Performance Analysis of Rate-based Congestion Control Scheme and Choice of High and Low Thresholds

Yuan-Cheng Lai Ying-Dar Lin
Department of Computer and Information Science
National Chiao Tung University
1001 Ta Hsuch Rd., Hsinchu, Taiwan

Abstract

The paper presents a performance analysis of a rate-based congestion control mechanism. The switching capacity of the buffer is finite to reflect real conditions. Using a differential equation approach, we get the closed-form equations of cell loss probability and utilization. Numerical results are given to show that our analysis is correct.

In rate-based control, the important issue is how to determine congestion occurrence and congestion relief. The most common method is to set two thresholds of queue length, a high threshold and a low threshold. The values of these two thresholds seriously influence the system performance. Hence, we present the concept of best area to determine how to set the high and low thresholds to guarantee good performance, i.e., cell loss probability is zero and utilization is one, if it is possible. When good performance is not achieved due to too many connections or too large propagation delay, some rules are also given to prevent unnecessary cell-loss and under-utilization.

1 Introduction

The rate-based flow control for available bit rate(ABR) traffic is defined in the ATM forum Traffic Management Specification Version 4.0 to provide a wide range of non-real time applications [1]. In this control scheme, the allowed cell transmission rate of each ABR connection is dynamically regulated by feedback information from the network [1-3]. If the network is congested, the source end decreases its cell transmission rate when it receives congestion indication. Also, the source end increases its cell transmission rate when congestion is relieved. The rate-based control mechanism can efficiently control the connection flows and utilize the network bandwidth.

Recently several analyses and simulations have been conducted for rate-based control schemes [4-7]. These analyses assume that no cell loss occurs at the switch. Based on this assumption, the maximum queue length (buffer requirement) at the switch is derived. However, the switch capacity of a buffer is of finite size, which may be less than the maximum queue length derived. Actually the maximum queue length ought not to exceed the buffer size at the switch, and extra arrival cells will be lost when the buffer is full.

In this paper, we assume the switch capacity of the buffer is finite to reflect the real conditions. Also the performance issues that we are concerned with change from the maximum queue length to cell loss probability and utilization. The cell loss probability is an important parameter of QoS, and utilization is an important parameter of network performance. The contribution of this paper is the analysis of rate-based congestion control. We get the closed-form equations of cell loss probability and utilization.

In rate-based control, the common method is to set two thresholds of queue length to determine congestion occurrence and congestion relief. The values of these two thresholds vastly influence the system performance. For example, when the high threshold is large, the switch detects congestion later and thus, cell loss probability would be larger. Similarly, when the low threshold is small, the detection time of congestion relief would be late and cause buffer to underflow and utilization to drop. Hence some rules, which determine how to set the high and low thresholds, are needed to avoid unnecessary cell-loss and under-utilization.

2 Rate-based Control

First we briefly introduce the basic operation of a close-loop rate-based control algorithm. The source end system (SES) sends a forward Resource Management (RM) cell every N_{RM} data cells to probe the congestion status of the network. The destination end system (DES) returns the forward RM cell as a backward RM cell to the SES. Depending on the received backward RM cell, SES adjusts its allowed cell rate (ACR), which is bounded between Peak cell rate (PCR) and Minimum cell rate (MCR).

The RM cell contains a 1-bit congestion indication (CI) which is set to zero, and an explicit rate (ER) field which is set to PCR initially by the SES. Depending on the different ways to indicate congestion status, two types of switches are implemented. One is the Explicit Forward Congestion Indication (EFCI) switch, the other is the Explicit Rate (ER) switch. In the EFCI type, the switch in congestion status sets the EFCI bit to one (EFCI=1) in the header of each passing data cells. The DES, if a cell with EFCI=1 has been received, marks the CI bit (CI=1) to indicate congestion in each backward RM cells. In the ER type, the switch sets the EFCI bit of the RM cells to indicate whether there is congestion or not, and sets the ER field to indicate the bandwidth the VC should use.

