Anticipatory Mobility Management by Big Data Analytics for Ultra-Low Latency Mobile Networking

Che-Yu Lin*, Kwang-Cheng Chen, *Fellow, IEEE*†, Dilranjan Wickramasuriya‡, Shao-Yu Lien§ and Richard D. Gitlin, *Life Fellow, IEEE*‡

**Graduate Institute of Communication Engineering, National Taiwan University. Email: r04942097@ntu.edu.tw*

†Department of Electrical Engineering, University of South Florida, Tampa, FL 33620. E-mail: kwangcheng@usf.edu

‡Department of Electrical Engineering, University of South Florida, Tampa.

§National Chung-Cheng University, Taiwan.

***Abstract*—Massive deployment of autonomous vehicles, unmanned aerial vehicles, and robots, brings in a new technology challenge to establish ultra-low end-to-end latency mobile networking to enable holistic computing mechanisms. With the aid of open-loop wireless communication and proactive network association in vehicle-centric heterogeneous network architecture, anticipatory mobility management relying on inference and learning from big vehicular data plays a key role to facilitate such a new technological paradigm. Anticipatory mobility management aims to predict APs to be connected in the next time instant and in a real-time manner, such that ultra-low latency downlink open-loop communication can be realized with proactive network association. In this paper, we successfully respond this technology challenge using big data analytics with location-based learning and inference techniques, to achieve satisfactory performance of predicting APs. Real traces of big vehicular movement data have been used to verify that the proposed prediction methods are effective for the purpose of anticipatory mobility management and thus ultra-low latency mobile networking.**

***Index Terms*—big data, data analytics, mobile networks, ultra-low latency, fog computing, vehicular networks, autonomous vehicles, machine learning, virtual networks, 5G**

1. INTRODUCTION

Fifth generation (5G) mobile communications will come into our lifes in next few years, with wide deployment of more advanced mobile broadband communication and diverse Internet of Things (IoT) applications [1]. Another major goal of 5G, low latency is also expected to be reduced from hundreds of milliseconds in 4G. In light of massive deployment of autonomous vehicles (AVs) in the next decade, heterogeneous cellular network structure to support the emerging need of critical control, command, and management are very much wanted in the near future, as an emerging application scenario of IoT [2]. However, ultra-low latency, in the order of 1 msec, is required for such critical messages in the massive operation of AVs [3], which is much beyond the low latency in 5G. Due to high mobility of AVs, such requirements suggest even more challenging technology than Tactile Internet [4]. Therefore, achieving ultra-low latency mobile networking requires a new technological paradigm.

Leveraging edge/fog computing to reduce end-to-end latency for the purpose of control and management suggests achieving low-latency by a heterogeneous network architecture [5], [6] with a two-tier structure. Recent LTE based vehicular radio access targets at integrating 3GPP vehicle-to-everything (V2X) connection, licensed-assisted access (LAA) and device-to-device (D2D) proximity services to address the particular issues of sidelink shared channel collision and resource uncertainty on uplink, downlink and sidelink. Subsequent IEEE 802.11ax based vehicular direct radio access emphasizes providing a quality-of-service (QoS) guaranteed sidelink access scheme, to enable direct vehicle data exchange, which paves the way to deploy V2X in unlicensed spectrum [7]. In the meantime, open-loop communication [8], [9] as a technology for physical transmission can also significantly reduce latency as long as appropriate error control is in place [10]. The final piece of technology to achieve ultra-low latency networking is the proactive network association and corresponding anticipatory mobility management (AMM) at the edge of network infrastructure, such that mobile stations like a vehicle can communicate with the fog computing network elements at the minimal latency without the need of complicated handover procedure. Following the research efforts by [11] and [12], predictive capability emerges as a very much needed technology in the architecture of Section 2 of this paper.

Therefore, AMM realized by prediction of APs in use for each vehicle from big vehicular datasets emerges as a technologically open problem. In this paper, prediction of access points to ultra-low latency mobile networks based on big vehicular data analytics will be investigated, under the criterion and constraint of near real-time computation to differ from the scenarios of most machine learning and data analytics. Our proposed prediction completes the network function design of ultra-low latency mobile networks and paves the avenue to autonomous vehicles, unmanned aerial vehicles, and robotics. This might be the first in-depth exploration in wireless networks that ”require” big data analysis and machine learning [13] to fulfill the technological innovation in wireless and vehicular networks.

