Internode Mobility Correlation for Group Detection and Analysis in VANETs

Yujin Li, *Student Member, IEEE*, Ming Zhao, *Member, IEEE*, and Wenye Wang, *Senior Member, IEEE*

***Abstract*—Recent studies on mobility-assisted schemes for routing and topology control and on mobility-induced link dynamics have presented significant findings on the properties of a *pair* of nodes (e.g., the intermeeting time and link life time) or a *group* of nodes (e.g., network connectivity and partitions). In contrast to the study on the properties of a set of nodes rather than individuals, many works share a common ground with respect to node mobility, i.e., *independent* mobility in multihop wireless networks. Nonetheless, in vehicular ad hoc networks (VANETs), mobile devices installed on vehicles or held by humans are not isolated; however, they are *dependent* on each other. For example, the speed of a vehicle is influenced by its close-by vehicles, and vehicles on the same road move at similar speeds. Therefore, the gap between our understanding of the impact of *independent mobility* and our interest in the properties of *correlated mobility* in VANETs, along with the real systems altogether, declare an interesting question. How can we *measure the internode mobility correlation,* such as to uncover the node groups and network components, and explore their impact on link dynamics and network connectivity? Bearing this question in mind, we first examine several traces and find that node mobility exhibits *spatial locality* and *temporal locality* correlations, which are closely related to node grouping. To study the properties of these *groups on the fly*, we introduce a new metric, i.e., *dual-locality ratio* (DLR), which quantifies mobility correlation of nodes. In light of taking spatial and temporal locality dimensions into account, the DLR can be used to effectively identify stable user groups, which in turn can be used for network performance enhancement.**

***Index Terms*—Correlated mobility, group detection, vehicular ad hoc networks.**

I. Introduction

In mobile ad hoc networks (MANETs), mobile devices communicate by wireless links through multihop networking. The mobility of wireless devices leads to connection and disconnection of communication links and routes, and grouping and dispersion of nodes, which in turn greatly affect network performance. Thus, mobility has been the focus of researchers' extensive studies that include modeling node mobility [2], examining the impact of mobility (e.g., in [3]–[6]), and designing mobility-based schemes (e.g., [7], [8]).

Manuscript received September 18, 2012; revised February 17, 2013 and May 11, 2013; accepted May 17, 2013. Date of publication May 22, 2013; date of current version November 6, 2013. This work was supported by the National Science Foundation under Award CCF-0830680 and Award CNS-1018447. Parts of this paper appeared in the *Proceedings of the IEEE 2011 Global Communications Conference* [1]. The review of this paper was coordinated by Prof. V. W. S. Wong.

The authors are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606 USA (e-mail: yli27@ncsu.edu; mingzh@cisco.com; wwang@ncsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2264689

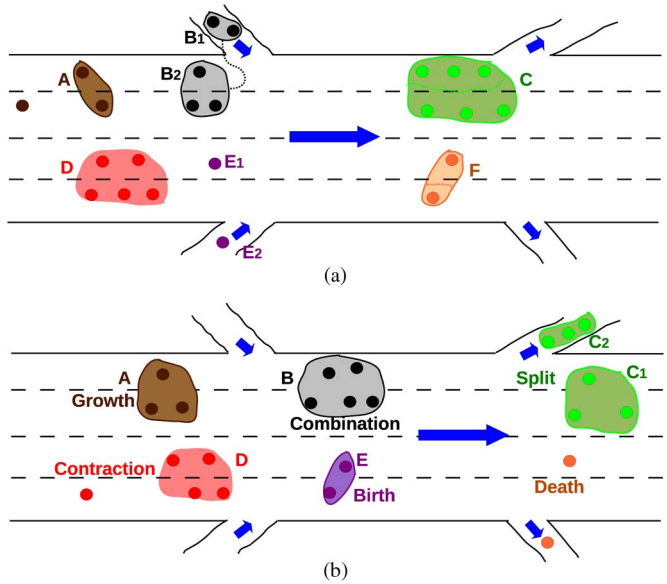


Fig. 1. Group evolutions. Birth (individual nodes E_1 and E_2 form new group E), death (nodes in group F disperse to move individually), growth (number of nodes in group A increases from 2 to 3), contraction (number of nodes in group D decreases from 5 to 4), split (group C splits into two groups C_1 and C_2), and combination (Groups B_1 and B_2 combines into a bigger group B). (a) $t = t_1$. (b) $t = t_2$.

Existing studies have focused on independent node mobility because of its simplicity for analysis and simulation. However, investigations of node mobility in real traces have observed that mobility of wireless devices, which is associated with mobile humans (e.g., pedestrians or drivers), exhibits a significant degree of correlation. In corporate and campus wireless local area network (WLAN) traces [9], [10], nodes can be classified into different groups according to their home locations where they spend most of their time. In vehicular ad hoc networks (VANETs), the mobility of cars is correlated as their movements are constrained by road layout, speed limit, and speed of nearby cars. This correlation of vehicle mobility leads to vehicle groups (i.e., connected network components), such as groups C and D shown in Fig. 1, which is known as the phenomena of *group mobility* in wireless communication networks. In addition, node mobility results in node groups changing over time, including the groups' growth, contraction, split, and combination shown in Fig. 1. In summary, correlated mobility directly determines the architectures and dynamics of node groups (i.e., connected network components).

Presence of groups due to correlated mobility means that nodes have unequal abilities to relay data to other parts of the network and could be exploited to improve information

dissemination and topology management. For example, two neighboring vehicles with similar speeds (i.e., high mobility correlation) are less likely to be out of each other's communication range (i.e., long link lifetime) and are thus more suitable for building routes with high stability. As another example, in a bus-based delay-tolerant network (DTN), a bus carrying a message may prefer sending copies to encountered buses that run on different routes (i.e., different mobility patterns) to increase the opportunity of meeting with a destination bus. Apparently, knowledge of *mobility correlations* among nodes is helpful for establishing stable routes, increasing packet delivery ratio, and reducing delay.

Being aware of the correlated node mobility in reality and its potential benefits in enhancing network performance, existing works have explored group mobility in improving routing performance [7], [11], predicting network connectivity and partition [12], [13], and assisting information dissemination [4], [8], [14]. These studies have assumed *preconfigured* group mobility behavior and group membership, such as reference point group mobility (RPGM) [15] and its variances [5], [12], [13].

Nevertheless, in a real-world setting, groups of wireless users are dynamically changing and evolving due to the autonomous human mobility and social behaviors, except in mission-oriented applications. Fig. 1 shows the group evolutions in a typical VANET scenario. The main occurring phenomena include group birth and death, growth and contraction, and split and combination [16]. For instance, groups B_1 and B_2 at time t_1 combine to a larger group B at t_2, whereas group C at t_1 splits into two groups C_1 and C_2 at t_2 because of vehicle movements. The dynamic movement behaviors of vehicles mean that group mobility is not a prior knowledge in wireless communication networks or in group structures. The dynamics of node groups make the insights and benefits of group mobility claimed by applications based on predefined group mobility and structure questionable. Therefore, the open question is *how to measure internode mobility correlation and identify dynamic user groups* on the fly such that we can leverage group mobility for efficient network design, performance enhancement, and topology management.

Because correlated mobility is still underexplored, in this paper, we aim to characterize the mobility correlations among nodes and identify mobile group structures in VANETs. To begin with, we examine how node mobility is correlated using real traces. We observe that the mobility of vehicles is not only correlated with the space domain (e.g., adjacent locations and similar speeds) but is also related with the time domain (e.g., overlapping movement paths and a shared visitation schedule to community sites). On one hand, vehicle mobility at a certain time depends on geographic surroundings (e.g., road layout and speed limit), which means that vehicle mobility shows *spatial locality*. On the other hand, being associated with mobile humans, mobile devices likely appear at several community sites (such as houses, offices, and shopping centers) with different frequencies and sojourn times at different times. In other words, node mobility, such as human mobility [17], exhibits *temporal locality* because of a user's movement schedule. Moreover, there is a close relationship between node mobility correlation in spatial and temporal localities and group structures, which can be explored to identify user groups on the fly.

A follow-up and challenging issue is how to *quantify* this internode mobility correlation in spatial and temporal localities to identify node groups. As spatial locality has two attributes (i.e., location and speed), and the similarity in spatial locality of two nodes is characterized by their *relative speed* and *distance*. Because temporal locality is characterized by the mobility schedules to execute social activities in community sites, which has a similar function with "caves" in ancient times, we consider a *cave profile* to model visiting frequencies, orders, and sojourn times of community locations in the network. By taking an *entropic* approach, correlation degree in temporal locality is measured as the *closeness* of cave profiles. Therefore, by incorporating both spatial locality similarity (SLS) and temporal locality similarity (TLS), we introduce a new metric, i.e., *dual-locality ratio (DLR)*, to quantify mobility correlation between a pair of nodes. A high DLR means that two nodes tend to join together as a group, whereas a low DLR means individual movements with little correlation.