When the SES receives a backward RM cell, it modifies its ACR using additive increase and multiplicative decrease. The new ACR is computed as follows, depending on CI:

$$ACR = \max(\min(ACR + N_{RM} \cdot AIR, ER), MCR), \text{if CI=0},$$

$$ACR = \max(\min(ACR \cdot (1 - \frac{N_{RM}}{RDF}), ER), MCR), \text{if CI=1},$$

where AIR is the additive increase rate and RDF is the rate decrease factor.

In this paper, we focus on the EFCI switch and use a simple model as shown in Fig. 1. There are N_{VC} homogeneous traffic sources sharing a bottleneck link where the bandwidth is BW. We assume that each SES always has cells to send. This assumption allows us to investigate the performance of an EFCI switch in the most stressful situations.

Congestion condition is determined by the switch according to its queue length. There are two values, high threshold Q_h and low threshold Q_l, which decide whether congestion occurs or not. When the queue length exceeds Q_h, the EFCI bit of passing data cells is set to one to indicate congestion. Congestion is relieved when the queue length drops below Q_l.

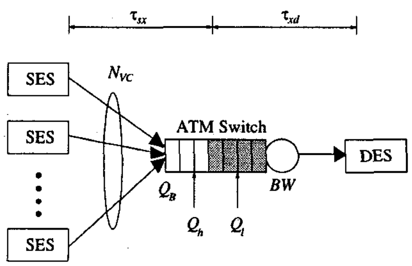

Figure 1: Analytic model for the rate-based congestion control

We define τ_{sx} as the propagation delay between the SES and the switch, and τ_{xd} as the propagation delay between the switch and the DES. Also, the feedback propagation delay from the switch to the SES is denoted by τ_{xds} and the round trip propagation delay is denoted by τ. Thus we get the relation $\tau_{xds} = \tau_{sx} + 2\tau_{xd}$ and $\tau = 2(\tau_{sx} + \tau_{xd})$. The propagation delay is a critical parameter of system performance.

Note that the buffer at the switch is of finite capacity denoted by Q_B. Due to this fact, the arrival cells shall be lost when the buffer is full. The act of cell loss causes the derivation of the dynamic behavior of the allowed cell rate, $ACR(t)$, and the buffer size, $Q(t)$, to have some differences with other previous studies.

3 Basic description

We describe the four elementary phases and four different cycles in this section. A cycle is composed of some elementary phases. Depending on the various compositions of the phases, four different cycles are constructed.

3.1 Elementary phases

The evolution of the ACR can be characterized into four elementary phases differentiated by (1) the increase or decrease of the ACR, and (2) the speed of the increase or decrease. The speed of adjustment is determined by the rate RM cells are received in the SES. Meanwhile, this rate is influenced by the status of the buffer τ_{xds} time units before. So actually the evolution of the ACR is partitioned into four phases depending on the $ACR(t)$ and $Q(t - \tau_{xds})$.

• *Phase 1*: $ACR(t) \uparrow$, $Q(t - \tau_{xds}) > 0$

During this period, each SES receives the backward RM cells with CI=0 at constant rate $BW/N_{VC}N_{RM}$. When a backward RM cell is received, the SES increases its ACR by $N_{RM}AIR$, to a rate not greater than PCR. We use a continuous time approximation to model this discrete increase for simplicity. This approximation is accurate to characterize the dynamic behavior of rate-base control [7]. Thus the following differential equation for $ACR_1(t)$ is given by

$$\frac{dACR_1(t)}{dt} = \frac{BW \cdot AIR}{N_{VC}},$$

which gives

$$ACR_1(t) = min(ACR_1(0) + \frac{BW \cdot AIR}{N_{VC}} t, PCR). \quad (1)$$

• *Phase 2*: $ACR(t) \downarrow$, $Q(t - \tau_{xds}) < Q_B$

During this period, each SES receives the backward RM cells with CI=1 at constant rate $BW/N_{VC}N_{RM}$. When a backward RM cell is received, the SES reduces its ACR by $N_{RM}ACR/RDF$, to a rate not lower than MCR. We thus have the following continuous approximation for $ACR_2(t)$.

