The 37th Multi-Service Networks workshop (MSN 2025)

Pontus: A Memory-Efficient and High-Accuracy Approach for Persistence-Based Item Lookup in High-Velocity Data Streams

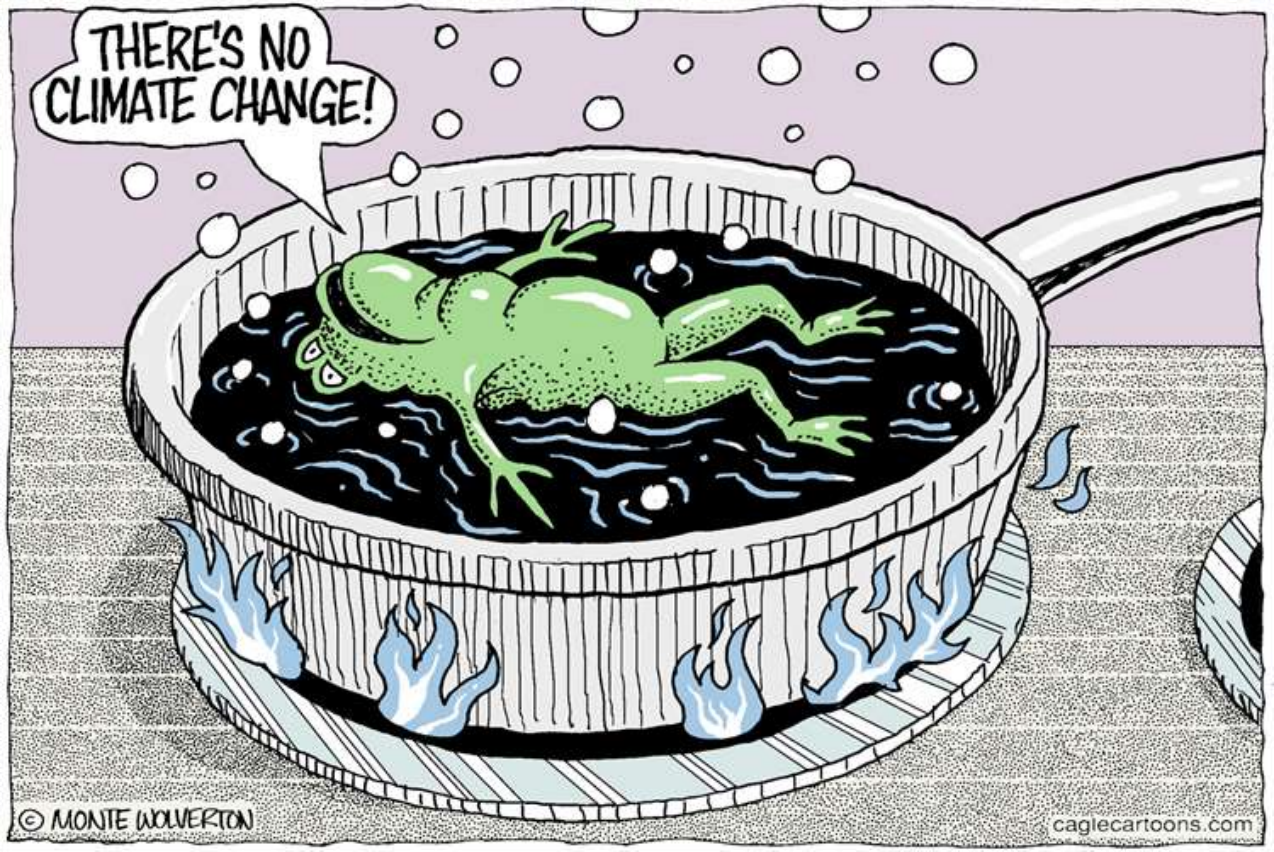
Weihe Li[1], Zukai Li[1], Beyza Bütün[2], Alec F. Diallo[1], Marco Fiore[2], Paul Patras[1]

