Application of the Stochastic Fluid Flow Model for Bottleneck Identification and Classification

Markus Fiedler† and Kurt Tutschku‡

†Blekinge Institute of Technology, Department of Telecommunications and Signal Processing, SE-371 79 Karlskrona, Sweden. Phone: +46 708 537339, Fax: +46 455 385657,
email: markus.fiedler@bth.se

‡Department of Distributed Systems, Institute of Computer Science, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. Phone: +49 931 8886641, Fax: +49 931 8886632,
email: tutschku@informatik.uni-wuerzburg.de

Keywords: Bottleneck identification, bit rate statistics, stochastic fluid flow model, network performance management.

Abstract:
Network performance management is facing the challenge of provisioning advanced services with stringent delay and throughput requirements. For this reason, shortage of network capacity implying delay or loss, so-called bottlenecks, have to be identified and to be classified. The latter tasks imply the need for tractable analytical performance models. We identify the *stochastic fluid flow model*, which is based on bit rates and its statistics, as a possible candidate of being capable of describing qualitative behaviour of bottlenecks. In this work, we show how total and individual bit rate statistics at the output of a bottleneck are calculated via the stochastic fluid flow model. From this, we deduce some general behaviours and classification criteria for bottlenecks.

1 Introduction

Advanced network services and network applications, such as Voice-over-IP or e-commerce, challenge IP network operation and IP traffic engineering considerably. They generate data streams which are increasingly sensitive to specific throughput and delay requirements. If the requirements are not met, the services may degrade substantially and might become of no use.

In packet-oriented networks, the traffic streams contend for finite resources provided by the communication network. They interfere with each other as well as with network entities. If the capacity limits of the network entities are too tight, the entities might change the characteristics of the data streams too rigorously. The requirements of the application are missed. A major task in network performance management is to eliminate this destructive behaviour. The first step in this procedure is to identify the limiting capacities, commonly denoted as *bottlenecks*. The bottlenecks have to be characterized in terms of location and bandwidth.

Typical network performance management and capacity planning cycles [8] are based on observing the network state. The state is typically determined by passive flow, throughput or load measurements [9]. The observed performance values are usually averages which are obtained on long time scales. The typical duration of observation intervals is in the order of 5 min, e.g. the widely used performance monitoring tool MRTG [12] considers this time as default value. On the other hand, many of today's network performance models for identifying bottlenecks are either not accurate enough or too complex for daily use. Bottleneck identification procedures based on active measurements, cf. [3] and [10], are of limited use since they inject a high volume of maintenance traffic. The additional traffic changes the network load and may interfere inadequately with regular traffic. An approach based on passive measurements is to be more favorable since the network is exposed to minimal extra load.

Many analytical performance models describe packet processes and focus on the analysis of packet delay and packet loss. The transfer of these results into throughput values, which can be observed from network management tools, is in general difficult. As a consequence, many network administrators base their performance management actions on their experience rather than on performance models. To facilitate model-based bottleneck identification, it is necessary to align measurement-based approaches with tractable analytical performance models. In particular, an analytical model should directly address throughput and load measurements. They should reliably predict the qualitative behaviour of the network.

Taking these requirements into account, the *stochastic fluid flow model* emerges as an interesting candidate for an analytical performance model for bottleneck identification. The fluid flow model considers averages of bit rates on small time scales. These averages can easily be measured, and thus, bit rate statistics for data streams entering and leaving a bottleneck can be obtained. The fluid flow model helps us to obtain analytical estimates on the impact of network entities on data streams and their statistics, i.e., efficiently predict the effects of bottlenecks on data streams.

During the years, the fluid flow model has mostly been used for performance evaluation, dimensioning and call admission control purposes, see e.g. [6] for a recent measurement-based work. The fluid flow model has even been successfully applied for describing the output rate process of a buffer [1]. But even though the source model was quite simple – a number of homogeneous on-off sources with exponential autocorrelation were assumed, forming a one-dimensional Markov process – the formulation of the output process became very complicated: the result was a three-dimensional Markov process.

In this work, we present a much simpler method of deriving output bit rate statistics for individual traffic flows and their superposition assuming a general fluid flow model. From the numerical results, which are obtained from standard stochastic fluid flow analysis for Markov-Modulated Rate Processes (MMRP) [2, 7], we analyze how the bottleneck qualitatively changes the total and individual bit rate statistics of data streams.

