韓國經營科學會誌
第24卷 第4號
1999年 12月

사용자수 제한을 갖는 개방형 다중계층구조의 대기행렬 네트워크 분석에 관한 연구*

이 영**

An Analysis of a Multilayered Open Queueing Network with Population Constraint and Constant Service Times*

Young Rhee**

■ Abstract ■

In this paper, we consider a queueing network model, where the population constraint within each subnetwork is controlled by a semaphore queue. The total number of customers that may be present in the subnetwork can not exceed a given value. Each node has a constant service time and the arrival process to the queueing network is an arbitrary distribution.

A major characteristics of this model is that the lower layer flow is halted by the state of higher layer. We present some properties that the inter-change of nodes does not make any difference to customer's waiting time in the queueing network under a certain condition. The queueing network can be transformed into a simplified queueing network. A dramatic simplification of the queueing network is shown. It is interesting to see how the simplification developed for sliding window flow control, can be applied to multi-layered queueing network.

1. Itroduction

In a typical queueing network, a customer may have traverse several layers of flow controlled mechanisms before it comes out from the network. In this paper, we present a model for an-

* Supported in part by a grant from IBM-RTP and Keimyung University
** Department of Industrial Engineering, Keimyung University

alyzing the delays introduced by such multilayered flow mechanism. The number of customers in each subnetwork is controlled by a semaphore queue.

Reiser [6] modeled a computer communication system consisting of many virtual routes with end-to-end window flow control, as a closed multichain queueing network under the assumption of a loss system. Pennotti and Schwartz [4] and Schwartz [9] analyzed a virtual route as a closed tandem queueing network under the same assumption. Reiser [5] observed that in a real situations, packets that arrives to find a full window are not lost, but are queued in an input queue. Reiser [5] and Thomasian and Bay [11], use a flow equivalent server technique to model the sliding window link as a single server queue with state dependent service rate. In this approach, the effect of delays due to all sequence numbers in use is accounted for in the delivery service time of the equivalent server. Varghee, Chou and Nilsson [12] and Gihr and Kuehn [2], presented a similar approach to the above. Varghee, Chou and Nilsson [12] analyzed an open queueing network without an acknowledgment delay using the approximation method. Gihr and Kuehn [2] obtained the characteristics of the physical transmission process using hierarchical decomposition and aggregation methods. Recently, Rhee and Perros [7] modeled an open tandem queueing network with population constraint and constant service times. The total number of customers that may be presented in the network can not exceed a given value k. Customers arriving at the queueing network when there are more than k customers are forced to wait in an external queue. The arrival process to the queueing network is assumed to be arbitrary.

For an analysis of multi-layered communication network, the communication functions are partitioned into a vertical set of layers. Each layer performs a related subset of the functions required to communicate with another system. It relies on the next lower layer to perform more primitive functions and to conceal the details of those functions. It provides services to next higher layer. Mitchell and Lide [3] presented a general framework to model sliding window flow control from the closed queueing network models. Fdida, Perros and Wilk [1] presented a methodology for analyzing nested and tandem configurations of sliding window controlled networks. Each layer of sliding window control is reduced to a state dependent infinite server queue without acknowledgment using a flow-equivalence methodology. A single-hop OSI structured network with multiple layers of sliding window flow control and packet fragmentation between layers is analyzed by Shapiro and Perros [10]. They presented a hierarchical method to analyze nested sliding window flow controlled layers. Each layer with sliding window control is reduced to a single queue with state dependent service rate.

In this paper, we present a multilayered open tandem queueing network controlled by semaphore queue. This type of queueing networks have application in diverse area, such as pallet based production system, computer sharing and multiprogramming systems, communication network model and semaphore controlled software in an operating system. A major characteristic of this model is that the lower layer flow is halted by the state of higher layer. However, we focus on the behavior of the queueing network in terms of customer's waiting time. We present some properties that the inter-change of nodes does not make any difference to customer's waiting time in the queueing network under a certain condition. This paper is a sequel to earlier two papers by Rhee and Perros [7, 8].

This paper is organized as followed : Section 2 presents the model for a two-layer queueing network with semaphore queue. In section 3, some characteristics of a semaphore controlled queueing network properties of inter-change among nodes, is presented to analyze the customer's waiting time. In section 4, we consider a population constrained queueing network with more than two-layer. Finally, the conclusion is presented at section 5.

