A QoS CONTROL MODEL IN OVERLAY CONVERGENCE TRANSMISSION

Binghui XU[1], Wensheng YAN[2]

In order to provide effective delay guarantee in the IP network, this paper proposes a QoS control model in overlay convergence transmission. In order to ensure the delay requirements of the convergent flow in the service overlay network, this paper has designed a simple and effective adaptive control mechanism. It can provide full service in the circular sequence and can guarantee a variety of sudden requests of delay generated by the convergent flow. At the same time, this paper uses Markov chain and probability generating function to analyze the relevant performance and verify the analysis results through simulation experiments.

Keywords: QoS, SON, streaming media

1. Introduction

With the explosive growth of Internet users and multimedia services, Quality of Service (QoS) requirements will become more and more common in future network applications. ‘All IP-based’ solutions have become the first choice for the next generation of Internet technology, and the importance of IP QoS in multimedia applications has been continuously demonstrated. Multimedia applications require end-to-end QoS guarantee, which includes effective control of QoS in many links such as end systems, local area networks, backbone networks and access networks [1]. In order to achieve IP QoS, Internet Engineering Task Force (IETF) has proposed several solutions in recent years, which are IntServ, DiffServ and MPLS [2]. However, there are two key conditions for these solutions. Firstly, all routers along the route need to have scheduling and cache management functions [3]. Secondly, proper incentive mechanisms are needed to encourage Internet service providers to provide support for these functions [4]. However, these requirements are still difficult to achieve, and best-effort is still a common service model on the Internet. So far, Internet QoS is still an open research subject.

On the other hand, the Internet has evolved into a basic network for service distribution. Many value-added services and multimedia content distribution services are implemented by overlay networks instead of IP layers, as

[1] Sr. Eng., Logistics Department, Taizhou Vocational & Technical College
[2] Assoc. Prof., College of Information Technology Engineering, Taizhou Vocational & Technical College

mentioned in literature [5-7]. In view of this, QoS guarantee technology is particularly important, and many beneficial attempts have been made at home and abroad on this subject. Literature [8-9] studies the problem of load balancing of multimedia streams based on P2P architecture. Literature [10-11] discusses the QoS performance evaluation and real-time streaming modeling of the overlay multicast network. The Service Overlay Network (SON) architecture described in [7] is a typical QoS solution for service integration networks. SON consists of a series of service gateways, which mainly implement relay and control functions of service data. The logical connection between the two service gateways is provided by the underlying network, and the bandwidth is guaranteed to a certain extent. This QoS guarantee is only limited to bidirectional Service Level Agreements (SLAs) between the SON and the underlying network. With bidirectional SLAs, SON can provide end-to-end QoS sensitive business services through proper service resource management [12]. The underlying network can aggregate the SONs services and provide flow control and QoS control based on the corresponding SLAs. However, this QoS control only considers the bandwidth of different aggregated services and does not involve the delay [13].

1.1. Motivation

In the network transmission, the transmission of high-quality video services often requires high bandwidth, low latency or low cost. In the traditional Qos control mode, when the video is transmitted on a bandwidth-constrained network or in a best-effort mode, the end-to-end bandwidth requirements are often not met and the quality of service cannot be guaranteed. In addition, the end-to-end network status is comprehensively determined by the status of each node in the network, and thus is affected by the dynamic changes of flow control rules. Once some links are congested and the available bandwidth is small, a large number of data packets will be lost when the video data passes through the link, and the packet loss rate will increase significantly and even serious network congestion will occur, thus greatly reducing the data transmission effect. This puts forward higher requirements for the QoS control mechanism and requires it to have a resource management capability for different services so that it can adaptively adjust the network status in a complex network environment.

Due to the lack of sufficient flexibility and self-management and self-adjustment capabilities in traditional Internet architecture and QoS models, it is difficult to achieve efficient and self-adaptive management control and QoS guarantee in a complex network environment. In the design of future network architecture, QoS technology research faces challenges including:

- Organic integration of QoS model and new network architecture.
- Performance pressure of packet QoS processing and forwarding mechanism.

- The adaptive ability of QoS model to dynamic network and business environment.

In view of the importance of QoS guarantee technology in the field of future network research, this paper proposes a QoS control model in overlay convergence transmission. It has made research attempts in the following aspects：

- It proposes a new QoS architecture that matches the current and next-generation Internet architecture.
- It provides efficient QoS guarantee mechanisms and algorithms for high-speed differentiated processing of packets.
- It solves the problems of complex configuration and high management overhead of traditional QoS models and provides adaptive and dynamic adjustment capabilities to support personalized user needs in complex network environments.

