SketchFeature: High-Quality Per-Flow Feature Extractor Towards Security-Aware Data Plane

Sian Kim[1], Seyed Mohammad Mehdi Mirnajafizadeh[2], Bara Kim[3]
Rhongho Jang[2]✤, DaeHun Nyang[1]✤

1 Ewha Womans University
2 Wayne State University
3 Korea University

✤ Corresponding authors.

EWHA,
THE FUTURE
WE CREATE

- Background and Challenges
- Contributions of SketchFeature
- Analysis and Evaluations
- Conclusion

AI-enhanced In-network Defense

- **Importance of Data for In-network Defense**
  - Security systems rely on high-quality network data to detect threats.

    → Incomplete, low-resolution features lead to misclassification and blind spots.

    → Real-time data visibility is critical for proactive threat mitigation.
  - Feature extraction in the data plane remains a significant challenge.

    ※ Hardware constraints (e.g., limited computational resources)
- **Feature Extraction**
  - 1st-order features: Flow-agnostic, per-packet (e.g., TTL, packet size).
  - 2nd-order features: Flow-aware, per-flow statistical features (e.g., mean, variance).
  - 3rd-order features: Per-flow distribution features (e.g., inter-packet delay, packet size distribution).

    → Higher-order features provide superior attack detection capacities

    ※ High computational complexity

Symptom Detectors

- **Symptom Detector**
  - Selectively escalates only a subset of flows, identified as suspicious, to the control plane.
  - Instead of performing full per-flow feature extraction, it filters flows based on predefined symptoms.
- **FlowLens**
  - Selects only top-K bins

    → Attackers evade detection by shifting packet size patterns
- **NetWarden**
  - Uses low-resolution quantization

    → High false positives and control plane flooding attacks

Challenges in Feature Extraction

- **Constraints of Data Plane**
  - Only simple arithmetic operations (e.g., addition, subtraction) are supported.
  - Limited memory resources → Difficult to store high-dimensional per-flow data.
- **Challenges**
  - Eliminate Selective Measurement → Extract features for all flows rather than a subset.
  - Enable Full-Range, High-Resolution Features → No top-K bin or coarse quantization limitations.
  - Ensure Line-Rate Processing in the Data Plane → Efficient memory use and low computational overhead.

