Performance Analysis of Probabilistic Multipath Transmission of Video Streaming Traffic over Multi-Radio Wireless Devices

Wei Song, *Member, IEEE*, and Weihua Zhuang, *Fellow, IEEE*

***Abstract*—Popular smart wireless devices become equipped with multiple radio interfaces. Multihoming support can be enabled to allow for multiple simultaneous associations with heterogeneous networks. In this study, we focus on video streaming traffic and propose analytical approaches to evaluate the packet-level and call-level performance of a multipath transmission scheme, which sends video traffic bursts over multiple available channels in a probabilistic manner. A probability generation function (PGF) and z-transform method is applied to derive the PGF of packet delay and any arbitrary moment in general. Particularly, we can obtain the average delay, delay jitter, and delay outage probability. The essential characteristics of video traffic are taken into account, such as deterministic burst intervals, highly dynamic burst length, and batch arrivals of transmission packets. The video substream traffic resulting from the probabilistic flow splitting is characterized by means of zero-inflated models. Further, the call-level performance, in terms of flow blocking probability and system throughput, is evaluated with a three-dimensional Markov process and compared with that of an always-best access selection. The numerical and simulations results demonstrate the effectiveness of our analysis framework and the performance gain of multipath transmission.**

***Index Terms*—Video streaming, multipath transmission, performance analysis, packet delay, call admission control, multi-radio wireless devices.**

I. Introduction

Nowadays, the proliferating wireless infrastructure offers a variety of broadband access options such as cellular networks, IEEE 802.16 wireless metropolitan area networks / WiMAX (worldwide interoperability for microwave access), IEEE 802.20 broadband wireless access / Mobile-Fi, IEEE 802.11 wireless local area networks (WLAN) / Wi-Fi, and IEEE 802.15 wireless personal area networks (WPAN). To improve the system throughput, the multi-channel and multi-radio capabilities can be exploited to enable concurrent transmissions over multiple wireless links [1]. For instance, the mainstream smart phones such as RIM's BlackBerry and Apple's iPhone already have built-in Wi-Fi (over IEEE 802.11) and Bluetooth (over IEEE 802.15.1) in addition to a regular cellular radio. With the rapid development and breakthrough of wireless technologies, mobile video will generate most of the mobile traffic growth through 2015 as predicted by Cisco [2]. The statistics collected by leading mobile operators worldwide in 2010 [3] also show video streaming accounts for 37% of mobile data usage, which is the largest fraction next to file sharing (30%) and Web browsing (26%). To deliver high-quality video services, it becomes vital to consider the heterogeneous wireless infrastructure and multi-radio capability of mobile devices.

Manuscript received July 15, 2011, revised November 28, 2011, and accepted January 23, 2012. W. Song is with the Faculty of Computer Science, University of New Brunswick, Fredericton, Canada (Email: wsong@unb.ca). W. Zhuang is with the Department of Electrical and Computer Engineering, University of Waterloo, Canada (Email: wzhuang@uwaterloo.ca).

To coordinate the heterogeneous wireless access for multi-radio devices, network selection is one of the major issues that are researched intensively in the literature. Many centralized and distributed network selection algorithms are proposed to automatically assign an incoming traffic flow to a best available network [4,5]. Basically, access selection aims to share the heterogeneous network resources at the flow-level time scale. Taking one step further, we can exploit the multihoming and multi-streaming support to aggregate available bandwidth over diverse wireless links. Multihoming enables a wireless device to maintain multiple simultaneous associations with more than one attachment point. Multi-streaming allows data to be partitioned into multiple streams and delivered independently to the application at the receiver. Multi-streaming can prevent the head-of-line blocking problem that occurs in the transport control protocol (TCP). If multihoming and multi-streaming capabilities are enabled for multi-radio devices, a traffic flow can be split into multiple streams and delivered simultaneously over multiple network interfaces. As such, the access selection problem is addressed from a different perspective.

In this paper, we develop a framework to analytically evaluate the video streaming performance with flow splitting and multipath transmission. Taking advantage of multihoming capability of multi-radio devices, a simple multipath transmission scheme is considered to make use of fractional bandwidth available in integrated wireless networks. Although multiple access networks are available for multi-radio devices, it becomes challenging if none has sufficient bandwidth to accommodate a video flow experiencing large bursts of traffic. On the other hand, video frames actually arrive in batch due to forward, backward, or bidirectional prediction in video coding and compression. Exploiting such traffic patterns, we can split a bandwidth-demanding video flow in a simple probabilistic manner towards multiple links. The contribution of this paper is several-fold.

- First, we take into account the batch arrival nature and fixed inter-arrival time of video traffic in our delay performance analysis. As observed in the testing of real video traces [6], it is essential to preserve such characteristics to have accurate performance statistics.
- Second, we propose an analytical framework to evaluate the delay performance of a probabilistic flow splitting

scheme for multipath transmission. A closed-form probability generating function (PGF) of packet transfer delay is derived to obtain primary delay metrics.

- Third, we conduct numerical and simulation experiments with video traces to validate the feasibility and accuracy of the proposed framework. The packet-level and call-level performance of flow splitting and multipath transmission is also demonstrated with numerical results.

The rest of this paper is organized as follows. In Section II, we introduce related work on video performance analysis and multipath streaming. Section III gives the network model and traffic model for this study. An analytical framework is presented in Section IV for video streaming. Numerical results are presented in Section V, followed by conclusions and future work in Section VI.

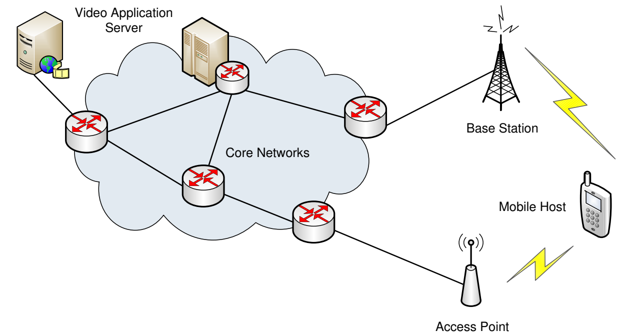


Fig. 1. System model of integrated heterogeneous wireless networks.

II. Related Work

In the literature, there has been extensive work to analyze video streaming performance [7]. Many studies focus on the modelling and analysis at the video frame level, group of pictures (GoP) level, and fluid flow level. Traditionally, video traffic can be viewed as a fluid flow and modelled with a Markov-modulated process by neglecting the traffic discreteness. In [8], the IPTV performance is investigated according to a two-level Markovian traffic model which captures both the GoP-level and frame-level characteristics of video traffic. In [9], Sarkar *et. al.* propose a Markov-modulated Gamma-based framework, which models the video frame size at each state with an axis-shifted Gamma distribution. For such Markov-based models, the video performance can be evaluated with a fluid flow analytical approach [10]. Accordingly, data loss probability and effective bandwidth of a video flow can be derived from the leaky bucket algorithm [11] by simulating data transmission with bucket leaking.

In practice, video bitstreams are packetized for transmission. The packet-level performance is of great interest and importance to have a complete perception of video quality. Multipath transmission may lead to more performance dynamics due to flow splitting among multiple links. Hence, it is necessary to closely examine the packet-level performance so as to evaluate the effects of multipath transmission. For analysis simplicity, most previous works assume Poisson packet arrivals [12]. The study in [13] relaxes the assumption and deals with inter-arrival time of video packets as an arbitrary distribution. However, the batch arrival [6] structure is not addressed.

