# MODELING AND MANAGEMENT OF TRAFFIC IN WIRELESS IP NETWORKS

**T. Janevski**

Faculty of Electrical Engineering, University "Sv. Kiril i Metodij"
Karpos 2 bb, P.O. box 574, 1000 Skopje, Macedonia
tonij@cerera.etf.ukim.edu.mk

**Abstract:** This paper presents traffic classification and proposes management framework for Quality of Service (QoS) provisioning in wireless IP networks. Classification of IP traffic is made into two main classes, A and B, where class-A is targeted to traffic with QoS support, and class-B is best-effort traffic. Class-A is further divided into three subclasses, which are targeted to conversational, streaming and interactive services, respectively. Upon the results of traffic analysis, the paper proposes a traffic management framework, including packet differentiation, scheduling and admission control in wireless IP access networks.

**Keywords:** Wireless Networking, Quality of Service, Traffic, Modeling, Management

## 1 Introduction

Nowadays we are facing a transition between the second-generation (2G) mobile networks based primarily on the voice service, and third-generation (3G) mobile networks that are targeted to introduction of new services and packet-based communication besides the traditional circuit switching. There are different standards created for 3G, such as Universal Mobile Telecommunication System (UMTS) in Europe and CDMA2000 in Americas. UMTS has defined two main access techniques for the air interface, i.e. CDMA-TD using combination of CDMA and TDMA, and WCDMA as a combination of CDMA, FDMA and TDMA. On the other side CDMA2000 is created as an upgrade of 2G mobile systems called IS-95, but with three times wider spectrum. Also, wireless local area networks (WLANs) are gaining momentum via the popularity IEEE 802.11b WLAN. While 3G mobile networks are capable to provide data rates up to 2 Mbps for users with low mobility, current IEEE 802.11b WLAN provides 11 Mbps but without any QoS support.

Also, other wireless technologies are in the development phase, such as Wireless Personal Area Networks (WPAN), Broadband Radio Access Networks (BRAN) etc. The main conclusion is that we have many different wireless access technologies. The future wireless cellular networks (e.g., 4G, NextG) should be unified wireless IP with heterogeneous access and multimode terminals capable to transit from one network to other anytime, anywhere, and for any service [1]. In such case we need programmable

network nodes that can adapt to different traffic classes and access technologies. In this paper we propose a management framework for wireless IP networks.

The paper is organized as follows. In next section we describe statistical characteristics of the most common traffic types in future wireless networks. Traffic management model is presented in section 3. Finally, section 4 concludes the paper.

## 2 Statistical Analysis and Modeling of IP Traffic

In wireless IP networks we may have different traffic types, such as voice traffic, video and audio streaming, web-browsing, gaming, chat, e-mail, file transfer etc. While in second-generation mobile systems we have only single traffic class and circuit-switched service (i.e. voice service), in third generation mobile systems are defined four different traffic classes considering the QoS requirements, and they are: conversational class, streaming class, interactive class and background class [2]. Examples of conversational class are: voice, videophone; typical streaming services are audio, video and multimedia streaming; World Wide Web (WWW), e-mail, messaging, belong to interactive class; and background class is planned to be used for Short Message Service (SMS), fax, inter-server communication etc. First two groups of classes, conversational and streaming, can be classified into real-time services, and the last two, interactive and background, can be referred to as non-real-time services. Third generation mobile systems support circuit-switched services (e.g., voice) and packet-based services. Mobile networks beyond 3G (e.g., 4G) are expected to be all-IP networks (i.e. end-to-end IP) that should provide higher bandwidth and new killer services, such as personal services (e.g., location-based services, on-line shopping), business services (e.g., on-line stock trading), entertainment services etc. Hence, we refer to wireless IP networks in our analysis and design approach in this paper. However, statistical properties of IP traffic differ from traditional voice traffic found in previous generations cellular networks (e.g., 1G, 2G).

