APPROXIMATE SOLUTION OF QUEUEING NETWORKS BASED ON
EXACT AND BOUNDED AGGREGATION TECHNIQUE

S. Balsamo

Dipartimento di Informatica *
University of Pisa
Pisa, Italy

Solution methods for queueing networks based on subnetwork aggregation
provide approximate results when applied to networks with general service time
distribution. We derive the conditions under which aggregation method applied
to markovian networks provides exact results, by defining the characteristics of
the single composite node which represents the aggregated subnetwork in terms
of subnetwork parameters and conditional probabilities. An approximated
aggregation method is presented where these conditional probabilities are
approximated by applying a bounded aggregation technique.

1. INTRODUCTION

Queueing networks are used as model of computer and communication systems and have been
proved to be a powerful tool for performance analysis and prediction [11,12,16]. If the system
model belongs to the restricted class of product-form queueing networks [4], it is possible to
compute performance measures efficiently by some computational algorithms [12,16].
Unfortunately, there are many features of systems whose representation in queueing network results
a non-product-form model. These includes, for example, general service time distribution at FCFS
service centers, priority scheduling disciplines, simultaneous resource possession of more than one
resource by a job, and blocking phenomenon due to limited buffer size.

The solution of a non-product-form network model, i.e., the computation of performance
measures, can be obtained by solving the associated Markov process, whose state space dimension
grows exponentially with the number of system components. Thus, as this number increases, the
problem dimension becomes unmanageably large and recourse to approximate solutions is
necessary.

Approximate solution methods for markovian network models include numeric solution of very
sparse linear systems [19], matrix-geometric solution methods [14], and many techniques based on
the aggregation principle; these techniques include both aggregation methods defined on the process

* Corso Italia, 40, 56100 Pisa (I). This work has been partially supported by M.P.I. Research Project Found.
the process matrix \([5,7,8,17,18,21]\) and aggregation methods directly defined on the queuing network \([1,2,3,6,7,9,12,13,16,17]\).

Approximation methods based on subnetwork aggregation can be used in a hierarchical decomposition modelling approach. In this case, by using subnetwork aggregation (also called Norton's theorem or flow-equivalence method) \([6]\), the network is first partitioned into a set of smaller disjoint subnetworks, then each of these subnetworks is analyzed and evaluated in isolation, and finally the individual results are combined in order to obtain the overall solution for the original network model. The recombination (or aggregation) process is realized by using a special type of service center called flow-equivalent, or composite node. This technique can be applied to obtain exact solution for a product form network, by replacing a subnetwork of queues by a single exponential composite node with queue length dependent service rate \([6,2]\), while, when applied to non-product-form networks, it provides approximate results with unknown accuracy. Many approximation iterative methods based on subnetwork aggregation have been proposed in the literature to analyze networks characterized by a variety of features including, for example, general service time distribution at FCFS service centers \([7,12,13,16,17]\). Except the special case of nearly completely decomposable systems for which a subnetwork aggregation method is given in \([7]\), and for which an extensive error analysis is provided, the error analysis of subnetwork aggregation methods applied to general non-product-form networks is an open problem.

Indeed, for these network models it is only known that, as proved by \([20]\), in order to obtain exact results, the single composite node (which represents the aggregated subnetwork) must be defined as a function of the state of the remaining portion of the network. Since \([20]\) provide a non-constructive existence and unicity proof, the problem arises of how the equivalent composite node must be defined.

In this paper we provide an explicit expression of the composite node characteristics, as a function of subnetwork parameters and conditional probabilities. This expression allows us to define completely the exact aggregated network of a non-product-form markovian network obtained by applying a subnetwork aggregation technique. An approximated aggregation method is presented where the conditional probabilities are approximated by applying a bounded aggregation technique \([8]\). Moreover, the presented result can provide a basis for a comparison and an error analysis of the approximation methods based on subnetwork aggregation, and can be helpful in the search of new improved approximations.

The paper is organized as follows. The queuing network model and the associated Markov process are defined in section 2. Section 3 deals with exact subnetwork aggregation technique both for product-form and for non-product-form markovian networks, and the main result is derived. In section 4 an approximated method based on exact and bounded aggregation is presented.

