Multi-priority Scheduling Using Network Calculus: Model and Analysis

Jun Huang*, Zi Xiong*, Yanbing Liu†, Qiang Duan‡, Yunlong He*, Juan Lv*, and Jianyu Wang*
*School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing, China 400065
Email: xiaoniuadmin@gmail.com
†School of Computer Science
Chongqing University of Posts and Telecommunications, Chongqing, China 400065
Email: liuyb@cqupt.edu.cn
‡Info. Science and Technology Department
The Pennsylvania State University, Abington, Pennsylvania 19001
Email: qduan@psu.edu

***Abstract*—Network Calculus (NC) is a powerful means to provide deep insight for flow problems in network performance analysis. Multi-priority scheduling as one of the fundamental models in NC has become an active research topic recently. However, existing works consider neither the arrival interval of the flows nor their arrival orders; thus limiting their applications to only a few delicate scenarios. In this paper, we address these two issues and propose a novel multi-priority model based on the non-preemptive priority scheduling. We derive the theoretical formulation for the service curve under this model, and then obtain the upper bounds of delay and backlog for multi-priority scheduling. We also use two representative case studies to show the correctness and effectiveness of the proposed model. The theoretical analysis is further validated by the numerical experiments. In addition, we discuss the factors that may affect the delay and backlog bounds according to the numerical results.**

I. Introduction

Recent rapid development of network services, such as video, voice, e-health, etc, together with the increasing demand of network communications has brought a great challenge to network modeling and performance analysis. Network Calculus (NC) has been developed as a powerful means of addressing such a challenge. NC employs the envelope method to describe both network arrival traffic and service capacity and adopts the systematic method to analyze network performance. NC provides an efficient tool for calculating the bounds of end-to-end Quality-of-Service (QoS) parameters such as delay and backlog. It not only can study the QoS in a more accurate manner but also can predict and analyze the network performance locally or globally as well. As a result, network calculus has drawn a lot of research interest and become an active area.

As the Internet evolves toward the global multi-service network of the future, a key consideration is to deliver services with guaranteed QoS. However, the current Internet are designed on the basis of "best-effort" model, which may not support QoS offering. To overcome this limitation, multi-priority scheduling has emerged as one of the most efficient ways for service delivery. It differentiates the flows by classifying traffic into several flow types and scheduling them according to their priorities. For example, classifying live video traffic into a high-priority queue can guarantee the QoS of the traffic while achieving efficient network utilization. Therefore, multi-priority scheduling plays a crucial role in enabling QoS and improving network performance, and it has been embraced in NC as a fundamental model.

While a model with two priorities scheduling has been reported in NC, the extension on the multi-priority model with three or more priorities is still a challenge problem that has attracted a great amount of research interest. Qian et al. [1] developed an algorithm to automate the delay bound calculation. They compared and analyzed the Strict Priority Queuing (SPQ) and weighted Round Robin (WRR) scheduling. Sofack et al. [2] addressed the quality of service for a residual flow in a context of aggregation with non-preemptive fixed priority scheduling, in which they mainly considered the strict residual service curve. Zhang et al. [3] explored various scheduling policies in Avionics Full Duplex Switched Ethernet (AFDX) protocol. They calculate the end-to-end virtual link delays using Network Calculus: Packet Generalized Processor Sharing (PGPS) and Static Priority (SP) scheme on scheduler of End System (ES) and FIFO and SP on output port of AFDX switch. Li et al. [4] considered Generic Cell Rate Algorithm (GCRA) and used a hierarchical scheduling policy of Fixed Priority combined with Round Robin (FP-RR) to derive end-to-end delay bounds. Yu et al. [5] provided end-to-end delay bounds of transmitting heterogeneous flows using Deficit Round Robin (DRR) scheduling policy on switch output, and compared the DRR, FIFO, static priority, in terms of transmission delays and fairness. Zhang et al. [6] studied how to guarantee a deterministic transfer delay for cyclic control data in switched industrial Ethernet, in which real-time data is assigned the priority to access network resources. Moreover, leaky bucket regulator was introduced at source nodes to constrain the traffic entering the network. Schmitt et al. [7] made a case for a simple alternative in providing QoS in packet-switched networks based on strict priority queuing. They perceived

978-1-4799-1353-4/13/$31.00 ©2013 IEEE

based on strict priority queuing to regain importance in packet networks due to schemes like DiffServ that provides a less expensive alternative for offering performance guarantees in so-called class of service networks. Also, Schmitt et al. [8] derived basic worst-case properties for the case where strict priority queuing is used as packet scheduling mechanism and proposed simple admission control rules under strict class-based priority queuing which facilitate deterministic delay and bandwidth guarantees for each flow.

