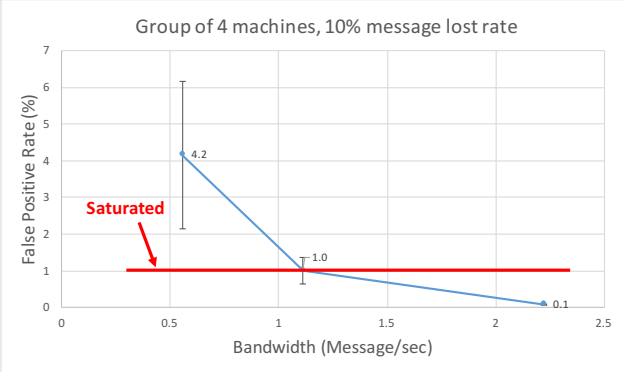
CS425 MP2 Report

Group 4: Yi-Hsin Chen(yihsinc2), Che-Lin Huang (clhuang2)

Design

Failure Detection:

Since we need to use virtual ring as the topology of the system, we design our system that each node in the system will pick **4 neighbors (2 successors/2 predecessors)** to send heartbeat signals to ensure **completeness up to 4 failures**. To make sure that a **failure must be detected within 2 seconds**, we set the **timeout** equals to the upper limit: **2 seconds**. Moreover, to **compromise between bandwidth and false positive rate** given the timeout, we set the heartbeat period to be 1 second, as shown below:



Each node will maintain a membership list locally, whose format is shown below:

Node ID	Heartbeat Counter	Local Time (msec)
Node_A	1234	5678
Node_B	456	5640

where **Node ID** contains the information of timestamp and node's IP address, **Heartbeat Counter** is used to decide the priority between different messages (the local membership list will be updated accordingly only if the node receives a message whose heartbeat counter is larger than the value stored in the local membership list), and **Local Time** is used for failure detection.

Dissemination:

To ensure that our system could be scalable to large number *(N)* of machines, we adopt the **gossip** algorithm, by which the message will reach almost all the nodes in *O(logN)* with high confidence.

According to the equation of infected nodes of gossip algorithm at $t = clog(N)$:

$$y \approx (N+1) - \frac{1}{N^{cb-2}}$$

we set c = 1, b = 4 to make sure that only a small portion of nodes don't receive the message. For a system with $N=10$, $t = 2.3$. Thus, we set $TTL = 3$ for each gossip message.

Marshaled Message Format:

To ensure platform independent message, we encode our message into **string** before sending out. There are 4 kinds of message in our design.

- **JOIN**: once a new node wants to join the group, the node will send a message to one of the introducers. The message format is **"LocalTimeStamp_IP Address"**.
- **HEARTBEAT**: once a node is in the group, it will send out heartbeat signals periodically. The message format is **"0_Node ID_Heartbeat Counter"**, where **"0"** is the flag to indicates this is a heartbeat message, **"Node ID"** is the node's ID, and **"Heartbeat Counter"** is used to prioritize message.
- **ADD:** once a new node joins the group, the introducer will send out this message to inform others to add this new node to their membership list. The message format is **"1_Node ID_ADD_Heartbeat Counter_TTL"**, where **"1"** is the flag to indicates that this message is a gossip one, and **"TTL"** is the lifetime of this package.
- **REMOVE:** once a node voluntarily leaves the group, or it detects a failure of other node, the node will send out the message to inform other nodes to remove the leaving/failed node from their membership list. The message format is **"1_Node ID_REMOVE_Heartbeat Counter_TTL"**, the meaning of different fields in the message is similar with **ADD** message.

Usage of MP1

We use MP1 to grep the log files from different VMs, which tremendously enhance the debugging efficiency since we don't need to go through all the VMs one by one to check logs.

Experiment

We define the bandwidth as **number of message per second** for the following reported values.

i. The background bandwidth of 4 machines (only includes heartbeats): 12

ii. The measured extra bandwidth (excludes heartbeat signals) of a node leaves/joins/fails a group with 4 members is: 13/28/39, which matches the calculated results. Let $X = \min(b, N)$ where b is the number of targets to relay the gossip, and N is the number of nodes in the group. When one node joins/leaves: the extra bandwidth is $\frac{1}{T}\sum_{k=1}^{TTL} X^k$. When one node fails, it will be detected by all its 3 neighbors, which will all send out gossip messages, so the extra bandwidth is: $\frac{3}{T}\sum_{k=1}^{TTL} X^k$. In our setting, b=4, N=5/3 for a node joins/leaves, we got the same results.

iii. The results are shown below. As expected, the higher the message loss rate, the higher the false positive rate.