This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

Cooperative Caching in Vehicular Content Centric Network based on Social Attributes and Mobility

Lin Yao, Yuqi Wang, Xin Wang and Guowei Wu

Abstract—Communications in vehicular ad-hoc network (VANET) are subject to performance degradation as results of channel fading and intermittent network connectivity. The emerging Vehicular Content Centric Network (VCCN) is promising in supporting the needs of contents and alleviating the communication problems in VANET. Specifically, to improve the cache hit ratio and reduce the access delay of content retrieval, it helps to choose the appropriate vehicles to cache the frequently accessed data items. In this paper, we propose a Cooperative Caching scheme based on Social Attributes and Mobility Prediction (CCSAMP) for VCCN. CCSAMP is based on the observation that vehicles move around and are liable to contact each other according to drivers' common interests or social similarities. A caching node sharing more social attributes with the content requester is more likely to be interested in the same contents and distribute the contents to others with similar interests. Furthermore, a caching node that frequently meets other nodes is a better candidate to keep cache copies. To increase the network performance, CCSAMP also exploits the regularity of vehicle moving behaviors to predict the chance for a vehicle to reach hot zones based on Hidden Markov Model (HMM). We evaluate CCSAMP through the ONE simulator to demonstrate its higher cache hit ratio and lower content access delay compared to other state-of-the-art schemes.

Index Terms—Social Similarity, Mobility Prediction, Cooperative Caching, VCCN

1 Introduction

Vehicular Ad-hoc Network (VANET), a special type of mobile ad hoc network, is formed with a set of moving vehicles equipped with communication facilities and Road Side Units (RSUs). VANET is often subject to frequent communication disruption as results of harsh propagation environment [1] and the mobility of vehicles thus the dynamic network topologies. The difficulty of maintaining end-to-end connections in VANET causes the performance degradation of data dissemination. Besides the communication challenge, the conventional host-oriented Internet transmission format can not support the application needs of VANET. Vehicular applications are normally information-oriented, and the identity of the content provider is not important. In addition, as the store-and-forward routing format is often taken by VANET to combat the constant disruption of communication links in VANET, it would be more efficient to retrieve contents from the most convenient providers to reduce the data latency and network traffic.

To better support the application needs, the Vehicular Content Centric Network (VCCN) is proposed to offer commercial and entertainment services to drivers and passengers by applying the CCN model to the vehicular environment [2]. The focus of CCN is on contents but not the actual carriers of the contents. Rather than taking the traditional address-centric communication format, CCN takes the content-centric transmission format, where a user retrieves a given content directly using the 'name' without caring about the identity or IP address of the content provider [3]. More specifically, each mobile user requests the desired content by sending an Interest packet with the content name inserted. If the content can be provided by an intermediate node, it will return the content in a Data packet without forwarding the Interest further; Otherwise, the Interest will be forwarded until reaching a source provider or the limit of communication hop-count. Different from the push-based communication model often used in conventionally VANET, the use of pull-based information retrieval in VCCN (Fig. 1) makes it attractive to exploit in-network caching. Particularly, VCCN can benefit from cooperative caching, taking advantage of the cooperation among the caching nodes to achieve a higher cache hit ratio, lower content access latency and less network traffic.

Depending on where a content chunk is cached, cooperative caching mechanisms can be classified into on-path caching and off-path caching [4]. In the on-path caching, a content chunk is cached at some nodes along the reverse path from the requester to the provider. In the off-path caching, a content chunk is cached at some critical nodes of the network. No matter which type of caching format is taken, cooperative techniques proposed for CCN cannot be directly employed in VCCN, because the topology of VANET is characterized with dynamics as a result of its short-lived communication links.

- *Lin Yao, Yuqi Wang and Guowei Wu are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China. Lin Yao is also with the International School of Information Science & Engineering, Dalian University of Technology, China. Yuqi Wang and Guowei Wu are also with the School of Software, Dalian University of Technology, China.(e-mail: yaolin@dlut.edu.cn)*
- *Xin Wang is with the Department of Electrical and Computer Engineering, Stony Brook University, New York State, U.S.A. (e-mail: x.wang@stonybrook.edu).*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

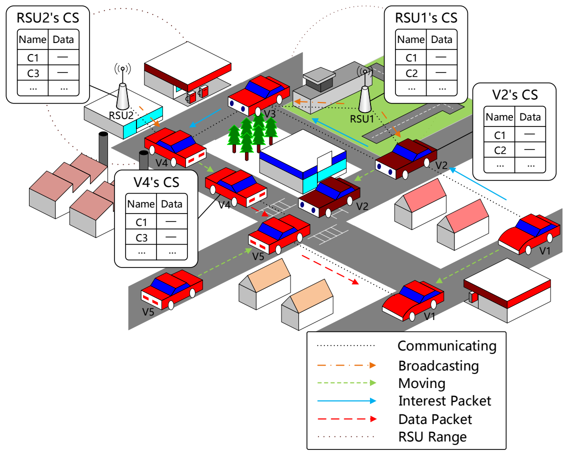


Fig. 1: VCCN Network Model

Recently, social interactions among vehicles are considered in VANET to support better vehicular communications for safety and entertainment applications [5]. People with more common interests such as a group of enthusiasts in the same club may go to attend similar events and have a higher probability of encountering each other. A node with a higher frequency of encountering other vehicles can provide more connectivity links for the network, and is considered as an important component in the network. Since people's social relationship generally has long-term characteristics and is less volatile than their mobility [6], we expect that the stability of social relationship can be exploited to better support content dissemination in VANET. As some example applications, a fan can get a new pop song of Taylor from other enthusiasts by sending a request, and a driver can request a detailed map and road conditions of an area on the path. Our analysis of the vehicular data traces of cambridge/haggle confirms that the encounter frequency increases with the growing number of common social features in [7]. Therefore, we introduce social relationship of drivers into the design of a cooperative caching scheme for VCCN.

The communication in VANET is influenced by mobility patterns of vehicles, which are impacted by both driver behaviors and driving paths [8]. Vehicles often have regular visits of places such as shopping malls over the weekend and workplaces during a weekday, and thus have regular contacts during daily activities. Numerical analyses in [8] have proved that traffic patterns can provide social interactions. As an instance, an area with high vehicle density (e.g. around the shopping mall) is a popular social place for drivers to get contents. Consequently, it is very helpful to consider traffic patterns when determining the caching nodes. By exploiting the regularity of mobility, we can more effectively predict the future location of a vehicle based on its past trace.

Based on above discussions, we consider a node to be a good candidate for caching contents in VANET if it has the following features: a higher social similarity with the requesting node, a bigger bridging centrality and a tendency of visiting hot zones which possess higher vehicle density. By sharing social features with the requester, the candidate caching node is likely to be interested in the same contents. In addition, cached contents are easier to be accessed by other vehicles for three reasons: 1) Request flows are likely to reach the node with a bigger bridging centrality, 2) The node which has higher social similarity with the content requester is likely to meet others which share the common interests, and 3)The candidate node has higher frequency in visiting hot zones. Although social features have been considered in facilitating caching in cellular networks and delay tolerant networks, the social features in VANET are unique and impacted by all above factors along with vehicle mobility patterns. It calls for a method to coherently integrate different factors into the effective design of cooperative caching strategy. In this paper, we propose a Cooperative Caching scheme based on Social Attributes and Mobility Prediction (CCSAMP) for VCCN. The main contributions of our paper are listed as follows:

(1) We propose a novel caching strategy with the integrated consideration of social properties of vehicles as well as the vehicular mobility and traffic patterns. To the best of our knowledge, we are the first to incorporate both social attributes and hot-zone visiting probability into the design of a cooperative caching scheme for VCCN.
(2) We further propose a cache replacement policy that evaluates the content popularity by combining the social similarity and time interval of two consecutive requests.
(3) We evaluate the performance of our schemes with comprehensive trace-driven simulations. Compared with other caching schemes, **DPC** [9], **LDCC** [10] and **DAC** [11], our **CCSAMP** has a superior performance with high successful ratio in finding contents in cache and low content access delay.