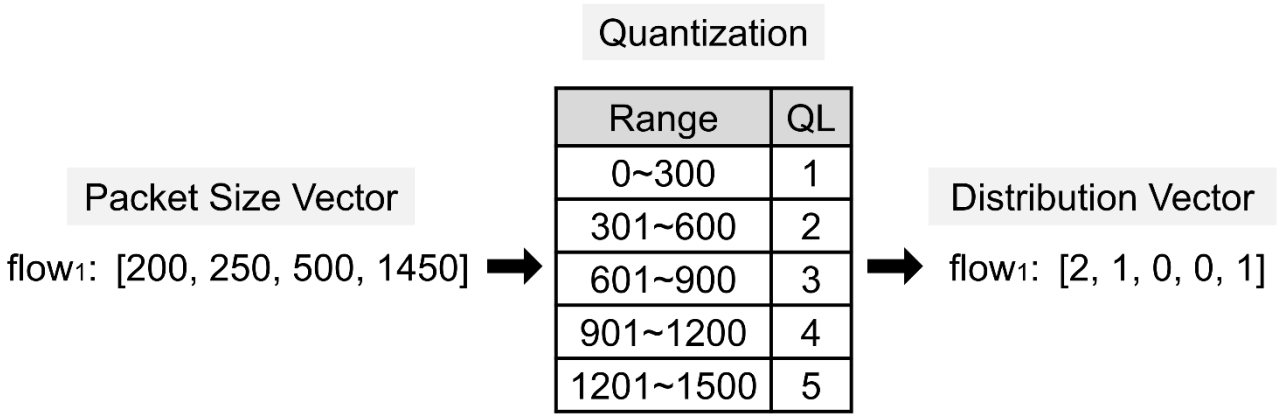
Main Contributions of SketchFeature

- **Ensuring HAF (High-Resolution, All-Flow, Full-Range) Feature Measurement**
  - Eliminates Top-K Bias → Captures full feature distribution rather than selected bins.
  - Detects a wider range of attacks with better granularity and accuracy.
- **Addressing Technical Challenges for Sketch Evolution**
  - Sketch Virtualization to extend sketch capacity.
  - Introduces Membership Test to ensure that only real encoded features are queried.
- **Theoretical & Experimental Validation**
  - Proves error bounds & probability guarantees for 3rd-order feature extraction.
  - Evaluates per-flow feature accuracy in multiple attack scenarios.
- **Advancing Security Applications**
  - Evaluates SketchFeature's security effectiveness through various use cases.
  - Successfully deployed in a commercial switch (Tofino fabric).

Building Blocks of SketchFeature

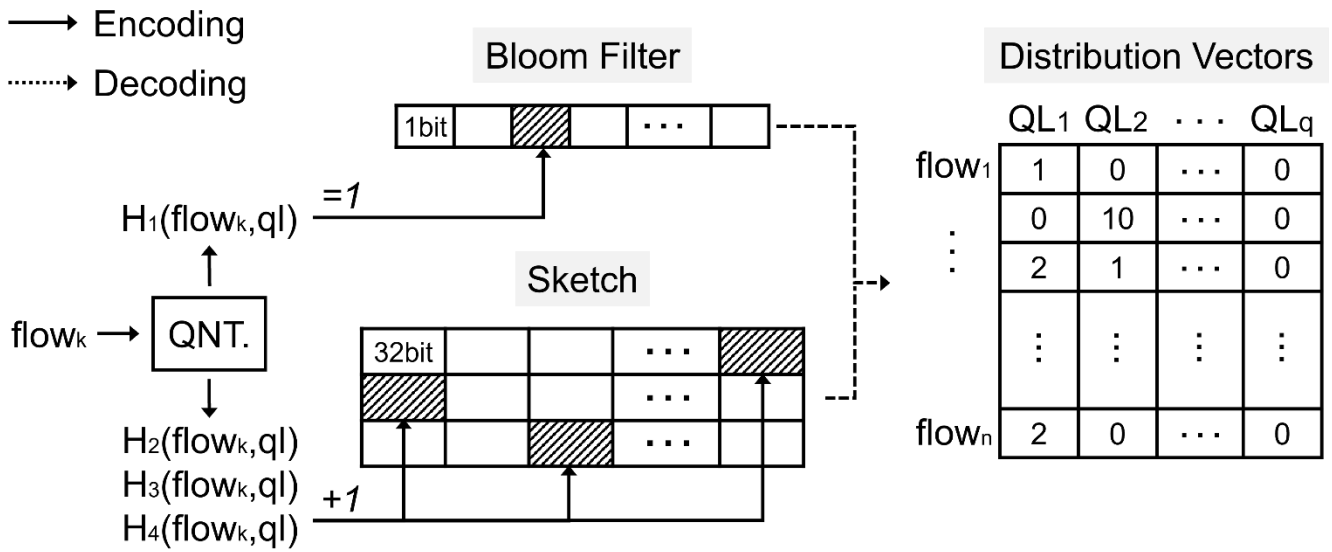
- **Complexity reduction of per-flow distribution**
  - 3rd-order features involve continuous values (e.g., inter-packet delay, packet size distributions).
  - ASIC-based switches cannot store every possible value due to memory constraints.

    → Divides feature values into discrete bins.



