BlowFish

Dynamic Storage-Performance Tradeoff in Data Stores

Cheng Zhao

December 6, 2017

Layer Sampled Array

Sample rate: $\alpha = 8$

Idx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Values	9	11	15	2	3	1	0	6	12	13	8	7	14	4	5	10

LayerID	Exists Layer?
8	1

	9								12							

LayerID	8								8							
LayerIdx	0								1							

LayerID	8
Count	1

Layer Sampled Array

Sample rate: $\alpha = 4$

Idx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Values	9	11	15	2	3	1	0	6	12	13	8	7	14	4	5	10

LayerID	Exists Layer?															
8	1	9								12							
4	1					3								14			

LayerID	8				4				8				4			
LayerIdx	0				0				1				1			

LayerID	8	4
Count	1	1

Layer Sampled Array

Sample rate: $\alpha = 2$

Idx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Values	9	11	15	2	3	1	0	6	12	13	8	7	14	4	5	10

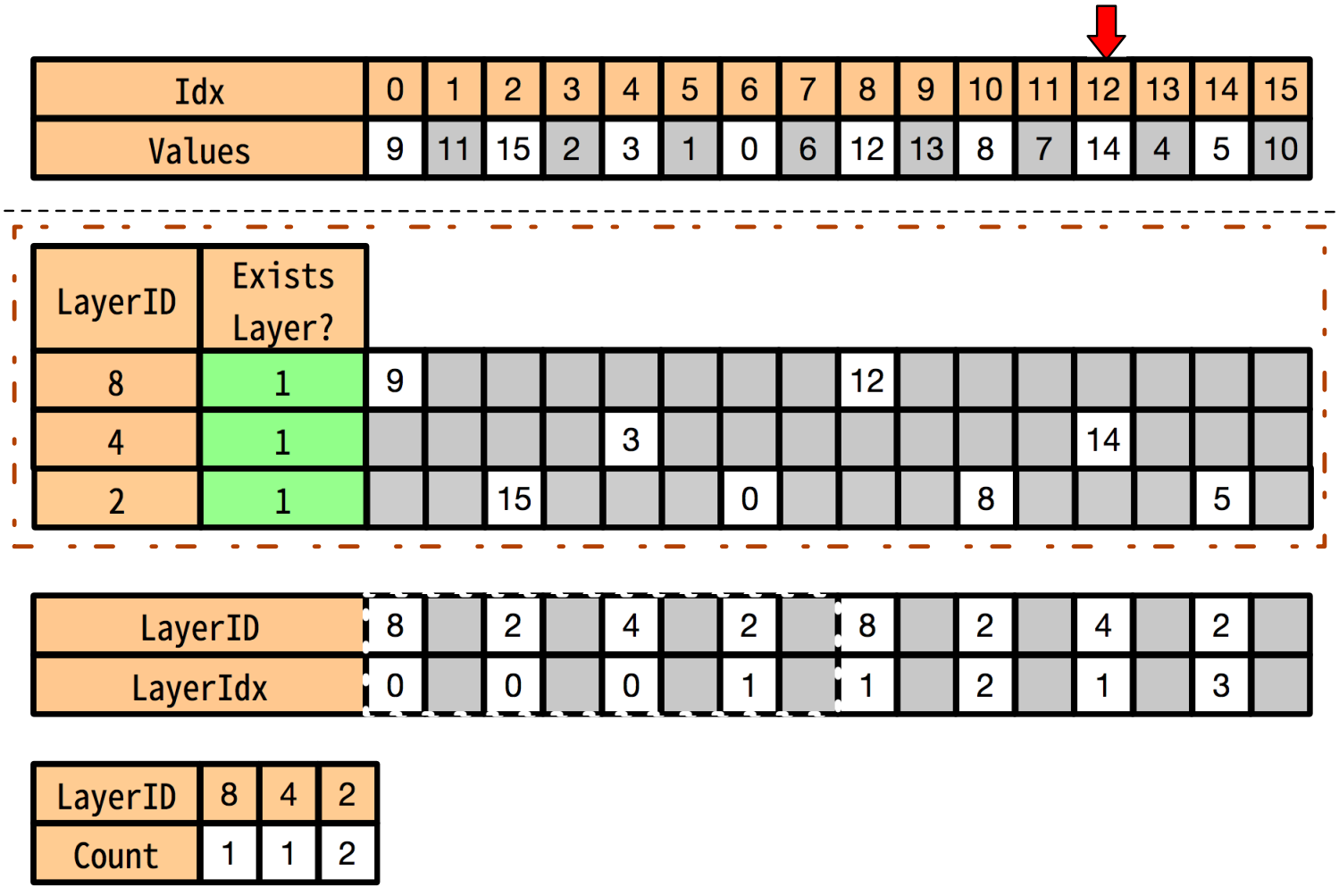
LayerID	Exists Layer?															
8	1	9								12							
4	1					3								14			
2	1			15				0				8				5	

LayerID	8		2		4		2		8		2		4		2	
LayerIdx	0		0		0		1		1		2		1		3	

LayerID	8	4	2
Count	1	1	2

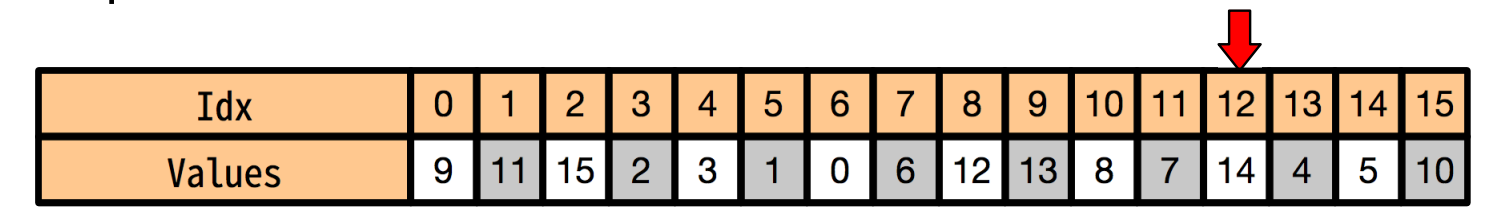
Layer Sampled Array LookUp

Sample rate: $\alpha = 2$



Layer Sampled Array LookUp

Sample rate: $\alpha = 2$



Idx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Values	9	11	15	2	3	1	0	6	12	13	8	7	14	4	5	10

LayerID	Exists Layer?															
8	1	9								12							
4	1					3								14			
2	1			15				0				8				5	

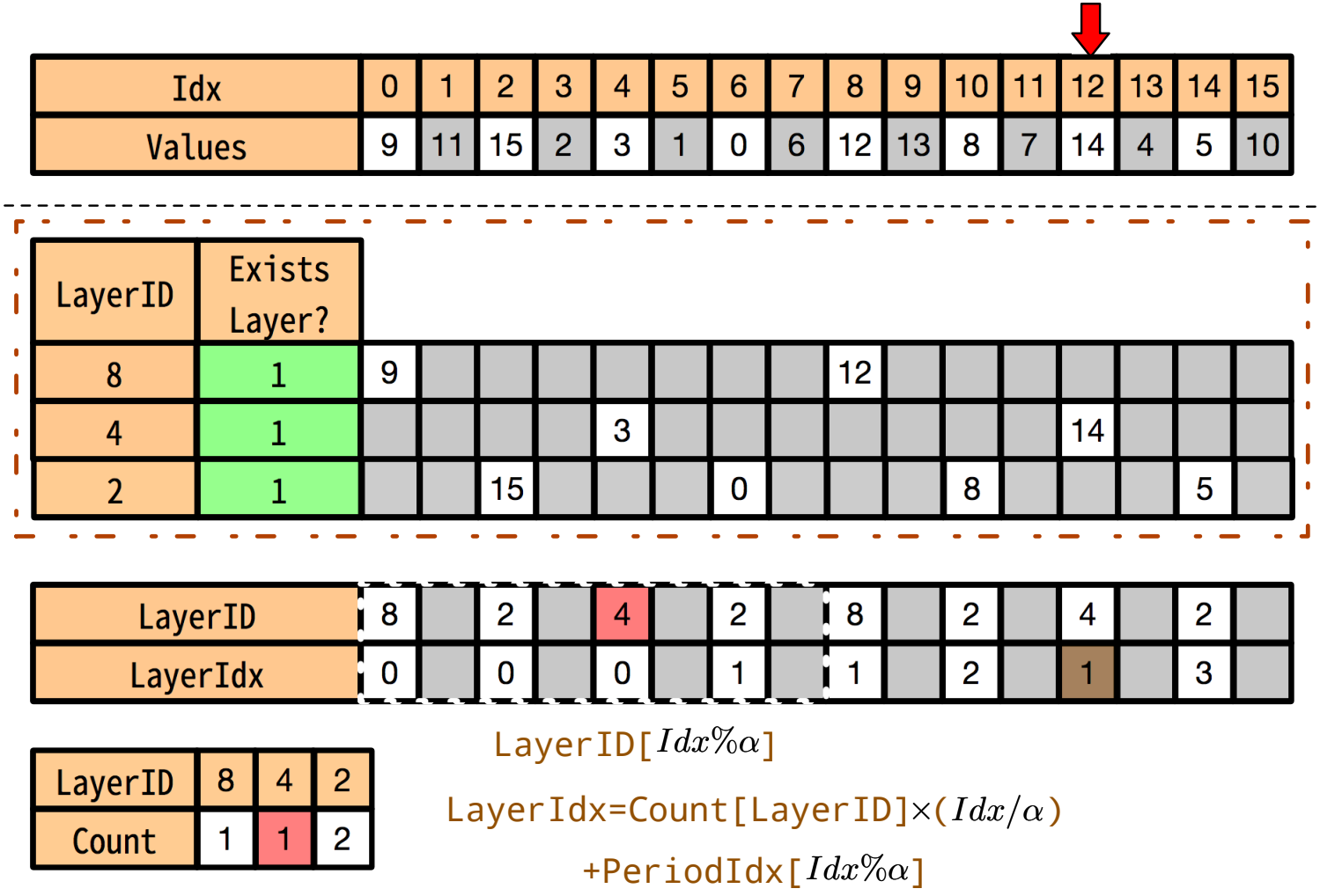
LayerID	8		2		4		2		8		2		4		2	
LayerIdx	0		0		0		1		1		2		1		3	