This work focuses on the design of cache management schemes in vehicular content centric networks. The information flows naturally follow the content retrieval paths, which is different from conventional packet forwarding schemes where each node needs to make sophisticated forwarding decision. The potential of performance improvement with different forwarding strategies is beyond the scope of this paper.

The remainder of this paper is organized as follows. In Section 2, we discuss the related work. Network model and problem statement are given in Section 3. In Section 4, we present the details of our approach. We evaluate the performance of CCSAMP in Section 5, and conclude the work in Section 6.

2 Related Work

In this section, we review some literature work related to our proposed method.

CCN- Based on whether a content chunk is cached along its delivery path, cooperative caching schemes in the CCN can be classified into on-path caching and off-path caching [4]. Off-path caching aims to replicate a content within a network in order to increase its availability regardless of the delivery path taken [12], while on-path caching aims to reduce the network traffic and the delay in the CCN by caching the content along the path from the provider to the requester. Although off-path caching is easy to be provided as an overlay by the third party, only a small number of off-path caching solutions are proposed [13] [14].

As contents can be easily cached along the transmission path, on-line caching has attracted a lot of research attentions. Targeting for solving the web caching problem, Breadcrumbs [15] considers that each router in the network has a local cache to store files passing by. As a content router does not inform its cached contents to its neighbors, a request may be forwarded multiple times between two neighboring routers. To reduce the bandwidth loss and content access latency, each content router in [16] periodically exchanges its local cache summaries with its one-hop neighbors. In [17], a content router advertises its available contents to its k-hop neighbors. In [18] [19], a central content router is responsible for maintaining all the cache states in the network. Rather than advertising the cache states to other content routers, Psaras et al. proposed a probabilistic caching scheme where the capability of each router is estimated and only popular contents are cached [20] [21]. In WAVE [22], the downstream routers determine which chunk should be cached based on the content popularity recommended by the upstream routers. In [23], a cooperative caching strategy is designed to handle large video streams by combining the traditional hash-based and directory-based cooperative caching schemes. In [9], caching decisions are made by each node independently with the consideration of three factors: users' demands mined from the collected Interests, relative movement of the receiver and the sender, and the importance of vehicles based on degree centrality and betweenness centrality.

VANET, MANET, DTN- To improve the data access opportunity and reduce the overhead caused by the global network flooding, VANET caching schemes mainly rely on the cooperative approach with a group of nodes caching some contents together [24]. A caching scheme for mobile ad hoc network (MANET) was first proposed in [25], where the cross-layer design is exploited to improve the caching performance with the cooperative caching. To reduce the hop count and response latency, a resource efficient caching scheme was proposed to distribute the data items among nodes according to the data requirements within the network [26]. In [27], a fuzzy hybrid caching scheme for MANET was proposed to minimize the duplicated caching of data between neighbors and improve the network performance based on the utility and the access similarity of data items. In [28], to balance data accessibility and caching overhead, the nodes located at some key locations are chosen for caching in the Delay Tolerant Network (DTN). In [29], CLIR was proposed to improve the retrieval of information in MANET by maintaining the locations of the documents and distributing the requests along the network. Suno et al. [30] proposed a cooperative caching invalidation scheme along with its enhancement for VANET, in which an invalidation report on some contents is sent to home agents by a server and then the report is sent to the gateway agents by home agents. Instead of blindly broadcasting it to all vehicles, the gateway agents are responsible for answering the validity of the requested data to reduce the query delay. Similarly, the gateways in different regions also cooperate to maintain the invalidate contents so as to reduce the query delay for urban vehicles in [31]. In [10], the mobility of vehicles is considered, and ones with a higher probability of staying within an area will be selected as caching nodes in a small time duration. In [32], the Network Central Location Cooperative Caching (NCLCC) scheme identifies several network central locations to cache popular contents. In [33], Khawaga et al. proposed an administrative cluster-Based cooperative caching scheme. The cluster heads are in charge of maintaining the cluster cache information. A backup node is used to enhance the data availability within the cluster, reduce the delay and improve the bandwidth utilization. In [11] [34], social attributes such as contact patterns and relationship are used to choose caching nodes. In [35], a centralized base station estimates the popularity based on the requests observed and applies it to control the caching probability. In [36] [37] [38] [39], authors proposed to perform proactive caching based on the information extracted by a base station from users social interactions over social network overlay leveraging Device-To-Device (D2D) communications. The use of central control to facilitate information abstraction and transmission is very different from VANET communications. User social interactions have also been exploited in delay-tolerant network [40] with D2D communications. Data disseminations and social relationship in vehicular networks are impacted by driving paths and driver behaviors, which are different from other types of networks. A caching strategy needs to fully exploit vehicle network features.

Summary- To improve the cache hit ratio and access delay, the critical challenge is to select appropriate caching nodes. Though the mobility of nodes has been considered in some caching schemes, none has taken a full advantage of trajectory records to predict the future moving paths of vehicles. Furthermore, the social attributes are not considered in selecting the appropriate caching nodes in VANET. Compared with the caching schemes in the literature, we apply the CCN model into the vehicular environment. Based on the observation that the frequency of encounters among mobile users increases with the growing number of common social features in [7], we consider the impact of social attributes when choosing a cache node. Furthermore, we take into account the future trajectory of each vehicle. A vehicle which is likely to go to a hot zone and possess a higher centrality will be chosen as the caching node. In summary, our proposed CCSAMP concurrently considers three major factors in the design of an effective caching strategy for VCCN: *social similarity to exploit the social relationship among nodes, bridging centrality to select nodes important for information dissemination, and probability of hotzone visiting based on the path predicted with the trajectory records*.

3 Network Model and Problem

3.1 Network Model

VCCN has emerged as a future network technology for VANET. Different from the push-based communication model in the conventional VANET, VCCN usually takes the pull-based approach. The conventional TCP/IP layer is replaced by CCN at the network layer [2] in Fig. 2, and the communication between vehicles is shifted from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

host centric to the information centric. Without carrying the source and destination addresses, Interests are directly forwarded to their reachable neighbors. Interest packets are stored, carried and forwarded by vehicles or RSUs until their time-to-live (TTL) timers expire or they reach a cache node which possesses the content. As illustrated in Fig. 1, VCCN consists of RSUs and vehicles, where vehicles can carry and relay data and RSUs are responsible for collecting the trajectory data of vehicles and pushing new contents to vehicles.

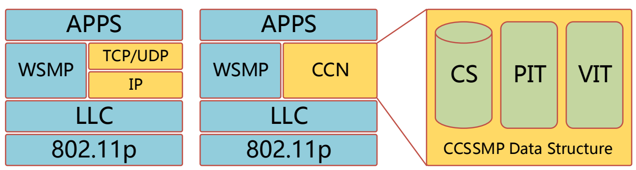


Fig. 2: Applying CCN in VANET

Each vehicle in our paper maintains three data structures shown in Table 1, Content Store (CS) to keep a record of each cached content along with its name, Pending Interest Table (PIT) to keep track of forwarded Interests, and Vehicle Information Table (VIT) to store the vehicle's trajectory data and social attributes. Each RSU only maintains a content store.

