Abstract—Mobility in vehicular networks naturally leads frequent handoffs and reassociations between the vehicles and roadside access points. The overhead due to these reassociation comes in the form of the bandwidth and delays associated with the handoff related packets that are exchanged between a vehicles and the access point. This paper derives an information theoretic lower bound on the Medium Access Control (MAC) layer overhead associated with reassociations caused due to node mobility. An efficient MAC protocol is then proposed that reduces the handoff delays. Simulations are used to demonstrate the superior performance of the proposed protocol in terms of the throughput and packet latency.

I. INTRODUCTION

Vehicular networks comprise of vehicle-to-vehicle and vehicle-to-roadside communications based on local area networking technologies [12]. Such networks require secure, low latency communications in the presence of mobility, and are typically built in an ad-hoc manner with short durations of contact between nodes [12]. Vehicular mobility introduces a number of challenges for providing seamless connectivity in these environments with frequent handoffs being a problem of particular concern. In this paper we first provide a bound on the reassociation overhead caused due to handoffs. Then we propose a simple MAC layer technique to reduce the reassociation times in vehicle-to-roadside networks.

The infrastructure behind the vehicle-to-roadside networks targeted in this paper consists of fixed roadside access points. Such points may be located at street corners, co-located with street signals, placed in parking lots, rest areas or emergency phones on a highway etc. The need for frequent handoffs and reassociations with the access points and the associated overhead degrades the performance of many applications. Consequently, considerable research has been done on reducing the handoff times. This paper provides a lower bound on the reassociation rate required to ensure that a vehicle is connected to an AP in its range when it has data to send or receive, irrespective of the handoff mechanism employed. This lower bound is applicable to all proposals that have been proposed in literature to reduce the handoff times.

Both centralized as well as distributed MAC protocols have been proposed for mobile and vehicular networks. Among these, protocols based on centralized assignment of resources are less flexible in adapting to the changing set and number of nodes in vehicular networks, making it difficult to dynamically allocate the slots, channels or codes. Decentralized MAC protocols may thus perform better in such environments [21]. Also, all protocols have to specifically address the problem of efficient handoffs in vehicular networks while being adaptive to the network dynamics. Existing work on reducing the handoff times focus mainly on preemptive probing or scanning for APs and predictive schemes. To complement these approaches, we propose a mechanism that aims to reduce the channel access time of the handoff related packets in order to reduce the overall reassociation time. Our methodology can be used in conjunction with any of the existing schemes for reducing the overall handoff delay.

The main contribution of this paper is an information theoretic bound on the MAC layer overhead due to node reassociations resulting from node mobility. This analysis is independent of the MAC protocol used in the network. In addition we propose a MAC protocol that combines aspects of centralized and decentralized protocols to reduce the channel access times of handoff related packets to reduce the handoff times. The parameter selection process minimizes the packet delays while ensuring fairness. The proposed protocol eliminates simultaneous data transmissions by hidden nodes and provides priority access for reassociation or disassociation. Simulation results using realistic mobility models on actual city maps are used to evaluate and demonstrate the superior throughput and delay characteristics of the proposed protocol.

The rest of the paper is organized as follows. In Section II we discuss the related work. Section III presents an information theoretic lower bound on the reassociation overhead. Section IV presents the proposed protocol and its parameter selection process. Finally, Section V presents the simulation results and Section VI concludes the paper.

II. RELATED WORK

There has been considerable work on techniques to reduce the handoff delays associated with various MAC protocols, with IEEE 802.11 receiving the greatest attention. Handoff delay reduction techniques have been proposed based on continuous probing and scanning for base stations [14], using location information of the base stations [16], using probe relays [16], middleware [1], using cooperating nodes [5], and predictive techniques [6]. However, these techniques only reduce the time required to locate a new access point and not the MAC layer delay required to access the channel for
transmitting the handoff related packets. In contrast, our goal is to reduce the MAC layer channel access time for the handoff packets.

Many existing MAC protocols like reservation-ALOHA, CSMA, CSMA/CA and IEEE 802.11 have been applied to roadside to vehicle networks [12]. ALOHA based protocols suffer from the risk of instability in the case of many participating nodes and frequent reservation attempts [12]. The token ring based approach for vehicular communications in [11] relies on the physical layer of IEEE 802.11 and has not been evaluated in networks with mobility. MAC protocols that combine CDMA with random channel access have been proposed in [13] but suffer from multi-access interference resulting in secondary collisions and have not been evaluated under mobility. The repetition based MAC protocol in [21] repeats each message in a number of slots to ensure reliability, at the cost of redundant transmissions.

The basic aim of the proposed MAC protocol is to reduce the handoff delays by reducing the channel access times of handoff related packets while being adaptive to the changing conditions in a vehicular network. Now, MAC protocols using demand based assignment exist in literature where nodes use contention based schemes to reserve future time slots for data transmission [9]. While the proposed protocol has some similar features, we introduce mechanisms specifically for vehicular networks such as priority mechanisms for handoffs and collision based inference to determine the parameter settings for delay minimization.

III. REASSOCIATION OVERHEAD

This section presents an information theoretic bound on the overhead due to node mobility induced handoffs or reassociations. We obtain the minimum required reassociation rate so that the probability that each node is associated with an AP in its range when it has data to send or receive, is greater than an arbitrary value $1 - \epsilon$. The overhead evaluated in this paper arises due to the mobility of the vehicles and characterizes the frequency of handoffs. This overhead does not depend on the actual sequence of packets exchanged during the handoff which may be either bidirectional between the user and the AP (e.g. IEEE 802.11) or unidirectional (e.g. IEEE 802.11p). Thus this analysis is applicable to any MAC protocol where handoffs occur when a vehicle moves out of range of its current AP such as the protocol proposed later in this paper, in addition to IEEE 802.11, IEEE 802.11p, and all other MAC protocols and techniques to reduce the handoff latencies listed in the previous section.

