

UNIVERSITY OF MELBOURNE

DISTRIBUTED SYSTEMS

COMP90015

Assignment 2: Group JWP

Authors:

William CAMPBELL

(WCAMPBELL)

Justin GUNN (GUNNJK)

Peng ZHAO (PZHAO2)

Instructor:

Prof Aaron HARWOOD

Student number:

756142

989347

899632

Emails

W.CAMPBELL@STUDENT.UNIMELB.EDU.AU

GUN-

NJK@STUDENT.UNIMELB.EDU.AU

PZHAO2@STUDENT.UNIMELB.EDU.AU

February 26, 2019

1 Introduction

The project improved upon our first iteration of a distributed group communication system. The new system is designed to provide availability and eventual consistency in the presence of server failures and network partitioning. The main protocols implemented were

- Replication of Message Queue and a global clock from the server side to achieve eventual delivery of messages in the order they were sent.
- Failure protocol to reconnect partitioned servers and continue sending replicated Message Queues.
- Reconfigure login, so that users can login from anywhere

We achieved a functioning scalable and dynamic distributed application. Providing high availability and eventual consistency in its message passing service. The messages have a guarantee to be sent to all members connect to the network at that time. Additionally messages sent from a client will be delivered to receivers in the order they were sent.

2 failure model

Our system achieved high availability and eventual consistency by implementing a "self healing" tree structure in our servers. Each server uses a message queue to retrieve messages from its clients. When a client sends a message it is put in this queue. The server then processes jobs FIFO from this queue to pass them to all other servers. This queue is replicated and put in that nodes parent node, then updated each time the message queue is. Updates occur before messages are processed. If a server fails, connections to that server are re-routed to where the backup message queue is held. Then this queues messages are again replicated and pushed to that nodes parent and then processed until the queue is empty. This way all messages are eventually sent to all clients.

Whenever a server fails or a connection to a server fails (even if transiently) servers connected to that server are rerouted to that servers parent node. In the special case that the route node is killed, it puts it replicated message queue in the first server to connect to it. We implemented this logic

by giving each server a dead-pool list. This stores the connection information of their parent nodes replication node. When the socket to a parent has failed, they access their dead-pool and make a run-time connection to the new server to initiate replication message queue passing. This arrangement is illustrated in figure 2 and figure 3 below, demonstrating initial dead-pool lists, reconnect and re-establishing dead-pool lists.

Figure 1: Initiated server tree

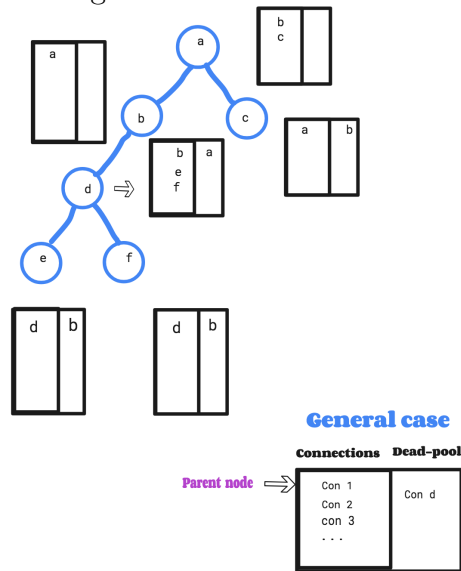
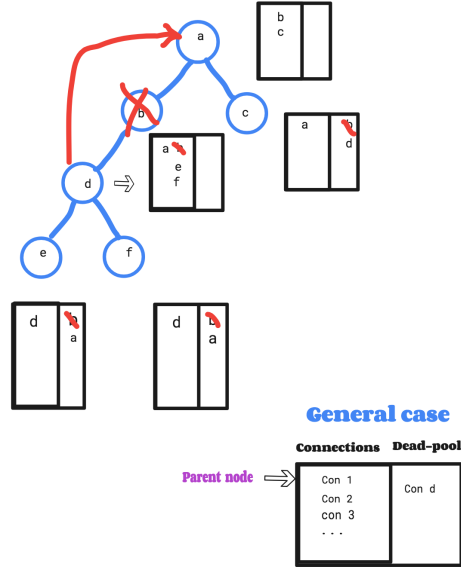


Figure 2: Server b failure, reconnect of d to a



3 Protocols and messages

AUTHENTICATION_SUCCESS

When a server receives an authentication message from another. They send back an AUTHENTICATION_SUCCESS message, containing its parent nodes information, for the servers dead-pool. If they have no parent node, the message is left blank.