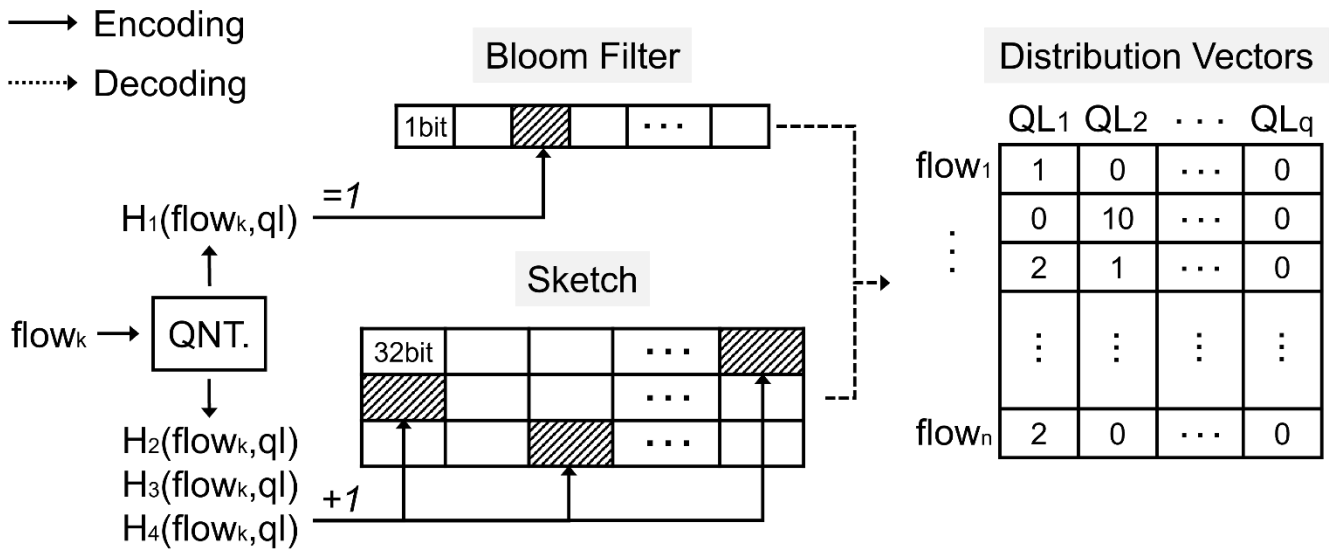
Building Blocks of SketchFeature

- **Data Structure**
  - Sketch (Count-Min Sketch with Sketch Virtualization)
  - Bloom Filter (Solving Phantom Decoding issue)



Building Blocks of SketchFeature

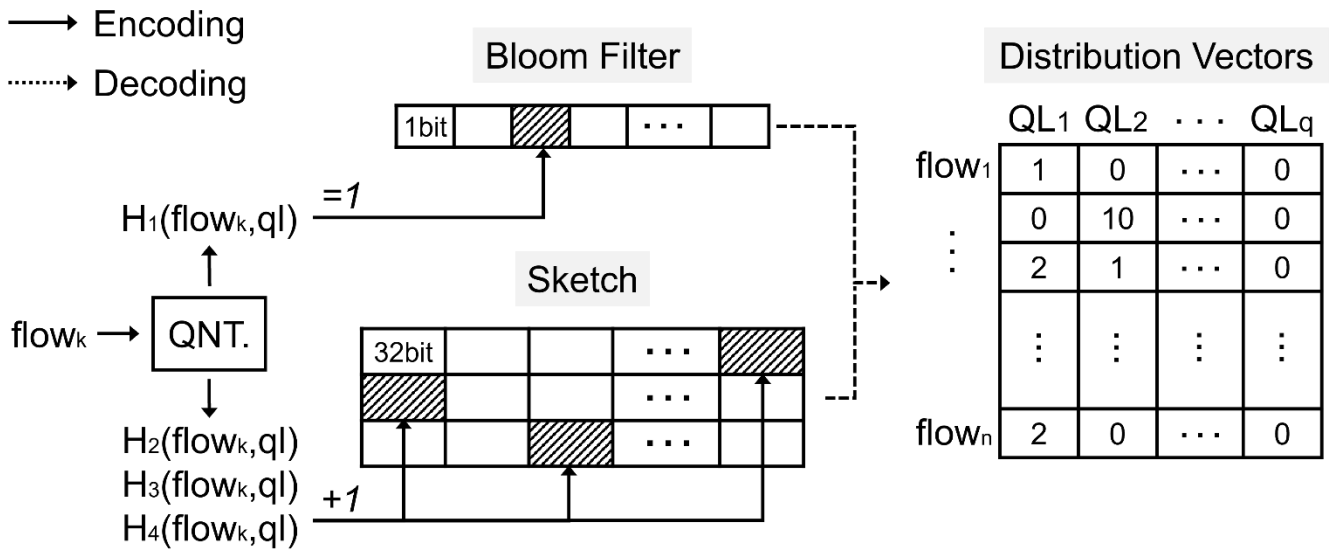
- **Encoding**
  - Convert feature value into a quantized bin (QNT).
  - Update Bloom Filter → Mark the feature as recorded.
  - Hash feature into Virtual Sketch → Update the relevant counters.



