RESEARCH   **Open Access**

Resource reservation for mobile hotspots in vehicular environments with cellular/WLAN interworking

Wei Song

Abstract

Nowadays, the pervasive wireless networks enable ubiquitous high-rate wireless access from everywhere. There have been extensive studies on interworking of complementary wireless technologies in an indoor (residential or business) environment such as offices, hotels, and airport terminals. Nonetheless, there are ever-increasing demands for systematic deployment of moving networks in a vehicular environment such as public transits (e.g., a bus, train, or airplane). Due to the high mobility, it is very challenging to deliver smooth and high-quality video services for such a vehicular network. In this article, we focus on a two-hop moving network integrating both the cellular network and wireless local area network. By means of proper bandwidth reservation, we can satisfy the quality of service constraints of video applications with respect to data loss rate and packet delay. Analytical approaches are introduced to effectively estimate the achievable performance and derive the required bandwidth. To characterize video traffic, a sigmoid function is proposed to model video flows as a Markov-modulated process and fluid-flow analysis is feasible to evaluate data loss rate. At a finer packet level, the batch structure of packet arrivals is captured in the queueing analysis and packet delay is evaluated. On the other hand, for aggregate traffic multiplexed at a local gateway for the vehicular network, we use a fractional Brownian motion process to model the self-similar traffic and estimate data loss rate and packet delay. According to the performance evaluation, we can derive the required channel bandwidth for such a mobile hotspot. Numerical results are presented to demonstrate the application for bandwidth reservation, which is especially useful in case of handover.

Keywords: vehicular networks, vehicular interworking, video transmission, handoff management, resource reservation, cellular/WLAN integration, quality of service

1 Introduction

Nowadays, mobile communications and wireless networks are ushering in a new era. The pervasive wireless networks enable ubiquitous high-rate wireless access from everywhere, such as the third-generation (3G) cellular networks, IEEE 802.16 wireless metropolitan area networks/WiMAX (worldwide interoperability for microwave access), IEEE 802.20 broadband wireless access/MobileFi, IEEE 802.11 wireless local area networks (WLAN)/Wi-Fi, and IEEE 802.15 wireless personal area networks (WPAN). With well-entrenched infrastructure, the cellular networks provide ubiquitous coverage but relatively low data rates, whereas WLANs

support high data rates with cost-effective deployment over smaller geographic areas. To enable ubiquitous services, there have been intensive studies on the interworking between the complementary 3G cellular networks and WLANs by means of vertical handoff, access selection, and load balancing [1,2]. Most of the previous studies on cellular/WLAN interworking focus on a simple scenario with static WLAN deployment in an indoor environment such as offices, hotels, and airport terminals. In addition to such slow-moving or static scenarios in an indoor (residential or business) environment, there are ever-increasing demands for systematic deployment of moving networks in a vehicular environment such as public transits (e.g., a bus, train, or airplane). For mobile hotspots in a vehicular environment, it is usually not feasible to take advantage of the overlay

Correspondence: wsong@unb.ca
Faculty of Computer Science, University of New Brunswick, Fredericton, Canada

structure of the cellular/WLAN integrated network in an indoor environment. The high speed of vehicles poses more stringent constraints for fast, smooth, and reliable handoff. *Gogo In-flight Internet* by Aircell (http://www.aircell.com) is one successful story under an extremely high mobility condition, where in-flight broadband access is provided for planes flying at an altitude above 10,000 ft and at a speed of 500 miles per hour. Gogo is a ground-to-air system using the 3G cellular technology evolution-data optimized (EV-DO) integrated with 802.11 WLAN for last-hop access. However, at a high moving velocity, when the plane is crossing over ground radio towers, data rates degrade due to handoff between neighboring cells. As a result, slow playback and even halting of video streaming has been observed in previous Gogo tests [3].

It is known that mobile video will generate most of the mobile traffic growth through 2015 as predicted by Cisco [4]. The statistics collected from leading mobile operators worldwide in 2010 [5] also show video streaming accounts for 37% of mobile data usage, which is the largest part next to file sharing (30%) and Web browsing (26%). There are two major types of video services, i.e., conversational video (e.g., videoconference) and video streaming [6]. Conversational video is characterized by very stringent end-to-end latency constraint and two-way traffic with a bursty pattern due to the use of live video encoder. In contrast, video streaming usually only involves with one-way downlink traffic. Video streams can be pre-stored at application servers and allow for a pre-rolling delay of a few seconds before the start of video playback. As such, video streaming applications are more concerned with playback smoothness and the delay constraint is relaxed in a sense.

Due to high vehicular mobility, it becomes very challenging to deliver good-quality video applications for a mobile hotspot. Usually, a multi-mode wireless gateway can be deployed to connect end users in the vehicle with the backbone network. In case that a handoff is required between neighboring cells, the ongoing traffic can be multiplexed at a local gateway and handed over as a whole [7], since all end users are moving together. The multiplexing gain is thus exploited to reduce bandwidth demands. Considering the high vehicular mobility and intensive bandwidth demands of video traffic, it is essentially important to reserve bandwidth in advance for handoff traffic so as to minimize service interruptions. In order to estimate bandwidth requirement of video flows, we need to appropriately evaluate the achievable quality of service (QoS) with a given bandwidth from the two-hop vehicular network. User perceived video quality is subject to stringent QoS constraints in terms of data loss rate and packet delay.