2. ARCHITECTURE OF COMPUTING AND NETWORKING TO ACHIEVE ULTRA-LOW LATENCY

Traditionally, centralized managed wireless network allocated physical channels (radio resource units) to each mobile node (a vehicle in our case), and a handover mechanism is required when the mobile node connects from one AP (or base station) to another. Each AP or base station (BS) needs to serve multiple mobile stations (MSs), which relies on fast adapting and complicated closed loop control signaling between BS/AP and each MS, such as power control in 3G cellular and channel estimation in 4G cellular systems. Although various techniques can improve the latency such as tunneling protocols, a new technology paradigm becomes a must to dramatically reduce end-to-end networking latency in the tens of milliseconds for connected vehicles and further down to milliseconds for autonomous vehicles [3].

A break through idea is to treat each vehicle as a virtual cell, that is, there is only one mobile station in each virtual cell and multiple APs cooperatively serve this mobile node. Each AP designates a network/radio slice to this virtual cell, and serves multiple virtual cells at the same time. Consequently, each AP and the subsequent Anchor Node (AN) must run network virtualization in software defined networking (SDN) to facilitate the idea. To realize ultra-low latency in each radio transmission, proactive network association and open loop communication have to be adopted. The scenario is depicted in Figure 1. Due to proactive network association and open-loop communication, real-time and centralized mobility management to precisely determine APs to serve virtual cell of high mobility in next moment is not available. In other words, how can anchor nodes and fog/edge networks determine, in the next time instant, which APs will serve a virtual cell in the downlink?

Therefore, the success of ultra-low latency networking must rely on the technology that the AN is capable of facilitating the AMM to predict likely and appropriate APs for each mobile node to proactively connect in next time instant, particularly for ultra-low latency downlink communication. This is a fundamentally new technology challenge and machine learning on big vehicular data appears an attractive approach. Though applying machine learning to enhance the performance of wireless networks has attracted recent research interest [13], it might be the very first effort to develop wireless networking relying on machine learning and big data analysis. In this paper, we focus on the AMM technology by proper prediction of APs connected by proactive network association, and focus on big data analytics with considering AN governing APs application scenario only. In an unlikely but possible scenario that no AP is predicted to cover, the networking falls back to high power and longer latency cellular base station(s).

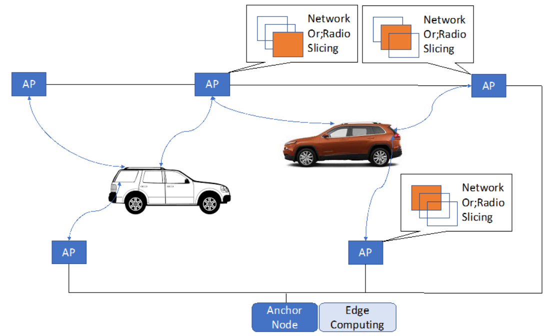


Figure 1. Radio slicing and network slicing to serve virtual cells (*i.e.* vehicles) to achieve ultra-low latency networking. The orange vehicle is the center of the virtual cell and communicating with three APs. To minimize networking latency, proactive network association has been adopted, which allows the virtual cell to select proper APs (*i.e.* network slice) to access and to proceed uplink transmission, implemented by cooperative communication like coordinated multiple-point (CoMP) transmission and reception, and are realized as open-loop communication without feedback acknowledgement, via selected radio slice. In the downlink, cooperative communication similar to CoMP proceeds, while each AP allocates an appropriate radio slice to the virtual cell. An anchor node is under the instruction of edge/fog computing and sends packets to those APs associated with the virtual cell, again by open-loop communication.

3. PREDICTION OF ACCESS POINTS VIA DATA ANALYTICS IN FOG

AMM in the uplink communication is straightforward since a vehicle just connects to APs in range via proactive network association. However, for ultra-low latency packets in the downlink, AMM must predict APs to be connected by the vehicle in next time instant. Both uplink and downlink require multi-path error control as [14]. The AMM consists of prediction by machine learning [13] by the anchor node in the fog, and falling back mechanism to high power node(s) in the heterogeneous network if no AP is successfully predicted. The extremely simple AMM operation except challenging prediction warrantees ultra-low latency mobile networking. [12] demonstrates the successful accomplishment of ultra-low latency by assuming AMM. In the following, we explore the design of AMM, with focus on the prediction from big data analysis.

3.1. Prediction Using Fog Computing/Networking

To facilitate ultra-low latency mobile networking empowered by proactive network association and open-loop communication, we adopt fog computing at the edge of network and this fog networking to form a heterogeneous network structure with cellular network. An AN in the fog to govern a number of APs for the purpose of the short-range communication accompanies the fog computing facility that analyzes the data which may include the map, recently associated APs, big historical data of AP association patterns, or side information regarding global moving patterns such as knowledge of source-destination, localization, *etc.*, to predict APs for a virtual cell to associate at next time instant.

