# On the Way to a Wireless Network Calculus – The Single Node Case with Retransmissions

Hao Wang and Jens B. Schmitt
disco | Distributed Computer Systems Lab
University of Kaiserslautern, Germany

Technical Report No. 375/10

**Abstract**

During the last two decades, starting with the seminal work by Cruz, network calculus has evolved as a new theory for the performance analysis of networked systems. In contrast to classical queueing theory, it deals with performance bounds instead of average values and thus has been the theoretical basis of quality of service proposals such as the IETF's Integrated and Differentiated Services architectures. Besides these it has, however, recently seen many other applications scenarios as, for example, wireless sensor networks, switched Ethernets, avionic networks, Systems-on-Chip, or even to speed-up simulations, to name a few.
In this paper, in an attempt to improve the versatility of network calculus, we extend its reach to error-prone wireless links employing ARQ schemes. This is based on a stochastic extension of data scaling as introduced in [17]. Modelling the single node case with retransmissions results in a set of equations which are amenable to a fixed point solution. This allows to find the arrival constraints of each flow corrsponding to a certain number of retransmissions. Based on the description of each retransmission flow, probabilistic performance bounds for a wireless system with ARQ can be derived. To illustrate the actual procedures, a numerical example wraps up the paper.

*Keywords*: Performance bounds, network calculus, wireless channel, stochastic scaling, ARQ.

# 1 Introduction

## 1.1 Motivation

Network calculus is a min-plus system theory for deterministic queuing systems which builds on the calculus for network delay in [12], [13]. The important concept of *service curve* was introduced in [14, 27, 9, 24, 2]. The service curve based approach facilitates the efficient analysis of tandem queues where a linear scaling of performance bounds in the number of traversed queues is achieved as elaborated in [11] and also referred to as pay bursts only once phenomenon in [25]. A detailed treatment of min-plus algebra and of network calculus can be found in [4] and [25], [10], respectively.

Network calculus has found numerous applications, most prominently in the Internet's Quality of Service (QoS) proposals IntServ and DiffServ, but also in other scenarios as, for example, wireless sensor networks [28, 23], switched Ethernets [30], Systems-on-Chip (SoC) [7], or even to speed-up simulations [22]. Hence, besides queueing theory it has established as a valuable methodology.

However, as a relatively young theory, compared to e.g. traditional queueing theory, there is also a number of challenges network calculus still has to master. To name a few: recently there has been much interest and progress with respect to stochastic extensions (see [11], [18], [20] for recent advances); tool support for network calculus has been addressed by [29], [5] and brings about new interesting perspectives. A very tough challenge is also found in applying network calculus in wireless scenarios, where servers may be unreliable and some part of the data is not delivered at all, i.e. we face a loss system. Most often, the unreliability of wireless channels is compensated by using ARQ techniques, i.e., retransmission of lost data. This brings another challenge for a network calculus model as this essentially introduces feedback into the system. In this paper, we attack these challenges for the case of a single wireless link that operates under an ARQ scheme using a stochastic extension of data scaling as introduced in [17]. Being able to solve the resulting model using a fixed-point approach makes us believe that we are on a promising way to a *wireless network calculus.*

## 1.2 Related Work

While not so much previous research tried to transfer network calculus concepts into the domain of wireless networks, there is some related work to be discussed here: Remarkably, one of the earliest research on stochastic extensions of network calculus [16] already introduced a "channel impairment process" to model a time-varying channel capacity as well as an ARQ component. The model, however, is very much tied to a specific scheduling algorithm (SCED) to operate on the link. At the time, the concept of scaling, which we use in this paper to capture the error characteristics of the wireless transmission, had not yet been introduced in network calculus. Another related work [1] investigates a network calculus model of a wireless link using the concept of a clipper [15] to capture data loss. The issue of operating the wireless link with an ARQ scheme is briefly discussed but not investigated in further detail, in particular the influence on the overall service capacity is not taken into account. The paper also remains unclear how the deterministic clipper component can actually capture the stochastic behaviour of wireless transmissions errors. In [32], the authors introduce a so-called effective capacity model of a wireless link, in analogy to the notion of effective bandwidth. The goal was to capture the time-varying characteristics of a fading channel. Losses and corresponding ARQ schemes were not considered. Similarly, though following a different analytical approach, [19] focused on the characterization of a fading channel with memory. In this work, a network calculus based on moment-generating functions was used to calculate probabilistic performance bounds. Again, the issue of data loss and ARQ schemes were not subject of this research. In a recent study [33], data loss has explicitly been taken into account and has been abstracted as a stochastic process. As in [16], the introduced error process was composed into the server, in order to propose the so-called error server model. We believe this approach to lose flexibility compared to using scaling for modelling the data loss process as done in this paper. With respect to the integration of ARQ schemes, the paper makes brief mention but it is not central to it. In our work, we focus on losses due to wireless transmission errors and the influence of ARQ schemes to compensate for these losses.

### 1.3 Outline

The remaining sections are organized as follows. In the next section we recall some important concepts and theorems of network calculus. In section 3, deterministic scaling is extended to a stochastic version and some examples of stochastic scaling are given. In section 4, the model of a wireless link with ARQ using network calculus is presented and the way to analyse it is derived. In section 5, a numerical example is given for illustrative purposes.

## 2 Preliminaries on Network Calculus and Data Scaling

In this section, we provide the necessary background on deterministic and stochastic network calculus. Furthermore, previous work of ours on data scaling in network calculus is reviewed as it provides the basis for the work in this paper.

### 2.1 Deterministic Network Calculus

As network calculus is built around the notion of cumulative functions for input and output flows of data, the set $\mathcal{F}$ of real-valued, non-negative, and wide-sense increasing functions passing through the origin plays a major role:

$$\mathcal{F} = \left\{ f : \mathbb{R}^+ \to \mathbb{R}^+, \forall t \geq s : f(t) \geq f(s), f(0) = 0 \right\} .$$

In particular, the input function $F(t)$ and the output function $F'(t)$, which cumulatively count the number of bits that are input to, respectively output from, a system $\mathcal{S}$, are in $\mathcal{F}$.

There are two important min-plus resp. max-plus algebraic operators:

DEFINITION 1: (Min-plus and Max-plus Convolution and Deconvolution) The min-plus resp. max-plus convolution and deconvolution of two functions $f, g \in \mathcal{F}$ are defined to be (here $\wedge$ denotes the minimum and $\vee$ the maximum operator)

$$(f \otimes g)(t) = \underset{0 \le s \le t}{inf} \{f(t-s) + g(s)\}, \ (\wedge, + \ convolution)$$

$$(f \oslash g)(t) = \underset{u \ge 0}{sup} \{f(t+u) - g(u)\}, \ (\wedge, + \ deconvolution)$$

$$(f \overline{\otimes} g)(t) = \underset{0 \le s \le t}{sup} \{f(t-s) + g(s)\}, \ (\vee, + \ convolution)$$

$$(f \overline{\oslash} g)(t) = \underset{u \ge 0}{inf} \{f(t+u) - g(u)\}, \ (\vee, + \ deconvolution)$$

It can be shown that the triple $(\mathcal{F}, \wedge, \otimes)$ constitutes a dioid [25]. Also, the min-plus convolution is a linear operator on the dioid $(\mathbb{R} \cup \{+\infty\}, \wedge, +)$, whereas the min-plus deconvolution is not. Similar statements can be made for max-plus systems. These algebraic characteristics result in a number of rules that apply to those operators, many of which can be found in [25], [10].

