

Improving QoS in High-Speed Mobility Using Bandwidth Maps

Jun Yao, Salil S. Kanhere, *Member, IEEE*, and Mahbub Hassan, *Senior Member, IEEE*

Abstract—It is widely evidenced that location has a significant influence on the actual bandwidth that can be expected from Wireless Wide Area Networks (WWANs), e.g., 3G. Because a fast-moving vehicle continuously changes its location, vehicular mobile computing is confronted with the possibility of significant variations in available network bandwidth. While it is difficult for providers to eliminate bandwidth disparity over a large service area, it may be possible to map network bandwidth to the road network through repeated measurements. In this paper, we report results of an extensive measurement campaign to demonstrate the viability of such bandwidth maps. We show how bandwidth maps can be interfaced with adaptive multimedia servers and the emerging vehicular communication systems that use on-board mobile routers to deliver Internet services to the passengers. Using simulation experiments driven by our measurement data, we quantify the improvement in Quality of Service (QoS) that can be achieved by taking advantage of the geographical knowledge of bandwidth provided by the bandwidth maps. We find that our approach reduces the frequency of disruptions in perceived QoS for both audio and video applications in high-speed vehicular mobility by several orders of magnitude.

Index Terms—Location-dependent and sensitive, mobile computing, mobile environments, mobile communication systems.

1 INTRODUCTION

THANKS to the rapid developments and deployments of the Wireless Wide Area Network (WWAN) technology, e.g., 3G, HSDPA, and WiMAX, ubiquitous high-speed mobile data services have become more available and affordable to mobile users. It is now possible for commuters in automobiles, buses, and trains to enjoy the entire spectrum of online network services that were available to residential and enterprise users.

While service providers have devoted significant effort in improving WWAN service rate and coverage, the stability and the uniformity of the WWAN bandwidth are not guaranteed. Unlike wired networks, which have relatively stable bandwidth capacity, the WWAN bandwidth experienced by mobile users is largely dictated by the location-dependent wireless signal quality and the time-varying load dynamics in each cell. In a high-speed vehicular environment, a mobile user experiences frequent and significant bandwidth fluctuations [1], [2], [3], as a moving vehicle continuously changes its geographical location. It is well known that the location-dependent radio propagation characteristics, such as path loss, fading, and cochannel interference, affect the signal quality received by the mobile terminal [1]. To adapt to the varying radio quality, WWAN technologies employ adaptive modulation and coding schemes [4], which inherently vary the data rate of the wireless channel. The ensuing bandwidth fluctuations during a vehicular trip can seriously compromise the quality of service (QoS) experienced by network applications [2]. For

example, a sudden drop in the bandwidth can cause significant packet loss, thus distorting the quality of real-time interactive communication such as a Skype call [5].

In this paper, we seek to alleviate the aforementioned problem by exploiting the strong correlation between geographic location and WWAN bandwidth [6], [7], [8]. Our idea is based on a simple intuition that the WWAN bandwidth may hold a *Past Tells More Than Present* (PTMTP) property. That is, the summary, e.g., the mean value, of the past bandwidth observations at a particular location is a better indicator of the actual bandwidth experienced at that location, as compared to the bandwidth at the previous location along the route. To confirm this intuition, we conduct an eight-month long vehicular measurement campaign to collect bandwidth data on three different WWAN providers over two different routes in Sydney, Australia. Our analysis of the bandwidth traces overwhelmingly confirms the existence of the PTMTP property. This finding implies that the past bandwidth knowledge may help mobile applications to foresee the impending bandwidth fluctuations during a vehicular trip and take proactive action. We also find that the time of day influences mobile bandwidth, but location is the most dominant factor. Hence, we exclusively focus on the correlation between bandwidth and location in this research.

We store past bandwidth performance data in the form of *bandwidth maps*, by mapping the average value of historical network bandwidth observations to the existing road network. The bandwidth samples are collected by repetitive measurements. These maps provide information such as “if you are near the intersection of Street X and Street Y, you can expect an average bandwidth of 1.2 Mbps from network provider A.” Various aspects of on-board communication systems, e.g., admission control, data prefetching, multihoming scheduling, and applications such as multimedia streaming and video conferencing, may benefit from this information. In this paper, we use two

• The authors are with the School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia.
E-mail: {jyao, salilk, mahbub}@cse.unsw.edu.au.

Manuscript received 20 Apr. 2010; revised 21 Nov. 2010; accepted 28 Jan. 2011; published online 29 Apr. 2011.

For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2010-04-0181. Digital Object Identifier no. 10.1109/TMC.2011.97.

concrete examples to explain our idea. In the first instantiation, we interface the bandwidth maps with the rate adaptation agents of adaptive video streaming applications that are run on personal mobile devices. Our trace-driven simulations highlight that our approach can help streaming applications to achieve fast and smooth adaptation to the varying bandwidth quality when the vehicle is on the move. We find our scheme effectively improves the user perceived QoS by up to *five folds*, as compared to the state-of-the-art reactive adaptation techniques. In the second example, we demonstrate that bandwidth maps can improve the QoS offered by a vehicular gateway equipped with multiple WWAN interfaces. The proposed principle enables the network traffic schedulers of vehicular gateways to intelligently use the location-specific bandwidth information in making scheduling decisions. Using audio streaming as an example, our simulation results show that this approach can improve the overall QoS by up to *four folds*, in comparison with current schemes.

The rest of the paper is organized as follows: we review related work in Section 2. In Section 3, we study the PTMTP property and its implications on mobile applications, using the empirical bandwidth traces collected from our vehicular measurement campaign. In Section 4, we present bandwidth maps as the means to capture the location-bandwidth dependency of WWAN networks. In Sections 5 and 6, we demonstrate the usefulness of bandwidth maps through two different representative vehicular network applications. Finally, Section 7 concludes the paper.

2 RELATED WORK

Our work is inspired by recent findings that highlight the strong correlation between geographic location and network performance. Global Internet maps such as [9] illustrate that the statistical properties of wired Internet bandwidth vary as a function of geographical region (e.g., country and continent). Residential broadband ISPs, e.g., [10], often provide coverage maps that indicate the expected service data rate depending on the customer's street address. This is even more true for wireless networks. WWAN and Wi-Fi service providers often conduct network measurements within their service domains and release coverage maps to indicate the expected service quality at different regions. However, these maps only provide coarse information of the network performance. Empirical studies [1], [2], [7], [8] have revealed that the performance of the last-hop wireless link can vary significantly from location to location, even at the granularity of a few hundred meters on the same street. Independent organizations, e.g., [11], now provide street-level coverage maps of WWAN and Wi-Fi networks. These maps are typically profiled from the measurement data reported by voluntary smartphones running specific measurement applications. The maps are freely available online, which provide an unbiased comparison of different network carriers and aid consumers in selecting carriers. In this paper, we are interested in *personalized* bandwidth maps, which store historical bandwidth measurements of individual mobile users that are relevant to their mobility patterns. We seek to capitalize on the bandwidth information provided by such maps for

improving the QoS experienced by mobile Internet applications that are run from moving vehicles.

Several empirical studies [1], [2], [3] have investigated the WWAN bandwidth performance in a vehicular scenario. A common observation is that the WWAN bandwidth deteriorates significantly under vehicular mobility as compared to a stationary environment. The goal of our measurement campaign is different from these prior studies. We intend to investigate if past bandwidth knowledge at a certain location can be a useful indicator of the current bandwidth experienced at that location. In order to present statistically conclusive results, we conduct a large set of *repeated* measurements along two distinct routes in Sydney. To the best of our knowledge, this is the first empirical study to present results of WWAN network performance from several repeated measurements.

Recent works [12], [13] have proposed to profile the performance of roadside Wi-Fi APs using historical observations. By examining the profiles of all "visible" APs at a given location, the mobile device can intelligently select the "best" AP to associate with, thus optimizing the handoff decision. Our work shares the same design philosophy, i.e., capitalizing on the strong correlation between wireless network performance and location. However, unlike prior research, we seek to characterize the user-experienced WWAN bandwidth conditions from past observations, while driving along urban roads. Another novel aspect of our work is that we use the *a priori* bandwidth information at each location to proactively react to any impending bandwidth fluctuations and thus improve the QoS of mobile Internet applications.

3 WHAT'S PAST IS PROLOGUE

It is well known that history repeats itself. There are several examples of this phenomenon, both in nature (e.g., weather) and human influenced activities (e.g., human mobility patterns, Internet usage). In this section, we investigate if the downlink WWAN wireless bandwidth in a high-speed vehicular environment exhibits a similar trend. We present an overview of our empirical measurement campaign for collecting wireless bandwidth traces under typical driving conditions. We analyze the data and show that the average value of past bandwidth observations collected from the same location is a better indicator of the current bandwidth experienced at that location than the bandwidth encountered at the previous location during the current trip. Finally, we explain how this property can be used by mobile applications to maintain QoS under the varying bandwidth conditions experienced while driving.

We undertook a measurement campaign for profiling the WWAN wireless bandwidth in a high-speed vehicular environment. We developed a simple client-server measurement system (see Fig. 1) using off-the-shelf hardware. The server was housed in our lab at the University of New South Wales (UNSW). The client (as shown in Fig. 2) comprised of two Soekris Net4521 boards interconnected via 10 Mbps Ethernet and configured in master/slave mode. The boards were enclosed in a protective casing, connected to a power generator and housed in the boot of a car. Three PCMCIA cellular modems were housed in the

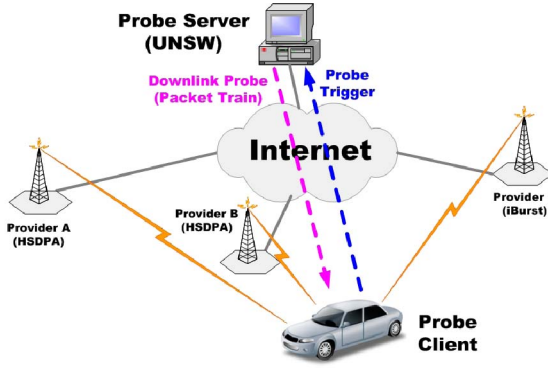


Fig. 1. Measurement architecture.

system (two in the master board and one in the slave board). To account for network and technology diversity, we simultaneously conducted measurements over two HSDPA [4] networks (A and B) offering different service rates and a network (C) supporting a pre-WiMAX proprietary standard, iBurst [14]. The advertised peak rates for A, B, and C were 7.2, 3.6, and 1 Mbps, respectively. To enhance the wireless signal reception, the cellular modems were connected to external antennas mounted on the car windshield. A Garmin GPS18 GPS sensor was connected to the client for recording the vehicle location.

