Deanonymizing Internet Traffic with Website Fingerprinting

Nate Mathews

nate.mathews@mail.rit.edu

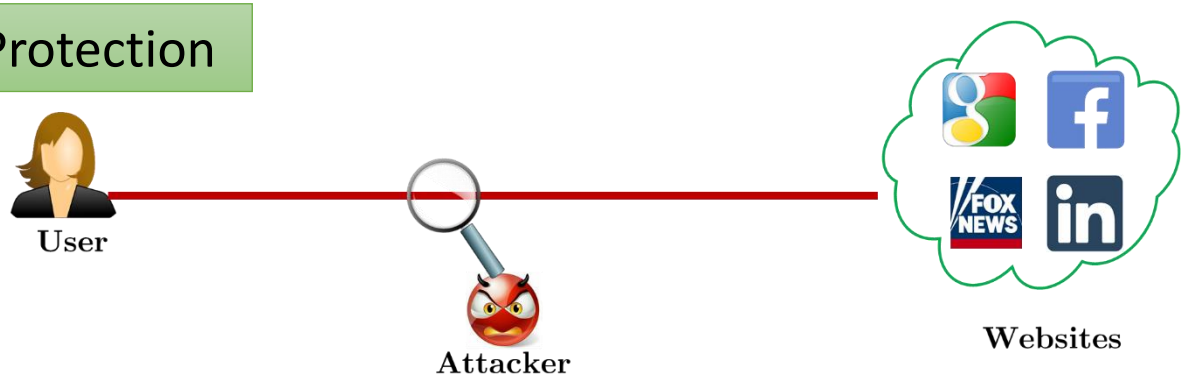
Advisor: Dr. Matthew Wright

matthew.wright@rit.edu

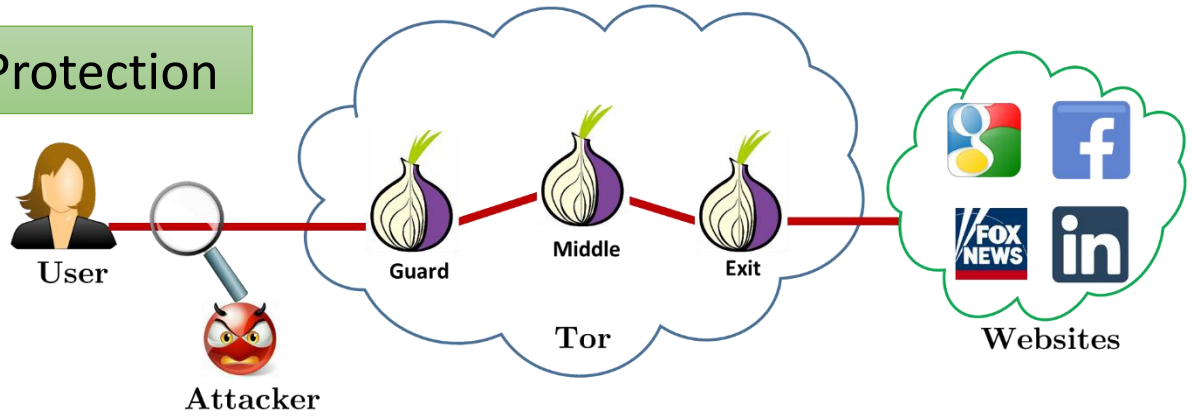
Rochester Institute of Technology
Global Cybersecurity Institute

