Evaluating Mobile Signal and Location Predictability along Public Transportation Routes

Hatem Abou-zeid*, Hossam S. Hassanein†, Zohaib Tanveer*, and Najah AbuAli‡
*Electrical and Computer Eng. Dept., Queen's University, Canada, {h.abouzeid, 12zt4}@queensu.ca
†School of Computing, Queen's University, Canada, hossam@cs.queensu.ca
‡College of Information Technology, UAE University, UAE, najah@uaeu.ac.ae

***Abstract*—Emerging mobility-aware content delivery approaches are being proposed to cope with the increasing usage of data from vehicular users. The main idea is to forecast the user locations and associated link capacity, and then proactively counter service fluctuations in advance. For instance, a user that is heading towards low coverage can be prioritized and have video content prebuffered. While the reported gains are encouraging, the results are primarily based on assumptions of perfect prediction. Investigating the predictability of mobility and future signal variations is therefore imperative to evaluate the practical viability of such predictive content delivery paradigms. To this end, this paper presents a large-scale measurement study of 33 repeated trips along a 23.4 km bus route covering urban and sub-urban areas in Kingston, Canada. We provide a thorough analysis of the collected traces to investigate the effects of geographical area, time, forecasting window, and contextual factors such as signal lights and bus stops. The collected dataset can also be used in several other ways to further investigate and drive research in predictive vehicular content delivery.**

I. Introduction

The dramatic increases in mobile network traffic are a constant burden on cellular operators. Striking a balance between coverage, consistent data rates, and infrastructure costs is a significant challenge. Meanwhile, the widespread adoption of Smartphones is increasing the traffic from vehicular users, particularly buses and trains. An alternative to expansion is therefore needed to provide mobile services at sustainable costs. As such, novel content delivery approaches are receiving increasing interest. In particular, *predictive* resource allocation techniques that exploit user mobility have been recently proposed to improve throughput and fairness [1], [2], video streaming delivery [3]–[5], and transmission energy [6]. This is accomplished by leveraging the knowledge of the future link capacity users are expected to experience, and then performing *long-term* Resource Allocation (RA) plans over several seconds. By doing so, Base Stations (BSs) can prioritize users headed to poor channel conditions and for example, proactively prebuffer additional future video content.

The underlying assumption of such anticipatory schemes is the predictability of the users' future channel states as they traverse a road network. While the gains reported in [1]–[6] are encouraging, the results are primarily based on assumptions of perfect prediction. However, there are two primary sources of uncertainty that can affect the reported gains significantly: 1) location prediction errors and 2) signal strength prediction errors. Investigating the predictability of mobility and future signal variations is therefore imperative to evaluating the practical viability of the emerging predictive content delivery paradigms. To this end, this paper presents a large-scale measurement study of bus trips covering urban and sub-urban neighborhoods in Kingston, Canada during July 2014. In this paper, we focus on public transportation vehicles, as they are attractive candidates for predictive content delivery. This is because: 1) their routes and stops are known, and 2) they generate large amounts of mobile traffic which can benefit significantly from long-term optimization. We have measured the signal strength variations and geographical coordinates along a popular bus route. The trips are also made at three different times of the day to investigate the temporal variations in both location and signal strength predictability.

This work was made possible by NPRP grant no NPRP4-553-2-210 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

We summarize the key contributions of this paper in the following:

- To the best of our knowledge, this is the first large-scale mobile signal and location study along a *public transportation* route. The dataset includes approximately $475,200$ logs collected over 33 hours covering a total of 759 km. This dataset can be used to 1) analyze predictability and propose predictive models that capture the measured dynamics, and 2) practically evaluate the recent predictive delivery schemes [2]–[6] with real data.
- We provide an analysis of the collected measurements and investigate the effects of 1) the forecasting window duration, 2) the geographical context (urban vs. sub-rural), and 3) time of day, on the predictability of the location and signal strength. We also show that modeling prediction *uncertainty* is paramount due to the high variability observed in the measurements.
- We investigate the joint effects of location and signal strength errors on the signal strength predictability. Our findings indicate that errors in the predicted locations can undergo sudden increases due to the uncertainties around stopping at bus stops and traffic lights. These imperfections significantly impact signal strength predictability.

