

* Submitted to the Workshop on *Computer Networks*

Bound on Waiting Time for Multiplexing Traffic Enforced/Shaped by GCRA in ATM Networks

Fang-Jang Kuo

kfj@ilantech.edu.tw

Tel: 886-3-9357400, FAX: 886-3-9369507

Department of Electronic Engineering, National I-Lan Institute of Technology,
1, Sheng-non Road, Section 1, I-Lan, Taiwan, 26041

Abstract

The Generic Cell Rate Algorithm (GCRA) is recommended by the ATM Forum to perform Usage Parameter Control at the User Network Interface in ATM network. In order to facilitate the Call Admission Control and resource allocation procedure, it is important to investigate the characteristics of the model that GCRA-enforced sources are merged together by a multiplexer. Such a multiplexer may stand for the one set in front of a switch so as to concentrate user traffic and reduce the required input ports. It may also stand for the logical multiplexer at the output port of a switch that collecting cells routed from different input ports. Moreover, it may represent the service function of the edge router between the integrated-services (IntServ) networks and the backbone networks that provides differentiated-services (DiffServ).

In this paper, the worst traffic pattern based on the criteria of average waiting time in the multiplexer is found out. The average waiting time is expressed as the function of the GCRA parameters and number of connections. Especially, the analysis also takes the speed-up function into consideration, which is a widely-used technique for ATM multiplexers and switches. The results are the same if the GCRA enforcer is replaced by the GCRA shaper with the same parameter set.

Keywords: ATM, multiplexer, switch, GCRA, waiting time

1. Introduction

Call Admission Control (CAC) and Usage Parameter Control (UPC) are two important steps for congestion control in ATM networks [1]. During the CAC phase, the network user declares the source traffic parameters and the required Quality of Service (QoS), so as for the network to decide whether to accept the connection or not. After a connection is admitted, some UPC scheme must be utilized to monitor and control traffic by detecting whether it conforms to the declared parameters. At the edge node of the network, traffic enforced by the UPC scheme is usually multiplexed by a multiplexer before entering the network. Such a multiplexer may stand for the one set in front of a switch (private or public) so as to concentrate user traffic and reduce the required input ports. On the other hand, it may stand for the logical multiplexer at the output port of a switch that collecting cells routed from different input ports. It may also represent the service function of the edge router between the broadband connection-oriented access network that provides integrated services (IntServ) and the backbone network that provides differentiated services (DiffServ) [2, 3, 4].

In the past, the Leaky Bucket (LB) is the most popular UPC scheme due to its simple algorithm [5, 6]. Based on the cell loss ratio at the multiplexer, [7] has discussed whether the full rate periodic on/off pattern is the worst pattern. Based on the criteria of average queueing delay at the multiplexer, [8] found out the worst pattern under the assumption that multiple cells can pass through a LB at the same time as long as there are sufficient tokens in the pool.

Based on the principle of LB, the ITU-T Recommendation I.371 [1] utilized the Generic Cell Rate Algorithm (GCRA) to define the traffic parameters Peak Cell Rate (PCR) and Cell Delay Variation Tolerance (CDV Tolerance) of an ATM connection. The ATM Forum UNI-specification [9] even applied the algorithm to define Sustainable Cell Rate (SCR, the upper bound of average rate) and Burst Tolerance (BT) so as to facilitate UPC function. There are two pairs of parameters based on GCRA for conformance testing at the

User Network Interface (UNI). The PCR (represented by the peak interarrival time I_p) and the CDV Tolerance (denoted as L_p) are tested by GCRA(I_p, L_p). In the same way, the SCR (represented by the mean interarrival time I_s) and the BT are tested by the GCRA(I_s, L_s), where L_s is the sum of BT and the specified CDV Tolerance (L_p) at the UNI. While the parameter BT is converted by $BT = (MBS - 1)(I_s - I_p)$, where MBS is the expected Maximum Burst Size that can be transmitted at peak cell rate.

So far, the performance characteristic about the multiplexer fed with worst traffic pattern after the GCRA enforcer has not been fully discussed in the literature. This paper is devoted to find out upper bound of the average waiting time at the multiplexer.

2. Description of system model

The model under discussion is shown in Figure 1. Before the N_s traffic sources are merged by the multiplexer, every source is enforced by a dual-stage GCRA enforcer with a parameter set (I_p, L_p, I_s, L_s) . In this article, the parameters are normalized to the unit cell time η_I corresponding to the input link and they may be non-integer. The speed of the output link may be higher than that of the input one. The corresponding unit cell time, which is represented as η_O , is also normalized to η_I . The flow chart of the dual-stage GCRA enforcer is shown in Figure 2. It comprises two single-stage GCRA UPCs.

For the sake of convenient manipulation, the following assumptions are made for the system.

- (1). For the system to be stable, the total sustainable cell rate must be less than the output link capacity, i.e., $\frac{N_s}{I_s} < \frac{1}{\eta_O}$.
- (2). For the purpose of nontrivial discussing, the relation, $\frac{N_s}{I_p} > \frac{1}{\eta_O}$, is assumed to ensure the existence of bursty level congestion.
- (3). The enforcer only discards non-conforming cells, while keeps transparent to conforming cells.

- (4). The transmission time of the source, the processing time of the enforcer and the propagation delay time are neglected. So, any cell departing from a source would be immediately sent out of the enforcer and be multiplexed to the buffer as long as it is a conforming one. It will be explained in section 5 that these items of delay time would not effect the average waiting time.
- (5). Any cell found at the input port of the multiplexer is stored in the buffer. If more than one cells arrive at the multiplexer at the same time, the arrival times of these cells at the buffer are treated as the same.

3. Most Clumping Pattern (MCP)

The first step for discussing the upper bound of delay is to find out that how soon can an arrival process passes through a dual-stage GCRA.

As the flow chart shown in Figure 2, if there are some non-conforming cells before the k -th arriving cell, then $k > l$. However, the non-conforming cell would neither pass through the enforcer nor bring a new $TAT_p(l+1)$ or $TAT_s(l+1)$. As a result, for convenient discussion, it is assumed that the sources only send out conforming cells. That is to say, we may ignore the blocks enclosed by the dashed lines by substituting k for l and assuming that

$$t(k) = \max\{TAT_p(k) - L_p, TAT_s(k) - L_s\}. \quad (1)$$

As we can see from Figure 2,

$$TAT_p(k+1) = \begin{cases} TAT_p(k) + I_p, & \text{if } t(k) \leq TAT_p(k) \\ t(k) + I_p, & \text{if } t(k) > TAT_p(k) \end{cases}$$

