A Dynamic Rate Allocation Technique for Wireless Communication Systems

Lorenzo Caponi, Francesco Chiti, Romano Fantacci

Dipartimento di Elettronica e Telecomunicazioni - Università di Firenze

via di S. Marta 3, I-50139 Firenze, Italy

tel. : +390554796271 - fax : +390554796485

e-mail: {caponi,chiti,fantacci}@lenst.det.unifi.it

Abstract— In this paper, a Medium Access Control (MAC) technique for wireless systems that efficiently manages Adaptive Modulation and Coding (AMC) schemes is presented with the aim of improving the downstream data rate, together with preserving Quality of Service (QoS) requirements. Basically, the proposed access scheme relies on physical channel monitoring and state mapping into a variable rate packet format. A suitable analytical model is proposed herein in order to predict the system performance, i.e., the average packet queuing delay or, equivalently, the average queue length. In particular, the non-stationary transmission channel is modeled as a Markov chain and each state is associated with a particular Modulation and Coding Scheme (MCS) for a specified data rate. Comparisons between analytical predictions and simulation results are provided to validate our model and to highlight the improvement of the proposed technique with respect to a static rate allocation scheme, especially in the presence of heavy tailed traffic conditions, like those arising in remote file transfer typical of a Web connection.

I. INTRODUCTION

The growing demand for integrated data services compels telecommunication networks to be designed and managed according to extremely efficient policies. As to typical third generation (3G) networks Mobile Subscriber (MS) needs, it is worth mentioning a complete access to web applications such as HTTP, SMTP, FTP protocols, as explained in [1]. This can be accomplished by means of an efficient internetworking with IPvx backbones and by providing a different data rate with regard to the user equipment capabilities and the communication channel conditions. In particular, UMTS services will be available with a data rate up to 2 Mbps in an indoor environment.

Because of the increasing market interests in offering services to mobile users within a wireless indoor scenario, as an open space or hot spot neighborhood, UMTS standard is expected to continuously update its access schemes in order to satisfy the future requests for packet data services and to compete with already existing wireless networks standards, as IEEE 802.11x. Moreover, the asymmetric data traffics feature leads to a particular optimization in the downlink. As a consequence, several enhanced access schemes, have been proposed and are currently crucial items under investigation within IMT2000, as 1xEVDO for cdma2000 standard [2] and High Speed Downlink Packet Access (HSDPA)for 3th Generation Partnership Project (3GPP), within Release 5 [3].

The main goal of HSDPA is to allow Wideband Code Division Multiple Access (WCDMA) technology to support downlink peak data rate up to approximately 8-10 Mbps for best effort packet data services, that is considerably beyond the basic capacity of 2 Mbps and close to the typical wired LAN data rate. Furthermore, HSDPA should also enhance the quality of the services provided in terms of lower delay and improved capacity.

The main approaches currently investigated in order to efficiently allow UMTS to supply such a services are the following [4]:

- **Adaptive Modulation and Coding** (AMC) schemes in which the modulation level and the code rate could be selected on a dynamic basis according to the Signal to Interference plus Noise Ratio (SINR) measurement or estimation, in cooperation with the mobile terminals.
- **Hybrid Automatic Repeat reQuest** (H-ARQ) algorithms according to which the excessive discarding of the erroneous packets can be avoided and successive retransmission of the same informative packet can be combined in order to reduce the packet error rate.
- **Fast Scheduling** (FS) techniques that dynamically select the user with the best channel and network conditions to maximize communication throughput.
- **Fast Cell Selection** (FCS) algorithms which dynamically assign to an active downlink user the cell with the best Signal to Interference Ratio (SIR), thus compensating the fast fading and interference effects.
- **Multiple Input Multiple Output** (MIMO) architectures where the presence of multiple antenna at both the transmitter and the receiver ends is considered to increase data throughput.

This paper deals with HSDPA schemes based on the AMC principle, thus adapting the main transmission parameters (modulation order and code rate) to channel conditions in order to achieve the maximum channel capacity.

In particular, according to the proposed scheme the channel monitoring is performed at the beginning of each slot, named Transmission Time Interval (TTI), and, hence, the best transmission format - i.e., Modulation and Coding Scheme (MCS) - is selected in order to optimize the system performance, as

suggested in [5]. In particular, this implies a packet transmission rate variability slot by slot in order to optimize the network throughput.