Example:

```
"secret" : "aaa",
"port" : "3789",
"hostname" : "127.000.101"
"empty": "F"
```

The server will:

- Set their deadpool with this address, stored in their Settings class under backSecret, backPort and backHost
- if this server dies, stored as Settings. parentServer. They initiate a connection to their deadpool using the same protocols as the initial AUTHENTICATE, updating their deadpool in the process.

Revised client login, login from anywhere

LOGIN_REQUEST

Broadcast from a server to all other servers (between servers only) to indicate that a client is trying to login as a username with a given secret. Example:

```
"command" : "LOGIN_REQUEST",  
"username" : "aaron",  
"secret" : "fmmpp3ai91qb3gc2bvs14g3ue"
```

A username must be present. The value of username may be anonymous in which case no secret field should be given (it should be ignored). Otherwise a secret must be given.

A server that receives this message will:

- Broadcast a LOGIN_DENIED to all other servers (between servers only) if the username is not known to the server and

it is a terminal server within the server tree (i.e. the last server on a branch).

- Broadcast a LOGIN_ALLOWED to all other servers (between servers only) if the username and secret is known to the server. The server will record this username and secret pair in its local storage.
- INVALID_MESSAGE in any case where the message is incorrect
- Broadcast another LOGIN_REQUEST to remaining network branches and await their responses if the server is not a terminal server.

LOGIN_DENIED

Broadcast from a server to all other servers (between servers only), if the server received a LOGIN_REQUEST and to indicate that the username and password is not known, the original client should not be logged in. Example:

```
"command" : "LOGIN_DENIEDT",  
"username" : "aaron",  
"secret" : "fmmpp3ai91qb3gc2bvs14g3ue"
```

LOGIN_ALLOWED

Broadcast from a server to all other servers (between servers only), if the server received a LOGIN_REQUEST and to indicate that the username and password already known, the original client should be logged in. Example: Example:

```
"command" : "LOGIN_ALLOWEDT",  
"username" : "aaron",  
"secret" : "fmmpp3ai91qb3gc2bvs14g3ue"
```

LOGIN_DENIED

Broadcast from a server to all other servers (between servers only), if the server received a LOGIN_REQUEST and to indicate that the username and password is not known, the original client should not be logged in. Example:

```
"command" : "LOGIN_DENIEDT",  
"username" : "aaron",  
"secret" : "fmmpp3ai91qb3gc2bvs14g3ue"
```

Clock server, time stamping for order and delivery guarantee

LOGIN_ALLOWED

Broadcast from a server to all other servers (between servers only), if the server received a LOGIN_REQUEST and to indicate that the username and password already known, the original client should be logged in. Example: Example:

```
"command" : "LOGIN_ALLOWEDT",  
"username" : "aaron",  
"secret" : "fmmpp3ai91qb3gc2bvs14g3ue"
```

Improvements To improve upon this model, we would need to implement the same message queue replication for received messages as well. However this proved above our programming skills to implement.

4 assumptions

The above implementation relies on assuming 2 failures do not happen to both a server and where its backup is at the same time. If this was not true we would need to hold additional replicas of each, however we assumed 2 failures would not happen at once for our system. Additionally our clock server is designed to never fail, this would need to be implemented using some form of RAID replication, guaranteeing that the clock is always on. This was outside our scope of abilities.

