

# Byzantine and Multi-writer K-quorums

Amitanand S. Aiyer

Lorenzo Alvisi

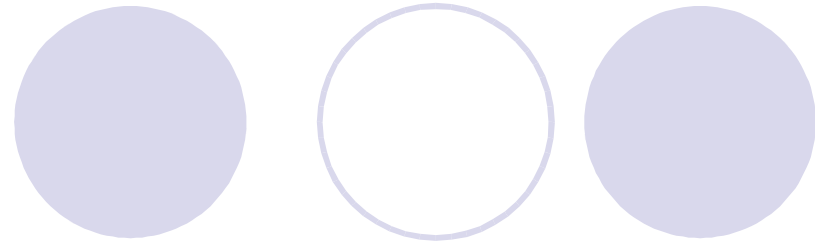
Rida A. Bazzi

*UT-Austin,*

*UT-Austin,*

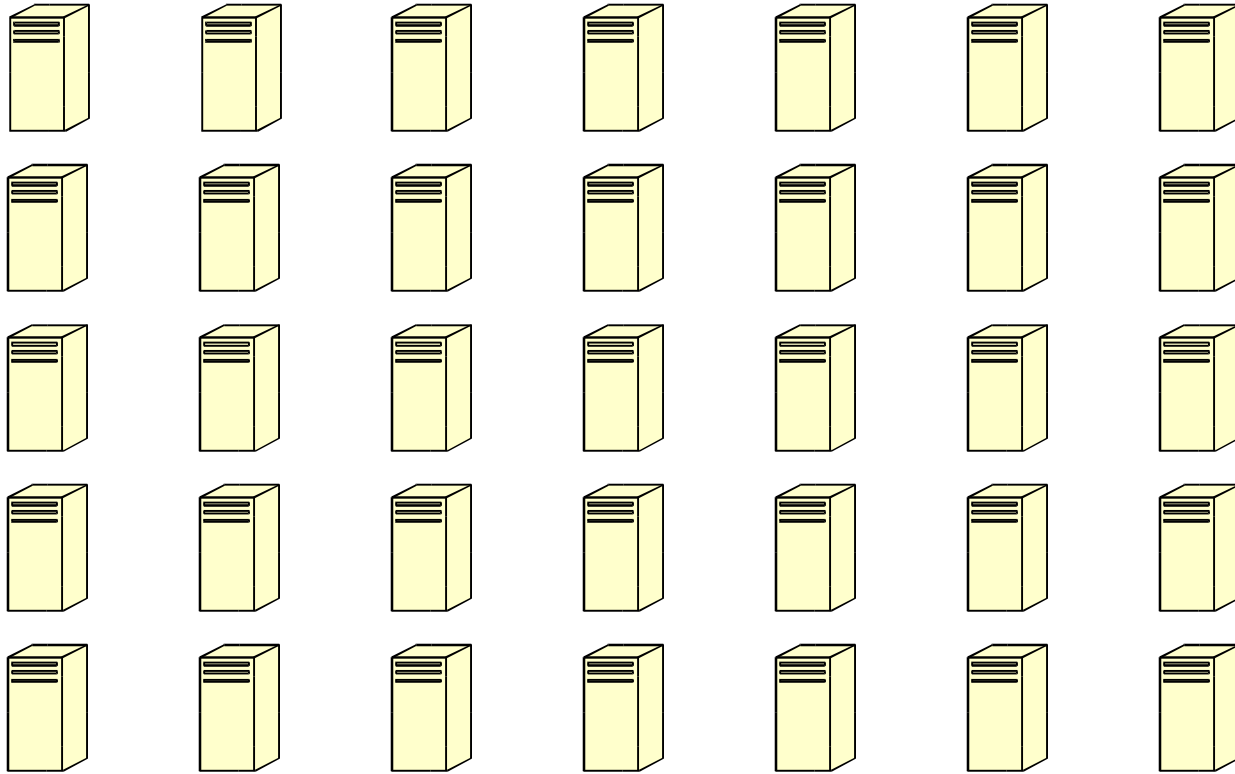
*ASU*

# Quorum Systems

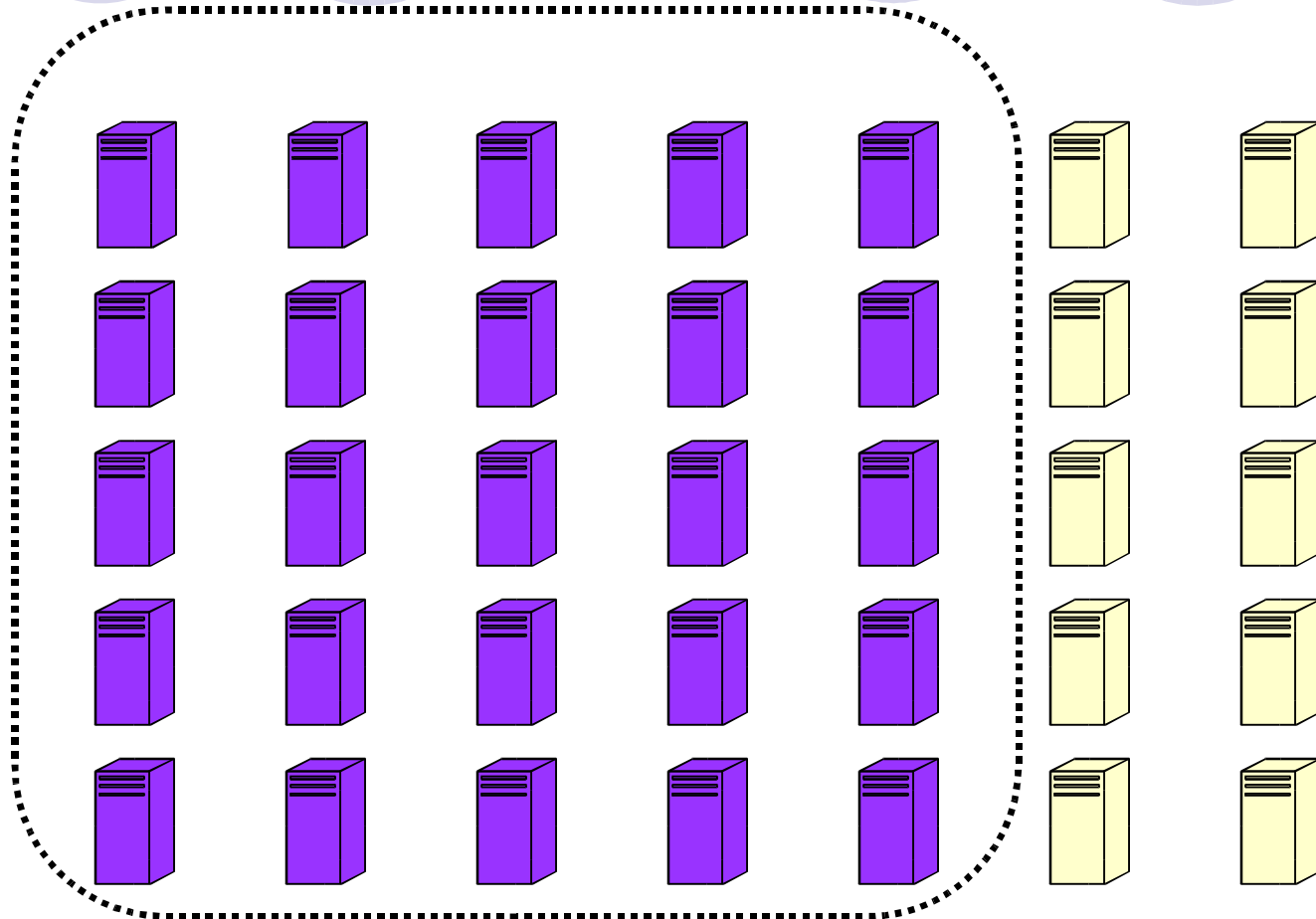


- Collection of servers, ██████████ organized into sets called quorums
- Write Operation
  - Write to all servers in a write quorum
- Read Operation
  - Read from all servers in a read quorum

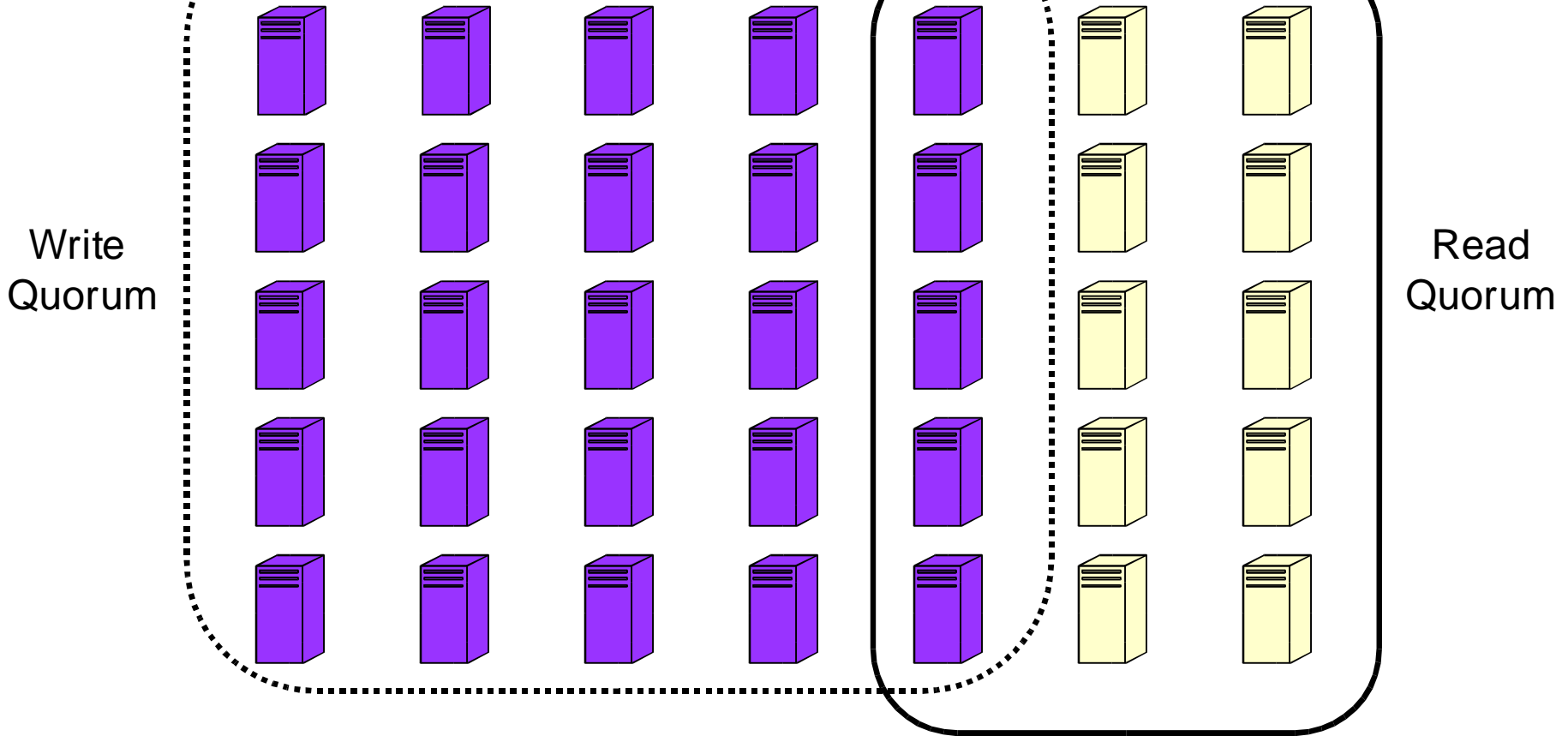
# Strict Quorum System



# Strict Quorum System



# Strict Quorum System



# Strict Quorum Systems



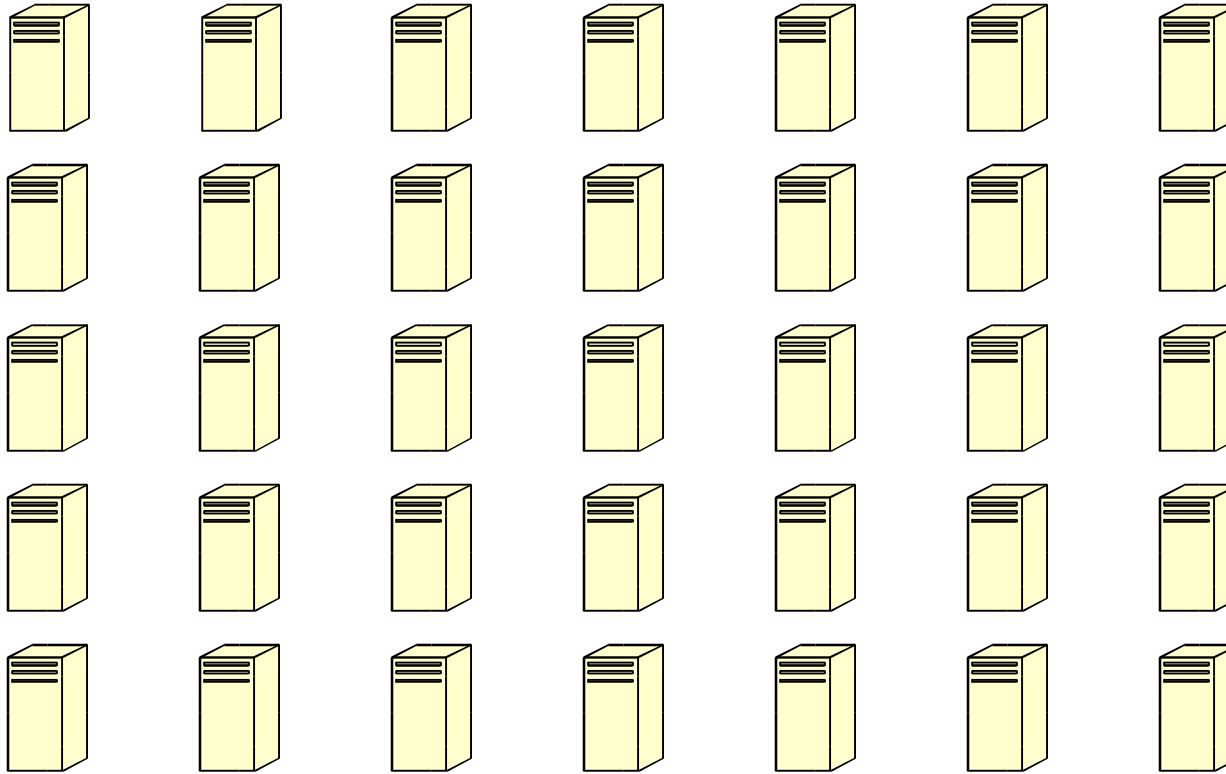
- Benefits

- Any read and write quorums intersect
  - Reader guaranteed to read the latest value

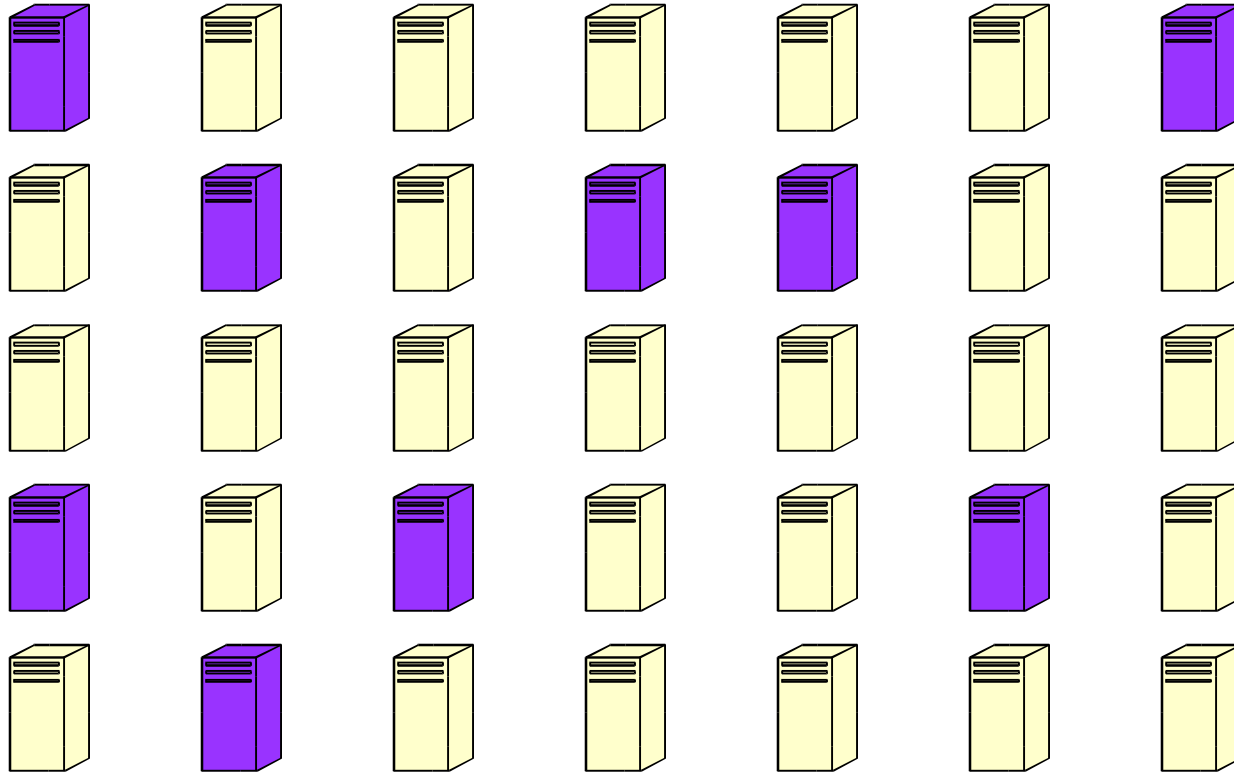
- Limitations

- Low availability

# Probabilistic Quorum System

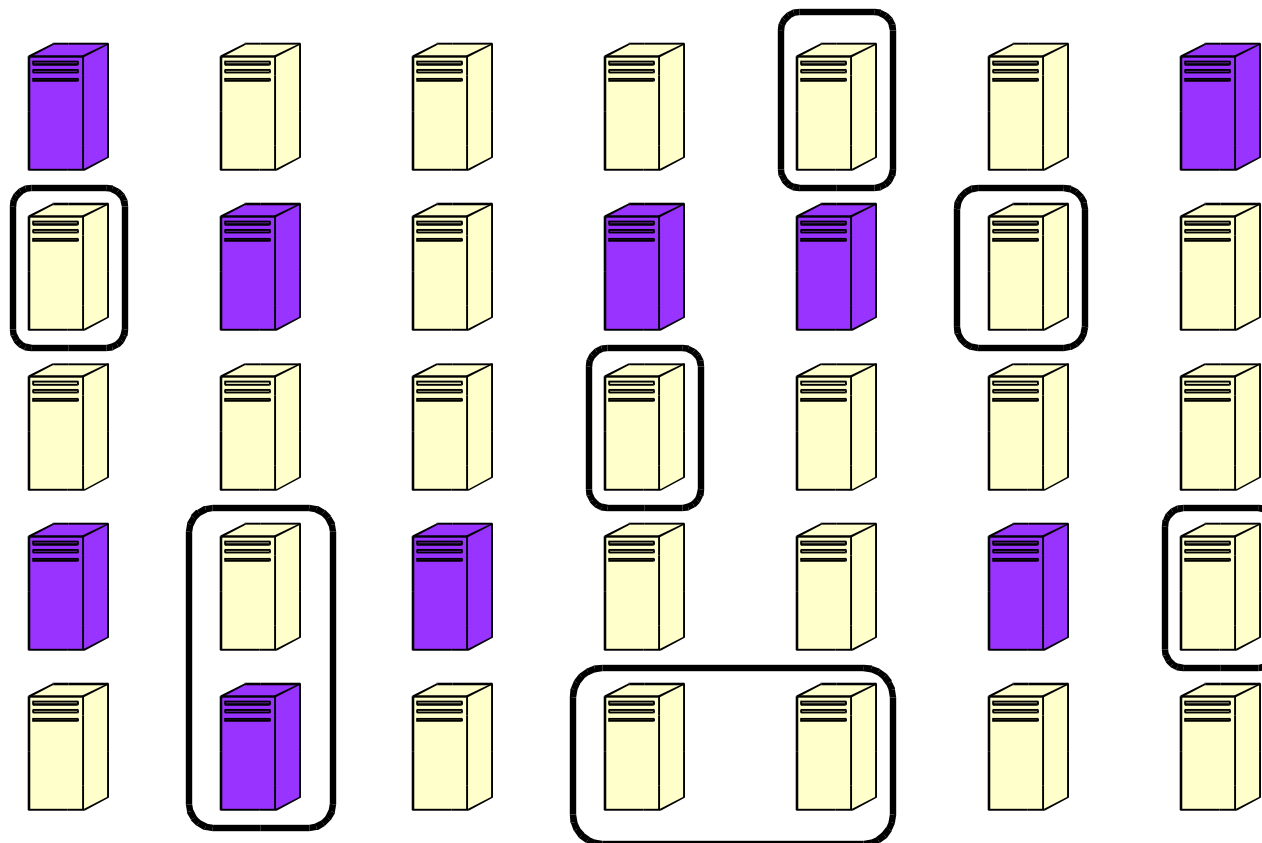


# Probabilistic Quorum System

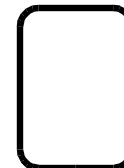


Write  
Quorum

# Probabilistic Quorum System



Write  
Quorum



Read  
Quorum



# Probabilistic Quorum System

- Benefits

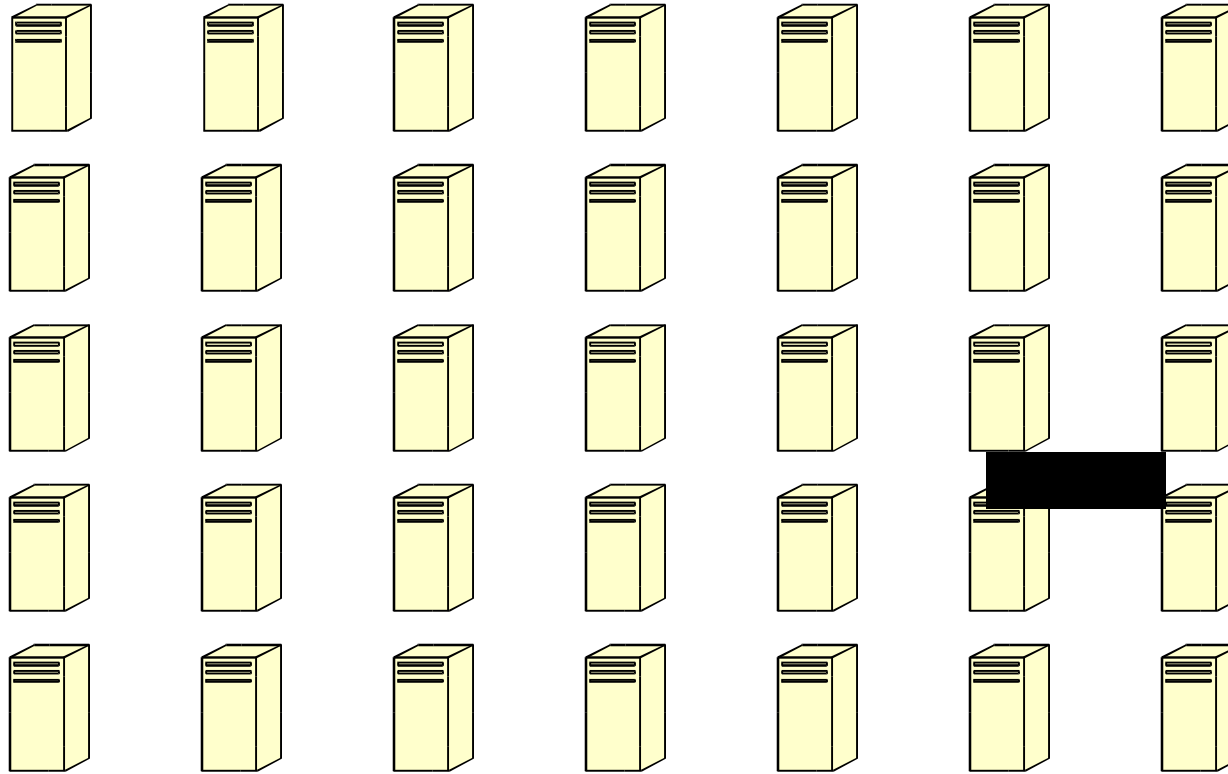
- High availability

- Limitations

- Only probabilistic intersection
  - Readers may return old values
- No deterministic bound on staleness of values
- Adversarial scheduler

# K-Quorum System

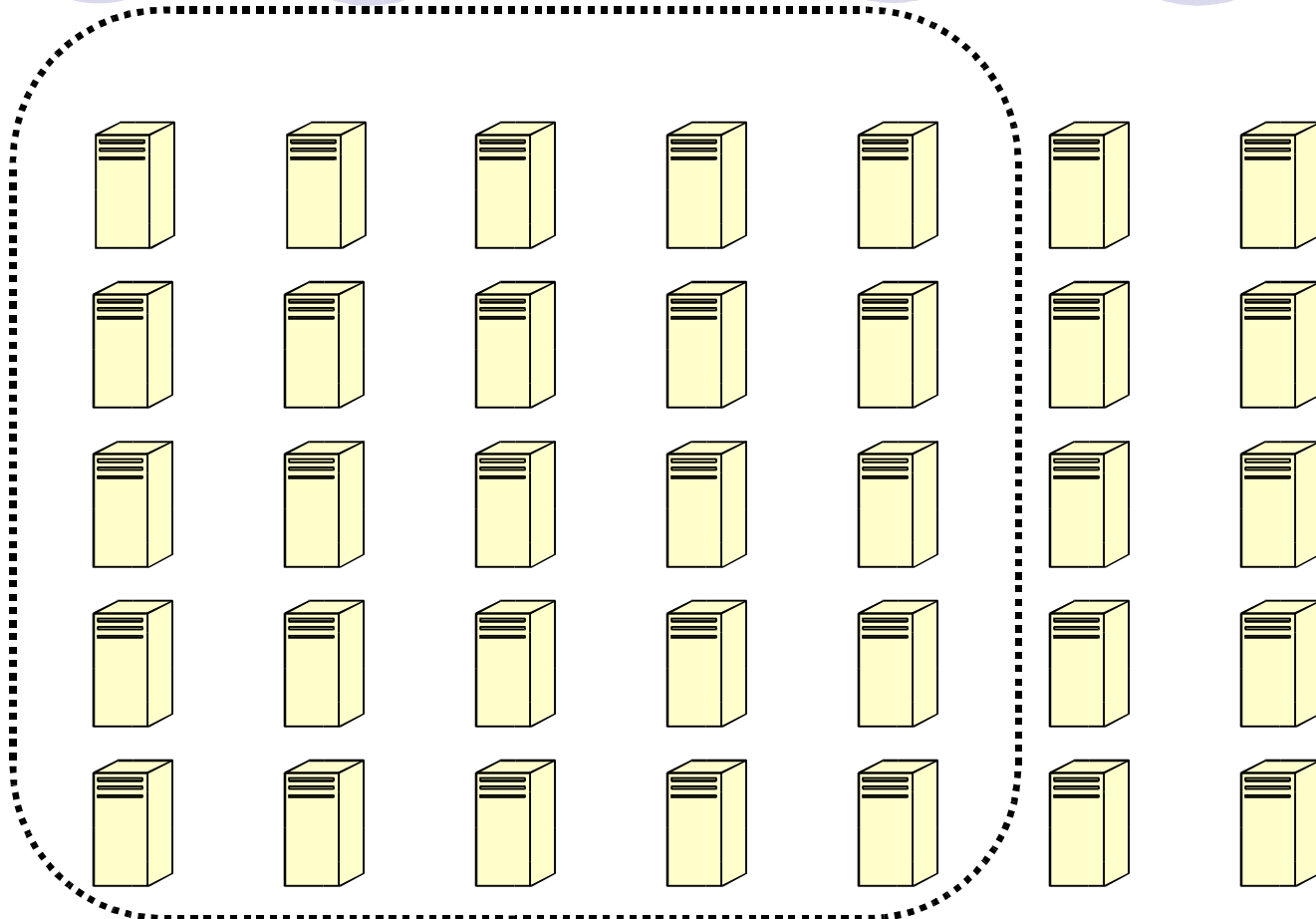
[DISC 2005]