3.2. AP Deployment and Association

The ideal deployment of APs shall be along the roadside as a sort of roadside units (RSUs) in traditional vehicular networks. However, for the purpose of applying heterogeneous cellular networks for ultra-low latency mobile networking, we practically and economically assume that the APs have multi-purpose roles in networking, say wireless broadband, smart city, and other IoT applications. Therefore, APs may be deployed in an area other than roadside. The coverage of an AP is similar to micro base stations or WiFi such that (quasi) real-time computing is possible. We further assume the APs are randomly deployed, which might not perfectly match the reality but provides the worst-case scenario for wireless networking. More precisely, the APs are deployed as a random geometric graph (RGG) with density λ_{AP} per km^2 in this paper, while realistic deployment can provide a better networking scenario. Please note that RGG deployment implies the non-trivial possibility of no coverage for a virtual cell under reasonable density of APs, which suggests worse performance than engineering practice in the Section 5 numerical evaluation (λ_{AP}=69.44/km^2).

Each virtual cell can proactively associate to K APs. Of course, for K the larger the better, but is limited by hardware and software capability to the value of K_{max}. Here, we pay attention to the case $K_{max} = 3$ in this paper. This means, for a virtual cell (*i.e.* a vehicle) that

- It associates with K_{max} APs of the strongest SINR (or other signaling to indicate suitability) if more than K_{max} APs are in radio range.
- It associates with K APs, if $K \leq K_{max}$ APs are in radio range.

For the initial research in this problem, a linearly decayed SINR that covers a circular area of diameter, say 200m, is considered, ignoring issues such as power fading and interference temporally. Accordingly, for multiple APs covering a specific location, the connecting priority would be inversely proportional to the distance in between directly.

3.3. Representation of Knowledge

The prediction of associated APs is actually a problem of inference on heterogeneous data [15], [16]. A very different aspect is the representation of APs as the target of inference. Suppose each AP has an ID. Given the rule of association, we could obtain time series representation by defining AP association vector as

$$\begin{aligned} X(t) &= [X^{(1)}(t)X^{(2)}(t)...X^{(d)}(t)] \in \{0,1\}^d \\ X^{(i)}(t) &= \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ AP is connected} \\ 0, & \text{if the } i^{\text{th}} \text{ AP is not connected} \end{cases} \end{aligned} \tag{1}$$

where d is the number of APs considered in learning and inference. For an example if $d = 16$ and time is indexed by positive integers, an association vector looks like

$$\begin{array}{cc} t_n & X(t_n) \\ n = 1, 2, ... & [0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0] \end{array} \tag{2}$$

3.4. Problem Formation of Prediction on APs

After defining AP association and the representation of AP association vector, we are able to define the prediction of associated APs for a specific vehicle.

Problem. *Considering one single vehicle driving on roads, given a series of time $t_1, ..., t_n$ and the corresponding AP association vector $X(t_1), ..., X(t_n)$ of the vehicle, predict $X(t_{n+1})$ for $t_{n+1} > t_n$.*

While the ultimate goal is to provide a perfect scheduling for the overall association between APs and vehicles, a simpler problem with one vehicle only is considered here such that there would be no queueing or serving capacity issue in connection to APs. Further more, as initial exploration of such prediction by machine learning techniques to support ultra-low latency networking, we restrict ourselves to the smallest information set containing the historical association vectors only, for there is no guaranteed access to other informative messages, due to the privacy concern.

4. PREDICTIVE METHODOLOGY

Although it appears that many machine learning techniques [13], [17], [18] are applicable to this scenario, there exists some serious issues that limits these well known techniques. Firstly, prediction of suitable APs from one time instant to the next must be executed in almost real-time. For example, training recurrent neural networks appears fitting the prediction problem but our experiment shows unsatisfactory performance due to this real-time concern [19]. Consequently, in this paper, we only consider machine learning techniques that can be executed near real-time by proper computing facilities. Regression or adaptive filtering may fit this constraint [20], while popular deep learning [21] might better fit cloud computing to assist. Secondly, we are not simply predicting the association vector, but "when" will it be "what state." Since the connecting status is based on the location of the vehicle, different time duration implies different mobile communication range, even with similar trajectory and speeding, and predicting the association vector only with dropping the timing information is thus unsuitable. Thirdly, the capability of learning from a batch of data means the existence of a fixed pattern, in other words, the process should be stationary. As what gives the association vector is the location while instantaneous GPS information of which can not be obtained by fog computing, its underlying mechanism is affected by multiple exogenous factors such as the traffic lights, the power of the vehicle, the time being rush or off-peak, *etc*. Without consideration of these factors, the distribution of the location is apparently non-stationary, not to mention the depended association vector. Consequently, AP prediction is not as straightforward as it appears, and deliberate considerations about predicting both "when" and "what state" under non-stationary environment in a real-time manner is thus needed.