We formulate the problem of obtaining the handover overhead as a rate distortion problem. For each vehicle, the probability that it is currently associated with an AP within its transmission radius (as opposed to a stale association with an AP no longer in its range), is used as the distortion measure. As in rate distortion problems, we then find the minimum reassociation rate required (i.e. rate of information exchange in rate distortion) to ensure that the distortion measure is greater than a desired value. We assume that an arbitrary set, $\mathcal{N}$, of vehicles is randomly and uniformly distributed on a two dimensional (2-D) plane and their movement is governed by a 2-D random walk in continuous time. This assumption is justified from [4] which shows that the lengths of roads in an urban environment is Rayleigh distributed, as is the case for displacement in a 2-D Brownian motion. The Brownian motion model may not be applicable for vehicle movement in highways or grid-type roads. The position of node $j$ at time $t$ is denoted by $x_j(t), y_j(t)$ and the distance between two nodes $i$ and $j$ at time $t$ is given by $\Delta_{ij}(t) = \sqrt{(x_i(t) - x_j(t))^2 + (y_i(t) - y_j(t))^2}$. All vehicles and APs are assumed to have a transmission range of $r$.

Consider an arbitrary AP, say AP $s$ and denote by $N_s(t) = \{j : \Delta_{sj}(t) \leq r, j \in \mathcal{N}\}$ the set of vehicles associated with it that are actually in its range at time $t$ and by $\hat{N}_s(t)$ the set of vehicles that are associated with it and perceive themselves to be in its range. A node may perceive itself to be an AP’s range when it is not, or vice versa, due to use of outdated association information. Define

$$Z_{sj}(t) = \begin{cases} 1 & \text{if } j \in N_s(t) \\ 0 & \text{otherwise} \end{cases} \quad \hat{Z}_{sj}(t) = \begin{cases} 1 & \text{if } j \in \hat{N}_s(t) \\ 0 & \text{otherwise} \end{cases}$$

as variables to indicate whether node $j$ actually is or perceives to be AP $s$’s neighbor or not. The difference

$$E_{sj}(t) = Z_{sj}(t) - \hat{Z}_{sj}(t)$$

(1)

denotes the accuracy of the association information of node $j$. It is desired that $E_{sj}(t) = 0$ at all times for all $j$. We now state the minimum reassociation rate problem in terms of $E_{sj}(t)$.

Minimum reassociation rate problem: What is the minimum rate at which a vehicle has to reassociate with APs such that

$$P[E_{sj}(T^k_j) = 0] \geq 1 - \epsilon, \quad \forall \, j \in \mathcal{N}, \quad 1 \leq k < \infty$$

(2)

where $T^k_j$ is the instance when the $k$-th packet to be sent to, or from node $j$ is generated.

We formulate the minimum reassociation rate problem as a rate distortion problem. Our analysis is based on extending the results of [17] which considers the problem of protocol overhead in location based routing. We denote by $Z_{sj}^N$ and $\hat{Z}_{sj}^N$ the vectors

$$Z_{sj}^N = \{Z_{sj}(T^j_s), Z_{sj}(T^2_s), \ldots, Z_{sj}(T^N_s)\}$$

$$\hat{Z}_{sj}^N = \{\hat{Z}_{sj}(T^j_s), \hat{Z}_{sj}(T^2_s), \ldots, \hat{Z}_{sj}(T^N_s)\}$$

and denote by $P_N(\epsilon)$ the family of joint probability distribution functions of $Z_{sj}^N$ and $\hat{Z}_{sj}^N$ such that $P[E_{sj}(T^k_j) = 0] \geq 1 - \epsilon, \quad \forall \, j \in \mathcal{N}$ and $1 \leq k < \infty$. We also denote by $R_N(\epsilon)$ the minimum reassociation rate such that $P[E_{sj}(T^k_j) = 0]$ and as per the discussion on rate distortion in [7], it is given by

$$R_N(\epsilon) = \min_{P_N(\epsilon) \in P_N(\epsilon)} \frac{1}{N} I_P(N, \hat{Z}_{sj}^N)$$

(3)

where $I_P(N, \hat{Z}_{sj}^N)$ is the mutual information between $Z_{sj}^N$ and $\hat{Z}_{sj}^N$. The minimum reassociation rate, $R(\epsilon)$, is then

$$R(\epsilon) = \lim_{N \to \infty} \min R_N(\epsilon)$$

(4)

We now obtain a bound for $R_N(\epsilon)$ and consequently $R(\epsilon)$ by evaluating a bound for $I_P(N, \hat{Z}_{sj}^N)$.
Claim 1: The minimum reassociation rate $R(\epsilon)$ satisfies
\[ R(\epsilon) \geq R_1(\epsilon) \]  

Proof: To prove Eqn. (5), we use the following relationship proved in [18], governing the mutual information between $Z_s^N$ and $\hat{Z}_s^N$:
\[ \inf_{P_{x_N} \in P_N(\epsilon)} I_{P_{x_N}}(Z_s^N; \hat{Z}_s^N) \geq N R_1(\epsilon) \]

Note that the above expression uses the assumption that $Z_s(T_j^1)$ and $Z_s(T_j^{k-1})$ are independent of each other. To prove the claim we note that the definition of the rate distortion function gives us
\[ R_N(\epsilon) = \min_{P_{x_N} \in P_N(\epsilon)} \frac{1}{N} I_{P_{x_N}}(Z_s^N; \hat{Z}_s^N) \geq \frac{1}{N} N R_1(\epsilon) = R_1(\epsilon) \]

and thus
\[ R(\epsilon) = \lim_{N \to \infty} R_N(\epsilon) \geq R_1(\epsilon) \]

which proves Eqn. (5) holds and thus proves the claim.