Internet Anonymity

No Protection



Tor Protection



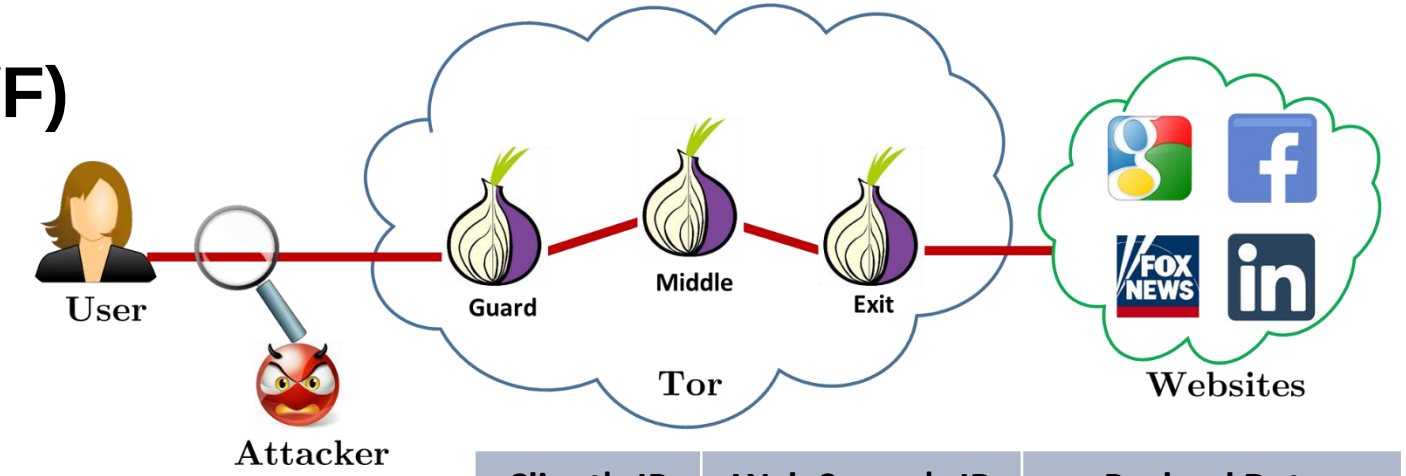
Tor Anonymity System

- Incrementally creating a circuit
- Sophisticated encryptions
- **No individual node has the complete path information**

The attacker fails to link user to the actual website she is visiting

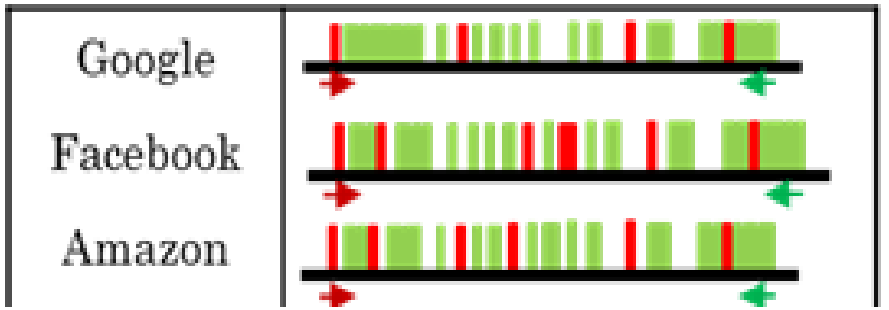
Internet Anonymity

- **Website Fingerprinting (WF)**
  - Try to link **the client** to **the website**



Concealed by Tor

Network Traffic



Classify

Information Leak
- Number of Packets Statistic
- Burst of packets

* Unique for each website

Website Fingerprinting

- **Experimental design**
  - Closed-world
    - Benchmark
  - Open-world
    - Comparable to real-world

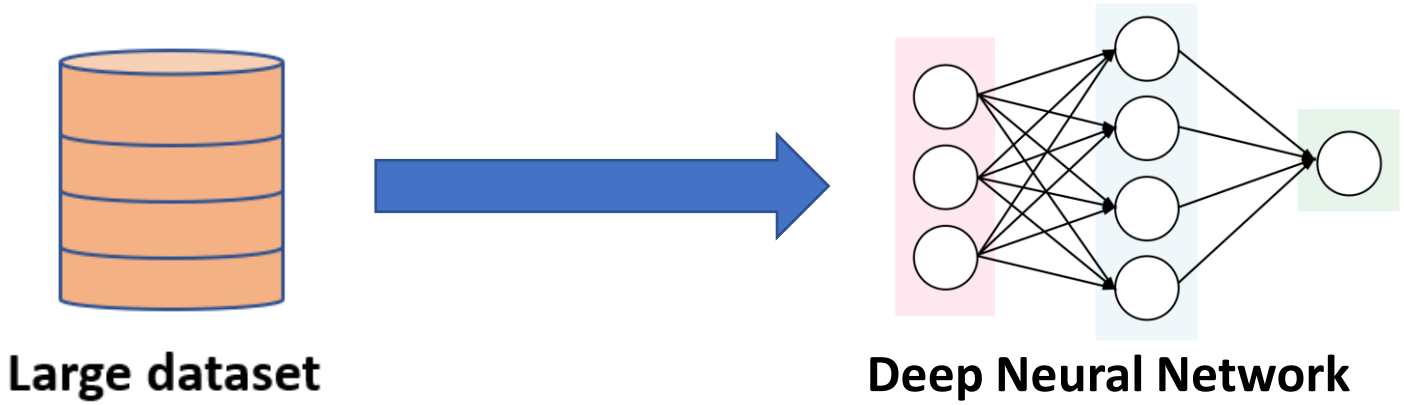
Website Fingerprinting

- **WF attacks using hand-crafted features** *[Panchenko et. al, Hayes et. al]*
  - Designed features
  - Machine learning classifiers
    - SVM, Random Forest, k-NN

- Panchenko et al. Website fingerprinting at internet scale, NDSS 2016
- Hayes and Danezis. k-Fingerprinting: A robust scalable website fingerprinting technique, USENIX 2016.