# K-Quorum System

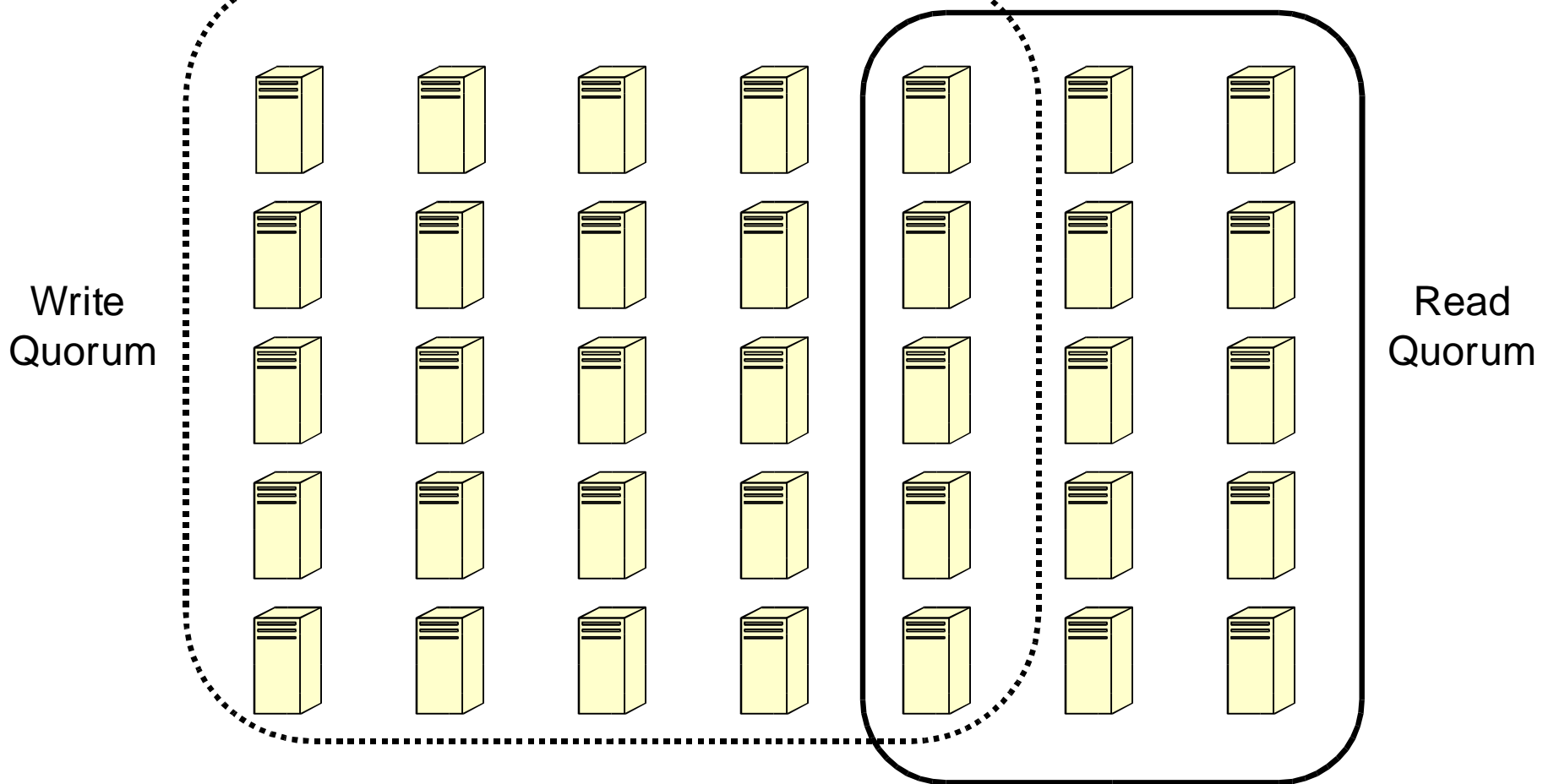
[DISC 2005]

Write  
Quorum



# K-Quorum System

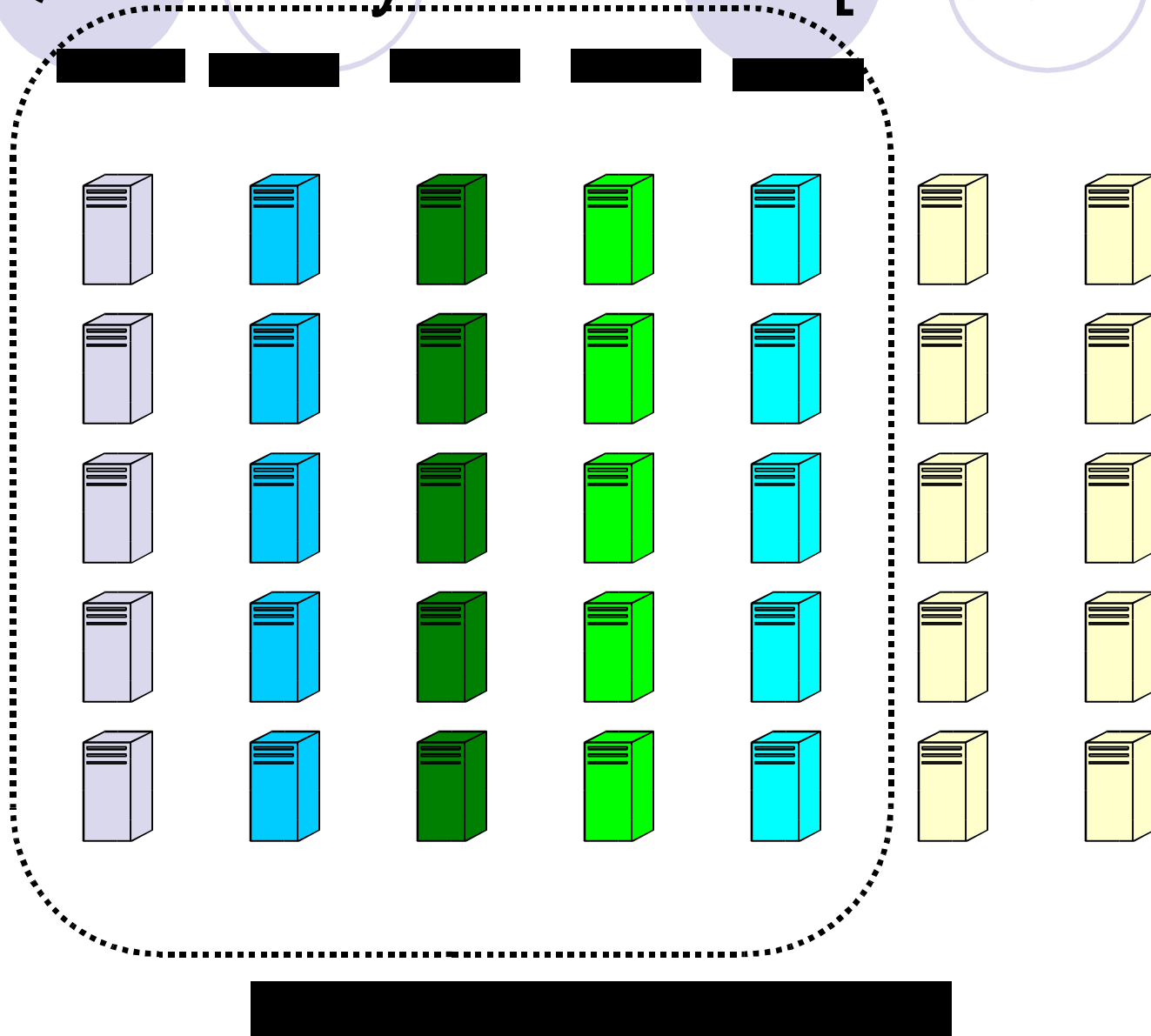
[DISC 2005]



# K-Quorum System

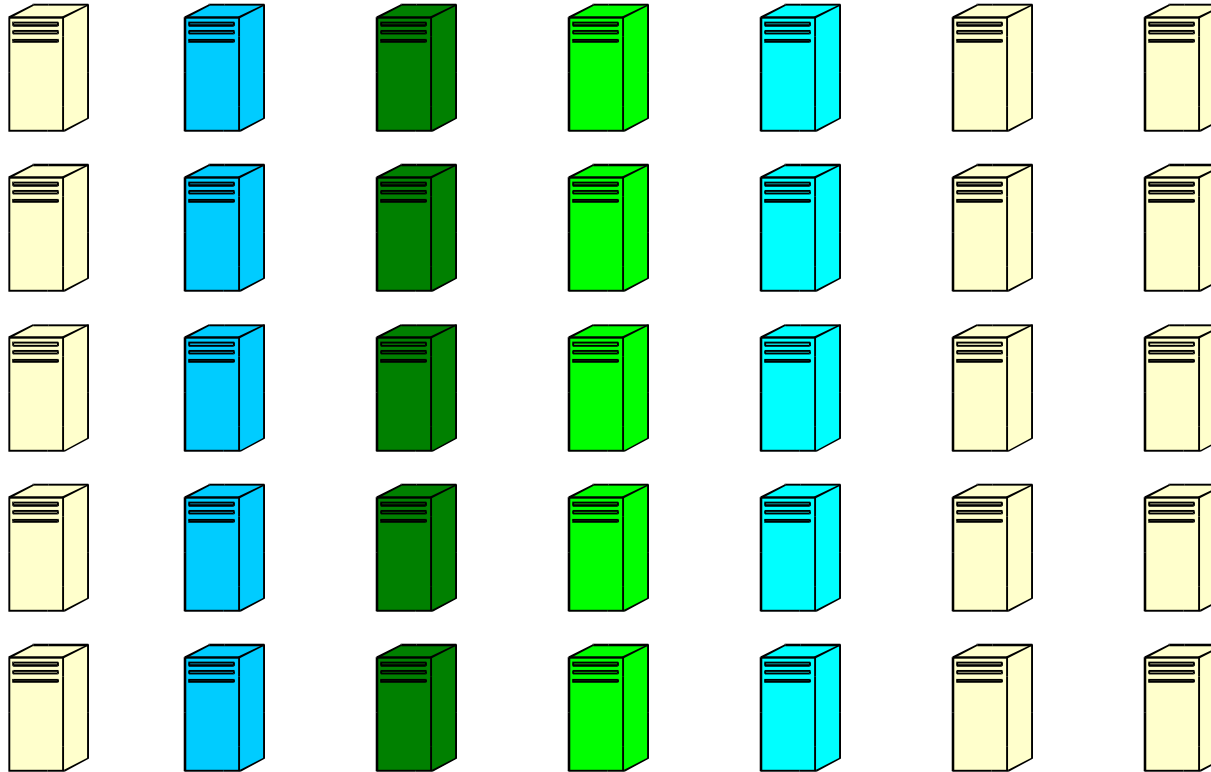
[DISC 2005]

Write  
Quorum



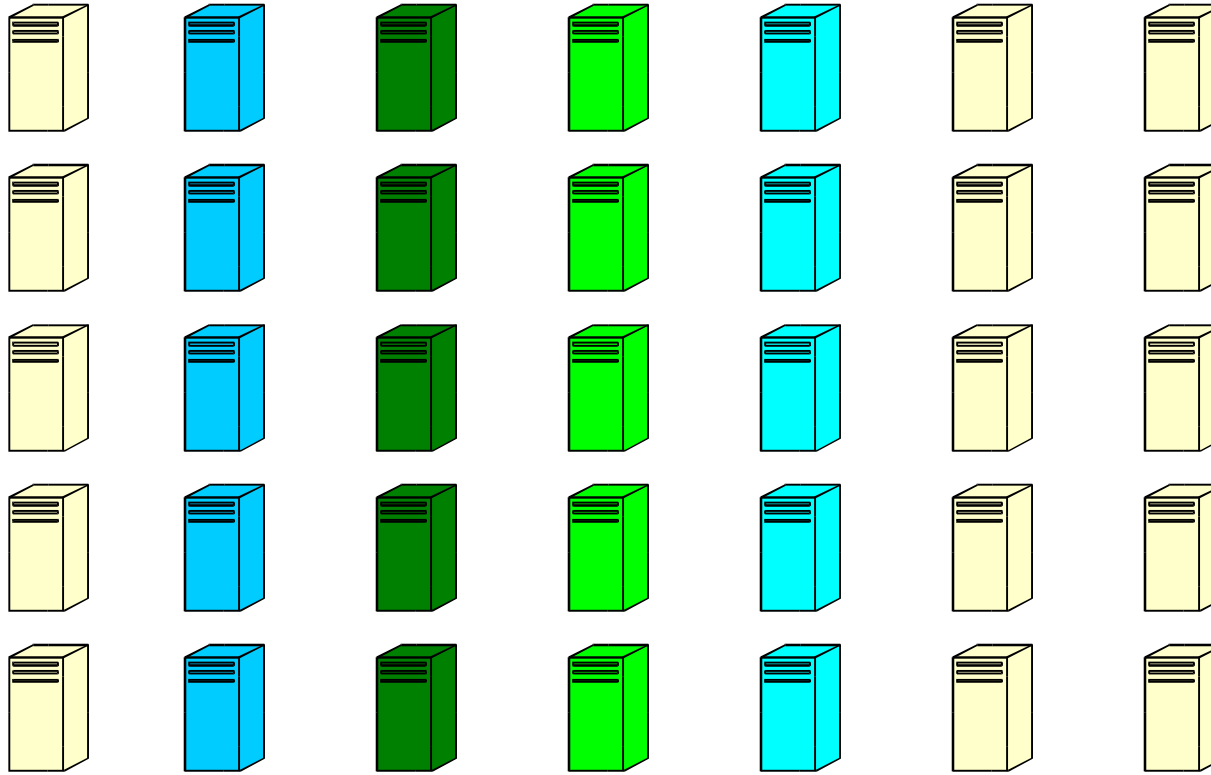
# K-Quorum System

[DISC 2005]



# K-Quorum System

[DISC 2005]



Eligible Servers

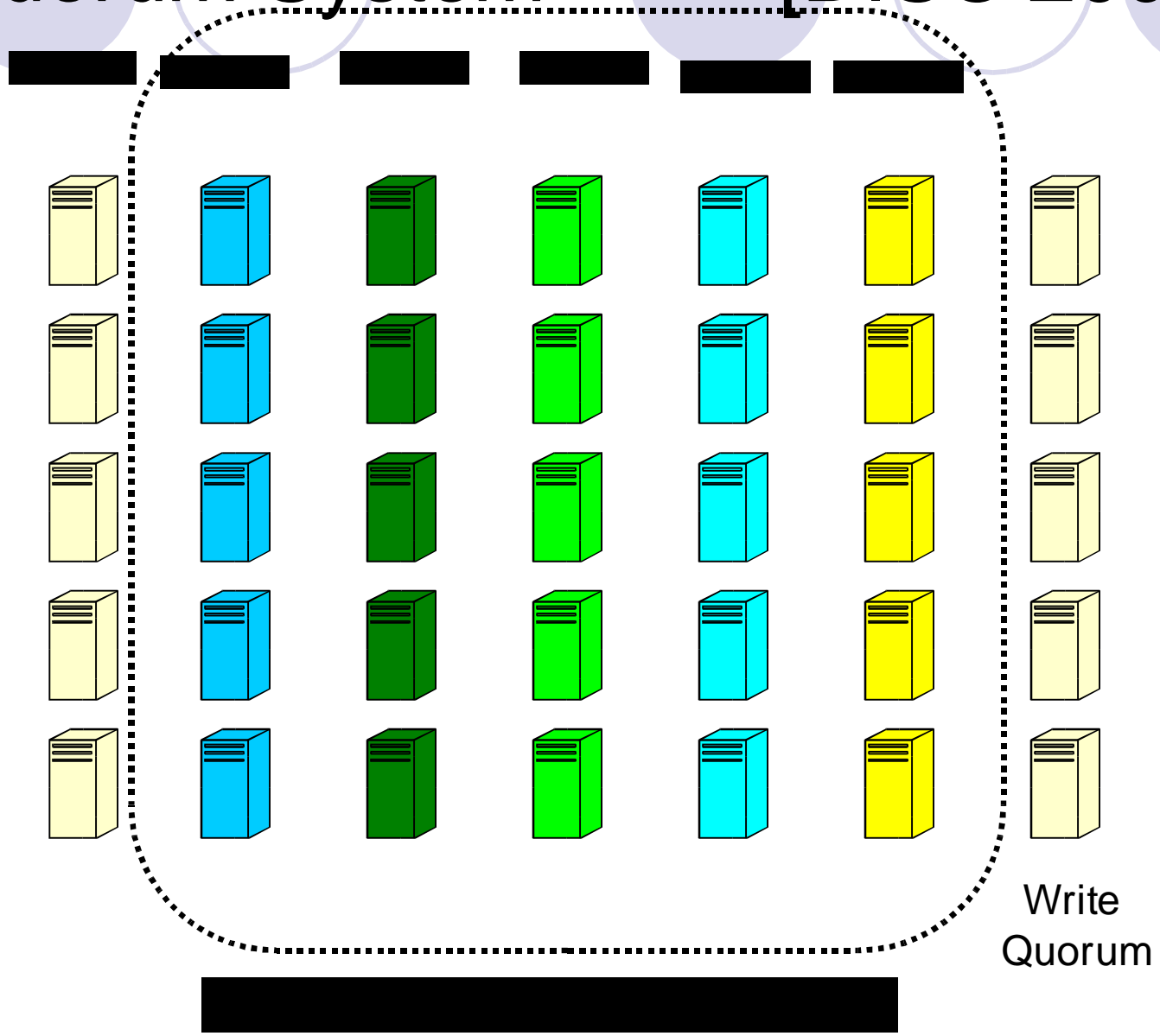


Eligible Servers



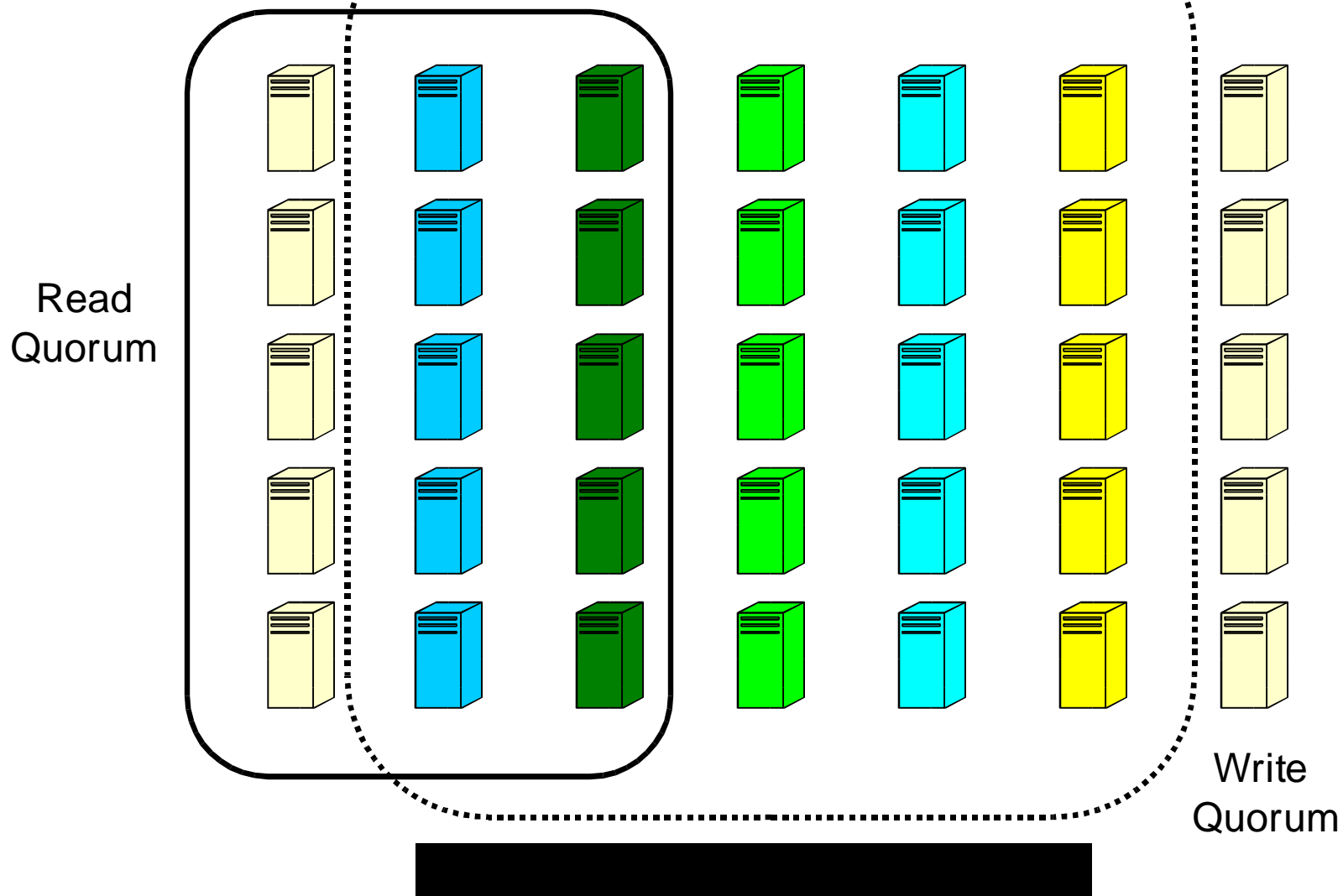
# K-Quorum System

[DISC 2005]



# K-Quorum System

[DISC 2005]



# K-Quorum System

[DISC 2005]

- Benefits

- Can provide high availability
  - When writes are infrequent
- Guarantees bounded staleness

- Limitations

- Only applicable to a single writer scenario
- Only applicable to non-malicious failures

# In this paper

- K-quorums tolerating Byzantine servers
- Multi-writer K-quorums for  $w$  writers
  - Built over Single writer K-quorums
    - With staleness bound of  $\Delta$
  - Staleness bound of  $\Delta$
- Lower Bound
  - Staleness of at least  $\Delta$

# In this paper

- K-quorums tolerating Byzantine servers
- Multi-writer K-quorums for  $w$  writers
  - Built over Single writer K-quorums
    - With staleness bound of  $\frac{w}{K}$
  - Staleness bound of  $\frac{w}{K}$
- Lower Bound
  - Staleness of at least  $\frac{w}{K}$

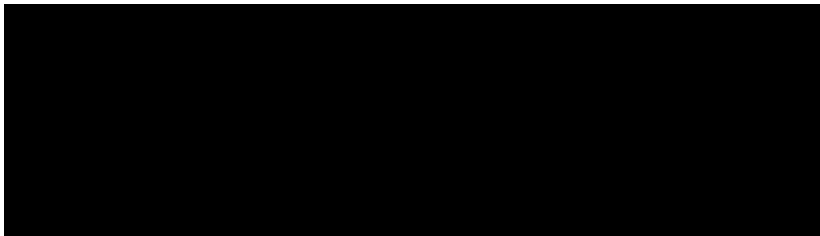
# Byzantine tolerant K-quorums



- Any  $k$  consecutive *partial-write-quorums* form a *write quorum*
- *Write quorum* and *Read quorum* intersect in at least  $k$  nodes

# Write Operation

- Writes  $\lfloor \frac{n}{2} \rfloor$  tuples,
  - One for each of the last  $\lfloor \frac{n}{2} \rfloor$  writes.
- Waits for acks from a *partial-write-quorum*



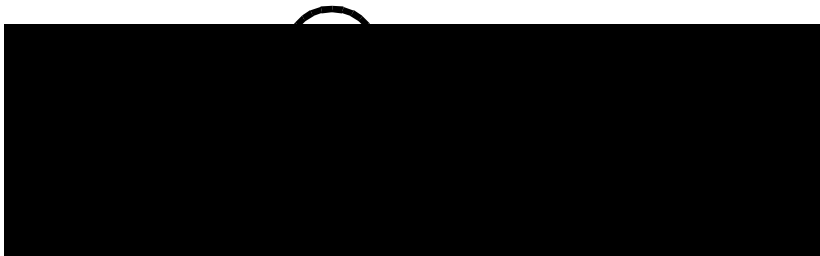
# Write Operation

- Writes  $\lfloor \frac{n}{2} \rfloor$  tuples,
  - One for each of the last  $\lfloor \frac{n}{2} \rfloor$  writes.
- Waits for acks from a *partial-write-quorum*



# Write Operation

- Writes  $\lfloor \frac{n}{2} \rfloor$  tuples,
  - One for each of the last  $\lfloor \frac{n}{2} \rfloor$  writes.
- Waits for acks from a *partial-write-quorum*



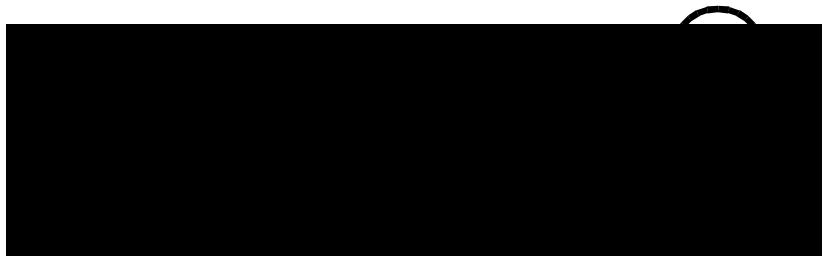
# Write Operation

- Writes  $\lceil \frac{n}{2} \rceil$  tuples,
  - One for each of the last  $\lceil \frac{n}{2} \rceil$  writes.
- Waits for acks from a *partial-write-quorum*



# Write Operation

- Writes  $\lceil \frac{n}{2} \rceil$  tuples,
  - One for each of the last  $\lceil \frac{n}{2} \rceil$  writes.
- Waits for acks from a *partial-write-quorum*

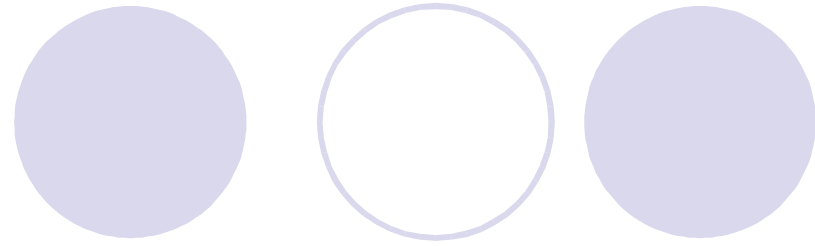


# Write Operation

- Writes  $\lfloor \frac{n}{2} \rfloor$  tuples,
  - One for each of the last  $\lfloor \frac{n}{2} \rfloor$  writes.
- Waits for acks from a *partial-write-quorum*



# Read Operation

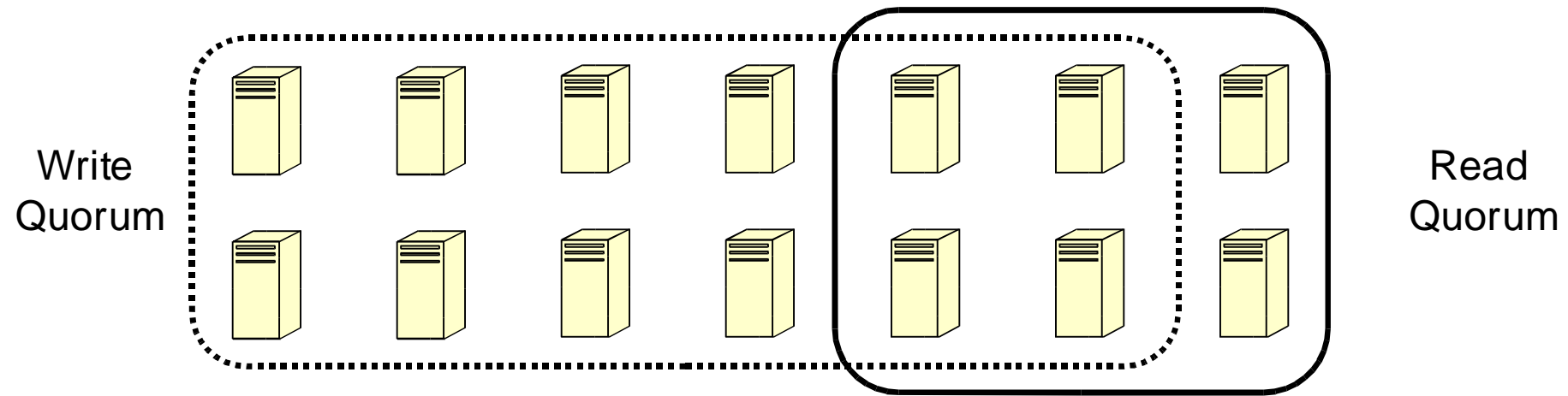
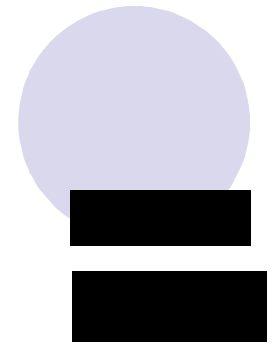


- Contact a *Read quorum*
- Collect responses from *Read quorum*
- Identify a recently written set of  $\lfloor \frac{n}{2} \rfloor$  tuples
- Write back the value to a *partial-write-quorum*

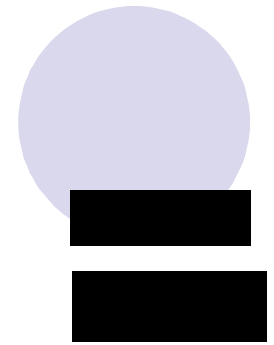
# Byzantine K-Quorum Example

- 14 nodes
- Staleness bound  $\leq 1$
- Byzantine failures  $\leq 2$
- Write Quorum Size = 12
- Read Quorum Size = 6

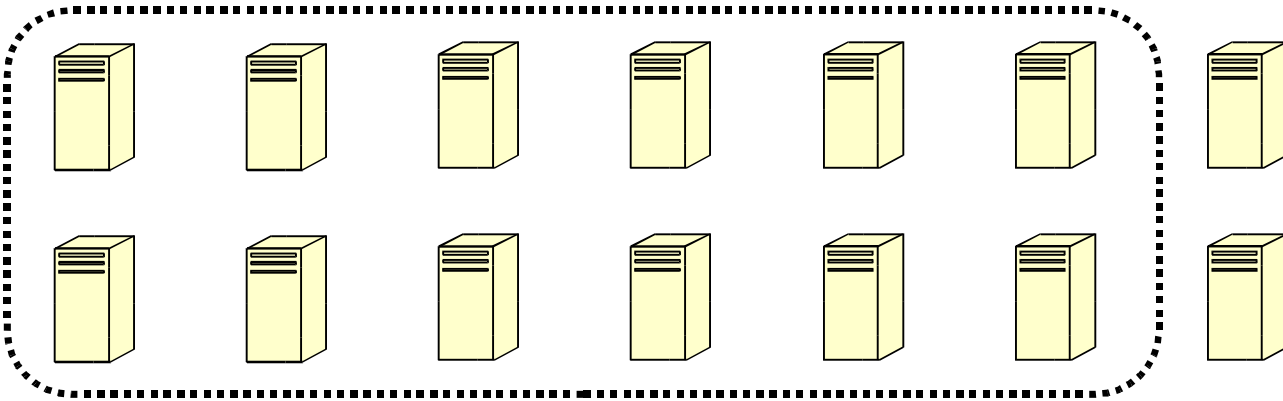
# Byzantine K-Quorum Example



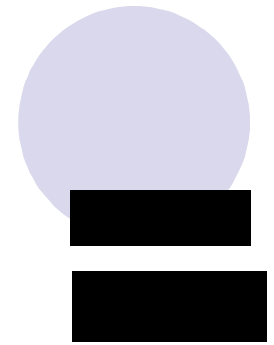
# Byzantine K-Quorum Example



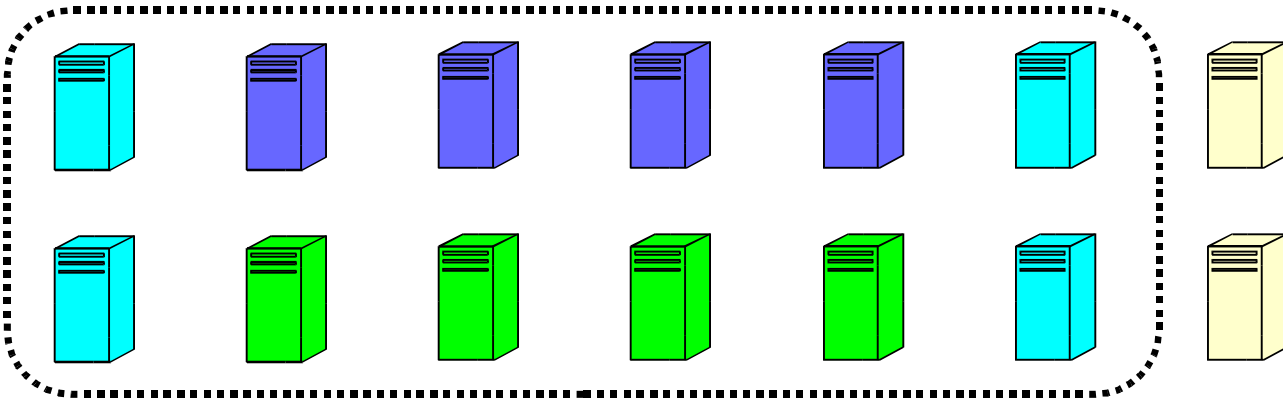
Write  
Quorum



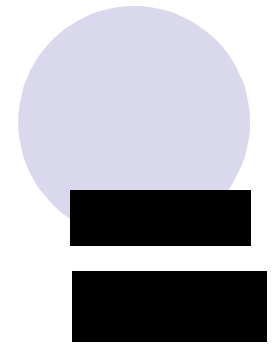
# Byzantine K-Quorum Example



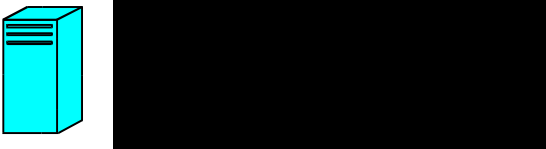
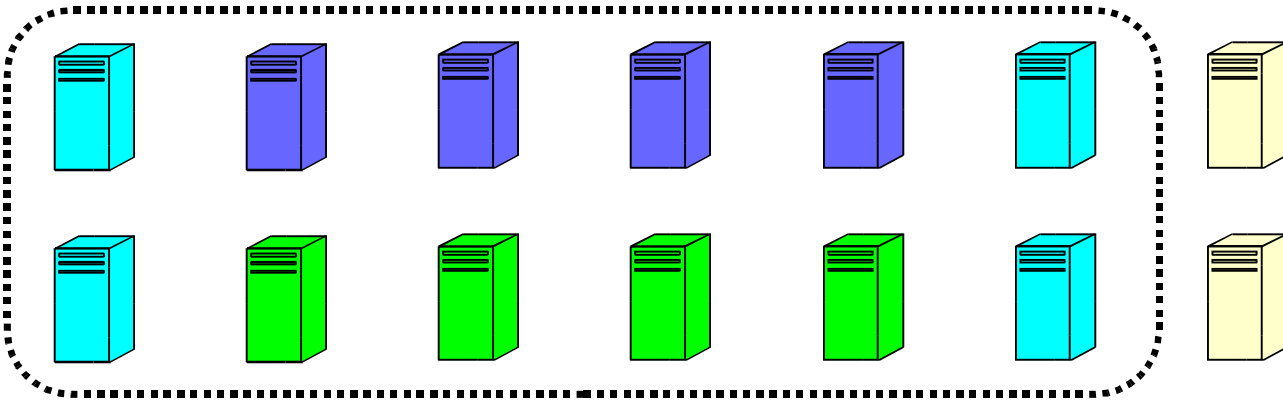
Write  
Quorum



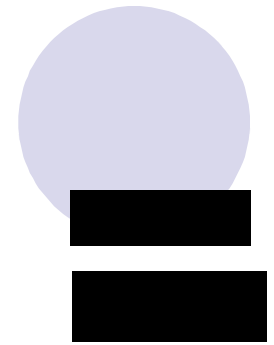
# Byzantine K-Quorum Example



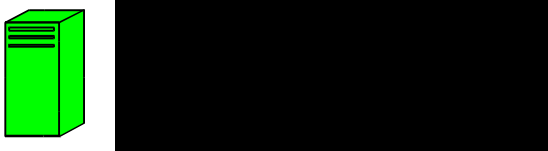
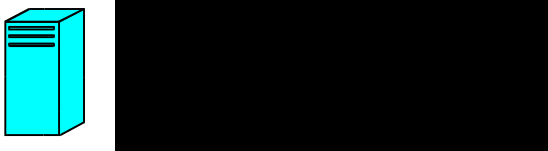
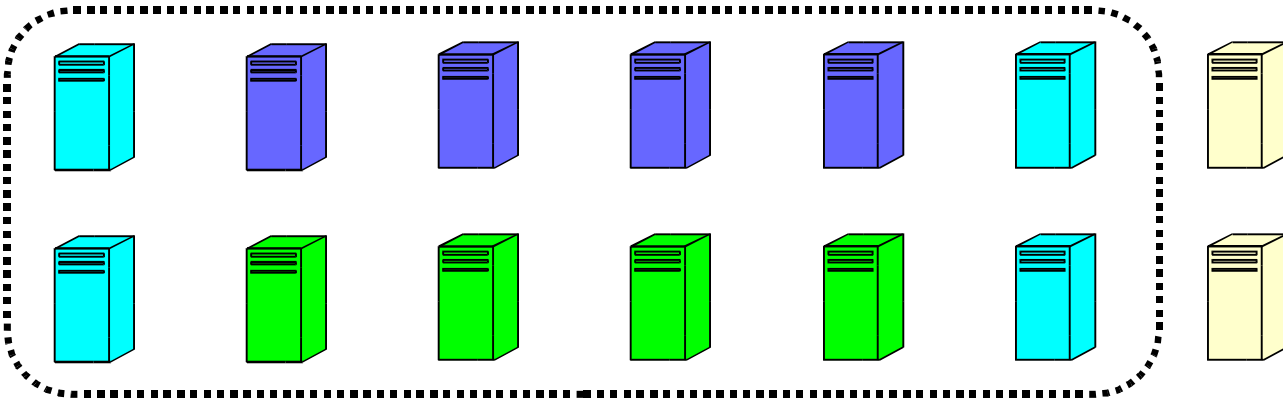
Write  
Quorum



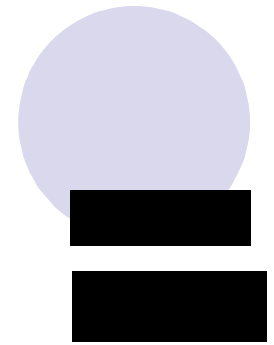
# Byzantine K-Quorum Example



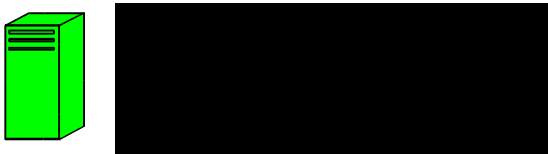
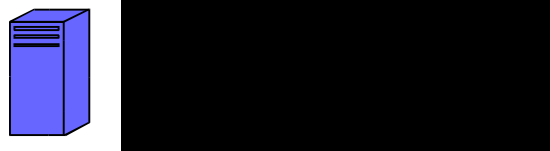
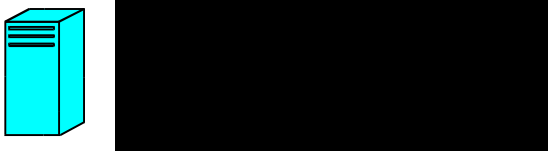
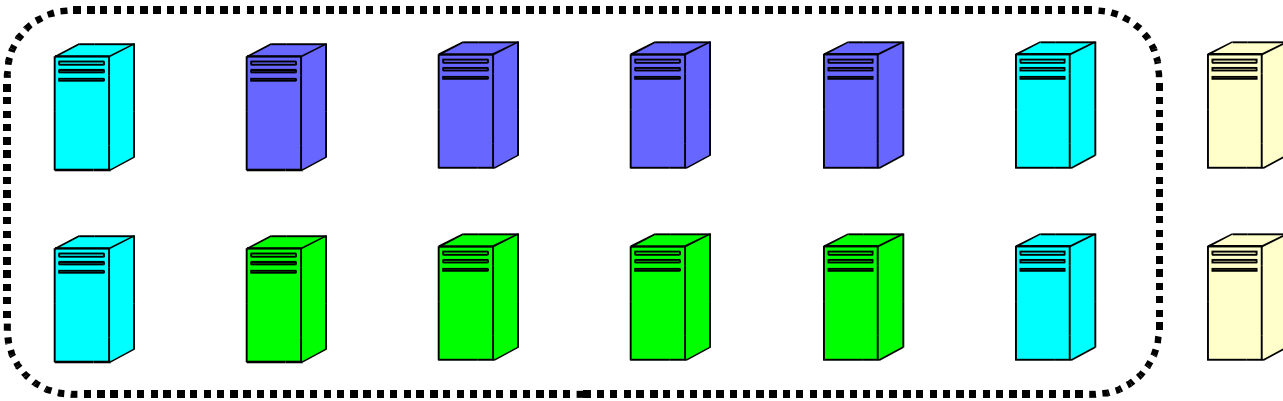
Write  
Quorum



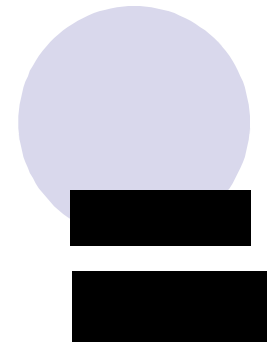
# Byzantine K-Quorum Example



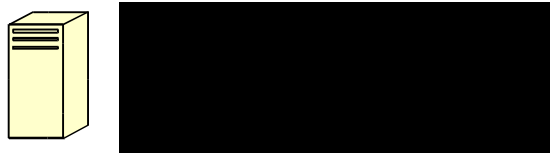
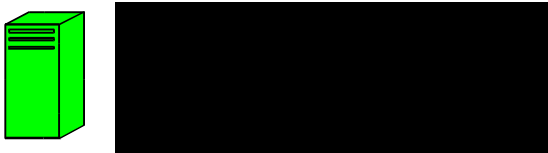
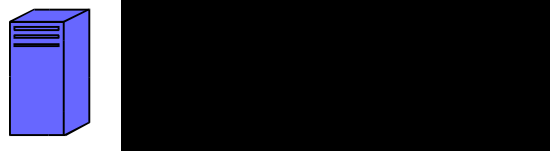
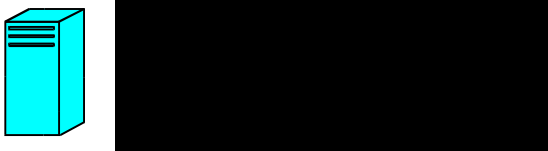
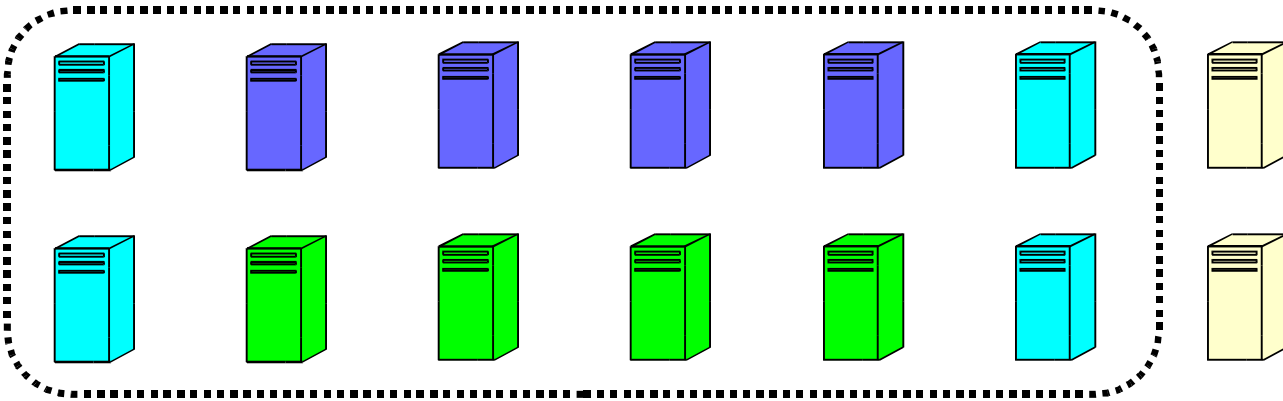
Write  
Quorum



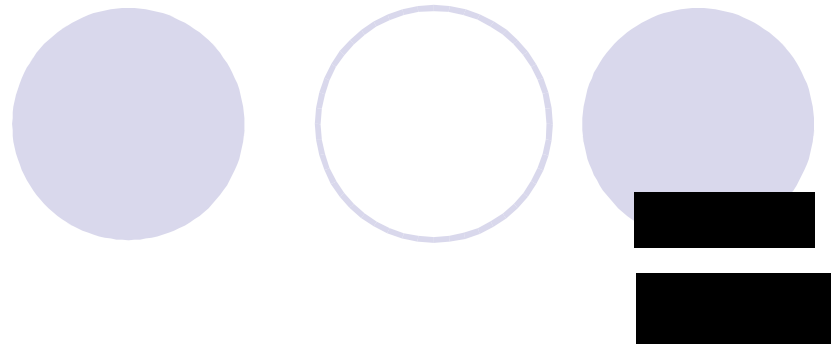
# Byzantine K-Quorum Example



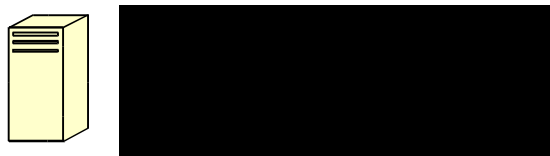
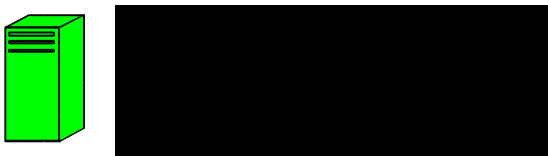
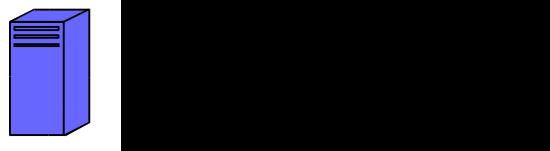
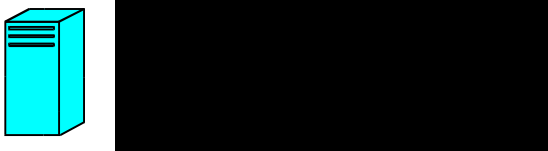
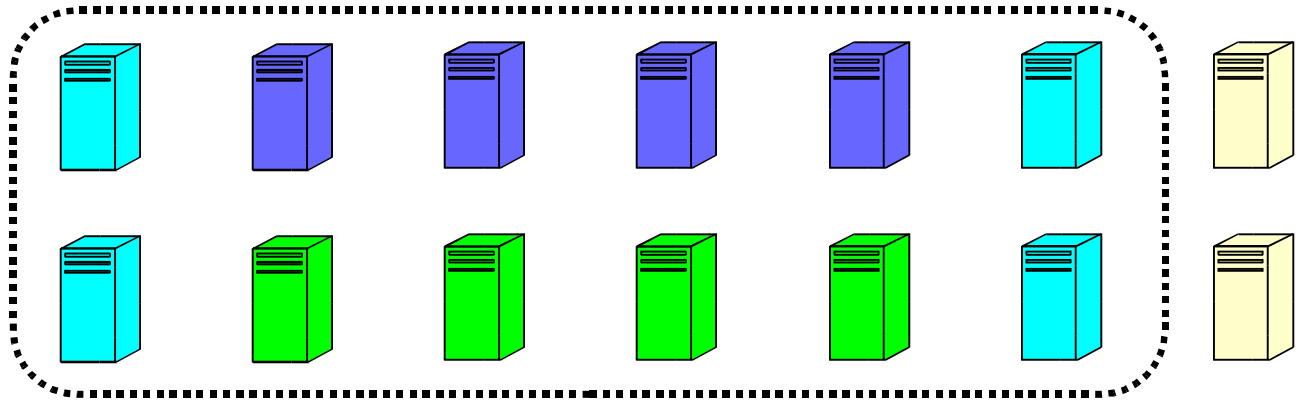
Write  
Quorum



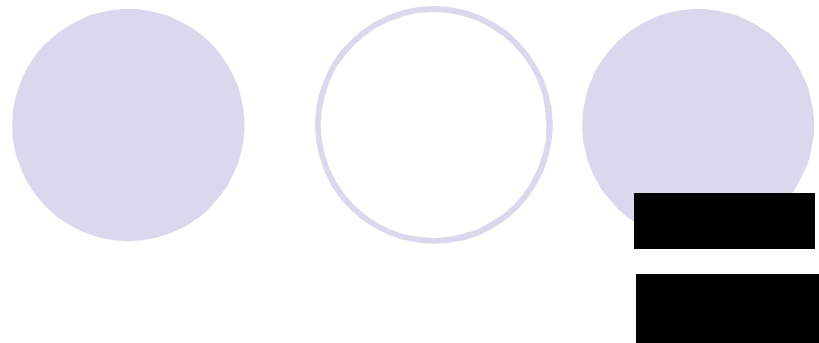
# Read Operation



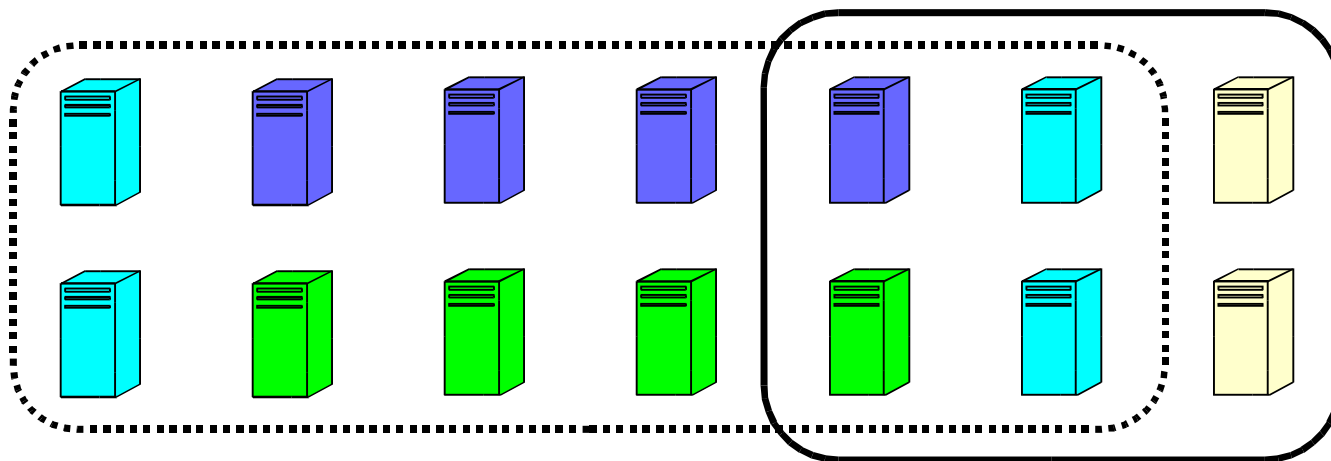
Write  
Quorum



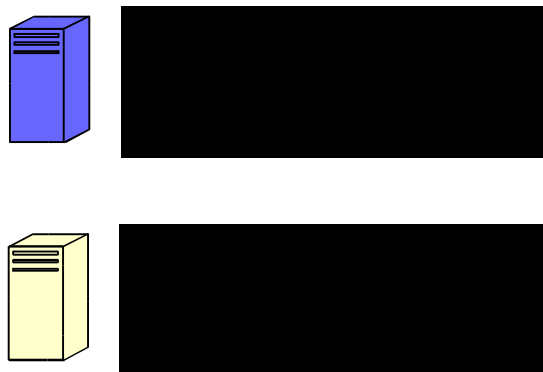
# Read Operation



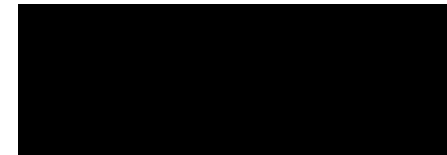
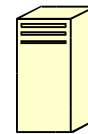
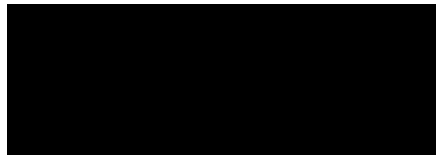
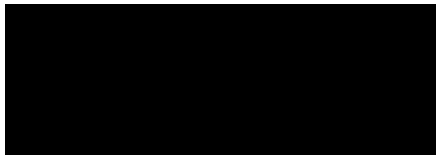
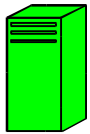
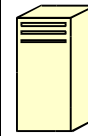
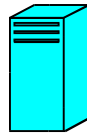
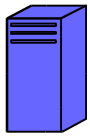
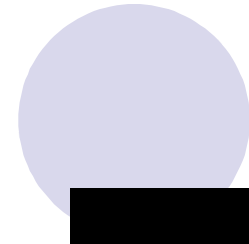
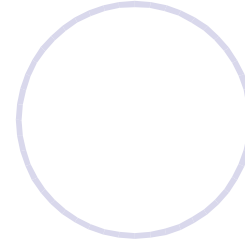
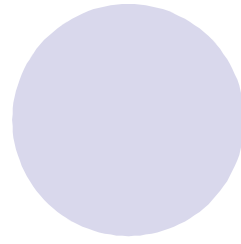
Write  
Quorum