As a result, for the k -th cell to minimize $TAT_p(k+1)$, $t(k)$ must be subject to

$$t(k) \leq TAT_p(k). \quad (2)$$

Similarly, for the k -th cell to minimize $TAT_s(k+1)$, $t(k)$ is subject to

$$t(k) \leq TAT_s(k). \quad (3)$$

Besides, the inherent limitation on the input transmission line is

$$t(k) = t(k-1) + \eta_I \quad (4)$$

Combining Eqs (1), (2), (3) and (4), the earliest time for the k -th cell to be a conforming cell and to minimize the allowable arrival time of the $(k + 1)$ -th cell can be expressed as

$$t_e(k) = \max\{TAT_p(k) - L_p, TAT_s(k) - L_s, t(k-1) + \eta_I\}. \quad (5)$$

Without loss of generality, we may let the arrival time of the first cell be $t(1) = 0$ and set the initial value $TAT_p(1) = TAT_s(1) = 0$ such that the first arrival cell is conforming. It results in $TAT_p(2) = I_p$ and $TAT_s(2) = I_s$. Similar as the above derivation, we may let the arrival time of the second cell be

$$t(2) = \max\{I_p - L_p, I_s - L_s, \eta_I\}. \quad (6)$$

Here comes out the problem of determining the maximum of the three terms. Similar problem would be encountered when determining the arrival time of the following cells. It involves the comparison among the GCRA parameters and η_I . It is complicated due to too many conditions. However, we may simplify it by the following procedures.

At first, we define

$$X(\omega) = \omega\eta_I,$$

$$Y(\omega) = \omega I_p - L_p \quad \text{and}$$

$$Z(\omega) = \omega I_s - L_s, \quad \text{where } \omega \in R^+.$$

The inherent relation between the mean interarrival time, peak interarrival time and the unit cell time is

$$I_s \geq I_p \geq \eta_I. \quad (7)$$

They respectively correspond to the slopes of $X(\omega)$, $Y(\omega)$ and $Z(\omega)$ with respect to ω . Three possible relationship among $X(\omega)$, $Y(\omega)$ and $Z(\omega)$ are shown in Figure 3. Indeed, Figure 3-(c) would not appear because $L_s = BT + L_p > 0$. Let

$A(I_p, L_p, I_s, L_s)$: the minimum positive integer such that $Y(\omega) \geq X(\omega)$ and $Y(\omega) \geq Z(\omega)$,

$B(I_p, L_p, I_s, L_s)$: the minimum positive integer such that $Z(\omega) \geq X(\omega)$ and $Z(\omega) \geq Y(\omega)$.

Evidently, B does exist. However, A may not always exist. Now, the relationship between the GCRA parameters and η_I may be divided into two classes. For class ‘a,’ A does exist and for class ‘b,’ A does not exist. As a result, Figure 3-(a)'s is classified as class ‘a’ while

Figure 3-(b)'s is classified as class 'b.'

Class a:

Because both A and B exist, we may describe the discrete characteristics in this class as

$$\begin{cases} X(W) \geq Y(W) \text{ and } X(W) \geq Z(W), 0 \leq W \leq A-1, \\ Y(W) \geq X(W) \text{ and } Y(W) \geq X(W), A \leq W \leq B-1, \text{ where } W \in N \\ Z(W) \geq X(W) \text{ and } Z(W) \geq X(W), B \leq W, \end{cases}$$

So, in this class, we can obtain the result of Eq (6) as $t_a(2) = \eta_l$ (i.e., $W = 1$), where the subscriptor 'a' means class 'a'. From Eq (7) and figure 2, we have

$$TAT_p(3) = TAT_p(2) + I_p = 2I_p \text{ and}$$

$$TAT_s(3) = TAT_s(2) + I_s = 2I_s.$$

Then, according to Eq (5), let the third cell arrives at the earliest time ($W = 2$), i.e.,

$$t_a(3) = \max\{2I_p - L_p, 2I_s - L_s, 2\eta_l\} = 2\eta_l.$$

Similarly, let the i -th cell arrive at the earliest time ($W = i-1$), i.e.,

$$\begin{aligned} t_a &= \max\{(i-1)I_p - L_p, (i-1)I_s - L_s, (i-1)\eta_l\} \\ &= (i-1)\eta_l, 1 \leq i \leq A, i \in N. \end{aligned}$$

Following the same principle we obtain the earliest arrival time of other cells and express them as,

$$t_a(i) = \begin{cases} (i-1)\eta_l, 1 \leq i \leq A, \\ (i-1)I_p - L_p, A+1 \leq i \leq B, \\ (i-1)I_s - L_s, B+1 \leq i. \end{cases}$$

The corresponding pattern is illustrated in Figure 4-(a). If we denote the interarrival times

as $\Delta_a(i) = t_a(i+1) - t_a(i)$, then

$$\Delta_a(i) = \begin{cases} \eta_l, & 1 \leq i \leq A-1, \\ AI_p - L_p - (A-1)\eta_l, & i = A, \\ I_p, & A+1 \leq i \leq B-1, \\ BI_s - L_s - (B-1)I_p + L_p, & i = B, \\ I_s, & B+1 \leq i, \end{cases}$$

where $\eta_l < AI_p - L_p - (A-1)\eta_l < I_p$ and $I_p < BI_s - L_s - (B-1)I_p + L_p < I_s$. For an arrival pattern with such a sequence of interarrival time, all cells (except the first one) arrive with

the shortest allowable time apart from the first one. So, such a pattern is named “Most Clumping Pattern (MCP) of class a.”

Class b:

Because only B exists as shown in Figure 3-(b), we may describe the discrete characteristics in this class as

$$\begin{cases} X(W) \geq Y(W) \text{ and } X(W) \geq Z(W), 0 \leq W \leq B-1, \\ Z(W) \geq X(W) \text{ and } Z(W) \geq Y(W), B \leq W, \end{cases} \quad W \in N.$$

Similar as the derivation for class a, we may assign the earliest arrival time of each cell as

$$t_b(i) = \begin{cases} (i-1)\eta_I, & 1 \leq i \leq B, \\ (i-1)I_s - L_s, & B+1 \leq i, \end{cases}$$

The corresponding pattern is illustrated in Figure 4-(b). Similarly,

$$\Delta_b(i) = t_b(i+1) - t_b(i) \text{ and } \Delta_b(i) = \begin{cases} \eta_I, & 1 \leq i \leq B-1, \\ BI_s - L_s - (B-1)\eta_I, & i = B, \\ I_s, & B+1 \leq i, \end{cases}$$

For an arrival pattern with such a sequence of interarrival time, all cells (except the first one) arrive with the shortest allowable time apart from the first one. So, such a pattern is named “Most Clumping Pattern (MCP) of class b.”

4. Average waiting time

In this section, we shall derive the upper bound for the average waiting time in the buffer of figure 1. Before that, we shall define some notations and introduce two lemmas.

$\alpha(t)$: It is a function representing the number of arriving cells in $[0, t]$. It also stands for the arrival process.

N_b : The number of busy periods.

$\bar{W}_u(\alpha(t))$: It is the average waiting time of the cells in the μ -th busy period as a result of the arrival process $\alpha(t)$.

