An IoT Data Communication Framework for Authenticity and Integrity

Xin Li*, **Huazhe Wang***, **Ye Yu†**, **Chen Qian***

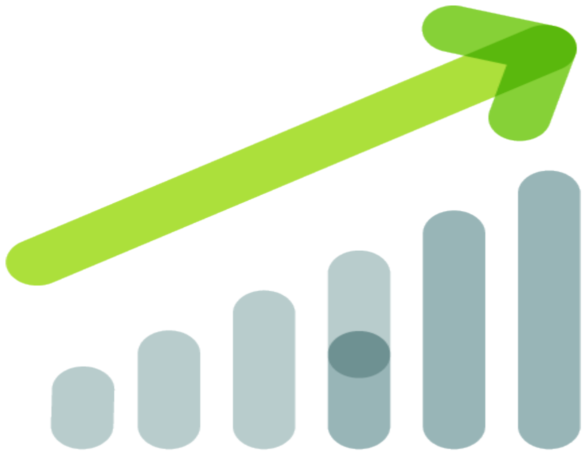
*University of California Santa Cruz

†University of Kentucky

IoT is ubiquitous

IoT trends

By 2020



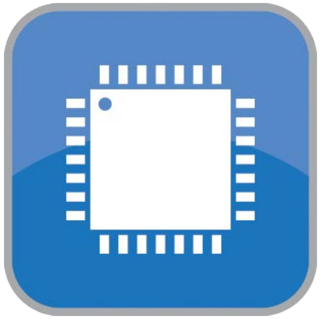
4 Billion

Connected People

4 Trillion

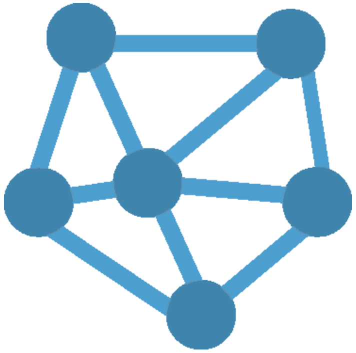
Revenue

25+ Billion



Device

50 Trillion GB



Traffic

IoT Data Applications: data consumers

Analytics

- Environment monitoring
- Traffic estimation
- Business decision making

Prediction

- Whether forecast
- Electricity load forecast

Real-time control

- Autonomous car
- Manufacturing
- Smart lighting

IoT device hardware platforms

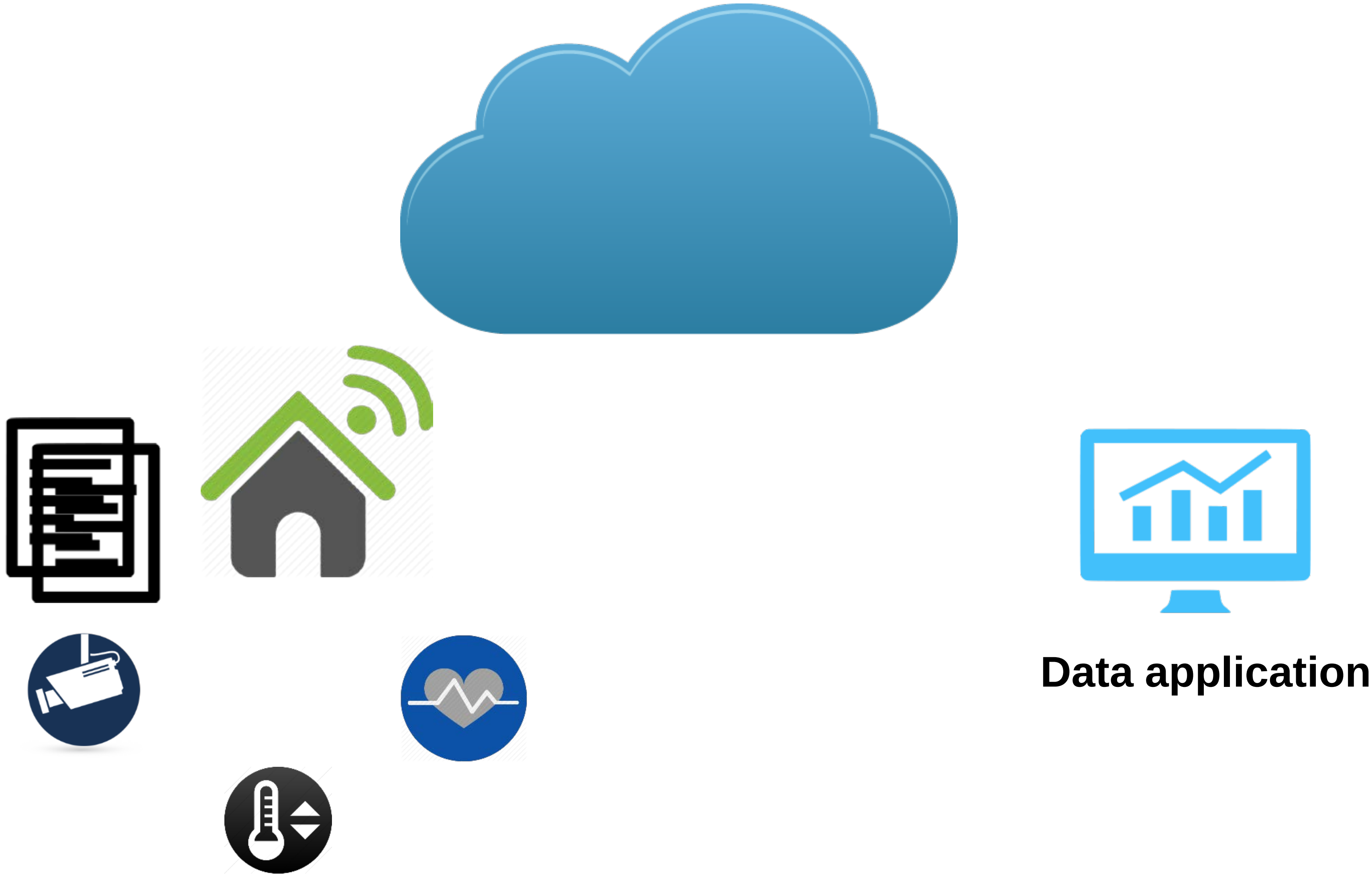
ARDUINO

intel Edison

SODAQ

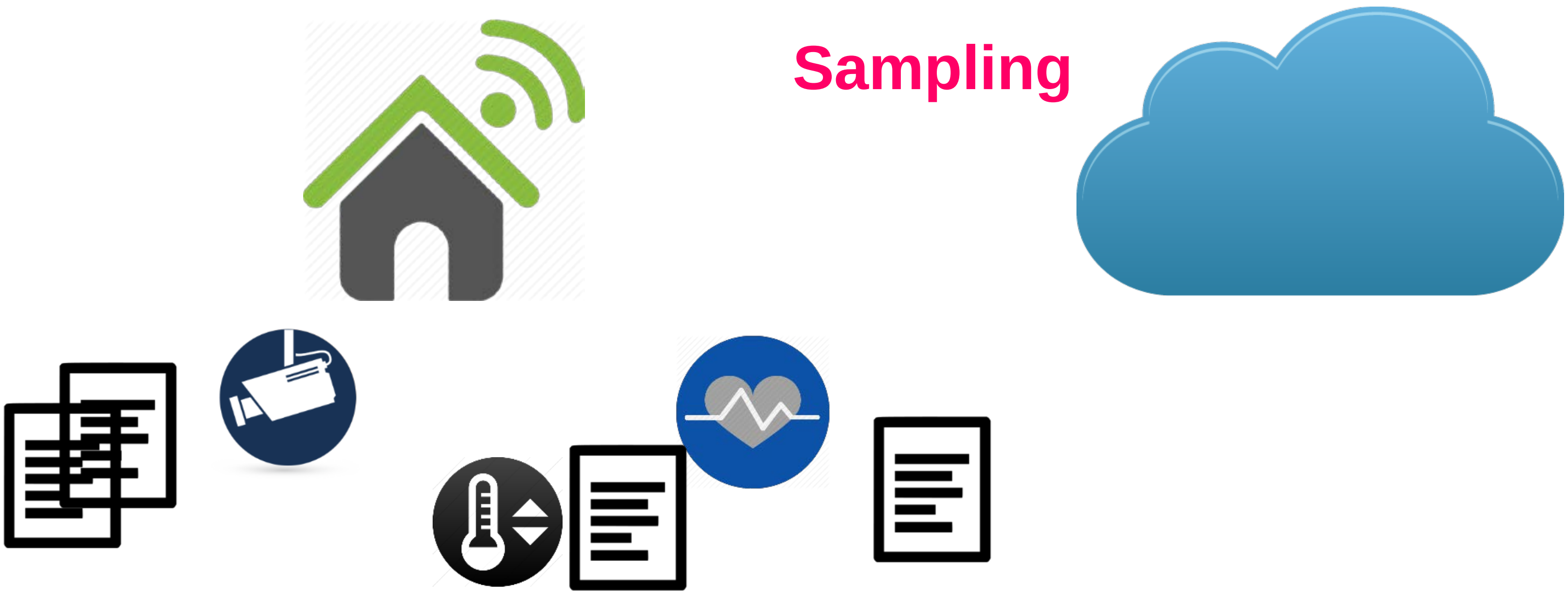
Problem:
Limited computation and storage capacity

IoT Communication Framework



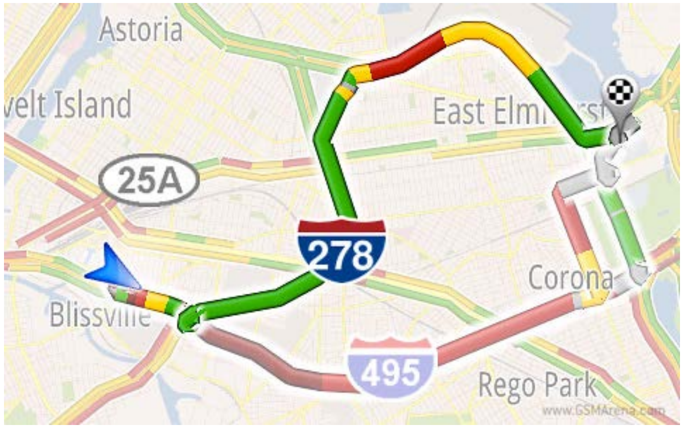
Requirements and challenges

- Data sampling
  - ❖ Due to bandwidth, storage quota limits