To offer video services on multi-radio devices, the heterogeneous wireless access can be exploited with network selection for *always-the-best connectivity*. Previous studies on access selection [4,5] mainly focus on assignment of traffic flows during admission and dynamic reassignment via vertical handover for efficient resource sharing and quality-of-service (QoS) enhancement. On the other hand, multi-streaming is a promising approach that takes advantage of the multihoming capability of multi-radio devices and modifies the treatment of access selection by allowing more alternatives. The aggregation of available bandwidth of multiple networks is especially beneficial for wireless networks that may offer limited, varying, and distinct capacities [14]. The solutions for load splitting differ from the link layer to the application layer [14,15]. Basically, a link-layer solution is constrained to a local network and sensitive to channel rate variations. The network-layer load splitting challenges the transport-layer control by involving spontaneous timeouts due to disparate round-trip time (RTT) and unnecessary fast retransmissions with persistent out-of-order packets. The stream control transmission protocol (SCTP) is one of the well-known transport-layer approaches. The original SCTP is designed to improve throughput and reliability by exploiting multiple paths. A primary path is selected for transmission of data chunks, while a secondary path is used only for retransmission of lost data units or as a backup for the primary path. However, the transport-layer solutions may sacrifice compatibility with the pervasive TCP. An application-layer solution can minimize modifications to the existing network infrastructure. Nonetheless, the stripping and merging of data streams further complicate the design of user applications.

Many multipath streaming approaches focus on maximizing throughput [15]. For real-time applications, delay is another key performance metric along with throughput [12]. The trade-off between throughput and delay is generally explored for wireless networks in [16]–[18]. To satisfy QoS requirements of a specific application such as video streaming, it is essential to minimize the delay while achieving a high throughput, i.e., to optimize the delay-constrained throughput. In this work, we exploit video streaming traffic characteristics at the application layer and apply a probabilistic flow splitting to improve user perceived QoS in terms of delay and throughput.

III. System Model

A. Network Model

Consider a heterogeneous wireless network infrastructure integrating multiple access options, as illustrated in Fig. 1. The mobile host is equipped with multiple radio interfaces and multihoming capable. Multipath transmission can be enabled to deliver traffic between the multi-homed mobile host and the application server. A middleware is deployed at both sides to deal with splitting and merging of traffic flows across available networks. The middleware employs the application-layer and network-layer information in load splitting. In particular, the available bandwidth over each associated network can be

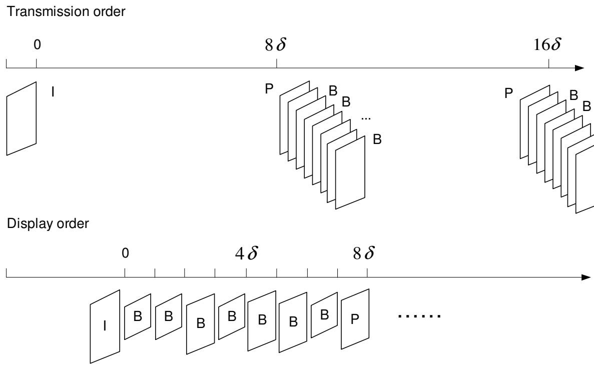


Fig. 2. Transmission and display orders of video frames.

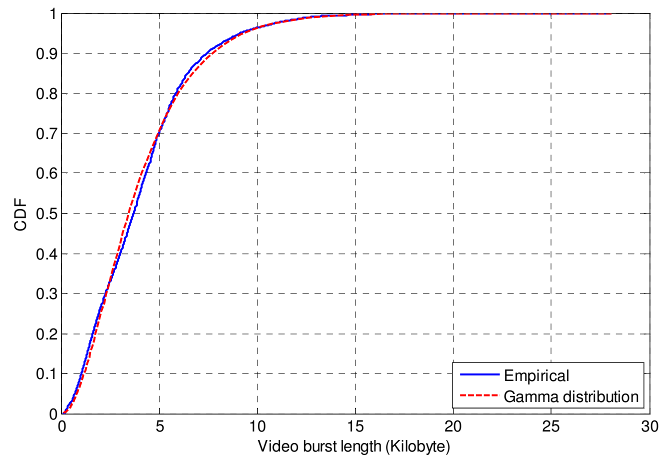


Fig. 3. Empirical CDF of video burst length from real traces and CDF of Gamma distribution.

estimated by packet probing over each interface of the mobile host. It is known that wireless access is most likely to be the bottleneck for end-to-end transmission due to the limited bandwidth and time-varying channel. Hence, we focus on the wireless access domain in Fig. 1.

B. Video Streaming Traffic

It is known that video traffic is inherently long-range dependent and highly correlated due to compression coding. In cellular networks, H.264 Advanced Video Coding (AVC) is recommended for high-quality video. 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, traffic transmission follows the batch arrivals shown in Fig. 2. Here, the GoP follows a structure of size 16 such as "$I_0P_8B_1B_2\ldots B_7P_{16}B_9B_{10}\ldots B_{15}\ldots$". In contrast, video frames are decoded and displayed at the receiver in a reorganized order. Hence, B frames are subject to a more stringent delay constraint than I and P frames.

In the literature, there has been extensive work modeling the varying rate and frame size of video traffic. A Markov-modulated Gamma-based model is proposed in [9] to capture video frame size variations and auto-correlation by grouping video clips of a stream into a small number of classes. The sizes of I, P, and B frames in each class are modeled by an axis-shifted Gamma distribution. On the other hand, as observed in [6], given that the batch structure and fixed inter-arrival time are preserved, the use of a hypothesized and independent distribution for frame sizes approximates the trace behavior fairly well and gives close performance statistics. Motivated by the above observation, we model the length of video burst between two key I/P frames (denoted by S) by a Gamma distribution with a probability density function (PDF)

$$f_S(x) = \frac{x^{\alpha-1}e^{-x/\eta}}{\Gamma(\alpha)\eta^{\alpha}}, \qquad \alpha > 0, \quad \eta > 0 \tag{1}$$

where α and η are the shape parameter and scale parameter, respectively.

To check the feasibility of the hypothesized distribution for video burst length, we compare the empirical cumulative distribution function (CDF) of video traces and that of a Gamma distribution. The shape parameter and scale parameter are set by matching the mean and variance of the Gamma distribution, denoted by $\overline{S} = \alpha\eta$ and $\sigma_S^2 = \alpha\eta^2$, respectively. As an example, we consider a video trace, *NBC News*, from the video trace library of Arizona State University [19]. The video sequence takes a common intermediate format (CIF) resolution (352×288), a fixed frame rate at 30 frames/s, a GoP size of 16 with 7 B frames between I/P key pictures, and a quantization step-size indexed at 38. The quantization level varies with the step-size, and a higher quantization index (between 0 and 51) results in a lower encoding bit rate. As shown in Fig. 3, the empirical CDF of video burst length obtained from the trace is very close to that of the hypothesized Gamma distribution.

C. Discrete-Time Traffic Model

As seen in Fig. 2, video frames are generated in burst according to a coding and compression algorithm. Each video burst consists of an I or P frame and a number of B frames between two key I/P frames. A number of B frames are generated for each video burst depending on a target encoding bit rate. For example, a video trace with a GOP size of 16 can have 0 to 15 B frames in a traffic burst. Given a constant frame rate f, the traffic burst rate is $g = f/(J+1)$, where J is the number of B frames between two key I/P frames. For analysis purpose, 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. At the finer packet level, we use a batch arrival process to model the video traffic. Video frames generated in a burst are considered as a batch and fragmented into a random number of transmission packets of fixed size Δ. As the length of video burst is modelled with a Gamma distribution, we characterize the number of packets in a "packet train" with a negative binomial distribution $NB(r, p)$, which is a