TABLE 1: Three Data Structures

(a) Content Store

Notation	Description
$name$	Content name
$content$	Cached content

(b) Pending Interest Table

Notation	Description
$name$	Content name
$time$	Time of the last request

(c) Vehicle Information Table

Notation	Description
VID	Vehicle's ID
SA	Social attribute sequence of the vehicle
$\mathcal{L}$	Trip sequence of the vehicle

When a vehicle receives an Interest, it will search for the content name in its CS, and return the corresponding content if cached. If the content is not in the cache, the vehicle will add the Interest into its PIT; If the content is in the cache, it will update the receiving time of the stored Interest to that of the new request and then discard the Interest to avoid the duplicate storage. When receiving a Data chunk, each vehicle will make its own decision on whether to cache it or not based on a caching scheme (such as CCSAMP proposed in this work). Thus, a given content may be provided by multiple nodes, rather than being cached on a particular one in VCCN. The reply path and query path are not necessarily the same because of the dynamic network topology shown in Fig.1.

3.2 Problem and Design Considerations

As contents are queried and delivered by users in vehicles, our algorithm is applied to determine which vehicles should cache the contents. In VCCN, as vehicles have different moving trajectories, it is important to properly choose nodes to cache contents. For example, a vehicle which tends to drive towards an area with sparse vehicle distribution is not a good option for being a caching node, because it may encounter fewer nodes. Similarly, a node with less connectivity with others is not suitable either, because it may be isolated from remaining vehicles. In VCCN, vehicles tend to have more contacts if their drivers have some common interests, while a driver has a lower probability of sharing the hobby with members in a different community. Given the challenge in selecting caching nodes in VCCN, we would like to answer the following questions in this paper: 1) How to select caching nodes taking advantage of the regularity of moving patterns of vehicles to improve the cache hit ratio and reduce the access delay of the requested contents? 2) How to exploit the social attributes to select the caching nodes for more efficient cache usage?

Our algorithm is used by a candidate caching node to determine what types of contents to cache. However, not every node in the network needs to serve as caching node. A vehicle can decide whether to act as a caching node based on different factors, such as its buffer space and its interest. Some nodes may be configured as caching nodes by the system, such as some public transportation vehicles. Without caching contents, a node can still help forward Interests or Data packets. The providing of policy to prescribe which nodes can serve as caching nodes is beyond the focus of this paper.

4 Cooperative Caching based on Social Attributes and Mobility Prediction

CCSAMP scheme chooses suitable caching nodes based on three metrics, social similarity, bridging centrality, and future trajectory. We first give an overview of CCSAMP, and then present our algorithm on how to compute these metrics. The frequently used notations are listed in Table 2.

TABLE 2: Frequent Notations

Notation	**Description**
v_i	The i-th vehicle
SA_i	The social attribute sequence of v_i
w_i	The weight of the i-th social attribute.
n_{e_i}	The number of encounters of v_i
n_{l_i}	The number of links not through v_i
n_v	The number of vehicles in the network

4.1 Overview

As illustrated in Fig. 3, CCSAMP includes three major modules to determine social similarity and bridging centrality, and to predict the trajectory. The social similarity is compared between the forwarding node and the requesting node, the bridging centrality determines the linkage of a vehicle with other nodes, and the future trajectory of a vehicle is predicted based on its historical mobility pattern and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

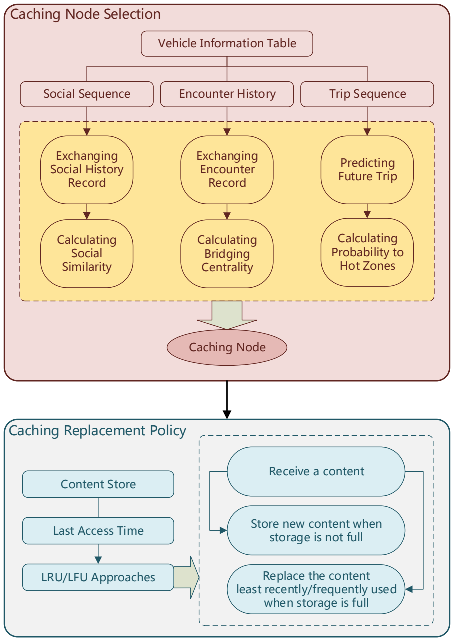


Fig. 3: Architecture of our CCSAMP

current link information. To facilitate the finding of social similarity, each Interest contains the social features of the requesting node, and each Data packet contains the social features of the corresponding Interest. In the content based network, node identity is not included with the packet, which helps to protect the privacy of each node.

CCSAMP works with the following procedures:

1) Each vehicle uploads its trajectory records to the nearest RSU within its communication range periodically. RSUs synchronize all the records to determine the hot zones that have a higher vehicle density.
2) A vehicle sends a beacon frame containing its social attributes periodically to neighbors around or upon encountering new vehicles. The receiver updates its social history record to track the number of encounters that own a specific social attribute and then discards the message.
3) Upon receiving an Interest packet, a vehicle searches its CS for a hit. If the requested content is found inside its cache, the node will encapsulate the content into a Data packet to send towards the requesting node; Otherwise, the vehicle will forward the Interest to its neighbors and update its PIT.
4) After receiving a Data packet, a vehicle calculates the *social similarity* based on the source's social attributes embedded in the Data packet and *bridging centrality* with its social history record, and applies Hidden Markov Model (HMM) to predict its future trajectory based on its mobility trace to determine whether it can arrive at a hot zone.
5) The node will opt to cache the content if it intends to go to a hot zone and also has higher social similarity and bridging centrality.

4.2 Caching Node Selection

CCSAMP chooses the caching nodes based on three factors: social similarity, bridging centrality, and future trajectory. We introduce our method in determining each factor.

4.2.1 Social Similarity

To determine whether a node is suitable for being a caching node, we consider the social similarity between the forwarding node and the requesting node. Our CCSAMP is motivated from several social contact networks, such as the 2009 cambridge/haggle dataset, where socially-similar people tend to share the common interests and people with more common social attributes are liable to meet or come in contact with each other more frequently [41]. Furthermore, we have used the dataset from the cambridge/haggle trace to illustrate that the encounter frequency increases with the growing number of common features between nodes [7]. With the popularity of social networks, a lot of social attributes are posted online or shared among members of a social network. The content centric nature of VCCN allows the sharing of social attributes among vehicles without associating node IDs with packets.

If a forwarding node shares many social attributes with the requester, it may be interested in the same content and also likely to meet other nodes with similar interests, thus it is a good candidate to cache contents. We thus exploit social similarity as one component of our caching metric and determine it based on encounter history. First, we introduce some basic terminologies:

- ***Social Attribute (SA) Sequence*:**
$$\boldsymbol{SA_i} = \langle sa_{i_1}, sa_{i_2}, \cdots, sa_{i_j}, \cdots, sa_{i_r} \rangle.$$
Social attributes of the vehicle v_i form a sequence, where sa_j represents the j-th social attribute and there are r attributes. For example, the sequence of social attributes in Table 3 is $\langle$*Country*, *City*, *Nationality*, *Languages*, *Affiliation*, *Position*, $\cdots\rangle$, which is from the cambridge/haggle traces and often seen in an online social network such as Linkedln. Each attribute may have multiple possible values.
- ***Social History Record*:** Each vehicle exchanges its social attributes with neighbors to form the social history record, and maintains a table to record encounters' social attributes. Taking Table 3 as an example, the social attribute $Country$ has 3 values, China, Japan and America. "China = 3" represents this vehicle has met 3 neighbors from China. The social history record is updated based on the received social profiles from all neighboring nodes.
- ***Social Similarity*:** This parameter is used to evaluate the common social attributes between two nodes.