The paper is structured as follows. In Section 2 we define the meaning of the term *bottleneck* in this work. In Section 3 we describe a bottleneck analysis using the stochastic fluid flow model. Section 4 illustrates the findings of the previous section with a numerical example. Section 5 provides a short summary and outlook to future work.

2 Bottleneck Characteristics

Packet-oriented networks rely on providing resources on-demand. Due to the stochastic nature of the data streams, the finite resources of the network entities passed by these streams may be for some time or in general not sufficient to accommodate the requirements. This view provides directly a definition of the term bottleneck:

Definition: A *bottleneck* within a communication facility is a temporary or permanent lack of capacity compared to the requirements of information streams to be handled. A bottleneck alters the characteristics of an information stream passing through it by introducing delay and/or loss.

The capacity of networks is in general determined by the forwarding capacity of switches, the relaying capacity of routers, or the link capacity which is typically measured in *bits per second* (bps). This capacity is not always known beforehand. In practice, it is strongly influenced by the current configuration, load conditions or malfunctioning components. In our investigation we are primarily interested in bottlenecks with constant capacities and their qualitative impact on passing data streams. We distinguish between two types of bottlenecks: *a)* bottlenecks with small buffers; these kind of bottlenecks will introduce medium and limited delay but quite large packet loss, and *b)* bottlenecks with large buffers; these bottleneck are expected to cause large packet delays but only few packet losses.

The next section will outline how these types of bottlenecks can be modeled and identified with the stochastic fluid flow model.

3 Bottleneck Analysis Using the Fluid Flow Model

Being one performance model amongst others, the fluid flow model — no matter whether is of deterministic or stochastic, time-continuous or time-discrete

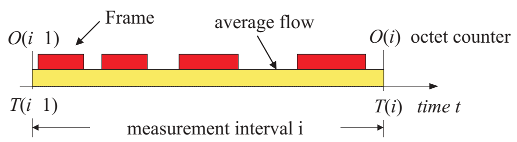


Figure 1: Frames during a measurement interval and corresponding fluid.

nature — has several appealing properties in the context of our study.

The model that has been discussed extensively in literature, see e.g. [2, 13, 7], models the intensity of flows of packets or frames rather than their interarrival time and length processes. Each packet or frame, depending on the layer of modeling, contributes to a stream of fluid particles, which is illustrated by Figure 1. Bit rates can be measured in a simple way using standard network management protocols like SNMP [5]. Assume $O(i)$ to be an octet counter that is measured at time $T(i)$, which denotes the end of interval i. Two examples of such counters are the SNMP MIB variables `ifInOctets` and `ifOutOctets` that count "the aggregated number of octets received/transmitted on the interface, including framing characters" [11]. Then, the average flow intensity during interval i is given by $R(i) = \frac{O(i)-O(i-1)}{T(i)-T(i-1)} \times 8\,\text{bit}$. For an explicit discussion on such measurements, please see [4].

In a fluid flow bottleneck, disturbances of the streams are calculated from comparing required rates with available rates. As long there is enough capacity, incoming packets pass the bottleneck as if it wasn't there at all. Short-term queuing of packets waiting for their turn to be transmitted is hardly covered. The model focuses rather on the consequences of temporary overload situations when the demand for capacity exceeds the available capacity of the bottleneck, which is completely in line with the focus of our study.

In the following, we are looking at how bit rate distributions of data streams are changed when passing a bottleneck. We do this by investigating the flow of fluid particles through the buffer over time.

3.1 Basic Assumptions and Definitions

We denote an individual traffic stream by index i and assume σ states $S = 0, \ldots \sigma - 1$ during which the *aggregate input bit rate* $r_s = \sum_i r_s^{(i)}$ and the *individual input bit rates* $r_s^{(i)}$ are constant. The mean state durations are denoted by $\mathbf{E}[T_s]$ and the state probabilities by $\pi_s = \Pr\{S = s\}$, respectively. The bit rate statistics for a superposition of independent streams, for instance at the entrance of a bottleneck, is obtained from the convolution of the bit rate statistics of the individual streams.