Let us now turn to the performance characteristics of flows which can be bounded by network calculus means:

DEFINITION 2: (Backlog and Virtual Delay) Assume a flow with input function $F$ that traverses a system $\mathcal{S}$ resulting in the output function $F'$. The *backlog* of the flow at time $t$ is defined as

$$b(t) = F(t) - F'(t).$$

Assuming *FIFO* delivery, the *virtual delay* for a bit input at time $t$ is defined as

$$d(t) = \inf \{\tau \ge 0 : F(t) \le F'(t+\tau)\}.$$

Next, the arrival and departure processes specified by input and output functions are bounded based on the central network calculus concepts of arrival and service curves:

DEFINITION 3: (Arrival Curve) Given a flow with input function $F$, a function $\alpha \in \mathcal{F}$ is an arrival curve for $F$ iff

$$\forall t, s \ge 0, s \le t : F(t) - F(t-s) \le \alpha(s) \Leftrightarrow F = F \otimes \alpha.$$

A typical example of an arrival curve is given by an affine arrival curve $\gamma_{r,b}(t) = b + rt$, $t > 0$ and $\gamma_{r,b}(t) = 0$, $t \le 0$, which corresponds to token-bucket traffic regulation.

Definition 4: (Service Curve) If the service provided by a system $\mathcal{S}$ for a given input function $F$ results in an output function $F'$ we say that $\mathcal{S}$ offers a minimum resp. maximum service curve $\beta$ resp. $\gamma$ iff

$$F' \geq F \otimes \beta, \text{ resp.}$$

$$F' \leq F \otimes \gamma.$$

A typical example of a service curve is given by a so-called rate-latency function $\beta_{R,T}(t) = R(t-T) \cdot 1_{\{t>T\}}$, where $1_{\{cond\}}$ is 1 if the condition *cond* is satisfied and 0 otherwise.

Using those concepts it is possible to derive tight performance bounds on backlog, delay and output:

Theorem 1: (Performance Bounds) *Consider a system $\mathcal{S}$ that offers a minimum and maximum service curve $\beta$ and $\gamma$, respectively. Assume a flow $F$ traversing the system has an arrival curve $\alpha$. Then we obtain the following performance bounds:*

$$\begin{aligned} &\text{backlog: } \forall t : b(t) \leq (\alpha \oslash \beta)(0) =: v(\alpha, \beta), \\ &\text{delay: } \forall t : d(t) \leq \inf \{t \geq 0 : (\alpha \oslash \beta)(-t) \leq 0\} \\ &\qquad\qquad\qquad\qquad =: h(\alpha, \beta), \\ &\text{output (arrival curve } \alpha' \text{for } F'\text{): } \alpha' = (\alpha \otimes \gamma) \oslash \beta. \end{aligned}$$

## 2.2 Stochastic Network Calculus

In recent years, many efforts towards a stochastic network calculus have been made (see e.g., [8, 34, 16, 31, 6, 3, 11, 18, 21, 26]). Many different definitions of stochastic extensions of arrival and service curves have been proposed and discussed. In particular, to provide a stochastic service curve definition that still allows for a favourable concatenation has shown to be a hard problem for some time. In this section, we simply provide the necessary definitions and basic results as they pertain to the work in this paper, without delving into the deep discussions on alternative definitions. Our definitions are mainly based on [6] and can be seen as direct generalizations of the deterministic network calculus counterparts.

Definition 5: (Stochastic Arrival Curve) Given a flow with input function $F$, a function $\alpha^\epsilon \in \mathcal{F}$ is called a stochastic arrival curve for $F$ iff

$$P(F = F \otimes \alpha^\epsilon) \geq 1 - \epsilon.$$

Note that this definition provides a sample path bound as for example discussed in [6], where it is also called sample-path effective envelope.

Definition 6: (Stochastic Service Curve) If the service provided by a system $\mathcal{S}$ for a given input function $F$ results in an output function $F'$ we say that $\mathcal{S}$ offers a stochastic service curve $\beta^\epsilon$ iff

$$P(F' \geq F \otimes \beta^\epsilon) \geq 1 - \epsilon$$

This definitions follows again [6], where it is called statistical or effective service curves. Based on these definitions, the following stochastic performance bounds can be derived (see again [6] for the proof).

THEOREM 2: (Stochastic Performance Bounds) *Consider a system $\mathcal{S}$ that offers a stochastic service curve $\beta^{\epsilon_\beta}$, respectively. Assume a flow $F$ traversing the system has an arrival curve $\alpha^{\epsilon_\alpha}$. Then we obtain the following stochastic performance bounds:*

$$\begin{aligned}
&\textit{backlog}:\ \forall t : P(b(t) \leq v(\alpha^{\epsilon_\alpha}, \beta^{\epsilon_\beta})) \geq 1 - \epsilon_\alpha - \epsilon_\beta,\\
&\textit{delay}:\ \forall t : P(d(t) \leq h(\alpha^{\epsilon_\alpha}, \beta^{\epsilon_\beta})) \geq 1 - \epsilon_\alpha - \epsilon_\beta,\\
&\textit{output}:\ \alpha' = \alpha^{\epsilon_\alpha} \oslash \beta^{\epsilon_\beta}\\
&\quad \textit{with}\ P(F' = F' \otimes \alpha') \geq 1 - \epsilon_\alpha - \epsilon_\beta.
\end{aligned}$$

It should be noted that under the stochastic service curve definition being used here the concatenation of nodes is problematic without further assumptions. In particular, the violation probabilities for concatenated service curves are time-dependent and can therefore be made equal to one, which makes the guarantees of the concatenated service curve void. Several resorts have been proposed in the literature, the most obvious being the introduction of time-scale bounds which avoids the degeneration of the service curve guarantee for large time durations. We refer the reader to the very good discussion about these issues in [6]. In this paper, we stay with the straightforward definitions.

## 2.3 Data Scaling in Network Calculus

In this subsection, we provide the some definitions and results for introducing scaling elements into network calculus models as presented in [17].

DEFINITION 7: (Scaling Function) A scaling function $S \in \mathcal{F}$ assigns an amount of scaled data $S(a)$ to an amount of data $a$.

As can be seen from the definition of scaling functions, they are a very general concept for taking into account transformations in a network calculus model. Note, however, that they do not model any queuing effects – scaling is assumed to be done infinitely fast. Queuing related effects are still modeled in the service curve element of the respective component.

DEFINITION 8: (Scaling Curves) Given a scaling function $S$, two functions $\underline{S}$, $\overline{S} \in \mathcal{F}$ are minimum and maximum scaling curves of $S$ iff

$$\begin{aligned}
\underline{S} &\leq S \bar{\oslash} S,\\
\overline{S} &\geq S \oslash S.
\end{aligned}$$

The following corollary states the effect scaling has on arrival constraints of a traffic flow.