A requesting node sends an Interest packet with its social attributes encapsulated. Once a node receives an Interest, the node will send the requested content in a Data packet towards the requester if there is a cache hit. To facilitate other nodes to make the caching decision, this node also attaches the social attributes of the requesting node (extracted from the Interest packet) with the content returned. When a forwarding node v_i receives a Data packet, it first records the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

TABLE 3: *An Example of Social History Record*

Attribute	**Value1**	**Value2**	**Value3**	$\cdots$
Country	China(3)	Japan(2)	America(3)	$\cdots$
City	Beijing(2)	Shanghai(3)	Tokyo(1)	$\cdots$
Nationality	Chinese(3)	Japanese(2)	American(1)	$\cdots$
Languages	Chinese(4)	English(5)	Japanese(2)	$\cdots$
Affiliation	Tsinghua(3)	Tencent(2)	Citibank(1)	$\cdots$
Position	Engineer(2)	Professor(2)	Manager(3)	$\cdots$
Name	Allen (1)	Daniel (1)	George (2)	$\cdots$
Email	allen@ live.com	daniel@ msn.com	russell@ gmail.com	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

social profile of the requesting node v_s. Then, v_i computes the social similarity between itself and v_s. In this paper, we adopt the Jaccard similarity coefficient, a statistic value used for comparing the similarity and diversity of sample sets. It is defined as the size of the intersection divided by the size of the union of the sample sets [42]. Different from the original Jaccard similarity coefficient, each attribute is assigned with a weight, and we have the social similarity between v_i and v_s as

$$SS_i = J(v_i, v_s) = \frac{\sum_{j \in SA_i \cap SA_s} w_j}{\sum_{j=1}^{r} w_j}, \tag{1}$$

where r is the total number of social attributes and w_j is the weight of sa_j belonging to v_s.

Social attributes are commonly used in the analysis of social networks, such as link prediction and community detection [43]. User attributes could be static (e.g., school, major, employer and city) and derived from user profiles, or dynamic (e.g., online interest and community groups). Dozens of demographic attributes have appeared in the mobile social network and Twitter network [44]. To illustrate the computation of SS_i, we take the six attributes, ⟨*Country*, *City*, *Nationality*, *Languages*, *Affiliation*, *Position*⟩ from Table 3 as an example. Other types of social attributes can be also used. For example, a sequence of ⟨*Location*, *Destination*, *Moving Direction*, *Community*⟩ is used to compute the social similarity in [45]. Without the need of carrying node identity in VCCN, it helps to protect the privacy of nodes. A node could also choose to hide some sensitive attributes.

When v_s has the following attribute sequence, ⟨"China", "Shanghai", "Chinese", "Chinese", "Tsinghua", "Professor"⟩ and v_i has the attribute sequence ⟨"China", "Beijing", "Chinese", "Chinese", "Dalian University", "Professor"⟩, the intersection attribute $SA_i \cap SA_s$ is ⟨"China", "Null", "Chinese", "Chinese", "Null", "Professor"⟩, where "Null" means v_s and v_i have different values corresponding to that attribute. To find SS_i in Equation (1), we first need to determine w_j, which depends on the ratio r_j between the number of vehicles with a certain value of the attribute sa_j and all the encountered vehicles. We define $r_j = \frac{n_{j_k}}{\sum_{k=1}^{l_j} n_{j_k}}$, where l_j is the number of value types for the attribute sa_j. Each specific value is owned by n_{j_k} vehicles encountered, with $k = 1, 2, \cdots, l_j$. For sa_1 *Country*, it has 3 different values in Table 3, and the corresponding numbers of encounters are 3, 2, and 3. For the value "China", we can get $r_1 = \frac{3}{3+2+3} = \frac{3}{8}$. For v_s's attribute sequence, the ratio sequence is $\langle r_1, r_2, r_3, r_4, r_5, r_6 \rangle = \langle \frac{3}{8}, \frac{1}{2}, \frac{1}{3}, \frac{4}{11}, \frac{1}{2}, \frac{2}{7} \rangle$. For the four attributes of $SA_i \cap SA_s$, the ratio sequence is $\langle r_1, r_3, r_4, r_6 \rangle = \langle \frac{3}{8}, \frac{1}{3}, \frac{4}{11}, \frac{2}{7} \rangle$. We define $w_j = r_j^{-1}$. A bigger r_j represents that more nodes share the same attribute with v_s and v_s should have a higher probability of serving as a caching node. Finally, we can get

$$SS_i = \frac{\frac{8}{3} + 3 + \frac{11}{4} + \frac{7}{2}}{\frac{8}{3} + 2 + 3 + \frac{11}{4} + 2 + \frac{7}{2}} = \frac{143}{191} \approx 0.749.$$

If they share the same attribute values, SS_i will be 1.

4.2.2 Bridging Centrality

In social network analysis, the centrality metric is often used to identify the most influential node(s) or key infrastructure node(s) in a social network, where the social relationship of nodes or ties is often applied to identify the importance of nodes. As VCCN is dynamic and the network topology constantly changes with the node mobility, a node which has a higher encounter frequency can provide more connectivity links for the network. To identify how important a vehicle is, we use bridging centrality as the metric, which is calculated based on degree centrality and betweenness centrality. Before showing the method of determining bridging centrality, we first introduce how to calculate degree centrality and betweenness centrality based on the encounter records from neighbors.

- ***Degree Centrality*:** It is defined as the number of links between the vehicle and its neighbors. A vehicle has a higher degree centrality if it encounters other nodes more frequently. The degree centrality of v_i is defined as:

$$DC_i = \frac{n_{e_i}}{n_v - 1}. \tag{2}$$

  where n_{e_i} and n_v represent the number of vehicles that v_i has encountered and the total number of vehicles that it has counted based on the encounter records from neighbors.

- ***Betweenness Centrality*:** It is a measure of the extent to which a node has a control over the information flowing among others [9]. In CCSAMP, we apply betweenness centrality to examine the extent to which a vehicle stands in the path between two vehicles which are not directly connected. A higher value of betweenness centrality indicates that v_i plays a key role in facilitating the connection of other vehicles.

$$BC_i = 1 - \frac{n_{l_i}}{(n_{e_i} - 1) n_{e_i}}, \tag{3}$$

  where n_{l_i} represents the number of links that do not pass v_i. For example, v_i has two neighbors, v_m and v_{m+1}. Both v_m and v_{m+1} exchange their encounter records with v_i. If v_m exists in the record of v_{m+1}, it means that these two nodes can communicate with each other even without the help of v_i. Bigger n_{l_i} represents that more nodes can communicate with each other directly. Higher BC_i means v_i plays a larger role in providing a connection between two nodes.

- ***Bridging Centrality***: This metric is used to identify how important role a vehicle plays in the network to form the connection between nodes. We apply betweenness centrality and degree centrality to measure bridging centrality [5]:

$$BRC_i = DC_i \times BC_i = \frac{(n_{e_i} - 1)n_{e_i} - n_{l_i}}{(n_v - 1)(n_{e_i} - 1)}. \tag{4}$$

By taking into account bridging centrality in our caching decision, we would like to ensure that the caching node has a higher chance of staying on the path between a content requester and a content provider.

4.2.3 Trajectory Prediction

A vehicle's near future path can be predicted according to its past mobility trace with Hidden Markov Model (HMM) [46]. By applying a Forward-Backward Algorithm to train HMM, we can make full use of the past mobility patterns to find the maximum probability that a vehicle arrives at a destination. Before explaining our scheme, we first introduce the following terminologies:

- ***Trip Sequence:*** A trip sequence consists of a set of link points, $\mathcal{L} = \langle \ell_1, \ell_2, \cdots, \ell_i, \cdots \rangle$. It is the collection of movement records of a vehicle. Each trajectory point ℓ_i is composed by a triple (x_i, y_i, t_i), representing that the vehicle is located at (x_i, y_i) at time t_i. In our work, each vehicle records its trajectory point at each sample time.