2. THE MODEL

Consider a markovian network model \(W\) with \(M\) service centers. Service time distribution can be represented by a coxian \([10]\) or a phase-type distribution, and any service discipline, independent of service time, is allowed. For the sake of simplicity, we consider single class closed networks
with N customers, and FCFS service discipline. Extensions to multiple classes, open networks, and to other disciplines are based on the same scheme.

Let us introduce some network parameters definitions. Let $S = (S_1, \ldots, S_M)$ denote the state of the network where $S_i = (n_i, c_i)$ denotes the state of a coxian-$K_i$ node $i (i=0,\ldots,M)$, when it contains $n_i$ customers ($n_i=0,\ldots,N$) and the customer in service is currently in stage $c_i (c_i=1,\ldots,K_i)$. The coxian service distribution of node $i$ is represented by the set of $K_i$ exponential stages with state dependent service rates $\mu_i(S_i)$, and by probability $b_i(c_i)$ that the service in node $i$ is completed just after stage $c_i (n_i=0,\ldots,N$, $c_i=1,\ldots,K_i$, $b_i(K_i)=1)$. We assume that if $n_i = 0$, then $c_i = 1$ and $\mu_i(S_i) = 0$. Finally, let $R=\|r_{ij}\|$ denote the routing matrix ($i,j=1,\ldots,M$). The state space, denoted by $E$, is given by:

$$E = \{S = (S_1, S_2, \ldots, S_M) \mid \sum_{i=1}^{M} n_i = N , 0 \leq n_i \leq N , 1 \leq c_i \leq K_i , i = 1,\ldots,M \} \quad (1)$$

Then the network can be modeled as an irreducible continuous-time Markov chain defined on state space $E$. The steady-state probability distribution $\pi = \{\pi(S), S \in E\}$ can be derived by solving the linear system $\pi = \pi Q$, where matrix $Q$ is the process transition rate matrix.

For the special class of product-form networks the steady-state solution $\pi$ can be obtained by the following closed form expression [4]:

$$\pi(S_1, \ldots, S_M) = \frac{1}{G(N)} \prod_{i=1}^{M} f_i(S_i) \quad (2)$$

where $G(N)$ is a normalizing constants, and function $f_i(S_i)$ definition depends on the node $i$ characteristics; specifically, it depends on state $S_i$, on service rates $\mu_i(S_i)$, and on relative throughput $\lambda_i$ ($i=1,\ldots,M$), which can be easily derived as a solution of the homogeneous M-dimensional linear system $\lambda = \lambda R$, with $\lambda = (\lambda_1 , \ldots, \lambda_M)$. In this case the computational complexity of solution $\pi$ is reduced to $O(MN)$ [12,16].

On the other hand, since the solution of non-product-form networks is proportional to state space $E$ dimensions, which grows exponentially with the number of system components ($M,N,K_i$, $i=1,\ldots,M$), then exact analysis becomes soon prohibitively expensive.

Therefore, in order to reduce state space dimension, a hierarchical decomposition approach based on subnetwork aggregation technique can be applied. We shall now define subnetwork aggregation method and the condition under which it provides exact results both for product-form and for non-product-form network models.

## 3. EXACT SUBNETWORK AGGREGATION

Consider the network partition into two subnetworks $\sigma_1$ and $\sigma_2$. Then subnetwork aggregation technique can be applied to network $W$ by replacing subnetwork $\sigma_1$ with a single composite node $C$, as shown in Figure 1.

In other words, let the $M (M=s+r)$ service centers of network $W$ be partitioned into subnetworks $\sigma_1 = \{1,\ldots,s\}$ and $\sigma_2 = \{s+1,\ldots,s+r\}$ (fig. 1a). The network partition can be related to a state space $E$ partition as follows. Let us rewrite the network state as $S = (S_{\sigma_1}, S_{\sigma_2})$, where $S_{\sigma_1}$...
Fig. 1 - Subnetwork aggregation technique applied on network \( W \) for subnetwork \( \sigma_1 \):

(a) network \( W \) partitioned into subnetworks \( \sigma_1 \) and \( \sigma_2 \);
(b) aggregated network \( W' \).

