ISSN No. 0976-5697

Volume 3, No. 4, July- August 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Modified Election algorithm in Distributed System using election commission & comparision with original bully algorithm & ring algorithm

Sachi Choudhary*
Computer Science & Engineering
Raipur Institute of Technology
Raipur (C.G.), India
sachi.choudhary1617@gmail.com

Mr. Dipesh Sharma
Reader, Computer Science & Engineering
Raipur Institute of Technology
Raipur (C.G.), India
download.dks@gmail.com

Abstract: In this paper an election algorithm is presented that can be used in distributed systems to elect a coordinator. Many distributed algorithms need one process to act as a leader or coordinator to make synchronization between different processes. There are various algorithms that elect a particular node as coordinator from all the different nodes in the distributed network. This paper will present a modified version of bully election algorithm using a new concept called election commission. This approach will not only reduce the number of messages per election but also minimize total number of elections and hence it will minimize network traffic, redundant elections.

Keywords: Coordinator; Election Commission; Distributed system; Bully election; Ring election; Messages; Complexity.

I. INTRODUCTION

A distributed system consists of a collection of autonomous computers, connected through a network and distribution middleware, which enables computers to coordinate their activities and to share the resources of the system, so that users perceive the system as a single, integrated computing facility[2][3]. In a distributed computing system, a process is used to coordinate many tasks. It is not an issue which process is doing the task, but there must be a coordinator that will work at any time. So it is sometimes necessary that,from a set of processes, a process must be selected as a leader or coordinator. There are various algorithms that elect a particular node or process as leader from all the different nodes in the distributed network. These algorithms are responsible for maintaining a leader in the distributed network like Bully algorithm. In this paper we are presenting a modified version of Bully algorithm which not only overcome the limitations of Bully algorithm but also provides a mechanism for node failure detection. This algorithm will reduce the network traffic, message complexity, time complexity as well as redundant election in the network.

II. DISTRIBUTED SYSTEM

A distributed system is a collection of processors interconnected by a communication network in which each processor has its own local memory and other peripherals and the communication between them is held by message passing over the communication network [2][3].

Features of Distributed System:

a. Inherently distributed applications
b. Information sharing among distributed users
c. Resource sharing
d. Better price performance ratio
e. Shorter response times and higher throughput
f. Higher reliability
g. Extensibility and incremental growth
h. Better flexibility in meeting users needs

Several distributed algorithms require that there be a coordinator node in the entire system that performs some type of coordination activity needed for the smooth running of other nodes in the system. As the nodes in the system need to interact with the coordinator node, they all must unanimously who the coordinator is. Also if the coordinator node fails due to some reason (e.g. link failure) then a new coordinator node must be elected to take the job of the failed coordinator [3].

III. ELECTION ALGORITHM

Coordinator election algorithm[1][2][7] is a classical problem in distributed system as data is distributed among different node which is geographically separated. For maintaining co-ordination between the node, coordinator or leader node have to be selected. This leader node is responsible for proper maintenance and functioning of the network. There are various algorithms that elect a particular node as leader from all the different nodes in the distributed network. These algorithms are responsible for maintaining a leader in the distributed network. There can be various events like leader node going down, previous leader node recovering from a crash, etc. It is the duty of the leader election algorithms to always maintain a leader in the distributed network (by electing a new leader if current leader crashes). Thus the leader election algorithm has to be good enough to perform all its tasks efficiently.

Election algorithms are based on the following assumptions [4][5]:

A. Every node has a unique ID; e.g.
   a. the network address
   b. a process number

B. Every node in the system should know the values in the set of ID numbers, although not which node is up or down.
C. The node with the highest ID number will be the new coordinator.
D. Node groups satisfy these requirements.
E. Any node can serve as coordinator
F. Elections may be needed when the system is initialized, or if the coordinator crashes or retires.
G. Any node can "call an election" (initiate the algorithm to choose a new coordinator).

