Pacific Northwest NATIONAL LABORATORY

Denial of Service Attack Detection via Differential Analysis of Generalized Entropy Progressions

May 21, 2024

Omer Subasi, **Joseph Manzano**, **Kevin Barker**
Pacific Northwest National Laboratory (PNNL)
Richland, WA, USA

U.S. DEPARTMENT OF ENERGY BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy

Introduction and Motivation

- **Denial-of-Service (DoS) attacks** are one of the most common and consequential cyber attacks in computer networks.
- A plethora of detection methods, yet the problem of **detecting DoS attacks remains an open problem**:
  - Detection approaches based on **hyperparameters**, such as **thresholds**, typically perform poorly.
  - Low **scalability** and low **cost**.
    - ✓ We treat **low cost** as having computational or memory complexity that is **lower than quadratic**, i.e., less than $O(N^2)$, and **no requirement of large amount of training data**.
  - High **false positives** and/or **false negatives**.
  - Differentiation between **flash events** and **actual DoS attacks** is non-trivial.
  - Misleading performance metrics: **Standard metrics** such as **accuracy** may be misleading.

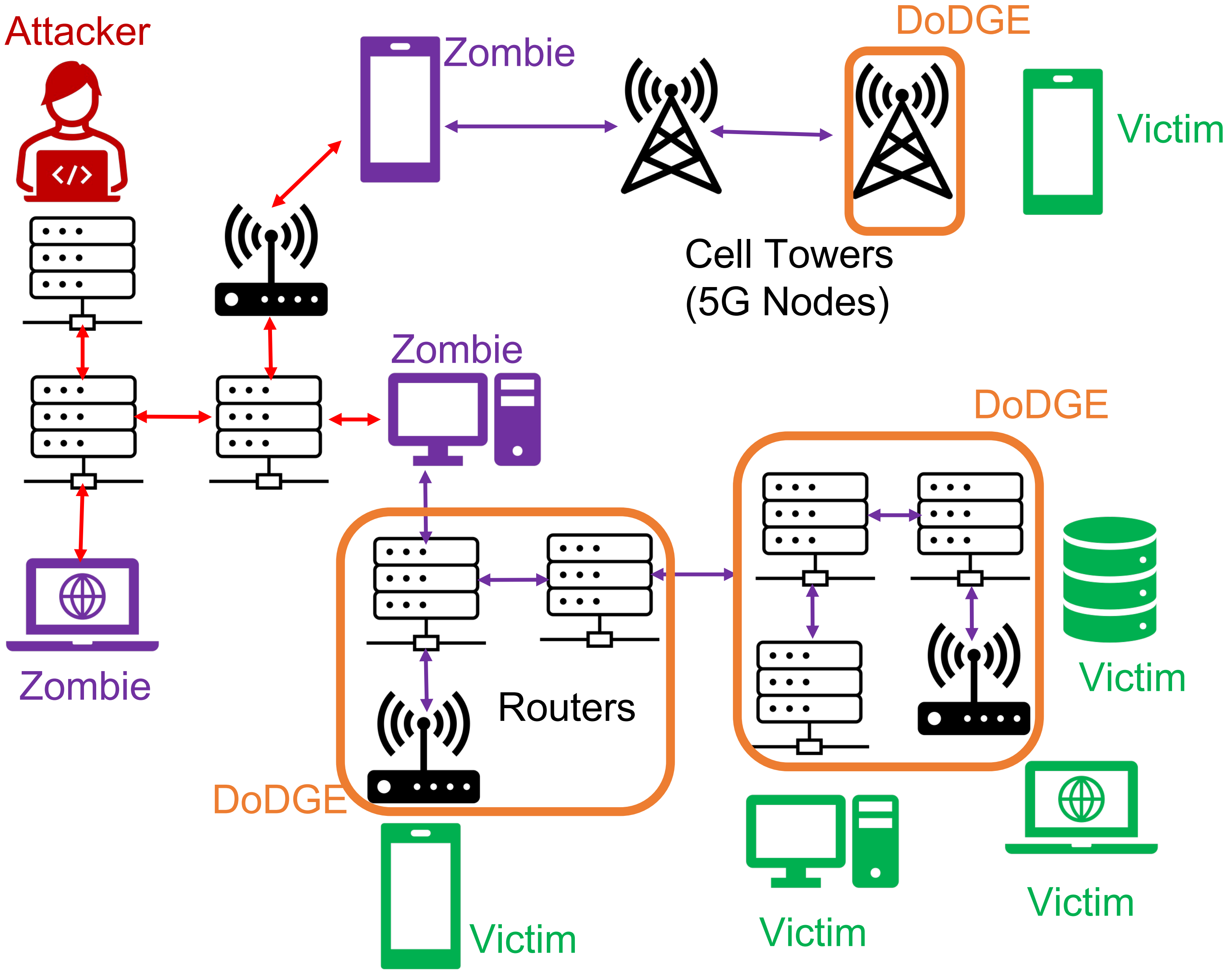
Our Proposal: DoDGE

- **DoDGE**:
  - A more **general entropy formulation** (**Tsallis**) than Shannon entropy.
    - ✓Improves detection accuracy
  - **Removes thresholds**:
    - ✓Instead, uses standard deviation of entropy progression derivatives
    - ✓Improves detection accuracy
  - Leverages **the asymmetric entropy behavior at target and source addresses to distinguish flash events and DoS attacks.**
  - Computations on **local data** (or nearby locations).
    - ✓**Low-cost**
  - Deployed on **5G edge nodes or Internet routers**
    - ✓Making DoDGE embarrassingly **parallel and scalable**

Background: Entropy

- **Entropy** appears in many areas such as thermodynamics, information theory and statistical mechanics.
- It generally refers to **a measure of disorder, randomness, and uncertainty**.
- In information theory, the most well-known entropy is **Shannon entropy**:
  - $\mathrm{H(X)} = \sum_i^N p_i \log(p_i)$ where X is a discrete random variable which has possible outcomes x_i with probability p_i.
- More **general formulations** exist such as:
  - **Renyi**: $R_a(X) = \frac{1}{1-\alpha}\log\left(\sum_i^N p_i^\alpha\right)$
  - **Tsallis**: $S_q(X) = \frac{1-\sum_i^N p_i^q}{q-1}$.
  - Used in complex dynamical systems having multifractality, systems with long range forces, and entanglement in quantum systems.
  - Such system require generalized entropy measures with weaker assumptions than Shannon's entropy such as non-additivity.

Threat Model



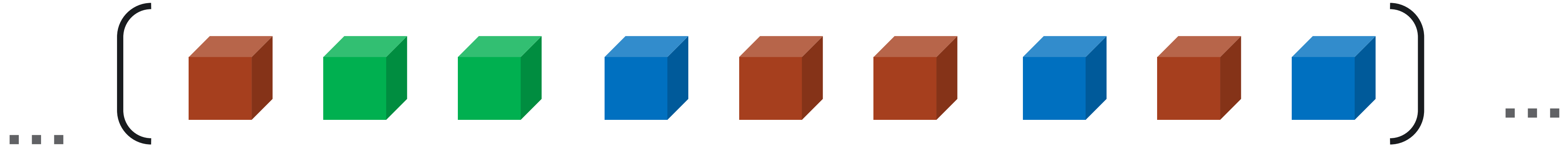
- An attacker exploits Internet and launches a DoS attack.

