Forwarding Performance Model in ForCES System under FBM-based Traffic Arrivals

Chuanhuang Li[1,2] Sanyuan Zhang[1] Weiming Wang[2]

[1] College of Computer Science and Technology, Zhejiang University
Hangzhou, Zhejiang 310058, China
syzhang@zju.edu.cn

[2] College of Information and Electronic Engineering, Zhejiang Gongshang University
Hangzhou, Zhejiang 310018, China
{Chuanhuang_li,wmwang}@zjsu.edu.cn

Received 1 February 2013; Revised 15 April 2013; Accepted 16 April 2013

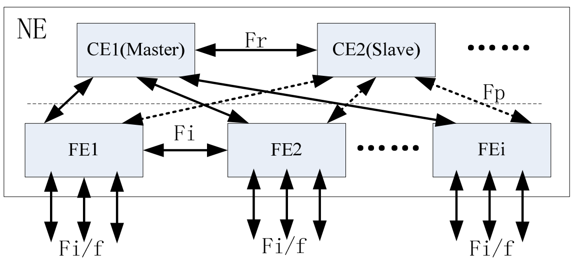
Abstract. In the NE (Network Element) of ForCES (Forwarding and Control Element Separation) architecture, there may be hundreds of FEs (Forwarding Element). These FEs' performance directly reflects the NE's ability of processing data and providing services. Their topologies exist in various forms. In this paper, the common features of these topologies are studied, and a performance solving method that abstracting FE as GPS scheduler is proposed. Under the self-similar FBM (Fractional Brownian Motion) traffic, the performance bounds of multiple flows processed by a single FE are derived. In order to guarantee the performance of the through traffic, based on the performance model, the requirements of processing weight in each FE are analyzed

Keywords: ForCES performance model, Forwarding performance, GPS abstraction, Stochastic network calculus

1 Introduction

Forwarding and Control Element Separation (ForCES) Working Group (WG) in IETF Routing Area, is one of the most influential research organizations in open programmable network research area. The WG specializes in the architecture and protocol standards of open programmable IP Network Element (NE, Such as, router, firewall, or load balancer, etc.). As shown in Fig.1, the IP NE can be constituted by more than one CE and FE which are connected by ForCES protocol [1]. CE can do dynamic configuration and management to the resources of the FEs via the ForCES messages. The resources within the FE are expressed as different Logical Functional Blocks (LFB) [2].

Flows those entered into the NE will be processed by one or more FEs. These FEs may be heterogeneous, or isomorphism. That is, they can be constituted with the same or different LFBs, and they also can be implemented in different platforms, such as network processor, ASIC etc. FEs with the same or different functions are interconnected together through some kind of physical network. To form logical topology and implement various network services, their logical connections can be dynamically changed by CE



Fi: FE-FE interface   Fr: CE-CE interface
Fi/f: FE external interface

Fig. 1. Architecture of ForCES NE

There are many FE topology types. From the topology shape, there are ring, chain and full mesh topologies. From the processing direction of flow, there are feed-forward and non feed-forward topologies. The performance of different data channels of different topologies directly reflects the capability that the NE can provide.

Many researchers have paid attention to the forwarding behavior and forwarding performance of the network nodes, such as, Xiong [3] studied a scalable fast forwarding approach for IP networks to optimize the cost and performance of routers, and he also introduced two LIC-based fast forwarding schemes for explicit routing with scalability [4]; Chang [5] studied the issues of geographic forwarding in MANET; Niu [6] introduced a P2P Query algorithm based on betweenness centrality forwarding in opportunistic networks. The linear topology is studied widely in current research area of network performance. If the envelope of cross flow is known, the service curve of through flow will be directly determined by the system scheduling policy: such as, Mao [7] introduced how to solve the service curve under multiplexing, Kumar [8] studied the service curve of FIFO multiplexing, Blake [9] and Liebeherr [10] introduced the service curve under EDF scheduling policy. Fidler [11][12] introduced a method to solve the performance model of feed-forward topology with no loops. In his studies, the nodes were marked according to the processing sequence in the system. For each node, first, they calculated the envelope of cross traffic, got the leftover service curve of through traffic, and then derived the departure envelope process of through traffic. When carrying out the above process to the through flow in the first node, its departure envelope can be obtained, and then the arrival process is known in the node with higher label. That is because the arrival process of the through flow in one node is exactly the departure process in its previous node. By the stepwise method, the end-to-end performance of linear system can be obtained. Lenzini [13] and Kim [14] introduced a method to solve the end-to-end service curve in tree topology. Lenzini [15] also introduced a method named LUDB (Least Upper Delay Bound) based on network calculus to solve the end-to-end delay. The transmission of the flow in the network was generalized and extended processed, and the end-to-end delay could be derived without considering the specific internal path of flow in the network topology. At the same time, the author divided the topology into nested tandem network and non-nested tandem network, and transformed the nested network into tree topology, then used the method introduced by Lenzini [13] and Kim [14] to solve the performance.