Website Fingerprinting

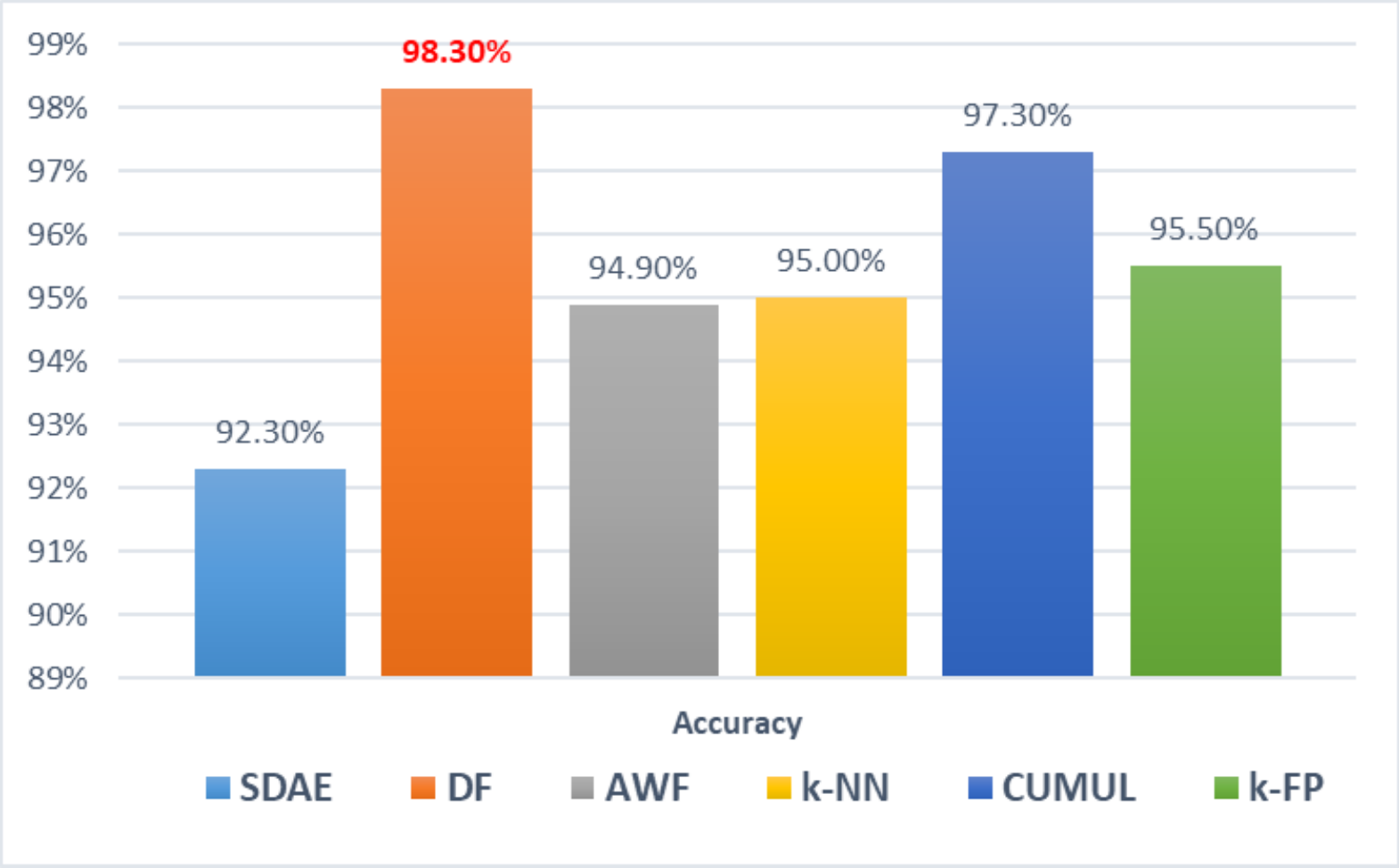
- **WF attacks using deep learning** *[Sirinam et. al, Bhat et al.]*
  - Automated feature learning
  - Higher performance
    - Larger data requirements



