ACCURACY OF ESTIMATING CONGESTION QUANTITIES FROM SIMULATION RESULTS

L. KOSTEN
Technical University, Delft, The Netherlands

ABSTRACT

The analytical determination of congestion quantities in large service systems mostly is impossible owing to complexity of the system. Mostly, simulation is used. Determination of the accuracy of the measurement from variance in results of consecutive runs may result in too optimistic results owing to correlation (the so-called "cluster-effect").

In the past this effect has been analysed for the standard M/M/c cases. A more general theory is presented here. The older results (which are nearly inaccessible) are reviewed at hand of the present theory.

1. INTRODUCTION

Many large service systems, such as occur in telecommunication and datahandling [1,7] may be described as aggregates of flows* of arrivals (demands for service), a set of servers and possibly queues. From a traffic point of view such a system may be specified by a large ergodic stochastic system, supposedly stationary. The state of occupation of the servers and the number of items in the queues together define a state of the system. When the queues are of limited size, the number of states of the system is finite.

During periods in which arrivals are relatively frequent and/or in which service-times are relatively large, arrivals may experience some (mild or severe) form of hindrance, such as: (i) not being served at all (they are lost, the system is said to be blocked for those arrivals); (ii) delay of service; (iii) the test procedure for finding a free server lasts too long. Such hindrance will in general be called congestion. In order to judge the system's performance with a view to congestion, some measure is needed, to be called overall congestion C. It may refer to all arrivals or only to those from those flows* whose service-times are relatively large, arrivals may experience some hindrance will in general be called congestion.

Most definitions of overall congestion C are either: (i) the arrival-expectation of the number of queued arrivals when the system is in state i, the number of tests necessary (the hunt) is denoted by Ea*{9k} (put into a queue; \( 9_k = 0 \)) or delayed (\( 9_k = \infty \)), the overall congestion C now is the probability of delay.

b. Consider a delay system. When the state is \( \mathcal{L} \), an arrival of flow \( A_k \) is either given service immediately (put into a queue; \( 9_k = 0 \)) or delayed (\( 9_k = \infty \)). The overall congestion C now is the probability of delay.

c. Consider again a delay system. Let \( q_i \) be the total number of queued arrivals when the system is in state \( \mathcal{L} \). Let be \( C \equiv E[q_i] \). Then \( C/p \) is the expected waiting-time.

d. Consider some (possibly very complex) blocking system (e.g. a "grading", [1,2]). Arrivals of flow \( A_k \) test (in telephone parlance "hunt") a subset \( H_k \) of \( q^* \) servers in a certain order, until either a free server has been found or else \( H_k \) is found not to contain a free server. When the system is in state \( \mathcal{L} \), the number of tests necessary (the hunt) is denoted by \( 9_k \). Then \( C \equiv E^*\{9k\} \) is the expected hunt.

Figure 1, which is considered self-explanatory, shows some simple cases.

In principle the states \( i \) can be markovised, when necessary, by the introduction of a sufficient number of supplementary variables. The solution of the Chapman-Kolmogorov equations then may yield C. Practically this mostly is precluded by the large size of the system. Hence, it is customary to assume all arrival flows to be Poisson processes and all service-times to have exponential, Erlang-k or hyperexponential distributions. This reduces the total process to a Markov process \( M \) with say \( m \) states. Let \( A \) be the infinitesimal generator and the stationary probability vector. Then:

\[
\begin{align*}
\lambda^A &= 0 \quad ; \quad \mu^e = 1 \\
{\begin{array}{c}
{\begin{array}{c}
\lambda^A &= 0 \\
\mu^e &= 1
\end{array}}
\end{array}}
\end{align*}
\]

(1.1)

Now, consider the case of C being defined as \( E^m\{9_k\} \). As the flows are Poissonian, the probability of an arrival stemming from \( A_k \) is \( \lambda^A/p \), independent of the state \( \mathcal{L} \). Hence, the arrival-expectations of \( 9_k \) and of \( 9_k \) defined by:

\[
q_k \triangleq \lambda^A/p \quad \mu^e \quad C_k \triangleq \frac{q_k}{\mu^e}
\]

(1.2)

are equal. Moreover, they equal the time-average of \( 9_k \). With:

\[
q_k \triangleq (q_1, \ldots, q_m)
\]

(1.3)

the following equalities exist:

* Vectors are vectors in state-space. The vector \( \xi \) is the unitvector: \( \xi = (1, \ldots, 1) \). Products of vectors are scalars. Hence, no distinction need be made between row- and column-vectors.