$$\frac{dACR_2(t)}{dt} = -ACR_2(t)\frac{BW}{N_{VC}RDF},$$

which gives

$$ACR_2(t) = max(ACR_2(0)e^{-\frac{BW}{N_{VC}RDF}t}, MCR). \quad (2)$$

• *Phase 3*: $ACR(t)\uparrow, Q(t-\tau_{xds})=0$

When the buffer is empty, the switch is not fully utilized. In order not to waste the bandwidth with idle time, we ought to avoid the occurrence of this phase. In this phase, RM cells arrive at the SES depending on its own rate τ time before. i.e. $ACR_3(t-\tau)$. So the differential equation for $ACR_3(t)$ is given by

$$\frac{dACR_3(t)}{dt} = ACR_3(t-\tau)\cdot AIR,$$

which gives

$$ACR_3(t) = min(ACR_3(0)e^{\beta t}, PCR), \qquad (3)$$

where β is given as the root of the equation $\beta = AIRe^{-\beta\tau}$.

• *Phase 4*: $ACR(t)\downarrow, Q(t-\tau_{xds})=Q_B$

When the buffer is full, the arrival cells are discarded at the switch. In order not to waste the bandwidth with retransmission of the lost cells, we also hope not to enter this phase. In this phase, although the cell loss happens, the number of data cells between two consecutive RM cells can be approximated as N_{RM} because data cells and RM cells are both discarded in the switch when buffer is full. The approximation causes the easier analysis, and it is verified as a good approach [8]. Thus the arriving rate of RM cells at the SES is $BW/N_{VC}N_{RM}$. The behavior of ACR during this phase is given by

$$\frac{dACR_4(t)}{dt} = -ACR_4(t)\frac{BW}{N_{VC}RDF},$$

which gives

$$ACR_4(t) = max(ACR_4(0)e^{-\frac{BW}{N_{VC}RDF}t}, MCR). \qquad (4)$$

Although the equations of phase 2 and phase 4 are the same, we differentiate between them since there is different meaning.

Note the MCR and PCR are not considered in the analysis below since we assume $MCR = 0$ and $PCR = BW$. The analysis of $MCR > 0$ or $PCR < BW$ can be expanded from our analysis, but the equations are more complex.

3.2 Various cycles

According to the elementary phases described in the last subsection, four different cycles are constructed. In all four cycles, phase 1 and phase 2 are the essential phases. The cycles are differentiated by whether phase 3 and phase 4 occur or not.

• *Best cycle* :

This perfect cycle is composed of phase 1 and phase 2. They occur alternatively as Fig. 2. When the queue length is below the low threshold, i.e., at the time t_{Q_l}, congestion is relieved in the switch and this congestion relief signal will arrive at the SES after the propagation delay τ_{xds}. Then the SES increases its ACR and phase 1 begins. For convenience in explaining the time shift between $ACR(t)$ and $Q(t)$ caused by the propagation delay, we use the t^+ and t^- to denote the "after τ_{xds}" and "before τ_{sx}" time units from the time t, respectively. So the phase 1 begins at $t^+_{Q_l}$, and similarly, the beginning time of phase 2 is $t^+_{Q_h}$, where t_{Q_h} is the time that queue length exceeds Q_h.

Let t_0 and t_1 be two time stamps for our analysis. At the time t_0 and t_1, ACR is equal to BW/N_{VC}.

