

Linearizability

Zhaoguo Wang

Some slides are adapted from Prof. Jinyang Li's DS class

Outline

Recap VR

Linearizability

Outline

Recap VR

Linearizability

Correctness: servers execute the same sequence of log



No two different ops are committed at the same log position



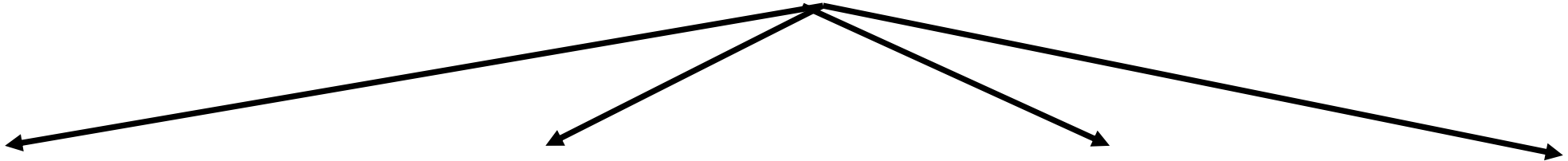
All committed ops in view $V - 1$ must be known to primary in view V



In any majority, there exist a backup has all committed requests in its log.
New primary collects safe log from a majority.



A primary needs to commit a request with a majority.
In the same view, a backup's log is always the prefix of a primary's log.



At the beginning of a view,
backup's logs are identical
with primary's log.

Request is committed in the
same position of log in a
majority of servers.

Before put a request into
the log, wait until all earlier
log entries are appended to
the log

Recover by replacing
the log from a
majority

Normal Process

Primary sends the req with curV to all



On receiving Prepare msg, server must:

- ignore if its curV $\neq$ msg.curV or status $\neq$ NORMAL
- wait until it has entries in its log for all earlier requests
- append req to log at position msg.op-number



Primary considers a req committed after getting majority OKs.

Recovery

A healed server sends Recovery request to all



On receiving *Recovery* msg, server must:

- Ignore if status $\neq$ NORMAL
- Reply <curV, x, log, op-number, commit-number>



On receiving *RecoveryReply* from a majority:

- Replace log and curV with replies of latest view
- Change its status to NORMAL

ViewChange

A candidate sends a ViewChange request to all



On receiving ViewChange msg, server does:

- reject if its curV $>$ msg.V
- set curV = msg.V, status= VIEWCHANGE
- return its log and the lastV.



On receiving OK from majority:

S1 decides on the newLog for V4 with following rules:

1. Rule 1: Pick the log w/ biggest lastV
2. Rule 2: if >1 logs exist in rule-1, pick the longest one



New primary sends StartView message to all



Backups replace log with primary's log, update status and lastV

Exercise I

curV lastV status

1	1	NORMAL	S1	OP1
1	1	NORMAL	S2	OP1
1	1	NORMAL	S3	OP1
1	1	NORMAL	S4	OP1
1	1	NORMAL	S5	OP1

What is the eventual view number?
What is the content of the primary's log?

1. Network Partitions S1, S2 from Others 2. S1 receives OP2, OP3, OP4 3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1 5. S3 crashes 6. Network heals 7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed 9. S4 crashes 10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	NORMAL	S1	OP1
1	1	NORMAL	S2	OP1
1	1	NORMAL	S3	OP1
1	1	NORMAL	S4	OP1
1	1	NORMAL	S5	OP1

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	NORMAL	S1	OP1	OP2	OP3	OP4
1	1	NORMAL	S2	OP1			
1	1	NORMAL	S3	OP1			
1	1	NORMAL	S4	OP1			
1	1	NORMAL	S5	OP1			

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	NORMAL	S1	OP1	OP2	OP3	OP4
1	1	NORMAL	S2	OP1			
3	3	NORMAL	S3	OP1	OP5	OP6	
3	3	NORMAL	S4	OP1	OP5	OP6	
3	3	NORMAL	S5	OP1	OP5	OP6	

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	NORMAL	S1	OP1	OP2	OP3	OP4
1	1	NORMAL	S2	OP1	OP2	OP3	OP4
3	3	NORMAL	S3	OP1	OP5	OP6	
3	3	NORMAL	S4	OP1	OP5	OP6	
3	3	NORMAL	S5	OP1	OP5	OP6	