We developed a lightweight packet-train program to measure the WWAN bandwidth, which achieves fast convergence and generates minimal network traffic. We refer readers to [7] for further details about the program and validations. We collected one bandwidth sample for approximately every 200 m section of the route (by adjusting the sampling interval according to the vehicle speed as reported by the GPS). The samples are tagged with location coordinates and time, and stored in a repository. On occasions, some packets in the train were lost, leading to some missing samples. To deal with these missing samples, we used 500 m as the smallest location granularity (note that, no two successive samples were missing in our data). Thus, the bandwidth for a segment is represented by the average of all samples collected within it.

We collected bandwidth samples by driving the car along two distinct routes in the Sydney metropolitan area. Fig. 3 depicts the trajectory of the chosen routes. The starting point of the *inbound* (7 Km) route is UNSW in the eastern suburbs, with the final destination being Sydney CBD. The *outbound* (16.5 Km) route runs from Sydney CBD to Macquarie University (MQ) located in the northwestern suburbs. The chosen routes are typical representations of daily commute.

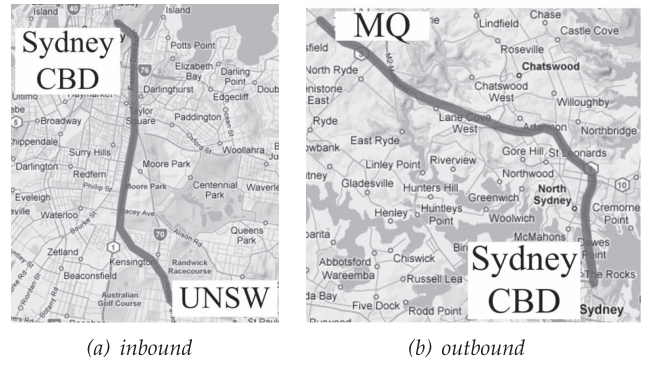


Fig. 3. Trajectory of chosen routes.

Recall that our goal is to determine if historical bandwidth measurements can be helpful for future trips. To draw meaningful conclusions, we need to ensure that the samples collected are sufficient to estimate the actual bandwidth characteristics at each location. In our measurements, each trip only yields one bandwidth sample for each location along a route. Collecting a large number of samples requires making repeated trips along the selected routes. However, conducting such measurements involves significant monetary cost (fuel, mobile bandwidth), manpower, and time. Thus, an important question is—“How many samples (i.e., repeated trips) are sufficient?”

The issue of sample sufficiency is a well-studied problem in measurement and survey studies [15]. The standard strategy is that the sample size is deemed sufficient when the margin of error E of the estimate on the statistic of interest (e.g., the mean) is small enough [15]. Recall that, E is equal to one half the width of the Confidence Interval (CI), and reflects the precision of the estimate. For example, “an E at $(1 - \alpha) \times 100\%$ confidence level” means that one can be $(1 - \alpha) \times 100\%$ confident that the estimate will not differ from the true value of the statistic by more than E . Clearly, the smaller the margin of error, the more precise the estimate. Thus, the sample size required in our measurement campaign largely depends on how precisely we want to estimate the WWAN bandwidth statistics. We are interested in estimating the mean bandwidth μ_l at each location l using the sample mean \bar{b}_l^n observed from the past n trips. For this, we calculate the margin of error for estimating μ_l . Let us denote s_l as the sample standard deviation at location l , n as the sample size, and $t_{n-1, 1-\alpha/2}$ as the upper critical value for $(1 - \alpha) \times 100\%$ CI from the t -distribution with $n - 1$ degree of freedom. Then, from the definition of margin of error [15], we have

$$E = t_{n-1, 1-\alpha/2} \times \frac{s_l}{\sqrt{n}}. \quad (1)$$

Further, we normalize the margin of error by computing its ratio to the sample mean \bar{b}_l^n , i.e.,

$$\epsilon = \frac{E}{\bar{b}_l^n}. \quad (2)$$

After completing each measurement trip (which yields one sample at each location), we recompute ϵ at each location for each provider, using (2). As is common in statistical studies, we select 95 percent as the confidence level, i.e.,

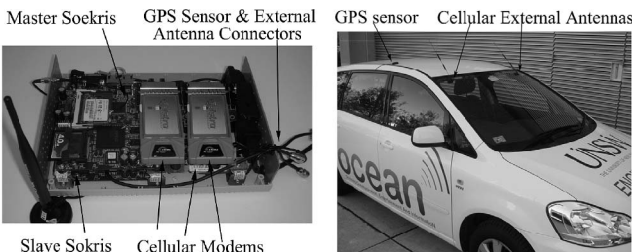


Fig. 2. Measurement setup.

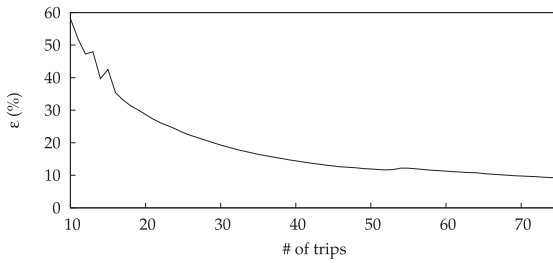


Fig. 4. Margin of error as a function of # of trips made (at location #12 of the inbound route for provider B).

$\alpha = 0.05$. Fig. 4 plots the evolution of ϵ at one particular location for provider B, as the sample size grows. The graph shows that with the availability of more samples, ϵ initially decreases sharply. However, after about 70 samples, the improvement in ϵ is marginal with each additional sample. Further, the ϵ drops below 10 percent, which implies that \hat{b}_i^n is now very close to the true mean μ_i . Based on similar observations at all other locations for all providers, and given the high costs associated with the measurements, we decided to cease the measurement after making 75 repeated trips. To account for the impact of different time of day on the WWAN bandwidth, we conducted the measurements at different times including peak and off-peak commuting hours and different days of the week. Collectively, our measurements spanned over the eight-month period from August 2007 to April 2008, clocked more than 60 driving hours, and covered over 1,600 Km.

It has been reported that WWAN bandwidth fluctuates in both time and space [8] in a stationary environment. We first investigate if similar observations can be confirmed in mobile environments by analyzing our empirical bandwidth traces. To begin with, we study the impact of location on bandwidth. We compute the average bandwidth at each location, assuming 500 m as the location granularity. As a representative example, Fig. 5 plots the average bandwidth for each location (loc. avg) along the inbound route for provider C. Note that, we observe similar results (omitted for brevity) for the outbound route and other providers. It is evident from this graph that the WWAN bandwidth is strongly influenced by location. This can be attributed to the location-dependent characteristics of the wireless channel such as path loss, shadowing, cochannel interference, and fading. Another observation is that the bandwidth conditions between consecutive locations can vary significantly (e.g., location #8 and #9). This highlights the fact that a fast-moving vehicle can experience drastic changes in bandwidth.

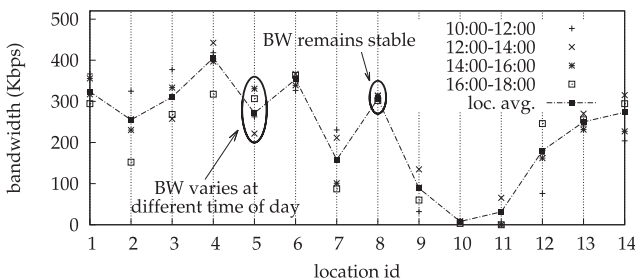


Fig. 5. Average bandwidth along the inbound route (provider C).

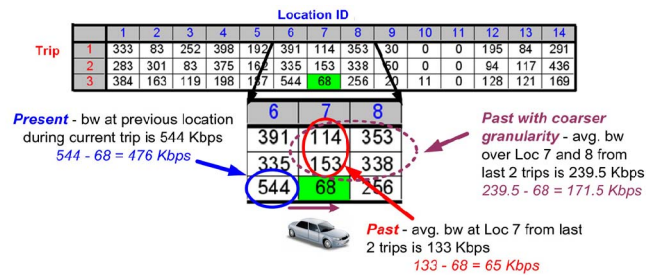


Fig. 6. Past versus Present (Provider C, inbound route).

Next, we investigate if the time of day influences the WWAN bandwidth at a given location. Recall that, our measurement trips were randomly conducted at different time of day. We partition the bandwidth samples into different time intervals using a granularity of 2 hours. This results in four nonoverlapping intervals, i.e., 10:00-12:00, 12:00-14:00, 14:00-16:00, and 16:00-18:00. Next, for each time interval, we compute the corresponding average bandwidth at each location (assuming 500 m granularity as before) and plot the same in Fig. 5. At some locations such as location #5, the WWAN bandwidth does appear to vary noticeably with time. This can be attributed to the time-varying characteristics of the wireless channel such as fading and interference and also the network load dynamics (e.g., number of active users). However, for the majority of locations, the impact of time is not as pronounced. We make similar observations from the data from other trips and providers (not shown for reasons of brevity).

In summary, the above analysis indicates that there is a strong correlation between WWAN bandwidth and location. Time of day also influences the bandwidth, but location is the most dominant factor. Hence, in this paper, we focus on how best we can leverage the association between mobile bandwidth and location, since we expect that it will achieve the most significant gains. That said, our ideas can be readily extended to include time of day.