[1]University of Edinburgh, Edinburgh, United Kingdom

[2]IMDEA Networks Institute, Madrid, Spain

THE UNIVERSITY *of* EDINBURGH

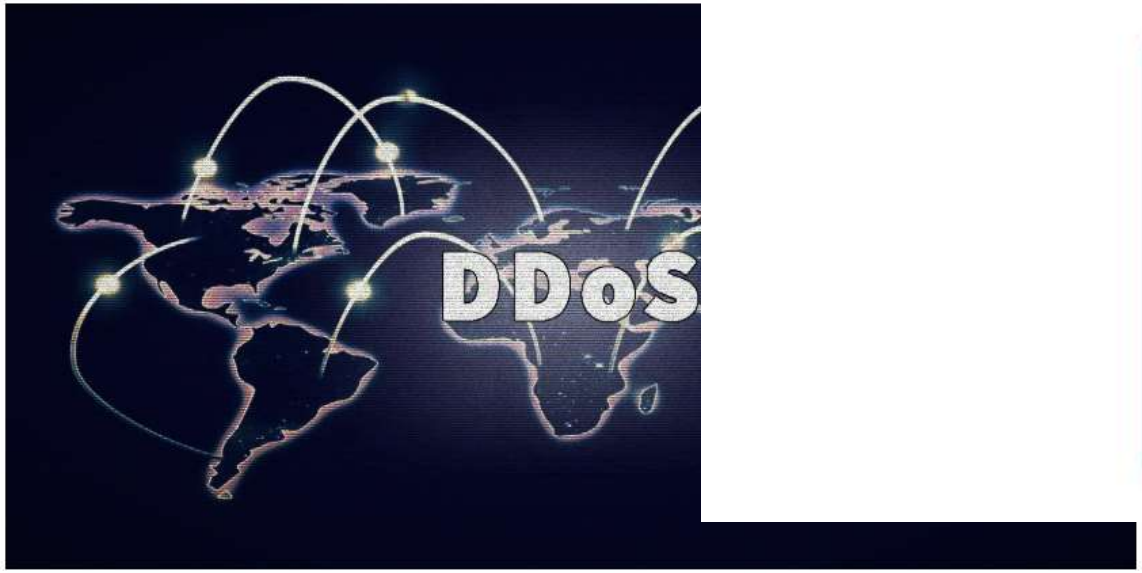
Did you know? Many of the most dangerous cyberattacks and financial frauds are not “one-off strikes,” but rather “***boiling frog***” scenarios.



Home > News > Security > Italian CERT: Hacktivists hit govt sites in 'Slow HTTP' DDoS attacks

Italian CERT: Hacktivists hit govt sites in

By Bill Toulas



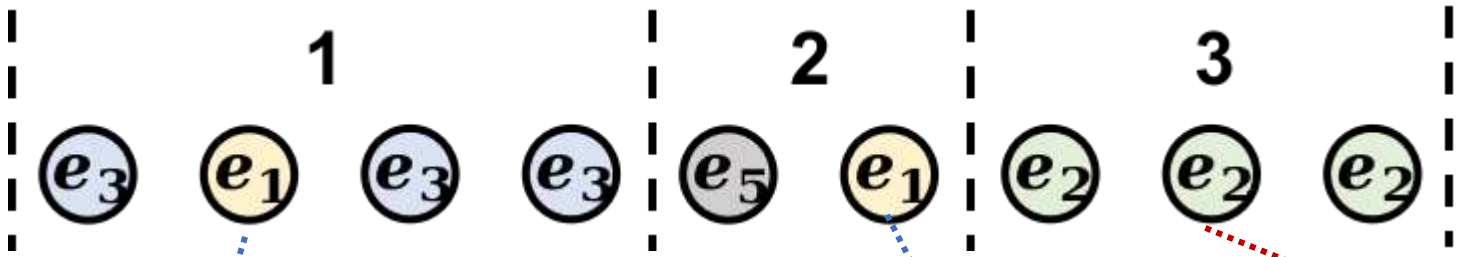
Attacks And Vulnerabilities | Critical Infrastructure | Industrial Cyber Attacks | Malware, Phishing & Ransomware | News | Reports | Threat Landscape

RSA 2025: Trellix CyberThreat reveals 136% surge in APT attacks on US in Q1 2025 as threat landscape intensifies

APRIL 30, 2025

Neglecting persistent item detection is like leaving a door unlocked for attackers -- it creates long-term vulnerabilities in the system.

Persistence



Appeared, **persistence = 1**

Appeared, **persistence = 2**

Appeared
persistence = 1

Item	Persistence
e_1	2
e_2	1
e_3	1
e_5	1

Challenges

- **Fast processing speed**

e.g. 10 Gb/s data stream: each item every 67 ns

- **Limited fast memory**

L1 Cache: around 64KB[1]

Infeasible to store information for all items

[1] Li, W. and Patras, P. Tight-sketch: A high-performance sketch for heavy item-oriented data stream mining with limited memory size. ACM CIKM 2023.

