A Martingale-Envelope and Applications

Felix Poloczek and Florin Ciucu
T-Labs / TU Berlin

ABSTRACT
In the framework of stochastic network calculus we present a new envelope-based approach which uses martingales to characterize a queueing system. We show that this setting allows a simple handling of multiplexing and scheduling; whereas multiplexing of several sources results in multiplication of the corresponding martingales, per-flow analysis in a scheduled system can be done by shifting the martingales to a certain point in time. Applying this calculus to Markov Arrival Processes, it is shown that the performance bounds can become reasonably tight.

1. INTRODUCTION
The Stochastic Network Calculus (SNC) arises as an alternative to classical queueing theory. Its main advantage is its amenability to scheduling and multi-hop scenarios (see [4]). However, the price of this advantage is that the resulting performance metrics, like the distribution of the stationary queue size and the delay, are not computed exactly but only estimated via upper and/or lower bounds.

As most of these bounds in the existing literature are based on techniques from large deviations theory, which are known to render loose bounds (see, e.g., [2]), the relevance of SNC itself is often questioned. This looseness stems from the following fact: When computing bounds on some performance metrics (e.g., buffer overflow probability), the crucial fact is to estimate the supremum of a stochastic process, per-flow analysis in a scheduled system can be done by shifting the martingales to a certain point in time. Applying this calculus to Markov Arrival Processes, it is shown that the performance bounds can become reasonably tight.

1. INTRODUCTION

The Stochastic Network Calculus (SNC) arises as an alternative to classical queueing theory. Its main advantage is its amenability to scheduling and multi-hop scenarios (see [4]). However, the price of this advantage is that the resulting performance metrics, like the distribution of the stationary queue size and the delay, are not computed exactly but only estimated via upper and/or lower bounds.

As most of these bounds in the existing literature are based on techniques from large deviations theory, which are known to render loose bounds (see, e.g., [2]), the relevance of SNC itself is often questioned. This looseness stems from the following fact: When computing bounds on some performance metrics (e.g., buffer overflow probability), the crucial point is to estimate the supremum of a stochastic process, say \( X_n \), in terms of the tail probability: \( P(\sup_n X_n \geq \sigma) \).

The tool mainly used in the literature so far is Boole's inequality:

\[
P(\sup_n X_n \geq \sigma) \leq \sum_n P(X_n \geq \sigma) .
\]  

(1)

It is no surprise that this inequality is far away from being tight and is thus responsible for the weak results mentioned above: As the right hand side of (1) treats every time instance separately, it does not account for dependencies among them and is thus unable to capture major properties of the process \( X_n \).

In this paper we advocate a different approach to estimate the left hand side of (1). Let us consider the analogous problem for (one-dimensional) random variables instead of stochastic processes: In this case it is well known that the tail distribution of \( X \geq 0 \) can be easily estimated by Markov's inequality

\[
P(X \geq \sigma) \leq \frac{\mathbb{E}[X]}{\sigma^2} .
\]

(2)

which, when applied to the exponential r.v. \( e^{\theta X} \), leads to the so called Chernoff bound, which in many cases is sharp. Although in general the direct extension of (2) to arbitrary stochastic processes does not hold, there is one class where it actually does, namely supermartingales. For a supermartingale \( X_n \), a simple application of the optimal stopping theorem gives a variant of Doob's maximal inequality

\[
P(\sup_n X_n \geq \sigma) \leq \mathbb{E}[X_0]\sigma^{-1} .
\]

(3)

Besides this more technical advantage of providing a conceivably sharper inequality than Eq. (1), there is also a conceptual similarity between supermartingales and queueing systems: A supermartingale roughly is a process such that for a given point in time we expect any state in the future to be less than the current. The same is true for the backlog-process in a queueing system: in order to guarantee finite performance metrics, the average arrivals have to be strictly less than the capacity, so that the increments are negative on average. This is typically ensured by a stability condition, like the one of Loynes.

By these two reasons we develop a calculus of queueing systems being described by certain martingale-envelopes, which can be regarded as stochastic bounds on suitable representations of the queueing systems. A major advantage of such a representation is that basic operations involving queueing systems, such as multiplexing and scheduling, correspond directly to operations on the martingale-envelopes themselves, i.e., multiplication and shifting at a certain point in time, respectively.