* Flow is equivalent to point-process.
Figure 1: some examples of congestion models.

\[ C \triangleq E^*[J_{ik}] = E^*[J_k] = E^*[1_{ik}] \quad \text{(I+II)} \]

I, II and III can be considered three equivalent definitions of \( C \) in this case. When III or II is the original definition, \( C \) can be given a meaning by taking \( \mathbb{P}[i] \triangleq \mathbb{P}[i] \). \( \mathbb{P}[i] \).

Now, even after reduction of the process to a Markov process, its size mostly prohibits analytical or numerical treatment. The only possibility then is simulation. By simulation a realisation of the Markov process is formed for the interval \((0,t]\) starting at time \( t \) for a certain extent. For the said interval the order of events (i.e. transitions in the Markov process and arrivals not resulting in transitions) as well as inter-event durations then are known. From the three definitions of \( C \) we immediately obtain three possible estimates of \( C \), viz. I) as the observed arrival-average of \( J_{ik} \); II) ditto of \( J_k \); and III) as the observed time-average of \( J_k \) for the interval \((0,t]\).

The three estimates use different parcels of knowledge contained in the realisation. Method I uses the values of \( J_{ik} \) at arrivals only. Method II also uses the values of \( J_{ik} \) that would have been found when the actual arrivals had been replaced by arrivals from other flows (blended in the correct proportions). Finally, method III in addition also uses the inter-event times. Hence, method III will be the most accurate, method I the least. Method III should always be preferred. There may, however, be impediments. In general, the system will be so large that it is impossible to construct lists of \( J_{ik} \) or \( J_k \). Hence, they should be constructed ad hoc from the configuration and the stored conditions of servers and queues. Now, the construction of \( J_{ik} \) in principle takes \( m \) times as long as that of \( J_k \) (cf. 1.2). This sometimes will rule out methods II and III. This leaves us with I and III as useful alternatives, to be called the count-method and the continuous observation method.

Simulation is time-consuming and hence expensive. Estimation of the accuracy of results is utterly important. The accuracy of the estimates can be judged from their coefficients of variation (i.e. standard deviation divided by expectation; abbreviation c.o.v.). As simulation is undertaken in view of the impossibility of obtaining analytical results for expectations, analytical investigation of variances and c.o.v.'s as a rule is precluded a fortiori. Hence, accuracy mostly is judged from variances in series of results for consecutive simulation runs. As congestion phenomena tend to occur in bursts or clusters, those run results may be heavily correlated. Neglecting this clustering effect may result in a far too optimistic estimate of the accuracy. It is worth while to have the disposal of an analysis of this effect in some simple cases. The c.o.v.'s of \( C_e(t) \) and \( C_a(t) \) equal those of

\[ N(t) \triangleq \text{the number of arrivals in } (0,t] \]
\[ R(t) \triangleq \sum_{N(t)} \frac{J_{ik}}{N(t)} \text{ all } N(t) \text{ arr.} \]

Then the observed arrival congestion is defined by:

\[ C_e(t) \triangleq \frac{R(t)}{N(t)} \quad \text{or} \quad C_e(t) \triangleq \frac{R(t)}{pt} \]

These are two estimates of \( C \), the second one unbiased, the first one probably the more accurate.

CONTINUOUS OBSERVATION METHOD. Let \( J(t) \) be the state at time \( t \).

When we define:

\[ U(t) \triangleq \int_0^t J(t') dt' \]

the observed time congestion \( C_e(t) \):

\[ C_e(t) \triangleq \frac{U(t)}{t} \]

again is an (unbiased) estimate of \( C \).
...and investigated the accuracy of the C(t) methods in general systems; and (ii) the accuracy of continuous observation results for adjacent intervals. The older results have been reviewed by Kosten as a sideline in another context [10]. The present paper contains: (i) an extension to count methods in general systems by a limiting procedure applied to (i). Moreover, it also deals with higher order statistics and with correlation of observation results for adjacent intervals. The older results [7,8] will be reviewed at hand of the present theory. During the preparation of the present paper a report containing a rather general approach [10] was received. The present paper, though less general, enters in more analytical detail.