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	NORMAL	S1	OP1 OP2 OP3 OP4
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	3	ViewChange	S4	OP1 OP5 OP6
4	3	ViewChange	S5	OP1 OP5 OP6

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

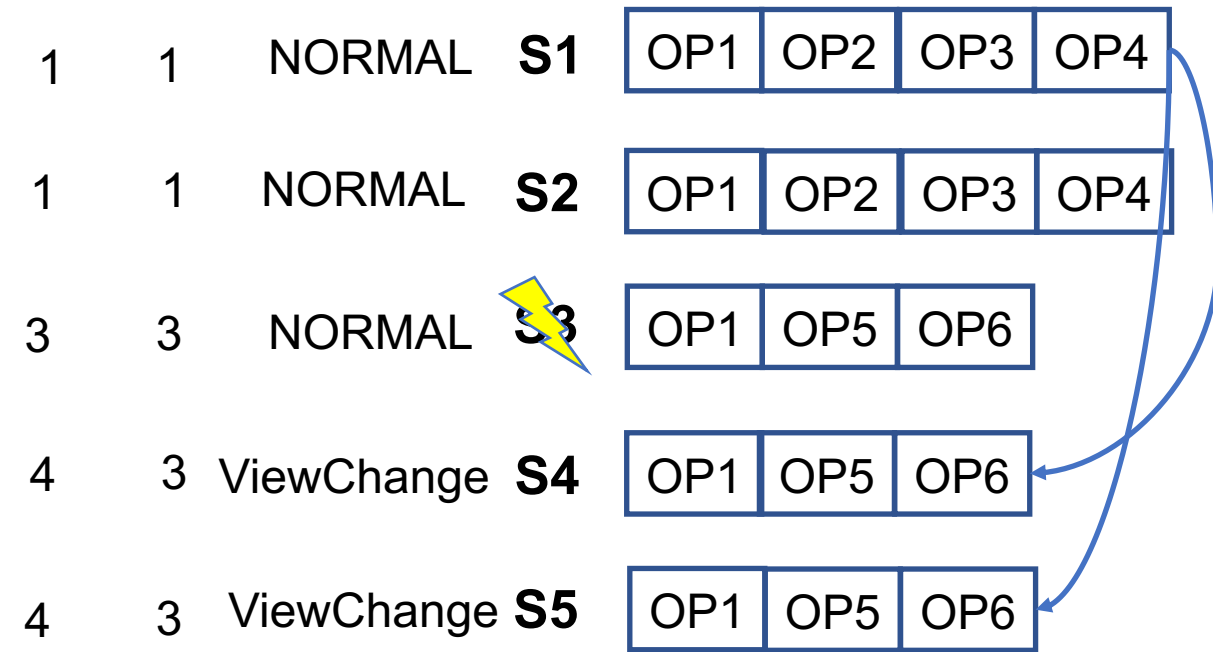
curV lastV status

1	1	NORMAL	S1	OP1 OP2 OP3 OP4
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	3	ViewChange	S4	OP1 OP5 OP6
4	3	ViewChange	S5	OP1 OP5 OP6

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status



1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	Recovery	S1	OP1 OP2 OP3 OP4
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	3	ViewChange	S4	OP1 OP5 OP6
4	3	ViewChange	S5	OP1 OP5 OP6

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

1	1	Recovery	S1	OP1 OP2 OP3 OP4
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	4	NORMAL	S4	OP1 OP5 OP6
4	4	NORMAL	S5	OP1 OP5 OP6

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

4	4	NORMAL	S1	OP1 OP5 OP6
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	4	NORMAL	S4	OP1 OP5 OP6
4	4	NORMAL	S5	OP1 OP5 OP6

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

4	4	NORMAL	S1	OP1 OP5 OP6 OP7
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	4	NORMAL	S4	OP1 OP5 OP6 OP7
4	4	NORMAL	S5	OP1 OP5 OP6 OP7

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

4	4	NORMAL	S1	OP1 OP5 OP6 OP7
1	1	NORMAL	S2	OP1 OP2 OP3 OP4
3	3	NORMAL	 S3	OP1 OP5 OP6
4	4	NORMAL	 S4	OP1 OP5 OP6 OP7
4	4	NORMAL	S5	OP1 OP5 OP6 OP7