The set of bottleneck parameters consists of the constant capacity C and the buffer size K. The capacity C determines the drift $d_s = r_s - C$ of the buffer content X ($0 \leq X \leq K$) in state s. A state belongs to the set of *overload states* $\mathcal{S}^o$, if $d_s > 0$; during such a state, a buffer fills up at rate d_s until it gets full, but is never empty. A state with $d_s < 0$ belongs to the set of *underload states* $\mathcal{S}^u$; during such a state, the buffer content sinks at rate $|d_s|$ until the buffer is empty. Finally, a state with $d_s = 0$ belongs to the set of *equilibrium states* $\mathcal{S}^e$; during such a state, the buffer content remains unchanged. The set of all states is denoted as $\mathcal{S} = \mathcal{S}^o \cup \mathcal{S}^e \cup \mathcal{S}^u$.

In general, buffers in bottlenecks are of finite size. Consequently, traffic streams passing them may experience delay and loss. Loss is avoided in *infinite/unlimited* buffers, while delay is maximized. The queue is stable as long as the average load does not reach 100 %. All the traffic entering the bottleneck will leave. The temporal characteristics of the traffic flows are changed due to queuing. Turning to very small buffers, used for packet synchronization but unable to cope with overload situation, we observe maximum loss and minimal delay.

In the fluid flow model, such buffers are modeled by setting the buffer size to *zero*. In this so-called *bufferless case*, the characteristics of the traffic passing the bottleneck are altered by loss rather than by delay. As the impact of a bottleneck with finite buffer on traffic streams is to be found somewhere in-between the impacts of infinite and zero buffer, we focus our study on these extreme cases and leave finite buffer sizes for further study.

The bottleneck is *saturated* in case of capacity shortage that appears during or after situations of overload. In this case, the *aggregate output bit rate* $R^o(t)$

is delimited by C, and thus, the throughput characteristics of the individual streams are changed. We assume that both buffer and outlet may be used by the participating streams in a fair-share manner, which is reflected by the factor $\gamma_s^{(i)} = r_s^{(i)}/r_s$ in the *individual output bit rate* $R^{(i),\mathrm{o}}(t) = \gamma_s^{(i)} C$. We denote ν_s as the joint probability that the bottleneck is saturated and the fluid leaving the buffer originates from state s.

In the *non-saturated* case, the bottleneck is more or less invisible to the traffic streams. Its buffer is empty, and both, aggregate and individual, bit rates are unchanged, i.e. $R^{(i),\mathrm{o}} = R^{(i)}$. We denote ε_s as the joint probability that the bottleneck is not saturated and the fluid leaving the buffer originates from state s.

The probabilities ν_s and ε_s with

$$\sum_{s\in\mathcal{S}} \nu_s + \varepsilon_s = 1 \qquad (1)$$

are the key for calculating the output bit rate statistics of R^{o} and $R^{(i),\mathrm{o}}$, respectively. Their calculation will be discussed in the next sections.

3.2 Infinite Buffer

Let the joint buffer content distribution be denoted by $\vec{F}(x)$, $F_s(x) = \Pr\{X \le x \wedge S = s\}$ and its complement by $\vec{G}(x)$, $G_s(x) = \Pr\{X > x \wedge S = s\}$. Depending on the source model, these values can be obtained by standard fluid flow analysis, see [2, 13, 7]. Of special interest in our context are the probabilities that the buffer is empty in state s, $F_s(0)$, and its complement $G_s(0) = \pi_s - F_s(0)$.

We now introduce some notation used in the sequel, see Figures 2 to 4. Let state s begin at time t_s and at buffer level x_s with the first fluid particle arriving to the buffer. This particle leaves the buffer at time t_s^{o}, i.e. experiences a delay τ_s^{o}. At time $t_{s'} = t_s + T_s$, state s ends at a buffer level $x_{s'}$. The last particle from state s leaves the buffer at $t_{s'}^{\mathrm{o}}$, i.e. with a delay of $\tau_{s'}^{\mathrm{o}}$. Fluid particles from state s that have been entering the buffer during T_s need $\tau_s = t_{s'}^{\mathrm{o}} - t_s^{\mathrm{o}}$ to escape from the buffer. The latter time interval may consist of two sub-intervals: The time interval τ_s^ν covers the time during which fluid originating from state s flows through the outlet while the buffer is *non-empty*, whereas the time interval τ_s^ε addresses the case when the buffer is *empty*.