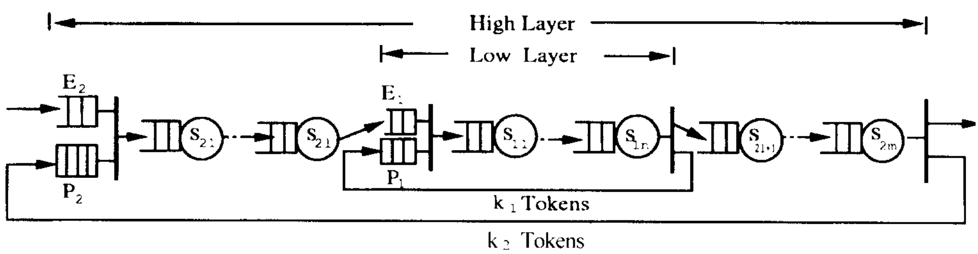
We note that throughout this paper, we interpret the waiting time of a customer as the total time a customer spends queueing up in the queueing network, rather than the total time it takes to traverse the queueing network which also includes service times.

2. A model for two-layer queueing network

Let us consider for a moment a queueing network with population constraint and constant service times [7]. The population constraint of the queueing network is controlled by a semaphore. The semaphore is a mechanism that consists of a pool of k tokens and an external queue. Customers arrive at the external queue. If there is a token available at the token queue, a customer takes the token and proceed to the queueing network. Upon the completion of service the customer leaves the network and the token is returned to the external queue. Customers that arrive at the external queue when the token queue is empty are forced to wait in the external queue. For this queueing network, the waiting time of a customer remains the same even though the order of the service times is rearranged. In view of this, we can represent the queueing network by simpler two node queueing network as is shown in Rhee and Perros [7].

Now, let us consider a two-layer open tandem queueing network with population constraint and constant service times as shown in [Figure 1]. The population constraint of the queueing network is controlled by a semaphore. For presentation purpose, we shall refer to the outside window flow control as the high layer or level 2 layer, and to the inside window flow control as the low layer or level 1 layer.

An arriving customer takes a token from the high layer token pool P_2 and enters the high layer queueing network. The customer holds this token until it leaves the high layer queueing



[Figure 1] A two-layer open tandem queueing network

network. Customers arrive through external queue E_2 on the high layer. The customer proceeds to the high layer queueing network until the low layer external queue E_1. In order to enter the low layer queueing network, the customer needs another token from the low layer token pool P_1. The customer is then subjected to the low layer window flow control. Upon service completion in the low layer queueing network, the customer returns it's token immediately to the token pool P_1 and proceed to the rest of the node in the high layer queueing network. Again upon service completion in the high layer queueing network, the token is returned to the token pool P_2 in zero time. Customers that arrive during the time when the token pool is empty, are forced to wait in either external queue E_1 or E_2 respectively. The first customer in the external queue, enters the queueing network as soon as a token is returned to it's corresponding token pool. The arrival process to the queueing network is assumed to be an arbitrary general distribution with a rate λ.

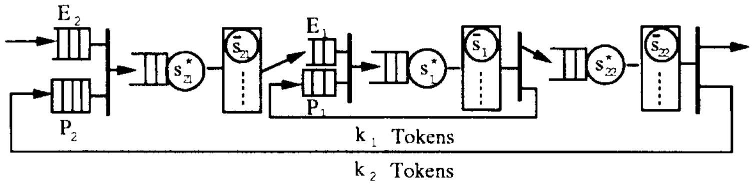
Let s_{ij} be the constant service time at node j and i^{th} layer queueing network. Let $s_1^* = max\{s_{1j}, j = 1, \ldots n\}$, $s_{21}^* = max\{s_{2j}, j = 1, \ldots l\}$, and $s_{22}^* = max\{s_{2j}, j = l+1, \ldots m\}$. Also, let $T_1 = \sum_{j=1}^{n} s_{1j}$, $T_{21} = \sum_{j=1}^{l} s_{2j}$ and $T_{22} = \sum_{j=l+1}^{m} s_{2j}$. Finally, let k_2 and k_1 be the number of tokens or window size for the high layer and the low layer queueing network respectively. We assume that $k_2 > k_1$. This is because if $k_2 \le k_1$, there is no waiting customer in the external queue E_1. Hence, the population constraint of the low layer queueing network can be relaxed. We assume that $k_1 s_1^* \le T_1$ for the low layer queueing network. By letting $s^* = max\{s_1^*, s_{21}^*, s_{22}^*\}$ and $T = T_1 + T_{21} + T_{22}$, we also assume that $k_2 s^* \le T$ for the high layer queueing network see Corrolary 1 and 2 in Rhee and Perros [7].

Although the above described model has only two layer of population constrained queueing network, it is of sufficient generality to demonstrate the method of analysis proposed here. We focus on the reduction of an arbitrary number of nodes of window flow control to the simpler queueing network which represents the same performance characteristics of the multi-layered queueing network.