A. Related Work

There have been a number of recent works investigating the signal strength and bandwidth predictability of mobile networks along roads. A measurement campaign was recently made in [2] for different car trips. Yao *et al.* [7] also analyze bandwidth traces collected from two independent cellular providers for routes running through different radio conditions including terrestrial and underwater tunnels. Their findings confirm the correlation between user rates and location. Han *et al.* [8] also conduct an interesting measurement study, and addresses other contextual factors such as user speed, time of day, and humidity to predict the available bandwidth more accurately. Riiser *et al.* [9] also conduct a small measurement campaign of throughput along a metro, tram, bus, and ferry to illustrate how bandwidth varies. However, the traces are not intended to assess predictability, and signal strength values were not recorded.

While these works reveal the correlation between location and network capacity, they do not address joint location and signal strength predictability along *public transportation* routes. Further, the scope of the collected data does not facilitate developing models that can capture the time dependencies and geographical dynamics in public transportation routes.

B. Paper Organization

In the following section, we present an overview of the collected data set. Section III discusses the location predictability of the bus trips. Therein we investigate the effects of time and the forecasting window on the prediction. In Section IV, we present the signal strength measurements, and investigate the effects of geographical context, time, and location awareness on the predictability of signal strength. Finally, in Section V we summarize our findings and future directions.

II. The Dataset

The measurements were conducted along a popular bus route in Kingston, Canada shown in Fig. 1. The logs include a timestamp, longitude and latitude coordinates, and average signal strength in dBm, recorded every second. Each trip has the same start point, end point and direction. As this is an express route, there are only six stops along the route and a major transfer point at the Cataraqui Centre (which includes the major mall in Kingston) shown in Fig. 1. The route from the start point to the Cataraqui Centre is primarily urban, while after that it is primarily sub-urban and low density urban. The bus typically stops for a few minutes at the Cataraqui Centre, so we have measurements at a stationary point as well.

The trips were made at three different times of the day: 12 pm, 6 pm, and 7 pm. This was to account for both road traffic differences and varying interference and mobile network connectivity levels. In total we have surveyed 33 hours covering a total of 759 km arriving at $475,200$ logged data points.

Fig. 2 shows the latitude and signal strength variation with time for a sample log. The data was filtered to account for any anomalies in the recorded measurements (particularly that of the GPS coordinates). We can see that signal strength variations are quite rapid between the starting point and the Cataraqui Centre, where there are fluctuations even though the bus is stationary. After that, the signal strength remains at a relative low in the sub-urban area. This is followed with clear gradual increases and decreases in the signal strength due to line of sight in the fields preceding point Y and the waterfront road shown in Fig. 1.

Fig. 1. The 23 km trajectory of the bus route in Kingston, Canada.

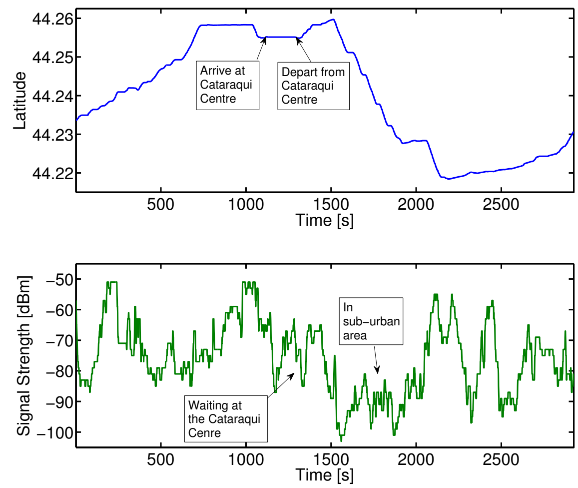


Fig. 2. Sample latitude and signal strength measurements of a bus trip.