In light of taking spatial and temporal localities into account, we find that the DLR can be used to identify group structures in two citywide vehicle traces. In the *cab-based* trace in which groups are likely formed or broken down due to *geographic* constraints, the DLR identifies cab groups according to locations and speeds of nodes. Whereas in the *bus-based* trace in which the trajectories indicate different levels of companionship because of the discrepancy in itineraries, the DLR detects bus groups consistent with bus routes. This attests that the DLR is able to catch correlations among nodes in spatial and temporal localities. Simulation results show that the DLR can recognize *stable* groups in *dynamic* network environments in real time under group mobility.

Identified group structures allow nodes to make routing decisions that can improve network performance, such as reducing packet loss. Simulation results show that DLR-assisted data forwarding can increase the packet delivery ratio. In addition, mobility correlation can be used to organize nodes into stable clusters, which is one of the most general applications of topology control. Simulations also show that the DLR-assisted clustering greatly reduces the changing rate of a cluster head.

Overall, this paper makes three contributions. First, it presents a new metric called DLR that quantitatively measures mobility correlation in spatial and temporal localities. Second, it shows that DLR can effectively identify real-time stable group structure in VANETs. Third, it incorporates DLR in assisting data forwarding and clustering.

The remainder of this paper is organized as follows. In Section II, we give a succinct summary of existing work on group mobility and the motivation of this paper. Section III presents our observations of spatial and temporal localities in node mobility by investigating two citywide traces. In Section IV, we quantify mobility correlation using a new metric, i.e., *DLR*. In Section V, we evaluate this metric through two traces and simulations by showing that the DLR can effectively capture the mobile group structures in wireless networks. Section VI gives applications of the DLR in assisting data forwarding and clustering. Section VII concludes this paper.

II. Related Work

In contrast to the random mobility modeled by popular mobility models (such as random walk), the moving behaviors of mobile users usually follow some mobility patterns and exhibit significant degree of correlations, which leads to overlapping movement trajectories of vehicles driven by humans [18]. As observed in corporate/campus WLAN traces [9], [10], mobile users spend most of their time at their home locations, where node gathering yields connected components. Traces in [19] show that nodes belonging to one community have frequent contacts and long contact duration. The correlated vehicle mobility (also called group mobility) clearly leads to node grouping on the road or at community sites where humans perform tasks or social activities.

As a result, many research works have elaborated upon the impact of group mobility. In particular, simulations in [3] show that routing protocols (i.e., dynamic source routing, destination-sequenced distance vector routing, and ad hoc on-demand distance vector routing) achieve the highest throughput and the least overhead with RPGM compared with Freeway and Manhattan models. The authors of [18] observe a significantly reduced packet delivery ratio when employing the realistic trace simulator to control mobility of nodes. The work in [5] reveals that correlated node movements have a huge impact on asymptotic throughput and delay, and can sometimes lead to better performance than that achievable under independent node movements.

Existing studies on correlated mobility heavily rely on group mobility models with simplified node grouping behaviors, such as RPGM and virtual track models [20]. It is commonly assumed that nodes are partitioned into several groups beforehand and group memberships either never change (e.g., RPGM) or evolve according to a certain stochastic process. For instance, the virtual track model [20] for a vehicular network scenario binds nodes' group movements on edges in a graph, and group split and merge only happen at vertices. The work in [21] assumes that groups of nodes merge or split according to a Markov chain process.

Nevertheless, in a spontaneously deployed ad hoc network with no preconfigurations, mobile nodes have no prior knowledge about the mobility groups and their memberships. Moreover, rather than the simplified node grouping behaviors in existing group mobility models, node groups in reality evolve not only in various ways (such as growth, contraction, combination, and split, as shown in Fig. 1) but also according to complex mechanisms due to the autonomous human mobility and complicated social behaviors of mobile users. As group mobility is still underexplored, correlated mobility needs to be studied to identify dynamic node groups.

In this paper, we try to identify dynamic node groups that are induced by correlated mobility of vehicles. We achieve this by describing mobility correlation between any two vehicles through a novel metric, i.e., DLR, which incorporates mobility correlations of nodes in both spatial and temporal domains. The effectiveness of DLR is evaluated using traces and simulations. In addition, we demonstrate how DLR can be leveraged to assist data forwarding and clustering.

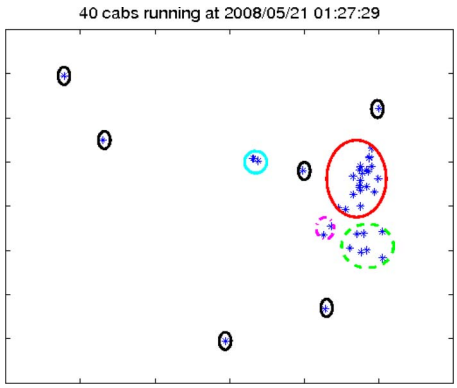


Fig. 2. Geographical grouping in SFCAB.

III. Correlated Mobility in Empirical Traces

In vehicle-to-vehicle networks, mobile devices are mostly carried by drivers/passengers or installed on vehicles that are driven by humans, i.e., mobile devices are associated with mobile humans. Vehicle mobility, in contrast to random independent and identically distributed mobility model, is restricted by geographic surroundings and dependent on each other, which are examined using real traces here.

A. Spatial Locality of Node Mobility

To find out *whether there exist correlations among mobile nodes in the real world*, we start with a taxicab trace in San Francisco (SF) from the Cabspotting Project [22]. This trace is chosen for our study because the customers may have quite different destinations and independent movements, which do not show obvious mobility correlations. The SF cab trace (SFCAB) contains GPS logs of 536 yellow cabs for over 30 days. Cab location is updated almost every minute if the cab stays online. In this cab-based VANET, each cab is taken as a mobile node. We extract locations of 40 cabs running at 01:27 on May 05, 2008, as shown in Fig. 2.

By fixing the time, Fig. 2 shows node mobility by a cab's location and speed. These two attributes of vehicle mobility are constrained by road layout and speed limit, which demonstrate the *spatial locality* of node mobility. Although cabs likely move independently as their destinations span a broad area, we still observe several cab clusters with different sizes, i.e., there exist mobility correlations among cabs. In Fig. 2, there are two larger groups (in red and green) and two smaller groups (in blue and pink), along with six individual nodes. In the largest group (in red circle), 22 cabs are located in downtown SF, which means that these 22 users have similar spatial locality properties (i.e., adjacent locations and possibly similar speeds). This node grouping is probably caused by vehicles moving on the same road, around attraction of hot spots (such as airports and shopping centers). In other words, nodes with similar spatial locality tend to form user groups.

Remark 1: The snapshot of the cab trace shows that node mobility exhibits spatial locality and that mobile users with similar spatial locality can establish connected network components.

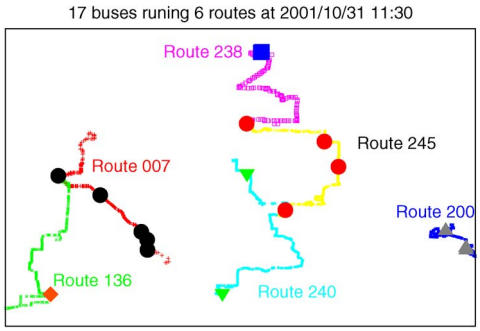


Fig. 3. Social grouping in STBUS.

B. Temporal Locality of Node Mobility

In addition to the observation of mobility correlation in spatial locality aspect, we turn to another trace, i.e., a bus trace in Seattle (ST) from the Ad Hoc City Project [23], to explore the mobility pattern of vehicles. This trace is considered because the movements of buses are not as independent as the cab-based trace, serving people who may share similar stops and timing patterns. The Seattle bus trace (STBUS) contains GPS logs of more than 1200 buses running 239 routes for around 20 days. Bus location is updated about every 2 min while the bus is running. Because there are frequent holes in the STBUS trace, we observe 17 buses running six routes at 11:30 A.M. on October 10, 2001, shown in Fig. 3. The curves are trajectories of bus routes and the points are current locations of buses.

In the Ad Hoc City Project, bus stops are important locations in the network. The important locations in the network are compared with *caves* in ancient times when people have social activities in caves of tribes. Over time, a bus demonstrates its mobility pattern through its *frequent* visits to the stops on its running route while it seldom travels to others, it has a long *sojourn time* at bus stations and short stay time at stops on roadside, and its *visit order is* from one stop to another (e.g., in-town or out-town). This dependence of mobility pattern on caves in the network reveals the *temporal locality* of node mobility.

Fig. 3 shows that buses running on the same routes tend to form groups (e.g., blue route 200 and red route 007). Clearly, buses on the same route have identical movement trajectory and thus likely meet frequently (e.g., the three black dots on the lower part of Route 007). In other words, buses can be grouped with respect to stops for different purposes, such as shopping centers and industry parks, which are related to social behavior and similar temporal locality. Notice that buses running the same route do not always locate near to each other, e.g., the four buses (in red spots) on Route 245 are almost evenly distributed en route. That means that correlation in temporal locality is different from correlation in spatial locality as temporal locality is induced by social behaviors, whereas spatial locality is induced by geographic limitations.