LayerID	8	4	2
Count	1	1	2

`LayerID[`$Idx\%\alpha$`]`

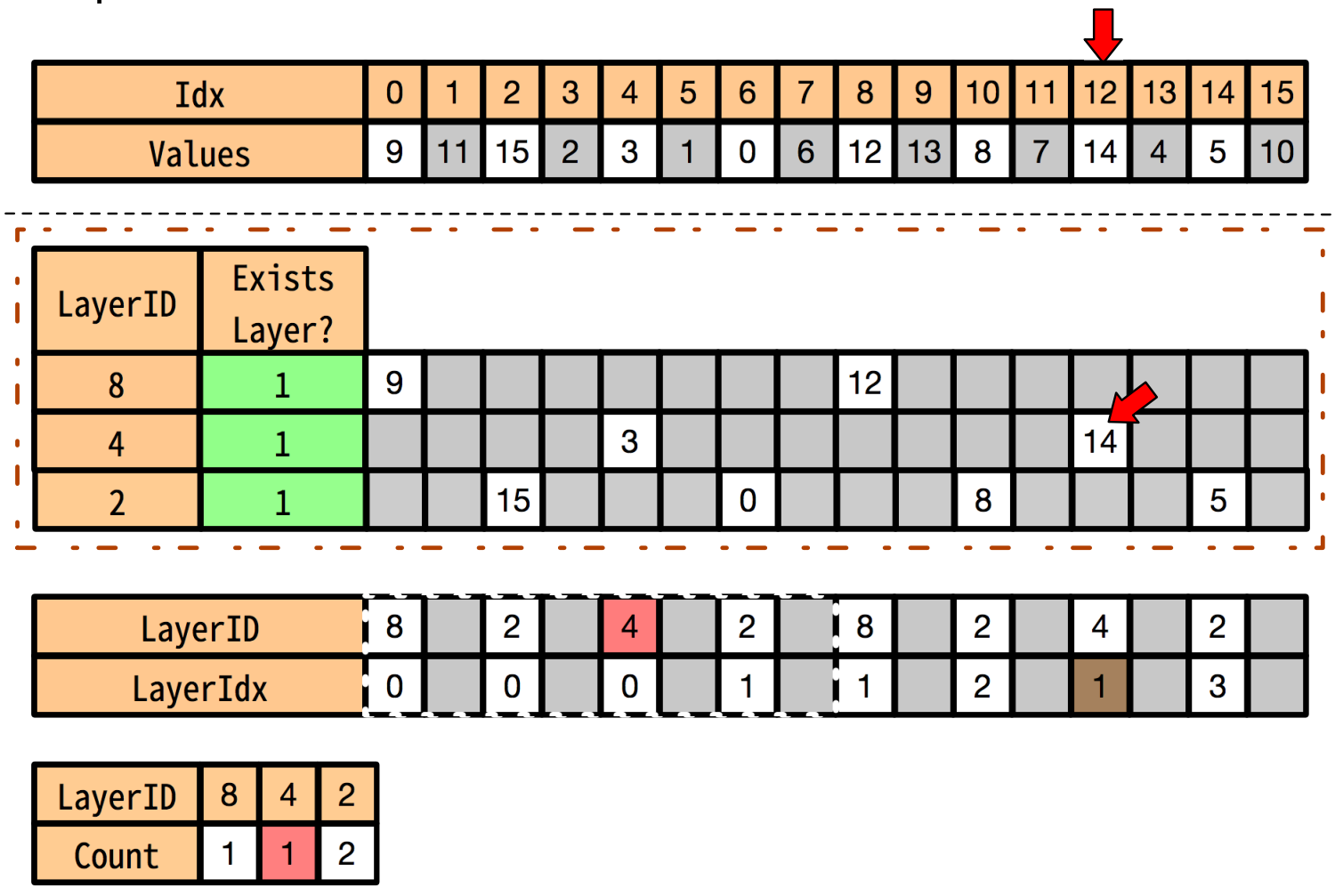
Layer Sampled Array LookUp

Sample rate: $\alpha = 2$



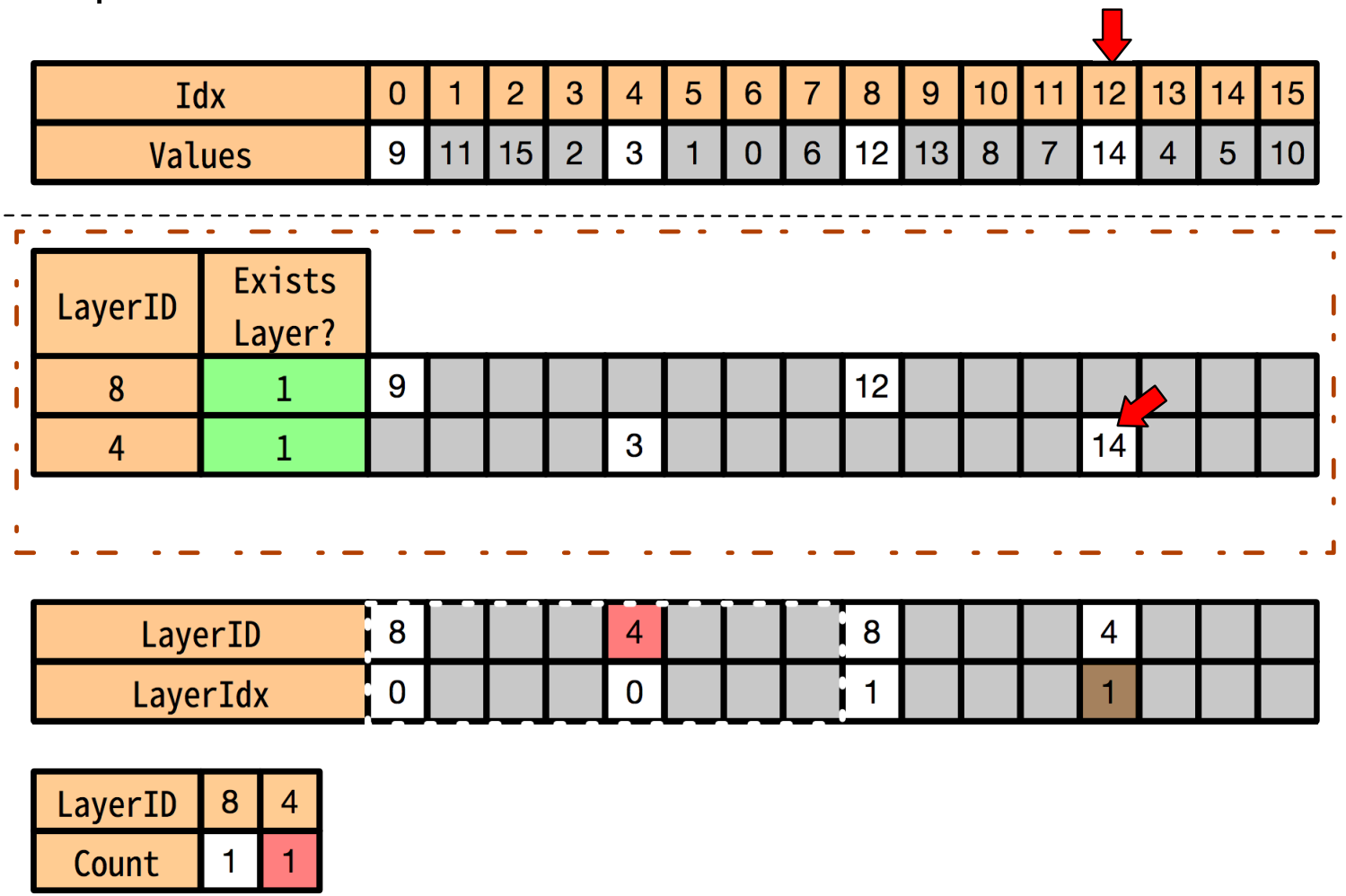
Layer Sampled Array LookUp

Sample rate: $\alpha = 2$



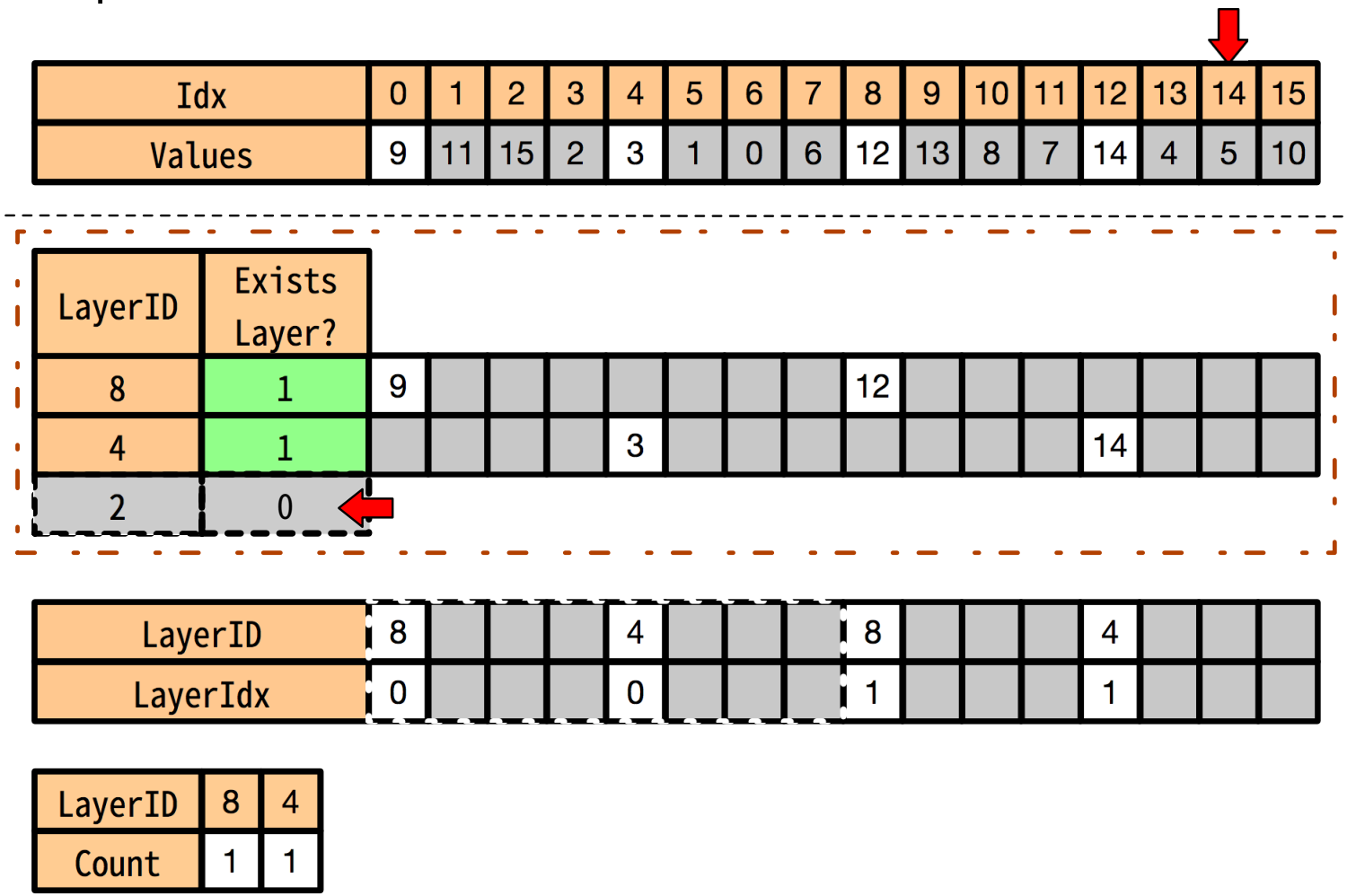
Opportunistic Layer Construction

Sample rate: $\alpha = 4$



Opportunistic Layer Construction

Sample rate: $\alpha = 4$



Opportunistic Layer Construction

Sample rate: $\alpha = 2$

Idx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Values	9	11	15	2	3	1	0	6	12	13	8	7	14	4	5	10