- Sirinam et al. Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning, CCS 2018
- Bhat et al. Var-CNN: A Data-Efficient Website Fingerprinting Attack Based on Deep Learning, PoPETS 2019

Website Fingerprinting

- Closed-world performance

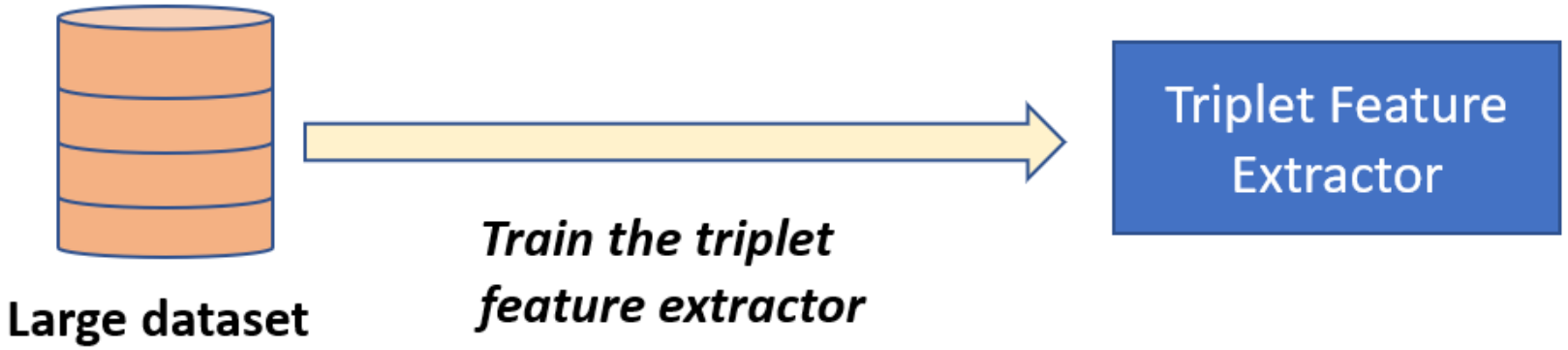


Website Fingerprinting

- **New directions in WF attacks**
  - Improve performance in open-world
  - Improve attacker assumptions
    - *Lower data requirements*
    - *Webpage vs. Website fingerprinting*

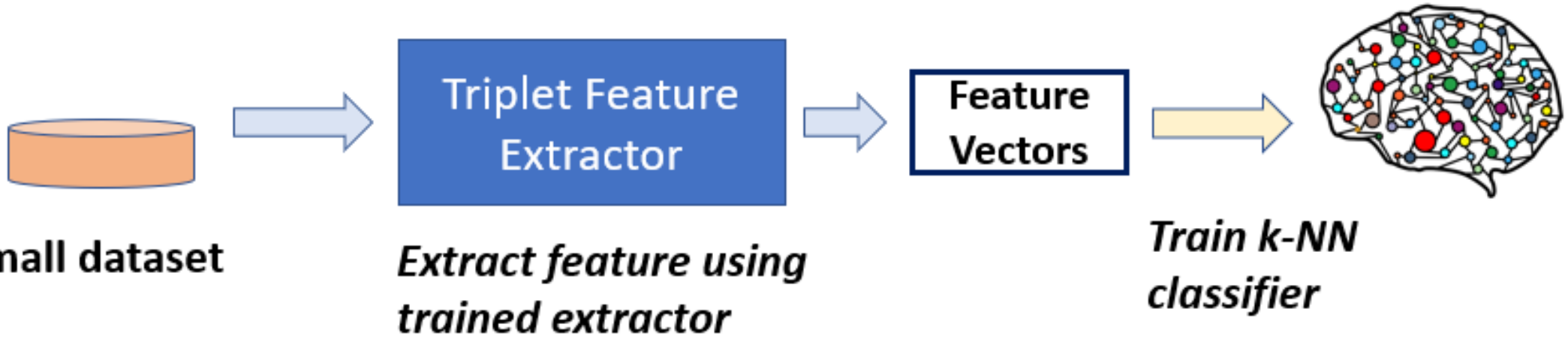
Recent-work: Triplet Fingerprinting

1. **Pre-training step**
   - Train triplet network as feature extractor
   - Large, preexisting dataset
   - Nontargeted