COROLLARY 1: (Arrival Constraints under Scaling) *Let $F$ be an input function with arrival curve $\alpha$ that is fed into a scaling function $S$ with maximum* *scaling curve $\overline{S}$. An arrival curve for the scaled output from the scaling element is given by*

$$\alpha_S = \overline{S}(\alpha).$$

# 3 Stochastic Data Scaling

We base our model of an error-prone wireless link on data scaling. In particular, lost data is modelled as a scaled version of the output flow of the wireless link. That means, we first pretend the service to be perfectly deterministic and then take out some data units according to a scaling process. Obviously, this scaling process is of stochastic nature and can usually not be bounded deterministically in a useful manner. Therefore, in this section, along the footprints "arrival curve→stochastic arrival curve and service curves→stochastic service curves", we extend scaling curves to their stochastic counterparts.

## 3.1 Stochastic Scaling Curves

We provide a straightforward probabilistic interpretation of scaling curves, which will nevertheless allow us to model realistic error process from wireless channels.

Definition 9: (Stochastic Scaling Curves) Consider a scaling function $S$. Any two functions $\underline{S}^{\underline{\varepsilon}} \in \mathcal{F}$ and $\overline{S}^{\overline{\varepsilon}} \in \mathcal{F}$ are said to be a minimum resp. maximum stochastic scaling curve of $S$ if for all $b \geq 0$ it holds that

$$\begin{aligned} Pr((S\overline{\oslash}S)(b) &\geq \underline{S}^{\underline{\varepsilon}}(b)) \geq 1-\underline{\varepsilon}, \\ Pr((S\oslash S)(b) &\leq \overline{S}^{\overline{\varepsilon}}(b)) \geq 1-\overline{\varepsilon}. \end{aligned}$$

Here, $\underline{\epsilon}$ and $\overline{\epsilon}$ denote the violation probabilities for stochastic minimum and maximum scaling curves, respectively.

Note that the stochastic scaling curve properties are defined over sample paths as realized by the respective scaling function. In the context of stochastic arrival curves this has also been coined as sample-path effective (see Section 2.2).

In the following corollary we state the influence of stochastic scaling on the arrival constraints of a flow.

Corollary 2: (Arrival Constraints under Stochastic Scaling) *$F$ is an arrival function which is input to a scaling function $S$ with $\overline{S}^{\overline{\varepsilon}}$ and let $\alpha^{\epsilon_\alpha}$ be a stochastic arrival curve of $F$. A stochastic arrival curve of the stochastically scaled output arrival function $F_S$ is given by $\alpha_S = \overline{S}^{\overline{\varepsilon}}(\alpha^{\epsilon_\alpha})$ with probability $\geq 1-\overline{\epsilon}-\epsilon_\alpha$.*

Proof: Consider a sample path for which neither $\overline{S}^{\overline{\varepsilon}}$ nor $\alpha^{\epsilon_\alpha}$ is being violated. For all such sample paths we obtain from Corollary 1 that $\alpha_S = \overline{S}^{\overline{\varepsilon}}(\alpha^{\epsilon_\alpha})$ is an arrival curve. The probability for being on such a sample path can be computed as

$$\begin{aligned} & P(F_S(t) - F_S(s) \leq \overline{S}^{\overline{\epsilon}}(\alpha^{\epsilon_\alpha})) \\ \geq \; & P(\{\alpha^{\epsilon_\alpha} \; not \; violated\} \wedge \{\overline{S}^{\overline{\epsilon}} \; not \; violated\}) \\ \geq \; & 1 - \overline{\epsilon} - \epsilon_\alpha \end{aligned}$$

due to Boole's inequality, which establishes the claim of the corollary.□

Because scaling is considered to have no delay and backlog, the definitions of backlog and virtual delay bounds of stochastic scaled server are the same as theorem in 2.1 resp. 2.2.

## 3.2 Binary Symmetric Channel as Stochastic Scaling

To illustrate how the error process of a wireless channel can be captured with stochastic scaling, we assume provide the case of binary symmetric channel (BSC) as an example. More complicated channel models can also be captured, however for the ease of description we restrict the discussion here on the BSC.

The bit loss process of a BSC is an i.i.d. Bernoulli process with parameter $\theta$, the crossover probability of the BSC. The probability of $k$ bits loss for $n$ arrival bits can then be calculated as binomial probability

$$\theta^k(1-\theta)^{n-k}\begin{pmatrix} n \\ k \end{pmatrix}.$$

Based on this formulation, a stochastic upper resp. lower bound of scaling functions can be formulated as follows

$$\begin{aligned} S_{upper}^{\overline{\varepsilon}}(n) &= \sum_{k=0}^{n} 1_{\left\{ \sum_{i=0}^{k} \theta^i (1-\theta)^{n-i} \begin{pmatrix} n \\ i \end{pmatrix} < 1-\overline{\varepsilon} \right\}} \\ S_{lower}^{\underline{\varepsilon}}(n) &= [(\sum_{k=0}^{n} 1_{\left\{ \sum_{i=0}^{k} \theta^i (1-\theta)^{n-i} \begin{pmatrix} n \\ i \end{pmatrix} \leq \underline{\varepsilon} \right\}}) - 1]^+ \end{aligned}$$

With $n$-axis and $k$-axis, let us imagine these two curves in Figure 1. The whole sample path space of scaling functions is shown in dotted lines. Note, each segment of scaling function has a rate equaling 0 or 1, which means "not lost" or "lost". The meaning of the curve is: for each $n$, if we cumulate the probabilities from 0 to $k$ until the requirement "prob.$\geq 1-\overline{\varepsilon}$ resp. prob.$\geq 1-\underline{\varepsilon}$" is satisfied, we can then achieve a number $k$; connecting such $k$ for each $n$ we obtain the curves of the bounds.

With the formulation of $S_{upper}^{\overline{\varepsilon}}(n)$ and $S_{lower}^{\underline{\varepsilon}}(n)$, two probabilities are introduced and the form of these two formulations seems similar to the definition of stochastic scaling curves (definition 3.1). Are they really stochastic scaling curves? We analyze the upper bound for example. The analysis about the lower bound follows along the same lines. From the meaning of the curve together with another important argument - BSC is memoryless, we have

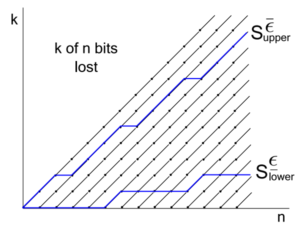


Figure 1: Stochastic scaling curves for BSC.

$$Pr\{S(0,\, n) \leq S^{\overline{\varepsilon}}_{upper}(n)\} \geq 1 - \overline{\varepsilon}$$

$$\overset{BSC\ is\ memoryless}{\Longleftrightarrow}$$

$$Pr\{\forall x:\ S(x,\, x+n) \leq S^{\overline{\varepsilon}}_{upper}(n)\} \geq 1 - \overline{\varepsilon},$$

where $S(x+n)-S(x)$ is written as $S(x, x+n)$. The latter formulation establishes $S^{\overline{\varepsilon}}_{upper}(n)$ as a maximum stochastic scaling curve according to definition 9.

It should be noted that, in fact, we can also assume the bit pass behavior of a BSC as the scaling on the other hand. Such a scaling focuses on the data flow traversing through and can be defined as the "complementary scaling" of the bit loss scaling. Then the calculation of stochastic scaling curve will be based on the probability of $k$ bits pass of $n$ arrival bits.