1.2 Research contents and contributions

In response to the above problems, this paper innovatively proposes a QoS control model in overlay convergence transmission. A control mechanism is added to the SON architecture to provide differentiated delay guarantee for different aggregation services. Considering that different types of multimedia services have different burstiness, for example, the video service has a greater burstiness than the voice service [14], the control model consists of an asymmetric polling system that includes two types of queues. The polling rules use full service, which integrates the traditional control level and management level. It fully guarantees the consistency of the decision-making process. The control model is located on the SON service gateway as an independent functional unit. It can adaptively provide lower queuing delay for service flows with higher bursts without priority setting. At the same time, through direct control of the entire network, it avoids the view interaction between the traditional control mechanisms, which is better guaranteed the stability of the network transmission service.

The main contribution of this paper is to fully integrate the traditional QoS control model and use the control theory to propose a QoS control model in overlay convergence transmission, which can provide lower queuing delay for services with higher burstiness without setting priority in advance. It can effectively improve transmission efficiency while ensuring stability. In addition, this paper embeds Markov chain and probability generating function method to theoretically analyze the model and obtains the analytical expression of the average queuing delay, which provides a new idea for studying the Qos adaptive control mechanism.

The rest of this article is organized as follows: section 2 describes the control model in detail, section 3 analyzes the model in detail, section 4 performs numerical analysis on the model and performs simulation experiments to verify, and section 5 sums up.

2. Control model

The control model consists of two queues, and the control center queries and serves each queue in turn according to the full-service rules, as shown in Fig. 1.

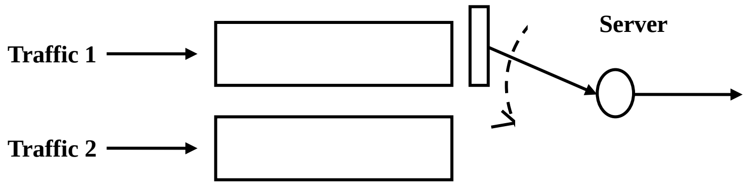

Fig. 1. Control model

As different multimedia services have different burstiness, the multimedia service flow can be divided into two types of queues according to the burstiness, one queue corresponds to the group with larger burstiness, and the other corresponds to the smaller burstiness. According to the full-service rule, when the central processing unit polls a queue in turn, it not only processes the packets waiting in the queue, but also processes all the packets that arrive during the service. Then, after a conversion time interval, the next queue is polled [15].

Each queue in the system model involves three independent random processes, which are packet arrival process, service process of central processing unit for the queue, and the conversion process between the two queues [16].

3. Model analysis

In the control model proposed in the previous section, each queue has three independent random processes, which are packet arrival, queue service and queue conversion. QoS requirements involve many parameters, such as delay, delay jitter, and packet loss rate. This paper focuses on the queuing delay that occurs in the nodes of the overlay network.

3.1 Related basic knowledge

(i). Poisson process. Poisson process refers to the most basic independent incremental process that accumulates the number of random events. In this article, the system accumulates the received queue packets over time, which forms a Poisson process.

(ii). FCFO. FCFO refers to First-Changed First-Out. It uses the highest priority number first and first-in-first-served algorithms to manage and schedule processes separately.

(iii). define the following random variables:

$u_1(n)$: The time interval between the transition from queue Q_i to queue Q_{i+1} at time t_n.

$v_1(n)$: The service time of the queue Q_i at time t_n.

$\mu_j(u_i)$: Number of packets arriving in the queue Q_j in time u_i.

$\eta_j(v_i)$: Number of packets arriving in the queue Q_j in time v_i.

$\xi_i(n)$: Number of packets in the queue Q_j in time t_n.

3.2 Basic definition

Assuming that the system is discrete, the buffer space is large enough to prevent packet loss. The packet arrival process of each queue is independent of each other. The service process of one queue is independent of the service process of other queues and the arrival process of all queues. The following parameters are defined as follow:

(iv). The random arrival process of packets in queue Q_i is an independent Poisson process with an arrival rate of λ_i. The corresponding probability generating function is $A_i(z)$, and its first and second moments are respectively λ_i and $A_i''(1)$.

(v). The grouping service time variable in the queue Q_i follows the general distribution of the parameter β_i, and the corresponding probability generating function is $B_i(z)$. The first and second moments are respectively β_i and $B_i''(1)$, and the queue service is in accordance with FCFO.

(vi). When the central processing unit has finished servicing the queue Q_i, it will turn to query the service queue Q_{i+1}. The conversion time variable of the period follows the general distribution, and the corresponding probability generating function is R_i.

