

Parameterized and Runtime-tunable Snapshot Isolation in Distributed Transactional Key-value Stores

Hengfeng Wei, Yu Huang, Jian Lu

Nanjing University, China

September 16, 2017



Parameterized and Runtime-tunable Snapshot Isolation

RVSI: Relaxed Version Snapshot Isolation

1 Experimental Evaluation

Parameterized and Runtime-tunable Snapshot Isolation

RVSI: Relaxed Version Snapshot Isolation

1 Experimental Evaluation

Impacts of RVSI specification on the *transaction abort rates* in various scenarios

Impacts of RVSI specification on the *transaction abort rates* in various scenarios

Performance?

- ▶ Not done yet in this work
- ▶ CHAMELEON prototype is ...

Transactions abort for two reasons:

- ▶ “vc-aborted”: RVSI version constraints violated
- ▶ “wcf-aborted”: the WCF property violated

Transactions abort for two reasons:

- ▶ “vc-aborted”: RVSI version constraints violated
- ▶ “wcf-aborted”: the WCF property violated

Transaction abort rates due to “vc-aborted” are *sensitive* to different values of k_1 , k_2 , or k_3 ,

Transactions abort for two reasons:

- ▶ “vc-aborted”: RVSI version constraints violated
- ▶ “wcf-aborted”: the WCF property violated

Transaction abort rates due to “vc-aborted” are *sensitive* to different values of k_1 , k_2 , or k_3 , but those due to “wcf-aborted” are not.

Transactions abort for two reasons:

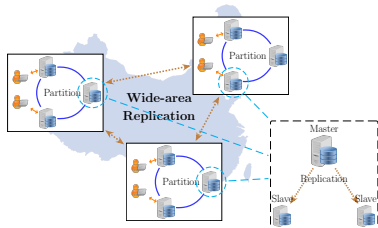
- ▶ “vc-aborted”: RVSI version constraints violated
- ▶ “wcf-aborted”: the WCF property violated

Transaction abort rates due to “vc-aborted” are *sensitive* to different values of k_1 , k_2 , or k_3 , but those due to “wcf-aborted” are not.

$$h \in \text{RVSI} \iff h \in k_1\text{-BV} \cap k_2\text{-FV} \cap k_3\text{-SV} \cap \text{WCF}.$$

CHAMELEON prototype on Aliyun:

- ▶ 3 datacenters ¹
- ▶ 3 nodes in each datacenter
- ▶ Partition & Replication
- ▶ Clients in our lab ²



¹ Located in East China, North China, and South China, respectively.

² Located in East China.

³ <https://github.com/hengxin/aliyun-ping-traces>

CHAMELEON prototype on Aliyun:

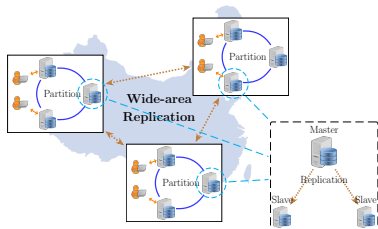
- ▶ 3 datacenters ¹
- ▶ 3 nodes in each datacenter
- ▶ Partition & Replication
- ▶ Clients in our lab ²

(One-way) delays among nodes ³:

Within datacenter: 1 ~ 2ms

Across datacenters: 15 ~ 25ms

Clients to nodes: 15 ~ 20ms



¹Located in East China, North China, and South China, respectively.

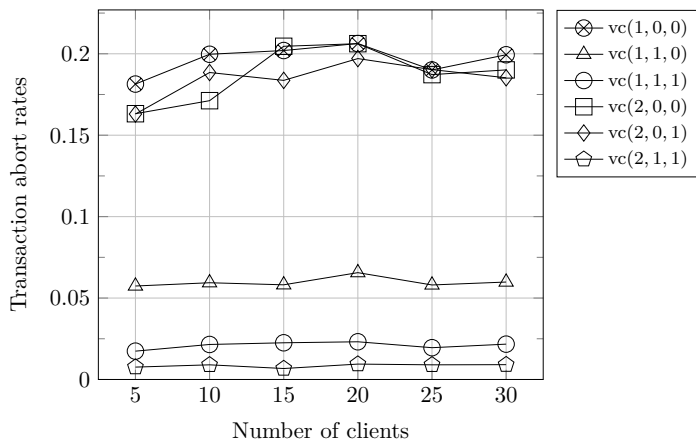
²Located in East China.