# 4 A Network Calculus Model of a Wireless Link Employing ARQ

In this section, we propose a model for a wireless link that employs an ARQ scheme to recover from data loss due to channel impairments. The model builds upon the novel concept of stochastic scaling as introduced in the previous section. Based on this model and existing results of network calculus we provide a method to derive the arrival curves of the retransmitted flows. This method applies a fixed point approach from which we also obtain stability criteria for the overall system. Knowing the arrival curves of all the flows in the system, we are able to derive bounds on backlog, delay, and output. These bounds are probabilistic in nature due to the stochasticity of the scaling.

## 4.1 Basic Model and Assumptions

Wireless links are inherently error-prone and, thus, data loss is a common event. Very often, wireless links, therefore, employ ARQ schemes to compensate for data loss by retransmissions. The data loss process can be captured in a network calculus model by a stochastic scaling element as elaborated for the example of a BSC channel in Section 3.2. Now, what is lost has to be retransmitted, so we

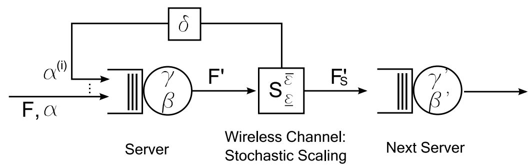


Figure 2: Network calculus of wireless link with ARQ.

feedback the lost data units to the entrance of the server, whose transmission resourcesare modelled using maximum and mimumum service cuves. This fedback retransmission flow would in reality be acknowledgment packets that indicate which data units are missing. Hence, a delay, modelled by a service curve $\delta_T$, for this feedback may be included into the model. Note that retransmitted data units may have to be retransmitted again, resulting in further retransmitted flows. We make the assumption, which holds true in most practical implementations, that there is a limit on how often data is being retransmitted. This model of a wireless link employing ARQ is depicted in Figure 2.

In this model, we distibguish between different retransmission flows, those consisting of data units being retransmitted once, twice, and so on. Correspondingly, we denote all flows in the system as $\alpha^{(0)} = \alpha, \alpha_{\varepsilon_1}^{(1)}, ..., \alpha_{\varepsilon_i}^{(i)}$, where we set $i$ as the limit for the number of retransmission for a single data unit. Note that the arrival curves, apart from the original flows, are stochastic due to the stochasticity of the scaling element that models the loss process (according to Corollary 2). All flows are effectively multiplexed over the wireless link. We make the practical assumption that retransmission flows with higher index, i.e. with data units that have been lost more often so far, are given priority over flows with lower index. Note, however, that this assumption is not essential for what follows.

In the next section, we explicitly derive the arrival curves for all retransmission flows. We do this under a deterministic interpretation of all arrival, service, and scaling curves under investigation. Since, originally only the scaling behaviour is bounded stochastically, this means the arrival curves only apply for those sample paths of the overall system for which the scaling curves apply. In Section 4.3, we then show how we can derive probabilistic bounds based on this observation.

## 4.2 Arrival Curves of Retransmission Flows

If we consider that retransmission flows with higher indices have higher priority, we can use existing network calculus results on priority multiplexing to obtain the following formulations for the service and arrival curves of each retransmis-

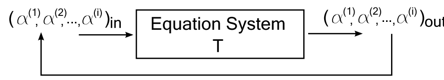


Figure 3: Self-feedback equation system.

sion flow:

$$
\begin{aligned}
\beta^{(0)} &= [\beta - \sum_{k=1}^{i} \alpha^{(k)}]^{+} & \alpha^{(0)} &= \alpha \\
\beta^{(1)} &= [\beta - \sum_{k=2}^{i} \alpha^{(k)}]^{+} & \alpha^{(1)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(0)} \oslash \beta^{(0)}) \oslash \delta_T \\
\beta^{(2)} &= [\beta - \sum_{k=3}^{i} \alpha^{(k)}]^{+} & \alpha^{(2)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(1)} \oslash \beta^{(1)}) \oslash \delta_T \\
&\ldots & &\ldots \\
\beta^{(i)} &= \beta & \alpha^{(i)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(i-1)} \oslash \beta^{(i-1)}) \oslash \delta_T
\end{aligned}
$$

If we combine the two sets of equations and, for the sake of simplicity, ignore the retransmission delay, i.e., $\delta_T = \delta_0$, we obtain the following formulations:

$$
\begin{aligned}
\alpha^{(0)} &= \alpha \\
\alpha^{(1)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(0)} \oslash [\beta - \alpha^{(1)} - \alpha^{(2)} - \ldots - \alpha^{(i)}]^{+}) \\
\alpha^{(2)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(1)} \oslash [\beta - \alpha^{(2)} - \alpha^{(3)} - \ldots - \alpha^{(i)}]^{+}) \\
&\ldots \\
\alpha^{(i)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(i-1)} \oslash [\beta - \alpha^{(i)}]^{+}).
\end{aligned}
$$

Our goal is to find explicit formulations of $\alpha^{(1)}$, $\alpha^{(2)}$, ..., $\alpha^{(i)}$. The alert reader will have realized that each equation implies a self-feedback problem, i.e., the arrival curves of the retransmission flows depend on themselves. We follow a fixed-point approach to resolve this issue of self-feedback. In fact, we can abstract the above equations as a mapping as shown in Figure 3. By dint of fixed point theory, we can interpret our equation system as a mapping $T$, which takes the input $\alpha_n = (\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(i)})_n$ and produces the output $\alpha_{n+1}$. So the problem is transformed into finding a fixed point of $T$, following the recursive rule $\alpha_{n+1} = T\alpha_n$. So, if $T$ makes $\alpha_n$ convergent, we can consequently solve for all the arrival curves of retransmission flows.

To solve the fixed-point problem in general is hard, if not even impossible without further assumption. In this paper, we follow a pragmatic approach where we assume affine arrival curves, i.e. token bucket functions, and rate-latency functions as service curves. Under these assumptions, assessing the convergence of the mapping $T$ is much simplified, yet the this choice of arrival and service curves still covers a lot of practical applications. Similarly, we also

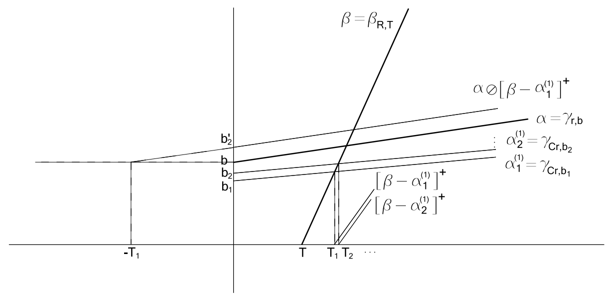


Figure 4: Illustration of the calculation for one retransmission flow.

restrict the scaling curve $\overline{S}^{\overline{\varepsilon}}$ to be an affine curve without too much loss of generality. In particular, we set $\overline{S}^{\overline{\varepsilon}}(x) = Cx + B$, where we assume $B \geq 0$ and $0 \leq C < 1$. The latter condition is actually necessary for the mapping $T$ to be convergent, since otherwise the scaling would not act as a contractor but an expander such that $\alpha_n$ would diverge to infinity. The investigation of more complicated arrival, service and scaling curves is left for your future work, but we believe it can follow a similar line of argument.

For illustrative purposes, we demonstrate the derivation of the arrival curves of retransmission flows for three cases: a single retransmission, two retransmissions, and the general case of $i$ retransmissions. The case of a single retransmission flow we look at in a very detailed, whereas for the other cases we mainly outline the differences and generalizations.