The scheduling operation in particular demonstrates the benefits of integrating martingale techniques in the SNC framework, which decouples scheduling from computing performance metrics [4]. Moreover, by numerically showing that the new SNC bounds are reasonably tight, we demonstrate that the mentioned weakness of previous bounds are not inherent to SNC itself but instead to inappropriate inequalities such as Eq. (1).

2. A CALCULUS WITH MARTINGALES

We consider a queueing scenario consisting of a flow, given by the bivariate arrival process \( A(m, n) = \sum_{k=m+1}^{n} a_k \), and a single server with constant capacity \( C > 0 \). For simplicity, we assume throughout that the increment process \( (a_n) \) is stationary and that \( A \) is reversible. Using the service curve
representation $S(m, n) = C(n - m)$, the corresponding departure process $D(m, n)$ satisfies

$$D(n) \geq A \ast S(n) := \inf_{k \leq n} \{A(k) + S(k, n)\}.$$ 

As usual $X(n) := X(0, n)$ for any bivariate process $X(m, n)$. We are interested in two performance metrics: the tail distribution of the backlog

$$Q := \sup_{n \in \mathbb{N}} \{A(n) - Cn\}, \quad (4)$$

and of the virtual delay,

$$W(n) := \inf\{k \in \mathbb{N} \mid A(n - k) \leq D(n)\}.$$ 

We next introduce our characterization of a queueing system by a certain supermartingale. 

**Definition 1.** For a monotonically increasing function $h$ and $\theta > 0$, we say the flow $A$ admits a $(h, \theta, C)$-martingale-envelope if for $n \geq m$

$$M(n) := h(a_n) e^{\theta(A(m, n) - C(n - m))}$$

is a supermartingale\(^1\).

This envelope loosely captures the long-term arrival rate of $A$ through some larger value $C$, which can be interpreted as the flow’s allocated capacity. In turn, $\theta$ and $h$ capture the correlation structure of $A$. The next definition characterizes another important property of an $h$.

**Definition 2.** For a flow $A$ and a capacity $C$ we define the threshold

$$\tau_{A, C} := \inf\{x > C \mid P(a_k \in [x, \infty)) > 0\}.$$ 

Intuitively, $\tau_{A, C}$ is the smallest instantaneous arrival $x$ such that $x - C$ is positive. The importance of $\tau_{A, C}$ lies in the following: If $n$ denotes the first point in time where the supremum in Eq. $(4)$ is attained then clearly $a_n - C > 0$, and thus $h(a_n) \geq h(\tau_{A, C})$ by the monotonicity assumption on $h$.

**Theorem 3.** If the flow $A$ admits a $(h, \theta, C)$-martingale-envelope, then we have the following upper bound on the backlog

$$P(Q \geq \sigma) \leq \frac{E[h(a_0)]}{h(\tau_{A, C})} e^{-\theta \sigma}$$

and for the virtual delay

$$P(W(n) \geq k) \leq \frac{E[h(a_0)]}{h(\tau_{A, C})} e^{-\theta C k}.$$ 

**Proof.** For the backlog we have:

$$P(\sup_{n \in \mathbb{N}} \{A(n) - Cn\} \geq \sigma) = P(\sup_{n \in \mathbb{N}} e^{\theta (A(n) - Cn)} \geq e^{\theta \sigma})$$

$$\leq \sup_{n \in \mathbb{N}} \frac{E[h(a_n)] e^{\theta (A(n) - Cn)}}{h(\tau_{A, C}) e^{\theta \sigma}}$$

$$\leq \frac{E[h(a_0)]}{h(\tau_{A, C})} e^{-\theta \sigma},$$

where we used the monotonicity of $h$ and Doob’s maximal inequality. The proof for the virtual delay is similar. \(\square\)

---

\(^1\)A similar martingale representation has been recently considered in [3], which however restricts to a single class of (continuous-time) arrivals; by proposing an envelope representation, the approach herein aligns to the network calculus principle of addressing broad classes of arrivals in a uniform manner.

---

We next show how martingale-envelopes behave when two arrival flows $A_1$ and $A_2$ are multiplexed. A technical Definition is necessary: For two monotonically increasing functions $h_1, h_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, we define the $(\min, x)$-convolution

$$h_1 \otimes h_2(t) := \inf_{0 \leq s+t \leq t} h_1(s) h_2(t-s).$$

It is easy to check that $h_1 \otimes h_2$ is monotonically increasing as well. Another useful property is

$$h_1 \otimes h_2(a+b) \leq h_1(a) h_2(b)$$

for all $a, b$.