$\bar{W}(\alpha(t))$: It is the average waiting time of the cells as a result of the arrival process $\alpha(t)$.

Lemma 1.

For a constant rate server with sufficient buffer size,

$$\bar{W}(\alpha(t)) \leq \max_{1 \leq u \leq N_b} \bar{W}_u(\alpha(t)).$$

This lemma has been proved in lemma 2 of [8]. It allows us to find the upper bound of average waiting time only by observing the worst busy period. In the following, we shall investigate how to maximize the average waiting time. Similarly, we consider only the worst busy period which starts at time $t = 0$.

This lemma reveals that the arrival pattern with larger arrival process would result in longer mean waiting time.

Lemma 2.

Assume an arrival process $\alpha(t)$ feeding to a constant-rate server with sufficient buffer size and resulting in a unique busy period $[0, t_{prd}]$. If any of the cell arrival time (except the first one) is shifted forward, and the resulted arrival process is represented as $\alpha^{fw}(t)$, then

$$\bar{W}(\alpha^{fw}(t)) \geq \bar{W}(\alpha(t)).$$

Proof:

Lemma 1 in [10] has shown that the total delay of all cells of $\alpha^{fw}(t)$ is larger than that in $\alpha(t)$. Due to the same total transmitted cells, average waiting time of $\alpha^{fw}(t)$ is also larger than that of $\alpha(t)$.

4.1 Synchronous Clumping Condition (SCC)

We first define the condition that all sources synchronously send out MCP as Synchronous Clumping Condition (SCC). Under this condition, all of the transmitted cells would transparently pass through the dual-stage GCRA. The N_s cells that simultaneously arrive at the multiplexer are called a ‘‘bulk.’’ For such an aggregated traffic process, we would like to find that how many bulks must arrive at the buffer so as to maximize the average waiting time. The derivation must include two classes corresponding to that of

MCP.

4.1.1 Class a

To find the maximum of average waiting time for class-a arrival pattern, we shall divide the arrival pattern into three parts according to interarrival time. For convenience, we let the arrival time of the first bulk be $t=0$ and denote that of the i -th bulk as $t_{a,1}(i)$, where the subscript 'a' means class 'a.'

Part 1: For the beginning A bulks, the arrival time of the i -th bulk is

$$t_{a,1}(i) = \begin{cases} 0, & i = 1 \\ (i-1)\eta_t, & 2 \leq i \leq A \end{cases}$$

The waiting time of the n -th cell of the i -th bulk, $W_{a,1}(i, n)$, is obtained as

$$W_{a,1}(i, n) = [(i-1)N_s + (n-1)]\eta_o - (i-1)\eta_t$$

The total waiting time for the i -th bulk is

$$d_{a,1}(i) = \sum_{n=1}^{N_s} W_{a,1}(i, n) = (i-1)N_s^2 \cdot \eta_o + \frac{N_s-1}{2} N_s \eta_o - (i-1)N_s \cdot \eta_t$$

The total waiting time of the beginning i bulks in this part is

$$\sum_{k=1}^i d_{a,1}(k) = \frac{(i-1)}{2} i \cdot N_s^2 \cdot \eta_o + \frac{N_s-1}{2} N_s \cdot \eta_o \cdot i - \frac{(i-1)}{2} N_s \cdot \eta_t \cdot i \quad (8)$$

The corresponding average waiting time can be obtained by dividing Eq (8) by $(N_s \cdot i)$. It is evident that the average waiting time of a cell is increasing with i . So, the maximal value for this part of arrival pattern must occur at the A -th bulk.

Part 2: The bulk arrival time in this part is expressed as

$$t_{a,2}(i) = (i-1)I_p - L_p, \quad A+1 \leq i \leq B$$

Thus the waiting time of the n -th cell in the i -th bulk is obtained as

$$W_{a,2}(i, n) = [(i-1)N_s + n - 1]\eta_o - I_p(i-1) + L_p, \quad \text{for } A+1 \leq i \leq B, 1 \leq n \leq N_s$$

The total waiting time of the all cells in the i -th bulk is

$$d_{a,2}(i) = \sum_{n=1}^{N_s} W_{a,2}(i, n) = (i-1)N_s^2 \cdot \eta_o + \frac{N_s-1}{2} N_s \cdot \eta_o - [(i-1) \cdot I_p - L_p] N_s \quad (9)$$

The average waiting time of a cell in this part can be obtained by dividing Eq (9) by N_s .

Because we assume that $N_s \eta_o - I_p > 0$, it can be easily proven that the average waiting time of a cell in this part is increasing with i . Besides, it is larger than that of the cell in the A -th bulk, which is the maximum value of part 1. As a result, the maximum average waiting time up to the B -th bulk happens at the B -th bulk and it is expressed as

$$\bar{\omega}_{a,B} = \left[\sum_{k=1}^A d_{a,1}(k) + \sum_{k=A+1}^B d_{a,2}(k) \right] / (B \cdot N_s)$$

Part 3: From the $(B+1)$ -th bulk on, the arrival time of the i -th bulk is $t_1(i) = (i-1)I_s - L_s$.

To maintain a busy period, any bulk must arrive before its previous bulk is completely served. Thus, we have $[(i-1)I_s - L_s] \cdot \eta_l \leq (i-1)N_s \cdot \eta_o$ and obtain the maximum number of bulk of a busy period as

$$\varepsilon(I_p, L_p, I_s, L_s) = \left\lfloor \frac{L_s \cdot \eta_l}{I_s \cdot \eta_l - N_s \cdot \eta_o} \right\rfloor + 1$$

The waiting time of the n -th cell in the i -th bulk is obtained as

$$W_{a,3}(i, n) = [(i-1)N_s + n - 1]\eta_o - I_s(i-1) + L_s, \text{ for } B+1 \leq i \leq \varepsilon \text{ and } 1 \leq n \leq N_s$$

The total waiting time of the all cells in the i -th bulk in part 3 is

$$d_{a,3}(i) = \sum_{n=1}^{N_s} W_{a,3}(i, n) = (i-1)N_s^2 \cdot \eta_o + \frac{N_s - 1}{2} N_s \cdot \eta_o - [(i-1) \cdot I_s - L_s] N_s \quad (10)$$

The average waiting time of a cell in the i -th bulk in part 3 can be obtained by dividing Eq (10) by $(N_s \cdot i)$. Since we assumed $N_s \eta_o < I_s$, it can be proven to decrease with i . As a result, the maximum average waiting time in this part is happen at the $(B+1)$ -th bulk, and is expressed as

$$\bar{\omega}_{a,B+1} = \left[\sum_{k=1}^A d_{a,1}(k) + \sum_{k=A+1}^B d_{a,2}(k) + d_{a,3}(B+1) \right] / [(B+1) \cdot N_s]$$

Comparing the result in part 2 and part 3, whether the maximum value happens at the B -th or $(B+1)$ -th bulk depends on the relation between $\bar{\omega}_{a,B}$ and $\bar{\omega}_{a,B+1}$, which depends on the GCRA parameters and the system parameters. So, the maximum of average waiting time under the synchronous clumping condition (SCC) is expressed as

$$\tilde{W}_a = \begin{cases} \bar{\omega}_{a,B}, & \text{if } \bar{\omega}_{a,B} \geq \bar{\omega}_{a,B+1} \\ \bar{\omega}_{a,B+1}, & \text{otherwise} \end{cases} \quad (11)$$

4.1.2 Class b

To find the maximum of average waiting time corresponding to class b pattern, we shall divide the arrival pattern into two parts for discussion.

