High-Fidelity Application-Centric Evaluation Framework for Vehicular Networks[*]

Yi Yang, Maneesh Varshney, Shrinivas Mohan, Rajive Bagrodia
Mobile Systems Lab
Computer Science Department
University of California, Los Angeles
Los Angeles, California 90095-1596
{yangyi, maneesh, shrini, rajive}@cs.ucla.edu

ABSTRACT

This paper proposes an evaluation framework for vehicular networks to achieve accurate, scalable, flexible and repeatable performance studies through the utilization of a hybrid emulation testbed and incorporation of high fidelity protocol and environment models. The proposed framework not only addresses the unique challenges of vehicular networks but also enables new types of network analyses via the capability of conducting application-centric evaluation. Compared to traditional network-centric evaluation, case studies show scenarios where network-level statistics do not clearly discriminate between the two routing protocols while significant performance differences were observed using the application-level metrics. The paper also identifies future research directions that are enabled by the evaluation framework.

Categories and Subject Descriptors

I.6.m [**Computing Methodologies**]: Simulation and Modeling

General Terms

Experimentation, Performance

Keywords

Emulation, Application Centric, Performance Evaluation, Vehicular Networks

1. INTRODUCTION

With the rapid evolution of wireless communication technologies, an increasing number of car manufacturers are now acting to equip vehicles with on-board computing and wireless communication devices, in-car sensors and GPS systems, in preparation for the deployment of large-scale vehicular networks to the general public due to the potential improvements in safety, highway efficiency and driver convenience. Various services and applications are enabled through communication among the mobile vehicles on the road, between vehicles and fixed access points (APs) along the roadside, as well as network linkings from these entities to public transportation agencies and commercial service providers. With the continuous increase and diversification of applications and services, consequentially, vehicular networks are required to optimize performance simultaneously in the many domains that these networks are expected to support, as the quality of service requirements of various applications and services can be significantly different. For example, *traffic alert* requires very low latency and authenticated communication within a local area; while a *movie downloading* service requires high bandwidth and low jitter communication from the roadside to vehicles.

Based on the evaluation technique used, the past performance studies in the context of vehicular networks can be divided into two broad categories: measurements and simulation. Using small-scale physical testbeds, a large set of measurement studies [12, 2, 23, 5] examine the wireless communication characteristics of packet transmission among vehicles and between vehicles and roadside APs. Although physical testbeds allow real implementations to be tested, the resource investments

---

[*] This work is supported in part by U.S. Army Research Office under grant number W911NF-05-1-0246 *DAWN* and National Science Foundation under NRT grant ANI-0335302 *WHYNET*. We thank Valery Naumov et al from ETH Zurich for making the vehicle movement traces available.

VANET'07, September 10, 2007, Montréal, Québec, Canada.

needed to deploy a large scale physical testbed make it prohibitive to use physical testbeds as the tool to evaluate vehicular networks in a scalable manner. Further, it is difficult to provide repeatable experiments for a given input configuration via physical testbeds, particularly with diverse operating conditions like vehicles on a highway. In contrast, past simulation studies [22, 17, 6, 10, 1] are conducted to investigate the performance of various vehicular network protocols, mainly running at network and MAC layer. While simulation offers a flexible and scalable approach, models adopted in simulation evaluation may fail to capture network deployment and vehicle movement patterns in a very faithful fashion. Take mobility for example, the commonly-used models are directly borrowed from studies of wireless ad hoc networks, which inaccurately represent real-life vehicle movement patterns. Further, simulation tools do not provide the flexibility of executing operational softwares (e.g. real application implementations). These observations indicate that in order to perform accurate, scalable, flexible and repeatable evaluation of vehicular networks, an evaluation framework that utilizes the relative benefits of both physical testbeds and simulation is required to be in place.

In this paper, we propose an evaluation framework for vehicular networks that addresses the above issues. Specifically, we use a hybrid emulation testbed TWINE [25], which combines simulation, emulation and physical networks in an integrated testbed for evaluation of wireless networks. Using TWINE, evaluation of large scale vehicular networks can be performed in lab environment. Further, we incorporate into the framework high fidelity models of lower protocol layers and physical environments that are specific to vehicular networks (e.g. channel, mobility, deployment data), which provide realistic details as well as the flexibility to support diverse network conditions. The advantages of the proposed evaluation framework are manifested not only in the scalability and accuracy of the obtained results, but also in the types of analyses that are enabled. The latter part, referred to as *application-centric* evaluation in this paper, provides an understanding of the performance of vehicular networks at the application level, which is essential to improve the experience of end users, i.e. their satisfaction of perceived services provided by the underlying vehicular networks.

There are three key points that we hope to demonstrate in this paper.

(1) Propose an evaluation framework for vehicular networks which, by utilizing a hybrid emulation testbed, addresses the unique challenges of these networks and provides an accurate, scalable, flexible and repeatable evaluation tool.

(2) Show the utility and value of this evaluation paradigm. By using TWINE, real applications specific to vehicular networks can be executed directly such that network performance can be measured at the application layer using application-specific metrics. In this way, the extent to which the quality of service of various classes of applications (e.g. safety applications, commercial services) provided by the current generation of vehicular networks can be investigated effectively. In our study, by comparing the Peak-Signal-to-Noise-Ratio (PSNR) of a video streaming application delivered by two routing protocols (AODV and GPSR) and their respective network-level performance, we were able to show that application-level metrics are more directly related to end user experience and thus provide more reliable performance results. The study also highlights scenarios where network-level statistics including throughput, delay and jitter fail to discriminate between the two routing protocols while significant performance differences were observed using application-level metrics, i.e. order of tens of dB on PSNR improvement and a 38.3% reduction on mean square root of error achieved by AODV over GPSR.

