Highly Available Dynamic Deployment Algorithm Based on Leader-Follower Mode in SDN

Peng Kuang¹, Yiwen Qiu², Hailan Kuang²* and Xinhua Liu²
¹Wuhan Middle School, Wuhan, China;
²School of Information Engineering, Wuhan University of Technology, Wuhan, China
Email: *Kuanghailan@whut.edu.cn

Abstract. Controller is the core of the Software Defined Network (SDN) network architecture, how to ensure the stability of the network operation process by dynamically deploying controllers is one of the important tasks of the SDN network. This paper proposes a highly available dynamic deployment algorithm based on the leader-follower mode. First, leader and follower controllers are determined through a neighborhood leader controller election algorithm based on Paxos protocol; secondly, the leader controller performs a time optimal allocation algorithm based on load constraint to determine the distribution plan of the switches under the failed controller; finally, the slave controllers perform the switch migration operation according to the distribution result to complete network failure recovery. The simulation results show that the proposed algorithm ensures the controller not overloaded during the fault recovery, and has a lower controller average response delay than other similar algorithms.

1. Introduction
Software Defined Network (SDN) was proposed as a new type of network architecture in 2008[1], it separates the control power from the network equipment and sends it to the centralized controller for unified management. Controller in the SDN network system, the failure of the controller will cause the switches it manages to fail in properly forward data packets, so how to deal with the controller failure during the operation of the SDN network is a crucial issue in SDN. The controller deployment algorithm was first proposed by Heller et al.[2] in 2012. In order to reduce the transmission distance between the switch managed by the failed controller and the relocated controller after failure recovery, literature [3] completed the deployment of the controller by optimizing the distance between the controller and its backup controller during network initialization. Hock et al.[4] proposes a shortest path algorithm (SP) in the case of a controller failure in a network system. Literature [5] proposes a proximity-based heuristic algorithm (PBH) and a residual capacity-based heuristic algorithm (RCBH). The RCBH algorithm migrates the switches managed by the failed controller to the lightest controller in the current network system. This paper proposes a highly available dynamic deployment algorithm based on the leader-follower mode (HADD-LFM) to deal with the controller failure during the operation of the network system. It makes the following contributions: (1) proposed a neighborhood leader controller election algorithm based on Paxos protocol to determine the leader and follower controllers. (2) designed a time optimal allocation algorithm based on load constraint, and let the main controllers execute this algorithm to complete the distribution of switches under the failed controller. (3) For the case that the fault controller returns to normal operation, we have designed a network map restoration mechanism to re-map the switches and the fault controller.

2. System model
In this paper, the SDN network topology is represented as $G(S,E,C)$, where $S$ denotes the switch set of the network topology, $E$ denotes the link set of the network topology, and $C$ denotes the controller
set of the network topology; so that the size of the switch set is $N$, and the switches are respectively $s_0, s_1, \ldots, s_{N-1}$. The size of the controller set is $K$, and each controller is $c_0, c_1, \ldots, c_{K-1}$. The meaning of each parameter is shown in Table 1.

### Table 1. The Meaning of Parameters

| symbol | definition |
|--------|------------|
| $c_i$  | the $i$-th controller |
| $s_j$  | the $j$-th switch |
| $U_i$  | the rated load of $c_i$ |
| $L_i$  | the current load of $c_i$ |
| $p_{ul}$ | the load limit, $p_{ul} \in [0,1]$ |
| $a_{ij}$ | the number of Packet-In messages sent from $s_j$ to $c_i$ in unit time |
| $\delta$ | average computational overhead for the controller to process a single Packet-In message |
| $u_i$  | the effect of $s_j$ on the load of $c_i$ in unit time |
| $d_{ij}$ | the control link distance between $s_j$ and $c_i$ |
| $S_i$  | the switches managed by $c_i$ |
| $N_i$  | the size of the switches managed by $c_i$ |
| $D_i$  | the neighborhood controllers of $c_i$ |
| $w_{ij}$ | the working status of $c_i$ |
| $r_{ij}$ | the mapping between $c_i$ and $s_j$ |

The impact on the load per unit time depends on the number of Packet-In messages, as shown in (1).

$$u_{ij} = \delta \times a_{ij}$$

The goal of the algorithm in this paper is to migrate the switches managed by the failed controller to other valid controllers in the event of a controller failure in the network system, and to satisfy the capacity constraints of the move-in controller during the fault recovery process. At the same time, the transmission distance between the switch and the immigration controller is also minimized. Which can be described as (2), where $w_{ij} = 0$, that is, $c_i$ is in a fault state.

