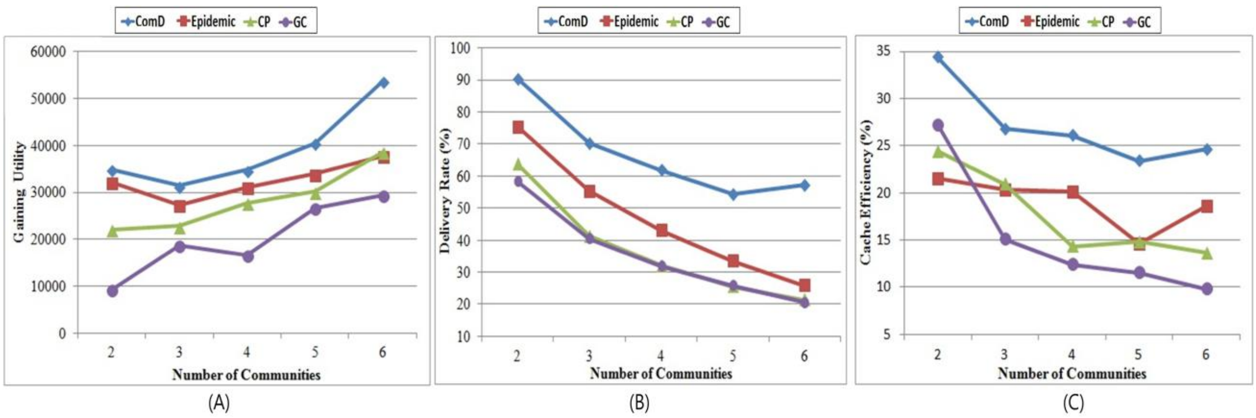


Fig. 4. Simulation results under various numbers of passing communities.

As we can discover in the Fig. 3 and Fig. 4, the ComD has the best results in gaining utility, delivery rate and cache efficiency, even when the number of group members and passing communities are increased. Furthermore, in the Fig. 3(c), when the number of group members is increasing, the cache efficiency is degraded. That is because all schemes start to carry low-utility data when the cache size is large enough. Since nodes are more likely to request high-utility data, these low-utility data might not be requested which degrades the cache efficiency. However, even if the cache efficiency decreases, ComD obtained better results than other schemes. As in Fig. 4(a), when the number of passing communities increases, ComD always has higher gaining utility than other schemes. That is because ComD considers the utility of the whole system Rather than the destination only, nodes will carry data for all passing communities if it can obtain higher utility. However, as more passing communities are in a trip, the number of data that nodes can carry for each community will be decreased, and thus the delivery rate and cache efficiency also decrease as in the Fig. 4(b) and Fig. 4(c). But, ComD still performed well in these situations.

In addition, the average utility per cache space is also described in the Table II.

TABLE II: AVERAGE UTILITY PER CACHE (14 COOPERATING NODES)

| **Scheme** | ComD | CP | Epidemic | GC |
|---|---|---|---|---|
| **Average Utility (Utility/MB)** | 20.07 | 10.87 | 12.66 | 6.09 |

The average utility per cache space of ComD is better than other methods, which indicates that ComD makes better use of the cache space. That is because ComD exploits the community information to decide which data will be stored in the cache. Thus, the limited cache space can be used to carry the high-utility data.

## V. CONCLUSIONS

In this paper, a community-based data dissemination scheme is proposed to efficiently deliver data in opportunistic networks. By exploiting the locality of demand property, data can be disseminated to appropriate areas. Moreover, the use of cooperative cache technique also alleviates the data duplication problem. Simulation results show that ComD can obtain better system utility, delivery rate and cache efficiency than existing data dissemination schemes.

## REFERENCES

[1] L. Pelusi, A. Passarella, and M. Conti, “Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks,” *IEEE Communications Magazine*, vol. 44, pp. 134-141, Nov. 2006.

[2] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” *Tech. Rep. CS-2000-06*, Department of Computer Science, Duke University, Durham, NC, 2000.

[3] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: An Efficient routing scheme for intermittently connected mobile networks,” in *Proc. the ACM SIGCOMM Workshop on Delay-Tolerant Networking*, Aug. 2005.

[4] E. Talipov, Y. Chon, and H. Cha, “Content sharing over smartphone-based delay-tolerant networks,” *IEEE Trans. on Mobile Computing*, vol. 12, pp. 581-595, Mar. 2013.

[5] S. Suman and S. Henning, "BBS-ONE: Bulletin board and forum system for mobile opportunistic networks," *Wireless Communications, Networking and Information Security (WCNIS)*, pp. 370-375, 2010.