Different services have different requirements on QoS parameters, such as loss, delay and jitter, as well as different demands for data rates. For example, real-time services can tolerate some Frame Error Ratio (FER, where FER refers to frames transmitted in the wireless channel that experience bit errors), while non-real-time services are not tolerable to data loss. For example, in [2] is specified FER<3% for voice in 3G mobile systems, and FER<1% for videophone and streaming services. To provide no-loss for non-real-time applications we usually use retransmissions and error correction techniques, as well as their combinations. On the other side, requirements on delay are opposite, i.e. real-time services have stringent demands on transport delay, while non-real-time services are more tolerable to delay. For example, recommended on-way delay for voice communication is 150 ms, but also one can tolerate delays up to 400 ms. Streaming services can experience longer delays up to 10 seconds due to the unidirectional character of the communications and possibility for buffering of data at the receiving end. But, interactive services have stringent demands on delay that should be kept within the range of 2-4 seconds. Considering the jitter (i.e. delay

variations), only conversational services are sensitive to jitter (the jitter should be less than 1 ms for voice), but other services are not.

In today's telecommunications networks voice is still the main service. On the other side, in Internet (wired or wireless) the largest traffic share goes to WWW traffic, which accounts for ¾ of all bytes in Internet [3]. Video traffic (e.g., video streaming) is a service with highest bandwidth demands. Hence, we choose to analyze these three traffic types as the most representative ones.

Considering the statistical characteristics there is difference between the voice service on one side, and video streaming and WWW on the other. In the following subsections we analyze statistical properties of each of these three services, and model each of them.

### 2.1 Voice Traffic

Let first consider the IP telephony in wireless IP networks. Voice traffic is well described by Poisson call arrivals and exponentially distributed call holding times. Hence, we assume Poisson arrival processes in a given cell. With $\lambda_n$ and $\lambda_h$ we denote new call arrival rate and handover arrival rate in the cell. An ongoing voice connection completes at rate $\mu_c$ or departs the cell at rate $\mu_h$. We do use statistical multiplexing of voice sources and other non-voice traffic over the wireless link. If we denote with $\lambda$ the total call arrival rate, then:

$$\lambda = \lambda_n + \lambda_h \qquad (1)$$

Similarly, we denote with $\mu$ the total call departure rate:

$$\mu = \mu_c + \mu_h \qquad (2)$$

Then we obtain a birth-death process as shown in Fig. 1, where $N$ is the maximum number of active users in the cell. So, a user may be an active voice source or idle.

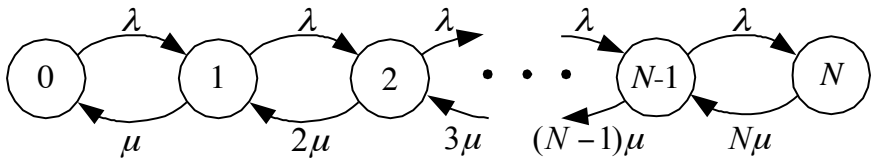


Fig. 1: Markov chain model

It is already a practice to model packet-based telephony by two-state Markov chain with one state representing the talk spurt (ON) and other state representing the silent period (OFF). During ON (talk) periods the source is transmitting IP packets back to back. Most encoding schemes have fixed bit rate and fixed packetization delay. During off (silence) periods the source sends no packets. We assume that ON and OFF periods are exponentially distributed, which is well analyzed in [4]. The voice sources can be viewed as two state birth-death processes with birth rate $\alpha_{on}$ (arrival rate for on-periods) and death rate $\alpha_{off}$ (ending rate for on-periods). Then, $1/\alpha_{on}$ and $1/\alpha_{off}$ are durations of talk period and silent-period of a voice source, respectively. The talk

spurt has a constant bit rate that we denote as $b_{ts}$. Then, average voice connection bit rate is:

$$b_{av} = \frac{T_{on}}{T_{on} + T_{off}} b_{ts} = \frac{\alpha_{off}}{\alpha_{on} + \alpha_{off}} b_{ts} \qquad (3)$$

The superposition of the voice sources can be also viewed as a birth-death process, where total incoming rate is sum of incoming rates of individual sources. A convenient model in teletraffic theory for a superposition of many ON-OFF voice sources is Markov Modulated Poisson Process (MMPP). For voice sources with talk spurts and silent periods (without packets on link) it is more convenient to use the special case of MMPP, i.e. IPP (Interrupted Poisson Process), which is in fact special case of Coxian distributions (i.e. Cox-2). By a definition, state *j* means *j* sources are ON.