The subscript i of the queue Q_i is i = 1, 2, which corresponds to the multimedia service 1 and the multimedia service 2 respectively, and so are the other subscripts.

3.3 Average waiting delay analysis

First of all, lemma 1 is proposed:

Lemma 1: $G_i(z_j, z_k)$ describes the steady-state joint probability distribution generating function of the system when the queue Q_i is queried, expressed as:

$$G_1(z_1,z_2)=R_2(A_1(z_1)A_2(z_2))G_2(z_1,B_2(A_1(z_1)F_2(A_1(z_1)))) \quad (1)$$

$$G_2(z_1,z_2)=R_1(A_1(z_1)A_2(z_2))G_1(B_1(A_2(z_2)F_1(A_2(z_2))),z_2) \quad (2)$$

In the above formula, F_1 and F_2 are expressed respectively as follows:

$$F_2(z)=A_2(B_2(zF_2(z)))$$

$$F_1(z)=A_1(B_1(zF_1(z)))$$

The proof is listed in Annex.

Next, we introduce theorem 1:

Definition 1: The first moment of the queue Q_j is related to the system state of the queue Q_i when it receives service, and it satisfies the following equation:

$$g_i(j)=\lim_{z_z,z_2\to 1}\frac{\partial G_i(z_1,z_2)}{\partial z_j},\ i=1,2 \quad j=1,2 \quad (3)$$

Definition 2: When the queue Q_i receives service, the steady-state second moment of the system is defined as follows:

$$g_i(j,k)=g_i(k,j)=\lim_{z_1,z_2\to 1}\frac{\partial^2 G_i(Z_1,Z_2)}{\partial z_j \partial z_k},\ \ i=1,2 \quad j=1,2 \quad k=1,2 \quad (4)$$

Theorem 1: The average waiting time of the packets in the queue Q_i satisfies the following expression:

$$\overline{w_i}=\frac{g_i(i,i)}{2\lambda_i g_i(i)}+\frac{(2\rho_i^{\,2}+\rho_i-1)A_i''(1)}{2\lambda_i^{\,2}(1-\rho_i^{\,2})}+\frac{\lambda_i B_i''(1)}{2(1-\rho_i)},\quad i=1,2 \quad (5)$$

The proof is listed in Annex.

4. Numerical results and simulation

In this section, we will perform numerical analysis according to theorem 1 and expressions (10)-(13) and establish a computer simulation model for experimental verification. We will analyze the correlation between average waiting delay, average queue length and packet arrival rate, service strength.

In order to verify the stability of the QoS model designed in this article in dealing with complex network environments, this article conducted two experiments, Simulation 1 and Simulation 2. By separately testing and comparing the effects of waiting delay and queue length changes on the packet arrival rate, simulate the adaptability and stability of the test system in the actual complex network environment. In order to verify whether the model has performance advantages when dealing with complex network environments, two experiments, Simulation 3 and Simulation 4 are carried out in this paper. By separately testing and comparing the effects of waiting delay and queue length changes on the average service time, simulate test whether the performance of the system under the actual complex network environment is effective.

This article uses Wireshark software to capture real-time business flow. The pre-processing program is developed based on C++ on the Microsoft Visual Studio platform to test the waiting time delay, queue length, service intensity and other parameter values of the business flow, and to simulate the model parameters in different network environments for simulation experiments. Based on the data obtained by the above method, Matlab is used to verify the effectiveness of the model proposed in this paper.

The hardware configuration environment of the experiment is Windows 7 Professional(64bit)，Intel(R) Core(TM)i5-6500 @3.2Ghz，8G DDR3 memory。

The specific simulation parameters are shown in the Table 1.

Table 1

Simulation parameters

Parameters	Value
Simulation scene size（m×m）	20×20
Number of nodes	100
Simulation time /min	5
Channel transmission rate /（Mb•s-1）	2
Node transmission distance	2
Node buffer capacity /packets	100
MAC protocol	IEEE 802.11 DCF

The experiment uses the real-time data set captured by Wireshark software, which contains various common business flows.

In numerical analysis and simulation experiments, the service conversion time parameter is $\gamma_1 = \gamma_2 = 1$.

(i). Simulation 1

Set β_1 and β_2 to take the same value of 1, and λ_2 takes a fixed value of 0.01, as shown in Fig. 2.