EWHA,
THE FUTURE
WE CREATE

Building Blocks of SketchFeature

- **Decoding**
  - Check Bloom Filter → Prevent Phantom Decoding.
  - Query Virtual Sketch → Retrieve feature values.



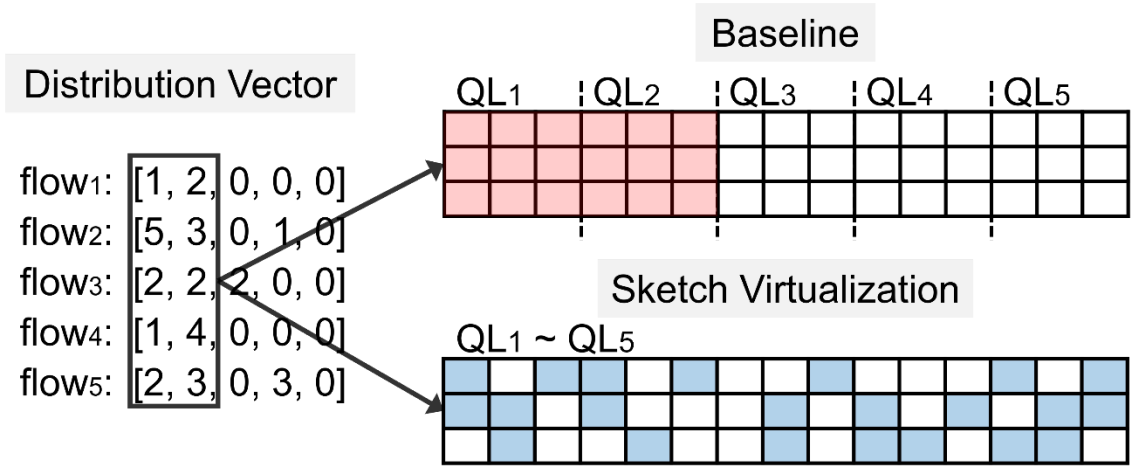
EWHA,
THE FUTURE
WE CREATE

Design Logistics of SketchFeature

- **Sketch Partitioning (Baseline)**
  - Count-Min Sketch is great for scalar values but fails for vector (distribution) features.
  - Existing methods (FlowLens, NetWarden) use hard partitioning

    ※ Causes inefficiency
- **Sketch Virtualization**
  - Single sketch memory is shared across all bins.
  - Different hash functions for each bin.