Next, we find a bound on $R_1(\epsilon)$ in order to bound $R(\epsilon)$. We consider two cases: (1) $Z_s(0) = 1$ and (2) $Z_s(0) = 0$ depending on whether node $j$ is initially in the range of AP $s$ or not. $R_1(\epsilon)$ is then bounded by the maximum of the rate distortion functions for these two cases.

Case 1: Denote by $L_j$ the region in space of possible positions for node $j$ at time $t = 0$ such that $Z_s(0) = 1$, i.e. $L_j = \{x_j, y_j : \sqrt{(x_j^2 + y_j^2)} \leq r\}$.

Claim 2: The rate distortion function in case 1 (C1), $R_{1,C1}(\epsilon)$ is bounded by
\[ R_{1,C1}(\epsilon) \geq \max_{x_j(0), y_j(0) \in L_j} H(Z_s(T_j^1)) + \epsilon \log \left( \frac{\epsilon}{2} \right) + (1 - \epsilon) \log(1 - \epsilon) \]

Proof: From the definition of mutual information
\[ I_{P_{x_N}}(Z_s(T_j^1); \hat{Z}_s(T_j^1)) \geq H(Z_s(T_j^1)) - H(E_s(T_j^1)) \]

Since $Z_s(T_j^1)$ and $E_s(T_j^1)$ take on a value of either 0 or 1, its probability mass function can be written in terms of some $p_1$, $p_2$ and $p_3$ as
\[ E_s(T_j^1) = \begin{cases} -1 & \text{w.p. } p_1 \\ 0 & \text{w.p. } p_2 \\ 1 & \text{w.p. } p_3 \end{cases} \]

where $P[E_s(T_j^1) = 0] = p_2 \geq 1 - \epsilon$ and $p_1 + p_2 + p_3 = 1$. Thus we have $p_1 + p_3 \leq \epsilon$. The entropy of $E_s(T_j^1)$ is then given by $H(E_s(T_j^1)) = -p_1 \log p_1 - p_2 \log p_2 - p_3 \log p_3$ which is maximized when $p_2 = 1 - \epsilon$ and $p_1 = p_3 = \epsilon/2$. This maximum entropy is given by
\[ H(E_s(T_j^1)) = -\epsilon \log \left( \frac{\epsilon}{2} \right) - (1 - \epsilon) \log(1 - \epsilon) \]

$H(Z_s(T_j^1))$ depends on the position of node $j$ at $t = 0$ and the reassociation rate should account for the initial location that results in the maximum entropy. Substituting Eqn. (10) in Eqn. (8) we then have
\[ I_{P_{x_N}}(Z_s(T_j^1); \hat{Z}_s(T_j^1)) \geq \max_{x_j(0), y_j(0) \in L_j} H(Z_s(T_j^1)) + \epsilon \log \left( \frac{\epsilon}{2} \right) + (1 - \epsilon) \log(1 - \epsilon) \]

To obtain $H(Z_s(T_j^1))$ we note that if $P[Z_s(T_j^1) = 1] = \delta$ then $H(Z_s(T_j^1)) = -\delta \log \delta - (1 - \delta) \log(1 - \delta)$. We now obtain the probability $P[Z_s(T_j^1) = 1]$ for this case by obtaining $P[Z_s(T_j^1) = 1 | \Delta_s(0) = l, T_j^1 = \tau]$ with $l \leq r$ and then unconditioning on $\tau$. Since the node $j$ follows a two dimensional random walk with variance $\alpha$, unconditioning the result of Appendix case 1, Eqn. (22), on packet interarrival times (which have a pdf $f_\tau(\tau)$), we have
\[ P[Z_s(T_j^1) = 1 | \Delta_s(0) = l] = \int_0^\infty \int_0^{l} \frac{2 \pi e^{-x^2} e^{-\frac{x^2}{2\alpha}}} {\pi \alpha} f_\tau(\tau) d\tau \]

Claim 3: The rate distortion function in case 2 (C2), $R_{1,C2}(\epsilon)$ is bounded by
\[ R_{1,C2}(\epsilon) \geq \max_{x_j(0), y_j(0) \in L_j} H(Z_s(T_j^1)) + c \log \left( \frac{\epsilon}{2} \right) + (1 - \epsilon) \log(1 - \epsilon) \]

Proof: The proof is identical to that for Case 1.

Using the results of the Appendix Case 2, Eqn. (23),
\[ P[Z_s(T_j^1) = 1 | \Delta_s(0) = l] = \int_0^\infty \int_{l-r}^{l+r} 2 \cos^{-1} \left( \frac{-r^2 + 2x^2 + \alpha^2}{2\alpha x} \right) e^{-\frac{x^2}{2\alpha}} f_\tau(\tau) d\tau \]

The maximum $H(Z_s(T_j^1))$ is achieved when $l = r$ and we denote this entropy by $H_{c2}(Z_s(T_j^1))$. Then
\[ R_{1,C2}(\epsilon) \geq H_{c2}(Z_s(T_j^1)) + c \log \left( \frac{\epsilon}{2} \right) + (1 - \epsilon) \log(1 - \epsilon) \]

The lower bound on the reassociation rate is then
\[ R(\epsilon) \geq R_1(\epsilon) \geq \max \{ R_{1,C1}(\epsilon), R_{1,C2}(\epsilon) \} \]

IV. MAC LAYER SUPPORT FOR REDUCING HANDOFF DELAYS

This section describes the proposed protocol which aims to reduce the reassociation delays during handoffs. All nodes and APs are assumed to be equipped with a single radio and can transmit on a single channel at any given time. Neighboring APs transmit on different channels while individual nodes transmit on the channel of the AP they are associated with and scan channels to find new APs during handoff.