- DoDGE is placed at 5G nodes or cell towers and Internet routers.
  - At 5G nodes, DoDGE operates completely local (non-communicating)
  - At routers, DoDGE messages among a small group of neighbors (3-4).

- Majority vote.

Entropy Calculation

Let a window be 9 packets.



1. Compute **the frequencies of the packets** having the same color:

The frequency of the brown address is $\frac{4}{9}$.

The frequency of the green address is $\frac{2}{9}$.

The frequency of the blue address is $\frac{3}{9}$.

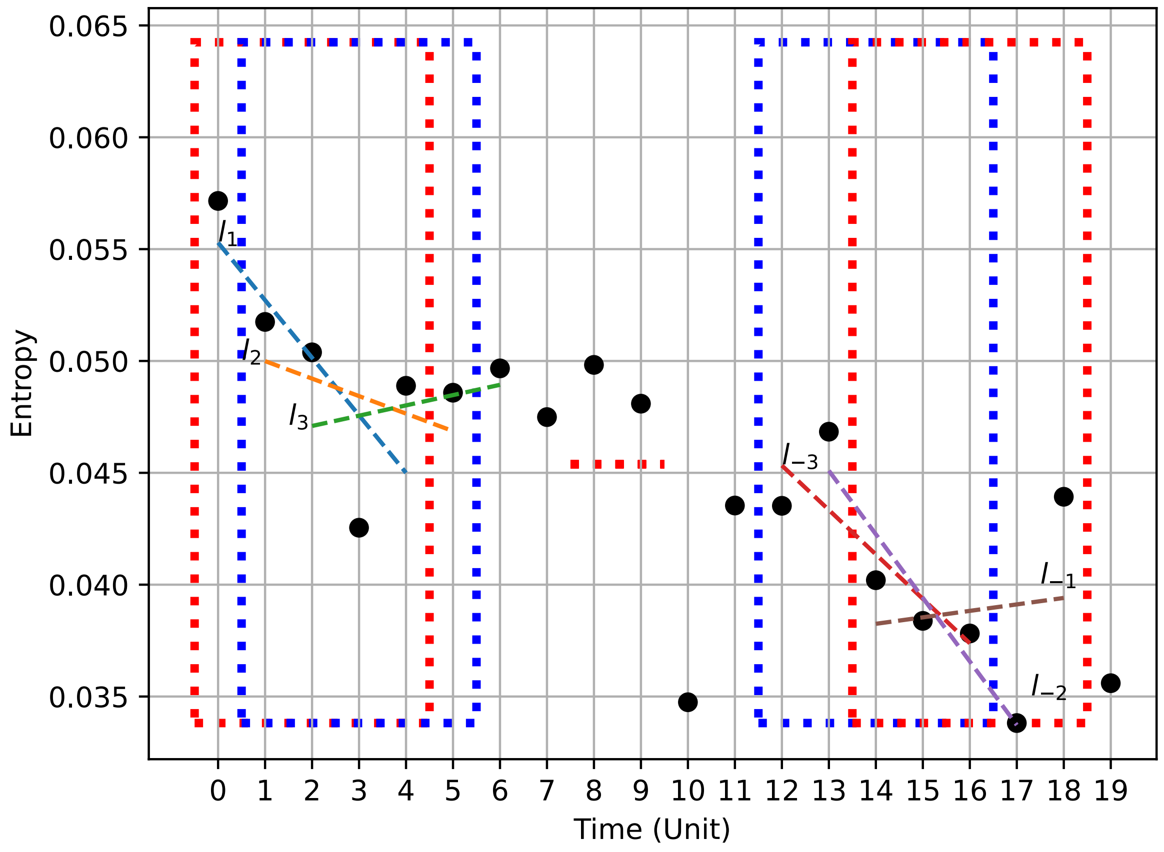
Same color = Same address

2. **Take the frequencies as the probabilities of the addresses** and compute the entropy for this window:

$$S_{q=8} = \frac{1-\sum_i^N p_i^q}{q-1} = \frac{1-\sum_i^N p_i^8}{8-1} = \frac{1-\left(\frac{4}{9}\right)^8-\left(\frac{2}{9}\right)^8-\left(\frac{3}{9}\right)^8}{7} = 0.1426$$

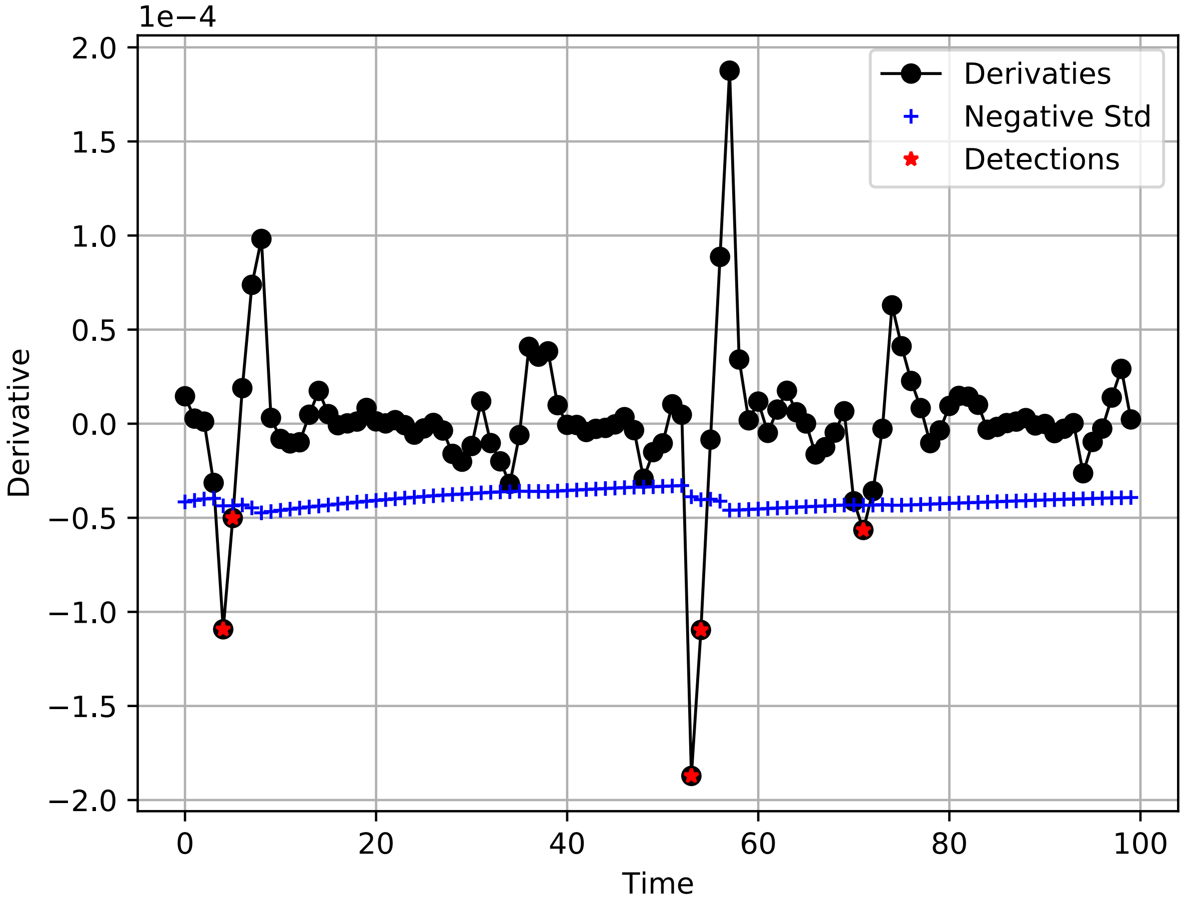
Differential Analysis of Generalized Entropy Progressions: Key Ideas I