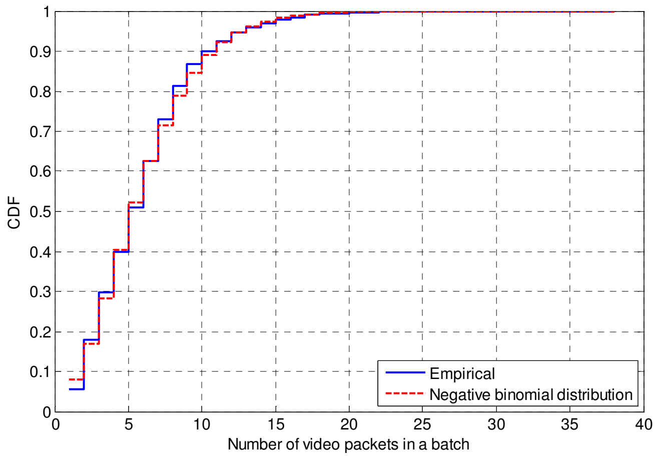


Fig. 4. Empirical CDF of the number of video packets in a batch from real traces and CDF of negative binomial distribution.

discrete analogue of Gamma distribution. The probability mass function (PMF) of the number of packets in a batch (denoted by A) is given by

$$\mathrm{P}[A = k] \triangleq f_A(k) = \binom{k+r-1}{r-1}(1-p)^r p^k \qquad (2)$$
$$k = 0, 1, \ldots, \quad r > 0, \quad 0 < p < 1.$$

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

$$\overline{A} = \frac{rp}{1-p}, \qquad \sigma_A^2 = \frac{rp}{(1-p)^2}. \qquad (3)$$

Similarly, we plot the empirical CDF of video batch size from the video trace and that of the corresponding negative binomial distribution. Here, we choose a packet size of 750 bytes for transmission packetization. As seen in Fig. 4, the negative binomial distribution provides a close approximation for the batch size. Our following analysis can also be extended to a general distribution for the number of packets in a batch.

IV. Performance Analysis for Multipath Video Streaming

A. Probabilistic Video Flow Splitting

Suppose there are multiple access networks available for a multi-radio device. In a heavy traffic situation, none of the available access may be able to solely provide sufficient bandwidth to support the video flow. Nonetheless, it is possible to accommodate the video flow by multiple networks simultaneously. A video streaming flow can be split into multiple substreams and delivered through different networks simultaneously. According to a video codec, each video traffic burst is generated over fixed intervals and consists of an I or P frame and a number of B frames. A basic idea of flow splitting is to randomly dispatch each video burst to an access network i with a probability ϑ_i, $\sum_i \vartheta_i = 1$. To guarantee overall QoS specified in delay metrics for all video substreams, we should derive the splitting probability ϑ_i according to the effective channel rate b_i of the network i.

B. PGF of Packet Delay with Batch Arrivals

Video streaming plays back video content at the receiver during the delivery. A stringent delay requirement needs to be satisfied to prevent data overflow and depletion at the playout buffer. Particularly, as multipath transmission may result in a high variation to delivery performance, it is essential to effectively evaluate the packet transfer delay of video streaming traffic.

According to the probabilistic flow splitting, each video burst is dispatched to an available access network i with a probability ϑ_i. Let h_i denote the average number of time units to transmit a packet of size Δ through network i. The equivalent channel data rate to the video application is then $b_i = \Delta/(h_i \cdot \tau)$. For analysis simplicity, h_i is assumed to be an integer. Here, we focus on the wireless access domain, which is usually the bottleneck of the end-to-end path. The packet transmission time (denoted by H_i) depends on the radio access links as well as the service traffic loads in different wireless access networks. In other words, the characteristics of different transmission paths and service loads can be captured in H_i. For presentation clarity, we omit the subscript i for an available access in the following.

For a wireless channel, the channel data rate or equivalent packet transmission time is random due to channel fading and/or access contention. For instance, as investigated in [12,20], the packet transmission time over the 802.11 WLAN link is accurately approximated by an exponential distribution. Given that the packet transmission time over heterogeneous wireless networks is properly characterized, our following approach provides a generic analytical framework for the delay performance. Generally, effective numerical evaluation of derivatives and definite integrals is required for a complex channel characterization. We also present specific derivation assuming a constant packet transmission time. It is observed in Section V.A that we obtain accurate results even if the assumption is relaxed.

Given the aforementioned batch structure and constant burst interval, the delay (denoted by T) experienced by a tagged packet in a video batch 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_Q; and 3) the transmission time of the tagged packet, which is h time units on average.

Firstly, W_G is the waiting time for a video batch to become the head of the queue. When the first packet in the batch is queued for transmission, the batch of packets all experience the waiting time W_G before the transmission starts for the batch. To evaluate W_G with a queueing system, each video batch can be regarded as a single customer whose service time is the total transmission time of all packets in a batch. An analytical approach is introduced in [21] for the waiting time of a $D/G/1$ queue, whose inter-arrival time is deterministic and service time follows a general distribution. However, due to multipath transmission, traffic arrivals are not exactly deterministic but follow a zero-inflated model [22]. That is, for each constant burst interval, there is no traffic arrival with a probability $1-\vartheta$, whereas there is an incoming video batch with a probability ϑ. Hence, we modify the batch service time (denoted by G) to incorporate zero-sized batches. The probability generating function (PGF) of G is then

$$G(z)=\sum_{k=0}^{\infty}\mathrm{P}[G=k]z^k=(1-\vartheta)+\vartheta B(z) \tag{4}$$

where $B(z)$ is the PGF of the total service time for a video batch.

As defined in (2), the number of packets generated from each burst of the original video clip follows a negative binomial distribution. The PGF of the distribution $NB(p,r)$ can be easily obtained as

$$A(z)=\sum_{k=0}^{\infty}f_A(k)z^k=\left(\frac{1-p}{1-pz}\right)^r. \tag{5}$$

Letting H denote the packet transmission time, we have the PGF of the total service time of a video batch as follows

$$\begin{aligned}B(z)&=\sum_{k=0}^{\infty}\mathrm{P}[B=k]z^k\\&=\sum_{n=0}^{\infty}\mathrm{P}[A=n]\sum_{k=0}^{\infty}\mathrm{P}[H_1+...+H_n=k|A=n]z^k\\&=\sum_{n=0}^{\infty}\mathrm{P}[A=n]\cdot[H(z)]^n=A(H(z)).\end{aligned} \tag{6}$$

In particular, if the packet transmission time is constant at h time units, the PGF of the batch service time can be scaled accordingly, given by

$$B(z)=\Big(\frac{1-\tilde{p}}{1-\tilde{p}z}\Big)^{\tilde{r}} \tag{7}$$

$$\tilde{p}=1-\frac{\overline{A}h}{\sigma_A^2h^2}=1-\frac{1-p}{h},\qquad \tilde{r}=\overline{A}h\Big(\frac{1-\tilde{p}}{\tilde{p}}\Big)=\frac{rp}{\tilde{p}}.$$

Then, using the approach in [21], we can obtain the PGF of the waiting time, given by

$$W_G(z)=\frac{\Phi\cdot(z-1)\prod_{k=1}^{N-1}(z-z_k)}{z^N-G(z)} \tag{8}$$

where z_1, ..., z_{N-1} are the unique roots of $z^N-G(z)=0$ within the unit circle $|z|<1$, and Φ is a normalization constant. When N is large, we can apply the Muller method [23] to find the roots z_k. The normalization constant Φ is calculated by

$$\Phi=\lim_{z\to1^-}\frac{z^N-G(z)}{(z-1)\prod_{k=1}^{N-1}(z-z_k)}. \tag{9}$$

Based on (8), the mean and variance of W_G are respectively given by

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

$$\begin{aligned}\sigma_{W_G}^2=&-\sum_{k=1}^{N-1}\frac{z_k}{(1-z_k)^2}-\frac{N(N-1)(N-2)-G'''(1^-)}{3[N-G'(1^-)]}\\&-\frac{N(N-1)-G''(1^-)}{2[N-G'(1^-)]}+\left[\frac{N(N-1)-G''(1^-)}{2[N-G'(1^-)]}\right]^2.\end{aligned}$$