The operation of the protocol is divided in cycles of variable lengths. A cycle begins with the transmission of a beacon by the AP. The beacon contains the AP’s identity as well
as information on the number of backoff slots in the cycle. The beacon is followed by backoff slots, termed ASC slots, reserved for nodes that wish to reassociate or disassociate with the AP. The ASC slots are followed by the data contention slots. Data contention slots are used by nodes to convey bandwidth reservation requests to the APs. Actual data transmissions follow once the data contention slots are over.

1) Reassociation and disassociation: When a node overhears a beacon from an AP that it wants to reassociate or disassociate with, it notes the number of ASC slots, say $n$, specified in the beacon. It then picks one of the $n$ slots at random to transmit its reassociation or disassociation packet. In case no other node transmits in the same slot, the AP successfully receives the packet and replies with an ACK, completing the reassociation or disassociation. In case of a collision, the node has to wait for the next beacon.

2) Data transmission: When a node wishes to send a data packet, it first waits for a beacon and the following ASC slots to pass. Then, using the number of data contention slots specified in the beacon, it selects a slot at random to send a medium reservation (or bandwidth) request to the AP. If there is no collision at the AP due to other nodes selecting this slot, the AP responds with an ACK confirming the reservation and specifying the time when the node may transmit. The node waits till this time and then transmits its data (if a node does not receive an ACK for a reservation request, it assumes that there was a collision at the AP and waits for the next cycle to retransmit the request, again by selecting a contention slot at random). The node expects an immediate ACK for the data from the AP and in case it is not received, assumes that the data was lost and attempts a retransmission in the next cycle (a limit may be placed on the number of retransmission attempts). The node defers from transmitting at all other times, unless it is sending an ACK for a data frame sent to it by the AP.

3) Transmissions from the AP: The AP transmits the beacons and ACK frames for reassociation/disassociation requests, bandwidths requests and on the successful reception of data packets. The AP may also send out downstream data packets, both unicast and broadcast. Data packets may be either piggybacked to the ACKs that it sends to a node or sent at any point during the data transmission phase of the cycle. The AP accounts for the downstream traffic while assigning the transmission times to nodes with bandwidth requests. The AP expects an ACK for its unicast data but not for broadcast data. Since power is less of a concern in vehicular networks, nodes may keep their radios on at all times (e.g. for obtaining real-time traffic information). Thus the AP does not waste bandwidth on specifying the timings for downstream traffic. Finally, the AP dynamically updates the number of ASC and data contention slots, as per the procedure in Section IV-B.

4) Bandwidth Allocation and Robustness: An AP may allocate bandwidth to the nodes based on its fairness or priority considerations. In the default operation, a single slot is assigned in a frame to each node with a successful contention. In addition, nodes with delay sensitive packets may be scheduled at the beginning of the data transmission part of the frame. The basic operation of the MAC protocol is independent of the priority or fairness mechanism used. Also, the operation of the protocol is critically dependent on the reception of the beacon frame by the nodes. To provide robustness against link errors, the beacon may be transmitted at a lower rate (or protected with error correcting codes). The communications in the ASC slots may be similarly protected to avoid delays in the reassociations.

5) Example: Figure 1 presents an example of the protocol’s operation over one cycle. The beacon specifies three ASC and five data contention slots. Node 1 wants to reassociate with the AP and picks ASC slot 2 at random. The AP replies with an ACK confirming the reassociation. Nodes 1 and 2 want to send packets to the AP and they pick slot 2 and 4 respectively for their requests, which are responded with ACKs by the AP. Once nodes 1 and 2 transmit their data, the AP sends them ACKs confirming the reception. The AP then sends a data packet to node 3 which replies with an ACK.

A. Protocol Features

Priority for Handoff Packets: Nodes requiring handoff are provided with special, exclusive slots (ASC slots), significantly reducing the likelihood of collisions. Also, ASC slots occur at the beginning of a cycle and nodes that successfully reassociate with an AP in a cycle are eligible to compete for bandwidth reservation in the same cycle. This further reduces the latency of data packets in the presence of handoffs.

Hidden Terminals: In the proposed protocol, nodes are not allowed to transmit data unless the channel has been specifically assigned to it during that time. This eliminates the possibility of two mutually hidden nodes transmitting data simultaneously to the AP and thus collisions due to hidden nodes.

Fairness: In the proposed protocol, the number of contention slots available to any node associated with an AP is the same. Thus all these nodes have statistically similar throughput and delay parameters which ensures fairness. Also, the AP can compare the contention success rate of a node with the overall collision rate in contention slots to determine if a node has been cheating while sending its requests.

B. Parameter Setting

This section describes the methodology used by APs to estimate the number of active nodes in the network and using it to obtain the optimal number of data contention slots.
1) Estimation of the Number of Active Nodes: In each cycle, APs monitor all data contention slots and mark them as idle, successful or collision and then use this information to estimate the number of active users. Prior work has used active probing [15] and collision based inference using Kalman and non-linear filters [2], [8] to estimate the number of active nodes. These techniques have drawbacks such as the assumption of Gaussian observations and divergence due to non-linear dependence on state. We next develop an estimation technique without the drawbacks of the existing schemes.