**Theorem 4.** (Multiplexing) Assume two independent arrivals $A_1$ and $A_2$ admitting $(h_1, \theta, C_1)$ and $(h_2, \theta, C_2)$ martingale-envelopes, respectively. Then the aggregate flow $A_1 + A_2$ admits a $(h \otimes \theta, \theta, C_1 + C_2)$-martingale-envelope.

**Proof.** Clearly,

$$h_1 \otimes h_2(a_n^1 + a_n^2) e^{\theta (A_1(m, n) + A_2(m, n) - (C_1 + C_2)(n - m))}$$

is the product of two independent martingales is again a martingale. \(\square\)

If the arrivals are not homogenous, in $\theta$, then the following transform of martingale-envelopes can be invoked.

**Lemma 5.** If $A$ admits a $(h, \theta, C)$-martingale-envelope and $\theta' < \theta$, then $A$ admits a $(h, \theta', C')$-martingale-envelope as well.

**Proof.** The function $\varphi(t) = t^{\frac{\theta}{\theta'}}$ is concave and monotonically increasing for $t \geq 0$, and Jensen’s inequality for conditional expectations yields the result. \(\square\)

We now turn to the situation where we are only interested in the performance of a single flow, say $A_1$. The key element is the following theorem:

**Theorem 6.** (Scheduling) In the scenario from Theorem 4, the following sample-path bound holds for all $l \geq 0$ and $\sigma > 0$

$$P\left( \sup_{0 \leq m \leq n - l} \{A_1(m, n - l) + A_2(m, n) - C(n - m)\} \geq \sigma \right)$$

$$\leq \frac{E[h_1(a_0)] E[h_2(a_0)]}{h_1 \ast h_2(\tau_{A_1 + A_2, C_1 + C_2})} e^{-\theta (\sigma + C_1 l)}.$$ 

**Proof.** Clearly, the process

$$(h_1(a_n^1) h_2(a_n^2) e^{\theta (A_1(l, n-l) - C_1 (n-l))} e^{\theta (A_2(0, n) - C_2 n)})_{n \geq 1}$$

is a supermartingale as well. The rest is similar to the proof of Theorem 4. \(\square\)

In the theorem, the shifting parameter $l$ depends on the scheduling policy. Using the service processes for $A_1$ (see also [3]), we obtain the following bounds on the virtual delay (i.e., on $P(W(n) \geq k)$):

- **FIFO:**
  $$\leq \frac{E[h_1(a_0)] E[h_2(a_0)]}{h_1 \ast h_2(\tau_{A_1 + A_2, C_1 + C_2}) e^{-\theta C_k}}$$

- **SP:**
  $$\leq \frac{E[h_1(a_0)] E[h_2(a_0)]}{h_1 \ast h_2(\tau_{A_1 + A_2, C_1 + C_2}) e^{-\theta C_1 k}}$$

- **EDF:**
  $$\leq \frac{E[h_1(a_0)] E[h_2(a_0)]}{h_1 \ast h_2(\tau_{A_1 + A_2, C_1 + C_2}) e^{-\theta (C_k - C_2 \min(k,y))}}.$$
For SP, $A_1$ has lower priority than $A_2$. For EDF, $d_1$ and $d_2$ are the relative deadlines of $A_1$ and $A_2$, and $y := d_1 - d_2$ is assumed positive; the case $y < 0$ can be treated similarly.

3. APPLICATIONS

3.1 Processes with independent increments

The simplest model is clearly given by a process with independent increments: Let $a_1, a_2, \ldots$ denote nonnegative i.i.d. random variables. The arrival process is thus $A(m, n) = \sum_{k=m+1}^{n} a_k$. Let a capacity $C > 0$ satisfying

$$E[a_1] < C < \sup a_1,$$

to avoid trivial scenarios. Now, for $\theta > 0$,

$$E[e^{\theta(A(n+1)-(n+1)C)} | a_1, \ldots, a_n] = e^{\theta(A(n)-nC)} E[e^{\theta a_1}] e^{-\theta C}.$$  

Due to the first stability condition we know that

$$\frac{d}{d\theta} E[e^{\theta a_1}] |(0) = E[a_1] < C = \frac{d}{d\theta} e^{\theta C}(0).$$

Due to the second condition, $E[e^{\theta a_1}]$ will eventually become larger than $e^{\theta C}$, so that there is $\theta_0$ such that both functions are equal, and $e^{\theta(A(n)-nC)}$ is a martingale. Thus, $A(n)$ admits a $(1, \theta_0, C)$-martingale-envelope.