(3) The proposed framework facilitates the investigation of cross-layer interactions across the protocol stack which includes applications. This is demonstrated in our study by examining the effects of the interaction among video streaming application, two network-layer routing protocols and a MAC-layer rate adaptation protocol. The evaluation framework enables the study of possible correlation between application-level and network-level performance, which is crucial in identifying appropriate network-level metrics that are capable of accurately reflecting application layer performance.

The rest of the paper is organized as follows. In Section 2, existing evaluation techniques and performance studies in vehicular networks are discussed. Our approach of realizing an application-centric evaluation framework with realistic vehicular network environment settings is described in Section 3. Section 4 uses a case study to demonstrate the utility and benefits of the proposed evaluation framework. The future research directions enabled by the framework are discussed in Section 5. Finally the paper is concluded in Section 6.

2. RELATED WORK

The evaluation tools employed by the past performance studies in the context of vehicular networks mainly include physical testbeds and simulators. The physical testbeds that have been set up are relatively small scale containing less than ten nodes. Among the measurement studies using these testbeds, a large set [12, 2, 23, 5] are devoted to examining the quality of wireless communication to be expected when a moving vehicle connects to a fixed roadside station using short range wireless communication technologies such as IEEE 802.11 and DSRC. These studies are constrained to a single wireless link and report statistics including scan, association and IP acquisition latency, connectivity and disconnectivity duration, AP coverage, packet loss and TCP throughput. Characteristics of vehicle-to-vehicle communication have also been studied, focusing on examining the UDP/TCP throughput as the number of hops increases in the route.

Most of the existing simulation studies in vehicular networks are performed using network simulators adopted by other types of wireless networks, examples being NS-2 [11] and Qulnet [15]. To have realistic settings, a common approach used by these studies is to input mobility traces produced by traffic simulators [3, 17] offline into the network simulator to generate movements of vehicles. These simulation studies evaluate the performance of various vehicular network protocols [22, 17, 6, 10, 1]. The metrics used in these studies include packet delivery ratio, aggregate throughput, per-flow delay and jitter. Centering on such network-level metrics, these performance evaluations target at studying the dynamics of communication, i.e. packet transmissions, and examine how such dynamics evolve over time or how they are influenced by scenario or system parameters. Though acclaimed to be applicable for vehicular network operation in general, the results obtained are agnostic of upper layer applications.

In the existing literature, measurement of application-level performance is constrained to studies that aim at designing better applications or application-layer adaptations and optimizations. In [21] and [8], rate-distortion performances of various video coding standards are compared in light of video-specific metrics including PSNR and Just Noticeable Difference (JND) [9]. Little work has been done to incorporate application-level metrics into network performance evaluation and conduct network analysis from the application-centric perspective. In few network studies, application-level metrics are introduced; however, the target applications are simple and these metrics do not involve or reflect any end user subjectivity, making the measuring of them easy to realize. For example, in [4], application latency metrics such as packet inter-reception time (IRT) and cumulative number of packet reception are defined to measure the performance of a collision warning application.

3. APPLICATION-CENTRIC EVALUATION FRAMEWORK FOR VEHICULAR NETWORKS

Our primary objective in proposing a new evaluation framework for vehicular networks is to address the unique issues of these networks such as large scale and vehicle mobility without possibly losing the benefits offered by existing evaluation techniques including physical testbeds and simulation. Given that vehicular networks are inherently large scale, the evaluation framework should be able to handle scenarios where hundreds or even thousands of vehicles exist in the network in a scalable and efficient manner. Next, the evaluation framework is expected to provide high degree of control in the sense that for a given input configuration, particularly with diverse operating conditions like vehicles on a highway, experiments can be easily repeated and results reproduced. Further, in order to conduct meaningful analysis, the settings used to investigate protocols should represent realistic environmental conditions of vehicular networks including network topology, vehicle mobility and wireless channel. Finally, the proposed framework is expected to effectively enable the investigation of vehicular network performance at the application level, by measuring metrics that are specific to target applications. Some applications, such as web browsers and FTP sessions, can be modeled with relative ease, which are already part of the current suite of network simulators. For a large set of human-in-the-loop applications, however, abstraction of these applications into models is very complicated. Examples of such applications include video streaming, IP telephony and video conferencing. In addition, application-layer metrics can be subjective to human interpretations and therefore, demand direct human involvement.

The above challenges are addressed in our framework by a three-fold approach, each of which will be described in detail in one of the later subsections.

- A hybrid emulation testbed TWINE [25] is utilized to enable evaluation of vehicular networks in an accurate, scalable, repeatable and flexible manner. TWINE allows the execution of real applications in a simulated network environment, which effectively averts the difficulty in modeling applications by using operational applications.

- Objective application-level metrics are identified and used in the performance evaluation, which not only effectively capture the overall user perceived performance but also facilitate the development of practical approaches to measure them, as compared to the commonly-used time and resources consuming perceptual evaluation.

- Realistic vehicular network environment settings are created by integrating into the evaluation framework high fidelity mobility traces and real deployment data.

