The Load Balancing Research of SDN based on Ant Colony Algorithm with Job Classification

Wucai Lin[1,a] , Lichen Zhang[2,b]

[1]School of Computer Science,Guangdong University of Technology Guangdong, Guangzhou 510006, China.

[2]School of Computer Science,Guangdong University of Technology Guangdong, Guangzhou 510006, China.

[a]claytale@163.com

Keywords: SDN, Cloud computing, Load balance, ant colony algorithm, job classification.

Abstract. In this paper, we propose a dynamic load balancing algorithm conbined with job classification and ant colony algorithm in SDN network cloud computing environments. After scene analysis was done, we classify the server nodes that have the same processing capability of the CPU into a small cluster sub-network in SDN with the classify SDN network server nodes.With the help of SDN control plane and forwarding layer separation characteristics, we create a central controller whose work is to monitor the entire network the network, and there is a sub-controller int each sub-network. Ant colony algorithm runs in each controller.When a job comes, the central controller first finds the corresponding sub-network based on the job demand for CPU performance and send the job to the sub-network controller,then in the sub-network, the ant colony algorithm can calculate the minimum load Link Road with the real-time network load provide by the controller.In the end,the controller send forwarding policy to the switch. We build the SDN environment with the Mininet and OpenDayLight controllers, and through the algorithm experiment with other load-balancing algorithm ,the experimental results proves: the optimized ant colony algorithm can achieve better load balancing in each link which results in improving the utilization of resources and enhancing the usability of the system.

1. Introduction

With the rise in recent years and the explosive growth of the Internet, e-commerce, big data and other information services, in proportion to the number of users increases, which makes the current system increases, the number of system resources and species are increasingly more, and how to take advantage of this valuable system to manage these resources become a hot topic in today's iT industry. Wherein the load balancing technology is an important part of the research. Load balancing directly affects the availability of applications and system services. Load balancing purposes is to improve resource utilization, thereby improving the throughput of the system and shorten the response time of the system and improve the user experience, to avoid the situation that some part of the server and the server load idle.

SDN (Software-defined networking), namely, software-defined network, is the clean slate project by Stanford University study group[3] proposed a new network architecture innovation, its core technology through the OpenFlow network device control and data planes separated open to [4-6], enabling flexible control of network traffic, innovation and application of the core network provides a good platform. When the data stream arrives at a time Open Flow switch, the switch will first query Flow table Flow table query forwarding path. In SDN network, each OF switch has a flow meter, which is mainly a collection of process data streams for all actions, such as looking up and forward, the main flow table contains headers, counters and actions three fields in which actions field is represented forwarding rules, and flow meters is updated once every so often. Through the above know the SDN controller determines the traffic forwarding rules, we can achieve load balancing algorithm in the controller, so that load balancing of each link.

2. The Definition and architecture of SDN

SDN (software-defined networking)[7] is a new computer network architecture, a project about UC Berkeley 2008 years of Stanford University and its origin. In SDN through OpenFlow protocol layer of the network is divided into forwarding, control and service layers , the basic architecture shown in Figure 1. Communication between each layer, service layer and control layer of the communication interface via an interface called the NBI, and forwarding the control interface layer between the layers of southbound interface.

The main characteristics of SDN network architecture differs from traditional network control layer and data layer separation, although the control layer and data layer heat traditional IP network equipment is isolated, but its physical structure or tightly coupled mode independent of one another, and between the two communicate via the bus. SDN is used in the control logic centralized mode, network equipment overall control module pulled out, concentrated SDN network node - controller. Such control layer and data layer can not be in the same position, but the most important is the global SDN controller node can control the network, ensuring effective forward data streams OF switches, but also improve the SDN network on the physical structure scalability, enhanced network flexibility

3. Improvement of CAP algorithm

Above we introduced to the job classification, here mainly according to the requirements of CPU classification. The same CPU for the need for a class, labeled as J class (j=1,... C). Under normal circumstances, the operation is divided into static and dynamic operation. Static job refers to the operation does not need to be calculated directly from the memory or hard drive to read, for example, a picture of the web server to get the request is a static job. Since the execution time of the static request is proportional to the rate of the return result of the request, we can classify it by the size of the flow. And the demand for CPU is very little. When a static operation request arrives, classification module from operation parameters in parsing out the job request file name and according to the type of file query operation type table work, then the the type parameter is added to the operation parameters, as behind the load distribution and scheduling with.