- ***Hot Zone:*** A hot zone is an area with higher vehicle density. Each RSU counts the number of vehicles within its range and broadcasts its hot-zone list to other RSUs and vehicles around periodically. Once receiving the message, each updates its own list of hot zones $\mathcal{Z} = \langle z_1, z_2, \cdots, z_j, \cdots \rangle$.

We adopt HMM to predict the future locations by exploiting the trip sequences. HMM can be denoted by $\varphi = (\pi, A, B)$, where:

$\pi = \{\pi_i\}$ is the set of initial hidden state probabilities, with $\pi_i = P(S_i)$.

$A = \{a_{ij}\}$ is the set of transition probabilities between the hidden states S_i and S_j, with $a_{ij} = P(S_j|S_i)$;

$B = \{b_j(k)\}$ is the set of probabilities of the observable states O_k in the hidden state S_j, with $b_j(k) = P(O_k|S_j)$;

$\zeta_t(i,j)$=$P(S_i, S_j|O_1O_2...O_T, \varphi)$ is the probability of transiting from the hidden state S_i at the time t to the hidden state S_j at the time $t+1$, given the model φ and the observation sequence.

$\eta_t(i)$=$P(S_i|O_1O_2...O_T, \varphi)$ is the probability of the hidden state S_i at the time t, given the model φ and the observation sequence.

We can obtain the accurate model by calculating the corresponding variables in Equation (5),

$$\begin{aligned} \overline{\pi_i} &= \eta_t(i), \\ \overline{a_{ij}} &= \frac{\sum_{t=1}^{T} \zeta_t(i,j)}{\sum_{t=1}^{T} \eta_t(i)}, \\ \overline{b_j(k)} &= \frac{\sum_{t=1,O_k}^{T} \eta_t(j)}{\sum_{t=1}^{T} \eta_t(j)}. \end{aligned} \tag{5}$$

where $\overline{\pi_i}$ represents the probability that the vehicle stays at the hidden state S_i at t, $\overline{a_{ij}}$ represents the probability of transition from S_i to S_j, and $\overline{b_j(k)}$ represents the probability of observing the state O_k when the hidden state is S_j.

In our scheme, ℓ_k represents the observed state O_k in the k-th time slot and the hidden state S_{k+1} represents the predicted location in $(k+1)$-th time slot. Combining three equations in Equation (5), each vehicle can determine the probability that it will reach the hot zones with the following matrix:

$$\begin{bmatrix} p_{i\to 1} & p_{i\to 2} & \cdots & p_{i\to j} & \cdots & p_{i\to n} \end{bmatrix},$$

where $p_{i\to j}$ represents the probability of v_i entering z_j and n represents the number of hot zones. The probability that v_i entering hot zones can be represented as $\sum_{j\in\mathcal{Z}} p_{i\to j}$.

4.2.4 Caching decision

According to the importance of the social similarity, bridging centrality and future path, the probability P to cache a data packet in v_i is:

$$P = \alpha \cdot SS_i + \beta \cdot BRC_i + \gamma \cdot \sum_{j\in\mathcal{Z}} p_{i\to j}, \tag{6}$$

where the impact of these three factors on the choice of a caching node is considered. In Equation (6), α is the weight of social similarity, β is the weight of bridging centrality, γ is the weight of the probability of entering some hot zones, $\alpha + \beta + \gamma = 1$ holds and P belongs to $[0, 1]$. Higher SS_i means that v_i has more common social attributes with the requesting node, which indicates that they have a higher chance of sharing the interest in the same content. If a vehicle has a higher probability of entering some hot zones, it will meet more vehicles and have a higher chance of sharing its contents with others, while bridging centrality shows how likely that a node stays in-between requesters and providers and sends contents it caches to the requester. Besides, it has been proved [7] that a node sharing more social features with the destination is more likely to travel close to the latter in the near future and should be chosen as the next-hop forwarder. Compared with the metrics of bridging centrality and future trajectory, social similarity should possess a higher weight. We will verify this conclusion in our performance studies.

4.3 Cache Replacement

In our CCSAMP, we design the cache replacement policy based on the content popularity. If the storage is full, only contents with higher popularity are cached. For each Interest received, a node records the sequence of social attributes of the requester into VIT, and inserts the content ID and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

request time into PIT. Periodically, the content popularity ρ_{c_i} is updated for each named content as follows:

$$\rho_{c_i} \Leftarrow \rho_{c_i} \cdot e^{-\lambda \Delta t/(SS+1)} + \theta, \tag{7}$$

where Δt represents the time interval between the current time and the last time of receiving the same request on c_i, SS is the social similarity between the caching node and the source requester, λ is an exponential decay constant and θ is the popularity increase constant. If the corresponding Interest cannot be received in a time slot, Δt is set to be the time interval between the ending time of the current slot and the last time of receiving the request. If an Interest for the new content is received for the first time, the popularity ρ_{c_i} will be set to θ. Our popularity calculation captures both the request frequency and the freshness of the requests. We also consider the social position of the caching node. A higher SS allows the caching node to better provide c_i for other members which share social attributes, and we preferably let the node to keep the content.

Since RSUs do not have social attribute sequences, they will use the popularity increase constant θ without considering SS. The popularity is calculated as follows:

$$\rho_{c_i} \Leftarrow \rho_{c_i} \cdot e^{-\lambda \Delta t} + \theta, \tag{8}$$

In Equation (7) and Equation (8), ρ_{c_i} will be updated upon receiving a new Interest. If no Interest for c_i is received upon the arrival of a new content to cache while the cache space runs out, Δt will be set as the difference between the current time and the time of the last request to update the popularity. If no request for a content is received in several consecutive time slots, the corresponding content may be replaced by popular contents in the cache upon space constraint.

5 Performance Evaluation

In order to evaluate the performance of our caching scheme, we conducted our simulations over the Opportunistic Network Environment (ONE) simulator [47]. ONE includes several mobility models, from simple Random Waypoint to more realistic Map-Based Movement. We use the 2009 cambridge/haggle dataset in our simulation, including the social attributes of 85 people. In our design, there are 335 nodes, including 30 RSUs and 305 users (85 pedestrians, 100 buses, and 120 taxis), distributed in the map. All nodes move following the Working-Day-Movement Model in ONE with a daily routine, which mainly consists of staying at home, working in the office, going to the gymnasium and so on. Instead of using all social features in the dataset, we adopt 6 informative features based on their entropies [41]: ⟨*Country*, *City*, *Nationality*, *Languages*, *Affiliation*, *Position*⟩, which forms the social attribute sequence to record in its VIT. Each vehicle exchanges its social sequence with its neighbors to form and maintain a social history record, which tracks the number of encounters associated with a specific social attribute. A person drives a car with a chance of p_v, or s/he must take the bus or taxi to reach different destinations. Buses follow the Route-Based Movement model in ONE and interact with passengers through a bus control system. Taxis run by the Random Waypoint Model in ONE. All nodes have the same range of moving speeds, transmission range and data rate. At the beginning, we collected the social history records and trip sequences of vehicles in 5 working weeks for the training purpose and applied them to different caching schemes. In the actual evaluation stage, every person periodically generated content requests following the Zipf's law distribution [48], $f(k;s,N) = \frac{1/k^s}{\sum_{n=1}^{N}(1/n^s)}$, where N is the number of elements (number of contents in our paper), k is the rank of contents, and s is the exponent characterizing the distribution. Each vehicle records its trajectory points every 10 seconds and insert them as a sequence into its VIT. It only keeps trajectory points taken within the past 50 seconds, which are used to predict the probability of reaching a hot zone. We list important simulation parameters in Table 4.