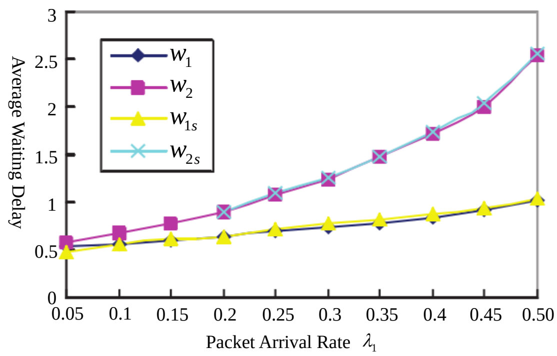

Fig.2. Average waiting delay-packet arrival rate

According to the results of numerical analysis, w_1, w_2 indicate the correlation between the average waiting delay and the burstiness of multimedia services. And then carry out computer simulation experiments according to the control model. w_{1s} 、 w_{2s} correspond to w_1, w_2 respectively. Simulation results are consistent with numerical analysis. The results show that the average waiting delay of packets with a higher packet arrival rate is lower than that with a lower arrival rate, and both increase with the increase of λ_i. The reason is that as the system load increases, the value is $\rho_i = \beta_i \lambda_i$.

(ii). Simulation 2

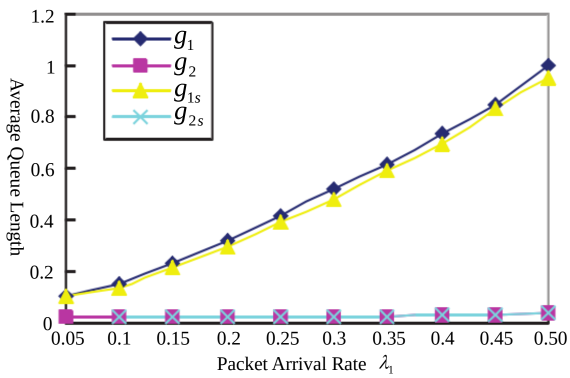

Fig.3. Average queue length-packet arrival rate

Set β_1 and β_2 to take the same value of 1, and λ_2 takes a fixed value of 0.01, as shown in Fig. 3.

According to the result of numerical analysis, g_1, g_2 indicate the correlation between the average queue length and the burstiness of multimedia services. In the computer simulation experiments, g_{1s} and g_{2s} correspond to g_1 and g_2 respectively. Simulation results are consistent with numerical analysis.

The results show that the average queue length of the service queue with a higher packet arrival rate is lower than the service queue with a lower arrival rate, and both increase with the increase of λ_1. Business queues with higher packet arrival rates require higher buffer space.

(iii). Simulation 3

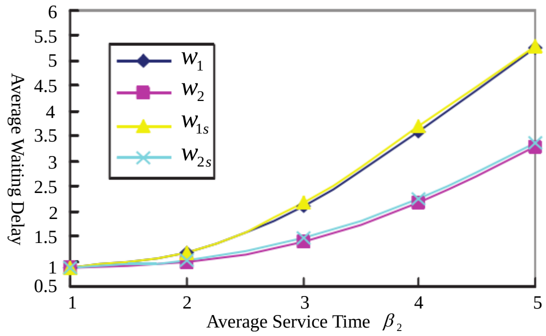

Fig.4. Average waiting delay-average service time

Set λ_1 and λ_2 to take the same value of 0.1, and β_1 takes a fixed value of 1, as shown in Fig. 4.

According to the result of numerical analysis, w_1, w_2 indicate the relationship between the average waiting time and service strength of different service queue packets. In the computer simulation experiments, w_{1s} and w_{2s} correspond to w_1 and w_2 respectively. Simulation results are consistent with numerical analysis. The results show that business queues with larger β values have lower average packet waiting time. The reason is that the services received by the relevant queues have been effectively enhanced.

(iv). Simulation 4

Set λ_1 and λ_2 to take the same value of 0.1, and β_1 takes a fixed value of 1, as shown in Fig. 5.

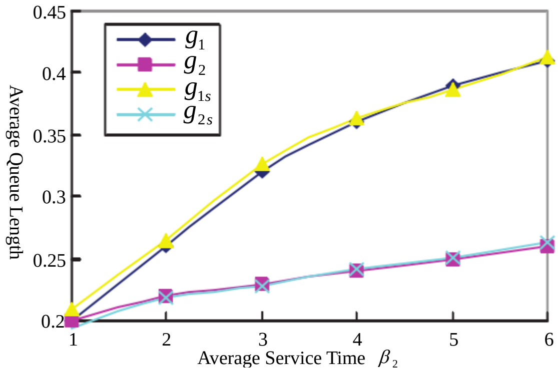

Fig.5. Average queue length-average service time

Based on the simulation results, the following conclusions can be obtained: The control model designed in this paper can provide lower queuing delay for services with higher burstiness without the need to preset priority. Moreover, the change in service strength β can affect the length of the queuing delay. This means that the value of the packet length can effectively affect the queuing delay, which has a positive significance for multimedia applications in IP networks.