- **Key Ideas I:**
  - We keep track of the **entropy progression** which is the time series of the entropies computed based on source or destination addresses.
  - To detect a decrease in entropy, we check if the **derivative** of the entropy progression is negative.
  - To calculate the derivative, we use the simplest model: **line of best fit**. The slope of this line is the derivative of the progression. If the **derivative** is **negative**, then the **entropy** is **decreasing**.



Differential Analysis of Generalized Entropy Progressions: Key Ideas II

- **Key Ideas II:**
  - We also compute **dynamically** the **standard deviation** of the entropy progression to increase the precision of attack detection.
  - We **avoid** using **thresholds**.
  - An attack is signaled when the **derivative** of the progression is **less than** the **negative of** the **standard deviation**.



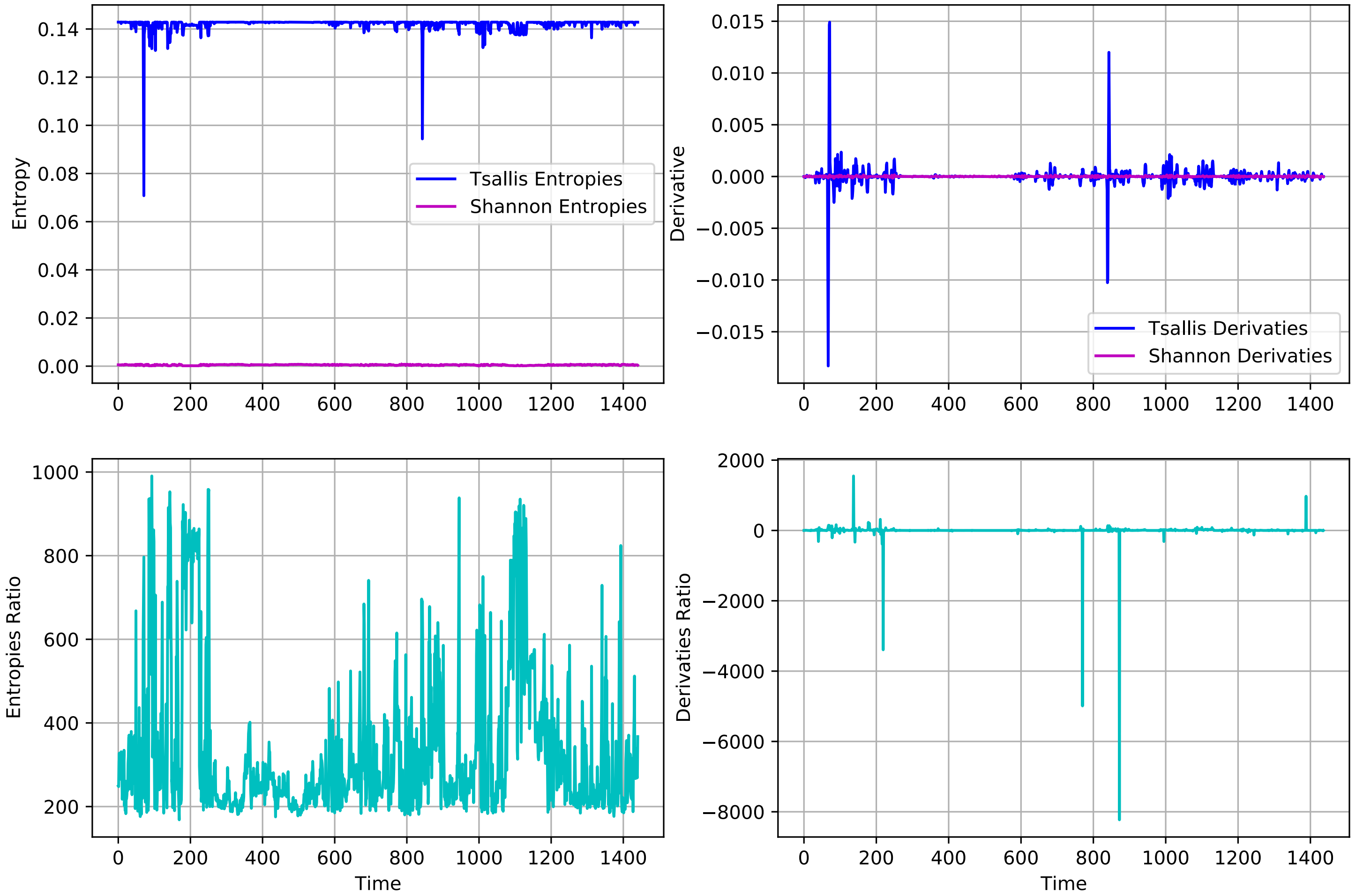
Differential Analysis of Generalized Entropy Progressions: Key Ideas III

- **Key Ideas III:**
  - We use **generalized entropies** to **amplify** the **magnitude** of the computed entropy. This improves the precision and accuracy of attack detection.