***(1) $i = 1 \leftrightarrow$ one retransmission flow***

We know $\alpha^{(0)} = \alpha$, $\alpha^{(1)} = \overline{S}^{\overline{\varepsilon}}(\alpha^{(0)} \oslash \beta^{(0)}) = \overline{S}^{\overline{\varepsilon}}(\alpha \oslash [\beta - \alpha^{(1)}]^+)$. It is to be checked whether the mapping for $\alpha^{(1)}$ is convergent and what is the fixed point of the mapping $\alpha_\infty^{(1)}$. Put differently, the task is to check whether there is a convergent limit $b_\infty$ for $b_1$ and what is its value, if we set the initial input of $\alpha^{(1)}$ as $\alpha_1^{(1)} = \gamma_{Cr,b_1}$. Note, that we set the rate of $\alpha^{(1)}$ as $Cr$. This is because whatever rate of $\alpha^{(1)}$ we set, after the deconvolution $\alpha \oslash [\beta - \alpha^{(1)}]^+$, the rate will be limited to the rate of $\alpha$ i.e. $r$; and after the invokation of the scaling curve, $\overline{S}^{\overline{\varepsilon}}(x) = Cx + B$, the rate of $\alpha^{(1)}$ will finally always become $Cr$.

What we do next is a step-by-step calculation of the formulation $\alpha^{(1)} = \overline{S}^{\overline{\varepsilon}}(\alpha \oslash [\beta - \alpha^{(1)}]^+)$ until we achieve enough information to assess its convergence and areable to calculate the fixed point value $b_\infty$ of $b_1, b_2, \ldots.$ This process is depicted in Figure 5 and explained in the following steps.

(a) Calculate curve $[\beta - \alpha_1^{(1)}]^+ = \beta_{R-Cr,T_1}$ with $T_1$ according to

$$\frac{CrT_1 + b_1}{T_1 - T} = R \implies T_1 = \frac{RT + b_1}{R - Cr}.$$

(b) Calculate $\alpha \oslash [\beta - \alpha_1^{(1)}]^+$. Draw a line at point $(-T_1, b)$ with rate equal to the rate of the arrival curve $\alpha$, which is $r$. Now calculate $b'_2$ as

$$\frac{b'_2 - b}{T_1} = r \implies b'_2 = rT_1 + b.$$

(c) At last, calculate $\alpha_2^{(1)} = \overline{S}^{\overline{\varepsilon}}(\alpha \oslash [\beta - \alpha_1^{(1)}]^+)$:

$$\alpha_2^{(1)} = \gamma_{Cr,b_2} \text{ with } b_2 = Cb'_2 + B = C(rT_1 + b) + B.$$

Repeat (a), (b) and (c) for $\alpha_2^{(1)}$, and we obtain

$$\begin{aligned} b_3 &= Cb'_3 + B \\ b'_3 &= rT_2 + b \implies b_3 = C(rT_2 + b) + B \\ \frac{CrT_2 + b_2}{T_2 - T} &= R \implies T_2 = \frac{RT + b_2}{R - Cr}. \end{aligned}$$

Repeat again to obtain $b_4 = C(rT_3 + b) + B$ and so on. That means the convergence of $\alpha^{(1)}$ depends on the sequence of $T_1, T_2, T_3$, ... We can write this sequence as follows:

$$\begin{aligned} T_1 &= \frac{RT + b_1}{R - Cr} \\ T_2 &= \frac{RT + b_2}{R - Cr} = \frac{RT + C(rT_1 + b) + B}{R - Cr} \\ T_3 &= \frac{RT + b_3}{R - Cr} = \frac{RT + C(rT_2 + b) + B}{R - Cr} \\ &\cdots \\ T_i &= \frac{RT + b_i}{R - Cr} = \frac{RT + C(rT_{i-1} + b) + B}{R - Cr} \\ &= \frac{Cr}{R - Cr} T_{i-1} + \frac{RT + Cb + B}{R - Cr}. \end{aligned}$$

And for the bucket depths we obtain:

$$b_i = (R - Cr)T_i - RT.$$

For the deconvolution $\alpha \oslash [\beta - \alpha^{(1)}]^+$ to exist, it must hold that

$$\begin{aligned} \lim_{t \to \infty} \frac{[\beta - \alpha^{(1)}]^+(t)}{t} \geq \lim_{t \to \infty} \frac{\alpha(t)}{t} &\implies R - Cr \geq r \\ &\implies \frac{Cr}{R - Cr} \leq C < 1. \end{aligned}$$

This result constitutes a stability criterion for the system, which is also very intuitive. In fact, $R > r + Cr$ means the long-term capacity of the server can

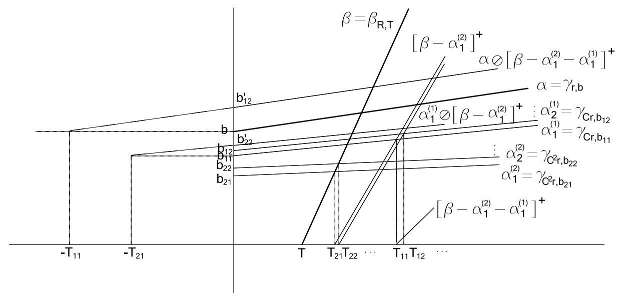


Figure 5: Illustration of the calculation for two retransmission flows.

satisfy the long-term needs of the original and the retransmitted flow. Applying this stability condition $\frac{Cr}{R-Cr} < 1$ to $T_i = \frac{Cr}{R-Cr}T_{i-1} + \frac{RT+Cb+B}{R-Cr}$, we can conclude that the sequence of $T_i$ is convergent, i.e., there is a fixed point $T_\infty$ with

$$T_\infty \quad = \quad \frac{RT + Cb + B}{R - 2Cr}.$$

Finally, we can calculate the arrival curve of the retransmission flow

$$\begin{aligned} \alpha^{(1)} &= \gamma_{Cr,b_\infty}, \ where \\ b_\infty &= (R - Cr)T_\infty - RT = (R - Cr)\frac{RT + Cb + B}{R - 2Cr} - RT. \end{aligned}$$

***(2) i = 2 ↔ two retransmission flows***

In this case, the basic equations are as follows:

$$\begin{aligned} \alpha^{(0)} &= \alpha \\ \alpha^{(1)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(0)} \oslash \beta^{(0)}) = \overline{S}^{\overline{\varepsilon}}(\alpha \oslash [\beta - \alpha^{(1)} - \alpha^{(2)}]^+) \\ \alpha^{(2)} &= \overline{S}^{\overline{\varepsilon}}(\alpha^{(1)} \oslash \beta^{(1)}) = \overline{S}^{\overline{\varepsilon}}(\alpha^{(1)} \oslash [\beta - \alpha^{(2)}]^+) \end{aligned}$$

As initial input we set $\alpha_1^{(1)} = \gamma_{Cr,b_{11}}$ and $\alpha_1^{(2)} = \gamma_{C^2r,b_{21}}$. Again, it has to be checked whether the mapping for $(\alpha^{(1)}, \alpha^{(2)})$ is convergent and what are is the fixed point of the mapping $(\alpha_\infty^{(1)}, \alpha_\infty^{(2)})$. What we do is illustrated Figure 5. Since the calculation process of this case is in principle very similar to the case of one retransmission flow, the calculation steps (a), (b), and (c) are not given here again. The main difference is that $\alpha_1^{(1)}$ and $\alpha_1^{(2)}$ affect each other.