  - ❖ Requirement: uniformity



Requirements and challenges

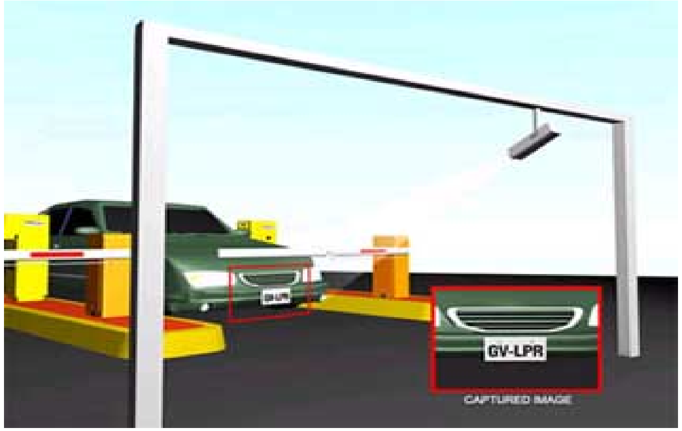
- Partial Data Retrieval
  - ❖ Different granularity requirements
  - ❖ Requirement: partial data retrieval, uniformity



Corse-grained traffic estimation

Requirements and challenges

- Partial Data Retrieval
  - ❖ Different granularity requirements
  - ❖ Requirement: partial data retrieval, uniformity



License plate recognition for toll way billing

Security threat

The Dirty, Little Secret of the Data Center — Data Corruption White Pa

COMPUTERS & ELECTRONICS SCIENCE & SOCIETY

Computer hackers take to the cloud

Online data storage services often — and unknowingly — host malicious software

Wrong Decisions!

Security threat

Security threat

Scope: Authentication and integrity.
Privacy is orthogonal.

Digital signature preliminary

H()

Alice

Expensive
& Slow

Bob

Alice's private key

Alice's public key

Digital signature preliminary

H()

Alice

Bob

Alice's private key

Alice's public key

Digital signature preliminary

H()

Expensive
& Slow

H()