Next, we evaluate the waiting time W_Q due to the transmission of other packets prior to a tagged one in a video batch. After a batch queues for the time W_G to start transmission, a specific packet in the batch further waits for the time W_Q until the packets queued before it are transmitted. Only after that can a tagged packet capture the channel for transmission. To evaluate W_Q, we need to focus on batches of non-zero sizes. According to the probabilistic flow splitting and the original batch size distribution of the video clip, with a probability $1-\vartheta$, there is no video burst dispatched to a given network; and with a probability ϑ, a batch of video packets are delivered towards the designated network and the number of packets (or batch size) follows a negative binomial distribution. Let A_N denote the number of packets in the non-zero sized batch that arrives to an access network after flow splitting during each video burst interval. We have the PGF of A_N as follows

$$A_N(z)=\frac{A_0(z)-p_0}{1-p_0} \tag{10}$$

where $A_0(z)$ is the PGF of the size of zero-inflated batches resulting from flow splitting and p_0 is the probability that the zero-inflated process produces a zero-sized batch. Clearly, after every burst interval of N time units, the probability that there is no video batch arrival or the batch size is zero is given by

$$p_0=1-\vartheta+\vartheta\cdot(1-p)^r. \tag{11}$$

In (10), $A_0(z)$ is the PGF of the batch size after zero-inflated with flow splitting. Obviously, we have

$$A_0(z)=1-\vartheta+\vartheta\cdot A(z)=1-\vartheta+\vartheta\left(\frac{1-p}{1-pz}\right)^r \tag{12}$$

where $A(z)$ is is the PGF of the number of packets in a burst of the original video clip given in (5).

As defined, W_Q is the waiting time of a tagged packet to transmit all the other packets that are generated in the same video batch of the tagged packet but queued before it. Clearly, W_Q depends on the non-zero batch size ($A_N>0$) and the position of the tagged packet. According to the analysis in [24], the probability that an arbitrary tagged packet falls within a batch of a size k is given by

$$u_k=\frac{k\cdot\mathrm{P}[A_N=k]}{\overline{A}_N},\ \mathrm{E}[A_N]\triangleq\overline{A}_N=\lim_{z\to1^-}\frac{dA_N(z)}{dz}. \tag{13}$$

If n packets from the same video batch as the tagged packet are queued prior to it, the batch size must be no less than n.

Hence, we obtain the PGF of the number of packets queued before the tagged one as

$$Y(z) = \sum_{n=1}^{\infty} z^n \sum_{k=n+1}^{\infty} u_k \cdot \frac{1}{k} \tag{14}$$
$$= \sum_{n=1}^{\infty} z^n \sum_{k=n+1}^{\infty} \frac{k \cdot \mathrm{P}[A_N = k]}{\overline{A}_N} \cdot \frac{1}{k} = \frac{1 - A_N(z)}{\overline{A}_N(1-z)}.$$

Given the PGF $H(z)$ of the packet transmission time, the PGF of the waiting time W_Q is then

$$W_Q(z) = Y(H(z)). \tag{15}$$

Particularly, if the packet transmission time is deterministic and equal to h time units, we have

$$H(z) = z^h, \qquad W_Q(z) = Y(z^h). \tag{16}$$

Hence, the k^{th} factorial moment of W_Q can be obtained from (15) as follows

$$\mathrm{E}\big[(W_Q - 1)...(W_Q - k + 1)\big] = \lim_{z \to 1^-} \frac{d^k W_Q(z)}{dz^k}. \tag{17}$$

The average and variance of W_Q are then

$$\overline{W}_Q = \lim_{z \to 1^-} \frac{dW_Q(z)}{dz}$$
$$\sigma^2_{W_Q} = \lim_{z \to 1^-} \frac{d^2 W_Q(z)}{dz^2} + \overline{W}_Q - (\overline{W}_Q)^2.$$

Since the overall packet delay T consists of the three independent components, the corresponding PGF is obtained from (8) and (15) as

$$T(z) = H(z) \cdot W_Q(z) \cdot W_G(z). \tag{18}$$

The average packet delay and delay jitter can be evaluated accordingly by

$$\overline{T} = h + \overline{W}_G + \overline{W}_Q, \ \psi = \left| \sqrt{\sigma^2_{W_G} + \sigma^2_{W_Q} + \sigma^2_H} - \overline{T} \right|. \tag{19}$$

C. Evaluation of Packet Delay Outage Probability

As seen in Fig. 2, video frames are encoded and transmitted in burst, whereas video frames are decoded and displayed at the receiver in the GoP order. Because video frames are captured and coded in constant intervals, there is a deadline to play back a designated video frame at the receiver. If a frame to play has not been completely delivered to the receiver buffer at fetch time, the playback is interrupted. Let L denote the packet lifetime, beyond which the resulting video frame expires for playback. The probability that the packet delay exceeds the packet lifetime is referred to as *delay outage probability*. An expired packet can be discarded since it becomes useless for the receiver playback. To ensure continuous and smooth playback, we need to upper-bound average packet delay, delay jitter, as well as delay outage probability.

Consider the video traffic structure illustrated in Fig. 2. Suppose the k^{th} burst of video frames are transmitted at $t_{k,0}$, the display deadlines of the B frames are $t_{k,1}$, $t_{k,2}$, ..., $t_{k,J-1}$, respectively, and the display deadline of the I/P key frame is $t_{k,J}$. Here, J is the number of B frames between two key I/P frames. Since video streaming enables simultaneous delivery and playback, there is usually a short pre-roll delay (denoted by s_0) in the order of $5-15$ seconds between the start of delivery and the beginning of playback at the receiver [25]. Referring to the coding structure in Fig. 2, we have $t_{k,1} = s_0 + t_{k,0}$, $t_{k,j+1} = t_{k,j} + \delta$, and $t_{k,J} = t_{k,1} + (J-1)\delta$, where δ is the frame interval. As each video burst is segmented into a batch of video packets, we can obtain the statistics of packet lifetime $L = t_{k,j} - t_{k,0}$.

Due to the video burst arrivals and queueing effects through the end-to-end path, both the packet delay and packet lifetime are random. As studied in [26], the packet lifetime follows an exponential distribution. In the following analysis, we consider a geometric distribution for the packet lifetime, which is a discrete analogue of the exponential distribution. The analysis is also valid for a general packet lifetime. Denoting the packet lifetime by L, we obtain the PGF of a geometric-distributed lifetime as

$$L(z) = \frac{\beta z}{1-(1-\beta)z}, \quad 0 < \beta \le 1, \quad |z| < \frac{1}{1-\beta}. \tag{20}$$

Given the two independent random variables, T and L, for packet delay and lifetime, the PGF of the difference between the two, denoted by $F = T - L$, is obtained as

$$F(z) = T(z)L(z^{-1}) \tag{21}$$

where $T(z)$ and $L(z)$ are derived in (18) and (20), respectively. Based on the statistics of F, the delay outage probability ξ can be calculated by

$$\xi = 1 - \mathrm{P}\big[F \le 0\big] = 1 - \sum_{k=-\infty}^{0} \mathrm{P}\big[F = k\big]. \tag{22}$$

Letting $f_F(k) \triangleq \mathrm{P}\big[F = k\big]$ and $x(n) \triangleq \sum_{k=-\infty}^{n} f_F(k)$, we have

$$\xi = 1 - x(0). \tag{23}$$

Although it is not straightforward to directly obtain $x(0)$, we resort to the z-transform of $x(n)$, given by

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^n = \frac{1}{1-z} F(z). \tag{24}$$

In general, $x(n)$ can be derived from $X(z)$ via a contour integral at C_γ, which is the border of a circle centered at zero with a radius γ [27], that is

$$x(n) = \frac{1}{2\pi i} \oint_{C_\gamma} \frac{X(z)}{z^{n+1}} dz. \tag{25}$$

From the convergence regions of $T(z)$ and $L(z)$, we have $1 - \beta < \gamma < 1$. Numerical inversion is also possible [28] by setting $z = \gamma e^{i\phi}$. Eq. (25) can then be rewritten as

$$x(n) = \frac{1}{2\pi\gamma^n} \int_0^{2\pi} X(\gamma e^{i\phi}) e^{-in\phi} d\phi. \tag{26}$$

Thus, the delay outage probability is obtained from (26) as

$$\xi = 1 - \frac{1}{2\pi} \int_0^{2\pi} X(\gamma e^{i\phi}) d\phi. \tag{27}$$

D. Call-Level Performance with Admission Control

Among multiple access networks, it is likely that no individual network in a heavy load situation can satisfy the overall bandwidth requirement and delay constraints. An end-to-end delay over 250 ms is in general unacceptable for video delivery quality [29]. Since our study focuses on the wireless access domain, we can reserve certain margin for the core network while setting the delay bound T_B for wireless packet transmission. Although flow splitting and multipath transmission can aggregate fractional available bandwidth over multiple access networks, there is a control overhead for flow splitting and reassembly. A two-phase procedure can be applied to make an *admission control* decision for an incoming video flow.

