Latency-Aware Routing with Bandwidth Assignment for Software Defined Networks

Qiongyu Zhang*, Liehuang Zhu*, Meng Shen*†, Mingzhong Wang‡, Fan Li*
*Beijing Engineering Research Center of High Volume Language Information Processing and Cloud Computing Applications, School of Computer Science, Beijing Institute of Technology, P. R. China
†Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, P. R. China
‡Faculty of Arts and Business, University of the Sunshine Coast, Queensland, Australia
Email: {biterzqy, liehuangz, shenmeng, fli}@bit.edu.cn; mwang@usc.edu.au

***Abstract*—Reducing the flow latency is of great importance in traffic management, which benefits both service providers and end users. Routing design and flow scheduling are typical ways to improve the flow transmission efficiency. However, existing studies usually consider them separately, due to the complexity of joint consideration. Here, we propose a latency-aware routing scheme with bandwidth assignment in the Software Defined Networks, which can efficiently reduce the flow latency with a moderate complexity, to combine the routing and scheduling together.**

I. Introduction

Flow latency is crucial for network service efficiency, since the low flow latency is the basic of high service efficiency. Both routing and flow scheduling are the common ideas to reduce latency. However, in the existing researches, they are independent with each other. The main reason of this status is the restriction of the computing complexity of the solution [1][2]. Unfortunately, cooperation between flow scheduling and routing can obtain higher flow transmission efficiency. In this paper, we try to find the trade off between the forwarding latency and the solution complexity based on the divide-and-conquer strategy.

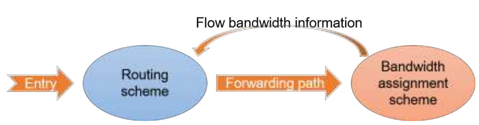

Fig. 1: Data flow between routing and flow scheduling

We propose a latency-aware routing scheme to find out a low latency forwarding path by utilizing the global flow bandwidth information, and a bandwidth assignment scheme, which is one kind of flow scheduling, to update the global flow bandwidth information based on the forwarding path from the latency-aware routing scheme. The Fig. 1 shows the relationship between routing and flow scheme in our solution.

The main contributions of this paper are: 1. Based on the divide-and-conquer strategy, joint the routing and flow scheduling to reduce flow latency without high computing complexity. 2. Propose a latency-aware routing scheme reduce the network-wide latency. 3. Propose a bandwidth assignment scheme based on the latency-aware routing scheme to improve the effect of the routing scheme.

The research work reported in this paper is supported by National Natural Science Foundation of China under Grant No. 61402037.

II. Latency-aware routing with bandwidth assignment scheme

A. Latency-aware routing scheme

In the routing scheme, our goal is to find out a forwarding path to reduce both the latency of current flow and the latency influences of existing flows. When routing for flow $F_c = \{M_c, B_c, T_c, L_c, P_c\}$, where M_c is the data size of the flow F_c, and B_c is the allocated bandwidth of the flow F_c, T_c is the start transferred time of the flow F_c, L_c is the deadline of the flow F_c, P_c is the forwarding path of the flow F_c. The goal of latency-aware routing scheme is to find a forwarding for F_c, which meets following equation:

$$\begin{aligned} &\min \quad (d_c + \sum_{f_i \in S} w_i(d_i' - d_i)) \\ &s.t. \quad d_c < L_c \end{aligned} \tag{1}$$

where S means the set of existing flows, which occupy some links with P_c. d_i, d_i' are the latency of flow f_i before T_c and after T_c respectively. d_c is the latency of f_c. and

$$w_i = \begin{cases} \frac{L_i - d_i'}{L_i} & L_i > d_i' \\ 1 & L_i \le d_i' \end{cases}$$

Assuming that the bandwidth of each link will be equally allocate to all of the flows on the link. In the candidate path p, the latency of current new flow is :

$$d_c = \frac{M_c}{B_c} = \frac{M_c}{\min_{e \in p} \frac{b_e}{|F_e|}}. \tag{2}$$

where B_c is the bandwidth of current new flow, b_e is the bandwidth capacity of link e, and F_e is the set of the flows on the link e.

For every link e in the path p, the latency of the affected flow f_i is:

$$d_i' = \frac{M_i - B_i \times (T_n - T_i)}{B_i'} = \frac{M_i - B_i \times (T_n - T_i)}{min \quad (\frac{b_e}{|F_e|}, B_i)} \tag{3}$$

where T_n is the current time. The latency of affected flow f_i before the new flow entered can be calculated as

$$d_i = \frac{M_i}{B_i} - (T_n - T_i) \tag{4}$$

978-1-4673-8590-9/15/$31.00 ©2015 IEEE

where B_i' is the bandwidth of existing flows f_i and B_i is the old bandwidth of flow f_i.

The latency-aware routing scheme problem is a simple optimal problem without constraints. We can utilize the branch and bound strategy to solve it. and treat $f(p) = d_c + \sum_{f_i \in S} w_i(d_i' - d_i)$ as the utility function of the forwarding path p , $d_c < L_c$ as the pruning condition.