and \( S_{\sigma_2} \) represent the state of subnetwork \( \sigma_1 \) and \( \sigma_2 \), respectively. Then state space \( E \), defined by (1), can be partitioned in subsets \( a(n, S_{\sigma_2}) \) defined as follows:

\[
a(n, S_{\sigma_2}) = \{ S \in E | S = (S_{\sigma_1}, S_{\sigma_2}), \sum_{i \in \sigma_1} n_i = n \} \tag{3}
\]

Let \( W' \) be the new closed markovian network with \( s+1 \) service centers and \( N \) customers, obtained by network \( W \) by aggregating subnetwork \( \sigma_1 \) in a single composite node \( C \). Therefore aggregated network \( W' \) is formed by subnetwork \( \sigma_2 \) and single composite node \( C \) (fig. 1b). State space \( E' \) of \( W' \) is formed by states \( a(n, S_{\sigma_2}) \) corresponding to sets of partition (3):

\[
E' = \{ a(n, S_{\sigma_2}) | n + \sum_{i \in \sigma_2} n_i = N, 0 \leq c_i \leq K_i, i \in \sigma_2 \} \tag{4}
\]

Since each set \( a(n, S_{\sigma_2}) \) includes all the states of \( E \) in which there is a total of \( n \) customers in subnetwork \( \sigma_1 \) and subnetwork \( \sigma_2 \) is in state \( S_{\sigma_2} \), each aggregate \( a(n, S_{\sigma_2}) \) corresponds to a state of the aggregated network \( W' \) [2].

Let routing matrix \( R' = (r_{ij}) \) (for \( i,j = C,s+1,\ldots,s+r \)) of network \( W' \) be defined as a function of routing matrix \( R \) of network \( W \), and of relative throughputs \( \lambda_i \), \( i=1,\ldots,M \) (as for product-form networks, [1]) as follows:

\[
\begin{align*}
    r'_{Cj} &= \frac{\sum_{i=1}^{s} \lambda_i r_{ij}}{\sum_{k \in \sigma_2} \sum_{i=1}^{s} \lambda_i r_{ik}} \quad \forall j \in \sigma_2, \quad r'_{CC} = 0 \tag{5} \\
    r'_{ij} &= r_{ij} \quad \forall i,j \in \sigma_2 \\
    r'_{iC} &= \sum_{j \in \sigma_1} r_{ij} \quad \forall i \in \sigma_2
\end{align*}
\]

Characteristics of each node of subnetwork \( \sigma_2 \) in \( W' \) are defined as in the original network \( W \). In order to complete network \( W' \) definition we have to specify composite node \( C \) characteristics. Moreover, we have to define the conditions under which subnetwork aggregation technique provides exact results.

**Definition**: Subnetwork aggregation technique is said to be exact when applied to network \( W \) if it defines an aggregated network \( W' \) whose steady-state solution \( \pi' \) is equal to steady-state
solution $\pi$ of the original network $W$, related to subnetwork $\sigma_2$, i.e., if $\pi'(S_{\sigma_2}) = \pi(S_{\sigma_2})$, $\forall S_{\sigma_2}$.

In other words, since network $W'$ parameters can be expressed in terms of original network $W$ parameters, subnetwork aggregation is an exact state space reduction and aggregated network $W'$ is also said to be equivalent to network $W$ [3].

3.1. Exact Aggregation in Product-Form Networks

Let us first consider subnetwork aggregation method applied to a product-form queueing network model [4], i.e., whose steady-state solution is given by expression (2). In order to obtain exact results, subnetwork $\sigma_2$ is replaced, in network $W'$, by an exponential single server composite node $C$ with FCFS (or PS) discipline, and state dependent service rate $\mu_C(n_C)$ [6,2]. (Note that the state $S_i$ of an exponential node $i$ is completely defined by its first component, $n_i$). Let $\tau_i(n)$ denote the throughput of service center $i$ in network $W$ when there are $n$ jobs in subnetwork $\sigma_i$. Then service rate $\mu_C(n_C)$ can be defined as follows:

$$\mu_C(n_C) = \sum_{i \in \sigma_1} \tau_i(n_C) \sum_{j \in \sigma_2} r_{ij}$$

By denoting with $G_{\sigma_1}(n)$ the normalizing constant in formula (2) only related to nodes in $\sigma_1$ and when there is a total of $n$ customers in subnetwork $\sigma_1$, one can rewrite formula (6) as follows:

$$\mu_C(n_C) = \frac{G_{\sigma_1}(n_C - 1)}{G_{\sigma_1}(n_C)} \sum_{i \in \sigma_1} \lambda_i \sum_{j \in \sigma_2} r_{ij}$$

