Correlated Arrival Processes in Switching Systems and some Applications

A. Brandt$^a$ and D. Weber$^b$

$^a$Humboldt-Universität zu Berlin, Institut für Operations Research, Spandauer Straße 1, D-10178 Berlin, Germany

$^b$Siemens AG - PN ST A 11, Hofmannstraße 51, D-81359 Munich, Germany

Abstract

Due to the specific interface between user and system, correlated arrival processes are the natural input stream offered to switching systems. As various publications in the past have shown, these correlations cannot be neglected when performance characteristics have to be computed. In our paper we derive - motivated by the description of so-called load profiles in switching systems - a second moment calculation for special correlated input streams and their superpositions. These type of processes can be used for better accuracy in approximations of system characteristics in switching systems. An example in the area of overload control performance is given.

1. INTRODUCTION

For predicting relevant performance characteristics of switching systems it is very important to use adequate models for the incoming stream of messages. Whereas the service process can be described very accurate on a message-by-message basis supported by appropriate measurement techniques [15], the input process is very much influenced by the user behaviour. Therefore it is necessary to describe calls from the user point of view. A typical call looks like as in Figure 1.

After the initial event with which the whole call is initiated (e.g. off hook) some events are generated to control the call setup (e.g. dialing of digits). When the call is properly set up (e.g. by B party going off hook) a relatively long period – the call holding time – follows. After that, some events are necessary to control the clear down phase of the call.

According to the fact that the number of users is relatively high, the assumption of a Poisson input for the initial event stream can be justified. But a Poisson assumption for the whole stream of incoming events generated by the users leads to underestimations for the performance characteristics of the system (c.f. [10, 12]). Therefore it is necessary to take into account the correlations of the event stream due to the dependency generated by the users.
Figure 1. Arrival process generated by a call

As described in more detail in [14], we characterize the input process of a switching system by a superposition of several call types 1, ..., M each of them consisting of $N_i$ $(i = 1, \ldots, M)$ events. For typical applications we have M about 6 and max $N_i$ about 20. As mentioned above the time instants of the initial events of type i calls are a Poisson process $\Phi_1^{(i)} = \{ T_{n+1}^{(i)} \}_{n=1}^{\infty}$ with intensity $\lambda_i$ $(i = 1, \ldots, M)$. The sequence $T_{n+1}^{(i)} < \ldots < T_{n}^{(i)}$ of events generated by the type i call arrival at $T_{n}^{(i)}$ is given by the spacings $\Delta_{n,j}^{(i)} = T_{n+1}^{(i)} - T_{n}^{(i)}$, $1 \leq j \leq N_i - 1$, where we assume that they may have a general distribution $F_{i,j}(t) = P(\Delta_{n,j}^{(i)} \leq t)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i - 1$ and that for fixed $i, j$ the sequence $\Delta_{n,j}^{(i)}$, $n = 1, 2, \ldots$ are i.i.d. Further $\Phi_1^{(i)}, \Delta_{n,j}^{(i)}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i - 1$ are independent random elements. Clearly the Point process (PP) $\Phi^{(i)} = \sum_{n=1, j=1}^{\infty} \delta_{T_{n,j}^{(i)}}$, where we have adopted the counting measure description for PP’s cf. [8], describes the input process initiated by the type i calls and the superposition process $\Phi_S = \Phi^{(1)} + \ldots + \Phi^{(M)}$ describes the complete arrival process, which is of a very complicated structure.

In the next section we derive the Laplace Stieltjes transform and in particular the second moment for the stationary interarrival-distribution of the events of the complete input process.

At the end of this paper we give, by approximating the arrival process by a renewal process and using a limit theorem of Dobrushin, an application in the field of practical performance analysis for SPC systems.