5 Functioning with failure demonstration

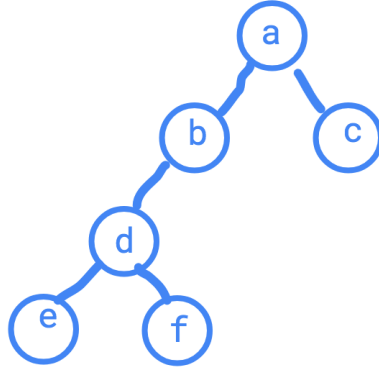
The "self healing tree" that we have implemented functions as described in our failure model section continues to function in the face of failure by maintaining a list of its parents parent. We have called this our dead pool (backPort and backServer in stored in settings). For every node below the root nodes children. This is a special case and only the 2nd child maintains a list of the 1st child's address. This allows the tree to "heal" if the root node dies. See diagrams in failure model section for further clarification. Where if one of the other nodes dies, the other child will still be connected to the root and our tree is maintained.

Bellow these children, each new node is passed the address of the parent of the node they connected to. Upon a nodes parents failure, they initiate a new socket with the node stored in their dead-pool. Those reconnecting a partitioned network and providing continued functionality for the system. Additionally we have implemented replicated Ques with every message a node will process. Once a message from a client comes it is put in this queue, the queue is replicated (if this is the first message) or sent to a replicated queue at this nodes parent. Meaning if one of the nodes fails mid message passing. Those messages have been copied and reside in their parent node. If they die the network will heal and reconnect, once this happens that queue will finish sending out any pending messages. This guarantees all messages sent by a client to a server will eventually reach their recipients, even in the face of server failure.

System test: Bellow are trees describing the connections that maintain system availability.

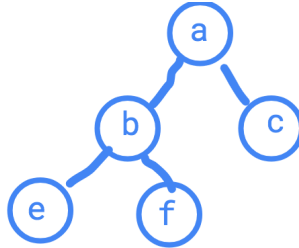
- Six servers are started up in a tree format as dictated in the bellow image. Where port numbers of each node are as follows.
 - a = 3781
 - b = 3782
 - c = 3783
 - e = 3784
 - f = 3785
 - g = 3786

Figure 3: Six server tree.



- Server b fails, leading to d reconnecting to server a and the tree bellow.

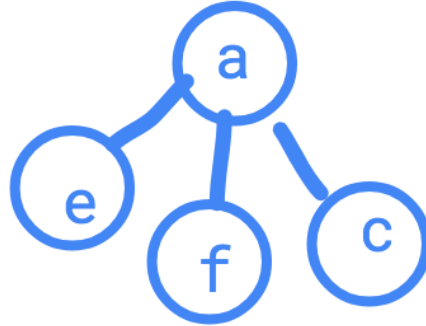
Figure 4: Six server tree.



- Server d fails lead to e and f to reconnect to a and the tree bellow.

Special cases Unfortunately the system implemented still contains some bugs and corner cases where the dead-pool list is not updated properly. Still strictly occurs when the pool of servers has grown over three. Then through failures the 2nd server added, then the first server added are eliminated in that

Figure 5: Six server tree.



order. In this case the remaining servers will fail to have an updated dead-pool list and when the root server dies they will try to connect to that 2nd server, however it is dead. Saying this the logic could be implemented, however we were not able to in the time span.

The above trees were established under the bellow sequence of steps running our program.

1. start a clock server to guarantee client message deliver running ClockServer and -lp 3779
2. Create the tree seen in figure 3, by starting servers using the following sequence of server start ups.
 - (a) -lp 3781 -lh localhost -ms aaa
 - (b) -lp 3782 -lh localhost -rh localhost -rp 3781 -s aaa -ms bbb
 - (c) -lp 3783 -lh localhost -rh localhost -rp 3782 -s aaa -ms ccc d
 - (d) -lp 3784 -lh localhost -rh localhost -rp 3783 -s bbb -ms ddd e

(f) -lp 3786 -lh localhost -rh localhost -rp 3785 -s ddd -ms
fff

4. Kill server b

[illegible][illegible][illegible]

Figure 9: Server d initial.

[illegible]

Figure 10: Server e initial.

[illegible]

Figure 11: Server f initial.

[illegible]

Figure 12: Server f after d death.

[illegible]