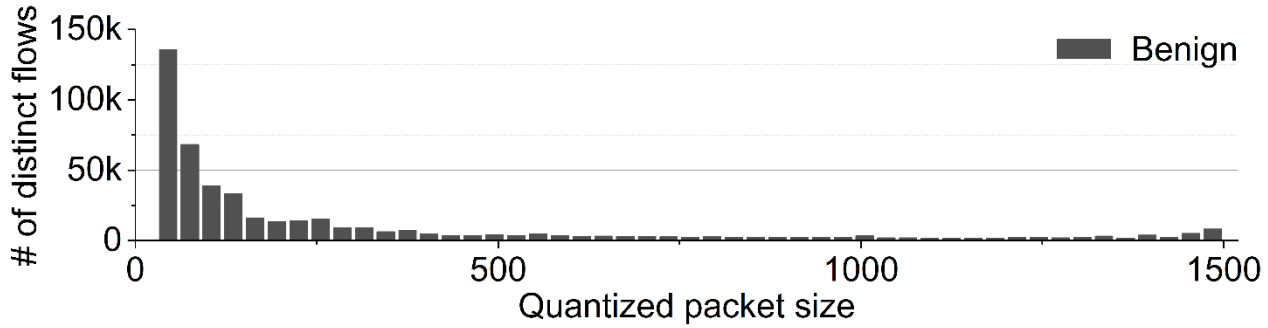
    → Ensures load balancing



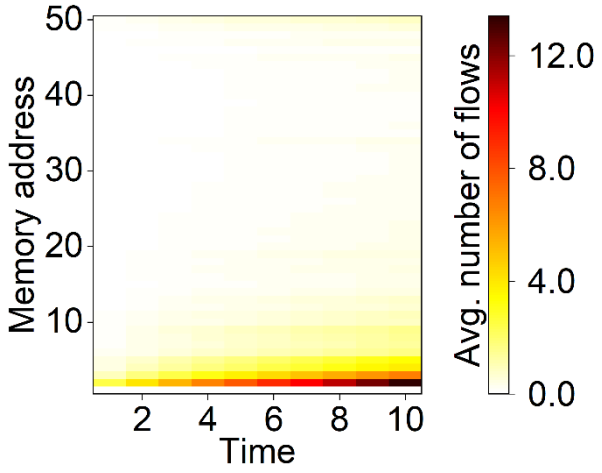
EWHA,
THE FUTURE
WE CREATE

Design Logistics of SketchFeature

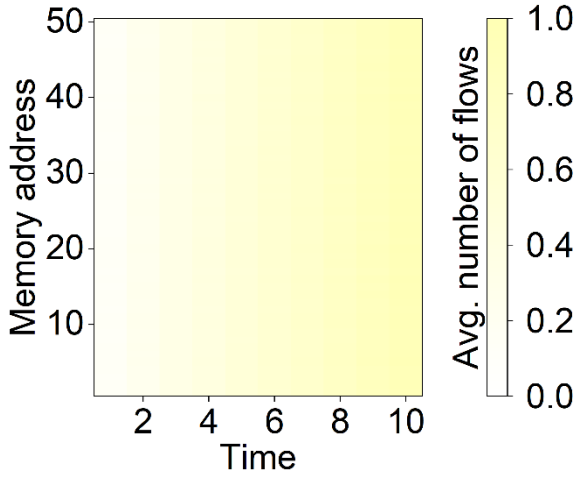
- **Packet size distribution of CAIDA Trace (Benign Trace)**