The strong correlation between location and bandwidth observed above suggests that the average of past measurements at a particular location can be a better indicator of the current bandwidth at that location than the bandwidth experienced at the previous location during the current trip. For this, we use the bandwidth observations collected for provider C during the first three trips along the inbound route (see Fig. 6). Note that, the location granularity is 500 m. As is evident, the bandwidth fluctuates significantly during each individual trip. Further, observe the strong correlation between bandwidth and location. For example, there is a constant outage at location #11. On the other hand, the bandwidth experienced at location #1 is consistently close to 300 Kbps. We focus on the third trip, just as the car is about to enter location #7, as shown in Fig. 6. Observe that, there is large difference between the current bandwidth (at location #7), 68 Kbps and the bandwidth experienced at the previous location (#6), 544 Kbps. In contrast, the average of the past two bandwidth observations at location #7, 133 Kbps, is much closer to the current bandwidth. This example suggests that, in a high-speed vehicular scenario, the *past* tells us more than the *present* about WWAN bandwidth. We formally define the above property as follows:

TABLE 1
Statistics Demonstrating PTMTP

Prov.	Locations with PTMTP		Average statistics				
	Inbound	Outbound	$\Delta_{present}$	$\Delta_{past-500m}$	$\Delta_{past-1000m}$	$\Delta_{past-2000m}$	$\Delta_{past-time}$
A	94%	85%	438 Kbps	350 Kbps	365 Kbps	388 Kbps	348 Kbps
B	93%	81%	245 Kbps	196 Kbps	202 Kbps	220 Kbps	195 Kbps
C	100%	100%	331 Kbps	218 Kbps	225 Kbps	240 Kbps	200 Kbps

Definition 1 (Past Tells More Than Present). Let b_l^n denote the bandwidth at location l during trip n . Let \bar{b}_l^{n-1} denote the mean bandwidth value based on past $n-1$ observations (during trips 1 to $n-1$) for location l . The PTMTP property holds if the difference between b_l^n and \bar{b}_l^{n-1} , Δ_{past} , is smaller than the difference, between b_l^n and b_{l-1}^n , $\Delta_{present}$, i.e.,

$$\Delta_{past} < \Delta_{present}. \quad (3)$$

Instead of using the absolute difference as in the illustrative example, we employ the more robust Root Mean Square (RMS) difference in our definition. Thus,

$$\Delta_{past} = \sqrt{\frac{\sum_{n=2}^{n=N} (\bar{b}_l^{n-1} - b_l^n)^2}{N-1}}. \quad (4)$$

Similarly,

$$\Delta_{present} = \sqrt{\frac{\sum_{n=1}^{n=N} (b_l^n - b_{l-1}^n)^2}{N}}. \quad (5)$$

We use the bandwidth samples collected from our measurements to evaluate if WWAN bandwidth exhibits the PTMTP property. With the default 500 m granularity, we compute $\Delta_{present}$ and $\Delta_{past-500m}$ for each location along both routes and for all three providers and compare the corresponding values. Since, we have collected 75 samples at each location, $N = 75$. Table 1 illustrates the percentage of locations, which exhibit the PTMTP property for different providers and routes. The data clearly show that at a vast majority of locations, WWAN bandwidth does indeed satisfy the PTMTP property.

Next, we investigate if the location granularity has an influence on the PTMTP property. We first explore this in the context of the example in Fig. 6. Consider that the bandwidth information from the first two trips is recorded at a coarser granularity of 1,000 m (cf. 500 m as in the original example). Location #7 and #8 are now part of the same 1,000 m segment and hence the bandwidth samples recorded during the individual segments are averaged to represent the bandwidth for the entire 1,000 m segment. Thus, the bandwidth recorded during trips 1 and 2 for this segment would be 233.5 and 245.5 Kbps. As in the previous case, consider that the car is about to enter location #7 during the third trip. The average from the past bandwidth observations at location #7 is now 239.5 Kbps, which is still significantly closer to the current bandwidth, 68 Kbps than the bandwidth at the previous location (#6), 544 Kbps. This suggests that the PTMTP property still holds even if the past data are maintained at a coarser location granularity. However, the difference between the past and present increases with an increase in the location

resolution (65 Kbps at 500 m as compared with 171.5 Kbps at 1,000 m). This suggests that finer granularity data would be more useful.

To quantitatively study the effect of the location granularity at which past bandwidth information is maintained on the PTMTP property, we compute $\Delta_{past-500m}$, $\Delta_{past-1,000m}$, and $\Delta_{past-2,000m}$, where the past bandwidth samples \bar{b}_l^{n-1} are recorded at different granularities of 500, 1,000, and 2,000 m, respectively. Recall that, 500 m is the smallest location granularity available from our measurement data. From Table 1, we observe that irrespective of the location granularity, the past bandwidth statistics are consistently a better indicator of the current bandwidth. In other words, the PTMTP property holds unanimously for different location resolutions. However, Δ_{past} is greater when the location granularity is coarser. This implies that maintaining past bandwidth statistics at finer granularity leads to a better approximation of the current bandwidth.

However, considering the practicalities of the real-world environment, which in the context of this research is a fast-moving vehicle, the location granularity cannot be arbitrarily small. For example, under typical urban driving speeds of 60 Kph, it only takes 3 s for a vehicle to travel 50 m. As such, the location granularity should be large enough such that changes made by mobile applications and protocols (see examples in Sections 5 and 6) in response to changes in location have had sufficient time to take effect. The location granularity is also constrained by the accuracy of the state-of-the-art localization techniques. Current GPS receivers report location coordinates with a typical error of about 10 m. However, the error can increase to up to 100 m in dense built-up metropolitan areas [16]. Hence, the location granularity must be large enough to mask the effects of positioning errors. Based on all of the above, we have used 500 m as the smallest granularity in this paper.

Finally, we investigate the impact of time of day on the PTMTP property. For this, we separate the bandwidth samples into the four 2-hour intervals used in our earlier analysis (see Fig. 5). Next, we compute $\Delta_{past-time}$ according to (6), where $\bar{b}_{t,l}^{n-1}$ denotes the mean bandwidth at location l from the past $n-1$ observations collected during the time interval t and $b_{t,l}^n$ denotes the current bandwidth at location l during the same time interval t .

$$\Delta_{past-time} = \sqrt{\frac{\sum_{n=2}^{n=N} (\bar{b}_{t,l}^{n-1} - b_{t,l}^n)^2}{N-1}}. \quad (6)$$

The $\Delta_{past-time}$ results listed in Table 1 are averaged over all locations (assuming 500 m granularity) from both routes. Comparing with $\Delta_{past-500m}$, we observe that considering the time of day as additional context does lower the

estimation error. However, the improvement is clearly marginal. The results confirm our earlier observation that the influence of time on mobile bandwidth is not as pronounced as location.

Now, we discuss how the PTMTP property can be leveraged by mobile applications. As is evident from our sample data in Fig. 6, WWAN bandwidth can fluctuate frequently when moving at vehicular speeds. To maintain QoS, network applications and in particular multimedia applications monitor the current bandwidth and react accordingly by varying certain application settings. For example, adaptive multimedia streaming applications [17] reactively adjust the streaming rate in response to the observed path conditions. These applications deploy monitoring agents (at the server and/or client) for continuously monitoring the bandwidth [18]. However, in the context of vehicular mobility, such reactive schemes are constantly playing catchup. This is because by the time the agent has learned about the current state of the network, the vehicle is likely to move to the next location, where the network conditions may be radically different from the previous location. The PTMTP property indicates that the average value from the past bandwidth measurements is a fair approximation of the present bandwidth at a location. As a result, if this a priori information is available to the application, then it can proactively take appropriate action to facilitate a smoother transition to the new conditions. Further, the monitoring agent can learn about the current state of the network quicker than when this information is not available. Let us again consider the example in Fig. 5. Assume that the vehicle is about to move to location #7 during the third trip. Further, assume that the monitoring agent has learned of the bandwidth at location #6, which is 544 Kbps. Observe that, there is a large difference in the bandwidth at these two adjacent locations. As a result, the monitoring agent requires significant time to learn about the conditions at location #7. This impacts the application QoS during this entire duration. In the worst case, the vehicle may already move to the next location before the learning period is over. Note that, the mean value of past bandwidth observations at location #7 is 133 Kbps. If the monitoring agent knows this, it can rapidly converge to the current bandwidth (68 Kbps) and thus ensure minimal disruption of QoS. We present two concrete examples of this idea in Sections 5 and 6.

4 BANDWIDTH MAPS

In the previous section, we demonstrated using empirical data that WWAN bandwidth exhibits the PTMTP property. We showed that historical knowledge of location-specific bandwidth can be leveraged by mobile applications to rapidly adapt to the bandwidth fluctuations that are inherent in a high mobility scenario. The obvious question is: How to collect and represent location-specific bandwidth information? We propose the creation of *bandwidth maps*, as a means for representing bandwidth data, by superimposing summary statistics of past bandwidth observations on road network maps. Mobile applications can use these maps and learn about the expected bandwidth conditions at future locations. In this section, we discuss methods for creating such maps.

The bandwidth samples can be collected either using active or passive measurement tools. Active measurements can be conducted using a setup similar to the one employed in our measurement campaign (Section 3). The active measurement client can be readily deployed on the mobile device (vehicular gateway, mobile phone, etc.). The measurement server can be hosted by a volunteer organization or the WWAN service providers. The advantage of the active approach is the ability to control the measurement process, e.g., the sampling frequency. One drawback of this method is that it introduces overheads, since probe packets are actively injected into the network. However, the bandwidth usage of the probes can be kept to a minimum by appropriately tuning the measurement parameters [7]. In the passive measurement approach, a simple client can be installed on the mobile device, which can infer the bandwidth by monitoring the network activity over the wireless interface [19]. Since passive techniques do not inject any traffic, they do not incur overheads. However, the measurement can only be conducted when there is network activity, thus providing little control. The bandwidth samples collected (using either approach) must be tagged with the corresponding location coordinates and time and stored in a repository (either at the client or server) for later processing. When the mobile client has built-in GPS receiver, the location coordinates required for geotagging can be directly collected from the receiver. Alternatively, Wi-Fi or cellular localization techniques [20] could be employed, albeit with a coarser location granularity.

Recall that, in Section 3, we discussed the importance of collecting a sufficient number of bandwidth samples. In our measurement campaign, we intentionally made repeated vehicular trips along the chosen routes with the objective of collecting a large set of samples. While we cannot expect ordinary citizens to do the same, the data collection process can certainly leverage the fact that human mobility patterns are fairly repetitive. It is well known that human beings are creatures of habit. Analysis of empirical human mobility traces [21] has revealed significant similarity in our day to day movement patterns. This is quite intuitive; for example, on weekdays, we travel to and from work along the same route. This is even more so true of public transport vehicles such as buses and trains, since they follow specific routes according to fixed timetables. Thus, if the measurements are conducted continuously whenever the mobile device is on the move, there is a high probability of collecting multiple samples for each location along the visited routes. The resulting bandwidth maps would be personalized, in the sense, that they would contain bandwidth statistics along the routes frequented by the device carrier. As a result, the bandwidth maps have greater relevance since future trips are very likely to include the locations contained in the maps. The maps could also be shared with other users via websites as is popular with Wi-Fi wardriving communities. Public transport providers can maintain a large repository of maps which are collected by all their vehicles. These maps may also be of interest to network providers who may offer incentives (e.g., free credits) to users for contributing their data.