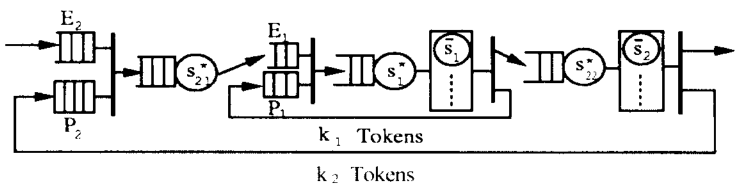
3. Characteristics of a semaphore controlled queueing network

Let us consider the model shown in [Figure 1]. The nodes of the two layer queueing network can be classified into three parts. The opening nodes of the high layer queueing network until s_{2l}, the low layer queueing network and the rest of nodes in the high layer queueing network. Using the results of Theorem 1 in Rhee and Perros [7], each part of the queueing network can be reduced into two-node queueing network without losing its performance characteristic, such as customer's traverse time in the queueing network. In view of this, we can represent the queueing network in [Figure 1] by a simpler six-node queueing network as shown in [Figure 2].

In [Figure 2], $\bar{s}_{21}$, $\bar{s}_1$ and $\bar{s}_{22}$ represent, $T_{21} - s_{21}^*$, $T_1 - s_1^*$, and $T_{22} - s_{22}^*$, respectively. For presentation purposes we shall refer to these six nodes as the first node, the second node, the third node and so on. The number of parallel



[Figure 2] A six-node two-layer queueing network



[Figure 3] A five-node two-layer queueing network

servers is infinite at the second node, the forth node and the sixth node respectively. There is no queue on the second node and the sixth node in the high layer queueing network. Therefore, the second and sixth node can be merged.

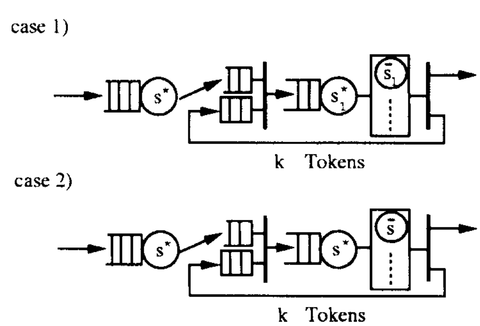
Let $\bar{s}_2 = \bar{s}_{21} + \bar{s}_{22}$. We can represent the queueing network in [Figure 2], by the simpler five-node queueing network as shown in [Figure 3]. A customer's waiting time in the five-node queueing network is the same as in the original queueing network under the study. In the configuration of [Figure 3], the study of interactions and interchangeability among the nodes give us how much we can improve the simplification of the given queueing network.

Theorem 1. Let us consider the open queueing network as shown in [Figure 4]. If we assume $s_1^* + \bar{s}_1 = s_1^* + \bar{s}$ and $s^* \geq + s_1^*$, a customer's waiting time in either case is the same.

Proof. For the first k arriving customers in either case, an arriving customer always finds a token in the token pool. There is no queue at the second node. This is because the a customer's inter-departure time from the first node is always greater than or equal to s^* (see in [7]). Letting a_i be the interarrival time between the i^{th} and the $(i-1)^{st}$ customer to the queueing network, a customer's waiting time in the queueing network, w_i is

$$w_i = \max\{0, w_{i-1} + s^* - a_i\} \quad \text{for} \quad 2 \leq i \leq k \quad (1)$$

Let w_i^j be the i^{th} arriving customer's waiting time at the node j. For the $(k+1)^{st}$ arriving customer in case 1, the waiting time at the first



[Figure 4] The open tandem queueing network

node is

$$w_{k-1}^i = max\{0, w_k^i + s^* - a_{k+1}\} \quad (2)$$

However, the $(k+1)^{st}$ arriving customer may wait in the external queue until a token returns to the token pool P. This is because every k^{th} arriving customer uses the same token. Thus, the $(k+1)^{th}$ arriving customer will take the first customer's token. The waiting time at the external queue, w^e_{k+1} is the time between the departure time of the first arriving customer, d_1 and the arrival time of $(k+1)^{st}$ customer to the external queue E. Since $d_1 = s^* + s^*_1 + \overline{s}_1$ and the arrival time of $(k+1)^{st}$ customer to the external queue is $\sum_{i=1}^{k+1} a_i + w^1_{k-1} + s^*$, the external queue waiting time of $(k+1)^{st}$ customer, w^e_{k+1} is

$$w^e_{k-1} = max\left\{0,\ d_1 - \left(\sum_{i=1}^{k+1} a_i + w^1_{k-1} + s^*\right)\right\}$$
$$= max\left\{0,\ s^*_1 + \overline{s}_1 - \left(\sum_{i=1}^{k+1} a_i + w^1_{k-1}\right)\right\} \quad (3)$$

The $(k+1)^{st}$ arriving customer's waiting time in the queueing network, w_{k+1} is equal to $w^1_{k+1} + w^e_{k+1}$.

