IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 15, 2019, accepted March 29, 2019, date of publication April 11, 2019, date of current version April 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910238

Reducing the Overhead of Multicast Using Social Features in Mobile Opportunistic Networks

XIA DENG[1], (Member, IEEE), LE CHANG[2], (Member, IEEE), JUN TAO[3], (Member, IEEE), AND JIANPING PAN[4], (Senior Member, IEEE)

[1]School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China
[2]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[3]Key Laboratory of CNII, MOE, School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China
[4]Department of Computer Science, University of Victoria, Victoria, BC V8P 5C2, Canada

Corresponding author: Le Chang (lechang@gdut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702127, in part by the Science and Technology Program of Guangzhou under Grant 201804010461, in part by the Hundred Young Talents Plan Project of Guangdong University of Technology under Grant 220413618, in part by the Natural Sciences and Engineering Research Council of Canada, in part by the Canada Foundation for Innovation, and in part by the British Columbia Knowledge Development Fund.

ABSTRACT Through utilizing user mobility and short-range device-to-device communication techniques, mobile opportunistic networks (MONs) enable end-to-end message delivery without the dependence on reliable network infrastructures. Examples of these networks include mobile social networks and vehicular networks. Multicast in MONs is used to disseminate information to a group of mobile nodes, which have attracted considerable attention due to high resource utilization. In this paper, we focus on reducing the overhead of multicast in MONs without compromising the delivery performance, through utilizing static social features of nodes and time-varying social behaviors. We first conduct a trace data analysis using the information entropy theory to identify the most important and representative social features in a popular trace, the Infocom06 trace. Based on these static social features, we propose a **S**ocial **P**rofile-based **M**ulticast (SPM) routing scheme, that supports efficient multicast message dissemination with a small maintenance overhead, i.e., little cost on maintaining the historical records. Furthermore, by exploring the time-varying social behaviors during daytime and nighttime, we verify that a small number of forwardings during that daytime is sufficient to achieve a desirable delivery ratio. We thus propose an improved overhead-reducing scheme **S**ocial **P**rofile-based **M**ulticast-**O**verhead **R**educing (SPMOR) that restricts the number of forwardings during the daytime. The extensive trace-driven simulations show that SPM achieves desirable delivery performance with small maintenance and transmission cost, and SPMOR can further reduce the transmission overhead under diurnal user behaviors. At last, we conduct a similar study on a campus-based MON trace, SocialBlueConn. We find that the main conclusions and performance results on SocialBlueConn are consistent with the Infocom06 trace, which verifies our methodology is scalable.

INDEX TERMS Mobile opportunistic networks (MONs), multicast, social behaviors, overhead reduction.

I. INTRODUCTION

Mobile opportunistic networks, as an emerging networking paradigm without the dependence on reliable network infrastructures (e.g., 4/5G), have proved to be useful in disaster recovery, opportunistic social interaction, on-site advertising, news and weather forecasts, traffic alerts, etc [1]–[3]. These networks exhibit specific characteristics, e.g., intermittent link connections, frequent node movement, large transmission delay, limited node buffer space and power, and severe network security vulnerabilities [4]–[10].

Multicast in mobile opportunistic networks is widely used for group communication in a variety of scenarios. For instance, people sharing the same interests at conference venues or shopping malls may form a group and exchange messages of common interests. To support such kind of applications, like unicast, data forwarding in multicast routing also adopts the ''store-carry-forward'' mode. The basic process of multicast routing is illustrated in Fig. 1. A dashed circle represents that the nodes inside are within the transmission range

The associate editor coordinating the review of this manuscript and approving it for publication was Burak Kantarci.

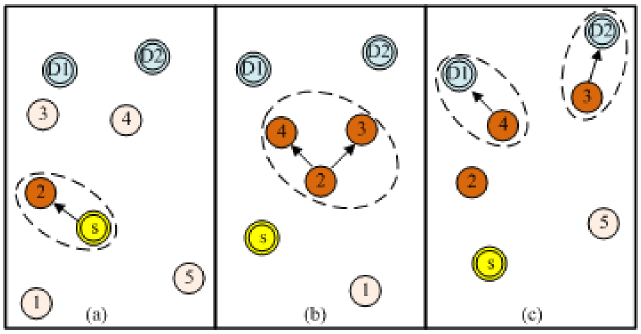


FIGURE 1. The data forwarding process of multicast routing.

of each other, referred to as a "contact" or "encounter". An arrow means a message is transmitted from one node to another, referred to as a "forwarding". In the figure, the source, node S, wants to send data to multicast group D, which contains two multicast members, D1 and D2. S forwards data to D1 by first encountering node 2 and then through node 4, which are called "relay nodes". Similarly, data can be forwarded to D2 through relay node 2 and 3. It is well-accepted that such multicast routing can improve resource utilization in mobile opportunistic networks, which has attracted great concern and research efforts in recent years [11]–[17].

Multicast members in the same multicast group often have a close relationship between each other, which can be **dynamic**, such as frequent contact records; or **static**, e.g., similar social features such as from the same organizations, speaking the same language, or sharing common interests [18]–[20]. In node trajectory or contact history-based approaches [21]–[23], it is usually expensive to maintain the contact records at each individual node as nodes move frequently. In contrast, static social features or profiles are easy to maintain and retrieve, and can be utilized to efficiently identify desirable relay nodes that deliver messages to destinations [18], [24]–[27]. Such static social information can be obtained and refined from user profiles when they register for an event, such as conference, disaster recovery, online social network platforms (e.g., Facebook and Twitter). This can be implemented by asking the users to simply tap a mobile device and write down the social profile, which is a common practice for mobile social network systems nowadays.

The efficiency of routing schemes based on social profile/features has been validated in unicast scenarios [18], [24]. In this paper, we focus on multicast and study how static social features help reduce the overhead while still achieving good delivery performance. The *overhead* we aim to minimize in this paper includes: (1) the *maintenance overhead*, which refers to the cost of maintaining necessary information for the routing schemes, e.g., social features for social profile-based routing; (2) the *transmission overhead*, i.e., the redundant forwardings in the entire network. We first conduct a trace-based study using the information entropy theory, to identify the most important and representative social features in popular MON traces, e.g., the Infocom06 trace. Then we verify that such social features in fact indicate the contact probability of nodes, through refining the trace data and calculating the number of node contacts associated with these features. Based on such a property, we design a multicast routing scheme, SPM, to achieve efficient delivery performance with small maintenance and transmission overhead. Next, we study the diurnal behaviors in the trace, i.e., the different contact frequency between the daytime and nighttime, and propose an improved scheme, SPMOR, by restricting the number of forwardings during the daytime. Our contributions are highlighted as follows.

1) Through the analysis of the Infocom06 trace data [28] using the information entropy and mutual information index, we determine the most important and representative social features, i.e., *affiliation* and *language*. These features are consistent with two typical multicast scenarios, i.e., affiliation-based and language-based multicast. We validate based on the trace data that nodes with similar affiliation or language labels are more likely to contact each other. We then apply this method to a campus-based trace, SocialBlueConn [29], and the results suggest that our method is applicable.
2) We develop a social profile-based multicast routing scheme, SPM, utilizing these social features, which indicates a small maintenance overhead, e.g., only two features in the Infocom06 trace. We select relay nodes which have a small average affiliation distance or large common language ratio to multicast destinations. Unlike the dynamic contact history information, individual nodes can obtain and maintain the static social features easily.
3) We further investigate the contact behaviors of nodes during the daytime and nighttime in the traces. We find most of the contacts happen during the daytime, providing more choices for a node to select relays. Aiming at further reducing the transmission overhead, we thus propose an overhead-reducing multicast routing scheme, SPMOR, which restricts the number of forwardings during the daytime.
4) The opportunistic network simulation environment (ONE) [30] is extended to support multicast, and trace-driven simulations are conducted to evaluate the proposed schemes using the Infocom06 and SocialBlueConn trace data. The simulation results show that both SPM and SPMOR achieve near-optimal performance in terms of data delivery ratio and delay with low transmission overhead and energy consumption at the same time, which outperform other best-known multicast schemes. Especially, SPMOR can further reduce the transmission overhead and energy consumption compared with SPM.

In the early conference version of this paper, we studied the social features in the Infocom06 trace and proposed SPM: a social-profile based multicast scheme utilizing social features [13]. Compared to the conference version, this paper proposes a new multicast scheme SPMOR to further reduce the transmission overhead and energy consumption by considering the diurnal behaviors of nodes. The trace analysis method and routing scheme design (SPM and SPMOR) are also extended to a new trace, SocialBlueConn, to verify their extensiblity. Besides, more simulation scenarios are included such as varying buffer sizes, which demonstrate the comprehensive performance of the proposed approaches. The energy efficiency is also investigated and discussed.