Part 1: For the beginning B bulks, the arrival time of the i -th bulk is

$$t_{b,1}(i) = \begin{cases} 0, & i = 1 \\ (i-1)\eta_I, & 2 \leq i \leq B \end{cases}$$

Because the arrival pattern in this part is the same as that in part 1 of class a, we have $W_{b,1}(i, n) = W_{a,1}(i, n)$ and $d_{b,1}(i) = d_{a,1}(i)$, except that the maximum value of i is B rather than A . As a result, the average waiting time of a cell is increasing with i and is maximized at the B -th bulk. The average waiting time of the beginning B bulks is obtained as

$$\bar{\omega}_{b,B} = \sum_{k=1}^B d_{b,1}(k) / (B \cdot N_s) = \frac{(B-1)}{2} N_s \cdot \eta_O + \frac{N_s - 1}{2} \eta_O - \frac{(B-1)}{2} \eta_I$$

Part 2: The bulk interarrival time in this part is denoted as

$$t_{b,2}(i) = (i-1)I_s - L_s, \text{ for } B+1 \leq i \leq \varepsilon.$$

Because the arrival pattern in this part is the same as that in part 3 of class a, we have $W_{b,2}(i, n) = W_{a,3}(i, n)$ and $d_{b,2}(i) = d_{a,3}(i)$. Similarly, the average waiting time of a cell in the i -th bulk of this part is decreasing with i . So, the maximum average waiting time in this part is happened at $B+1$, and is expressed as

$$\bar{\omega}_{b,B+1} = \left[\sum_{k=1}^B d_{b,1}(k) + d_{b,2}(B+1) \right] / [(B+1) \cdot N_s]$$

As a result, the maximum of average waiting time for class b pattern can be expressed as

$$\tilde{W}_b = \begin{cases} \bar{\omega}_{b,B}, & \text{if } \bar{\omega}_{b,B} \geq \bar{\omega}_{b,B+1} \\ \bar{\omega}_{b,B+1}, & \text{otherwise} \end{cases} \quad (12)$$

4.2 Any condition other than SCC

We assume the first cell of all sources arrives at the reference origin of time ($t = 0$),

while the first cell of source n arrives at time $d_n \geq 0$. Recalling the SCC, because all of the MCP's are sent out synchronously, we may denote and express the arrival time of the i -th cell of n -th source as

$$t^S(i, n) = t(i),$$

where the subscript 'S' denotes SCC and $t(i)$ may be $t_a(i)$ or $t_b(i)$, depending on the class.

Let's now consider one of the general conditions other than SCC. For convenience, we denote it as a G condition. Then, we can denote and express the arrival time as

$$t^G(i, n) = d_n + t(i).$$

Evidently, $t^G(i, n) \geq t^S(i, n)$, $1 \leq i \leq \varepsilon$, $1 \leq n \leq N_s$, $1 \leq i \leq \varepsilon$, $1 \leq n \leq N_s$. It follows that the relation between the corresponding arrival processes of n -th source is

$$\alpha^G(t, n) \leq \alpha^S(t, n). \quad (13)$$

From the viewpoint of the multiplexer, the aggregated arrival process is expressed as

$$\alpha^G(t) = \sum_{n=1}^{N_s} \alpha^G(t, n) \quad \text{and} \quad \alpha^S(t) = \sum_{n=1}^{N_s} \alpha^S(t, n).$$

From Eq (13), $\alpha^G(t) \leq \alpha^S(t)$. According to Lemma 2, we may say that the maximum of average waiting time resulted by G condition is upper bounded by SCC.

Now we consider any of the condition similar as G condition, except that some of the sources does not send out MCP's. As we have described, MCP is the pattern that all the cells arrive with the shortest interval apart from the first cell. As a result, if the n -th source does not send out MCP, then the corresponding arrival time of the i -th cell would be larger than $t^G(i, n)$. The resulted arrival process at the multiplexer is less than $\alpha^G(t)$ and $\alpha^S(t)$ accordingly. From Lemma 2, we may say that the maximum of average waiting time for the system in Figure 1 is upper bounded by \tilde{W}_a of Eq (11) for class-a arrival pattern or by \tilde{W}_b of Eq (12) for class-b arrival pattern, which is obtained under SCC.

5. Numerical results and discussion

According to the results obtained in the last section, we run some data for different conditions. The maximal average waiting time with respect to the parameter I_s is shown in figure 5-(a). In this figure, the arrival pattern may be classified to class ‘a’ or ‘b’ depending on the value of I_s . The ‘x’ mark is made on the x-axis if the pattern belongs to class ‘a.’ Besides, the maximum value may be resulted by the B -th bulk or the $(B+1)$ -th bulk. If the maximum value is resulted by the $(B+1)$ -th bulk, there is a dot in the circle mark of the maximum value. It is predictable that the waiting time decreases with I_s , which represents the mean interarrival time.

Fig 5-(b) shows the result when the output server speeds up to double ($\eta_o = 0.5$). The corresponding waiting time is a little less than half of that of the normal speed (i.e., equal to that of the input link). However, the values of I_s that result in maximum waiting time at the $(B+1)$ -th in figure 5-(a) do not always result in the similar result. Also, as we can expect, the values of A and B are independent of the speed of the output server.

The waiting time with respect to the parameter L_s is shown in figure 6-(a). Because larger L_s means longer maximum burst size of the arrival process, the waiting time is longer accordingly. Similar as that of figure 5-(b), figure 6-(b) shows the result when the server speeds up to double. The corresponding waiting time is a little less than half of that of the normal speed.

The more general pattern is shown in figure 7-(a), where the values of B are much larger than that of A . The speed-up characteristic shown in figure 7-(b) is similar to that of figure 5-(b) and 6-(b).

Figure 8-(a) and 8-(b) show the waiting time corresponding to I_p , where I_s being 60 and 40, respectively. With different I_s , the maximum waiting time may occur at different bulks – B -th bulk or $(B+1)$ -th bulk.

So far, the GCRA which is utilized in the system of Figure 1 is an enforcer rather than a shaper [11]. For an invalid cell, the shaper would store it in the buffer and then forward it as soon as it can be judged as a conforming cell. However, no matter whether the

GCRA shaper or enforcer is employed, the worst patterns that may appear at the output ports are the same.