IV. EXISTING ALGORITHMS

a. Bully Algorithm by Garcia Molina[1][4][7]
b. Ring Algorithm by Silberschatz And Galvin[2][3]
c. Modified Election Algorithm by M.S. Kordafshari et al.[7]

A. Bully algorithm by Garcia Molina:

Bully algorithm[1][4][7] is one of the most applicable election Algorithms which was presented by Garcia_Molina in 1982. The bully algorithm is a method in distributed computing for dynamically selecting a coordinator by process ID number. In this algorithm each process has a unique number to distinguish them and each process knows other's process number. In this algorithm processes don't know which ones are currently up and down. The aim of election Algorithm execution is selecting one process as leader (Coordinator) that all processes agree with it. (i.e. process with the highest id number)[4][5].

a. Assumptions:

This algorithm is established on some basic assumptions which are [1][4][5][7]:

a) It is a synchronous system and it uses timeout mechanism to keep track of coordinator failure detection .
b) Each process has a unique number to distinguish them .
c) Every process knows the process number of all other processes .
d) Processes do not know which processes are currently up and which processes are currently down .
e) In the election, a process with the highest process number is elected as a coordinator which is agreed by other alive processes
f) A failed process can rejoin in the system after recovery.

In this algorithm, there are three types of message and there is an election message (ELECTION) which is sent to announce an election, an reply(OK) message is sent as response to an election message and a coordinator (COORDINATOR) message is sent to announce the new coordinator among all other alive processes.

b. Algorithm:

Suppose that the process P finds out the coordinator crashed. This algorithm has the following steps: (As figure 1)

Step1- when a process, P, notices that the coordinator crashed, it initiates an election algorithm

a) P sends an ELECTION message to all processes with higher numbers respect to it.
b) If no one responses, P wins the election and becomes a coordinator.

Step2- when a process receives an ELECTION message from one of the processes with lower numbered response to it:

a) The receiver sends an OK message back to the sender to indicate that it is alive and will take over.
b) The receiver holds an election, unless it is already holding one.
c) Finally, all processes give up except one that is the new coordinator.
d) The new coordinator announces its victory by sending a message to all processes telling them, it is the new coordinator.

Step3- Immediately after the recovery of the crashed process is up, it runs bully algorithm.

c. Advantages :

a) The advantages of Bully algorithm are that this algorithm is a distributed method with simple implementation.
b) This method requires at most five stages, and the probability of detecting a crashed process during the execution of algorithm is lowered in contrast to other algorithms. Therefor other algorithms impose heavy traffic in the network in contrast to Bully algorithm.
c) Another advantage of this algorithm is that only the processes with higher priority number respect to the priority number of process that detects the crash coordinator will be involved in election, not all process are involved.

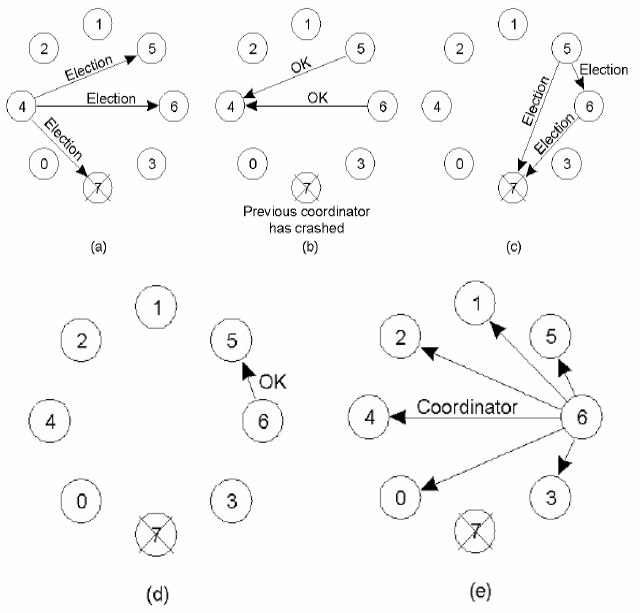

Figure 1: Bully algorithm example (a) process 4 detects coordinator is failed and holds an election, (b) process 5 and 6 respond to 4 to stop election, (c) each of 5 and 6 holds election now, (d) process 6 responds to 5 to stop election, (e) process 6 winds and announces to all.