III. Location Predictability

In this section, we investigate the location predictability of the bus trips at the different times of the day. Fig. 3 shows the latitude changes for sample trips at different times of the day. The variance between the trips is due to traffic lights, stopping at the bus stops, road traffic, and driver behavior. We can see the trips at 12 pm exhibit the highest variations as there is more road traffic and bus passengers, adding to the uncertainty in the bus location. The second half of the 7 pm trips (after the Cataraqui Centre transfer point) also show a high variation but this is partially attributed to the different departure times from the transfer point as highlighted in Fig. 3(c). The trips made at 6 pm are the most consistent. The longitude recordings show a similar behavior but are omitted due to limited space.

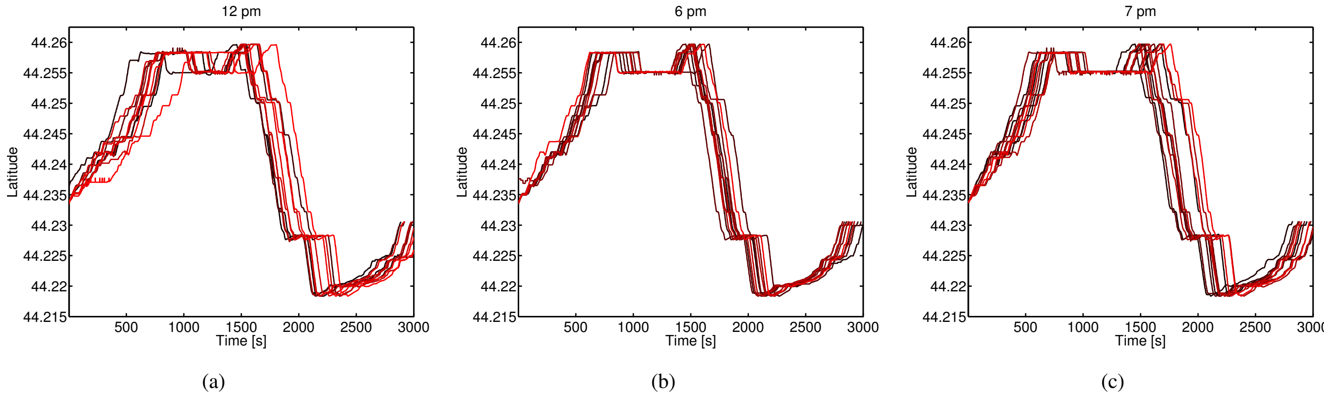


Fig. 3. Latitude variation per second for sample bus trips at (a) 12 pm, (b) 6 pm, and (c) 7 pm.

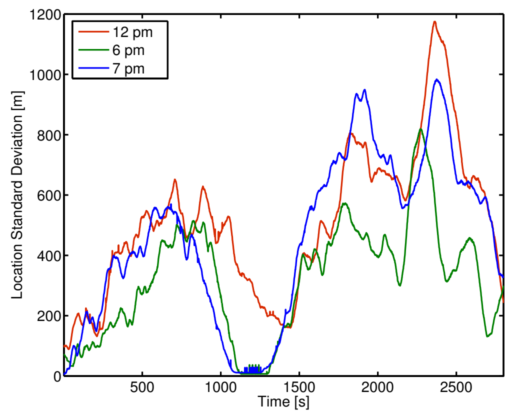


Fig. 4. Location standard deviation for different times of the day.

A. Location Variability

In order to quantify the bus location predictability we compute the location standard deviation at each second from the start of the trip. We need to calculate the distances between the average latitude and longitude measurements, and the individual trip measurements. This is accomplished using the Haversine formula [10] which is known to provide computationally precise results, as follows:

$$a = \sin^2(\Delta\mathrm{Lat}/2) + \cos(\mathrm{Lat}_1)\cos(\mathrm{Lat}_2)\sin^2(\Delta\mathrm{Lon}/2), \quad (1)$$

$$c = 2\tan^{-1}\sqrt{\frac{a}{1-a}}, \quad (2)$$

$$d = \mathrm{R}\cdot c, \quad (3)$$

where ΔLat and ΔLon are the latitude and longitude differences respectively, Lat_1 and Lat_2 are the trip latitude and average trip latitude respectively, and $\mathrm{R} = 6378.137$ m, is the radius of the Earth.