[6] S. M. Allena, M. J. Chorleya, G. B. Colomboa, E. Jaho, M. Karaliopoulos, I. Stavrakakis, and R. M. Whitaker, "Exploiting user interest similarity and social links for micro-blog forwarding in mobile opportunistic networks," *Pervasive and Mobile Computing*, vol. 11, pp. 106-131, 2014.

[7] P. Nicopolitidis, G. I. Papadimitriou, and A. S. Pomportsis, "Exploiting locality of demand to improve the performance of wireless data broadcasting," *IEEE Transactions on Vehicular Technology*, vol. 55, pp. 1347-1361, July 2006.

[8] B. Chiara, C. Marco, and P. Andrea, "Content place: Social-aware data dissemination in opportunistic networks," in *Proc. ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM)*, Oct. 2008, pp. 203-210.

[9] N. Jianwei, Z. Xing, W. Kongqiao, and Ma. Jian, "A data transmission scheme for community-based opportunistic networks," *International* in *Proc. Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, Sep. 24-26, 2009, pp. 1-5.

[10] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, "A socio-aware overlay for publish/subscribe communication in delay tolerant networks," in *Proc. ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM)*, Oct. 2007, pp. 225-234.

[11] Y.-W. Ting and Y.-K. Chang, "A novel cooperative caching scheme for wireless ad hoc networks: Group caching," *IEEE International Conference on Networking, Architecture, and Storage (NAS)*, pp. 62-68, July 29-31, 2007

[12] V. Lenders, G. Karlsson, and M. May, "Wireless ad hoc podcasting," *IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pp. 273-283, June 18-21, 2007.

[13] P. C. Cheng, K. C. Lee, M. Gerla, and J. Härri, "GeoDTN+Nav: Geographic DTN routing with navigator prediction for urban vehicular environments," *Mobile Networks and Applications*, vol. 15, pp. 61-82, Feb. 2010.

[14] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, New York: Springer-Verlag, 2004.

[15] S. Martello and P. Toth, "Heuristic algorithms for the multiple knapsack problem," *Computing*, vol. 27, pp. 93-112, June 1980.

[16] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *Proc. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, vol. 1, Mar. 1999, pp. 126-134.

**Chun-Chih Lo** received his BSc. in computer science (2004) from University of Johannesburg, South Africa and a MSc. in computer science (2007) from Tunghai University, Taiwan. He is currently working toward the PhD degree in computer science and information engineering at National Cheng Kung University.

His research interests revolve around wireless communication, with particular attention to wireless networking problems presented by vehicular networking.

**I-Hsun Chuang** received B.S. degree in applied mathematics from National Chiao Tung University in 2004, and received M.S. and Ph. D degree in computer science and information engineering from National Cheng Kung University in 2006 and 2014, respectively.

His research interests include cognitive radio, network security, wireless networks and opportunistic networks.

**Pey-Wen Lai** received B.S. degree in computer science and information engineering from National Chung-Cheng University in 2010, and received M.S. degree in computer science and information engineering from National Cheng Kung University in 2012.

Her research interests include wireless networks and opportunistic networks.

**Bing-Jie Guo** received B.S. degree in computer science and information engineering from National Cheng Kung University in 2014. He is currently working toward M.S degree in computer science and information engineering from National Cheng Kung University. His research interests include wireless networking and network security.

**Yau-Hwang Kuo** received Ph.D. degree in computer engineering from National Cheng Kung University (NCKU), Tainan, Taiwan in 1988.

He is currently the dean, College of Science, and a distinguished professor, Department of Computer Science, National Chengchi University. He is also a distinguished professor, Department of Computer Science and Information Engineering at NCKU, and a co-director on ICT-enabled service area of national networked communication program in Taiwan. In his career, he is persistently active in the fields of academia, education accreditation, and government policy planning. He has served as the deputy executive secretary of Science & Technology Advisory Group in the Executive Yuan (Cabinet), the director of Computer Center in the Ministry of Education, the director of Computer Science and Information Engineering Program in National Science Council (NSC), and the director of Engineering & Technology Promotion Center (ETPC), NSC. From 1999 to 2000, he was elected as a president of Taiwanese Artificial Intelligence Association. He has also served as an editor for several international journals, and consulted for several research institutes and hi-tech companies.

Prof. Kuo has published more than 300 papers and 42 patents. His research topics include wireless broadband communication, cloud computing, computational intelligence, intelligent information analytics, and context-aware computing. He is a member of the IEEE.