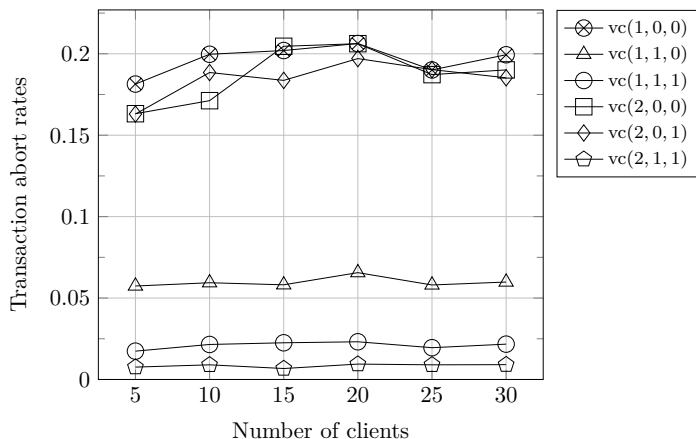
³<https://github.com/hengxin/aliyun-ping-traces>

Table: Three categories of workload parameters for experiments on Aliyun.

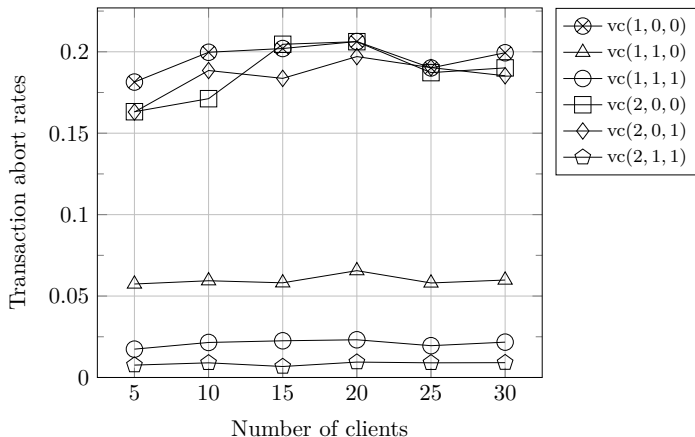
Parameter		Value	
Transaction-related	#keys	$25 = 5 \text{ (rows)} \times 5 \text{ (columns)}$	
	#clients	5, 10, 15, 20, 25, 30	
	#txs/client	1000	
	#ops/tx	$\sim \text{Binomial}(20, 0.5)$	n
	rwRatio	1:2, 1:1, 4:1	
	zipfExponent	1	
Execution-related	minInterval	0ms	
	maxInterval	10ms	
	meanInterval	5ms	
RVSI-related	(k_1, k_2, k_3)	$(1,0,0) (1,1,0) (1,1,1)$ $(2,0,0) (2,0,1) (2,1,1)$	