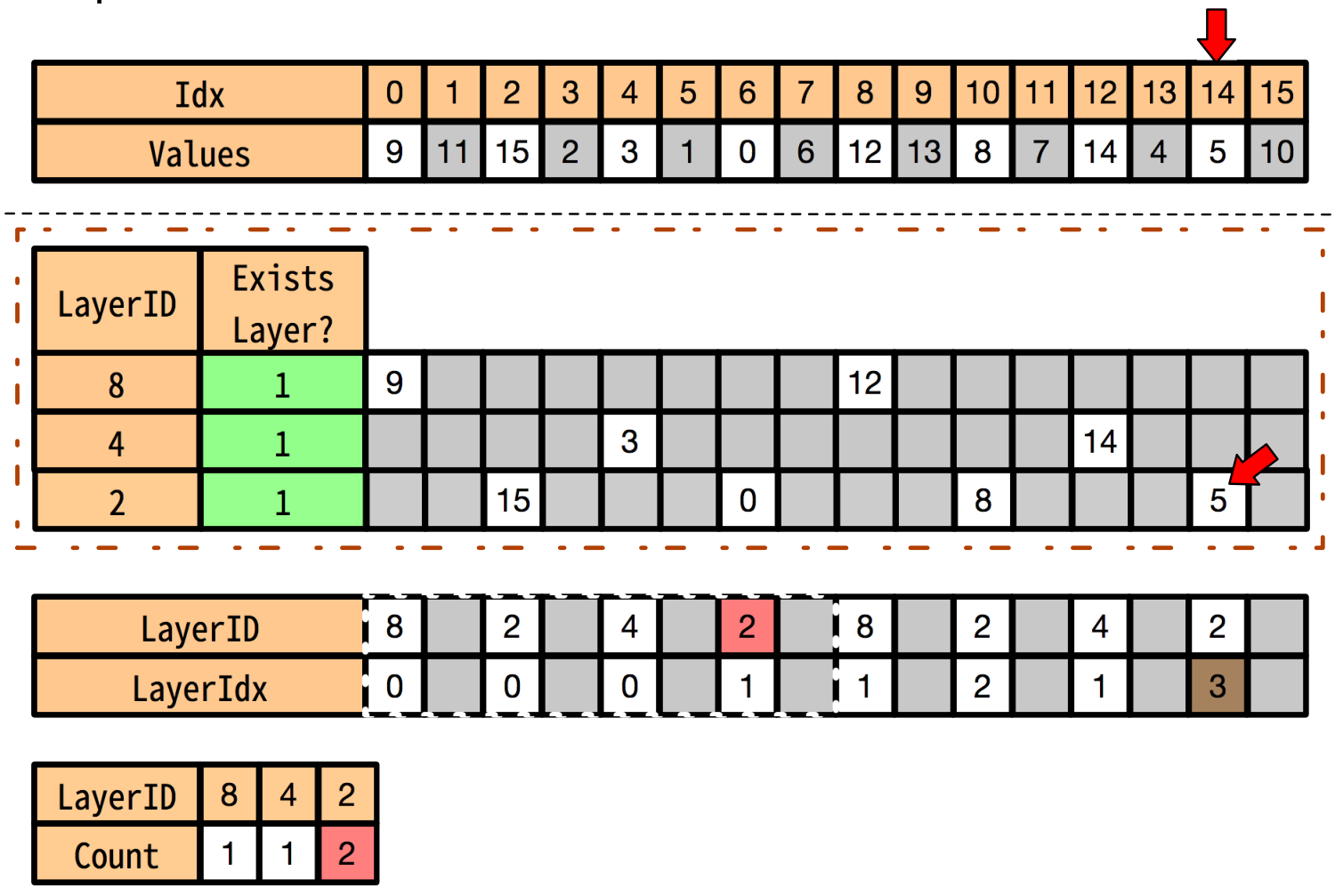
LayerID	Exists Layer?															
8	1	9								12							
4	1					3								14			
2	1			15				0				8				5	