d. Limitations :

a) The main limitation of bully algorithm is the highest number of message passing during the election and it has order O (n^2) which increases the network traffic.
b) When any process that notices coordinator is down then holds a new election. As a result, there May n number of elections can be occurred in the system at a same time which imposes heavy network traffic.
c) As there is no guarantee on message delivery, two processes may declare themselves as a coordinator at the same time. Say, P initiates an Election and didn't get any reply message from Q, where Q has a higher process number than P. At that case, p will announce itself as a coordinator and as well as Q will also initiate new election and declare itself as a coordinator if there is no process having higher process number than Q.
d) Again, if the coordinator is running unusually slowly (say system is not working properly for some reasons) or the link between a process and a coordinator is broken for some reasons, any other process may fail to detect the coordinator and initiates an election. But the coordinator is up, so in this case it is a redundant election

B. Ring Algorithm by Silberschatz And Galvin:

This election algorithm is based on the use of a ring[2][3]. We assume that the processes are physically or logically ordered, so that each process knows who its successor is. In this algorithm the processes are arranged in a logical ring, each process knows the structure of the ring.

a. Assuptions :

a) All the nodes in the system are organized as a logical ring.
b) The ring is unidirectional in the nodes so that all the messages related to election algorithm are always passed only in one direction.

b. Algorithm :

a) A process initiates an election if it just recovered from failure or it notices that the coordinator has failed.
b) Initiator sends *ELECTION* message to closest downstream node that is alive.
i. *ELECTION* message is forwarded around the ring
ii. Each process adds its own ID to the *ELECTION* message.
c) When *ELECTION* message comes back, initiator picks node with highest ID and sends a *COORDINATOR* message specifying the winner of the election.
d) *COORDINATOR* message is removed when it has circulated once.
e) Multiple elections can be in progress.

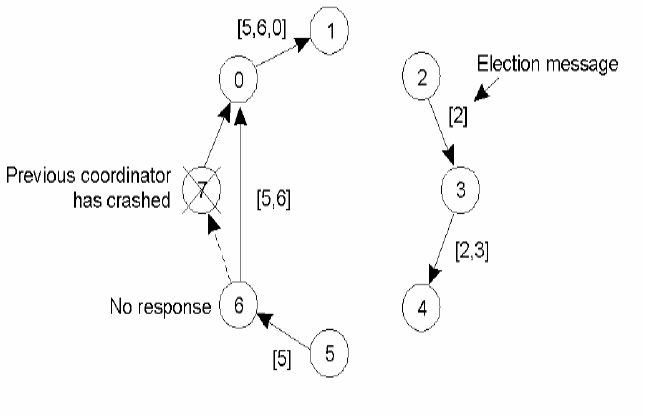

Figure: 2 Ring algorithm example.

c. Limitations :

a) The flow of message in ring is unidirectional which increases the time complexity of the election algorithm.
b) The worst-case scenario occurs when the counter-clockwise neighbor has the highest id.

C. Modified Election algorithm by M.S. Kordafshariet al.[7]:

Modified Bully algorithm by Quazi Ehsanul Kabir *Mamun et al*. Quazi Ehsanul Kabir Mamun et al. described an efficient version Bully algorithm to minimize redundancy in electing the coordinator and to reduce the recovery problem of a crashed process.

a. Assumption :

There are five types of message. An election message is sent to announce an election, an ok message is sent in response to an election message, on recovery, a process sends a query message to the processes with process number higher than it to know who the new coordinator is, a process gets an answer message from any process numbered higher than it in response to a query message and a coordinator message is sent to announce the number of the elected process as the new coordinator.

b. Algorithm :

a) When a process p notices that coordinator is down, it sends an election message to all processes with higher number. If no response, p will be the new coordinator.
b) If p gets ok message, it will select the process with highest process number as coordinator and send a coordinator message to all process.
c) When a crashed process recovers, it sends query message to all process with higher process number than it.
d) And if it gets reply then it will know the coordinator and if it doesn't get any reply it will announce itself as a coordinator.

c. Limitations :

Although this algorithm reduces redundant election on some extent, it still has some redundant elections and also has

high message complexity. Some of the limitations are given below:

a) On recovery, it sends query message to all processes with higher process number than it, and all of them will send answer message if they alive. Which increases total number of message passing and hence it increases network traffic.
b) It doesn't give guarantee that any process p will receive only one election message from processes with lower process number. As a result there may be q different processes with lower process number can send election essage to p and p will send ok message to all of them. This increases number of election and also number of message passing.
c) It doesn't give any idea if p will crash after sending an election message to all processes with higher process number.
d) It also doesn't give any idea if a process with the highest process number will crash after sending ok message to p.