Fig. 4 shows the resulting location standard deviation for the different trip times. Referring to Fig. 3, we can see that the plots match the overall behavior of the trips. There are two major peaks of deviation, one before the Cataraqui Centre and one after. Between 1100 s and 1275 s the bus is waiting at the Cataraqui Centre transfer point, so the location is known at 6 and 7 pm. However, this is not the case at 12 pm, as the bus may or may not arrive on time due to congestion and traffic. Note that the large uncertainty at approximately 2500 seconds is due to significant *longitude* variations as the bus moves along the waterfront (Front street). Our speculation for the low deviation at 6 pm is that traffic is more consistent as it is at the end of rush hour, and before the more random bus stops and traffic in the evening.

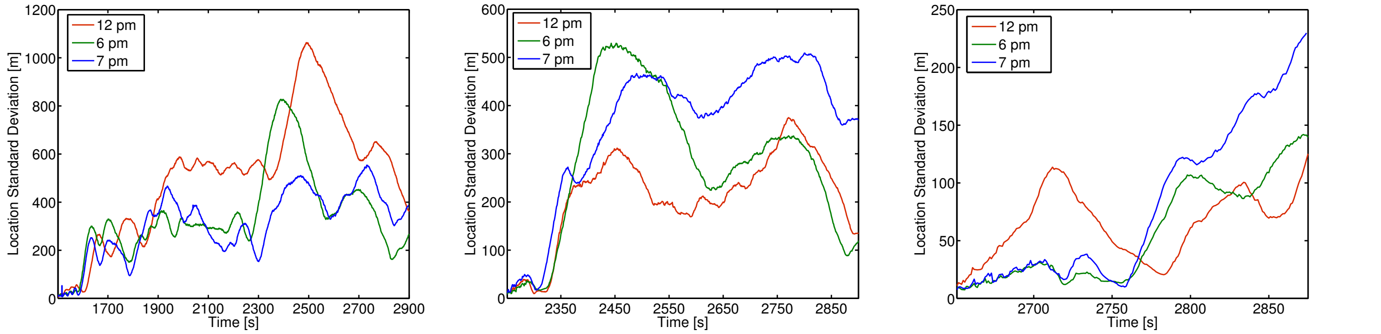
B. Effect of the Forecasting Window

The results in Fig. 4 show a very high uncertainty for the bus location after the Cataraqui Centre. However, these results are assuming that no feedback is provided throughout the trip on the bus location. As the bus traverses the suburban area, it covers large distances in small time durations. Therefore, even slightly different departure times from centre will significantly impact the location predictability. To study the effect of periodic location updating, we include three points denoted by X, Y, and Z, in Fig. 1 where the bus makes a location update. The corresponding results of the location uncertainty *after* the location update are illustrated in Fig. 5. Point X corresponds to the departure from the Cataraqui Centre and we now see a large reduction in the uncertainty between $1500-2300$ s compared to that measured in Fig. 4. A similar reduction in the location standard deviation is observed with the location updates at Y and Z in Fig. 5(b) and Fig. 5(c) respectively. However, there are still sudden increases in the location uncertainty, which are likely to arise due to bus stops and traffic lights. The results also indicate that it is possible to determine the bus location with considerably high accuracy for approximately 100 s, after which location updates are needed. Note that the location standard deviation is expected to decrease with more sophisticated predictors that examine the actual speed of the bus and can infer acceleration/deceleration as well as bus stop locations.

IV. Signal Strength Predictability

A. Results at a Glance

Fig. 6 shows the signal strength variations over time made at different times of the day. The different shades of red



(a) (b) (c)

Fig. 5. Location standard deviation after location updates at points (a) X, (b) Y, and (c) Z (denoted in Fig. 1).