LayerID	8		2		4		2		8		2		4		2	
LayerIdx	0		0		0		1		1		2		1		3	

LayerID	8	4	2
Count	1	1	2

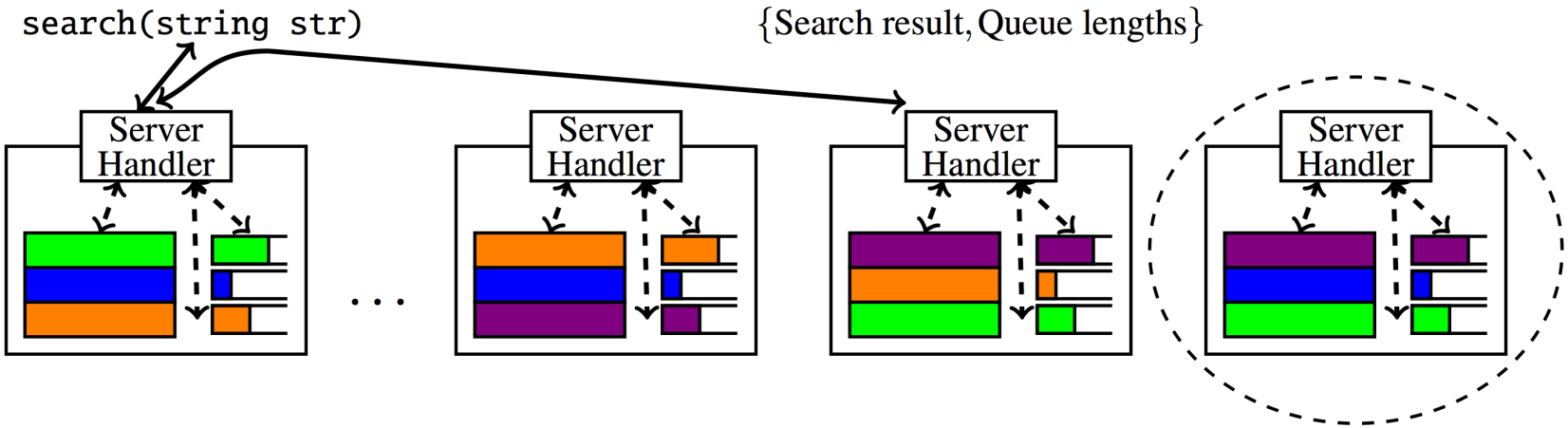
Opportunistic Layer Construction

Sample rate: $\alpha = 2$



Scheduling and Load Balancing

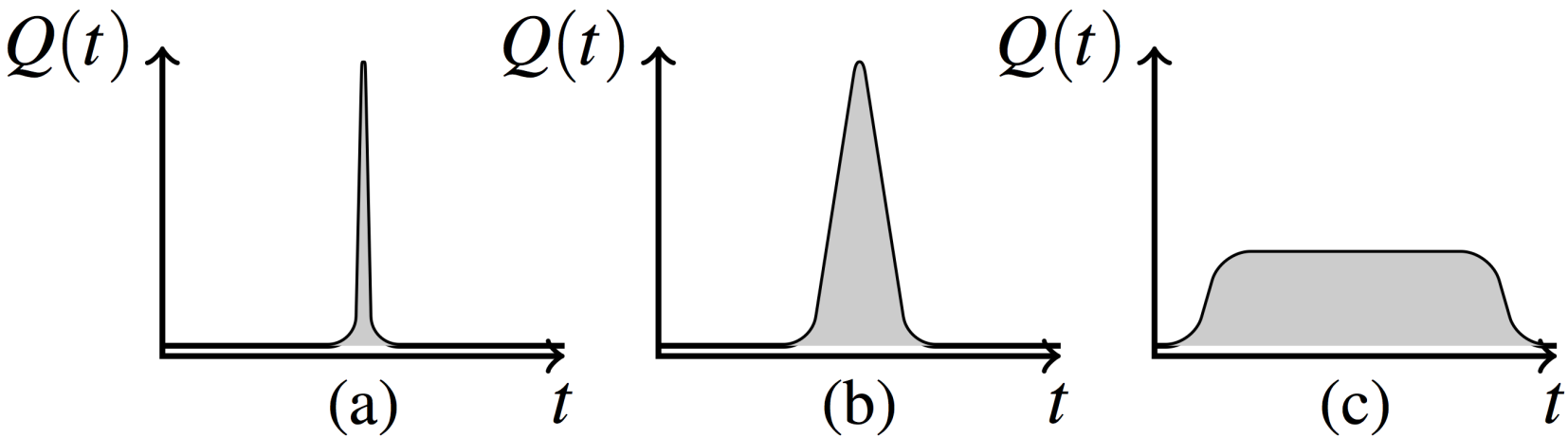
- Architecture



- Different replicas of the same shard may have different query execution time for the same query.
- **Join-the-shortest-queue** mechanism.

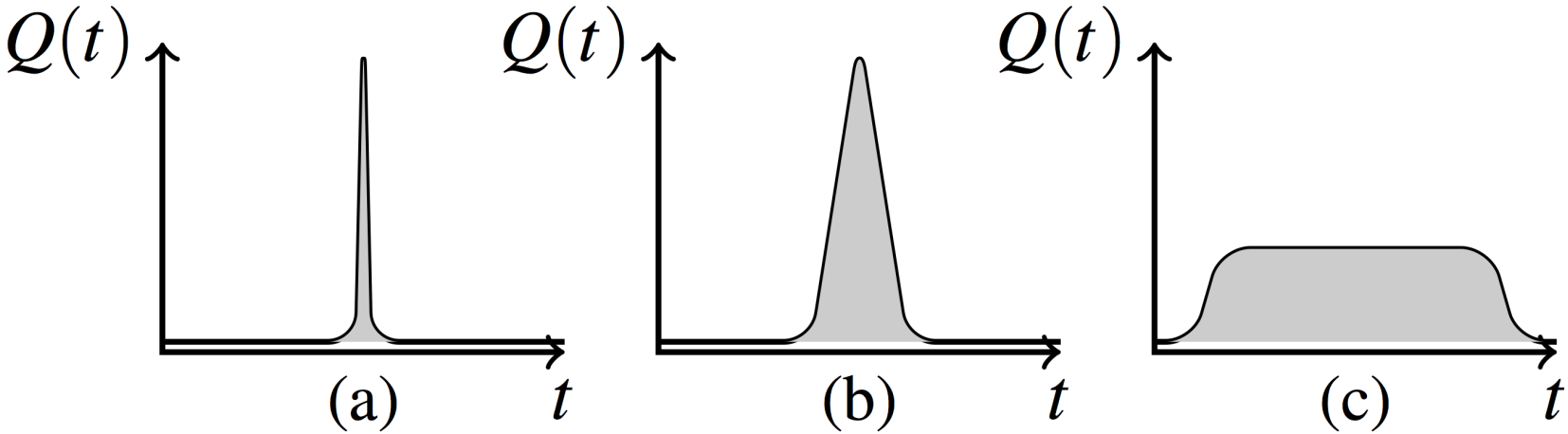
Dynamically Navigating the Tradeoff

- Three different scenarios of queue length ($Q(t)$) variation with time (t).



- a) Should be ignored.
- b) & c) Should adapt queue length.