2. THE INTERARRIVAL TIME DISTRIBUTION OF THE CORRELATED ARRIVAL PROCESS

Since we are interested in the stationary interarrival distribution of the events it is convenient to start with a time stationary model of the input, which can be obtained by extending the Poisson processes $\Phi_1^{(i)}$ to the whole real line and the $\Delta_{n,j}^{(i)}$ to the negative $n$. Then, by construction and using the same symbols as in Section 1 we see that $\Phi^{(i)} =$
\( \{T_n^{(i)}\}_{n=\infty}^{\infty}, \ldots < T_0^{(i)} < 0 < T_1^{(i)} < \ldots \) and \( \Phi_S = \{T_n\}_{n=\infty}^{\infty}, \ldots < T_0 < 0 < T_1 < \ldots \) are stationary PP on the real axis. They can be considered as the time stationary model of the input, for details cf. [1], [5].

If we want to approximate \( \Phi_S \) by a renewal process, what we are doing later, we need the stationary inter event distribution of \( \Phi_S \), i.e. we need the distribution of the PP \( \Phi_S \) from an arbitrary chosen point of view. This distribution is given by the Palm distribution \( P^0 \) which describes the arrival process from a typical event point of view, cf. [1] or [5]. Let \( F(t) \) be the interarrival time distribution in the event stationary model and \( F(t) = P(T_1 \leq t) \) the time of the first event after zero in the time stationary model. From the inversion formula, cf. [5], p. 27 formula (1.2.21), it follows
\[
F(t) = \lambda_{\Phi_S} \int_0^t (1 - F^0(x)) \, dx, \quad m_{F^0} = \int_0^\infty x \, dF^0(x) = (\lambda_{\Phi_S})^{-1},
\]

where the intensity \( \lambda_{\Phi_S} \) of the PP \( \Phi_S \) is given by \( \lambda_S = \lambda_{\Phi_S} = \sum_{i=1}^M \lambda_i N_i \). (Note that (1) is the same relation as in case of a stationary renewal process and its ordinary version \( P^0 \).)

For the Laplace-Stieltjes transforms indicated by a star it follows
\[
F^*(s) = (1 - F^{0^*}(s))/ \left( sm_{F^0} \right).
\]

In particular we get for the second moment \( m_{F^0}^{(2)} = \int_0^\infty x^2 \, dF^0(x) \)
\[
m_{F^0}^{(2)} = 2m_{F^0} m_F, \quad m_F = ET_1 = \int_0^\infty P(T_1 > t) \, dt
\]
i.e. for a computation of \( m_{F^0}^{(2)} \) we need \( m_F \). Since \( \Phi_1, \ldots, \Phi_M \) are independent we have
\[
P(T_1 > t) = P(T_1^{(1)} > t) \cdots P(T_1^{(M)} > t)
\]
and thus we have to calculate the distributions \( F_i(t) = P(T_i^{(i)} \leq t) \).

2.1. The distribution \( F_i(t) \)

The distribution \( F_i(t) \) can be computed by exploiting a correspondence to a particular series network as follows: (\( i \) is considered as fixed) Let \( \delta_{T_{n,j}}^{(i)} \) be the PP of all \( j \)-th events belonging to type \( i \) calls. Consider the series network \( \Phi_1^{(i)} | F_{i,1}| \infty \rightarrow | F_{i,2}| \infty \rightarrow \ldots \rightarrow | F_{i,N_i-1}| \infty \rightarrow \) consisting of \( N_i - 1 \) infinite server nodes with service time distribution \( F_{i,j} \) at the \( j \)-th node and the arrival process \( \Phi_1^{(i)} \) of the calls at node number one. In view of \( T_n^{(i)} = T_n^{(i)} + \Delta_n \) the PP \( \Phi_{j+1}^{(i)} \) corresponds to the departure process at node \( j \). From this and since \( \Phi_1^{(i)} \) is a Poisson process it follows that \( \Phi_{j}^{(i)} \) and inductively that the \( \Phi_{j}^{(i)} \), \( j = 2, \ldots, N_i \) are Poisson processes with intensity \( \lambda_i \) too, cf. e.g. [11], [13]. In particular the partial superposition processes \( \Phi_{k,m}^{(i)} := \Phi_k^{(i)} + \ldots + \Phi_m^{(i)}, 1 \leq k < m \leq N_i \) are of
the same nature as the original PP \( \Phi_k^{(i)} \), but with the arrival process \( \Phi_k^{(i)} \) and spacing distributions \( F_{i,k}, \ldots, F_{i,m-1} \).