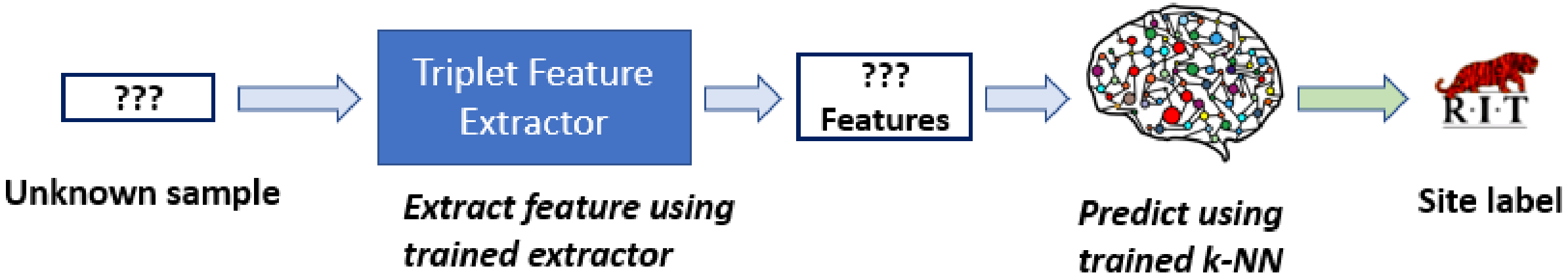
Recent-work: Triplet Fingerprinting

2. **Training step**
   - Collected targeted data.
   - Process into features and train classifier



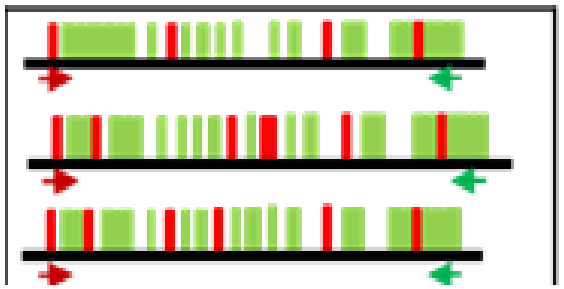
Recent-work: Triplet Fingerprinting

3. **Attack step**
   - Capture unknown sample.
   - Predict with trained classifier.



Working Towards a Defense

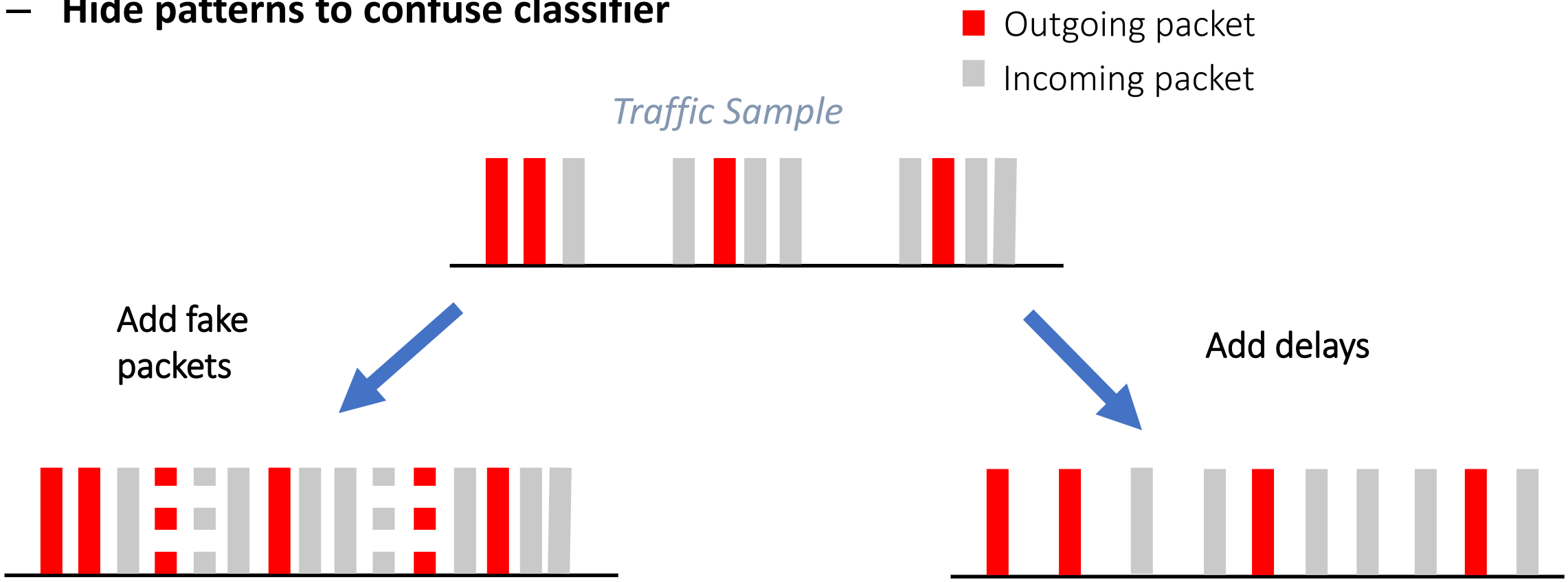
Why?