As the status of the association vector is uniquely determined by the location of the vehicle according to (1),

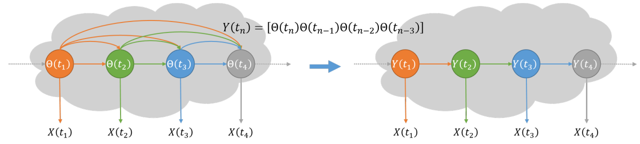


Figure 2. Illustration of the hidden Markov process that generates the association vectors. The Θ's denote the unobserved locations. The association vectors observed only depend on the corresponding hidden Θ's, according to the connecting rules, and a Markov property of unknown order exists in the series of Θ. For the order being 3, for example, $Y(t_n) = [\Theta(t_n)\Theta(t_{n-1})\Theta(t_{n-2})\Theta(t_{n-3})]$ together with X's form a simple hidden Markov chain.

the series of locations together with the corresponding association vectors actually form a hidden Markov process [18] but of unknown order of time dependence, as shown in Figure 2, due to the conditional independence property of the association vector given the location. From this viewpoint, we take inferring and predicting the latent location Θ as the problem, and learning the non-stationary transition probability between locations is thus needed.

In the problem of tracking, Bayes filtering [28], [29] gives the optimal solution theoretically, provided that the hidden process is 1st order Markovian. Though the order of temporal dependence in between is generally high, a modified representation could suggest the 1st order Markov process as reasonable. As a result, we adopted the concept of Bayes filter for solving the problem, which is also possible to support real-time inference.

4.1. Recursive Bayesian Estimation

Probabilistically, what we want to learn is

$$P_{X_{n+1}|X_{1:n}}(x_{n+1}|x_{1:n}) \tag{3}$$

where X_i, $X_{i:j}$ are shorthand notations for $X(t_i)$ and $[X(t_i)X(t_{i+1})...X(t_j)]$ respectively. Taking the location Θ into consideration, (3) can be rewritten as

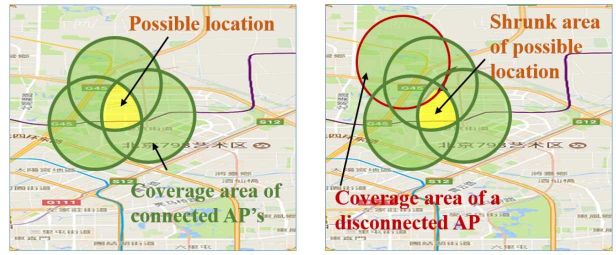
$$\int P_{X_{n+1},\Theta_{n+1}|X_{1:n}}(x_{n+1},\theta_{n+1}|x_{1:n})d\theta_{n+1} \tag{4}$$

and based on the conditional independent property of X given θ and the conditional probabilistic factorization, the integrand in (4) can be further written as

$$\begin{aligned}&P_{X_{n+1},\Theta_{n+1}|X_{1:n}}(x_{n+1},\theta_{n+1}|x_{1:n})\\ =&P_{X_{n+1}|\Theta_{n+1}}(x_{n+1}|\theta_{n+1})P_{\Theta_{n+1}|X_{1:n}}(\theta_{n+1}|x_{1:n})\end{aligned} \tag{5}$$

While the former term $P_{X_{n+1}|\Theta_{n+1}}(x_{n+1}|\theta_{n+1})$ on the right hand side of (5) follows the associating rules in Section 3.2 which is known in advance, the later term $P_{\Theta_{n+1}|X_{1:n}}(\theta_{n+1}|x_{1:n})$ turns out to be the knowledge to learn, making the prediction of association vector into the prediction of the unobserved location.

To perform estimation and prediction, based on the Markovian assumption, there is a recurrence relation between the posterior belief, $P_{\Theta_n|X_{1:n}}(\theta_n|x_{1:n})$, and the prior belief, $P_{\Theta_n|X_{1:n-1}}(\theta_n|x_{1:n-1})$, satisfying



(a) Possible positioning basing on information from connected APs

(b) Possible positioning with additional information from disconnected APs

Figure 3. Illustration of inferring location from association vector. In case 3(a), with considering connected APs, the enclosed area that the vehicle possibly locates is the intersection of the coverage area of the connected APs (yellow coloured). In case 3(b), in addition to the connected APs, some of the disconnected APs also provide extra information about the possible location.