In the literature, there has been extensive work modeling the varying data rate and frame size of video traffic [8,9]. Traditionally, video traffic can be viewed as a fluid flow and characterized with a Markov-modulated process by neglecting traffic discreteness. To capture both frame size variation and auto-correlation, we extend the Markov-modulated Gamma-based model (MMG) proposed in [8] for performance evaluation. As discussed above, because of the group mobility feature, the ongoing traffic within a vehicular network can be multiplexed at a local gateway and handed over together. Due to the long-range dependency of video traffic, the resulting aggregate traffic exhibits self-similarity and high variability over a wide range of time scales [10]. The fractional Brownian motion (FBM) process is a powerful tool to analyze self-similar data flows [11]. Based on such traffic models, we analytically evaluate data loss rate of video traffic at the fluid-flow level and average transfer delay at the packet level for a two-hop vehicular network. Accordingly, the bandwidth requirement for a backhaul connection can be estimated to satisfy the QoS constraints of video traffic.

The rest of the article is organized as follows. Section 2 gives the two-hop relay network structure and video traffic models considered in this study. In Section 3, we introduce the approaches to analyze video performance over a two-hop wireless channel. Numerical results are presented in Section 4, followed by conclusions and future study in Section 5.

2 System model
2.1 Two-hop network structure
As shown in Figure 1, we focus on a vehicular network with a two-hop relay structure, which is supported by the cellular/WLAN interworking. In a vehicular environment within mobile ambulances or public transits, user

Figure 1 Two-hop relay structure for vehicular moving networks.
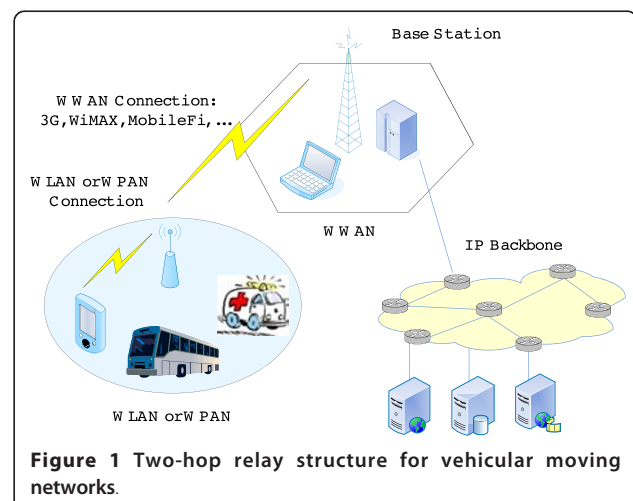
traffic can be first delivered and multiplexed at a local multi-mode WLAN router, which further relays the aggregate traffic through a cellular backhaul connection. As such, an end user in the vehicular network is served by a two-hop wireless relay channel via heterogeneous technologies. The wide-area coverage of cellular networks provides continuous access via seamless handoff. During a handoff between neighboring cells, the ongoing traffic from the moving network can be buffered at the WLAN router and handed over as a whole. It is known that the knowledge of moving patterns can assist handoff process. In the scenario of Figure 1, it is worth noting that the routes of public transits are usually predefined and known *a priori*. Taking advantage of this feature, we can reserve resources in advance for the group handoff traffic within the vehicular network.

2.2 Video traffic model

It is known that video traffic is inherently long-range dependent and highly correlated due to compression coding. In the 3G cellular networks, H.264 Advanced Video Coding (AVC) is recommended for high-quality video [12]. To remove temporal redundancy, intracoded (I) frames are interleaved with predicted (P) frames and bidirectionally coded (B) frames. I frames are compressed versions of raw frames independent of other frames, whereas P frames only refer preceding I/P frames and B frames can refer both preceding and succeeding frames. A sequence of video frames from a given I frame up to the next I frame comprise a group of pictures (GoP). Because P and B frames are encoded with reference to preceding and/or succeeding I/P frames, the transmission traffic follows the batch arrivals shown in Figure 2. Here, 3 B frames are coded between two key I/P frames and the GoP follows a structure of size 16 such as "$I_0 P_4 B_1 B_2 B_3 P_8 B_5 B_6 B_7 P_{12} B_9 B_{10} B_{11} I_{16} B_{13} B_{14} B_{15} ...$" In

contrast, video frames are decoded and displayed at the receiver in a reorganized order.

In the literature, there has been extensive work modeling the varying rate and frame size of video traffic [8,9]. The Markov chain-based models enable tractable queue analysis with the well-established fluid-flow analytical framework [13]. A MMG framework is proposed in [8] to model highly correlated video frame size. For a video flow consisting of frame bursts, video clips are grouped into a small number of shot classes depending on the burst size. As shown in Figure 3, state transitions between video shot classes are characterized by a Markov process. The size of I, P, and B frames for each shot class is modeled by an axis-shifted Gamma distribution, whose probability density function is given by

$$f_G(x) = \frac{x^{\sigma_k - 1} e^{-x/\theta_k}}{\Gamma(\sigma_k)\theta_k^{\sigma_k}}, \qquad \sigma_k > 0, \ \theta_k > 0, \ k = 1, 2, \ldots, K \quad (1)$$

where K is the number of video classes, $\Gamma(\cdot)$ is the Gamma function, and σ_k and θ_k are the shape parameter and scale parameter, respectively, for shot class k.

In the original MMG model, the GOP size boundaries for classification are geometrically separated. As observed in [14], the size of video frames based on H.264 exhibits heavy-tailed property. That is, extremely large frames exist with a non-negligible probability. The experimental results in [15] show that the boundaries between shot classes should be set appropriately so that the video statistics are captured accurately and video frames are balanced among the classes. To discern differences of large-size video clips in classification, we propose to use the following sigmoid function to determine the class boundaries:

$$s_k = \frac{1}{1 + e^{-\alpha(k - \beta)}}, \qquad k = 1, 2, \ldots, K + 1. \quad (2)$$