V. PROPOSED ALGORITHM

A. Election Commission (EC):

Election Commission is an electoral administrative body established to deal with leader election mechanism in a distributed computing system. It is constructed by a group of special processes in distributed system. It is authorized to handle the whole election process. It defines the rules and regulations for attending in an election process in a distributed computing system. It has few Election Commissioners. If any of the commissioners failed, Election Commission will recover that commissioner immediately and other processes do not have concern of that. An Election Commission has a unique group ID. Other processes in the system communicate with Election Commission using this group ID. As a result, if the commissioners is down, there will be not any problem in election. It has a reliable failure detector (FD) [4]. If Election Commission does not get any reply from a process within T time, then FD of Election Commission will report that requested process is down. As like as FD[4], Election Commission has another component named helper (HP), the function of HP is to find out the process with the highest process number using sending alive message. It knows process number of all processes of the system.

B. Election Commissioner:

Election Commissioner is a member of Election Commission. It is a special kind of process. Any Election Commission in a distributed system will have a few numbers of Election Commissioners.

VI. MODIFIED APPROCH OF ELECTION ALGORITHM

A. Algorithm :

a) When process P notices that the coordinator is down, it sends an ELECTION message to Election Commission.
b) FD of Election Commission verifies ELECTION message sent by P. If the sending notice of P is not correct, then Election Commission will send a COORDINATOR message to P with process number of the current coordinator.
c) If the sending notice of P is correct and if the highest process number is P, then Election Commission will send a COORDINATOR message to all processes with process number of P as a new coordinator. If the highest process number is not P, Election Commission will simply find out the alive process with the highest process number using HP and sends a COORDINATOR message to all processes with the process number of that process as a new coordinator.
d) If any process including last crashed coordinator is up, it will send a QUERY message to the Election Commission. If the process number of the newly entranced process is higher than the process number of the current coordinator, Election Commission will send a COORDINATOR message to all processes having the process number of new coordinator.
e) If not, Election Commission will simply send a COORDINATOR message to newly entranced process having process number of the current coordinator.
f) If more than one process sends ELECTION message to Election Commission at the same time, then Election Commission will consider the process with higher process number which ensure less message passing to find out the highest process number using HP.

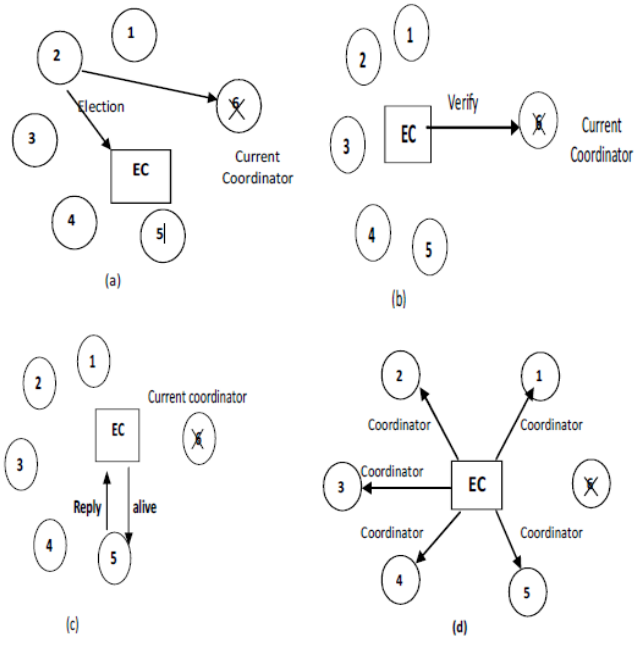

Figure 3 : Election Procedure: (a) Process 2 detects current coordinator is down and sends an election message to EC, (b) EC verifies either the coordinator is really down or not,(c) EC finds the alive process with highest number using alive message,(d) EC sends coordinator message to all process having process number of currently won.