In general, the j^{th} customer's waiting time in the network, w^1_j and w^e_j are respectively

$$w^1_j = max\{0,\ w^1_{j-1} + s^* - a_j\} \quad (4)$$

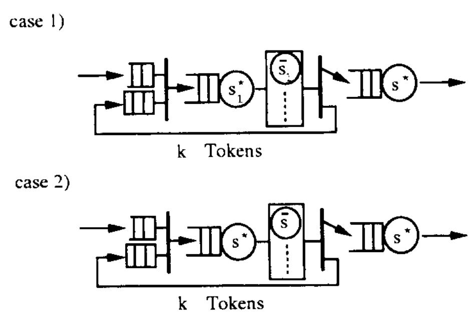
$$w^e_j = max\left\{0,\ d_{j-k} - \left(\sum_{i=0}^{j} a_i + w^1_j + s^*\right)\right\} \quad (5)$$

Now, let us consider the case 2 in [Figure 4]. For the first k arriving customers, a customer's waiting time is equal to (1). This is because the first node has the longest service time. For the $(k+1)^{st}$ arriving customer, the waiting time at the first node is also equal to (2), and the external queue waiting time is the time difference between the departure time of the first arriving customer and the arrival time of the $(k+1)^{st}$customer to the external queue E. The departure time of the first arriving customer is $s^* + s^* + \overline{s}$. Since $s^*_1 + \overline{s}_1 = s^* + \overline{s}$, w^e_{k+1} for the case 2 is exactly same with (3).

For the j^{th} arriving customer, the customer's waiting time in the first node is equal to (4). And the customer's waiting time in the external queue is also equal to (5), since the departure time of $(j-k)^{th}$ is the same with the case (1). We can prove recursively that the customer's waiting time in either case is identical. □

Theorem 2. Let us consider the open queueing network as shown in [Figure 5]. Let assume $s^*_1 + \overline{s}_1 = s^* + \overline{s}$ and $s^* \geq s^*_1$. Then, a customer's waiting time in either case is identical.

Proof. For the first k arriving customers in either case, the customer's waiting in the queueing network is the same. This is because the waiting time in the tandem queueing network



[Figure 5] The open tandem queueing network

with constant service time is decided by the longest service time. So that, the customer's departure time from the queueing network is identical in either case.

Let us consider the $(k+1)^{st}$ arriving customer. The customer's arrival time to the queueing network is $\sum_{i=1}^{k+1} a_i$. The $(k+1)^{st}$ arriving customer will use the token returned by the first cus tomer. Let d_1 be the i^{th} customer's departure time from the semaphore queueing network.

For the case 1, let w_i^j be the j^{th} customer's waiting time at node $j, j=e,1,2,3$. Since the second node has an infinite number of servers, $w_i^2=0$. Hence the i^{th} customer's waiting time w^i is $w_i^e+w_i^1+w_i^3$. Let a_i^j be the interarrival time between $(i-1)^{th}$ and i^{th} customer to the node j. Undoubtedly, $a_i^2=a_i^3$. The external queue waiting time w_{k+1}^e is $max\{0, d_1-\sum_{i=1}^{k+1} a_i\}$. The $(k+1)^{st}$ customer's interarrival time to the first node a_{k+1}^1 can be changed into $a_{k+1}+w_{k+1}^e-w_k^e$. The waiting time at the first node, w_{k+1}^1 is $max\{0, w_k^1+s_1^*-a_{k+1}^1\}$. The interarrival time to the second node, a_{k+1}^2 depends on w_{k+1}^1. If $w_{k+1}^1>0$, $a_{k+1}^3=s_1^*$. And if $w_{k+1}^1=0$, $a_{k+1}^3=a_{k+1}^1-w_k^1$. Since the interarrival time to the third node does not change, the waiting time at the third node w_{k+1}^3 is $max\{0, w_k^3+s^*-a_{k+1}^2\}$.

$$w_{k+1}^e = max\left\{0,\ d_1-\sum_{i=1}^{k+1}a_i\right\} \tag{6}$$

$$w_{k+1}^1 = max\{0,\ w_k^1+s_1^*-a_{k+1}^1\} \tag{7}$$

$$w_{k+1}^3 = max\{0,\ w_k^3+s^*-a_{k+1}^3\} \tag{8}$$

So that, w_{k+1} can be expressed in terms of w_k, s^* and a_j.

