On Cache Invalidation for Internet-based Vehicular Ad Hoc Networks

Sunho Lim† Soo Hoan Chae‡ Chansu Yu§ Chita R. Das¶

Abstract

Internet-based vehicular ad hoc network (IVANET) is an emerging technique that combines a wired Internet and a vehicular ad hoc network (VANET) for developing a next generation of ubiquitous communication infrastructure and improving universal information and service accessibility. A key optimization technique in IVANETs is to cache frequently accessed data items in a local storage of vehicles. Since vehicles are not critically limited by the storage space, it is a less of a problem which data items to cache. Rather, a critical design issue is how to invalidate them when data items are updated. This is particularly a concern due to vehicles’ high-speed mobility. In this paper, we propose a novel cache invalidation algorithm that takes advantage of the underlying location management scheme to reduce the number of broadcast operations and the corresponding query delay. Numerical results indicate that the proposed scheme significantly reduces the communication cost, and thus is proven to be a viable solution for IVANETS.

Keywords— Internet-based vehicular ad hoc network, cache invalidation, mobile IP.

1. Introduction

Vehicular ad hoc network (VANET) consists of a set of high-speed mobile vehicles equipped with communication facilities [1]. It allows either inter-vehicle or vehicle-to-roadside communications through a multi-hop message relay without the assistance of any fixed infrastructure. Numerous of the applications built for VANETS will be available in the near future as drivers demand information regarding, for example, safety and traffic. More specifically, they can monitor/share a real-time traffic condition, download files, send/receive an emergency warning message, and avoid an accident proactively such as an intersection collision or a chain collision [2, 3].

In terms of network architecture, drivers are able to access the Internet service and information on wheels [4] through a road side unit (e.g. an access point (AP), an infostation [5, 6], or a message relay box [7]) located along the road that acts as a gateway to an infrastructure network. Thus, it is imperative to consider integration of VANET with wireless infrastructure, where a wireless local area network (e.g. IEEE 802.11) and a wireless wide area network (e.g. 3G) are co-existed. It is envisaged that such a Internet-based vehicular ad hoc network, or IVANET, will provide a flexible connectivity, accessibility, and service in near future.

A key optimization technique in IVANET is to cache frequently accessed data items in a local storage. In IVANET, it is less of a problem to determine which data items to cache because the storage space in a vehicle is not critically limited. A critical design issue is cache invalidation for applications requiring a strong data consistency with the server. When a data item in a server is updated, it is necessary to make sure that the cached copies of this data item should be validated before they can be used.

A large number of cache invalidation schemes have been proposed in the literature, most of which is based on invalidation report (IR) [8, 9, 10, 11, 12, 13, 14, 15] and are targeting cellular networks or MANETS. A server periodically broadcasts an IR, which includes a list of updated data items. Then mobile nodes (later in short, nodes) that receive the IR invalidate cached data items. An important design goal of these traditional cache invalidation techniques is to achieve a certain level of energy conservation, but it is not an issue in IVANETS because a vehicle is supported by its own built-in battery. Rather, our concerns in developing a cache invalidation scheme in the context of IVANETS are mobility and communication cost.

- Due to high-speed mobility, vehicles reside in a coverage area for a short period of time and travel exten-
sive areas. For example, the connection time within a coverage area ranges from 5 to 24 seconds at city driving [16]. Since vehicles are scattered over the multiple coverage areas, multiple servers are involved to a broadcast operation and thus, the cost of broadcasting IRs becomes non-negligible.

- Since a vehicle has a low probability of finding common data items in adjacent vehicles in a real IVANET environment, the same content of IRs broadcasted to different vehicles is wasteful because most of contents are not relevant to the vehicles.
- When a server broadcasts an IR, it is assumed that every node within a coverage area receives the IR. However, we need to relax this assumption in IVANETS, where vehicles move across the multiple coverage areas. Thus it is essential to keep track of the current coverage areas of each vehicle for validation. The more knowledge of vehicles’ whereabouts is available, the less validation cost is achieved.