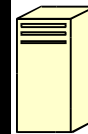
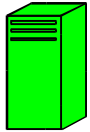
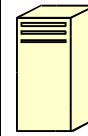
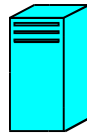
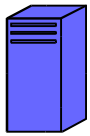
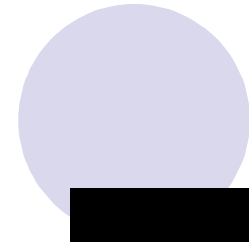
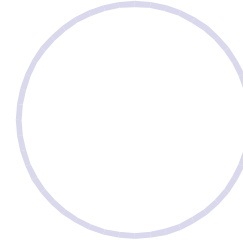
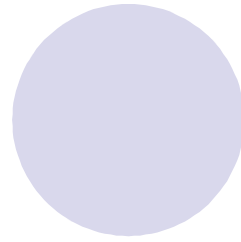
Read  
Quorum



# Reader's Values



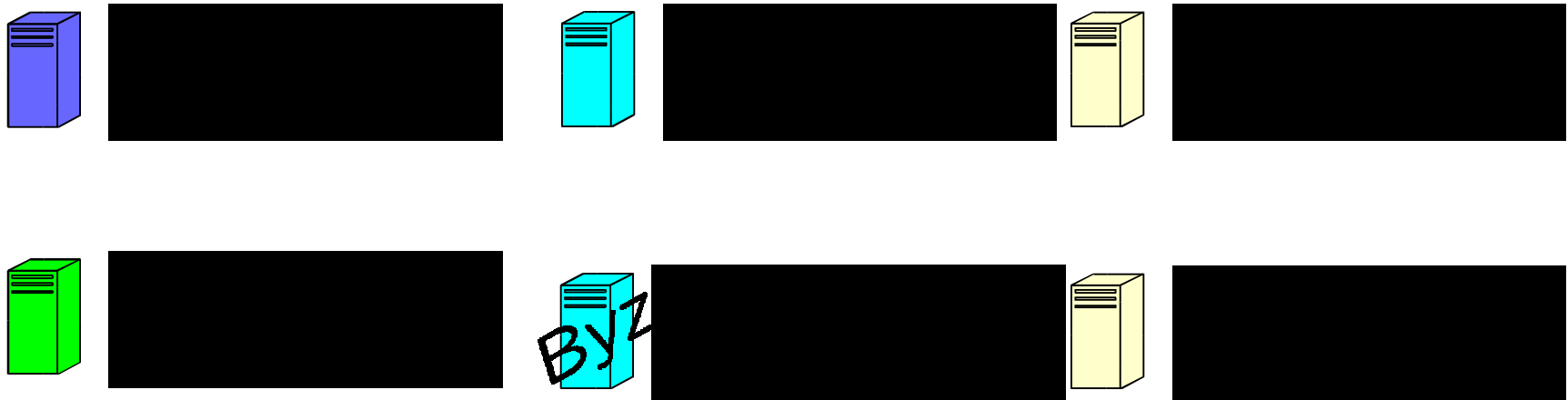
# Reader's Values



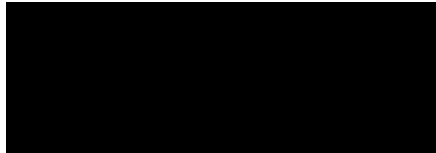
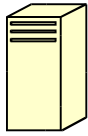
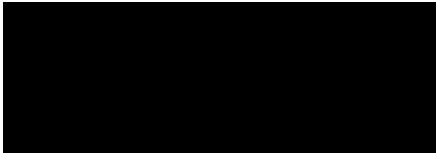
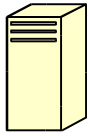
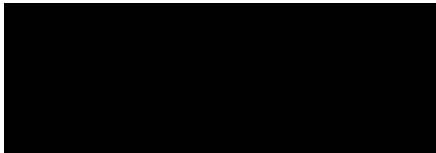
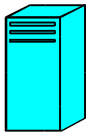
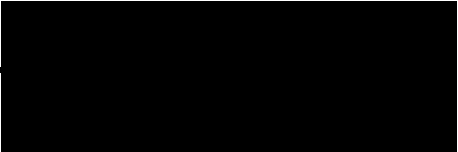
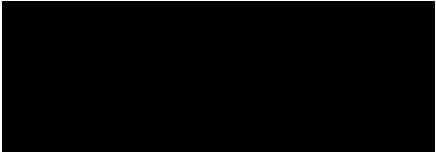
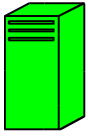
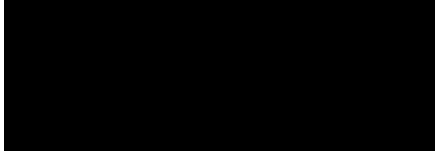
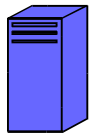
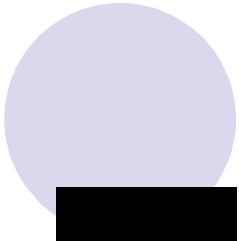
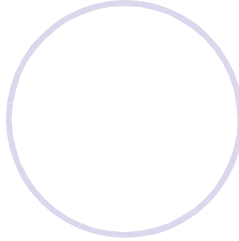
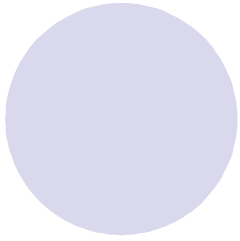
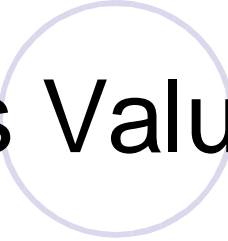
# Reader's Values



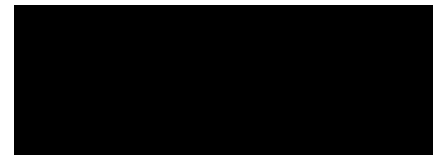
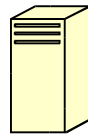
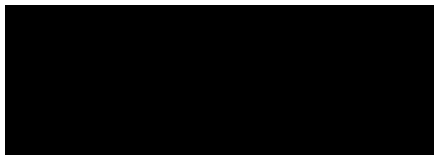
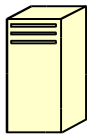
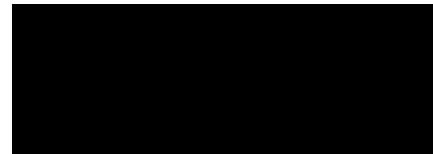
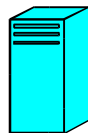
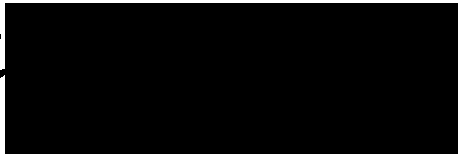
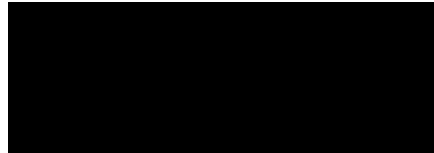
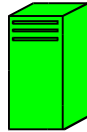
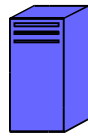
- Sort responses wrt. timestamp



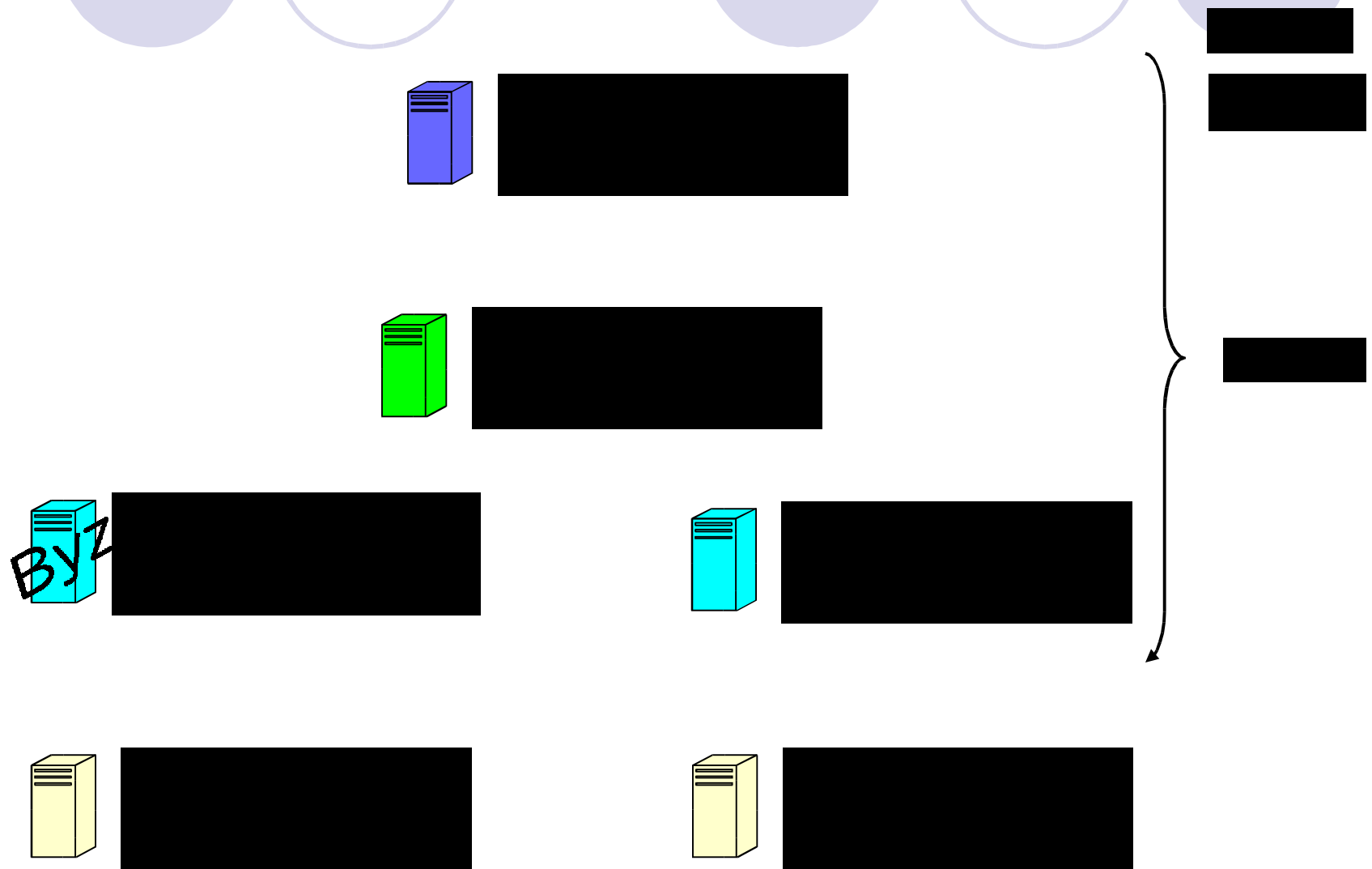
# Reader's Values



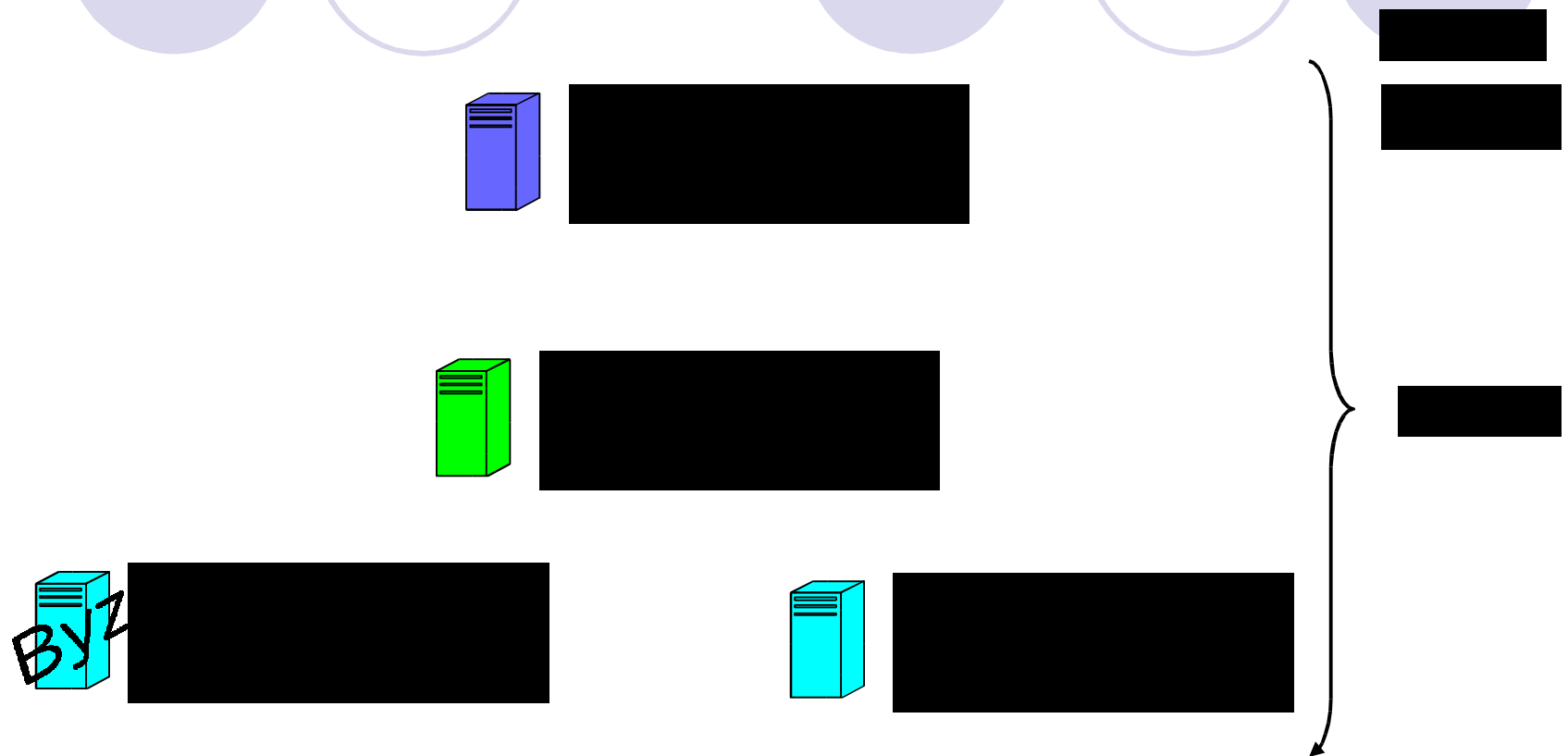
# Removing possibly old values



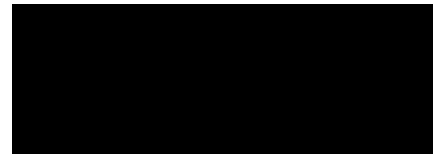
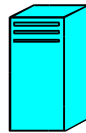
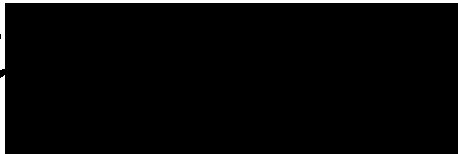
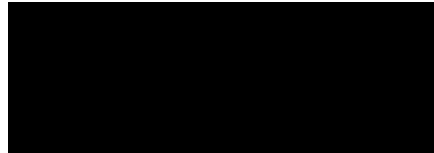
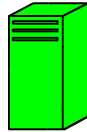
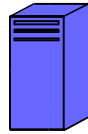
# Removing possibly old values



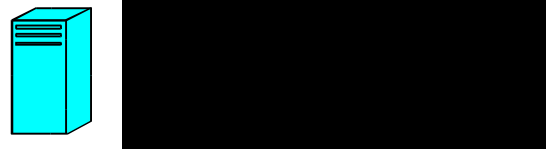
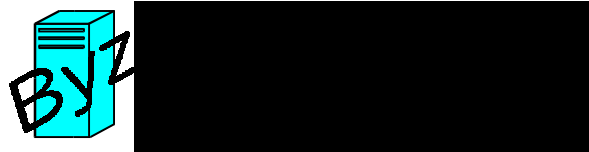
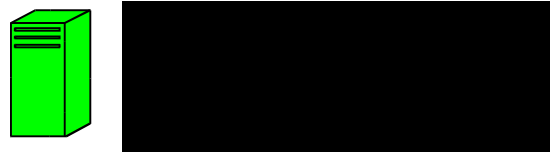
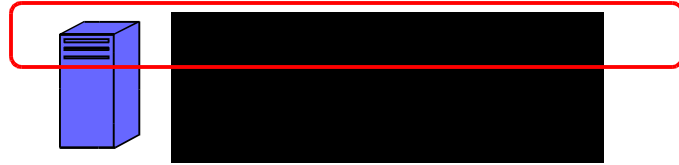
# Removing possibly old values



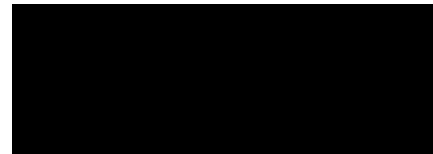
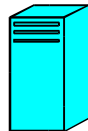
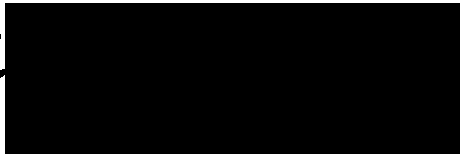
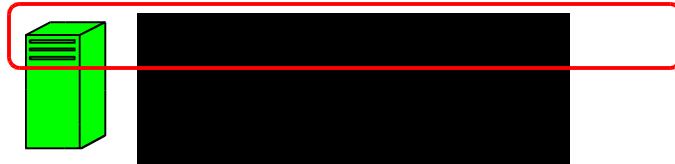
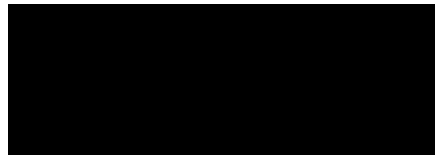
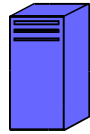
# Removing possibly malicious tuples



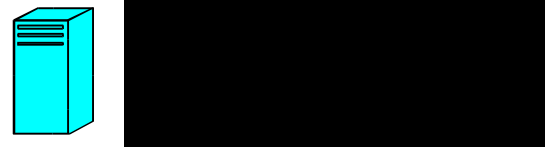
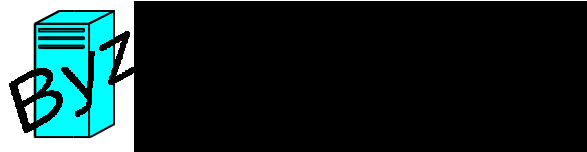
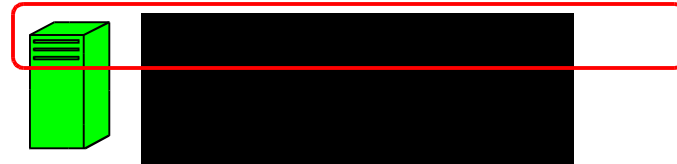
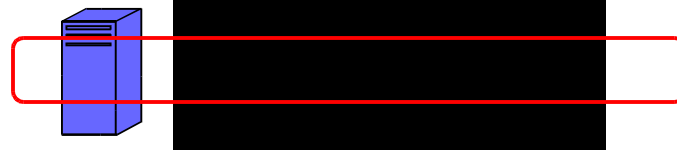
# Removing possibly malicious values



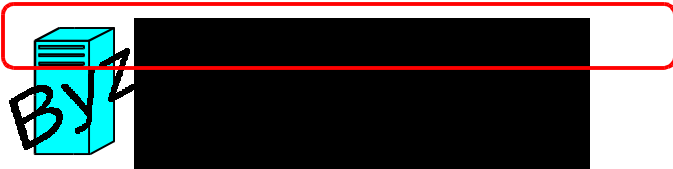
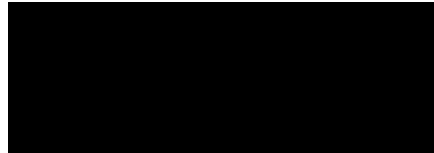
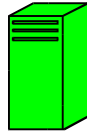
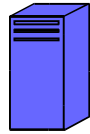
# Removing possibly malicious values



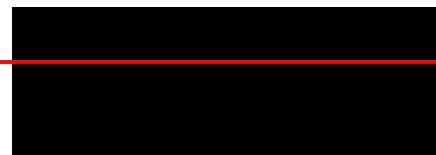
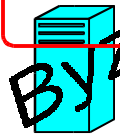
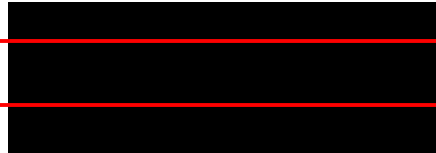
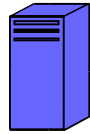
# Removing possibly malicious values



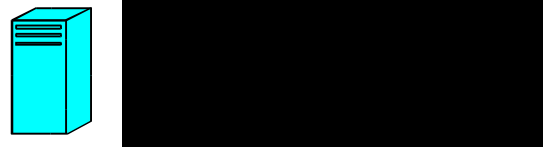
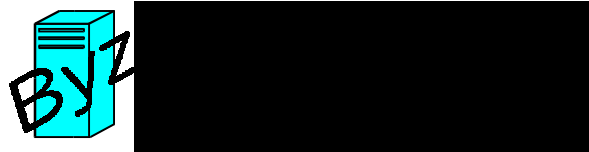
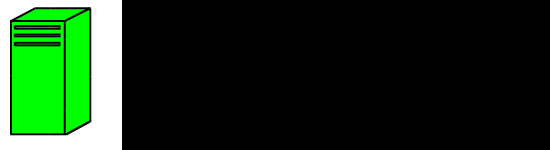
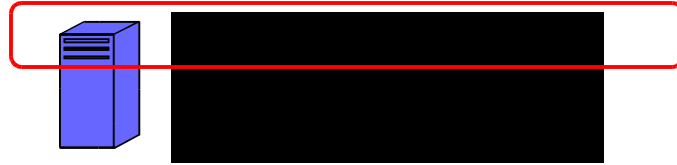
# Removing possibly malicious values



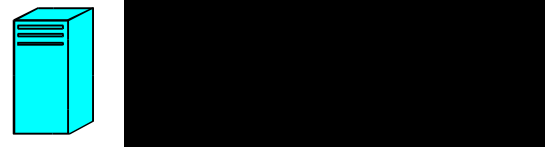
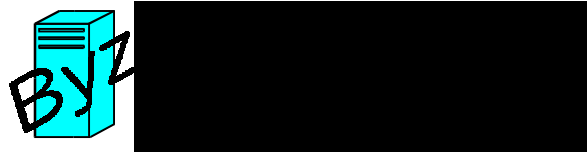
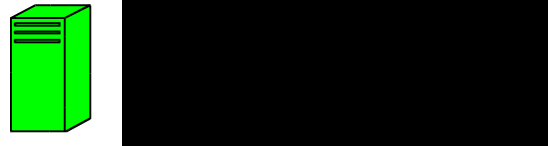
# Removing possibly malicious values



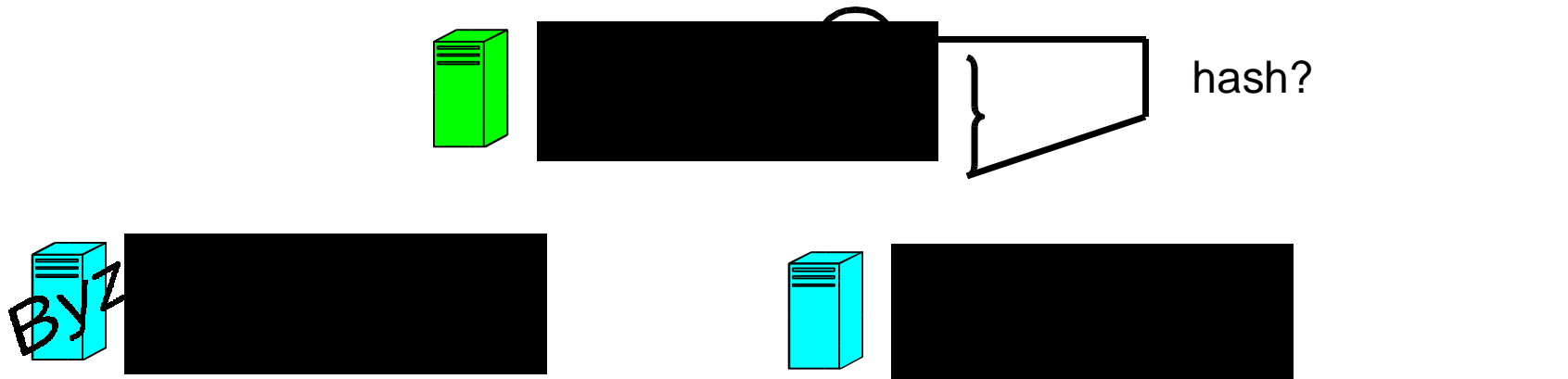
# Removing possibly malicious values



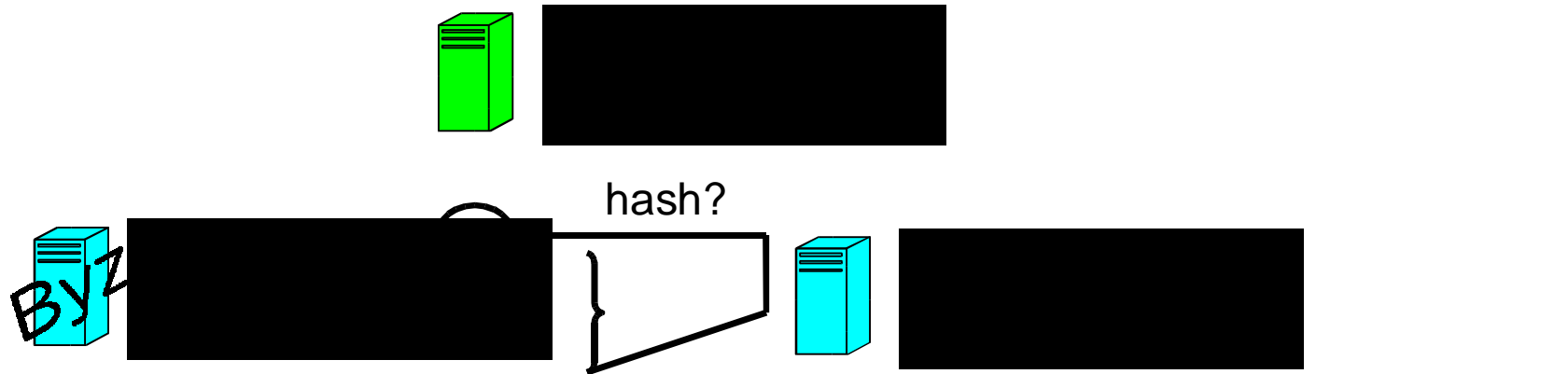
# Removing possibly malicious values



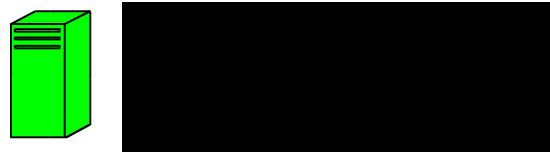
# Removing incorrect histories



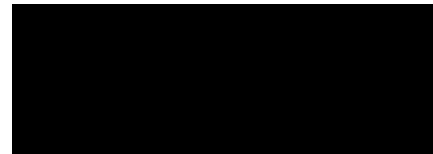
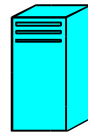
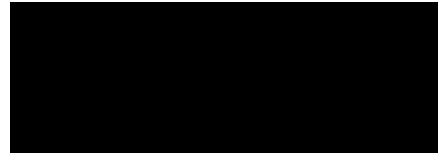
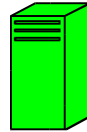
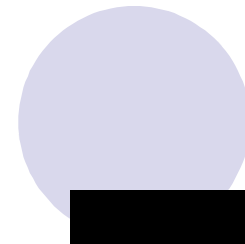
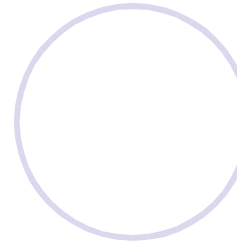
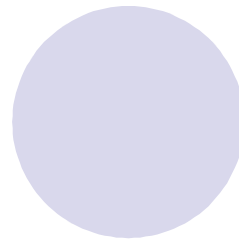
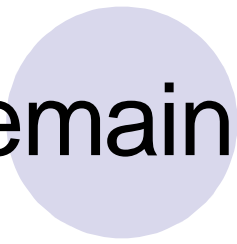
# Removing incorrect histories