It is shown in the following that the proposed scheme outperforms the classical scheme where a static rate allocation is adopted, especially for high loaded links and bursty traffics. The achieved gain allows to maximize the link-to-link throughput together with lowering queuing delay or increasing the number of active connections.

II. PHYSICAL CHANNEL MODEL

The physical channel has been modeled in our analysis as a time discrete two state Markov chain, commonly referred in the literature as the Gilbert Elliot model [6], [7]. According to our assumptions, the states are labeled as GOOD and BAD. This model relies on the fact that in a common Wireless Packet Access Network (WPAN) the radio communications have to face an air interface characterized by reflection, refraction and scattering phenomena. As a result, a multipath propagation is commonly involved and each signal replica is affected by a random attenuation within each path, leading to the so called multipath fading propagation conditions. Moreover, the characteristics of this transmission channel are typically time varying, because of the relative mobility between end points.

By taking into account that a specified Quality of Service (QoS) for data communications is usually related to a minimum value for the received power, the channel is assumed herein to be in the BAD state whenever the power of the received signal is below this threshold; on the contrary, a GOOD state is assumed. In the former case, the most efficient Signal to Noise ratio (SNR) modulation scheme has to be used to lower the data throughput degradation. Conversely, in the other case a more efficient bandwidth efficient modulation scheme and/or a lower code rate are chosen in order to increase the data throughput. Time is assumed to be divided into slots, whose duration is τ. The channel state transitions are considered to occur, only at the end of a slot. In the following, we denote as $r_{i,j}$ the transition probability from state i to state j, where BAD state is named 0 state, whilst GOOD state is 1 state [1].

III. PROPOSED ADAPTIVE MAC PROTOCOL

The proposed scheme belongs to the class of the Medium Access Control (MAC) sub-layer protocols suitable for a generic radio mobile, though in our analysis we focus on an UMTS access network with particular regard to the HSDPA scheme.

Being the considered traffic streams extremely asymmetric, our investigation is focused on the downlink communications for which there is a higher need for an access optimization. We assume in our analysis that the propagation time is negligible with respect to the slot duration and the packet arrivals are considered to occur only at the beginning of a slot. Further,

we assume that a packet is always correctly received even if it contains errors, being the service a real time.

The proposed MAC protocol can be described as it follows. The data traffics is composed of variable size messages that are stored in the NodeB shared buffer, whose size is assumed as infinite to not consider the dropping effect. At the beginning of a slot, packets are selected for transmission according to a First In First Out (FIFO) policy. The modulation and coding schemes to be used are chosen on the basis of knowledge of the channel state, i.e., the propagation conditions. A typical approach to reach this purpose is to periodically monitor each user pilot signal. In our analysis we have assumed that a data rate equal to one packet per slot is allowed within the BAD state, while, whenever the channel is detected to be in the GOOD state, two packets per slot are transmitted by means of an appropriate AMC scheme. This policy is made feasible by means of the MCS variation at the physical layer that implies a rate variation, while maintaining the same QoS requirements in terms of *residual error rate*. This means that the first MCS is adopted under bad propagation conditions, while the other MCS is used when the transmission channel experiments good propagation conditions. Finally, no priority within the data traffics as well as no mobility management for the users have been considered, being the cell isolated.

In the next section, a suitable analytical model for the proposed MAC scheme is presented in order to derive the average queued packet number $\overline{N}$ or, equivalently, the average packet delay $\overline{T}$, the latter being a widely adopted QoS parameter for a communication network. The predicted system performance has to be compared with the basic static scheduling algorithms, properly modeled as a G/D/1 system, since a constant bit rate (assumed equal to one packet per slot) is provided, always relating to the channel worst state.

IV. SYSTEM MODEL

Our investigation begins by defining the system under consideration as a vectorial state (j, k) in which j is the state of the channel at the beginning of a slot, hereinafter named *phase*, according to [8] and k is the number of queued packets, both considered at the beginning of a slot. Whenever the system state transitions are observed only at service completion time a discrete time Markov model can be derived. In particular, vectorial state transitions are described by a Markov chain as shown in Fig. 1 where state transitions occur at the end of each slot.