The location granularity can directly impact the performance of the applications that leverage the information



Fig. 7. A portion of the bandwidth map (provider A, outbound route).

offered by the bandwidth maps. If one employs coarse granularity, an entire geographical region such as a suburb would correspond to a location segment. We refer to this representation as *region level*. On the other hand, one could use an extremely fine grained representation, *segment level*, wherein a small section of the road (e.g., 500 m section) corresponds to a location segment. Between these extremes, is the granularity of a route, i.e., a sequence of road segments, referred to as the *route-level* representation. We have compared the three in our earlier work [22] and shown that segment-level statistics provide the best performance. In the rest of this paper, we will assume that a location corresponds to a 500 m road segment.

We have constructed bandwidth maps from the bandwidth traces collected in Section 3. The map for each route depicts the average bandwidth at each location for each WWAN provider. These maps are used in the simulations in Sections 5 and 6. Recall that, in the illustrative example in Section 3, we used the mean of the past bandwidth samples as the summary statistic. However, service providers are continually upgrading their network software and hardware for improving coverage and bandwidth. Thus, it is desirable to give more emphasis to recent samples and phase out old data. Hence, we use the exponential weighted moving average (EWMA) of the bandwidth samples as the summary statistic. The average bandwidth (\bar{b}_l^n) at location l is updated when a new sample for that location, b_l^n is available as follows:

$$\bar{b}_l^n = \alpha \times b_l^n + (1 - \alpha) \times \bar{b}_l^{n-1}. \quad (7)$$

Our experiments (excluded for reasons of brevity) have shown that a suitable value for the smoothing factor, α is 0.125. Fig. 7 depicts a fraction of the bandwidth map of the outbound route for provider A, assuming a granularity of 500 m. Note that, higher order stochastic models, such as fixed/variable order-N Markovian, Lezi and PPM [21], [23] can be used in place of the average. However, our simulations in later sections demonstrate that even using simple statistics such as the average bandwidth leads to significant performance gains. We intend to investigate higher order models in our future work.

5 ADAPTIVE STREAMING

In this section, we show how bandwidth maps can be used with adaptive video streaming running on personal mobile

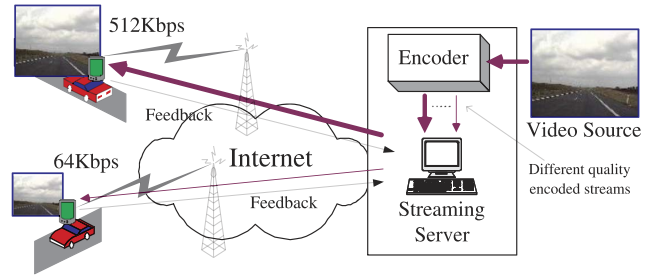


Fig. 8. Adaptive video streaming.

devices. We first identify the problems experienced by the state-of-the-art in video rate adaptation techniques in a mobile environment. Next, we explain the principle and design of our enhancement using bandwidth maps. We evaluate the performance of our scheme through simulations, using our empirically collected bandwidth traces (from Section 3).

5.1 Background

Adaptive streaming is a popular technique that dynamically adjusts the quality of a video stream during playback, based on changing network conditions. Fig. 8 illustrates a typical scenario, wherein, an adaptive streaming server encodes a video file (or live event) at multiple resolutions and bit rates. During playback, the server can transparently change the streaming rate to seamlessly adapt to the changes in the available bandwidth, by switching the stream between the pre-encoded videos with different qualities. For example, the server can switch to a lower rate stream when it detects that the available bandwidth reduces, thus graceful degrading the viewing quality.

To dynamically determine the appropriate streaming bit rate and quality, the server needs to closely track the underlying bandwidth conditions. Several such rate adaptation algorithms have been proposed for streaming applications [17], [24], [25], [26]. These algorithms are reactive in nature, in that, they monitor the video playout buffers or certain end-to-end network parameters (i.e., delay and packet loss) and adapt the streaming rate depending on the changes in the observed values. Among these algorithms, TCP Friendly Rate Control (TFRC) [26] is a popular choice [27]. Thus, we use TFRC as an illustrative example to explain our idea in the rest of this section.

TFRC maintains a similar average sending rate as TCP running under comparable network conditions, while providing a relatively smooth sending rate, which helps packets to meet the real-time constraints required by streaming media [26]. TFRC inherits the TCP slow-start mechanism. Once the receiver reports a packet loss event, the TFRC sender enters the congestion avoidance state. In this state, the TFRC sender controls the sending rate based on the following simplified TCP throughput model:

$$TFRC_{rate} = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RTO}\sqrt{\frac{2p}{3}}p(1 + 32p^2)}. \quad (8)$$

In the above, p denotes the loss event rate, which is received as feedback from the receiver, t_{RTO} refers to the TCP retransmission time-out, and s is the packet size. When used in conjunction with an adaptive streaming algorithm,

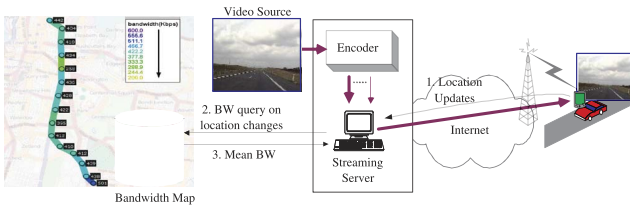


Fig. 9. Interfacing bandwidth maps with adaptive video streaming.

the streaming bit rate is directly controlled by the sending rate $TFRC_{rate}$.

However, TFRC is known to be slow to react to frequent changes in bandwidth [28], [29]. Consider a scenario where a user in a moving vehicle is streaming a video from an Internet server, wherein the streaming rate is adaptively controlled with TFRC. In this case, our simulation studies in Section 5.3 show that the slow convergence of TFRC leads to suboptimal streaming quality. For example, consider the case when the WWAN bandwidth increases suddenly as the vehicle enters into a new location. Now, the TFRC sending rate, computed using (8), is slow to increase to the actual bandwidth at the new location. The duration when the TFRC rate is lower than the actual bandwidth represents missed opportunities, because the streaming server could have possibly streamed with higher quality. A similar behavior is expected during the initial TCP-like slow-start period, since the TFRC rate slowly ramps up at the start of a session. When the bandwidth drops significantly at a new location, the TFRC sending rate is also slow in reacting to the change, thus remaining greater than the actual WWAN bandwidth. During this period, it leads to congestion over the WWAN link, which potentially leads to packet loss. Lost packets in turn lead to lost frames, which affect the streaming quality.

5.2 Enhancement with Bandwidth Maps

We refer to our enhancement for TFRC as *BW-MAP-TFRC*. Note that, using the same design principle, bandwidth maps can be readily incorporated in other rate adaptation algorithms.

Our idea is to proactively *bootstrap* the TFRC sending rate to the location-specific mean bandwidth (obtained from bandwidth maps) when the vehicle enters into a new location. As a consequence of the PTMTP property, it is highly likely that this proactive operation allows the TFRC sending rate to quickly converge to the actual bandwidth as compared to the purely reactive strategy adopted by plain vanilla TFRC (recall the illustrative example from Fig. 6). As shown in Fig. 9, our idea requires minimal changes to the video streaming infrastructure. We assume that the streaming server has access to bandwidth maps. We assume that the streaming client running on the user's mobile device is aware of its current location. The client updates the server when it changes its location. The server looks up the bandwidth map and determines the historical average bandwidth at that location. The rate adaptation agent, BW-MAP-TFRC sender, at the server is forced to change its sending rate to this mean value. This is achieved by freezing the sender state for a short period and disabling the normal operation of TFRC. Once this rate change is effected, normal TFRC operation resumes.

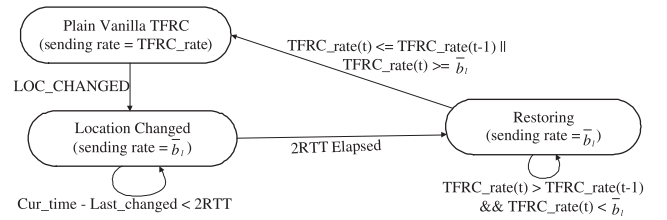


Fig. 10. State machine of BW-MAP-TFRC.

We use a state machine (Fig. 10) to explain BW-MAP-TFRC in detail. The normal operation of TFRC is represented by the *Plain Vanilla TFRC* state. In this state, the sending rate is determined as per (8). BW-MAP-TFRC introduces two additional states at the sender side: *Location Changed* and *Restoring*. The receiver remains unchanged. When the BW-MAP-TFRC receiver moves to a new location l , the sender transitions to the *Location Changed* state. The sending rate is forced to \bar{b}_l , the mean bandwidth at l as reported by the bandwidth map. The BW-MAP-TFRC sender remains frozen in this state for a period to ensure that the receiver starts to provide feedback based on the new rate. Note that, in order to ensure the successful rate enforcement, the duration of the freezing time should be long enough to ensure reliability by filtering out the effects of delay jitter. Our experiments (not reported for reasons of brevity) have found that a time period of $2RTT$ is sufficient for this purpose. Following this, the sender enters the *Restoring* state, wherein the sending rate is still maintained at \bar{b}_l . However, unlike in the *Location Changed* state, the BW-MAP-TFRC sender resumes to estimate the $TFRC_{rate}$ according to (8). Note that, even though the average bandwidth from the bandwidth map is close to the actual link bandwidth, it will never be exactly equal. Hence, it is important that the sender begins to monitor the feedback from the receiver to gauge if the current sending rate \bar{b}_l is either causing excessive congestion (i.e., if $TFRC_{rate}$ decreases) or underutilization (i.e., if $TFRC_{rate} \geq \bar{b}_l$) of the link. In either situation, the sender transitions to normal TFRC operation.

5.3 Simulation and Results

Simulation setup. We use the *ns-2* simulator in our experiments. Our simulation setup is illustrated in Fig. 9. We assume that the streaming server always has knowledge of the vehicle location. The adaptive streaming server is connected via a high-speed wired link to the Internet. The client from a moving vehicle streams a video from the server via a HSDPA link (provider B from our measurement campaign). We assume that a streaming session lasts throughout the whole duration of a particular trip. We split our 75-trip bandwidth traces into two groups. The first group of 35 trips forms the training set, which are used as historical observations to create the bandwidth maps. After the maps are created, we assume that they are updated as each trip progresses using the new bandwidth samples. The second group of 40 trips constitutes the evaluation set. To simulate mobility, we vary the bandwidth of the HSDPA link by playing back the empirical bandwidth traces for a particular trip from the evaluation set. Since the wired links in the Internet have sufficiently high bandwidth and small

TABLE 2
Mapping between PSNR and MOS

PSNR (dB)	MOS	Impairment
> 37	5	Imperceptible
31-37	4	Perceptible, not annoying
25-31	3	Slightly annoying
20-25	2	Annoying
< 20	1	Very annoying

delays, we abstract the Internet as a 100 Mbps wired link with a small propagation delay (10 ms).