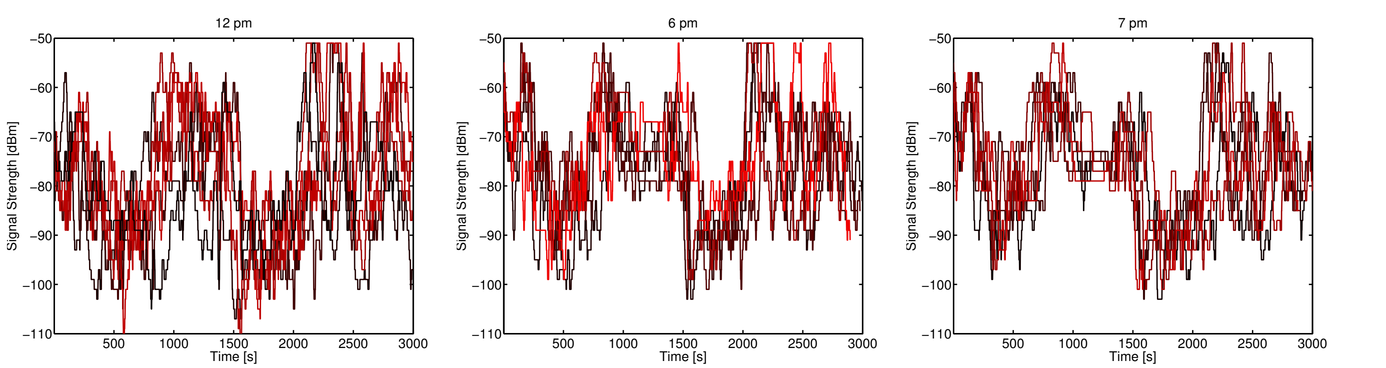
indicate different sample trips taken from our database. We can clearly see the high variability of signal strength at 12 pm, both within a single trip and between different trips. The variability is considerably less at 6 pm and even less at 7 pm. We also plot the signal strength distributions at each time in Fig. 7. From these plots we can infer that at 12 pm the signal strength exhibits lower signal strength values with higher probabilities. This is possibly due to more interference, bus passengers, network load, and road traffic. On the other hand, the distribution for 7 pm has both a lower variance and higher signal strength values. This is confirmed in the cumulative signal strength density function in Fig. 8.

B. Constructing Geographical Signal Strength Maps

Although the results in Fig. 6 appear to vary significantly between different trips, a closer look reveals that in many parts a *time translation* would reduce the variability significantly. This is due to the location variations observed in Fig. 3. Therefore, in order to evaluate the *geographic* signal strength variability, we construct signal strength maps along the bus route. To do so, the map is divided into small rectangular zones measuring 80 m longitude and 110 m in latitude (corresponding to 0.001 degrees). The signal strength measurements are then mapped to the nearest rectangular zone and the average and variance of the measurements at each zone are computed. Fig. 9 illustrates the resulting average signal strength map at 7 pm. The periodicity of signal strength variations with time are apparent with major peaks and dips along the route. Additionally, the sub-urban area suffers from a relatively long period of low signal. In addition to the average, we have also generated variance maps to investigate the geographical and temporal variance of the signal strength. The results in Fig. 10 show that at 12 pm the variance is significantly higher than at 7 pm, with particular geographical areas being affected the most.