Through similar steps as for the one retransmission flow case and again letting $n \to \infty$, we obtain the following equation system

$$\begin{pmatrix} R^2 - 2(C^2 + C)r & -C^2 r \\ -C^2 r & R - 2C^2 r \end{pmatrix} \begin{pmatrix} T_{1,\infty} \\ T_{2,\infty} \end{pmatrix} = \begin{pmatrix} RT + Cb + C^2 b + B + CB + B \\ RT + C^2 b + CB + B \end{pmatrix}.$$

Then we can use Cramer's Rule to assess if there are roots in $\mathbb{R}^+$ for both $T_{1,\infty}$ and $T_{2,\infty}$. Positive roots imply convergence and can serve as a stability criterion for the overall system. Finally, if the fixed point exists, we get the arrival curves of $\alpha^{(1)}$ and $\alpha^{(2)}$ as

$$\begin{aligned} \alpha^{(1)} &= \gamma_{Cr, b_{1,\infty}}, \ with \\ &\quad b_{1,\infty} = CrT_{1,\infty} + Cb + B, \\ \alpha^{(2)} &= \gamma_{C^2 r, b_{2,\infty}}, \ with \\ &\quad b_{2,\infty} = C^2 r T_{2,\infty} + C^2 r T_{1,\infty} + C^2 b + CB + B. \end{aligned}$$

***(3) $i = k \leftrightarrow k$ retransmission flows***

Because the calculation process for $i > 2$ is essentially the same as for $i = 2$, we ignore the details here and provide directly the results. First, for stability, the intuitive condition $R > (1 + C + C^2 + ... + C^k)r = \frac{1-C^{k+1}}{1-C} r$ applies. It may be noted that this inequality may be used as a tool to adjust the server capacity or the input flow rate. The equation system to be solved for the fixed point solution becomes

$$A \times (T_{1,\infty}, T_{2,\infty}, ..., T_{j,\infty}, ..., T_{k,\infty})^t = \phi, \text{ where}$$

$$A = \begin{pmatrix} R - 2r\sum_{i=1}^{k} C^i & -r\sum_{i=2}^{k} C^i & \dots & -r\sum_{i=k}^{k} C^i \\ -r\sum_{i=2}^{k} C^i & R - 2r\sum_{i=2}^{k} C^i & \dots & -r\sum_{i=k}^{k} C^i \\ \vdots & \vdots & \ddots & \vdots \\ -r\sum_{i=k}^{k} C^i & -r\sum_{i=k}^{k} C^i & \dots & R - 2r\sum_{i=k}^{k} C^i \end{pmatrix}$$

$$\phi = \begin{pmatrix} RT + b\sum_{i=1}^{k} C^i + B \cdot \sum_{p=0}^{k-1}\sum_{q=0}^{p} C^q \\ RT + b\sum_{i=2}^{k} C^i + B \cdot \sum_{p=1}^{k-1}\sum_{q=0}^{p} C^q \\ \vdots \\ RT + b\sum_{i=k}^{k} C^i + B \cdot \sum_{p=k-1}^{k-1}\sum_{q=0}^{p} C^q \end{pmatrix}.$$

Hence, we use Cramer's Rule to derive $T_{1,\infty} = \frac{det(A_1)}{det(A)}, T_{2,\infty} = \frac{det(A_2)}{det(A)}, ..., T_{k,\infty} = \frac{det(A_k)}{det(A)}$, where $A_i$ is the matrix $A$ with the $i$th column of A replaced by $\phi$. If all roots are positive, then a fixed point exists. In this case, the arrival curves of all $k$ retransmission flows are given as follows for $j = 1, \ldots, k$

$$\begin{aligned}\alpha^{(j)} &= \gamma_{C^j r, b_{j,\infty}}, with \\ & \quad b_{j,\infty} = C^j r(T_{j,\infty} + T_{j-1,\infty} + ... + T_{1,\infty}) \\ & \quad\quad +C^j b + (C^{j-1} + ... + C + 1)B\end{aligned}$$

## 4.3 Performance Bounds

In the previous section, we have demonstrated how to derive the arrival curves for each retransmission flow $\alpha^{(i)}$ and, consequently, also the service curves $\beta^{(i)}$ as seen by each of these flows. As discussed above the arguments were given under a deterministic interpretation of arrival, service, and scaling curves. However, for the derivation of performance bounds we now need to take into account the stochastic nature of the wireless channel and therefore of the scaling curve, which captures this stochastic behaviour.

At first, let us assume we are on a sample path of the overall system, where the scaling of each of the flows, original and retransmissions, does not violate the scaling curve $\bar{S}^{\bar{\epsilon}}$. In this case, we can derive for each flow its delay bound as $h(\alpha^{(i)}, \beta^{(i)})$. Since a particular data unit may encounter up to $i$ retransmissions we obtain a delay bound as

$$\forall t : d(t) \leq \sum_{j=0}^{i} h(\alpha^{(j)}, \beta^{(j)})$$

Note again that this delay bound only necessarily applies if the scaling as perceived by each of the flows does not violate the scaling curve $\bar{S}^{\bar{\epsilon}}$. Now each of the flows only experiences a partial realization of the scaling function, which we may call $S_j$. If we further denote $\overline{S}_j^{\bar{\epsilon}_j}$ as stochastic scaling curve for $S_j$, it can be shown that given stationarity of the increment process of $S$, i.e. stationarity of the error process of the channel, $S_j$ has the same stochastic properties as $S$ and therefore we can set $\overline{S}_j^{\bar{\epsilon}_j} = \bar{S}^{\bar{\epsilon}}$. Since most error processes can be considered stationary, this assumption is a mild one and we can calculate a probabilistic delay bound as follows

$$\begin{aligned}P(d(t) &\leq \sum_{j=0}^{i} h(\alpha^{(j)}, \beta^{(j)})) \geq P(\bigwedge_{j=0}^{i} \{\overline{S}_j^{\bar{\epsilon}_j} \; applies\}) \\ & \geq 1 - \sum_{j=0}^{i} \bar{\epsilon}_j = 1 - (i+1)\bar{\epsilon},\end{aligned}$$

where we used Boole's inequality again.

Similar reasoning we can apply to compute a probabilistic backlog bound for all of the flows as

$$P(b(t) \leq v(\alpha^{(0)} + \alpha^{(1)} + ... + \alpha^{(i)}, \beta)) \geq 1 - (i+1)\bar{\epsilon}.$$

## 5 Numerical Example

In order to illustrate the application of the model, let us go through a simple numerical example in this section. First, the necessary assumptions are given.

**Assumptions:** *Consider the scaler to be a BSC with bit loss probability* $\theta = 0.1$*. The scaler has two complementary scaling behaviors:* $S$ *and* $S_{compl}$*, which represent bit loss scaling and bit pass scaling, respectively. The arrival curve of the input flow is* $\alpha = \gamma_{r,b} = \gamma_{0.5,3}$*. The minimum service curve is assumed to be* $\beta = \beta_{R,T} = \beta_{0.8,3}$*. Let the violation probabilities of the scaling curves be* $\overline{\epsilon} = 0.01$ *and* $\underline{\epsilon} = 0.01$*, respectively. We also assume a maximum of two retransmission attempts for each data unit, i.e.,* $i = 2$*.*

**Target of Calculations:** *The arrival curves of all the retransmission flows; performance bounds.*

Basically, the calculation is divided into three parts: (i) analyze the wireless channel and construct the stochastic scaling curves, (ii) calculate the arrival curves of all retransmission flows, and (iii) calculate the performance bounds.