- **Sketch Partitioning**



- **Sketch Virtualization**



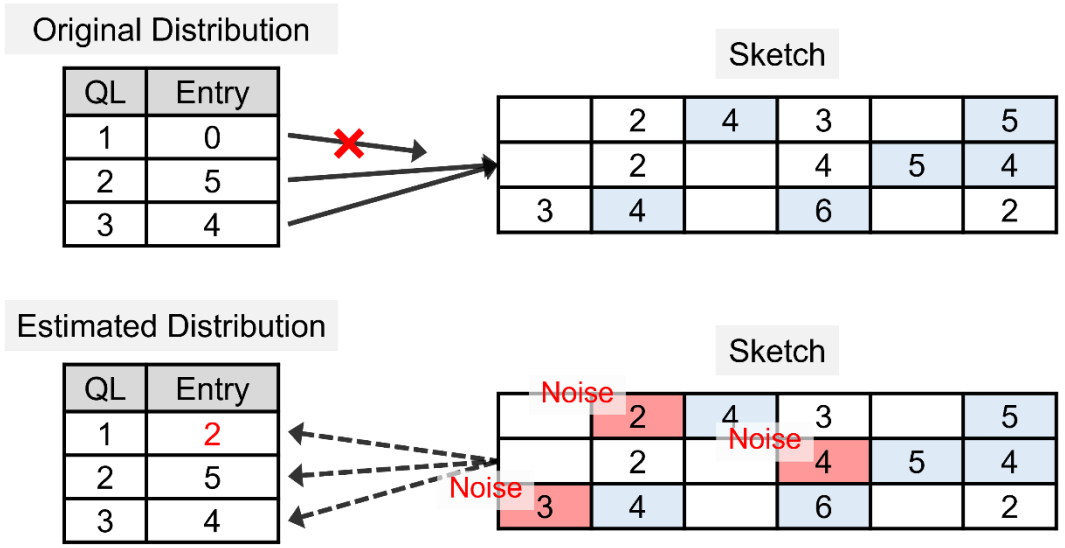
EWHA,
THE FUTURE
WE CREATE

Design Logistics of SketchFeature

- **Phantom Decoding Issue**
  - Standard sketches cannot differentiate between real and non-existent flow features.

    ※ Causes false positives, affecting feature accuracy
- **Membership Test using Bloom Filter**
  - Checks whether a feature actually exists before querying it.
  - Prevents Phantom Decoding.