$$\begin{aligned}&P_{\Theta_n|X_{1:n-1}}(\theta_n|x_{1:n-1})\\ =&\int P_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1})P_{\Theta_{n-1}|X_{1:n-1}}(\theta_{n-1}|x_{1:n-1})d\theta_{n-1}\end{aligned} \tag{6}$$

and

$$P_{\Theta_n|X_{1:n}}(\theta_n|x_{1:n}) = \frac{P_{X_n|\Theta_n}(x_n|\theta_n)P_{\Theta_n|X_{1:n-1}}(\theta_n|x_{1:n-1})}{\int P_{X_n|\Theta_n}(x_n|\theta_n)P_{\Theta_n|X_{1:n-1}}(\theta_n|x_{1:n-1})d\theta_n} \tag{7}$$

which forms the basis of the optimal Bayesian solution. For a more general situation that the Θ's exhibit higher order Markovian property as shown in Figure 2, we can still apply the recurrence (6) and (7), with only replacing Θ's by

$$Y_n = [\Theta_n\Theta_{n-1}...\Theta_{n-l}] \tag{8}$$

where l is the order of the Markovian property. It is easily seen that Y's forms a Markov chain, and the conditional independence property of X given Y still holds.

Based on this framework, we are now going to describe the detail implementation about how to obtain the posterior belief, (7), and how to make prediction, (6), under non-stationary vehicular data.

4.2. Posterior Belief of the Location

Though it is supposed to update the posterior belief with (7), based on Section 3.2, the information about the location at one time instant is uniquely contained in the corresponding association vector and is derived as follows. For APs with state 1, the possible location would only lie inside the intersection of their coverage area, as shown in Figure 3a. For APs with state 0, the possible location would lie outside their coverage area, or inside their coverage area but with farther distance from the center, comparing to APs with state 1. Accordingly, the possible location of a particular vehicle would narrow down if there exists overlapping between the intersection and the coverage area of disconnected APs, as shown in Figure 3b.

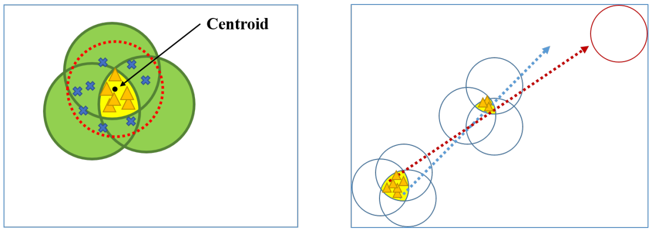


Figure 4. Illustration of utilizing Monte Carlo method for obtaining a set of points for representing the area of possible location and making prediction accordingly. For randomly spread points, we retain the ones in the yellow coloured area only for representing the possible location. Based on these points, we may estimate the moving velocity accordingly, and make use of it to predict the future location with (10). The red and blue dashed lines represent two different possible directions, leading to different locations.

Figure 5. Demonstration of the predicted results. The figure shows a example of AP prediction with the ordering of time corresponding to the frames arranged from left to right. The ellipses in the figure indicate the coverage area of APs. The blue colored ones means that we predict the AP to be connected and it is indeed connected, the green ones means that we predict the AP to be disconnected but it is actually connected, and the red ones means that we predict the AP to be connected but it is actually disconnected. The orange marker in the figure represents the predicted location of the car.

In this way, we can graphically enclose an area indicating the possible location for each of the observed association vectors. However, these areas are generally irregularly shaped, representative descriptions are generally not available. Consequently, we resort to the Monte Carlo method for finding sets of points as the representations of the areas, as shown in Figure 4. One simple way of achieving this is to randomly spread points around the centroid of the connected APs within region of size the same as a coverage area and retain points satisfying the restriction derived from the association vector as in 3b.

4.3. Future Location Prediction

With the posterior belief of the locations, we are now able to predict the future location with (6) and (8). In common usage of the Bayes filtering, the transition probability ($P_{\Theta_n|\Theta_{n-1}}(\theta_n|\theta_{n-1})$ or $P_{Y_n|Y_{n-1}}(y_n|y_{n-1})$) is prerequisite and can be estimated from a batch of data in advance. However, the mentioned non-stationary issue makes it unlikely. In dealing with the non-stationarity, different procedures should be adopted for the different causes [30], and for prediction, the varying velocity of the vehicle and the variable time duration makes θ non-stationary. To illustrate, following the law of motion,

$$\theta(t_{n+1}) = \theta(t_n) + \int_{t_n}^{t_{n+1}} \theta'(s)ds \tag{9}$$

where $\theta'(s)$ is the first order derivative of θ, representing the velocity. If we further approximate the movement within a short period of time with a constant velocity, then

$$\theta(t_{n+1}) \approx \theta(t_n) + v \cdot (t_{n+1} - t_n) \tag{10}$$

where v denotes the approximated velocity. While the duration is in our control (to predict different "when") with no fixed length, the v which is affected by multiple exogenous factors as mentioned previously is exact the reason for θ being non-stationary, and estimation of v with considering the non-stationary issue is thus what needed.