3.2 Markov-modulated on-off processes

We consider the following situation: The source is driven by a Markov chain $(a_n)$ jumping between the two states 0 and 1 with probabilities $p$ and $q$, respectively. While in state 1 it transmits data at rate $R$. The arrival process is thus given by $A(m, n) = \sum_{k=m+1}^{n} R_{a_k}$. To ensure the stationary increment process we assume that $(a_n)$ is in steady-state, i.e.,

$$\pi_0 := P(a_1 = 0) = \frac{q}{p+q}, \quad \pi_1 := P(a_1 = 1) = \frac{p}{p+q}.$$

Again, the flow arrives at a constant-rate server with capacity $C$. To avoid trivial scenarios, we assume that $R > C$ and $\pi_1 R < C$.

For $\theta > 0$, we consider a modification of the transition matrix $T$ of $(a_n)$ (see [1]):

$$T_\theta := \begin{pmatrix} 1-p & pe^{\theta R} \\ q & (1-q) e^{\theta R} \end{pmatrix},$$

where $\lambda(\theta)$ denotes the maximal positive eigenvalue of $T_\theta$, and $v = (v_0, v_1)$ is a corresponding positive eigenvector. It can be shown (cf. [1]) that $v_0 < v_1$ if and only if $p + q < 1$, so that the function $h$ defined by

$$h(0) := v_0 \text{ and } h(R) := v_1$$

is monotonically increasing.

**Lemma 7.** Let $A$, $C$, $\lambda(\theta)$ and $h$ as above. Further assume that $\theta$ can be chosen such that

$$\lambda(\theta) = e^{\theta y}.$$  

Then $A$ admits a $(h, \theta, C)$-martingale-envelope.

The proof is based on a martingale construction from [5]. We now consider the case of multiplexing $N$ such queueing systems $(A_i, C)$. The threshold obviously changes to

$$\tau_{\Sigma_{i=1}^{N} A_i, NC} = R[NCR^{-1}].$$

![Figure 1: CCDF of the packet delay with $N_1 = \frac{1}{2}N = 10$, $p = 0.1$, $q = 0.5$, $R = 1$, utilization $\rho = 0.75$, and $d_1 = 10$, $d_2 = 1$.](image)

**Figure 1**: CCDF of the packet delay with $N_1 = \frac{1}{2}N = 10$, $p = 0.1$, $q = 0.5$, $R = 1$, utilization $\rho = 0.75$, and $d_1 = 10$, $d_2 = 1$. so that

$$h^{\otimes N}(\tau_{\Sigma_{i=1}^{N} A_i, NC}) = v_0^{-1}N^{CR^{-1}} v_1^{NCR^{-1}}.$$  

with

$$\kappa := \frac{\pi_0 v_0 + \pi_1 v_1}{v_0^{-1}N^{CR^{-1}} v_1^{NCR^{-1}}}.$$  

We thus have:

**Corollary 8.** In the multiplexed queueing system it holds for the aggregate flow

$$P(W(n) \geq k) \leq \kappa e^{-\theta NCk},$$  

and for a single flow comprising $N_1 < N$ subflows under scheduling

**FIFO:** $P(W(n) \geq k) \leq \kappa e^{-\theta NCk}$

**SP:** $P(W(n) \geq k) \leq \kappa e^{-\theta N_1 C k}$

**EDF:** $P(W(n) \geq k) \leq \kappa e^{-\theta(N_1 C (N-N_1)\min(k,y))}$,  

where $y := d_1 - d_2 \geq 0$.

We point out that while the bounds for the aggregate flow have been obtained in [1], the single flow bounds represent the contribution of the presented SNC approach. In Figure 1 simulations and the corresponding bounds for SP and EDF are displayed (the Martingale bounds are from Corollary 8 and the Standard bounds are based on Eq. (1)).

4. REFERENCES