Although much progress has been made in this area, the above mentioned researches consider neither the arrival interval of the flows nor their arrival orders; thus limiting their applications to only a few delicate scenarios. In this paper we investigate multi-priority model in NC from a general perspective. The major contributions made are threefold.

- We present, for the first time, a multi-priority scheduling model under the condition that flows arrive in different arrival interval and arrival order. The presented model allows to gain tighter bounds compared to the existing models.
- We use two representative cases to examine the correctness and effectiveness of the proposed model. We then derive the theoretical formulations for the service curve and obtain the upper bounds of delay and backlog in each case.
- We further validate the model based on the case studies. We show that the model is general and flexible, thus can be applied in various networking systems. We also discuss the factors that may affect the delay and backlog bounds.

The rest of the paper is organized as follows. Section II gives an introduction to network calculus. Section III presents a model for multi-priority scheduling system and provides the service curve and delay and backlog bounds formulation for the system. In section IV, we show the numerical results and discuss the factors that may affect delay and backlog bounds, and we conclude the paper in Section V.

II. Preliminaries

In this section, we introduce the definitions and notations that will be used in the rest of the paper.

The theory of network calculus was first developed by Chang [9] and Cruz [10], [11] and extended by others (e.g. [12]–[15]).The theory formulates the multi-priority model based on the following definitions.

Definition 1: (Wide-sense Increase Function). Given a function f defines for $\forall s,t \in (R \cup \{+\infty\})$ we say that f is wide-sense increase function if and only if $s \leq t$: $f(s) \leq f(t)$

Definition 2: (Wide-sense increasing sequences). If $F = \{f(t)|f(t) = 0, \forall t < 0; f(0) \geq 0; f(s) \leq f(t), \forall s \leq t, s, t \in [0, +\infty]\}$, we say that F is wide-sense increasing sequences.

Definition 3: (Min-Plus Convolution). Let f and g be two functions or sequences of F. The min-plus convolution of f and g is the function $(f \otimes g)(t) = \inf_{0 \leq u \leq t}[f(u) + g(t-u)]$, if $t < 0, f \otimes g = 0$.

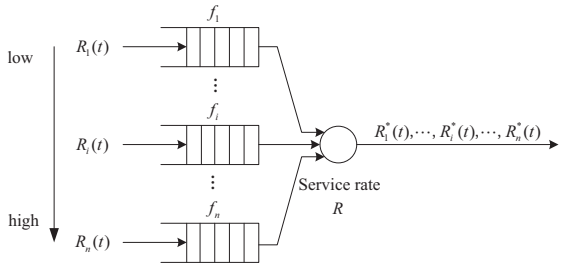


Fig. 1. Multi-priority Modeling

Definition 4: (Min-Plus Deconvolution). Let f and g be two functions or sequences of F. The min-plus convolution of f and g is the function $(f \emptyset g)(t) = \sup_{0 \leq u}[f(t+u) - g(u)]$.

Definition 5: (Arrival Curve). Given a wide-sense increasing function α defined for $t \geq 0$, we say that a flow R is constrained by α if and only if for all $s \leq t$ such that $R(t) - R(s) \leq \alpha(t-s)$. We say that R has α as an arrival curve, or also that R is α-smooth.

Since the arrival curve is often regulated by the leaky buckets, we will use affine arrival curves $\gamma_{r,b}$, defined by: $\gamma_{r,b}(t) = rt + b$ for $t > 0$ and 0 otherwise. Parameters b and r are called the burst tolerance and rate.