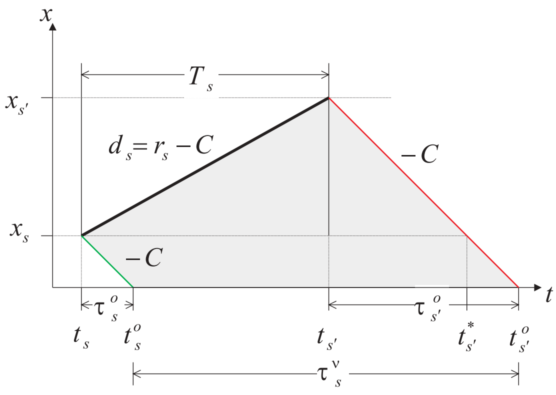


Figure 2: Buffer behaviour in overload state.

Overload states. Figure 2 shows how the buffer content changes over time during a typical overload state. The buffer content rises ($x_{s'} > x_s$) because of $d_s > 0$. Particles originating from state s pass the outlet during the time $\tau_s = \tau_s^\nu > T_s$. From Figure 2, we see that $\tau_s^\nu = t_{s'}^\star - t_s$. This leads to

$$\tau_s^\nu = \frac{r_s}{C} T_s\,, \qquad (2)$$

which is valid independently of x_s. During this time, the bottleneck is saturated. We now proceed by taking average over time, which leads to

$$\mathbf{E}[\tau_s^\nu] = \frac{r_s}{C}\,\mathbf{E}[T_s]\ . \qquad (3)$$

Due to $F_s(0) = 0$, there is no chance that the input stream gets to the output right away ($\mathbf{E}[\tau_s^\varepsilon] = 0$). As state s appears with probability π_s, the probabilities that the bottleneck is saturated/non-saturated from fluid injected in state s are obtained as:

$$\nu_s = \frac{r_s}{C}\,\pi_s \quad \forall s \in \mathcal{S}^o \qquad (4)$$

$$\varepsilon_s = 0 \quad \forall s \in \mathcal{S}^o \qquad (5)$$

Underload states. Figure 3 illustrates how the buffer content changes over time during a typical underload state. We obtain:

$$\tau_s^\nu = \frac{r_s}{C}\,\hat{\tau}_s \qquad (6)$$

$$\tau_s^\varepsilon = T_s - \hat{\tau}_s \qquad (7)$$

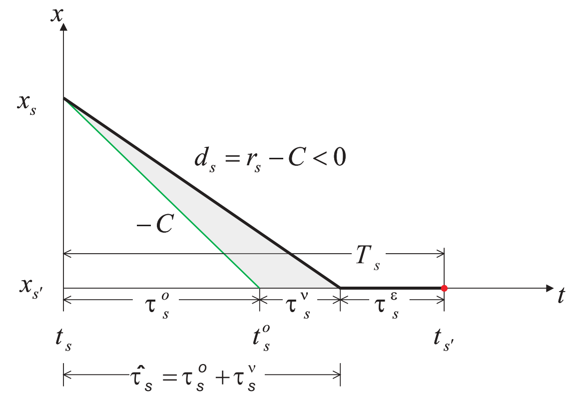


Figure 3: Buffer behaviour in underload state.

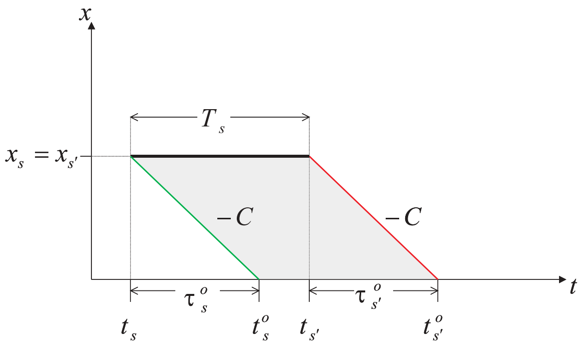


Figure 4: Buffer behaviour in equilibrium state.

Substituting and taking expectations leads to:

$$\mathbf{E}[\tau_s^\nu] = \frac{r_s}{C}\left(\mathbf{E}[\tau_s^o] + \mathbf{E}[\tau_s^\nu]\right) \quad (8)$$

$$\mathbf{E}[\tau_s^\varepsilon] = \mathbf{E}[T_s] - \mathbf{E}[\tau_s^o] - \mathbf{E}[\tau_s^\nu] \quad (9)$$