Because the control model designed in this article does not need to pre-set priorities for services, it effectively avoids the traditional method that requires all routers along the route to have scheduling and buffer management restrictions. Only by setting up routers with scheduling and cache management functions distributed in key nodes according to the actual situation, a relatively effective management strategy can be realized. When the router does not have complicated scheduling and cache management strategies, the data packets can be processed in the FCFO mode. Node routers with relatively advanced scheduling and queue caching strategies need to manage cache queues, monitor network status, and provide corresponding congestion indications to reduce network delays by reducing the impact of packet loss and retransmission, while effectively reducing construction cost.

5. Compare study on properties

In this section, we will test the QoS control model proposed in this article with the more commonly used Intserv and Diffserv in a practical application scenario. For each QoS control model, we take three business flows with different priorities for testing. The service flow parameters are as follows:

(i). Service flow 1 has an average rate of 1000 Kbit/s, a data packet of 1000 bytes, and a priority of 14.

(ii). Service flow 2 has an average rate of 500 Kbit/s, a data packet of 500 bytes, and a priority of 11.

(iii). Service flow 3 has an average rate of 500 Kbit/s, a data packet of 500 bytes, and a priority of 8.

The link speed is 2Mbit/s, and the propagation delay is 1m.

After testing, Table 2 shows the test results of the time delay of the three QoS control models.

Table 2

Time delay of three QoS control models

Business flow	Intserv	Diffserv	Model in this paper
1	132.37ms	187.26ms	135.64ms
2	225.28ms	212.68ms	162.53ms
3	/	523.47	219.73ms

The experimental results show that in service flow 1, when the rate is half of the link rate, the IntServ model can be used to obtain a lower delay, so that the other half of the bandwidth can be shared by other service flows, resulting in service congestion. This makes the transmission of the service flow 2 have a higher delay than the other two models and causes the transmission of the service flow 3 to be unable to be completed smoothly.

6. Conclusion

The QoS guarantee of Internet network multimedia applications is an important and complicated subject. The existing IP QoS guarantee mechanisms such as Inerserv and Diffserv have encountered many difficulties in the implementation process. On the contrary, the use of overlay network technology to provide QoS-guaranteed multimedia value-added services on the Internet is a relatively promising solution that has received more and more attention. A simple and effective adaptive control mechanism designed in this paper can be located on the SON service gateway as an independent functional unit. It can provide full service in the cyclic sequence and can guarantee the delay of various convergent flows due to burst request.

The research in this article focuses on solving the following two problems in the actual network QoS guarantee technology.

- In response to the demand for high-speed packet processing in complex networks, this article studies a high-performance QoS model suitable for future network information rates. This article proposes a solution with higher time and space efficiency, which has significant practical application value.
- In order to solve the QoS guarantee problem in the actual intricate network environment, this article applies the autonomous network technology to the design of the QoS model to solve the problems of complicated configuration and high management overhead of the

original QoS model. It enables the network and QoS model to have a high degree of self-adaptation and dynamic adjustment capabilities and can support future complex network environments and highly personalized user needs.

The main original contribution of this paper is to use control theory to propose a QoS control model in overlay convergence transmission. The model is theoretically analyzed using the embedded Markov chain and the probability generating function method, and the analytical expression of average queuing delay is obtained. This paper provides a new idea for the study of Qos adaptive control mechanism, and the validity of the analysis is verified through related computer simulation experiments. The results show that the control model can adaptively provide lower queuing delay for the service flow with higher burst without priority setting. This hybrid structure is conducive to achieving end-to-end QoS guarantee.

Acknowledgments

This work was supported by the Major and Special Project of Taizhou Vocational & Technical College under Grant 2020HGZ03.

REFERENCES

[1] L. Zuo, M. M. Zhu, C. Q. Wu. Fast and Efficient Bandwidth Reservation Algorithms for Dynamic Network Provisioning. Journal of the Network and Systems Management. 2015. 23(3):420–444.

[2] Chen Z, Lin T, Wu C. Decentralized learning-based relay assignment for cooperative communications. IEEE Transactions on Vehicular Technology, 2016, 65(2): 813-826.

[3] Mubashar Mushtaq, Toufik Ahmed. Hybrid. Overlay Networks Management for Real-Time Multimedia Streaming over P2P Networks. Lecture Notes IN Computer SCIENCE. 2008.7(47):1-13

[4] Khalil A, Mbarek N, Togni O. Self-Configuring IoT Service QoS Guarantee Using QBAIoT[J]. Computers. 2018. 7(4): 1-26.