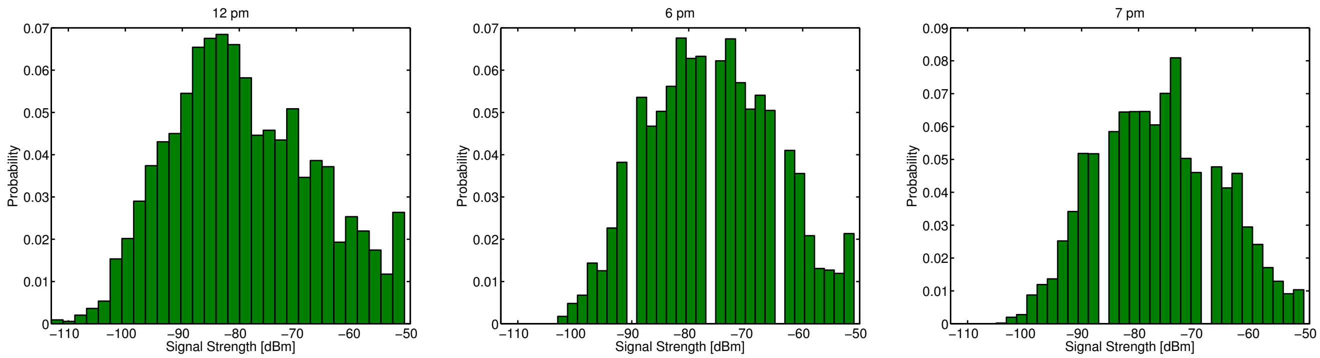
C. Effects of Geographical Context

We now divide the bus route into three geographical segments. The first is between the start point and the Cataraqui Centre, which we refer to as the *urban* segment. The second is during the wait at the Cataraqui Centre before departure, which we refer to as the *waiting* segment. The third is comprised of the remaining route after the Cataraqui Centre, and we call this the *sub-urban* segment. Next, we investigate the mean square error (MSE) of the measured signal strength for each segment, at the three different times of the day. For now, we assume perfect location information, i.e. the variability is only due to the variance in the signal strength between the different trips. In other words, we assume that the location is known, and then *reconstruct* a predicted signal strength based on the average



(a) (b) (c)

Fig. 6. Signal strength variation per second for sample bus trips at (a) 12 pm, (b) 6 pm, and (c) 7 pm.



(a) (b) (c)

Fig. 7. Distributions of signal strength along the bus route at (a) 12 pm, (b) 6 pm, and (c) 7 pm.

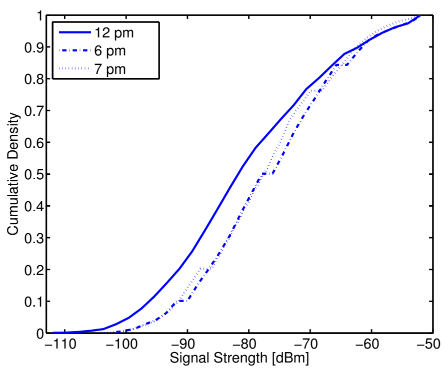


Fig. 8. Cumulative signal strength density for different times of the day.

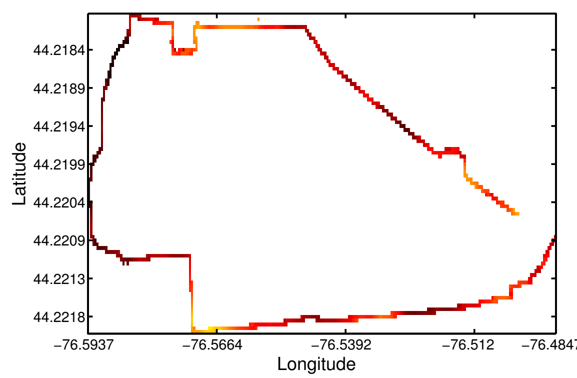


Fig. 9. Constructed signal strength map of the bus route at 7 pm.

signal strength maps shown in Fig. 9. Then, we compare the predicted signal strength to the actual traces and compute the MSE. The results are depicted in Fig. 11(a), from which we can make several observations:

- The measurements at 12 pm exhibit the highest signal strength variability, followed by 6 pm and 7 pm.
- The time waiting at the Cataraqui Centre has the highest MSE, which we suspect is due to the high volume of buses and people at the transfer point and in the shopping mall. This is supported by the observation that it does not decrease even at 7 pm.
- The sub-urban segment has the lowest MSE which was expected due to the line-of-sight areas and road along the waterfront.
- At 7 pm, the MSE for the urban segment decreases considerably. Our speculation is that road traffic and network usage is much less at this time, leading to lower interference levels.