For the case 2, let $\hat{d}_i$ be the i^{th} customer's departure time from the semaphore queueing network. Since there is no waiting time after the first node, let $\hat{w}_i^1$ be the i^{th} customer's waiting time at first node. The $(k+1)^{st}$ arriving customer's external queue waiting time, $\hat{w}_{k+1}^e = \hat{d}_1-\sum_{i=1}^{k+1}a_i$. The relationship between $\hat{d}_1$ and d_1 is $\hat{d}_1=d_1+w_1^3$. Since $w_1^3=0$, $\hat{d}_1=d_1$. The $(k+1)^{st}$ customer's interarrival time to the first node, a_{k+1}^1 has changed into $a_{k+1}+w_{k+1}^e$. The waiting time at the first node, $\hat{w}_{k+1}^1$ is $max\{0,\ \hat{w}_k^1+s^*-a_{k+1}^1\}$.

$$\hat{w}_{k+1}^e = max\left\{0,\ d_1-\sum_{i=1}^{k+1}a_i\right\} \tag{9}$$

$$\hat{w}_{k+1}^1 = max\{0,\ \hat{w}_k^1+s^*-a_{k+1}^1\} \tag{10}$$

.

We can verify easily that w_{k+1} and $\hat{w}_{k+1}$ are identical.

For an arbitrary j^{th} arriving customer in the case 1,

$$w_j^e = max\left\{0,\ d_{j-k}-\sum_{i=1}^{j}a_i\right\} \tag{11}$$

$$a_j^1=a_j+w_j^e-w_{j-1}^e \tag{12}$$

$$w_j^1=max\{0,\ w_{j-1}^1+s_1^*-a_j^1\} \tag{13}$$

$$a_j^2=a_j^1-w_{j-1}^1$$

$$\text{if } w_j^1=0,\ \text{that is, } w_{j-1}^1+s_1^*-a_j^1\le 0 \tag{14}$$

$$a_j^2=s_1^* \text{ if } w_j^1>0 \tag{15}$$

$w_j^2=0$ since the second node has infinite servers

$$a_j^3=a_j^2$$

$$w_j^3=max\{0,\ w_{j-1}^3+s^*-a_j^3\} \tag{16}$$

The j^{th} customer's waiting time w_j is

$$\begin{aligned} w_j &= w_j^e+w_j^1+w_j^3 \\ &= max\{0,\ w_{j-1}^e+w_{j-1}^1+w_{j-1}^3+s^*-a_j\} \quad (17) \\ &= max\{0, w_{j-1}+s^*-a_j\} \end{aligned}$$

Now, for an arbitrary j^{th} arriving customer in the case 2,

$$\widehat{w}_j^e = max\left\{0,\ \widehat{d}_{j-k} - \sum_{i=1}^{j} a_i\right\} \tag{18}$$

$$a_j^1 = a_j + \widehat{w}_j^e - \widehat{w}_{j-1}^e \tag{19}$$

$$\widehat{w}_j^1 = max\{0,\ w_{j-1}^1 + s^* - a_j^1\} \tag{20}$$

The j^{th} customer's waiting time $\widehat{w}_j$ is

$$\begin{aligned}\widehat{w}_j &= max\{0,\ \widehat{w}_j^e + \widehat{w}_j^1\} \\ &= max\{0,\ \widehat{w}_{j-1} + s^* - a_j\}\end{aligned} \tag{21}$$

Below we shall show that (17) and (21) are identical in terms of the customer's waiting time. We assume that $w_i = \widehat{w}_i,\ i=1,2,,,j-1$ to prove $w_j = \widehat{w}_j$. Therefore

$$\begin{aligned} w_{j-k} &= w_{j-k}^e + w_{j-k}^1 + w_{j-k}^3 \\ \widehat{w}_{j-k} &= \widehat{w}_{j-k}^e + \widehat{w}_{j-k}^1 \\ \widehat{w}_{j-k}^e + \widehat{w}_{j-k}^1 &= w_{j-k}^e + w_{j-k}^1 + w_{j-k}^3 \end{aligned} \tag{22}$$

The j^{th} arriving customer will take the token returned by $(j-k)^{th}$ customer. So that, the departure time of $(j-k)^{th}$ customer, d_{j-k} or $\widehat{d}_{j-k}$ is the same with the token returning time to the token pool.

$$\begin{aligned} d_{j-k} &= \sum_{i=1}^{j-k} a_i + w_{j-k}^e + w_{j-k}^1 \\ \widehat{d}_{j-k} &= \sum_{i=1}^{j-k} a_i + \widehat{w}_{j-k}^e + \widehat{w}_{j-k}^1 \end{aligned} \tag{24}$$

Since the arrival time of the j^{th} customer is $\sum_{i=1}^{j} a_i$, the j^{th} customer may wait in the external queue until the token arrives. Therefore, the j^{th} customer's external queue waiting time is

$$\begin{aligned} w_j^e &= max\left\{0,\ d_{j-k} - \sum_{i=1}^{j} a_i\right\} \\ &= max\left\{0,\ w_{j-k}^e + w_{j-k}^1 - \sum_{i=j-k+1}^{j} a_i\right\} \end{aligned} \tag{25}$$