TABLE 4: Simulation Parameters

Parameter Description	Value
Caching Buffer of each pedestrian	1600MB
Caching Buffer of each taxi	3200MB
Caching Buffer of each bus	6400MB
Request Interval	[50minutes, 100minutes]
Message TTL	10minutes
Network Area	10km×7.5km
Simulation Time	15weeks
RSU Number	30
RSU Transmission Range	500m
RSU Transmission Speed	10Mbps
Vehicle Speed	[7m/s, 10m/s]
Vehicle Transmission Range	50m
Vehicle Transmission Speed	2Mbps
Exponent s	0.8
Each content size	25MB
Number of contents	2000
Popularity Increment θ	0.1
p_v	0.3

5.1 Performance Metrics

Our main comparisons are made between the proposed CCSAMP scheme, and reference schemes DPC [9], LDC-C [10] and DAC [11]. In CCSAMP scheme, we calculate the social similarity between the forwarding node and the requester and the bridging centrality of the forwarding node to analyze the importance of each vehicle, and predict the vehicle's future locations using HMM to obtain the probabilities of arriving at the hot zones. Thus, we choose the probabilistic caching scheme DPC and LDCC as the references. In DPC, each node makes its own caching decision based on users' demand and its importance including degree centrality and betweenness centrality. In LDCC, a node with the highest probability staying in a valid scope is chosen as the caching node. DAC also considers the social attribute such as contact pattern and relationship to choose the caching node. The following metrics are used to compare these schemes:

- ***Cache Hit Ratio***: The probability of obtaining a cache hit from a caching node, which is defined as the ratio of the number of cache hits to the total number of receiving Interests.
- ***Average Access Delay***: The average delay of obtaining responses in successful queries.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

- ***Average Hop Count*:** The average hop count between the requester and provider in successful queries.
- ***Standard Deviation of Storage Usage* σ:** The standard deviation of storage usage among all nodes in the network.
- ***Average Storage Usage,* Γ_B, Γ_T, Γ_C:** The average storage usage of different nodes: Buses, Taxis, and private Cars, respectively, in the network.

In the remaining sections, we provide a number of studies to evaluate the performance of our proposed scheme.

5.2 Effect of three factors

Equation (6) shows that the probability for a vehicle to be the caching node is decided by a combination of the following three factors, social similarity, bridging centrality and future trajectory. We first evaluate the impact of each factor separately. Table 5 shows that any factor can improve the cache hit ratio compared with no factor adopted, and the social similarity plays the most important role. Table 5 also shows that the improvement of cache hit ratio is very close when either bridging centrality or future trajectory is adopted. Thus, we set β=γ=$\frac{1-\alpha}{2}$ in our later simulations.

TABLE 5: Effect of each factor

Factor	Cache Hit Ratio
Social Similarity	0.392
Bridging Centrality	0.354
Future Trajectory	0.348
Three Factors	0.416
No Factor	0.312

Besides, we evaluate the effect of α on the cache hit ratio in Fig. 4. At α=0, only two factors, the bridging centrality and future trajectory, are considered. At α=1, only the social similarity is considered. When α is between 0 and 0.5, the bridging centrality and probability of entering the hotzone play a more important role in deciding what contents to cache. When α is between 0.5 and 1, social similarity plays a more important role. The cache hit ratio keeps increasing with α until α reaches 0.6, which indicates that the social similarity plays the most critical role. In our later simulations, α is set to 0.6.

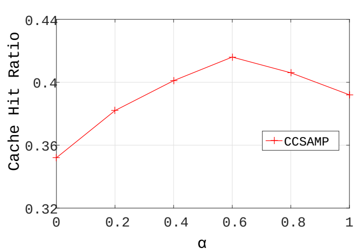


Fig. 4: Effect of α on CCSAMP

We also evaluate the effect of the number of common social attributes on the cache hit ratio in Fig. 5. The cache hit ratio increases as nodes share more social attributes, which also shows that the social similarity gradually plays a more important role.

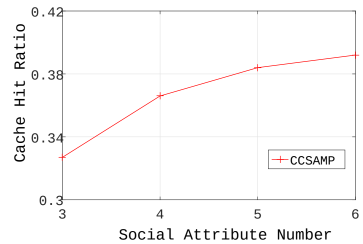


Fig. 5: Effect of social attribute number on CCSAMP

5.3 Effect of p_v

p_v is the probability that a person drives a car. Alternatively, the person can take the bus or taxi to reach different destinations. We evaluate the impact of p_v on the cache hit ratio of CCSAMP in Fig. 6. As p_v increases, more people choose to drive cars, making the mobility patterns less predictable. Consequently, the cache hit ratio decreases with p_v.

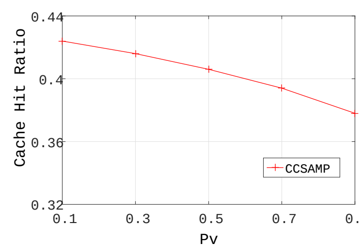
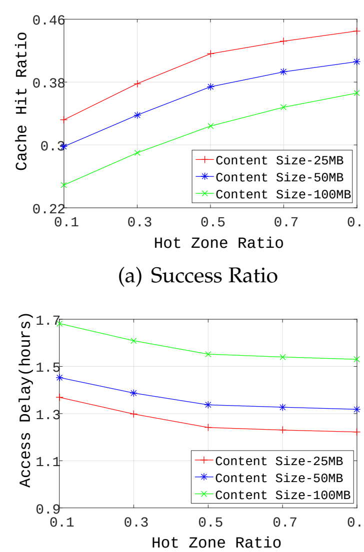


Fig. 6: Effect of p_v on CCSAMP

5.4 Effect of Hot Zone Ratio



(a) Success Ratio

(b) Average Access Delay

Fig. 7: Effect of Hot Zone Ratio on CCSAMP

As discussed before, a vehicle tending to go to a hot zone is liable to become a caching node. In Fig. 7, we evaluate the effect of hot zone ratio on the performance of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2019.2944829, IEEE Transactions on Mobile Computing

CCSAMP. Fig. 7(a) shows that the cache hit ratio can achieve about 34.04%-46.98% gain as the hot zone ratio increases. This indicates that it is important to consider the impact of hotzone when designing a caching strategy for VANET. This is because more vehicles are likely to enter hot regions, and be chosen as caching nodes. Accordingly, the access delay decreases with the hot zone ratio. When it exceeds 0.5, the cache hit ratio keeps relatively stable. In our later simulations, we set the hot zone ratio to be 0.5

5.5 Effect of Content Size

We vary the content size from 25MB to 125MB and compare cache hit ratio, average access delay and average storage usage of different schemes.

As the content size increases, each node caches fewer contents, causing lower cache hit ratio and higher access delay in Fig. 8(a) and Fig. 8(b). Our CCSAMP has the best performance with the highest cache hit ratio (up to 29.66% gain) and lowest access delay (up to 26.74% drop), because CCSAMP has taken into account social relationship and node mobility. Among the four schemes, DAC only considers the social attribute without predicting the future location of each vehicle, making it possess the biggest delay in Fig. 8(b). LDCC has the least cache hit ratio when the content size is larger than 90MB in Fig. 8(a) because the valid dwelling time in the future location is also considered to design the cache replacement policy, making the cache update slow. As the content size increases, fewer contents and slower update time cause the smallest cache hit ratio.

In Fig. 8(c) and Fig. 8(d), CCSAMP needs slightly more storage space compared to others, because it needs to store both location and social data to facilitate the selection of caching node. In DAC, only the social attributes are necessary, bringing the least storage. Though DPC and LDCC predict the future location, LDCC integrates other factors like cache size, access frequency, energy consumption, etc. to design the replacement policy, causing more storage than DPC.