Now let \( N_j(t) \) be the number of busy servers at \( j \) and \( X_j(t) = (X_{j,1}(t), \ldots, X_{j,n_j(t)}) \), \( j = 1, \ldots, N_i - 1 \) the vector of the residual service times at node \( j \) provided \( N_j(t) = n_j \) (we suppress here the index \( i \)). The product form solution for this series network cf. e.g. [11], [13] yields for each \( n = (n_1, \ldots, n_{N_i-1}) \) and \( x = (x_1, \ldots, x_{N_i-1}), x_j = (x_{j,1}, \ldots, x_{j,n_j}) \):

\[
P(X_j(t) \geq x_j, N_j(t) = n_j, j = 1, \ldots, N_i - 1) = \prod_{j=1}^{N_i-1} \left( \frac{\rho_{ij}^{n_j}}{n_j!} e^{-\rho_{ij}^{n_j}} \prod_{k=1}^{n_j} \tilde{F}_{ij}^i(x_{jk}) \right),
\]

where \( \rho_{ij} = \lambda_i \cdot m_{F_{ij}}, m_{F_{ij}} = \int_0^\infty xdF_{ij}(x), F_{ij}^i(u) = 1 - \tilde{F}_{ij}^i(u) = \frac{1}{m_{F_{ij}}} \int_0^u (1 - F_{ij}(s))ds \) and \( x \geq y \) iff \( x_j \geq y_j \) for \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \). The first jump of the right continuous stationary state process of the series network \( X(t) = (X_1(t), \ldots, X_{N_i-1}(t)) \) provides \( T_i^{(i)} : T_i^{(i)} = \inf\{ t \geq 0 : X(t) \neq X(t-) \} \). Then we get from (5) and since the arrival process is Poissonian:

\[
\tilde{F}_i(t) = 1 - F_i(t) = P(T_i^{(i)} > t) = e^{-\lambda_i t} \prod_{n=1}^{N_i-1} \sum_{j=1}^{n_j-1} \frac{\rho_{ij}^{n_j}}{n_j!} e^{-\rho_{ij}^{n_j}} \left( \tilde{F}_{ij}^i(t) \right)^{n_j}.
\]

\[
= e^{-\lambda_i t} \prod_{j=1}^{N_i-1} \sum_{n_j=0}^{\infty} \frac{\rho_{ij}^{n_j}}{n_j!} e^{-\rho_{ij}^{n_j}} = \exp \left( -\lambda_i t - \sum_{j=1}^{N_i-1} \rho_{ij}^{n_j} \tilde{F}_{ij}^i(t) \right).
\]

\[\text{(6)}\]

### 2.2. The distribution \( F(t) \)

From (4) and (6) we get

\[
\tilde{F}(t) = 1 - F(t) = e^{-\lambda t - g(t)}
\]

where

\[
g(t) = \sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \rho_{ij} \tilde{F}_{ij}^i(t), \quad \lambda = \sum_{i=1}^{M} \lambda_i.
\]

From (7) one can derive the Laplace-Stieltjes transform \( F^*(s) \) and hence \( F^{0*}(s) \), cf. (2):

\[
F^*(s) = \int_0^{\infty} (\lambda + \sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \frac{\rho_{ij}}{m_{F_{ij}}} (1 - F_{ij}(t))) e^{-\lambda s + g(t)} dt.
\]

For the second moment \( m_{F_0}^{(2)} \) we get from (3), (1) and (7)

\[
m_{F_0}^{(2)} = \frac{2}{\lambda s} \int_0^{\infty} e^{-\lambda t - g(t)} dt = \frac{2}{\lambda s} \lambda \int_0^{1} e^{-st - \frac{ln u}{\lambda}} du
\]

### (8)
where we have used the substitution $u = e^{-\lambda t}$. Since the integrand in (8), which we denote by $f(u)$, is monotonously increasing, we get for the lower and upper Riemann-sums

$$S_n = \sum_{k=0}^{n-1} \frac{1}{n} f\left(\frac{k}{n}\right), \quad \bar{S}_n = \sum_{k=1}^{n} \frac{1}{n} f\left(\frac{k}{n}\right),$$

respectively, the following estimation $\bar{S}_n - S_n = \frac{1}{n} (f(1) - f(0)) = \frac{1}{n} (1 - \exp(-\sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \rho_{ij}))$. Thus we have an estimate for the error if we apply a numerical integration in (8).