This paper proposes a new cache invalidation scheme for IVANETS, which can effectively deal with the vehicles’ high-speed mobility without incurring a large volume of traffic for cache invalidation.

- First, we suggest a novel network model of cache invalidation for IVANETS and bring several realistic issues with respect to scalability of algorithm and capturing of high-speed mobility. This has not been studied elsewhere in the literature to the best of the authors’ knowledge.
- Second, the proposed cache invalidation scheme is integrated with a mobile IP based location management. Here, a server asynchronously sends an IR to a home agent (HA) rather than blindly broadcasts it to the vehicles. Then the HA judiciously refines and distributes the IR to appropriate gateway foreign agents (GFAs), where they can answer a validity of the queried data item to reduce an unavoidable query delay witnessed in most IR-based approaches.

We develop a simulation model and observe the effect of mobility, query interval, data item update interval, and data item size on system performance. The overall numerical results show that the proposed scheme is superior in all aspects, and thus is proven to be a viable approach for IVANETS.

The rest of paper is organized as follows. The prior study is reviewed in Section 2. The system model is introduced in Section 3, and the proposed scheme is presented in Section 4. Section 5 is devoted to performance evaluation and comparison with other schemes. Finally, we conclude the paper with future directions in Section 6.

2. Related Work

Various cache invalidation schemes [8, 9, 12, 15] have been suggested for cellular networks, where the nodes are one hop away from a server. Barbara and Inielinksi [8] proposed a scheme for a stateless server, where the server broadcasts the invalidation report (IR). Since the IR is broadcasted periodically, the node operates in a doze mode and wakes up during the server broadcast to save the battery power. Depending on the IR size and broadcast interval, there is a tradeoff between the efficiency of cache invalidation and query latency. Since the node should wait for the IR before it answers a query, updated invalidation report (UIR) [9] is proposed to further reduce the latency by adding updated items between IRs. In addition, an asynchronous cache invalidation scheme [12] is proposed, where the server broadcasts an IR whenever a data item is updated.

In [17, 18], location-dependent cache invalidation schemes are proposed, where the value of data item is determined by the location where it is related and the requested data item is replied by a server with the valid scopes attached. With the attached validation information, nodes can check the validity of cached data items without even connecting to the server. Ren et al [19] proposed a semantic caching scheme for location-dependent data items, in which a node maintains semantic descriptions of data items. When a node generates a query, it analyzes the descriptions of the query and produces results or partial results based on the appropriate cached data items. Then the node tailors the query and sends a request to the server for the rest of results.

For mobile ad hoc networks (MANETS), Hara suggested a replication scheme [11] for periodically updated data items based on the previously suggested replica allocation method [10]. Replicated data items are periodically re-allocated depending on the access frequency, remaining time to next update, and network topology. However, estimation of optimal reallocation period and remaining time to the next update are not practically feasible. There are few invalidation approaches [13, 14] for Internet-based MANETS, called IMANETS, where MANET technologies are combined with an Internet to provide a flexible information accessibility and availability for users. In [13], several push and pull-based cache invalidation schemes are proposed based on the aggregate cache [20], in which the aggregated local caches of the nodes can be considered as an unified large cache for the IMANET.

In summary, none of prior techniques can be directly applied to an IVANET, where the vehicles’ high-speed mobility and communication cost are a primary concern.
3. System Model

As illustrated in Fig. 1, an IVANET consists of access points (APs), gateway foreign agents (GFAs), home agents (HAs), and servers of data item source. First, vehicles are equipped with communication facilities such as an IEEE 802.11-based dedicated short range communication (DSRC) transceiver. Thus, they can either communicate with other vehicles or connect to the Internet flexibly. Since providing a location tracking capability in vehicles is becoming popular, we assume that vehicles are aware of their current location. Also vehicles have a built-in navigation system via GPS, in which a digital map is loaded to show roads around current location and direction, the shortest path to the destination, traffic conditions, location-dependent information, and so on. In addition, unlike cellular and MANET environments, where nodes move without restrictions, vehicles are bounded to the underlying fixed roads with speed limits and traffic lights in VANETS. Thus, it is relatively easier to predict a future movement of vehicles. For example, a vehicle is moving in the east direction along the road, where there is no exit or intersection for next several miles.