**(i) Construction of Scaling Curves**. The wireless channel is a BSC with bit loss probability $\theta = 0.1$ and the violation probability for the maximum stochastic scaling curve is set to $\overline{\epsilon} = 0.01$. Recall the formulation $\overline{S}^{\overline{\epsilon}}(n)$ from Section 3.2. We want to draw its graph. Since it is rather hard to compute the curve for each $n$, we may do sampling for $n$ and draw $\overline{S}^{\overline{\epsilon}}(n)$ approximately. Connecting all the sampled values for $n$, we may derive a violated maximum stochastic scaling curve. If, however we always make sure to only use points for which $\overline{S}^{\overline{\epsilon}}(n) > \overline{S}^{\overline{\epsilon}}(n-1)$, by searching in the neighbourhood of a sample point, we can, when connecting all such points, derive a non-violated maximum stochastic scaling curve. Together with the complete scaling curve, the respective sampled scaling curves are depicted in Figure 6.

For further calculation, we should further simplify this maximum stochastic scaling curve as an affine function $\bar{S}^{\bar{\epsilon}}(x) = Cx + B$. Certainly, the approximate scaling curve should simulate the original curve as accurately as possible. In fact, many methods to calculate $C$ and $B$ can be conceived. In this example we only do some simple comparisons to obtain a solution, resulting in $C = 0.1250$ and $B = 5.8750$ (when connecting points $(n, k) = (185, 29)$ and $(193, 30)$).

**(ii) Calculation of Arrival Curves of Retransmission Flows**. It is assumed that the number of retransmission flows $i$ is given, we let $i = 2$. In accordance with the notations introduced in Section 4.2, we have

$$A = \begin{pmatrix} 0.6594 & -0.0078 \\ -0.0078 & 0.7844 \end{pmatrix}, \ \phi = \begin{pmatrix} 15.3063 \\ 9.0563 \end{pmatrix}.$$

By solving this equation system, we finally obtain $\alpha^{(1)}$ and $\alpha^{(2)}$ as

$$\begin{aligned} \alpha^{(1)} &= \gamma_{Cr,b_{1,\infty}} = \gamma_{0.0625,\,7.7096} \\ \alpha^{(2)} &= \gamma_{C^2r,b_{2,\infty}} = \gamma_{0.0078,\,6.9307}. \end{aligned}$$

Compare the rate of $\alpha$ and $\alpha^{(1)}$: the rate of $\alpha^{(1)}$, $Cr$, is much smaller than the rate of $\alpha$. Clearly, this is because $\alpha^{(1)}$ is the retransmission flow caused

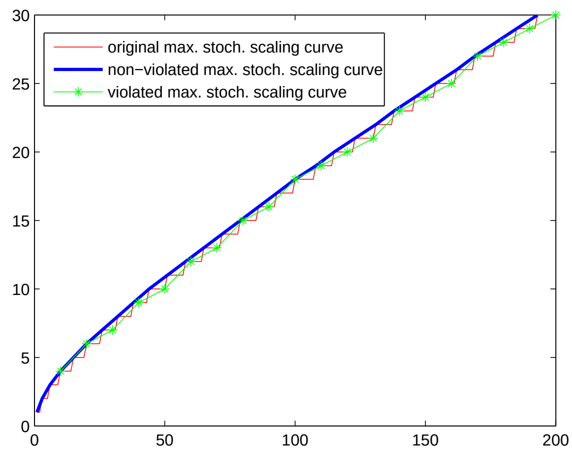


Figure 6: Getting non-violated maximum stochastic scaling curve.

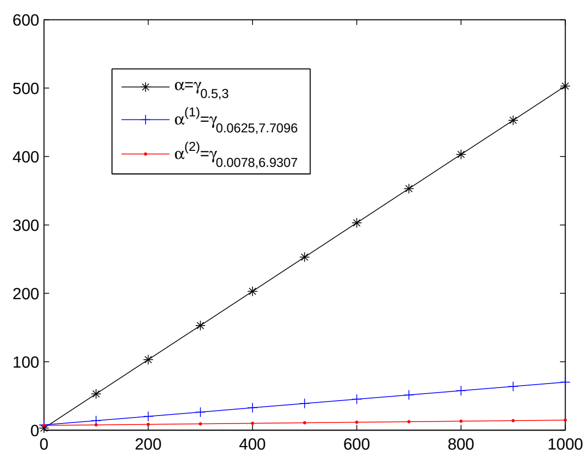


Figure 7: Arrival curves of original and retransmission flows.

by data loss and the probability of data loss is not too high, which means that the effect of the retransmission flow will be rather weak. Consequently, the retransmission flow of $\alpha^{(1)}$, $\alpha^{(2)}$ is even weaker. Let us compare $b$ with $b_1$ and $b_2$: $b_1 = 7.7096 > b_2 = 6.9307 > b = 3$. Essentially, $b$ as the original flow's parameter should intuitively be greater than $b_1$ and $b_2$. Yet, the way the parameters $C$ and $B$ were determined causes this not to be the case and actually the parameter choice could be formulated as an optimization problem. We leave this for future work, however, it is not too hard to conceive at least a numerical optimization procedure. All arrival curves are depicted in Figure 7.

**(iii) Calculation of Performance Bounds**. Based on the arrival curves of all the flows in the system, original and retransmission, we can now derive probabilistic performance bounds.

*Delay Bounds:*

With $\alpha^{(1)}$ and $\alpha^{(2)}$, the straightforward delay bounds for each flow, original and retransmission, are easy to calculate:

$$\begin{aligned} d_0(t) &\leq \bar{d_0} = h(\alpha, [\beta - \alpha^{(1)} - \alpha^{(2)}]^+) = 27.4642, \\ d_1(t) &\leq \bar{d_1} = h(\alpha^{(1)}, [\beta - \alpha^{(2)}]^+) = 21.5104, \\ d_2(t) &\leq \bar{d_2} = h(\alpha^{(2)}, \beta) = 11.6634. \end{aligned}$$

As more interest is usually in the delay of a given data unit, which may be retransmitted, we calculate the following probabilistic delay bound which applies to all data units for all times $t$:

$$P(d(t) \leq \bar{d_0} + \bar{d_1} + \bar{d_2} = 60.6380) \geq 1 - 3 \cdot \bar{\epsilon} = 0.97$$

This delay bound is certainly conservative: the scaling curve approximation already is conservative and, furthermore, many data units will not be retransmitted twice, such that the violation probability should actually be smaller than above. Optimizations along these lines are left for future work.