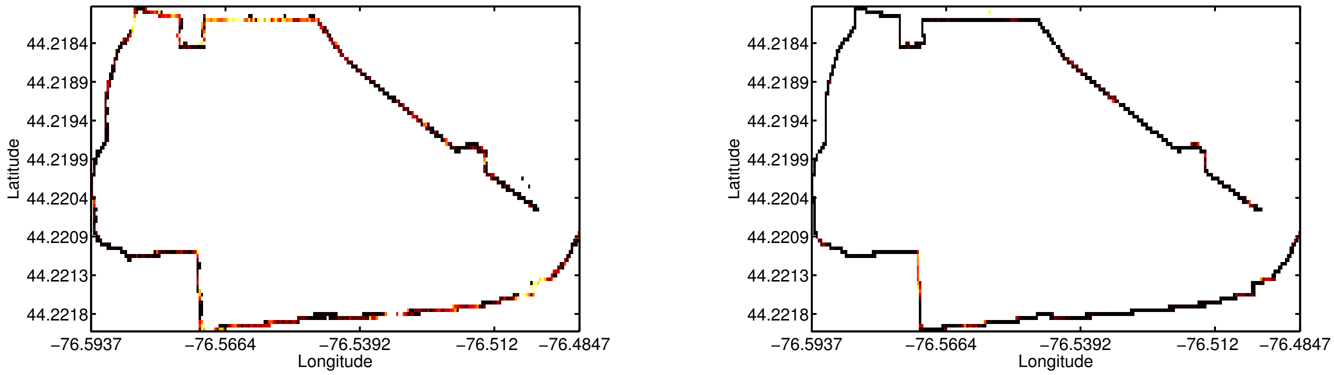
D. *Effects of Location Predictability*

In order to investigate the joint effect of location and signal strength variability, we compute the MSE with both an average location estimate and average signal strength map. The results for the *sub-urban* segment are shown in Fig. 11(b), where we can see the dramatic effect of location errors on the predicted signal accuracy. However, note that location variance of the sub-urban segment corresponds to that shown in Fig. 5(a) which has a forecast window of 1400 seconds, with no location updates. The high location uncertainty in Fig. 5(a) also matches the results of Fig. 11(b). Typically, one would not make forecasts for such a long duration without intermediate updates. Nevertheless, these results indicate that location-awareness is *key* to facilitating accurate signal strength predictions.

V. Conclusions

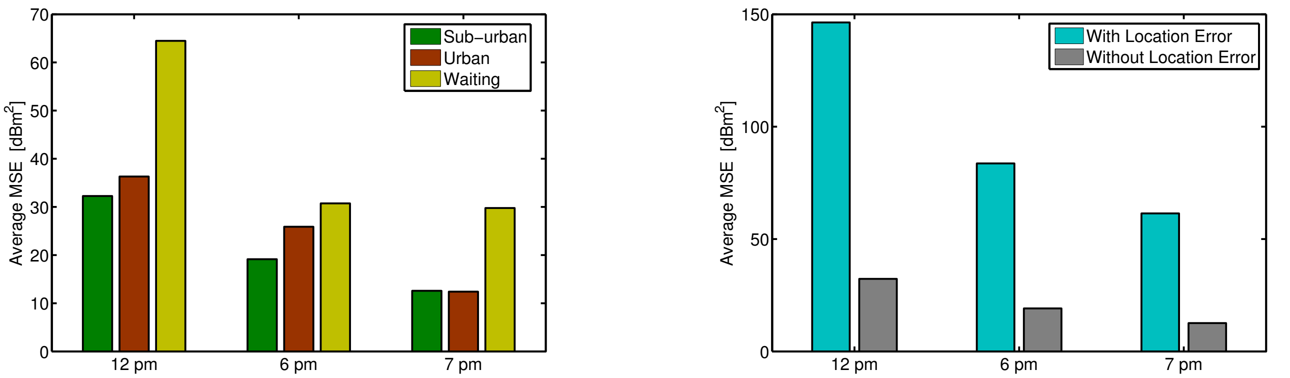
We hope that the conducted measurements and initial analysis in this paper can be used to further investigate and drive research into predictive vehicular content delivery. We now summarize our major findings and their implications.

