View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Creative Repository of Electro-Communications

国立大学法人電気通信大学 / The University of Electro-Communications

Imbalance state resolving considering flow types

著者（英）	Takuna Kaiwa, Nattapong Kitsuwan
journal or publication title	IEICE Communications Express
volume	9
number	2
page range	66-71
year	2020
URL	http://id.nii.ac.jp/1438/00009639/

doi: 10.1587/comex.2019XBL0132

Imbalance State Resolving Considering Flow Types

Takuna Kaiwa and Nattapong Kitsuwan
Department of Computer and Network Engineering,
The University of Electro-Communications

a) *tkaiwa@uec.ac.jp*

Abstract: This paper proposes a scheme to resolve the imbalance state in a network in which flow types, mice and elephant flows, are considered. A combination of link utilized rate and transmission delay of each link are considered as a link cost. In the proposed scheme, the load imbalance state is resolved by dividing the elephant flow into several subflows and injecting each subflow into multiple paths. The maximum utilization rate of the proposed scheme decreases 38.9%, compared to a conventional scheme.

Keywords: Elephant flow, Mice flow, routing, scalability
Classification: Network management/operation

References

[1] S. Hegde, S. G. Koolagudi, and S. Bhattacharya, "Scalable and fair forwarding of elephant and mice traffic in software defined networks," Computer Networks, vol. 92, pp. 330-340, 2015.
[2] X. Liu , Y. Xiang, "Disturbance based Dynamic Load Balancing Routing Mechanism under SDN," Proc. International Conference on Computer Science and Network Technology (ICCSNT), 2016.
[3] Jing Liu, Jie Li, Guochu Shou, Yihong Hu, Zhigang Guo and Wei Dai "SDN Based Load Balancing Mechanism Elephant Flow in Data Center Networks," Proc. International Symposium on Wireless Personal Multimedia Communications (WPMC), 2014.
[4] N. Kitsuwan, E. Oki, N. Kishi and T. Miki, "Nonreservation Fair Bandwidth Allocation Based on Dynamic Replacement in Optical Burst Ring Networks," IEEE/OSA J. Optic. Commun. Netw., vol. 1, iss. 5, pp. 416-427, Oct. 2009.
[5] K. Leung, V. O.K. Li, and D. Yang, "An Overview of Packet Reordering in Transmission Control Protocol (TCP): Problems, Solutions, and Challenges," IEEE Transactions on Parallel and Distributed Systems, vol. 18, iss. 4, pp. 522-535, 2007.

1 Introduction

In computer network, flow type is classified into an elephant flow and mice flow. The elephant flow is an enormously large continuous flow, while the mice flow is a small flow. The number of elephant flows in the network is

© IEICE 2019

less than 10% but the elephant flow occupies 40% of the amount of traffic in the network [1]. A load balancing-based routing method must be adopted for each flow type to avoid link congestion and to protect packet processing delay for the mice flow.

Load balancing considering flow types in a network has been studied. An algorithm that combines both a static and dynamic load balancing was introduced [2]. The static load balancing computes the routing in case of no imbalance state. Once the imbalance state occurs, the dynamic load balancing put the elephant flow to a low priority queue. The elephant flow is transmitted when there is no remaining flow with high priority queue in a network node. Head-of-line blocking occurs for the mice flow at the output port when the elephant is transferring. It results in long tail latency for the mice flow. In [1], an elephant flow is split into tiny flows according to the ratio which is inversely proportional to the link load. Each tiny flow is transmitted to its destination with different paths. Both elephant and mice flows are routed with the same routing. As a result, the head-of-line blocking still occurs.

This paper proposes a scheme to resolve the imbalance state by reducing the network congestion and head-of-line blocking for the mice flow by using different path calculations for both the elephant and mice flows in a TCP/IP network. The proposed scheme calculates a path cost based on a transmission delay and link utilization rate. The elephant flow is divided into multiple subflows. Each subflow is transmitted in a different path. Computer simulation shows that 38.9% reduction of the maximum utilization in the proposed scheme is achieved.

2 Related works

2.1 Scalable and fair forwarding of elephant and mice traffic in network

The elephant flow is divided by the number of routes and processed using a source routing method as same as the mice flow [1]. Therefore, the load on the original route that the elephant flow passes before dividing the elephant flow can be reduced. A source node adds the routing to the destination. The controller only updates the routing for the divided elephant flows at the source node. The routing at the source node is not flexible.

2.2 Disturbance based dynamic load balancing

The load balancing process is split into two parts, static load balancing routing (S-LBR) and disturbance based rerouting (D-LBR) [2]. S-LBR determines a static routing for a request by considering the normalized residual bandwidth W_{ij}^{ω} and normalized link utilization rate η_{ij}^{ω}, which are defined as follows.

$$\eta_{ij}^{\omega} = \frac{(\eta_{ij} - \eta^{min})}{(\eta^{max} - \eta^{min})} \quad (1)$$

$$W_{ij}^{\omega} = \frac{(W_{ij} - W^{min})}{(W^{max} - W^{min})}, \quad (2)$$

© IEICE 2019

where W_{ij}, W^{max}, and W^{min} are a residual bandwidth of link (i, j), the maximum, and minimum residual bandwidth of entire links in the network, respectively. η_{ij}, η^{max}, and η^{min} represent a link utilization rate of link (i, j), the maximum, and minimum link utilization of entire links in the network, respectively.