In case of exponentially distributed spacings, i.e. $F_{ij}(t) = F_{ij}^e(t) = 1 - e^{-\nu_{ij} t}$ it follows in particular

$$m_{F_0}^{(2)} = \frac{2}{\lambda_S \lambda} \int_0^1 \exp \left(- \sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \rho_{ij} (1 - u^{\frac{1}{\nu_{ij}}}) \right) du. \quad (9)$$

### 2.3. Asymptotically independence

Let $r_i$ be the indices of the spacings corresponding to the holding times of the type $i$-calls, i.e. by assumption the $\Delta_{n,\xi}^{(i)}$ are of a magnitude larger than the other $\Delta_{n,j}$, $j \neq r_i$. In this case formula (8) is not well suited for a numerical integration.

It can be shown (cf. [2]) that then the arrival processes

$$\xi_1^{(i)} = \Phi_1^{(i)} + \ldots + \Phi_{r_i}^{(i)} \quad \text{and} \quad \xi_2^{(i)} = \Phi_{r_i+1}^{(i)} + \ldots + \Phi_{N_i}^{(i)}$$

are "nearly" independent. More precisely they are asymptotically independent if one considers a series scheme $\Delta_{n,r_i}(l)$ with $F_{r_i}^{(e)}(t) = P(\Delta_{n,r_i}(l) \leq t) \rightarrow 0$, $t \in \mathbb{R}_+$. One gets in this case by using the "nearly" independence of $\xi_1^{(i)}$, $\xi_2^{(i)}$ and (7) for each of the $\xi_1^{(i)}$, $\xi_2^{(i)}$ separately

$$m_{F_0}^{(2)} \approx \frac{2}{\lambda_S} \int_0^\infty (1 - F(t)) dt$$

$$= \frac{2}{\lambda_S} \int_0^\infty \exp \left(- \lambda t - \sum_{i=1}^{M} \sum_{j=1}^{r_i-1} \rho_{ij} F_{ij}^e(t) \right) \cdot \exp \left(- \lambda t - \sum_{i=1}^{M} \sum_{j=r_i}^{N_i-1} \rho_{ij} F_{ij}^e(t) \right) dt$$

$$= \frac{2}{\lambda_S} \int_0^\infty \exp \left(- 2\lambda t - \sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \rho_{ij} F_{ij}^e(t) \right) dt$$

$$= \frac{1}{\lambda_S \lambda} \int_0^1 \exp \left(- \sum_{i=1}^{M} \sum_{j=1}^{N_i-1} \rho_{ij} F_{ij}^e(-\frac{\ln u}{2\lambda}) \right) du \quad (10)$$

which can be computed numerically well. The approximation (10) can also be obtained by letting $l \rightarrow \infty$ in (8).
3. APPLICATIONS: OVERLOAD CONTROL

In this section we would like to show briefly how the derived results can help to get better solutions to practical planning problems in switching systems. It is clear that with this approach all results for two-moment approximations in queueing systems are available, such as e.g. GI/G/1 delay formulae (see [9]). We would like to focus in the following to a special planning problem in the area of overload control mechanisms.

Overload control mechanisms in SPC systems have to follow general performance objectives [6]. One attribute of such overload control strategies is the so-called selectivity. Selectivity tells how precise a given strategy can distinguish a real overload situation from "normal" load fluctuations in time. An ideal overload control would react if and only if a load situation occurs that will not be tractable for the system with given grade of service criteria e.g. limitation for specific response times. Because of the random nature of the underlying processes a given overload strategy reacts with a certain probability at any load. The task is to determine the probability of predicting overload under normal or planning load level. This is what we call selectivity.