The main contribution of this paper is that a performance solving method that abstracting FE as GPS scheduler is proposed. the performance bounds of multiple flows processed by a single FE are derived under the self-similar FBM (Fractional Brownian Motion) traffic. In order to guarantee the performance of the through traffic, based on the performance model, the processing weight requirements in each FE are analyzed.
The paper is structured as follows. Section 2 studies the FE topology and GPS abstraction. Section 3 proceeds with the derivation of statistical performance bounds on backlog and delay of the FE. Section 4 analyzes the processing weight requirements to guarantee the performance of through flow. Simulation and validation are shown in section 5. Conclusions and discussing our future work are presented in section 6.

2 GPS scheduling abstraction for FE

When there are multiple FEs in ForCES NE, flows arrived from one FE can be processed and left from another FE. These flows are transmitted on the Fi reference point in ForCES architecture (see Fig. 1). FEs can learn the current FE topology by this reference point. There are various forms of FE topology, as shown in Fig. 2.

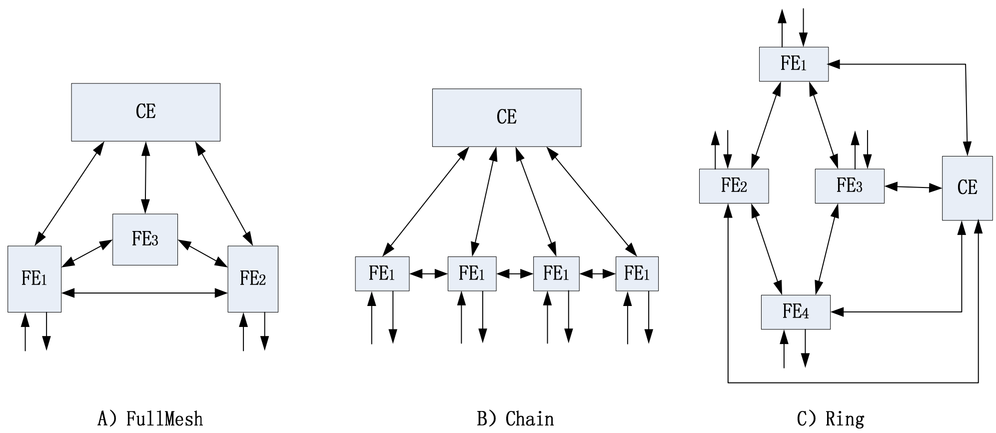


Fig. 2. The forms of FE topology (1)

No matter what form the FE topology is, based on the processing direction of flow between FEs, all FE topology can be divided into two categories: the feed-forward and non feed-forward, as shown in Fig. 3. Loops are easily occurred in non feed-forward topology, and a method by using spanning tree protocol is usually used to avoid these loops. So when analyzing the performance of a specific flow, we can convert all topologies into the feed-forward type.

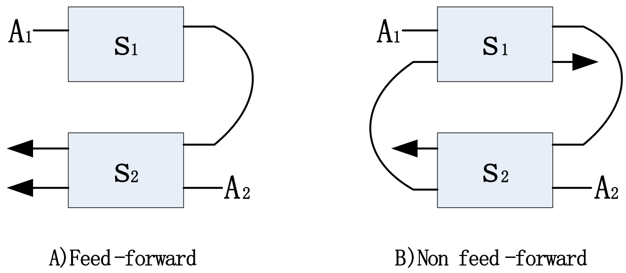


Fig. 3. The forms of FE topology (2)

In ForCES system, there are FEs with various functions, such as, media input processing, Qos processing, forwarding routing, media output processing, load balancing and so on. From the abstract perspective, each FE can be considered to be a single service system with certain rate. And in this system, different types of flow can obtain different levels service guarantee in each FE. So, the flows received process by every FEs, can be treated as they undergoing a GPS scheduling process, as shown in Fig. 4.