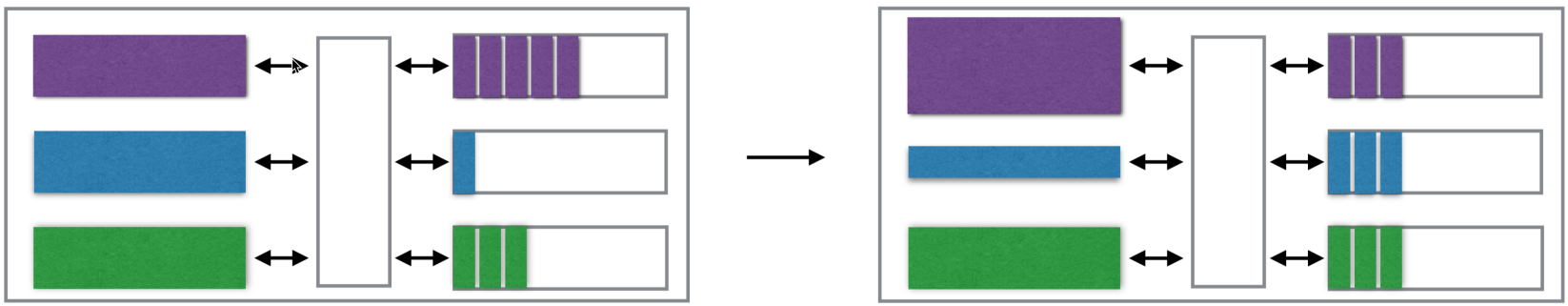
Dynamically Navigating the Tradeoff



- $$Q_t^{avg} = \beta \times Q_t + (1 - \beta) \times Q_{t-\delta}^{avg}$$
- β is the weight of recent queue length values.
- δ is the periodicity.

Dynamically Navigating the Tradeoff

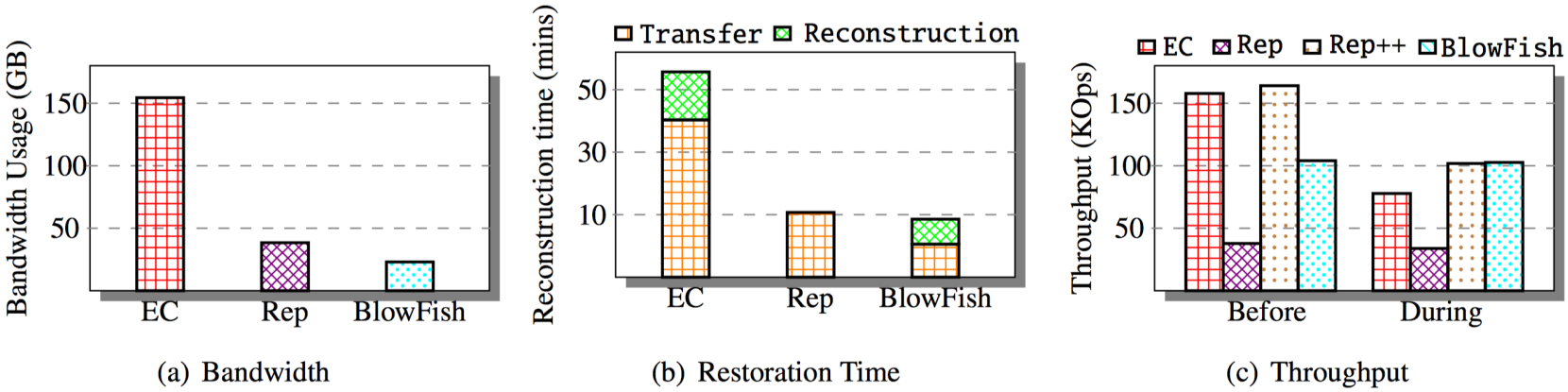
- Shards on the same server:



- Shard replica across servers: **Join-the-shortest-queue** mechanism.

Compare with EC & 3-Replication

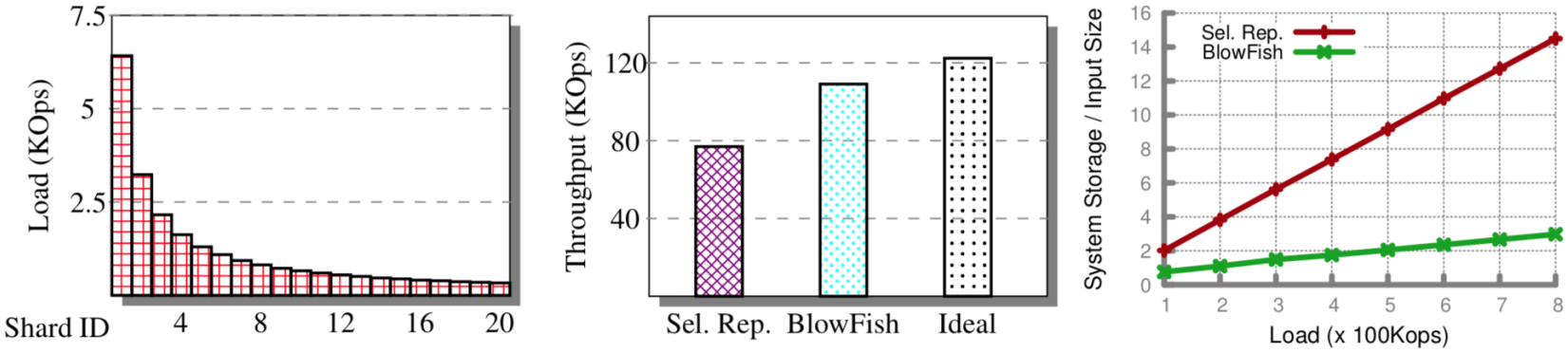
	Erasure Codes	Replication	BlowFish
Storage	1.2×	3×	1.8×
Repair Bandwidth	10×	1×	1×



(a) Bandwidth (b) Restoration Time (c) Throughput

- b) EC: the access link at the server where the data is being collected becomes the network bottleneck.
- 3-Replication and BlowFish is limited by load-balance.
- Rep++ is only one replica.

Skewed Workloads



(a) Load distribution across shards (b) Throughput for a fixed storage (c) Storage required to sustain load

- Load: Zipf distribution with skewness 0.01 (heavily skewed) across the shards.