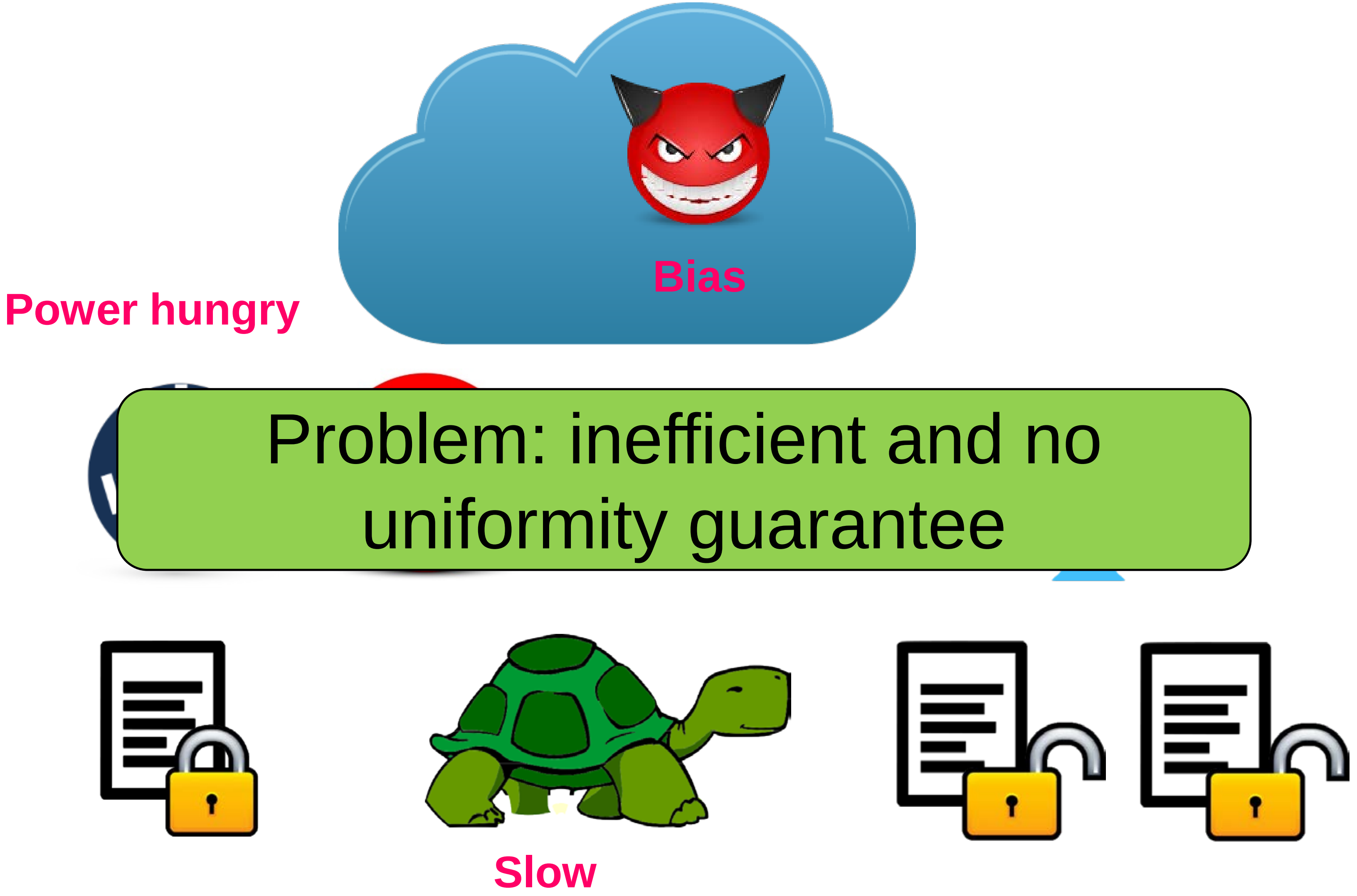
Alice

Alice's private key

Bob

Alice's public key

Digital signature scheme: sign-each



Problem: inefficient and no uniformity guarantee

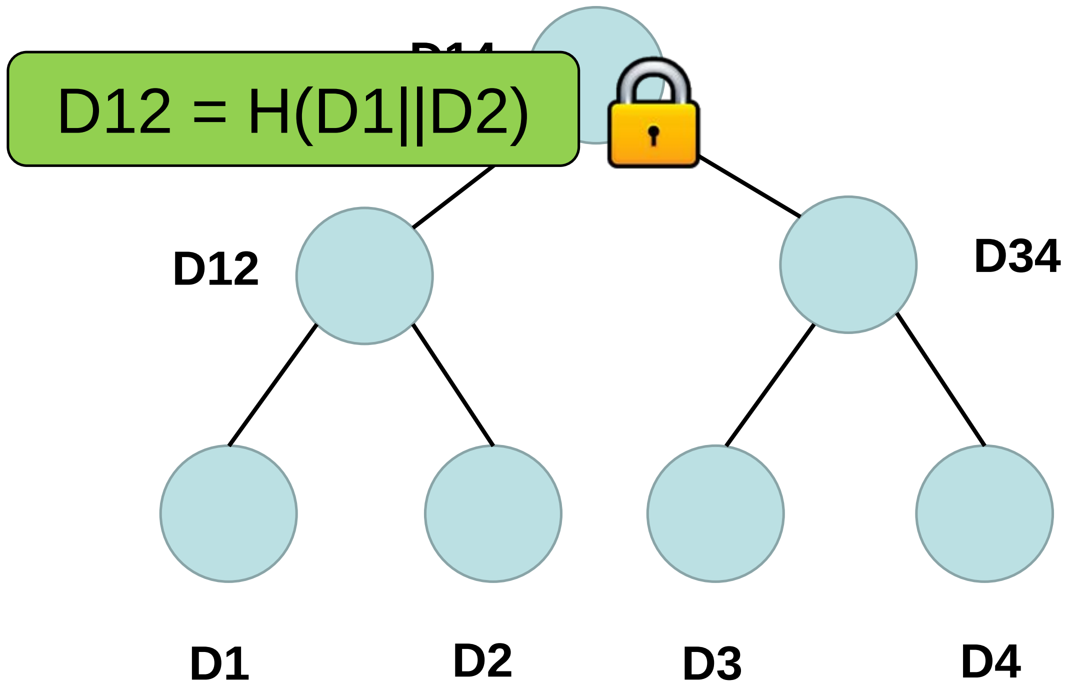
Digital signature scheme: concatenate

Problem: partial data retrieval
not supported

UC SANTA CRUZ

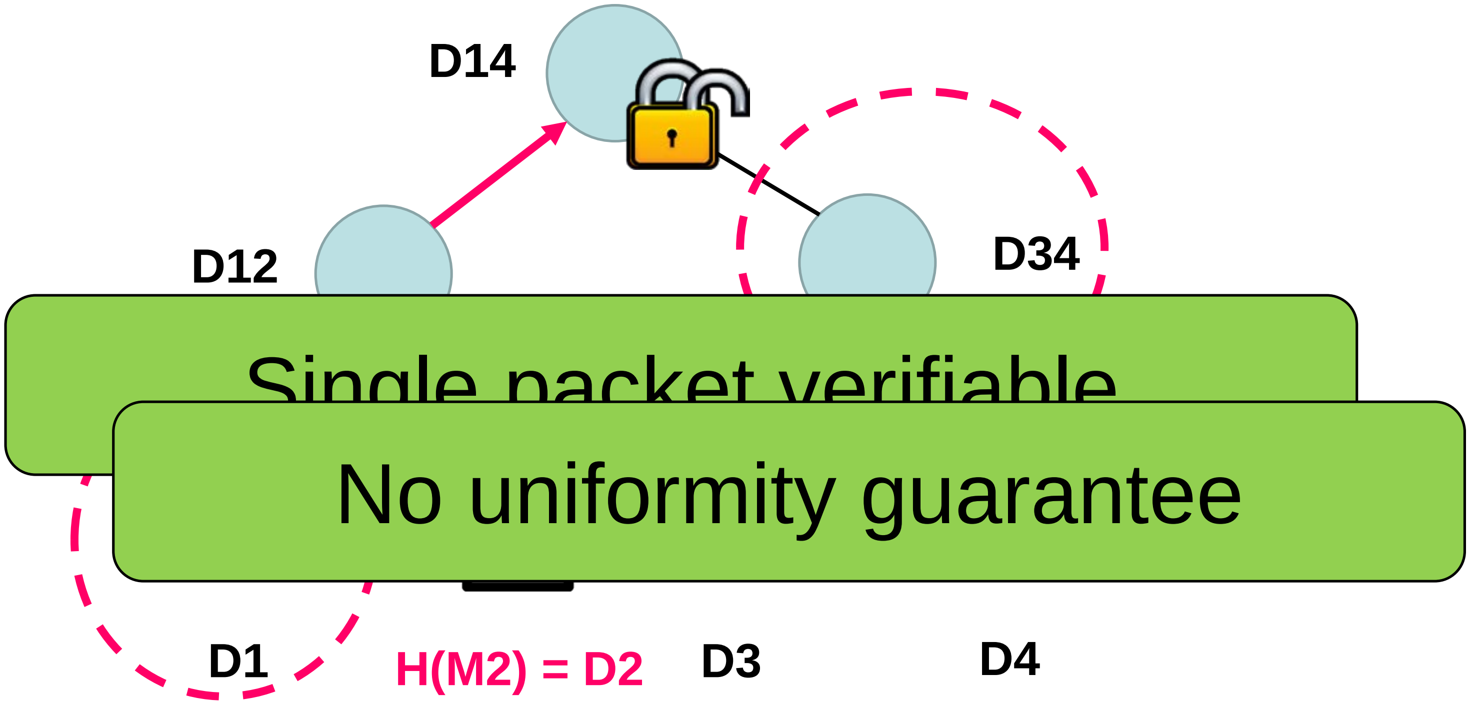
Digital signature scheme: Merkle tree