It is worth noticing that, whenever the channel is BAD, the state assume the form $(0, k)$, thus only one packet can be sent out. Conversely, if the channel is GOOD at the beginning of a slot, the system state is $(1, k)$ and, therefore, two packets can be transmitted.

Let a_k be the probability of k packets arrival at the NodeB buffer within a slot. Then, the steady state probability $p_{j,k}$, representing the probability of having the system in the state (j, k) at the beginning of a slot, can be defined [9]:
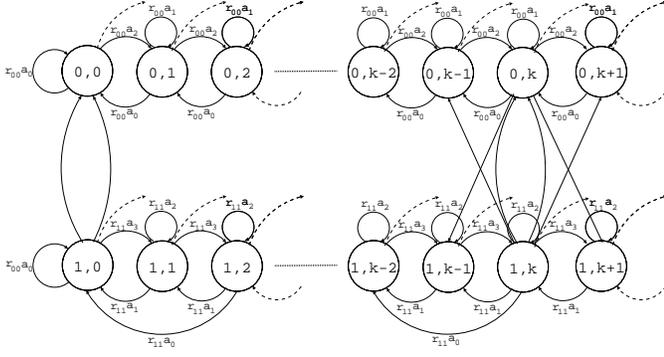
---

[1]Although the channel state transitions continuously occur in time, our simplified assumption has no severe impact on the accuracy of our analysis, by assuming τ lowser then the channel coherence time.

Fig. 1. Markov chain model

$$p_{0,k} = r_{0,0} \left[\sum_{i=0}^{k} a_i p_{0,k+1-i} + a_k p_{0,0} \right] +$$
$$+ r_{1,0} \left[\sum_{i=0}^{k} a_i p_{1,k+2-i} + a_k p_{1,1} + a_k p_{1,0} \right] \quad (1)$$

$$p_{1,k} = r_{1,1} \left[\sum_{i=0}^{k} a_i p_{1,k+2-i} + a_k p_{1,1} + a_k p_{1,0} \right] +$$
$$+ r_{0,1} \left[\sum_{i=0}^{k} a_i p_{0,k+1-i} + a_k p_{0,0} \right] \quad (2)$$

Let us introduce the arrival probabilities matrix $\mathcal{A}$, whose generic element $\mathcal{A}_{i,j}$ represents the probability of having j arrivals within a slot and the vectors π_0, π_1 whose element are $p_{0,k}$ and $p_{1,k}$, respectively. In order to evaluate the steady state probability $p_{j,k}$ by means of a linear system solution, a finite m size buffer has to be taken into account. Under this assumption, vectors π_0 and π_1 become:

$$\pi_0 = \begin{bmatrix} p_{0,0} & p_{0,1} & \dots & p_{0,m} \end{bmatrix} \quad (3)$$

$$\pi_1 = \begin{bmatrix} p_{0,1} & p_{1,1} & \dots & p_{1,m} \end{bmatrix} \quad (4)$$

and matrix $\mathcal{A}$ has the following expression:

$$\mathcal{A} = \begin{bmatrix} r_{0,0}\mathcal{A}_0 & r_{0,1}\mathcal{A}_0 \\ r_{1,0}\mathcal{A}_1 & r_{1,1}\mathcal{A}_1 \end{bmatrix} \quad (5)$$

where:

$$\mathcal{A}_0 \doteq \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & (1 - \sum_{k=0}^{m-1} a_k) \\ a_0 & a_1 & a_2 & \cdots & (1 - \sum_{k=0}^{m-1} a_k) \\ 0 & a_0 & a_1 & \cdots & (1 - \sum_{k=0}^{m-2} a_k) \\ 0 & 0 & a_0 & \cdots & (1 - \sum_{k=0}^{m-3} a_k) \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & a_0 & (1 - a_0) \end{bmatrix} \quad (6)$$

and

$$\mathcal{A}_1 \doteq \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & (1 - \sum_{k=0}^{m-1} a_k) \\ a_0 & a_1 & a_2 & \cdots & \cdots & (1 - \sum_{k=0}^{m-1} a_k) \\ a_0 & a_1 & a_2 & \cdots & \cdots & (1 - \sum_{k=0}^{m-1} a_k) \\ 0 & a_0 & a_1 & \cdots & \cdots & (1 - \sum_{k=0}^{m-2} a_k) \\ 0 & 0 & a_0 & \cdots & \cdots & (1 - \sum_{k=0}^{m-3} a_k) \\ \vdots & \vdots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & a_0 & a_1 & (1 - a_0 - a_1) \end{bmatrix} . \quad (7)$$