# Removing incorrect histories

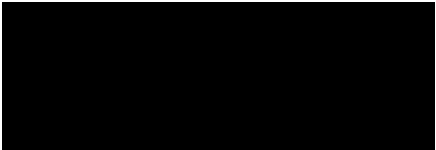
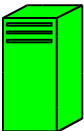
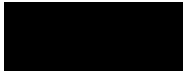
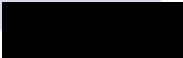
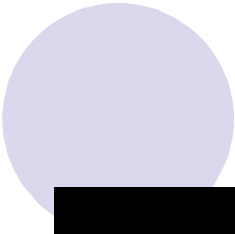
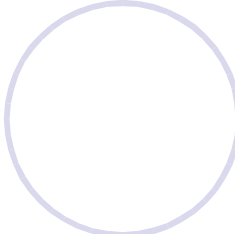
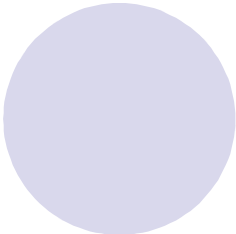
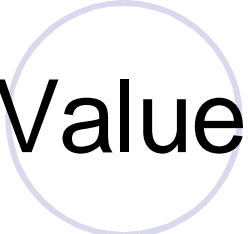
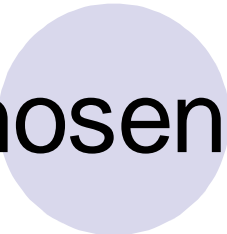


# Remaining Values



Chosen

Value



# In this paper

- K-quorums tolerating Byzantine servers
- Multi-writer K-quorums for  $w$  writers
  - Built over Single writer K-quorums
    - With staleness bound of  $\Delta$
  - Staleness bound of  $\Delta$
- Lower Bound
  - Staleness of at least  $\Delta$

# Multi-writer K-quorum

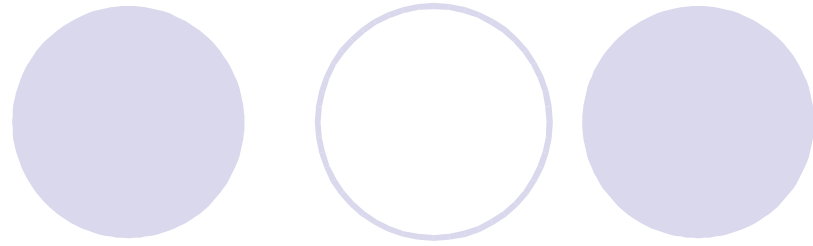
- Store **value-timestamp** pairs
  - one for each writer
- A writer only updates its value
- A reader reads all values
  - Will receive one of the latest **values**
  - Not possible to identify the latest one


# Write Protocol



- Read timestamps from each of the ■ writers
- Evaluate the approximate vector timestamp, by
  - incrementing the local timestamp for self
  - finding the maximum seen timestamp for each of the remaining writers
- Write the value and the generated timestamp to a *partial-write-quorum*

# Read protocol



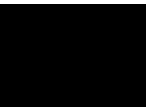
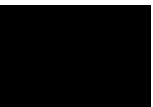
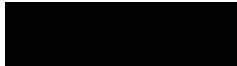
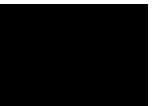
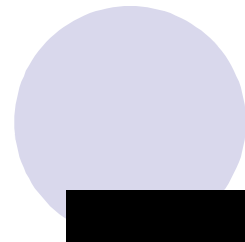
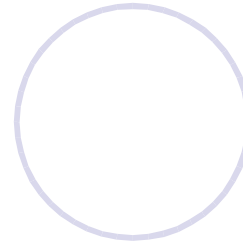
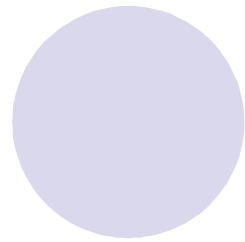
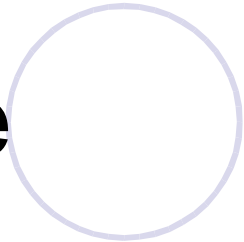
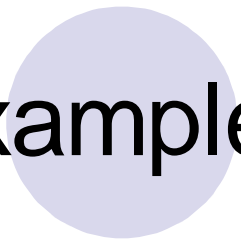
- Read value and timestamp from all the  single-writer registers.
- Reject older values.
- Return one of the remaining values.

# Example

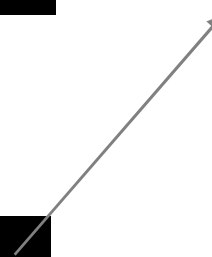
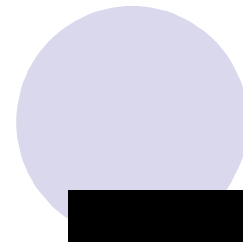
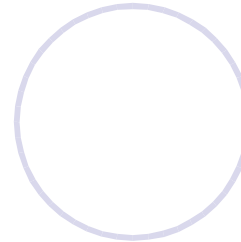
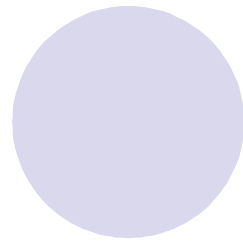
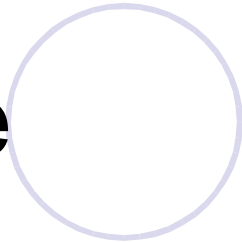
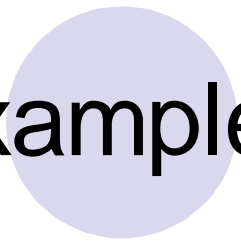
- Multi-Writer K-quorums

- 3 writers
- Staleness bound of 8
- Built over single writer K-quorums
  - Staleness bound of 2

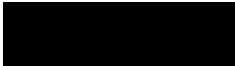
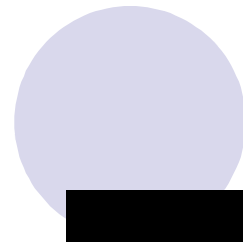
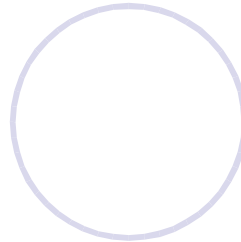
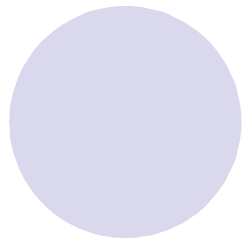
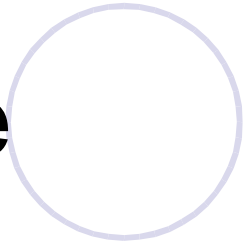
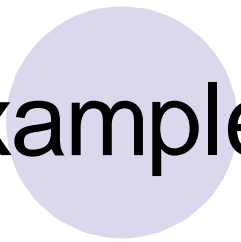
Example



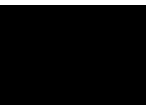
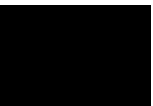
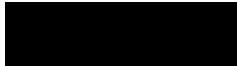
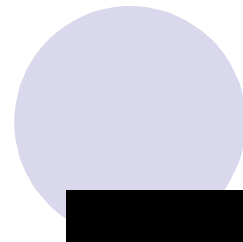
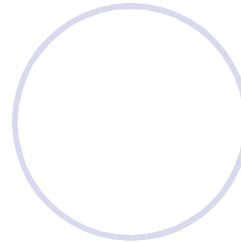
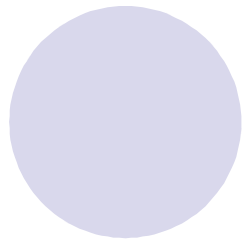
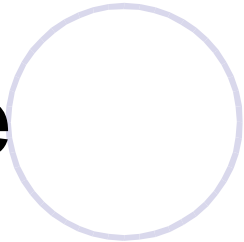
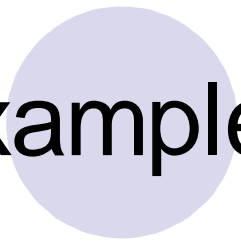
Example



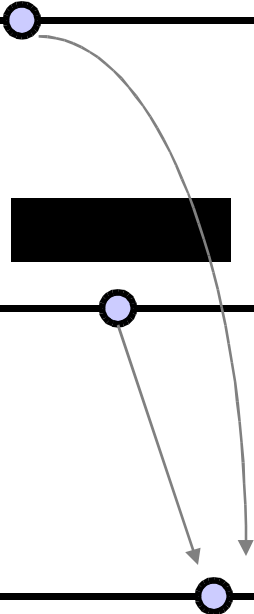
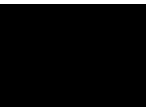
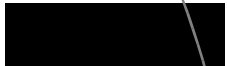
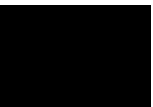
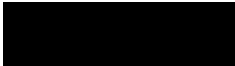
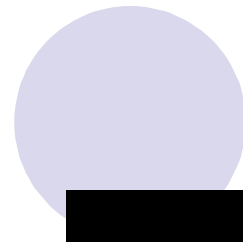
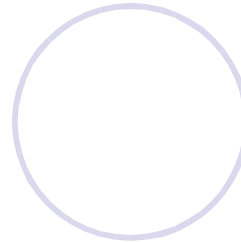
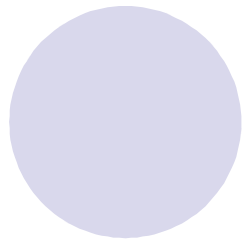
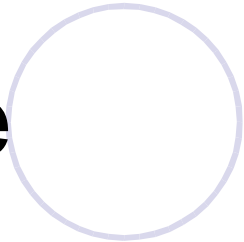
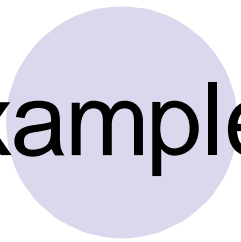
Example



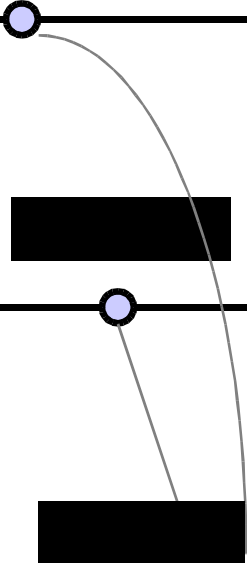
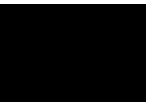
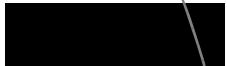
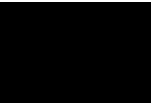
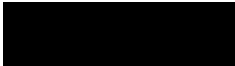
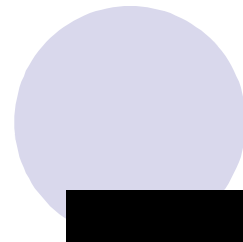
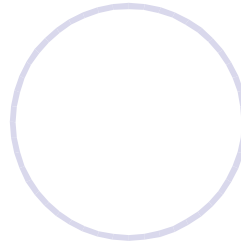
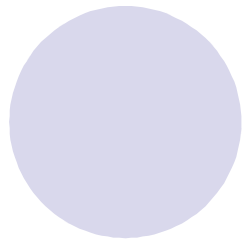
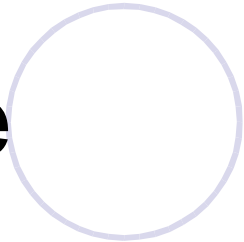
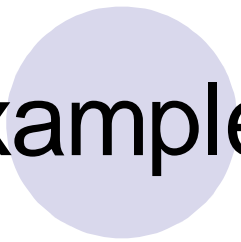
Example



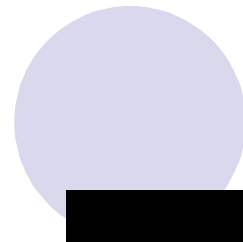
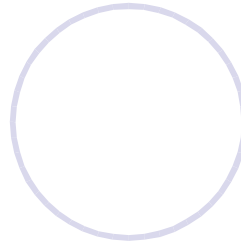
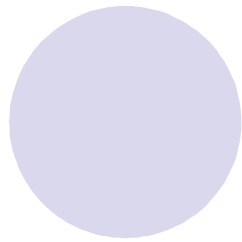
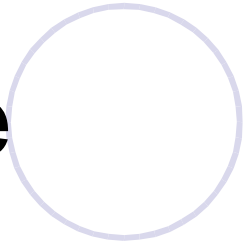
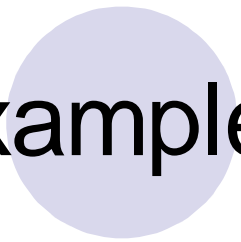
Example



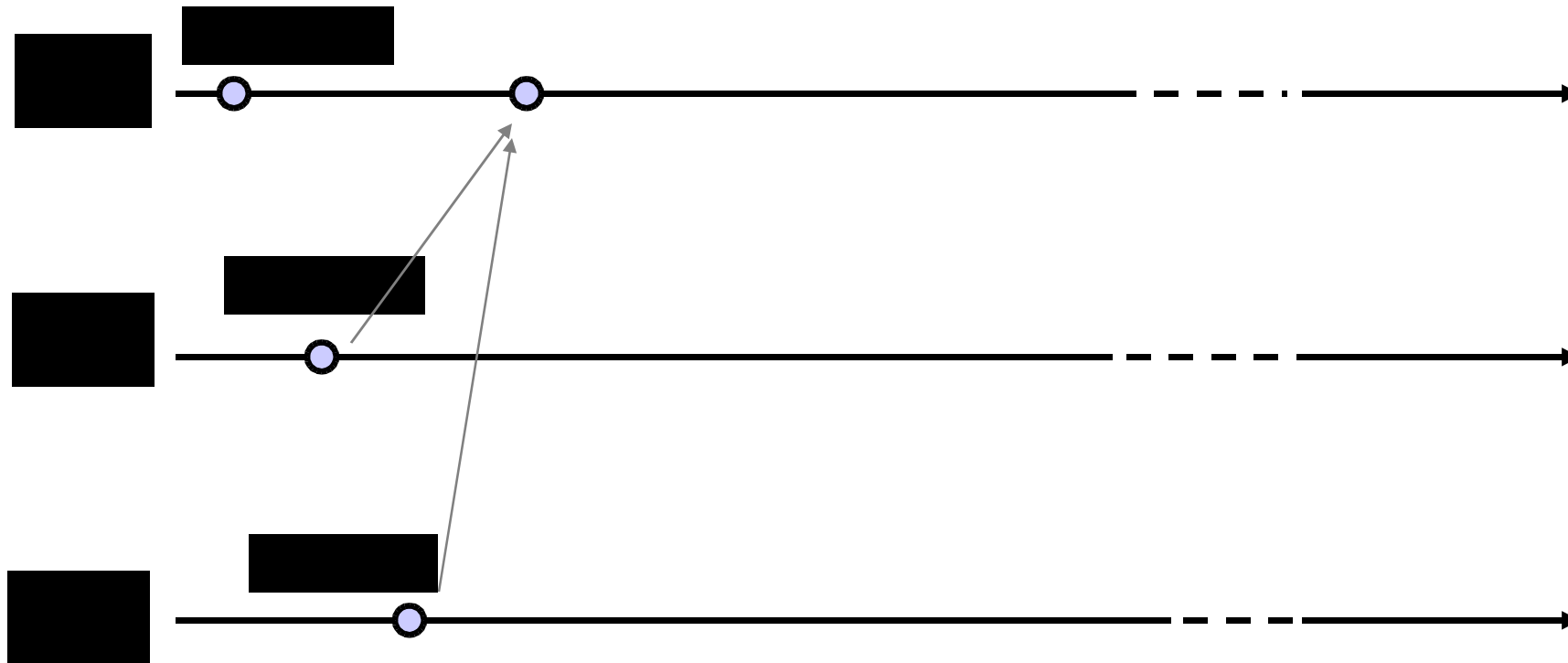
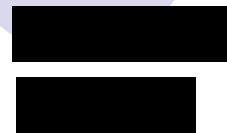
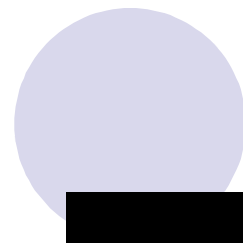
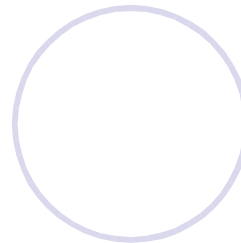
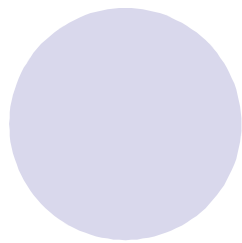
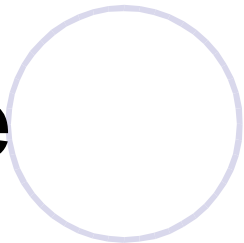
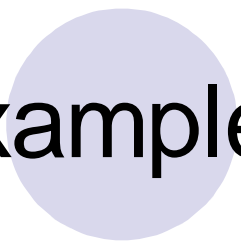
Example



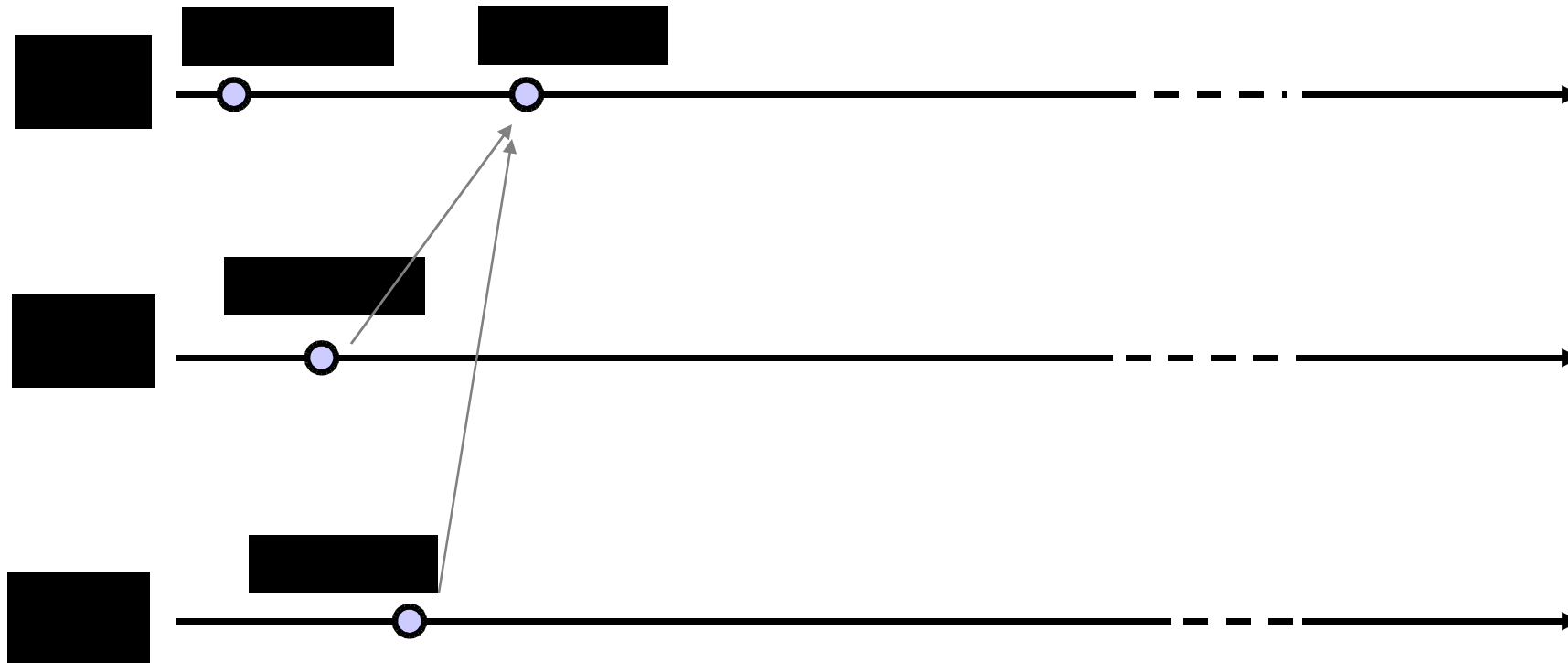
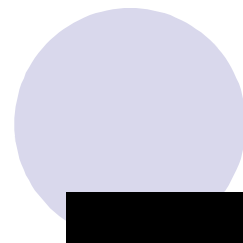
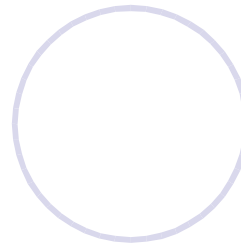
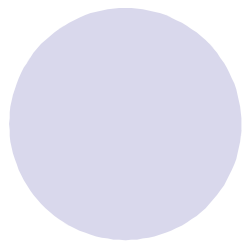
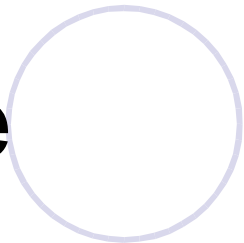
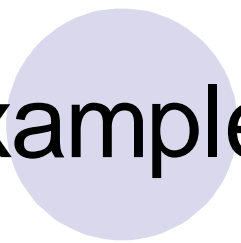
Example



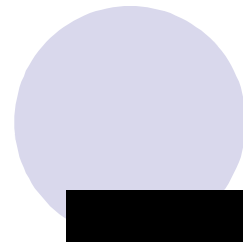
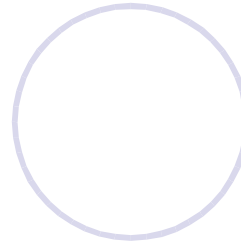
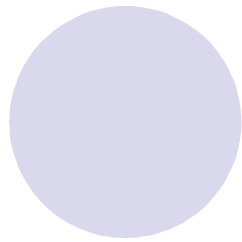
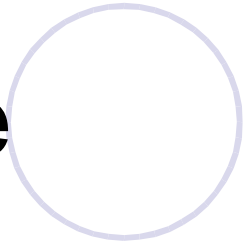
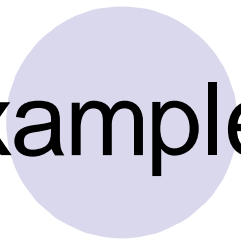
Example



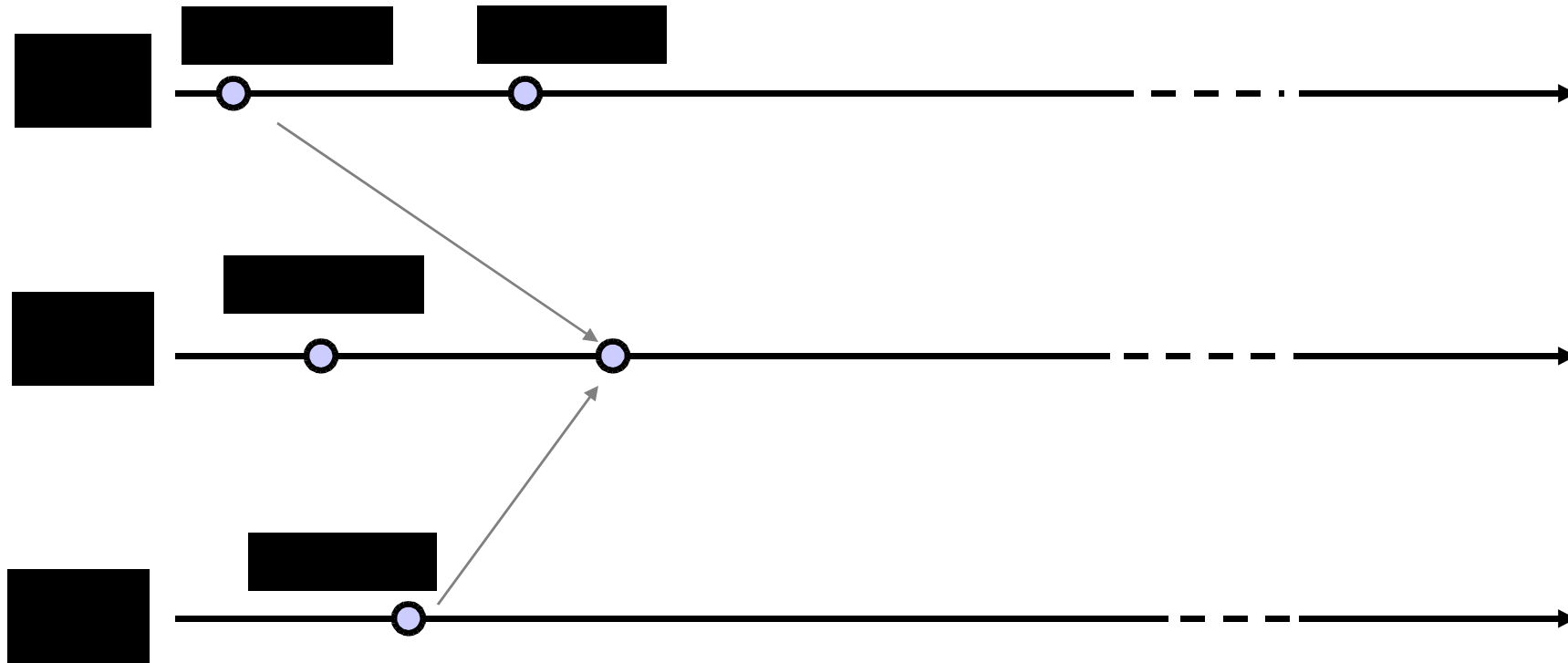
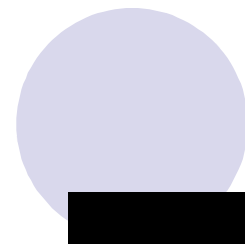
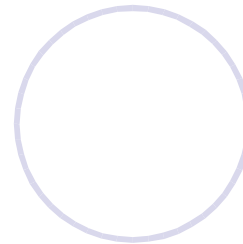
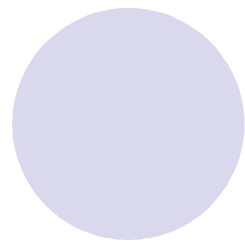
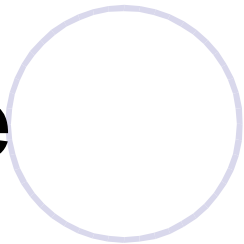
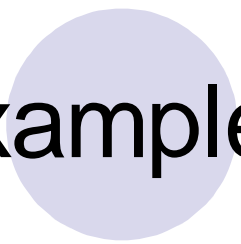
Example