Dynamic job is CPU intensive, the content of the request is not able to know in advance, may need to get through the calculation or from the database. For dynamic job classification, here mainly through the operation of the time to classify. The number of job execution time and CPU performance and CPU core core, CPU performance is by the number of instructions per second said MIPS , number of CPU instructions and dynamic job execution, we can estimate the dynamic job about execution time.

Server classification is also carried out in accordance with CPU. Here to each type of server J defines a weight W (J), said the J class server compared to other types of servers, the average CPU critical value (J), the minimum critical value of all server type Nm can be expressed as follows:

$$N_m = \min\left\{N_c(1), \ldots, N_c(j), \ldots, N_c(C)\right\} \tag{1}$$

The minimum critical value is related to the high demand for CPU, so we can define the weight of W (J) as the server type J:

$$w(j) = \frac{N_c(j)}{N_m}, j \in \{1, \ldots, C\} \tag{2}$$

In this paper, the link bandwidth usage is used to represent the link load, and the traffic information in the SDN network is monitored by IPEF software, and the bandwidth usage of the link in the SDN network is obtained. The action table, the switch, and after analysis, can know the bandwidth utilization rate is high, link load, link overhead becomes large, delay and packet loss rate also increased, the load will lead to a network of switches between the "distance" of different, the controller integrated in ant colony algorithm to calculate from the source node to the destination node between the shortest path, and also is the least loaded path through the flow table and notice of

data flow forwarding strategy of switch according to the corresponding flow forwarding data flow, which can reach the network in each link of the load is relatively balanced, the network more stable.

The ant colony algorithm is defined as the link in the SDN network, which is defined as the link load. When the link usage rate is higher, the pheromone concentration is lower, the relationship between the pheromone concentration and the link utilization is given.

$$\tau_{ij}(t) = k \circ \frac{1}{load_{ij}(t)} \qquad (3)$$

Which represents the t time node i and node j between the link between the pheromone, which represents the link between the node i and node j, K for the link between the pheromone and the conversion constant.

The probability formula of ant m from node i to node j in ant colony algorithm is as follows.

$$p_{ij}^{m} = \frac{[\tau_{ij}(t)]^{\alpha} \bullet [\eta_{ij}(t)]^{\beta}}{\sum_{j \in L} [\tau_{ij}(t)]^{\alpha} \bullet [\eta_{ij}(t)]^{\beta}} \qquad (4)$$

Among them, the set L is ants m may want to choose a route set, said routing link pheromone of ant the influence factor, the greater the value, indicating that ants of information element more depend on, the smaller the value of, said the pheromone of ant smaller. Said the ants to explore new path of the opening rounds, the greater the value that ants are increasingly interested in the new path, the value smaller said ants on a new path to less interested in [17], in order,. In which the ant to explore the new path of the heuristic function, the formula is as follows.

$$\eta_{ij}(t) = \frac{1}{load_{ij}} \qquad (5)$$

Which represents the link between the node i and node j load, and inversely proportional, the greater the smaller, the smaller the ant wants to explore the new path.

4. Simulation experiment and result analysis

Experimental test environment for the Intel 14.04 core i5 Ubuntu processor virtual machine configuration for 1GB memory, Gigabit Ethernet card and SDN OpenDayLight controller, combined with the use of Mininet to build the SDN network environment. Mininet software can quickly define complex network topology, and simulate the different configuration of the node servers and switches, set the link bandwidth. By using the OpenDayLight controller, the ant colony algorithm can be embedded into the controller, and the simulation experiment is carried out.
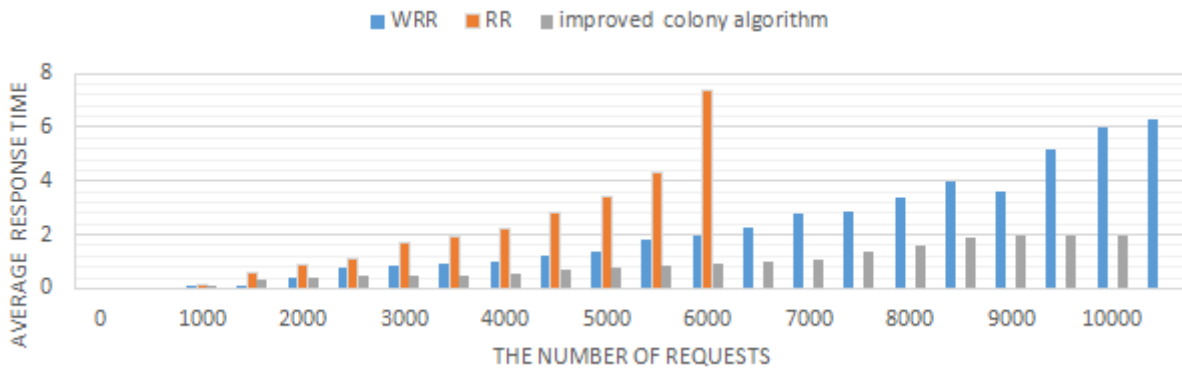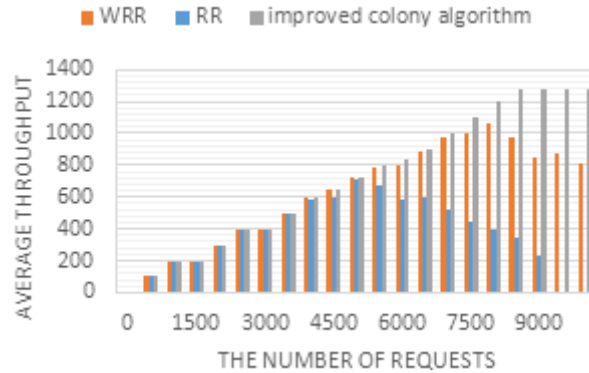