As this sigmoid function takes values within (0,1), we map the total size of video frames in a burst to the range $[S_{min}, S_{max}]$ such that

$$s_1 = \frac{S_{min}}{\delta \cdot S_{max}}, \qquad s_{K+1} = \frac{1}{\delta} \quad (3)$$
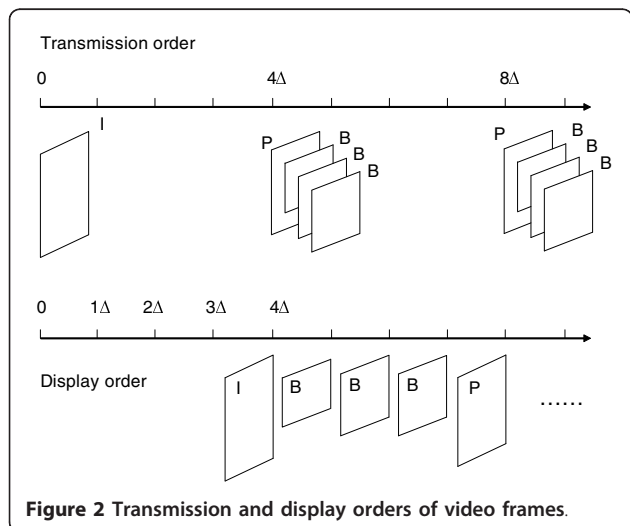
where δ ($0 < \delta < 1$) is a scale factor.

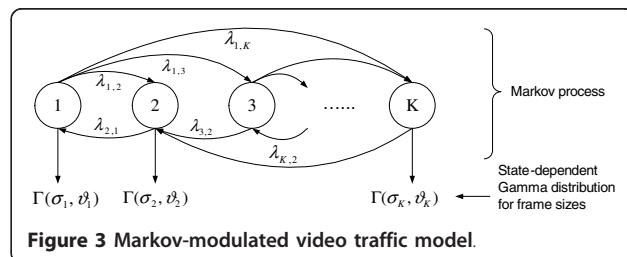**Figure 2 Transmission and display orders of video frames**.

Figure 3 Markov-modulated video traffic model.

As an example, we take a video sequence *Tokyo Olympics* coded with single layer H.264/AVC from the video trace library of Arizona State University [16]. The video trace has a common intermediate format (CIF) resolution (352 × 288), a fixed frame rate at 30 frames/s, a GoP size of 16 with 3 B frames between I/P key pictures, and a quantization step size indexed at 24. Figure 4 shows the size boundaries to classify video clips following a geometric function or a sigmoid function, respectively. As seen, the S-shaped size boundaries can also differentiate differences when the GOP size is very large.

According to the size boundaries, video clips are classified into shot classes. The transition probability p_{ij} from class i to class j can be estimated from the normalized relative frequency of transitions:

$$p_{ij} = f_{ij}/f_i \qquad (4)$$

where f_{ij} is the total number of transitions from state i to j and f_i is the total number of transitions out of state i. The resulting $K \times K$ matrix of transition probabilities, denoted by **P**, can be translated into a corresponding infinitesimal generating matrix in a continuous-time domain, denoted by **M**, as follows

$$\mathbf{M} = g(\mathbf{P} - \mathbf{I}) \qquad (5)$$

where g is the arrival rate of video bursts and **I** is the identity matrix. Let J denote the number of B frames between two key I/P pictures and f the constant frame rate. Then, the video burst rate $g = f/(J + 1)$.

2.3 FBM model for aggregate traffic

As shown in Figure 1, traffic flows from a vehicular network can be first multiplexed at a local gateway and then relayed to a remote cellular base station. Due to the long-range dependency of video flows, the aggregate traffic exhibits self-similarity and high variability over a
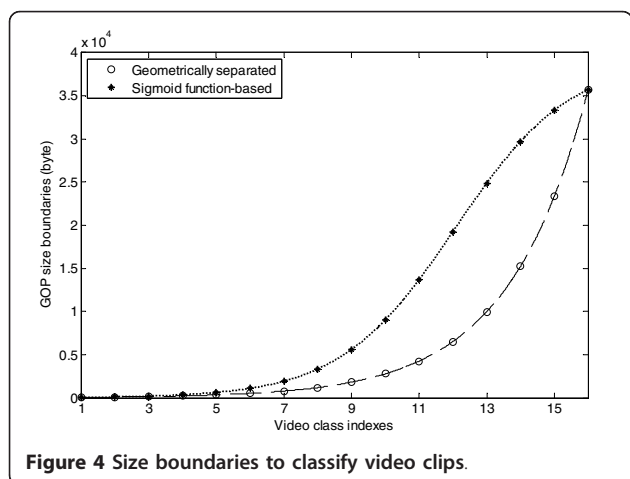

Figure 4 Size boundaries to classify video clips.

wide range of time scales [10]. In this study, we use the well-known FBM model [11] to characterize the self-similar data traffic. In particular, the cumulative arrival process of an aggregate traffic flow, denoted by $\Lambda(t)$, can be approximated by an FBM process as follows

$$\Lambda(t) = mt + \sqrt{am}Z(t) \qquad (6)$$

where m is the mean data arrival rate (the average number of data units arrived in a time interval), a is the variance coefficient (the ratio of variance to mean of data arrivals), and $Z(t)$ is a *normalized* FBM process with Hurst parameter $H \in [0.5, 1)$. The normalized FBM process $Z(t)$ processes the following properties [17]: (1) $Z(t)$ is a Gaussian process with stationary and ergodic increments; (2) $Z(0) = 0$ and $E[Z(t)] = 0$ for all t; (3) $Z(t)$ has continuous paths; and (4) $\text{Var}[Z(t)] = t^{2H}$.