2. THE COUNT METHOD

In the following \( J \) denotes the state of the system prior to an arrival and \( f \) the state afterwards. At an arrival from \( A_k \) the state \( I = J \) determines: (i) the congestion weight \( g_J \) and (ii) the new state \( f = A_k \). Let \( g_J \) be the total density of arrivals of which, when occurring during state \( J \), cause transition to state \( f \) whilst having a congestion weight \( g (\cdot, \cdot, \ldots, \cdot) \).

When \( J = j \) the congestion weight \( g \) is assumed to be \( > 0 \). We define \( g \) in the matrix:

\[
G = \{ g_{j,k} \} \quad (g = 1, \ldots, m)
\]

(2.1)

Let \( I(t) \) be the state of the original Markov process at time \( t \) and \( F(t) \) the total count during \( (0, t] \). Now, consider the compound Markov process \( \{X(t), Y(t) \} \) with state space \( \{0, 1, \ldots, \} \times \{0, 1, \ldots, \} \). Of course \( X(t) = 0 \), whilst the distribution of \( Y(t) \) is the stationary one. Let:

\[
\int f(t, e) = F(R(t)) = \int \int \int f(t, e) I(e) \quad (i = 1, \ldots, m)
\]

(2.3)

and:

\[
\int f(t, e) = \{ f(t, e), \ldots, f_m(t, e) \}
\]

(2.4)

The Chapman-Kolmogorov equations together with the initial conditions can be written in matrix-vector notation:

\[
\frac{df}{dt} = \{ f(t, t) + \sum_{j=1}^{\infty} \{ f(t, e, f_j(t, e)) \} \}
\]

(2.5)

We introduce the Generating Function:

\[
F(x, t) = \sum_{n=0}^{\infty} \{ f(t, e) x^n \}
\]

and subsequently the Laplace Transform:

\[
\mathcal{L}[F(x, t)] = \int_0^{\infty} e^{-ax} F(x, t) dt
\]

(2.7)

Equations (2.5) then yield (\( I = \) identity matrix):

\[
\mathcal{L}[\{ f(t, e) \}] = \sum_{j=1}^{\infty} \{ f(t, e) x^j \} \mathcal{L}[\{ f(t, e) \}]
\]

(2.10)

with:

\[
H^j = \sum_{k=0}^{\infty} \{ f(t, e) \} \mathcal{L}[\{ f(t, e) \}]
\]

(2.11)

The quantities \( f_j(t) \), \( j \geq 0 \) are determined recursively.

After the vectors \( \beta_j(t) \) have been determined, the Laplace transforms of the factorial moments \( \{ \mathcal{L}[x^j \{ f(t, e) \}] \} \) of \( f(t, e) \) are given by:

\[
L \{ \mathcal{L}[x^j \{ f(t, e) \}] \} = \{ \beta_j(t) \} \quad (j = 1, 2, \ldots)
\]

(2.13)

By induction it follows from (2.10, 12) that \( \beta_j(t) \) is rational, the rightmost pole being at \( -a^j \) with multiplicity \( 1 + j \). Hence, for small \( t \) values the following expansion exists:

\[
\beta_j(t) = \sum_{k=0}^{\infty} \frac{1}{k!} t^k \mathcal{L}[\{ f(t, e) \}]
\]

(2.15)

On account of (2.12) we have:

\[
\beta_j(t) = p^j \quad (j = 0, 1, \ldots)
\]

(2.16)

Insertion of (2.15) into (2.10) yields a relation between the coefficients of \( t^m \) of:

\[
Y(t, A) = \sum_{k=0}^{\infty} \{ x^k \} Y(t, e) H^k
\]

(2.17)