Definition 6: (Service Curve). Consider a system S and a flow through S with input and output function R and R^*. We say that S offers to the flow a service curve β if and only if β is wide sense increasing, $\beta(0) = 0$ and $R^* \geq R \otimes \beta$.

Theorem 1: (Performance Bounds). Assume a flow constrained by arrival curve α, traverses a system that offers a service curve of β, the performance bounds can be derived as follows.

Delay Bound. The delay bound $D(t)$ is expressed as

$$D(t) \leq h(\alpha, \beta) \tag{1}$$

$$h(\alpha, \beta) = \sup_{t \geq 0}\{\inf\{d \geq 0, \alpha(t) \leq \beta(t+d)\}\} \tag{2}$$

Backlog Bound. The delay bound $Q(t)$ is expressed as

$$Q(t) = R(t) - R^*(t) \leq \sup_{s \geq 0}\{\alpha(s) - \beta(s)\} \tag{3}$$

III. The Proposed Multi-priority Model

A. The Main Results

We present a multi-priority model as shown in Fig. 1, in which f_i denotes the flow with priority i, $R_i(t)$ and $R_i^*(t)$ denote the input and output accumulation functions, and $\alpha_i(t)$ is the arrival curve, $\beta(t)$ is the service curve.

In this model, we assume that f_i and f_j are the two flows, the arrival interval and arrival order of all the flows are arbitrary, the backlogged period of flow f_i start at S_i such that $R(S_i) = R^*(S_i)$. In addition, suppose $\beta(t)$ follows a rate-latency function $\beta_{R,T}(t) = R[t-0]^+$, and the scheduling for flows is non-preemptive. We use a leaky bucket model to

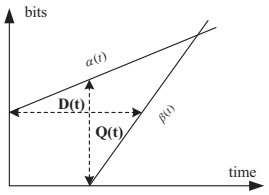


Fig. 2. Arrival rates constraint

constrain the arrival curve, and define $A_i(t) = \sum_{i<j} m\alpha_j(t) = \sum_{i<j} m(r_j t + b_j)$, $m = 0, 1$ as the sum of arrival curve for flow f_j, where f_j has higher priority than f_i. The parameter m can be calculated by the following two cases

- S_j falls in the interval $(S_i, t]$, then $m = 1$.
- S_j is not in the interval $(S_i, t]$, then f_j is independent of f_i, $m = 0$.

Denote $l^j_{\max}$ as the maximum packet size of f_j, and $L_i = \sum_{i>j} kl^j_{\max}, k = 0, 1$ as the sum of maximum packets size of f_j, who has lower priority than f_i.

The parameter k is calculated by the following three cases

If f_j is served before f_i

- S_j falls in the interval $(S_i, t]$, $k = 1$.
- S_j does not in the interval $(S_i, t]$, f_j is independent of f_i, $k = 0$.

If f_j is served after f_i

- $k = 0$.

Therefore, the service curve of f_i can be now derived as

$$\begin{aligned}\beta_{R_i,T_i}(t) &= [\beta_{R,T}(t) - L_i - A_i(t)]^+ \\ &= [Rt - \sum_{i>j} kl^j_{\max}] - (\sum_{i<j} mr_j \times t + \sum_{i<j} mb_j)]^+ \\ &= (R - \sum_{i<j} mr_j)[t - \frac{\sum_{i>j} kl^j_{\max} + \sum_{i<j} mb_j}{R - \sum_{i<j} mr_j}]^+\end{aligned} \tag{4}$$

where $k = 0, 1$ and $m = 0, 1$, the service rate is $R_i = R - \sum_{i<j} mr_j$, and the latency is $T_i = \frac{\sum_{i>j} kl^j_{\max} + \sum_{i<j} mb_j}{R - \sum_{i<j} mr_j}$.

Now we consider the constraint of arrival rates for n flows as shown in Fig. 2. We assume that $r_1 + \cdots r_i + \cdots + r_n \leq R$, according to Theorem 1, the delay upper bound, over the interval $(S_i, t]$, can be obtained by

$$D_i \leq \frac{b_i}{R - \sum_{i<j} r_j} + \frac{\sum_{i>j} kl^j_{\max} + \sum_{i<j} mb_j}{R_i} \tag{5}$$

where $k = 0, 1$ and $m = 0, 1$.