Figure 1. response time under different algorithms

Figure 2. throughput under different algorithms

Figure 2 shows the relationship between the number of jobs and the average response time of the system using different algorithms. From the figure we can draw less than 3000 the number of jobs, three algorithms of system response time difference, when the number of jobs for more than 3000, the gap becomes big, obvious static load balancing algorithm WRR due to not considering the state of nodes in the system, the response time is the longest, RR algorithm also did not consider the system load information, and improved ant colony algorithm according to real-time link load and multi classification operation, which makes the system load balancing, to raise the utilization rate of the resources in the cloud, response time than the short RR and WRR algorithm system. Typically, the shorter the system response time, the more work can be processed at the same time.

5. Summary

In a dynamic, multidimensional cloud computing in virtual environment, load balancing is used to improve the utilization rate of resources and shorten the response time and improve system throughput and ensure effective guarantee system stability and availability of the system. An ant colony of job classification adaptive load balancing algorithm based on, in the load distribution, not only based on the specific needs. Moreover, it analyzes the system nodes link real-time load information is presented in this paper. Due to the different working demand for resources is also different, we classify according to the operation of the CPU needs to work and give different kinds of jobs with different priorities, priority is higher, the probability of resource allocation is higher. In the end, it can avoid the occurrence of overload and some nodes are idle. So that we can make full use of resources, taking into account the various resources of specific load, the load distribution, the cloud environment all resources are load balancing, improve resource utilization rate, shorten the system response time.

References

[1] M. Dong, T. Kimata, K. Sugiura, and K. Zettsu, "Quality-ofExperience (QoE) in emerging mobile social networks," IEICE Transactions on Information and Systems, vol. 97, no. 10, pp. 2606–2612, 2014.

[2] D. Zhang, H. Huang, J. Zhou, F. Xia, and Z. Chen, "Detecting hot road mobility of vehicular Ad Hoc networks," Mobile Networks and Applications, vol. 18, no. 6, pp. 803–813, 2013.

[3] H. Kim and N. Feamster, "Improving network management with software defined networking," IEEE Communications Magazine, vol. 51, no. 2, pp. 114–119, 2013.

[4] S. Sezer, S. Scott-Hayward, P. Chouhan et al., "Are we ready for SDN? Implementation challenges for software-defined networks," IEEE Communications Magazine, vol. 51, no. 7, pp. 36– 43, 2013.

[5] M. Dong, H. Li, K. Ota, and J. Xiao, "Rule caching in SDNenabled mobile access networks," IEEE Network, vol. 29, no. 4, pp. 40–45, 2015.

[6] M. Channegowda, R. Nejabati, and D. Simeonidou, "Softwaredefined optical networks technology and infrastructure: enabling software-defined optical network operations [invited]," Journal of Optical Communications and Networking, vol. 5, no. 10, Article ID 6645122, pp. A274–A282, 2013.

[7] M. K. Shin, K. H. Nam, and H. J. Kim, "Software-defined networking (SDN): a reference architecture and open APIs," in Proceedings of the International Conference on ICT Convergence (ICTC '12), pp. 360–361, IEEE, October 2012.