Formula (7) completes network $W'$ definition, which can be proved to be equivalent to the original network $W$ in terms of steady-state probabilities of subnetwork $\sigma_2$ [6,2]. Since the proof of exact subnetwork aggregation is based on closed-form expression (2), then it can not be directly extended to non-product form queueing network models.

3.2. Exact Aggregation in Non-Product-Form Networks

In this section we derive the conditions under which subnetwork aggregation method provides exact results for a non-product-form network model.

In this case, unlike the previous model, the throughput produced by subnetwork $\sigma_1$ on subnetwork $\sigma_2$, defined by (6), cannot be rewritten as by formula (7). For this class of queueing network models it has been proved [20] that, if subnetwork $\sigma_2$ is formed by only one node, i.e., $r=1$ and $\sigma_2 = \{s+1\}$, there will always exist a unique markovian composite node $C$ with state dependent service rate $\mu_C(S_{\sigma_2})$ which is flow-equivalent to subnetwork $\sigma_1$. In other words, marginal probabilities $\pi(S_{\sigma_2}), \forall S_{\sigma_2}$, computed in network $W$ are identical to $\pi'(S_{\sigma_2})$ computed in network $W'$, where composite node $C$ service rate is equal to the throughput of $\sigma_1$ on subnetwork $\sigma_2$. This result can be easily extended to any subnetwork $\sigma_2$ formed by $r$ nodes ($r \geq 1$) [15]. Since [20] provide a non-constructive existence and unicity proof, the problem is to determine how the equivalent composite node $C$ must be defined.

Let $\lambda_i(S_i)$ be the throughput on node $i$, currently in state $S_i$, from the remaining portion of the
network (i.e., related to network partition into subnetworks $\sigma_2 = \{i\}$ and $\sigma_1=W-\sigma_2$). By considering state space partition (3), an expression for $\lambda_i(S_j)$ is derived in [17] as:

$$\lambda_i(S_j) = \sum_{j=1}^{M} \sum_{n_j=1}^{N} \sum_{c_j=1}^{K} \mu_j(S_j) b_j(c_j) r_{ij} P\{S_j | S_i\}$$

where $P\{S_j | S_i\}$ is the marginal probability of state $S_j$ in node $j$ conditioned to state $S_i$ in node $i$.

We shall now extend formula (8) to express the throughput produced by any subnetwork $\sigma_1$ on the remaining subnetwork $\sigma_2$, denoted by $\lambda_{\sigma_2}(S_{\sigma_2})$. Moreover, by defining composite node $C$ service rate in network $W'$ as equal to this throughput, we provide exact subnetwork aggregation method for non-product-form networks.

**Theorem** : Let the markovian network $W$ partitioned into two subnetworks $\sigma_1$ and $\sigma_2$. Then subnetwork aggregation method provides exact results when subnetwork $\sigma_1$ is aggregated in single composite node $C$ whose state dependent service rate $\mu_C(S_{\sigma_2})$ is equal to throughput $\lambda_{\sigma_2}(S_{\sigma_2})$ produced by subnetwork $\sigma_1$ on subnetwork $\sigma_2$ in $W$ when subnetwork $\sigma_2$ is in state $S_{\sigma_2}$, which is defined as follows:

$$\mu_C(S_{\sigma_2}) = \lambda_{\sigma_2}(S_{\sigma_2}) = \sum_{i \in \sigma_2} \sum_{j \in \sigma_1} \sum_{n_j=1}^{N} \sum_{c_j=1}^{K} \mu_j(S_j) b_j(c_j) r_{ij} P\{S_j | S_{\sigma_2}\}$$

where $P\{S_j | S_{\sigma_2}\}$ is the marginal probability of state $S_j$ in node $j$ in $\sigma_1$ conditioned to state $S_{\sigma_2}$ of subnetwork $\sigma_2$.