We equate \( Y(t, e) \) to zero when \( j < 0 \) or \( k < 0 \). As the
row-sums of $A$ are zero, the sum of the LHS members of (2.17) vanishes. The same must apply to the RHS members. Changing $A$ into $A^r$ this reads:

$$y_{ij} = \sum_{i=1}^{\infty} y_{i,j} - e_i^* h_j$$

The system of equations (2.17) has a defect 1. Equation (2.18) completes the system. The $y_{ij}$ are determined in the index order $n_1, n_2, n_3, \ldots$, $e_i^*$ being the largest $e$-value needed. For the sequel we define:

$$y^* h_j = \varphi (e) \quad (Y \text{ arbitrary vector}) \quad (2.19)$$

An asymptotic expansion for $\mu_{ij} \{R(t)\} \text{ as } t \to \infty$, may be obtained from (2.14, 15). As the $y_{ij}$ are rational, so are the LT in (2.13). The rightmost pole (i.e. $z = 0$) determines the asymptotic behaviour. We need only insert (2.15) into (2.14), delete the non-negative powers of $z$, and replace $z^{-2}$ by $e^{z^*/(2 \pi \alpha)}$. After some reshuffling this yields:

$$\mu_{ij} \{R(t)\} \sim \frac{1}{t^2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} y_{i,j} - e_i^* h_j$$

When $\varphi$ is the count at an arbitrary arrival, it follows from the definitions (2.1, 11) of $\varphi$ and $h_j$ that $\varphi \sum \{z \} p_{\varphi}$ equals the expectation of $\varphi$ subject to $X = e$. Hence:

$$E \{ \varphi \} = \varphi \sum \{ z \} p_{\varphi} = \varphi \beta (\alpha) \quad (2.21)$$

Especially:

$$\beta (\alpha) = \sum \{ z \} p_{\varphi} = \beta (\alpha) \quad (2.22)$$

is the average count per unit of time. The number of linear systems of equations (2.17, 18) to be solved can be reduced by some relations to be derived now. From (2.17, 18; $t \to \infty$ ) and (2.16) it follows by induction that:

$$y_{ij} \to \alpha^i \beta^j \quad (2.23)$$

From (2.17, 18, 23) it follows that:

$$y_{ij} A = \alpha^i \beta^j \{ \alpha J - H \} \quad (2.24)$$

The last member is replaced by zero for $j > 1$. By induction:

$$y_{ij} = \alpha^i \beta^j \{ 1 + \alpha^j \beta^j \} \quad (2.25)$$

By the use of (2.19, 22, 23, 25) it follows from (2.20) that:

$$\varphi \{ R(t) \} = \alpha t$$

$$\varphi \{ R(t) \} \sim \alpha^2 \beta^2 \{ 1 + \alpha + \beta \} \quad (2.26)$$

The first quantity is the correct value of the expected count in $\{ R(t) \}$. From the factorial moments the cumulants $\kappa_\nu \{ R(t) \}$ may be obtained, which are generated by (c.f. Wilks [13]):

$$\sum_{\nu=0}^{\infty} \kappa_\nu \{ R(t) \} e^t = E \{ e^{R(t)} \} \quad (2.27)$$

They are multinomials of the $\mu_{ij} \{ R(t) \}$ [17]. As errors in the asymptotic expressions (2.26) are finite sums of negative exponentials, asymptotic expressions for those cumulants may be obtained by simply introducing the relations (2.26) into the multinomials mentioned. This yields:

$$\kappa_\nu \{ R(t) \} = \alpha^\nu \beta^\nu \{ R(t) \} \quad \text{as } t \to \infty$$

It should be observed that the 'simplificatory relations' (2.23, 25) are needed for establishing the fact that those asymptotic expressions for $\alpha$ and $\beta$ do not contain higher powers of $t$. We especially state the asymptotic expressions:

$$\varphi \{ R(t) \} \sim \{ \alpha + \beta \} t$$

From (2.21) we have:

$$E \{ \varphi \} = \beta (\alpha) \quad \alpha \neq 0$$

When all counts were mutually independent samples of $\varphi$, the process at hand would reduce to a Poisson process with multiple occurrences (Cox and Miller [3]). The variance of $\varphi(t)$ then would be:

$$\text{var} \{ \varphi \} \sim \{ \alpha + \beta \} t$$

For $t \to \infty$ the ratio $\text{var} \{ \varphi \} / \{ \alpha + \beta \} t$ tends to a constant, to be denoted by $\eta^\phi$. The quantity $\eta^\phi$ will be called the coefficient of the cluster effect, as it clearly describes the reduction in accuracy of the measurement due to the effect of clustering of high congestion values $\varphi$. We eventually find:

$$\eta^\phi = 1 + 2 \frac{1}{\alpha + 2 \beta \varphi} \quad (2.32)$$

The definition of the coefficient of cluster effect adopted here differs slightly from the original one [6, 7]. Originally, the number of arrivals (counts) in the "independent" process had been taken to be constant $\varphi$; it is to be conjectured that $\eta^\phi$ is larger than 1, in any case for "normal" cases.