S-LBR finds paths with the minimum hop count from a source to a destination. η_{ij} is considered as a link cost. A path with the minimum cost is selected as a transmission path. If there are several paths, the best path is selected by considering the distance and the residual bandwidth.

D-LBR, which makes adjustments to balance the load by rerouting the paths in S-LBR when the following condition is satisfied.

$$\overline{\eta(t)} \times \varepsilon < \eta_{ij}(t), \tag{3}$$

where $\overline{\eta(t)}$ represents the average of all link utilization rates at time t, and $1 < \varepsilon < 2$. First, a flow is categorized into elephant flow and mice flow. Second, a flow that passes a congestied link is detoured to reduce the link utilization rate of the congested link by selecting a path with the minimum hop on the detoured path. If there is no candidate for the detoured path, the mice flow is placed in a high priority queue, and the elephant flow is placed in a low priority queue. The mice flow may be routed on the same path with the elephant flow. The transmission delay for the mice flow is long.

3 Proposed scheme

The proposed scheme performs the different path computations between the elephant and mice flows. The elephant flow is split into multiple subflows. The subflows are forwarded along the computed paths as a max-min fairness policy [4]. The mice flow is forwarded along the shortest path. As a result, the proposed method can suppress the rapid increase of the maximum link utilization rate caused by the elephant flow and a congested links. A packet reordering problem occurs in the proposed scheme since the elephant flow is split. This problem can be solved by state reconciliation or threshold adjustment algorithms [5].

3.1 Rerouting mice flow and elephant flow

The proposed method considers the normalized transmission delay, d^{ω}_{ij}, and normalized link (i, j) utilization rate , η^{ω}_{ij} as a cost. d^{ω}_{ij} at time t is defined as follow.

$$d^{\omega}_{ij}(t) = \frac{(d_{ij}(t) - d^{min})}{(d^{max} - d^{min})}, \tag{4}$$

where d^{min} and d^{max} are the minimum and maximum transmission delays of the entire network link, respectively. The cost of link (i, j), l^{c}_{ij}, is defined as follow.

$$l^{c}_{ij} = \alpha \times d^{\omega}_{ij} + (1 - \alpha) \times \eta^{\omega}_{ij}, \tag{5}$$

© IEICE 2019

where α is a constant value and $0 \leq \alpha \leq 1$. We use two processes to decide the rerouted flow.

1. Selecting of flow rerouting
   The proposed method specifies the forwarding route for each flow passing through a congested link. A variance of link utilization rate for traffic pattern q, σ_q^2, is used as a load splitting value of the entire network. The load splitting value of the entire network is calculated as follows.

$$\sigma_q^2 = \frac{\sum_{k=1}^{K}(\eta_k(q) - \overline{\eta(q)})^2}{K-1}, \tag{6}$$

   where K represents the number of links. A traffic pattern q that achieves the minimum load splitting value σ_q^2 is selected to be rerouted.

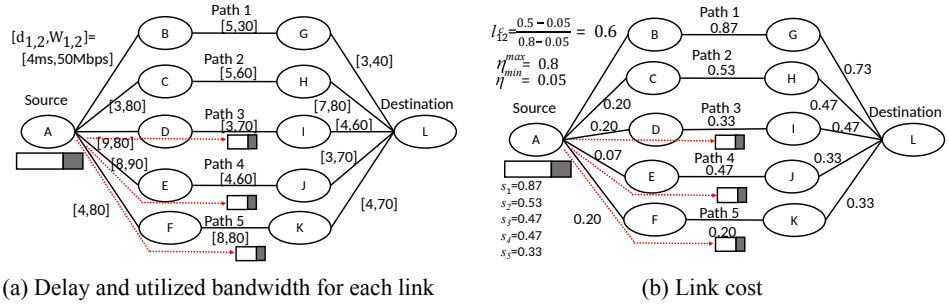
2. Flow rerouting
   If the selected flow is the mice flow, forward the flow along the shortest path. Otherwise, split the flow as a max-min fairness policy and forward the split flows along the computed paths.

3.2 Flow splitting algorithm for elephant flow

A flow requests bandwidth B from a source to a destination. Let P be a set of splitting paths for the requested flow, where $|P|$ is given. B is split into $|P|$ paths, each path takes the bandwidth b_p, where p is an index of element in P. A residual bandwidth for path p, which is the minimum W_{mn} for links (m, n) on path p, is defined as $W^p = \min_{(m,n)} W_{mn}$, where $W_{mn} = C_{mn} - (\eta_{mn} \times C_{mn})$ and C_{mn} is a capacity of link (m, n). The process of flow allocation with flow splitting is as follows.