Sketches

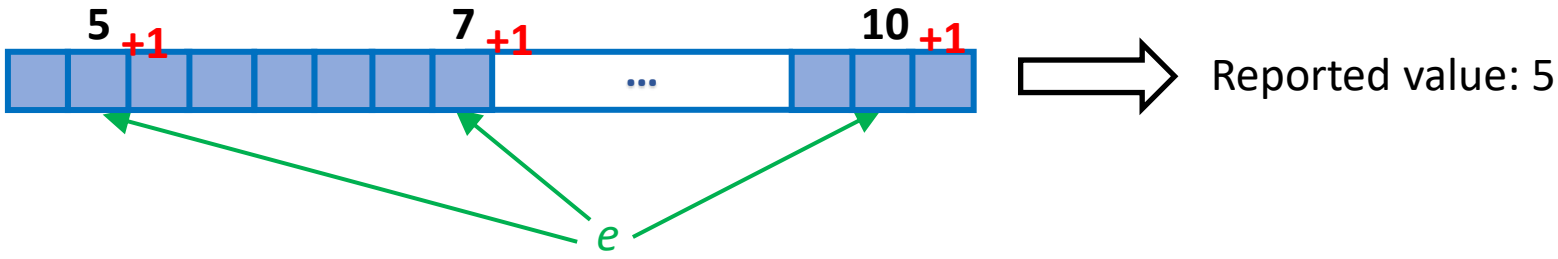
- **Sketches**: **compact data structure by hashing**
  - Idea: hash data into limited space

Sketches

- **Sketches**: **compact data structure by hashing**
  - Idea: hash data into limited space

Insertion: when a new item *e* comes

Query: query for the frequency of the item *e*

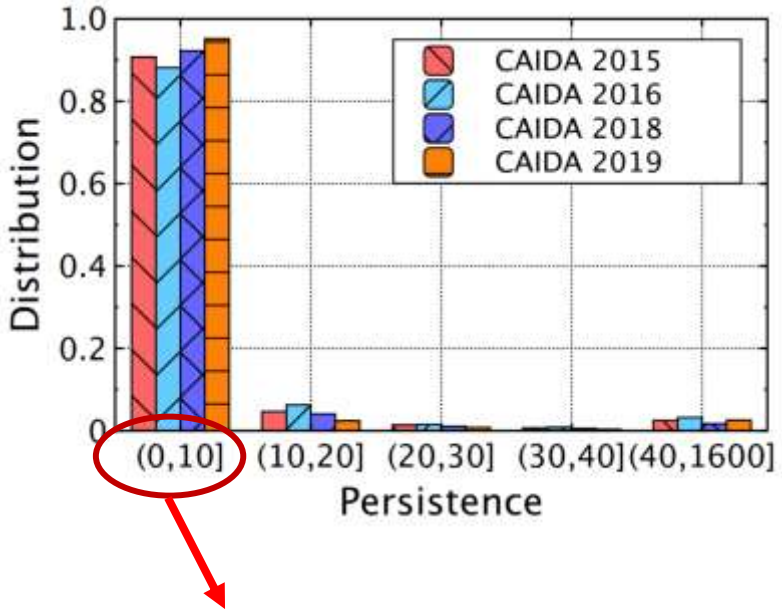


[2] Cormode, Graham, and Shan Muthukrishnan. "An improved data stream summary: the count-min sketch and its applications." *Journal of Algorithms* 55.1 (2005): 58-75.

Limitation of Existing Sketch Methods

- **Low detection accuracy under limited memory budgets**

  Persistent flows being evicted from the bucket by non-persistent ones due to the **highly skewed** traffic distribution.



Dropped Significantly!

Memory Size (KB)	16	32
F1 Score	0.27	0.3
64	**128**	**256**
0.32	0.94	0.99

On-Off Sketch[3]

CAIDA 2018 (1500 windows, threshold: 0.4)