In order to evaluate each $p_{j,k}$ and, consequently, $\overline{N}$ we have to solve the following linear system:

$$\pi\mathcal{A} = \pi \quad (8)$$

where $\pi \doteq [\pi_0, \pi_1]$, by introducing the normalization constraint, that is, $\pi e = 1$. Once the generic steady state probabilities $p_{j,k}$ have been derived, the average queued packets number $\overline{N}$ can be defined as:

$$\overline{N} = \sum_{j=0}^{1} \sum_{k=0}^{\infty} k\, p_{j,k}. \quad (9)$$

Therefore, by applying the Little formula, it is possible to derive $\overline{T}$, according to [10], as::

$$\overline{T} = \frac{\overline{N}}{\lambda} + \frac{\tau}{2}\left(p_{0,0} + p_{1,0}\right). \quad (10)$$

V. TRAFFIC MODELS

A complete solution of (8) can be obtained once the arrival process statistics are known. In this paper, three different input traffic distributions are taken into account, among these one is close to bursty characteristics of data traffic following heavy tailed distribution of the packet size. This choice can be motivated by the circumstance that HSDPA schemes have to be optimized with respects to applications particularly involved with bursty packet arrivals [11], like Web applications, data base queries and files downloading.

A. Poisson arrivals

The first type of traffic model is characterized by classical Poisson arrivals distribution. It is well known that the generic term a_k results to be:

$$a_k = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau} \quad (11)$$

where λ represents the average number of packet arrivals per time unit. This distribution is well-suited mainly for a memoryless traffic such as voice communications, as assumed in [12].

B. Batch arrivals: modified geometric message length

A more accurate statistical model [11] is represented by a compound Poisson arrivals process according to which the data traffic is characterized by Poisson message arrivals, whilst each message is composed by a modified geometric distributed number of packets. In this case we have:

$$A(z) \doteq \sum_{k=0}^{\infty} a_k z^k = e^{-\lambda\tau[1-M(z)]} \qquad (12)$$

where $M(z)$ is the generating function of message length such that $M(z) \doteq \sum_{i=0}^{\infty} m_i z^i$. By relying on the previous assumptions, we have:

$$M(z) = \frac{(1-p)z}{1-pz} \qquad (13)$$

where $(1-p)^{-1}$ is the average message length in packet. In this case, we can express (12) by means of (13) as:

$$A(z) = e^{\lambda\tau\left(\frac{z-1}{1-pz}\right)} \qquad (14)$$

from which the expression of each a_k can be derived as:

$$a_k = \sum_{s=1}^{k} \left[e^{-\lambda\tau} \frac{(\lambda\tau)^s}{s!} (1-p)^s p^{k-s} \frac{1}{(s-1)!} \prod_{j=1}^{s-1} (k-s+j) \right]. \qquad (15)$$

C. Batch arrivals: Pareto message length

A more accurate model that approximate a realistic data communication assumes heavy tailed distribution as to the message length. In particular, this can be described by means of a truncated Pareto distribution, as it happens whenever the file byte length within a web page is considered, [11]. In particular, the truncated Pareto probability density function (pdf) is defined as:

$$f_{\boldsymbol{x}}(x) \doteq \begin{cases} \frac{\alpha k^\alpha}{x^{\alpha+1}}, & k \le x < m \\ \frac{k^\alpha}{m^\alpha}, & x = m \\ 0, & otherwise \end{cases} \qquad (16)$$

where k and m represent, respectively, the minimum and the maximum file sizes and α is the so-called *shape* parameter.

An analytical expression for a_k can be derived by means of (12), by approximating each m_i as:

$$m_i \approx F_{\boldsymbol{x}}(x_i) - F_{\boldsymbol{x}}(x_{i-1}) \qquad (17)$$

where $F_{\boldsymbol{x}}(x)$ represents the truncated Pareto cumulative density function (cdf) and x_i is multiple of an elementary packet length assumed equal to 1500 bytes, i.e., the maximum IP packet size.