For implementing adaptive streaming, we use the *Evlavid-RA* framework [17]. In this framework, a video is pre-encoded at 31 different rates, denoted by 1-31, where 1 represents the highest quality [17]. Each of the encoded video streams is then partitioned into a sequence of small “chunks,” e.g., Group of Pictures (GoP). The streaming server dynamically varies the streaming quality and bit rate by switching between different quality GoPs from the corresponding encoded streams. We use a medium motion media sequence (“Foreman” QCIF sequence) encoded at 30 frames/second using the popular MPEG-4 codec. According to one of the recommended approaches [17], we conceal a lost frame by replacing it with the last successfully received frame.

We compare the performance of BW-MAP-TFRC with that of plain vanilla TFRC. To implement the former, we have modified the stock TFRC implementation in *ns-2* according to Fig. 10. We quantify the objective video quality using Peak Signal-to-Noise Ratio (PSNR). PSNR is calculated (in dB) for each frame, which represents the pixel-by-pixel mean square error (MSE) comparison of the decoded frame to the corresponding source frame. For example, the *MSE* between the pixels Y_s from the source frame and pixels Y_d from the destination frame is expressed as

$$MSE = 1/(m \cdot n) \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (Y_s(i, j) - Y_d(i, j))^2, \quad (9)$$

where m and n denote the number of rows and columns of pixels in a frame, respectively. Then, PSNR of the decoded frame is

$$PSNR = 20 \cdot \log_{10}(255/\sqrt{MSE}). \quad (10)$$

However, to understand the human perceived quality, we need subjective quality metrics, such as Mean Opinion Score (MOS) [30]. MOS ranges from 1 to 5, where 5 represents the best rating of user experience. An approximate estimate of MOS [31], [32] can be obtained from the PSNR using the PSNR-MOS mappings in [17] (as in Table 2).

Table 2 suggests that the picture quality is good when $PSNR > 31$ dB. It is reported that humans can perceive a drop in the streaming quality, when the PSNR remains below 31 dB consistently for 1 s or longer [33]. We refer to such an event as a *video glitch*. Clearly, reducing the occasions in which a viewer suffers from glitches directly improves the QoS of a video streaming session. We define the *glitch duration*, T_g , as the cumulative time that a user suffers from glitches during the streaming period, T_s (e.g., the time to travel through a section of the road). We use the

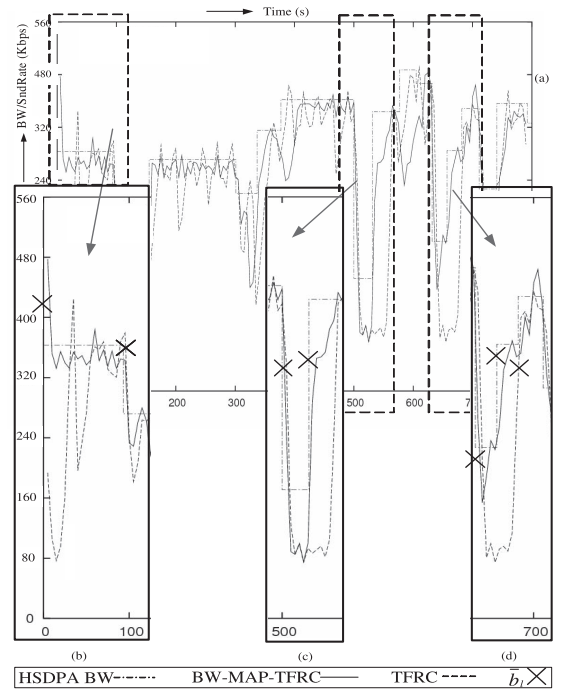


Fig. 11. Sending rate and HSDPA bandwidth dynamics for trip #19 of inbound route.

normalized glitch duration, which is defined as T_g/T_s , as a metric in the evaluations. We present mean results averaged from 40 trips. The standard error achieved is on average less than 5 percent of the mean value.

Results. Fig. 11 plots the instantaneous HSDPA bandwidth and the sending rates of TFRC and BW-MAP-TFRC as a function of time for a particular trip from the inbound route. Fig. 12 shows the corresponding instantaneous PSNR values. We are particularly interested in understanding

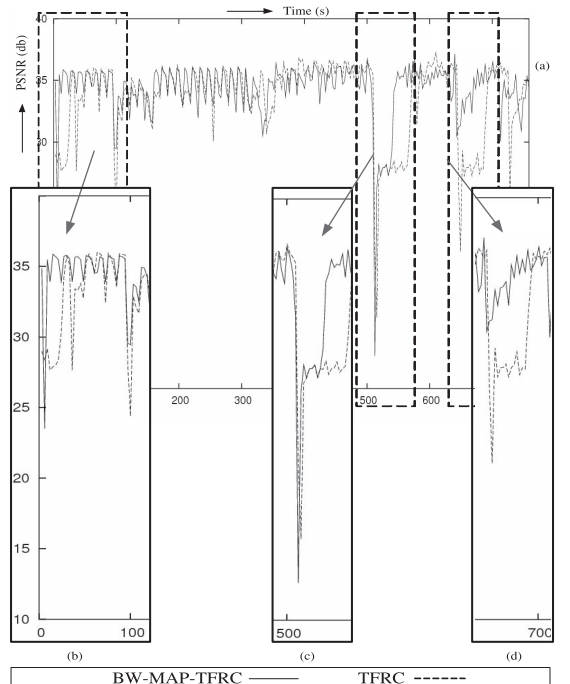


Fig. 12. PSNR dynamics for trip #19 of inbound route.

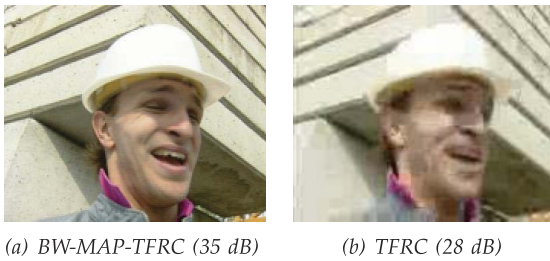


Fig. 13. Picture quality comparison for frame #581 in trip #19 of inbound route.

how these schemes behave during location transitions. Hence, we focus on three such instances in Figs. 11b, 11c, and 11d and Figs. 12b, 12c, and 12d. Note that, Figs. 11 and 12 are representative of the performance observed in other trips from both routes.

Fig. 11b focuses on the start of the trip. Since TFRC employs TCP's slow-start algorithm, it requires considerable time to ramp up its sending rate. Further, observe that the exponential increase during slow start leads to packet loss following which, the sender enters congestion avoidance. On the contrary, BW-MAP-TFRC bypasses slow start completely. The sending rate is bootstrapped at the past mean bandwidth at the initial location. As a result, the sending rate converges to the current bandwidth within 10 s. The improved convergence achieved by BW-MAP-TFRC is reflected in the PSNR results. Fig. 12b shows that the PSNR of BW-MAP-TFRC quickly converges to 35 dB and remains steady around that value. On the contrary, TFRC significantly underutilizes the bandwidth and thus misses out the opportunities to transmit a better quality stream. As a result, the picture quality is poor, wherein users suffer consecutive glitches. Fig. 13 offers a snapshot of the different viewing experience for both schemes by displaying the decoded frame #581, which is played at 20 s. Note that, the picture quality with TFRC is pixelated, whereas the frame delivered by BW-MAP-TFRC is crystal clear.

Fig. 11c captures bandwidth fluctuations due to two consecutive changes in location, at 500 and 530 s, respectively. When the bandwidth drops from 440 to 160 Kbps at 500 s, Fig. 12c shows that both schemes suffer from poor streaming quality. This is because that the instantaneous bandwidth is so low that good quality streaming is not possible. At 530 s, the bandwidth recovers. Fig. 11c shows that TFRC is conservative in increasing its rate, as it is affected by the sudden increase in packet loss and delay during the previous location transition. However, BW-MAP-TFRC does not suffer from this, since it treats each location independently and uses the past mean bandwidth as an initial estimate for the sending rate. Fig. 12c shows that BW-MAP-TFRC can avoid glitches by quickly restoring the PSNR to 35 dB at 540 s, which is nearly 20 s faster than TFRC.

Fig. 11d shows that the mobile bandwidth drops sharply from 470 to 240 Kbps at 630 s. Again, BW-MAP-TFRC is quicker in reacting to this sharp decrease as compared to TFRC. Fig. 12d illustrates that BW-MAP-TFRC is still able to maintain reasonable streaming quality. On the contrary, excessive packet loss causes TFRC to cut the rate more aggressively, thus reducing the streaming quality. When the

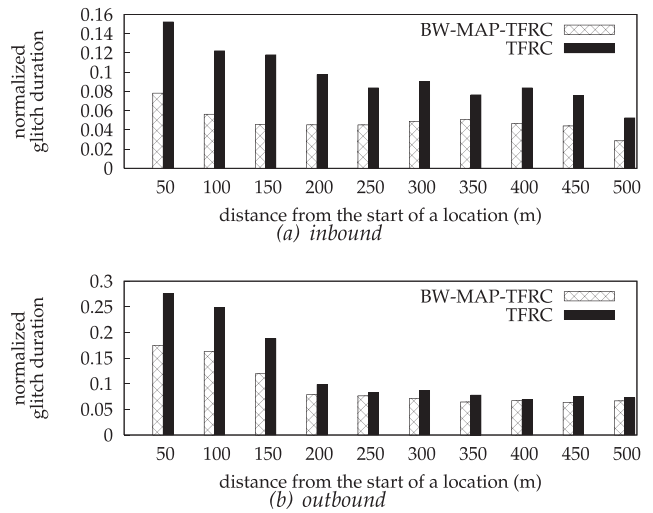


Fig. 14. Normalized glitch duration in different sections of a location.

bandwidth recovers at 660 s, BW-MAP-TFRC converges twice as fast as TFRC. Further, observe that by the time TFRC has successfully converged (680 s), the client has already moved to the next location.