The probabilities of interest are obtained as:

$$\nu_s = \frac{r_s}{C} G_s(0) \quad \forall s \in \mathcal{S}^u \quad (10)$$

$$\varepsilon_s = F_s(0) \quad \forall s \in \mathcal{S}^u \quad (11)$$

Equilibrium states. In a state s with an input rate matching the capacity, the buffer content remains unchanged ($x_{s'} = x_s$), see Figure 4. We see right away that $t_{s'}^o - t_s^o = t_{s'} - t_s = T_s$. Taking expectations and going over to probabilities, we have to distinguish two cases depending on the initial buffer content x_s:

$$\nu_s = G_s(0) \quad \forall s \in \mathcal{S}^e \quad (12)$$

$$\varepsilon_s = F_s(0) \quad \forall s \in \mathcal{S}^e \quad (13)$$

3.3 Bufferless Case

The derivation of the probabilities ν_s and ε_s gets much simpler in the bufferless case, as there is no chance of transferring any fluid into the domain of an adjacent state ($x_s = x_{s'} = 0$).

Overload states. The capacity of the bottleneck is shared by all participating streams:

$$\nu_s = \pi_s \quad \forall s \in \mathcal{S}^o \quad (14)$$

$$\varepsilon_s = 0 \quad \forall s \in \mathcal{S}^o \quad (15)$$

Underload and equilibrium states. There is enough capacity for all streams so that the input rates appear unchanged at the output:

$$\nu_s = 0 \quad \forall s \in \mathcal{S}^u \cup \mathcal{S}^e \quad (16)$$

$$\varepsilon_s = \pi_s \quad \forall s \in \mathcal{S}^u \cup \mathcal{S}^e \quad (17)$$

3.4 Bit Rate Statistics

Now that the joint probabilities ν_s and ε_s have been obtained, the bit rate statistics at the output of the buffer can readily be calculated.

Aggregated output bit rate statistics. The bit rate statistics for the aggregate output bit rate is composed by summing up all joint probabilities for a specific rate r:

$$\Pr\{R^o = r\} = \sum_{s \in \mathcal{S}^u \cup \mathcal{S}^e:\, r = r_s} \varepsilon_s + \sum_{s \in \mathcal{S}:\, r = C} \nu_s \quad (18)$$

Individual output bit rate statistics. These statistics that actually may differ for each stream i are composed in the same way as above:

$$\Pr\left\{R^{(i),o} = r\right\} = \sum_{s \in \mathcal{S}^u \cup \mathcal{S}^e:\, r = r_s^{(i)}} \varepsilon_s + \sum_{s \in \mathcal{S}:\, r = \gamma_s^{(i)} C} \nu_s \quad (19)$$

4 Numerical Example

Up to now, we assumed a generic stochastic fluid flow model from which we can derive the probabilities $F_s(0)$ and $G_s(0)$ for each state s, respectively. In order to obtain numerical results, however, we restrict to the well-established time-continuous stochastic fluid

r	$\Pr\{R^{\mathrm{o}} = r\}$						
	$C = 11$ Mbps	$C = 9$ Mbps		$C = 7$ Mbps		$C = 5$ Mbps	
		$K = 0$	$K \to \infty$	$K = 0$	$K \to \infty$	$K = 0$	$K \to \infty$
1 Mbps	5.63e-2	5.63e-2	5.63e-2	5.63e-2	5.63e-2	5.63e-2	5.62e-2
2 Mbps	1.88e-1	1.88e-1	1.88e-1	1.88e-1	1.88e-1	1.88e-1	1.82e-1
3 Mbps	2.82e-1	2.82e-1	2.82e-1	2.82e-1	2.82e-1	2.82e-1	2.62e-1
4 Mbps	2.50e-1	2.50e-1	2.50e-1	2.50e-1	2.50e-1	2.50e-1	2.07e-1
5 Mbps	1.46e-1	1.46e-1	1.46e-1	1.46e-1	1.45e-1	2.24e-1	2.93e-1
6 Mbps	5.84e-2	5.84e-2	5.84e-2	5.84e-2	5.65e-2	—	—
7 Mbps	1.62e-2	1.62e-2	1.62e-2	1.97e-2	2.25e-2	—	—
8 Mbps	3.09e-3	3.09e-3	3.07e-3	—	—	—	—
9 Mbps	3.86e-4	4.16e-4	4.38e-4	—	—	—	—
10 Mbps	2.86e-5	—	—	—	—	—	—
11 Mbps	9.54e-7	—	—	—	—	—	—

Table 1: Aggregate output bit rate distributions, bufferless ($K = 0$) and infinite buffer ($K \to \infty$) case.

flow model with exponential autocorrelation. We apply spectral analysis [2, 13] with a special treatment of the indifferent states [7].

