Towards an Edge-Located Time-Series Database

Timothy Krentz, Abhishek Dubey, Gabor Karsai
Institute for Software Integrated Systems
Vanderbilt University

Tel (615) 343-7472 Fax (615) 343-7440
1025 16th Avenue South | Nashville, TN 37212
www.isis.vanderbilt.edu

VANDERBILT UNIVERSITY

Emerging Trends

- IoT/Edge will continue to grow exponentially
  - 125B devices by 2030 [1]
- Smart cities and infrastructure create spatio-temporal data
  - Promote systems needing multi-dimensional queries
  - Systems like RIAPS [2] recovering from a fault may need time-series data
- Many existent time-series databases
  - InfluxDB
  - BTrDb
- Some systems may need cheap/ephemeral timeseries storage
  - Group of drones decide to pool memory temporarily

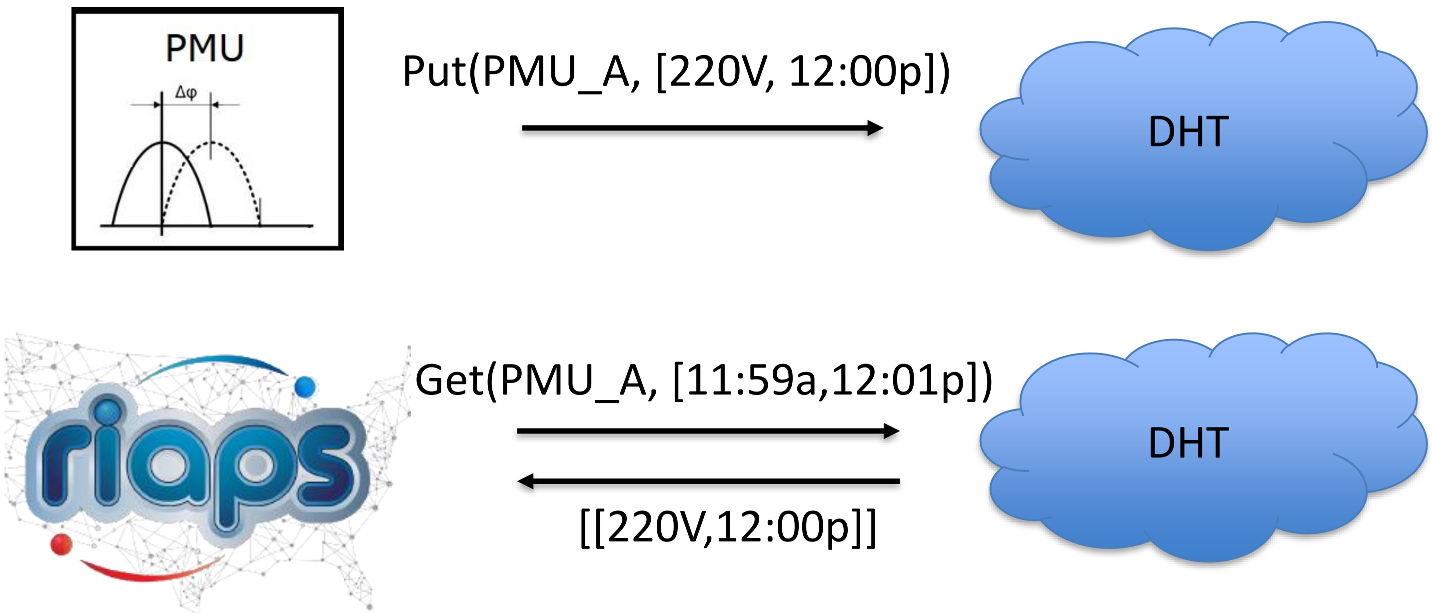
Goals

Desired solution should be:

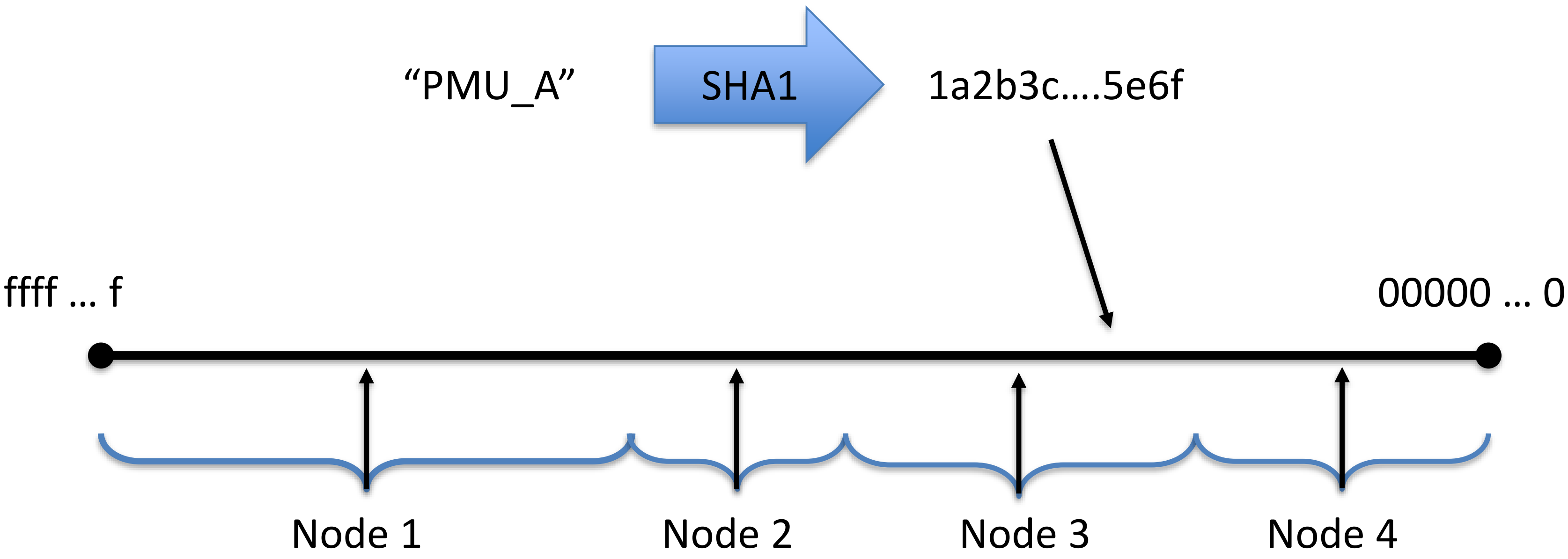
- Distributed
- Lightweight
- Simple API
  - Put(key,[value, timestamp])
  - Get(key,[timestamp_begin, timestamp_end])

DHTs

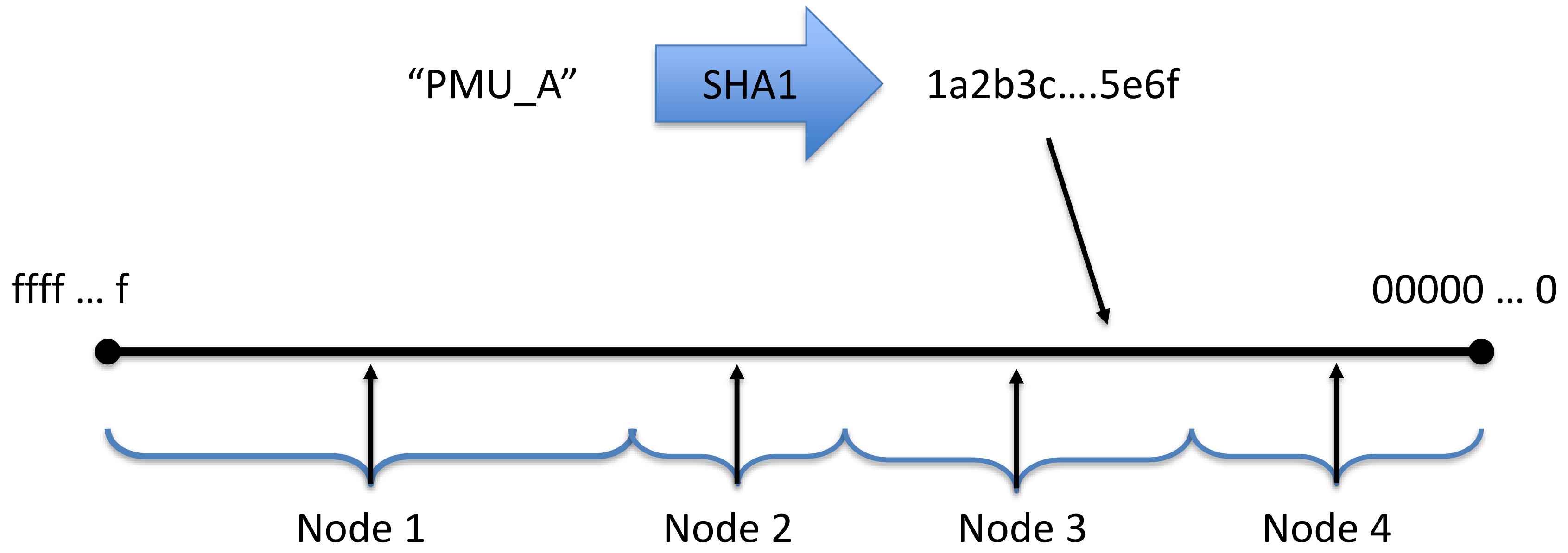
- Distributed Hash Tables (DHTs) provide distribution, are simple and lightweight, and have established use in these systems.