As we have revealed in section 3, many arrival patterns may result in the same departure pattern out of the GCRA enforcer. So, the SCC is not the only way for the sources to construct the maximum queue length in the buffer. Similarly, if a GCRA shaper is utilized, due to the shaping function, many arrival patterns may result in the same maximum queue length at the buffer.

The derivations in this article are based on the assumption that the cell transmission time, GCRA processing time and propagation delay are neglected. If these quantities are not neglected and equal to T_D , the same result would be obtained if the sent time of every cell is shifted forward by T_D . As a result, this assumption has no effect on the obtained maximum of average waiting time.

6. Conclusion

In this article, we have found out the maximum of average waiting time for multiplexing sources enforced by the GCRA policing/shaping scheme. It is obtained when the N_s sources send out cells under synchronous clumping condition (SCC). Although the relation between the GCRA parameters and unit slot time is so complicated, we find out the result by simply classifying as two classes only. Deterministic delay bounds for these two-class arrival patterns were obtained. Especially, the analysis also includes the speed-up index of the multiplexer. The results can be helpful to Call Admission Control and resource allocation. So far, they are obtained in the worst case of homogeneous environment (i.e., the parameters of all GCRA's are the same). Although it is very hard in our opinion, the further work is to find the probability density functions of waiting time even in the heterogeneous environment.

References

- [1] ITU-TSS Recommendation I.371, "Traffic Control and Congestion Control in B-ISDN," Frozen Issue, March 1995.
- [2] Y. Bernet et al. "A framework for Integrated services operation over Diffserv networks, RFC 2998, Nov. 2000.
- [3] B. Nandy et al, "Aggregate flow control: Improving assurances for differentiated services network," Infocom 2001.
- [4] S. Blake et al, "An Architecture for Differentiated Services," RFC 2475, December 1998.
- [5] G. Niestegge, "The leaky bucket policing method in the ATM network," International Journal of Digital and Analog Communications Systems, vol. 3, pp. 187-197, 1990.
- [6] N. Yamanaka, Y. Sato and K. Sato, "Performance limitation of bucket algorithm for usage parameter control and bandwidth allocation methods," IEICE Trans. Communication., vol. E75-B, no. 2, pp. 82-86, Feb. 1992.
- [7] B. T. Doshi, "Deterministic rule based traffic descriptors for broadband ISDN: worst case behavior and connection acceptance control," IEEE GLOBECOM'93, pp. 1759-1764.
- [8] D. C. Lee, "Effect of Leaky Bucket parameters on the average queue delay: worst case analysis," IEEE INFOCOM'94, pp. 482-489.
- [9] The ATM Forum, ATM User-Network Interface (UNI) Specification Version3.1, Prentice-Hall, 1995.
- [10] J. S. Wu and F. J. Kuo, "Performance analysis of an ATM multiplexer for GCRA-enforced traffic," Proc. of National Science Council R.O.C, pp. 466-475, vol. 21, no. 5, 1997.
- [11] F. J. Kuo, "Design and evaluation of GCRA shaper and enforcer architecture in ATM networks," International Journal of communication system, vol.9, pp 13-22, 1996.