Signing



Digital signature scheme: Merkle tree

Verifying



Signature scheme comparison

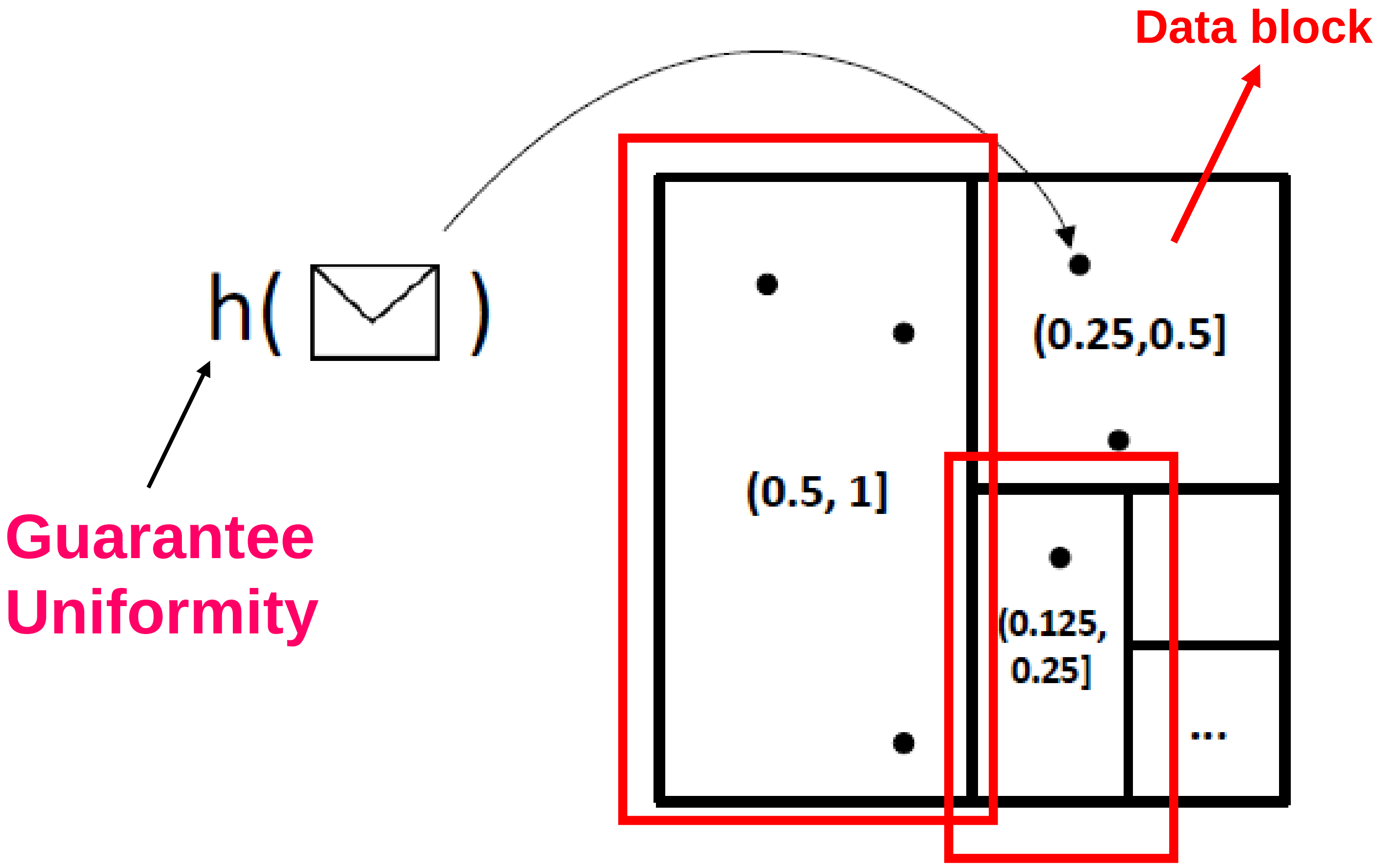
Signature Scheme	Computation Efficiency	Partial Data Retrieval	Uniformity
Sign each	✗	✓	✗
Concatenate	✓	✗	N/A
Merkle tree	✓	✓	✗
GSC	✓	✓	✓

Geometric star chaining

- Intuition: any fraction number can be represented or approximated by a few bits