3.1 Utilize Hybrid Emulation Testbed TWINE

TWINE [25] is a hybrid emulation testbed developed at UCLA Mobile Systems Lab (MSL), which combines simulation, emulation and physical networks in an integrated testbed for evaluation of wireless protocols and systems. Emulation refers to evaluation techniques where the protocol stack for an emulated host in the network contains real implementations of the higher layers while lower layers such as PHY and MAC are simulated. The integration of emulation and simulation in TWINE enables perceptual evaluation of real applications over a wide range of wireless network scenarios. TWINE is also highly scalable to permit evaluation of large-scale target networks like vehicular networks in lab environment. It contains high fidelity models of lower layers and physical environment (e.g. channel), which provide realistic details particularly at wireless physical layer as well as the flexibility to support diverse network conditions. Experiments can be repeated with the same set of controlled parameters so as to support fair comparison among different network protocols and systems.

Provided with the advantages of TWINE, our first approach is to move the vehicular network evaluation paradigm from simulation or testbeds to emulation-based evaluation. This emulation-based framework is integrated into our evaluation paradigm and tailored for vehicular network performance studies. More specifically, in order to assess application-level metrics, a client or end-host is emulated, i.e. being operational to run real applications that interface with the operating system. This client or end-host communicates with other hosts in the network using an emulated network interface. In addition to being able to run operational applications, realizing a client or end-host in emulation facilitates the capture of real time interactions between applications and lower layers. Simulation is used to model the rest of the network, the operation of which does not need to be modeled at the same level of fidelity as the emulated hosts, thus enabling evaluation of large scale vehicular networks. When prototype or implementation of a specific network protocol or system is available, a subnet of operational nodes can be created through emulation, where the actual code can be evaluated. High-fidelity channel and host mobility models are integrated as well to create realistic environment settings for the vehicular network under investigation.

3.2 Measure Objective Application-Level Metrics

Objective application-level metrics refer to the set of measures, the estimation of which does not require an end user's subjective assessment but can be directly computed from statistics collected at the application layer. In order to expedite the evaluation process and improve efficiency, objective application-level metrics that closely reflect user-perceived performance are employed in our evaluation. For example, we propose the use of PSNR as the application-level objective metric for streaming video, which is highly correlated with the subjective discernment of humans, thus serves as a good candidate of measures of the user-perceived video quality. PSNR is defined as the ratio of the maximum possible power of a video signal and the power of corrupting noise that affects the fidelity of the signal's representation in the received video. Our focus lies in developing methods so that the measurement of objective metrics such as PSNR can be effectively integrated into network performance studies. This is demonstrated in the following paragraph via an example in which PSNR for streaming video is measured. Other work-in-progress applications include IP telephony and video conferencing.

In our measurement of PSNR, VLC [19] is used as a representative application of video streaming. Two emulated hosts serve as streaming video server and client, running VLC in server and client mode respectively. Streaming video originates from the operational server, travels through a simulated multi-hop vehicular network, and is finally delivered at the client. The received video is played at the operational client and can be visually monitored to make judgement of the user-percieved video quality. Video frames can be dropped by the simulated network as well as by VLC client due to late arrival. The final displayed video is captured and stored at the client for post-processing. Both the original and final displayed video clips are divided respectively into a sequence of video frames using software VirtualDub [18] so that PSNR of each user-oberserved video frame can be computed. Each pair of the original and final displayed video frames with the same frame index are then passed as input parameters to Wavelet (a class library for wavelet transforms on images [20]) which computes the distance of the video image pair in PSNR. In this process, statistics of other application-level metrics such as root mean square of error, corrupted frames, number of corrupted intervals, average duration of corrupted intervals, corrupted sec per minute of video and dropped frames can also be easily collected.

3.3 Represent Realistic Vehicular Network Environment

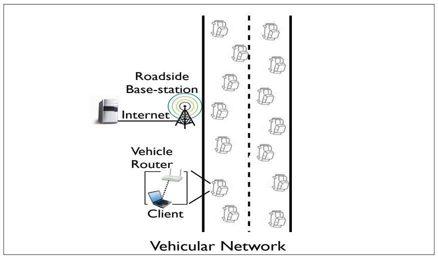


Figure 1: Vehicular network scenario

Past research has shown that the models of network environment have a significant impact on the results of wireless network analysis [17, 24]. To achieve accurate performance evaluation of vehicular networks, it therefore becomes imperative that environment settings reflect realistic scenarios. In order to create realistic environment settings of vehicular networks, high fidelity models of vehicle movements as well as real deployment data of roadside APs, are integrated into our evaluation framework to enable meaningful analysis.

To configure the target vehicular network in a realistic manner, real roadmaps of selected regions are used in our studies. Locations (e.g. GPS coordinates) of the currently deployed roadside APs are obtained through the AP locator website *www.jiwire.com* and used as the AP positions in the experiments. In order to produce close to reality vehicular movement patterns, we aim at integrating vehicular traffic simulators (e.g. CORSIM [3], TRANSIMS [16], VISSIM [14], MMTS [17]) into the network simulator Qualnet [15]. At the current stage, we have successfully incorporated vehicle mobility traces [17] into our evaluation framework. The mobility traces are obtained from a detailed vehicular movement simulation over real road maps using MMTS [17]. It contains a 24-hour movement pattern (coordinates, moving directions and speeds) of a total number of $259,978$ vehicles in Switzerland (area of $41,559km^2$). The traces are further parsed to obtain movement data for selected regions and scenarios (e.g., city, highway) and with desired vehicle density and speed.