$$\begin{aligned}
\text{min} & (\sum_{c_i \in D_j} \sum_{s_j \in D_i} r_{ij} \times d_{ij}) \\
\text{s.t.} & \sum_{c_i \in D_j} \sum_{s_j \in D_i} r_{ij} = N_i \\
& c2: L_j + \delta \times \sum_{a \in D_i} r_{ij} \times a_{ij} \leq p_{ul} \times U_j, \forall c_j \in D_i
\end{aligned}$$

Since the migration operation can be completed in a short period of time, this article assumes that the request rates of the switches before and after the migration will not change much. Therefore, the condition $c2$ of (2) can be expressed as shown in (3), that is $a_{ij} = a_{1a}$, where $a_{1a}$ can be obtained by the synchronization of the information between the controllers.

$$c2: L_j + \delta \times \sum_{a \in D_i} r_{ij} \times a_{1a} \leq p_{ul} \times U_j, \forall c_j \in D_i$$

### 3. Algorithm

HADD-LFM algorithm first selects the leader controller from the fault controller’s neighbor when controller failure occurs. Then, the leader controller executes a time optimal allocation algorithm based on load constraint, and distributes the distribution results of the switches to each neighboring domain controller. When the neighboring domain controller receives the node allocation result, it performs a specific switch migration operation to complete node failure recovery. In addition, after the fault
controller returns to normal operation, the HADD-LFM algorithm implements the network mapping restoration mechanism, which allows it to take over the switches managed before the fault occurs, to reduce the load of each of the inbound controllers and to reduce the controllers’ average response delay.

3.1. Leader controller election in the neighborhood

In order to reduce the transmission delay between the switch and the immigration controller after the fault recovery, this paper only uses the neighboring controllers as a candidate to migrate to the controller set. However, the controller domain often has multiple neighbor controllers, so the consistency of the migration status between neighbor controllers must be maintained during fault recovery. This paper adopts the leader-follower mode to maintain the consistency of the migration status, which means the leader and follower controllers need to be determined by the election algorithm among the neighbor controllers of the faulty controller. At present, the algorithms used for the election problem mainly include the hegemonic algorithm\(^6\), the ring election algorithm\(^7\), and the Paxos protocol\(^8\). The Paxos protocol is coherent and effective to solve distributed consistency problems. Therefore, this paper proposes a leader controller election algorithm based on Paxos protocol. The controller's election status includes:

![Flow Chart of The Neighborhood Leader Election](image)

- **NoActive**: The controller is not in the election process;
- **Looking**: The controller is in the election process but not the leader or follower;
- **Leader/Follower**: The controller is the leader or follower controller.

**Figure 1. The Flow Chart of The Neighborhood Leader Election**
Looking: The controller is in the decision-making phase of the election process and serves as the proposer and acceptor roles defined by the Paxos protocol. The voting consists of the identification, the remaining load, the transaction number, the logic clock, and the election status.

Leader: The controller is responsible for completing the distribution of the switch set to be distributed;

Follower: The controller is responsible for performing specific switch migration operations. When controller fail, the neighbor controller of the fault controller will execute the neighboring leader controller election algorithm. The algorithm flow is shown in Figure 1.

3.2. Node failure recovery
The neighborhood leader controller election algorithm determines the domain controllers of the failed controller. The leader controller is responsible for calculating the allocation plan of the switch set managed by the fault controller, and the slave controller node is responsible for performing the migration operation of the switch set according to the distribution scheme delivered by the leader controller node. The specific steps for node failure recovery are shown in Figure 2:

3.3. Network mapping restore
Network mapping restoration refers to restore the mapping relationship between switches and the controller to the state before the controller failure occurs, to further reduce the average response delay of the controller. HADD-LFM persists the switches and corresponding status information, and updates the network status when it changes. When the failed controller recovers, the required information can be obtained from the file system to perform the network map restoration mechanism. However, even if the fault controller can obtain the switch information before the fault occurs, in order to ensure the security of the switch, the switches needs to be treated differently, this paper proposes an active four-stage smooth migration protocol based on the four-phase switch migration protocol proposed in [3] to ensure the security before and after the switch is migrated. The protocol is shown in Figure 3.