3 Resource reservation for aggregate handoff traffic

3.1 Video quality indicators

As specified in [18], video quality is indicated by two components, namely, the loss rate and delay factor. Video distortion is a most direct QoS metric from the end user's perspective. It depends on data loss rate resulting from channel errors and encoding parameters that provide error resilience features. According to the R-D model proposed in [19], video distortion due to channel errors is defined as

$$D = \left(\frac{a}{1 - b + \psi b} \right) \left(\frac{PL}{1 - PL} \right) E[\mathcal{F}(k, k - 1)] \qquad (7)$$

where PL is the data loss rate, b is a parameter that describes the motion randomness of the video scene, a is the energy loss ration of the encoder filter, ψ is the percentage of intra-coded frames, and $E[\mathcal{F}(k, k - 1)]$ is the average value of the successive frame difference $\mathcal{F}(k, k - 1)$. As seen from (7), data loss rate is an important factor determining perceived video quality.

In Figure 1, a two-hop wireless relay via heterogeneous technologies is applied for a vehicular network to facilitate ubiquitous access, alleviate power constraint, and enhance transmission rate. An end user within a vehicle is served by both a local wireless router connected with a remote cellular radio tower. In contrast to wired networks, wireless channel is highly time-varying and supports a relatively low data rate. Hence, the performance over the two-hop wireless channel contributes significantly to the end-to-end QoS. Hence, video distortion perceived at the receiver is closely related to the data loss rate over the last-hop WLAN, denoted by PL_w, and that of the cellular relay channel, denoted by PL_c.

Moreover, because video frames are captured and coded in constant intervals, there is a deadline to play

back a designated video frame at the receiver end. If a frame to play has not been completely delivered to the buffer at fetch time, the playback is interrupted. Hence, the transfer delay is an important factor to ensure playback smoothness and continuity. The mean wireless access delay (denoted by T) should be upper-bounded by Δ, i.e.,

$$T = T_w + T_c \leq \Delta \qquad (8)$$

where T_w is the mean transfer delay over the last-hop WLAN and T_c is that between the WLAN router and the cellular base station. For 802.11 WLAN, the wireless channel is shared in a contention-based access manner. It is quite limited in QoS provisioning. On the other hand, in the cellular network, reservation-based resource allocation is enabled with a centralized infrastructure. As such, the aggregate traffic of a mobile hotspot can be delivered with dedicated cellular channels. The mean transfer delay is thus dependent on the reserved bandwidth of the cellular channel.

3.2 Fluid-flow data loss rate

As shown in Figure 3, an individual video flow within the mobile hotspot is modelled by a continuous-time Markov-modulated process, whose infinitesimal generating matrix M can be obtained from video traces according to a sigmoid classification function. The data rate for each shot class k is denoted by R_k (bps), which equals the average of total frame sizes of video bursts in the state over the burst duration $(1/g)$. Without loss of generality, the shot classes (states) are arranged in an ascending order with respect to date rates R_k. Viewing the video stream as a fluid flow, we can derive the equilibrium queue length distribution at the local wireless router. Let $F_k(x)$ denote the stationary probability that the buffer occupancy is less than or equal to x, given that the video source is in state k.

It is known that channel access in a WLAN follows a contention-based random access protocol, i.e., the carrier sensing multiple access with collision avoidance (CSMA/CA) algorithm and binary exponential backoff. As shown in [20], there exists an optimal operating point for the WLAN in the unsaturated case, beyond which the packet delay increases dramatically and the throughput drops quickly. When a WLAN operates in the unsaturated range, packet collision probability is quite small and each packet sees an approximately constant service rate. Assuming a constant channel service rate C_L (bps) for the last-hop WLAN and a Markov-modulated fluid process for the video flow, we can derive the equilibrium queue length distribution following the fluid-flow analytical approach [13]. That is,

$$\frac{d\overrightarrow{F}(x)}{dx}\mathbf{B} = \overrightarrow{F}(x)\mathbf{M} \qquad (9)$$

where $\overrightarrow{F}(x) = [F_1(x), F_2(x), \ldots, F_K(x)]$, **B** is a diagonal matrix defined by $\mathbf{B} = \mathbf{R} - C_L\mathbf{I}$, $\mathbf{R} = \mathrm{diag}[R_1, R_2, \ldots, R_k]$, and **M** is the infinitesimal generating matrix of the underlying Markov process.

Alternatively,

$$\frac{d\overrightarrow{F}(x)}{dx} = \overrightarrow{F}(x)\mathbf{M}', \qquad \mathbf{M}' = \mathbf{M}\mathbf{B}^{-1}. \qquad (10)$$

Based on the above set of first-order linear differential equations, we can obtain the solution to $\overrightarrow{F}(x)$:

$$\overrightarrow{F}(x) = \overrightarrow{\pi} + \sum_{k:\mathrm{Re}[z_k]<0} a_k \overrightarrow{\Phi}_k e^{z_k x} \qquad (11)$$

where $\overrightarrow{\pi}$ is the steady-state distribution of the continues-time Markov chain and satisfies $\overrightarrow{\pi}\mathbf{M} = \overrightarrow{0}$, z_k and $\overrightarrow{\Phi}_k$ are the *kth negative* eigenvalue and the corresponding eigenvector of $\mathbf{M}\mathbf{B}^{-1}$, that is

$$z_k\overrightarrow{\Phi}_k = \overrightarrow{\Phi}_k\mathbf{M}\mathbf{B}^{-1}. \qquad (12)$$

It is worth mentioning that $\overrightarrow{\Phi}_k$ are the *row eigenvectors* of $\mathbf{M}'$, which are equivalent to the transpose of the *column eigenvectors* of $(\mathbf{M}')^T$, i.e., the transpose of $\mathbf{M}'$. The coefficients a_k can be obtained from the boundary conditions, i.e., $F_k(0) = 0$ for $R_k > C_L$.