[5] Zhao Yuhui, An Yuyan, Wang Hongjun, Gao yuan. Live Streaming Overlay Multicast Service Network. Computer Science. 2007.2(34):19-21

[6] Di X, Fan Y, Shen Y. Local martingale difference approach for service selection with dynamic QoS. Computers and Mathematics with Applications. 2011. 61(9): 2638-2646.

[7] Z. Dan, Z. L. Zhang, and T. Hou. Service Overlay networks: SLAs, QoS and Bandwidth Provisioning. Proc. of 10th IEEE International Conference on Network Protocols. Paris, France. 2002.11

[8] Sharma V, Kumar R. Estimation-based queue scheduling model to improve QoS for end users in manets. Computing & Informatics. 2017.35(5): 1079-1109.

[9] Wang C, Guo S, Yang Y. An optimization framework for mobile data collection in energy-harvesting wireless sensor networks. IEEE Transactions on Mobile Computing. 2016.15(12): 2969-2986.

[10] Choi J. Effective Capacity of NOMA and a Suboptimal Power Control Policy With Delay QoS. IEEE Transactions on communications. 2017.65(4):1849-1858.

[11] K. Kinoshita, M. Aihara, S. Kono, et al. Joint Bandwidth Scheduling and Routing Method for Large File Transfer with Time Constraint and its Implementation. Ieice Transactions on Communications. 2018.101(3):452-471.
[12] P. Mieghem, F. Kuipers. On the Complexity of Qos Routing. Computer Communications. 2003.26:376–387.
[13] P. Dharam. Qos Routing for Big Data Transfer in Software-defined Networks. The University of Memphis Memphis. 2014:1–79.
[14] Kamimura A, Tomita K. A Self-Organizing Network Coordination Framework Enabling Collision-free and Congestion-less Wireless Sensor Networks. Journal of Network and Computer Applications. 2017. 93: 228-244.
[15] Bavio J, Marrón B. Properties of the Estimators for the Effective Bandwidth in a Generalized Markov Fluid Model. Open Journal of Statistics. 2018.8(1): 69-84.
[16] A. Hou, C.Q.Wu, D.Fang. Bandwidth Scheduling for Big Data Transfer Using Multiple Fixed Node-disjoint Paths. Journal of Network and Computer Application. 2017.85:47–55.

Annex

(i)The proof of Lemma 1:

First define the following random variables:

$u_1(n)$: The time interval between the transition from queue Q_i to queue Q_{i+1} at time t_n.

$v_1(n)$: The service time of the queue Q_i at time t_n.

$\mu_j(u_i)$: Number of packets arriving in the queue Q_j in time u_i.

$\eta_j(v_i)$: Number of packets arriving in the queue Q_j in time v_i.

$\xi_i(n)$: Number of packets in the queue Q_j in time t_n.

In the above definition, the subscript values are i = 1, 2, j = 1, 2.

At the moment t_n, the system status is $[\xi_1(n), \xi_2(n)]$ and the corresponding probability distribution is $P[\xi_i(n) = x_i; i = 1,2]$

When $\sum_{i=1}^{2} \beta_i \lambda_i = \sum_{i=1}^{2} \rho_i < 1$, the system steady-state probability is

$$\pi_i(x_1, x_2) = \lim_{n\to\infty} p[\xi_i(n) = x_i, i = 1,2]$$

Therefore, the system probability distribution function is

$$G_i(z_1, z_2) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \pi_i(x_1, x_2) z_1^{x_1} z_2^{x_2}, i = 1,2$$

With the following formula:

$$\begin{cases} \xi_j(n+1) = \xi_j(n) + \mu_j(u_i) + \eta_j(v_i), j \neq i \\ \xi_i(n+1) = \mu_i(u_i) \end{cases} \tag{6}$$

The probability generating function of the system at time t_{n+1} can be obtained as follows:

$$G_{i+1}(z_1,z_2)=\lim_{n\to\infty}E[\prod_{j=1}^{2}z_j^{\xi_j(n+1)}]=E[\prod_{j=1,\neq i}^{2}z_j^{\xi_j(n)+\eta_j(v_i)}]E[\prod_{j=1}^{2}z_j^{\mu_j(u_i)}] \quad (7)$$

The formulas (1) and (2) can be obtained from the formulas (6) and (7).