4.1 Description

We model the aggregate behavior of a couple of bursty sources as a time-continuous Markov-Modulated Rate Process (MMRP) with σ states with aggregate rates r_s and individual contributions $r_s^{(i)}$ as well as transition rates $\lambda_{s \to s'}$.

We assume the following scenario. A bottleneck of C Mbps is shared by constant and variable bit rate traffic. As input to the bottleneck, we have one constant bit rate source with $R^{(\mathrm{c})} \equiv 1$ Mbps and ten variable bit rate sources of exponential on-/off type, each with peak rate $r_{on} = 1$ Mbps and transition rates $\lambda_{\mathrm{on}\to\mathrm{off}} = 3\lambda_{\mathrm{off}\to\mathrm{on}}$, which implies a bit rate distribution of $\Pr\{R^{(\mathrm{v})} = 0 \text{ Mbps}\} = 0.75$ and $\Pr\{R^{(\mathrm{v})} = 1 \text{ Mbps}\} = 0.25$, respectively.

4.2 Aggregate Output Bit Rate Distribution

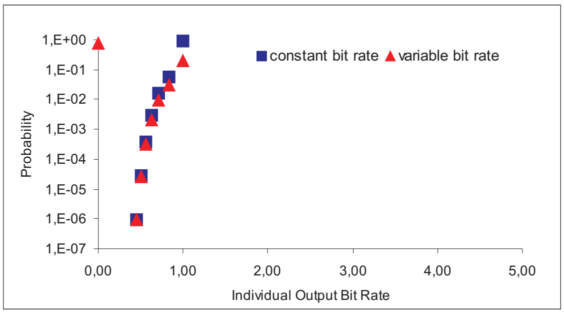
Table 1 compares aggregate output bit rate distributions for bottlenecks with different capacities and buffer sizes. The value of $C = 11$ Mbps implies that there is enough capacity at all times. No bottleneck is visible in the fluid flow model, and hence, the bit rate distribution at the output matches the one at the input.

For $C < 11$ Mbps, the aggregate output bit rate distributions are cut at $R^{\mathrm{o}} = C$. The lower the bottleneck capacity C (i.e. the worse the bottleneck) and the larger the buffer, the larger the deviation between input and output bit rate distributions, and the larger the peak in the distribution at $R^{\mathrm{o}} = C$. Furthermore, we observe that the probabilities of aggregate bit rates $C > R^{\mathrm{o}} = 4, 3, \ldots$ Mbps) are smaller than their counterparts at the input, which makes the peak at the capacity even better visible.

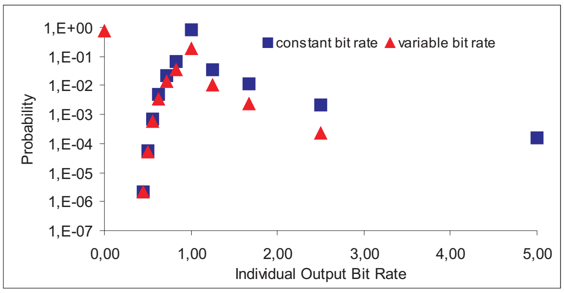
Consequently, one may assume that bottlenecks might be found from comparisons of the bit rate statistics of the aggregate output stream with the convolution of the bit rate statistics of (uncorrelated) input traffic streams especially if the bottleneck is equipped with a buffer of significant size.

4.3 Individual Output Bit Rate Distributions

Figure 5(a) compares the individual bit rate statistics of the constant bit rate stream with those of *one* variable bit rate stream. They show input and output bit rate statistics of a bufferless bottleneck with $C = 5$ Mbps, i.e. a load of 70 %. In the individual output bit rate statistics, speeds lower than the original 1 Mbps appear at the output of the bottleneck. These “slow-downs” origin from the overload states ($R^{\mathrm{o}} > C$) in which the capacity is shared between the competing streams. Interestingly, the output statistics of constant and variable bit rate streams are very sim-



(a) bufferless case



(b) infinite buffer case

Figure 5: Rate distributions at output for individual streams

ilar.