The cumulative distribution function (CDF) of the queue length is then

$$F(x) = \sum_{k=1}^{K} F_k(x). \qquad (13)$$

Given certain delay constraint, the queue occupancy is required to be less than a threshold Ψ_w, beyond which newly arrived data are very likely to be dropped at the receiver due to expiration for playback. Such data loss due to an intolerable long delay directly affects video distortion perceived by end users. Based on (13), we can estimate data loss rate PL_w for the last-hop WLAN caused by backlogging:

$$PL_w \approx 1 - F(\Psi_w). \qquad (14)$$

When a large number of data flows are multiplexed at the local gateway and forwarded toward their final destinations, the aggregate flow presents self-similar properties. The cumulative arrival process of aggregate traffic can be captured by a FBM process given in (6). Based on the analytical method in [21], we can evaluate the tail distribution of the queue length (Q_A) over the second-hop relay channel, given by

$$G(x) \triangleq P[Q_A > x] = \left[2\pi\left(1 + \sqrt{\mathcal{H}}x^{1-H}\right)^2\right]^{-\frac{1}{4}} \exp\left(-\frac{1}{2}\mathcal{H}x^{2-2H}\right) \quad (15)$$

where

$$\mathcal{H} = \frac{[(C_A - m)(1 - H)]^{2H}}{am(1 - H)^2 H^{2H}} \quad (16)$$

and C_A (bps) is the transmission rate of cellular relay channel for aggregate traffic and m is the mean arrival rate of the aggregate traffic flow per second. Similar to (14), data loss rate PL_c for the second-hop relay channel can be estimated from (15) as

$$PL_c \approx G(\Psi_c). \quad (17)$$

3.3 Packet delay performance

As shown in the R-D model in (7) for video distortion, data loss rate directly affects user perceived video quality. Packet delay is another important factor, which not only determines playback smoothness but also implicitly influences video distortion as expired packets violating delay constraints might be dropped. Assume that video frames are segmented into packets of fixed-size L for transmission. In (6), the self-similar aggregate traffic flow is modelled by an FBM process. Letting the data unit changed from bits to *packets*, we have the channel transmission rate (in *packets per second*) given by $\tilde{C}_A = C_A/L$. The mean and variance coefficient of data arrivals become $\tilde{m} = m/L$ and $\tilde{a} = a/L$, respectively. As empirically observed in [21], packet size has a relatively slight effect on the self-similarity of traffic flows. At the packet level, the traffic flow is still self-similar and the same Hurst parameter is preserved. Hence, similar to the derivation in (15), we can evaluate the mean packet delay over the second-hop relay channel as follows:

$$T_c = \frac{L}{C_A} + \frac{1}{\tilde{m}} \int_0^\infty P[\tilde{Q}_A > x]dx. \quad (18)$$

On the other hand, it is more complicated to evaluate the packet transfer delay (T_w) over the hop between end users and the local gateway. As shown in Figure 2, video frames actually arrive in bursts due to forward, backward, or bidirectional prediction in video coding and compression. If video frames are fragmented and encapsulated into fixed-size transmission packets, a random number of packets are actually generated as a batch for each video burst. Although video traffic correlation can be modelled by a Markov process described in Figure 3, it is observed in [14] that the use of a theoretical and independent distribution for frame sizes approximates the trace behavior fairly well and gives close performance statistics. The key is to preserve the batch

structure of packet arrivals and fixed inter-arrival time. As the sizes of video frame bursts for different shot classes are modelled by axis-shifted Gamma distributions, we approximate the number of packets in a "packet train" with a negative binomial distribution NB (r, p), which is a discrete analog of Gamma distribution. The probability mass function (PMF) of the number of packets in a batch (denoted by A) is given by

$$P[A = k] \triangleq f_A(k) = \binom{k+r-1}{r-1}(1-p)^r p^k, \quad r > 0, \quad 0 < p < 1, \quad k = 0, 1, \dots \quad (19)$$

where the binomial coefficient

$$\binom{k+r-1}{r-1} = \frac{(k+r-1)(k+r-2)\dots(r)}{k!}. \quad (20)$$

The parameters r and p can be obtained by fitting the mean and variance of the batch size:

$$\overline{A} = \frac{rp}{1-p}, \qquad \sigma_A^2 = \frac{rp}{(1-p)^2}. \quad (21)$$

To analyze the packet delay over the last-hop WLAN, we use a small time unit τ to discretize the time scale. The inter-arrival time of video bursts is then $N = 1/(g \cdot \tau)$ time units, where g is the video burst rate. Given fixed packet size L, the packet transmission time over the WLAN channel is assumed to be h time units, where $h = L/(C_L \cdot \tau)$. With the aforementioned batch structure and fixed burst inter-arrival time, the average delay experienced by a tagged packet in a video batch (T_w) consists of three independent components: (1) the waiting time of the first packet of that batch to be served, denoted by W_G; (2) the waiting time due to the transmission of the packets of that batch queued before the tagged packet, denoted by W_R; and (3) the transmission time of the tagged packet, which is assumed to be deterministic and equal to h time units.

To evaluate the component W_G with a queueing system model, a batch can be viewed as a single customer whose service time is the total transmission time of all packets in a batch. An analytical approach is introduced in [22] for the waiting time of a $D/G/1$ queue, whose inter-arrival time is deterministic and service time follows a general distribution. Let G denote the batch service time and $G(z)$ the corresponding probability generating function (PGF). The average waiting time for W_G is derived by

$$\overline{W}_G = -\frac{N(N-1) - G''(1^-)}{2[N - G'(1^-)]} + \sum_{k=1}^{N-1} \frac{1}{1 - z_k} \quad (22)$$

where N is the inter-arrival time of video batches, z_1, ..., z_{N-1} are the unique roots of $z^N - G(z) = 0$ within the

unit circle $| z | < 1$, and $G'(1^-)$ and $G''(1^-)$ are the first-order and second-order derivatives of $G(z)$ at $z = 1^-$.