Firstly, it should be checked if any single access can carry the entire video stream while satisfying prescribed delay constraints. Based on the packet-level delay analysis in Section IV, the minimum average packet transmission time h_s can be determined based on (19) and (27). Accordingly, we can obtain the minimum required channel rate $b_s = \Delta/(h_s \cdot \tau)$. Take a wireless network based on orthogonal frequency-division multiplexing (OFDM) as one example. A subset of subcarriers are grouped to form a subchannel, which is the basic frequency-resource unit allocated by the base station or access point. Further taking into account the upper-layer transmission overhead, we can obtain the number of required subchannels to support the minimum channel rate b_s. Given that the data rates of available networks b_i are known, we can determine whether the video flow is able to be streamed over a single access.

Secondly, if no single access can provide sufficient bandwidth to carry a video stream, flow splitting can be activated among multiple paths. Nonetheless, it is cost-effective to activate multipath transmission only when the resource occupancy of the available access networks exceeds an upper threshold ϱ ($0 < \varrho < 1$). A video flow is split into multiple substreams by dispatching each video burst towards an access network i with a probability ϑ_i. The flow splitting probabilities can be determined with a heuristic search algorithm. Starting with the access channel of the largest available bandwidth, we can find the corresponding flow splitting probability ϑ_1 that minimizes delay outage provability ξ_1 and satisfies average packet delay requirement $\overline{T}_1 \le T_B$ [30]. Then, the remaining flow zero-inflated with a probability $1 - \vartheta_1$ is considered to derive the flow splitting probability ϑ_2 for the channel of the second largest bandwidth. The procedure continues until all flow splitting probabilities are determined. Due to bursty video packet arrivals, the constraint on delay outage probability can be much more stringent than the bound for average packet delay. Therefore, we check the overall delay outage probability at the end and only accept flow splitting if overall delay outage is bounded as well. A video flow is then admitted with multipath transmission if all related substreams are accepted by corresponding access networks.

We can quantitatively evaluate the performance gain of multipath transmission at the call level. According to the statistics of online streaming media [31], the video clip duration is in the order of several minutes and independent of the occupied bandwidth. Assume that video flows arrive as a Poisson process of a mean rate λ and have an exponentially distributed duration with a mean $1/\mu$. As an example, consider a scenario with dual-mode wireless devices. Let C_1 and C_2 denote the total number of subchannels in two access networks. As flow splitting and merging involve a processing overhead, it is cost-effective if multipath transmission is activated only when the resource occupancy of the two networks exceeds the upper bounds $R_1 = \lceil \varrho \cdot C_1 \rceil$ and $R_2 = \lceil \varrho \cdot C_2 \rceil$, respectively, where $0 < \varrho < 1$ is a threshold. If both networks have sufficient bandwidth, the less loaded network is selected. We can derive the number of required subchannels v to support a complete video flow. Given flow splitting with probabilities ϑ_1 and $\vartheta_2 = 1 - \vartheta_1$ to the two networks, the number of required subchannels are v_1 and v_2 for the two substreams, respectively. Obviously, $v_1 < v$ and $v_2 < v$. Thus, we formulate a three-dimensional Markov process to model the call admission control procedure, in which a state (m_1, m_2, ℓ) indicates that the number of used subchannels in the two networks are m_1 and m_2, respectively, and there are ℓ multipath streaming flows in progress. The state space is defined by [1]

$$\begin{aligned}\Omega =\Big\{&(m_1, m_2, \ell) : m_1 \in \mathbb{N}_0, m_2 \in \mathbb{N}_0, \ell \in \mathbb{N}_0, 0 \le m_1 \le C_1,\\ &0 \le m_2 \le C_2, 0 \le \ell \le \min\left(\lfloor C_1/v_1 \rfloor, \lfloor C_2/v_2 \rfloor\right),\\ &m_1 \ge \ell v_1, m_2 \ge \ell v_2, (m_1 - \ell v_1) \mid v, (m_2 - \ell v_2) \mid v\Big\}\end{aligned}$$

where $\mathbb{N}_0$ refers to the set of natural numbers (non-negative integers). Then, we have the rates of outgoing transitions from an arbitrary state (m_1, m_2, ℓ) shown on top of next page.

The steady-state probabilities of the above Markov chain, denoted by $\pi(m_1, m_2, \ell)$ can be obtained by solving a sparse linear system of balance equations. The overall flow blocking probability with multipath transmission, denoted by P_b, is then derived by

$$\begin{aligned} P_b &= \sum_{(m_1,m_2,\ell)\in\Lambda} \pi(m_1, m_2, \ell) \qquad (29)\\ \Lambda &= \big\{(m_1, m_2, \ell) : (m_1 + v_1, m_2 + v_2, \ell + 1) \notin \Omega,\\ &\qquad (m_1 + v, m_2, \ell) \notin \Omega, (m_1, m_2 + v, \ell) \notin \Omega\big\}.\end{aligned}$$

The average number of single-stream flows (denoted by N_s) and multipath streaming flows (denoted by N_m) can also be obtained from the steady-state probabilities as follows

$$\begin{aligned} N_s &= \sum_{(m_1,m_2,\ell)\in\Omega} \left(\frac{m_1 - \ell v_1}{v} + \frac{m_2 - \ell v_2}{v}\right) \pi(m_1, m_2, \ell)\\ N_m &= \sum_{(m_1,m_2,\ell)\in\Omega} \ell \cdot \pi(m_1, m_2, \ell). \qquad (30)\end{aligned}$$

Considering that video packets received beyond deadlines are useless for playback, we can evaluate the effective system throughput by

$$\Psi = N_s \cdot \chi \cdot (1 - \xi_s) + N_m \cdot \chi \cdot (1 - \xi_m) \qquad (31)$$

[1] We use the notation $a \mid b$ to indicate a is divisible by b.

$$\begin{array}{lll} (\boldsymbol{m_1+v, m_2, \ell}): & \lambda & \text{if } m_1+v<R_1, m_1<m_2 \text{ or } m_2+v>R_2, (m_1+v,m_2,\ell)\in\Omega \\ (\boldsymbol{m_1, m_2+v, \ell}): & \lambda & \text{if } m_2+v<R_2, m_2<m_1 \text{ or } m_1+v>R_1, (m_1,m_2+v,\ell)\in\Omega \\ (\boldsymbol{m_1+v_1, m_2+v_2, \ell+1}): & \lambda & \text{if } m_1+v>R_1, m_2+v>R_2, (m_1+v_1,m_2+v_2,\ell+1)\in\Omega \\ (\boldsymbol{m_1-v, m_2, \ell}): & \mu\cdot(m_1-\ell v_1)/v & \text{if } (m_1-\ell v_1)>0, (m_1-\ell v_1)\mid v \\ (\boldsymbol{m_1, m_2-v, \ell}): & \mu\cdot(m_2-\ell v_2)/v & \text{if } (m_2-\ell v_2)>0, (m_2-\ell v_2)\mid v \\ (\boldsymbol{m_1-v_1, m_2-v_2, \ell-1}): & \ell\mu & \text{if } \ell>0. \end{array} \tag{28}$$

where χ is the incoming video flow rate, ξ_s is the delay outage probability of single-stream flows, and ξ_m is the delay outage probability when flow splitting is enabled.