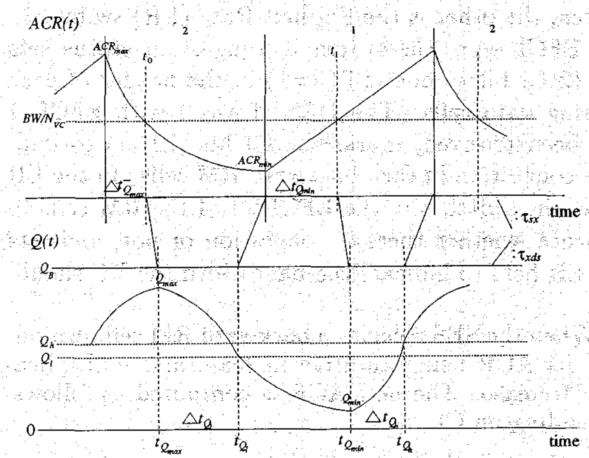

Figure 2: Evolution of ACR and Q in best cycle

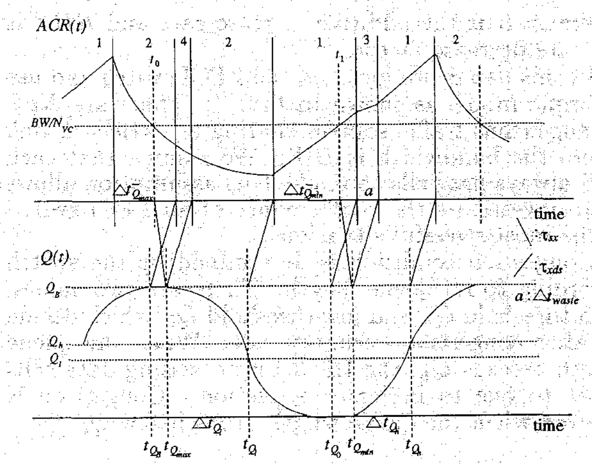

Figure 3: Evolution of ACR and Q in worst cycle

• *Worst cycle* :

Phase 3 and 4 occurs in this worst case. The cycle behaves as in Fig. 3. In this figure, phase 4 interrupts phase 2 and phase 3 interrupts phase 1.

When the queue length exceeds Q_B, cell loss starts. Hence phase 4 begins at the $t^+_{Q_B}$, and consequently, the ending time of phase 4 is after τ from t_0. This is because the aggregate cell arrival rate at the switch is below the bandwidth BW at time t_0, and the queue starts to decrease after the propagation delay between the SES and the switch, τ_{sx}. Similarly, the beginning time of phase 3 is $t^+_{Q_0}$, where t_{Q_0} is the time that queue length reaches zero, and the ending time of phase 3 is τ time units after the time t_1.

• *Cell-loss cycle* :
During this cycle, phase 4 occurs and phase 3 does not occur. From Fig. 2 and 3, the sequence of this cycle is phase 1-2-4-2. Also, the beginning time of each phase is obtained in the same way as before.
• *Under-utilization cycle* :
Phase 3 occurs and phase 4 does not occur in this cycle. The sequence of this cycle is phase 1-3-1-2.

4 Steady State Analysis

If we do not carefully choose Q_h and Q_l, not only do the phase 3 and 4 occur, but also the duration of each becomes longer, i.e., more bandwidth is wasted. In this section, we investigate how to choose the Q_h and Q_l to prevent the system falling into the worst cycle. First we study the minimum queue length, Q_{min}, and maximum queue length, Q_{max}.

4.1 Minimum and maximum queue length

In this subsection, the queue length at the switches, $Q(t)$, is not limited, no matter what the value of exact buffer size and minus queue length is, i.e., the derived maximum queue might be larger than the buffer size, and the derived minimum queue size might be less than zero.

• *Minimum queue length*
At time t_0, the aggregate cell arrival rate at the switch is below bandwidth, BW, and queue starts to decrease after the propagation delay between the SES and the switch, τ_{sx}. After the time Δt_{Q_l}, the buffer content reaches the low threshold, Q_l, and the ACR at the SES stops decreasing at $t^+_{Q_l}$. Although this period crosses phase 2 and phase 4, the same dynamic behavior exists for phase 2 and phase 4. Thus we have

$$\int_0^{\Delta t_{Q_l}} BW(1 - e^{-\frac{BW}{N_{VC}\cdot RDF}t})dt = \min(Q_{max}, Q_B) - Q_l.$$