- Step 1: Compute l_{ij}^c for all links in the network.
- Step 2: Find all possible set of paths, Q, from source to destination. Compute cost of all paths in Q, $s_q = \max_{(i,j)}\{l_{ij}^c\}$, where link (i, j) is on path q and $q \in Q$.
- Step 3: Among the possible paths, select k paths that have the minimum value of s_q. Put those k paths into P, and sort the elements in P in ascending order.
- Step 4: If $\sum_{p \in P} W^p < B$, splitting process fails.
- Step 5: Set $p = 0$ and $a = |P|$.
- Step 6: If $B > 0$, set $b = \left\lfloor \frac{B}{a} \right\rfloor$. Otherwise, finish the process.
- Step 7: If $W^p \geq b$, set $b_p = b$. Otherwise, set $b_p = W^p$.
- Step 8: Assign b_p to path p.
- Step 9: $B = B - b_p$.

© IEICE 2019



(a) Delay and utilized bandwidth for each link (b) Link cost

Fig. 1. Example of rerouting elephant flow of proposed scheme.

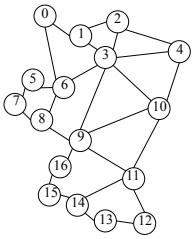
- Step 10: Increase p by one, decrease a by one and repeat from step 6.

Figure 1 shows an example of flow splitting for the elephant flow of the proposed scheme. $\alpha = 0$ is assumed. And all link's capacity are 100 Mbps. It is assumed that five paths, paths 1 to 5, are available. Figure 1(a) shows remain bandwidth of every link. Figure 1(b) shows the computed link cost of every link. s_1 to s_5 are 0.87, 0.53, 0.47, 0.47, and 0.33, respectively. Assume that $B = 100$ Mbps is split into three paths. The algorithm selects $P = \{\text{Path } 5, \text{Path } 3, \text{Path } 4\}$ as split paths. b on each path is 33 Mbps. The algorithm assigns 33 Mbps to s_5 and s_3, and 34 Mbps to s_4.

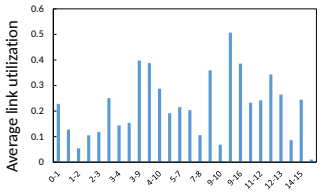
4 Performance evaluation

We evaluate the maximum utilization of the proposed scheme comparing to the conventional scheme by simulation. The German17 topology which consists of 17 nodes and 26 undirected links, as shown in Fig. 2(a), and the USIP topology which consists of 24 nodes and 41 undirected links, as shown in Fig. 3(a), are used in the simulation. 30 flows are generated. Three of them are the elephant flows and 27 of them are the mice flows. The size of the elephant flow is assumed between 100 Mbps and 200 Mbps, the flow size that less than 100 Mbps is assumed as the mice flow. Source and destination of each flow are randomly selected. Each link capacity is randomly set between 500 Mbps and 800 Mbps. The number of simulation is repeated for 1,000 times. If the elephant flow is split into subflows, the maximum size of each subflow is assumed to 30 Mbps. The subflows are split as a per-packet manner. We determine the path of the flow using SLBR before network operation, and perform load balancing when the imbalance state occurs. The strict source and record route is used in the simulation. Packet reordering is omitted.

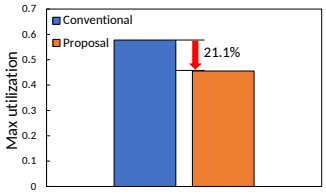
© IEICE 2019



(a) German17 topology

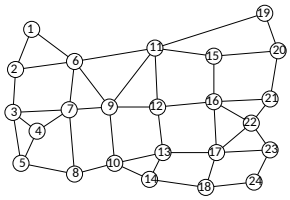


(b) Average link utilization before load balancing

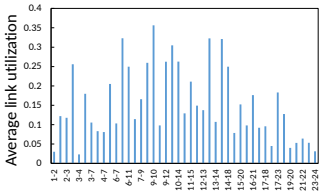


(c) Maximum utilization

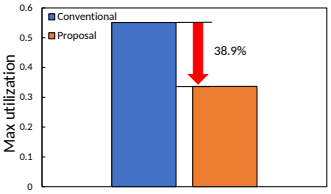
Fig. 2. Simulation with German17



(a) USIP topology



(b) Average link utilization before load balancing



(c) Maximum utilization

Fig. 3. Simulation with USIP

Figure 2 and 3 compare the maximum utilization rates of the conventional and proposed schemes in the German17 and USIP topologies, respectively. Figures 2(b) and 3(b) show the number of congestions of each link. A link with the high number of congestions implies that the bottleneck easily occurs. Figures 2(c) and 3(c) show that the maximum utilization rate of the proposed scheme achieves 21.1% and 38.9% reduction for German17 and USIP, respectively.

5 Conclusion

A scheme to resolve the imbalance state in a network was proposed. In this scheme, a flow to be rerouted is determined by a traffic pattern. If the rerouted flow is the mice flow, it is rerouted through the shortest path. If the rerouted flow is the elephant flow, the flow is split as max-min fairness. Each split flow is routed on a different path. The simulation results showed that the proposed scheme reduces 38.9% of the maximum bandwidth for our examined topologies, compared to the conventional scheme.

© IEICE 2019