Remark 2: In addition to the mobility correlation in spatial locality due to geographic constraints, correlation in temporal locality induced by social gatherings or similar activity schedules also has impact on grouping of nodes. This adds another modality to explore the internode mobility correlation.

Because correlated node mobility results in dynamic node groups and network structures, measuring mobility correlation is needed for identifying network structures and exploring its impact on network performance and application in network design. Two nodes in each other's transmission range maintain a communication link, and a group of nodes that locate and move closely (high correlation in spatial locality) would form a stable group (i.e., connected network component). Nodes with similar temporal locality are likely to meet frequently and stay together for a long period, resulting in short intermeeting time and long contact duration. Therefore, mobility correlation in both spatial and temporal localities should be measured to identify node groups on the fly.

However, existing studies fail to capture both modalities in measuring mobility correlation. The works in [24] and [25] proposed metrics that measure mobility correlation based on nodes' relative *location* and *speed* (i.e., spatial locality) at a certain time while neglecting correlation in mobility patterns (i.e., temporal locality) over time. On the other hand, many community detection algorithms [26] measure mobility correlation based on contact information between nodes. This approach tends to detect long-term community structure but is unable to capture the group evolutions in real time. To fully understand the properties of a group or network component, it is desirable to capture mobility correlation in both spatial and temporal locality domains.

IV. Measuring Internode Mobility Correlation

Based on our observations in Section III that vehicle mobility is correlated in spatial and temporal localities, the challenge is how to quantify this dynamic mobility correlation. Here, we first investigate correlations in spatial and temporal localities separately and then present a new metric, i.e., *DLR*, that combine these two factors to measure the mobility correlation.

A. SLS

As location and speed are two factors of the spatial locality of vehicle mobility, *distance* $d_{i,j}(t)$ and *relative speed* $v_{i,j}(t)$ provide information for the similarity of spatial locality between two nodes at time t. The SLS measure $\mathrm{SLS}_{i,j}(t)$ is time varying to capture dynamic mobility correlations. $\mathrm{SLS}_{i,j}(t)$ should increase when either the distance $d_{i,j}(t)$ or relative speed $v_{i,j}(t)$ decreases. To make sure that the spatial and TLS measures have the same order of magnitude, we also require $\mathrm{SLS}_{i,j}(t)$ to be normalized within $[0, 1]$.

With the statistical distance measure defined in statistical analysis [27], we are able to quantify the SLS of two vehicles with their relative movement information. Regarding location and speed as the two attributes of the 2-D spatial locality observation for a node, the statistical distance between the two 2-D observations of two nodes (i, j) at time t is $\sqrt{w_1 d_{i,j}^2(t) + w_2 v_{i,j}^2(t)}$ where $d_{i,j}(t)$ and $v_{i,j}(t)$ are their relative distance and speed, respectively, and w_1 and w_2 are the weight coefficients for each attribute. Note that $v_{i,j}(t) = \sqrt{|v_i(t)|^2 + |v_j(t)|^2 - 2|v_i(t)||v_j(t)|\cos\theta}$, where $|v_i|$ and $|v_j|$ are the magnitude levels of nodes v_i and v_j's speeds, and θ is the angle between their moving directions. Let $r_{\max}$ denote the maximum transmission range and $v_{\max}$ denote

the maximum speed. Because two neighboring nodes satisfy $0 \le d_{i,j}(t) \le r_{\max}$ and $0 \le v_{i,j}(t) \le 2v_{\max}$, we can set $w_1 = 1/(r_{\max})^2$ and $w_2 = 1/(2v_{\max})^2$ so that $d_{i,j}(t)$ and $v_{i,j}(t)$ can be scaled to have the same order of magnitude (i.e., be within the range of $[0, 1]$ for a neighboring node pair). Therefore, we define the SLS between a neighboring node pair (i, j) as

$$\mathrm{SLS}_{i,j}(t) = \frac{1}{1 + \sqrt{\left(\frac{d_{i,j}(t)}{r_{\max}}\right)^2 + \left(\frac{v_{i,j}(t)}{2v_{\max}}\right)^2}}. \quad (1)$$

$\mathrm{SLS}_{i,j}(t) = 0$ for two nonneighboring nodes i and j as no communication link exists between them.

Clearly, this SLS measure satisfies all our requirements. $\mathrm{SLS}_{i,j}(t)$ is symmetric and $0 \le \mathrm{SLS}_{i,j}(t) \le 1$. When $d_{i,j}(t)$ and $v_{i,j}(t)$ decrease to 0, which means that the two nodes move together at the same speed, similar to two people sitting in the same car, we have $\mathrm{SLS}_{i,j}(t)$ approaches to 1, implying the highest SLS. On the contrary, when the distance and the relative speed between these two nodes, i.e., $d_{i,j}(t)$ and $v_{i,j}(t)$, respectively, are large, then the value of $\mathrm{SLS}_{i,j}(t)$ becomes very small, even to 0, which indicates very little or no SLS between this pair of nodes. In other words, the more adjacent the locations and similar the speeds of nodes are, the higher SLS they have, which indicates stable communication links and local groups.

B. TLS

In addition to correlation in the spatial locality domain, correlation in temporal locality is also critical for understanding the mobility correlation of nodes. Here, we take an *entropy-based* approach to measure the degree of similarity of temporal locality. Specifically, we first present a mathematical model, i.e., a *cave profile*, to characterize the temporal locality of an individual node. Then, we measure the closeness of different cave profiles based on the idea of *relative entropy*.

1) Cave Profile Modeling: Existing studies on mobility patterns of mobile users [9], [10], [17], [28] have shown that, in contrast to the random trajectories predicted by the prevailing Levy flight and random-walk models [29], movements of mobile users show a high degree of *temporal locality*. More specifically, a mobile user not only visits different communities with *different frequencies* and stays there for *unequal amounts of time* but it also visits these sites at *different times*. As vehicles (e.g., buses, cabs, personal cars) have become the main transportation tool for mobile users (i.e., nodes in a VANET are associated with mobile humans), node mobility in VANETs also exhibits temporal locality. Therefore, we model vehicle's temporal locality by characterizing the *frequency* and *sojourn time* in each location and the *order* in which the locations are visited [30].

Assume that N nodes move in a network with M communities, which could be hot spots (such as campuses, houses, and offices) extracted from a real map or partitions of the network area. Mobile users visit different communities to perform different social activities, which is similar to people attaching to communities in caves in ancient times. Thus, communities in the network are referred to as "*caves.*" Let $\Omega = \{c_1, \ldots, c_M\}$ be the set of all caves in the network. Further, we assume that mobile nodes can record their location, either by GPS or other localization methods in [31]. Suppose that the time is slotted with each time interval Δt (e.g., $\Delta t = 1$ h). Let $X_i(t)$ be the cave at which node n_i presents at time slot t and $T_i(t) = \{X_i(1), X_i(2), \ldots, X_i(t)\}$ denote the historic caves visited over the past t time slots. In fact, $T_i(t)$ includes information on n_i's visiting *frequency*, *sojourn time*, and visiting *order* of community sites and can be thus used to extract temporal locality information.

We also notice that prior study [30] found that human mobility can be predicted with high accuracy ($\ge 93\%$) based on its movement trajectory because humans tend to repeat their daily schedules, such as working at offices during daytime and staying at their home in the evening. As nodes in VANETs are associated with mobile humans, node mobility would also show predictability. Moreover, they found that the best mobility predictability is achieved through the entropy measure based on $P(T_i')$, which is the probability of finding a particular time-ordered subsequence T_i' in the trajectory T_i, as it captures the full spatiotemporal order present in a person's mobility pattern. Therefore, we can estimate the likelihood of n_i targeting cave c_m as the next destination, which is denoted as $p_i^t(c_m)$, by the probability of finding the time-ordered subsequence $T_i(t-k, t)c_m$ conditioned on finding subsequence $T_i(t-k, t)$ in node's visiting history $T_i(t)$.

Definition 1: The *cave profile* of node n_i at time t is defined as $P_i^k(t) = \{p_i^k(c_1, t), \ldots, p_i^k(c_M, t)\}$, i.e.,

$$p_i^k(c_m, t) = P\left(X_i^{t+1} = c_m | T_i(t)\right) = \frac{N\left(T_i(t-k, t)c_m, T_i(t)\right)}{N\left(T_i(t-k, t), T_i(t)\right)} \quad (2)$$

where $T_i(t-k, t) = \{X_i(t-k), \ldots, X_i(t)\}$ is the substring of $T_i(t)$ recording the visitation history in the recent k time slots, and $N(s', s)$ denotes the number of times that the substring s' occurs in s. $N(T_i(t-k, t), T_i(t))$ and $N(T_i(t-k, t)c_m, T_i(t))$ are the numbers of times that time-ordered sequences of visited caves $\{X_i(t-k), \ldots, X_i(t)\}$ and $\{X_i(t-k), \ldots, X_i(t), c_m\}$ appear in movement history $T_i(t)$, respectively, which are determined by visiting frequencies, sojourn times, and visiting orders of the caves in the network. $P_i^k(t)$ depends only on the k most recent locations; thus, we refer to it as $O(k)$ *cave profile*. As (2) estimates the likelihood of appearing at cave c_m in the next time slot based on all three attributes of temporal locality, the cave profile mathematically represents the temporal locality.