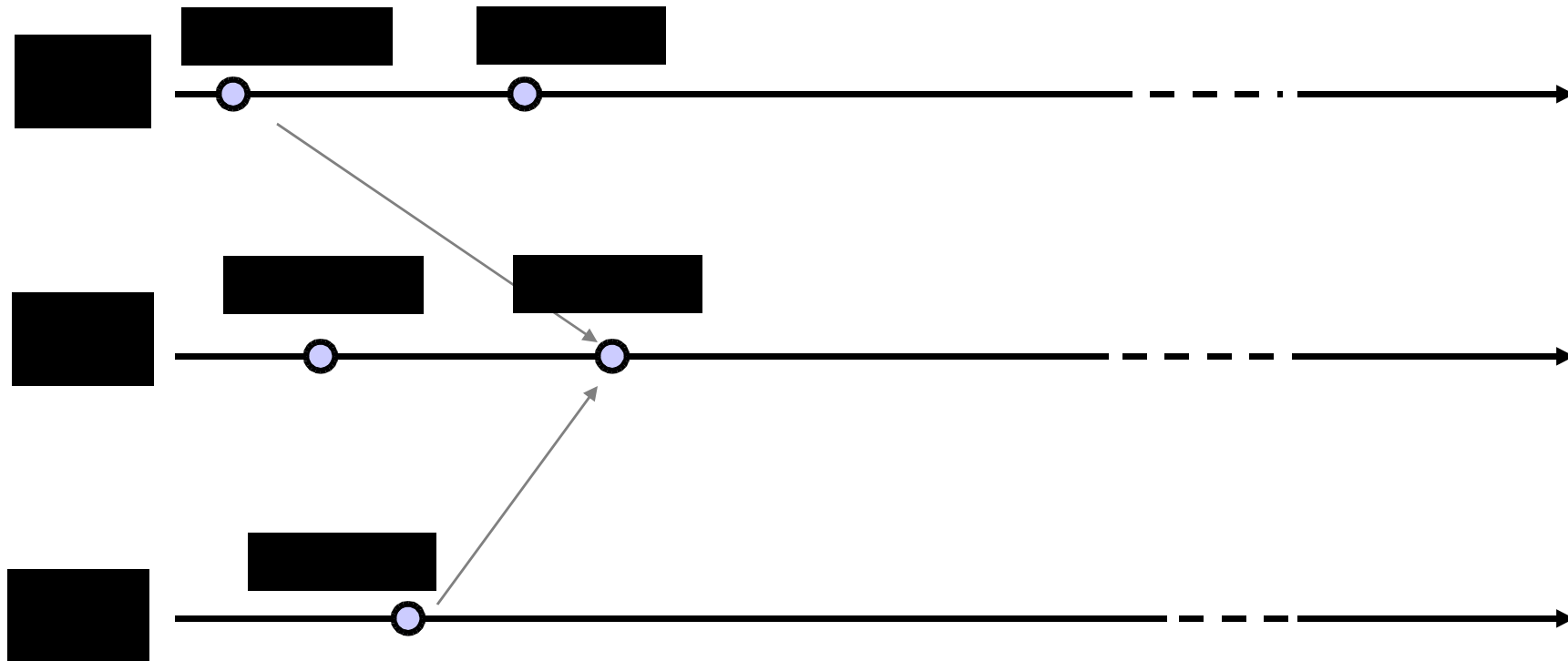
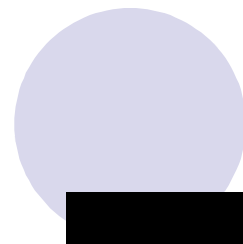
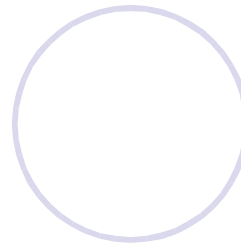
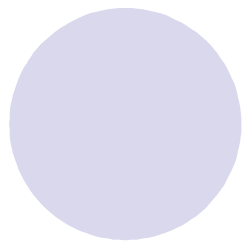
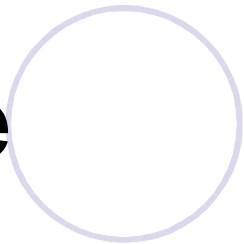
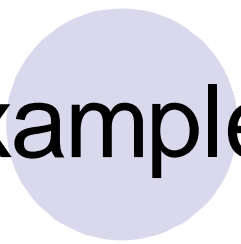
Example



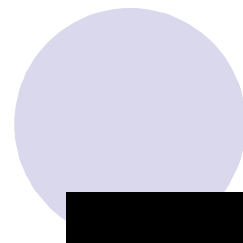
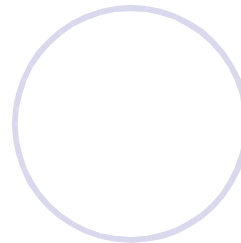
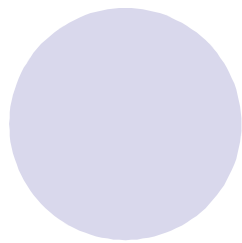
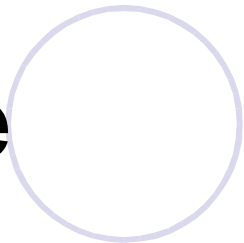
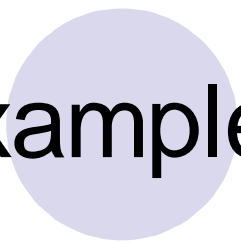
Example



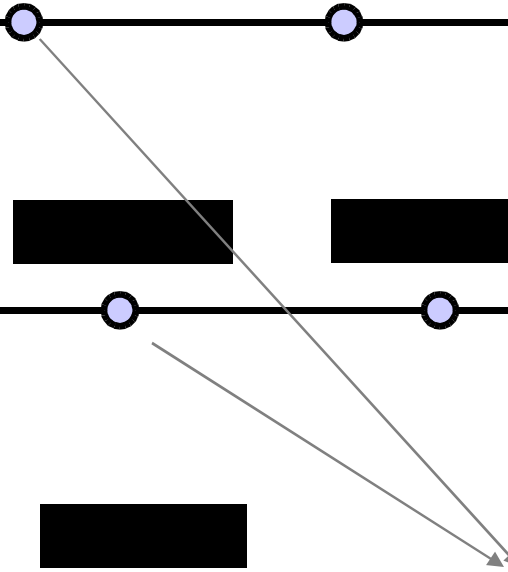
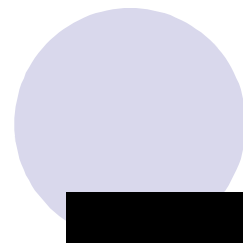
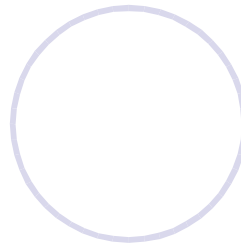
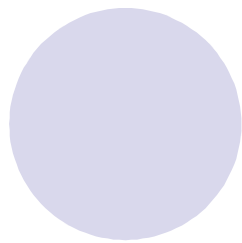
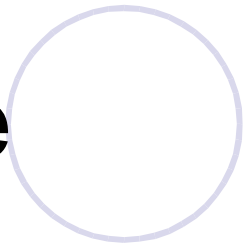
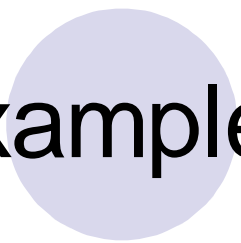
Example



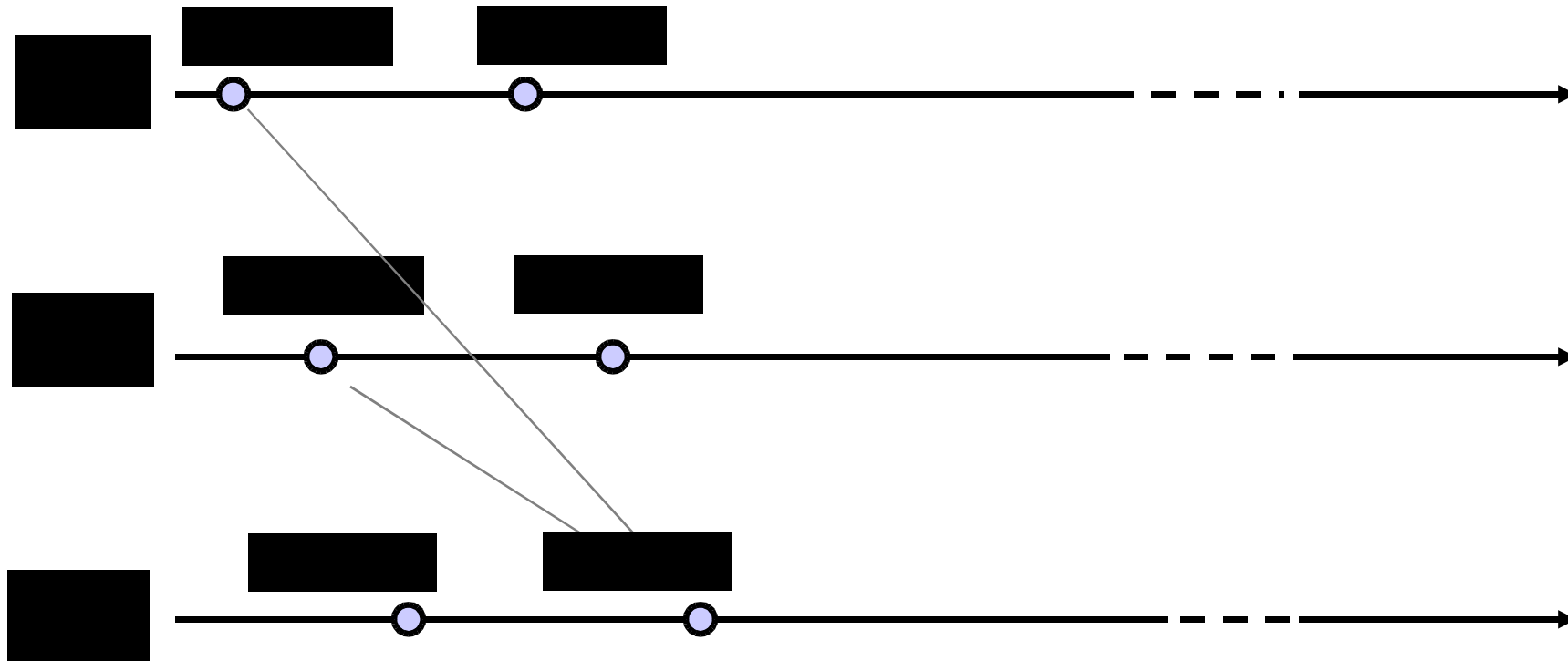
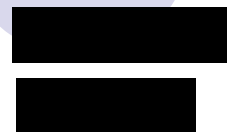
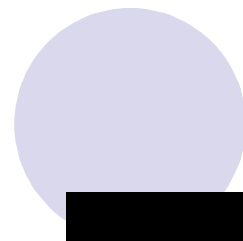
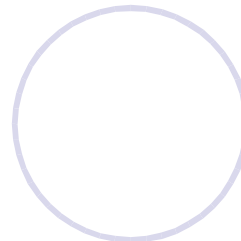
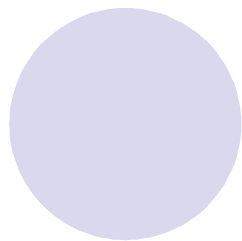
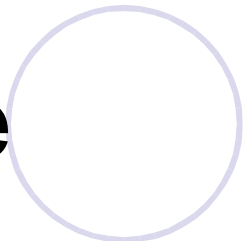
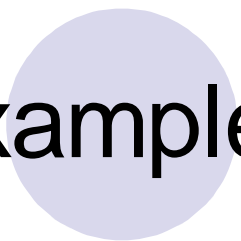
Example



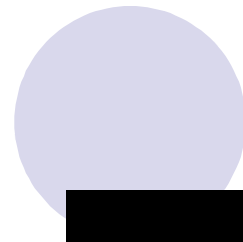
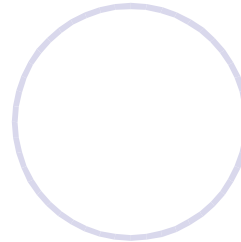
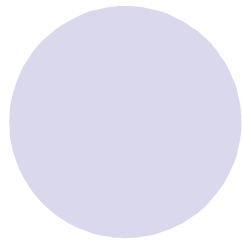
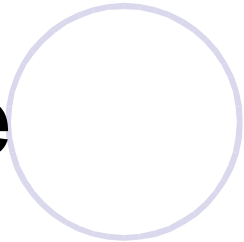
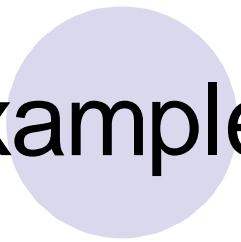
Example



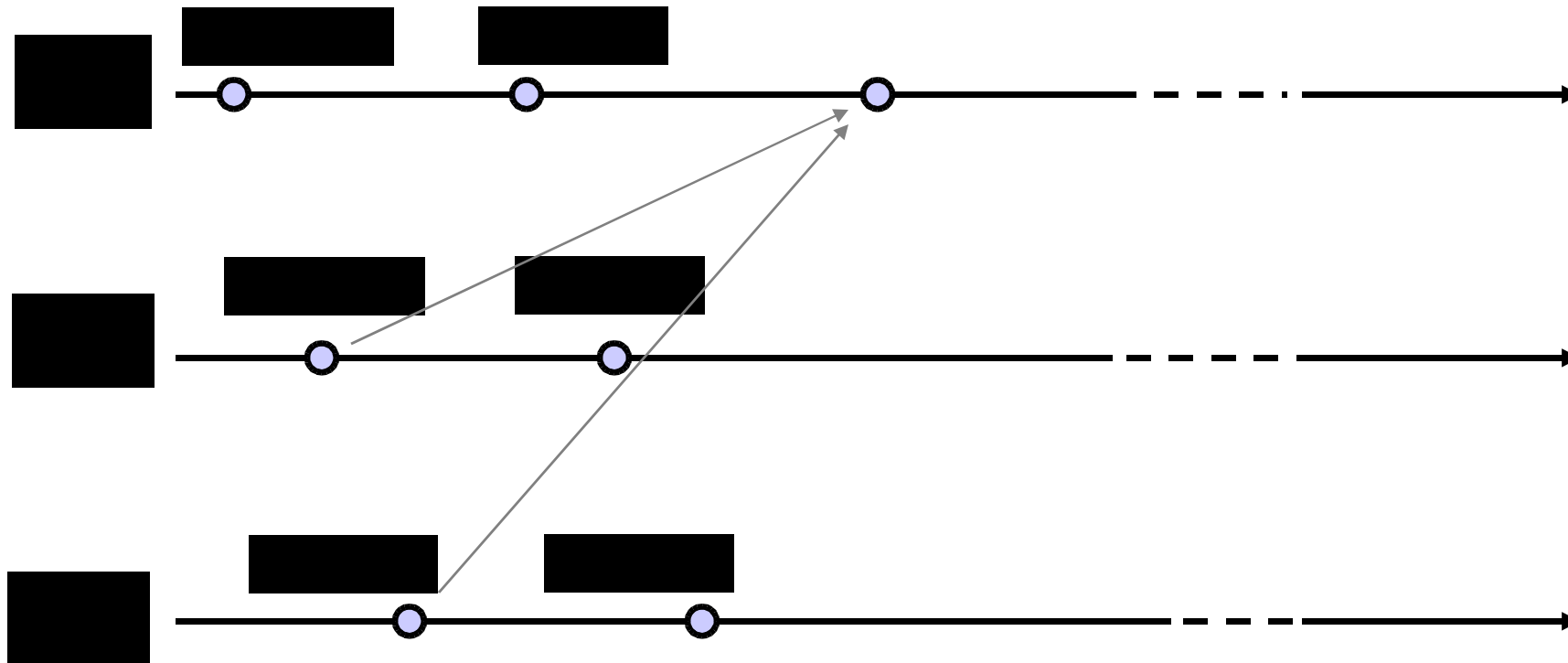
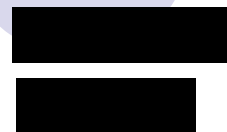
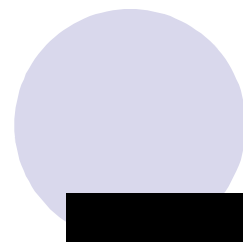
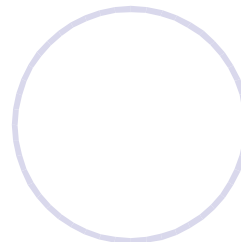
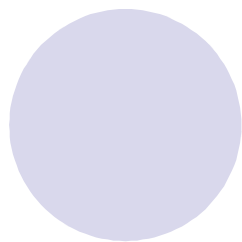
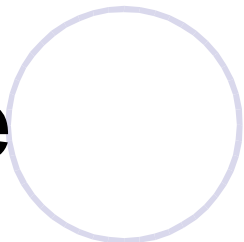
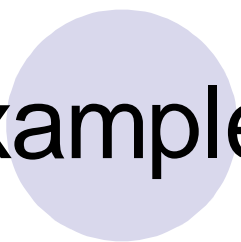
Example



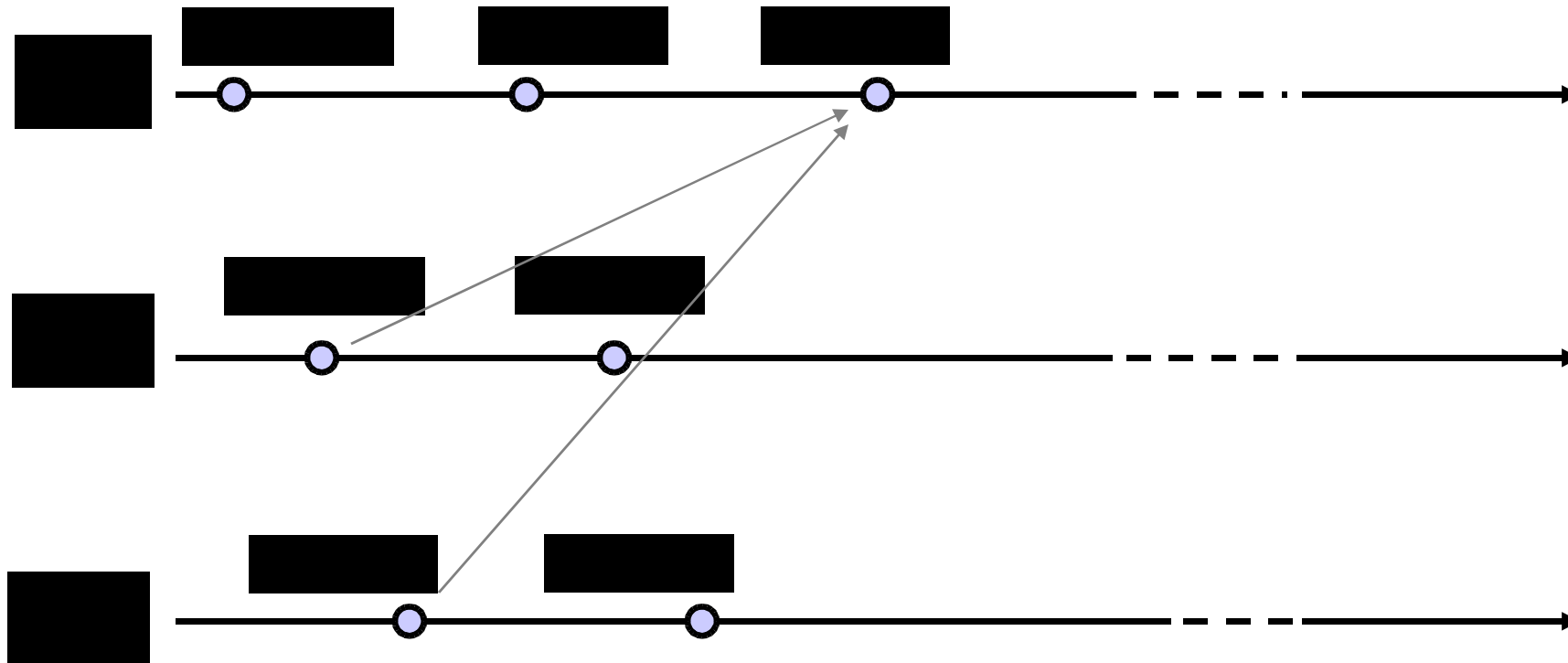
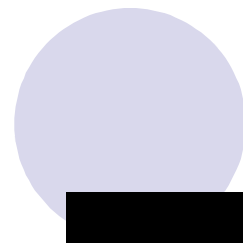
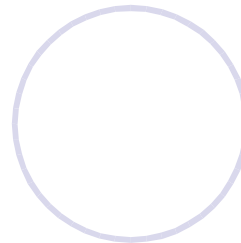
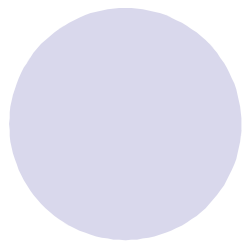
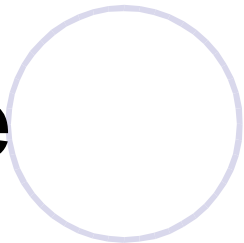
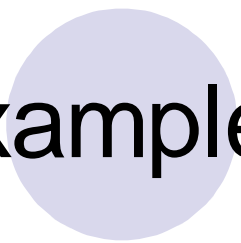
Example



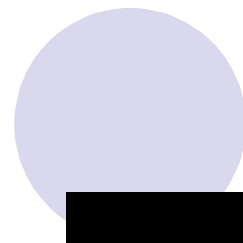
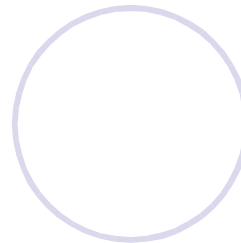
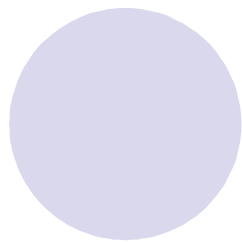
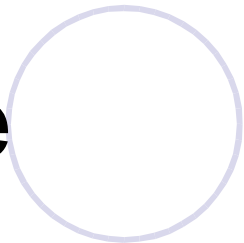
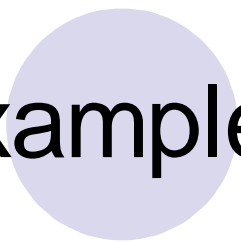
Example



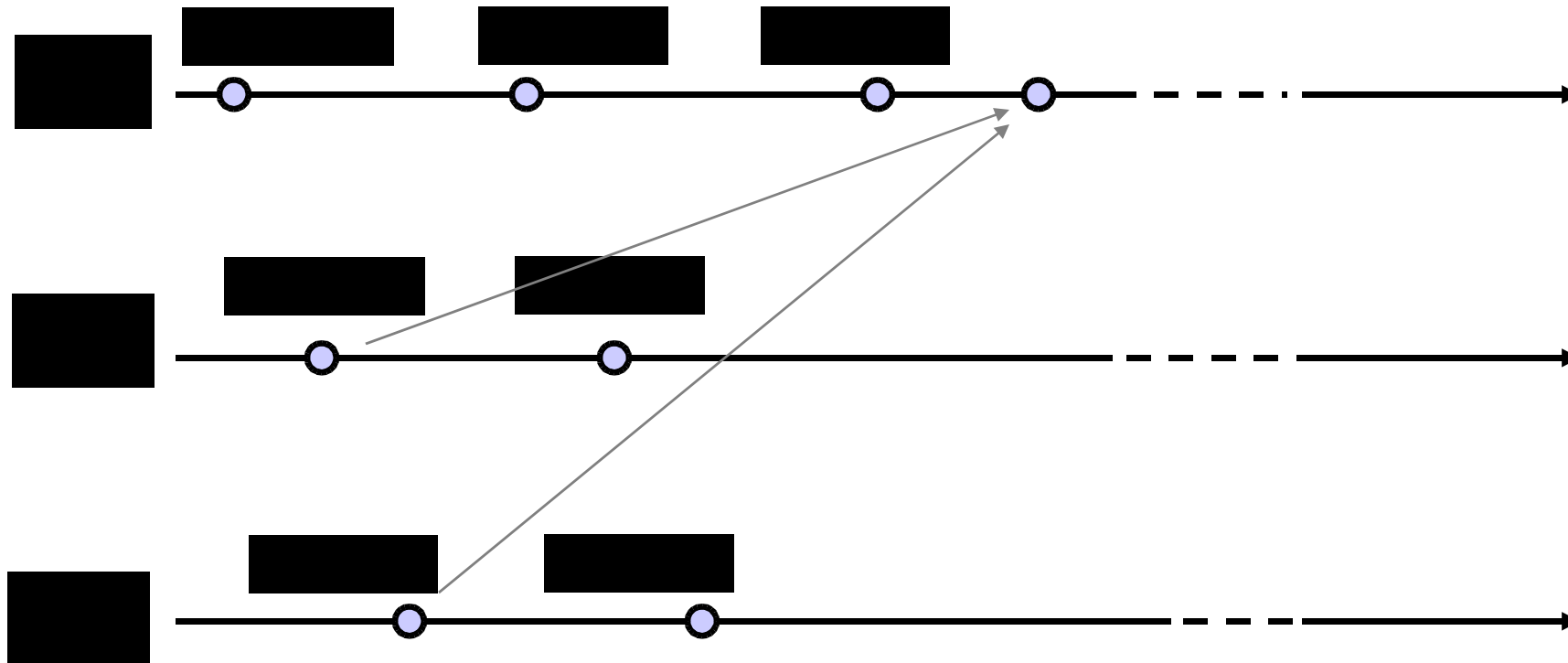
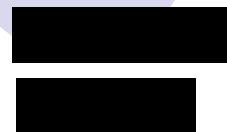
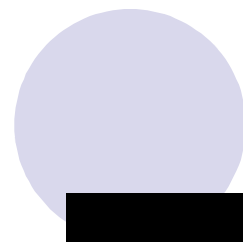
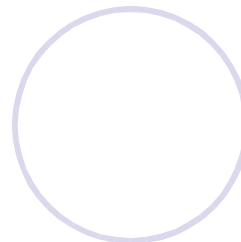
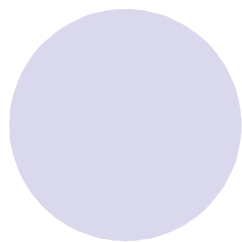
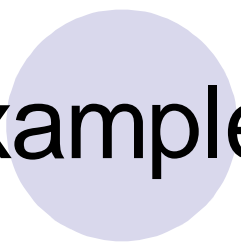
Example