Given that there are $M$ data contention slots in a cycle, let $n_0$, $n_1$ and $n_c = M - n_0 - n_1$ be the number of empty (e), successful (s), and collision (c) slots, respectively. Let the outcome of the $M$ slots be represented by a vector $y_t$ with $M$ elements, each of which can be e, s or c. An active node transmits in any one of the $M$ slots with equal probability, $1/M$. Then given that there are $x_t$ active nodes, the probability of observing the slot occupancy given by vector $y_t$ is given by

$$p(y_t | x_t = i) = \begin{cases} 0 & n_c > \lfloor \frac{i-n_1}{2} \rfloor \\ \sum_{c_1, \ldots, c_{n_c} \geq 2} \frac{c_1! \cdots c_{n_c}!}{c!} \frac{1}{M^i} & n_0 < M - i \end{cases}$$

(12)

where $c_1, c_2, \ldots, c_{n_c}$ are dummy variables representing the number of nodes in each collision. The problem of estimating the number of active nodes is then to estimate $x_t$ based on the observation of the slot occupancy. We use a Hidden Markov Model (HMM) to estimate the number of active nodes, $x_t$, in the network at any time. The sequence of the slot occupancy observations $y_t$ till time $t$ is denoted by $y_t = [y_1, y_2, \ldots, y_t]$ and the network state sequence upto time $t$ is $x_t = [x_1, x_2, \ldots, x_t]$. The HMM is governed by

$$x_t \sim M(\pi, A), \quad y_t \sim B(x_t)$$

(13)

where $M(\pi, A)$ denotes a Markov chain with initial probability distribution $\pi$ and transition probability matrix $A$. Both $\pi$ and $A$ are unknown. Based on the observations $y_t$, we wish to determine the $x_t$ that yields the maximum a posteriori probability $p(x_t | y_t)$. To obtain the $x_t$ that maximizes $p(x_t | y_t)$, we use an approximate maximum a posteriori (MAP) algorithm developed in [20] that is a modification of the Viterbi algorithm. From Bayes’ theorem:

$$p(x_t | y_t) = p(y_t | x_t) p(x_t | y_{t-1}) p(x_{t-1} | y_{t-1}) p(x_{t-1} | y_{t-1})$$

The approximate MAP approach aims to recursively maximize $p(x_t | y_t)$ with respect to $x_t$. To achieve this the Viterbi algorithm uses $\delta_t(i) = \max_{x_{t-1}} p(x_{t-1} | x_{t-1} = i) p(x_t | y_t)$, or equivalently

$$\delta_t(i) = \max_{x_{t-1}} p(y_t | x_t = i) \max_{x_{t-1}} p(x_{t-1} | y_{t-1})$$

(14)

that can be computed recursively if the transition matrix, and thus $p(x_t | x_{t-1}, y_{t-1})$ is known. Since this matrix is unknown here, we make the approximation that $p(x_{t-1} | y_{t-1})p(x_t | x_{t-1}, y_{t-1})$ is maximized when $p(x_{t-1} | y_{t-1})$ is maximized. An approximation $\hat{\delta}_t(i)$ of $\delta_t(i)$ can then be computed recursively as

$$\hat{\delta}_t(i) = p(y_t | x_t = i) \max_j \left[ \hat{\delta}_{t-1}(j) p(x_t = i | x_{t-1}, y_{t-1}) \right]$$

$$= p(y_t | x_t = i) \max_j \left[ \hat{\delta}_{t-1}(j) \frac{\alpha_{j,t-1}^{(j)}}{\sum_{k=1}^{\max_j} \alpha_{j,k,t-1}^{(j)}} \right]$$

(15)

where $p(y_t | x_t = i)$ is given in Eqn. (12), $x_{t-1}^{(j)}$ is the retained path ending at $x_{t-1} = j$ and $\alpha_{j,t-1}^{(j)}$ is the corresponding sufficient statistic for the probability of transition from state $j$ to state $i$, updated using

$$\alpha_{j,i,t-1} = \alpha_{j,i,t-2} + I(x_{t-2} - j) I(x_{t-1} - i)$$

(16)

with $I(x) = 1$ if $x = 0$ and $I(x) = 0$ otherwise. Our estimate of $x_t$ at time $t$ is the state that maximizes $\hat{\delta}_t(i)$.

2) The Optimal Number of Contention Slots: While a larger number of contention slots decreases the chances of collisions, it also decreases the throughput since less time is now spent on actually transmitting data packets. To obtain the optimal number of contention slots, we need to characterize the expected packet delay as a function of the number of contention slots, $M_{est}$, in a cycle. The expected number of new nodes that reassociate with an AP in a cycle is usually small and thus a fixed number of slots, say one or two is sufficient for this purpose. Thus in this section we only focus on data contention slots. We denote the estimated number of active nodes in the cycle, as obtained in the previous subsection, by $\hat{x}_t$. A node picks one of the $M_{est}$ slots to transmit its request with equal probability of $1/M_{est}$. Now any other node does not pick the same slot with probability $1 - 1/M_{est}$. The collision probability for an active node, $p_c$, is then

$$p_c = 1 - \frac{1}{M_{est}}$$

(17)

In case the reservation is successful (with probability $1 - p_c$), the packet is successfully transmitted in the same cycle and the delay experienced is $T_{cycle}$, the expected length of a cycle. If the reservation request experiences a collision (with probability $p_c$), it is retransmitted in the next cycle. The number of transmission attempts is thus geometrically distributed and the expected number of transmission attempts is

$$E[\text{attempts}] = \sum_{i=1}^{\infty} i p_c^{i-1} (1 - p_c) = \frac{1}{1 - p_c}$$

(18)