*Backlog Bounds*:

$$\begin{aligned} b_0(t) &\leq v(\alpha, [\beta - \alpha^{(1)} - \alpha^{(2)}]^+) = 14.6764 \\ b_1(t) &\leq v(\alpha^{(1)}, [\beta - \alpha^{(2)}]^+) = 8.4457 \\ b_2(t) &\leq v(\alpha^{(2)}, \beta) = 6.9541 \end{aligned}$$

The cumulative probabilistic backlog bound is computed as

$$P(b(t) \leq v((\alpha + \alpha^{(1)} + \alpha^{(2)}), \beta) = 19.3512) \geq 0.97.$$

*Output Bound:*

The calculation of an output bound applies the complementary scaling function, i.e., the bit pass scaling. In principle, the calculation process is similar to the computation of the arrival curves of retransmission flows. The respective results are omitted here.

## 6 Conclusion

In this paper, we believe to have made a promising first step on the way to a wireless network calculus. Based on the stochastic extension of data scaling in conventional network calculus we showed how to model the error-prone transmission behaviour of a wireless link. Moreover, we also integrated the typical handling of transmission errors in wireless links by ARQ mechanisms. To solve the model for a single node case involved a fixed-point analysis yielding stability conditions as well as probabilistic performance bounds. In a numerical example we illustrated how to apply the theoretical results.
For future work many open issues remain. Let us mention two of the most challenging ones: Clearly, the multiple node case will be an interesting challenge if the concatenation principle is to be carried over to a tandem of wireless links with ARQ. Also, as briefly discussed in Section 4, there is potential for an improved stochastic analysis to arrive at lower violation probabilities for the performance bounds.

## References


[1] F. Agharebparast and V. Leung. Modeling wireless link layer by network for efficient evaluations of multimedia qos. In *IEEE International Conference on Communications*, volume 2, pages 1256–1260. IEEE, 2005.

[2] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan. Performance bounds for flow control protocols. *IEEE/ACM Transactions on Networking*, 7(3):310–323, June 1999.

[3] S. Ayyorgun and W.-C. Feng. A systematic approach for providing end-to-end probabilistic qos guarantees. In *Proceedings of the 13th IEEE International Conference on Computer Communications and Networks (ICCCN), Chicago, IL*, October 2004. Also available as Technical Reports LA-UR-03-3668 and LA-UR-03-7267, Los Alamos National Laboratory.

[4] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity: An Algebra for Discrete Event Systems*. Probability and Mathematical Statistics. John Wiley & Sons Ltd., 1992.

[5] A. Bouillard and Eric Thierry. An algorithmic toolbox for network calculus. *Discrete Event Dynamic Systeme*, 18(1).

[6] A. Burchard, J. Liebeherr, and S. D. Patek. A min-plus calculus for end-to-end statistical service guarantees. *IEEE Transactions on Information Theory*, 52(9):4105–4114, September 2007.

[7] S. Chakraborty, S. Kuenzli, L. Thiele, A. Herkersdorf, and P. Sagmeister. Performance evaluation of network processor architectures: Combining simulation with analytical estimation. *Computer Networks*, 42(5):641–665, 2003.

[8] C.-S. Chang. Stability, queue length and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

[9] C.-S. Chang. On deterministic traffic regulation and service guarantees: A systematic approach by filtering. *IEEE Transactions on Information Theory*, 44(3):1097–1110, May 1998.

[10] C.-S. Chang. *Performance Guarantees in Communication Networks*. Telecommunication Networks and Computer Systems. Springer-Verlag, 2000.

[11] F. Ciucu, A. Burchard, and J. Liebeherr. A network service curve approach for the stochastic analysis of networks. In *Proc. ACM SIGMETRICS*, pages 279–290, June 2005.

[12] R. L. Cruz. A calculus for network delay, Part I: Network elements in isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, January 1991.

[13] R. L. Cruz. A calculus for network delay, Part II: Network analysis. *IEEE Transactions on Information Theory*, 37(1):132–141, January 1991.

[14] R. L. Cruz. Quality of service guarantees in virtual circuit switched networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1048–1056, August 1995.

[15] R. L. Cruz and M. Taneja. An analysis of traffic clipping. In *Conference of Information Sciences and Systems, Princeton, New Jersey*, March 1998.

[16] R.L. Cruz. Quality of service management in integrated services networks. In *Proceedings of the 1st Semi-Annual Research Review, Center for Wireless Communications, UCSD*, 1996.

[17] M. Fidler and J. Schmitt. On the way to a distributed systems calculus: An end-to-end network calculus with data scaling. In *ACM SIGMETRICS/Performance 2006 (SIGMETRICS'06)*, pages 287–298, St. Malo, France, June 2006. ACM.

[18] Markus Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *Proc. of IEEE IWQoS*, pages 261–270, Jun 2006.

[19] Markus Fidler. A network calculus approach to probabilistic quality of service analysis of fading channels. In *Proceedings of IEEE Globecom*, Nov 2006.

[20] Y. Jiang. A basic stochastic network calculus. In *Proc. ACM SIGCOMM*, pages 123–134, September 2006.

[21] Y. Jiang. A basic stochastic network calculus. *ACM Computer Communication Review*, 36(4):123–134, October 2006. (Proc. ACM SIGCOMM 2006).

[22] H. Kim and J.C. Hou. Network calculus based simulation: theorems, implementation, and evaluation. In *Proc. IEEE INFOCOM*, March 2004.

[23] A. Koubaa, M. Alves, and E. Tovar. Modeling and worst-case dimensioning of cluster-tree wireless sensor networks. In *Proc. 27th IEEE International Real-Time Systems Symposium (RTSS'06)*, pages 412–421, Rio de Janeiro, Brazil, 2006. IEEE Computer Society.

[24] J.-Y. Le Boudec. Application of network calculus to guaranteed service networks. *IEEE Transactions on Information Theory*, 44(3):1087–1096, May 1998.

[25] J.-Y. Le Boudec and P. Thiran. *Network Calculus A Theory of Deterministic Queuing Systems for the Internet*. Number 2050 in Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany, 2001.

[26] J.-Y. Le Boudec and M. Vojnović. Elements of probabilistic network calculus for packet scale rate guarantee nodes. In *Proc. of 15th Int'l Symp. of Mathematical Theory of Networks and Systems*.

[27] H. Sariowan, R. L. Cruz, and G. C. Polyzos. Scheduling for quality of service guarantees via service curves. In *Proc. IEEE ICCCN*, pages 512–520, September 1995.

[28] J. Schmitt and U. Roedig. Sensor network calculus - a framework for worst case analysis. In *Proc. Distributed Computing on Sensor Systems (DCOSS)*, pages 141–154, June 2005.

[29] J. Schmitt and F. Zdarsky. The DISCO Network Calculator - a toolbox for worst case analysis. In *Proc. of VALUETOOLS*. ACM, November 2006.

[30] T. Skeie, S. Johannessen, and O. Holmeide. Timeliness of real-time IP communication in switched industrial ethernet networks. *IEEE Transactions on Industrial Informatics*, 2(1):25–39, February 2006.

[31] D. Starobinski and M. Sidi. Stochastically bounded burstiness for communication networks. *IEEE Transactions on Information Theory*, 46(1):206–212, January 2000.

[32] D. O. Wu and R. Negi. Effective capacity: A wireless link model for support of quality of service. 2(4):630–643, July 2003.

[33] J. Xie and Y. Jiang. An analysis on error servers for stochastic network calculus. In *IEEE Conference on Local Computer Networks (LCN)*. IEEE, 2008.

[34] O. Yaron and M. Sidi. Performance and stability of communication networks via robust exponential bounds. *IEEE/ACM Transactions on Networking*, 1(3):372–385, June 1993.