If $N(T_i(t-k, t), T_i(t)) = 1$, $p_i^k(c_m, t) = 0$ for $m = 1, \ldots, M$. To avoid this case, (2) degrades to $O(k-1)$. If node n_i visits $X_i(t)$ for the first time, (2) degrades to the $O(0)$ cave profile, in which

$$p_i^0(c_m, t) = N\left(c_m, T_i(t)\right) / |T_i(t)| \quad (3)$$

where $|T_i(t)|$ is the length of $T_i(t)$. Clearly, the $O(0)$ cave profile characterizes the temporal locality based on visiting frequency and sojourn time. To find the value of k, we take the suggestion from an earlier work [32] that human mobility can be well predicted depending on the two most recent locations.

Therefore, we recommend the $O(2)$ cave profile for the simplicity of analysis and practical applications. That means

$$p_i^2(c_m, t+1) = \frac{N\left(X_i(t-1)X_i(t)c_m, T_i(t)\right)}{N\left(X_i(t-1)X_i(t), T_i(t)\right)}. \quad (4)$$

In addition, we normalize the cave profile to ensure that $\sum_{m=1}^{M} p_i^k(c_m, t) = 1$ by rewriting

$$\hat{p}_i^k(c_m, t) = \frac{p_i^k(c_m, t)}{\sum_{j=1}^{M} p_i^k(c_j, t)}. \quad (5)$$

The cave profile is able to model the temporal locality of *individual* mobility. Then, we move on to the measurement of TLS between *two* nodes.

2) Measuring TLS: Among various similarity/distance measures that compare two probability distributions (see [33]), the *Kullback–Leibler divergence* (KLD) (also relative *entropy*) is a well-known method of measuring the difference between two probability vectors in information theory. The KLD is well defined for both discrete and continuous distributions, and is always nonnegative. However, the KLD is a nonsymmetric measure and is sensitive to quantization effects in the histogram computation. The *Jenson–Shannon divergence* (JSD), which is the symmetrized and smoothed version of the KLD, is an empirically derived divergence that is numerically stable and is also robust in the presence of noise. Thus, we choose the JSD to measure the similarity between mobility patterns.

Definition 2: The JSD between $\hat{P}_i^k(t)$ and $\hat{P}_j^k(t)$ is

$$\begin{aligned}\mathrm{JSD}&\left(\hat{P}_i^k(t) \| \hat{P}_j^k(t)\right) \\ &= \frac{1}{2}\sum_{m=1}^{M} \hat{p}_i^k(c_m, t) \log_2 \frac{2\hat{p}_i^k(c_m, t)}{\hat{p}_i^k(c_m, t) + \hat{p}_j^k(c_m, t)} \\ &\quad + \frac{1}{2}\sum_{m=1}^{M} \hat{p}_j^k(c_m, t) \log_2 \frac{2\hat{p}_j^k(c_m, t)}{\hat{p}_i^k(c_m, t) + \hat{p}_j^k(c_m, t)}. \quad (6)\end{aligned}$$

For the sake of simplicity, we refer to node n_i's cave profile as $\hat{P}_i(t)$ in the following. The JSD measure of TLS is thus defined as

$$\mathrm{TLS}_{i,j}(t) = 1 - \mathrm{JSD}\left(\hat{P}_i(t) \| \hat{P}_j(t)\right). \quad (7)$$

$0 \le \mathrm{TLS}_{i,j}(t) \le 1$ and $\mathrm{TLS}_{i,j}(t)$ is symmetric. $\mathrm{TLS}_{i,j}(t) = 1$ if and only if $\hat{P}_i(t) = \hat{P}_j(t)$. The closer the cave profiles of a pair of nodes are, the larger their TLS is, and *vice versa*.

3) Illustrative Example: To better understand the cave profile model and TLS measure, let us take an example of four mobile users with five-site options, i.e., $\mathcal{M} = \{1, 2, 3, 4, 5\}$. The location history for each user is shown in Table I. The historic observations, although short, mimic the skewed location visiting preferences of human mobility. For instance, n_1 visits communities c_1 and c_2 much more often than the other three communities, whereas n_3 mostly stays at c_2 and c_5. In contrast to the sequence of site visits of n_1, n_2 goes to c_2 after staying at c_1.

By applying (4) for the $O(2)$ cave profile, the likelihood of n_i visiting cave m after visiting caves c_3 and c_4 can be calculated, e.g., $p_1^2(1, t) = N(\{341\}, T_1)/N(\{34\}, T_1) = 1/2$ and $p_1^2(i, t) = N(\{34i\}, T_1)/N(\{34\}, T_1) = 0$ for $i = 2, 3, 4$.

TABLE I
EXAMPLE OF USER CAVE PROFILES

User	Location history T_i	Cave Profile
n_1	111122234111122234	{1, 0, 0, 0, 0}
n_2	222111134222111134	{0, 1, 0, 0, 0}
n_3	555522234555522234	{0, 0, 0, 0, 1}
n_4	111133334111133334	{1, 0, 0, 0, 0}

Accordingly, the normalized $\hat{P}_i^2$ is obtained using (5), and the resulting cave profiles are shown in Table I.

Using (6) and (7), the TLS between each user pair can be calculated, which is $\mathrm{TLS}_{1,4} = \mathrm{TLS}_{4,1} = 1$, and $\mathrm{TLS}_{i,j} = 0$, otherwise. The results indicate that nodes n_1 and n_4 have high TLS and thus likely move together as both of them probably will go to *Cave 1*. It is interesting to see that even n_1 and n_2 have the same historical probability of visiting each cave according to (3), they are not likely be moving together because they seldom appear at one location at the same time. In other words, $O(0)$ cave profile is inadequate for measuring TLS.

Remark 3: Based on the mobility history, the cave profile can be used to estimate the likelihood that each community is chosen as a user's next destination. By measuring the closeness of cave profiles, TLS shows the likelihood of two users visiting the same cave during the next time slot and thus tells the tendency of two users moving as a group.

C. DLR

Thus far, we have investigated SLS and TLS, both of which are essential in characterizing internode mobility correlation. By introducing a tune-up parameter α to jointly consider the given observations, we propose a new metric, i.e., *DLR*, to measure internode mobility correlation.

Definition 3: $\mathrm{DLR}_{i,j}(t)$ between two nodes (i, j) at time t is given by

$$\mathrm{DLR}_{i,j}(t) = (1-\alpha)\mathrm{SLS}_{i,j}(t) + \alpha\mathrm{TLS}_{i,j}(t) \quad (8)$$

where $0 \le \alpha \le 1$, and $0 \le \mathrm{DLR}_{i,j}(t) \le 1$.

Through adjusting the value of α, the weights of SLS and TLS can be adapted for different network scenarios. For networks that most nodes move independently, the DLR with small α can represent the mobility correlation in spatial locality, whereas for nodes with clear mobility patterns, the DLR with large α can manifest the similarity of nodes' temporal locality. The DLR can also accommodate different applications by choosing different α. Large α (even $\alpha = 1$) is suitable for mobility pattern recognition or communiction detection. On the other hand, small α (even $\alpha = 0$) is fit for link or path duration estimation.

The given definition of DLR measures the mobility correlation between a *pair* of nodes. In the following, we further show that DLR can be used to study properties of a *group* of nodes, such as group structure and stability.

V. IDENTIFYING NETWORK COMPONENTS ON THE FLY

A *group* means a number of nodes bounded together as being related in some way.

Definition 4: Let $\mathrm{DLR}_{\mathrm{th}}$ be the required *grouping threshold* for two users to belong to a group. Two neighboring nodes (i, j) are in the same group if $\mathrm{DLR}_{i,j} \geq \mathrm{DLR}_{\mathrm{th}}$. According to this definition, nodes can identify whether encountered nodes belong to a group. Each node i first obtains information on cave files, locations, and speeds from its neighbors to calculate DLR. When $\mathrm{DLR}_{i,j}$ exceeds $\mathrm{DLR}_{\mathrm{th}}$, node i will consider node j as its group member. Through further exchanging each other's group member information among encountered nodes, group structures can be uncovered. As nodes change information among neighbors, identified groups are connected network components.