Each cycle consists of: the beacon, ASC slots denoted by $M_{asc}$, $M_{est}$ data contention slots and the time spent on transmitting the data. Since the reservation requests of each of the $\hat{x}_t$ active nodes is successful in the cycle with probability $1 - p_c$, on an average, a time of $(1 - p_c)\hat{x}_t T_{data}$ is spent on transmitting data, where $T_{data}$ is the average time required to transmit a data packet. Finally, with the duration of a beacon, an ASC slot and a data contention slot denoted by $T_B$, $T_{ASC}$ and $T_{data}$ respectively, we have

$$T_{cycle} = T_B + M_{asc} T_{ASC} + M_{est} T_{data} + (1 - p_c)\hat{x}_t T_{data}$$

(19)

For each transmission attempt required for the reservation request, the packet experiences a delay of $T_{cycle}$. Thus the
information of major cities in the USA obtained from the MAC, and affect each MAC protocol equally. The scenarios considered for the simulations are from accurate roadmap data contention slots based IEEE 802.11p and the reservation based ADHOC MAC protocol as well as IEEE 802.11p and ADHOC [1], [5], [6] because they can be used in conjunction with the proposed protocol as well as IEEE 802.11p and ADHOC [3]. We do not consider any constraints we only show the results for AP C. The simulations use a default packet size of 1040 bytes and each ACK, association and reassociation packet is of 14 bytes. Each packet is preceded by a preamble of duration 20μsec, the length of a backoff slot is 20μsec, the length of a polling period in the proposed protocol is 300μsec, the length of a slot in ADHOC MAC corresponds to the time to send a packet and an ACK, and the DIFS and SIFS durations are 50μsec and 10μsec respectively. The channel data rate was 1Mbps, the length of each simulation run was 800 seconds and the average node speed was 55 miles per hour. Finally, while 802.11p allows for multiple service classes, the results reported correspond to minimum and maximum contention window sizes of 15 and 1023 respectively and AIFS=9 since they led to the best throughput and delay performance.

We first evaluate the effectiveness of the methodology of Section IV-B1 for estimating the number of active nodes. Table I compares our mean squared error in the estimate of the number of nodes with that of the Kalman filter based technique of [2] and the Density Inference Protocol (DIP) of [15]. In addition to the city section based random trip mobility model, we consider two other mobility models: random waypoint and random walk with wrapping [10]. The proposed method outperforms the other two.

Figures 3 and 4 compare the per node throughput and average packet delay for the proposed protocol, the ADHOC MAC protocol and IEEE 802.11p. The results show that the TIGER database of the US government [19]. Movement of the vehicles in the city roads was generated according to the realistic, random trip model developed in [10]. Simulation results were generated for Washington DC, Boston, Massachusetts and Houston, Texas. To avoid repetitive results, the paper only reports the simulation results for a section of Houston, an aerial map of which is shown in Figure 2. The simulator used for the results of all protocols was developed by us and written in C. Physical layer effects such as fading are not simulated since they affect transmissions of all MAC protocols equally.

The simulations use a default packet size of 1040 bytes and each ACK, association and reassociation packet is of 14 bytes. Each packet is preceded by a preamble of duration 192μsec, the length of a backoff slot is 20μsec, the length of a polling period in the proposed protocol is 300μsec, the length of a slot in ADHOC MAC corresponds to the time to send a packet and an ACK, and the DIFS and SIFS durations are 50μsec and 10μsec respectively. The channel data rate was 1Mbps, the length of each simulation run was 800 seconds and the average node speed was 55 miles per hour. Finally, while 802.11p allows for multiple service classes, the results reported correspond to minimum and maximum contention window sizes of 15 and 1023 respectively and AIFS=9 since they led to the best throughput and delay performance.

We first evaluate the effectiveness of the methodology of Section IV-B1 for estimating the number of active nodes. Table I compares our mean squared error in the estimate of the number of nodes with that of the Kalman filter based technique of [2] and the Density Inference Protocol (DIP) of [15]. In addition to the city section based random trip mobility model, we consider two other mobility models: random waypoint and random walk with wrapping [10]. The proposed method outperforms the other two.

Figures 3 and 4 compare the per node throughput and average packet delay for the proposed protocol, the ADHOC MAC protocol and IEEE 802.11p. The results show that the TIGER database of the US government [19]. Movement of the vehicles in the city roads was generated according to the realistic, random trip model developed in [10]. Simulation results were generated for Washington DC, Boston, Massachusetts and Houston, Texas. To avoid repetitive results, the paper only reports the simulation results for a section of Houston, an aerial map of which is shown in Figure 2.

The second derivative of Eqn. (20) is positive ensuring the existence of a minima.

\[ E[D] = \frac{T_{cycle}}{1-p_c} = \frac{\tau_B + M_{asc} \tau_{ASC} + M_{est} \tau_{data}}{(1 - \frac{1}{M_{est}}) \hat{x}_{t-1}} + \hat{x}_t \tau_{data} \]

\[ dE[D] = (\tau_B + M_{asc} \tau_{ASC} + M_{est} \tau_{data}) + \frac{\tau_{data}}{M_{est} \left[1 - \frac{1}{M_{est}}\right] \left(1 - \hat{x}_t\right)^{-1} \hat{x}_t^{1/2} \hat{x}_t - 1} \]

\[ dM_{est} = M_{est}^2 \left[1 - \frac{1}{M_{est}}\right] \left(1 - \hat{x}_t\right)^{-1} \hat{x}_t + \frac{\tau_{data}}{2 \hat{x}_t \tau_{data} + 4 \tau_{data} (\tau_B + M_{asc} \tau_{ASC}) (\hat{x}_t - 1)} \]

\[ M_{opt} = \frac{\hat{x}_t}{2} + \frac{\sqrt{\hat{x}_t^2 \tau_{data}} + 4 \tau_{data} (\tau_B + M_{asc} \tau_{ASC}) (\hat{x}_t - 1)}{2 \tau_{data}} \]