VI. NUMERICAL RESULTS

In this section, numerical results are shown as far as the analytical predictions and practical simulations of $\overline{N}$ and $\overline{T}$ values. These values are compared with the static transmission technique performance, properly modeled as s M/D/1 system since only one packet per slot is transmitted independently of the channel state. The channel model assumed has been described in II, whilst the traffic models have been introduced in V. The slot (TTI) duration is assumed equal to 2 ms, as in [4]. In order to highlight the protocol performance, we focus on the transmission channel worst case in which the channel state transition probabilities $r_{0,1}$ and $r_{1,0}$ are assumed equal and, in particular, $r_{0,1} = r_{1,0} = 0, 2$. As a matter of fact, in this case the channel is affected by a fast fading and the system is mostly stressed, thus implying the maximum protocol overhead.

In Figs. 2 and 3, the analytical and simulated results, as to average queue length $\overline{N}$ and average queuing time $\overline{T}$ are shown for the case of batch Poisson arrivals process with a Pareto message length distribution, whose parameters are $k = 1858$ byte, $m = 5,000,000$ byte, $\alpha = 1.1$, according to [11]. These results are further compared with the performance of the basic M/D/1 system. A remarkable performance enhancement is to be highlighted in terms of better mean queue length or packet delay obtained by the proposed high-speed MAC protocol with respect to basic techniques. This advantage can be, alternatively, spent in order to enhance the network capacity in terms of the number of the effectively served users. Moreover, the previous figures highlight a perfect accordance between analytical and numerical values, thus validating our analytical approach.

Even if the proposed MAC protocol obtains a performance improvement whatever the input traffic, a comparison of the relative throughput gains with respect to the three traffic models proposed in section V is needed. In particular, it is further assumed that the average message length is equal to 25, as stated in [11]. This comparison is performed in Fig. 4, where a performance improvement is highlighted as the traffic burstiness grows. In particular, a throughput gain up to 50% is pointed out, especially for high loaded and extremely heavy tailed traffics. This circumstance renders our protocol a proper candidate in order to satisfy the HSDPA scheme requirements in a realistic indoor environment.

Moreover, in order to highlight this characteristic the buffer occupation distributions for the proposed rate adaptive MAC protocol in the presence of a load factor equal to 0.4 are presented in Fig 5. Since these distributions are similar to the packet arrivals statistics, it follows that the buffer is always empty or filled with at most one message, thus confirming the protocol remarkable efficiency.

VII. CONCLUSION

In this paper, the feasibility of packet-oriented services within a radio mobile network was investigated in order to allow high bit rate and delay intolerant applications that hold a particular relevance for future communication systems. Since

the transmission channel is highly time variable, a static radio resource allocation scheme could exhibit a lower throughput or a significant outage probability, whenever this is based on the worst case or on an intermediate case. In order to eliminate these drawbacks, a dynamic rate allocation protocol was proposed. According to our approach, once a discrete time channel monitoring is performed and the quantized observations are mapped into a set of channel state values, a proper transmission format has to be selected for each state with the aim of increasing the link throughput, while preserving QoS requirements.

This protocol was applied in the downlink due to the asymmetric data exchange within a non real time communication and tested under several input traffic models close to realistic data services in terms of burstiness and correlation.

Analytical expressions for the buffer occupation at the base station equipment and for the link to link delay were derived, showing that our protocol is particularly efficient, reaching a gain up to 50% in the presence of high loaded link and bursty traffics.

By means of simulation on a realistic radio mobile environment, these results were validated, and a remarkable gain was obtained with respect to traditional static approaches. As a result, effective services providing could be achieved, as far as the available bandwidth, delay and throughput are concerned, while reducing power wasting and increasing network capacity.