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

5	5	NORMAL	S1	OP1 OP5 OP6 OP7
5	5	NORMAL	S2	OP1 OP2 OP6 OP7
3	3	NORMAL	 S3	OP1 OP5 OP6
4	4	NORMAL	 S4	OP1 OP5 OP6 OP7
5	5	NORMAL	S5	OP1 OP5 OP6 OP7

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise I – Answer

curV lastV status

5	5	NORMAL	S1	OP1 OP5 OP6 OP7
5	5	NORMAL	S2	OP1 OP2 OP6 OP7
5	5	NORMAL	S3	OP1 OP5 OP6 OP7
4	4	NORMAL	 S4	OP1 OP5 OP6 OP7
5	5	NORMAL	S5	OP1 OP5 OP6 OP7

1. Network Partitions S1, S2 from Others
2. S1 receives OP2, OP3, OP4
3. OP5 and OP6 got committed
4. S2 receives OP2, OP3, OP4 from S1
5. S3 crashes
6. Network heals
7. S1 tries to commit OP2, OP3, OP4
8. OP7 got committed
9. S4 crashes
10. S3 recovers

Exercise II

Show a possible sequence of events which leads to this state and give value of each “?”s.

curV	lastV	status
------	-------	--------

?	?	?	S1	<table><tr><td>OP1</td><td>OP2</td><td>OP3</td></tr></table>	OP1	OP2	OP3
OP1	OP2	OP3					
?	?	?	S2	<table><tr><td>OP1</td><td>OP4</td></tr></table>	OP1	OP4	
OP1	OP4						
?	?	?	S3	<table><tr><td>OP1</td><td>OP4</td></tr></table>	OP1	OP4	
OP1	OP4						

Exercise II

Show a possible sequence of events which leads to this state and give value of each “?”s.

curV	lastV	status		
1	1	NORMAL	S1	OP1 OP2 OP3
2	2	NORMAL	S2	OP1 OP4
2	2	NORMAL	S3	OP1 OP4

Exercise III

curV	lastV	status		
?	?	NORMAL	S1	OP1OP2OP3
?	?	NORMAL	S2	OP1OP2
?	?	NORMAL	S3	OP1
?	?	NORMAL	S4	OP1
5	5	NORMAL	S5	OP1

Exercise III

curV	lastV	status		
?	?	NORMAL	S1	OP1 OP2 OP3
?	?	NORMAL	S2	OP1 OP2
?	?	NORMAL	S3	OP1
?	?	NORMAL	S4	OP1
5	5	NORMAL	S5	OP1

S1 replicates op1 on all servers.

Network partitions S1, S2 from others.

S5 becomes primary voted by S3, S4 and itself.

Exercise III

curV	lastV	status		
1	1	NORMAL	S1	OP1 OP2 OP3
1	1	NORMAL	S2	OP1 OP2
5	5	NORMAL	S3	OP1
5	5	NORMAL	S4	OP1
5	5	NORMAL	S5	OP1

S1 replicates op1 on all servers.

Network partitions S1, S2 from others.

S5 becomes primary voted by S3, S4 and itself.

Exercise IV

curV lastV status

1	1	NORMAL	S1	OP1
1	1	NORMAL	S2	OP1
1	1	NORMAL	S3	OP1
1	1	NORMAL	S4	OP1
1	1	NORMAL	S5	OP1

Will OP2 and OP3 be eventually committed?

If YES, please give the minimal committed latency.

(Assume the average latency between any two servers a 1 msv and there is **zero latency** for failure detection)

If NO, please tell me why?

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes

Exercise IV

curV	lastV	status		t_0
1	1	NORMAL	S1	OP1 OP2
1	1	NORMAL	S2	OP1
1	1	NORMAL	S3	OP1
1	1	NORMAL	S4	OP1
1	1	NORMAL	S5	OP1

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes

Exercise IV

curV	lastV	status		
				t_0
1	1	NORMAL	S1	OP1 OP2
1	1	NORMAL	S2	OP1
				t_0+1
1	1	NORMAL	S3	OP1 OP2
1	1	NORMAL	S4	OP1
				t_0+1
1	1	NORMAL	S5	OP1 OP2

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes

Exercise IV

curV	lastV	status			t_0
1	1	NORMAL	S1	OP1	OP2
1	1	NORMAL	S2	OP1	
					t_0+1
1	1	NORMAL	S3	OP1	OP2
1	1	NORMAL	S4	OP1	
					t_0+1
1	1	NORMAL	S5	OP1	OP2