The remainder of this paper is organized as follows. Related work is discussed in Section II. The trace study is introduced in Section III. We describe the proposed social profile-based multicast routing scheme (SPM) in Section IV. The improved overhead-reducing multicast scheme SPMOR is described in Section V. The trace-driven performance evaluation is presented with analysis in Section VI. At last, we conclude the paper and discuss the future work in Section VII.

II. RELATED WORK

A. TRADITIONAL APPROACHES

Zhao *et al.* analyzed the semantic models of mobile opportunistic networks and classified existing multicast approaches into four categories based on how they are implemented: unicast-based, tree-based, broadcast-based [31], and group-based approaches [32], [33]. They concluded that group-based multicast strategies perform the best. Furthermore, through exploring the inherent mobile characteristics, such as connectivity and contact probability, probability-based multicast approaches were proposed, e.g., EBMR [22] and CAMR [23] . Wang *et al.* considered the active level of nodes, the probability to multicast destinations, and the contact state information, and built a multicast non-replication tree with a small amount of relay nodes [21]. Wang *et al.* set the probability to reach the destination as the main multicast routing metric, and further compared three multicast strategies, i.e., single-copy, multi-copy and delegation forwarding. The results showed that delegation forwarding outperformed the other two strategies [34]. Lee *et al.* proposed RelayCast, which adopted a two-hop relay for DTN multicast and analyzed the throughput bounds [35]. Yang *et al.* further proposed a general cooperative multicast strategy using the two-hop relay in mobile ad hoc networks [17]. By utilizing the Markov chain model to analyze the packet delivery process, the expected packet delivery probability and cost were derived, and then the multicast strategy was validated to achieve low transmission overhead and high data delivery ratio.

B. SOCIAL-BASED APPROACHES

In mobile opportunistic networks, nodes move frequently and the link connection is intermittent, which brings a great challenge for data transmission. However, there exists social characteristics, e.g., nodes prefer to contact more frequently with those sharing similar interests or visit certain communities, have regular social behaviors. These social features are stable and have been utilized to assist data transmission in MONs [36], [37].

Social-based approaches have also been studied for multicast. Qin *et al.* investigated the effect of social relationship and group size on multicast in ad hoc networks [38]. A two-layer model including the social and network layer was proposed. Then a multicast routing strategy was proposed based upon the Euclidean minimum spanning tree with the probability density function (PDF) of the destination positions. The results showed a better scalability than traditional approaches. A community-based multicast routing scheme [39] was proposed by Gao *et al.*. They formulated the selection of the relay nodes as a unified knapsack problem, and chose the nodes with higher contact centrality in priority. Animesh *et al.* further proposed an energy-aware relay selection multicast scheme based on SDM [39] that considered both social centrality and the residual energy of nodes [40]. The scheme supported more multicast sessions and a longer lifetime than SDM. Chen *et al.* considered the proportion of the connections involving certain social feature values, and improved data relaying performance by utilizing the community structure in a compare-split scheme [26]. Zhang *et al.* developed a hierarchical multicast routing in mobile opportunistic networks by adopting the backbone from a social network perspective [41]. The network was described in a weighted contact graph which reflected the node importance and contact tightness. Then a mobile backbone was constructed for multicast to reduce the overhead. Galluccio *et al.* proposed an adaptive infection recovery scheme which considered social relationship among users for multicast routing [42]. The delivery reliability was also investigated through utilizing the network coding technique.

Especially, the social profile of mobile nodes usually contains static and light-weighted information that not only infers the user mobility pattern, but also can be easily maintained, e.g., *affiliation* and *language*. This makes social profile-based routing schemes demonstrate desirable delivery performance at a substantially lower maintenance cost, which has been verified in unicast routing [18], [24], [27]. In the conference version of this paper [13], we designed an efficient multicast routing scheme (SPM) utilizing static social features with small maintenance and transmission overhead.

Time-varying user behaviors have also been studied recently. A time-varying community structure based on human mobile networks was introduced in [43]. Gao *et al.* utilized the social transient contact patterns to improve the data forwarding efficiency under a short-time condition [44]. Zhou further predicted the future transient social contact patterns through considering the temporal closeness and centrality [45]. Then the TCCB data forwarding strategy was proposed to reduce the delivery cost. Our work in this paper further studies the contact diurnal behaviors and proposes SPMOR, which restricts the number of forwardings when plenty of contacts happen during daytime. Such a conservative forwarding setting can greatly reduce

TABLE 1. The entropy of the social features in the Infocom06 trace.

Social Feature	Affiliation	City	Nationality	Country	Language	Position
Entropy	4.60	4.50	4.11	3.60	3.02	1.53

the transmission overhead without compromising the delivery performance of SPM.

III. SOCIAL FEATURES: A TRACE-BASED STUDY

In this section, we describe our trace analysis method which is used to determine the most important and representative social features in the Infocom06 trace. We use the Shannon entropy, mutual information index, and correlation analysis between social features and node contact behaviors.

A. THE INFOCOM06 TRACE

The Infocom06 trace [28] we used in this study is a popular trace gathered from an experiment at the Infocom 2006 conference, in which attendees carried Bluetooth devices, iMote, with a transmission range of 30 m. The trace data contains both the social profile information of the attendees, and the contact records. The social profile information includes social features, i.e., *affiliation*, *language*, *country*, *city*, *position*, etc. Each feature is labeled by numbers, i.e., IDs. For example, affiliation represents the organization or institution an attendee belongs to, labeled from 1 to 35 in integer. Attendees with the same affiliation label come from the same organization. The node mobility and interaction in the trace spans 4 days with a granularity of 120 seconds. In this study we focus on the nodes providing social profiles, which results in 63 nodes and 39,270 relevant contact records.

B. FINDING IMPORTANT SOCIAL FEATURES

In order to find the most important social features from the trace, we extend the approach of the feature analysis in [18]. Our method contains two steps: (1) identify the most important features according to their Shannon entropy; (2) determine a minimum set of mutually independent features through calculating the mutual information indexes. Consider a network consisting of N nodes, the features of each node are denoted by F_j and $j = 1, \cdots, m$, i.e., m features for a node. The entropy of feature F_j is calculated using $E(F_j) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$, where $x_1, x_2, \cdots, x_n$ are all possible values of feature F_j, and $P(x_i)$ is the probability that x_i is chosen for feature F_j [46]. When a feature has a larger entropy, it means that the feature contains more information and thus is more important. We use the method to calculate the entropy for all social features in the Infocom06 trace. For nodes with multi-label features, our method only considers the most common label. For example, some nodes are labeled with multiple languages, and we choose the one that is most common among all nodes, except English. Table 1 demonstrates the entropy values of the six most useful features in the trace, which show similar results with [18]. By calculating the entropy, we quantify the importance of the social features. In conclusion, affiliation is the most important, and position is the least.

C. FINDING REPRESENTATIVE SOCIAL FEATURES

The social features may be inter-dependent. For example, a specific affiliation may indicate a specific city. Thus some of the important features listed in Table 1 are redundant and can be simply neglected. In order to determine a minimum set of representative features, the inter-dependence between social features is analyzed using mutual information indexes [46], defined as

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)},$$

where X and Y are two features, and $P(x)$ and $P(y)$ are the marginal probability of feature X and Y, respectively. The joint probability of X and Y is $P(x,y) = P(y|x)P(x) = P(x|y)P(y)$. If the mutual information index value of two features is larger, their inter-dependence is greater.

TABLE 2. The mutual information indexes in the Infocom06 trace.

Features	Affiliation	City	Nationality	Country	Language
City	4.24				
Nationality	3.42	3.34			
Country	3.50	3.57	2.80		
Language	2.68	2.57	2.48	2.28	
Position	0.98	1.01	0.70	0.62	0.47

The numerical results of the mutual information indexes between the six features are listed in Table 2. We find that city, nationality and country demonstrate a strong dependence on affiliation. This is consistent with the practice in real world. An organization usually locates in a specific city of a specific country, and the employees usually hold the nationality of that country. In contrast, as many organizations often have a multi-culture work environment nowadays, the language feature is relatively independent. Other than English, the employees with different cultural backgrounds speak a variety of different languages. Although position is also independent according to the table, its entropy is small, which implies little information.

In summary, we claim that **affiliation** and **language** are the most important and representative social features in the trace.