To support a global wireless access of Internet, mobile IP [21] is deployed, in which vehicles are able to access the Internet and maintain their on-going communications while they move at a high-speed. When a vehicle accesses the Internet, it searches the nearest located AP, which is a gateway to the Internet, to access any information and service. An AP is located in a communication area and can be mounted on the top of a signal light or a road lamp along the road. Also the AP plays an additional role as a foreign agent (FA). Thus, when a vehicle moves from one coverage area to another, it performs a location update/register operation that involves an AP, a GFA, and a HA.

A database may be attached to any components in the proposed system, e.g. an AP, a GFA, a HA or a database server. However, due to the consistency issue among replicas stored in multiple databases, we assume that the database is attached to a database server. We also assume that the database consists of a total $D$ data items. A data item ($d_x$) is the basic unit of an update or a query operation, where $1 \leq x \leq D$. The database is only updated by the server, while a query is generated by the vehicles for read-only requests. The server broadcasts an IR to the vehicles to ensure data consistency.

When a vehicle joins an IVANET, it needs to access what Internet services and information are available in the current area, and connects to the nearest AP. Since a set of available services and information changes with a time and location, it is too costly for the vehicle to keep track of them. Thus AP periodically broadcasts a beacon message, and vehicles located within the coverage area of the AP are notified the current available services and information.

4. A Location Management based Cache Invalidation Scheme

In this section, we first introduce the location management mechanism, which is a basis of the proposed scheme, and then present detailed cache invalidation operations in IVANETS.

4.1. The Location Management

Due to high-speed mobility of vehicles, it is essential to keep track of vehicles’ whereabouts and their coverage areas to reduce the impact of mobility on the cache invalidation operation. In this paper, we develop a location management mechanism based on mobile IP [21] for providing a reference to the server. Mobile IP is a simple mobility-enabling protocol, and it supports vehicles to maintain all on-going connections and to access global wireless Internet. To avoid frequent location register/update operations for fast moving vehicles, we use an IP micro-mobility support, in which vehicles report the HA only when a major change is occurred such as changes of a regional network.

First, when a vehicle moves into a coverage area within the same regional network, it sends the location update to the GFA. When a vehicle moves into a different regional network, however, it sends the location update to the HA for correct forwarding the packets through the GFA and AP. The AP periodically broadcasts an agent advertisement packet, and vehicles can find the nearest located AP by constantly monitoring the packet and receive a care-of address
(CoA). Since vehicles are powered by a re-chargeable battery, they do not operate a doze mode.

Second, on receiving the CoA, the vehicle registers with the HA by sending a registration request packet containing home, CoA, and HA addresses to the AP. Then the AP forwards the packet to the GFA, where it binds the CoA and AP, saves them into a location registry ($R_h$), and forwards the packet to the HA. Then the HA binds the vehicles’ home address and the current CoA, and saves them into a location registry ($R_h$). If the GFA receives a packet from different AP which is under the same regional network, then it does not forward the packet to the HA. Here, both GFA and HA have the most update binding information in their location registries.

Third, when a server sends a data item to a vehicle, it sends the data item attached with the vehicle’s IP address without knowing the vehicle’s whereabouts. Here, let us assume that the server knows the vehicle’s IP address before it sends the data item. Since the Internet does not know where to route the data item, it routes the data item to the home network of the vehicle based on the standard routing mechanism of the Internet [21]. The HA intercepts the data item and searches the IP address in its location registry, and forwards it to the corresponding GFA. On receiving the data item, the GFA forwards it to the corresponding AP, where the vehicle is currently located under its coverage area.

Finally, we do not deploy an advanced location management scheme such as [22] to optimize the communication cost for the sake of simplicity.