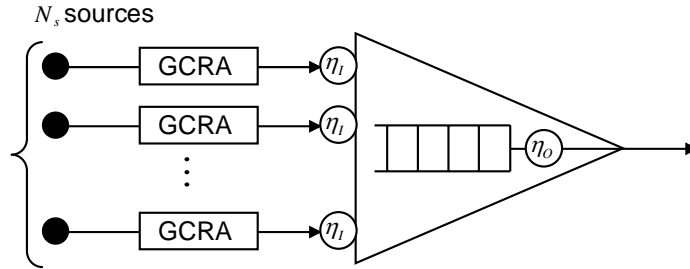


Figure 1. System model

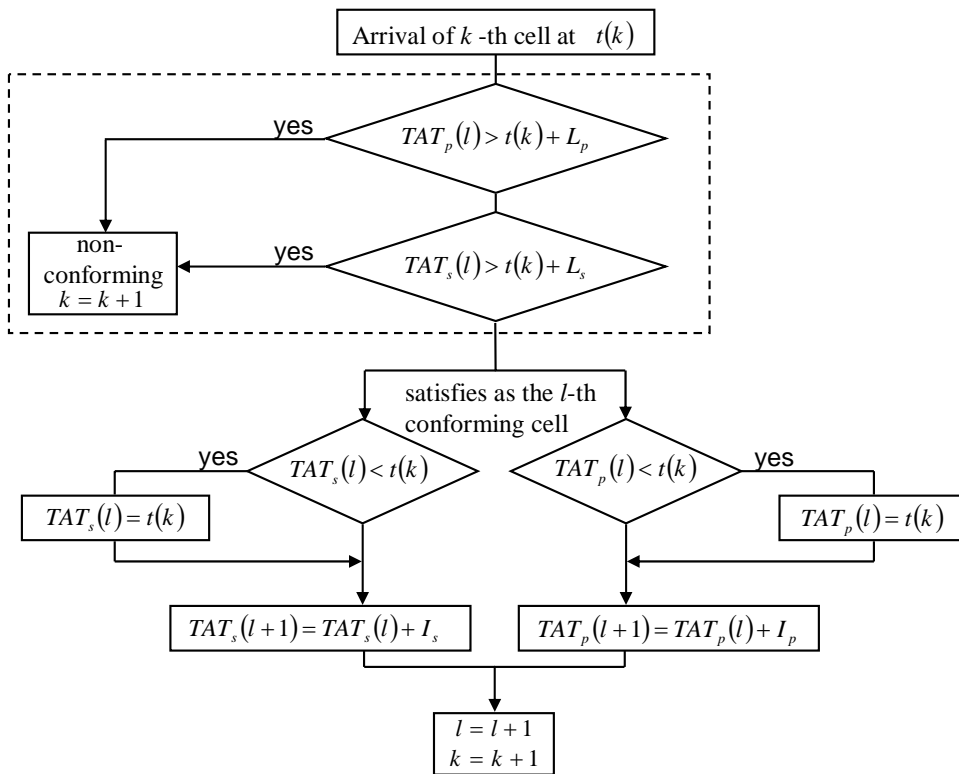


Figure 2. Dual-stage GCRA

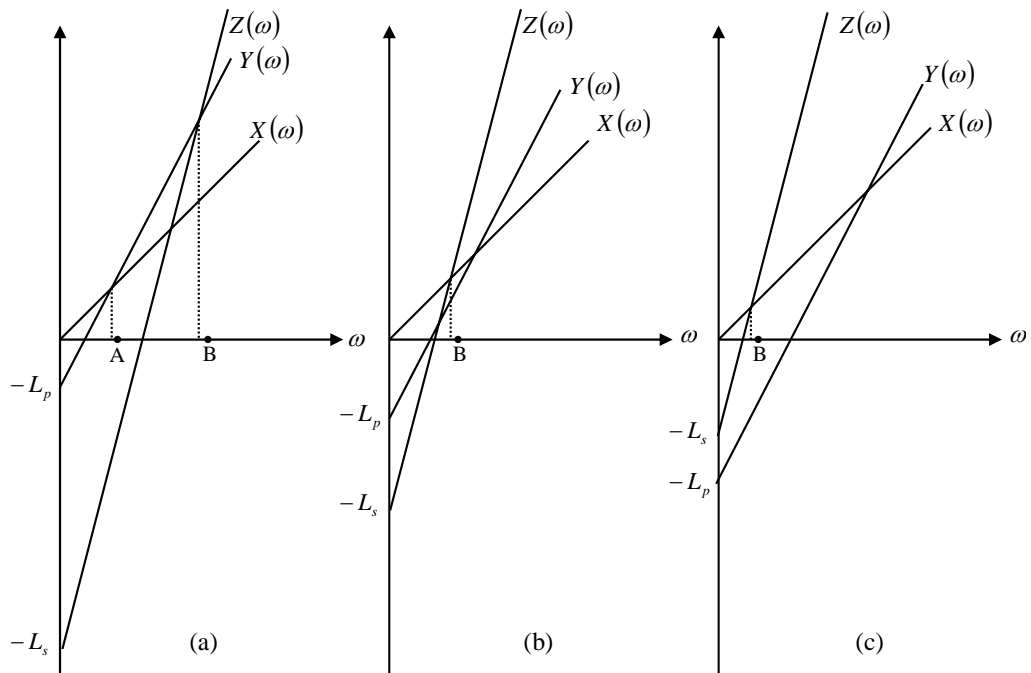


Figure 3. Three possible relationship among $X(\omega)$, $Y(\omega)$ and $Z(\omega)$

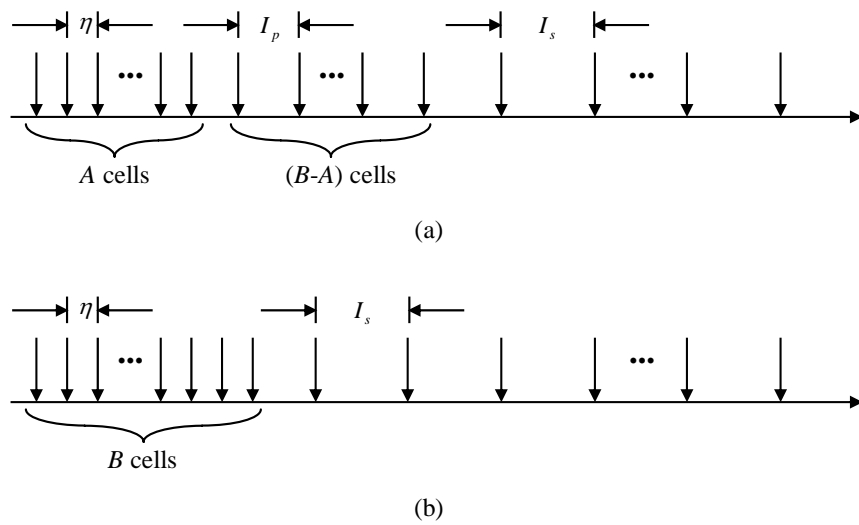


Figure 4. Most Clumping Patterns (MCP) of two classes

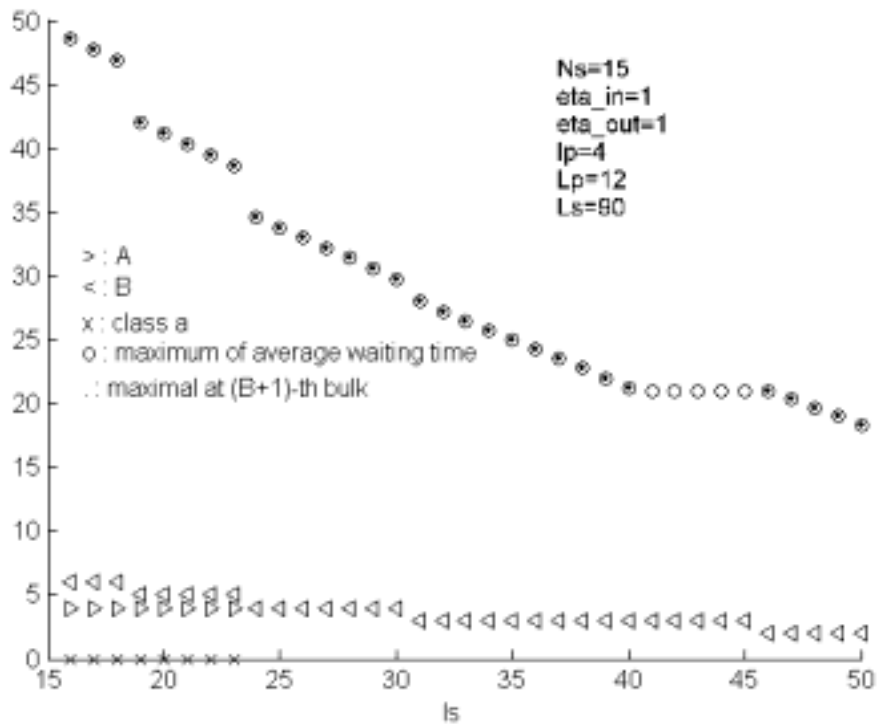


Figure 5-(a) Performance characteristics v.s. I_s (with normal speed o/p server)

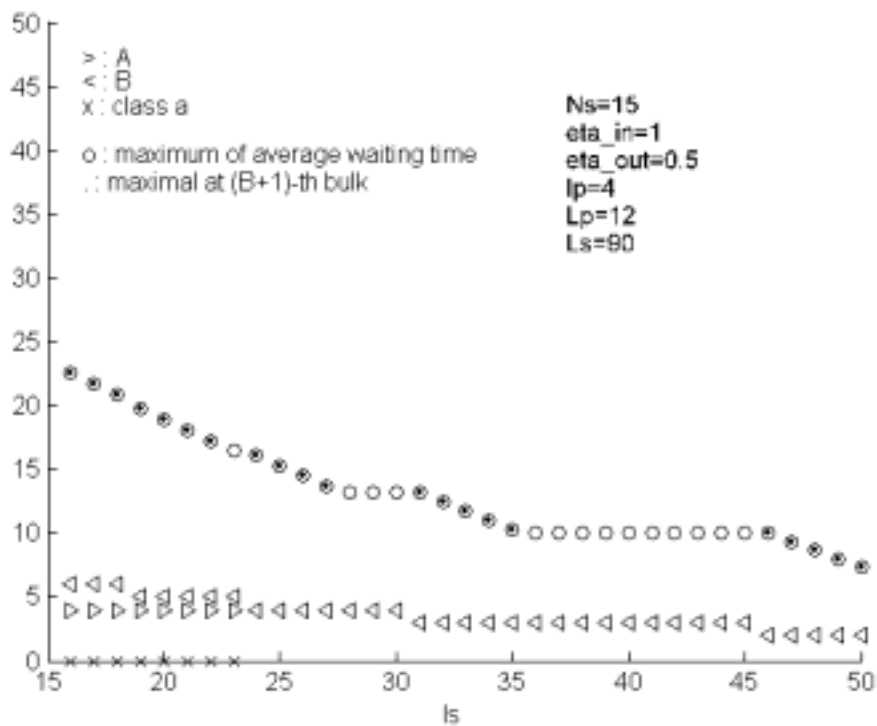


Figure 5-(b) Performance characteristics v.s. I_s (o/p server speed up to double)