Optimal velocity. Assuming a fixed relation between the velocity and the locations, *i.e.*, $v_i = v(\bar{\theta}_i, \bar{\theta}_{i-1}, ...)$, a function of the past locations, then the optimal $v(.)$ can be found with

$$v^* = \arg\min_{v(.)} \sum_{i=1}^{n} \lambda^{n-i} l(\bar{\theta}_i, \hat{\theta}_i) \tag{11}$$

where $l(.,.)$ denotes the loss function and

$$\hat{\theta}_i = \bar{\theta}_{i-1} + v(\bar{\theta}_{i-1}, ...) \cdot (t_i - t_{i-1}) \tag{12}$$

denotes the predicted θ_i. What behind (11) is that we wish to find an optimal $v(.)$ that minimizes the induced losses of all observed time instances. Here, $\lambda \in (0, 1)$ represents the forgetting factor for dealing with the non-stationary issue.

Remark. *As we have shown approximating the movement with constant velocity, higher order derivative can actually be introduced for more accurate approximation.*

4.4. Predicting Future Location and Association Vector

With the estimated velocity and $\tilde{\theta}$'s, the simulated sets of possible location, we are able to perform (6) now. Instead of yielding an analytic distribution function, the prior belief is the same presented with a set of points as what we've done for inferring the possible location. For each of the $\theta_n^{(i)} \in \tilde{\theta}_n$, a simulated set of predicting location for Θ_{n+1} could be obtained from

$$\{\theta : \theta = \theta_n^{(i)} + v_n(t_{n+1} - t_n),\ \forall \theta_n^{(i)} \in \tilde{\theta}_n\} \tag{13}$$

where v_n is the estimated optimal velocity. With (13), we can transform each of the θ in the predicted location set into the association vector easily according to Section 3.2, yielding a set of possible association vectors. The final decision is then the prediction with the highest votes.

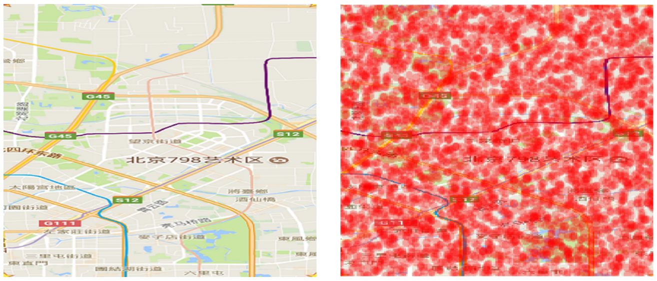


Figure 6. Real world region of consideration and randomized deployment of APs.

5. NUMERICAL VERIFICATIONS

5.1. Vehicular Dataset

To examine whether learning mechanisms can fit our purpose, testing with real dataset is needed. Though there is no perfect dataset available, the big data from large-scale taxi service or Uber appears to meet our initial need. We are lucky to obtain recent big data of Beijing taxi data to develop the prediction mechanism, which is the same dataset used in [12] and [24]. Such taxi data consists of 12,000 taxis operating over two months with 24-hour GPS trace for each taxi. After data cleaning and aligning with map, we can therefore construct a trajectory via GPS coordinate records from each taxi, to represent typical vehicular movement data in urban area. Each record contains the ID of the car, the latitude, the longitude and the recording time with resolution of every several seconds, which is of the form

ID	time	latitude	longitude
204806	20121101123159	116.5541328	40.02253789

In the numerical evaluations of this paper, we focus on the region of interest ranging from 116.435° to 116.505° (latitude) and 39.928° to 40.061° (longitude), roughly 7.8×6.6 km^2. Within this region, there are 3,575 APs in total with $\lambda_{AP} = 69.44$, and the GPS records are transformed into association vectors described in Section 3.2 as the representation form of (2).

5.2. Prediction Demonstration

In Figure 5, a series of frames demonstrating the consecutive prediction results is presented. Under the Bayes filter framework, online prediction is feasible. Although some of the time instances yield perfect prediction but some do not, the deviation from the centroid of the true connected APs to the predicted ones is actually not significant at all.