Similarly, the backlog upper bound can be derived by

$$Q_i \leq b_i + r_i \cdot \frac{\sum_{i>j} kl^j_{\max} + \sum_{i<j} mb_j}{R_i} \tag{6}$$

where $k = 0, 1$ and $m = 0, 1$.

These are the main results of our proposed model. Eq. (5) and (6) indicate that our proposed model produces tighter bounds in terms of both delay and backlog compared with previous results [1]–[9], [16]. Here we would like to uses two case studies, namely, the flows arrive in the short interval and the flows arrive in the random interval, to further make the model and its relevant analysis clear.

B. Case Study

Consider the scheduler that serves three flows f_L, f_M and f_H where f_L, f_M and f_H denotes flow with low priority, medium priority, as well as high priority respectively, and they are arrived in a short interval. Suppose $R_L(t)$, $R_M(t)$, $R_H(t)$ and $R^*_L(t)$, $R^*_M(t)$, $R^*_H(t)$ are the inputs and outputs accumulation functions for the three flows, $\alpha(t)$, $\beta(t)$ is the arrival curve and service curve respectively, the backlogged period of flow f_i starts at S_i such that $R(S_i) = R^*(S_i)$.

Fig. 3 gives an the scheduling example for the first case to show the flows arrive in a short period, where f_H is assumed to be served first, then the f_M and last the f_L.

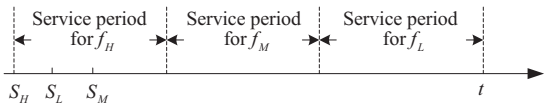


Fig. 3. Scheduling for Case 1

Let us first consider the low priority flow f_L. S_L falls in the period of f_H, i.e. the case of above example. Assume $\alpha_M = \gamma_{r_M, b_M}$ and $\alpha_H = \gamma_{r_H, b_H}$ be the arrival curves, namely, the f_M and f_H are constrained by the leaky buckets, to any time t and S_L, the system output during the interval $(S_L, t]$ is $R(t - S_L)$. Thus

$$\begin{aligned}R^*_L(t) - R^*_L(S_L) = R(t - S_L) - \\ [R^*_M(t) - R^*_M(S_L)] - [R^*_H(t) - R^*_H(S_L)]\end{aligned} \tag{7}$$

At time S_L, the backlog for f_L is empty, so $R^*_L(S_L) = R_L(S_L)$, therefore

$$\begin{aligned}R^*_L(t) = R_L(S_L) + R(t - S_L) - \\ [R^*_M(t) - R^*_M(S_L)] - [R^*_H(t) - R^*_H(S_L)].\end{aligned} \tag{8}$$

This suggests that

$$R^*_M(t) - R^*_M(S_L) \leq R_M(t) - R_M(S_L) \leq \alpha_M(t - S_L) \tag{9}$$

$$R^*_H(t) - R^*_H(S_L) \leq R_H(t) - R_H(S_L) \leq \alpha_H(t - S_L) \tag{10}$$

and

$$R^*_M(t) - R^*_M(S_L) \geq 0, R^*_M(t) - R^*_M(S_L) \geq 0 \tag{11}$$

Integrating equations (8), (9), (10), (11), we can obtain

$$\begin{aligned}R^*_L(t) \geq R_L(S_L) + [R(t - S_L) - \alpha_M(t - S_L) \\ -\alpha_H(t - S_L)]^+ = R_L(L) + S(t - S_L)\end{aligned} \tag{12}$$

where $S(u) = [Ru - \alpha_M(u) - \alpha_H(u)]^+$.

If $S(u)$ is wide-sense increasing, the service curve of f_L equals to the function $S(u)$, i.e. $\beta_{R_L,T_L}(t) = R_L(t - T_L)^+$, where $(x)^+ = \max(x, 0)$. This is the rate-latency service curve with service rate $R_L = R - r_M - r_H$ and latency $T_L = \frac{b_M + b_H}{R - r_M - r_H}$.