The component W_R is the average waiting time due to the transmission of other packets prior to the tagged one in a batch. Obviously, it depends on the statistics of the number of packets queued before a tagged packet. According to the analysis in [23], the PGF of the probabilities that n packets in the current batch of a tagged packet are to be served prior to the tagged one is given by

$$Y(z) = \sum_{n=1}^{\infty} z^n \sum_{k=n+1}^{\infty} \frac{k \cdot P[A = k]}{\overline{A}} \cdot \frac{1}{k} = \frac{1 - A(z)}{\overline{A}(1 - z)} \qquad (23)$$

where $A(z)$ is the PGF of the number of packets in each non-zero sized video batch according to (19), given by

$$A(z) = \frac{\left[\frac{p}{1 - (1 - p)z} \right]^r - (1 - p)^r}{1 - (1 - p)^r}. \qquad (24)$$

Since the packet transmission time is assumed to be deterministic with h time units, the PGF of the waiting time W_R due to the transmission of the packets in a batch before a tagged one is then

$$W_R(z) = Y(z^h). \qquad (25)$$

Therefore, the average waiting time W_R are obtained as

$$\overline{W}_R = \lim_{z \to 1^-} \frac{dW_R(z)}{dz}. \qquad (26)$$

The overall packet delay is then given by

$$T_w = h + \overline{W}_R + \overline{W}_G. \qquad (27)$$

Based on the evaluation of data loss rate in Section 3.2 and the delay analysis this section, we can see that the achievable video quality depends on the available bandwidth within the last-hop WLAN and that of the cellular relay channel. To satisfy the QoS constraints for video transmission, we can derive the required bandwidth to be reserved for video flows within the vehicular network.

4 Numerical results and discussion

In this section, we present numerical examples to illustrate the use of our analysis for bandwidth reservation. Table 1 gives the system parameters for numerical experiments.

4.1 Effective bandwidth for video flows

As discussed in Section 2.2, we extend MMG to characterize video traffic as fluid flows. A sigmoid function is applied to determine shot class boundaries. In this section, we give numerical examples for traffic modelling based on a Markov-modulated process. Here, we consider H.264/AVC video sequences of *Tokyo Olympics* from the video trace library [16]. These video sequences have a CIF resolution, a fixed frame rate at 30 frames/s, a GoP size of 16, and 7 B frames between two I/P key pictures. The quantization level varies with the step size and a higher quantization index (between 0 and 51) results in a lower encoding bit rate. Taking the number

Table 1 System parameters for numerical analysis

Symbol	Value	Definition
α	1.3484	Parameter of sigmoid function for video classification
β	5.2924	Parameter of sigmoid function for video classification
δ	1.1	Scale factor to segment video flows
f	30	Video frame rate (/s)
g	3.75	Video frame burst rate (/s)
J	7	Number of B frames between two key I/P frames
K	6	Number of video classes for traffic modelling
a	175.38	Variance coefficient of aggregate traffic (kbit/s)
$\bar{a}$	29.23	Variance coefficient of aggregate traffic (packets/s)
H	0.7	Hurst parameter
C_L	~3.0	One-way channel rate of last-hop WLAN channel (Mbit/s)
τ	0.006	Time unit (s)
L	6000	Transmission packet size (bits)
p	0.4424	Parameter of negative binomial distribution for video batch size A
r	2.5812	Parameter of negative binomial distribution for video batch size A
σ	1.4847	Shape parameter of Gamma distribution for video burst size G (kbits)
θ	27.56	Scale parameter of Gamma distribution for video burst size G (kbits)

of video classes $K = 6$, we use the classification function defined in (2) to determine the classes (states) of video clips. Table 2 gives the approximate matrix of transition probabilities **P** for the video sequence at the quantization level 34. To bound data loss rate PL_w at the flow level, the effective bandwidth can be derived from the analysis in Section 3.2. Figure 5 demonstrates the effective bandwidth requirements of video flows at different quantization levels when data loss rate is upper bounded by 0.01. For instance, at the quantization level 34, the effective bandwidth requirement of the video flow is 634.8 kbit/s.

4.2 Bandwidth reservation to ensure fluid-flow and packet-level QoS

As introduced in Section 3.1, video distortion is one of the most direct quality indicators for video applications. It depends on various factors including encoding parameters and transmission performance such as data loss rate. Packet delay affects not only playout smoothness but also video distortion level due to dropping of expired packets. In Sections 3.2 and 3.3, we have presented the analytical approaches we can leverage to estimate the achievable performance int terms of data loss rate and packet delay. In this section, we will give some numerical examples to demonstrate the use of the above analysis to derive required channel bandwidth. It is especially useful for the the handover of a mobile hotspot between neighboring cells.

Consider an H.264/AVC video sequence of *Tokyo Olympics* at the quantization level 34. The average bit rate of an individual video flow is 153.5 kbit/s. Suppose the effective one-way channel rate of the WLAN is around 3 Mbit/s. By changing the number of video flows in the mobile hotspot, we can obtain a varying traffic load to the WLAN channel. Figure 6 shows the data loss rate under different load conditions. As seen, the data loss rate increases slowly with a light traffic load but grows much faster when the traffic load becomes heavy. Based on the Markov-modulated video traffic model and fluid-flow analysis, we can effectively evaluate the data loss rate.