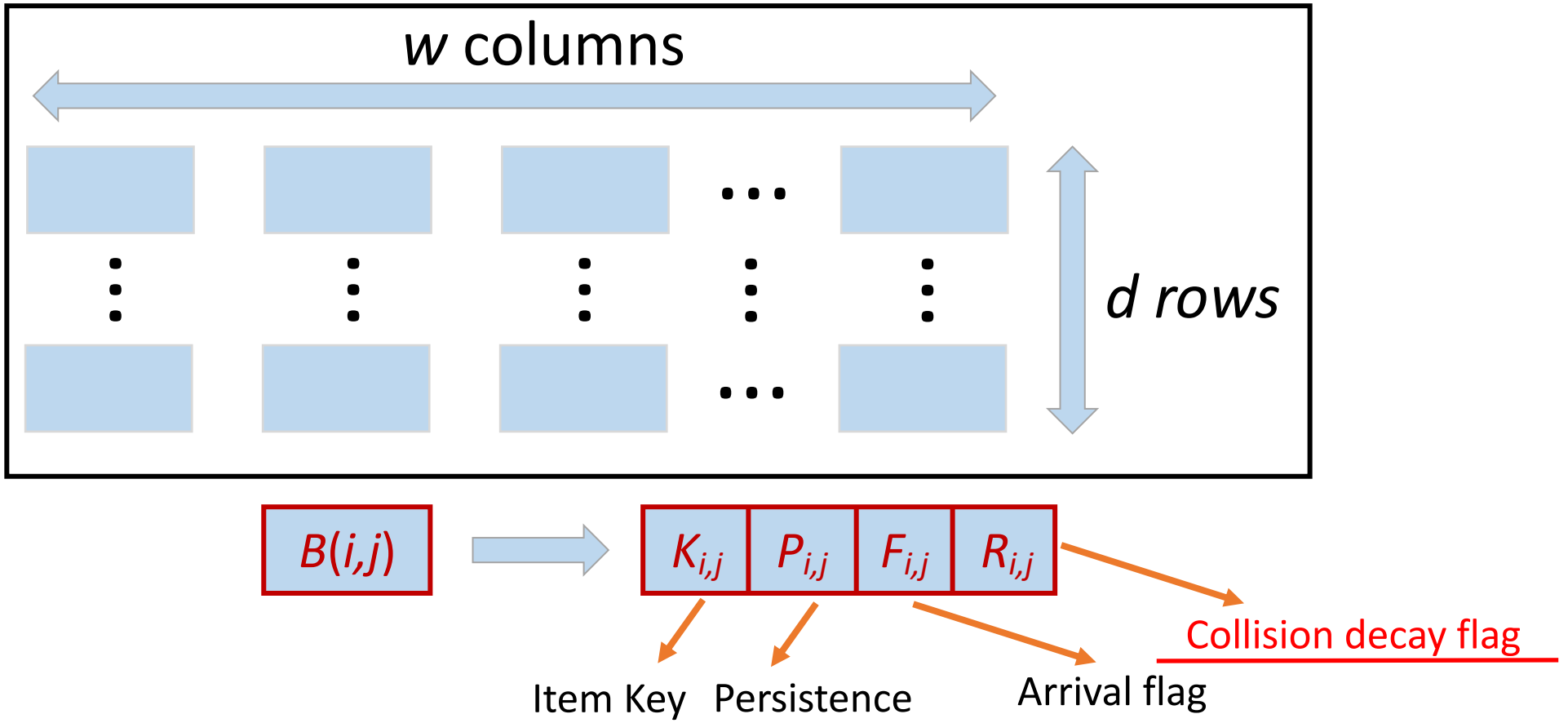
[3] Zhang, Yinda, et al. "On-off sketch: A fast and accurate sketch on persistence." Proceedings of the VLDB Endowment 14.2 (2020): 128-140.

Our Contributions

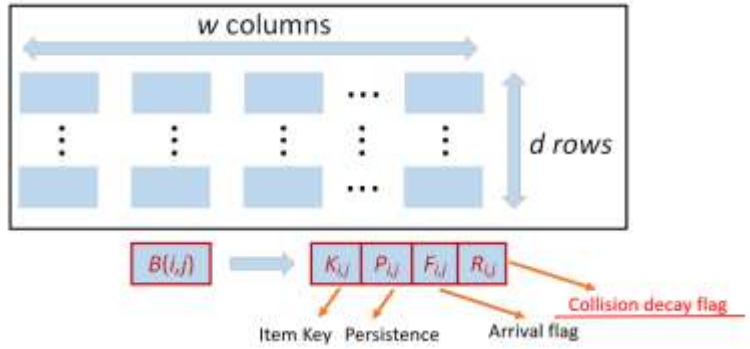
- **Pontus**
  - A novel method for persistent item lookup
  - **High accuracy, high memory-efficiency** and **fast processing speed**
  - Deployable on the practical hardware, Tofino programmable switch

[4] Li, W., Li, Z., Bütün, B., Diallo, A.F., Fiore, M. and Patras, P. "Pontus: A Memory-Efficient and High-Accuracy Approach for Persistence-Based Item Lookup in High-Velocity Data Streams." Proceedings of the ACM on Web Conference 2025.