B. Description :

Figure 3 represents regular election procedure of the proposed algorithm. Here, the system consists of six processes with process number 1 to 6. Current coordinator is the process 6. But it has just crashed and process 2 first notices this. So it sends an election message to the EC in Figure 3(a).In Figure3(b), EC sends verify message to the current coordinator to be sure about the election message sent by process 2. After verification, In Figure 3 (c), EC sends alive message to process 5 (the next highest process number) to check either the current highest process is alive or not. And EC gets a reply message from 5. In Figure 3(d), EC select 5 as new coordinator and sends coordinator message to all processes having 5 as a new coordinator of the system.

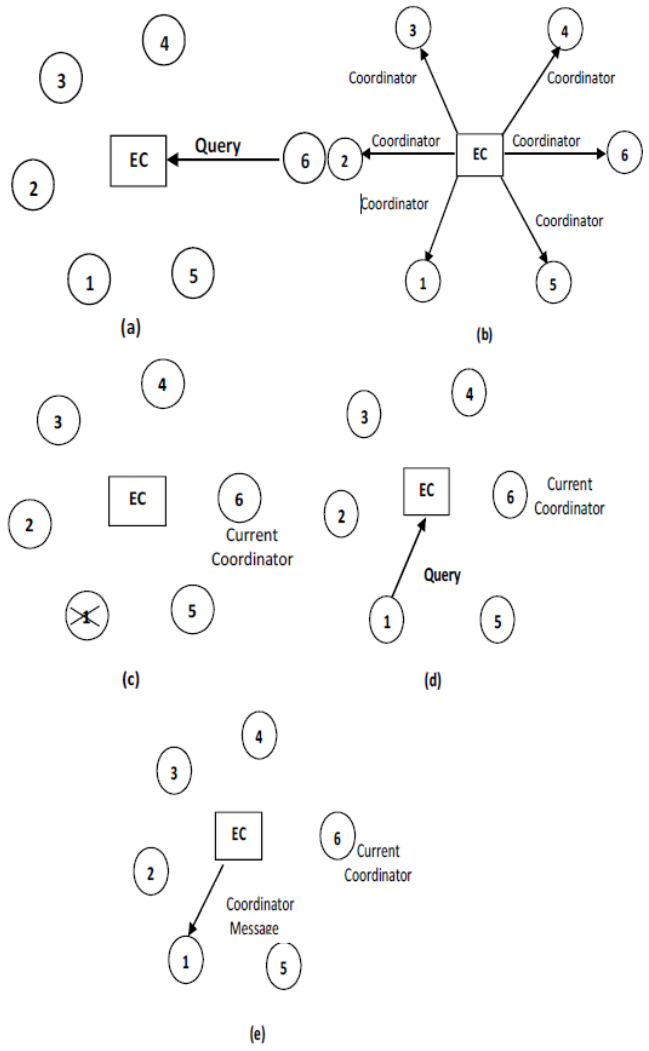

Figure 4. Query after Recovery: (a) Last crashed coordinator 6 is up and sends a query message to the EC, (b) EC selects 6 as new coordinator and sends coordinator message to all processes, (c) Now process 1 is crashed, (d) Again process 1 is up and sends query message to EC,(e) EC sends coordinator message to process 1 having the current coordinator).

Figure 4 represents the steps when a crashed process is up. In Figure 4 (a), last crashed coordinator 6 is up and sends a query message to EC. As process number of 6 is higher than the current coordinator of the system, in Figure 4 (b), EC sends coordinator message to all processes with process number 6 as new coordinator. In figure 4 (c), process 1 is now just Crashed. In figure 4 (d), process 1 is just up after crashed, and it sends a query message to EC. EC checks that process number of newly entranced is lower than the current coordinator. So in Figure 4 (e) EC sends coordinator message to only process 1 having the process number of current coordinator of the system.