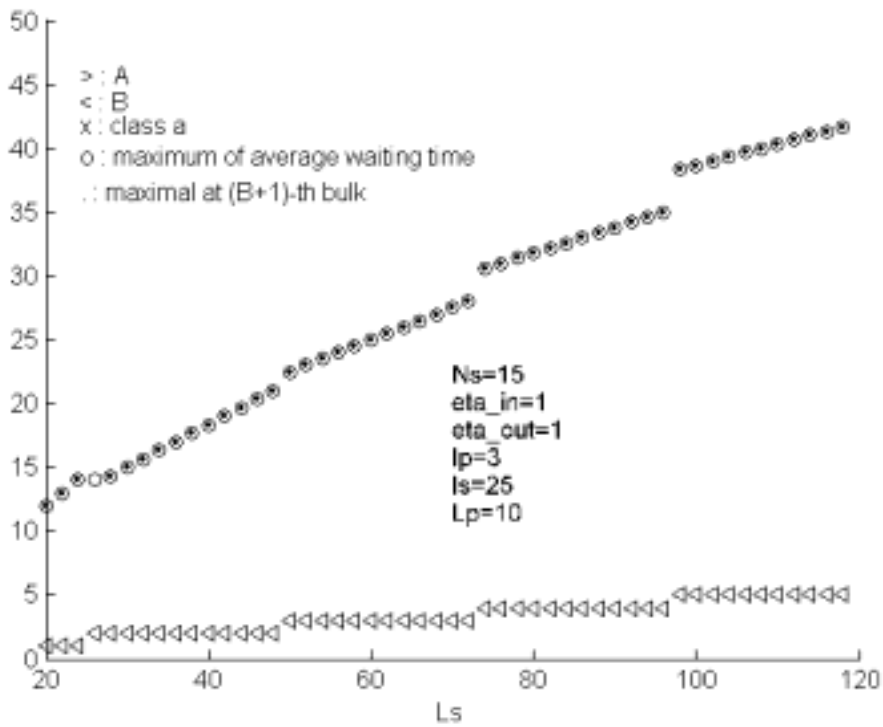


Figure 6-(a) Performance characteristics v.s. L_s (with normal speed o/p server)

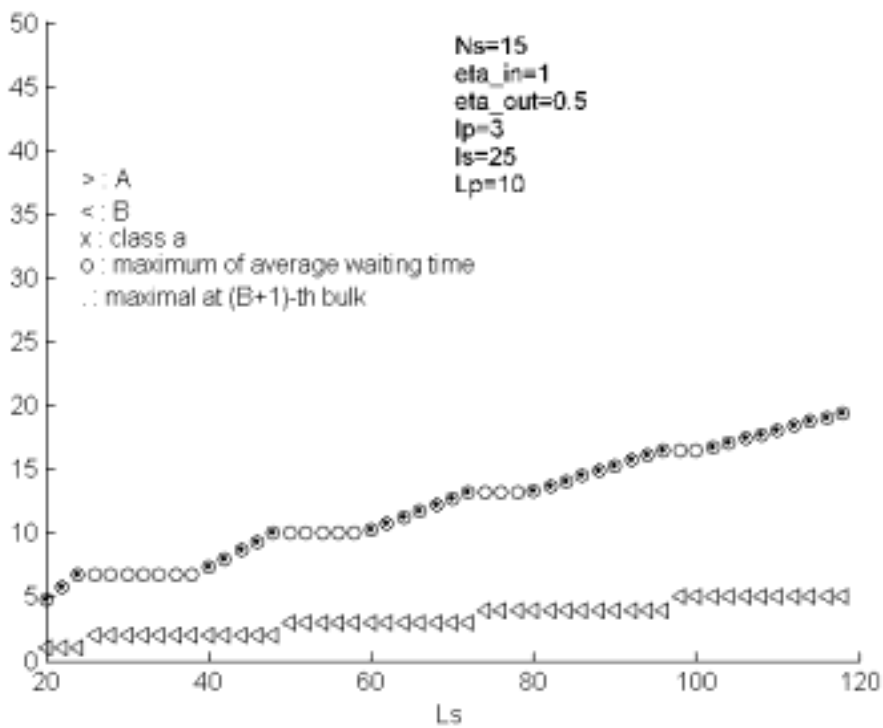


Figure 6-(b) Performance characteristics v.s. L_s (o/p server speed up to double)

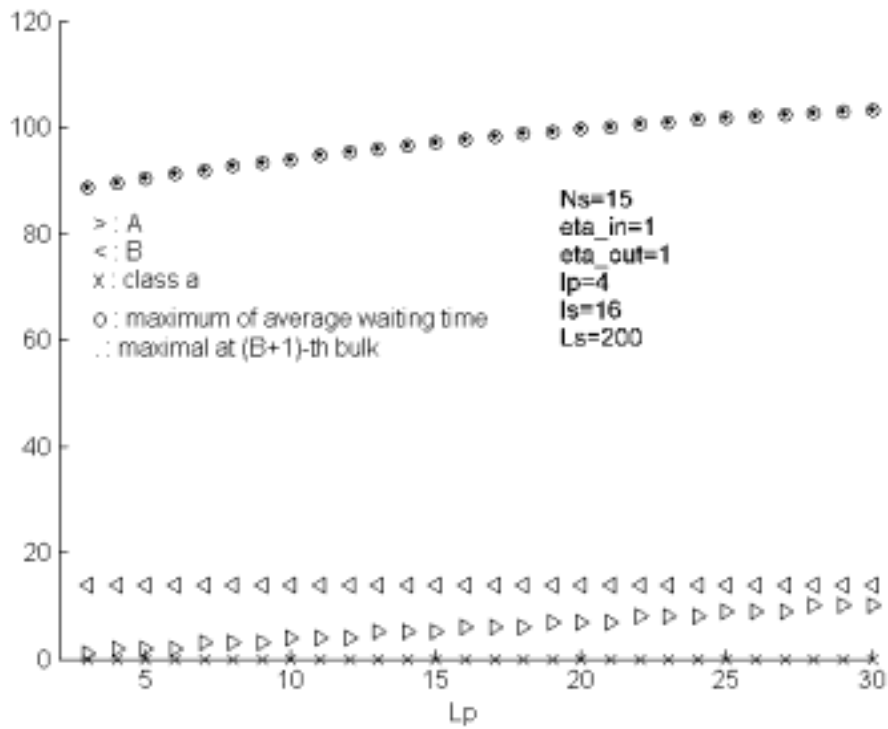


Figure 7-(a) Performance characteristics v.s. L_p (with normal speed o/p server)