1. S1 receives OP2 from client 2. S3 and S5 replicate OP2 3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3 5. Network heals 6. S2 crashes

Exercise IV

curV	lastV	status			t_0
1	1	NORMAL	S1	OP1	OP2
2	2	NORMAL	S2	OP1	OP2
					t_0+5
2	2	NORMAL	S3	OP1	OP2
					t_0+1
2	2	NORMAL	S4	OP1	
					t_0+1
2	2	NORMAL	S5	OP1	OP2

S2 send ViewChange to others t_0+2

S2 becomes primary with newLog t_0+3

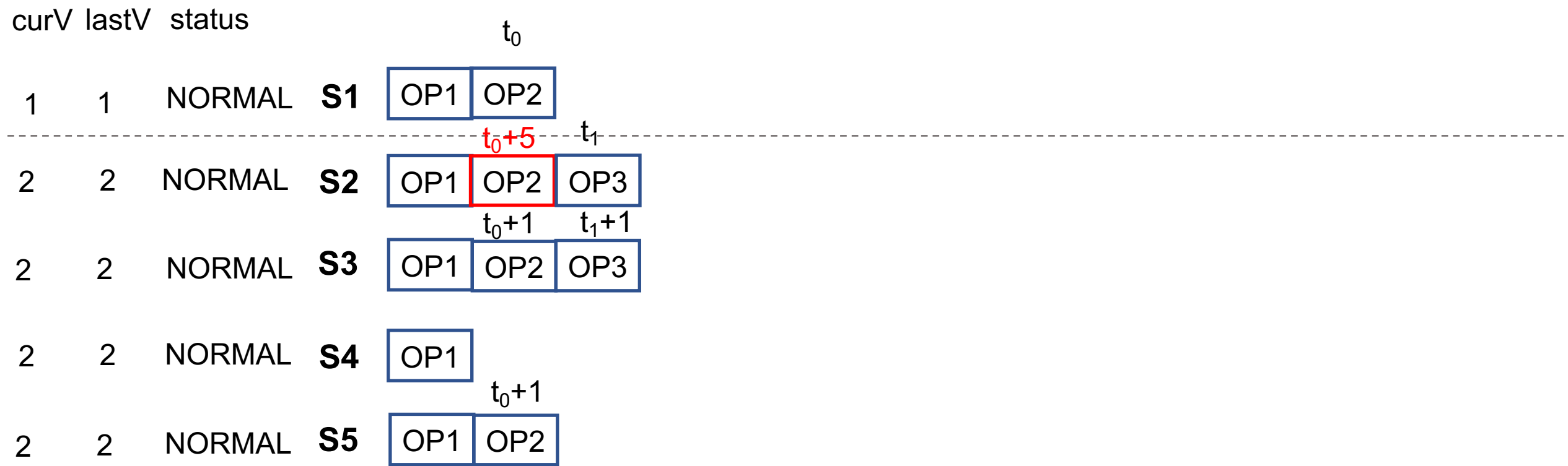
S2 sends prepare of OP2 to others t_0+4

Others reply prepare OK t_0+5

S2 commits OP2

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes

Exercise IV



1. S1 receives OP2 from client 2. S3 and S5 replicate OP2 3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3 5. Network heals 6. S2 crashes

Exercise IV

curV	lastV	status		
1	1	NORMAL	S1	t_0 OP1 OP2
2	2	NORMAL	 S2	t_0+5 OP1 OP2 OP3 t_1
2	2	NORMAL	S3	t_0+1 OP1 OP2 OP3 t_1+1
2	2	NORMAL	S4	OP1
2	2	NORMAL	S5	t_0+1 OP1 OP2

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes

Exercise IV

curV	lastV	status		
1	1	NORMAL	S1	t_0 OP1 OP2
2	2	NORMAL	 S2	t_0+5 t_1 OP1 OP2 OP3
2	2	NORMAL	S3	t_0+1 t_1+5 OP1 OP2 OP3
2	2	NORMAL	S4	OP1
2	2	NORMAL	S5	t_0+1 OP1 OP2

1. S1 receives OP2 from client
2. S3 and S5 replicate OP2
3. Network partitions S1 from others
4. S2 receives OP3, and replicate it at S3
5. Network heals
6. S2 crashes