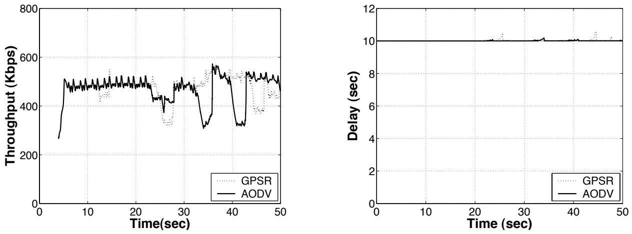
To summarize, the proposed evaluation framework extends the existing paradigms by “emulating" components of a vehicular network, in an environment that captures with high fidelity mobility and topology aspects, and provides efficient metrics to analyze the networks in their application/user-perceived performance.

4. DEMONSTRATION OF UTILITY AND BENEFITS OF APPLICATION-CENTRIC EVALUATION PARADIGM

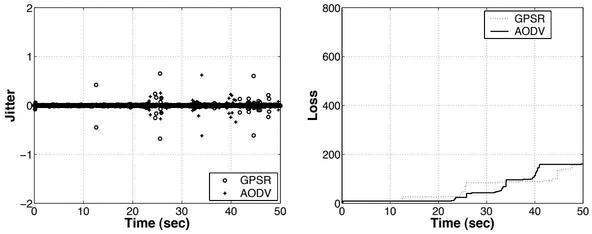
In this section the utility and benefits of the proposed application-centric evaluation paradigm are demonstrated through a case study. The experiments conducted in this study investigate the performance of a video streaming application running in a vehicular network. The network scenario has vehicles moving on a freeway which are equipped with wireless devices and thus able to communicate with each other (refer to Figure 1). Along the freeway there are APs that serve as gateways to the Internet. The vehicles can connect with these APs, possibly over multi-hop routes, to access the Internet. The application used is a media player that displays, on a client device (e.g. laptop, PDA) in a vehicle, a streaming video from a server on the Internet. A MacBook Pro is used as the client machine which runs VLC [19] streaming video client application. The streaming server runs VLC in server mode. For the routing protocol, AODV [13] and GPSR [7] are used, as they represent two large classes of ad hoc routing protocols: reactive non-geographic and geographic with greedy forwarding. Both protocols are well documented, tested in many research studies and shown to exhibit excellent performance in their respective class of routing protocols. For MAC and PHY, IEEE 802.11b is used, in which the APs and vehicles maintain a fixed transmission rate at 11Mbps, given the relatively high bandwidth requirement of video. The channel effects are modeled using two-ray pathloss model and Rayleigh fading model with varied maximum fading velocity.

At the start of each experiment, the client is located next to an AP so that it can communicate with the AP directly. As time advances, the vehicle drives away from the AP and needs to communicate through multiple hops. A default setting used to configure the experiments has 45 vehicles on the highway, moving in both directions at an average speed of 30 ± 2m/s. A single vehicle serves as the client and receives the streaming video from a roadside AP. The video is encoded at 112Kbps. Rayleigh fading is turned off in the default setting to isolate the effects of fast fading.

To effectively illustrate the advantages associated with the application-centric evaluation framework, as compared to the traditional network-centric evaluation, Figure 2 plots the network-level performance of GPSR and AODV in terms of throughput, delay, jitter and loss. It is observed that across all the metrics, the relative performance of GPSR and AODV varies over time. It is, however, hard to discriminate between GPSR and AODV in terms of the overall performance. Further, due to the lack of a thorough understanding of the correlation between application-level and network-level performance, given these results, it is not straightforward to estimate the streaming video quality to be expected by an end user.



(a) Throughput (b) Delay

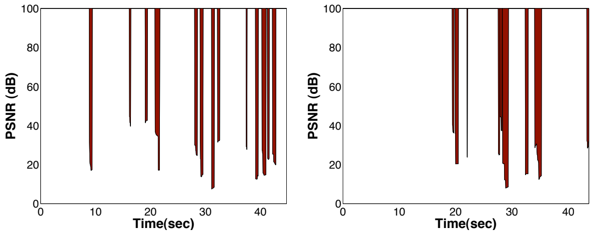


(c) Jitter (d) Loss

Figure 2: Network-level performance of GPSR and AODV in terms of throughput, delay, jitter and loss respectively in default setting

In contrast, the application-level performance of GPSR and AODV is shown in Figure 3 and Table 1. Figure 3 plots the PSNR for each video frame received at the client. The maximum value of PSNR, which corresponds to no distortion of video, is set to be 100dB. The shaded areas in the figure can be viewed as the time instances when the video is corrupted and the magnitude indicates the extent of corruption. Seen in the figure, it is obvious that overall, AODV delivers higher quality video compared to GPSR, especially towards the end of the experiment when the video is streamed through multi-hop routes from the roadside to the vehicle client. By looking at these figures, one can easily discern how the video application is performing over time.