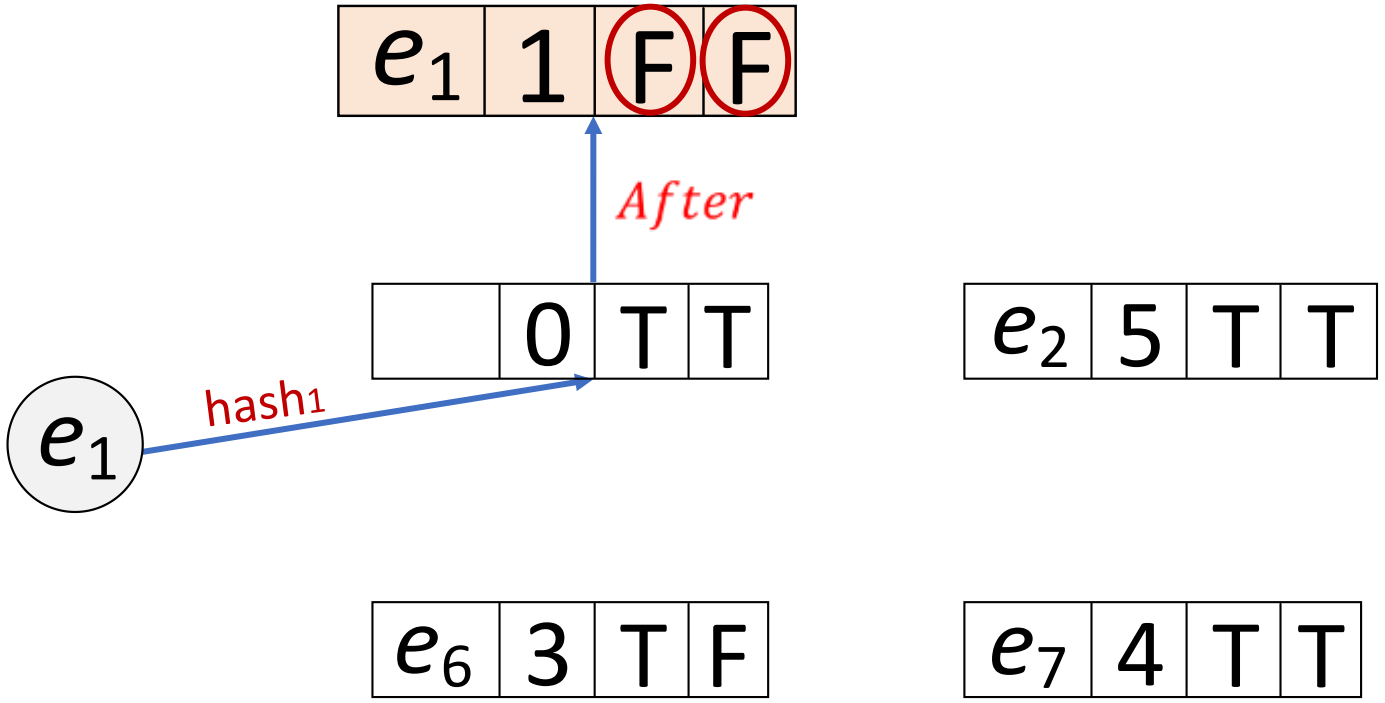
Data Structure



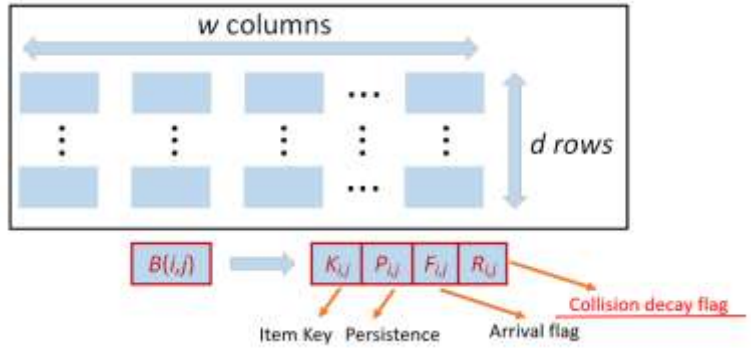
Update