To demonstrate the performance gain of multipath transmission, we can compare with an always-best access selection scheme, which only chooses the less loaded network to admit a video flow. The corresponding call-level performance can be evaluated as above by setting $R_1 = C_1 + 1$ and $R_2 = C_2 + 1$, which means multipath transmission is never activated.

V. Numerical Results and Discussions

In this section, we present example numerical results with our analytical approaches to evaluate the performance of the probabilistic multipath transmission scheme. Table I gives the system parameters for numerical analysis. The video traffic parameters are selected according to H.264/AVC video traces from the library of Arizona State University [19]. To address high traffic variations, we choose the video sequences of *Star War IV* and *Silence of the Lambs*, which contain many quick scene changes and fast motions. These video sequences have a CIF resolution, a fixed rate of 30 frames/s, a GoP size of 16, and 7 B frames between two I/P key pictures. The quantization step-size is indexed at 22, 24, 28, 34, and 38 to have a mean encoding bit rate ranging from 43.7 kbit/s to 330.5 kbit/s.

TABLE I

System parameters for numerical analysis.

Symbol	Value	Definition
f	30	Video frame rate (/s)
g	3.75	Video burst rate (/s)
J	7	Number of B frames between two key I/P frames
τ	0.006	Time unit (s)
Δ	6000, 9600	Transmission packet size (bits)
N	$17 \sim 45$	Video burst interval (time units)
h	$1 \sim 14$	Packet transmission time (time units)
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
$\overline{S}$	12288	Mean video burst length S (bits)
σ_S	11499	Standard deviation of video burst length S (bits)
α	1.1419	Shape parameter of Gamma distribution for video burst length S
η	10760.33	Scale parameter of Gamma distribution for video burst length S
ϑ	$0 \sim 1$	Flow splitting probability
λ	$0.02 \sim 0.12$	Average streaming flow arrival rate (/s)
μ	150	Average streaming flow duration (s)
ϱ	0.8	Resource occupancy threshold
C_1	24	Number of subchannels in access network 1
C_2	30	Number of subchannels in access network 2
v	4	Number of required subchannels to admit a video flow
v_1	2	Number of required subchannels to admit a video substream in access network 1
v_2	3	Number of required subchannels to admit a video substream in access network 2

A. Analysis Validation

In Section IV, we introduce an analytical approach to evaluate the packet delay T. The calculation of the PGF of T and corresponding delay metrics involves operations such as one-sided limit, derivatives, root-finding for complex-valued equations, and definite integrals. Because it may not be feasible to obtain symbolic solutions in any case, we need to resort to numerical evaluation when necessary. For example, Muller method [23] can be used to find the roots z_k in (8) if N is very large. We use the numerical algorithms in the Symbolic Math Toolbox of MATLAB 7.10.0 (R2010a) for one-sided limit, derivatives, and definite integrals. As numerical evaluation may introduce approximation errors, we conduct computer simulations to validate the feasibility and accuracy of the analytical approach. A discrete event-driven simulator is developed with *C++* following the traffic model in Section III.C and the probabilistic flow splitting in Section IV.A. For each simulation run, around 10^6 video packets are generated and transmitted to collect the delay statistics. The results of 5 simulation rounds are averaged to reduce randomness effect.

Fig. 5 shows the comparison of the simulation results and analytical results in terms of average packet delay $\overline{T}$, delay jitter ψ, and delay outage probability ξ. Here, we take a constant packet size $\Delta = 6000$ bits but a varying packet transmission time, i.e., h time units. As such, the experiments correspond to varying the channel bandwidth and effective data rate. The video traces have a quantization step-size indexed at 38. As seen in Fig. 5, the simulation results match well the analytical results for various scenarios. There are small approximation errors introduced in the calculation, which are $0.07\% - 1.4\%$ for average packet delay, $0.26\% - 4.9\%$ for delay jitter, and $0.15\% - 1.9\%$ for delay outage probability. This verifies the feasibility of our analytical approach and the accuracy of the numerical evaluation. The proposed analysis captures the essential characteristics of video traffic, such as deterministic burst intervals, highly varying burst length, and

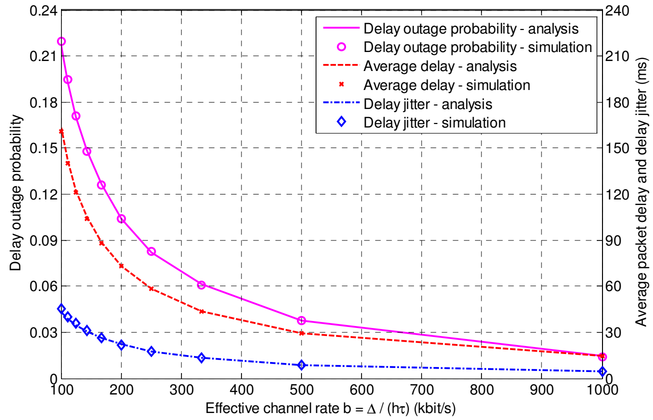


Fig. 5. Comparison of simulation and analysis results in terms of packet delay performance.

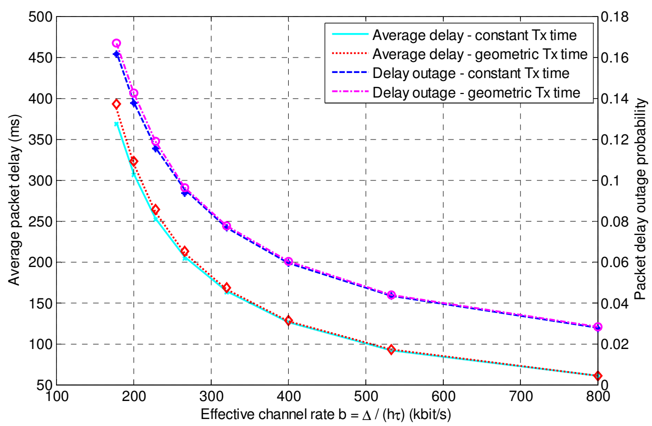
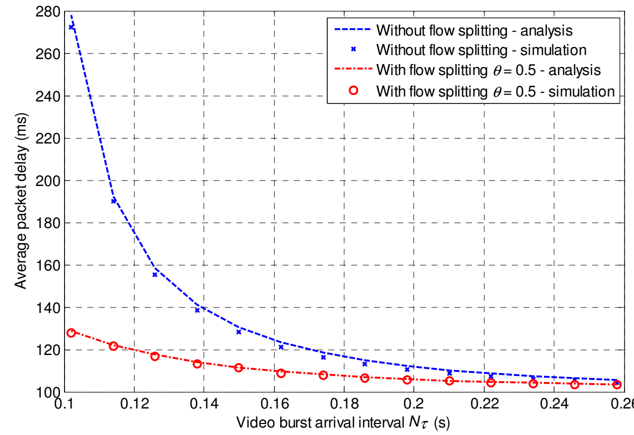
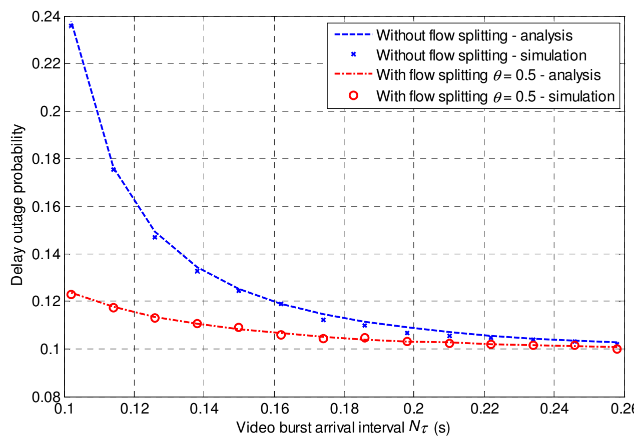


Fig. 6. Packet delay performance when packet transmission time H is constant and geometric distributed.