Let us proceed to examine the flow who has medium priority. Since the flows arrive in the short period, f_M will wait for the transmission of f_H until the queue is empty. To the arbitrary time t, the output of system during period $(S_M, t]$ is $R(t - S_M)$. Thus

$$R_M^*(t) - R_M^*(S_M) \geq R(t - S_M) - [R_H^*(t) - R_H^*(S_M)] \tag{13}$$

Likewise, we can rewrite above formula as

$$\begin{aligned} R_M^*(t) &\geq R_M(S_M) + [R(t - S_M) - \alpha_H(t - S_M)]^+ \\ &= R_M(S_M) + S(t - S_M) \end{aligned} \tag{14}$$

where $S(u) = [Ru - \alpha_H(u)]^+$. If $S(u)$ is wide-increasing, the service curve f_M equals to the function $S(u)$, the rate is $R_M = R - r_H$ and the latency is $T_M = \frac{b_H}{R - r_H}$.

Now we ultimately consider the high priority flow. Suppose f_H arrives at the system first. The flow of f_H would be accepted immediately as long as it arrives, until the queue for f_H becomes empty.

$$R_H^*(t) - R_H^*(S_H) = R(t - S_H) \tag{15}$$

Therefore we can obtain

$$\begin{aligned} R_H^*(t) &= R_H(S_H) + R(t - S_H) \\ &= R_H(S_H) + S(t - S_H) \end{aligned} \tag{16}$$

That is, the service curve of f_H is $\beta_{R_H,T_H}(t) = S(u) = R(t - 0)^+$, in which the service rate is $R_H = R$, and the latency is $T_H = 0$.

In the second case, we discuss the flows with random arrival interval. We consider the same service sequence, i.e., f_H, $f_M$and f_L. Fig. 4. shows such case where service curves of f_M and f_H do not change. So we only consider service curve of f_L.

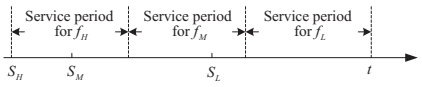


Fig. 4. Scheduling for Case 2

As illustrated in Fig. 4, flow f_L is served after both flows f_H and f_M. S_L are located in the serving period of f_M, that is to say, f_L is independent of f_H, then we have

$$R_L^*(t) - R_L^*(S_L) \geq R(t - S_L) - [R_M^*(t) - R_M^*(S_L)] \tag{17}$$

Hence, the service curve of f_L is $\beta_L(t) = (R - r_M)[t - \frac{b_M}{R - r_M}]^+$.

Now, we can obtain the service curve for f_L who is a rate-latency service curve, that is,

$$\beta_L(t) = (R - r_M - kr_H)[t - \frac{b_M + kb_H}{R - r_M - kr_H}]^+, k = 0, 1 \tag{18}$$

where if S_L does belong to a serviced period of $f_H, k = 1$, otherwise $k = 0$.

From the above discussion, we can see that both the arrival order and the arrival interval have a great impact on the delay and backlog bounds. Therefore, our model is more general that can be applied in various networking environments.

IV. Numerical Results

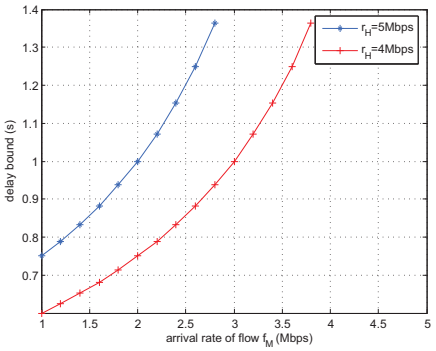
In this section, we conduct numerical experiments to examine the correctness and effectiveness for the proposed models based on the previous case studies. We also explore factors that affect the delay and backlog bounds. We use the analytical equations given by Section III.

In the numerical experiments, we consider a node that serves three flows. The burst parameter of leaky bucket is all set to be 1Mb. And the maximum packet is equal to $L_{\max} = 1500$ Bytes.

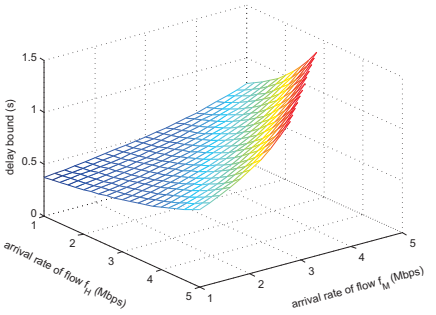
In the following, we derive the delay and backlog bounds for non-preemptive priority model that are given in Eq. (5) and Eq. (6). We discuss the relations between the delay, backlog, arrival rate and service rate of flows. The flows are compared which have short arrival interval with those which have random arrival. We assume Fig. 3 and Fig. 4 is illustrated on the condition that the flow of f_L is discussed alone.

Fig. 5 shows the delay bounds for case illustrated in Fig. 3, in which the arrival interval is short. In Fig. 5(a) and Fig. 5(b), we assume that the server rate is equal to $R = 10$ Mbps; therefore the service curve of the server is $\beta(t) = 10[t - 0]^+$. The relation between delay bounds of the flow f_H and arrival rates of the flow f_M is given in Fig. 5(a), in which the two curves respectively show f_H delay bounds when the flow has 4 Mbps and 5 Mbps arrival rate. This figure shows that f_H delay bounds increase with the arrival rate of f_M. For a given f_M arrival rate, higher r_H causes longer delay bounds for this flow. Fig. 5(b) shows the changing of delay bounds of the flow f_H with the arrival rates of flows f_H and f_M. From this figure we can see that the delay bounds increase with the arrival rate of each flow and the increasing speed rises significantly when the total arrival rate of the two flows approaches the server capacity (i.e., when $r_M + r_H$ is close to R). In Fig. 5(c) the two curves give the relation of delay bounds for the flow f_H, with arrival rate 4 Mbps and 5 Mbps respectively, and the server rate, which changes from 12 Mbps to 17 Mbps. The curves show that the delay bound drops when service rate increases and such delay decreasing is more obvious when the flow in question has a higher arrival rate. Therefore from Fig. 5 we can conclude that the delay bound of high priority flow is affected by the its own arrival rate, the arrival rate of the flow with mediate priority, and the service rate.

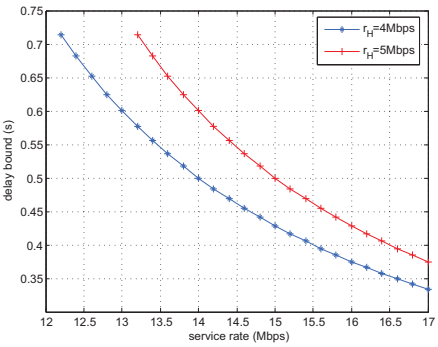
We consider delay bound when the flow arrival interval is random. The results are shown in Fig. 6. In Fig. 3, S_L belongs



(a) $R = 10, r_M \in [1, 5], r_L = 2, r_H = 4, 5$

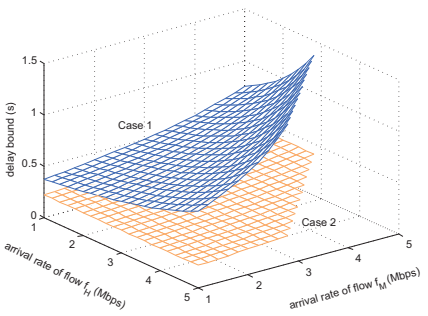


(b) $R = 10, r_M \in [1, 5], r_H \in [1, 5], r_L = 2$

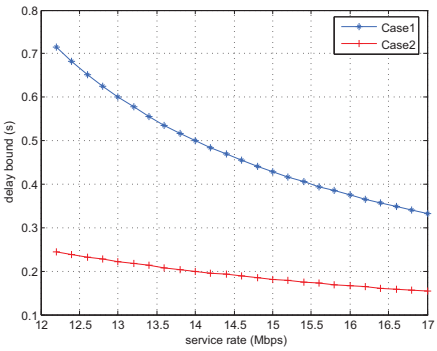


(c) $R \in [12, 17], r_M = r_L = 4, r_H = 4, 5$

Fig. 5. Delay bound vs. different parameters when the flow arrival interval is short.