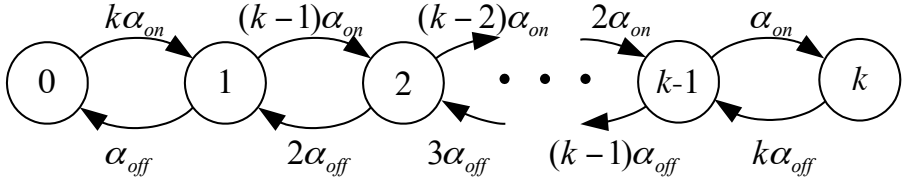


Fig. 2: State transition diagram for a superposition of *k* voice sources

In Fig. 2 we assume that all *k* connections are previously established. Now, we may link this chain with Markov chain for calls given in Fig. 1. To provide desired Quality of Service (QoS) for the voice calls we need to determine maximum number of active voice calls at a time. This number is denoted with *N* in Fig. 1. So, at a given moment of time we may have *k* voice connections, where $k \leq N$. If we have *k* active connections then we have the situation given in Fig. 2 considering ON and OFF periods of sources. IP packets arrive from *k* input lines/sources. Each of the *k* calls occupying the input lines i.e. wireless link alternates between talk spurts and silent periods. Assuming state-equilibrium in Fig. 2, we may derive the distribution of active voice sources, i.e. the probability that *j* sources are ON out of *k* active voice connections is:

$$P_j = \binom{k}{j} \left(\frac{\alpha_{off}}{\alpha_{on}}\right)^j \Bigg/ \left(1 + \frac{\alpha_{off}}{\alpha_{on}}\right)^k \qquad (4)$$

## 2.2 Video Traffic

Video traffic is likely to be also very popular type of application. However, it has different statistical characteristics compared to voice traffic.

Video coders use spatial and temporal coding to remove redundancy information within video frames. Most of video services in Internet are on-demand. Typical example is MPEG-4 standard, which support flexible video coding to adapt to error-prone environments or transmission channels with lower data rates. Hence, MPEG-4 is well suited to be used in a wireless network, which is characterized with higher bit

error ratio than wired networks. However, we should mention MPEG-1 and MPEG-2 standards that are targeted to local video retrieval (on CD-ROM, hard disk), and digital video broadcast (e.g., digital TV) or retrieval video (e.g., DVD), respectively.

Due to periodicity of video frames (frame rate is 25 Hz for PAL, and 30 Hz for NTSC), due to different contents from different scenes, and due to statistical coding, video traffic shows self-similar properties, such as long tailed autocorrelation and slow-decay variance [1]. There has a considerable effort from research community to find appropriate models for this traffic type [1]. Such models are MMPP, Auto Regressive (AR) models, Pareto models, Fractional Brownian Motion, etc. The simplest model to capture self-similar properties of the video is the Pareto model, which is presented in the following subsection.

### 2.3 Web Traffic

Web traffic is the dominant type of IP traffic nowadays. But, similar to video streaming, it shows self-similar characteristics. While session arrivals can be modeled by Poisson process, length of packets bursts exhibits long tails (low decaying autocorrelation). For modeling the size of a packet bursts we may use Pareto distribution due to its characteristic to have long tails as packet call sizes. The classical Pareto distribution with shape parameter $\alpha$ and location parameter $k$ has the probability density function [5]:

$$f_x(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}},\ x \geq k \qquad (5)$$

In [6] is mapped the $k$ parameter onto the $H$ (Hurst) parameter as follows:

$$k = 3 - 2H \qquad (6)$$

Because the range of $H \in [0.5, 1)$, it follows that $k \in (1, 2]$. For WWW traffic we use as a default value $k$=1.1, what leads to $H = (3-k)/2$=0.95. However, $H$ parameter of a single WWW session may vary over the whole range [0.5, 1). Therefore, we cannot dimension a network with stringent QoS guarantees for non-real-time services such as WWW, but we may provide a minimum guarantees, if the user demands them. So, we need information about the average packet call sizes over many WWW sessions during the highest network load. Then, we may guarantee the user minimum QoS for WWW services, or the network may reject the WWW call with QoS demand, and instead offer a best effort service to the user. Of course, different pricing schemes should be applied for each traffic class and service.