DHTs



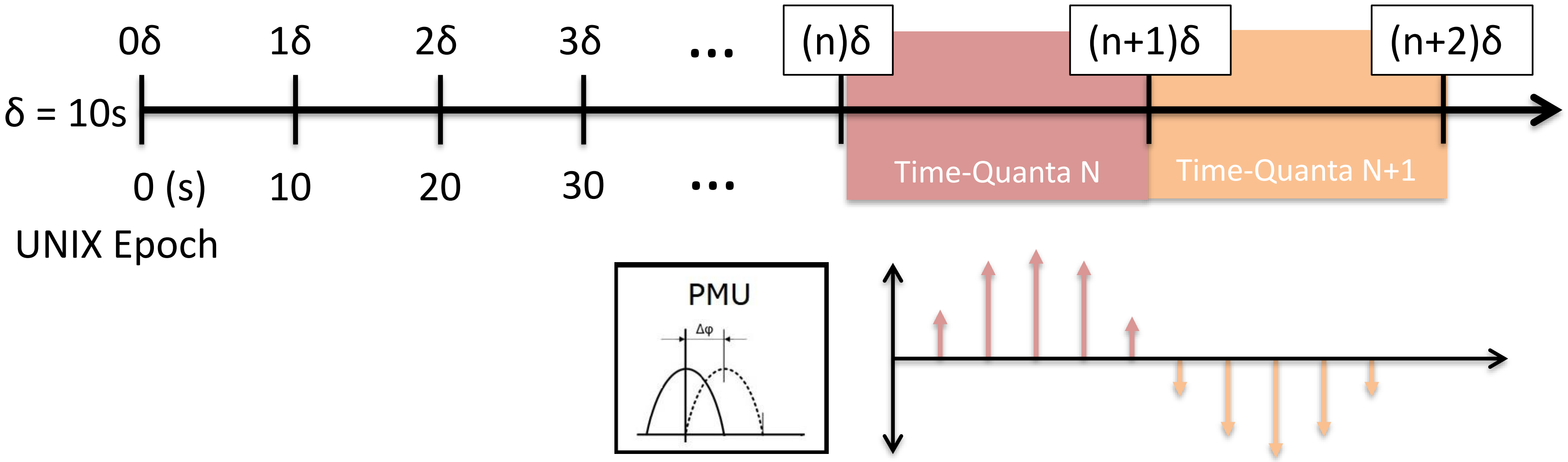
DHTs



- Problem: Access method relies on an exact key
- Including an exact timestamp with the application key hashes the datapoint to a random location in the table, making retrieval difficult
- Need a way to include time in storage and lookup without making lookup expensive