(a) $R = 10, r_M \in [1, 5], r_H \in [1, 5], r_L = 2$



(b) $R \in [12, 17], r_L = r_M = r_H = 4$

Fig. 6. Delay bound vs. different parameters when the flow arrival interval is random.

to a service period of f_H, whereas, in Fig. 4, S_L does not belong to a serviced period of f_H. Fig. 6(a) and Fig. 6(b) indicate that the delay bound in case of the condition of Fig. 4 (Case 2 in the Fig. 6) is obviously smaller than that of Fig. 3 (Case 1 in the Fig. 6). And when the arrival rate is close to service rate, the gap of the delay under those two conditions will emerge.

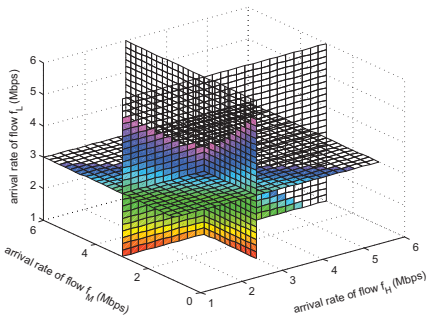
Fig. 7 shows the backlog bound on the condition illustrated in Fig. 3 in which the arrival interval is short. We find that both the arrival rate and service rate have significant effect on the backlog bound. The backlog bound is a increasing function of the arrival rate and a decreasing function of the service rate. We also observed that the arrival rate has stronger influence on backlog bound when it is close to the service rate.

Fig. 8 shows the backlog bounds with different arrival time (Fig. 3 and Fig. 4). The backlog bound illustrated in Fig. 4 is obviously smaller than that in Fig. 3. As well as the delay bound, when the arrival rate is close to service rate, the gap of the backlog under those two conditions will occur.

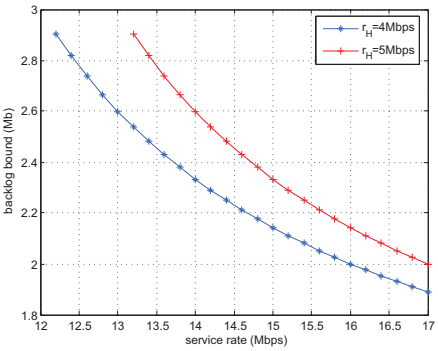
In summary, different arrival intervals could influence the delay and backlog of the system. While the arrival rate is close to service rate, the influence will be amplified. In other word, our model can faithfully reflect the practical network conditions.

V. Conclusion

Recent development in network service and communication has brought a great challenge to network modeling and performance analysis. Network Calculus (NC) has been developed as a powerful means of addressing such a challenge. In this paper we proposed a novel multi-priority scheduling model under the framework of Network Calculus (NC). By taking the arrival interval and arrival order of the flows into account, the presented model is allowed to obtain tighter performance bounds. We used two representative case studies to examine the correctness and effectiveness of the proposed model, then derived the theoretical formulations for the service curve and obtained the upper bounds of delay and backlog in each case. We further validated the model based on the case studies through numerical experiments, our evaluation results showed that our proposed model is general and faithfully reflect the practical network conditions, thus can be applied in various networking systems.

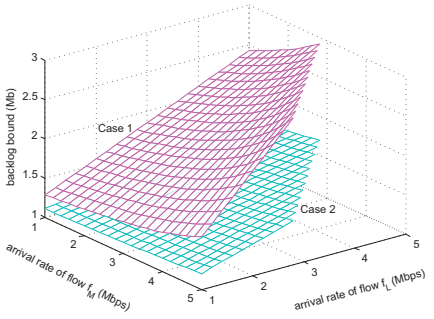


(a) $R = 10, r_L, r_M, r_H \in [1, 6]$

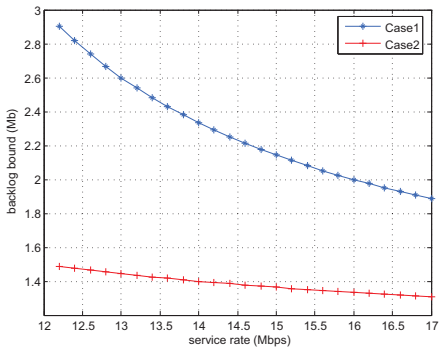


(b) $R \in [12, 17], r_L = r_M = 4, r_H = 4, 5$

Fig. 7. Backlog bound vs. different parameters when the flow arrival interval is short.