$$5/8 = (0.101)_2$$

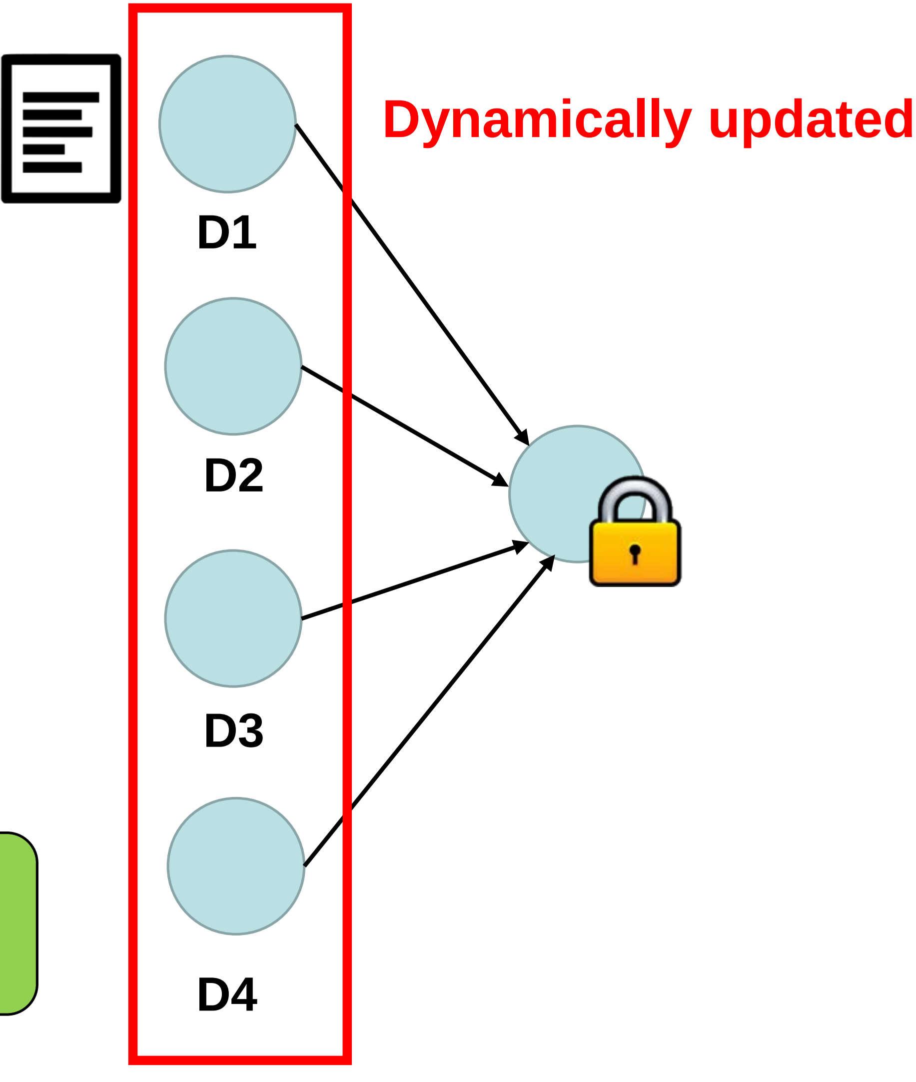
Geometric star chaining



$$5/8 = (0.101)_2$$

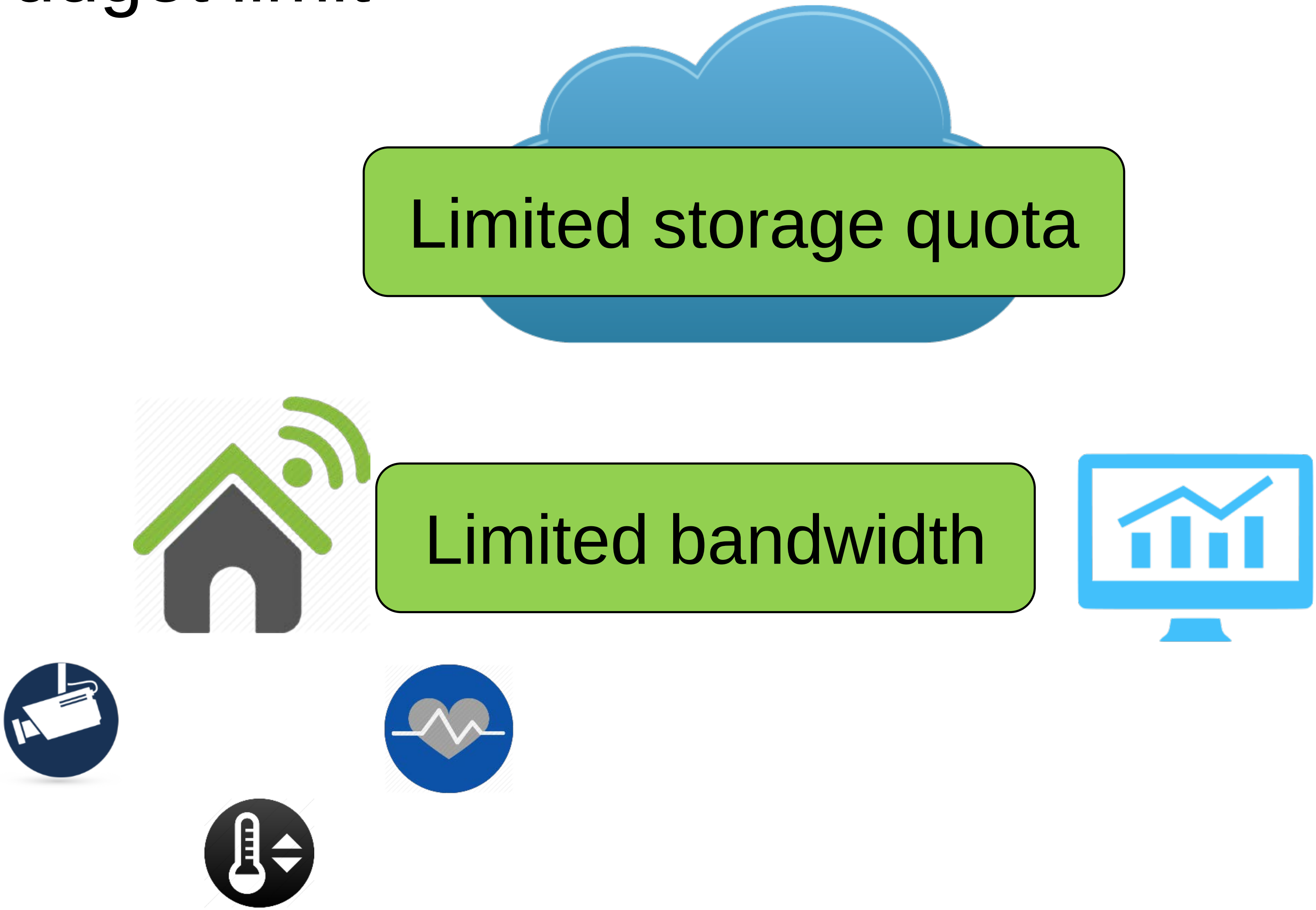
Geometric star chaining

D1 = H(m||D1)



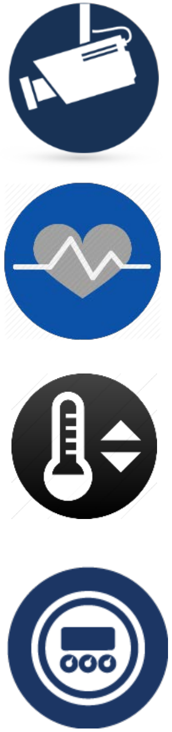
Constant!