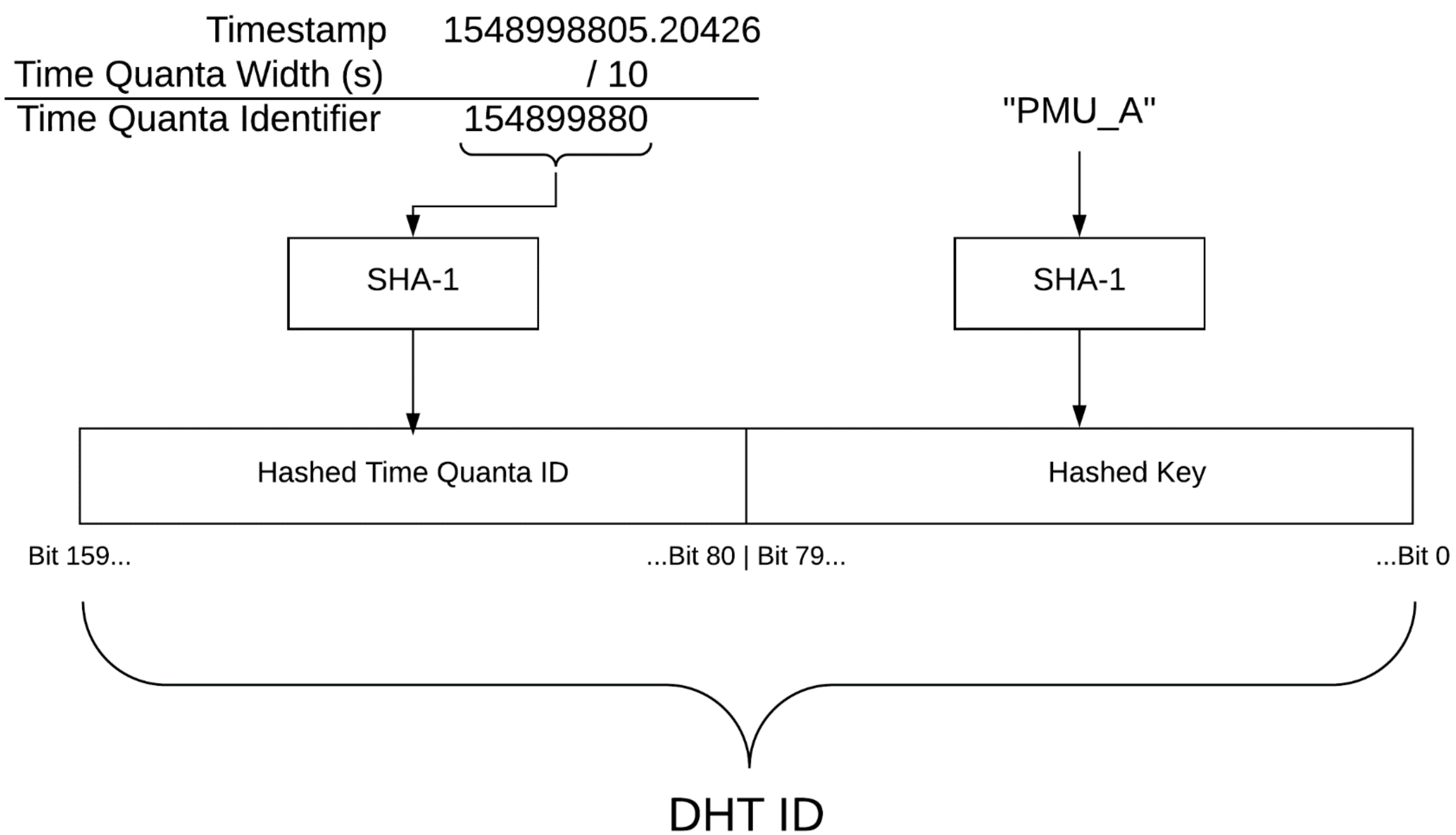
Time-Quanta

- Discretize continuous time into "time-quanta" of a constant width δ
- Timestamped samples are to be stored within that time quanta
- Time quanta are consistently addressable, providing granular access to datapoints