Shannon: $H(X) = \sum_{i}^{N} p_i \log(p_i)$ ✗

Renyi: $R_a(X) = \frac{1}{1-\alpha} \log(\sum_{i}^{N} p_i^{\alpha})$ ✗

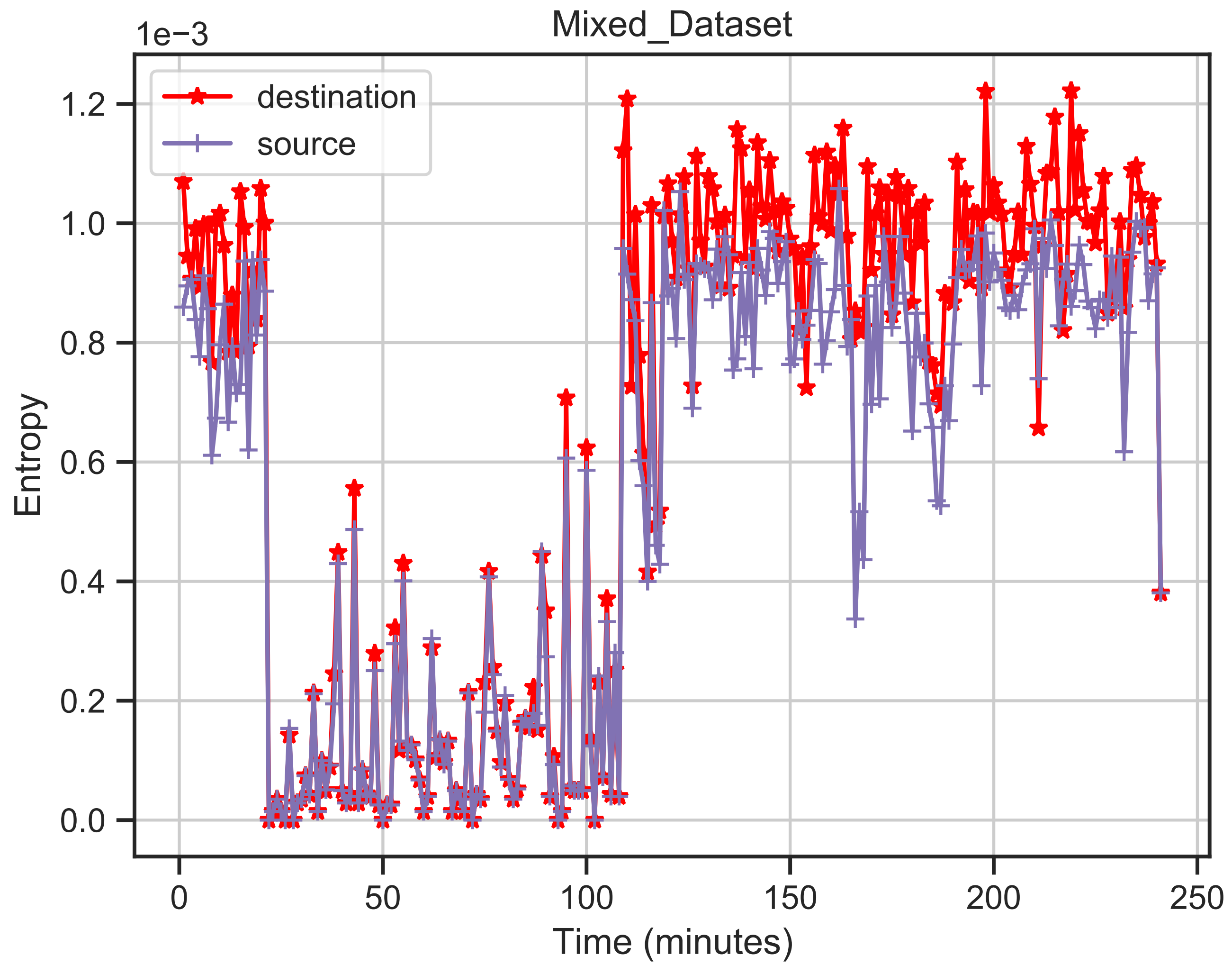
Tsallis: $S_q(X) = \frac{1 - \sum_i^N p_i^q}{q-1}$ ✓



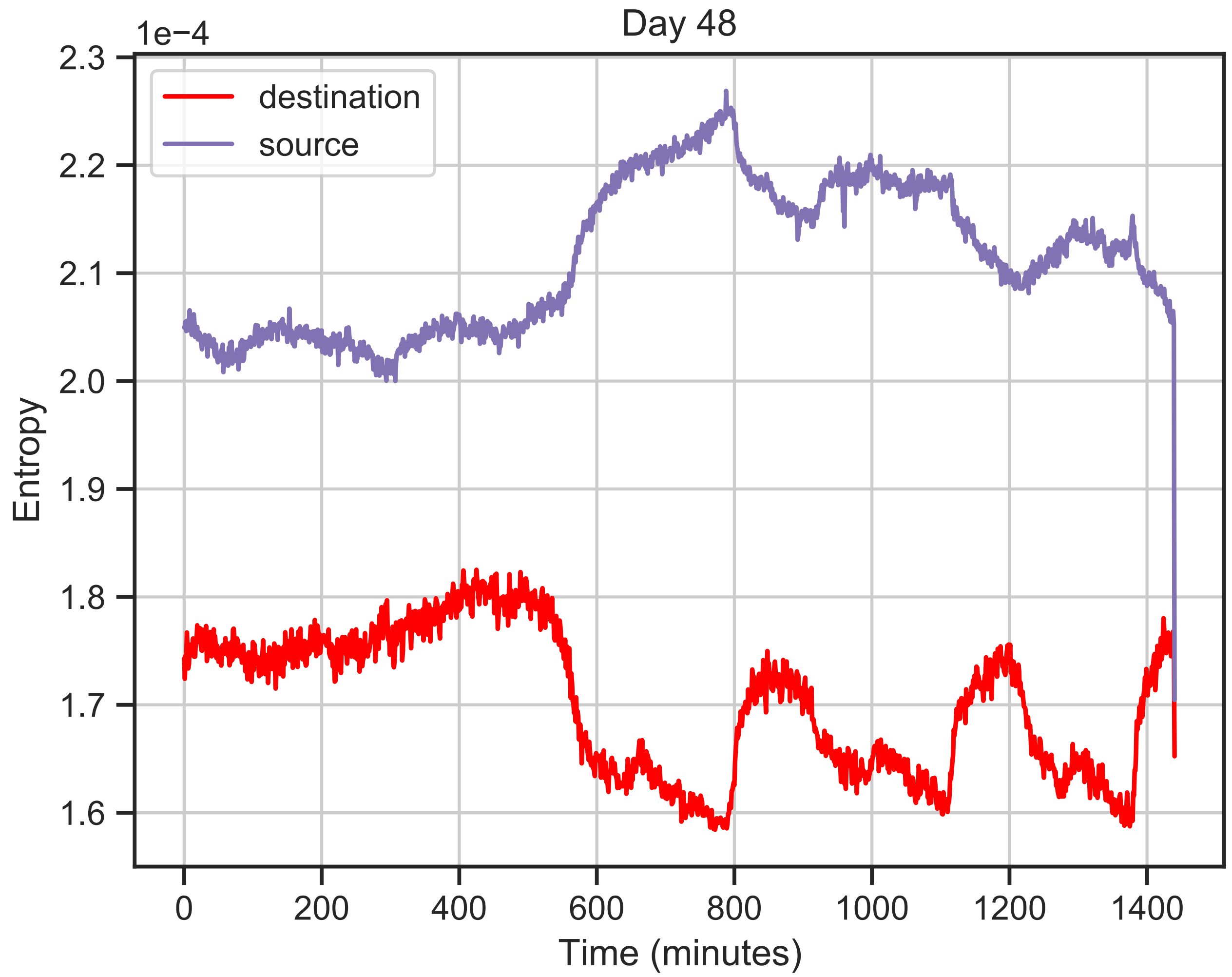
Differential Analysis of Generalized Entropy Progressions: Key Ideas IV

- **Key Ideas IV:**
  - Leveraging **the asymmetric entropy behavior in flash events**.