The main overhead of the computing DLR is due to the exchange of mobility information between neighboring mobile nodes. Two types of messages are generated and periodically broadcast by nodes to update: 1) spatial locality information; and 2) temporal locality information. The spatial locality message includes node's location, speed, and moving direction. The temporal locality message includes the node's cave profile (i.e., caves that the nodes visited over the past t time). Assume that the sizes of the spatial locality message and the temporal locality message are S_s and S_t, and the corresponding broadcast frequencies are f_s and f_t, respectively. The overhead at each node is $S_s f_s + S_t f_t$. Clearly, the spatial locality message has small S_s. The spatial locality update frequency is related to relative velocity and transmission range r of vehicles [34]. f_s can be set as $r/2v_{\max}$ so that the link that is established or broken can be captured. For example, f_s approximately equals to one message per 4 s when $r = 250$ m and $v_{\max} = 30$ m/s. On the other hand, f_t can be set as one message per 10 min, which is sufficient to capture the driver's visit to different communities. The temporal locality message only needs to record visited caves over 24 h as humans tend to repeat their daily movement schedules. The total overhead in the network is $(S_s f_s + S_t f_t)N$, where N is the number of nodes in the network.

A. Trace Evaluation

To use the DLR to identify groups, spatial and temporal localities need to be extracted from traces. Because locations of nodes are logged in both data sets, information on spatial locality can be easily obtained. Hence, we focus on how we extract cave profiles for characterizing temporal locality and the results of group identification.

1) Group Identification in SFCAB: First of all, we investigate whether there is traceable temporal locality feature in cab mobility. As SFCAB records whether a cab is carrying customers or not, we can extract locations where cabs pick up or drop off customers, i.e., locations of *stops*. It is worth noting that the customers are autonomous and take cabs without coordination, which means that there is no correlation among their destinations. Surprisingly, the stops of three cabs shown in Fig. 4 reveal that locations of stops are not uniformly distributed in the city area. For instance, *Cab 1* visits the western area more often than the eastern part of the city, whereas *Cab 2* prefers the eastern area. This may be because cab drivers prefer working in different areas or hot spots (e.g., airport or downtown), i.e., cab mobility exhibits temporal locality.

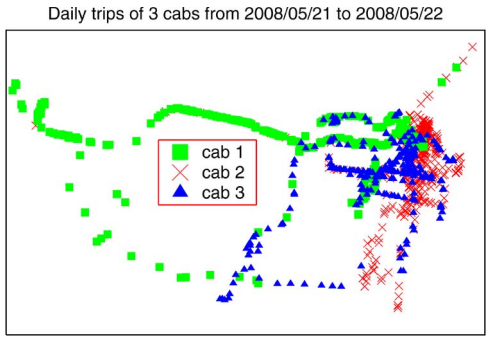


Fig. 4. SFCAB. Temporal locality in cab mobility.

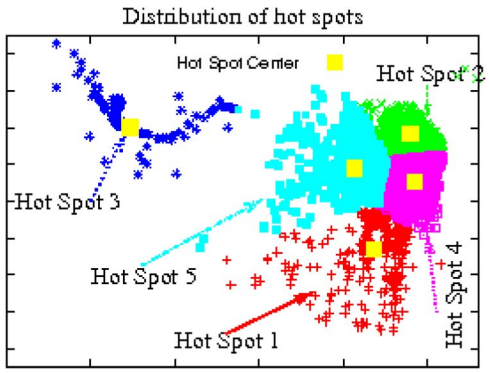


Fig. 5. SFCAB. "Caves" are visible.

To measure the TLS of two cabs, we then identify hot spots as "caves," where people frequently get on or off cabs. The stops of 100 cabs are clustered to five caves (hot spots) through the *k-means clustering* algorithm in Matlab, which is shown in Fig. 5. The stop locations of one cab in Fig. 5 show that this cab visits caves 2, 4, and 5 more frequently, whereas it occasionally visits caves 1 and 3. For simplicity, we use $O(0)$ cave profile, i.e., the probability that a cab stays in each cave. Car n_i's preference for cave c_m is

$$p_i^0(c_m) = \frac{\text{number of stops in cave } c_m}{\text{total number of stops of car } n_i}. \tag{9}$$

By parsing the data in the SFCAB trace file, we finally identify three groups (1, 4, and 6) with multiple members and nine additional single nodes, as shown in Fig. 6. The length and the direction of an arrow represent the speed and the moving direction of a node, respectively, and the dotted line between n_i and n_j means that $\mathrm{DLR}_{i,j} \geq \mathrm{DLR}_{\mathrm{th}}$, where $\mathrm{DLR}_{\mathrm{th}} = 0.2$ is the average DLRs among all node pairs. By taking a closer look, we further observe that node 5 does not join *group* 1 because node 5 is not within the transmission range of any node in this group. In addition, nodes 2 and 3 cannot be clustered to a group because they move at different speeds, i.e., the arrow length (speed) for node 3 is much longer (faster) than that of node 2. Nodes 9 and 10 are classified as two groups because of their opposite moving directions.

When we change $0 < \alpha < 1$ with different values, i.e., different weights of TLS and SLS, we observe a little difference in identified groups, as shown in Fig. 6. One plausible reason is that cabs mostly move independently and do not show

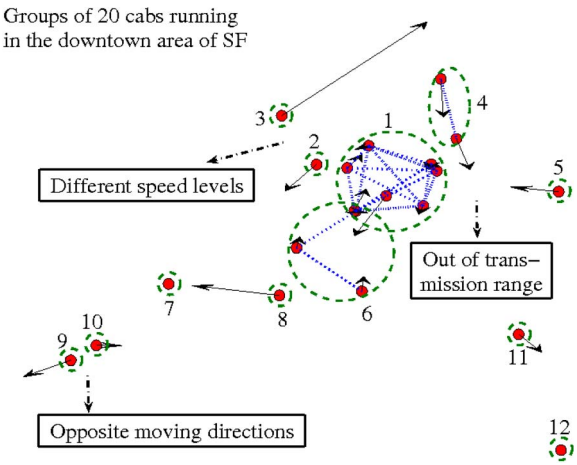


Fig. 6. SFCAB. Identified groups with $\alpha = 0.5$ and $\mathrm{DLR}_{\mathrm{th}} = 0.2$.

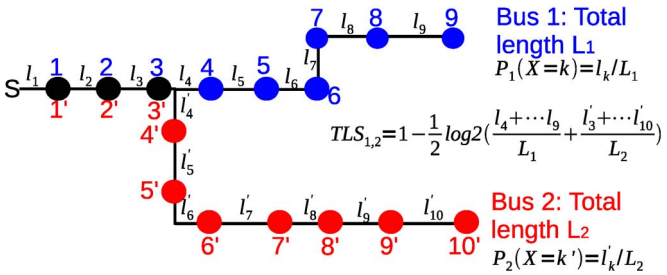


Fig. 7. TLS between two buses.

apparent correlation in temporal locality; thus, SLS dominates node grouping, and TLS has little impact on identifying group structures in the SFCAB.

Remark 4: Fig. 6 shows that the DLR can effectively identify groups, in which nodes have similar mobility features (such as location and velocity), have more connections among them than connections outside the group, and form connected network components. Therefore, DLR can properly measure the SLS among nodes based on their relative distance and speed.

2) Group Identification in STBUS: Assume that each bus stop is a "cave," and Ω is the set of all bus stops. We extract cave profile $P_i^0 = \{p_i^0(c_1), \ldots, p_i^0(c_M)\}$ by calculating $p_i^0(c_m)$ as the likelihood that bus i appears at bus stop c_m. For c_m on a running route of i, $p_i^0(c_m)$ is i's sojourn time around c_m, which can be approximated by l_m/L_i, where l_m is the length between c_m and its previous stop, and L_i is the length of the running route of bus i (see Fig. 7); otherwise, $p_i^0(c_m) = 0$. Using (6) and (7), the TLS between two buses running routes i and j is simplified as

$$\mathrm{TLS}_{i,j} = 1 - \frac{1}{2}\log_2\left(\frac{L_i - L_{i,j}}{L_i} + \frac{L_j - L_{i,j}}{L_j}\right) \qquad (10)$$

where $L_{i,j} = L_i \cap L_j$ is the overlapping length of two routes.

Note that, for two buses running on the same route, $\mathrm{TLS}_{i,j}$ is 1 as they have the same mobility pattern. For buses running on different routes, $\mathrm{TLS}_{i,j}$ depends on the proportion of overlapping trajectory over the total route length. The more two routes overlap, the more similar their temporal locality become.

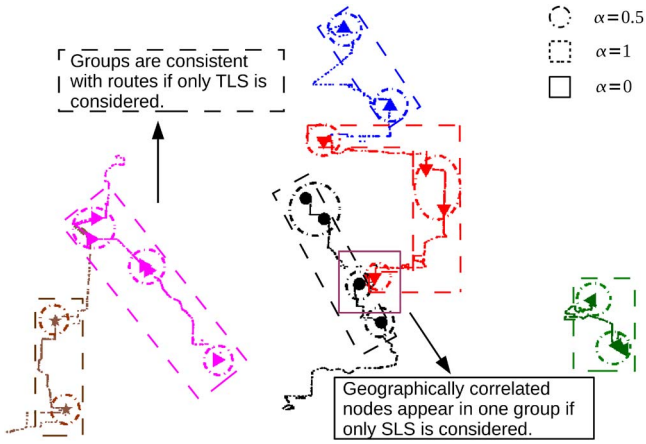


Fig. 8. STBUS. Identified groups with $\alpha = 0, 0.5$, and 1.