(a) $R = 10, r_L, r_M \in [1, 5], r_H = 2$



(b) $R \in [12, 17], r_L = r_M = r_H = 4$

Fig. 8. Backlog bound vs. different parameters when the flow arrival interval is random.

As for the future work, we would like to analyze the performance for three or more flows with different priorities, and explore the multi-priority model in the condition of preemptive scheduling.

Acknowledgment

This work is supported by NCET, NSFC (Grant No. 61272400), Chongqing Innovative Team Fund for College Development Project (Grant No. KJTD201310), Foundation of CEC (Grant No. KJ130523), and CQUPT Research Fund for Young Scholars (Grant No. A2012-79).

References

[1] Y. Qian, Z. Lu, and Q. Dou, "Qos scheduling for nocs: Strict priority queueing versus weighted round robin," in *2010 IEEE International Conference on Computer Design (ICCD)*, 2010, pp. 52–59.

[2] W. Sofack and M. Boyer, "Non preemptive static priority with network calculus," in *2011 IEEE 16th Conference on Emerging Technologies Factory Automation (ETFA)*, 2011, pp. 1–8.

[3] X. Zhang, X. Chen, L. Zhang, G. Xin, and T. Xul, "End-to-end delay analysis of avionics full duplex switched ethernet with different flow scheduling scheme," in *2011 International Conference on Computer Science and Network Technology (ICCSNT)*, vol. 4, 2011, pp. 2252–2258.

[4] J. Li, H. Guan, J. Yao, G. Zhu, and X. Liu, "Performance enhancement and optimized analysis of the worst case end-to-end delay for afdx networks," in *2012 IEEE International Conference on Green Computing and Communications (GreenCom)*, 2012, pp. 301–310.

[5] Y. Hua and X. Liu, "Scheduling design and analysis for end-to-end heterogeneous flows in an avionics network," in *Proceedings IEEE INFOCOM 2011*, 2011, pp. 2417–2425.

[6] Q. Zhang, Y. Cai, D. Gu, and W. Zhang, "Determine the maximum closed-loop control delay in switched industrial ethernet using network calculus," in *American Control Conference, 2006*, 2006, pp. 4870–4875.

[7] J. Schmitt and F. Zdarsky, "A case for simplicity in providing network quality of service: class-based strict priority queueing," in *Proceedings 12th IEEE International Conference on Networks, 2004. (ICON 2004).*, 2004, pp. 809–813.

[8] J. Schmitt, P. Hurley, M. Hollick, and R. Steinmetz, "Per-flow guarantees under class-based priority queueing," in *IEEE Global Telecommunications Conference, 2003. GLOBECOM '03.*, 2003, pp. 4169–4174.

[9] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913–931, 1994.

[10] R. Cruz, "A calculus for network delay. i. network elements in isolation," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–131, 1991.

[11] ——, "A calculus for network delay, part ii: Network analysis," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 132–144, 1991.

[12] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan, "Performance bonds for flow control protocols," *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, pp. 310–323, 1999.

[13] Y. Jiang and Y. Liu, *Stochastic network calculus*. Springerverlag London, 2008.

[14] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. Springer, 2001, vol. 2050.

[15] J.-Y. Le Boudec, "Application of network calculus to guaranteed service networks," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1087–1096, 1998.

[16] Z. Wu, T. Lv, X. Wang, and N. Huang, "The buffer size assignment of afdx based on network calculus," in *2011 9th International Conference on Reliability, Maintainability and Safety (ICRMS)*, 2011, pp. 1319–1323.