D. CONTACTS BETWEEN NODES WITH SIMILAR AFFILIATIONS

We now verify the impact of the affiliation feature on user behaviors. In the trace, the affiliation feature is labeled with an affiliation ID. We retrieve the contact probability on different affiliation distance, i.e., the absolute value of the arithmetic difference between two affiliation IDs. This distance measures how two organizations are "close" to each other, e.g., in terms of geography. A small distance means the two organizations are "close" to each other, and specifically, distance 0 means the same affiliation. It has been verified that individuals with similar social features in the Infocom06 trace tend to contact each other more often [27]. Moreover, the collectors of the Infocom06 trace data revealed that affiliation and country are mainly labeled by a number smaller than 13 (the maximum is 35) when they own the language label 5 and 10 (French and Spanish) [24]. We thus infer that the label values of the affiliation follow an approximate geographical order, i.e., Europe first and then other continents. To inspect how such affiliation distance impacts the contact probability, we extract the number of contacts in the trace and plot the contact probability for any affiliation distance in Fig. 2, which is the proportion of the contacts of any two nodes with that affiliation distance in the trace. Surprisingly, the relationship between the affiliation distance and contact percentage shows an approximate linear correlation, which indicates that nodes with smaller affiliation distances are more likely to contact each other. This verifies our inference that organizations in the same or neighboring countries are labeled with close IDs.

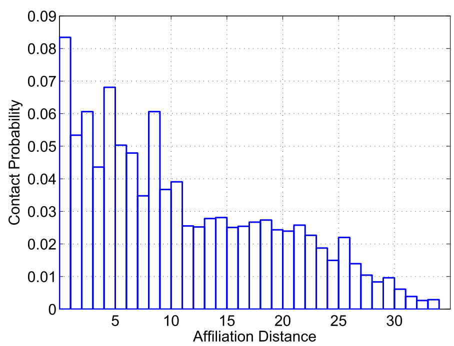


FIGURE 2. The contact probability with different affiliation distance.

E. CONTACTS BETWEEN NODES WITH COMMON LANGUAGES

We also study the role that the language feature plays in the human interaction in the trace. We exclude English in the analysis as English is the official language used by everyone during the conference. Existing work has verified that the inter-contact time of attendees follows a power-law distribution in the Infocom06 trace [24], [47]. The larger the power coefficient is, attendees will meet each other more frequently. In this study, we further analyze the inter-contact time between attendees with and without common languages, excluding English. Figure 3 shows the CCDF of the inter-contact time, where the curve with common languages has a larger power law coefficient than that without a common language. This validates that attendees with common languages will contact each other more frequently. For example, in the trace, attendees from Europe with language 5 will communicate more frequently with other people with the same label.

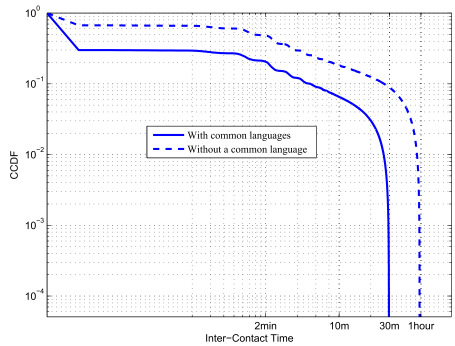


FIGURE 3. The CCDF of the inter-contact time with/out common languages.

F. INSIGHTS INTO THE INFOCOM06 TRACE DATA

Now we give a summary of the insights that we have revealed from the Infocom06 trace data. The most important and representative social features in the trace are affiliation and language. People tend to contact more frequently with those belonging to the same or close organizations. Moreover, attendees also prefer to connect with those having common languages. The insights revealed from the trace are consistent with two typical multicast applications nowadays: (1) the *affiliation-based* multicast, where people with similar affiliations form a multicast group; (2) the *language-based* multicast, in which a multicast group consists of people sharing common languages (other than English). In the rest of this paper, we focus on these two kinds of multicast applications.

IV. SPM: SOCIAL PROFILE-BASED MULTICAST ROUTING

This section first presents in detail the two typical multicast scenarios for the Infocom06 trace, i.e., the *affiliation* and *language*-based multicast. Then a social profile-based multicast scheme (SPM) is proposed, which can reduce both the maintenance and transmission overhead by utilizing these two static social features.

A. AFFILIATION AND LANGUAGE-BASED MULTICAST

1) AFFILIATION-BASED MULTICAST

Existing work has studied affiliation-based multicast, in which nodes with the same affiliation label belong to the same multicast group [24]. As observed in Section III-D, people from close organizations also communicate with each other frequently. We thus partition these nodes into multicast groups according to their affiliation label. There are many partition methods available, and we use the `K-means` clustering algorithm [48] mainly due to its universality and simplicity.

The `K-means` algorithm is to find k clusters such that the intra-cluster distance between the nodes and their cluster centers is minimized. Given N nodes with affiliation label set $\{x_1, x_2, \cdots, x_N\}$, we use `K-means` to classify the N nodes

TABLE 3. The clusters and affiliation labels.

Cluster	1	2	3	4	5	6	7	8
Affiliation Labels	31,32,33, 34,35	4,5, 6,7	1,2, 3	12,13,14, 15,16	8,9, 10,11	22,23, 24,25	17,18,19, 20,21	26,27,28, 29,30

into k ($k \ll N$) groups $S_1, S_2, \cdots, S_k$, with the minimum sum of the squared intra-cluster affiliation distances. Thus the problem can be formulated as $\min \sum_{i=1}^{k} \sum_{x_j \in S_i} ||x_j - \mu_i||^2$, where $\mu_i = \frac{1}{N_i} \sum_{x_j \in S_i} x_j$ is the average value of the node affiliation ID in S_i, and N_i is the number of nodes in S_i.

Next, we determine the number of groups. We utilize the R-square metric, R_s, which can be calculated as $R_s = R_{BS}/R_{TS}$, where $R_{BS} = \sum_{i=1}^{k} N_i||\mu_i - \overline{x}||^2$ is the sum of the squared inter-cluster affiliation distances, $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, and $R_{TS} = \sum_{i=1}^{N} ||x_i - \overline{x}||^2$ is the sum of the squared total cluster affiliation distances. If R_s is closer to 1, it means the nodes in the same cluster have a greater similarity. In Fig. 4, we plot the R_s value with varying numbers of clusters in the trace. It is observed that R_s is close to 0.98 when the cluster number $k = 8$, and R_s only increases slightly afterwards. We thus set the number of groups to $k = 8$, and the clustering results for the affiliation-based scenarios are shown in Table 3.

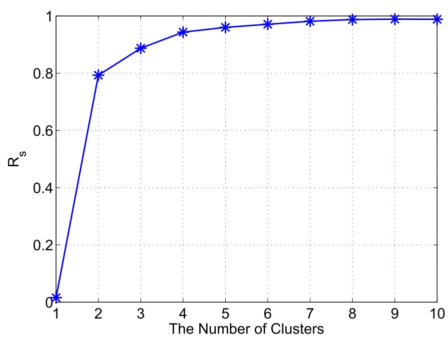


FIGURE 4. R_s with different cluster numbers.

We assume that the cluster ID of a node is fixed and known to it once the partitioning process is finished. This can be achieved through centralized administration, e.g., with the assistance of the staff at the registration desk. After the cluster is settled, a node may store its cluster ID to its local buffer, and the cost to carry such ID is negligible. In real applications, there can be some nodes with high social centrality that carry and disseminate the global clustering information to nodes.

2) LANGUAGE-BASED MULTICAST

We find that 26 out of 63 attendees have a common language labeled 5. These nodes communicated with each other frequently, and contributed to a large portion of the contacts in the trace. Therefore, we assume a multicast service for these attendees, and some multicast messages are sent to them, i.e., 26 attendees in one multicast group with language label 5.

B. THE PRINCIPLE AND SCHEME DESIGN OF SPM

Based on the insights from Fig. 2, nodes with smaller affiliation distances will contact each other more frequently. Thus the nodes with smaller affiliation distances to the multicast destinations are more likely to relay data to them, which should be chosen as the relay nodes. Moreover, according to Fig. 3, nodes sharing common languages (except English) also contact each other more frequently. Therefore, we can simply select a relay node if it shares a large number of common languages with the multicast destinations, excluding English.