Using DLR in (8) and setting $\mathrm{DLR}_{\mathrm{th}}$ as the average of DLRs over all neighboring node pairs, the group structure in STBUS is shown in Fig. 8. Groups in dashed rectangles are obtained by using $\alpha = 1$, which are consistent with bus routes. Based solely on SLS ($\alpha = 0$), buses running on different routes may be clustered into a group (the small solid square in the middle) as long as they geographically locate closely. By considering both SLS and TLS, e.g., $\alpha = 0.5$, identified groups (the dashed circles) only include buses running on the same route and moving closely.

Remark 5: The DLR can effectively detect user groups in STBUS, in which nodes have the same mobility pattern (i.e., traveling route) or closely located. Nodes belonging to the same group have more interactions among them than with other nodes (particularly nodes running on different routes), which implies a good measurement of mobility correlations among mobile nodes.

B. Simulation Evaluation

By examining real traces, we have observed how the DLR can be used in identifying groups with either apparent correlation in spatial locality (e.g., cabs) or temporal locality (e.g., buses). To the best of our knowledge, there is no trace with both spatial and temporal locality information available for public use; thus, we resort to simulations for a thorough evaluation.

1) TSC Mobility: To reproduce the skewed location preferences of human mobility, i.e., people spend most of their time at a few frequented caves and stay shortly at other caves, we implement a *time–space-varying caveman (TSC)* mobility model in OMNeT ++ [35].

At $t = 0$, each node assigns its cave preference profile $\{p_i(c_m), 1 \le m \le M\}$, according to truncated power-law (TPL) distribution to ensure the skewed location preferences. Denote by $W_i = \{w_i(c_1), w_i(c_2), \ldots, w_i(c_M)\}$ the pausing times of node i at M caves, where $w_i(c_m)$ is proportional to $p_i(c_m)$. Accordingly, nodes visit and stay in different caves with different *weights/probabilities* that enables hierarchical location preferences.

An example with five communities in the network is shown in Fig. 9. Node i at home location generates location profile

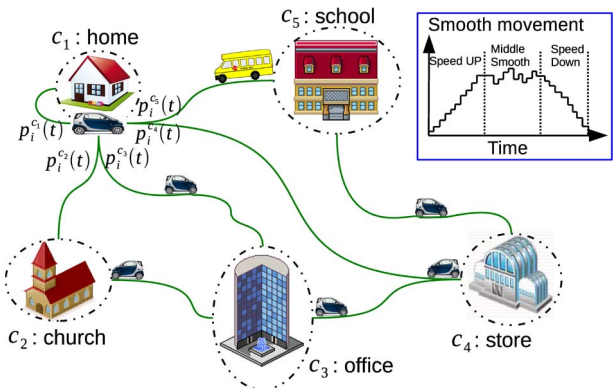


Fig. 9. TSC mobility. Locality modeling and smooth movement.

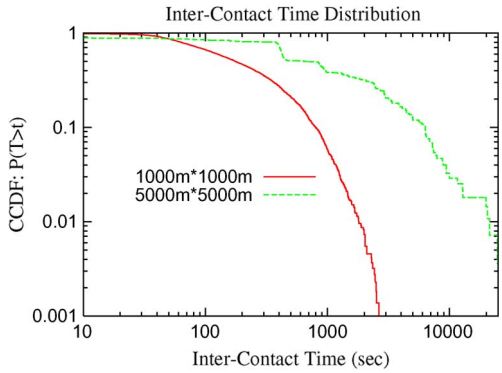


Fig. 10. Intercontact time under TSC mobility.

$\{p_i(c_m), 1 \le m \le 5\}$ and sojourn time W_i. Then, node i selects one of the M communities as its next target with likelihood $p_i(c_m)$ and randomly chooses a destination point around the target cave. The node moves to its destination using smooth movement that first speeds up, then moves at stable speed, and finally slows down before coming to a stop. The speed of smooth movement is proportional to the distance between the starting point and the destination [36]. When node i reaches its destination at c_m, it stays there for $w_i(c_m)$ period of time with pauses or small movements around c_m. Then, node i repeats this process again. All N nodes in the network continue their movements until the end of simulation.

Before using TSC mobility, we run simulations to make sure that TSC mobility exhibits characteristics of node mobility observed in empirical traces, which is the TPL decay of intercontact time [37]. Simulation runs for 24 h with 20 nodes moving in a five-community network area, as shown in Fig. 9. The transmission range is set as 250 m. Fig. 10 shows the complementary cumulative distribution function of the intercontact time, i.e., $\mathcal{P}\{T_I > t\}$, on a log–log scale with simulation area ranging from 1000 m $\times$ 1000 m to 5000 m $\times$ 5000 m. Consistent with the studies in [37], the intercontact time follows a TPL distribution, and for the 5000 m $\times$ 5000 m simulation area, the power-law behavior is dominant over up to $O(10^4)$ s, which is followed by a sharp decrease beyond that time scale.

Remark 6: The TSC constrains node movement on lines between caves, which reflects the road-topology-constrained car movements. Moreover, the TSC mobility model not only mimics skewed location preferences of human mobility but also yields power law and exponential decay dichotomy of intercontact time. Therefore, we use TSC to evaluate the DLR.

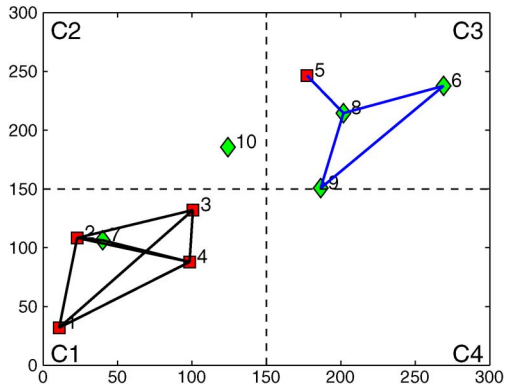


Fig. 11. Nodes with similar mobility patterns are identified as groups.

2) Group Identification Under TSC Mobility: Here, we first study group identification in a simple four-cave network and then in a general network. As a simple scenario, the network is partitioned into four caves $\{C1, C2, C3, C4\}$, with ten nodes moving for 48 h according to TSC mobility. Nodes update their mobility observations every 10 min. We introduce two mobility patterns: $O(0)$ cave profile for $\{n_1, \ldots, n_5\}$ is $\{0.85, 0.05, 0.05, 0.05\}$, whereas that for $\{n_6, \ldots, n_{10}\}$ is $\{0.05, 0.05, 0.85, 0.05\}$.

Fig. 11 shows that DLR identifies two groups, in which $\alpha = 0.5$ and $\text{DLR}_{\text{th}} = 0.5$ are used. *Group* 1 in $C1$ includes nodes $\{n_1, n_2, n_3, n_4, n_7\}$, and *group* 2 in $C3$ includes nodes $\{n_6, n_8, n_9\}$. Notice that node n_7 is loosely connected to other nodes in *group* 1 because it has a very different mobility pattern (i.e., cave file) from nodes in *group* 1, e.g., it has a higher preference of *cave* $C3$, whereas the nodes in *group* 1 spend most of their time in *cave* $C1$. The connections between node n_7 and n_2 in *group* 1 are due to correlation in spatial locality (i.e., adjacent locations). Node n_7 is disconnected to nodes n_1, n_3, and n_4 in *group* 1 because of different temporal and spatial localities. Node n_5 does not belong to *group* 2 because it neither has a similar mobility pattern nor locate closely to nodes n_6, n_8, and n_9 in *group* 2.

Remark 7: When node mobility show correlations in both spatial and temporal localities, DLR can identify user groups in which nodes either have similar mobility patterns or are closely located. Nodes with correlations in both spatial and temporal localities form *strongly connected components*.

3) Group Lifetime: To further show the benefits of DLR in group identification, we turn to evaluate the stability of DLR-identified groups. To proceed, we define the group lifetime.

Definition: Strict group lifetime is defined as follows. Let G be an identified node group and t_0 be the first time when it emerges. The strict group lifetime of G is the consecutive time intervals, in which G maintains the same set of group members.

According to this definition, group G dies when there is a node that joins in or there is a member that leaves (i.e., node switches between groups). In other words, strict group lifetime is the duration that a group survives without nodes switching in or out of the group. The longer the strict group lifetime is, the

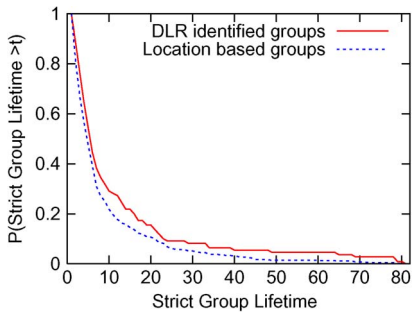


Fig. 12. Strict group lifetime. DLR-identified versus location-based groups.

more stable the group is and the stronger mobility correlations among group members are.