**Proof.** In order to prove the theorem it is sufficient to compare process matrix $Q'$ of aggregated network $W'$, where node $C$ service rate $\mu_C(S_{\sigma_2})$ is unknown, with matrix $Q_A$ obtained by applying an exact aggregation procedure on process matrix $Q$ of $W$ [17,21], related to partition (3). Due to network definition, aggregated matrix $Q_A$ yields a particular structure, and it can be defined in terms of network $W$ parameters and conditional probabilities (see [2], Theorem 1). By comparing matrices $Q$ and $Q_A$ one can show [15] that $Q = Q_A$ if and only if formula (9) holds. And this completes the proof.

Expression (9), which represents the throughput produced by subnetwork $\sigma_1$ on subnetwork $\sigma_2$ for markovian networks, completes network $W'$ definition, which can be proved to be equivalent to original network $W$ in terms of steady-state probabilities of subnetwork $\sigma_2$. In other words, we can state that in a markovian network it is possible to aggregate any subnetwork $\sigma_1$, without constraints on its dimension, in a single node $C$.

**Corollary 1** : Aggregated network $W'$ obtained by $W$ through subnetwork aggregation process of $\sigma_1$ in the single composite node $C$ with service rate $\mu_C(S_{\sigma_2})$, given by (9), and with routing matrix $R'$, given by (5), is equivalent to original network $W$, in terms of steady-state probabilities of subnetwork $\sigma_2$, i.e., $\pi(S_{\sigma_2}) = \pi'(S_{\sigma_2}), \forall S_{\sigma_2}$.

This equivalence result, and, specifically, expression (9) for node $C$ service rate, can provide a basis for comparison and error analysis of approximation methods based on subnetwork aggregation and can be helpful in the search of new improved approximations.

Note that state dependent service rate $\mu_C(S_{\sigma_2})$, given by (9), is defined in terms of network $W$ parameters and conditional probabilities $P\{S_j | S_{\sigma_2}\}, \forall j \in \sigma_1$, i.e., it is state $S_{\sigma_2}$ dependent. Therefore an exact computation of expression (9) requires the exact computation of conditional

4.1B.3.6
probabilities $P\{S_j|S_{\sigma_2}\}, \forall j \in \sigma_1$. Unfortunately, for general non-product-form networks, this computation is based on the solution of a large linear system whose complexity is comparable to the solution of the entire system. Therefore recourse to approximation is necessary. In the next section we present a subnetwork aggregation method obtained by exact aggregation where conditional probabilities are computed applying bounded aggregation [8].

4. APPROXIMATION BASED ON BOUNDED AND EXACT AGGREGATION

In order to approximate conditional probabilities $P\{S_j|S_{\sigma_2}\}, \forall j \in \sigma_1$, consider the following state space $E$ partition corresponding to network $W$ partition into subnetworks $\sigma_1$ and $\sigma_2$:

$$a(S_{\sigma_1}, n) = \{ S \in E | S = (S_{\sigma_1}, S_{\sigma_2}), \sum_{i \in \sigma_2} n_i = n \}$$  \hspace{1cm} (10)

This partition, which is dual to partition (3), corresponds to subnetwork $\sigma_2$ aggregation in a single composite node $C_2$. Let $W''$ be the new aggregate markovian network formed by subnetwork $\sigma_1$ and node $C_2$. State space of $W''$ is formed by states $a(S_{\sigma_1}, n)$ corresponding to sets of partition (10). Consequently, process matrix $Q''$ of network $W''$ can be obtained by aggregating process matrix $Q$ of network $W$, with relation to partition (10). By solving network $W''$ one obtains marginal probabilities $P\{S_j|n\}$ of state $S_j$ for any node $j \in \sigma_1$, conditioned to state $n$ in node $C_2$. These probabilities can be interpreted as marginal probabilities of node $j \in \sigma_1$, conditioned to a total of $n$ jobs in subnetwork $\sigma_2$.