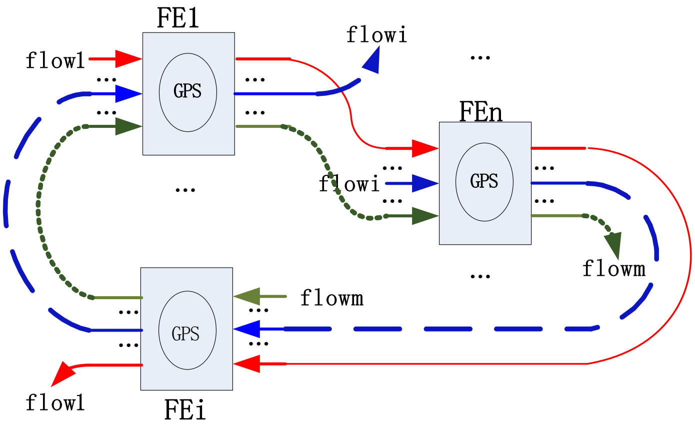


Fig. 4. The flows processed in FE topology with no loops

3 Performance model of GPS scheduling with FBM-based traffic

Consider a GPS scheduling system with service rate r and serving N flows. According the definition of GPS, we assign a positive parameter ϕ_i $(1 \leq i \leq N)$, called weight, to each flow. Let $S_i(\tau,t)$ denote the amount of traffic served in the time interval $[\tau,t]$ for flow i, then

$$\frac{S_i(\tau,t)}{S_j(\tau,t)} \geq \frac{\phi_i}{\phi_j} \quad ,j = 1,2,\cdots,N. \tag{1}$$

At the same time, we call $g_i = \frac{\phi_i}{\sum_{j=1}^{N}\phi_j} r$ is the guaranteed service rate for flow i in unit time.

A concept of "feasible ordering" for GPS was given in [16][17][18]: for a given set of input flows in a GPS system whose long-term average rate is ρ_i, $\sum_{i=1}^{N}\rho_i < r$, an ordering is called "feasible ordering" among the flows with respect to $\{\rho_i\}_{1\leq i\leq N}$ and $\{\phi_i\}_{1\leq i\leq N}$, if

$$\rho_i < \frac{\phi_i}{\sum_{j=i}^{N}\phi_j}\left(1-\sum_{j=1}^{i-1}\rho_j\right), 1 \leq i \leq N. \tag{2}$$

Generally, there are not only one feasible orderings for the flows in a given GPS system with $\{\rho_i\}_{1\leq i\leq N}$ and $\{\phi_i\}_{1\leq i\leq N}$.

According [23], the statistical sample path envelope of FBM traffic is

$$G'(t;\sigma) = (\rho+\theta)t+\sigma\,, \qquad \varepsilon'(\sigma) = Le^{-(\sigma/c)^{\beta}} \tag{3}$$

where $\beta = 2(1-H), c = \left(\frac{2}{\theta H}\right)^{\frac{1}{1-H}}, L = e\cdot max\left\{1, 4^H \frac{\eta/\theta+2-H}{2H(1-H)}\right\}$, θ is any positive number.

We first give two auxiliary theorems, before solving the GPS scheduling performance with FBM-based arrival traffic. Theorem 1 was appeared in [19] and [20]. We will proof theorem 2.

Theorem 1: For any positive numbers a_k and b_k, $k = 1,2,\cdots,K$, and any $x \geq 0$, we have

$$\inf_{x_1+\cdots+x_K=x} \sum_{k=1}^{K} a_k e^{-b_k x_k} = e^{\frac{-x}{w}} \prod_{k=1}^{K} (a_k b_k w)^{\frac{1}{b_k w}} \quad (4)$$

where $\mathrm{w} = \sum_{\mathrm{k}=1}^{\mathrm{K}} \frac{1}{\mathrm{b_k}}$.

Theorem 2: For any positive numbers a_k,b_k $(k = 1,2,\cdots,K)$, $0 < \beta < 1$, and any $x \geq 0$, we have:

$$\inf_{x_1+\cdots+x_K=x} \sum_{k=1}^{K} a_k e^{-b_k {x_k}^{\beta}} = e^{\frac{-K^{1-\beta}x^{\beta}}{w}} \prod_{k=1}^{K} (a_k b_k w)^{\frac{1}{b_k w}} \quad (5)$$

where $w = \sum_{k=1}^{K} \frac{1}{b_k}$.