As: (i) corresponding cumulants for independent stochastic variables are additive, and (ii) count-totals $\mathcal{R}(t)$ and $\mathcal{R}(t)$ for large adjacent intervals ($\alpha, \beta$) and ($\alpha, \beta \varphi$) are intuitively asymptotically independent, it may conjectured that all cumulants $\kappa_\nu \{ R(t) \}$ are asymptotically linear in $\varphi$.

3. THE CONTINUOUS OBSERVATION METHOD

The results of this section will be obtained from those of section 2. We construct a hypothetical model based on section 2. The flows $H_1, \ldots, H_m$, though retaining their functions as possible state-changers, do no longer cause counts; formally they are no longer arrival flows. Instead new virtual arrival flows $\tilde{W}_1, \ldots, \tilde{W}_m$ are created with densities $\tilde{W}_1, \ldots, \tilde{W}_m$ (2.1 for the definition of $\varphi$). Virtual arrivals (now the arrivals) do not change the state, but may cause unit counts, viz. when $\tilde{W}_k$. Marking quantities in this hypothetical system by tildes ($\tilde{\cdot}$) we have:

ITC-9 KOSTEN-4
\[
\begin{align*}
\mathcal{A} & \triangleq \mathcal{A}_1; \quad \text{hence:} \quad \mathbb{E} = \mathcal{P} \\
\mathcal{P} & \triangleq \mathcal{P}_1; \quad \mathcal{H}_1 \triangleq \text{diag}\{q_1, \ldots, q_m\}; \quad \mathcal{H}_1^* = 0
\end{align*}
\]

Hence:
\[
\mathcal{H}_{1,2} = \mathbb{I}; \quad \mathbb{P}_1 = \mathcal{A}_1
\]

The counting-process defined above now is a doubly stochastical Poisson process (cf. Cox and Lewis [22]). When the state is \( i \), the density is \( q_i \). For a fixed realisation of \( \mathcal{M} \) on \( \mathcal{O}(t) \) has a Poisson distribution with expectation
\[
\int_{\mathcal{O}(t)} e^{-t} dt \quad \text{(cf. 1.7)}.
\]

Let \( \mathcal{F}_n(t) \) be the unknown c.d.f. of \( \mathcal{U}_n(t) \). Then the p.f. of \( \mathcal{F}_n(t) \) is:
\[
\mathbb{P}\{\mathcal{F}_n(t) = k\} = \int_0^t e^{-r} dr \mathcal{F}(t) = \mathcal{A}_n(t) \]

and its \( k \)-th factorial moment:
\[
\mathbb{P}_n\{\mathcal{F}_n(t) = k\} = \mathcal{A}_n(t)^k e^{-\mathcal{A}_n(t) t} \mathcal{F}(t)
\]

When the quantities defined in (3.1, 2, 3) are used in section 2, the asymptotic expressions for \( \mathbb{P}_n\{\mathcal{U}_n(t)\} \) now follows from (2.26):
\[
\begin{align*}
\mathbb{P}_1\{\mathcal{U}_0(t)\} & = \mathcal{A}_1 t / \mathcal{P} \\
\mathbb{P}_2\{\mathcal{U}_0(t)\} & = \mathcal{A}_1^2 t^2 / \mathcal{P}^2 + \mathcal{A}_1 t / \mathcal{P} \\
\mathbb{P}_3\{\mathcal{U}_0(t)\} & = 2 \mathcal{A}_1^3 t^3 / \mathcal{P}^3 + 2 \mathcal{A}_1^2 t^2 / \mathcal{P} \\
& + 2 \mathcal{A}_1 t / \mathcal{P} + 1 / \mathcal{P}
\end{align*}
\]