Time-varying Workloads

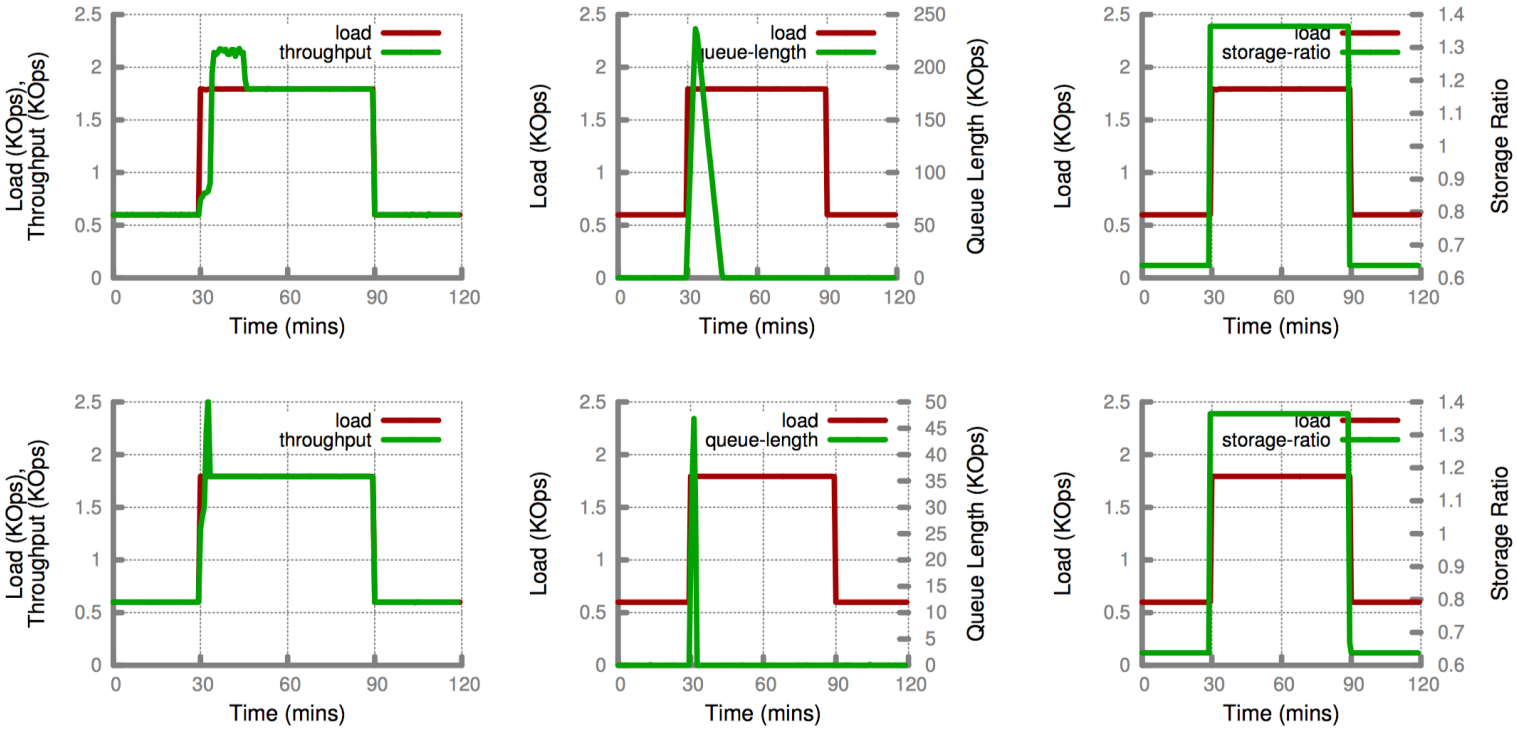


Figure 9: Opportunistic layer construction with spiked changes in load for uniform workload (top three) and skewed workload (bottom three). The figures show variation in throughput (left), request queue length (center) and storage footprint (right).

Time-varying Workloads

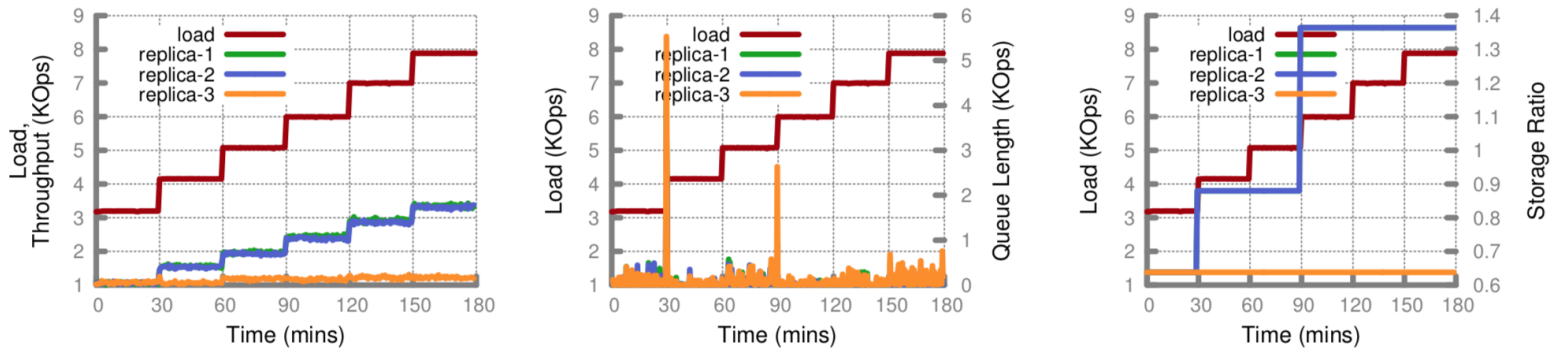


Figure 10: The effectiveness and stability of BlowFish's query scheduling mechanism in a replicated system (discussion in §4.4). Variation in throughput (left), request queue lengths (center) and storage-footprints (right) for the three replicas.