한국통신학회논문지(J-KICS) '13-11 Vol.38B No.11
http://dx.doi.org/10.7840/kics.2013.38B.11.934

Routing Relevant Data to Group Mobile Users by Mining Social Trajectory Pattern

Hyunjeong Cho◆, Yourim Park*, HyungJune Lee°

ABSTRACT

A routing scheme for a group of mobile users for wireless ad-hoc networks is presented. The proposed scheme mines social activity patterns from wireless traces, and exploits social user group for efficient data routing among users based on a data publish approach. Simulation results based on real-world wireless traces show that our routing scheme reduces routing cost for a large mobile user group with a factor of 1.8 compared to a baseline counterpart.

Key Words : Social Trajectory Pattern Mining, Data Publish, Group User Routing

Ⅰ. Introduction

As communication networks become distributed in D2D and M2M architecture, selective data delivery to mobile users based on their past preference and behavior in local areas gets more attention. To understand mobile users' unique behavior and preference for classifying data of their interests, a variety of opportunistic data delivery schemes such as SocialCast[1], DTN routing[2], and BubbleRap[3] have been proposed. However, these approaches rely only on opportunistic encounters of users, and do not effectively exploit any connection information to in-network stationary nodes.

Our work is different from previous literature in that we aim to extract social behavioral pattern from mobile nodes' visit history to stationary nodes for efficient routing to group mobile users. This letter addresses the problem of 1) classifying a group of mobile users of similar interests, 2) estimating relevant data, and 3) routing to the mobile user group through selected stationary relays in ad-hoc networks.

Ⅱ. Social Clustering

We present a social clustering algorithm that mines social activity patterns simply from wireless traces while moving around the physical space. The algorithm is divided into three phases below.

First, we construct a characteristic vector of a mobile user that embeds preference over particular locations. We use only wireless association traces where how long and how often the user visits a single mesh node. For each mobile user, we accumulate the association duration for each specific mesh node over the learning period, and use it as an element in the vector where the length of the vector is equal to the number of mesh nodes in the network as follows:

$$\overrightarrow{v_n}=[\frac{AssocDuration_{n,node1}}{TotalDuration},\frac{AssocDuration_{n,node2}}{TotalDuration},...,\frac{AssocDuration_{n,node\,M}}{TotalDuration}]^T$$

, which is the mobile user n's *characteristic vector* for stationary node 1 through node M.

Second, given a set of characteristic vectors of all mobile users, we calculate $L2$ Euclidean distance between all pairs of characteristic vectors from mobile users. Then, we apply a hierarchical clustering method for extracting hierarchical user relationship with respect to preference similarity. Depending on threshold levels, we can classify users into a group of roughly similar users, or up to a group of very similar users. We define the threshold level for this proposed social clustering as *group threshold(GT)*. Given a selected GT threshold, the hierarchical clustering method provides the number of clusters N, and each cluster's members as follows:

$$C_1=\left\{User_{k_11}, User_{k_12}, User_{k_13}, ...\right\}, C_2=\left\{User_{k_21}, User_{k_22}, User_{k_23}, ...\right\}, ...,$$

※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2013R1A1A1009854).
◆ First Author : KAIST 전산학과 컴퓨터 네트워크 연구실, hoonai14@kaist.ac.kr, 학생회원
° Corresponding Author : 이화여자대학교 컴퓨터공학과 지능형 네트워크 시스템 연구실, hyungjune.lee@ewha.ac.kr, 정회원
* 이화여자대학교 컴퓨터공학과 지능형 네트워크 시스템 연구실
논문번호 : KICS2013-10-454, 접수일자 : 2013년 10월 21일, 최종논문접수일자 : 2013년 11월 8일

$C_N = \{User_{k_N 1}, User_{k_N 2}, User_{k_N 3}, \dots\}$ where $\bigcup_{i=1}^{N} C_i$ is the set of all mobile users.