Hence the time interval Δt_{Q_l} is got as the root of above equation. The minimum rate, ACR_{min}, is given by

$$ACR_{min} = \frac{BW}{N_{VC}}e^{-\frac{BW}{N_{VC}\cdot RDF}(\Delta t_{Q_l}+\tau)}.$$

Let $ACR_{t_0}(t)$ be the evolution of the ACR from time t_0, which is given by

$$ACR_{t_0}(t) = \frac{BW}{N_{VC}}e^{-\frac{BW}{N_{VC}\cdot RDF}t}.$$

At the time that the ACR becomes minimum, the SES begins to increase its ACR and the system enters phase 1. After the time $\Delta t^-_{Q_{min}}$, the aggregate cell arrival at the switch is over BW. Hence we get

$$\Delta t^-_{Q_{min}} = \frac{BW - N_{VC}\cdot ACR_{min}}{BW\cdot AIR}.$$

Now the minimum queue length, Q_{min}, is given by

$$Q_{min} = \min(Q_{max}, Q_B) - \int_0^{\Delta t_{Q_l}+\tau} (BW - N_{VC}\cdot ACR_{t_0}(t))dt$$

$$-\int_0^{\Delta t^-_{Q_{min}}} (BW - N_{VC}\cdot ACR_1(t))dt.$$

Using $ACR_1(0) = ACR_{min}$, Δt_{Q_l}, $\Delta t^-_{Q_{min}}$, and $ACR_{t_0}(t)$ from the above equations, finally we have

$$Q_{min} = Q_l - \tau\cdot BW$$
$$+N_{VC}\cdot RDF\cdot e^{-\frac{BW}{N_{VC}\cdot RDF}\Delta t_{Q_l}}\cdot(1 - e^{-\frac{BW}{N_{VC}\cdot RDF}\tau})$$
$$-\frac{(BW - N_{VC}\cdot ACR_{min})^2}{2\cdot BW\cdot AIR}. \quad (5)$$

• *Maximum queue length*
Similarly, we can obtain the derivation of maximum queue length. First we get the equation of the interval, Δt_{Q_h}, which is the time that the queue length increases from Q_{min} to Q_h. Since the increasing rate of ACR at each SES is $BW\cdot AIR/N_{VC}$ from equation 1, we get

$$\Delta t_{Q_h} = \sqrt{\frac{2(Q_h - Q_{min})}{BW\cdot AIR}}.$$

The maximum rate, ACR_{max}, is given by

$$ACR_{max} = \frac{BW}{N_{VC}}(1 + AIR\cdot(\tau + \Delta t_{Q_h})).$$

At the time that the ACR becomes maximum, the SES begins to decrease its ACR and the system enters phase 2. After the time $\Delta t^-_{Q_{max}}$, the aggregate cell arrival at the switch is below BW, and the queue length begins to decrease. So we have

$$ACR_{max}e^{-\frac{BW}{N_{VC}RDF}\Delta t^-_{Q_{max}}} = BW/N_{VC},$$

which is solved as

$$\Delta t^-_{Q_{max}} = -\frac{N_{VC}\cdot RDF}{BW}log\frac{BW}{N_{VC}\cdot ACR_{max}}.$$

Now the maximum queue length, Q_{max}, is given by

$$Q_{max} = Q_{min} + \int_0^{\Delta t_{Q_h}+\tau} (BW\cdot AIR\cdot t)dt$$

$$+\int_0^{\Delta t^-_{Q_{max}}} (N_{VC}\cdot ACR_2(t) - BW)dt.$$

Using $ACR_2(0) = ACR_{max}$, Δt_{Q_h}, $\Delta t^-_{Q_{max}}$ and $ACR_2(t)$ from above equation, finally we have

$$Q_{max} = Q_h + \tau\cdot\sqrt{2(Q_h - Q_{min})\cdot BW\cdot AIR}$$
$$+\frac{\tau^2\cdot BW\cdot AIR}{2}$$
$$+N_{VC}\cdot RDF(log(\frac{BW}{N_{VC}\cdot ACR_{max}}) + \frac{N_{VC}\cdot ACR_{max}}{BW} - 1). \quad (6)$$

4.2 Cell loss probability and utilization

In the last subsection the derived maximum queue might be larger than the buffer size, Q_B, and the derived minimum queue size might be less than zero. Now the cell loss probability and utilization of four various cycles, described in the subsection 3.2, are calculated.