Budget limit

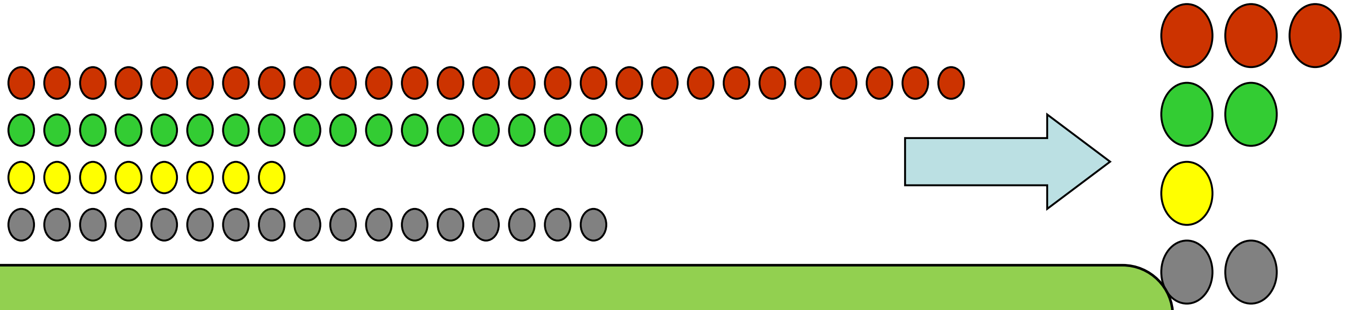


UC SANTA CRUZ

Budged-based distributed stream sampling



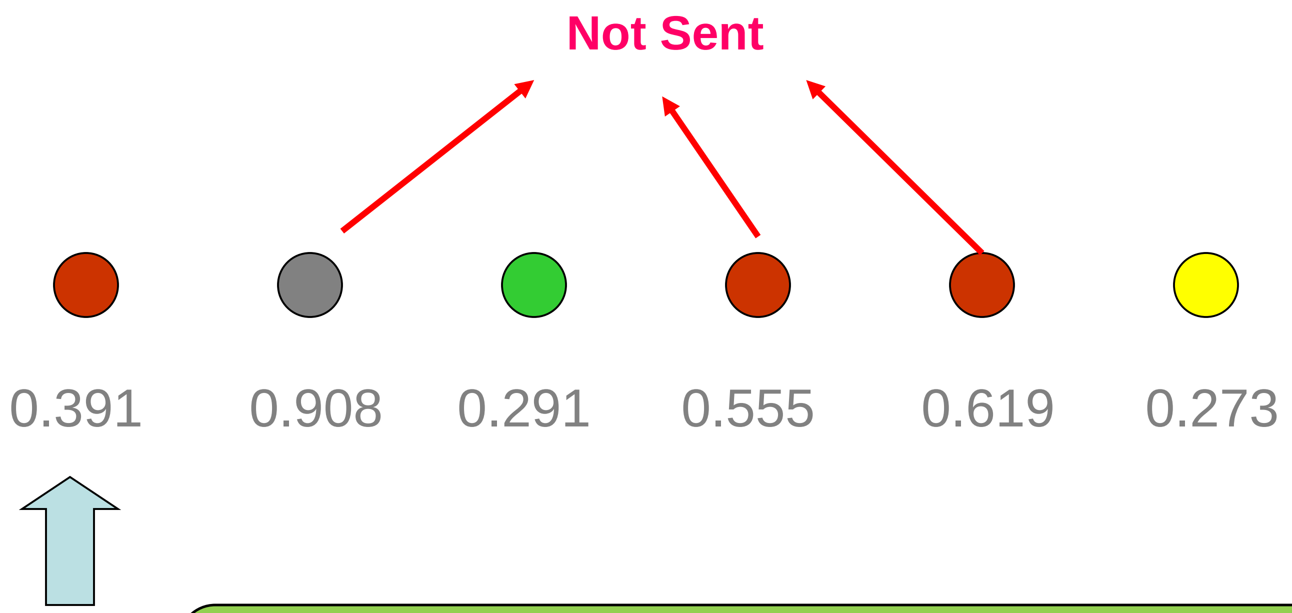
Each epoch: budget = 8



Store and sample -> Significant Space overhead

UC SANTA CRUZ

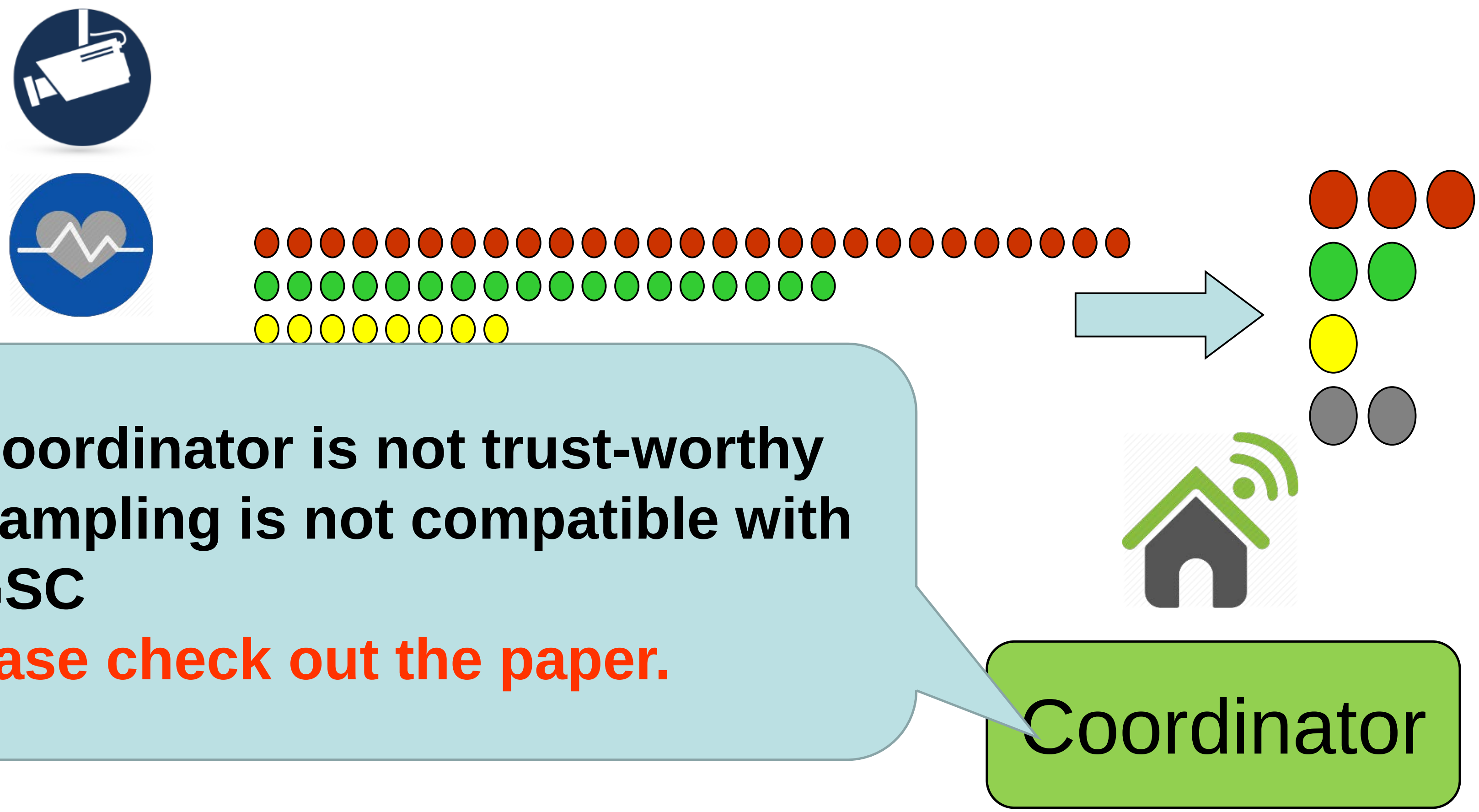
Min-wise sampling



smallest changes

Communication cost is reduced

Budged-based distributed stream sampling



- **Coordinator is not trust-worthy**
- **Sampling is not compatible with GSC**

Please check out the paper.

Sampling Protocol Design

Sensing device

Algorithm 1: SP at sensing device k in round j

foreach *event* e **do**
  $i \leftarrow \underset{x \in \mathbb{N}}{\text{argmin}}\{h(e) \geq 2^{-x-1}\}$;
  $l_i^k \leftarrow l_i^k + 1$;
  if $i \geq j$ **then**
    Forward e to the coordinator;
  else
    Discard e;
  end
end

Coordinator

Algorithm 2: SP at the coordinator in round j

foreach *event* e **do**
  $i \leftarrow \underset{x \in \mathbb{N}}{\text{argmin}}\{h(e) \geq 2^{-x-1}\}$;
  $k \leftarrow e.source$;
  if $i \geq j$ **then**
    $Q_i^k.\text{add}(e)$;
    $l'_i \leftarrow l'_i + 1$;
    $g \leftarrow g + 1$;
    while $g > B$ **do**
      Discard queues $\{\forall \hat{k}, Q_j^{\hat{k}}\}$;
      $g \leftarrow g - l'_j$;
      $j \leftarrow j + 1$;
      Broadcast j to all sensing devices;
    end
  else
    Discard e;
  end
end