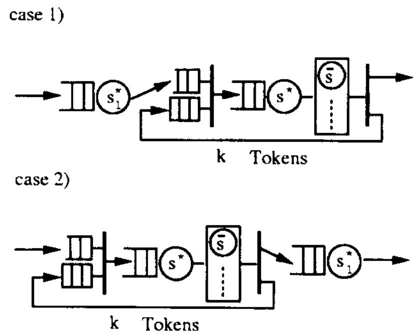
$$\begin{aligned} \widehat{w}^e &= max\left\{0,\ \widehat{d}_{j-k} - \sum_{i=1}^{j} a_i\right\} \\ &= max\left\{0,\ \widehat{w}_{j-k}^e + \widehat{w}_{j-k}^1 - \sum_{i=j-k+1}^{j} a_i\right\} \\ &= max\left\{0,\ w_{j-k}^e + w_{j-k}^1 + w_{j-k}^3 - \sum_{i=j-k+1}^{j} a_i\right\} \end{aligned} \tag{26}$$

From the comparison between (25) and (26), $\widehat{w}_j^e \geq w_j^e \geq 0$ since $w_{j-k}^3 \geq 0$. If $\widehat{w}_j^e = 0$, w_j^e should be 0. We can prove $w_j = \widehat{w}_j$ from (11) to (21). If $\widehat{w}_j^e > 0$, w_j^e should be greater than or equal to 0. We can easily prove $w_j = \widehat{w}_j$ from (11) to (21) when $\widehat{w}_j^e > 0$ and $w_j^e = 0$.

Finally, we shall show that the j^{th} arriving customer's waiting time when $w_j^e > 0$ and $\widehat{w}_j^e > 0$. Using (25) and (26), $w_j^e = \widehat{w}_j^e$ as long as $w_{j-k}^3 = 0$. So that, $w_j = \widehat{w}_j$ is obvious from (11) to (21). Now we consider the case 1 when $w_{j-k}^3 > 0$. $w_{j-k}^3 > 0$ means that the token returning interval time between $(j-k-1)^{st}$ and $(j-k)^{th}$ customer to the token pool is less than s^*. $w_{j-k}^3 > 0$ means also $\widehat{w}_j^e > w_j^e$. The $(j-1)^{st}$ customer will take the token returned by the $(j-k-1)^{st}$ customer. And the j^{th} customer will take the token returned by the $(j-k)^{th}$ customer. So that, the interarrival time between the $(j-1)^{st}$ and the j^{th} customer to the second node is less than s^*. This fact tells that some amount of waiting time for the j^{th} customer exists at least in the third node. Therefore, we can prove $w_j = \widehat{w}_j$ from (11) to (21) when $\widehat{w}_j^e > 0$ and $w_j^e > 0$. Furthermore, $w_j^e > 0$, $w_j^1 = 0$ and $w_j^3 = 0$ in the case 1 and $\widehat{w}_j^e > 0$ and $\widehat{w}_j^1 = 0$ in the case 2 can be happened as long as $w_{j-k}^2 = 0$. □

Further, the queueing network with constant service gives another theorem as follows.

Theorem 3. Let us consider the open queueing network as shown in [Figure 6]. For an arriving customer, if $s_1^* \le s^*$, then either case 1 or case 2 has the same waiting time in the queueing network.



[Figure 6] Equivalency between the queueing networks

Proof. For the first k arriving customers, obviously a customer's waiting times in both case 1 and case 2 are the same. In general, for the j^{th} arriving customer, the customer's waiting time in the queueing network depends on the $(j-k)^{th}$ customer's departure time from the queueing network and the $(j-1)^{st}$ customer waiting time in the queueing network. For the j^{th} arriving customer, the $(j-k)^{th}$ customer in case 1 returns its token s_1^* unit time later than the customer in case 2 when $k \le j \le 2k$. his is because no customer ever waits in the second node in case 2. Sequentially, we can prove that each arriving customer's waiting time in the queueing network is the same in both cases. Hence, we may assume that the waiting time for both cases is the same until the $(j-1)^{st}$ customer.