As is evident from the above example, with TFRC, the video stream is more likely to suffer from poor quality during the time period immediately following a change in the vehicle's location. This is due to the purely reactive nature of TFRC, which requires considerable time to adapt to the bandwidth at the new location, particularly when the change in location results in a significant change in the wireless bandwidth. To confirm this, we partition each 500 m location segment into 10 small sections, each of which is 50 m long, and evaluate the performance of the two schemes in each section. Fig. 14 plots the normalized glitch duration for each 50 m section. (In this case, T_s is the time required to traverse the corresponding 50 m section in each trip.) Note that, the results presented are averaged over all locations and 40 trips. Fig. 14 confirms our intuition that, with TFRC, there is substantial impairment to the video quality in the first few sections of the location. The glitches taper off as the vehicle progresses toward the end of the location. Further, observe that by relying on past bandwidth statistics, BW-MAP-TFRC can significantly (by about 50 and 40 percent in inbound and outbound routes, respectively) reduce the glitches encountered during the initial sections.

Fig. 15 plots the normalized glitch duration at different locations along both routes. Note that, the results are averaged over 40 trips. At the first location of both routes, BW-MAP-TFRC effectively achieves a *threefold* reduction in the normalized glitch duration. Recall that, this is because BW-MAP-TFRC bypasses the TFRC slow-start phase at the begin of the session. Further, by relying on the PTMTP property, BW-MAP-TFRC achieves over 50 percent reduction in the glitch duration at several other locations in both routes as compared to TFRC. At certain locations, e.g., location #3 in inbound route and location #20 and #21 in outbound route, BW-MAP-TFRC outperforms TFRC by more than a factor of 5. We also observe at a few locations, e.g., location #5 in inbound route and #11 in outbound route, both TFRC and BW-MAP-TFRC achieve similar

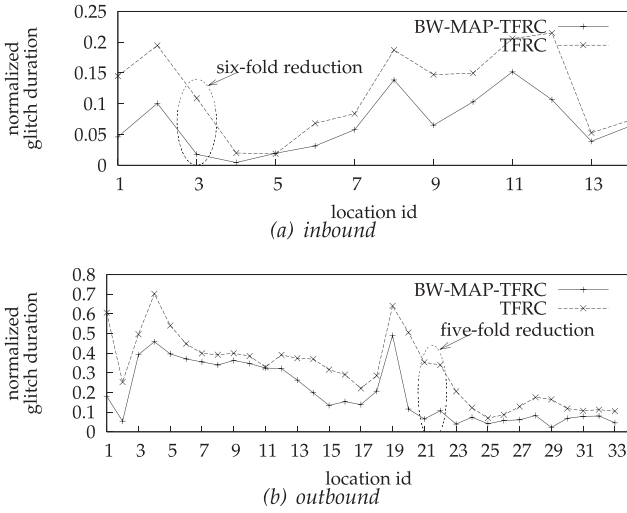


Fig. 15. Normalized glitch duration at each location.

performance. We attribute this to the fact that the bandwidth experienced at these particular locations is highly variable (i.e., the bandwidth varies significantly from one trip to another, which can be confirmed from our empirical data). As a result, the mean estimated from the past observations is not necessarily a good approximation of the current bandwidth.

6 MULTIHOMING TRAFFIC SCHEDULING

In this section, we present a second application of bandwidth maps, which enhances the QoS offered by an on-board communication network connected to the Internet via multiple WWAN links. We propose to interface bandwidth maps with the traffic scheduler of an on-board network. This allows the scheduler to use the average bandwidth from past observations at different locations in making load balancing decisions. We demonstrate the effectiveness of our technique for audio streaming using trace-driven simulations.

6.1 Background

A typical on-board communication network is illustrated in Fig. 16. Multiple user devices inside the vehicle are plugged into an on-board LAN (wireless/wired). A Mobile Router (MR) seamlessly connects the on-board LAN to the Internet using (typically) multiple WWAN services for enhanced capacity and reliability. The MR and the Home Agent (HA) transparently manage the mobility of the on-board network using the NEMO basic protocol [34]. In recent years, several such commercial systems have been deployed for providing Internet services in public transport vehicles (e.g., iComera¹) as well as personal automobiles (Autonet Mobile²).

According to the NEMO basic protocol, all traffic in the downlink direction is routed through the HA as illustrated in Fig. 16. Thus, the HA is responsible for scheduling all inbound individual traffic flows (user sessions) to the WWAN links. Similarly, MR schedules all uplink traffic

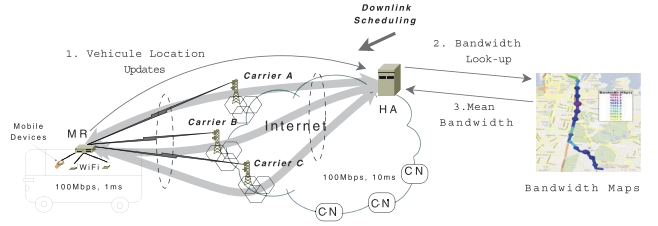


Fig. 16. Interfacing bandwidth maps with a multihomed scheduler.

flows. A scheduling decision that does not adapt to the underlying bandwidth conditions of WWAN links may lead to assigning too much traffic on some of the links while leaving the other links underutilized. In the former case, the excessive traffic can lead to congestion, thus increasing the delay and causing packet loss. In the latter, the link capacity is not used optimally. Therefore, the scheduling decisions made by traffic schedulers are crucial to the QoS offered by the on-board networks. However, since the available bandwidth of WWAN links may vary significantly from one road segment to another during the trip (confirmed by our empirical measurements presented in Section 3), it is challenging to dynamically optimize the flow allocation.

6.2 Enhancement with Bandwidth Maps

We demonstrate our idea in the context of downlink traffic, given that it makes up the bulk of the traffic in a typical on-board network. The principles described here can be readily incorporated into the uplink scheduler, housed at the MR.

As an example, we have chosen to use the Proportional Fair (PF) scheduling discipline [35] to demonstrate our idea. However, the same design philosophy can be used in conjunction with any other scheduling algorithm. We use flow-based scheduling, wherein the unit of scheduling is a flow. Thus, in the context of an on-board network, the number of traffic flows assigned to each link is in proportion to the bandwidth of the corresponding end-to-end path between the HA and MR. Since the last-hop WWAN link is known to be the bottleneck along this path, the bandwidth on the WWAN link dictates the end-to-end bandwidth. Assuming that the total number of active traffic flows in the system is λ , the number of flows λ_i assigned to link i by the PF scheduler is given by

$$\lambda_i = \frac{\hat{b}_i}{\sum_i \hat{b}_i} \times \lambda, \quad (11)$$

where \hat{b}_i represents the estimated bandwidth of link i . Thus, the rescheduling is triggered in the event that 1) the bandwidth of WWAN link changes, 2) a new flow enters the system, or 3) an existing flow departs.

Fig. 16 summarizes our idea. We use the location-specific bandwidth mean as reported by the bandwidth maps in conjunction with the downlink scheduler. When the vehicle reports a change in its location, the HA looks up the bandwidth maps for the mean bandwidth values $\bar{b}_{i,l}$ at the new location l for each WWAN link i . The scheduler then reschedules the traffic flows using the new bandwidth estimates. As a result, the total traffic allocated to link i is given by

1. <http://www.iComera.com>.
2. <http://www.autonetmobile.com>.

$$\lambda_i = \frac{\bar{b}_{i,l}}{\sum_i \bar{b}_{i,l}} \times \lambda. \quad (12)$$

The above principle can be readily incorporated in both closed-loop and open-loop schedulers:

Closed-loop scheduling. A closed-loop scheduler relies on a feedback mechanism that adapts to fluctuations in the link bandwidths. For each WWAN link, the MR passively monitors the link conditions and continually reports to the HA. From these updates, the HA estimates the link bandwidth and adjusts the traffic load assigned on the links. For example, the scheduler can detect a link is overloaded, when the estimated bandwidth of a link is less than the assigned traffic load onto it. A rescheduling event is triggered when any of the links is overloaded. Commercial on-board network providers are known to employ such schedulers [18], though the exact details are proprietary and not freely available. Despite their ability to learn about the current network conditions, such schedulers may be slow to react to the frequent WWAN bandwidth variations between successive locations in a fast-moving vehicular scenario.

Due to the PTMTP property, the information from the bandwidth maps can alleviate the above problem. When the vehicle moves to a new location, the scheduler retrieves the mean bandwidth values at the current location from the bandwidth maps for all WWAN links. Then, the scheduler intelligently initiates a rescheduling action using (12). Following this, the scheduler continues with its reactive closed-loop adaptation and makes further changes to fine-tune the flow allocation as required.

Open-loop scheduling. An open-loop scheduler does not monitor the network conditions in real time and is hence unable to adapt to changes in the link bandwidth. As an example, the scheduler may naively estimate the typical bandwidth of each link using provider advertised bandwidth statistics, e.g., the average values of the advertised bandwidth ranges from providers. Note that, these a priori estimates are location invariant. As a result, the scheduler always performs flow allocation according to these fixed bandwidth estimates at different locations. An open-loop scheduler that incorporates bandwidth maps uses the mean bandwidth values at each location from the maps in rescheduling traffic flows as per (12). As a result, it is able to adjust to the bandwidth variations and thus achieve better performance. For brevity, in this paper, we only focus on closed-loop schedulers. We refer the reader to [22] for further details about open-loop schedulers and evaluations.

6.3 Simulation and Results

Simulation setup. We conduct simulations in *ns-2*. We simulated an on-board NEMO network as illustrated in Fig. 16. The MR maintains three parallel WWAN connections, which correspond to providers A, B, and C from our empirical measurements in Section 3. We refer readers to Fig. 16 for the network parameters used in the simulation. As in Section 5, we simulate the vehicle mobility by playing back the corresponding bandwidth trace files from the three providers for a particular trip. We consider a scenario, wherein, several on-board users are streaming audio (e.g., radio talk shows and webinars) from different Internet servers, which are denoted as Corresponding Nodes (CN).

We simulate G711-encoded 64 Kbps constant bit rate (CBR) flows [36]. Flow arrival follows Poisson distribution with mean $\lambda = 10$ flows/minute. We assume the session duration is exponentially distributed with mean $\beta = 180$ s. Note that, these traffic load settings represent a scenario when the system load is close to saturation.