5.6 Effect of Caching Node Ratio

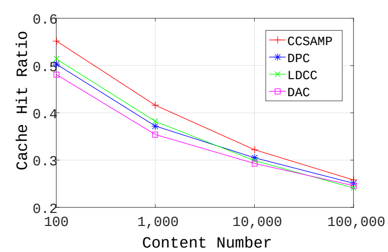
In our scheme, each caching node decides whether to cache some contents by considering the above three factors. However, not all nodes may be configured as cache nodes. Which nodes will serve for caching can be determined by network requirements or nodes' interests. We evaluate the effect of the ratio of cache nodes on the CCSAMP performance. In Fig. 9(a), as expected, the cache hit ratio improves with the caching node ratio. DAC does not adopt any prediction on the future location, so it has the lowest cache hit ratio. CCSAMP has the highest cache hit ratio by combing the prediction and social attributes. Meanwhile, Fig. 9(b) and Fig. 9(c) show that the average access delay and average hop count decrease with the increase of caching nodes and CCSAMP outperforms other three schemes consistently. Though each node decides whether to cache the contents based on its future location in DPC and LDCC, DPC does not consider the dwelling time in one region. DPC shows its advantage and achieves a better performance than LDCC as more nodes are chosen as the caching nodes in Fig. 9(d). The interesting convex shape of standard deviation in Fig. 9(d) might have suggested an "optimum" caching node ratio, in which cache hit ratio is close to the best and a collection of "core" nodes are used to cache popular contents.

5.7 Effect of Content Number

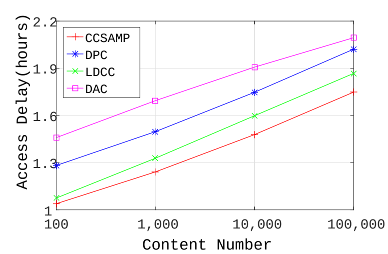
In Table 6, we show the cache hit ratio with different content numbers. It can be seen that the cache hit ratio decreases as the number of contents increases, because a larger number of content types increase the difficulty of cache hit. Fig. 10 shows that our CCSAMP scheme always achieves the best performance among the four schemes in terms of the cache hit ratio and the average access delay though the number of contents increases by a 100-fold.

TABLE 6: Effect of Content Number

Content Number	Cache Hit Ratio
100	0.552
$1,000$	0.416
$10,000$	0.322
$100,000$	0.258



(a) Success Ratio

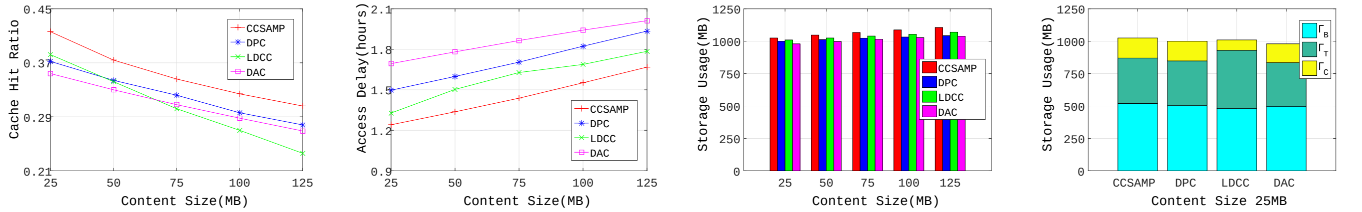


(b) Average Access Delay

Fig. 10: Effect of Content Number on Different Schemes

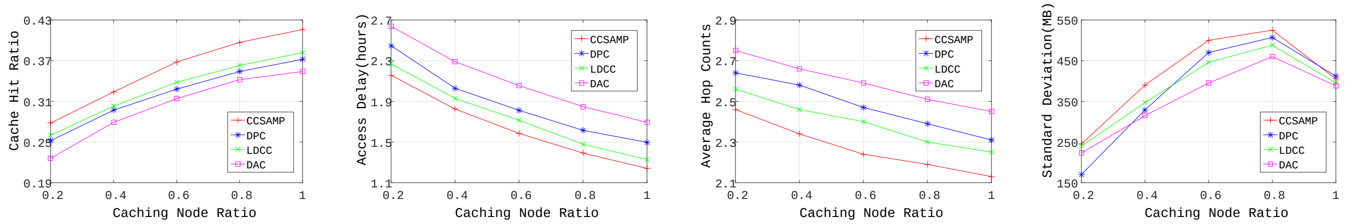
6 Conclusion

Vehicular Content Centric Network (VCCN) is expected to play an essential role in entertainment, advertisement, and other popular content delivery. In this paper, we propose a novel scheme, termed CCSAMP, to support such content caching, request and delivery. Utilizing the social attributes and trajectory history records of vehicles, CCSAMP calculates social similarities and bridging centralities of vehicles and adopts HMM to predict the probabilities of their next visits to hot zones in the area. Based on the above three factors, caching nodes will be chosen. In addition, we have performed extensive simulations to compare CCSAMP with several other state-of-the-art schemes in cache hit ratio, average access delay, average hop counts and average



(a) Success Ratio (b) Average Access Delay (c) Average Storage Usage (d) Storage Usage of Different Type of Nodes

Fig. 8: Effect of Content Size on Different Schemes



(a) Success Ratio (b) Average Access Delay (c) Average Hop Counts (d) Standard Deviation of Storage Usage

Fig. 9: Effect of Caching Node Ratio on Different Schemes

storage usage. It has been shown that CCSAMP possesses a better performance in all cases studied.

In our future work, we plan to optimize our algorithm to minimize the energy consumption of the system. Besides, we will design some incentive policy to encourage each vehicle to actively serve others by caching contents.

Acknowledgment

This research is sponsored in part by the National Natural Science Foundation of China (contract/grant numbers: 61872053), the Fundamental Research Funds for the Central Universities (DUT19GJ204) and US National Science Foundation (CNS 1526843).

References

[1] G. Deng, F. Li, and L. Wang, "Cooperative downloading in vanets-lte heterogeneous network based on named data," in *Computer Communications Workshops*, 2016, pp. 233–238.

[2] S. H. Bouk, S. H. Ahmed, and D. Kim, "Vehicular content centric network (vccn): a survey and research challenges," in *ACM Symposium on Applied Computing*, 2015, pp. 695–700.

[3] M. Amadeo, C. Campolo, and A. Molinaro, "Information-centric networking for connected vehicles: a survey and future perspectives," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 98–104, 2016.

[4] M. Zhang, H. Luo, and H. Zhang, "A survey of caching mechanisms in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1473–1499, 2015.

[5] A. M. Vegni and V. Loscrł, "A survey on vehicular social networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2397–2419, 2015.

[6] J. Shi, X. Wang, M. Huang, K. Li, and S. K. Das, "Social-based routing scheme for fixed-line vanet," *Computer Networks the International Journal of Computer & Telecommunications Networking*, vol. 113, no. C, pp. 230–243, 2017.

[7] L. Yao, Y. Man, Z. Huang, J. Deng, and X. Wang, "Secure routing based on social similarity in opportunistic networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 594–605, 2016.

[8] A. A. F. Loureiro, R. A. F. Mini, A. C. Vianna, and F. D. Cunha, "How effective is to look at a vehicular network under a social perception?" in *IEEE International Conference on Wireless and Mobile Computing, NETWORKING and Communications*, 2014, pp. 154–159.

[9] G. Deng, L. Wang, F. Li, and R. Li, "Distributed probabilistic caching strategy in vanets through named data networking," in *Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2016, pp. 314–319.

[10] E. Chan, Y. Wang, W. Li, and S. Lu, "Movement prediction based cooperative caching for location dependent information service in mobile ad hoc networks," *The Journal of Supercomputing*, vol. 59, no. 1, pp. 297–322, 2012.