Let us consider the j^{th} arriving customer for case 1. The notations are the same as in previous Theorem 1 and Theorem 2. The customer's waiting time at the first node, w_j^1 is

$$w_j^1 = max\{0,\ w_{j-1}^1 + s_1^* - a_j\} \qquad (27)$$

The interarrival time to the external queue, a_j^e, has changed as follows ;

$$a_j^e = \begin{cases} s_1^* & \text{if } w_j^1 > 0 \\ a_j - w_{j-1}^1 & \text{if } w_j^1 = 0 \end{cases}$$

The waiting time in the external queue depends on the departure time of the $(j-k)^{th}$ customer d_{j-k}, and the arrival time of customer j to the external queue. Since $d_{j-k} = \sum_{i=1}^{j-k} a_i + s_1^* + s^* + \bar{s} + w_{j-k}$ and the j^{th} customer's arrival time to the external queue is $\sum_{i=1}^{j} a_i + s_1^* + w_j^1$, we have

$$w_j^e = max\left\{0,\ w_{j-k} + s^* + \bar{s} - \sum_{i=j-k+1}^{j} a_i - w_j^1\right\} \qquad (28)$$

The interarrival time to the second node also has changed to $a_j^2 = a_j^e + w_j^e - w_{j-1}^e$. Therefore, the customer's waiting time at the second node, w_j^2 is

$$w_j^2 = max\{0,\ w_{j-1}^2 + s^* - a_j^2\} \qquad (29)$$

For case 2, an arriving customer only experiences queueing at the external queue and at the first node. Hence the external queue waiting time is

$$w_j^e = max\left\{0,\ w_{j-k} + s^* + \bar{s} - \sum_{i=j-k+1}^{j} a_i\right\} \qquad (30)$$

The interarrival time to the first node has changed to $a_j^1 = a_j + w_j^e - w_{j-1}^e$. The waiting time in the first node is

$$w_j^1 = max\{0,\ w_{j-1}^1 + s^* - a_j^1\} \qquad (31)$$

In case 1, the j^{th} customer's waiting time is the sum of (27), (28) and (29). The j^{th} customer's waiting time in case 2 is the sum of (30) and (31). We can express the j^{th} customer's waiting time, w_j using w_{j-1} and w_{j-k} for both case. Case 1 and case 2 have the same amount of waiting time when $w_j^1 = 0$ in (27). Without loss of generality, if $w_j^1 > 0$ in (27), at least one of (28) and (29) is positive. This is because the interarrival time in case 1 to the external queue is s_1^*. If $w_j^e > 0$ in (28), (30) becomes positive. So that, a customer's waiting time in both cases can be expressed in terms of w_{j-1}, w_{j-k}, s^* and a_j. If $w_j^e = 0$ in (28), (29) should be positive. This is because the interarrival time to the second node becomes $s_1^* - w_{j-1}^e$. From (27) and (29), the customer's waiting time in case 1 becomes $w_{j-1} + s^* - a_j$. Since (29) is positive, (31) becomes positive. Thus, the customer's waiting time in case 2 becomes $w_{j-1} + s^* - a_j$. Therefore, the waiting time in case 1 is the same as in case 2. □

4. A population constrained queueing network with more than 2-layer.

Let us consider the nested 5 node queueing network as shown in [Figure 3]. If we apply Theorem 1, 2 and 3 into the nested 5 node queueing network, we can make a simpler and an equivalent queueing network as far as customer's waiting time is concerned.

- When $s_{21}^* = max\{s_{21}^*, s_1^*, s_{22}^*\}$.

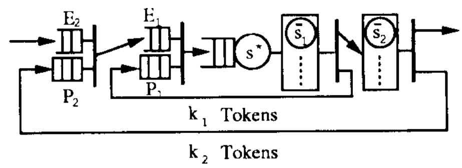
1. using Theorem 1, set s_1^* to s_{21}^* and adjust $\bar{s}_1$ to $T_1 - s_{21}^*$.
2. using Theorem 3, place the semaphore queueing network in the first place.

- When $s_1^* = max\{s_{21}^*, s_1^*, s_{22}^*\}$.

1. using Theorem 3, place the semaphore queueing network in the first place.

- When $s_{22}^* = max\{s_{21}^*, s_1^*, s_{22}^*\}$.

1. using Theorem 2, set s_1^* to s_{22}^* and adjust $\bar{s}_1$ to $T_1 - s_{22}^*$.
2. using Theorem 3, place the semaphore queueing network in the first place.



[Figure 7] 3-node two layered queueing network

Since the longest service time is placed at beginning of the 5 node nested queueing network, there is no queue after the first node. Therefore, the time that a customer spends in the level 2 queueing network is only the sum of the service times. Let s^* be the longest service in the original nested queueing network. Let $\bar{s}_1 = T_1 - s_1^*$ and $\bar{s}_2 = T_{21} + T_{22}$. Then, the original nested queueing network can be transformed into simple two layer queueing network having 3 nodes as shown in [Figure 7].

We know that a customer's waiting time in the queueing network depends on the number of the tokens inside. Since T_1 and T are already defined in section 3, we have the following corrolary.