No Flash Events



Flash Events



DoDGE Algorithm (Simplified)

Inputs: The Destination Progression $\{EP_{D_i}\}$, the Source Progression $\{EP_{S_i}\}$

...

while (True)

…

destination_slope = **line_of_best_fit**($\{EP_{D_i}\}$) //slope for destination entropies

source_ slope = **line_of_best_fit**($\{EP_{S_i}\}$) //slope for source entropies

σ = … // dynamically compute standard deviation for <u>destination derivatives</u>

if (destination_slope < −σ)

if (source_slope > 0)

Flash Event

else DoS Attack, Launch Mitigation

else Normal Traffic

DoDGE Complexity Analysis

- **Computational complexity:** For **N** number of network packets in **a single window**:
  - Entropy computation is **O(N)**.
  - Fitting the line of best fit to the entropy progression which has a fixed small number of entropies is **O(1)**.
  - Computing the standard deviation of the derivatives on-the-fly is **O(1)**.
  - Checking the detection condition is **O(1)**.
  - Therefore, the total computational complexity is **O(N)**.
- **Memory complexity:** For **N** number of network packets in the unit-time window:
  - The memory for the window is **O(N)**.
  - The memory for the temporary variables needed for the method is **O(1)**.
  - Therefore, the total memory complexity is **O(N)**.

Threshold- and Entropy-based DoS Attack Detection

- Thresholds can be static or dynamic.
- A static threshold would be to compute the average entropy for benign traffic **offline** and use it as a reference.
  - When a detection method is in use, it signals an attack if the current entropy is bigger than this reference value.
- Dynamic thresholds is computed when the detection method is running.
- Dynamic thresholds are average values over longer periods of time - **not** computed for each time window.

Threshold-based and Entropy-based DoS Attack Detection Continued

- **Bidirectional entropy**
  - Incorporates both source and destination traffic flows.
- **Short- and long-term entropies**
- **Thresholds**:
  - Can be static or dynamic.
  - Many possibilities for dynamically computed thresholds:
    - ✓ $Threhold_t = \frac{1}{k}\sum_{j=t-k}^{t-1} threshold_j$ for some k.
- A **decision strategy** is Boolean-valued function whose input is entropies and thresholds.
  - It is used to decide if there is an attack or not.
  - Example:

$$\Psi(dst_{ste}, dst_{lte}, dst_{thr}, \ldots) = dst_{ste} < dst_{thr} \,\&\, dst_{lte} < dst_{thr}$$

Evaluation Datasets

- **“Application” Dataset:** Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting HTTP-based Application Layer DoS attacks on Web Servers in the presence of sampling." Computer Networks, 2017.
- **“Benign” and “Mixed” Datasets:** Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), January 2018.
- **“UDP” and “TCP” Datasets:** Derya Erhan, October 9, 2019, "Boğaziçi University DDoS Dataset", IEEE Dataport.
- **A Labelled Dataset for ML Comparison:** I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, “Developing realistic distributed denial of service (ddos) attack dataset and taxonomy,” in International Carnahan Conference on Security Technology, 2019, pp. 1–8.
- **France World Cup 98 Dataset:** Internet traffic to www.france98.com during 1998 World Cup in France. It includes benign traffic with flash events occurring during match times. Randomly chosen Days 48, 63, 66, 69, and 78.

Performance Metrics: Standard Metrics

- TP = true positive
- FP = false positive
- TN = true negative
- FN = false negative

- **Standard metrics**:
  - $\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$
  - $\text{Precision} = \frac{TP}{TP+FP}$
  - $\text{Recall} = \frac{TP}{TP+FN}$

Standard metrics are suitable for balanced data.

In balanced data, different classes have similar number of instances.

Pacific Northwest NATIONAL LABORATORY

Performance Metrics: Balanced Accuracy

- TP = true positive.
- FP = false positive.
- TN = true negative.
- FN = false negative.
- TPR = true positive rate.
- TNR = true negative rate
- **TPR**$= \frac{TP}{TP+FN}$
- **FPR**$= \frac{TN}{TN+FP}$
- **Balance Accuracy** $= \frac{1}{2}(TPR + TNR) = \frac{1}{2}(\frac{TP}{TP+FN} + \frac{TN}{TN+FP})$

When data is highly unbalanced, standard metrics are not **suitable** and can be **misleading**.

In unbalanced data, different classes have very different number of instances.

Metrics, such as **balanced accuracy**, that take account the imbalance are needed to be used.

Performance Metrics: Balanced Accuracy Cont.

- In the test dataset we used, among **4.3 million** instances **only 35772** instances are **benign**. That is, **only 0.8% are benign**.
- Considering **ML models**, they tend to be **biased** toward the class(es) that have a **high number of instances.**
- Regardless of their performance for the classes with few instances, ML models' performance in terms of standard metrics will be close to 100%, especially if the imbalance is very high.
- This shows that the percentages with respect to standard metrics can be misleading.
- We see this in our evaluation.

Pacific Northwest NATIONAL LABORATORY

Comparison to ML

Algorithm	Accuracy	Precision	Recall	Balanced
SVC	99.20%	99.20%	99.90%	50.20%
DT	99.20%	99.40%	99.90%	61.60%
RF	99.30%	99.30%	99.90%	59.10%
KN	12.10%	97.40%	11.80%	37.10%
GB	99.20%	99.40%	99.80%	61.20%
LR	99.20%	99.20%	100%	50.10%
CONV	99.20%	99.20%	100%	50.00%
LSTM	99.20%	99.20%	100%	50.00%
GRU	99.20%	99.20%	100%	50.00%
ED	99.20%	99.20%	99.90%	49.90%
DoDGE	**75.70%**	**100%**	**75.50%**	**99.30%**