## 3 Traffic Management in Wireless IP Networks

From the statistical analysis it is straightforward that dimensioning and management of voice-based networks is easier due to low correlation of voice traffic and possibility to describe the traffic with a single parameter, and that is intensity (e.g., arrival call/packet intensity, departure call/packet intensity). On the other side, TCP-based traffic (e.g., WWW) and video streaming show self-similar properties. Because $H$

parameter of WWW and video sequences is close to 1, it is certain that we cannot apply closed-form analytical analysis for these traffic types as we did for voice in the past. Heavy-tailed autocorrelation of web and video traffic demands differentiation of voice service from these services (e.g., using Differentiated Services by exploiting the ToS/DS field in IPv4/IPv6 packet's headers). Following this discussion, we propose classification of IP traffic into two main classes [7]: class-A for traffic with QoS support; and class-B for traffic without any QoS guarantees i.e. best-effort traffic. Furthermore, class-A is divided into three subclasses, and they are: A1 for traffic with stringent QoS demands (e.g., voice), A2 for real-time traffic with variable bandwidth requirements (e.g., video/audio streaming), and A3 for non-real-time traffic with minimum QoS guarantees (e.g., web browsing). Four QoS classes defined for 3G mobile networks, i.e.: conversational, streaming, interactive and background; can be easily mapped on A and B traffic classes and subclasses. However, our traffic classification is not limited to 3G networks or existing services.

Wireless networks have different characteristics than wired networks, i.e.: 1) Mobility of the users, and 2) Location-dependent and time-variable bit errors in the wireless channel due to fading, shadowing etc. These unique characteristics of wireless networks should be dealt with to provide certain QoS.

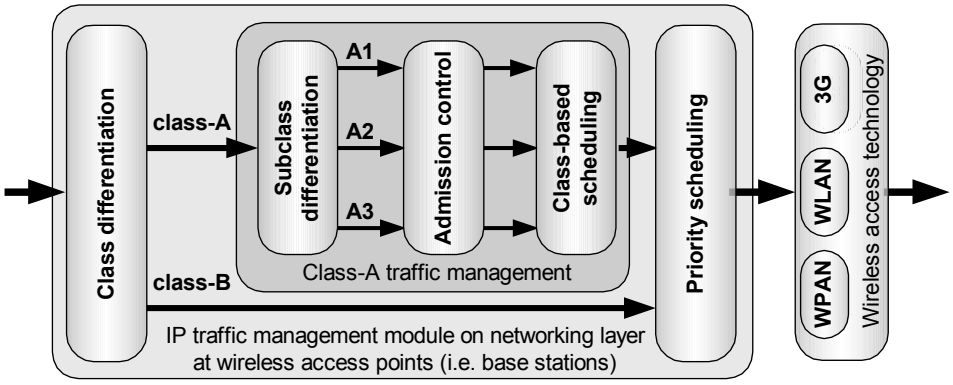


Fig.3: IP traffic management model for wireless access points in wireless IP networks

Traffic classification and wireless network characteristics are incorporated into the traffic management model presented in Fig. 3. In this model first differentiation module separates traffic flows with and without QoS support. The later is the best-effort traffic, i.e. class-B, which does not have any QoS demands. So, further traffic management is targeted to class-A. Besides differentiation of this traffic into separate queues for each subclass, we need to perform admission control and class-based scheduling as well. The admission control is targeted to provide QoS upon given constraints (e.g. packet delay, losses, jitter). Hence, A1 and A2 subclasses that are targeted to real-time services are serviced with higher priority compared to A3-subclass. However, to provide minimum QoS guarantees (on packet delay) for A3 flows we reserve small part of the bandwidth for this type of traffic only [8], with aim to avoid

monopolization of the wireless link by higher priority subclasses. The amount of the bandwidth reserved for A3 traffic should be a programmable parameter that is dependent upon network operator's policy. Furthermore, A1 traffic is targeted to conversational services (e.g., voice, videoconferencing etc.), which are very sensitive to delays, while A2 traffic is targeted to streaming services (e.g., video, audio or multimedia streaming) that are less sensitive to delays. In this case we may apply priority to A1-class with some bandwidth reservations for A2 class (to avoid link monopolization by A1 flows) or we can use some of many types of wireless fair queuing. However, within each subclass flows are scheduled using some of the wireless weighted fair queuing mechanisms [1] (e.g. Wireless Fair Service).

