Machine Programming 2 – Distributed Group Membership

Yu Tao, Jie Yin
yutao2@illinois.edu; jiey3@illinois.edu

Design

Nodes are located in the virtual ring. Each node sends the heartbeat to its four successors and monitors its four predecessors.

For join and rejoin, the potential nodes will firstly contact the introducer (VM1). Introducer will copy its own membership list to the potential nodes and then forward the "join message" to its four successors and the four successors will send this message to their successors and so on. Eventually, all nodes in the group can update their own membership.

For failure detection, when the heartbeat of a node is not received by its monitor within 5 times of the heartbeat duration, the corresponding node is marked as failure and the monitor sends the "failure message" to all its successors.

For voluntarily leave, the node which is going to leave will send the "leave message" to its four successors. Note that for node join, failure and leave, nodes in the group can receive the same message for a maximum of four times since there are four predecessors for each node. In our design, when the message is already documented, i.e. the membership has already been updated, the message will be ignored.

The addresses of nodes are uniquely hashed into integers ranged from 0-127. When N becomes large, the ring size can be enlarged and rehashing is applied, so it always scales to large N.

In our implementation, the type of messages sent over the group is string which comprised an integer, hostname, timestamp and hashed ID. The integer ranging from 0-5 stands for the following purposes:

Integer	0	1	2	3	4	5
Message type	heartbeat	join	membership list	new member	leave	failure

To debug MP2, MP1 plays a significant role. For each node, its executions are all logged, i.e. sending messages, receiving messages. Querying the logs of nodes allows us to trace a specific instruction, for example, whether the heartbeat is sent properly.

Analysis

(i) *The background bandwidth*

The messages over the network are heartbeats assuming no membership changes. The heartbeat period is 200ms, i.e. 5 times/second. For each node, it sends heartbeats to its three successors. So, the bandwidth is **5*3*4=60 messages/second**

(ii) *The expected bandwidth usage whenever a node joins, leaves or fails*

Assume 4 machines and heartbeats are not included, so that:

node joins: 1 request message sent by potential node, 4 messages containing introducer's membership list sent to potential node. 1 *3 * 4 "join message" send over the group. Note that messages sent using UDP is fast and upon the successor receiving messages, the messages will be further sent to successor's successors. Instead of giving the bandwidth, the number of messages usage for "join" is 17 for 4 machines.

node leaves: when one node leaves, it sends message to its three successors. Successors will forward the message to their successors including the leaving node. This takes 3 * 4 =12 messages.

node fails: when one node fails, its three monitors sends "failure message" to successors. This takes

3 * 2 = 6 messages.

(iii) *plot the false positive rate of your membership service when the message loss rate is 3%, 10%, 30%* Failure count within 10 minuets for each case. Assume that kills one node is manfully killed while other failures are caused by false positive.

Table 1 Raw Data

		reading 1	reading2	reading3	reading4	reading5
two nodes	3%	4	4	3	3	4
	10%	20	23	21	21	23
	30%	194	199	182	178	186
four nodes	3%	7	9	8	7	10
	10%	96	87	92	79	88
	30%	810	793	821	801	812

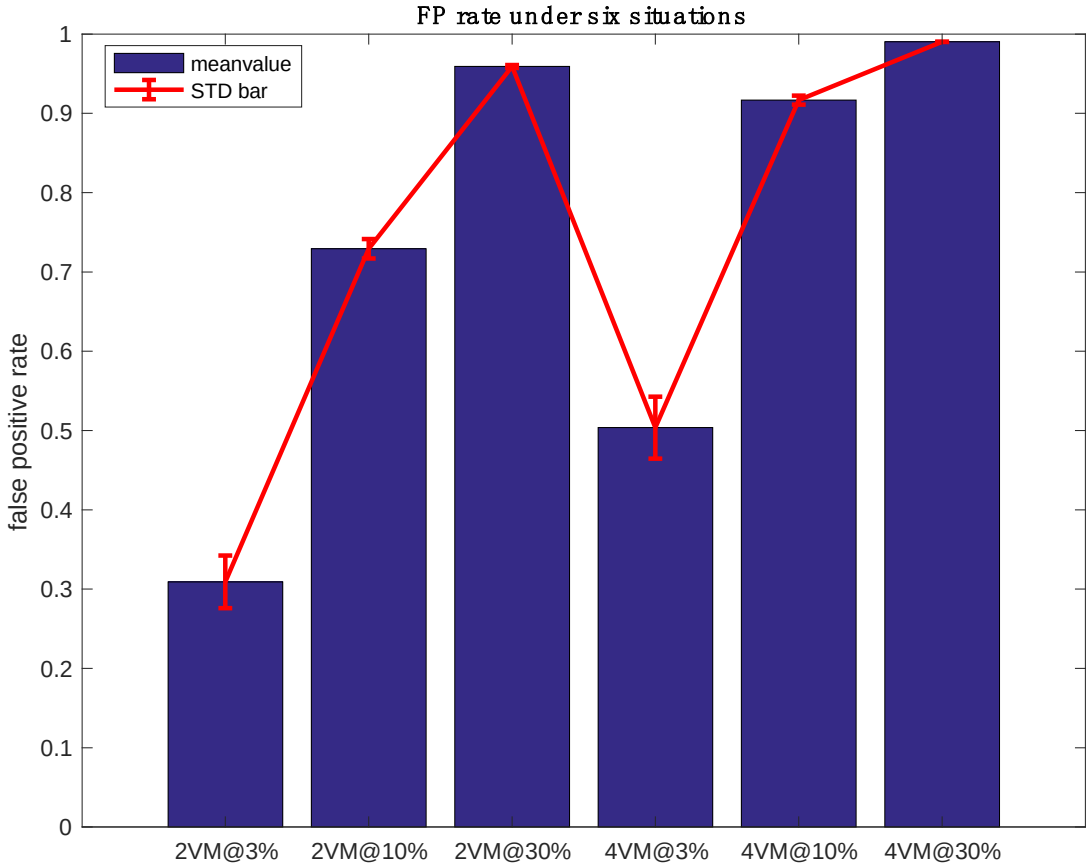

Figure 1 Average & STD Plot for 6 Situations

From above figure, the false positive rate increases with the increment of message loss rate and gradually approaches to one. When the number of VMs in the group increases, the false positive rate increases as well. It matches our expectation, because as the message loss rate increases, one node's time out probability also increases so that it is more likely to be marked as failure. Likewise, as the number of VM increases, the overall chance of being marked as failure increases as well.