- *Best cycle*: $Q_{max} \leq Q_B$ and $Q_{min} \geq 0$

In this cycle, the maximum queue length is less than the buffer size in the switch. Also the minimum queue length is larger than zero. In this perfect case, cell loss probability is zero, $P_{loss} = 0$, and utilization is one, $\rho = 1$.

- *Worst cycle*: $Q_{max} > Q_B$ and $Q_{min} < 0$

The events of cell-loss and under-utilization occur in this worst case. As we know, when the maximum queue length is larger than the buffer size, the system enters phase 4 and some cells are lost to keep the maximum queue length as buffer size. Similarly, when the minimum queue length is less than zero, the system enters phase 3 and some bandwidth is wasted because of the idle time. In order to fit this case, we need to re-derive the above equations. Due to lack of space, the derivation is skipped and the detail is shown in [8].

- *Cell-loss cycle*: $Q_{max} > Q_B$ and $Q_{min} \geq 0$

For the cycle where phase 4 occurs and phase 3 does not occur, utilization is one, $\rho = 1$, and cell loss probability is directly computed from the worst cases.

- *Under-utilization cycle* : $Q_{max} \leq Q_B$ and $Q_{min} > 0$

For the cycle where phase 3 occurs and phase 4 does not occur, cell loss probability is zero, $P_{loss} = 0$, and utilization is directly computed from the worst cases.

4.3 Choice of Q_h and Q_l

That Q_{min} is influenced directly by Q_l is obviously observed from above equations. Meanwhile, Q_{min} is also influenced by ACR_{min}, which is determined by $\min(Q_B, Q_{max})$. Hence we use the function f to denote the complex expressions. Similarly let g be the function to calculate Q_{max}.

$$Q_{min} = f(Q_l, \min(Q_B, Q_{max}))$$

$$Q_{max} = g(Q_h, \max(0, Q_{min}))$$

To keep the system operating in the best cycle, we define "best area" to determine how to set the high and low thresholds. When the point of high and low thresholds is located in the best area, good performance is guaranteed, namely, utilization is one and cell loss probability is zero. Note the best area may not exist if N_{VC} or τ is large.

First we get two thresholds Q_l^* and Q_h^* by solving the equation $0 = f(Q_l^*, Q_B)$ and $Q_B = g(Q_h^*, 0)$. The best area can be obtained according to Q_l^* or Q_h^*. Nonetheless, the same best area is obtained whether according to Q_h^* or Q_l^*. Below we only discuss the solution according to Q_h^* :

Case 1: $Q_h = x > Q_h^*$: In this case, Q_{min} must be larger than zero to avoid cell loss. That is, the expected value Q_{min}° is the solution of the equation $Q_B = g(x, Q_{min}^\circ)$. Also we get Q_l° by solving $Q_{min}^\circ = f(Q_l^\circ, Q_B)$. Therefore we get the best area $(Q_h = x, [Q_l^\circ, x])$.

Case 2: $Q_h = x < Q_h^*$: In this case, the main concern is to avoid the occurrence of under-utilization. We get Q_{max}° by solving the equation $Q_{max}^\circ = g(x, 0)$, and get Q_l° by solving the equation $0 = f(Q_l^\circ, Q_{max}^\circ)$. Therefore we get the best area $(Q_h = x, [Q_l^\circ, x])$.

Looking at figure 4, we partition space into five areas. If (Q_h, Q_l) is located in the area "Best", the system shall operate in the best cycle. If $Q_h > Q_h^*$ and $Q_l < Q_l^*$, that causes $Q_{max} > Q_B$ and $Q_{min} < 0$, the worst cycle occurs. Similarly, areas "Cell-loss" and "Under-utilization" are corresponding to the cell-loss cycle and under-utilization cycle, respectively. Area "Non-exist" does not exist because we can not set $Q_h < Q_l$.