As shown in Fig. 3 admission control is applied to all class-A traffic. In reality, admission control must be applied to A1 subclass. In this case maximum number of flows admitted into a cell can be calculated using Eq. (4). Further, for A2 and A3 traffic (e.g., video streaming and web traffic, respectively) we cannot obtain closed-form analytical tool for network dimensioning, due to self-similar nature of these traffic types (refer to subsections 2.2 and 2.3 in this paper). There are two options for them: 1) to avoid admission control for A2 and A3 traffic, and to leave adaptation to bandwidth fluctuations to higher level protocols at the end-peers of communication session; or, 2) to apply admission control where average buffer occupancy at the base stations will be used as a parameter. However, one may choose to apply admission control to A2 traffic (besides the A1) and not to A3 traffic.

Finally, class-B packets are served when all class-A queues are drained out (i.e. priority scheduling towards wireless link, as shown in Fig. 3). Fig. 3 shows the traffic management model in downlink direction, towards mobile terminals. The traffic in IP networks is asymmetrical for all services except for the conversational ones, and hence the downlink carries more traffic. In the uplink direction, the packets are scheduled at the mobile terminals, but even in this case virtual queues can be managed at the base stations using the same traffic management model.

## 4 Conclusions

In this paper we discussed modeling and management for wireless IP networks. We presented models for three main traffic types today: voice, video streaming and web browsing. It was concluded that voice over IP traffic is well described using the MMPP process, while video and web exhibit self-similar properties and therefore cannot be described in closed-form analytical relations. These properties of IP traffic led to definition of two main traffic classes: class-A for traffic with QoS support; and class-B for traffic without any QoS guarantees. However, such heterogeneous traffic requires appropriate traffic management. Hence, we proposed a framework for management of IP traffic at wireless access points. It considers differentiation, scheduling, and admission control. Admission control is applied only to class-A traffic because only this class has QoS constraints. Additionally, admission control is necessary only for A1 traffic (i.e. voice), while it is optional for streaming and interactive services.

This traffic management model can be applied separately in both, downlink and uplink direction, in wireless IP networks.

## 5 References

1. T. Janevski, *Traffic Analysis and Design of Wireless IP Networks*, Artech House Inc., Boston, USA, 2003
2. 3GPP TS 23.107, "Technical Specification Group Services and System Aspects; QoS Concept and Architecture (Release 5)", V5.3.0, January 2002.
3. K. Thompson, G.J. Miller, R. Wilder, "Wide-area Internet Traffic Patterns and Characteristics", *IEEE Network*, November/December 1997.
4. B. Ahlgren at el., "Dimensioning Links for IP Telephony", SICS, *CNA Laboratory*, Sweden.
5. 3GPP TR 25.881, "Improvement of RRM across RNS and RNS/BSS (Release 5)", V5.0.0, *3GPP Proposed Technical Report*, 12-2001.
6. F. Huebner, D. Liu and J.M. Fernandez, "Queuing Performance Comparison of Traffic Models for Internet Traffic", *GLOBECOM '98*, pp.1931-1936, Sydney, Australia, November 8-12, 1998.
7. T. Janevski, B. Spasenovski, "QoS Provisioning for Wireless IP Networks with Multiple Classes Through Flexible Fair Queuing", *GLOBECOM 2000*, San Francisco, USA, November 27-December 1, 2000.
8. T. Janevski, B. Spasenovski, "Admission Control for QoS Provisioning in Wireless IP Networks", *European Wireless 2002*, Florence, Italy, February 2002.