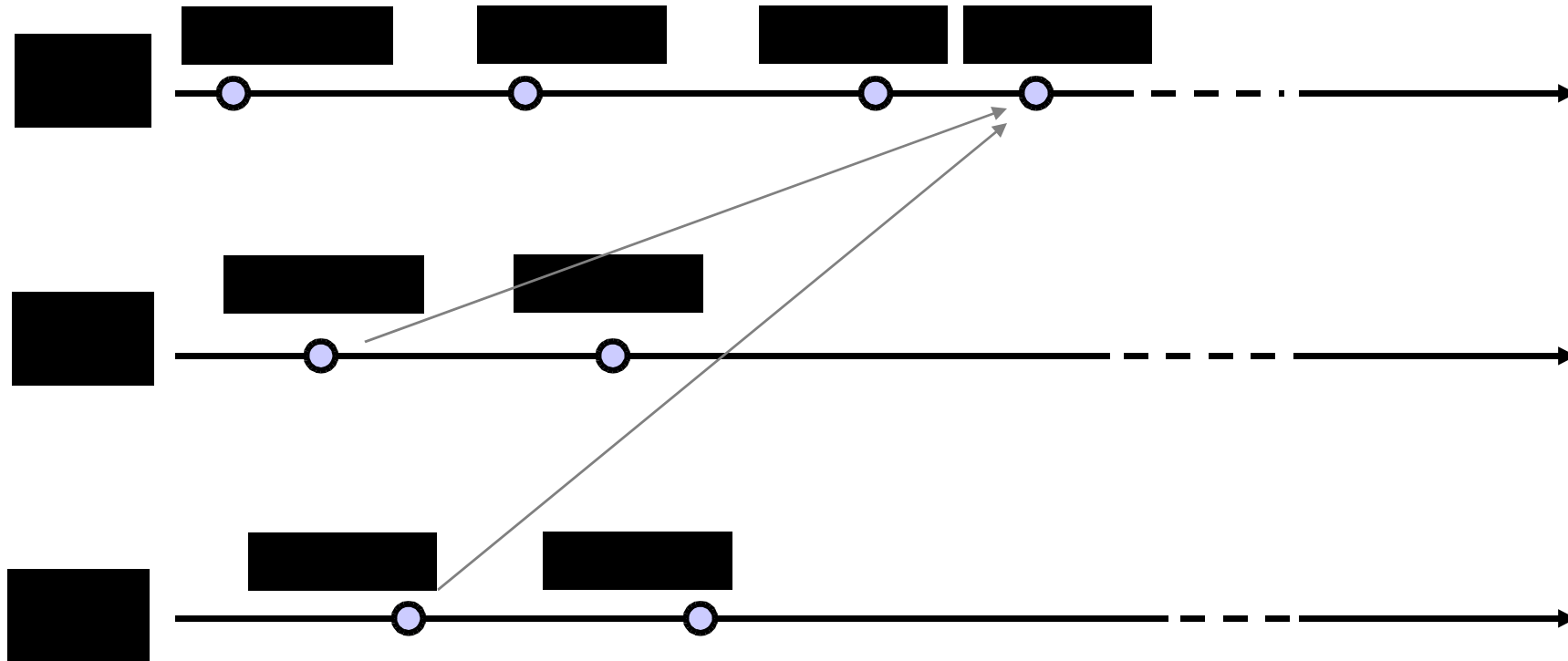
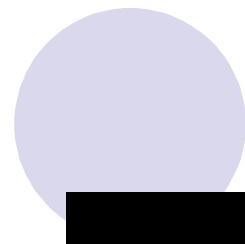
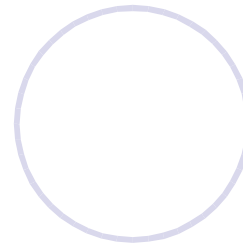
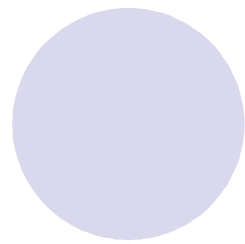
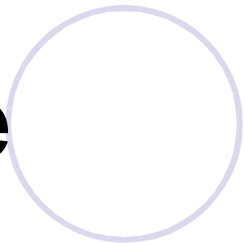
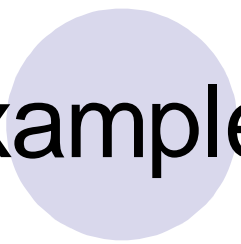
Example



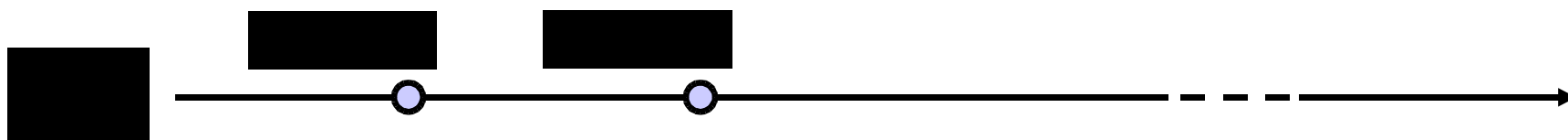
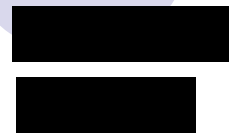
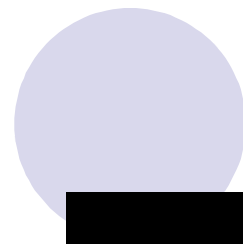
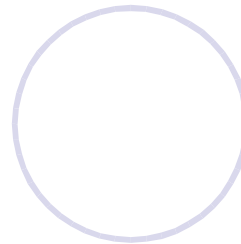
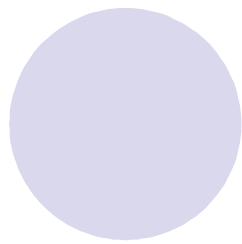
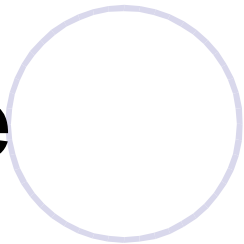
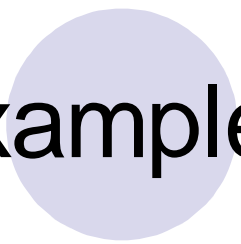
Example



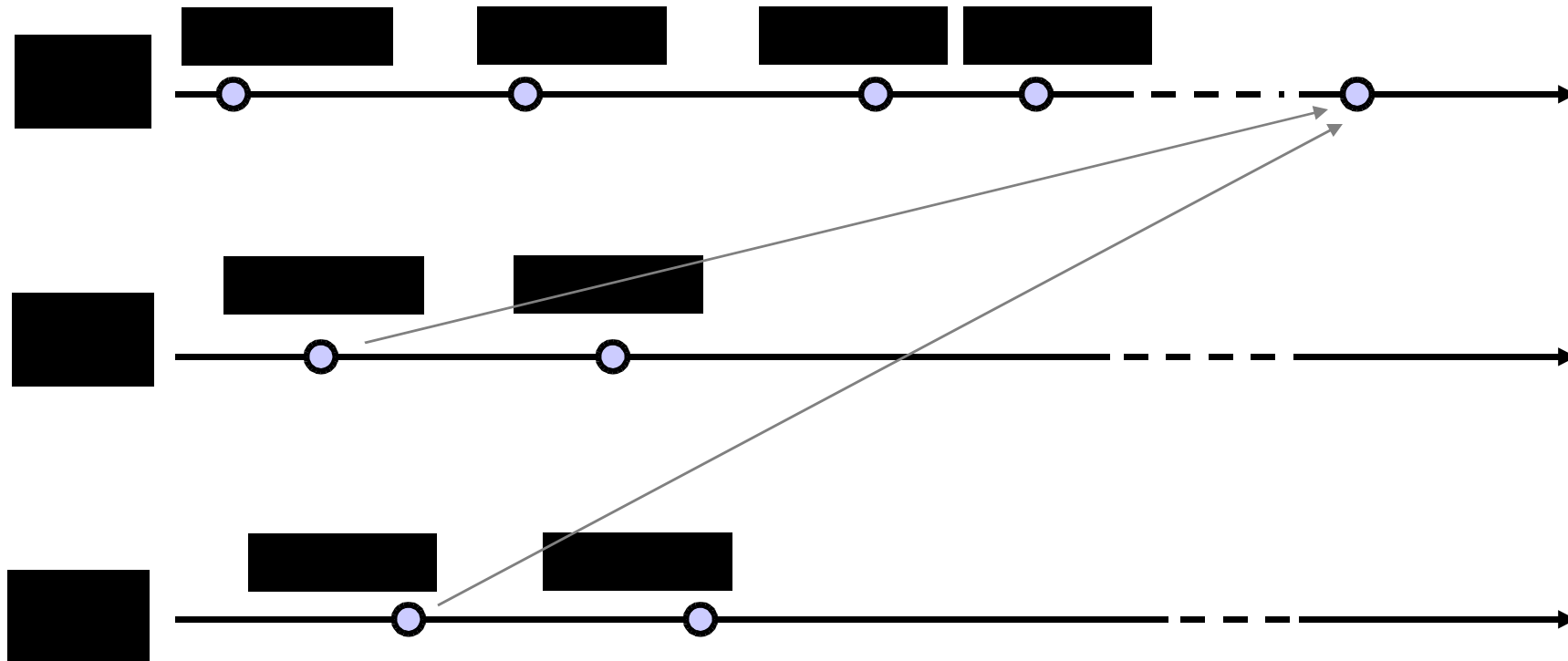
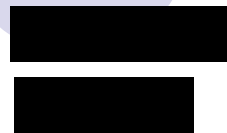
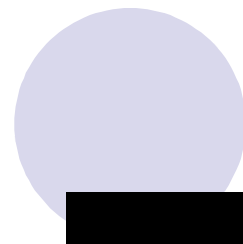
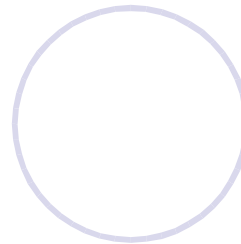
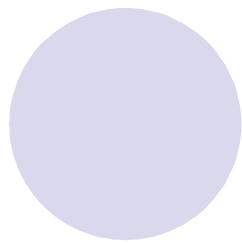
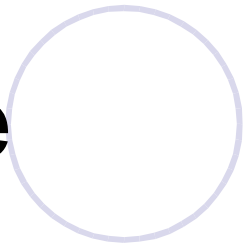
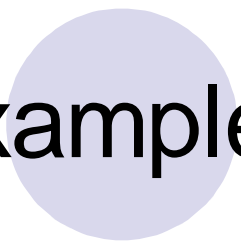
Example



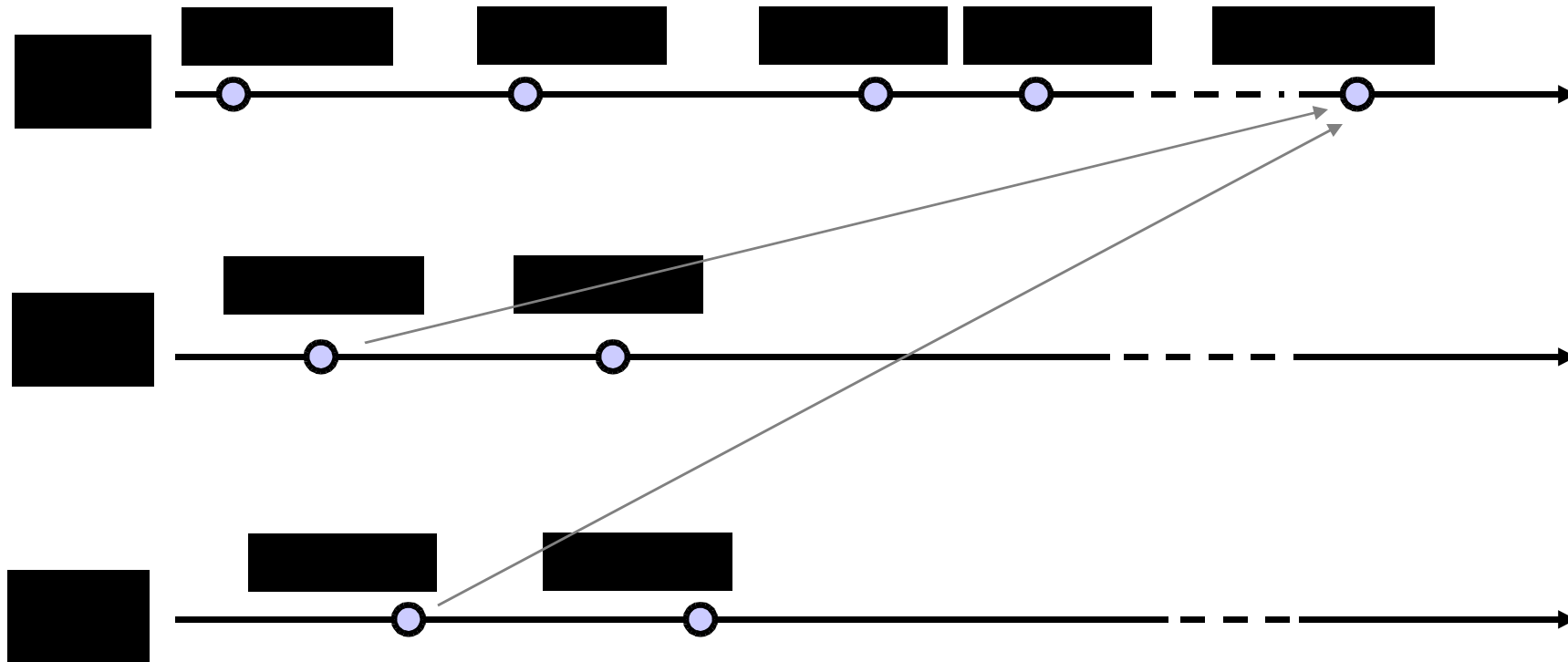
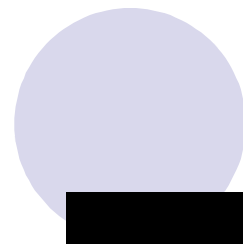
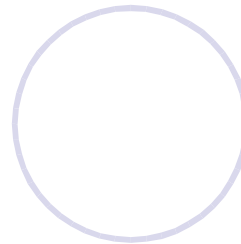
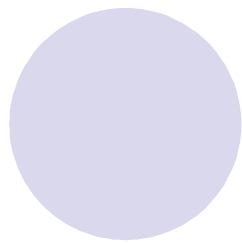
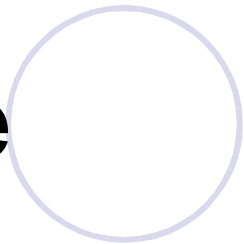
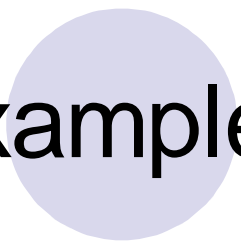
Example



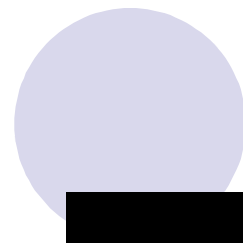
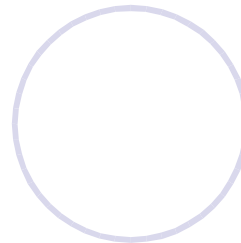
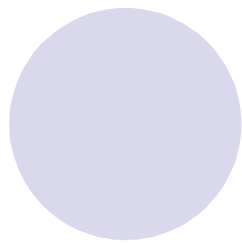
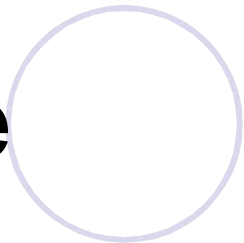
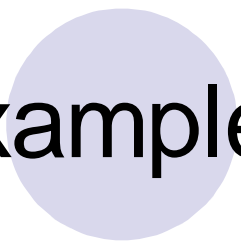
Example



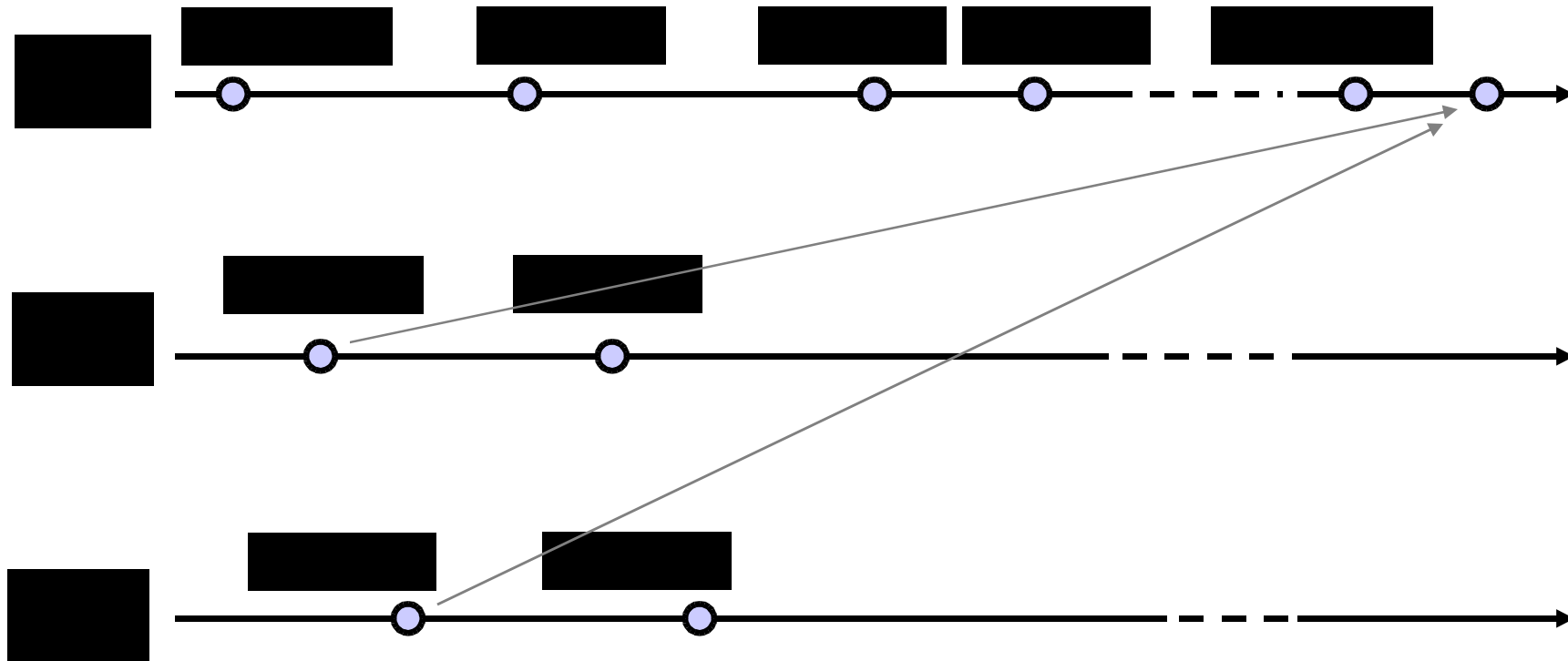
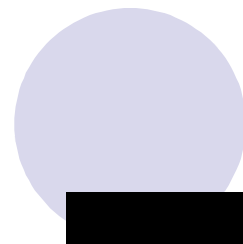
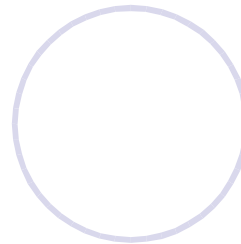
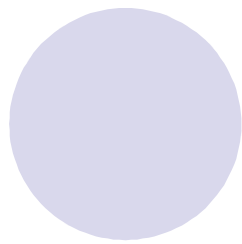
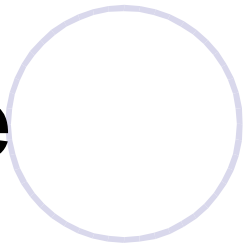
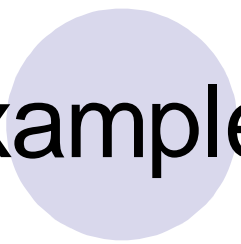
Example



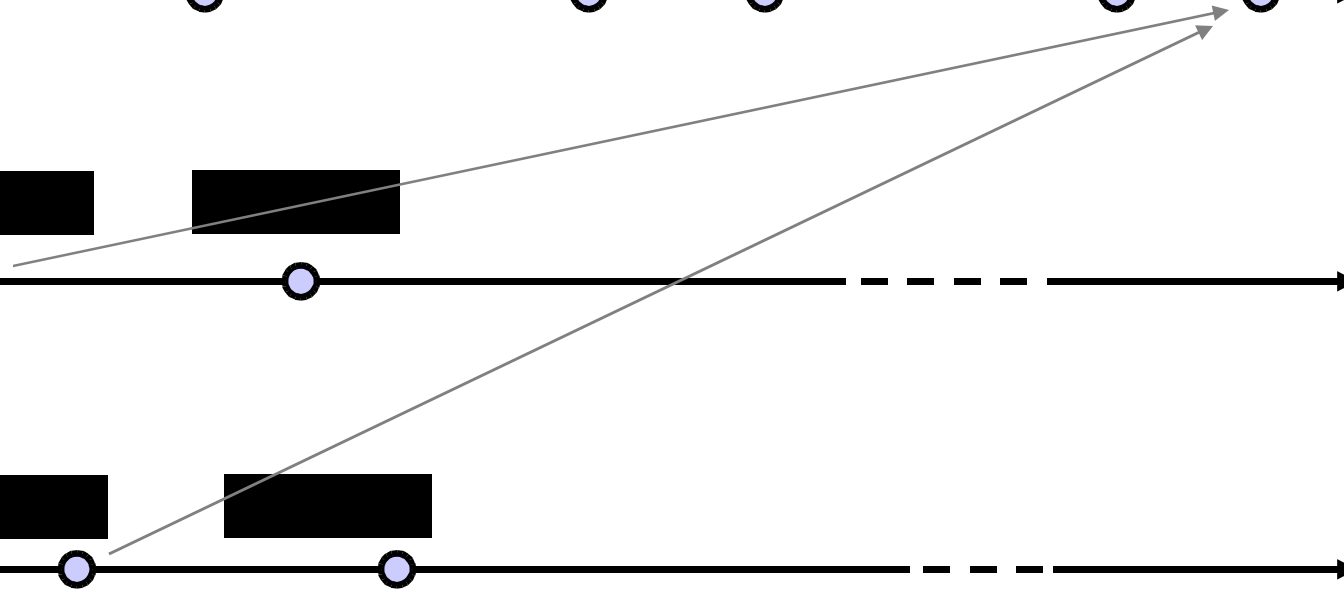
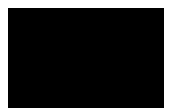
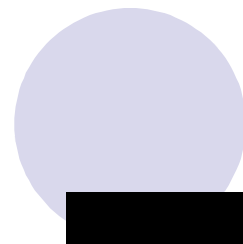
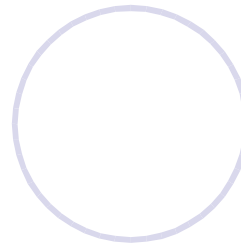
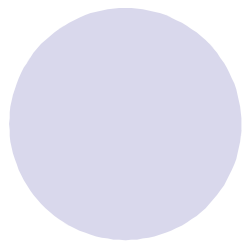
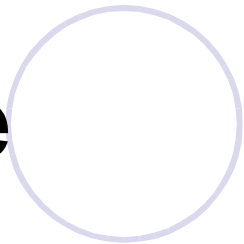
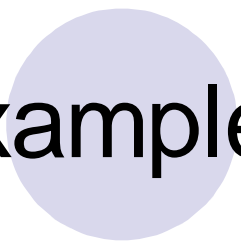
Example



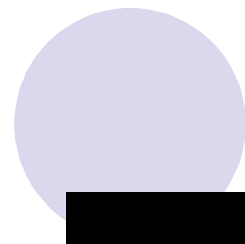
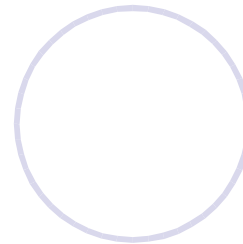
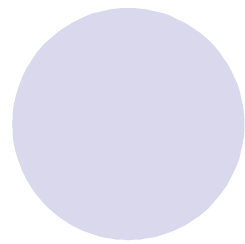
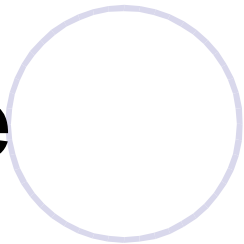
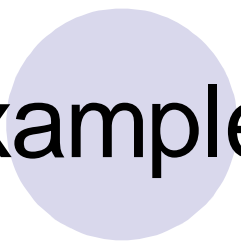
Example



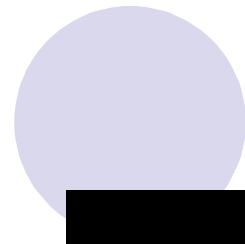
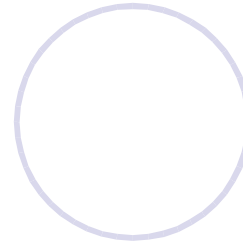
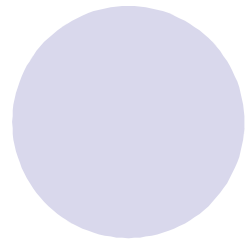
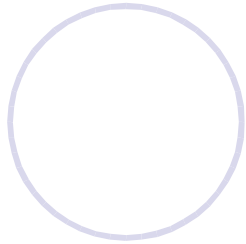
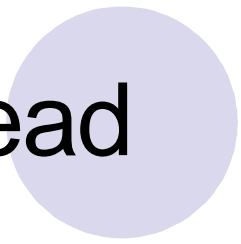
Example



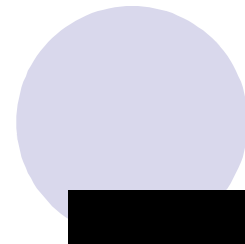
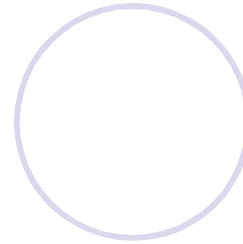
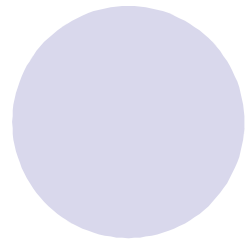
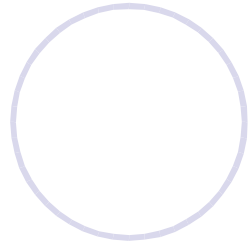
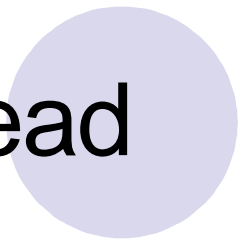
Example



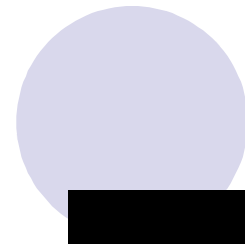
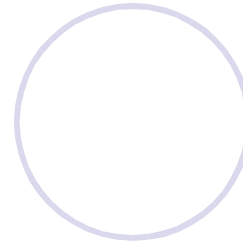
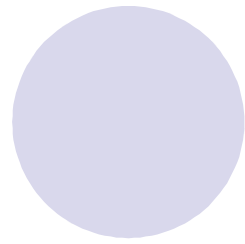
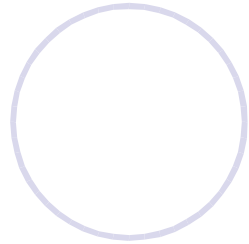
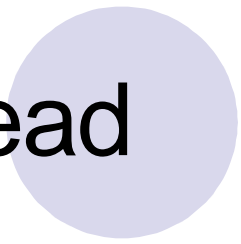
Read