In the simulations, nodes are moving in a six-cave network under the TSC mobility. The caves are simply obtained by partitioning a network area into two-row and three-column grids. As cave profiles are generated according to TPL distribution, the nodes have different cave profiles, which not only exhibit the skewed location preferences but also produce different levels of mobility correlations in temporal locality. For instance, a node with the cave profile (0.4, 0.35, 0.125, 0.125, 0, 0) will visit caves c_0 and c_1 more frequently and stay there for most of its time and occasionally visit caves c_2 and c_3. Nodes with the same type of a cave profile have very high TLS.

Node groups are identified based on: 1) mobility correlation (i.e., DLR value) that neighboring nodes with DLR $>$ 0.5 would be clustered as in a group; and 2) node location that nodes located in the same cave belong to a group. We compute the strict group lifetimes of DLR-identified groups and location-based groups and show the probabilities $P(\text{strict group lifetime} > t)$ in Fig. 12.

Fig. 12 shows that node groups identified based on DLR tend to maintain longer lifetime than groups clustered based on node locations. In other words, DLR-identified groups, in which nodes have high mobility correlations, are more stable than location-based groups. Therefore, DLR can be used to effectively identify stable groups.

4) Impact of α and DLR$_{\text{th}}$: Here, we provide guidelines on how to choose appropriate α and DLR$_{\text{th}}$.

α is an important parameter in DLR, which can adjust the weights of SLS and TLS. On one hand, if mobile nodes (such as buses) have relatively stable movement habits, DLR with large α tends to detect long-term communities of nodes with similar temporal locality. Thus, *large* α is preferred for networks with well-defined *communities*. On the other hand, DLR with high weight of SLS tends to cluster nodes into closely moving groups. Without taking into account SLS ($\alpha = 1$), the group structure may be unable to capture the dynamics of node movements. Therefore, for a network with high or random node mobility, such as a cab-based ad hoc network, *smaller* α may be preferred to capture network *dynamics*.

Although DLR$_{\text{th}}$ is an important parameter in identifying node groups, a universal DLR$_{\text{th}}$ may be hard to determine due to the dependence of group structures on *node mobility* and *application requirements*. An easy way to set DLR$_{\text{th}}$ is averaging DLRs between a node and its neighbors because nodes in a group are more closely related to each other. Another possible way to choose appropriate DLR$_{\text{th}}$ is adaptively changing DLR$_{\text{th}}$ through learning the stability of identified groups. If nodes move in or out the group frequently, increasing DLR$_{\text{th}}$ can help to identify more stable groups. To find a suitable DLR$_{\text{th}}$, application requirements also need to be considered. For reliable applications, such as sending control messages, we suggest using large DLR$_{\text{th}}$ to ensure the stability of links and group structures; otherwise, small DLR$_{\text{th}}$ is feasible.

VI. Dual-Locality Ratio-Based Applications

The internode correlation can be helpful in many applications in addition to detecting network components. For example, in mobility-aware routing, two nodes with a high DLR move closely and probably maintain a stable link; thus, they are more suitable for establishing routes with high stability. In data forwarding of DTNs, if a node currently carrying a message sends a copy to a node with different mobility patterns (i.e., low DLR), the chance of at least one of them meeting the destination could be increased. In mobility-aware clustering, as the node with high mobility correlations with its neighbors is much less likely to be disconnected from its neighbors, therefore, communication overhead for changing cluster head can be reduced if it is selected as the cluster head. Here, we present two applications to show the benefits of the DLR in assisting data forwarding and clustering.

A. DLR-Assisted Data Forwarding

As an application, we investigate how DLR can be used to enhance performance of data forwarding. We develop a DLR-assisted data-forwarding mechanism for information dissemination in DTNs. The relays are selected among the contacted neighbors of a message carrier based on their mobility correlation to increase delivery ratio. DLR-assisted data forwarding is compared with simple random forwarding that the source randomly chooses $k-1$ relays. This eliminates other factors' impact on network performance in more complicated routing schemes, thus highlighting the impact of DLR.

Specifically, we assume that neighboring nodes can exchange their locations, speeds, and cave profiles. Suppose that the neighbors of data source S are $N_S = \{n_1, \ldots, n_l\}$, which are relay candidates. We assume that D is not in the set N_S; otherwise, S can simply transmit the data to D immediately. A relay is selected according to the following steps.

1) S collects the spatial and temporal localities of its neighbors N_S by exchanging messages.
2) S calculates the mobility correlation DLR$_{i,S}$ between S and its neighbor n_i. S selects relay node R_1 with DLR$_{S,R_1} \leq$ DLR$_{\text{th}}$ and sends a message copy to R_1.
3) S repeats step 2 and sends message copies to $k-1$ distinct relays. The message is forwarded to D by the relay node set $R = \{S, R_1, R_2, \ldots, R_{k-1}\}$.

We implement DLR-assisted data forwarding in OMNeT ++ [35]. Fifty nodes with density 8 nodes/km^2 move according to TSC mobility. Node speed is within 10 to 20 m/s. Simulation update interval is 1 s. There are 25 caves in the network. The

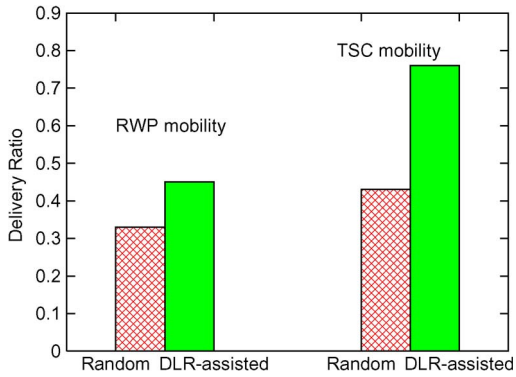


Fig. 13. Delivery ratio. SCF versus random forwarding.

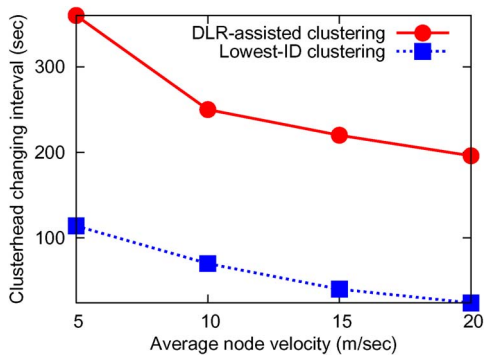


Fig. 14. DLR assists clustering under TSC mobility.

carrier frequency is 2.4 GHz. The medium-access-control-layer bit rate is 2 Mbps. Transmission power is 2 mW and the signal attenuation threshold is -25 dBm. The fading model is a path-loss model with a coefficient of 2. Then, the node transmission range is approximately 250 m. The performance metric is the delivery ratio, which is the percentage of destinations that receive messages from source nodes within 30 min.

Fig. 13 shows the delivery ratio of DLR-assisted data forwarding compared with random forwarding with a varying number of relays and $\mathrm{DLR_{th}}$. Clearly, the more relays are used, the higher the delivery ratio is. DLR-assisted data forwarding with $\mathrm{DLR_{th}} = 1$ is the same as random forwarding; thus, they have the same performance (red curve in Fig. 13). When $\mathrm{DLR_{th}} = 0.5$, DLR-assisted data forwarding achieves better performance than random forwarding because DLR is based on mobility similarity and a relay is selected when it has relatively different mobility pattern with the message carrier, which increases the probability of at least one relay meets the destination. Interestingly, further reducing $\mathrm{DLR_{th}}$ to 0.2 degrades DLR-assisted forwarding performance as relay nodes become hard to find. This indicates that there exists an optimal $\mathrm{DLR_{th}}$ producing best delivery ratio performance. Using DLR introduces overhead because of spatial and temporal locality information updates among nodes. To reduce overhead, we can use DLR reactively and only allow spatial and temporal locality information exchange between relay nodes and their neighbors. Suppose that the update frequencies of spatial and temporal locality messages are f_s and f_t, respectively. For example, under the given simulation scenario, $f_s = r/2v_{\max}$ (i.e., 1 message per 6 s), and $f_t = 1$ message per 10 min. The average number of spatial locality messages transmitted is lower bounded by $(k-1)$ and upper bounded by $\tau f_s \pi r^2 \lambda$, where k is the number of relay nodes, τ is the message delay tolerance, r is node's transmission range, and λ is the node density. The number of temporal locality messages transmitted is lower bounded by $k-1$ and upper bounded by $\tau f_t(N-2)$, where the total number of nodes $N = 50$. Hence, the total number of packets associated with DLR is lower bounded by $2(k-1)$ and upper bounded by $\tau f_s \pi r^2 \lambda + \tau f_t(N-2)$. Here, we briefly discuss the application of DLR in data forwarding. A more sophisticated DLR-based routing algorithm can be designed to achieve better network performance without too much overhead.

B. DLR-Assisted Clustering

We further utilize DLR in mobility-aware clustering, which is one of the most general applications of topology control, and show its benefit of lower cluster-head changing rate compared with the lowest-ID algorithm [38].

In the lowest-ID clustering algorithm, each node is randomly assigned an ID, and the node with the smallest ID among its neighbors acts as the cluster head. Because nodes may move at a high speed and travel different routes to their destinations, a node may be frequently disconnected with its randomly selected cluster head, i.e., clusters based on the lowest ID are unstable. The stability of clusters could be improved by selecting a cluster head based on average DLRs of nodes with their neighbors. Hence, we implement a DLR-assisted clustering algorithm that selects the node with the highest average DLR over DLRs between it and its neighbors as the cluster head.

Using the same simulation setting as in Section V-B1, the average cluster-head changing time is measured to indicate cluster stability. Fig. 14 shows that under various node speeds, changing the interval of the cluster head selected base on DLR is much longer than that based on the lowest ID. A node with a higher average DLR with its neighbors not only moves closely with its neighbors but also may share overlapping paths; therefore, it is much less likely to be disconnected from its cluster members if selected as the cluster head, i.e., the changing rate of the cluster head is reduced. Therefore, DLR can be used to obtain stable clustering for topology management in MANETs.

VII. Conclusion

In this paper, we have studied the internode mobility correlation in VANETs, which aims to fill the gap between the studies on *individual* mobility and performance evaluation based on the properties of a *group* of nodes. Spatial and temporal localities are observed in two vehicle traces with complementary properties. The SFCAB trace features diversified mobility and destination and the STBUS features arranged routes and schedules. We introduced a new metric, DLR, to quantify the internode correlation in spatial and temporal localities. To explore the applications of this metric, we have demonstrated how to use DLR to identify *stable groups* and to assist data forwarding and clustering.

REFERENCES

[1] Y. Li, M. Zhao, and W. Wang, "Intermittently connected vehicle-to-vehicle networks: Detection and analysis," in *Proc. IEEE GLOBECOM*, 2011, pp. 1–6.

[2] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. Mobile Comput., Special Issue Mobile Ad Hoc Netw., Res., Trends Appl.*, vol. 2, no. 5, pp. 483–502, Aug. 2002.

[3] F. Bai, N. Sadagopan, and A. Helmy, "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proc. IEEE INFOCOM*, 2003, pp. 825–835.

[4] J.-L. Huang and M.-S. Chen, "On the effect of group mobility to data replication in ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 5, pp. 492–507, May 2006.

[5] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad-hoc networks," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[6] Y. Li, W. Wang, and A. Duel-Hallen, "The latency of gaining -reliability for message dissemination in vehicle-to-vehicle networks," in *Proc. IEEE GLOBECOM*, 2012, pp. 5550–5555.

[7] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility," in *Proc. IEEE PerCom Workshops*, 2007, pp. 79–85.

[8] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. ACM MobiHoc*, 2008, pp. 241–250.

[9] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. ACM MobiSys*, 2003, pp. 303–316.

[10] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Comput. Netw.*, 2008, pp. 187–201.

[11] M. Thomas, S. Phand, and A. Gupta, "Using group structures for efficient routing in delay tolerant networks," *Ad Hoc Netw.*, vol. 7, no. 2, pp. 344–362, Mar. 2009.

[12] K. H. Wang and B. Li, "Group mobility and partition prediction in wireless ad-hoc networks," in *Proc. IEEE ICC*, 2002, pp. 1017–1021.

[13] J. M. Ng and Y. Zhang, "Impact of group mobility on ad hoc networks routing protocols," in *Proc. 8th ICACT*, 2006, pp. 1134–1139.

[14] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. ACM MobiHoc*, 2009, pp. 299–308.

[15] X. Hong, M. Gerla, G. Pei, and C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proc. ACM MSWiM*, 1999, pp. 53–60.

[16] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, Apr. 2007.

[17] M. C. Gonzalezl, C. A. Hidalgo1, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature Lett.*, vol. 453, no. 7196, pp. 779–782, Jun. 2008.

[18] V. Naumov, R. Baumann, and T. Gross, "An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces," in *Proc. ACM MobiHoc*, 2006, pp. 108–119.

[19] P. Hui and J. Crowcroft, "Human mobility models and opportunistic communications system design," *Phil. Trans. R. Soc. A, Math. Phys. Eng. Sci.*, vol. 366, no. 1872, pp. 2005–2016, Jun. 2008.

[20] A. Nandan, S. Tewari, S. Das, M. Gerla, and L. Kleinrock, "Adtorrent: Delivering location cognizant advertisements to car networks," in *Proc. IEEE/IFIP WONS*, 2006, pp. 1–10.

[21] S. Heimlicher and K. Salamatian, "Globs in the primordial soup the emergence of connected crowds in mobile wireless networks," in *Proc. ACM MobiHoc*, 2010, pp. 91–100.

[22] M. Piorkowski, N. S. Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *Proc. 1st Int. Conf. Commun. Syst. Netw.*, 2009, pp. 1–10.

[23] J. G. Jetcheva, Y.-C. Hu, S. PalChaudhuri, A. K. Saha, and D. B. Johnson, "Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture," in *Proc. 5th IEEE WMCSA*, 2003, pp. 32–43.

[24] K. Wang and B. Li, "Efficient and guaranteed service coverage in partitionable mobile ad-hoc networks," in *Proc. IEEE INFOCOM*, 2002, pp. 1089–1098.

[25] B. Gu and X. Hong, "Mobility identification and clustering in sparse mobile networks," in *Proc. IEEE MILCOM Conf.*, 2009, pp. 1–7.

[26] P. Hui, E. Yoneki, S.-Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proc. Sigcomm Workshop MobiArch*, 2007, pp. 1–8.

[27] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 5th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[28] D. Lelescu, U. Kozat, R. Jain, and M. Balakrishnan, "Model T ++: An empirical joint space-time registration model," in *Proc. ACM MobiHoc*, 2006, pp. 61–72.

[29] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the levy-walk nature of human mobility," in *Proc. IEEE INFOCOM*, 2008, pp. 924–932.

[30] C. Song, N. B. Z. Qu, and A.-L. Barabási, "Limits of predictability in human mobility," *Sci. Mag.*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.

[31] T. A. Alhmiedat and S.-H. Yang, "A survey: Localization and tracking mobile targets through wireless sensors network," in *Proc. PGNet Int. Conf.*, 2007, pp. 1–6.

[32] L. Song, D. Kotz, and R. Jain, "Evaluating location predictors with extensive WI-FI mobility data," in *Proc. IEEE INFOCOM*, 2004, pp. 1414–1424.

[33] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *J. Math. Models Methods Appl. Sci.*, vol. 1, no. 4, pp. 300–307, 2007.

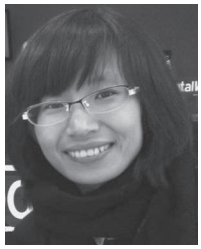
[34] I. Stojmenovi, "Location updates for efficient routing in ad hoc networks," in *Handbook of Wireless Networks and Mobile Computing*. New York, NY, USA: Wiley, 2002, pp. 451–471.

[35] A. Varga, Omnet ++ discrete event simulation system. [Online]. Available: http://www.omnetpp.org/

[36] A. Mei and J. Stefa, "Swim: A simple model to generate small mobile worlds," in *Proc. IEEE INFOCOM*, 2009, pp. 2106–2113.

[37] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law inter-meeting time in MANET," in *Proc. ACM MobiCom*, 2007, pp. 159–170.

[38] A. Ephremides, J. Wieselthier, and D. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proc. IEEE*, vol. 75, no. 1, pp. 56–73, Jan. 1987.



Yujin Li (S'11) received the B.S. and M.S. degrees in control science and engineering from Beijing Institute of Technology, Beijing, China, in 2007 and 2009, respectively. She is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA.

Her research interests include mobility modeling and management in wireless networks.

Ming Zhao (M'08) received the B.S. degree in electrical engineering from Harbin Institute of Technology, Harbin, China, in 1997; the M.S.E.E. degree from New York State University at Buffalo, Buffalo, NY, USA, in 2004; and the Ph.D. degree from North Carolina State University, Raleigh, NC, USA, in 2009.

Since 2008, he has been a Software Engineer with Cisco Systems, Research Triangle Park Campus, NC, USA. His research interests include mobility modeling, user mobility, network performance and management in wireless networks, and video multicast routing in Internet Protocol networks.

Wenye Wang (SM'12) received the B.S. and M.S. degrees from Beijing University of Posts and Telecommunications, Beijing, China, and the M.S.E.E. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, GA, USA, in 1999 and 2002, respectively.

She is an Associate Professor with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA. Her research interests include mobile and secure computing, network topology and architecture, and smart grids.

Dr. Wang received the National Science Foundation CAREER Award in 2006.