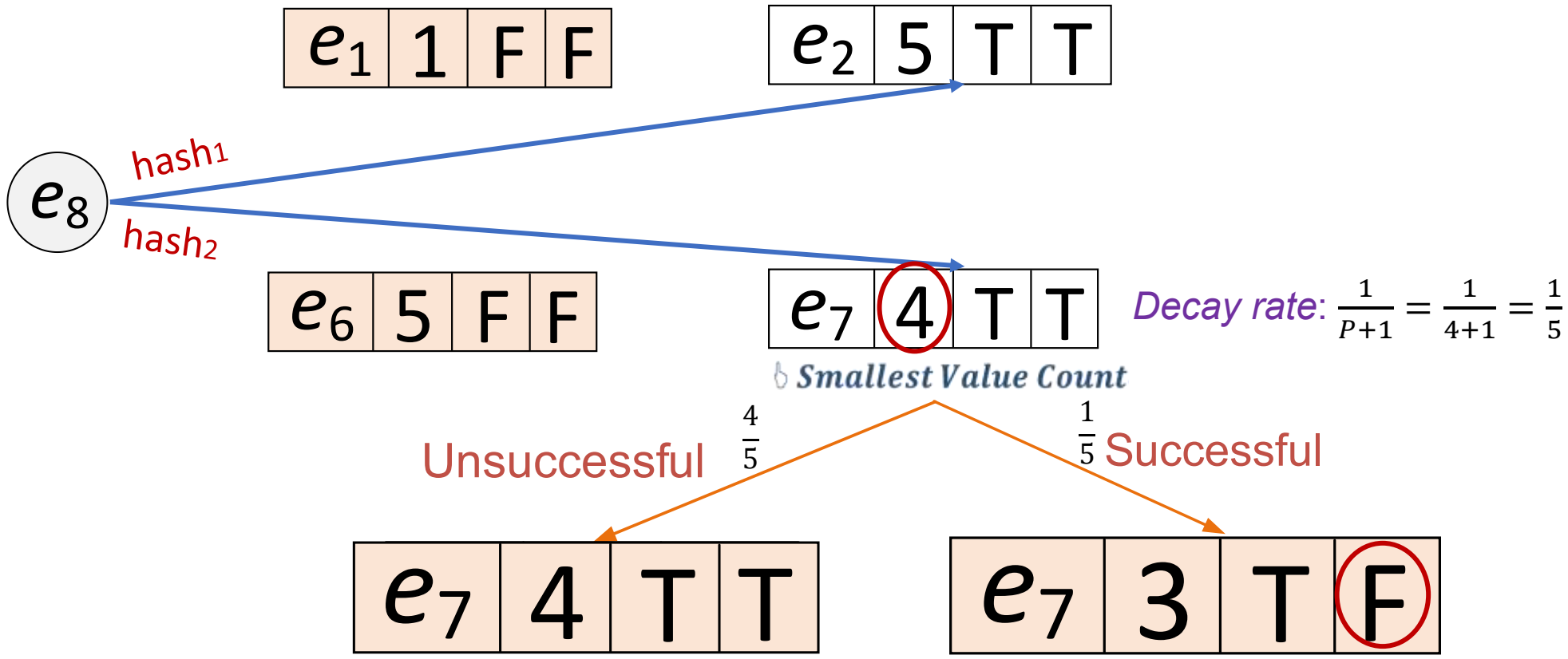
Case 1:



Update

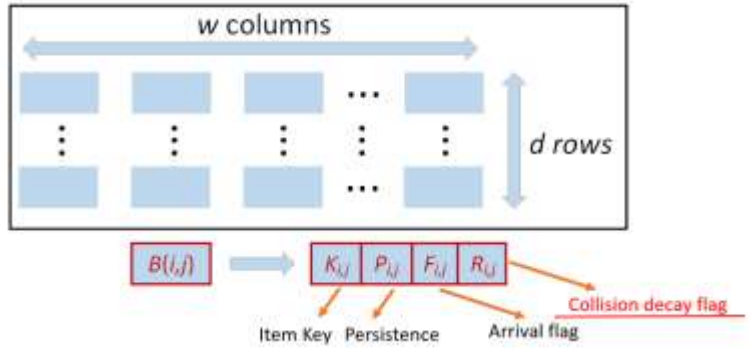


Case 2:

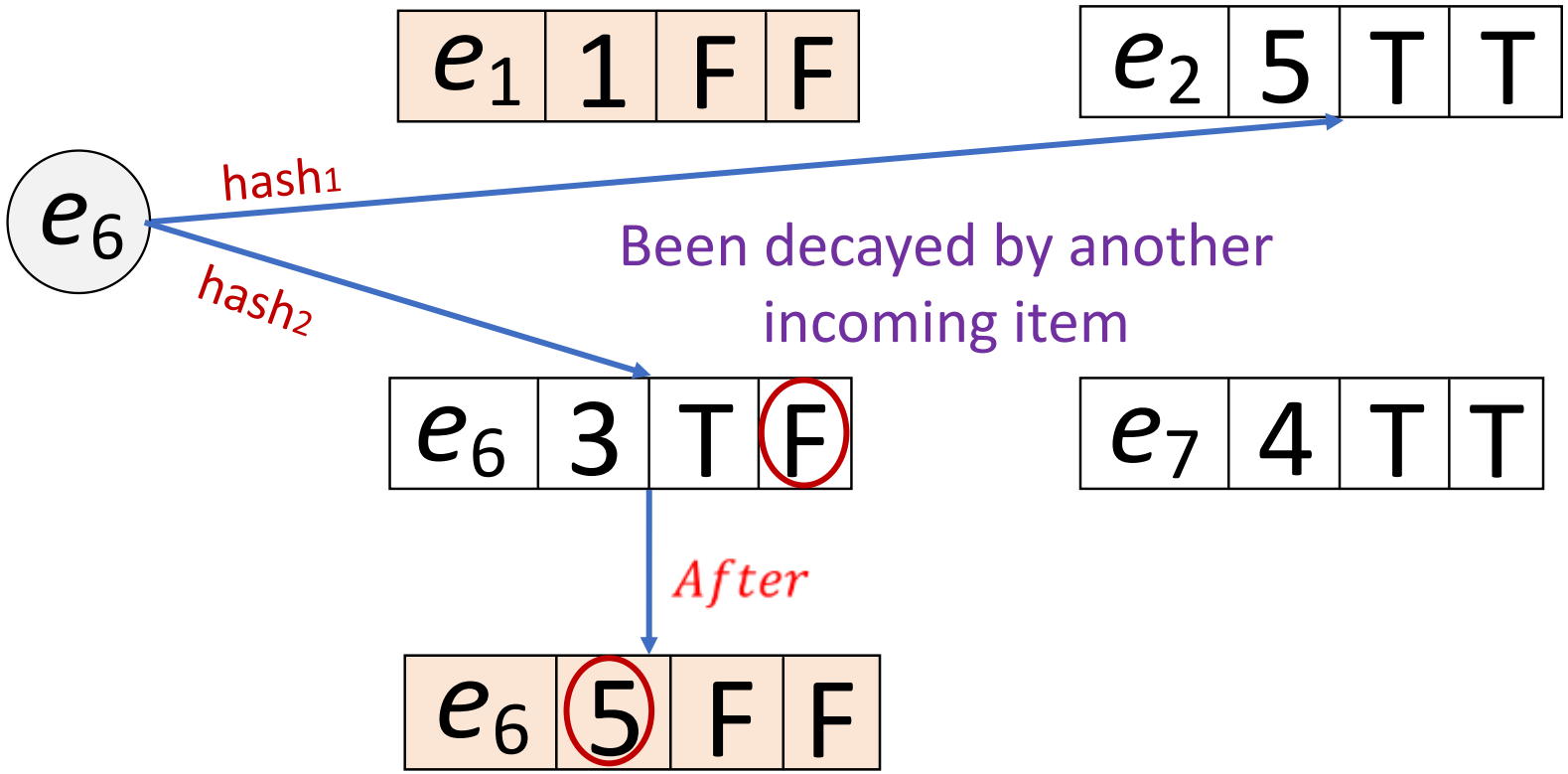


The incoming item can replace the tracked item only if its counter has decayed to **zero**.

Update

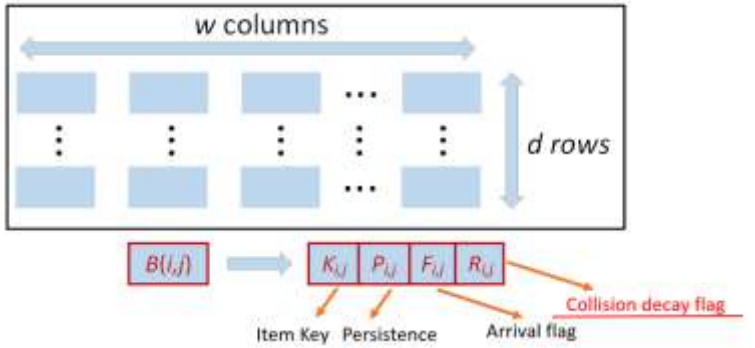


Case 3:



Increment by **2** instead of 1 for accurate persistence tracking

Query



Only a scan of all buckets is required to determine which bucket contains a value higher than the predefined threshold.

Evaluation

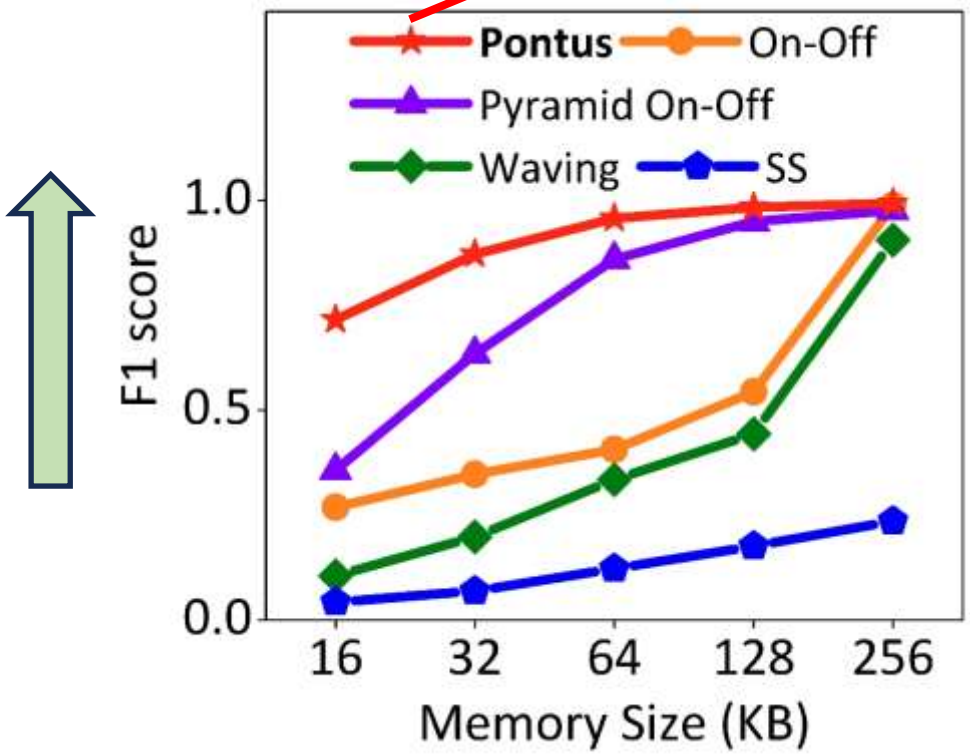
- **Software** -- CPU Platform (Intel(R) Core(TM) i5-1135G7 @ 2.40GHz processor, C++)
- **Hardware** -- Tofino Switch (P4)
- **Traces** -- CAIDA 2019[5]