Signal strength variability: The surveyed bus route exhibited several areas of low signal strength. Some of those were short-lived, while others were prolonged such as along the sub-urban area. This increases the need for predictive resource allocation schemes [2]–[6] to provide sustainable services. Our findings also demonstrate the importance of modeling signal strength *variability* across the routes at the different times of the day. This can provide a guideline as to *when* and *where* predictive transmission schemes can be applied.



(a) (b)

Fig. 10. Variance of the signal strength maps at (a) 12 pm, and (b) 7 pm.



(a) (b)

Fig. 11. MSE of the predicted signal strength (a) for the different road segments, (b) with location prediction errors along the sub-urban segment.

Developing mathematical models: Real measurements will be typically needed before predictive transmission schemes can be applied, due to the significant temporal and geographical variability observed. Mathematical models for signal strength will have to account for both the general statistics observed at the different times (as in Fig. 7), and the more specific geographical dependencies. The results also demonstrate that location accuracy affects the predictability significantly. Thus a primary challenge is to develop accurate location predictors that incorporate contextual factors of signal lights, bus stops and time of the day.

Optimizing the forecasting window: The optimal forecasting window duration is needed to control the tradeoff between prediction accuracy and the derived gains of predictive RA For this, measure of the joint uncertainty in signal strength and location predictability will be needed. Further, the results in Fig. 5 also indicate that contextual factors can influence the location predictability significantly, and thus it may be challenging to derive general solutions without real measurements.

References

[1] H. Abou-zeid, H. Hassanein, and S. Valentin, "Optimal predictive resource allocation: Exploiting mobility patterns and radio maps," in *Proc. IEEE GLOBECOM*, pp. 4714–4719, 2013.

[2] R. Margolies, A. Sridharan, V. Aggarwal, R. Jana, N. K. Shankaranarayanan, V. A. Vaishampayan, and G. Zussman, "Exploiting mobility in proportional fair cellular scheduling: Measurements and algorithms," in *Proc. IEEE INFOCOM*, pp. 1339–1347, 2014.

[3] Z. Lu and G. de Veciana, "Optimizing stored video delivery for mobile networks: The value of knowing the future," in *Proc. IEEE INFOCOM*, pp. 2806–2814, 2013.

[4] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2013 – 2026, 2014.

[5] H. Abou-zeid and H. S. Hassanein, "Toward green media delivery: location-aware opportunities and approaches," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 38–46, 2014.

[6] H. Abou-zeid and H. S. Hassanein, "Predictive green wireless access: Exploiting mobility and application information," *IEEE Wireless Commun.*, vol. 20, no. 5, pp. 92–99, 2013.

[7] J. Yao, S. S. Kanhere, and M. Hassan, "An empirical study of bandwidth predictability in mobile computing," in *Proc. ACM WiNTECH*, pp. 11–18, 2008.

[8] D. Han, J. Han, Y. Im, M. Kwak, T. T. Kwon, and Y. Choi, "MASERATI: Mobile adaptive streaming based on environmental and contextual information," in *Proc. ACM WiNTECH*, pp. 33–40, 2013.

[9] H. Riiser, P. Vigmostad, C. Griwodz, and P. Halvorsen, "Commute path bandwidth traces from 3g networks: analysis and applications," in *Proc. ACM Multimedia Syst. Conf.*, pp. 114–118, 2013.

[10] R. W. Sinnott, "Virtues of the haversine," *Sky and telescope*, vol. 68, pp. 158–159, 1984.