From these results the cumulants may be obtained. Especially:
\[
\begin{align*}
\text{var}\{\mathcal{U}_0(t)\} & \sim 2 \mathcal{A}_1^2 t / \mathcal{P}^2 \\
\text{cov}\{\mathcal{U}_0(t)\} & \sim 2 \mathcal{A}_1 t / \mathcal{P}^2
\end{align*}
\]

It is to be conjectured that the latter result is less than \( \text{c.o.v.} \{\mathcal{F}_n(t)\} \) (cf. 2.29). When probabilities of blocking are estimated, one has
\[
\mathcal{H}_1 = \mathcal{H}_1^*; \quad \mathcal{P}_1^* = \mathcal{A}_1^* / \mathcal{P}_1; \quad \mathcal{P}_1^* = 0
\]

\[
\text{cov}\{\mathcal{R}(t)\} = \text{c.a.r.}\{\mathcal{U}_0(t)\} + \mathcal{A}_1 / \mathcal{P}
\]

a result that can be obtained by simpler means.

4. CORRELATION OF RESULTS FOR DIFFERENT INTERVALS

Let \( \mathcal{R}_1, \mathcal{R}_2 \) and \( \mathcal{R}_3 \) be the counts for the intervals \( (0, t) \), \( (t_1, t_2) \) and \( (0, t_2) \), respectively. Then

\[
\begin{align*}
\mathcal{R} & = \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 \\
\text{var}\{\mathcal{R}\} & = \text{var}\{\mathcal{R}_1\} + \text{var}\{\mathcal{R}_2\} + 2 \text{cov}\{\mathcal{R}_1, \mathcal{R}_2\}
\end{align*}
\]

or:
\[
\text{cov}\{\mathcal{R}_1, \mathcal{R}_2\} = \frac{1}{2} \left[ \text{var}\{\mathcal{R}(t_2)\} - 2 \text{var}\{\mathcal{R}(t)\} \right]
\]

When the asymptotic expressions (2.28) for \( \text{var}\{\mathcal{R}\} \) is used, we obtain:
\[
\text{cov}\{\mathcal{R}_1, \mathcal{R}_2\} \sim - \lambda (t_1)
\]

In section 2 we introduced the coefficient of clustereffect (2.32) to describe the effect of clustering. It seems that this effect may also be described by the covariance mentioned.

When \( t \to \infty \), the variances increase asymptotically linearly. Hence, the relative influence of an asymptotically constant covariance becomes less and less.

In the same way, by considering the results \( \mathcal{R}_1, \mathcal{R}_2 \) and \( \mathcal{R}_3 \) for three consecutive intervals of duration \( t \) it may be shown that \( \text{cov}\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\} \) tends to zero for \( t \to \infty \).

5. THE STANDARD M/M/C CASES

The numerical implementation of the foregoing theory is very easy in the present computer age. Hence, further analysis for the standard M/M/c cases would hardly be advantageous, were it not that it offers the possibility of obtaining rules of thumb.

Let \( \mathcal{R}(t) \) be the number of lost arrivals in \( (0, t) \) for the M/M/c blocking system. There is one flow of arrivals \( \lambda = \lambda \) with density \( \lambda \). The service-time has an exponential distribution with average \( 1 \). When all \( c \) servers are occupied, new arrivals are lost without further consequence. The state of the system is indicated by the number of occupied servers \( v \in \{0, \ldots, c\} \).

The data are:

\[
A = \text{threeband matrix}
\]

\[
\begin{align*}
\text{upper:} & \quad \mathcal{A}_1 \quad \mathcal{A}_2 \quad \mathcal{A}_3 \\
\text{lower:} & \quad \mathcal{0} \quad \mathcal{A}_1 \quad \mathcal{A}_2 \quad \mathcal{A}_3 \quad \mathcal{0}
\end{align*}
\]

\[
\mathcal{H}_1 = \mathcal{P} \text{ diag}\{q_1, \ldots, q_n\}; \quad \mathcal{H}_1^* = 0
\]

The system of equations (1.1) now reads:
\[
\begin{align*}
\mathcal{P} \mathcal{A}_1 & + (\mathcal{P} - \mathcal{A}_1) \mathcal{A}_2 + (\mathcal{P} - \mathcal{A}_1) \mathcal{A}_3 + \mathcal{P} \mathcal{A}_3 = 0 \\
& + \ldots + \mathcal{P} \mathcal{A}_n = \mathcal{I}_v = 1
\end{align*}
\]