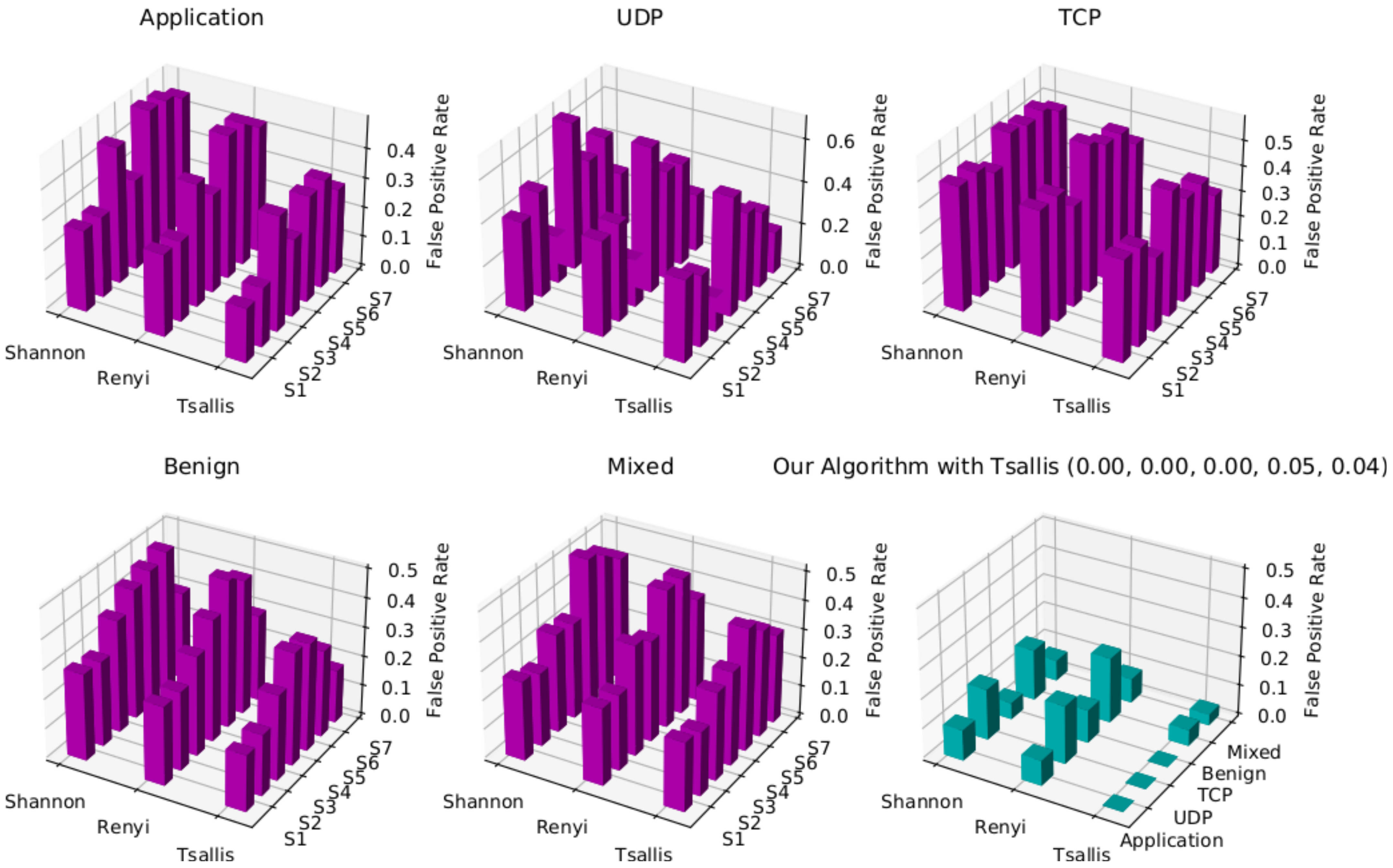
- **DoDGE has balanced accuracy of 99%.**
- **All 10 ML models have balanced accuracy < 62%.**
- **Average ML balanced accuracy is 52%.**

Support vector machines (SVC)
Decision Trees (DT)
Random Forest (RF)
K-Neighbors (KN),
Gradient Boosting (GB)
Logistic Regression (LR)
Convolutional Network (CONV),
Long Short-Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Encoder- Decoder (ED)

False Positive Rates for All Methods

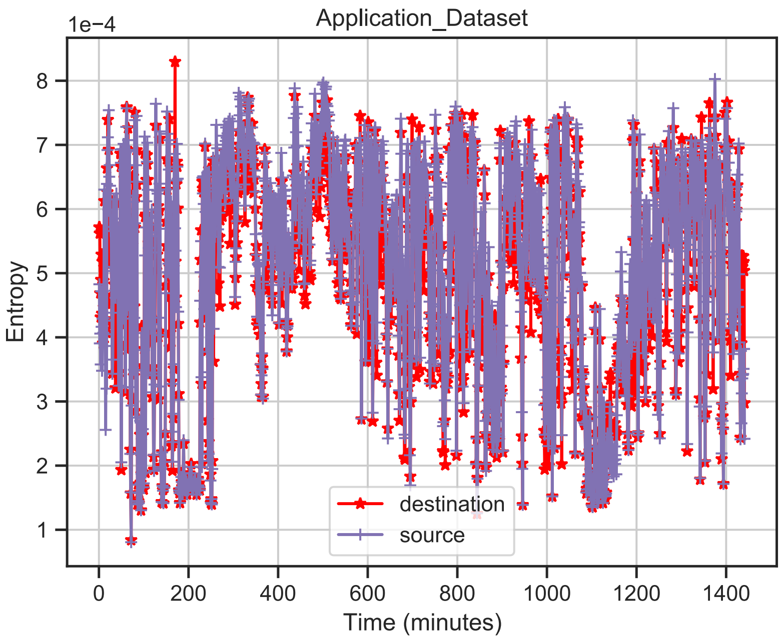
DoDGE outperforms threshold-based methods by two orders of magnitude for false positives on average.