5.3. Performance Comparison

5.3.1. Benchmark Performance. Although it is less suitable applying classification techniques to AP prediction, we still try with the naïve Bayesian approach, suggesting the benchmark mechanism and performance. As the details of the naïve Bayes classification could be found in [23], here the targets to predict are the APs, $X^{(k)}(t_{i+1})$ for $k = 1, ..., d$, and the corresponding predictive features are the three most recent association vectors, $X(t_i),\ X(t_{i-1}),\ X(t_{i-2})$. With this approach, a prediction is independently made for each AP and any correlations among groups of APs that frequently provide connectivity to a vehicle together due to close proximities are ignored. This deteriorates the performance, and thus, a slight improvement based upon the raw prediction was made by adding a second decision layer. On occasions where only two APs were predicted, a search was made in the training data to determine the most frequently occurring set of three APs of which the predicted APs were a subset. These APs were then chosen as the output instead of the raw predictions. The accuracy is expressed as a percentage of the number of times correct AP predictions are made and is summarized in Table 1. As we can see, the naïve Bayesian approach can not supply correct probability when a vehicle can only connect to a single AP. Though the situation of association with a single AP is not happening frequently, it is still desirable to make accurate enough prediction in this situation.

5.3.2. Proposed Method Performance. Based on the framework of recursive Bayesian estimation, we proposed a method to approximate the velocity that controls the hidden location by optimizing the loss. The performance of the method is shown in Table 2. With the loss function in (11) being set to $(\bar{\theta}_i - \hat{\theta}_i)^2$, the well known recursive least square (RLS) [17] algorithm can be applied for solving (11) efficiently. Comparing to the naïve Bayes classification, the proposed method is superior in not only the full connected situation, but also in the other two as well, which means the method is applicable even when λ_{AP} is of a low level. Such performance well satisfies the need of AMM for ultra-low latency networking in Section 2. Our results demonstrate (i) limited possibility of falling back to cellular (*i.e.*, no AP to be correctly connected) (ii) since we use random deployment of APs, there exists a non-trivial possibility in our numerical that no AP is actually in communication range of a vehicle, which can be easily corrected in practical deployment with road map as a reference.

6. CONCLUDING REMARKS

In this paper, via big vehicular data, we examine the naïve Bayes approach as the benchmark, then a Bayes-filter-based latent location tracking of a vehicle for predicting the connected APs to achieve the ultra-low latency mobile networking. Though this simple scenario is considered, it can be extended to a more complicated SINR setting following the same procedure of recursively performing posterior belief estimation (7) and making prediction according to (6). This research opens a new avenue for machine learning and big data analytics for mobile networks. Further real-time predictive techniques to effectively handle a substantial member of vehicles and even scheduling AVs [24] or techniques assisted by deep learning in the cloud definitely

Improved Naïve Bayes classification				
# of connected APs	# of correctly predicted APs			
	0	1	2	3
1	86.36%	13.63%		
2	3.91%	27.90%	68.19%	
3	1.25%	4.11%	11.80%	82.84%

TABLE 1. IMPROVED NAÏVE BAYES CLASSIFICATION

Method of Calculating Optimal Velocity				
# of connected APs	# of correctly predicted APs			
	0	1	2	3
1	26.68%	73.32%		
2	8.49%	13.68%	77.83%	
3	1.25%	2.12%	2.74%	93.89%

TABLE 2. OPTIMAL VELOCITY ESTIMATION

merits further investigation. There are further efforts to complete the design of AMM in heterogeneous networks, such as clustering APs under anchor nodes, precise falling back to cellular mechanism compatible with 3GPP standards, more mature error control in multi-path operation, and precise network virtualization for ultra-low latency packets and broadband multimedia traffic.

References

[1] M. R. Palattella, M. Dohler, A. Greico, G. Rizzo, J. Torsen, T. Engel and L. Ladid, "Internet of things in 5G era: enablers, architecture, and business models," *IEEE Journ. on Selected Areas in Commun.*, vol. 34, no. 3, pp. 520–527, Mar. 2016.

[2] K. Zhang, Q. Zheng, H. Yang, L. Zhao, L. Hou and P. Chatzimisios, "Reliable and efficient autonomous driving: the need for heterogeneous vehicular networks," *IEEE Commun. Mag.*, pp. 72–79, Dec. 2015.

[3] M. Chiang and T. Zhang, "Fog and IoT: an overview of research opportunities, *IEEE Internet of Things Journ.*, vol. 3, no. 6, pp. 854–864, Dec. 2016.

[4] G. P. Fettweis, "The tactile Internet: applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, March 2014.

[5] S.-Y. Lien, S.-C. Hung, K.-C. Chen and Y.-C. Liang, "Ultra-low latency ubiquitous connections in heterogeneous cloud radio access networks," Special Issue on Heterogeneous Cloud Radio Access Networks, *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 22–31, June 2015.