Evaluation

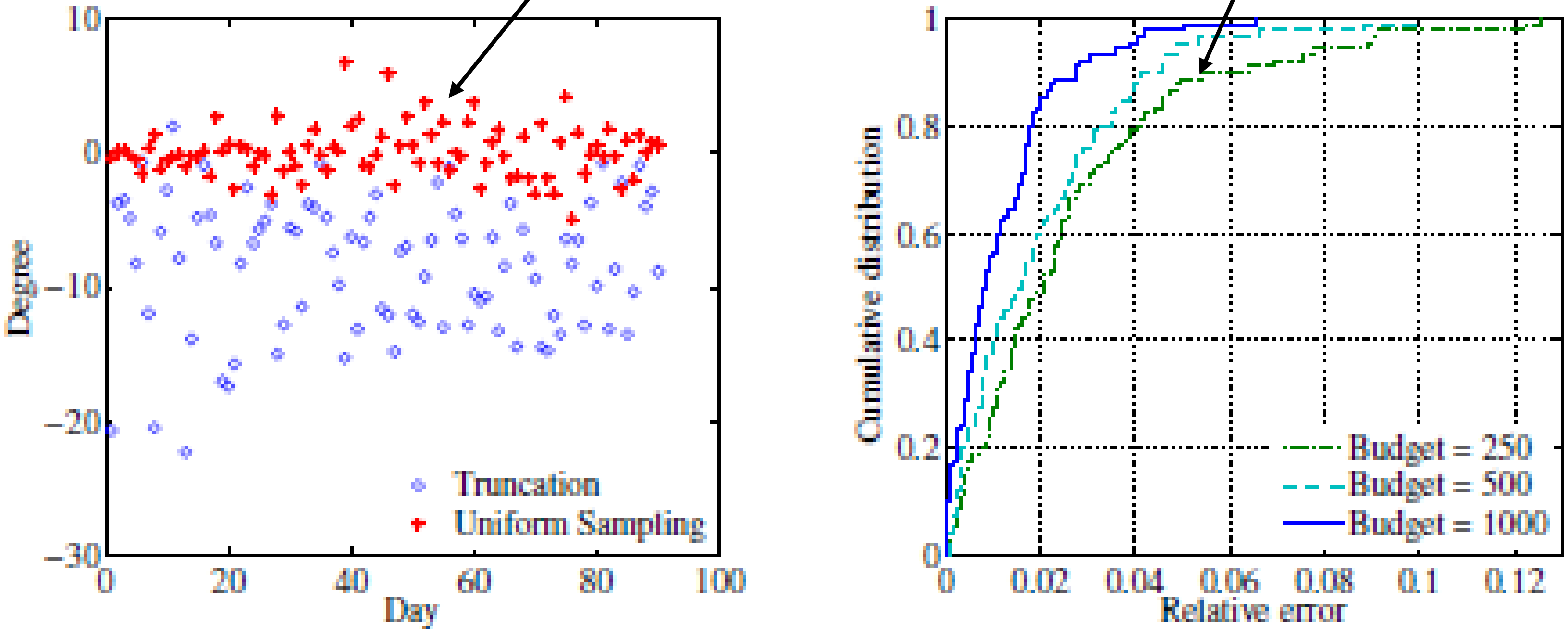
- Simulation and prototype emulation
  - Real dataset : 5 event-based sensing data
- Prototype emulation
  - DSA
  - MD5, SHA1, SHA256