Third, based on the social clustering algorithm, we find mobile users of special interests depending on data attribute (e.g., the mesh node where location-sensitive data are created). If a specific mesh node *k* (as data provider) has data to deliver to potential mobile user group, we traverse through characteristic vectors of all mobile users, and pick up the most relevant mobile user *h* with the highest value among all users, i.e., $User_h = \arg\max_n \overrightarrow{v_{nk}}$. Also, we find similar user group $C_s(\ni User_h)$ given a group threshold level described above (see Fig. 1).

Ⅲ. Data Publish for Routing

We present a data publish scheme for routing to a group of mobile users. By classifying into a group of mobile users with similar interests through social clustering algorithm, we propose to route the relevant data not directly to the target mobile users, but to send instead to data publisher that will be likely visited by mobile users.

In order to select data publishers given a data provider (where location-sensitive data are created), we consider 1) where the target mobile users will traverse in common based on their previous history and 2) routing cost from the data provider to a respective data publisher candidate. We define a popularity of mesh nodes given the target mobile user group as the sum of respective elements in characteristic vectors of the target mobile user group. The higher popularity of a mesh node means that the target mobile users will visit the mesh node with the higher chance in common. However, if the mesh nodes even with high popularity are located far from the data provider, it takes high routing cost to publish data to the selected mesh nodes. For this reason, we also need to take into account routing cost from data provider to data publisher for the selection.

We use a normalized popularity measure by routing cost (multicast hop count used in

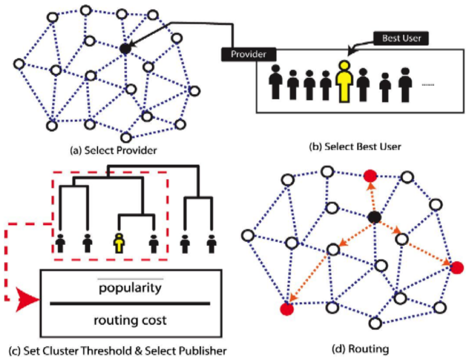


Fig. 1. Social clustering and data publish scheme.

experiments) from data provider *P* to publisher candidate node *i* given a group threshold GT:

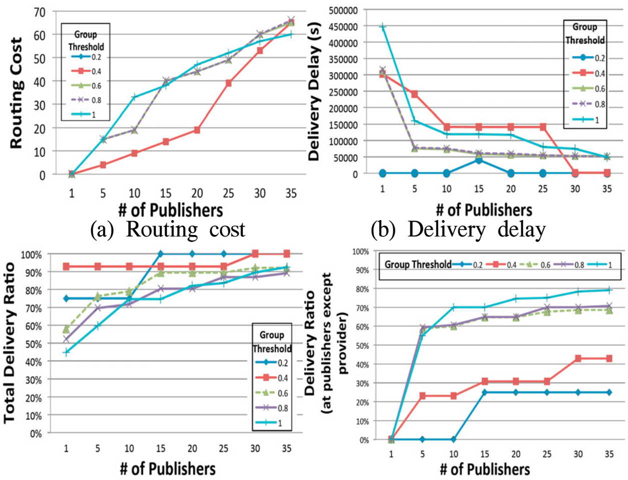
$$\overline{popularity}_{i,(P,GT)} = \frac{popularity_{i,(P,GT)}}{routingCost_{P \to i}}$$

Then, we select *n* publisher nodes with the n largest normalized popularity measures among all possible mesh nodes, *n DataPublishers* = $\{i \mid \arg\ n-largest\ \overline{popularity}_{i,(P,GT)}\}$ and then push multiple *n* copies of data to the selected publishers using Dijkstra's shortest-path routing algorithm. We apply the multicast delivery mechanism to reduce duplicate data transmission.

Ⅳ. Evaluation

We evaluate our proposed scheme using Dartmouth dataset[4]. The total number of mobile users is 67, and the number of access points (APs) in the network is 623. We use these APs as mesh nodes for our simulation. Since the mesh network of only these nodes is not fully connected, we append 35 additional mesh nodes on top of Dartmouth dataset. We implemented a packet-level simulator with a transmission range of 70 m in 802.11 mesh networks.