Figure 5 Effective bandwidth of video flows at different quantization levels.
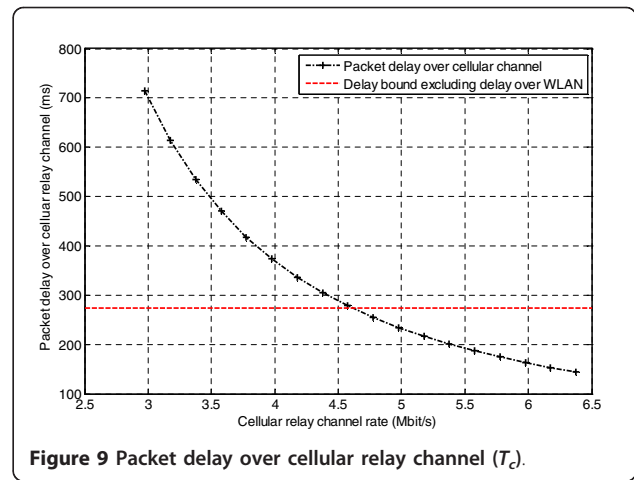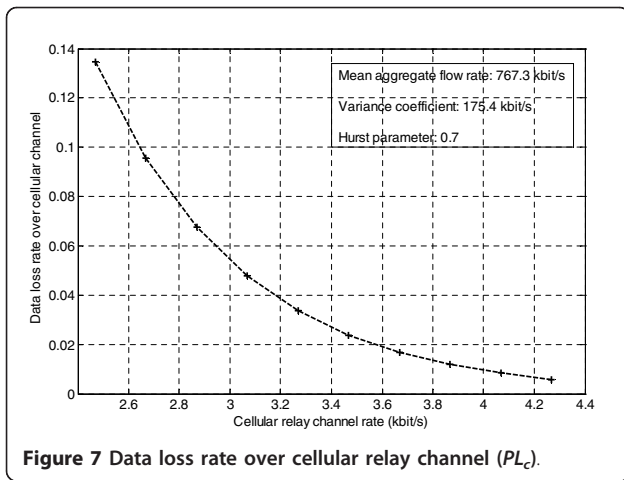
As shown in Figure 1, a two-hop wireless channel is applied for the moving network in a vehicular environment. In case of a handoff between neighboring cells, the ongoing traffic within the mobile hotsplot can be multiplexed at a local gateway and handed over together. Suppose we have a number of video flows in progress and the aggregate flow rate is around 767.3 kbit/s. Figure 7 illustrates the relationship between the bandwidth of cellular relay channel and the corresponding data loss rate over it. Due to self-similarity of the aggregate traffic flow, we use the FBM model in (6) to capture the data arrival process. Accordingly, the data loss rate over the cellular channel can be estimated. As seen from Figure 7, sufficient bandwidth should be reserved from the cellular network so as to bound the data loss rate by 0.01. In this example, the minimum bandwidth requirement is around 4.06 Mbit/s.

To enable real-time transmission of video packets over the two-hop scenario in Figure 1, the overall delay should also be upper bounded. Given the same

Table 2 Matrix of transition probabilities for a video flow

State	State					
	1	2	3	4	5	6
1	0.1866	0.5598	0.2493	0.0044	0	0
2	0.1384	0.2796	0.5345	0.0468	0.0007	0
3	0.0200	0.1811	0.6670	0.1308	0.0011	0
4	0.0007	0.0267	0.2459	0.6945	0.0322	0
5	0	0	0.0227	0.4838	0.4708	0.0227
6	0	0	0	0	0.7778	0.2222
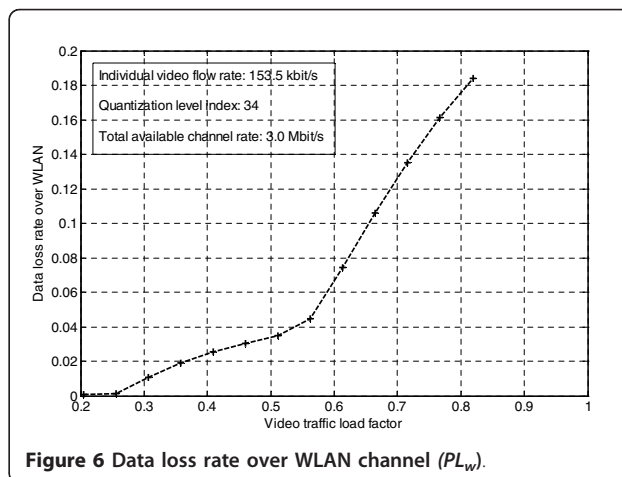
Figure 6 Data loss rate over WLAN channel (PL_w).

**Figure 7 Data loss rate over cellular relay channel (PL_c).**

parameters as above, Figures 8 and 9 show the delay performance at the packet level. For the last-hop WLAN, because of inter-coded frames, video traffic actually arrives in bursts and results in a batch of encapsulated packets for each video burst. Based on the queueing analysis in Section 3.3, packet delay can be evaluated according to a $D^{[A]}/G/1$ queueing system. The analysis preserves the essential video traffic features such as fixed burst inter-arrival time and batch structure for packet arrivals. The packet delay over the WLAN hop is approximated fairly well. As an example, we can see in Figure 8 that the average packet delay is around 75.7 ms when the load factor is 0.26. Because the WLAN channel is based on random access, it is hard to support bandwidth allocation. In contrast, sufficient bandwidth can be reserved for the relay channel from the cellular network so that the overall QoS constraints are satisfied. In Figure 9, the dashed red line shows the delay bound for the cellular relay channel excluding the WLAN-hop delay. The black dotted curve shows the variation of packet delay with cellular channel

Figure 8 Packet delay over WLAN channel (T_w).

Figure 9 Packet delay over cellular relay channel (T_c).

bandwidth. The packet delay is obtained according to (18). As seen, when the mean packet delay decreases with the increase of the cellular channel rate, we can obtain the channel bandwidth to be reserved by locating the intersection point with the line of delay bound. In this case, if a cellular channel with a bandwidth around 4.58 Mbit/s is reserved, the overall wireless access delay can be bounded by 350 ms on average.