Proof: Let ${x_k}^{\beta} = y_k$. Because $x_1 + \cdots + x_K = x$, according to Lagrange multiple method, we can get the maximum value of ${x_1}^{\beta} + \cdots + {x_k}^{\beta}$ is $K^{1-\beta}x^{\beta}$(when $x_1 = \cdots = x_k = \frac{1}{K}x$).

As a result, $y_1 + \cdots + y_K \leq K^{1-\beta}x^{\beta}$, then

$$\inf_{x_1+\cdots+x_K=x} \sum_{k=1}^{K} a_k e^{-b_k {x_k}^{\beta}} = \inf_{y_1+\cdots+y_K \leq K^{1-\beta}x} \sum_{k=1}^{K} a_k e^{-b_k y_k}$$

According Theorem 1 and with the fact that $e^{\frac{-x}{w}}$ is monotone decreasing, we have

$$\inf_{y_1+\cdots+y_K \leq K^{1-\beta}x} \sum_{k=1}^{K} a_k e^{-b_k y_k} = e^{\frac{-K^{1-\beta}x^{\beta}}{w}} \prod_{k=1}^{K} (a_k b_k w)^{\frac{1}{b_k w}}$$

The theorem is proofed.

We use horizontal decomposition method mentioned in [21] and [22] to solve the performance of GPS scheduling system. The GPS system is decomposed into several SSQs (Single-server queue), as shown in Fig. 5. Without loss of generality, we also think $1,2,\cdots,N$ is exactly a feasible ordering. $\delta_i(t)$ is the backlog of the ith SSQ.

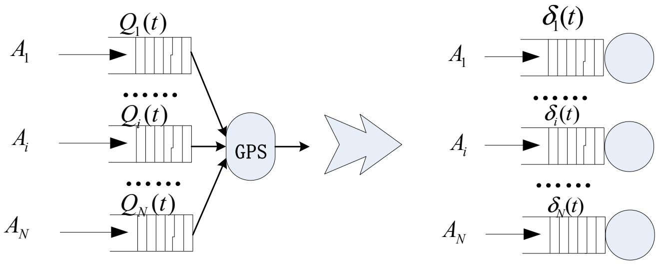


Fig. 5. Decomposing the GPS into **N** fictitious SSQs

From [16], we can know there is the relationship between the backlogs of un-decomposed and decomposed GPS system.

$$Q_{GPSi}(t) \leq \delta_i(t) + \varphi_i \sum_{j=1}^{i-1} \delta_i(t). \quad (6)$$

$$D_{GPSi}(t) \leq \frac{1}{g_i}\left(\delta_i(t) + \varphi_i \sum_{j=1}^{i-1} \delta_i(t) \right) \quad (7)$$

Where $\varphi_i = \frac{\phi_i}{\sum_{j=i}^{N}\phi_j}$, $g_i = \frac{\phi_i}{\sum_{j=1}^{N}\phi_j}C$.

From (6), we can get

$$\begin{aligned} Q_{GPSi}(t) &\leq \delta_i(t) + \varphi_i \sum_{j=1}^{i-1} \delta_i(t) \\ &= \delta_i(t) + \varphi_i\delta_1(t) + \cdots + \varphi_i\delta_{i-1}(t) \\ &= \delta_i(t) + \delta_1^{\varphi_i}(t) + \cdots + \delta_{i-1}^{\varphi_i}(t) \end{aligned} \quad (8)$$

where $\delta_j^{\varphi_i}(t) = \varphi_i\delta_j(t), j < i$.

We have the following theorem about the performance bound in GPS scheduling system with FBM-based traffic:

Theorem 3: Given a GPS scheduling work conserving system which serves N flows, the service rate is C , the weight of the ith flow is ϕ_i $(1 \leq i \leq N)$. $A_i(t)$ （$i = 1,2,\ldots N$）is the arrival process of each flow. The arrival processes are stochastically independent, and they are the FBM process with Hurst parameter H. Their average arrival rate is ρ_i, where $\sum_{i=1}^{N}\rho_i < C$. Assume that $1 \to 2 \to \cdots \to N$ is a feasible ordering with respect to $\{\rho_i + \theta_i\}_{1\leq i\leq N}$ and $\{\phi_i\}_{1\leq i\leq N}$. For each flow, the average backlog and delay has the following bounds:

$$Pr\{Q_i(t) \geq x\} < e^{\frac{-i^{1-\beta}x^{\beta}}{w}} \prod_{k=1}^{i} (L_k b_k w)^{\frac{1}{b_k w}} \quad (9)$$

$$Pr\{D_i(t) \geq d\} < e^{\frac{-i^{1-\beta}(g_i d)^{\beta}}{w}} \prod_{k=1}^{i} (L_k b_k w)^{\frac{1}{b_k w}} \quad (10)$$

where $b_k = \begin{cases} \frac{1}{\varphi_i{}^{\beta} c_k{}^{\beta}}, & k = 1,2, \dots (i-1) \\ \frac{1}{c_k{}^{\beta}}, & k = i \end{cases}$, $w = \sum_{k=1}^{i} \frac{1}{b_k} = c_i{}^{\beta} + \varphi_i{}^{\beta} \sum_{k=1}^{i-1} c_k{}^{\beta}$, $\beta = 2(1-H), c_i = \left(\frac{2}{\theta_i{}^{H}}\right)^{\frac{1}{1-H}}$,

$L_k = e \cdot max\left\{1, 4^{H} \frac{\rho_k/\theta_k + 2 - H}{2H(1-H)}\right\}$, $g_i = \frac{\phi_i}{\sum_{j=1}^{N} \phi_j} C$ 。

Proof: We first analyze the SSQ. The error function of probability is zero in a service system with constant service rate, so, for the ith SSQ，according Corollary 5.3 in [19], we have

$$Pr\{\delta_i(t) \geq x\} = \varepsilon_i(x) = L_i e^{-\left(x/c_i\right)^{\beta_i}} \quad (11)$$

Then in (8):

$$Pr\left\{\delta_j^{\varphi_i}(t) \geq x\right\} \leq Pr\left\{\delta_j(t) \geq \frac{x}{\varphi_i}\right\} \leq L_j e^{-\left(x/\varphi_i c_j\right)^{\beta_j}} \quad (12)$$

For the convenience of calculation，we think every traffic has the same H, then

$$Pr\left\{\delta_j^{\varphi_i}(t) \geq x\right\} \leq Pr\left\{\delta_j(t) \geq \frac{x}{\varphi_i}\right\} \leq L_j e^{-\left(x/\varphi_i c_j\right)^{\beta}} \quad (13)$$

Let $\varepsilon_j^{\varphi_i}(x) = L_j e^{-\left(x/\varphi_i c_j\right)^{\beta}} = L_j e^{-\left(1/\varphi_i c_j\right)^{\beta} x^{\beta}}$, then according to (6) and (8), we have

$$\begin{aligned} Pr\{Q_i(t) \geq x\} &\leq Pr\{\delta_i(t) + \delta_1^{\varphi_i}(t) + \dots + \delta_{i-1}^{\varphi_i}(t) \geq x\} \\ &< \varepsilon_i \otimes \varepsilon_1^{\varphi_i} \otimes \dots \otimes \varepsilon_k^{\varphi_i} \otimes \dots \otimes \varepsilon_{i-1}^{\varphi_i}(x) \end{aligned} \quad (14)$$

Combined with theorem 2, we can get (9). For SSQ, the service guarantee rate is $g_i = \frac{\phi_i}{\sum_{j=1}^{N} \phi_j} C$, So we can get (10).

Theorem is proofed.

4 Forwarding performance guarantee in ForCES NE

As a network node that can process various forms of flows, ForCES NE should provide the capability of differentiated service to different flows, such as the worst-case delay requirements. Generally, when analyzing the issues of performance guarantee by using the performance upper bound based on probability, an error probability should be specified firstly. And in order to achieve the guarantee to a flow, we need find the appropriate processing weight to the flow in all FEs. In the subsequent analysis, we assume the error probability is known. Through changing the weight, according to (10), we can get a different delay d. We need find a best weight to a spedific d.

If there is a ForCES NE with N FEs which serves M flows, and the service rate of the ith FE is C_i. $A_i(t)$ （$i = 1,2, \dots M$）is the arrival process of each flow. Their arrival processes are stochastically independent, and they are all self-similar FBM process with Hurst parameter H. Their average arrival rate is ρ_i, where $\sum_{i=1}^{M} \rho_i < C_i$. When a flow is processed by the ith FE, the delay is d_i. The total delay is d while the flow goes through the ForCES NE. The processing weight is $x_i(0 < x_i < 1)$ in the ith FE.