4.2. The Cache Invalidation

In this paper, we propose a cache invalidation scheme combining with a mobile IP based location management to reduce the number of broadcast operations and the query delay in VANETS. In the scheme, a server asynchronously sends an IR to a home agent (HA) rather than blindly broadcasts it to the vehicles. Then the HA judiciously refines and distributes the IR to appropriate gateway foreign agents (GFAs), where they can answer the validity of the queried data item. The rationale behind this is that since vehicles move at a high-speed and visit multiple number of coverage areas, if a prior IR-based approach is used, server may broadcast an IR to the wrong coverage area where the vehicle already has left. Thus the broadcast operation becomes expensive and inefficient. Also unlike a prior IR-based approach, where all nodes are one-hop away from the server and wait for the periodically broadcasted IR resulting in an unavoidable query delay, vehicles may fail to receive the IR because they move out of the coverage area while waiting for the IR. In addition, the same content of IR to different vehicles is wasteful in VANETS. Thus, it is our primary concern to consider the impact of high-speed mobility in developing the cache invalidation scheme. The detail mechanism is followed.

First, the server maintains information about which data item ($d_x$) is accessed by which vehicle ($v_y$) in a registry ($R_s$), $\{id(d_x), [\text{vid}(v_{x,y}), t_{x,y}]\}$, where $v_{x,y} = \{v_y\} | \text{request}(d_x) \land (y \in N)\}$, and $t_{x,y}$ is the access time of $d_x$ by $v_y$. Here, $N$ is a number of vehicles. Whenever a data item is updated, the server generates an IR (later IR(s)), $< id(d_x), \text{vid}(v_{x,y}), t_{\text{cur}} >$, and it is not directly broadcasted to the vehicles but sent to the HA. Here, $t_{\text{cur}}$ is a current timestamp. Upon receiving the IR(s), since the HA maintains location information about which vehicle is currently roaming under which GFA ($g_i$) in a registry ($R_h$), $\{\text{vid}(v_y), g_i\}$, it compares the $v_y$ in both received IR(s) and the $R_h$. Then the HA creates a new IR (later IR(h)) based on the matched $g_i$, $< id(d_x), t_{\text{cur}} >$, and sends it to the appropriate GFA. Depending on the number of matched GFAs, more than one IRs(h) can be created. Also each IR(h) may have different size, which is shorter than the original IR(s) received from the server. Upon receiving the IR(h), the GFA updates its registry and thus, it has the most recently updated IR with the server.

Second, when a vehicle ($v_y$) initially connects to a server, it sends all the ids of cached data items ($d_k$) and their timestamps, $[\text{vid}(v_y), id(d_k), t_k]$. Here, $1 \leq k \leq c$ and $c$ is a total number of cache slots. Then the server updates its registry. Whenever a vehicle sends a request packet for downloading/validating a data item, the server also updates its registry.

Finally, when a vehicle moves into the coverage area which is covered under different GFA, then the HA forwards the IR(h) to the new GFA again. The IR(h) may be forwarded to the appropriate APs to further reduce the query latency for valid cached copies. However, it is an overhead for the GFA when a vehicle frequently roams coverage areas under the several APs.

4.2.1. The Query Request Based on the proposed cache invalidation scheme, the detail operations for a query request are followed. First, when a vehicle ($v_y$) generates a query for a data item ($d_x$) which is not cached, it sends a request packet for requesting the data item to the server. The packet contains the vehicle’s CoA, an id of the requested data item, and a single bit flag representing whether the queried data item is cached or not, $[\text{CoA}_y, id(d_x), f_x]$, and it is sent to the nearest located AP. Here, $f_x$ is 0. As the packet is forwarded to the server through the AP and GFA, CoAs of the AP and GFA are appended in the packet header, $[\text{CoA}_{GFA}, \{\text{CoA}_{AP}, \text{CoA}_y, id(d_x), f_x\}]$, to keep the route information. Upon receiving the request packet, the server attaches the queried data item to the ack packet with the route information. The ack packet is replied back
to the AP via the GFA. Then the AP broadcasts the packet and thus, the corresponding vehicle can answer the query.

Second, when a query is generated which can be answered by a cached copy of data item, the vehicle sends a request packet for validation to the GFA through the nearest located AP. The GFA compares the id of queried data item with the IR(h) and replies an ack packet, if the cached copy is valid. If the cached copy is not valid, then the GFA forwards the packet to the server. When the server receives the request packet, it verifies the status of the queried data item and replies an ack packet with the route information, if it is a valid copy. If not, it is same as the procedure for requesting the data item which is not cached.