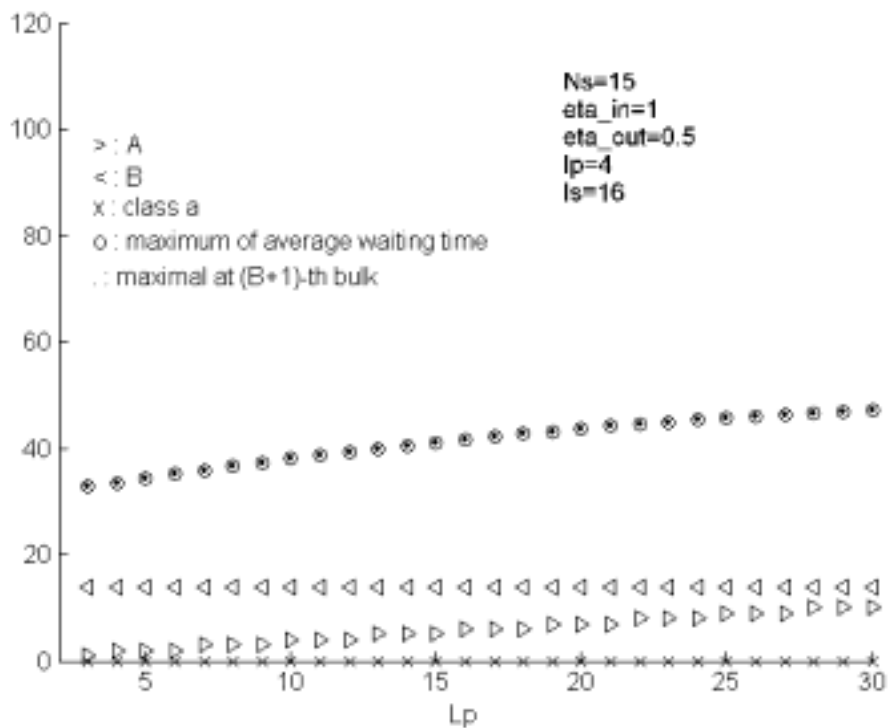


Figure 7-(b) Performance characteristics v.s. L_p (o/p server speed up to double)

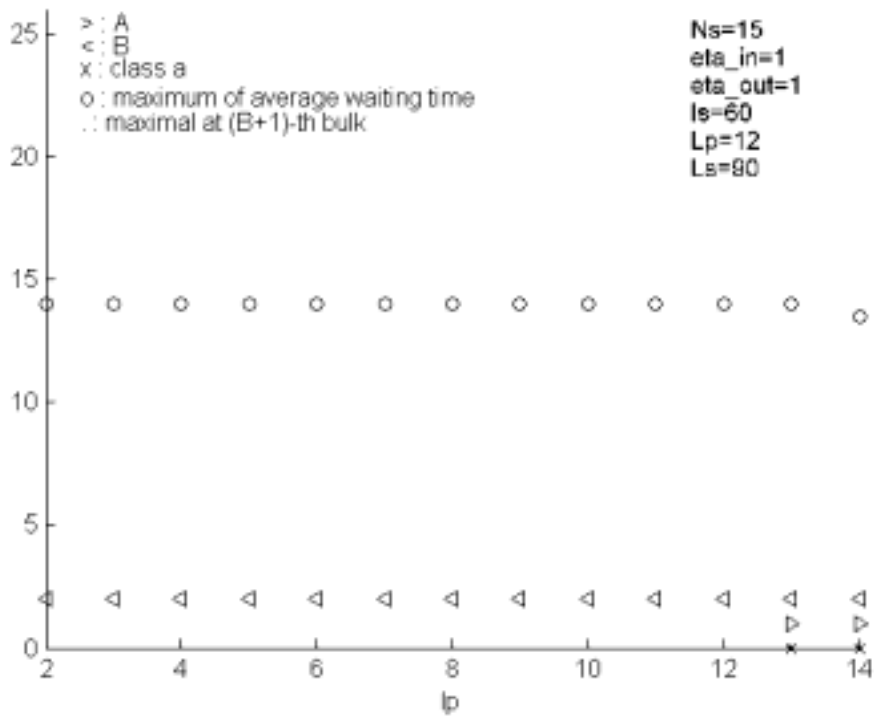


Figure 8-(a) Performance characteristics v.s. I_p (with $I_s=60$)

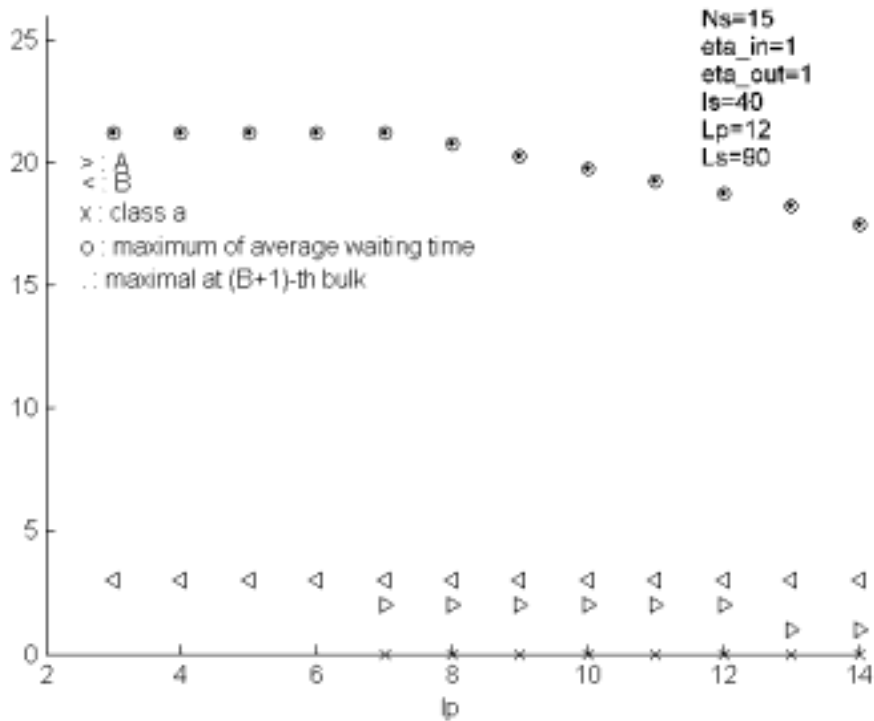


Figure 8-(b) Performance characteristics v.s. I_p (with $I_s=40$)