- **Rules**

1: When $Q_h^* \geq Q_l^*$, good performance can be achieved. We set (Q_h, Q_l) to locate in the best area. Also to provide more number of connections and larger propagation delay, we ought to set (Q_h, Q_l) to close the center point of the best area as much as possible.

2: When $Q_h^* < Q_l^*$, good performance can not be achieved. When we feel the effect of cell-loss is more serious than the effect of under-utilization, such as loss-sensitive applications, we set $(Q_h = Q_h^*, Q_l = Q_h^*)$ to avoid cell loss.

3: When $Q_h^* < Q_l^*$ and applications are delay-sensitive, we set $(Q_h = Q_l^*, Q_l = Q_l^*)$ to avoid under-utilization.

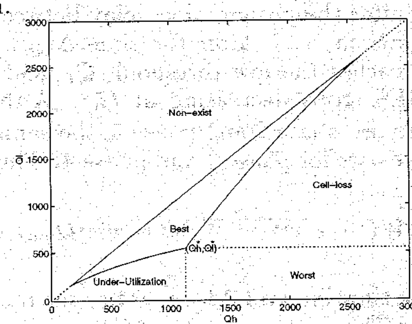

Figure 4: Five areas to determine Q_h and Q_l.

5 Simulation and Discussion

Some examples are presented to show the correctness of our analysis of cell loss probability and utilization. The parameters of the overall systems are set to N_{VC}=10, BW=155Mbps, PCR=155Mbps, ICR=PCR/20, AIR=PCR/2048, RDF=512, τ_{sx}:τ_{xds}=1:3, τ=0.2ms, N_{RM}=32, Q_B=3000, Q_h = 2500, and Q_l=500.

Figure 5 shows the cell loss probability as a function of the high threshold. There is a good agreement between the analytical and simulation results, which are shown by circles. The small gaps are due to the continuous time approach and the synchronous assumption in the analysis. When Q_h is set too high, P_{loss} increases. This is because the detection time of congestion is too late. Also as N_{VC} becomes larger, cell loss probability is larger. When N_{VC} becomes big, the increase speed of total allowed cell rate, $N_{VC} \cdot ACR$, does not change, that is, the speed is still $BW \cdot AIR$. Hence the maximum total ACR ($N_{VC} \cdot ACR_{max}$) does not change. However, the decrease speed of total ACR slows down. Therefore, Δt_{Q_l}, the interval of ACR between ACR_{max} and BW/N_{VC}, is longer. The longer Δt_{Q_l} causes P_{loss} to rise. Also, when τ becomes longer, ACR_{max} and Δt_{Q_l} grow. The growth leads P_{loss} to increase.

Actually the value of Q_h not only directly influences Q_{max}, but also indirectly influences Q_{min}. Hence we also observe the influence of Q_h on the utilization in Fig. 6. In the case with no cell loss, utilization drops quickly when Q_h increases. However, in the case with cell loss, ρ rises slightly when Q_h increases. As we know, the increment of Q_h directly causes Q_{max} to uprise. In the case with no cell loss, the increment of Q_{max} will cause Q_{min} and ρ to drop. But in the case with cell loss, Q_{min} is influenced by Q_B, not by Q_{max}. Hence the wasted bandwidth of under-utilization does not change. However, the increment of Q_h causes Δt_{Q_h} to last longer. Hence the total cycle time is longer and thus, utilization increases.

Note that the case of N_{VC}=20, τ=0.2ms has no cell loss. This is because Q_l^* is less than Q_l, which is 500 in this experiment. Similarly there is no cell loss in the case of N_{VC}=10, τ=0.02ms.