5 Conclusions and future work

In this article, we investigate the resource reservation issue for moving networks in a vehicular environment. A two-hop relay structure integrates heterogeneous technologies of the cellular network and wireless local area network. Focusing on the ever-increasing video applications, we have introduced feasible analytical approaches to effectively evaluate video performance in terms of data loss rate and packet delay. To satisfy video QoS constraints, we can derive the bandwidth to be reserved for handover. Our performance evaluation has taken into account the two-hop relay structure and the essential characteristics of video traffic. At the flow level, we use a sigmoid function to model video flows as a Markov-modulated process and calculate data loss rate with fluid-flow analysis. At a finer packet level, due to inter-coded frames, video traffic arrives in bursts and results in transmission packets as a batch for each video burst. Our queueing analysis for packet delay has considered this feature and provided fairly accurate estimate. On the other hand, video flows from end users within a vehicular network can be multiplexed at a local gateway. As the aggregate traffic exhibits self-similarity, we use a FBM process to characterize the aggregate traffic flow beyond the wireless gateway. The data loss rate and packet delay for the second hop are approximated accordingly. As demonstrated in the numerical examples, the performance evaluation can be applied to
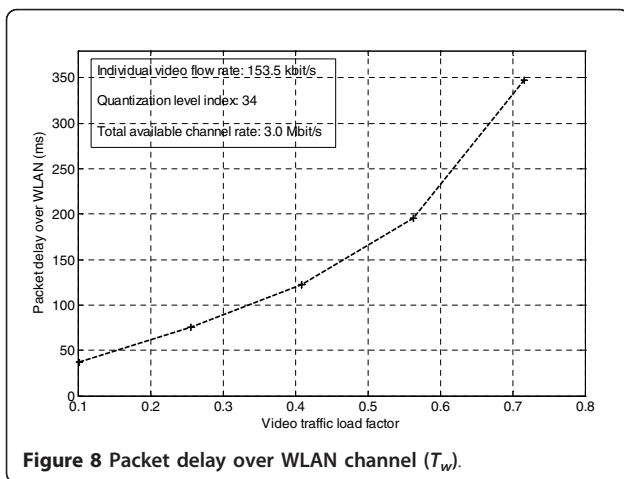
derive bandwidth requirement during handover. In our future study, we would examine the impact of varying channel rates due to radio fading and access contention. To optimize video delivery quality, we can develop an efficient handoff algorithm depending on the vehicular mobility and traffic variation.

Competing interests
The authors declare that they have no competing interests.

References
1. W Song, W Zhuang, Multi-service load sharing for resource management in the cellular/WLAN integrated network. IEEE Trans Wirel Commun. **8**(2), 725–735 (2009)
2. P Si, H Ji, FR Yu, Optimal network selection in heterogeneous wireless multimedia networks. ACM/Springer Wirel Netw. **16**(5), 1277–1288 (2009)
3. WS Mossberg, Internet-a-Gogo: Airlines to offer in-flight access. The Wall Street J http://online.wsj.com/article/SB121382851874286403.html (2008)
4. Cisco, Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015. (2011)
5. Allot, Allot mobile trends: Global mobile broadband traffic report. (2010)
6. S Wenger, H.264/AVC over IP. IEEE Trans Circ Syst Video Technol. **13**(7), 645–656 (2003). doi:10.1109/TCSVT.2003.814966
7. X Cai, F Liu, Network selection for group handover in multi-access networks. in *Proc IEEE ICC* 2164–2168 (2008)
8. UK Sarkar, S Ramakrishnan, D Sarkar, Markov-modulated Gamma-based framework. IEEE/ACM Trans Netw. **11**(4), 638–649 (2003). doi:10.1109/TNET.2003.815292
9. DP Heyman, The GBAR source model for VBR videoconferences. IEEE/ACM Trans Netw. **5**(4), 554–560 (1997). doi:10.1109/90.649513
10. WE Leland, MS Taqqu, W Willinger, DV Wilson, On the self-similar nature of Ethernet traffic (extended version). IEEE/ACM Trans Netw. **2**, 1–15 (1994). doi:10.1109/90.282603
11. Y Cheng, W Zhuang, X Ling, FBM model based network-wide performance analysis with service differentiation. in *Proc ICST Qshine* (2007)
12. 3GPP, Transparent end-to-end packet-switched streaming service (PSS); Protocols and codecs. (2010) 3GPP TS 26.234 V9.3.0
13. B Maglaris, D Anastassiou, P Sen, G Karlsson, J Robbins, Performance models of statistical multiplexing in packet video communications. IEEE Trans Commun. **36**(7), 834–844 (1988). doi:10.1109/26.2812
14. DMB Masi, MJ Fischer, DA Garbin, Video frame size distribution analysis. Telecommun Rev. **19**, 74–86 (2008)
15. F Wan, L Cai, TA Gulliver, A simple, two-level Markovian traffic model for IPTV video sources. in *Proc IEEE GLOBECOM* (2008)
16. P Seeling, M Reisslein, B Kulapala, Network performance evaluation with frame size and quality traces of single-layer and two-layer video: a tutorial. IEEE Commun Surv Tutorials. **6**(3), 58–78 (2004)
17. I Norros, On the use of fractal Brownian motion in the theory of connectionless networks. IEEE J Sel Areas Commun. **13**(6), 953–962 (2002)
18. J Welch, J Clark, RFC 4445: A proposed media delivery index (MDI). (2006)
19. Z He, J Cai, CW Chen, Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. IEEE Trans Circ Syst Video Technol. **12**(6), 511–523 (2002). doi:10.1109/TCSVT.2002.800313
20. H Zhai, X Chen, Y Fang, How well can the IEEE 802.11 wireless LAN support quality of service?. IEEE Trans Wirel Commun. **4**(6), 3084–3094 (2005)
21. X Jin, G Min, QoS analysis of queuing systems with self-similar traffic and heavy-tailed packet sizes. in *Proc IEEE ICC* 100–104 (2008)
22. LD Servi, D/G/1 queues with vacations. Oper Res. **34**(4), 619–629 (1986). doi:10.1287/opre.34.4.619
23. PJ Burke, Delays in single-server queues with batch input. Oper Res. **23**(4), 830–833 (1975). doi:10.1287/opre.23.4.830