We refer to the plain reactive closed-loop scheduling scheme as *React*. In our implementation, the MR implements a monitoring agent that passively estimates the throughput for each WWAN link over each 2-second interval and reports it to the HA. Note that, the reported throughput may not accurately represent the actual link bandwidth. For example, when a link is underutilized, the estimate computed by the MR reflects the actual traffic load assigned to the link rather than the link bandwidth. To infer the actual bandwidth, we use a simple Multiplicative Increase Multiplicative Decrease (MIMD) algorithm. When the reported throughput of a WWAN link is lower than the assigned traffic load, the HA assumes that the link is overloaded. In this case, the HA estimates the bandwidth as half of the assigned traffic load on the link. Otherwise, when the measured bandwidth of the link equals the assigned load, the link is deemed as underutilized. In this case, the bandwidth is estimated as $1.1 \times$ assigned traffic load. The parameters for the MIMD algorithm have been chosen based on our experiments (excluded for reasons of brevity). Based on the bandwidth estimates, the HA uses (11) to schedule flows. Rescheduling is also triggered with the traffic load in the on-board network changes due to a new flow starting or an existing flow ending.

We refer to our bandwidth map supplemented closed-loop scheme as *BW-MAP*. We assume that the HA is always aware of the current vehicle's location. (The MR is equipped with a GPS and sends periodic location updates to the HA.) Recall that, as described in Section 6.2, the key difference between *BW-MAP* and *React* is that, when the vehicle enters into a new location, *BW-MAP* uses the location-specific mean bandwidth values of all WWAN links to reschedule the flows using (12).

As in Section 5.3, we use the first 35 trips to create the initial bandwidth maps for each WWAN link. The bandwidth maps are updated as the simulation progresses. The remaining 40 trips are used in the trace-driven simulations for evaluating the performance of different schemes. We present averaged results from the 40 trips. The standard error achieved is on average less than 5 percent of the mean value that is reported.

We use MOS to evaluate the human perceived audio quality. To calculate MOS, we first calculate the *R factor* using the ITU E-Model [37]:

$$R = R_o - I_s - I_d - I_{e-eff}, \quad (13)$$

where R_o is the basic signal-to-noise ratio, I_s represents impairments to the source signal, I_d is the impairment due to delay and echo effects, and I_{e-eff} represents impairments due to packet losses. For G711 codec, R_o is defined as 93.2 [38]. We assume no impairments to the source signal, i.e., $I_s = 0$. We assume that clients have sufficiently large playout buffers, which filter out the effects of delay jitter, and thus $I_d = 0$. For each active flow, we measure the

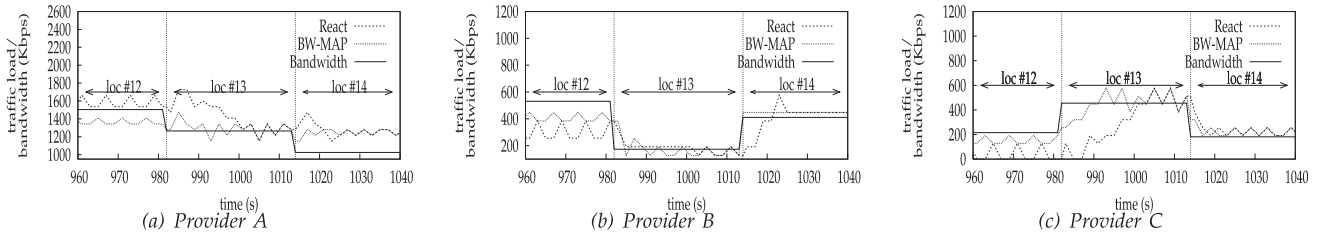


Fig. 17. A snapshot of WWAN bandwidth and aggregate network traffic load during trip #67 inbound route.

instantaneous packet loss rate (P_{pl}) and mean number of packets lost in a burst of lost packet ($BurstR$) at each second. I_{e-eff} is calculated as

$$I_{e-eff} = I_e + (95 - I_e) \cdot \frac{P_{pl}}{P_{pl}/BurstR + B_{pl}}, \quad (14)$$

where I_e is the impairment due to the codec used and B_{pl} is the robustness factor to random packet loss. $I_e = 0$ for the G711 codec and $B_{pl} = 25.1$ for the G711 codec with Packet Loss Concealment (PLC) [38]. The R factor can be converted into MOS using the following [37]:

$$MOS = \begin{cases} 1, & \text{for } R < 0, \\ 1 + 0.035 \cdot R + R \cdot (R - 60) \cdot (100 - R) \cdot 7 \cdot 10^{-6}, & \text{for } R \in [0, 100], \\ 4.5, & \text{for } R > 100. \end{cases} \quad (15)$$

We evaluate the instantaneous MOS for each active flow every second. When the MOS drops below 3, the audio quality is perceived as annoying [30]. As in Section 5.3, we refer to such an event as a *glitch*. Since there are multiple flows in the system at all time, we define *mean glitch duration*, \bar{T}_g , as the average value of the glitch duration T_g experienced by all active flows during a given streaming period T_s . Thus, the *normalized mean glitch duration* is defined as \bar{T}_g/T_s . Note that, the smaller this metric, the better the QoS.

Results. Fig. 17 plots a snapshot of the instantaneous WWAN bandwidth and the aggregate traffic load assigned to each WWAN link for one particular trip along the inbound route. Fig. 18 plots the corresponding values of the instantaneous packet loss rate and MOS, averaged over all active flows in the system. Fig. 17 shows that, in reaction to

the change in the bandwidth as the vehicle moves from location #12 to #13, both React and BW-MAP reduce the traffic load on providers A and B and reassign it to provider C. Further, observe that BW-MAP is quick to react to these changes as compared to React, which is more gradual. Fig. 18a shows that BW-MAP results in up to 70 percent lower packet loss as compared to React. This is because BW-MAP is more agile in reacting to the changes and thus avoids overloading the WWAN links. As a result, the average MOS with BW-MAP is consistently above the threshold of 3. In contrast, with React, the MOS dips below 3 after the location transition and only recovers after 997 s.

As in Section 5.3, we investigate the performance of the two schemes as the vehicle travels along different sections within the same location. Fig. 19 compares the results of normalized mean glitch duration of all active flows in each 50 m section within a location. Note that, the results are averaged over all locations from all 40 trips. As expected, the glitches are much higher in the first 50 m and reduce gradually as the vehicles progress further. However, observe that BW-MAP reduces the glitches encountered by nearly 40 percent as compared to React in the first 50 m section. The difference between the two becomes smaller as the vehicle approaches the end of the location, since React gradually learns and adapts to the present bandwidth. In the following discussion, we present results from the first 50 m segment of a location, as the effect of using bandwidth maps is most pronounced in this short instance after a change in the vehicle's location.

Fig. 20 plots the normalized mean glitch duration for the first 50 m of the locations, averaged over 40 trips. It is

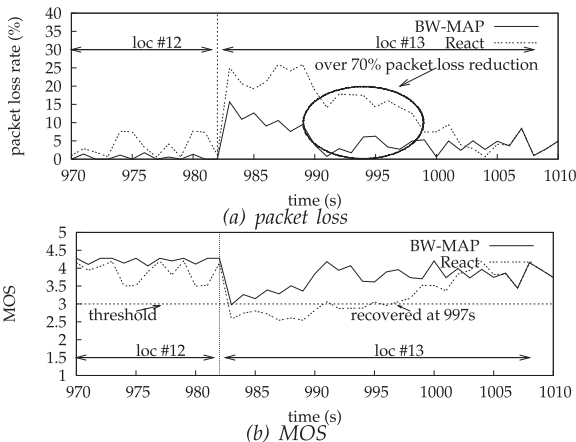


Fig. 18. Average packet loss rate and MOS as a function of time from trip #67 inbound route.

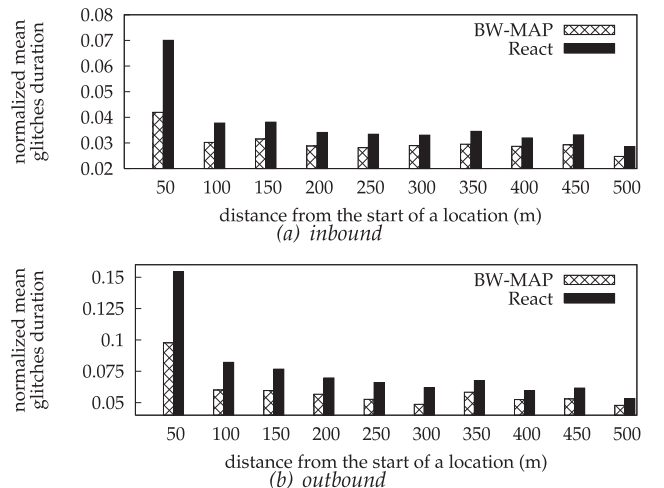


Fig. 19. Normalized glitch duration in different sections of a location.

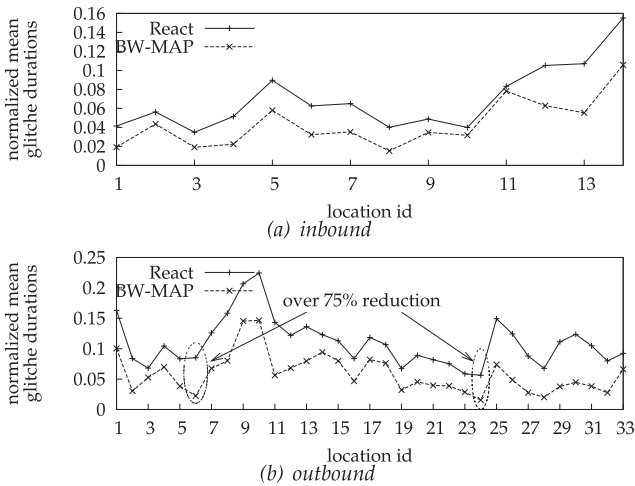


Fig. 20. Normalized mean glitch duration at individual locations.

obvious that the use of bandwidth maps consistently reduces the effect of glitches at all location. At certain locations, e.g., at location #6 and #24 in outbound route, BW-MAP reduces the glitches by up to *four* folds as compared to React. On occasion, e.g., location #10 and #11 along the inbound route, React can perform as well as BW-MAP. This is because the WWAN link bandwidths are fairly stable in that particular region (confirmed by our empirical data), which aids the purely reactive mechanisms used in React. The above results show results averaged across all flows. We are also interested to investigate how many flows are affected, since this would reflect the number of unsatisfied users in the on-board network. Fig. 21 plots the fraction of active flows that suffer glitches as a function of location and shows that BW-MAP reduces the number of flows that suffer from glitches across the board. At many locations, e.g., from location #19-#33 in the outbound route, the improvement is more than 50 percent.