VII. COMPARISION AND ANALYSI OF DIFFERENT ELECTION ALGORITHMS

In **Bully algorithm**,[4][7] when the process having the lowest priority number detects the coordinator's failure and initiates an election, in a system of n processes, altogether (n-2) elections are performed. All the processes except the active process with the highest priority number and the coordinator process that has just failed perform elections. So in the worst case, the bully algorithm requires O(n2) messages. When the process having the priority number just below the failed coordinator detects failure of coordinator, it immediately elects itself as the coordinator and sends n-2 coordinator messages. So in the best case, it has O(n) messages.

During recovery, a failed process must initiate an election in recovery. So once again, Bully algorithm requires O(n2) messages in the worst case, and (n-1) messages in the best case.

In **ring algorithm[2][3]**, on the contrary, irrespective of which process detects the failure of coordinator and initiates an election, an election always requires 2(n-1) messages. (n-1) messages needed for one round rotation of the ELECTION message, and another (n-1) messages for the COORDINATOR message. The algorithm proposed by Sandipan Basu has O(n) message efficiency.

During recovery, a failed process does not initiate an election on recovery, but just searches for the current coordinator. So ring algorithm only requires n/2 messages on average during recovery.

For the case of **modified bully algorithm[7]** there will be need of or O(n) message passing between processes. In worst case that is the process with lowest process number detects coordinator as failed, it requires 3n-1 message passing. In best case when p is the highest process number, it requires (n-p) + n messages.

For the case of **our proposed algorithm** there will be need of 1 election message to inform EC, 2 verify message to ensure the failure of coordinator, and say r is the highest alive process then alive and reply message to find out the highest alive process and so total or O (n) message passing between processes. If the process with lowest process number detects coordinator as failed it will not change total message. In worst case it may happen that our algorithm needs to check up process to p+1 to find out highest alive process. Only at that case it requires message passing between processes. However, in best case, our algorithm may find the highest alive process with only one alive and one reply message that is highest alive process in the system is process with process number n-1. In that case, our algorithm requires only 1+2+2+n messages. When p is the highest process number, it requires only 1+2 + n messages.

If a process crashes and recovers again, it sends a query message to all processes higher than that process to know the current coordinator which requires 2*(n-p) message passing. But in our algorithm, any process after recovery will only send query message to EC and EC will send a coordinator message having process number of current coordinator which requires only 2 messages passing.

VIII.CONCLUSION

In this paper, we modified bully algorithm using a new concept Election Commission (EC). We tried to overcome limitations of original bully algorithm and modified bully algorithm. Our comparison and discussion section prove that our algorithm is more efficient than bully algorithm and modified bully algorithm in respect of message passing, redundant election and network traffic.

IX. REFERENCES

[1]. Sandipian Basu "An Efficient Approach of Election Algorithm in Distributed Systems" Post graduate Department of Computer Science,St. Xavier's College, Kolkata INDIA. Indian Journal of Computer Science & engg. ISSN : 0976-5166 Vol. 2 No. 1

[2]. Tanenbaum A.S, Distributed Operating System, Pearson Education, 2007. PP. 140-142

[3]. Silberschatz Galvin Gagne, "Operating Systems Concepts " ,Seventh edition, page no. 265 topic : Election algorithm.

[4]. Heta Jasmin Javeri, Sanjay Shah "A Comparative Analysis of Election Algorithm in Distributed System" IP multimedia communication, A special issue from IJCA

[5]. Garcia – Molina "Elections in a distributed computing system", IEEE transactions on computers, vol C-31, No 1,pp 48-59., 1982.

– S. Mahdi Jameii, "A Novel Coordinator Selection Algorithm in Distributed Systems" Department of Computer-Engineering, Islamic Azad University, Shahr-e-Qods Branch, Tehran, Iran IJAEST Vol No. 9, Issue No. 2, 310 -313

[6]. Quazi Ehsanul Kabir Mamun, Salahuddin Mohammad Masum , **"**Modified Bully algorithm for electing coordinator in distributed system" . CD proceedings of 3rd WSEAS international conference on software engg. , parallel & distributed system (SEPADS) 2004 feb 13-15, Australia

[7]. "Message Efficient Leader Election in Synchronous Distributed System with Failure Detectors "Sung-Hoon Park School of Electrical and Computer Engineering, Chungbuk National Unvi. Cheongju ChungBuk 361-763, Korea