(ii)The proof of Theorem 1:

Let $\xi_i(m)$ denote the number of packets arriving in the queue i in time slot $\tau_m(\tau_m=t_{m+1}-t_m)$, and the lth packet $(l=1,\Lambda,\xi_i(m))$ will be sent at a certain time $t_{m'}$ within the queue i receiving service time $t_n(t_m<t_{m+1}<t_n<t_{m'})$. The average waiting time of the lth packet in the queue can be expressed as $w_i(n)=w_{i,1}(n)+w_{i,2}(n)$, and $w_{i,1}(n)=t_n-t_m$, $w_{i,2}(n)=t_{m'}-t_n$. According to the mathematical model established above, we can get:

$$w_{i,1}(n)=\lim_{T\to\infty}\frac{\sum_{n-1,n_s=1}^{T}(\theta_{i,1}(n)+\theta_{i,1}(n_s))}{\sum_{n=1}^{T}\xi_i(n)}=\frac{E[\theta_{i,1}(n)]+E[\theta_{i,1}(n_s)]}{E[\xi_i(n)]} \quad i=1,2$$

$$=\frac{g_i(i,i)}{2\lambda_i(1+\rho_i)g_i(i)}-\frac{(1-\rho_i-\rho_i^2)A_i''(1)}{2\lambda_i^2(1-\rho_i^2)}+\frac{\lambda_iB_i''(1)}{2(1-\rho_i^2)} \quad (8)$$

$$w_{i,2}(n)=\lim_{T\to\infty}\frac{\sum_{n=1,n_s=1}^{T}(\theta_{i,2}(n)+\theta_{i,2}(n_s))}{\sum_{n=1}^{T}\xi_i(n)}=\frac{E[\theta_{i,2}(n)]+E[\theta_{i,2}(n_s)]}{E[\xi_i(n)]} \quad i=1,2$$

$$=\frac{\rho_ig_i(i,i)}{2\lambda_i(1+\rho_i)g_i(i)}+\frac{\rho_i^2A_i''(1)}{2\lambda_i^2(1-\rho_i^2)}+\frac{\lambda_i\rho_iB_i''(1)}{2(1-\rho_i^2)} \quad (9)$$

In the above formula, $\theta_{i,1}(n)$, $\theta_{i,2}(n)$, $\theta_{i,1}(n_s)$ and $\theta_{i,2}(n_s)$ can be respectively defined as follows:

$$\theta_{i,1}(n)=\alpha_{i,1}(n)(\eta(n)-1)^++\alpha_{i,2}(n)(\eta(n)-2)^++\Lambda+\alpha_{i,(\eta(n)-1)}(n)^+ \quad i=1,2$$

$$\theta_{i,1}(n_s)=\alpha_{i,1}(n_s)(\eta(n_s)-1)^++\alpha_{i,2}(n_s)(\eta(n_s)-2)^++\Lambda+\alpha_{i,(\eta(n_s)-1)}(n_s)^+$$

$$i=1,2 \quad s=1,2,\Lambda$$

$$\theta_{i,2}(n)=v_{i,1}(n)+(v_{i,1}(n)+v_{i,2}(n))+\Lambda+\sum_{l=1}^{\xi_i(n)}v_{i,l}(n) \quad i=1,2$$

$$\theta_{i,2}(n_s) = v_{i,1}(n_s) + (v_{i,1}(n_s) + v_{i,2}(n_s)) + \Lambda + \sum_{l=1}^{\xi_i(n_s)} v_{i,l}(n_s) \quad i = 1,2 \quad s = 1,2,\Lambda$$

$\eta(n)$ indicates the time interval for querying queue i again at time t_n from the last time servicing queue i in last query cycle according to the full service rule. Its value meets the following conditions:

$$(\eta(n) - \varepsilon)^+ = \begin{cases} \eta(n) - \varepsilon & \eta(n) \geq \varepsilon \\ 0 & \eta(n) < \varepsilon \end{cases}$$

$\alpha_{i,m}(n)$ represents the number of packets that arrive in the queue i in the τ_m gap.

$v_{i,l}(n)$ represents the service time of the lth packet($i = 1,2$, $l = 1,2,\Lambda\ ,\xi_i(m)$).

$\eta(n_s)$ represents the time required to send all packets in the queue when a queue accepts the query.

$\alpha_{i,m}(n_s)$ represents the number of newly arrived packets in a queue during query service.