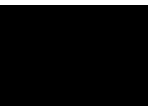
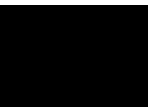
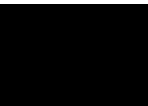
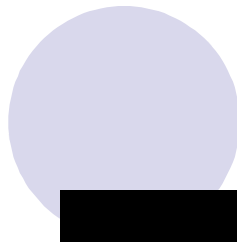
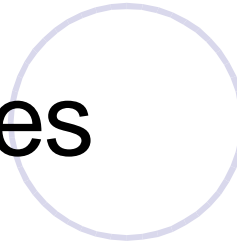
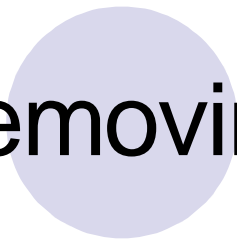
Read



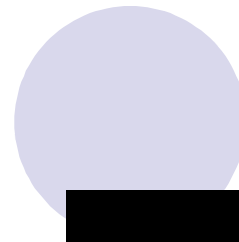
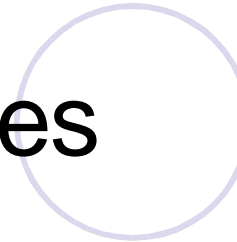
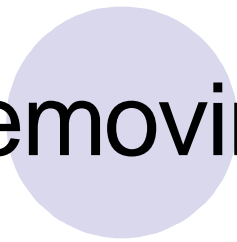
Read



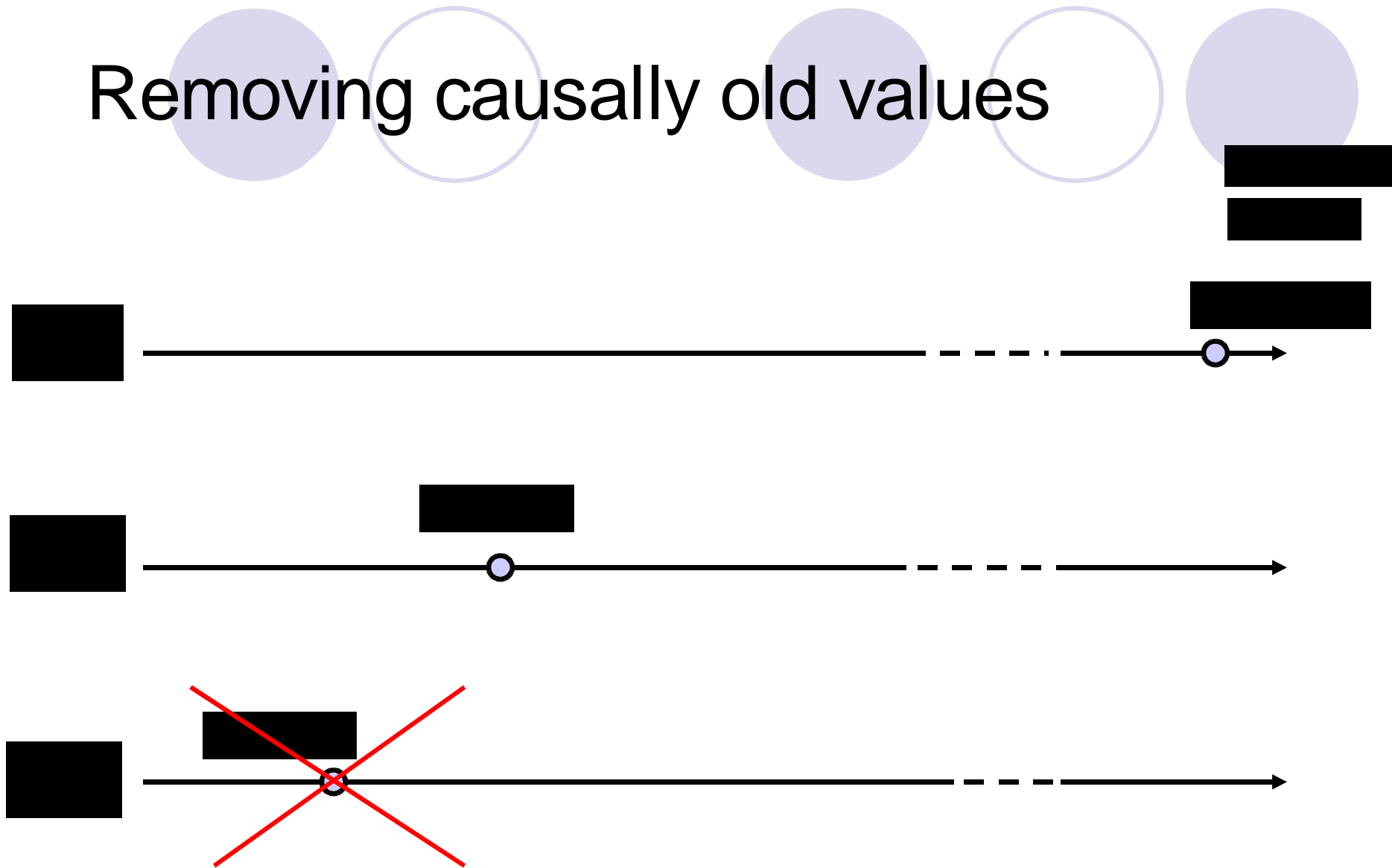
# Removing causally old values



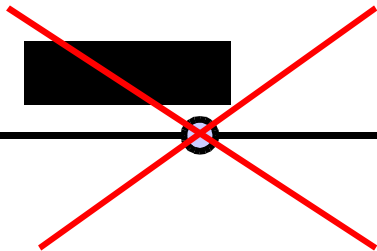
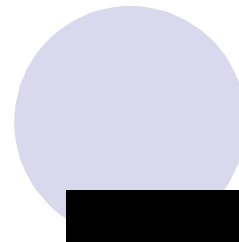
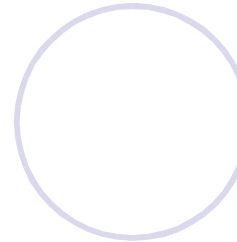
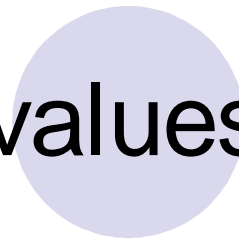
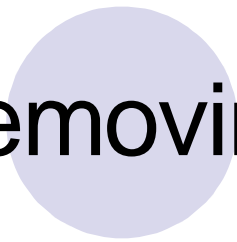
# Removing causally old values



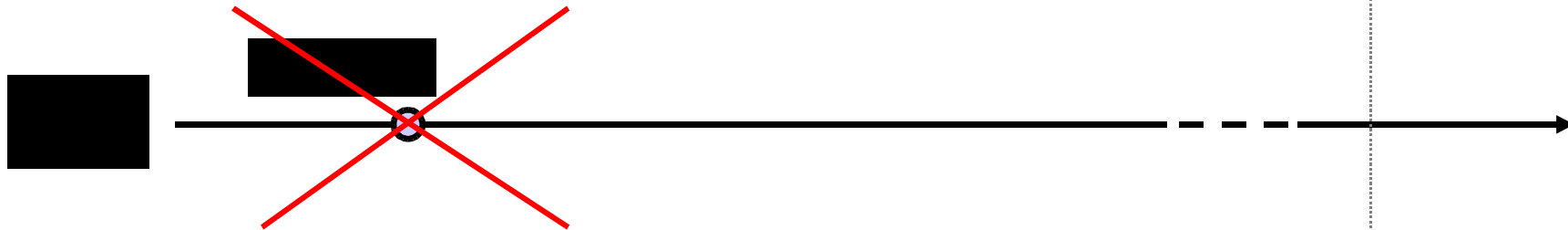
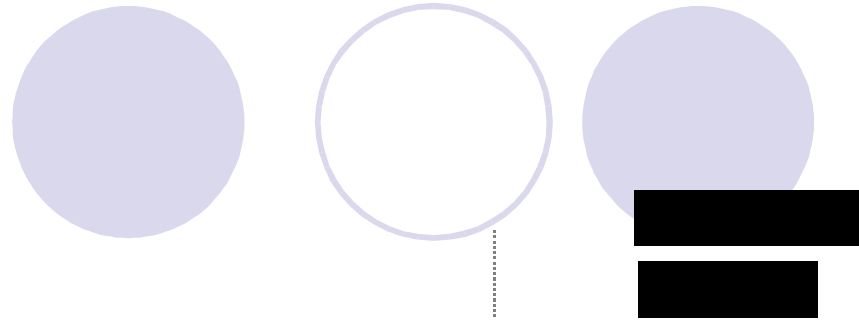
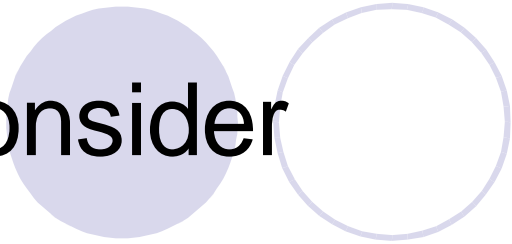
# Removing causally old values



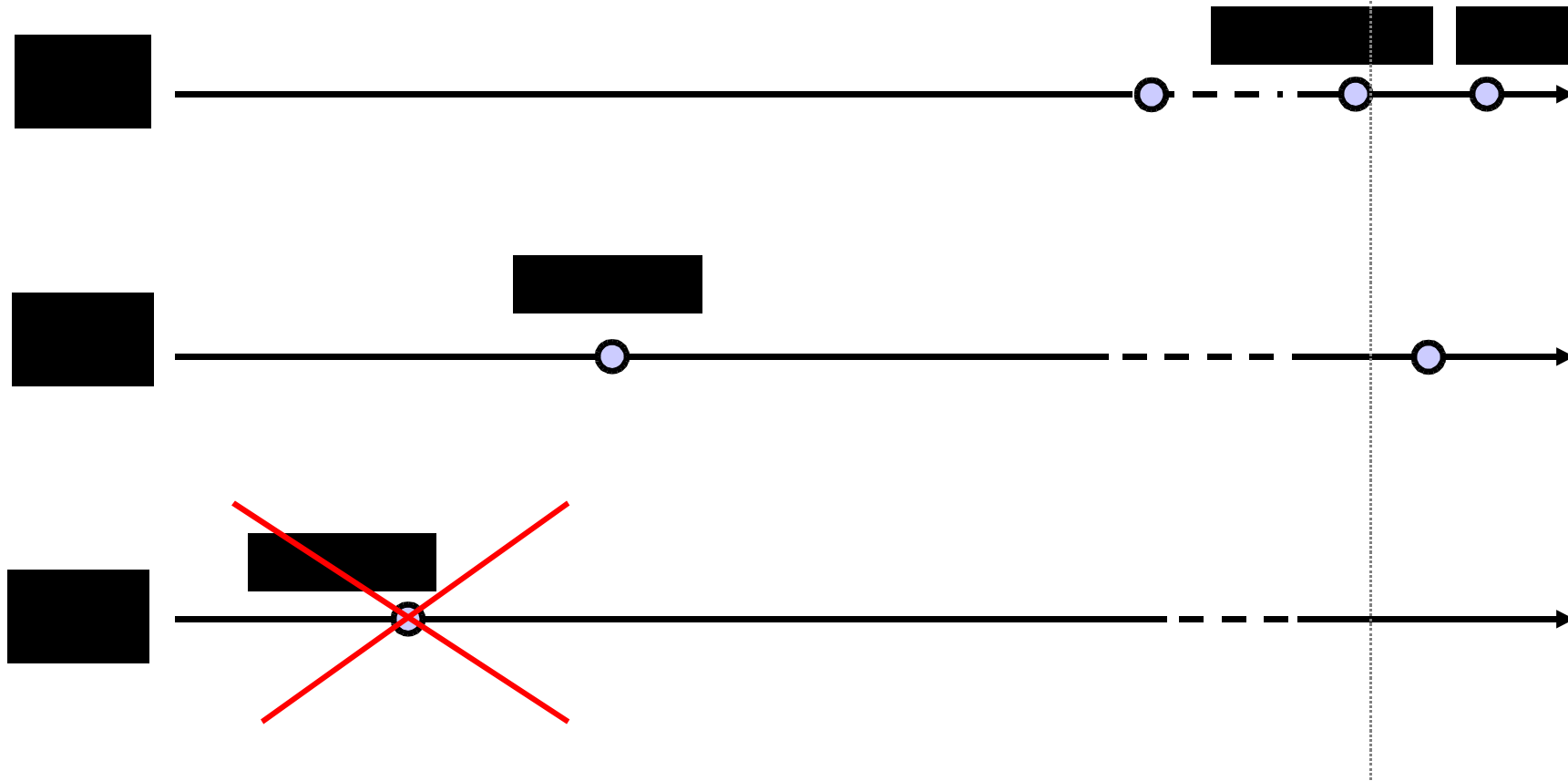
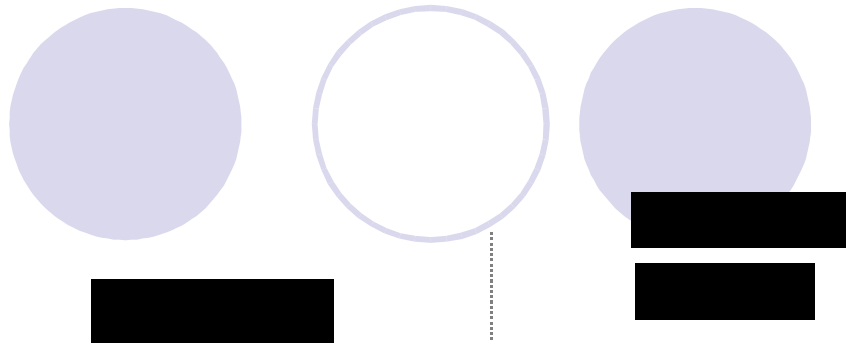
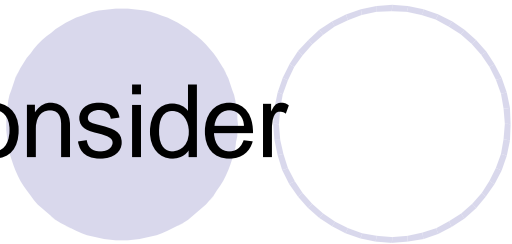
# Removing other old values



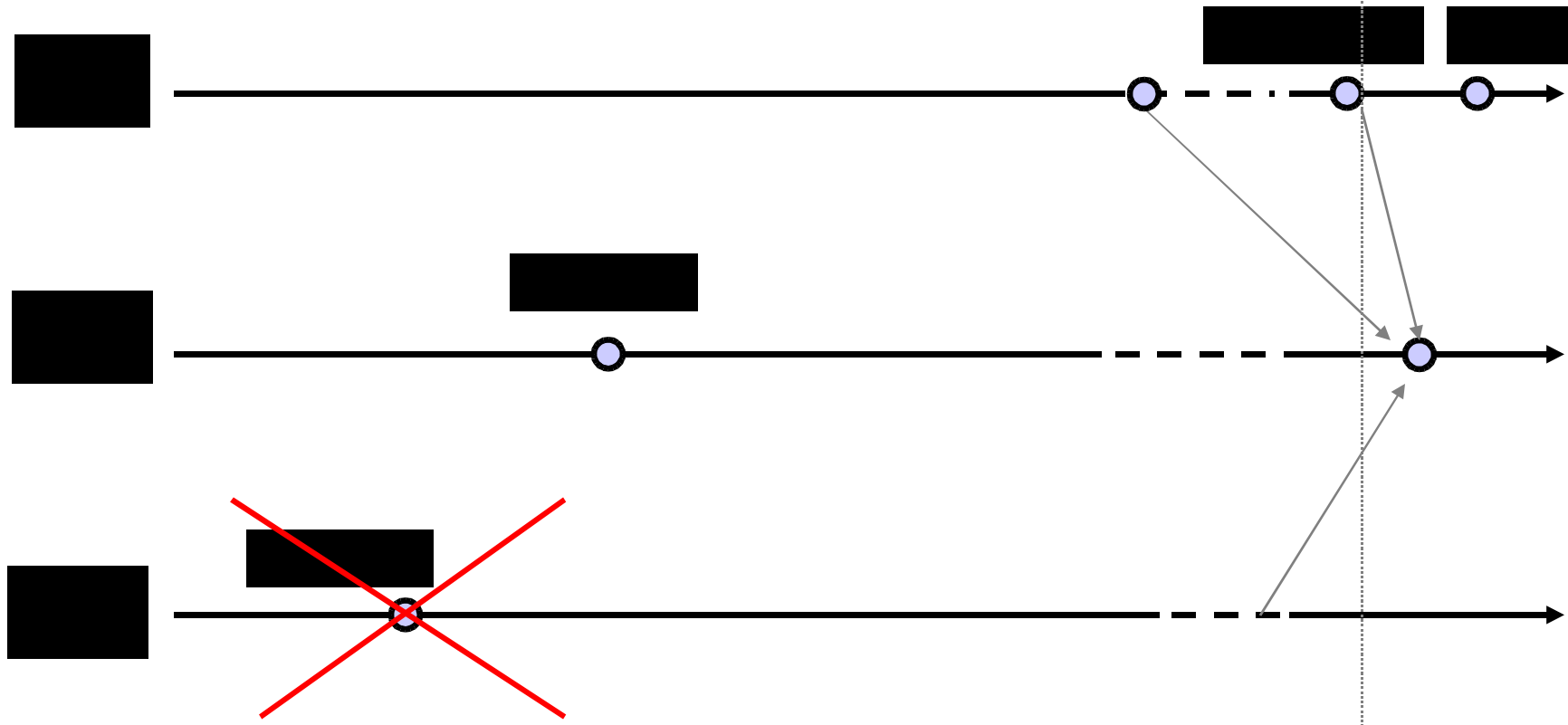
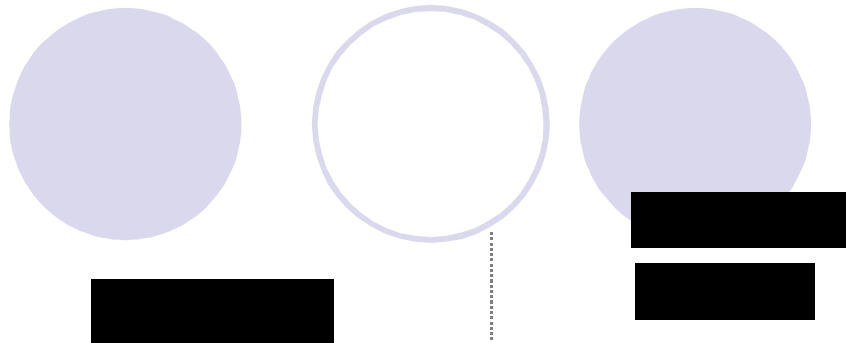
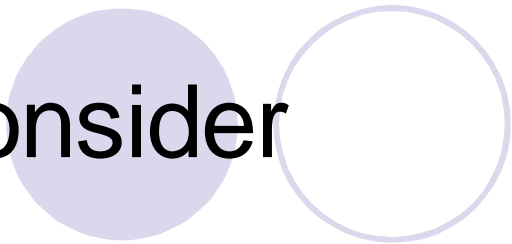
Consider



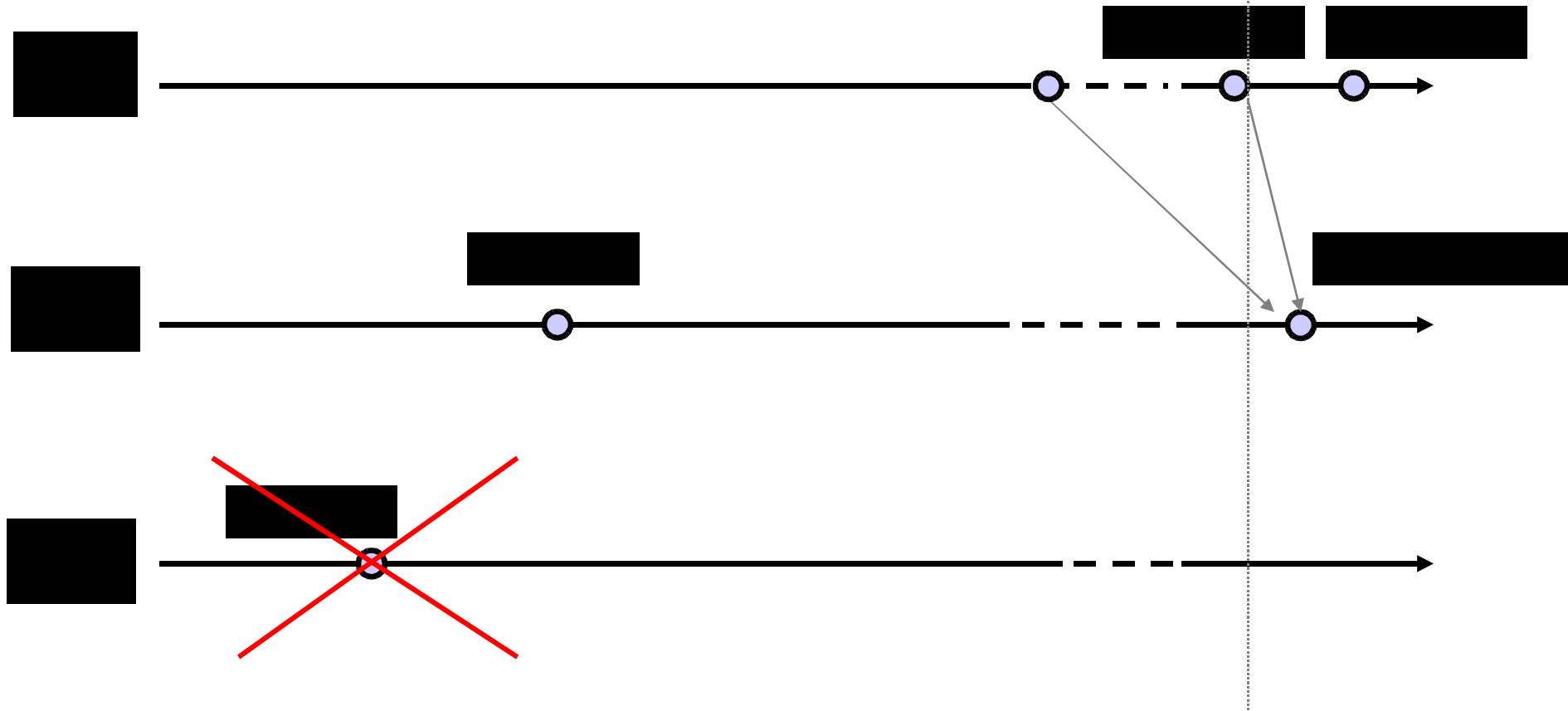
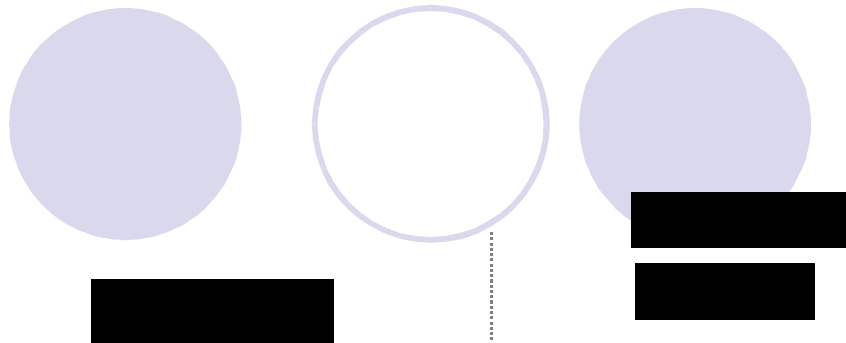
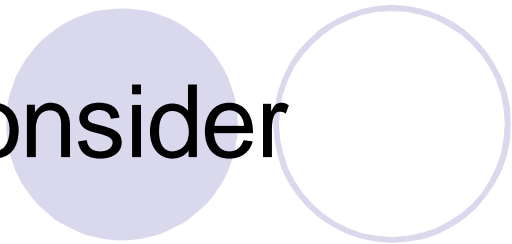
Consider



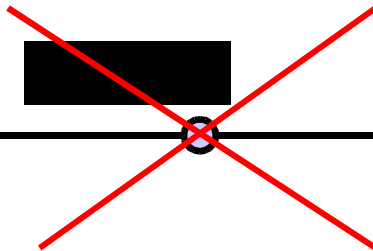
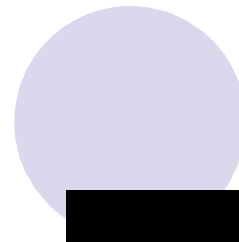
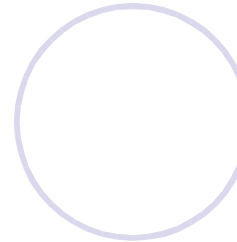
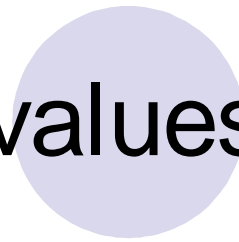
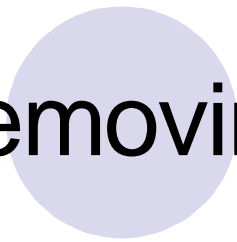
Consider



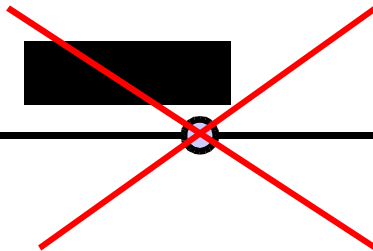
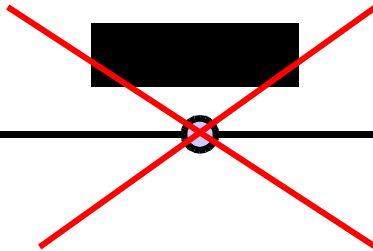
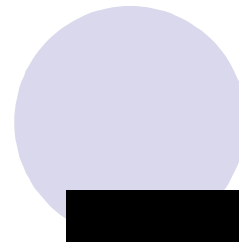
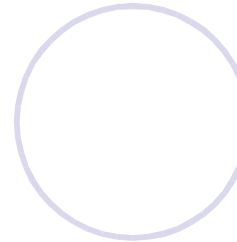
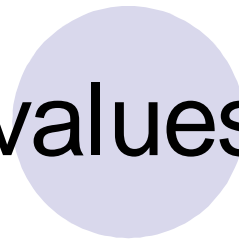
Consider



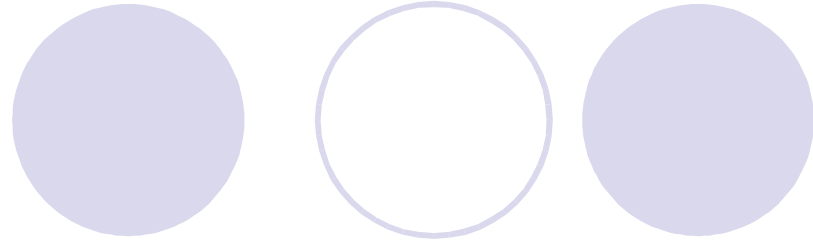
# Removing other old values



# Removing other old values



# Staleness bounds



- Any value that is more than



writes old will be rejected

- Lower Bound

  - Staleness of at least





# Conclusion

- Byzantine tolerant K-quorums
- Multi-writer K-quorum system
  - Built over single writer K-quorums
- Lower bound

The slide features five light purple circles. Two are solid and positioned in the lower-left quadrant. Three are hollow with a thin purple outline, arranged in a loose cluster in the upper-right quadrant. The text "Questions ?" is centered within one of the hollow circles.

**Questions ?**

**Thank You**