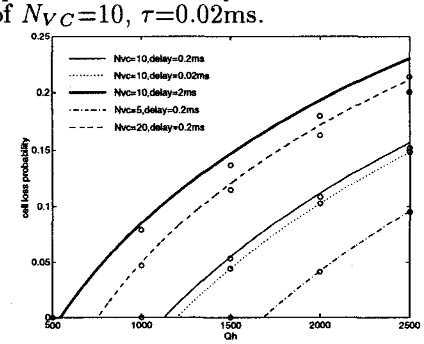

Figure 5: Influence of Q_h on P_{loss} (Q_l=500)

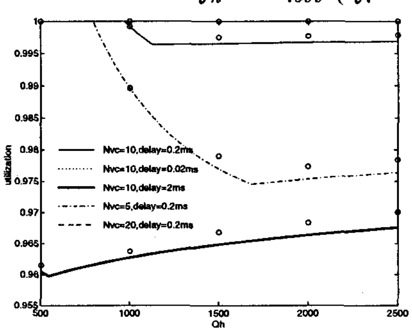

Figure 6: Influence of Q_h on ρ (Q_l=500)

We observe the influence of the number of connections and the round-trip propagation delay, on the best area in Fig. 7. When N_{VC} becomes large, the best area shrinks, and the center point of the area slightly drops. Similarly, when τ becomes large, the best area shrinks. In this case, however, the center point of the area does not drop. The best area of $N_{VC} = 10, \tau = 2ms$ is empty because of $Q_h^* < Q_l^*$.

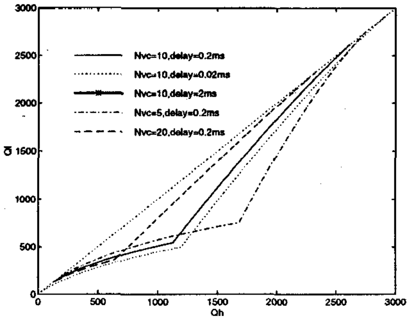

Figure 7: Comparison of best areas among different N_{VC} and τ

6 Conclusion

In this paper, an analysis for rate-based congestion control is provided. The equations of cell loss probability and utilization are derived. When the number of sources increases, the cell loss probability increases and utilization increases. However, when the propagation delay becomes large, the cell loss probability increases and utilization decreases.

The concept of "best area" help us to determine the high and low thresholds. By setting the high and low thresholds in the best area, the system can achieve good performance. When the number of connections becomes larger or the propagation delay becomes longer, the best area shrinks. When the best area shrinks to empty, cell-loss or under-utilization is unavoidable. Nevertheless, rule 2 and 3 are provided to prevent unnecessary cell-loss or under-utilization.

References

[1] ATM Forum, "ATM Forum Traffic Forum Traffic Management Specification Version 4.0", ATM Forum af-tm-0056.000, April 1996.

[2] B. Flavio and W. F. Kerry, *The Rate-Based Flow Control Framework for the Available Bit Rate ATM service.* IEEE Network Magazine, pp. 25-39, March/April 1995.

[3] P. Newman, *Backward Explicit Congestion Notification for ATM Local Area Network-s.* IEEE GLOBECOM'95, vol. 2, pp. 719-723, Houston, TX, Dec. 1993.

[4] N. Yin, *Analysis of a Rate-Based Traffic Management Mechanism for ABR Service.* IEEE GLOBECOM'95, pp. 1076-1082, Singapore, 1995.

[5] G. Ramamurthy and Q. Ren, *Analysis of the Adaptive Rate Control for ABR Service in ATM Networks.* IEEE GLOBECOM'95, pp. 1083-1088, Singapore, 1995.

[6] H. Ohsaki, M. Murata, H. Suzuki, C. Ikeda and H. Miyahara *Rate- Based Congestion Control for ATM Networks.* ACM SIGCOMM Computer Communication Review, pp. 60-72, April 1995.

[7] M. Ritter, *Network Buffer Requirement of the Rate-Based Control Mechanism for ABR Services.* IEEE INFOCOM'96, pp. 1190-1197, San Francisco, California, 1996.

[8] Yuan-Cheng Lai and Ying-Dar Lin, Performance Analysis of Rate-based Flow Control under a Variable Number of Sources. to appear in Computer Networks and ISDN systems.