Assume the flow is in the first place in the processing sequence, according the definition of feasible ordering, when $(\rho_1 + \theta_1) < x_i C_i$ ， that is $x_i > \frac{(\rho_1 + \theta_1)}{C_i}$, $1 \to (j)_{j=2,3 \dots N}$ is a feasible ordering with respect to $\rho_1 + \theta_1, \sum_{j=1}^{N-1}(\rho_j + \theta_j)\}$ and $\{x_i, (1 - x_i)\}$, where $\theta_j > 0$. And then $\varphi_1 = x_i$, $\varphi_2 = 1$, combined with (10), we can get

$$Pr\{D_{iGPS1}(t) \geq d_i\} < L_1 e^{-\left(\frac{d_i x_i C_i}{c_1}\right)^{\beta}} \quad (15)$$

Let the error probability $L_1 e^{-\left(\frac{d_i x_i C_i}{c_1}\right)^{\beta}} = \varepsilon$, then

$$d_i = \frac{c_1 \sqrt[\beta]{-\log(E/L_1)}}{x_i C_i} \quad (16)$$

There is $d_i > 0$, $\frac{\rho_1}{C_i} < x_i < 1$, and $d = \sum_{i=1}^{N} d_i$, so

$$d = c_1 \sqrt[\beta]{-\log(\varepsilon/L_1)} \sum_{i=1}^{N} \frac{1}{x_i C_i} \quad (17)$$

where $\beta = 2(1-H_1)$, $c = \left(\frac{2}{\theta^{H_1}}\right)^{\frac{1}{1-H_1}}$, $L_1 = e \cdot max\left\{1, 4^{H_1}\frac{\rho_1/_{\theta}+2-H_1}{2H(1-H_1)}\right\}$, θ is the optimization of parameter, and $(\rho_1+\theta_1) < x_iC_i$.

Given a ForCES NE with two FEs, the service rate $C_1 = 10MB/s$, $C_2 = 20MB/s$. An FBM flow with $H = 0.8$ is processed by these FEs, its average arrival rate $\rho_1 = 6MB/s$. The error probability ε is set to 0.01. When the delay that the flow goes through the NE is required less than 500ms or 800ms, according (17) and the condition of feasible ordering, we can get the processing weight relationship in two FEs, as shown in Fig. 6 and Fig. 7 respectively.

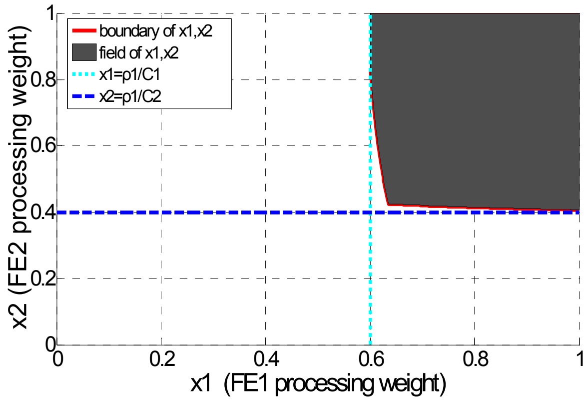


Fig. 6. The relationship of processing weight in FE1 and FE2 (**d = 800ms**)

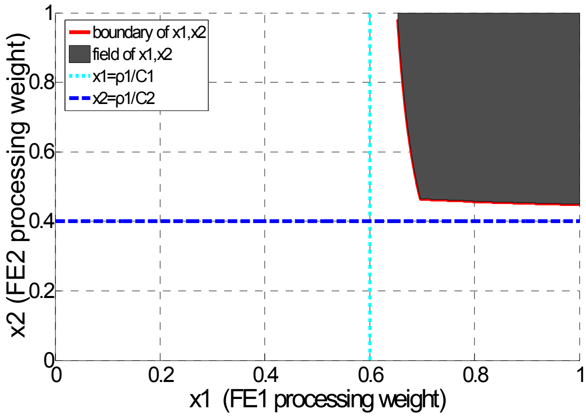


Fig. 7. The relationship of processing weight in FE1 and FE2 (***d* = 500*ms***)

From the figures, we can see, if the total worst-case delay requirement is smaller, the value field of x_1 and x_2 will be narrower. x_1 And x_2 are all have a minimum value. If the value of x_1 increases, the value of x_2 will be reduced correspondingly, that is, in the condition of total delay guarantee, when the flow's processing weight in one FE is increased, the processing weight in the other FE can be reduced.