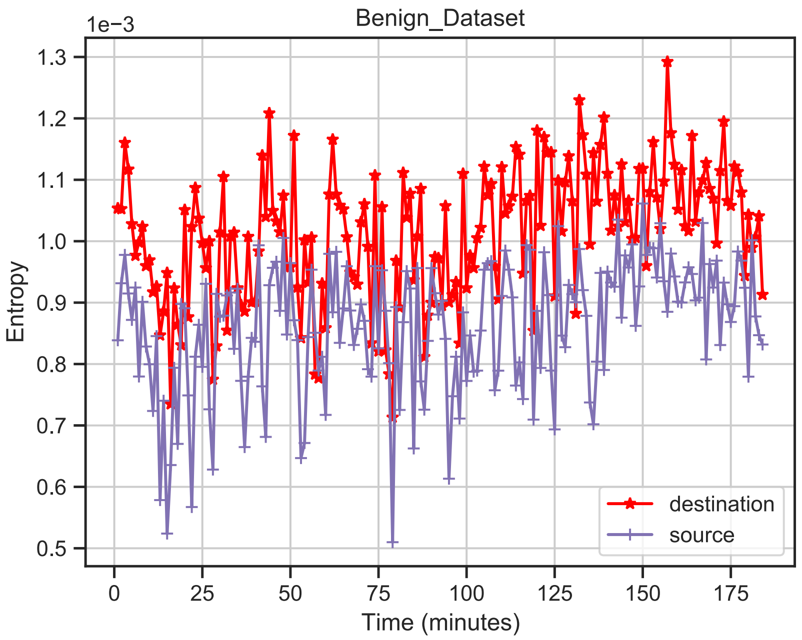
Purple: Thresholds
Green: DoDGE

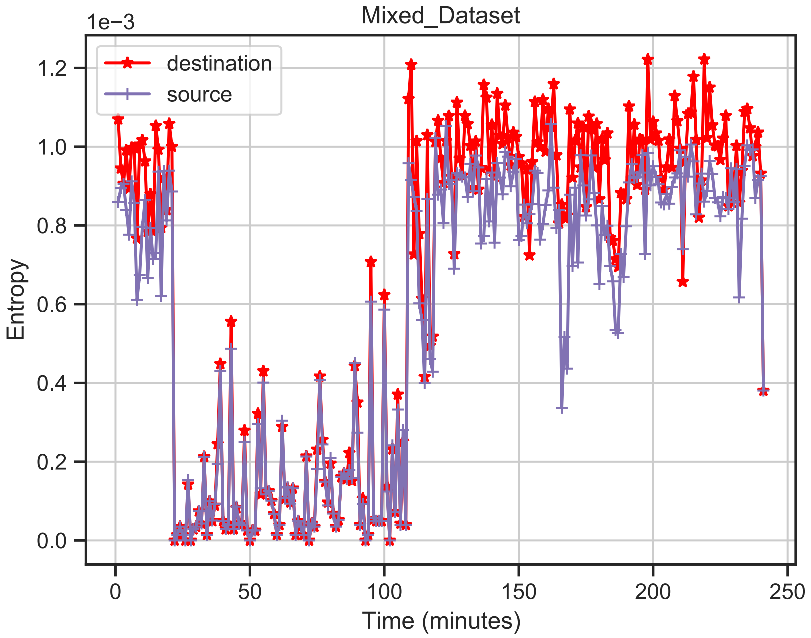


Flash Events: Entropy at Source and Destination Addresses

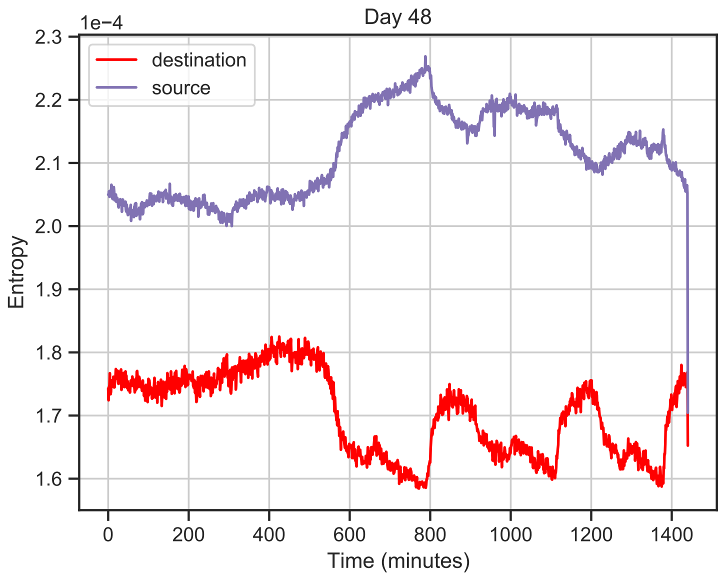
No Flash Events

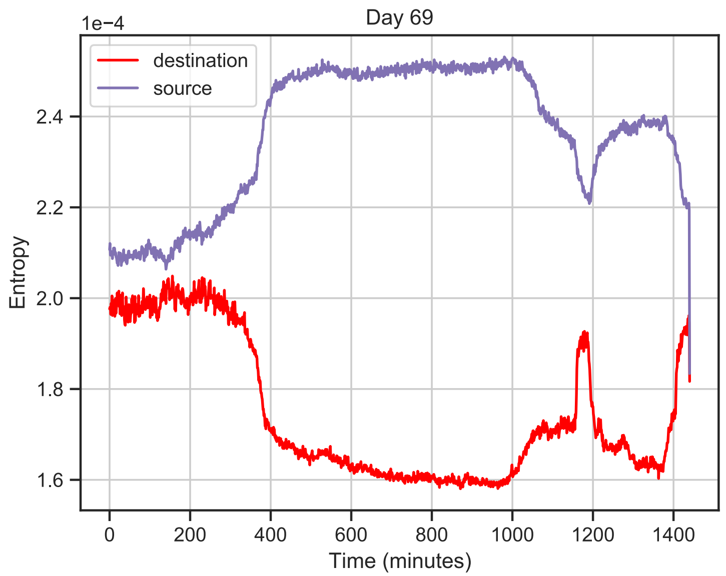


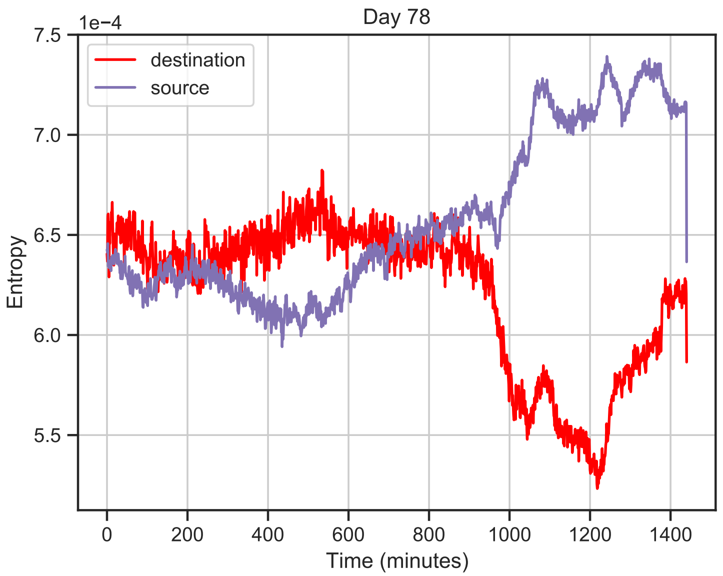




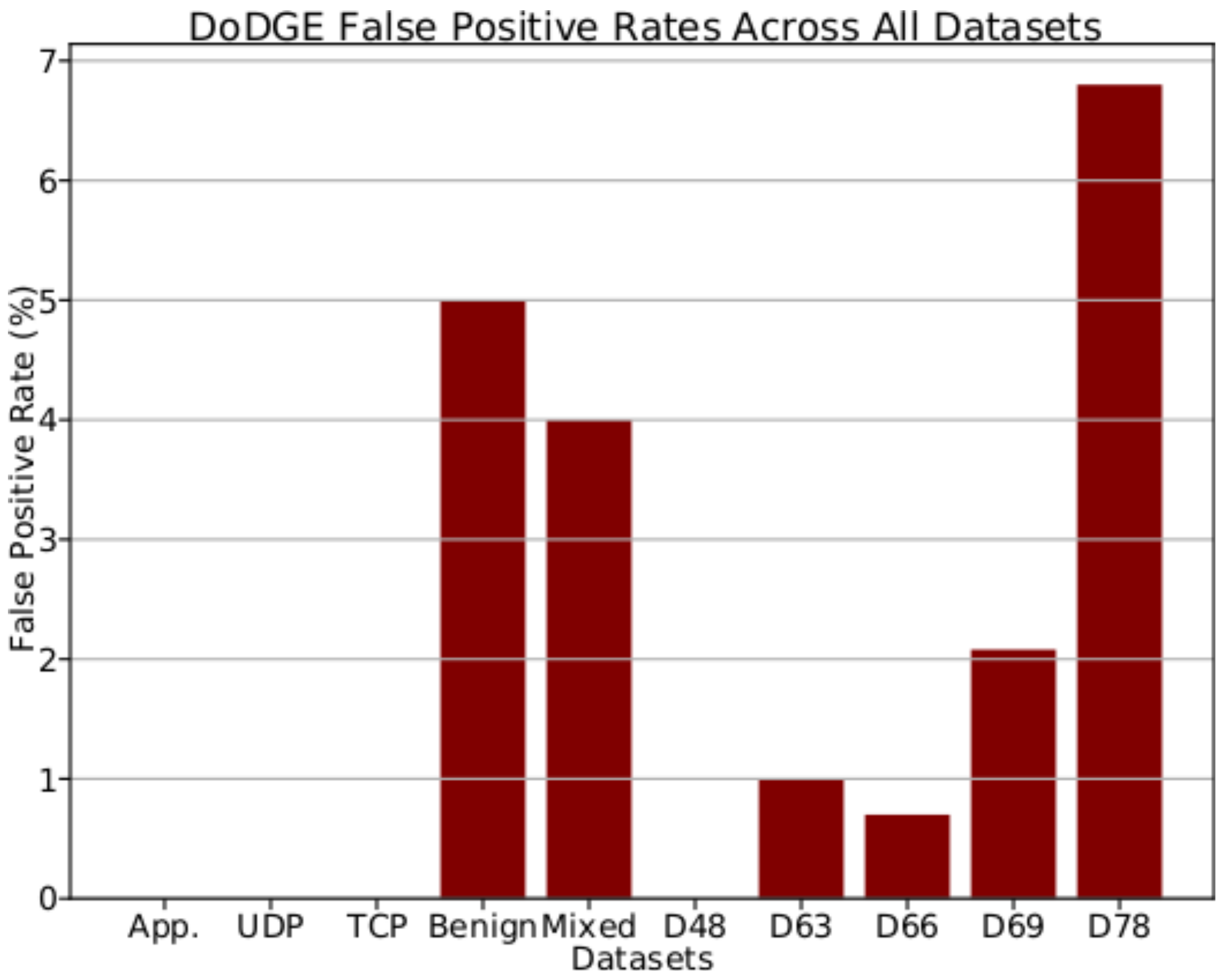
Flash Events

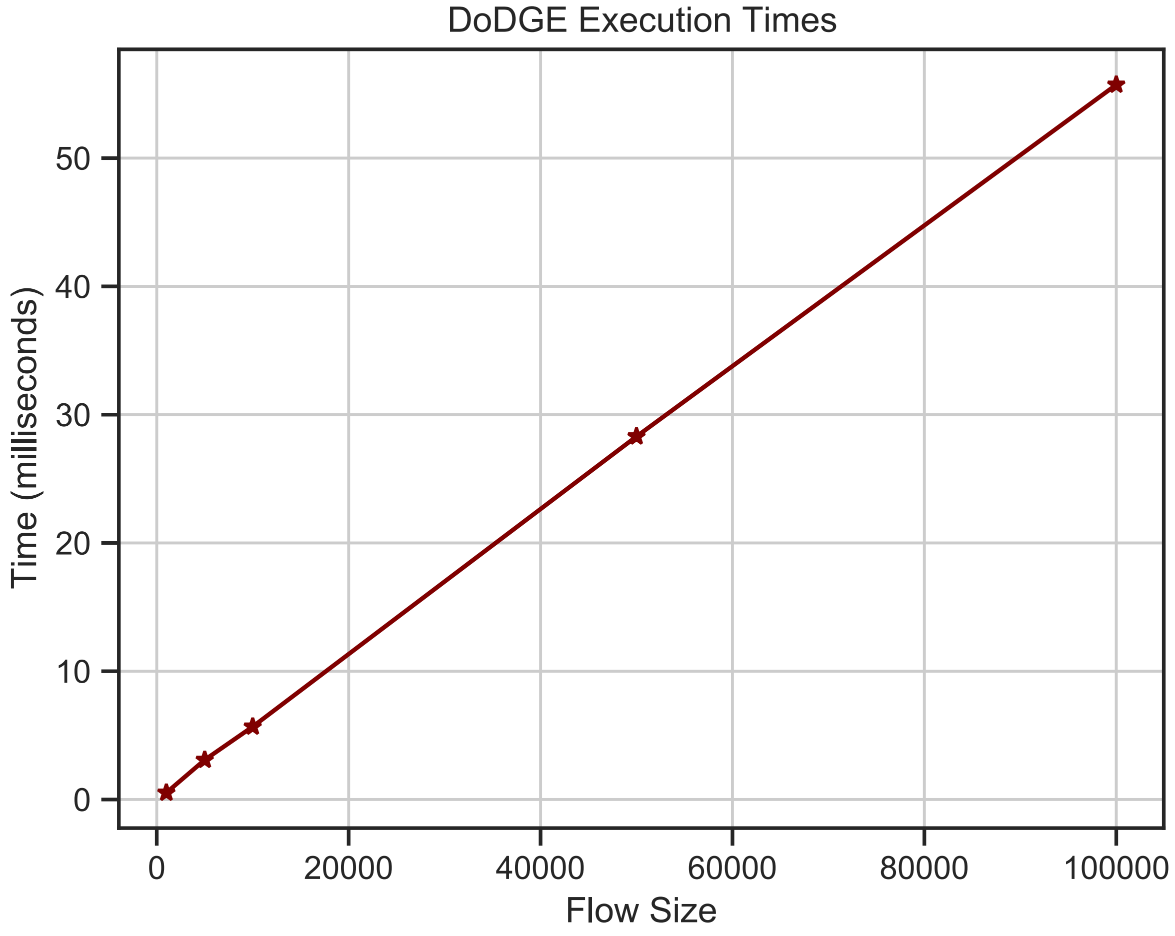






False Positive Rates and Scalability





DoDGE achieves low false positive rates.

DoDGE is lightweight and scalable.

Pacific Northwest
NATIONAL LABORATORY

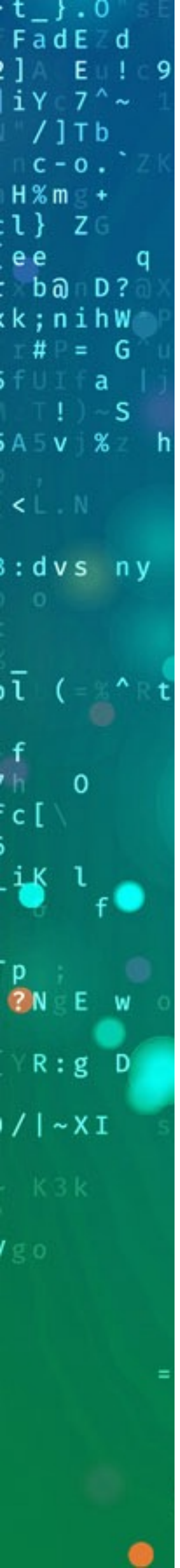
Conclusions

- A **D**o**S** attack detection method using **D**ifferential analysis of **G**eneralized **E**ntropy progressions - **DoDGE**.
- DoDGE outperforms threshold-based methods by two orders of magnitude in terms of false positives on average.
- DoDGE's balanced accuracy of 99% vs all 10 ML/DL models' balanced accuracy < 62%.
  - The average balanced accuracy is 52% for ML/DL.
- DoDGE successfully differentiates flash events and DoS attacks.
- DoDGE is
  - lightweight - linear time and memory complexity -,
  - scalable,
  - embarrassingly parallel.

Acknowledgement

- This work was supported by the U.S. DOE Office of Science, Office of Advanced Scientific Computing Research, under award 66150: “CENATE - Center for Advanced Architecture Evaluation” project. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.

- Link to our paper: https://ieeexplore.ieee.org/document/10224957



Thank you