Such temporal results can be aggregated into quantities that are listed in Table 1. Each quantity shown is an average of multiple runs of the same experiment configuration. Root mean square of error indicates the average distortion over all the received video frames. It is noted that the video delivered by GPSR has higher average distortion than AODV, which is consistent with the PSNR results. Metrics including corrupted frames, number of corrupted intervals, average duration of corrupted intervals, corrupted blocks per minute and corrupted sec per minute of video reflect the degree of distortion of the received video and the distribution of corrupted frames over time. To summarize the results on these metrics, the use of GPSR results in more corrupted frames and longer corrupted intervals. The last metric, average PSNR of corrupted intervals, represents how badly the frames are corrupted. It is interesting to see that although the overall performance of AODV is better than GPSR, in terms of corrupted frames, GPSR produces slightly higher quality.



(a) GPSR (b) AODV

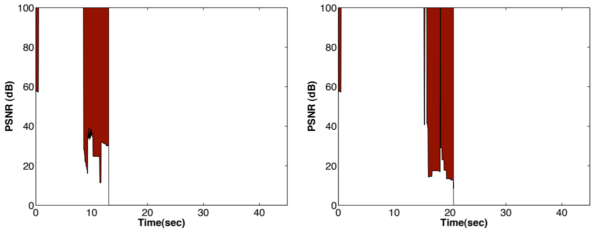
Figure 3: Application-level performance of GPSR and AODV in terms of PSNR in default setting

The results shown in the above figures and table demonstrate that application-level metrics are more directly related to end user experience and therefore facilitate the understanding of the quality of service of applications achieved by the operation of a target vehicular network. Another point illustrated by these results is that there exist cases where network-level performance represented by the commonly-used network-level metrics bears little correlation with application-level performance, i.e. such network-level metrics fail to discriminate between the two routing protocols while significant performance difference can be observed at the application layer. In general, the correlation of network-level and application-level performance relates closely to the specific nature and implementation details of the application. Depending on the application, it can be complicated to identify the appropriate set of network-level metrics that closely reflect the performance at the application layer. The potential relationship of network-level and application-level performance requires a significant amount of effort to be studied. This provides another argument to use application-level metrics in network performance analysis to expedite the evaluation process and obtain reliable results. In our case, by comparing the set of application-level metrics measured for video streaming, it is observed that PSNR constitutes the best candidate to effectively reflect the overall user-peceived video quality over time. In the rest of the paper, PSNR is used as the primary application-level metric.

To further illustrate the benefits of the evaluation framework, the impact of vehicular network environment on the application-level performance is studied. The two parameters used to vary the environment settings are wireless channel quality (i.e. maximum Rayleigh fading velocity) and vehicle density. Figure 4 plots the PSNR performance of GPSR and AODV with the maximum Rayleigh fading velocity set to 30m/s, which represents the realistic vehicle speed on a high way of fair

Application-level Metric	GPSR	AODV
Root Mean Square of Error	5.22	3.22
Corrupted Frames	18.44%	10.16%
Number of Corrupted Intervals	10.40	8.00
Avg Duration of Corrupted Intervals	0.70 sec	0.43 sec
Corrupted blocks per minute	14.02	11.13
Corrupted sec per minute of video	9.18	4.79
Avg PSNR in corrupted intervals	21.70 dB	21.55 dB

Table 1: Application-level performance of GPSR and AODV in terms of other application-layer metrics in default setting

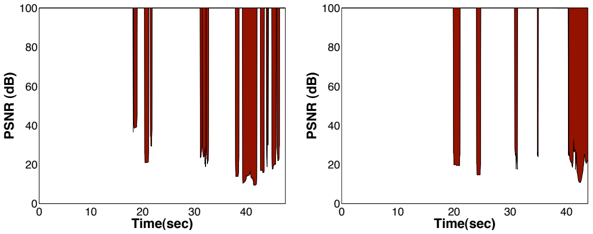


(a) GPSR (b) AODV

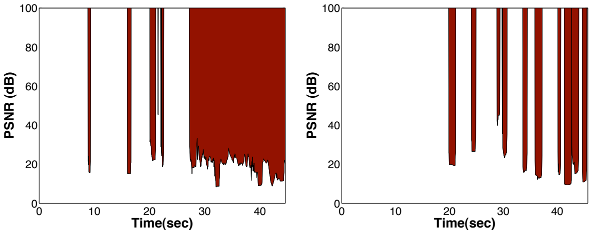
Figure 4: PSNR of GPSR and AODV with Rayleigh fading at max velocity 30m/s

vehicle density. It is seen that fading has a great impact on the performance of both protocols; neither GPSR nor AODV is capable of delivering video of fair quality when channel conditions become unfavorable. Such observation leads to the following implications. First, it is difficult to maintain high quality video when vehicles are moving very fast on the free way (as compared to a city scenario). Second, in order to provide video streaming services to vehicles on a highway, intelligent cross-layer design decisions should be made in order to efficiently respond to the changing channel.