Simulation

Uniformity

Uniformly drawn data reveals substantial information

2% data => 5% error

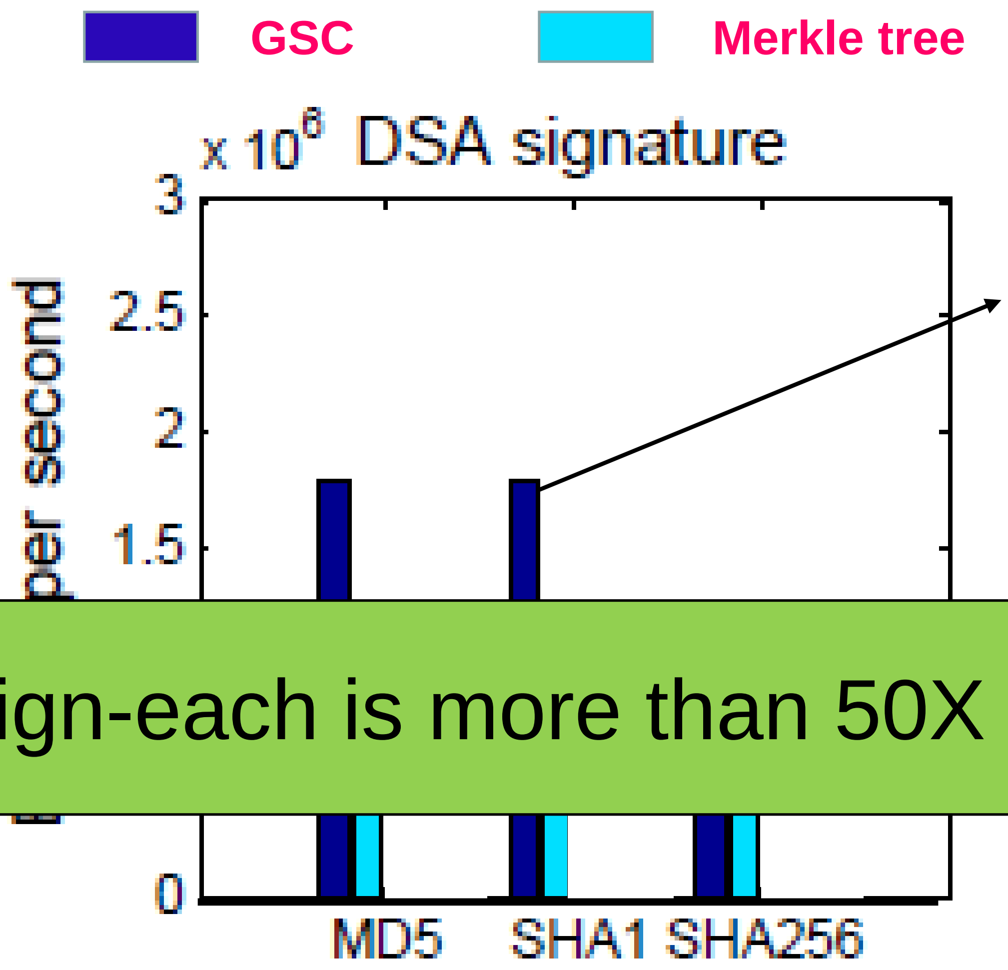


(a) Deviation from the ground truth

(b) Impact of budget limit

UC SANTA CRUZ

Prototype emulation

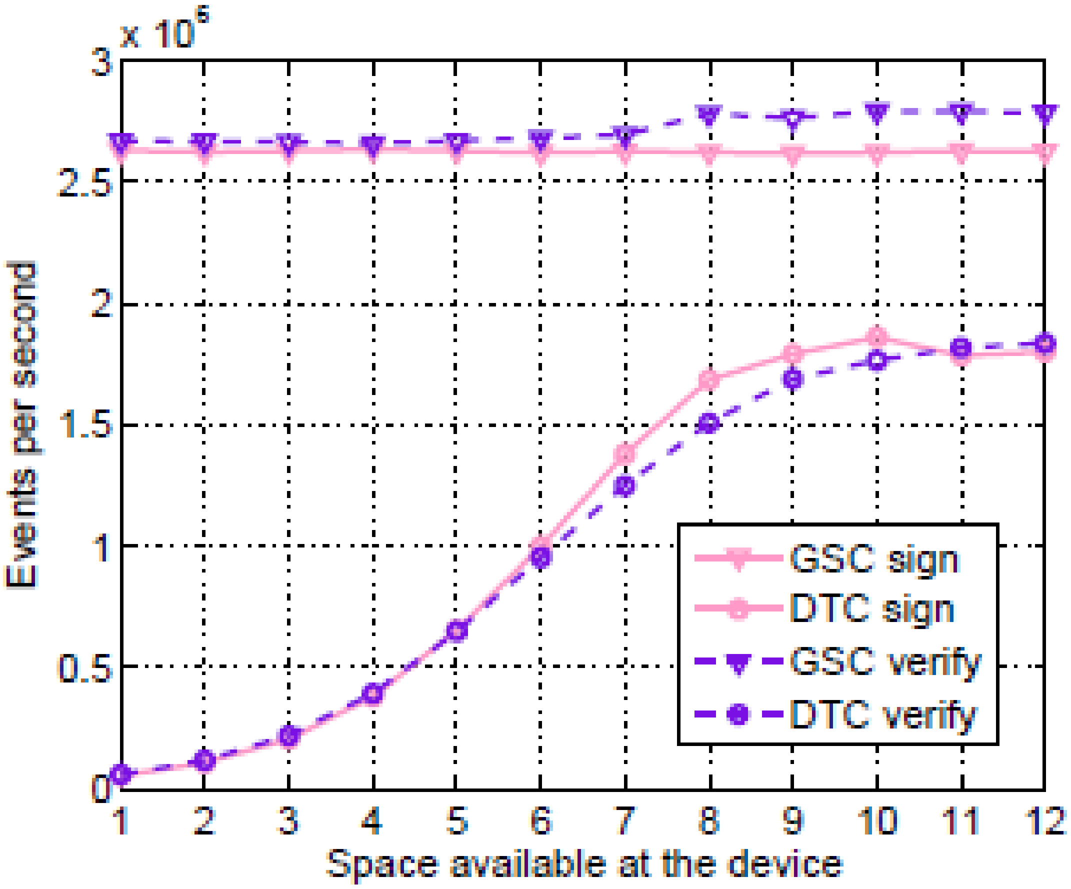


GSC is faster under all cases

Sign-each is more than 50X slower

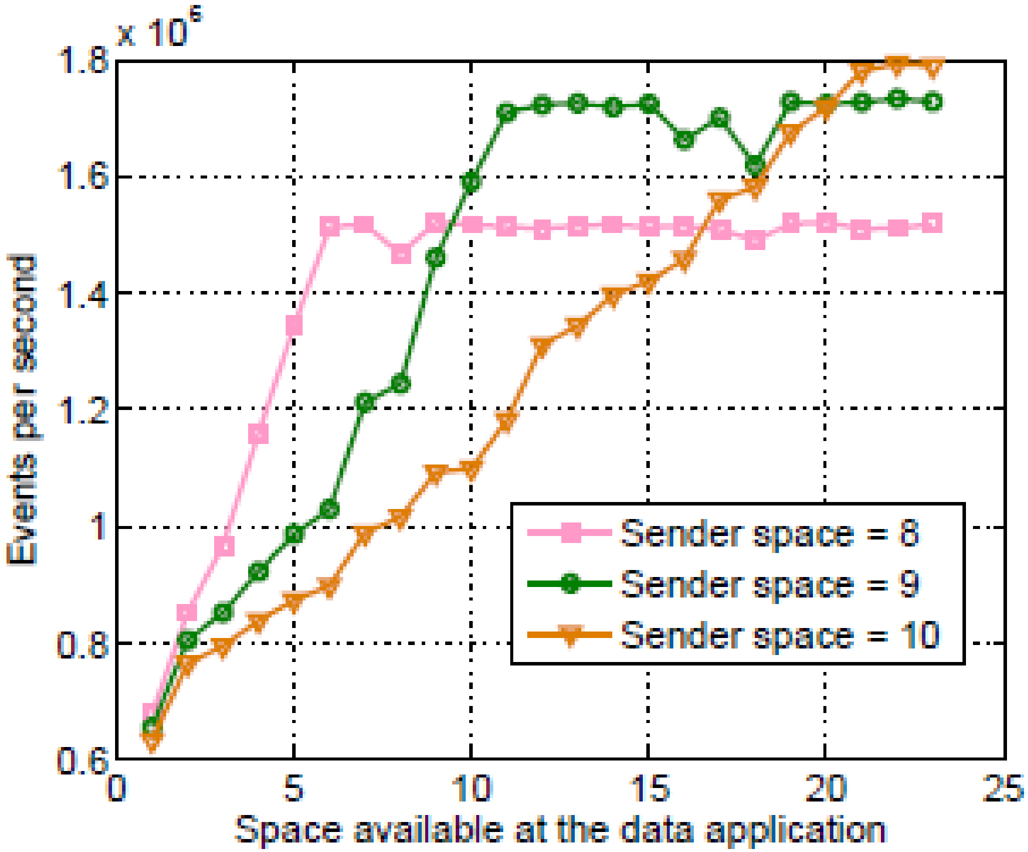
UC SANTA CRUZ

Prototype emulation



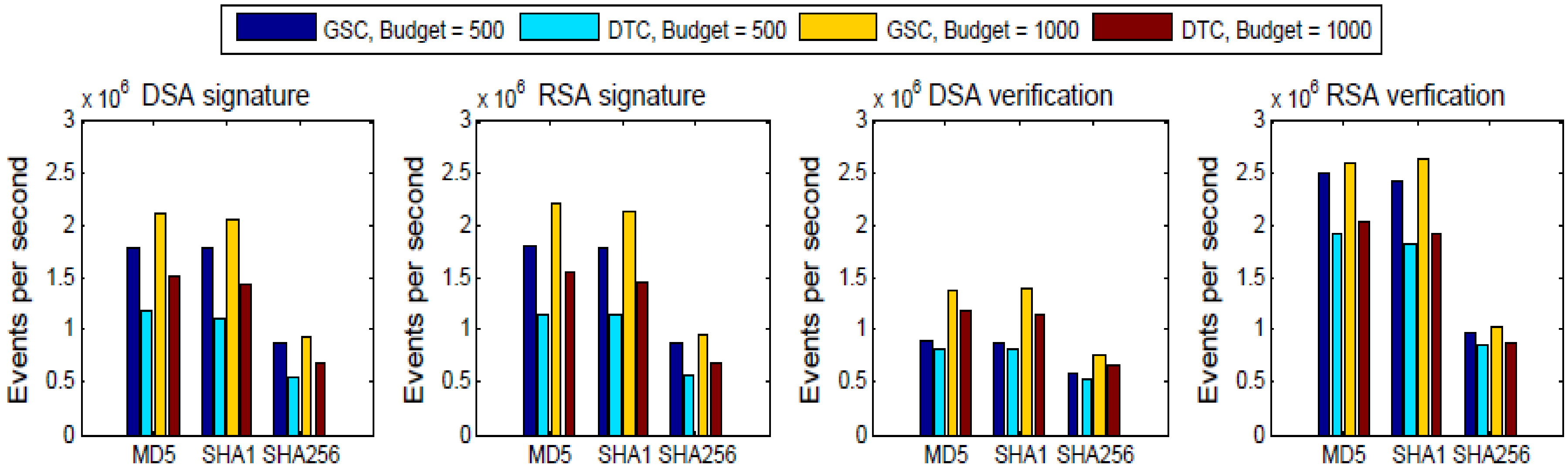
UC SANTA CRUZ

Prototype emulation



UC SANTA CRUZ

Prototype emulation



UC SANTA CRUZ

Conclusion

- Requirement of IoT communication
  - ❖ Computation efficiency
  - ❖ Uniformity
  - ❖ Partial data retrieval

- GSC is able to satisfy all these three requirements simultaneously.

Thank you !