We shall now approximate conditional probabilities $P\{S_j|S_{\sigma_2}\}$ with probabilities $P\{S_j|n\}, \forall j \in \sigma_1$. An approximation is introduced because we consider only the total number $n$ of customers in $\sigma_2$, disregarding the different influence of each subnetwork state $S_{\sigma_2}$. Exact computation of probabilities $P\{S_j|n\}$ can be still computational expensive for non-product-form networks. In order to reduce computational complexity (or, in some cases, to make the solution computation feasible) we can apply an approximation method. Consider bounded aggregation method, introduced by [8], which can be used as an approximation method characterized by known accuracy.

Bounded aggregation, provides lower and upper bound to the solution of a large non-negative irreducible matrix, approximated by a block decomposition-aggregation technique; lower and upper bounds on the steady-state probabilities are derived from either a lower or an upper bound to the process matrix and from the spectral radius, and they have been shown to be the tightest possible in this case [8,9]. This method can be simplified when applied to markovian networks, whose process matrix yields a particular structure [2]. By applying bounded aggregation to network $W$, related to partition (10), lower and upper bounds, $X_{\text{inf}}$ and $X_{\text{sup}}$, on conditional probabilities $P\{S_j|n\}$ can be obtained from bounds on the aggregated process matrix $Q''$ (see [2] for details). Since matrix $Q''$ can be partially computed in terms of networks parameters ($\mu_i(S_1), b_i(S_2), \forall i \in \sigma_1$ and $R$), lower and upper bounds matrices can be partially defined without error, as described in [2]. From bounds $X_{\text{inf}}$ and $X_{\text{sup}}$ approximated conditional probabilities $P^*\{S_j|n\}$ can be defined, for example, as the mean value.
To sumarize, the proposed approximation method applied to network W partitioned into subnetworks $\sigma_1$ and $\sigma_2$ in order to obtain aggregated network $W'$ consists of four steps:

1. Apply bounded aggregation to network $W$ related to partition (10), and compute bounds $X^{\inf}$ and $X^{\sup}$ on conditional probabilities $P(S_j|n)$, $\forall j \in \sigma_1$.

2. Define approximation $P^*(S_j|n)$ of $P(S_j|n)$ ($\forall j \in \sigma_1$) as the mean value of bounds $X^{\inf}$ and $X^{\sup}$.

3. Replace $P(S_j|n)$ with approximation $P^*(S_j|n)$ ($\forall j \in \sigma_1$) in formula (9) in order to define composite node service rate $\mu_C^*(S_{\sigma_2})$, approximation of $\mu_C(S_{\sigma_2})$, i.e.:

$$\mu_C^*(n) = \sum_{i \in \sigma_1} \sum_{j \in \sigma_2} \sum_{h_j=1}^{N} \sum_{c_j=1}^{K_i} \mu_j(S_j) b_j(c_j) r_{ji} P^*(S_j|n)$$

(11)

4. Formula (11) completes aggregated network $W'$ definition, whose routing matrix $R'$ is defined by (5), as described in section 3.

Network $W'$ solution provides an approximate solution for subnetwork $\sigma_2$ in network $W$. The presented subnetwork aggregation method based on exact and bounded aggregation can be used in a hierarchical decomposition approach. Specific characteristics of the network can be used to simplify the analysis.

The presented approximation has been tested on a class of non-product-form networks by varying network parameters, i.e., network dimensions $(M,N)$, service time distributions, and node utilizations. Approximation error has been defined as the maximum of errors on utilization factor, on normalized mean queue length, and on normalized mean queueing time. Comparison with exact solution showed a maximum error within 5% even for high utilizations [15].

5. CONCLUSIONS

Subnetwork aggregation is a basic step of many approximation methods for queuing network models. The conditions under which subnetwork aggregation method applied to markovian networks provides exact results has been derived by defining the composite node $C$ service rate by expression (9) in terms of subnetwork parameters and conditional probabilities. An approximated aggregation method has been defined, where these conditional probabilities are approximated by applying a bounded aggregation technique. Moreover, the presented result can provide a basis for a comparison and an error analysis of the approximation methods based on subnetwork aggregation and can be helpful in the search of new improved approximations.

REFERENCES