[11] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. La Porta, "Social-based cooperative caching in dtns: A contact duration aware approach," in *8th International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 92–101.

[12] A. Ioannou and S. Weber, "A survey of caching policies and forwarding mechanisms in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2847–2886, 2016.

[13] Z. Li and G. Simon, "Time-shifted tv in content centric networks: The case for cooperative in-network caching," in *IEEE International Conference on Communications*, 2011, pp. 1–6.

[14] V. Sourlas, L. Gkatzikis, P. Flegkas, and L. Tassiulas, "Distributed cache management in information-centric networks," *IEEE Transactions on Network and Service Management*, vol. 10, no. 3, pp. 286–299, 2013.

[15] E. J. Rosensweig and J. Kurose, "Breadcrumbs: Efficient, best-effort content location in cache networks," in *INFOCOM*, 2009, pp. 2631–2635.

[16] J. M. Wang, J. Zhang, and B. Bensaou, "Intra-as cooperative caching for content-centric networks," in *ACM SIGCOMM Workshop on Information-Centric NETWORKING*, 2013, pp. 61–66.

[17] S. Eum, K. Nakauchi, M. Murata, Y. Shoji, and N. Nishinaga, "CATT: potential based routing with content caching for ICN," in *Edition of the Icn Workshop on Information-Centric NETWORKING*, 2012, pp. 81–86.

[18] D. Rossi and G. Rossini, "On sizing CCN content stores by exploiting topological information," in *Proceedings - IEEE INFOCOM*, 2012, pp. 280 – 285.

[19] S. Saha, A. Lukyanenko, and A. Yla-Jaaski, "Cooperative caching through routing control in information-centric networks," in *Proceedings - IEEE INFOCOM*, 2013, pp. 100–104.

[20] I. Psaras, W. K. Chai, and G. Pavlou, "In-network cache management and resource allocation for information-centric networks,"

IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 11, pp. 2920–2931, 2014.

[21] I. Psaras, K. C. Wei, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *Edition of the Icn Workshop on Information-Centric NETWORKING*, 2012, pp. 55–60.

[22] K. Cho, M. Lee, K. Park, T. T. Kwon, Y. Choi, and S. Pack, "Wave:popularity-based and collaborative in-network caching for content-oriented networks," in *Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2012, pp. 316–321.

[23] Z. Li and G. Simon, "Cooperative caching in a content centric network for video stream delivery," *Journal of Network and Systems Management*, vol. 23, no. 3, pp. 445–473, 2015.

[24] M. Chaqfeh, A. Lakas, and I. Jawhar, "A survey on data dissemination in vehicular ad hoc networks," *Vehicular Communications*, vol. 1, no. 4, pp. 214–225, 2014.

[25] M. K. Denko and J. Tian, "Cross-layer design for cooperative caching in mobile ad hoc networks," in *Communications and NETWORKING Conference*, 2008, pp. 375–380.

[26] H. Dan and S. Madria, "A resource-efficient adaptive caching scheme for mobile ad hoc networks," in *IEEE Symposium on Reliable Distributed Systems*, 2010, pp. 64–71.

[27] I. H. Bae and S. Olariu, "Design and evaluation of a fuzzy cooperative caching scheme for manets," in *International Conference on Wireless Communications NETWORKING and Mobile Computing*, 2010, pp. 1–5.

[28] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Supporting cooperative caching in disruption tolerant networks," in *International Conference on Distributed Computing Systems*, 2011, pp. 151–161.

[29] F. J. González-Cañete, E. Casilari, and A. Triviño-Cabrera, "A cross layer interception and redirection cooperative caching scheme for manets," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 63, 2012.

[30] S. Lim, C. Yu, and C. R. Das, "Cache invalidation strategies for internet-based vehicular ad hoc networks," *Computer Communications*, vol. 35, no. 3, pp. 380–391, 2012.

[31] R. Tiwari and N. Kumar, "Cooperative gateway cache invalidation scheme for internet-based vehicular ad hoc networks," *Wireless Personal Communications*, vol. 85, no. 4, pp. 1789–1814, 2015.

[32] W. Gao, G. Cao, A. Iyengar, and M. Srivatsa, "Cooperative caching for efficient data access in disruption tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 3, pp. 611–625, 2014.

[33] S. E. El Khawaga, A. I. Saleh, and H. A. Ali, "An administrative cluster-based cooperative caching (accc) strategy for mobile ad hoc networks," *Journal of Network and Computer Applications*, vol. 69, pp. 54–76, 2016.

[34] T. Le, Y. Lu, and M. Gerla, "Social caching and content retrieval in disruption tolerant networks (dtns)," in *International Conference on Computing, NETWORKING and Communications*, 2015, pp. 905–910.

[35] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A learning-based approach to caching in heterogenous small cell networks," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1674–1686, 2016.

[36] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5g wireless networks." *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.

[37] ——, "Social and spatial proactive caching for mobile data offloading," in *IEEE International Conference on Communications Workshops*. IEEE, 2014, pp. 1–6.

[38] ——, "A transfer learning approach for cache-enabled wireless networks," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*. IEEE, 2015, pp. 161–166.

[39] ——, "Proactive caching in 5g small cell networks," *Towards 5G: Applications, Requirements and Candidate Technologies*, pp. 78–98, 2016.

[40] M. Karmoose, M. Cardone, and C. Fragouli, "Simplifying wireless social caching," in *IEEE International Symposium on Information Theory*. IEEE, 2016, pp. 410–414.

[41] J. Wu and Y. Wang, "Social feature-based multi-path routing in delay tolerant networks," in *Proc. of the IEEE INFOCOM*. IEEE, 2012, pp. 1368–1376.

[42] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of jaccard coefficient for keywords similarity," in *Iaeng International Conference on Internet Computing and Web Services*, 2013, pp. 1–5.

[43] N. Z. Gong, W. Xu, H. Ling, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of social-attribute networks: Measurements, modeling, and implications using google+," *In: IMC (2012*, pp. 131–144, 2012.

[44] W. Chen, J. Wang, Z. Jiang, Y. Zhang, and X. Li, "Hierarchical mixed neural network for joint representation learning of social-attribute network," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2017, pp. 238–250.

[45] Z. Li, C. Wang, S. Yang, C. Jiang, and X. Li, "Lass: Local-activity and social-similarity based data forwarding in mobile social networks," *IEEE Transactions on Parallel & Distributed Systems*, vol. 26, no. 1, pp. 174–184, 2014.

[46] W. Khreich, E. Granger, A. Miri, and R. Sabourin, "On the memory complexity of the forwardcbackward algorithm," *Pattern Recognition Letters*, vol. 31, no. 2, pp. 91–99, 2010.

[47] A. Keranen, J. Ott, and T. Karkkainen, "The one simulator for dtn protocol evaluation," in *International Conference on Simulation TOOLS and Techniques*, 2009, p. 55.

[48] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: evidence and implications," in *Proceedings - IEEE INFOCOM*, 1999, pp. 126 – 134.

Lin Yao is a professor in DUT-RU International School of Information Science & Engineering, Dalian University of Technology (DUT), China. Her research interests include privacy of big data and security of Vanet, CCN, and Social network.

Yuqi Wang is an M.E. candidate in School of Software, Dalian University of Technology (DUT), China. His research interests include security and privacy in social network.

Xin Wang is currently an associate professor of the department of Electrical and Computer Engineering, Stony Brook University, New York State, U. S. A. Her research interests include mobile and ubiquitous computing, wireless communications and network systems, networked sensing and fusion, detection and estimation.

Guowei Wu is a professor in the School of Software, Dalian University of Technology (DUT), China. His research interests include smart edge computing, software defined networks, and N-FV.