Obviously, because there are no other constraints, when all processing weight in each FE is assigned to a particular flow, its total delay will be minimal. So it is difficult to implement worst-case performance delay guarantees, when we pay attention to one flow in isolation.

5 Simulation and validation

We simulated and validated the results of section 4 in NS-2 (Network Simulator, Version 2). We inherited the WIRR (Weighted Interleaved Round) to approximate GPS scheduling. The simulation topology is shown in Fig. **8**. GPS scheduling was running on node 1 and node 2. s1 is through flow, and s2 is cross flow. We used RMD (Random Midpoint Displacement) algorithm to generate self-similar FBM traffic.

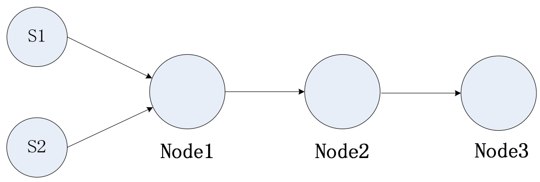


Fig. 8. Simulation topology

In the simulation, we recorded the arriving time of through flow on Node1 and Node2 (t1 and t3). Then the total delay is (t3-t1). All parameters used are same as those used in Fig. 6 and Fig. 7. When $x_1 = 0.65$, the results are shown in Fig. 9.

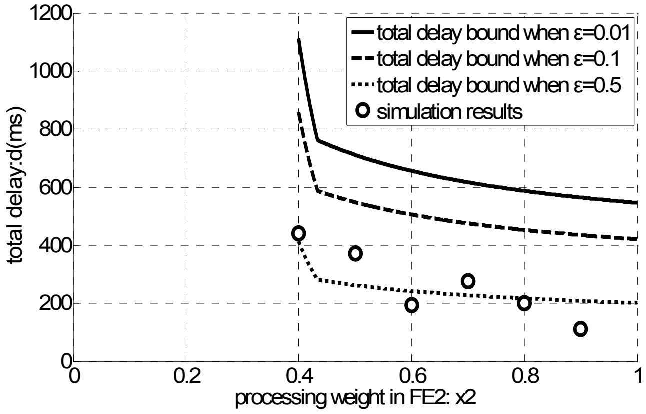


Fig. 9. The relationship of total delay with the processing weight in FE2（$\boldsymbol{x_1 = 0.65}$）

As shown in the figure, when $x_1 = 0.65$ and we fix an error probability, the worst-case total delay bound will be decreases with the increase of the value of x_2 theoretically, and the simulation results are basically in accordance with this statistical tendency. Moreover, the results are smaller than the theoretical value of the case that error probability is set to 0.1, and are close to the value with error probability 0.5.

6 Conclusion and future work

In this paper, we studied the FE forwarding performance model in ForCES NE. To establish the model, FE is abstracted as GPS scheduler. And in a stochastic network calculus framework, based on the self-similar FBM (Fractional Brownian Motion) traffic, the performance bounds of multiple flows processed by a single FE are derived. By using the performance model, in the ForCES system with multiple FEs, the processing weight requirements to guarantee the performance of business flows are analyzed. We also have given a simulation. Through comparing and analyzing the results, we can see the model can basically reflect the real system performance.

There are a number of issues to be investigated. For example, if there are multiple flows need to guarantee their performance simultaneously, how to optimize their processing weights in all FEs? At the same time, if there is parallel processing structure in the system, in this case, how to optimize the processing path according to the changes of performance caused by the dynamic changes of FBM parameters?

Acknowledgement

The authors would like to thank the anonymous reviewers for helpful comments. This work is supported by the National 973 Program (No. 2012CB315902, No.2009CB320804) of China, National Natural Science Foundation of China (No. 61272304).

References

[1] A. Doria, W. Wang, et al., "Forwarding and Control Element Separation (ForCES) Protocol Specification," IETF RFC5810, 2010.

[2] L. Yang, J. Halpern, et al., "Forwarding and Control Element Separation (ForCES) Forwarding Element Model," IETF RFC5812, 2010.

[3] K. Xiong, Z. Qiu, H. Zhang, C. Li, H-C Chao, "A Scalable Fast Forwarding Approach for IP Networks," *International Journal of Internet Protocol Technology*, Vol. 3, No. 2, pp. 119-127, 2008.