7 CONCLUSION

The emerging WWANs do not guarantee bandwidth uniformity over the geographical coverage. More precisely, at any given time, it is possible to receive significantly different bandwidth from the same network provider at different locations of the same street. We have found that the past bandwidth information is a good indicator of the actual bandwidth experienced at a given location. While time of day also influences the mobile bandwidth, our analysis of empirical traces has shown that location appears to have a far greater influence than time. We have shown that it is not difficult for vehicular users to capture the past bandwidth knowledge in the form of geographical bandwidth maps for part of the road network frequently traveled. We have further demonstrated the usefulness of these maps with two representative case studies, adaptive multimedia and mobile router-based vehicular Internet. Our work has shown that the use of past location-specific bandwidth knowledge can significantly improve the QoS of multimedia streaming applications in high-speed mobility.

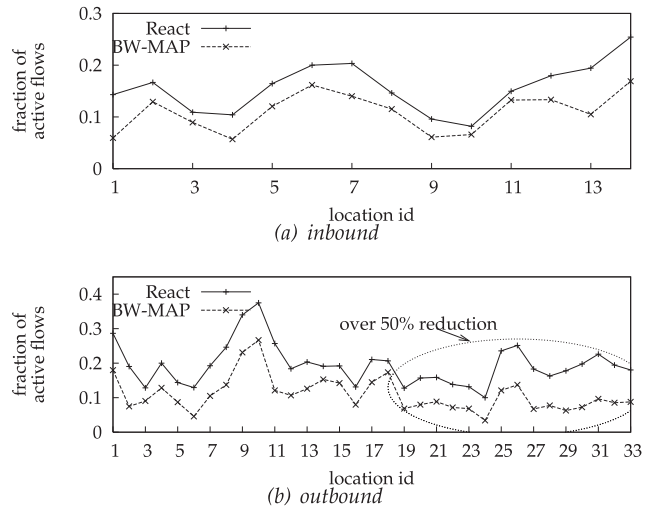


Fig. 21. Fraction of active flows suffering glitches at individual locations.

REFERENCES

- [1] J. Derksen, R. Jansen, M. Majjala, and E. Westerberg, "HSDPA Performance and Evolution," *Ericsson Rev.*, pp. 117-120, 2006.
- [2] F.P. Tso, L. Zhang, J. Tengy, W. Jia, D. Xuany, and F. Zhang, "An Empirical Evaluation on the Performance of Mobile HSPA Networks," technical report, Dept. of Computer Science, City Univ. of Hong Kong, July 2009.
- [3] K. Jang, M. Han, S. Cho, H.-K. Ryu, J. Lee, Y. Lee, and S.B. Moon, "3G and 3.5G Wireless Network Performance Measured from Moving Cars and High-Speed Trains," *Proc. ACM Workshop Mobile Internet through Cellular Networks (MICNET '09)*, Aug. 2009.
- [4] "High Speed Downlink Packet Access (HSDPA); Overall Description; Stage 2," <http://www.3gpp.org/ftp/specs/html-info/25308.htm>, 2011.
- [5] L.D. Cicco, S. Mascolo, and V. Palmisano, "Skype Video Responsiveness to Bandwidth Variations," *Proc. Int'l Workshop Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '08)*, May 2008.
- [6] J. Ormont, J. Walker, S. Banerjee, A. Sridharan, M. Seshadri, and S. Machiraju, "A City-Wide Vehicular Infrastructure for Wide-Area Wireless Experimentation," *Proc. ACM Int'l Workshop Wireless Network Testbeds, Experimental Evaluation and Characterization (WinTech '08)*, Sept. 2008.
- [7] J. Yao, S.S. Kanhere, and M. Hassan, "An Empirical Study of Bandwidth Predictability in Mobile Computing," *Proc. ACM Int'l Workshop Wireless Network Testbeds, Experimental Evaluation and Characterization (WinTech '08)*, Sept. 2008.
- [8] W.L. Tan, F. Lam, and W.C. Lau, "An Empirical Study on 3G Network Capacity and Performance," *Proc. IEEE INFOCOM*, May 2007.
- [9] TeleGeography, "Global Internet Map," http://www.telegeography.com/products/map_internet/index.php, 2011.
- [10] TPG, "TPG's DSLAM Coverage Maps," <http://www.tpg.com.au/maps>, 2011.
- [11] Root Wireless, <http://rootwireless.com/>, 2011.
- [12] A.J. Nicholson and B.D. Noble, "BreadCrumbs: Forecasting Mobile Connectivity," *Proc. ACM MobiCom*, Sept. 2008.
- [13] J. Pang, B. Greenstein, M. Kaminsky, D. Mccoy, and S. Seshan, "Wifi-Reports: Improving Wireless Network Selection with Collaboration," *Proc. ACM MobiSys*, June 2009.
- [14] Arracom Inc. "iBurst Broadband Wireless - System Overview," <http://www.arraycomm.com/docs/iBurstOverview.pdf>, 2011.
- [15] D.S. Moore and G.P. McCabe, *Introduction to the Practice of Statistics*, fifth ed. W.H. Freeman & Company, 2009.
- [16] J.A. Farrell, *The Global Positioning System and Inertial Navigation*. McGraw-Hill, 1998.
- [17] A. Lie and J. Klaue, "Evalvid-RA: Trace Driven Simulation of Rate Adaptive MPEG-4 VBR Video," *ACM/Springer Multimedia Systems J.*, vol. 14, pp. 33-50, Nov. 2007.
- [18] M. Karlsson, M. Bergek, M. Agervald, and K. Axelsson, "A System for Data Transmission via Several Communication Routes," European Patent EP1175757B1, Jan. 2002.

- [19] K. Lai and M. Baker, "Nettimer: A Tool for Measuring Bottleneck Link Bandwidth," *Proc. USENIX Symp. Internet Technologies and Systems*, Mar. 2001.
- [20] A. LaMarca, Y. Cheng, Y. Chawathe, and J. Krumm, "Accuracy Characterization for Metropolitan-Scale Wi-Fi Localization," *Proc. ACM MobiSys*, June 2005.
- [21] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating Location Predictors with Extensive Wi-Fi Mobility Data," *Proc. IEEE INFOCOM*, Mar. 2004.
- [22] J. Yao, S.S. Kanhere, and M. Hassan, "Geo-Intelligent Traffic Scheduling for Multi-Homed On-Board Networks," *Proc. Int'l Workshop Mobility in the Evolving Internet Architecture (MobiArch '09)*, June 2009.
- [23] A. Bhattacharya and S.K. Das, "LeZi-Update: An Information-Theoretic Framework for Personal Mobility Tracking in PCS Networks," *Proc. ACM MobiCom*, Aug. 1999.
- [24] S. Wenger, U. Chandra, M. Westerlund, and B. Burman, "Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF)," IETF RFC 5104, Feb. 2008.
- [25] V. Singh, J. Ott, and I. Curcio, "Rate Adaptation for Conversational 3G Video," *Proc. IEEE INFOCOM*, Apr. 2009.
- [26] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," IETF RFC 5348, Sept. 2008.
- [27] L. Xu and J. Helzer, "Media Streaming via TFRC: An Analytical Study of the Impact of TFRC on User-Perceived Media Quality," *Proc. IEEE INFOCOM*, Apr. 2006.
- [28] O. Mehani, R. Boreli, and T. Ernst, "Analysis of TFRC in Disconnected Scenarios and Performance Improvements with Freeze-DCCP," *Proc. Int'l Workshop Mobility in the Evolving Internet Architecture (MobiArch '09)*, June 2009.
- [29] J. Yao, S.S. Kanhere, and M. Hassan, "Quality Improvement of Mobile Video Using Geo-Intelligent Rate Adaptation," *Proc. IEEE Wireless Comm. and Networking Conf. (WCNC '10)*, Apr. 2010.
- [30] "Methods for Subjective Determination of Transmission Quality," ITU Recommendation P.800, Aug. 1996.
- [31] Q. Huynh-Thu and M. Ghanbari, "Scope of Validity of psnr in Image/Video Quality Assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800-801, 2008.
- [32] O. Nemethova, M. Ries, M. Zavodsky, and M. Rupp, "Psnr-Based Estimation of Subjective Time-Variant Video Quality for Mobiles," *Proc. Int'l Conf. Measurement of Audio and Video Quality in Networks (MESAQIN)*, June 2006.
- [33] Z. Orlov, "Network-Driven Adaptive Video Streaming in Wireless Environments," *Proc. IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm. (PIMRC '08)*, Sept. 2008.
- [34] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol," IETF RFC 3963, Jan. 2005.
- [35] F. Guo, J. Chen, W. Li, and T. cker Chiueh, "Experiences in Building a Multihoming Load Balancing System," *Proc. IEEE INFOCOM*, Mar. 2004.
- [36] "Pulse Code Modulation (PCM) of Voice Frequencies," ITU Recommendation G.711, Nov. 1988.
- [37] "The E-Model, A Computational Model for Use in Transmission Planning," ITU Recommendation G.107, Apr. 2009.
- [38] "Transmission Impairments Due to Speech Processing," ITU Recommendation G.113, Nov. 2007.



communication networks.

Jun Yao received the BE degree in automation from the South China University of Technology, Guangzhou, Guangdong, P.R. China, in 2002, and the MCompIT degree in inter-networking from the University of New South Wales, Sydney, Australia, in 2005. He is currently working toward the PhD degree in the School of Computer Science and Engineering at the University of New South Wales. His current research interests include mobile and vehicular



vehicular communication, mobile computing, and network security. He is a member of the ACM and a senior member of the IEEE.

Salil S. Kanhere received the BE degree in electrical engineering from the University of Mumbai, India, in 1998, and the MS and PhD degrees in electrical engineering from Drexel University, Philadelphia, in 2001 and 2003, respectively. He is currently a senior lecturer with the School of Computer Science and Engineering at the University of New South Wales, Sydney, Australia. His current research interests include wireless sensor networks,



Network, IEEE Communications Magazine, Journal of Supercomputing, and Real Time Imaging. His research interests are in mobile and wireless communication networks. He is a senior member of the IEEE.

Mahbub Hassan received the BSc degree from Middle East Technical University, Turkey, the MSc degree from the University of Victoria, Canada, and the PhD degree from Monash University, Australia. He is currently a full professor in the School of Computer Science and Engineering, University of New South Wales, Sydney, Australia. He is an area editor for the *Computer Communications* journal and previously served as a guest editor for *IEEE*

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.