[6] S.-C. Hung, H. Hsu, S.-Y. Lien and K.-C. Chen, "Architecture harmonization between cloud radio access networks and fog networks," *IEEE Access*, vol. 3, pp. 3019–3034, Dec. 2015.

[7] S.-Y. Lien, D.-J. Deng, H.-L. Tsai, Y. Lin and K.-C. Chen, "Vehicular radio access to unlicensed spectrum," to appear in the *IEEE Wireless Commun.* (early access).

[8] S.-C. Hung, S.-Y. Lien and K.-C. Chen, "Low latency communication for Internet of things," *Proc. IEEE Int. Conf. on Commun. in China (ICCC)*, Shenzhen, 2015.

[9] S.-Y. Lien, S.-C. Hung and K.-C. Chen, "Optimal radio access for fully packet switching 5G networks. IEEE International Conference on Communications, *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, pp. 3921–3926, 2015.

[10] E. Ayanoglu, C.-L. I., R. D. Gitlin and J. E. Mazo, "Diversity coding for self-healing and fault-tolerant communication networks," *IEEE Trans. on Commun.*, vol. 42, no. 11, pp. 1677–1686, Nov. 1993.

[11] A. Magnano, X. Fei, A. Boukerche and A. A. F. Loureiro, "A novel predictive handover protocol for mobile IP in vehicular networks," IEEE Trans. On Veh. Technology, vol. 65, no. 10, Oct. 2016.

[12] S.-C. Hung, H. Hsu, S.-M. Cheng, Q. Cui and K.-C. Chen, "Delay garanteed network association for mobile machines in heterogeneous cloud radio access networks, *IEEE Trans. On Mobile Computing* (in revision).

[13] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[14] I.-W. Lai, C.-H. Lee, K.-C. Chen and E. Biglieri, "End-to-end virtual MIMO transmission in Ad Hoc cognitive radio networks," *IEEE Trans. on Wireless Commun.*, vol. 13, no.1, pp. 330–341, Jan. 2014.

[15] K.-C. Chen, S.-L. Huang, L. Zheng and H.V. Poor, "Communication theoretic data analytics," *IEEE Journ. on Selected Areas in Commun.*, vol. 33, no. 4, pp. 663–675, Apr. 2015.

[16] K.-C. Chen, B. Mankir, S.-L. Huang, L. Zheng and H. V. Poor, "Communication theoretic inference on heterogeneous data," *IEEE Int. Conf. on Commun.*, 2016.

[17] S. Haykin, *Adaptive filter theory*, Prentice-Hall, 2002.

[18] C. M. Bishop, *Pattern recognition and machine learning*, Springer, Aug. 2006.

[19] R. Pascanu, T. Mikolov and Y. Bengio, "On the difficulty of training recurrent neural networks," *30th Int. Conf. on Machine Learning*, Atlanta, Georgia, USA, 2013.

[20] A. J. Smola and B. Schlkopf, "A tutorial on support vector regression," *Stat. and Computing*, vol. 14, pp. 199–222, 2004.

[21] W. Huang, G. Song, H. Hong and K. Xie, "Deep architecture for traffic flow prediction: deep belief network with multitask learning," *IEEE Trans. on Intell. Transportation Syst.*, vol. 15, no. 5, pp. 2191–2201, 2014.

[22] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning*, Springer, New York, 2001.

[23] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, 103–130, 1997.

[24] I-C. Lin, C.-Y. Lin, H.-M. Hung, Q. Cui and K.-C. Chen, "Autonomous vehicle as an intelligent transportation service in a smart city," IEEE 86th VTC, Toronto, Sept. 2017.

[25] C. Hsu, C. Chang and C. Lin, "A practical guide to support vector classification," Taiwan: Dept. of Comput. Sci. and Inform. Eng., Nat. Taiwan Univ., Apr. 2010.

[26] S. Haykin, *Neural metworks: A comprehensive foundation*, NJ, Englewood Cliffs: Prentice-Hall, 1999.

[27] G. H. Bakr, T. Hofmann , B. Schölkopf , Alexander J. Smola , Ben Taskar , S. V. N. Vishwanathan, *Predicting structured data*, MIT Press, Cambridge, 2007.

[28] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no.2, pp. 174–188, Feb. 2002.

[29] G. Welch and G. Bishop, "An introduction to the Kalman filter", 2000.

[30] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy and A. Bouchachia, "A survey on concept drift adaption," *ACM Computing Surveys*, vol. 46, no. 4, article 44, Mar. 2014.