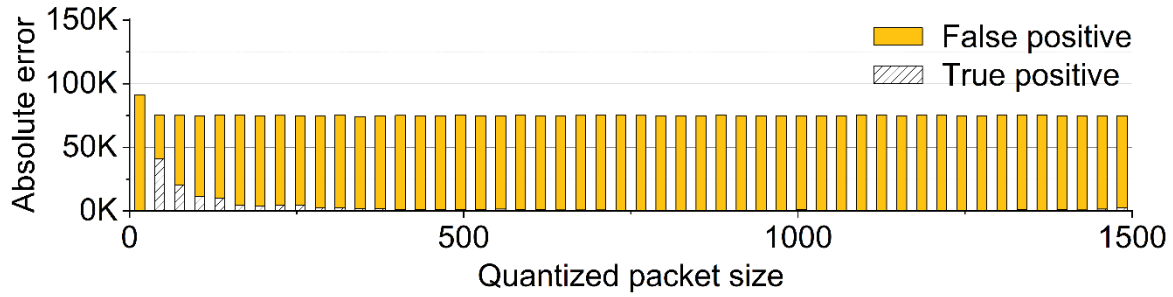
    → Reduces false positives



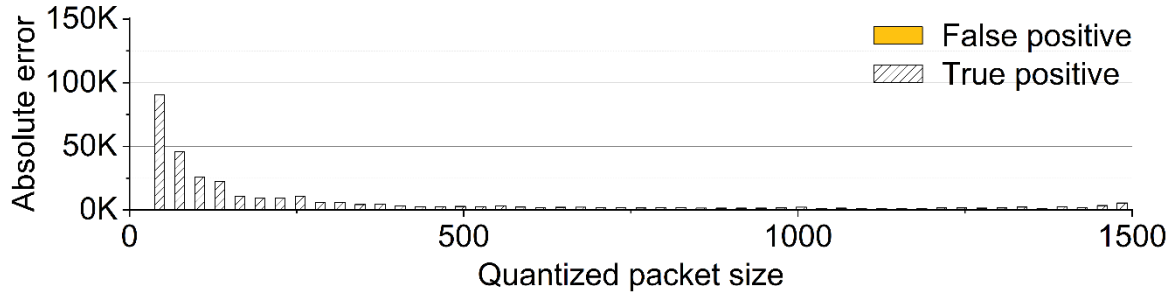
EWHA,
THE FUTURE
WE CREATE

Design Logistics of SketchFeature

- **Sketch virtualization without membership test**



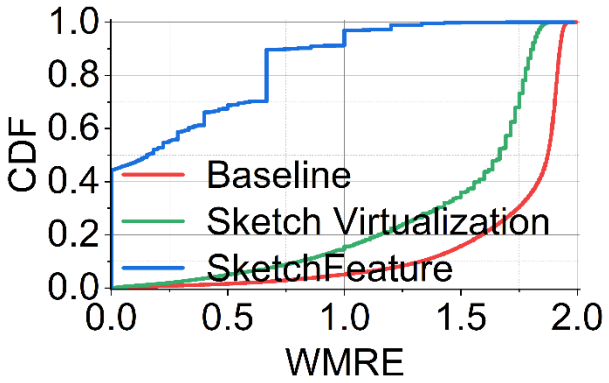
- **Sketch virtualization with membership test**



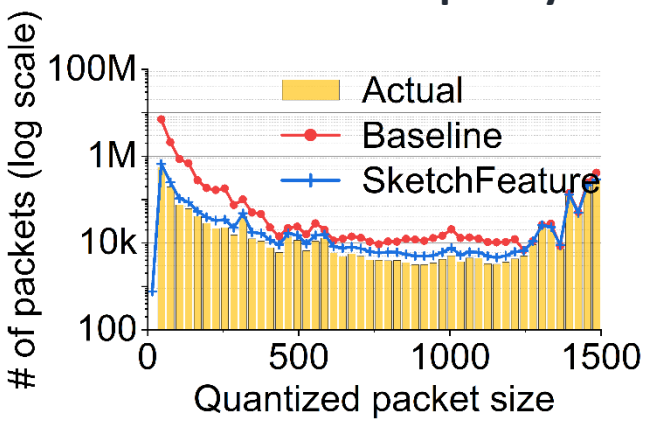
EWHA,
THE FUTURE
WE CREATE

HAF Feature Quality of SketchFeature

- **Per-flow distribution quality**



- **Traffic distribution quality**

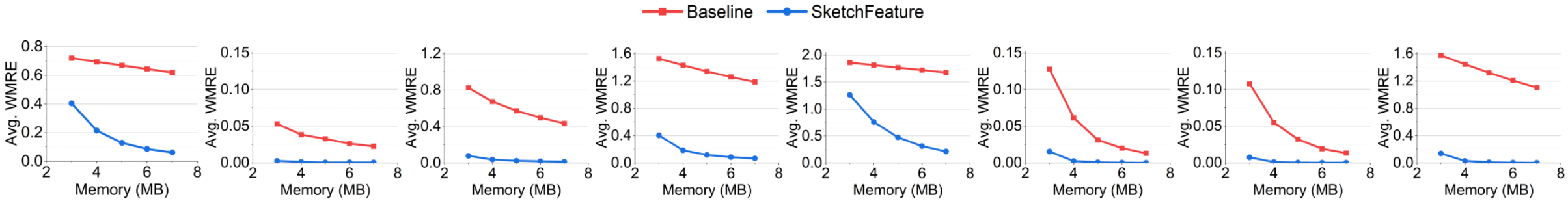


→ Accurate, efficient, and scalable for real-world in-network security applications

EWHA,
THE FUTURE
WE CREATE

Feature Quality Evaluation

- **Q1:** How accurate is SketchFeature's feature extraction varying memory size?



(a) Benign (IPD) (b) CSC (IPD) (c) DDoS (IPD) (d) Botnet (IPD) (e) Benign (PS) (f) CSC (PS) (g) DDoS (PS) (h) Botnet (PS)

Fig. 11: Comparison of 3rd-order (vector) feature quality. Average WMRE of distribution features (IPD and PS) varying memory.