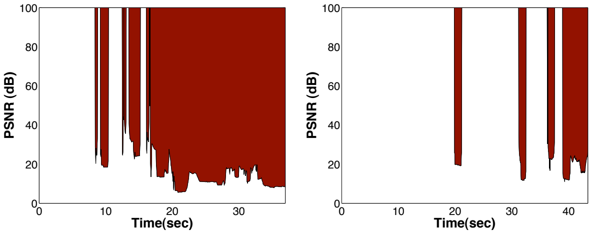
The impact of vehicle density on the delivered PSNR performance of GPSR and AODV is illustrated in Figure 5. The video is encoded at 256Kbps in this set of results. First it is noted that the performance of GPSR (see Figure 5(a), 5(c), 5(e)) deteriorates greatly when vehicle density grows. Such behavior can be explained by the fact that when there exist more vehicles in the network, the possibility of GPSR choosing an unstable long link increases. Correspondingly, the perceived video quality drops by a large extent. In contrast, it is seen that AODV exhibits more resilience to the change in vehicle density (see Figure 5(b), 5(d), 5(f)). The performance of AODV remains rather constant at various densities. These results show that the nature (positive or negative) and the extent of the impact on the application-level performance by vehicle density depend on the particular algorithm a routing protocol adopts. PSNR is effective in demonstrating



(a) GPSR, 23 vehicles (b) AODV, 23 vehicles



(c) GPSR, 45 vehicles (d) AODV, 45 vehicles



(e) GPSR, 91 vehicles (f) AODV, 91 vehicles

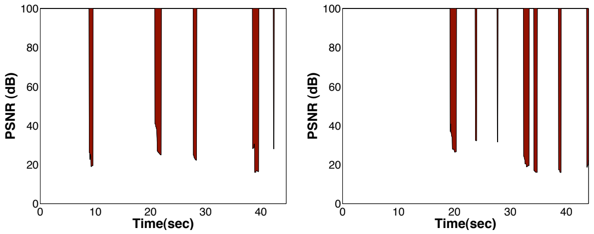
Figure 5: PSNR of GPSR and AODV at different vehicle densities with video rate 256Kbps

such impact at the application layer. In both cases, however, the rich connectivity provided by a large number of vehicles existing in the network does not help in improving the delivered video quality. This emphasizes that when designing a protocol, not only the possible impact of vehicular network environment should be taken into account but also the protocol should try to utilize the knowledge of the environment to further enhance its performance.

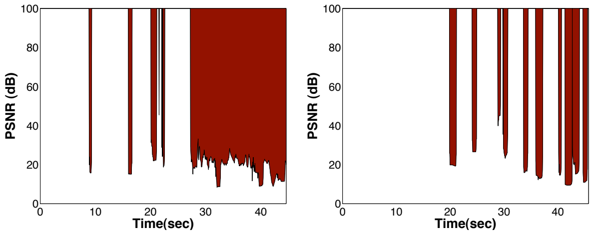
5. RESEARCH DIRECTIONS ENABLED BY APPLICATION-CENTRIC EVALUATION PARADIGM

5.1 Evaluate Application Design and Implementation in the Target Vehicular Network

As stated in the previous section, application-level performance is closely related to the implementation details of an application including application-layer adaptation, optimization, parameter configuration, etc. Through the proposed application-centric evaluation framework, the investigation of application design and implementation choices can be performed within the very context



(a) GPSR, 56Kbps video (b) AODV, 56Kbps video



(c) GPSR, 256Kbps video (d) AODV, 256Kbps video

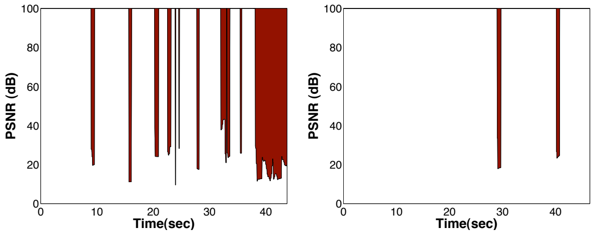
Figure 6: PSNR of GPSR and AODV at different video rates with fixed transmission rate 11Mbps

of the target vehicular network, instead of as a stand-alone process. In this section, the effects of using different video coding rates on the application-level performance of GPSR and AODV are studied. The results provide a basis to the future study of adaptive video streaming applications in a vehicular network. Figure 6 plots the PSNR performance of GPSR and AODV at two video rates, 56Kbps and 256Kbps. Together with Figure 3, it is clear that AODV outperforms GPSR in delivering higher quality video – the performance difference enlarges as the video rate increases. Such application-level performance discrepancy can be traced to the different loss performance of the two protocols (which are not shown due to space limit). *This implies that adaptation at the application layer should take into account the properties of the underlying protocols, in this case the loss behavior of a routing protocol.*

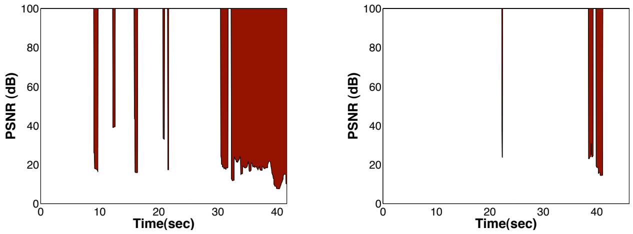
Application-level Metric	GPSR	AODV
Root Mean Square of Error	10.41	1.41
Corrupted Frames	28.85%	3.69%
Number of Corrupted Intervals	12.40	3.20
Avg Duration of Corrupted Intervals	0.40 sec	0.57 sec
Corrupted blocks per minute	17.42	4.16
Corrupted sec per minute of video	7.17	2.22
Avg PSNR in corrupted intervals	20.14 dB	19.06 dB

Table 2: Application-level performance of GPSR and AODV with ARF at video rate 112Kbps

