Computer Communications 44 (2014) 44–58

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Dynamic buffer sizing for wireless devices via maximum entropy

Andrés Vázquez-Rodas *, Luis J. de la Cruz Llopis [1], Mónica Aguilar Igartua [2], Emilio Sanvicente Gargallo [1]

Department of Telematics Engineering, Universitat Politècnica de Catalunya BarcelonaTech (UPC), Jordi Girona Street, 31, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 3 April 2013
Received in revised form 23 December 2013
Accepted 4 March 2014
Available online 14 March 2014

Keywords:
Buffer sizing
Wireless networks
Packet loss
Maximum entropy
Queuing systems

ABSTRACT

Buffer overflow is an important phenomenon in data networks that has much bearing on the overall network performance. Such overflow critically depends on the amount of storage space allotted to the transmission channels. To properly dimension this buffering capacity a detailed knowledge of some set of probabilities is needed. In real practice, however, that information is seldom available, and only a few average values are at the analyst disposal. In this paper, the use of a solution to this quandary based on maximum entropy is proposed. On the other hand, when wireless devices are taken into account, the transmission over a shared medium imposes additional restrictions. This paper also presents an extension of the maximum entropy approach for this kind of devices. The main purpose is that wireless nodes become able to dynamically self-configure their buffer sizes to achieve more efficient memory utilization while keeping bounded the packet loss probability. Simulation results using different network settings and traffic load conditions demonstrate meaningful improvement in memory utilization efficiency. This could potentially benefit devices of different wireless network technologies like mesh routers with multiple interfaces, or memory constraint sensor nodes. Additionally, when the available memory resources are not a problem, the buffer memory reduction also contributes to prevent the high latency and network performance degradation due to overbuffering. And it also facilitates the design and manufacturing of devices with faster memory technologies and future all-optical routers.

1. Introduction

Buffers are crucial elements of all kind of routers. They have a great impact in many performance evaluation parameters like packet loss probability, end-to-end delay, delay jitter, link utilization, and throughput. This impact is especially significant during congestion times. The prevention of packet losses has motivated the spread of excessively large buffering across a wide range of network devices and technologies, from Internet core routers to access devices on the edge networks. This phenomenon of buffer oversizing, known as bufferbloat [1], has been accelerated in recent years by the reduction in the memory cost. The bufferbloat brings as a direct consequence that end users experience excessive high latencies in their communications, independently of their access technology and bandwidth [2]. In such environment, the quality of service provided to real-time applications, which are very sensitive to delays, could be very low. Thus, new buffer sizing schemes can provide some benefits to modern network devices. For instance, a small buffer size is extremely valued in all-optical packet switching routers design and construction [3–5]. Nevertheless, too small buffers increase the packet losses and reduce the link utilization when TCP-alike protocols are used [6,7]. Dimensioning routers buffer size is therefore not an easy task and is an active research topic mainly for wired routers [6,8,9]. On the other hand, the growth in the use of mobile devices and their increasing computational capacity, together with the user ubiquitous access expectations, is causing a constant increase in the demand of wireless networking technologies (like wireless local area networks, mobile ad hoc networks, wireless mesh networks, etc.). However, less attention has been provided to the buffer sizing in wireless devices, where new challenging issues arise [10].

In the study of queuing theory, it is customary to begin assuming knowledge of the distributions of service and inter-arrival times, and from there the theory is constructed using Markov (and embedded Markov) chains, Laplace and Z transforms and other mathematical techniques [11]. In real practice, however, that detailed information is seldom available and, in fact, in most instances, the only information at our disposal is a few average values from which other parameter of interest, related to the system performance, must be provided. Obviously, this process

* Corresponding author. Tel.: +34 934017027; fax: +34 934011058.
E-mail addresses: andres.vazquez@entel.upc.edu (A. Vázquez-Rodas), ljcruz@entel.upc.edu (L.J. de la Cruz Llopis), maguilar@entel.upc.edu (M. Aguilar Igartua), telesg@entel.upc.edu (E. Sanvicente Gargallo).
[1] Tel.: +34 934016014; fax: +34 934011058.
[2] Tel.: +34 934015997; fax: +34 934011058.

http://dx.doi.org/10.1016/j.comcom.2014.03.003
0140-3664/

involves a risk since in so doing we are giving more information than we have. To be more specific, suppose that the average value of packets in a transmission system (buffer and server) and the server occupancy have been measured, using, for instance, moving averages. How could we, then, obtain, solely from those two averages, the percentiles of packets in the system? To produce these percentiles, the distribution of packets must be acquired, which is far beyond the available knowledge. The solution to this dilemma can be stated in general terms as follows: when in the course of a system evaluation, the data required to assess the system performance exceeds the available data, the extra information needed should be generated maximizing its entropy in a way compatible with the obtained measurements. This approach has been already presented by some researchers in a rigorous mathematical way [12]. In this paper, we include a new derivation of the method from an engineering point of view.

Besides, when dealing with the task of buffer sizing for wireless devices new problems arise. As it will be described, this is mainly due to the fact that the node transmission state does not depend only on itself, but also on the state of the other nodes inside the same collision domain. In this work we also extend the maximum entropy method for wireless devices working over shared channels. We provide to these devices the capability of dynamically self-configure its buffer sizes according to the traffic load variation and keeping bounded the packet loss probability. Extensive simulations have been done to verify the proper performance of the proposal. We evaluate different scenarios varying the network topology and load conditions. The analysis is done for two kinds of load variations. In the first scenario the variations are due to changes in the traffic generated precisely by the node whose buffer is being sized. In the second case the load fluctuation is due to the activation or deactivation of other nodes in the same collision domain.

In summary, the purpose of this work is to provide to devices which transmit over shared channels a straightforward method, based on easily measured parameters, to self-configure and efficiently manage their available memory. This is achieved by dynamically adapting their buffer sizes according to the traffic load variation during the network operation. The method is based on the application of the maximum entropy principle.

The rest of the paper is organized as follows. In Section 2 we report and analyze the related work. In Section 3 we present, in a practical way, our approximation to the solution of the G/G/1 and G/G/1/K queues via maximum entropy. Here, to make this paper more readable and self-contained, we start presenting some known expression relating the distribution of packets in the system to the distribution of packets found in the system by an arrival. We continue with the maximum entropy approach for the computation of the state probabilities. This section ends with the application that motivated this study, namely: buffer sizing for the G/G/1/K queuing system and its numerical evaluation. In Section 4 we extend the method for wireless devices transmitting over shared channels and present the results obtained from different simulation scenarios. Finally, in Section 5 we conclude our work and provide some lines for further study.

2. Related work

The task of dimensioning buffer sizes can be developed in two ways: once at the design stage or dynamically in order to adapt that size to the variability of the network and traffic conditions. A recent approach of the first type is presented in [13] where authors propose a large deviations framework to dimension the buffer size of delay-tolerant network nodes, as, for instance, VANETs (Vehicular Ad-hoc NETworks) or ICMNs (Intermittently Connected Mobile Networks) nodes. Here, each node is modeled as an M/M/1/B queue and large deviations theory is used to study the queue buffer sizes in terms of buffer overflow. Authors consider a buffer loss probability exponent as the configuration parameter and evaluate the performance of their approach in terms of delivery ratio, delivery delay and message loss ratio. They state that their sized buffer model, with the adequate configuration parameter, offers a performance statistically equivalent to the infinite buffer model. Another example of the importance of buffers in such wireless networks is given in [14]. Here, authors show the impact of the buffer size on packet loss probability, throughput and delay of IEEE 802.16 networks. They also observe that beyond a certain threshold, larger buffers do not improve none of these performance parameters.

On the other hand, several authors propose dynamic buffer sizing mechanisms. For instance, authors in [10] remark the necessity of such a dynamic mechanism for wireless local area networks in order to guarantee high throughput efficiency and reasonable low end-to-end delays. They analyze a simple adaptation of the classic bandwidth-delay product (BDP) rule [6]. This well-known rule, based on the dynamics of the TCP's congestion control mechanism, states that an internet router requires a buffer size B given by $B = C \times RTT$ to achieve a hundred percent utilization at the bottleneck links. Here, C represents the link data rate and RTT is the average round-trip time of a TCP flow passing through that link. Following this rule, and taking into account the currently high possible values of C, impractically large buffers may be obtained. This was first observed by Appenzeller et al. [8] where authors showed that a link with n long-lived or short-lived TCP flows requires only $B = (C \times RTT)/\sqrt{n}$ buffers, and further analyzed in [15] for congested links with different TCP flow types. Some open issues are presented by the same authors in [16]. The adaptation proposed by Leith and Malone [10] consists of an online measurement and actualization of the mean packet service time values. This is required since, in contrast to wired networks in which this value is constant, for 802.11-based wireless networks the service time depends on the number of active stations that contend for the channel (CSMA/CA mechanism stochastic effect) and on the varying modulation and coding scheme chosen by the physical layer (which in turn depends on the radio channel conditions). For this first approach, they use a maximum queuing delay as the configuration parameter. In second place they propose the Adaptive Limit Tuning (ALT) algorithm which main idea is to decrease the buffer size when it has been busy for a long time and increase the buffer size when it has been idle for a long period. The aim is to take advantage of the statistical multiplexing of TCP congestion window backoffs when multiple flows share the same link.

In the same line, authors in [17] propose a buffer sizing mechanism in order to reduce the queuing delays of TCP multi-hop flows while maintaining high network utilization inside 802.11-based WMNs. Their main idea is to consider a joint neighborhood buffer distributed over a set of nodes that contend for channel access within a collision domain. The cumulative buffer for the collision domain is also determined using the classical bandwidth-delay product. To distribute the collective buffer amongst neighborhood nodes, they establish a simple cost function that takes into account the fact that a queue drop close to the source node wastes fewer network resources than a queue drop in a node closer to destination.

Different approaches based on modern control theory are presented in [18,19]. In [18], authors show the monotonic relationship between the buffer size and packet loss rate and utilization. This relationship states that if the buffer size increases then the loss rate monotonically decreases, while the utilization monotonically increases. Based on these monotonic relationships the authors propose the so called Adaptive Buffer Sizing (ABS) mechanism. ABS