EWHA,
THE FUTURE
WE CREATE

Feature Quality Evaluation

- **Q2:** How well does SketchFeature handle high workloads?

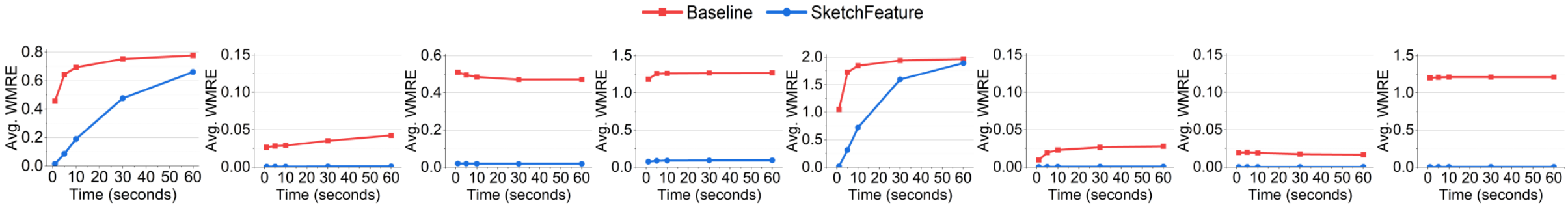


Fig. 12: Comparison of 3rd-order (vector) feature quality. Average WMRE of distribution features (IPD and PS) varying workload.

EWHA,
THE FUTURE
WE CREATE

End-to-end System Evaluation

- **Q3:** How does SketchFeature perform in security use cases?

Schemes	AUC	Acc.	F1	FPR	FNR
Covert Storage Channel (CSC)					
Perfect Measure	0.999	0.999	0.981	0.0	0.020
Baseline	0.986	0.999	0.978	0.0	0.027
NetBeacon [63]	0.995	0.981	0.983	0.018	0.018
FlowLens [6]	0.502	0.998	0.132	0.0	0.928
SketchFeature	**0.998**	**0.999**	**0.981**	**0.0**	**0.019**
Covert Timing Channel (CTC)					
Perfect Measure	1.0	1.0	1.0	0.0	0.0
Baseline	0.526	0.230	0.090	0.798	0.148
NetWarden [50]	0.999	0.999	0.999	0.0	0.0
FlowLens [6]	0.501	0.956	0.029	0.0	0.985
SketchFeature	**0.999**	**0.996**	**0.965**	**0.003**	**0.0**

Schemes	AUC	Acc.	F1	FPR	FNR
Distributed Denial of Service (DDoS)					
Perfect Measure	0.997	0.997	0.877	0.001	0.125
Baseline	0.561	0.991	0.197	0.0	0.875
NetBeacon [63]	0.738	0.995	0.636	0.0	0.521
FlowLens [6]	0.504	0.991	0.126	0.0	0.932
SketchFeature	**0.994**	**0.994**	**0.746**	**0.004**	**0.130**
Botnet					
Perfect Measure	0.922	0.997	0.845	0.001	0.142
Baseline	0.494	0.979	0.007	0.012	0.978
FlowLens [6]	0.498	0.992	0.106	0.0	0.942
SketchFeature	**0.914**	**0.995**	**0.715**	**0.003**	**0.172**

Conclusion

- **Extended Sketch Capacity**

  → Enabled 3rd-order feature extraction in the data plane

- **Resolved Technical Challenges**

  → Introduced membership tests to eliminate phantom decoding

- **Implementation in Tofino Switch**

  → Showed feasibility of SketchFeature in the data plane

- **Validated Theoretical and Experimental Performance**

  → Proved error bounds & feasibility

- **Advanced Security Applications**

  → Demonstrated high detection accuracy in real-world scenario

Thank You

EWHA,
THE FUTURE
WE CREATE