In summary, based on the underlying location management scheme, the server does not blindly broadcast an IR but sends it to the HA. Then the HA judiciously refines and distributes the IR to the GFAs, where the queried data items can be validated.

5. Numerical Results

In order to evaluate the performance of the proposed scheme, we develop a simulation model, in which the communication cost of location management and cache invalidation is considered as a key performance metric. The unit cost of transmission and process at network components in the system model are similarly defined to [22]. We use a one-dimensional wrap-around highway area, where 25 APs cover the area. We assume that each AP with a diameter 750 (m) is located in the center of coverage area. Also each GFA consists of five APs. Initially, the vehicles are randomly located in the area and move with a give velocity. Both query and update inter arrival times are assumed to be exponentially distributed. In addition, we assume that the total number of data items stored in the database is 1000 and they are classified into two subsets: cold and hot data items. Also 80% of the queries is for hot data items, and an update request is uniformly distributed within the hot and cold subsets. Vehicles cache 10% of data items of the database.

In this paper, we consider two more cache invalidation schemes based on the proposed system model for a comparison study, where our scheme is namely a location management based cache invalidation (LMC): a no cache (NC) and an on demand (OD). Although no cache is not practical in real because most applications cache data items what they have processed implicitly, we include this scheme for a comparison purpose. Queries generated from each vehicle should be forwarded to the server for validation. In the on demand, each vehicle caches data items but a server does not broadcast an IR. Like the no cache case, queries generated from each vehicle should be forwarded to the server for validation and access of the queried data item.

First we observe the effect of mobility and show the communication cost as functions of velocity (m/s, 19.4460 to 33.3360 (= 70 to 120 km/h)) and density (vehicles/m², 0.00002 to 0.00012 (= 20 to 120 vehicles/km²)). Since vehicles do not change their velocity or direction frequently, we use a fluid flow model [23] to capture the mobility in VANETS. The cost includes the home and regional location updates. In Figure 2 (a), as the velocity and density increase, the cost increases but it is more sensitive to the density.

Next, we examine the effects of query rate (λ_q), update rate (λ_u), and data item size (n_p) on the communication cost of cache invalidation. The cost includes answering queries, broadcasting IRs, invalidating cached data items, and updating IRs. Here, we assume that a data item consists of a variable number of packets depending on its size. In Figure 2 (b), as the query rate increases, the cost of all the schemes increases. Even though there is an overhead to update IRs stored in the GFAs compared to both no cache and on demand schemes, our scheme shows better performance than the others and almost 40% improvement is achieved when query rate is high. With caching, there is a high probability of the requested data items being validated in the GFAs before queries are further forwarded to the server, and thus less cost is achieved.

In Figure 2 (c), although the update rate increases, the cost has no or a little change in the no cache and on demand schemes, respectively. This is because queries generated under both schemes are validated only by the server. However, our scheme shows slightly higher cost than the others.
when the update rate is high. Due to the overhead of broadcast/update IRS, caching frequently updated data items does not contribute to a performance gain. In fact this is the disadvantage of caching frequently updated data items as observed in [24]. As shown in Figure 2 (d), our scheme outperforms the others and achieves the less cost for entire period. The performance gap with the no cache scheme is larger when the data item size increases. Due to the additional communication overhead with the server for validation and access of data items, both on demand and no cache schemes show the higher costs.

In summary, our approach provides the performance superiority in overall except the high update rate.

6. Concluding Remarks and Future Work

In this paper, we investigate the cache invalidation issue, and propose a novel system model and a cache invalidation scheme combining with a mobile IP based location management for IVANETS. We compare the proposed scheme with two other schemes and show better performance with respect to the communication cost. There are many interesting issues that need further investigation to exploit the full potential of the proposed scheme. Currently, we are developing an analytical model and examining a design tradeoff that balances the communication cost and complexity of algorithm.

References