[4] K. Xiong, Z.-D. Qiu, Y.-C. Guo, H.-K. Zhang, H.-C. Chao, "Two LIC-Based Fast Forwarding Schemes for Explicit Routing with Scalability, Flexibility and Security," *Journal of Internet Technology,* Vol. 12 No. 3, pp. 407-416, 2011.

[5] J.-M. Chang, C.-C. Hsu, H.-C. Chao, J.-L. Chen, "Design of End Avoidance Method for Geographic Forwarding in MANET," *Wireless Personal Communications*, Volume 61, No. 4, pp. 765-777, 2011.

[6] J. Niu, M. Liu, H.-C. Chao, "PQBCF: A P2P Query Algorithm based on Betweenness Centrality Forwarding in Opportunistic Networks," *Mobile Information Systems*, DOI:10.3233/MIS-130166, 2012.

[7] S. Mao, S. S. Panwar, "A survey Of Envelope Processes and Their Applications in Quality of Service Provisioning," *IEEE Communications Surveys & Tutorials*, Vol. 8, No. 3, pp. 2-20, 2006.

[8] A. Kumar, D. Manjunath, J. Kuri, *Communication Networking: An Analytical Approach*, Morgan Kaufmann Press, 2004.

[9] S. Blake, D. Black, M. Carlson, et al., *An Architecture For Differentiated Services*, IETF RFC2475, 1998.

[10] J. Liebeherr, D. E. Wrege, D. Ferrari, "Exact Admission Control for Networks with a Bounded Delay Service," *IEEE/ACM Transactions on Networking*, Vol. 4, No. 6, pp. 885-901, 1996.

[11] M. Fidler, V. Sander, "A Parameter Based Admission Control for Differentiated Services Networks," *Computer Networks*, Vol. 44, No. 4, pp.463-479, 2004.

[12] M. Fidler, *Providing Internet Quality of Service Based on Differentiated Services Traffic Engineering*, Mainz Press, 2004.

[13] L. Lenzini, L. Martorini, E. Mingozzi, et al., "Tight End-to-end Per-Flow Delay Bounds in FIFO Multiplexing Sink-tree Networks," *Performance Evaluation*, Vol. 63, No. 9, pp. 956-987, 2006.

[14] K I M Geunhyung, K I M Cheeha, "Deterministic Edge-To-Edge Delay Bounds for a Flow Under Latency Rate Scheduling in a DiffServ Network," *IEICE Transactions on Communications*, Vol. 88, No. 7, pp. 2887-2895, 2005.

[15] L. Lenzini, E. Mingozzi, G. Stea, "A methodology for computing end-to-end delay bounds in FIFO-multiplexing tandems," *Performance Evaluation*, Vol. 65 No. 11, pp. 922-943, 2008.

[16] Z. Zhang, D. Twosley, J. Kursose, "Statistical Analysis of the Generalized Processor Sharing Scheduling Discipline," *IEEE Journal on Selected Areas in Communications*, Vol.13, No. 6, pp. 1071-1080, 1995.

[17] A. K. Parekh, R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-node Case," *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp. 344-357, 1993.

[18] A. K. Parekh, R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case," *IEEE/ACM Transactions on Networking*, Vol. 2, No. 2, pp. 137-150, 1994.

[19] Y-M Jiang, Y. Liu. *Stochastic Network Calculus*, Springer-Verkag Press, 2008

[20] F. Ciucu, A. Burchard, J. Liebeherr, "A Network Service Curve Approach for the Stochastic Analysis of Networks," in *Proceedings of 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (ACM SIGMETRICS 2005)*, pp. 279-290, ACM Press, 2005.

[21] C. Li, S. Zhang, W. Wang, "Scheduling Model and Performance Analysis in Transport Mapping Layer of Control Element in Forwarding and Control Element Separation System," *International Journal of Communication Systems,* Vol. 26, pp. 395-411, 2013.

[22] C. Li, S. Zhang, W. Wang, "Statistical Performance Analysis of Strict Priority and Generalised Processor Sharing Two-Stage Scheduling System under Exponentially Bounded Burstiness Input Model," *IET Networks*, Vol. 1, No. 3, pp.146-154, 2012.

[23] J. Liebeherr, A. Burchard, F. Ciucu, "Delay Bounds in Communication Networks with Heavy-tailed and Self-similar Traffic," *IEEE Transactions on Information Theory*, Vol. 58, No. 2, pp. 1010-1024, 2012.