Algorithm 1 SPM: Social Profile-Based Multicast Routing

Require: The receivers D, Msg m, node X meets Y
Ensure: A_H, M_H, whether X forwards m to Y
if $(A(X, m) > A(Y, m)) \wedge (A(Y, m) < A_H)$ **then**
  forward Msg m to Y
  update $A_H \leftarrow A(Y, m)$
else if $(M(X, m) < M(Y, m)) \wedge (M(Y, m) > M_H)$ **then**
  forward Msg m to Y
  update $M_H \leftarrow M(Y, m)$
end if

Different from unicast, multicast destinations consist of a group of nodes. Given N_r multicast members in a multicast group $\mathcal{D} = \{D_1, D_2, \cdots, D_{N_r}\}$, any message m from a source node S will be sent to all other nodes in $\mathcal{D}$. According to the trace, multicast members connect with each other closely, as they are from close affiliations. We can adopt the average values of these metrics, e.g., the average affiliation distance to the members in a group. Assuming that node X carrying message m meets node Y, the routing scheme needs to decide whether Y should be the next relay, i.e., whether X should forward message m to Y. The algorithm running on X is described in Alg. 1, and the definitions of these metrics are described as follows.

(1) The *average affiliation distance* for a message m of X is $A(X, m) = \frac{1}{N_r} \sum_{i=1}^{N_r} d(X, D_i)$, where $d(X, D_i)$ is the affiliation distance (the absolute value of the arithmetic difference between two affiliation IDs) between node X and D_i. The smallest value of $A(X, m)$ encountered so far is denoted as A_H.

(2) The *common language ratio* for a message m of X is $M(X, m) = \frac{1}{N_r} \sum_{i=1}^{N_r} L(X, D_i)$, where $L(X, D_i) = 0$ if there is no common language between node X and D_i, and $L(X, D_i) = 1$ otherwise. The largest value of $M(X, m)$ encountered so far is denoted as M_H.

In SPM, the **average affiliation distance** and **common language ratio** are selected as the routing metrics, with the relationship as `or`. When Y satisfies any one of the criteria compared with X, the forwarding process will be triggered. Moreover, the delegation forwarding strategy is adopted to reduce the transmission overhead [34], where the encountering nodes whose routing metrics are the best so far will be selected as relays.

V. SPMOR: SOCIAL PROFILE-BASED MULTICAST ROUTING-OVERHEAD REDUCING

In this section, we will explore the time-varying social behaviors in the trace, i.e., different social patterns during the daytime and nighttime. Taking this into consideration, we propose SPMOR, an improved social-profile based multicast algorithm further reducing the transmission overhead and the energy consumption, through restricting the number of forwardings during the busy contact periods, i.e., the daytime.

A. TIME-VARYING SOCIAL BEHAVIORS

The human social behaviors change over time, which has been verified in [44]. We thus further study the time-varying social behaviors in the Infocom06 trace. The trace data were collected from April 23rd to April 27th. The Bluetooth devices were distributed to volunteers between 7:00 PM and 9:00 PM on April 23rd and collected back on April 26th and April 27th during the daytime. All the Bluetooth devices were set to start working at 5:01 PM on April 23rd. Once two nodes met, a contact occurred. Each contact has its start time and end time, and we treat the start time as the time of the occurrence of a contact. We define the contact percentage during a time period as the number of contacts happened during the period over the total number of contacts of all periods.

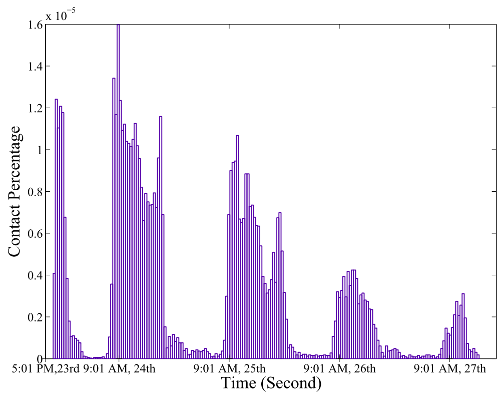


FIGURE 5. The contact percentage throughout the conference.

Figure 5 shows the contact percentage during all the time periods of the experiment of Infocom06. We observe that the contact percentage is much higher during the daytime than nighttime. This is because the opening session, keynote speeches and technical sessions were arranged from 9:00 AM to 6:00 PM. People attended these sessions and contacted each other at a high frequency. However, during the nighttime, if no activities are arranged, attendees are most likely to stay in separate small groups, resulting in a lower contact percentage.

B. THE PRINCIPLE AND SCHEME DESIGN OF SPMOR

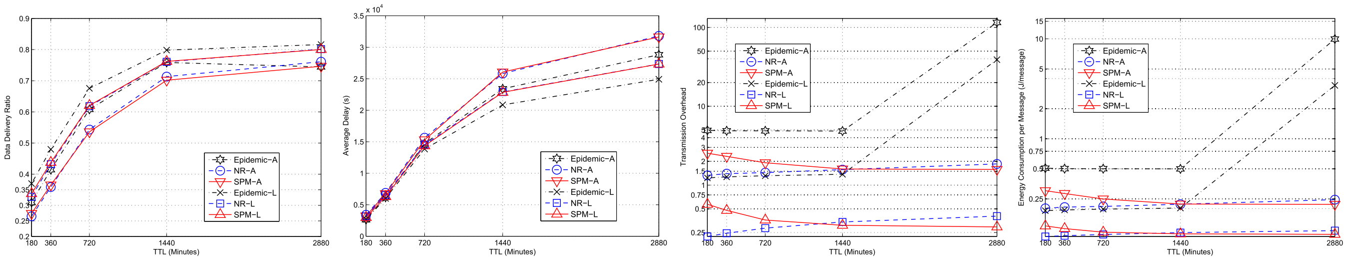
Based on the observations in Fig. 5 that most contacts occur during the daytime, we propose a simple yet efficient algorithm, SPMOR, to further reduce the transmission overhead and the energy consumption, without compromising the delivery performance. Compared with SPM, in SPMOR each node needs to store and maintain an extra metric, the number of forwardings, i.e., the number of messages that have been relayed successfully by the node. It is worth noting that maintaining this metric is quite easy at individual nodes, and incurs little maintenance cost.

Consider that people contact each other more frequently during the daytime, there will be better flexibility for each node, i.e., more nodes as relay candidates. Thus, a node may behave more conservatively in sending out messages in order to reduce the transmission overhead and the energy consumption. SPMOR simply compares the number of forwardings of two contacting nodes, and invokes forwarding only if the number of forwardings of the relay candidate is smaller than the message holder. The details of SPMOR are described in Alg. 2, and we briefly describe its main procedure as follows, assuming Node X needs to decide whether to transmit message m to node Y.

Algorithm 2 SPMOR: Overhead-Reducing Social Profile-Based Multicast Routing

Require: The receivers D, Msg m, node X meets Y
Ensure: A_H, M_H, whether X forwards m to Y
if current time is in [9:00 a.m., 6:00 p.m.] **then**
  if number of forwardings of Y < number of forwardings of X **then**
    RestrictFlag = **false**;
  else
    RestrictFlag = **true**;
  end if
else
  RestrictFlag = **false**;
end if
if *RestrictFlag* == **false then**
  if $A(X, m) > A(Y, m)$ **and** $A(Y, m) < A_H$ **then**
    forward Msg m to node Y
    update $A_H \leftarrow A(Y, m)$
  else if $M(X, m) < M(Y, m)$ **and** $M(Y, m) > M_H$ **then**
    forward Msg m to node Y
    update $M_H \leftarrow M(Y, m)$
  end if
end if

Step 1: determine whether message m should be forwarded based on the number of forwardings and contact time. During the daytime, more contacts are expected, so the forwarding will be restricted if the number of relayed messages on the candidate, node Y, is greater than the message holder, node X.



(a) (b) (c) (d)

FIGURE 6. The performance of SPM with different TTLs (Infocom06). (a) Data delivery ratio. (b) Average delay. (c) Transmission overhead. (d) Energy consumption per message.

The purpose is to reduce possible transmission overhead and preserve energy for better relays. Such restriction does not apply to contacts during nighttime, as there are few contacts during that time period.

Step 2: if the transmission is not restricted for the concern of the number of forwardings in step 1, make the final decision whether to forward message m to node Y based on the crucial social features: i.e., affiliation and language. Node X always selects the relays that are closer in terms of affiliation or with common languages. This step is the same with SPM.

VI. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our SPM and SPMOR schemes through trace-driven simulations. We first explain the setup of the simulations, and then present and analyze the simulation results on the Infocom06 and SocialBlueConn trace, respectively.