(a) Average packet delay.



(b) Delay outage probability.

Fig. 7. Delay performance with and without flow splitting when video burst intervals vary.

batch arrivals of transmission packets. Note that, given the PGF of the packet delay in (18), any arbitrary moment of the packet delay can be obtained.

Due to wireless channel variations, packet transmission time H may follow different statistics. In Fig. 6, we compare the delay metrics when the packet transmission time is assumed to be deterministic or follow a geometric distribution. Here, the quantization step-size of the video traces is indexed at 28. The packet size is taken to be 9600 bits. Video flows are streamed over two possible access networks with a probability $\vartheta = 0.5$. The pre-roll delay is taken to be 2 seconds. In Fig. 6, we use different line styles to show the analytical results and different-shaped markers to highlight the simulation results over the analytical curves. As seen, the simulation results match well the analytical results even when the packet transmission time is randomly distributed. It is also interesting to notice that the delay performance is very close to that with a constant packet transmission time. There is a difference around $0.44\% - 6.56\%$, which is slightly larger when the channel rate is relatively low. When a highly varying channel results in complex statistics for the packet transmission time, a close approximation is achievable at a much lower computation complexity by assuming a constant packet transmission time.

B. Impact of Flow Splitting and Multipath Transmission

Fig. 7 compares the delay performance when a video flow is split into two substreams with $\vartheta_1 = \vartheta_2 = 0.5$ and the delay performance without flow splitting. Fig. 7(a) and Fig. 7(b) show the average packet delay and delay outage probability, respectively. As the variation trend of delay jitter is observed to be similar to that of the average packet delay, the results for delay jitter are not presented due to space limitation. It is known that a smaller burst arrival interval $N\tau$ indicates a higher traffic load. As B frames refer to both preceding and succeeding frames, B frames can only be encoded after the succeeding frames, which results in burst traffic arrivals for transmission. In our experiments, we consider a frame rate of $f = 30$ frames/s and $J = 7$ B frames between two key I/P frames. To ensure timely playback at the receiver, we limit the transmission interval by $(J+1)/f$. As seen in Fig. 7, flow splitting can better transmit a video flow and achieve a smaller average packet delay and delay outage probability. On the other hand, the delay performance converges when the traffic rate is much lower with a larger burst arrival interval. That is, flow splitting offers a larger performance gain in a heavy load situation.

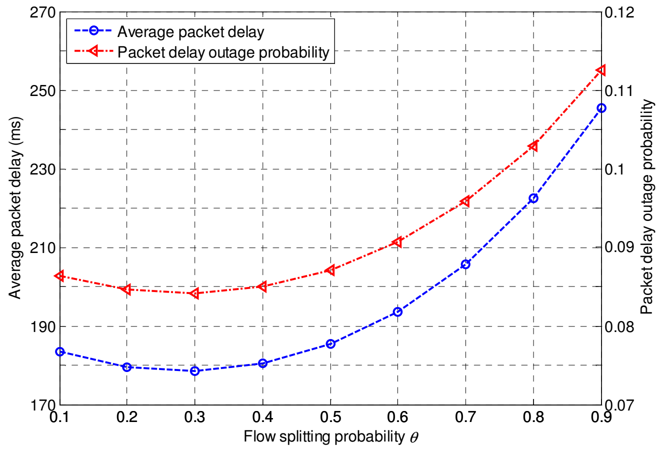


Fig. 8. Delay performance of multipath streaming over heterogeneous access networks.

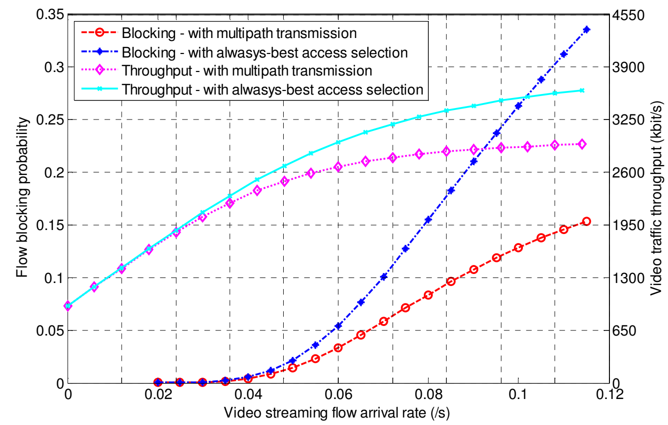


Fig. 9. Call-level performance in terms of flow blocking probability and system throughput for video streaming flows with multipath transmission and that with always-best access selection.

Fig. 8 shows the delay performance of multipath streaming over heterogeneous access networks to demonstrate the impact of flow splitting. Here, we consider video traces of a quantization step-size indexed at 22. The average video encoding rate is around 340.5 kbit/s. The packet size is taken to be 9600 bits. Based on our analytical approach, an effective channel rate of around 800 kbit/s is required to bound average delay and delay outage probability such that $\overline{T} \leq 200$ ms and $\xi \leq 0.1$. In Fig. 8, we show the delay performance of multipath streaming over two access networks of date rates 533 kbit/s and 640 kbit/s, respectively. The x-axis gives the splitting probability ϑ for the low-rate channel. Correspondingly, the probability to stream video bursts over the high-rate channel is $(1-\vartheta)$. As seen, there is an optimal ratio around 0.3 to split the video flow over the two access networks. A larger fraction around 0.7 of the video traffic is carried by the high-rate channel. The splitting probabilities should be properly selected to fully utilize available resources and avoid overloading in the meantime.

C. Call-Level Blocking Probability and Throughput

Performance gain with multipath transmission is also observed at the call level. In the following experiments, the resource occupancy threshold ϱ is taken to be 0.8. That is, when more than 80% of the subchannels in both access networks are being used, multipath transmission is allowed to split a video flow into two substreams. The two substreams require 2 and 3 subchannels, respectively, to satisfy the delay constraints. Without flow splitting, at least 4 subchannels are needed to accommodate a complete video flow. The total number of subchannels available in the two networks are assumed to be 24 and 30, respectively. The average streaming flow duration is set at 2.5 minutes [31]. With the Markov chain analysis in Section IV.D, Fig. 9 shows the flow blocking probability and system throughput of multipath transmission and that of the always-best access selection. As seen, the flow blocking probability is reduced by more than 50% with multipath transmission when the networks are heavily loaded. On the other hand, it is observed that the overall system throughput with multipath transmission is slightly lower than that of always-best access selection in a heavy load condition. To satisfy stringent delay constraints, the total bandwidth of two access networks to support two substreams can be larger than the minimum required bandwidth to carry a single complete flow. As a result, the overall system throughput on average is lower. However, the user perceived QoS such as packet delay is guaranteed, while a much lower flow blocking probability is achievable with multipath transmission. Depending on specific application requirements, we can further adjust the trade-off between delay and throughput by tuning the resource occupancy threshold for multipath transmission.

VI. Conclusions and Future Work

This paper introduces an analytical approach to evaluate the packet-level delay performance of video streaming traffic with multipath transmission. The analysis takes into account the essential characteristics of video traffic, such as deterministic burst intervals, highly varying burst length, and batch arrivals of transmission packets. To enable multipath transmission via multi-radio wireless devices, each video burst can be dispatched to an available wireless network according to a flow splitting probability. A resulting video substream is then characterized by a zero-inflated model. The PGF of packet transfer delay is derived to evaluate delay metrics such as average delay, delay jitter, and delay outage probability. Accordingly, we can obtain the minimum channel data rate to support a video flow/substream.