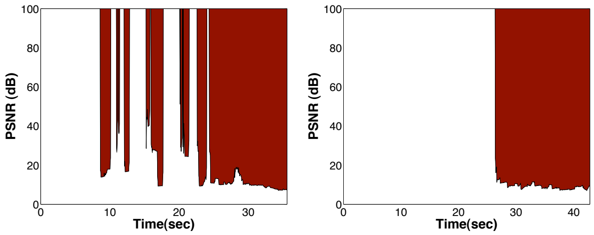
5.2 Investigate Cross-Layer Interaction across Protocol Stack Including Applications



(a) GPSR+ARF, 56Kbps (b) AODV+ARF, 56Kbps



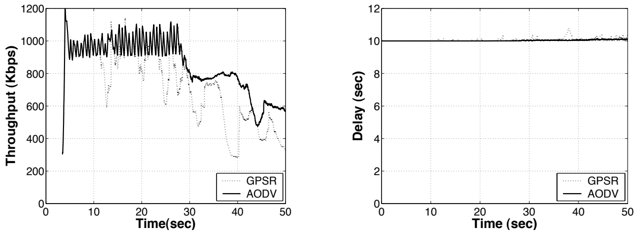
(c) GPSR+ARF, 112Kbps (d) AODV+ARF, 112Kbps



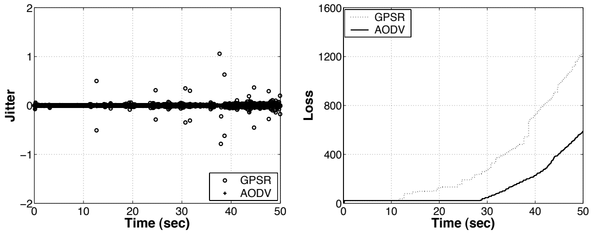
(e) GPSR+ARF, 256Kbps (f) AODV+ARF, 256Kbps

Figure 7: PSNR of GPSR and AODV with ARF at different video rates

The previous study shows the impact of network operation on the achieved user satisfaction and the need to study cross-layer interaction. These observations are further corroborated in this section by investigating the effects of the interactions among the video streaming application, the two routing protocols (GPSR and AODV) and a MAC-layer rate adaptation protocol (Auto Rate Fallback – ARF) on the user-perceived video quality. Figure 7 plots the PSNR of GPSR and AODV with ARF at three video rates. Comparing to Figure 3 and 6, it is observed that the performance of AODV is enhanced by ARF at video rates 56Kbps and 112Kbps; while at 256Kbps, the performance of AODV deteriorates largely with the existence of ARF. In contrast, the performance of GPSR drops significantly at all video rates when ARF is used. Such relative behavior is also observed on other application-level performance measures, i.e. comparing Table 2 to Table 1. These observations indicate that user-perceived video quality is the result of combined functioning of various protocols at different layers from the application down to PHY. When different protocols are used and their configurations vary, application-level performance changes. Fur-



(a) Throughput (b) Delay



(c) Jitter (d) Loss

Figure 8: Network-level performance of GPSR and AODV with ARF at video rate 256Kbps

ther, *the impact of a particular protocol on the application-level performance differs depending on the choice of protocols at other layers*. For instance, at the same video rate of 112Kbps, the performance of AODV is improved by ARF while the performance of GPSR impaired.

To understand the reason why the use of ARF is able to cause PSNR to drop, especially at high video rates, Figure 8 shows the network-level performance of GPSR and AODV with ARF at video rate 256Kbps. It can be seen that ARF improves delay and jitter, comparing to the case where a fixed transmission rate of 11Mbps is used, because ARF reduces the transmission rate when it sees packet drops. However, it is also observed that once ARF decreases the transmission rate, it stays at that lower rate for a relatively long period of time until sufficiently large number of successful transmissions are seen. This can be illustrated by the throughput performance of AODV (refer to Figure 8(a)), which stays at three levels as time advances, corresponding to the three higher transmission rates of 802.11b. Such an approach of ARF to trade throughput for reliability works effectively with low-rate videos, as the throughput requirement of the application can be well satisfied even at low transmission rates. However, when the video rate is high, using low transmission rates fails to meet the application requirement. As seen in Figure 8(d), a large number of packets are dropped. Due to high loss and low throughput, when video rate is high, e.g. 256Kbps, the application-level performance decreases with the use of ARF. These observations indicate that *a lower layer protocol should be aware of the application requirements in determining its adaptation strategy. When the application requirements change, the protocol should adjust its behavior responsively.* ARF ignores the application requirements and reduces transmission rate for higher reliability even when high throughput is demanded, resulting in its poor performance with high-rate videos.

5.3 Explore Correlation Between Application-level and Network-level Performance

Another research direction opened up by the application-centric evaluation framework is to study the correlation between application-level and network-level performance. In Figure 3, it is seen that the performance difference between AODV and GPSR becomes more prominent during the last 10 seconds of the experiment. Associating the PSNR performance with the network-level performance (see Figure 2), it is noted that as the network-level performance drops, the perceived video quality decreases correspondingly. This conforms with the general observation that in order to deliver high quality video, high throughput, low delay, low jitter and low loss are required. It is observed that with AODV, video frames are more likely to be dropped in bursts, as opposed to randomly with GPSR. Such behavior stems from the specific path selection algorithm used by the two routing protocols. In MPEG-1, a video clip consists of key frames that encode all the information of the current frame, and delta frames which only store the incremental change of the current frame from other relevant ones. When a delta frame is lost, the subsequent delta frames will apply the delta function using an incorrect frame as reference. If these frames are lost at regular intervals, such as in the case of GPSR, the errors may accumulate to show significant deviation in the received video from the original one. The above discussion indicates that with the current set of network-level metrics, it is hard to represent the correlation between application-level and network-level performance in a quantitative manner. *A new set of metrics, which examine the network-level performance in more detail than the first order, are required to be designed in order to effectively study the impact of network operation on end user experience*.