Figure 5(b) depicts the corresponding individual output bit rate statistics for a system with infinite buffer. As in the bufferless case, we see “slow-downs” stemming from the overload states. As opposite to the bufferless case, even “speed-ups” (bit rates exceeding the original peak bit rates at the input) appear. These have their origins in situations when the bottleneck is saturated in underload states. In this case, a stream may be served at a speed that even can reach bottleneck speed if it is the only stream in progress, cf. the maximal bit rate of $R^{(c)} = C = 5$ Mbps in Figure 5(b). Such “speed-ups” are typical for buffered bottlenecks. Again, the shapes of the distributions are the same for constant and variable bit rate traffic. In other words, the constant bit rate stream inherits the characteristics introduced by the variable bit rate streams.

Last but not least, it is worth mentioning that the individual bit rate statistics can be used in an end-to-end manner to reveal the overall bottleneck performance of a network path between two end nodes.

5 Conclusions and Open Issues

Motivated by the need for a simple and tractable queuing model in the context of network performance management, we have applied the *stochastic fluid flow model* for bottleneck identification and classification. We were able to identify qualitative criteria for bottlenecks with respect to capacity and buffer size based on comparisons of input and output bit rate statistics. This has been shown both for aggregate and individual traffic streams.

While the total bit rate statistics is basically shaped by the capacity of the bottleneck, the individual bit rate statistics at the output of the bottleneck reveal the interaction of traffic streams with different characteristics when passing the bottleneck. For instance, the bit rate pattern of streams with variable bit rate that are causing overload in the bottleneck is also visible in the pattern of streams with constant bit rate sharing the same resource. Moreover, the individual output bit rate distributions in particular contain clear indications about whether the bottleneck is equipped with significant buffering capabilities or not.

Further work will include finite buffers as well as simulations to validate our analytical results and to imitate real traffic conditions. Another important piece of future work will cover measurements of bit rate statistics and bottleneck identification and classification in real environments based on the qualitative results obtained in this paper.

Acknowledgement: The authors like to thank the management committee of COST 279 for supporting this work by funding it as a Short Term Scientific Mission, and the reviewers for their valuable comments.

References

[1] S. Aalto. Output from an A-M-S type fluid queue. In *Proceedings of the 14th International Teletraffic Congres, Antibes Juan-Les-Pins, France*, pages 421–443, 1994.

[2] D. Anick, D. Mitra, and M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *The Bell System Technical Journal*, 61(8):1871–1894, 1982.

[3] CAIDA. Bandwidth estimation: Issues and approaches. available at http://www.caida.org/ analysis/performance/bandwidth/, Cooperative Association for Internet Data Analysis (CAIDA).

[4] P. Carlsson, M. Fiedler, K. Tutschku, S. Chevul, and A. Nilsson. Obtaining reliable bit rate measurements in SNMP-managed networks. In *Proceedings of 15th ITC Specialist Seminar on Internet Traffic Engineering and Traffic Management, Würzburg, Germany*, 2002.

[5] J. Case, M. Fedor, M. Schoffstall, and C. Davin. Simple network management protocol (SNMP). RFC 1157 (Internet Standard STD15), May 1990, IETF.

[6] R. Casellas. Performance evaluation of MPLS load sharing. In *Proceedings of the 1th Workshop on the Design and Performance Evaluation of 3G Internet Technologies - IEEE MASCOTS 2002 Workshops, Fort Worth, USA*, 2002.

[7] M. Fiedler and H. Voos. New results on the numerical stability of the stochastic fluid flow model analysis. In *Proceedings of Networking 2000 Conference, Paris, France*, 2000.

[8] D. Grillo, R. S. nd S. Chia, and K. Leung. Teletraffic engineering for mobile personal communications in ITU-T work – the need for matching practice and theory. *IEEE Personal Communications*, 5(6):38–58, 1998.

[9] M. Grossglauser and J. Rexford. *The Internet as a Large-Scale Complex System*, chapter Passive Traffic Measurement for IP Operations. Oxford University Press, 2002.

[10] K. Harfoush, A. Bestavros, and J. Byers. Measuring bottleneck bandwidth of targeted path segments. Technical Report 2001-016, available at http://www.cs.bu.edu/, Boston University, 2001.

[11] K. McCloghrie and M. Rose. Management information base for network management of TCP/IP-based internets: MIB-II. RFC 1213 (Internet Standard STD17), March 1991, IETF.

[12] T. Oetiker. The multi router traffic grapher (MRTG). available at http://www.mrtg.org/, Swiss Federal Institute of Technology, Zürich, 1994-2002.

[13] T. Stern and A. Elwalid. Analysis of separable markov-modulated rate models for information-handling systems. *Advances in Applied Probability*, 23:105–139, 1991.