\[ \frac{\tau_{data}}{2} + \frac{\sqrt{\hat{x}_t^2 \tau_{data}} + 4 \tau_{data} (\tau_B + M_{asc} \tau_{ASC}) (\hat{x}_t - 1)}{2 \tau_{data}} \]

\[ \frac{\tau_{data}}{2} + \frac{\sqrt{\hat{x}_t^2 \tau_{data}} + 4 \tau_{data} (\tau_B + M_{asc} \tau_{ASC}) (\hat{x}_t - 1)}{2 \tau_{data}} \]
<table>
<thead>
<tr>
<th>Method</th>
<th>City Section</th>
<th>Mobility Model</th>
<th>N = 40</th>
<th>N = 60</th>
<th>N = 80</th>
<th>N = 40</th>
<th>N = 60</th>
<th>N = 80</th>
<th>N = 40</th>
<th>N = 60</th>
<th>N = 80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.01</td>
<td></td>
<td>0.02</td>
<td>0.07</td>
<td>1.62</td>
<td>0.04</td>
<td>0.36</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIP</td>
<td>1.48</td>
<td></td>
<td>1.22</td>
<td>2.14</td>
<td>5.12</td>
<td>1.87</td>
<td>2.90</td>
<td>4.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalman</td>
<td>0.54</td>
<td></td>
<td>0.56</td>
<td>0.67</td>
<td>1.17</td>
<td>0.63</td>
<td>0.78</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE I
MEAN SQUARE ERROR IN THE NUMBER OF ESTIMATED NODES FOR VARIOUS METHODS, MOBILITY MODELS AND TOTAL NODES IN THE NETWORK (N)

addition of ASC slots in the proposed protocol does not lead to an increase in the overall delay. This is because the proposed protocol optimally selects the channel contention parameters and is also better at suppressing hidden nodes (which are more likely to occur at high node densities). Thus the overall collision rates are lower and the packet delays are also smaller, compensating for the time added due to ASC slots. Table II shows the corresponding average delays experienced by reassociation packets. The proposed protocol achieves delays that are almost an order of magnitude lower than ADHOC MAC. This is because ADHOC MAC requires a node to wait for more than two frames before it can join an AP. Also, no results are shown for IEEE 802.11p as there is no need for explicit reassociation in this protocol. A node may associate with an AP as soon as it overhears a beacon and thus reassociation times may controlled by changing the beacon rate. While 802.11p may have lower reassociation times than the proposed protocol if beacon rates are made sufficiently high, the beacon overhead will further reduce the user throughput and increase the packet delays.\(^3\)

To further evaluate the performance of the proposed protocol in different settings, Table III compares the the three protocols under two additional mobility models on a 1400 × 1400 meter area: random waypoint and random walk with wrapping. The results follow the same trend as before. Next, Table IV compares the per node throughput, packet and reassociation delays in the case of bidirectional traffic. The results follow the same trend as those with unidirectional traffic.

To observe the effect of the per packet channel overhead on the protocol performance, Table V compares the performance of the three protocols in a network of 100 nodes for various packet sizes. For small packets, the time spent in the polling periods or channel contention becomes comparable to the time spent in transmitting data and the reservation based ADHOC MAC protocol thus performs better in terms of the throughput and delay, though its reassociation times are significantly higher. As the packet size increases, the proposed protocol outperforms the other two in all three aspects.

Finally Figure 5 compares the bound on the reassociation rate obtained from Eqn. (11) with simulations for practical reassociation schemes. In these results, the packet interarrival time is assumed to follow an exponential distribution (since it has the highest entropy among all continuous time distributions with base \([0, \infty)\) and a given mean). The periodic scheme (see figure caption) comes close to achieving the lower bound for small values of \(\alpha\). As \(\alpha\) increases, both schemes have the same reassociation overheads since the period required by the periodic scheme is so small that the vehicles essentially reassociate as soon as they move out of range of the previous AP. In contrast, the lower bound decreases as \(\alpha\) increases. This suggests that for very high variance levels, the locality of a mobile node may be established more efficiently by the absence of reassociation packets. However, such high variance levels are impractical in vehicular networks and are included here only to show the trend.

VI. CONCLUSIONS
This paper presents an information theoretic lower bound on the reassociation overhead of MAC protocols due to node mobility. An efficient MAC protocol for vehicle-to-roadside networks is also proposed that achieves low packet and handoff delays while maintaining fairness. Simulation results are used to verify the performance of the proposed protocol.

\(^3\)Our 802.11p simulations are for the best case where nodes join the AP immediately and the beacon overhead is neglected.
TABLE III
AVERAGE PER NODE THROUGHPUT (T IN KBPS), AVERAGE PACKET DELAY (D IN MSEC) AND AVERAGE REASSOCIATION DELAY (R IN MSEC) FOR THE RANDOM WAYPOINT AND RANDOM WALK WITH WRAPPING MOBILITY MODELS. RESULTS ARE PRESENTED FOR NETWORK SIZES OF 40, 60 AND 80 NODES.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N = 40</th>
<th>N = 60</th>
<th>N = 80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>D</td>
<td>R</td>
</tr>
<tr>
<td>Proposed</td>
<td>125.1</td>
<td>66.5</td>
<td>16.8</td>
</tr>
<tr>
<td>ADHOC</td>
<td>120.9</td>
<td>68.8</td>
<td>153.2</td>
</tr>
<tr>
<td>802.11p</td>
<td>114.7</td>
<td>72.5</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE IV
AVERAGE PER NODE THROUGHPUT, AVERAGE PACKET DELAY AND AVERAGE REASSOCIATION DELAY FOR VARIOUS NETWORK SIZES WITH BI-DIRECTIONAL TRAFFIC. THE MOBILITY MODEL IS RANDOM TRIP ON THE HOUSTON CITY SECTION MAP.