6. CONCLUSIONS

In this paper, we propose a new evaluation framework for vehicular networks, which addresses the unique challenges of these networks and the issues with the existing evaluation techniques – physical testbeds and simulation. Centering at the utilization of a hybrid emulation testbed TWINE and the incorporation of high fidelity models of vehicular network protocols and environments, the proposed evaluation framework not only provides accurate, scalable, flexible and repeatable performance studies of realistic large-scale vehicular networks but also enables application-centric evaluation, with the objective to enhance the experiences of end users. Using quantitative experiment data, the proposed evaluation framework is shown to be able to deliver more accurate and reliable results compared to the traditional network-centric evaluation. Further, shown by case studies, it opens up future research directions including but not limited to the evaluation of application design and implementation in the context of the target vehicular network environment, the study of cross-layer interaction and the investigation of correlation between application-level and network-level performance. It is our hope that the evaluation framework proposed in this paper offers a suitable platform to advance the research in vehicular networks in the fields outlined above.

7. REFERENCES

[1] M. M. Artimy, W. Robertson, and W. J. Phillips. Assignment of Dynamic Transmission Range Based on Estimation of Vehicle Density. In *VANET*, 2005.

[2] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden. A Measurement Study of Vehicular Internet Access Using In Situ Wi-Fi Networks. In *Proc. IEEE/ACM MobiCom*, 2006.

[3] CORSIM. FHWA, 2003.

[4] T. ElBatt, S. K. Goel, G. Holland, H. Krishnan, and J. Parikh. Cooperative Collision Warning Using Dedicated Short Range Wireless Communications. In *VANET*, 2006.

[5] R. Gass, J. Scott, and C. Diot. Measurements of In-Motion 802.11 Networking. In *Proc. IEEE WMCSA*, 2006.

[6] S. Jaap, M. Bechler, and L. Wolf. Evaluation of Routing Protocols for Vehicular Ad Hoc Networks in City Traffic Scenarios. In *Proceedings of the 5th International Conference on Intelligent Transportation Systems Telecommunications (ITST), Brest, France*, June 2005.

[7] B. Karp and H. T. Kung. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In *Proc. IEEE/ACM MobiCom*, 2002.

[8] P. Lambert, W. D. Neve, P. D. Neve, I. Moerman, P. Demeester, and R. V. D. Walle. Rate-Distortion Performance of H.264/AVC Compared to State-of-the-Art Video Codecs. In *IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 1*, Jan. 2006.

[9] J. Lubin. A Human Vision System Model for Objective Picture Quality Measurement. In *Sarnoff Corporation, USA*.

[10] R. Mangharam, D. S. Weller, D. Stancil, and R. Rajkumar. GrooveSim: A Topography-Accurate Simulator for Geographic Routing in Vehicular Networks. In *VANET*, 2005.

[11] S. McCanne and S. Floyd. Network simulator ns-2. *http://ww.isi.edu/nsnam/ns*.

[12] J. Ott and D. Kutscher. Drive-thru Internet: IEEE 802.11b for "Automobile" Users. In *Proc. IEEE Infocom*, 2004.

[13] C. Perkins, E. Royer, and S. Das. Ad Hoc On Demand Distance Vector (AODV) Routing. *draft-ietf-manet-aodv-10.txt, Internet Draft, IETF*, March 2002.

[14] PTV. VISSIM User Manual 4.10. 2005, PTV Planung Transport Verkehr AG: Karlsruhe, Germany.

[15] Qualnet. http://www.scalable-networks.com.

[16] TRANSIMS. http://transims.tsasa.lanl.gov/.

[17] R. B. Valery Naumov and T. Gross. An Evaluation of Inter-Vehicle Ad Hoc Networks Based on Realistic Vehicular Traces. In *Proc. ACM MobiHoc*, 2006.

[18] VirtualDub. http://www.virtualdub.org/.

[19] VLC Media Player. http://www.videolan.org/vlc.

[20] Wavelet. http://herbert.the-little-red-haired-girl.org/en/software/wavelet/.

[21] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan. Rate-Constrained Coder Control and Comparison of Video Coding Standards. In *IEEE Transactions on Circuits and Systems for Video Technology*, Jul. 2003.

[22] H. Wu, J. Lee, M. Hunter, R. Fujimoto, R. L. Guensler, and J. Ko. Simulated Vehicle-to-Vehicle Message Propagation Efficiency on Atlanta's I-75 Corridor. In *Transportation Research Record, Journal of the Transportation Research Board*, 2005.

[23] H. Wu, M. Palekar, R. Fujimoto, R. Guensler, M. Hunter, J. Lee, and J. Ko. An Empirical Study of Short Range Communications for Vehicles. In *ACM VANET*, 2005.

[24] J. Yoon, M. Liu, and B. Noble. Random Waypoint Considered Harmful. In *Proc. IEEE Infocom*, 2003.

[25] J. Zhou, Z. Ji, and R. Bagrodia. TWINE: A Hybrid Emulation Testbed for Wireless Networks and Applications. In *IEEE INFOCOM 2006*, April 23-29, 2006.