Machine Practice 2
CS425 Distributed System Networking
Prof. Indranil Gupta

Chen Zhu
Weiran Lin
Date: October 8, 2018

1 Objectives

- Implement a p2p Network and hold a membership list in different servers. The membership list will be updated in every server when a server join, leave or crash in the network.

2 System Design

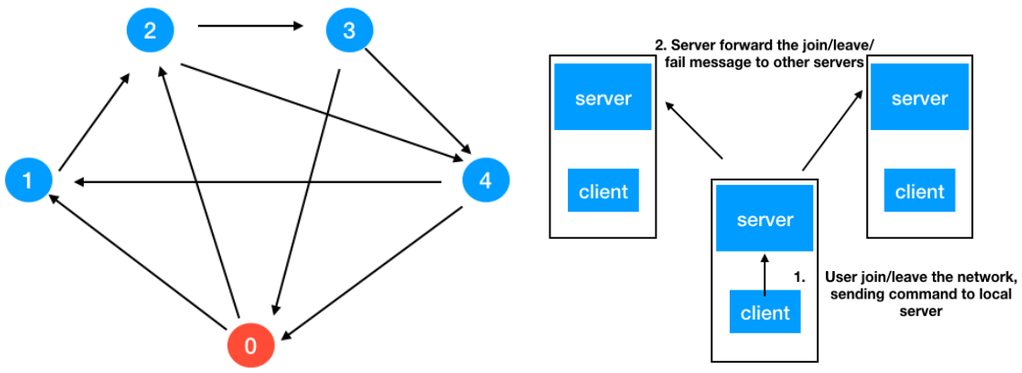


Fig. 1B.

Figure 1: Topology and Elementary Design. Figure in the left shows the simplified network topology of our system. Figure in the right shows the simplified procedure and elementary design.

2.1 Topology and Elementary Design

Figure 1 shows a simplified design of network topology. The first node is defined as the "introducer" node and should be added to the network first before other node joins. Each node is connected to the next k peer node in the virtual ring. (In this project we have 10 nodes and k=3)

The elementary design is also shown in the figure1. Each node is divided into server (daemon) and client process The joining and leaving process is as follows:

- **Join**: User input the join command in client. The client will send the command to local server. Then the server will gossip this joining message to the "introducer" and fetch the member list. After that, the "introducer" will gossip the joining message to its peer nodes
- **Leave**: User input the leave command in client. The client will send the command to local server. Then the server will gossip this leave message to its peer nodes actively

2.2 SWIM-like Fault Detection

All nodes in the system are arranged in a virtual ring. Every node maintains a peer list containing three nodes behind it by VM id.

Each node sends UDP packets to live nodes in its peer list every 0.5 second. If the receiver does not response in consecutive two times, the node will mark it as dead, and then multicast the message to nodes in its peerlist. These arguments are set because we hope a monitor can rapidly find if its UDP packets receiver is dead without occupying too much network bandwidth.

2.3 Message Format

We use gob package to marshal message send between servers and clients.

2.4 MP1 Usage

Each host will generate logs on network connection issues, membership list and peer list update, and received massage received, and MP1 is used for grep log message from all hosts.

3 Performance Analysis

3.1 Bandwidth When No Joining or Leaving

We will first determine the UDP traffics when there are no joining or leaving in the network, which means there are only fault detection traffics. We use **iptraf** software to determine number of incoming and outgoing bytes in one minute.

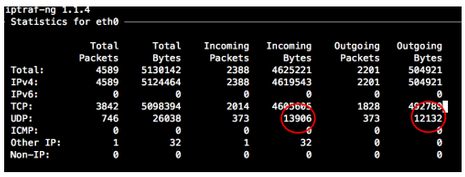


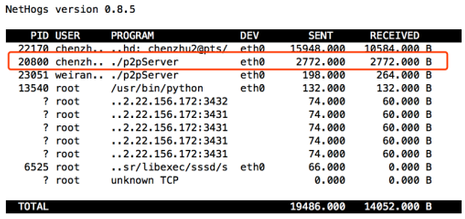
Fig. 2A. The iptraf, we count the total number of bytes sent and received in UDP protocol in 70 seconds. We can see that for 10 VMs and 3 peers, the total throughput (incoming+outgoing) is 26132, the average throuput is 373.3 bytes/s

VM No.	Number of Peers	Throughput (bytes/s)
0	3	373.1
1	3	372.5
2	3	371.2
3	3	374.1

Fig. 2B. The average throughput with different VMs. We can see that the throughput is almost the same. This is true because with the same number of peers, there will always be 6 peers to send/response packets

We can see that average throughput is around 372 bytes/s. This can be validated by mathematical proof. The UDP packets are sent every 500 ms. For each second, every node will receive packets from 3 peer nodes and send packets to 3 peer nodes. The packet length of each UDP packet (counting IP overhead) is around 30 bytes. Therefore, the throughput every second is 6 packets*2*30 bytes/packet=360 bytes/s

3.2 Bandwidth When Joining, Leaving or Failing



Msg Type	Sent (B)	Received (B)
Join	2309	2306
Leave	2772	2772
Fail	2306	2306

Figure 2: The traffic of the server process when there is a single join/leave/fail msg in certain node. The result is always precisely the same without any deviation. Every join/leave command will lead to 3 gossip messages to send and 3 messages to receive in each nodes. Therefore, the sent and received packet will be 2772/3=924 bytes long for "leave" message, 2309/3=769.67 bytes for leave message, and 2306/3=768.67 bytes for fail message.

3.3 False Positive Rate When Message Loss

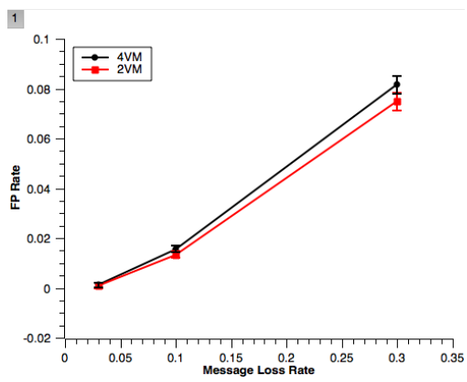


Figure 3: The false positive rates are theoretically 0.09, 0.01, 0.0003 when message losses are 30%, 10%, 3% with two VMs in the group. When the group has four VMs, false positive rate is higher because there are more chance to come across false positive cases with more members in peer list. The standard deviation in 3% message loss is higher because the data amount is less caused by longer false positive case interval.