under read-frequent workloads



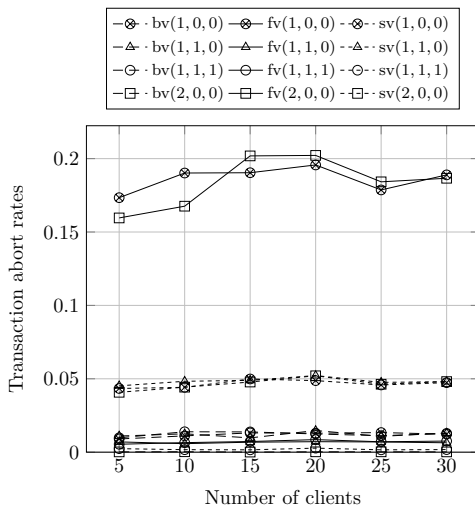


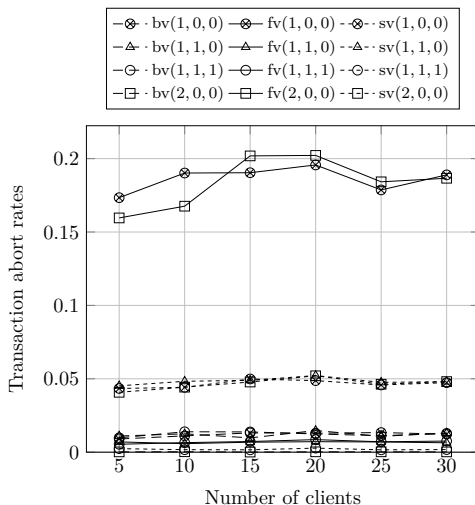
The transaction abort rates due to “vc-aborted”



The transaction abort rates due to “vc-aborted” can be **greatly reduced** by **slightly** increasing the values of k_1 , k_2 , or k_3 :

$$vc(1, 0, 0) = 0.1994 \implies vc(2, 1, 1) = 0.0091 \quad (\#clients = 30)$$





Most “vc-aborted” transactions abort because of violating $k_2\text{-FV}$.

$$fv(1, 0, 0) = 0.1889 \implies fv(2, 0, 0) = 0.1866 \implies fv(1, 1, 0) = 0.0064$$

Question: when does k_1 for k_1 -BV take effect?

It seems that k_1 -BV has *little* impact on the transaction abort rates.

Question: when does k_1 for k_1 -BV take effect?

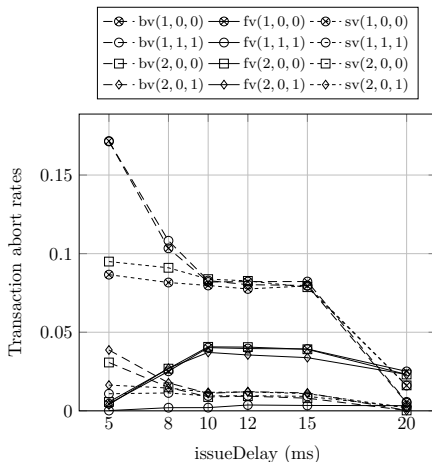
It seems that k_1 -BV has *little* impact on the transaction abort rates.

It may be the case in the Aliyun scenario.

What about other scenarios?

Three types of delays for **controlled experiments** on local hosts.

Types	Values (ms)	Explanation
issueDelay	5, 8, 10, 12, 15, 20	delays between clients and replicas
replDelay	5, 10, 15, 20, 30	delays between masters and slaves
2pcDelay	10, 20, 30, 40, 50	delays among masters



When the “**issueDelay**” gets shorter,
the impacts of k_2 -FV go weaker,
and the impacts of k_1 -BV have begun to emerge.

issueDelay = 20ms: $bv(1, 0, 0) = 0.0057$ $fv(1, 0, 0) = 0.0251$

issueDelay = 15ms: $bv(1, 0, 0) = 0.08225$ $fv(1, 0, 0) = 0.0393$

issueDelay = 5ms: $bv(1, 0, 0) = 0.1716$ $fv(1, 0, 0) = 0.0045$

Generally, RVSI helps to reduce the transaction abort rates when applications are willing to tolerate certain anomalies.

Generally, RVSI helps to reduce the transaction abort rates when applications are willing to tolerate certain anomalies.

k_2 -FV: On Aliyun, most transactions have been aborted because of violating k_2 -FV.

Generally, RVSI helps to reduce the transaction abort rates when applications are willing to tolerate certain anomalies.

k_2 -FV: On Aliyun, most transactions have been aborted because of violating k_2 -FV.

k_1 -BV: In controlled experiments,

Generally, RVSI helps to reduce the transaction abort rates when applications are willing to tolerate certain anomalies.

k_2 -FV: On Aliyun, most transactions have been aborted because of violating k_2 -FV.

k_1 -BV: In controlled experiments,

k_3 -SV:

