BRICKS
Basic Research in Informatics

www.bsik-bricks.nl

Performance of QoS-Differentiating Mechanisms

Pascal Lieshout[1], Michel Mandjes[2,3] and Sem Borst[4,5]

[1]CWI, Amsterdam, The Netherlands
[2]University of Amsterdam, Amsterdam, The Netherlands
[3]EURANDOM, Eindhoven, The Netherlands
[4]Bell Labs, Lucent Technologies, Murray Hill, NJ, USA
[5]Eindhoven University of Technology, Eindhoven, The Netherlands

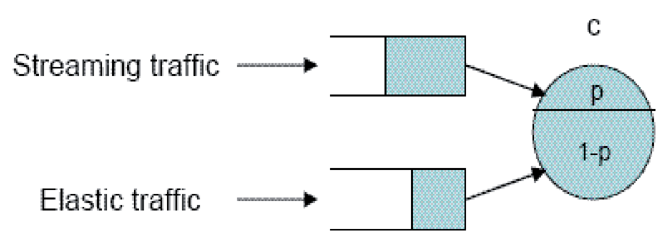
Email: lieshout@cwi.nl, mmandjes@science.uva.nl and s.c.borst@tue.nl

The communication networks of the nearby future are expected to support a wide range of heterogeneous services. Here one can think of traditional applications, such as data, video and voice, but also of sophisticated multimedia applications, such as gaming and remote surgery. These services may have different traffic characteristics, but in addition may also have different Quality-of-Service (QoS) requirements. Typical QoS requirements are constraints on loss probabilities (overflow probabilities) and delay. The integration of heterogeneous traffic flows thus raises the need for service differentiation. In this project we consider mechanisms that are capable of supporting different QoS-levels, and we focus on several performance measures of interest.

GPS

Generalized Processor Sharing (GPS) is a service discipline that is able to support different QoS levels. Since the majority of the traffic of the communication networks can broadly be categorized into *streaming* and *elastic* traffic, each one having its own QoS requirements, let us assume that we have two different traffic classes. Now GPS works as follows:

• We have a single server with a fixed service rate *c*.
• A fixed proportion *p* (weight)t of the service rate is reserved for streaming traffic.
• A fixed proportion *1-p* of the service rate is reserved for elastic traffic.
• In case the streaming (elastic) traffic class does not fully use its guaranteed rate *p* (*1-p*), then the excess rate becomes available to the elastic (streaming) traffic class.
• In case the traffic of a class exceeds the available service, it is stored in the buffer of that class.



Goal

One goal of the project is to solve two interesting open problems on GPS, namely:

• What are the delay asymptotics?, i.e., how does the delay probability of some class behave as the delay grows large?
• How should one select the weights in the GPS mechanism?

Gaussian inputs

In order to solve the two mentioned problems, we have to make some assumption concerning the input process of both classes. In this project we choose Gaussian inputs as input processes for the traffic of the two classes, as Gaussian inputs are a general and versatile class of input processes, covering a broad range of correlation structures. Importantly, Gaussian processes include both short-range and long-range dependent traffic. It is known that Gaussian processes describe real traffic well, as long as the level of aggregation, both in time and number of flows, is sufficiently large.

Results

With the assumption made above, we solved the two open problems:

• Using Schilder's sample-path large deviations, we obtained the exponential decay rate of the delay probability of a class. The results showed that the nature of the asymptotics depends on the GPS weights. In particular, we showed that three different kind of asymptotics are possible.
• In order to address the weight-setting problem, we first characterized the admissible region of the system for fixed GPS weights, i.e., all combinations of flows that satisfy the QoS requirements of both classes. Then we obtained the realizable region by taking the union of the admissible regions over all possible GPS weight values. The results indicate that, under a broad variety of traffic characteristics and QoS requirements, nearly the entire realizable region can be obtained by strict priority scheduling disciplines, i.e., the case where we assign all weight to a single class (*p=0* or *p=1*), such that the other class can only be served if there is no traffic of this single class queued. In addition, we showed how the buffer thresholds, QoS requirements and traffic characteristics of the two classes determine which class should be prioritized.

Part of this research has been funded by the Dutch BSIK/BRICKS project.

CWI Centrum voor Wiskunde en Informatica · TU Delft · NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek · Universiteit Utrecht · TU/e technische universiteit eindhoven · Universiteit Twente