In order to minimize the processing overhead for multipath transmission, we consider a restricted admission policy, where flow splitting is allowed only when the resource occupancy exceeds a threshold. The call-level performance can then be analyzed with a Markov process. As seen from the numerical results, our approach offers an accurate evaluation of delay performance. By aggregating fractional bandwidth available in multiple networks, multipath transmission significantly reduces the flow blocking probability and achieves a high resource utilization. Nonetheless, the overall throughput degrades slightly in a heavy load condition to satisfy delay constraints.

In this work, we have studied a simple probabilistic flow splitting scheme for multipath transmission. Extending this work, we will investigate adaptive flow splitting and multipath transmission schemes. Especially, in a high-mobility condition, channel capacities may fluctuate rapidly due to severe fading, frequent switching of network attachment points, or adaptive modulation and coding. An efficient algorithm is needed to dynamically determine the flow splitting probabilities. The adaptation should minimize performance fluctuation while satisfying QoS constraints.

References

[1] F. Hou, L. Cai, X. Shen, and J. Huang, "Asynchronous multi-channel MAC design with difference-set-based hopping sequences," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1728–1739, 2011.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015," Feb. 2011.

[3] Allot, "Allot mobile trends: Global mobile broadband traffic report," Sept. 2010.

[4] P. Si, H. Ji, and F. R. Yu, "Optimal network selection in heterogeneous wireless multimedia networks," *ACM/Springer Wireless Networks*, vol. 16, no. 5, pp. 1277–1288, Aug. 2009.

[5] W. Song and W. Zhuang, "Multi-service load sharing for resource management in the cellular/WLAN integrated network," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 725–735, Feb. 2009.

[6] D. M. B. Masi, M. J. Fischer, and D. A. Garbin, "Video frame size distribution analysis," *Telecommunications Review*, vol. 19, pp. 74–86, Sept. 2008.

[7] L. X. Cai, X. Shen, J. W. Mark, and L. Cai, "Supporting voice and video applications over IEEE 802.11n WLANs," *ACM/Wireless Networks*, vol. 15, no. 4, pp. 443–454, 2009.

[8] F. Wan, L. Cai, and T. A. Gulliver, "A simple, two-level Markovian traffic model for IPTV video sources," in *Proc. IEEE GLOBECOM*, Nov./Dec. 2008.

[9] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, "Modeling full-length video using Markov-modulated Gamma-based framework," *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 638–649, Aug. 2003.

[10] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, no. 7, pp. 834–844, July 1988.

[11] M. Schwartz, *BroadBand Integrated Networks*. Prentice Hall, 1996.

[12] H. Bobarshad, M. van der Schaar, and M. R. Shikh-Bahaei, "A low-complexity analytical modeling for cross-layer adaptive error protection in video over WLAN," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 427–438, Aug. 2010.

[13] T. H. Luan, L. X. Cai, and X. Shen, "Impact of network dynamics on user's video quality: Analytical framework and QoS provision," *IEEE Trans. Multimedia*, vol. 12, no. 1, pp. 64–78, Jan. 2010.

[14] A. Abd El Al, T. Saadawi, and M. Lee, "LS-SCTP: A bandwidth aggregation technique for stream control transmission protocol," *Computer Communications*, vol. 27, no. 10, pp. 1012–1024, June 2004.

[15] T. Goff and D. S. Phatak, "Unified transport layer support for data striping and host mobility," *IEEE J. Select. Areas Commun.*, vol. 22, no. 4, pp. 737–746, May 2004.

[16] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks - Part I: The fluid model," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2568–2592, June 2006.

[17] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks - Part II: Constant-size packets," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5111–5116, Nov. 2006.

[18] J. Abouei, A. Bayesteh, and A. K. Khandani, "Delay-throughput analysis in decentralized single-hop wireless networks," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 1401–1405.

[19] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation with frame size and quality traces of single-layer and two-layer video: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 6, no. 3, pp. 58–78, Third Quarter 2004.

[20] A. Abdrabou and W. Zhuang, "Service time approximation in IEEE 802.11 single-hop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 305–313, Jan. 2008.

[21] L. D. Servi, "D/G/1 queues with vacations," *Operations Research*, vol. 34, no. 4, pp. 619–629, Jul./Aug. 1986.

[22] N. Jansakul and J. P. Hinde, "Score tests for extra-zero models in zero-inflated negative binomial models," *Communications in Statistics - Simulation and Computation*, vol. 38, no. 1, pp. 92–108, 2009.

[23] S. D. Conte and C. de Boor, *Elementary Numerical Analysis: An Algorithmic Approach*. McGraw-Hill, New York, 1972.

[24] P. J. Burke, "Delays in single-server queues with batch input," *Operations Research*, vol. 23, no. 4, pp. 830–833, Jul.-Aug. 1975.

[25] J. G. Apostolopoulos, W. Tan, and S. J. Wee, "Video streaming: Concepts, algorithms, and systems," *HP Laboratories*, Sep. 2002.

[26] C.-W. Yuen, W. C. Lau, and O.-C. Yue, "CAPEL: A packet discard policy for real-time traffic over wireless networks," in *Proc. IEEE ICC*, 2007, pp. 158–163.

[27] M. L. Esquível, "Probability generating functions for discrete real-valued random variables," *Theory of Probability and its Applications*, vol. 52, no. 1, pp. 40–57, July 2004.

[28] J. Abate and W. Whitt, "Numerical inversion of probability generating functions," *Operations Research Letters*, vol. 12, no. 3, pp. 245–251, Oct. 1992.

[29] W. Tu and W. Jia, "APB: An adaptive playback buffer scheme forwireless streaming media," *IEICE Transactions on Communications*, vol. E88.B, no. 10, pp. 4030–4039, Oct. 2005.

[30] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992.

[31] J. Catone, "Any way you slice it: YouTube dominates online video," http://www.sitepoint.com/, July 2008.

Wei Song (M'09) received her Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2007. Since 2008, she has been supported by the Natural Science and Engineering Research Council (NSERC) of Canada and worked as a postdoctoral research fellow at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. In July 2009, she joined the Faculty of Computer Science, University of New Brunswick, as an Assistant Professor. She received a Harrison McCain Foundation Young Scholars Award in 2010, a Top 10% Award from IEEE Workshop on Multimedia Signal Processing (MMSP) 2009, and a Best Paper Award from IEEE WCNC 2007. Her current research interests include the interworking of cellular networks and wireless local area networks (WLANs), resource allocation for heterogeneous wireless networks, cooperative wireless networking, and cross-layer design for multimedia service provisioning.

Weihua Zhuang (M'93-SM'01-F'08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair in Wireless Communication Networks. Her current research focuses on resource allocation and QoS provisioning in wireless networks. She is a co-recipient of the Best Paper Awards from the IEEE Multimedia Communications Technical Committee in 2011, IEEE Vehicular Technology Conference (VTC) Fall 2010, IEEE Wireless Communications and Networking Conference (WCNC) 2007 and 2010, IEEE International Conference on Communications (ICC) 2007, and the International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine) 2007 and 2008. She received the Outstanding Performance Award 4 times since 2005 from the University of Waterloo, and the Premier's Research Excellence Award in 2001 from the Ontario Government. Dr. Zhuang is the Editor-in-Chief of IEEE Transactions on Vehicular Technology, and the Technical Program Symposia Chair of the IEEE GLOBECOM 2011. She is a Fellow of the IEEE, a Fellow of the Canadian Academy of Engineering (CAE), a Fellow of the Engineering Institute of Canada (EIC), and an elected member in the Board of Governors of the IEEE Vehicular Technology Society. She was an IEEE Communications Society Distinguished Lecturer (2008-2011).