Black box DL Model

WF Defenses

– **Hide patterns to confuse classifier**



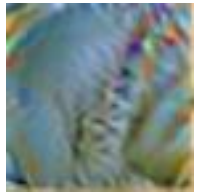
WF Defenses

- **Popular strategies**
  - Stuff trace with fake traffic
    - **High overheads harm network performance**
  - Create traffic pattern “collisions”
    - **Lower overheads**
    - **Mathematical guarantees**
    - **Cumbersome to implement**

Ongoing-work: Adversarial Patches

African-Elephant
(92% prediction)

+



Adversarial Patch

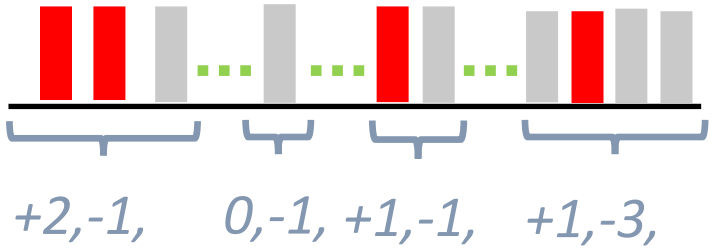
=

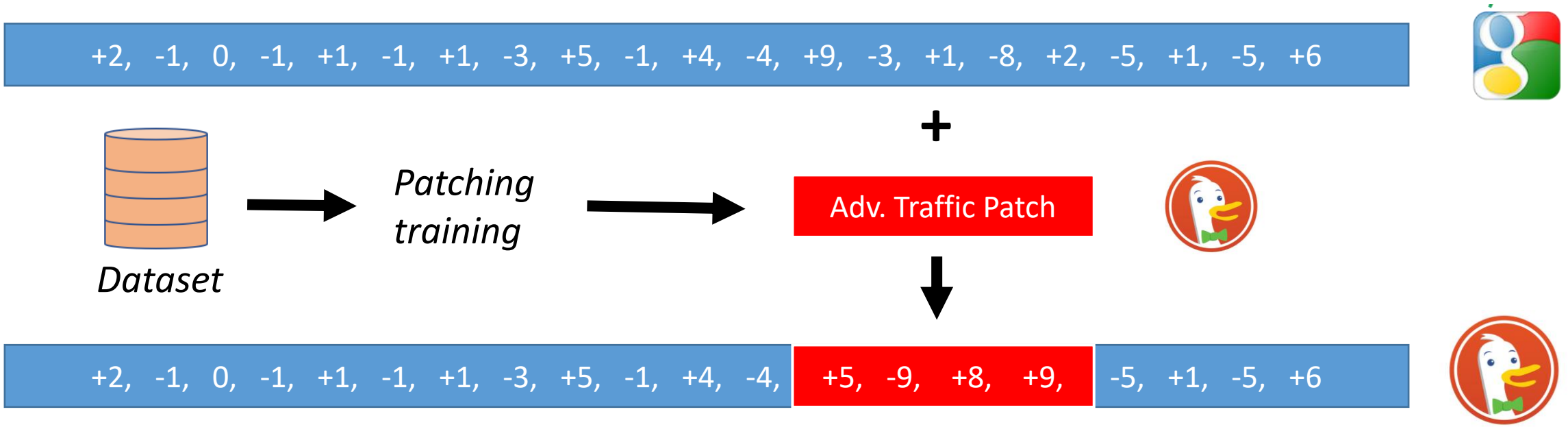
Baseball
(90% prediction)

Karmon et al. LaVAN: Localized and Visible Adversarial Noise, ICML 2018.

Ongoing-work: Adversarial Patches

Burst-based representation





WF Defense: Open Questions

- **How much defense is enough?**
- **Defending against future, unknown attack types.**

Thanks for listening!

Questions?