We consider in the following an overload control mechanism with processor load indication as described in [6]. Therefore overload is predicted whenever the average CPU utilization during the last $T$ seconds exceeds a given threshold $x$ (CPU utilization in this context only counts for activities processed with appropriate high priority levels). The problem is to determine the distribution of the average CPU utilization over a given time interval $T$.

Let us first consider the limiting cases $T \to 0$ and $T \to \infty$ and a planning load of 70 %. We get the following situation:

![Figure 2. Limiting cases $T \to 0$ and $T \to \infty$.](image)

For both cases we get a discrete distribution. For $T \to 0$, i.e. averages taken from an infinitesimal short time period yield a two-point distribution. We predict 100 % load with probability 0.7 and 0 % load with probability 0.3, because at any given point in time the CPU is either busy or idle. Therefore, any threshold $x$ yields a poor selectivity of 70 %. For $T \to \infty$, i.e. averages taken from an infinite long time period we get a one-point
distribution at planning load level 70%. Therefore any threshold \( x \geq 70\% \) yields to the ideal selectivity of 0 %.

In realistic applications it is not possible to average over very large time intervals because then a suddenly starting overload situation will be discovered too late. Therefore one has to look for an appropriate choice of \( T \) between 0 and \( \infty \). Let \( S_n \) be the work associated with the n-th event of \( \Phi_s \), which are assumed to be i.i.d. r.v.'s. Approximating the arrival process \( \Phi_s \) by a renewal process with the interarrival distribution \( F^0(t) \), we get from a general limit theorem for sums of random variables given by Dobrushin [4], cf. also [3], pp. 173 that the load \( L(T) = \sum_{n=1}^{\infty} S_n \mathbb{I}\{0 \leq T_n \leq T\} \) is approximately normally distributed with the mean \( \rho T \) and variance \( \rho m_b^{(1)} T (c_a^2 + c_b^2) \), where \( c_a^2 = \frac{m_a^{(2)}}{(m_P)^2} - 1 \), \( c_b^2 = \frac{m_b^{(2)}}{(m_b^{(1)})^2} - 1 \), \( m_i^{(i)} = ES_i^i, i = 1, 2 \), \( \rho = \lambda s m_b^{(1)} \) is the planning load and \( T \) is the observation interval.

For various load profiles this approximation has been tested against simulation, where the coefficient of variation for the arrival process is computed according to formula (8) of Section 2. Simulations were performed using specific GPSS simulation techniques. For details see [7].

For typical parameters (load range from 60 to 80 %, averages from 1 - 3 second time intervals and a call service rate from 5 to about 30 per observation interval) this approximation has been tested. The following figures show typical results for parameter settings as follows:

- load \( \rho = 0.71 \)
- call rate \( \lambda = 13/s \)
- number of events per call \( N_i = 8 \)
- spacings \( F_i \), exponentially distributed with mean:
  - 1.8; 0.6; 0.6; 0.6; 6; 50; 2s \( i = 1, \ldots, 7 \).

This refers to a digital internal call with four-digit dialling (see [14] for details).

![Figure 3. Density function of cumulated load values for 1s load averages](image)
Figure 4. Density function of cumulated load values for 3s load averages

For example $x = 0.95$ yields a selectivity of 3.7 % (appr.) and 3.2 % (sim.) for $T = 1s$ and $2.10^{-6}$ (appr.) and 0 (sim.) for $T = 3s$, respectively. In practical planning cases the approximation using Dobrushins formula shows good accuracy.

4. CONCLUSIONS

Motivated by the description of input processes on an event-by-event basis and their corresponding load profiles for the switching processor, we consider a general model for correlated input streams in switching systems. The Laplace Stieltjes transform and in particular the second moment for the stationary interarrival-distribution of the whole process are derived. As application in the field of practical performance analysis for SPC systems, we approximate the arrival process of an SPC system by a renewal process, use a limit theorem of Dobrushin and derive performance characteristics related to overload control mechanisms.

Acknowledgements

The computation and simulation efforts of Dr. Alexandr Makarow are greatly appreciated.

References