B. Bandwidth assignment scheme

Based on the forwarding path of f_c from the latency-aware routing scheme, we need to assign bandwidth to all of the flows influenced by f_c and f_c itself. The goal of bandwidth assignment scheme is to find a assignment, denoted by $\mathcal{A} = \{b_i | \forall f_i \in S \cup \{f_c\}\}$, for F_c, which meets following equation:

$$\begin{aligned} \min \quad & \sum_{f_i \in S \cup \{f_c\}} \frac{M_i - (T_n - T_i) \times B_i}{b_i} \\ s.t. \quad & \sum_{f_i \in e} b_i < b_e, \forall e \in p_i \wedge f_i \in S \cup \{f_c\} \end{aligned} \tag{5}$$

where T_n is the current time. In the constraint condition, b_i is the bandwidth of the flow f_i, b_e is the bandwidth capacity of link e, p_i is the forwarding path of the current new flows f_c, i.e. the output of latency-aware routing scheme.

Actually, the bandwidth assignment scheme problem is an optimal problem with constraints. We can utilize the genetic algorithm with constraints to find out the optimal solution. Genetic algorithm is a common optimal algorithm, adding the constraints to the genetic select phase to solve the optimal with constraints is a general method.

III. Scheme Implement

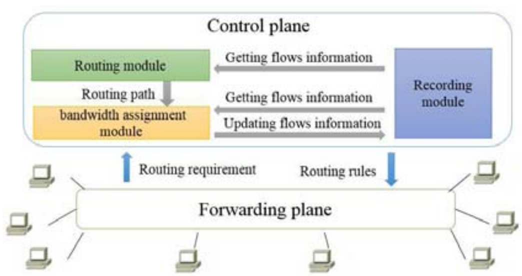

Fig. 2: System architecture

The architecture of our system is shown in Fig. 2, which is based on the OpenFlow framework. When the controller receives the routing request, the routing module will be called to get a routing path, utilizing the flows information from recording module. Based on the routing path, the bandwidth assignment module will be activated and output the bandwidth allocate for the new flow requiring routing and the affected flows. Finally, the flows information of current networks will be updated by the recording module.

Since the computing complexity of both the routing scheme and the bandwidth assignment scheme are $\Theta(n^2)$, the latency-aware routing with bandwidth assignment scheme is a polynomial-time scheme.

IV. Performance Experiments

We evaluate the efficiency of latency-aware routing with bandwidth assignment scheme with mininet and Floodlight controller. As to the routing scheme, Fig. 3 shows the relationship between flow latency and deadline. The spots are the distribution of flow latency, and the curve is the deadline, and the $M(F)$ is the size of flow. In comparison, the concentration zone of time cost in latency-aware routing scheme are below the tradition scheme. This means that the worse solution in our scheme is still better than the best of traditional routing scheme. What's more, we can see in all of the figures, the blue spots above the cures are more than the red spots. And the file size is larger, this phenomenon is more clear. Table. I shows the time cost reduction rate using assigning bandwidth.

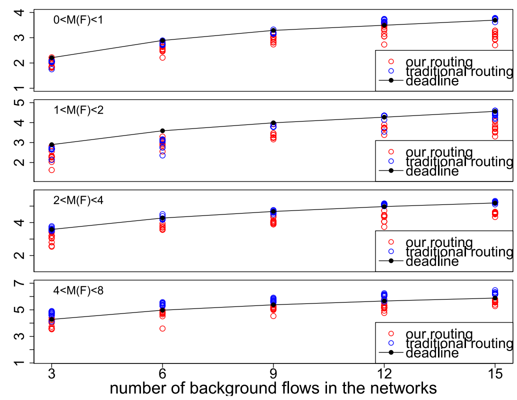

Fig. 3: Deadline comparison under different flow parallel level

TABLE I: Time cost reduction rate using bandwidth assignment

$0 < M(F) < 1$	$1 < M(F) < 2$	$2 < M(F) < 4$	$4 < M(F) < 8$
10.5%	3.2%	4.6%	2.2%

V. Conclusions

In this paper, we propose a latency routing with bandwidth assignment scheme for Software Defined Networks. Based on the divide-and-conquer strategy, we joint the routing and flow scheduling to reduce flow latency without high computing complexity. What's more, we propose a latency-aware routing scheme to reduce both the transmission latency of the current flow and the influence of other existing flow's latency. We also build a bandwidth assignment construction to reallocate the bandwidth of the flows for further reducing data delivery latency.

References

[1] L. Wang, F. Zhang, K. Zheng, et al, *Energy-Efficient Flow Scheduling and Routing with Hard Deadlines in Data Center Networks*. 2014 IEEE ICDCS 2014, 2014:248-257.

[2] S. Stiller, A. Wiese, *Increasing Speed Scheduling and Flow Scheduling. Algorithms and Computation*. 21st International Symposium, ISAAC 2010, 2010: 279-290.