4. Performance Evaluation
In order to demonstrate the effectiveness of the HADD-LFM algorithm, it is compared with Sp[4] and RCBH[5] with system fault tolerance as the optimization goal. In this section, the fault controller is $C_{\text{deg}}$.
its nearest neighbor controller is $c_{op}$, and the controller with the lightest load in the current network system is $c_{rb}$, and the load of $c_{deg}$ is $L_{deg}$, and the current load of $c_{op}$ is $L_{op}$, and the current load of $c_{rb}$ is $L_{rb}$. The controller calculation overhead can reflect the load situation of each controller, and the controller average response delay can reflect the overall performance of the network system. Let $c_{deg} = c_5$, and set the fault occurrence point to 20s, and set the failure recovery point to 40s. The experimental environment is shown in Table 2, and the specific experimental results are as follows from A-D:

### Table 2. Experiment environment

| Configuration item          | Configuration information                |
|----------------------------|-----------------------------------------|
| CPU                        | Intel Core i7                           |
| memory                     | 4GB                                     |
| hard disk                  | 1TB                                     |
| operating system           | Ubuntu 14.04 LTS                        |
| controller                 | Floodlight 1.2                          |
| topology simulation tools   | Mininet 2.3.0d1                         |
| Openflow protocol          | Openflow1.3                             |
| traffic simulation tools    | Iperf                                   |
| network topology           | Internet2 OS3E                          |

A. $L_{deg} < U - L_{op}$

Let each controller's Packet-In message arrival rate be $c_0=100KB/s$, $c_1=110KB/s$, $c_2=120KB/s$, $c_3=130KB/s$, $c_4=140KB/s$, $c_5=150KB/s$.

From Figure 4, it can be seen that all the three algorithms do not have an overload controller after failure recovery when $L_{deg} < U - L_{op}$. In Figure 5, the average response delays of the controllers are not significantly increased, but the controller average response delay of the HADD-LFM algorithm is 12.5% lower than the SP algorithm and 20.6% lower than the RCBH algorithm.

B. $U - L_{op} < L_{deg} < U - L_{rb}$

Let each controller's Packet-In message arrival rate be $c_0=120KB/s$, $c_1=130KB/s$, $c_2=140KB/s$, $c_3=150KB/s$, $c_4=160KB/s$, $c_5=170KB/s$.
From Figure 6, it can be seen that the SP algorithm has an overload controller after failure recovery when $U - L_C < L_{dmg} < U - L_{h}$. In Figure 7, the average response delay of the SP algorithm is significantly higher than the other two algorithms, which is 61.6% higher than the RCBH algorithm and 84.4% higher than the HADD-LFM algorithm.

Let each controller's Packet-In message arrival rate be $c_0 = 150\text{KB/s}$, $c_1 = 160\text{KB/s}$, $c_2 = 170\text{KB/s}$, $c_3 = 180\text{KB/s}$, $c_4 = 190\text{KB/s}$, $c_5 = 200\text{KB/s}$.

It can be seen from Figure 8 that when $c = 1$, both the SP and RCBH algorithms have an overload controller after failure recovery, in Figure 9, the average response delay of the SP algorithm is 82.8% higher and the RCBH algorithm is 88.3% higher.

5. Conclusion

This paper proposes a highly available dynamic deployment algorithm based on the leader-follower mode, aiming at the problem of controller failure in the process of network operation. The algorithm first determines the leader and follower controllers through the neighbor leader controller election algorithm. Then, the leader controller performs a time optimal allocation algorithm based on load constraint to determine the migration controller of the switches managed by the failed controller. Finally, the migration of the switches is performed by the allocation result calculated by leader controller. Comparing with the SP and RCBH algorithms, the HADD-LFM algorithm has a slightly longer runtime, but it can effectively avoid the overload situation of the inbound controller, and the average response delay of the controller obtained is also lower than that of the other two algorithms.

References

[1] McKeown N, Anderson T, Balakrishnan H, et al. OpenFlow: enabling innovation in campus networks[J]. ACM SIGCOMM Computer Communication Review, 2008, 38(2): 69-74.
[2] Heller B, Sherwood R, McKeown N. The controller placement problem[J]. ACM SIGCOMM Computer Communication Review, 2012, 42(4):473-478.

[3] Killi B P R, Rao S V. Controller placement with planning for failures in software defined networks[C]// IEEE International Conference on Advanced Networks and Telecommunications Systems. IEEE, 2017:1-6.

[4] Hock D, Hartmann M, Gebert S, et al. Pareto-optimal resilient controller placement in SDN-based core networks[C]// Teletraffic Congress. IEEE, 2013:1-9.

[5] Muller L F, Oliveira R R, Luizelli M C, et al. Survivor: An enhanced controller placement strategy for improving SDN survivability[C]// Global Communications Conference. IEEE, 2014:1909-1915.

[6] Garcia-Molina H. Elections in a Distributed Computing System[M]. IEEE Computer Society, 1982.

[7] Marchetti-Spaccamela A. New protocols for the election of a leader in a ring[J]. Theoretical Computer Science, 1987, 54(1):53-64.

[8] Lamport L. Paxos made simple[J]. ACM Sigact News, 2001, 32(4): 18-25.