We measure network performance of our social routing scheme in terms of routing cost (defined as multicast hop count in the shortest path tree from a data provider), delivery delay, and delivery ratio. Then, we compare the proposed group routing scheme in terms of routing cost with the point-to-point shortest-path distance vector routing protocol where a data provider sends data directly to the currently connected node of each mobile user, serving as the most optimized one-to-one routing with the lowest routing cost. We use the



(a) Routing cost (b) Delivery delay

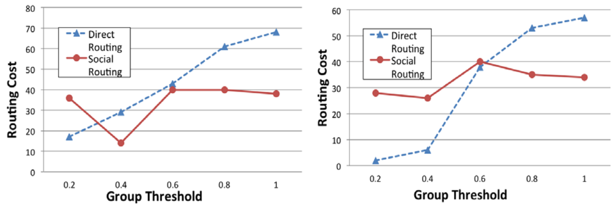
(c) Total delivery ratio (d) Delivery ratio(@non-providers)

Fig.2. Performance w.r.t group threshold and # of publishers.

data provider of AcadBldg18AP2 node (the access point SSID with the most popular traces).

First, we evaluate network performance with respect to group threshold. As the group threshold (corresponding to the number of group mobile users) increases, our proposed routing scheme does not incur much routing cost, reaching at a flat routing cost in Fig. 2(a). As the number of selected publishers increases, routing cost of our routing scheme increases. Regarding packet delivery delay, if the routing scheme decides to choose at least 15 publisher nodes, delivery delay (until mobile users pick up data at the publisher nodes) significantly drops (see Fig. 2(b)). The reason is that if more publishers are used, mobile users will more likely be connected to one of these publishers earlier. Fig. 2(c) shows that as the number of publishers increases, delivery ratio to mobile users reaches up to 100% (above 90% even for all of mobile users where GT = 1). Fig. 2(d) shows that delivery ratio via publishers other than the provider still reaches up to around 80%. This implies that even if we exclude the data provider itself as a data publisher, our proposed scheme achieves high delivery ratio by selecting common publishers used by other mobile users.

Finally, we compare our routing scheme (Social Routing) with the traditional direct routing approach (Direct Routing). Our social routing scheme uses 15 publishers, and routing cost is measured in case of the two most popular nodes



(a) Provider: AcadBldg18AP2 (b) Provider: LibBldg2AP15

Fig. 3. Routing cost compared to Direct Routing.

as data providers. As Fig. 3 shows, when a data provider delivers to a group of mobile users, our social routing scheme incurs flat routing cost due to the efficient usage of a constant number of publishers, whereas routing cost of the direct routing scheme increases proportionally with the number of mobile users. In case that a data provider needs to deliver to a larger mobile user group (where GT>0.6), the social routing scheme outperforms the direct routing with a factor of up to 1.8, leading to improved scalability due to the usage of only 15 selected publishers. Please note that there is a particular case at GT=0.4 in Fig. 3(a) that 15 publishers are selected closely each other, and thus, the Social Routing reduces (multicast) routing cost more significantly.

Ⅴ. Conclusion

This letter presented a routing scheme that delivers relevant data to a group of mobile users by mining social trajectory pattern from wireless traces. Evaluation based on real-world trace has demonstrated that our proposed routing scheme reduces routing cost for a group of many mobile users by selectively choosing publisher nodes.

References

[1] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 748-760, June 2008.

[2] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in Delay Tolerant Networks: A Social Network Perspective," in Proc. 10th ACM Int. Symp. Mobile Ad Hoc Networking Comput. (MobiHoc '09), pp. 299-308, New Orleans, U.S.A, May 2009.

[3] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," IEEE Trans. Mobile Comput., vol. 10, no. 11, pp. 1576-1589, Nov. 2011.

[4] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo, CRAWDAD trace dartmouth campus syslog (v. 2004-12-18), retrieved July 29, 2009 from http://crawdad.cs.dartmouth.edu/dartmouth/campus/syslog/01_04.