<table>
<thead>
<tr>
<th>Number of Nodes</th>
<th>Throughput (kbps)</th>
<th>Packet Delay (msec)</th>
<th>Reassociation Delay (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>ADHOC</td>
<td>802.11p</td>
</tr>
<tr>
<td>20</td>
<td>251.1</td>
<td>232.6</td>
<td>237.3</td>
</tr>
<tr>
<td>40</td>
<td>124.5</td>
<td>103.6</td>
<td>109.8</td>
</tr>
<tr>
<td>60</td>
<td>84.6</td>
<td>75.5</td>
<td>71.4</td>
</tr>
<tr>
<td>80</td>
<td>69.7</td>
<td>65.0</td>
<td>59.1</td>
</tr>
<tr>
<td>100</td>
<td>60.0</td>
<td>55.4</td>
<td>50.0</td>
</tr>
</tbody>
</table>

TABLE V
AVERAGE PER NODE THROUGHPUT, AVERAGE PACKET DELAY AND AVERAGE REASSOCIATION DELAY FOR VARIOUS PACKET SIZES (IN BYTES). THE NETWORK SIZE IS 100 NODES ARE THE MOBILITY MODEL IS RANDOM TRIP ON THE HOUSTON CITY SECTION MAP.

<table>
<thead>
<tr>
<th>Packet Size</th>
<th>Throughput (kbps)</th>
<th>Packet Delay (msec)</th>
<th>Reassociation Delay (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
<td>ADHOC</td>
<td>802.11p</td>
</tr>
<tr>
<td>256</td>
<td>44.2</td>
<td>53.3</td>
<td>45.2</td>
</tr>
<tr>
<td>512</td>
<td>55.3</td>
<td>59.2</td>
<td>50.4</td>
</tr>
<tr>
<td>750</td>
<td>59.9</td>
<td>61.8</td>
<td>52.1</td>
</tr>
<tr>
<td>1040</td>
<td>63.4</td>
<td>62.5</td>
<td>53.3</td>
</tr>
<tr>
<td>1400</td>
<td>65.7</td>
<td>63.1</td>
<td>54.7</td>
</tr>
</tbody>
</table>

VII. APPENDIX

Case 1: When \(x_j(0), y_j(0) \in L_j\), \(Z_{s_j}(T_j^1) = 1\) and \(Z_{s_j}(0) = 1\), node \(j\) is initially within a circular region of radius \(r\) centered at the AP’s location \(x_s(0), y_s(0)\). Then the probability \(P[Z_{s_j}(T_j^1) = 1]\) depends on the likelihood that node \(j\) is still within the circular region at \(t = T_j^1\). This probability can be evaluated by integrating the pdf of node \(j\)’s position at time \(T_j^1\) over the circular region. Figure 6 shows an example where the AP’s position is marked by A and node \(j\)’s initial position is marked by B. The probability that node \(j\) stays in the AP’s range can be obtained by first integrating the pdf of node \(j\)’s motion over the circle of radius \(r - l\) centered at B and then over arcs subtending an angle of \(2\pi - 2\theta\) at B as the radius sweeps over the range \(r - l \leq x \leq r + l\). Using elementary trigonometry: \(\theta = \pi - \beta = \pi - \cos^{-1}\left(\frac{-r^2 + l^2 + x^2}{2rl}\right)\). For 2-D Brownian motion with variance \(\alpha\tau\) in each dimension, the pdf of the distance \(x\) and angle \(\phi\) of node \(j\) at time \(\tau\) with respect to its origin at time 0 is \(p(x, \phi) = \frac{x}{2\pi\alpha}\frac{e^{-x^2/2\alpha\tau}}{\pi^{1/2}}\), with \(0 \leq \phi < 2\pi\), and \(0 \leq x < \infty\). Then

\[
P[Z_{s_j}(T_j^1) = 1 | \Delta_{s_j}(0) = l, T_j^1 = \tau] = \int_{r-l}^{r+l} \int_0^{2\pi/2} d\theta' \int_0^{\pi/2} x \frac{e^{-x^2/2\alpha\tau}}{\pi^{1/2}} dx d\theta' dx + \int_{r-l}^{r+l} \int_0^{2\pi/2} d\theta' \int_{\pi/2}^{\pi} x \frac{e^{-x^2/2\alpha\tau}}{\pi^{1/2}} dx d\theta' dx \tag{22}
\]

Case 2: When \(x_j(0), y_j(0) \in L_j\), \(Z_{s_j}(T_j^1) = 1\) and \(Z_{s_j}(0) = 0\), node \(j\) is initially outside the circular region of radius \(r\)
centered at the AP’s location but moves inside the circle at time $T^1_j$. In this case we integrate for arcs subtending an angle $2\theta$ as $x$ varies from $l - r$ to $l + r$. The angle $\theta$ is given by

$$\theta = \cos^{-1}\left(\frac{-l^2 + r^2 + x^2}{2lr}\right)$$

and thus the probability that node $j$, starting at distance $l$, $l > r$, from the AP at $t = 0$ becomes its neighbor at $t = T^1_j$ is

$$P(Z_{sj}(T^1_j) = 1 | \Delta_{sj}(0) = l, T^1_j = \tau] = \int_{l-r}^{l+r} \int_{0}^{2\pi} e^{-\frac{x^2}{2\alpha \tau}} d\theta dx$$  \hspace{1cm} (23)$$

REFERENCES