A. SIMULATION SETUP

In the simulations, we first import the contact history from the Infocom06 or SocialBlueConn traces, and reproduce the node movement and contact process using the opportunity network simulation environment (ONE) [30]. Besides our SPM and SPMOR schemes, two existing popular multicast schemes are also implemented in ONE as benchmarks, i.e., the Non-Replication tree (NR) scheme [21] and Epidemic scheme [31], and compared with SPM and SPMOR. In NR, a multicast tree is built according to the contact state information dynamically, in which leaf nodes are the multicast destinations. As a result, the NR multicast scheme needs to store and update the contact information at all nodes. In the Epidemic routing scheme, messages are flooded in the network. For the SocialBlueConn trace, we repeat a similar analysis and scheme design process with that of Infocom06 and conduct simulation-based experiments with similar configurations.

The performance is evaluated in terms of the following metrics.

(1) *Data Delivery Ratio*: the ratio of the number of messages arrived at destinations to the number of messages expected to arrive at destinations.

(2) *Transmission Overhead*: the total number of relayed messages (only successfully forwarded messages are included) minus the number of messages reaching the destinations, divided by the latter.

(3) *Average Delay*: the average time spent by all messages from the source to the destinations.

(4) *Energy Consumption per Message*: the energy consumption on sending/receiving all the messages over the number of successfully delivered messages. Different from transmission overhead, energy consumption includes the energy consumed on aborted transmissions, e.g., transmitting half size of a message and then aborted due to loss of contact. The energy consumption on a single message is calculated using $E(P) = i * u * t_p$, where $t_p = (P_h + P_d)/1\text{Mbps}$ is the time spent on transmitting the message, 1 Mbps is the transmission speed, i is the current value, u is the voltage value, and P_h and P_d are the sizes of the message header and payload, respectively. We refer to a Bluetooth 4.0 chip, CC2541 [49], to set the parameters in $E(P)$. When the transmission speed is 1 Mbps, i is 17.6 mA and 18 mA for receiving and transmitting data, respectively. u is 3 V.

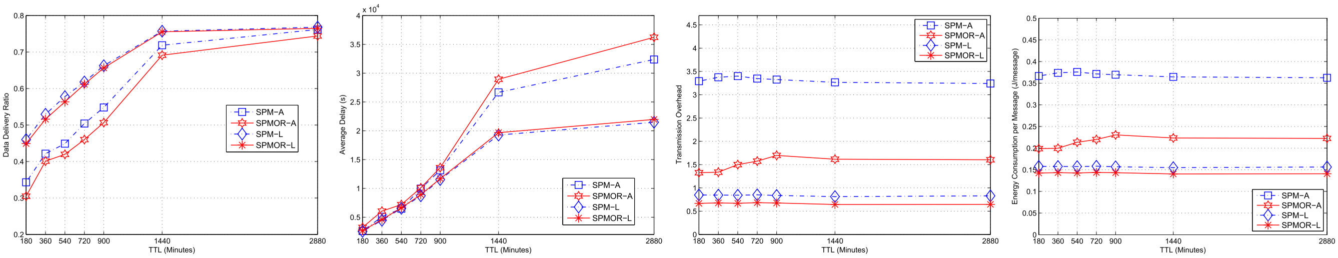
B. RESULTS ON THE INFOCOM06 TRACE

There are two scenarios considered for the Infocom06 trace. In the affiliation-based multicast scenarios, eight communication groups are set following Table 3. In the language-based multicast scenarios, one multicast group is selected containing 26 nodes, with a common language labeled 5. Let "$-A$" and "$-L$" denote the affiliation-based and language-based scenarios, respectively.

1) THE PERFORMANCE OF SPM

We first evaluate the performance of SPM in comparison with the NR and Epidemic schemes with varying TTL values on the Infocom06 trace. The message size is 100 KB, and the message generation interval is 300 seconds. The buffer size of any node is 30, 000 KB. Under these settings, in most cases (TTL < 1, 500 minutes), the number of distinct messages in the system will be less than 300. Considering the size of the buffer (30, 000 KB) and messages (100 KB), the node buffer is thus big enough to accommodate all the messages if TTL is set to one day (1, 440 minutes).

In Fig. 6, we evaluate the performance of the three schemes in terms of data delivery ratio, average delay, transmission overhead, and energy consumption per message with



(a) (b) (c) (d)

FIGURE 7. The performance of SPMOR with different TTLs (Infocom06). (a) Data delivery ratio. (b) Average delay. (c) Transmission overhead. (d) Energy consumption per message.

different TTLs. Figure 6a shows the data delivery ratio, which indicates that our SPM scheme achieves comparable performance with the history-based approach, NR, in either affiliation-based or language-based scenarios. This verifies the efficiency of selecting relay nodes based on static social features such as affiliation and language, which can be easily obtained and maintained compared with the contact history. It is worth-noting that Epidemic can be seen as the optimal strategy concerning delivery ratio and delay if messages are never discarded by nodes, i.e., with enough buffer size. Moreover, it is also observed that the performance of the three schemes under the language-based scenarios outperforms affiliation-based scenarios. The reason is: there is one single language group with 26 members under the language-based scenarios, which is much larger than affiliation-based groups (8 on average). A greater number of multicast members usually indicates a higher probability to share relay nodes, and thus a better delivery performance. The results concerning the average delay is shown in Fig. 6b. It can be observed that the SPM and NR schemes perform only slightly worse than Epidemic, i.e., a near-optimal performance. Note that Epidemic always demonstrates the shortest delay as it often selects the shortest path to destinations, yet with a great transmission overhead.

We then investigate the transmission overhead and energy consumption and plot the results in Fig. 6c and Fig. 6d, respectively. We observe a great reduction of the SPM and NR schemes compared with Epidemic in all scenarios. Specifically, our SPM scheme introduces the lowest cost when TTL is set to 2, 880 minutes, the most challenging case as the buffer is too small to accommodate all the messages. When TTL increases, the transmission overhead of Epidemic increases dramatically. This is because the buffer is insufficient when TTL becomes large and many messages are repeatedly copied and dropped before reaching the destination. The performance results of the energy consumption are consistent with the transmission overhead, in which SPM and NR perform better than Epidemic. This is because in these cases (e.g., message size = 100 KB, transmission rate = 1 Mbps), most of the packet transmissions can be completed during one encounter. Recall that one difference between transmission overhead and energy consumption is whether the aborted transmissions are counted. As a result, when transmission abortion rarely occurs, the energy consumption per message is in proportion to the transmission overhead.

2) THE PERFORMANCE OF SPMOR

We evaluate our SPMOR scheme with different TTLs and buffer sizes in both affiliation-based and language-based scenarios. In the Infocom06 conference, most activities were arranged during daytime when attendees contact each other more frequently than nighttime. To reflect this behavior, we thus set different message generation intervals for the daytime, night and midnight as 1, 000, 4, 000 and 8, 000 seconds, respectively.

We first fix the buffer size at a value that is large enough as 10, 000 KB. The size of a message is 100 KB, and we let TTL vary from 180 to 2, 880 seconds. Fig. 7a shows the delivery ratio of the SPMOR scheme, in comparison with the original SPM. We observe only marginal performance degradation in either affiliation-based or language-based scenarios. Similar phenomena are also observed for the delay performance in Fig. 7b. This demonstrates that although nodes are more conservative in forwarding messages during the daytime, the delivery performance is not compromised. Through utilizing the social features, SPMOR remains effective in the dissemination of messages. Fig. 7c and Fig. 7d demonstrate the transmission overhead and energy consumption of SPM and SPMOR. We observe that the overhead is reduced greatly in SPMOR compared with SPM in all scenarios, which also reflects great energy savings. For affiliation-based scenarios, the overhead is reduced by more than 50%, a quite significant decrease. The overhead reduction of SPMOR in language-based scenarios is slightly smaller than affiliation-based scenarios, but the performance improvement is still around 20%. Such a mass reduction of transmission overhead is due to restricting the number of forwardings during the daytime. There are a lot of activities during the daytime, and nodes move and contact each other frequently, where many forwardings are in fact redundant and do not contribute to improving the delivery ratio. These results verify the efficiency of SPMOR on reducing the transmission overhead and energy consumption of SPM through a conservative forwarding strategy.

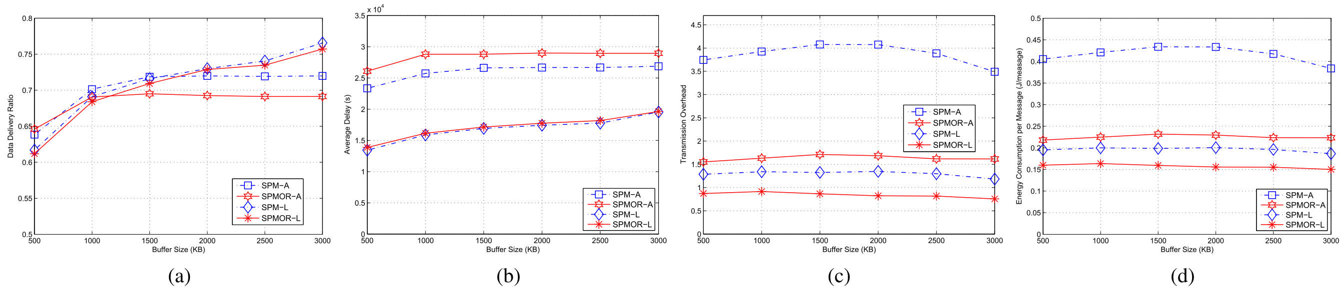


FIGURE 8. **The Performance of SPMOR with different buffer sizes (Infocom06). (a) Data delivery ratio. (b) Average delay. (c) Transmission overhead. (d) Energy consumption per message.**

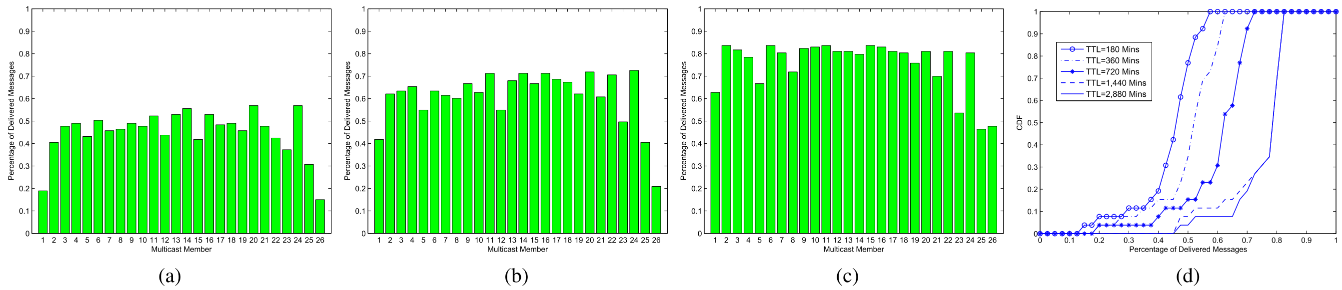


FIGURE 9. **The percentage of delivered messages at multicast members (Infocom06 with SPMOR-L). (a) TTL=180 minutes. (b) TTL=720 minutes. (c) TTL=1,440 minutes. (d) The CDF.**

We then fix TTL to 1, 440 seconds and vary the buffer size from 500 KB to 3, 000 KB. Note that the size of any message is 100 KB. We adopt small buffer sizes (e.g., 500 KB) to test the performance of our routing algorithms in possible cases of MONs where nodes may have limited storage space. Figure 8 shows the performance of SPMOR with different buffer sizes. First, we observe that the data delivery (Fig. 8a) and delay (Fig. 8b) performance remain at the same level with SPM, which is similar to the cases of varying TTL in the previous section. The delivery ratios of both SPM and SPMOR are greater than 0.6, which is quite desirable. Specifically, SPMOR demonstrates nearly the same delivery ratio with SPM with smaller buffer sizes (e.g., $<1,000$ KB), showing that the conservative forwarding policy adapts well to small buffers. Once again, similar to the cases with different TTLs, the transmission overhead has been greatly reduced by SPMOR, with an average decrease of more than 50% under affiliation-based and 35% under language-based scenarios, shown in Fig. 8c. For these cases, we observe no distinct impact of the buffer size on the overhead or energy consumption, shown in Fig. 8c and Fig. 8d. SPMOR can further reduce the energy consumption of SPM greatly. The performance results of the energy consumption are consistent with the transmission overhead in Fig. 7 and Fig. 8, and the reason is similar to the cases in Fig. 6. We also observe a better performance under language-based scenarios than affiliation-based scenarios, due to the bigger group size of the language-based multicast. This bigger group shares more relay nodes, and also makes the performance gap between SPM and SPMOR less distinct under these scenarios.

We further explore the percentage of delivered messages at individual members in the multicast group under the language-based scenarios, which is summarized in Fig. 9. There is one single multicast group containing 26 members in these cases, and these 26 members are the destinations of multicast messages. The percentage of delivered messages at a multicast member is the portion of the multicast messages successfully arrived at this node. From Fig. 9a to 9c, we list the percentage of delivered messages at each individual multicast member labeled from 1 to 26, when TTL=180, 720, and 1440 minutes, respectively. We observe that only a small number of members suffer from relatively lower delivery ratios. Moreover, with a greater TTL, the percentage of delivered messages increases. We thus plot Fig. 9d to verify such improvement with increasing TTL values, which is the CDF of the multicast members with different percentage of delivered messages. For example, when TTL=1440 minutes, the multicast group demonstrates a desirable delivery performance, where more than 80% members have a delivery ratio greater than 70%. The performance improves only marginally when TTL>1440 minutes.

C. RESULTS ON THE SOCIALBLUECONN TRACE

We extend our analysis method and routing design principles to another trace, the SocialBlueConn trace. Similar to our work on the Infocom06 trace, we first analyze the SocialBlueConn trace data and extract important and representative social features, and then define multicast groups based on these features. The corresponding SPM and SPMOR routing algorithms are designed using specific routing metrics

TABLE 4. The entropy of the interests (SocialBlueConn).

Interest	Politics	Literature	Cinema	Sport	Multimedia Entertainment
Entropy	2.1210	1.8874	1.4838	1.4138	1.3788

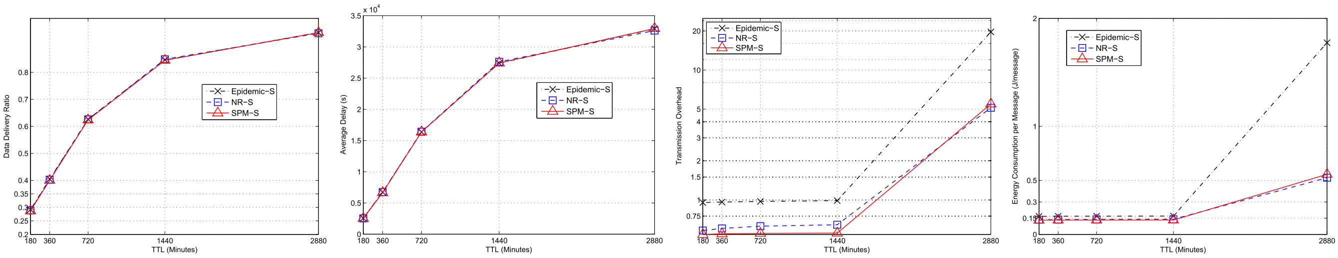


FIGURE 10. The performance of SPM with different TTLs (SocialBlueConn). (a) Data delivery ratio. (b) Average delay. (c) Transmission overhead. (d) Energy consumption per message.

associated with the social features of SocialBlueConn, and then tested using the ONE simulator.

1) THE SOCIALBLUECONN TRACE

The SocialBlueConn trace contains Bluetooth encountering records and social profiles of a set of 15 students through an ad-hoc Android application, SocialBlueConn, at University of Calabria [29]. The trace spans one week from January 28th 2014 to February 5th 2014, excluding non-working days. We refine the data and use 40, 136 contact records between the 15 students. The granularity is 180 seconds, and the transmission range is 10 m. Participants' interests were collected at the beginning of the experiment through a questionnaire, which contained a list of questions considering the participants' preferences in the following 9 categories: mobility, sport, music, cinema, literature, multimedia entertainment, politics, other hobbies and social networks.

TABLE 5. The mutual information indexes (SocialBlueConn).

Features	Politics	Literature	Cinema	Sport
Literature	0.8703			
Cinema	0.8513	0.5188		
Sport	0.2099	0.4838	0.4299	
Multimedia Entertainment	0.4606	0.3060	0.3060	0.1002

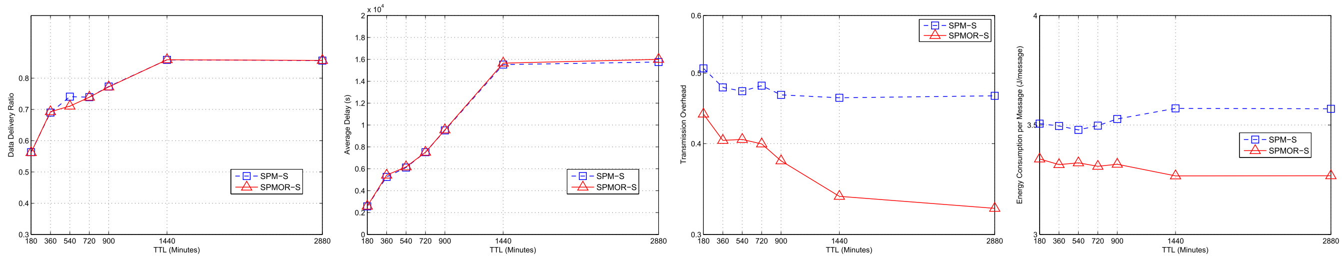
2) ANALYSIS ON THE TRACE DATA

Following the same analysis strategy with the Infocom06 trace, we calculate the participants' preference entropy, and list the five largest ones in Table 4: politics, literature, cinema, sport and multimedia entertainment. We then calculate the mutual entropy, and filter out literature as it is highly dependent on politics as their mutual entropy is large (in Table 5). In contrast, sport is independent of politics. Therefore, **Politics** and **Sport** are determined to be the most important and representative preference features in the SocialBlueConn trace. Concerning the multicast scenarios, we consider interest-based multicast such as sport group and select the nodes with soccer as the sport interest to form one multicast group, denoted by "$-S$". In the corresponding routing algorithm SPM and SPMOR for SocialBlueConn, we select the nodes with more common politics or sport interests with the multicast group as relay nodes. The design is similar to the routing strategies for the Infocom06 trace, so we omit the details here.

3) THE PERFORMANCE OF SPM AND SPMOR

Similar to the experiment settings on the Infocom06 trace, we set the message size to 100 KB and the buffer size to 30, 000 KB to evaluate the SPM scheme, where few transmissions are aborted during data forwarding. The performance of SPM with different TTLs on the SocialBlueConn trace is demonstrated in Fig. 10, where SPM shows nearly the same results with NR in terms of delivery performance and transmission cost, but with less maintenance overhead in most cases. It verifies that SPM achieves desirable performance when using the static social profile strategy.

We also evaluate SPMOR with greater sizes of messages. A greater size indicates it is more likely to abort a transmission within a short contact duration. We assign the size of a message a random value from [1 KB, 5 MB]. To reflect the diurnal human behaviors, we set different message generation intervals for the daytime, night and midnight to 300, 1, 200 and 2, 400 seconds, respectively. The buffer size is 1, 000 MB. Compared with SPM, SPMOR further reduces the transmission overhead and energy consumption without performance degradation in all cases as illustrated in Fig. 11. From Fig. 11c and Fig. 11d, we can see that the energy consumption per message generally follows a similar trend to the transmission overhead, but with small fluctuations, e.g., different patterns when TTL are 720 minutes and 900 minutes. This is due to aborting transmissions before completing the reception of packets. For a greater message size, e.g., in the range of [1 KB, 5 MB], the transmission rate, i.e., 1 Mbps, is insufficient to guarantee the completion of receiving a message within the contact duration in the trace. According to our simulation results, nearly 1/3 transmissions are aborted due to insufficient contact duration. In case of aborted transmissions, the energy consumed on sending and



(a) (b) (c) (d)

FIGURE 11. The performance of SPMOR with different TTLs (SocialBlueConn). (a) Data delivery ratio. (b) Average delay. (c) Transmission overhead. (d) Energy consumption per message.

receiving the transmitted part will be counted in calculating the total energy consumption, while such partial transmissions are ignored in calculating the transmission overhead. Therefore the energy consumption per message is not strictly in proportion to the transmission overhead when the message size is large. It is worth mentioning that the transmission overhead here focuses more on the transmission aspect, i.e., the routing and caching cost during the entire routing process, through calculating the number of replications and caching times among the relay nodes.

In summary, we observe that all the simulation results are consistent with the Infocom06 trace. This verifies the extensiblity of our analysis method and the SPM/SPMOR algorithms on different traces.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we focused on reducing the maintenance and transmission overhead of multicast using social features in mobile opportunistic networks. Our methodology can be summarized as follows. First, we used the information entropy theory to analyze the trace data, e.g., Infocom06 and SocialBlueConn, to identify a small set of important and representative static social features, e.g., affiliation and language in the Infocom06 trace. We then used these features as the routing metrics, e.g., the affiliation distance and common language ratio, to design a social profile-based multicast routing scheme, SPM. Furthermore, we analyzed the time-varying social behaviors during the daytime and nighttime, finding that a great number of contacts happened during the daytime, among which many were in fact redundant. By restricting the number of forwardings during the daytime, we proposed SPMOR, which further reduced the transmission overhead and energy consumption compared with SPM. Through trace-driven simulations, we validated that SPM achieved a near-optimal delivery performance with small overhead, and SPMOR further reduced the overhead in case nodes demonstrate diurnal behaviors. The results on the Infocom06 and SocialBlueConn traces verify that our approach is general and can apply to many scenarios of MONs.

Concerning the future work, we will consider a variety of other multicast applications, such as literature or music-based multicast applications in the SocialBlueConn trace. Many other traces also provide the social profile information and we can extend our SPM and SPMOR schemes to adapt to more environments. Furthermore, we will study cooperative buffer management strategies among multiple nodes for messages with greater sizes, rather than the simple FIFO strategy in our current work. Last but not the least, we will study the energy optimization techniques to further reduce the energy consumption of our approaches.

ACKNOWLEDGMENT

An early version of this paper, titled Social Profile-based Multicast Routing Scheme for Delay-Tolerant Networks, has appeared in the Proceedings of the 2013 IEEE International Conference on Communications (**ICC2013 Best Paper Award**).

REFERENCES

[1] N. Chakchouk, "A survey on opportunistic routing in wireless communication networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2214–2241, Mar. 2015.

[2] Y. Zhao and W. Song, "Survey on social-aware data dissemination over mobile wireless networks," *IEEE Access*, vol. 5, pp. 6049–6059, 2017.

[3] X. Hu, T. H. S. Chu, V. C. M. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. B. Chan, "A survey on mobile social networks: Applications, platforms, system architectures, and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1557–1581, 3rd Quart., 2015.

[4] S. Trifunovic, S. T. Kouyoumdjieva, B. Distl, L. Pajevic, G. Karlsson, and B. Plattner, "A decade of research in opportunistic networks: Challenges, relevance, and future directions," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 168–173, Jan. 2017.

[5] Q. Xu, Z. Su, and K. Yang, "Optimal control theory-based epidemic information spreading scheme for mobile social users with energy constraint," *IEEE Access*, vol. 5, pp. 14107–14118, 2017.

[6] Z. Wang, Z. Ma, S. Luo, and H. Gao, "Enhanced instant message security and privacy protection scheme for mobile social network systems," *IEEE Access*, vol. 6, pp. 13706–13715, 2018.

[7] C. Gao, Q. Cheng, X. Li, and S. Xia, "Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network," in *Cluster Computing*. New York, NY, USA: Springer, 2018, pp. 1–9.

[8] J. Dede *et al.*, "Simulating opportunistic networks: Survey and Future directions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1547–1573, 2nd Quart., 2018.

[9] P. Yuan, L. Fan, P. Liu, and S. Tang, "Recent progress in routing protocols of mobile opportunistic networks: A clear taxonomy, analysis and evaluation," *J. Netw. Comput. Appl.*, vol. 62, pp. 163–170, Feb. 2016.

[10] X. Kui, A. Samanta, X. Zhu, S. Zhang, Y. Li, and H. Pan, "Energy-aware temporal reachability graphs for time-varying mobile opportunistic networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9831–9844, Oct. 2018.

[11] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, pp. 1691–1700, Dec. 2010.

[12] M. Mongiovi, A. K. Singh, X. Yan, B. Zong, and K. Psounis, "Efficient multicasting for delay tolerant networks using graph indexing," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1386–1394.

[13] X. Deng, L. Chang, J. Tao, J. Pan, and J. Wang, "Social profile-based multicast routing scheme for delay-tolerant networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 1857–1861.

[14] J. Ren, G. Zhang, and D. Li, "Multicast capacity for VANETs with directional antenna and delay constraint under random walk mobility model," *IEEE Access*, vol. 5, pp. 3958–3970, 2017.

[15] J. Luo, J. Zhang, L. Yu, and X. Wang, "The role of location popularity in multicast mobile ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2131–2143, Apr. 2015.

[16] J. Hu, L. Yang, and L. Hanzo, "Distributed multistage cooperative-social-multicast-aided content dissemination in random mobile networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 3075–3089, Jul. 2015.

[17] B. Yang, Y. Shen, X. Jiang, and T. Taleb, "Generalized cooperative multicast in mobile ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2631–2643, Mar. 2018.

[18] J. Wu and Y. Wang, "Hypercube-based multipath social feature routing in human contact networks," *IEEE Trans. Comput.*, vol. 63, no. 2, pp. 383–396, Feb. 2014.

[19] 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 Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 556–578, 1st Quart., 2014.

[20] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *IEEE Commun. Surveys Tut.*, vol. 15, no. 1, pp. 387–401, 1st Quart., 2013.

[21] Y. Wang and J. Wu, "A dynamic multicast tree based routing scheme without replication in delay tolerant networks," *J. Parallel Distrib. Comput.*, vol. 72, no. 3, pp. 424–436, 2012.

[22] Y. Xi and M. C. Chuah, "An encounter-based multicast scheme for disruption tolerant networks," *Comput. Commun.*, vol. 32, no. 16, pp. 1742–1756, 2009.

[23] M.-C. Chuah and P. Yang, "Context-aware multicast routing scheme for disruption tolerant networks," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 4, no. 5, pp. 269–281, 2009.

[24] P. Hui and J. Crowcroft, "How small labels create big improvements," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2007, pp. 65–70.

[25] E. Luo, Q. Liu, J. H. Abawajy, and G. Wang, "Privacy-preserving multi-hop profile-matching protocol for proximity mobile social networks," *Future Gener. Comput. Syst.*, vol. 68, pp. 222–233, Mar. 2017.

[26] X. Chen, C. Shang, B. Wong, W. Li, and S. Oh, "Efficient multicast algorithms in opportunistic mobile social networks using community and social features," *Comput. Netw.*, vol. 111, pp. 71–81, Dec. 2016.

[27] A. Mei, G. Morabito, P. Santi, and J. Stefa, "Social-aware stateless routingin pocket switched networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 252–261, Jan. 2015.

[28] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. (2009). *Crawdad Trace Cambridge/Haggle/Imote/Infocom2006 (V.2009-05-29)*. [Online]. Available: https://crawdad.org/cambridge/haggle/imote/infocom2006

[29] C. Antonio, A. Socievole, and F. D. Rango. (2015). *Crawdad Dataset Unical/Socialblueconn (V.2015-02-08)*. [Online]. Available: https://crawdad.org/unical/socialblueconn/20150208

[30] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proc. Int. Conf. Simulation Tools Techn.* (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 55.

[31] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke Univ., Durham, NC, USA, Tech. Rep. CS-200006, 2000.

[32] W. Zhao, M. Ammar, and E. Zegura, "Multicasting in delay tolerant networks: Semantic models and routing algorithms," in *Proc. ACM SIGCOMM Workshop Delay-Tolerant Netw.* New York, NY, USA: ACM, 2005, pp. 268–275.

[33] S. Patra, S. Saha, V. Shah, S. Sengupta, K. G. Singh, and S. Nandi, "A qualitative survey on multicast routing in delay tolerant networks," in *Recent Trends in Wireless and Mobile Networks*. Berlin, Germany: Springer, 2011, pp. 197–206.

[34] Y. Wang, X. Li, and J. Wu, "Multicasting in delay tolerant networks: Delegation forwarding," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[35] U. Lee, S. Y. Oh, K.-W. Lee, and M. Gerla, "Relaycast: Scalable multicast routing in delay tolerant networks," in *Proc. IEEE Int. Conf. Netw. Protocols (ICNP)*, Oct. 2008, pp. 218–227.

[36] Z. Li and C. Wang, "Modeling data transport capacity of mobile networks for mobile social services," *IEEE Access*, vol. 5, pp. 12143–12157, 2017.

[37] Z. Lu, Y. E. Sagduyu, and Y. Shi, "Integrating social links into wireless networks: Modeling, routing, analysis, and evaluation," *IEEE Trans. Mobile Comput.*, vol. 18, no. 1, pp. 111–124, Jan. 2019.

[38] Y. Qin, R. Jia, J. Zhang, W. Wu, and X. Wang, "Impact of social relation and group size in multicast ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 15, no. 7, pp. 1661–1673, Aug. 2015.

[39] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.* New York, NY, USA: ACM, 2009, pp. 299–308.

[40] A. Roy, S. Bose, T. Acharya, and S. DasBit, "Social-based energy-aware multicasting in delay tolerant networks," *J. Netw. Comput. Appl.*, vol. 15, no. 87, pp. 169–184, 2017.

[41] D. Zhang, H. Ma, and D. Zhao, "Social-aware backbone-based multicast routing in mobile opportunistic networks," in *Proc. Int. Conf. Big Data Comput. Commun.*, Aug. 2017, pp. 31–38.

[42] L. Galluccio, B. Lorenzo, and S. Glisic, "Sociality-aided new adaptive infection recovery schemes for multicast DTNs," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3360–3376, May 2016.

[43] S.-Y. Chan, P. Hui, and K. Xu, "Community detection of time-varying mobile social networks," in *Proc. Int. Conf. Complex Sci.* Shanghai, China: Springer, 2009, pp. 1154–1159.

[44] W. Gao, G. Cao, T. L. Porta, and J. Han, "On exploiting transient social contact patterns for data forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 151–165, Jan. 2013.

[45] H. Zhou, V. C. M. Leung, C. Zhu, S. Xu, and J. Fan, "Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10372–10383, Nov. 2017.

[46] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[47] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mobile Comput.*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

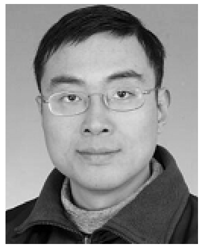
[48] C. Ding and X. He, "K-nearest-neighbor consistency in data clustering: Incorporating local information into global optimization," in *Proc. ACM Symp. Appl. Comput.* New York, NY, USA: ACM, 2004, pp. 584–589.

[49] (2013). *Texas Instruments Incorporated*. [Online]. Available: http://www.ti.com/lit/ds/symlink/cc2541.pdf

XIA DENG received the B.S. degree in electronic information engineering, and the M.S. and Ph.D. degrees in computer application technology from Central South University, Changsha, Hunan, China, in 2003, 2006, and 2015, respectively. In 2006, she joined the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China, where she is currently an Assistant Professor. She has published more than 10 journal and conference papers. Her current research interest includes protocol design and analysis in mobile and social networks. She received the Best Paper Award at IEEE ICC 2013.

LE CHANG received the B.S. and M.S. degrees in computer science from Central South University, Changsha, Hunan, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the University of Victoria, Victoria, BC, Canada, in 2013. From 2013 to 2017, he was a Research Scientist with the Central Research Institute at Huawei Technologies Co., Ltd. Since 2017, he has been an Assistant Professor with the School of Automation, Guangdong University of Technology, Guangzhou, China. He has published more than 20 journal and conference papers in peer-to-peer systems, VANETs, DTNs, and mobile computing systems. His research interests include various distributed systems and the design, and optimization of network protocols.

JUN TAO received the B.Sc. and M.Sc. degrees in computer science from Northeast University, China, in 1998 and 2001, respectively, and the Ph.D. degree in computer science from Southeast University, China, in 2005. He is currently a Professor of cyber science with Southeast University. His recent research interests include protocols and performance analysis of wireless ad hoc networks (MANETs/WSNs/VANETs).



JIANPING PAN received the bachelor's and Ph.D. degrees in computer science from Southeast University, Nanjing, China. He was a Postdoctoral Researcher with the University of Waterloo, Waterloo, ON, Canada. He was also with Fujitsu Labs and NTT Labs. He is currently a Professor of computer science with the University of Victoria, Victoria, BC, Canada. His area of specialization is computer networks and distributed systems. His current research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security. He received the Best Paper Awards of 2009 IEICE, 2011 WCSP, 2011 GLOBECOM, and 2013 ICC. He was the Ad Hoc and Sensor Networking Symposium Co-Chair of IEEE GLOBECOM 2012 and an Associate Editor of IEEE Transactions on Vehicular Technology. He has been serving on the technical program committees of major computer communications and networking conferences, including INFOCOM, ICC, GLOBECOM, WCNC, and CCNC. He is a Senior Member of ACM.