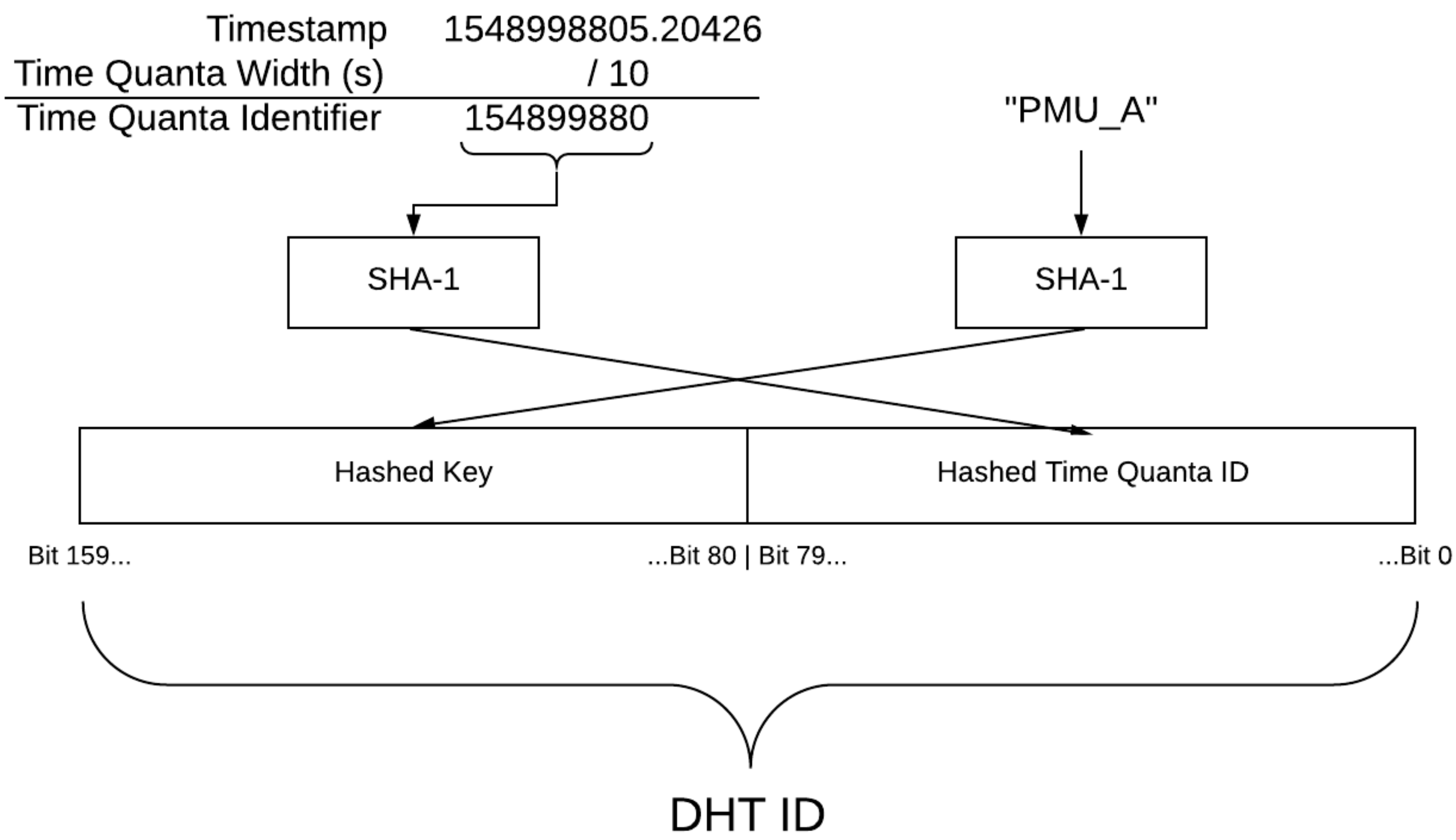
Time-Factored Key: Quanta-First ID (QFI)

DHT Query (Timestamp: 1548998805.20426, Key: "PMU_A")



Time-Factored Key: Key-First ID (KFI)

DHT Query (Timestamp: 1548998805.20426, Key: "PMU_A")



Expected outcomes of Quanta-First vs Key-First

- QFI
  - Samples within a measurement are distributed evenly throughout all participating nodes
  - Time-series lookups can be parallelized
- KFI
  - Samples within a measurement likely reside on the same nodes
  - Time-series lookups likely only need to contact a few nodes

Experimental Results

- 18 Beaglebone Blacks, Ubuntu 18.04
- DHT written in Go
- A single node stores 10,000 PMU measurements at 60 Hz
- Time-quanta width = 10s
- Measured read/write times for various factors of replication and time-series lengths requested

Experimental Results

- Write speed grows with replication, still can reach 7 replicas while being viable for 60 Hz data streams

- QFI read speed vs time-series request length levels off as expected, due to parallelized lookups. KFI reads showed similar behavior.

TABLE I
WRITE PERFORMANCE (MS/WRITE) VS REPLICATION FACTOR

Key Format	Replication Factor		
	1	4	7
QFI	4.54	10.9	14.1
KFI	5.05	10.1	13.4

TABLE II
READ PERFORMANCE (S/READ) VS REQUEST TIMESPAN

Key Format	Time-Series Length(s)		
	10	80	150
QFI	0.176	4.33	3.54
KFI	0.081	3.354	2.45

Moving forward

- Compare performance on the same hardware with pre-existing solutions
- Create better underlying storage for each time-quanta
- Spatio-temporal data applications need even higher-dimensional lookup schemes, DHTs using scalar “distance” limit them to a single dimension.

Thank you!

Tel (615) 343-7472 Fax (615) 343-7440
1025 16th Avenue South|Nashville, TN 37212
www.isis.vanderbilt.edu

VANDERBILT UNIVERSITY

References

1. Howell, J. (2017). *Number of Connected IoT Devices Will Surge to 125 Billion by 2030, IHS Markit Says - IHS Technology*. [online] Technology.ihs.com. Available at: https://technology.ihs.com/596542/number-of-connected-iot-devices-will-surge-to-125-billion-by-2030-ihs-markit-says [Accessed 1 May 2019].
2. S. Eisele, I. Madari, A. Dubey, and G. Karsai, "Riaps: Resilient information architecture platform for decentralized smart systems," in 2017 IEEE 20th International Symposium on Real-Time Distributed Computing (ISORC), May 2017, pp. 125–132.