Substitute expressions (8) and (9) into $w_i(n) = w_{i,1}(n) + w_{i,2}(n)$ to get expression (5). Finally, we need to use the first and second moments of the system probability generating function to calculate the average waiting delay which are $g_i(i)$ and $g_i(i,i)$. Therefore, substituting expressions (1) and (2) into expressions (6) and (7) and iterate, we can get $g_1(1)$, $g_2(2)$ and $g_1(1,1)$, $g_2(2,2)$ respectively, as follows:

$$g_1(1) = \lambda_1(1-\rho_1)\bar{\theta} \tag{10}$$

$$g_2(2) = \lambda_2(1-\rho_2)\bar{\theta} \tag{11}$$

In the above formula $\bar{\theta} = \dfrac{\gamma_1 + \gamma_2}{1 - \rho_1 - \rho_2}$,

$$\begin{aligned} g_1(1,1) = \frac{\bar{\theta}^2}{(\gamma_1+\gamma_2)^2(1-\rho_1-\rho_2+2\rho_1\rho_2)} \{ & \beta_2{}^2\lambda_1{}^2(1-\rho_1)^2(\gamma_1+\gamma_2)A_2''(1) \\ & + \lambda_1{}^2(1-\rho_1)(1-\rho_1-\rho_2)(1-\rho_1-2\rho_2+2\rho_1\rho_2+2\rho_2{}^2)R_2''(1) \\ + (\gamma_1+\gamma_2)\lambda_2\lambda_1{}^2(1-\rho_1)^2 B_2''(1) & + (\gamma_1+\gamma_2)[(1-\rho_2)^2(1-\rho_1)^3 + \rho_2{}^2\rho_1{}^3]A_1''(1) \\ & + \lambda_1{}^2(1-\rho_1)^2(1-\rho_1-\rho_2)R_1''(1) + (\gamma_1+\gamma_2)\lambda_1{}^3\rho_2{}^2 B_1''(1) \\ & + 2\gamma_1\gamma_2\lambda_1{}^2(1-\rho_1)(1-\rho_1-\rho_2+\rho_1\rho_2+\rho_2{}^2)(1-\rho_1-\rho_2) \\ & + 2\lambda_1{}^2[(1-\rho_2)(1-\rho_1)^2\gamma_1\rho_2 + (1-\rho_1)\gamma_2\rho_1\rho_2{}^2](1-\rho_1-\rho_2) \end{aligned}$$

$$+2(\lambda_1+\lambda_2)\lambda_1^{\,2}[(1-\rho_1)^2\gamma_1\rho_1\rho_2+(1-\rho_1)\gamma_2\rho_2(1-\rho_1-2\rho_2+2\rho_2\rho_1+\rho_2^{\,2})]$$
$$+\rho_2^{\,2}(1-\rho_1)(1-\rho_1-\rho_2+\rho_1\rho_2+\rho_1^{\,2})\Big\} \qquad (12)$$

$$g_2(2,2)=\frac{\bar{\theta}^2}{(\gamma_1+\gamma_2)^2(1-\rho_1-\rho_2+2\rho_1\rho_2)}\Big\{\beta_1^{\,2}\lambda_2^{\,2}(1-\rho_2)^2(\gamma_1+\gamma_2)A_1''(1)$$
$$+\lambda_2^{\,2}(1-\rho_2)(1-\rho_1-\rho_2)(1-\rho_2-2\rho_1+2\rho_1\rho_2+2\rho_1^{\,2})R_1''(1)$$

$$+(\gamma_1+\gamma_2)\lambda_1\lambda_2^{\,2}(1-\rho_2)^2B_1''(1)+(\gamma_1+\gamma_2)[(1-\rho_1)^2(1-\rho_2)^3+\rho_1^{\,2}\rho_2^{\,3}]A_2''(1)$$
$$+\lambda_2^{\,2}(1-\rho_2)^2(1-\rho_1-\rho_2)R_2''(1)+(\gamma_1+\gamma_2)\lambda_2^{\,3}\rho_1^{\,2}B_2''(1)$$
$$+2\gamma_1\gamma_2\lambda_2^{\,2}(1-\rho_2)(1-\rho_1-\rho_2+\rho_1\rho_2+\rho_1^{\,2})(1-\rho_1-\rho_2)$$
$$+2\lambda_2^{\,2}[(1-\rho_1)(1-\rho_2)^2\gamma_2\rho_1+(1-\rho_2)\gamma_1\rho_2\rho_1^{\,2}](1-\rho_1-\rho_2)$$
$$+2(\lambda_1+\lambda_2)\lambda_2^{\,2}[(1-\rho_2)^2\gamma_2\rho_2\rho_2+(1-\rho_2)\gamma_1\rho_1(1-\rho_2-2\rho_1+2\rho_2\rho_1+\rho_1^{\,2})]$$
$$+\rho_1^{\,2}(1-\rho_2)(1-\rho_1-\rho_2+\rho_1\rho_2+\rho_2^{\,2})\Big\} \qquad (13)$$