Its solution is:
\[
\mathcal{K}_v = \mathcal{K}^C \mathcal{P}_v = \mathcal{K}^C \mathcal{P}_v
\]

Hence:
\[
\alpha = \beta(1) = \mathcal{K}^C \mathcal{P}_v \quad \beta(1) = 0
\]

Replacing \( \lambda \) by \( \beta \) equations (2.24, \( j = 1 \)) now read:
\[
\begin{align*}
\mathcal{P} \mathcal{A}_1 & + (\mathcal{P} - \mathcal{A}_1) \mathcal{A}_2 + (\mathcal{P} - \mathcal{A}_1) \mathcal{A}_3 + \mathcal{P} \mathcal{A}_3 = \alpha \mathcal{K}^C \\
& + \ldots + \mathcal{P} \mathcal{A}_n = \mathcal{I}_v = 1
\end{align*}
\]

Routine use of Generating Functions yields:
\[
\mathcal{K}^C = \alpha \mathcal{K}^C \mathcal{P}^c - \alpha \mathcal{K}^C \mathcal{P}^c \sum_{i=0}^{\infty} \mathcal{P}^i
\]

Hence:
\[
\lambda(1) = \mathcal{P} \mathcal{A}_1 = \alpha \mathcal{K}^C \mathcal{P}^c - \alpha \mathcal{K}^C \mathcal{P}^c \sum_{i=0}^{\infty} \mathcal{P}^i
\]

The results (5.6, 10) are sufficient for calculating expectation and variance of \( \mathcal{R}(t) \) according to (2.29). They agree with previous ad hoc results [10]. Also the coefficient of clustereffect (2.32) can be obtained.

Now, consider the case of a large system \( (\mathcal{P}\gg\mathcal{K}) \) in which blocking is a rare event \( (\mathcal{K}/(1 - \mathcal{P}) \ll 1) \).

This is the case when e.g. \( \mathcal{P} = \mathcal{P} + 1/\mathcal{P} \). The last term in (5.9) then may be deleted. The quantity \( \mathcal{K}/(1 - \mathcal{P}) \) is the expectation of \( \mathcal{V}(t)/\mathcal{P} \). As the distribution function \( \mathcal{K}/(1 - \mathcal{P}) \) is sharply peaked around \( \mathcal{P} \) and \( \mathcal{V}(t) \) is a rather flat function in this peak region, said expectation may be
approximated by $\sqrt{(c-p)}$. Hence, $\frac{1}{V(c-p)} \approx \frac{x}{c-p}$. Then the coefficient of clustereffect may be approximated by:

$$\frac{f}{\sqrt{(c-p)}}.$$

The practical validity has been thoroughly tested.

Now, consider a M/M/c delay system with infinite queue. Let $A_0$ be either (i) the number of (new) arrivals during a delay, or (ii) the sum-total of queue lengths encountered by those arrivals. Division by $p/c$ yields estimates of the probability of delay and of the mean queue length, respectively. The c.o.v. of $R(t)$ for both cases has been obtained previously by ad hoc methods [7]. Those results have recently been verified by P. Kaland, using the present theory. Under the conditions for "rare" congestion $\frac{p}{c} > 1$, $\frac{q}{c} < 1$, the coefficient of clustereffect may be approximated by:

(i) measurement of prob. of delay (cf. [7]):

$$f \approx \frac{(c+p)}{(c-p)}$$  \hspace{1cm} (5.11)

(ii) measurement of mean queue length:

$$f \approx \sqrt{c+cp+p^2/(c-p)}$$  \hspace{1cm} (5.12)

easily obtainable from results in [7].

Results for the accuracy of $U(t)$ estimates in [6,7] are presently being verified.

### ROULETTE/SIMULATION

In Roulette Simulation [4,5,9] instead of a realisation of the continuous $A$-process a sequence-true series of events, embedded in this realisation, is built. The accuracy of this type of measurements is presently being considered.

### LIST OF LITERATURE

10. ----------, Parametrisation of Overflow Traffic in Telecommunication and Data-handling, 7th Conf. on Stoch. Proc., Enschede (Neth.), 1977.