Corrolary 1. The waiting time of a customer in the 3-node queueing network is independent of the number of tokens k_1, when $k_1 s^* \geq T_1$. As it were, the level 1 semaphore queue has no influence on the customer's waiting time in the queueing network. Therefore, the level 1 semaphore queue can be removed.

Corrolary 2. The waiting time of a customer in the 3-node queueing network is independent of the number of tokens k_2, when $k_2 s^* \geq T$. As it were, the level 2 semaphore queue has no influence on the customer's waiting time in the queueing network. Therefore, the level 2 semaphore queue can be removed.

Now, let us consider n-layer open tandem queueing network with population constraint and constant service times. Again each layer of the population is controlled by a semaphore. Using Theorem 1, 2 and 3 in section 3, it is evident that in order to be able to tackle any number of multilayers of sliding window flow control we need to be able to construct a simpler queueing network. The simplification of n-layer queueing network can be achieved using the following procedures.

1. Place the longest service time in the lowest layer.
2. Construct two-node queueing network in the lowest layer.
3. Adjust the equivalency of the sum of service times in the lowest layer.
4. For the next layer, place only one node having the sum of the service times in the corresponding layer.
5. Repeat until the highest layer.

The features of the simplified n-layer queueing network is similar to two-layer queueing network as shown in [Figure 7]. Therefore, we can represent n-layer queueing network having only $n+1$ nodes. The simplification process is specially limited for the constant service times. Once we obtain the simplified model, then the queueing network can be easy to implement and easy to analyze in terms of customer's waiting time in the queueing network. For the modeling purpose, we can reduce the network dimensionality of the queueing network.

5. Conclusion

In this paper, we have presented a multilayer queueing network model, where the population within each subnetwork is controlled by a semaphore queue. A major characteristic of this model is that the lower layer flow is halted by the state of the higher layer. We present some characteristics that the inter-change of nodes does not make any difference to customer's waiting time in the queueing network under a certain condition. This kind of works has not as yet been studied in the previous research presented in the published literature.

It is evident that in order to be able to tackle any number of multilayers of sliding window flow control we need to be able to construct a simpler queueing network. The simplification process is specially limited for the constant service times using Theorem 1, 2 and 3. A dramatic simplification of the queueing network is shown. Once we obtain the simplified model, then the queueing network can be easy to implement and easy to analyze in terms of customer's waiting time in the queueing network. It is interesting to see how the simplification developed for sliding window flow control, can be applied to multi-layered queueing network.

REFERENCES

[1] Fdida, S., H. Perros, and A. Wilk, "Semaphore queue : Modeling multilayered window flow control mechanisms," *IEEE Transaction on comm.*, Vol.38, No.3 (1990), pp.309-317.

[2] Gihr, O. and P. Kuehn, "Comparison of communication services with connection-oriented and connectionless data transmission," In *Proceedings of the international seminar on computer networking and performance evaluation*, Tokyo Japan, September (1985), pp.431-444.

[3] Mitchell, L. and D. Lide, "End-to-end performance modeling of local area networks," *IEEE J. Select. Comm.*, Vol.SAC-4, Sept. (1986), pp.305-320.

[4] Pennotti, M. and M. Schwartz, "Congestion control in store and forward tandem link," *IEEE Transaction on comm.*, Vol.23 (1975), pp.1434-1443.

[5] Reiser, M., "A queueing network analysis of computer communication networks with window flow control" *IEEE Transaction on comm.*, Vol.27 (1979), pp.1199-1209.

[6] Reiser, M., "Performance evaluation of data communication systems" *Proceeding of the IEEE*, Vol.70-2 (1982), pp.171-182.

[7] Rhee, Y. and H. Perros, "Analysis of an open tandem queueing network with population constraint and constant service times," *European Journal of O.R.*, Vol.92 (1996), pp.99-111.

[8] Rhee, Y. and H. Perros, "On the mean waiting time of a population constrained open tandem queueing network with constant service times," *IIE Transaction*, Vol.30 (1998), pp.973-979.

[9] Schwartz, M., "Performance analysis of the SNA virtual routing pacing control," *IEEE Transaction on comm.*, Vol.30 (1982), pp. 172-184.

[10] Shapiro, G.W. and H.G. Perros, "Nested sliding window protocols with packet fragmentation," *IEEE Trans. Comm.*, Vol.41, no.1 (1993), pp.99-109.

[11] Thomasian, A. and P. Bay, "Analysis of queueing network models with population size constraints and delayed blocked customers," In *ACM SIGMETRICS conf.*, Cambridge, (1984), pp.202-216.

[12] Varghee, G., W. Chou, and A. Nilsson, "Queueing delays on virtual circuits using a sliding window flow control scheme," In *ACM SIGMETRICS conf.*, Minneapolis (1983), pp.275-281.