[5] CAIDA. Anonymized Internet Traces. https://catalog.caida.org/dataset.

Evaluation – **Accuracy & Speed (CPU)**

Highest F1 score!

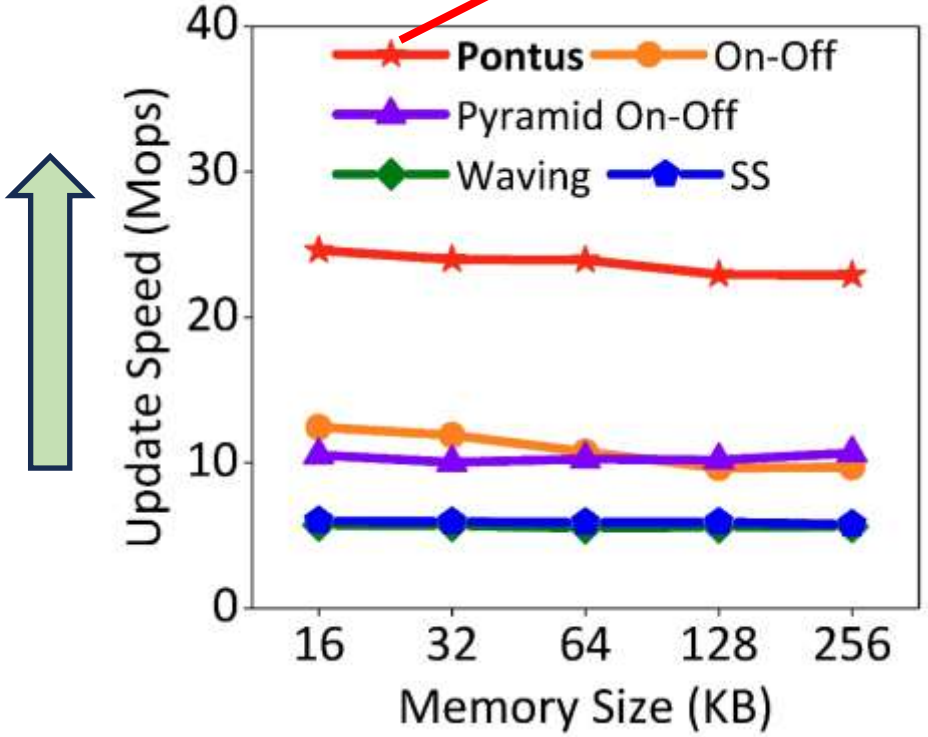
19.7% higher than Pyramid On-Off



CAIDA 2019

Highest Speed!

117.3% higher than On-Off



CAIDA 2019

Evaluation – Resource Usage (Tofino)

Resource	Usage
Hash Bit	5.7%
Gateways	16.7%
VLIW Instruction	7.3%
Match Crossbars	4.6%
Logical Table ID	21.9%
SRAM	4.3%
Total Average	**8.2%**

Limited Overhead

Summary

- **Pontus**
  - Versatile for multiple persistence-based tasks
  - High accuracy with small and static memory
  - Fast processing speed

- **Code**: https://github.com/Mobile-Intelligence-Lab/Pontus

This research was supported by the SNS JU and the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101139270 (ORIGAMI). Beyza Bütün is a Comunidad de Madrid predoctoral fellow (PIPF-2022/COM-24867). Weihe Li was partially supported by Cisco through the Cisco University Research Program Fund (Grant no. 2019-197006).