REFERENCES

[1] UMTS Forum, "The future mobile market: Global trends," Tech. Rep. 8, 03 1999.
[2] Technical Specification Group C, *C.S20024 cdma2000 High Rate Packet Data Interface Specification*, 3GPP2, v3.0 edition, 12 2001.
[3] Technical Specification Group Radio Access Network, *25.308 HSDPA Overall description*, 3GPP, release 2001 v5.1.0 edition, 12 2001.
[4] Technical Specification Group Radio Access Network, *25.855 HSDPA Overall UTRAN description*, 3GPP, release 2001 v5.0.0 edition, 09 2001.
[5] Technical Specification Group Radio Access Network, *25.331 RRC Protocol Specification*, 3GPP, release 1999 v5.0.0 edition, 03 2002.
[6] E.N.Gilbert, "Capacity of a burst-noise channel," *Bell Systems Tech. Journal*, , no. 39, pp. 1253–1266, 09 1960.
[7] E.O.Elliot, "Estimates of error rates for codes on burst-error channels," *Bell Systems Tech. Journal*, , no. 42, pp. 1977–1997, 09 1963.
[8] M.F.Neuts, *Structured stochastic matrices of M/G/1 type and their applications*, Marcel Dekker, 1989.
[9] D.Towsley, "The analysis of a statistical multiplexer with nonindependent arrivals and errors," *IEEE Transaction on Communications*, vol. COM-28, no. 1, 01 1980.
[10] L.Kleinrock, *Queueing Systems:Theory*, vol. 1, John Wiley & Sons, 1975.
[11] Technical Specification Group Services and System Aspects, *23.107 QoS Concept and Architecture*, 3GPP, release 2002 v.5.3.0 edition, 01 2002.
[12] P.T.Brady, "A model for on-off speech patterns in two way conversation," *Bell Systems Tech. Journal*, vol. 48, no. 7, pp. 2445–2472, 09 1969.
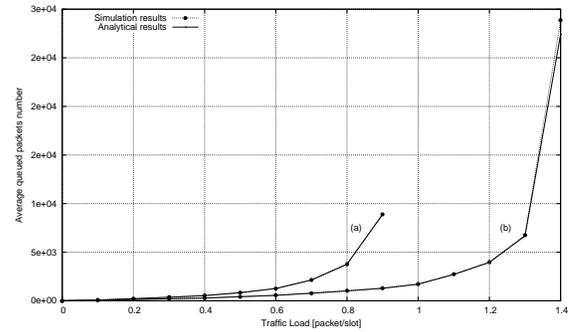
Fig. 2. $\overline{N}$ comparison for a a Pareto packet length arrival.

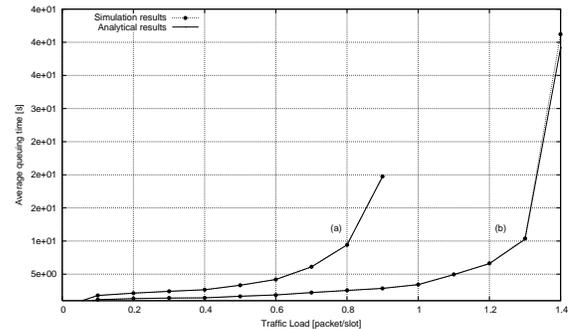Fig. 3. $\overline{T}$ comparison for a a Pareto packet length arrival.
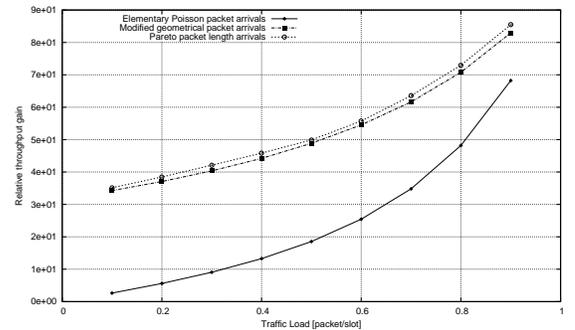

Fig. 4. Relative throughput gain with regard to G/D/1 system.
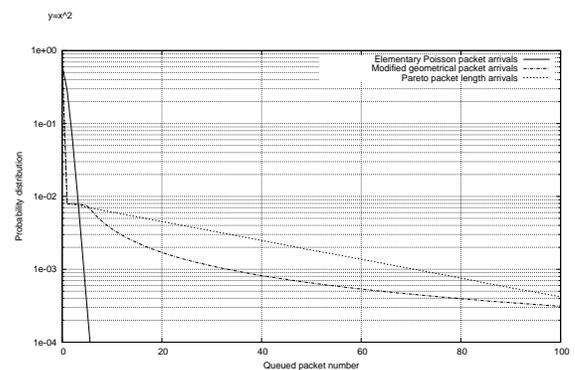

Fig. 5. Buffer occupation distribution comparison for the proposed protocol.