

# Machine Learning-Based Fingerprinting of Network Traffic Using Programmable Forwarding Engines

Greg Cusack, Oliver Michel, Eric Keller

## Goal

- Explore the efficacy of classifying large amounts of network traffic using PFE-generated, rich flow records in two separate applications
- Ransomware identification and classification
- Censorship circumvention traffic fingerprinting

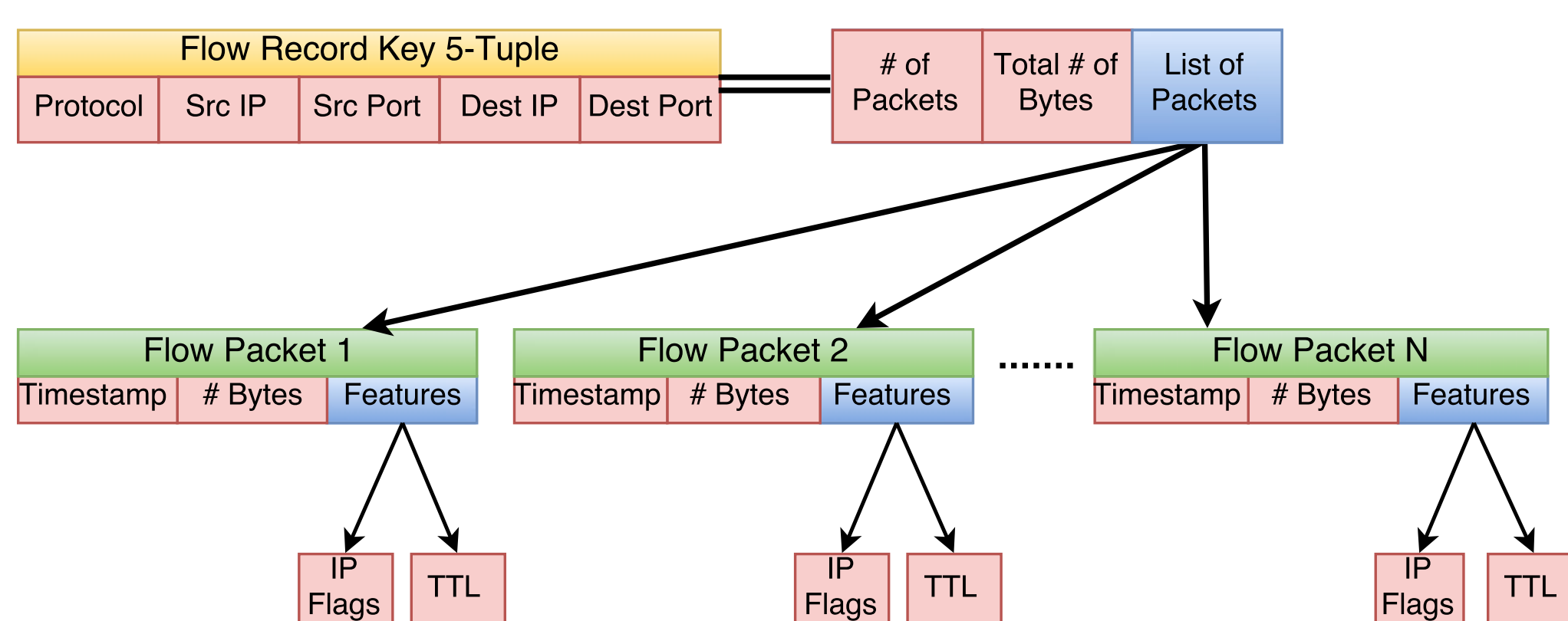
## Programmable Forwarding Engines (PFEs)

- Allow commodity network equipment to support the scalable generation of rich flow records
- Stream processing systems utilize PFE switch hardware to process network data at high-rates of speed and extract vital, per-packet flow information
- Provide system designers with the data and speed necessary for network, flow-based traffic analysis and fingerprinting

## Compact, per-packet flow records

### PFE flow record overview

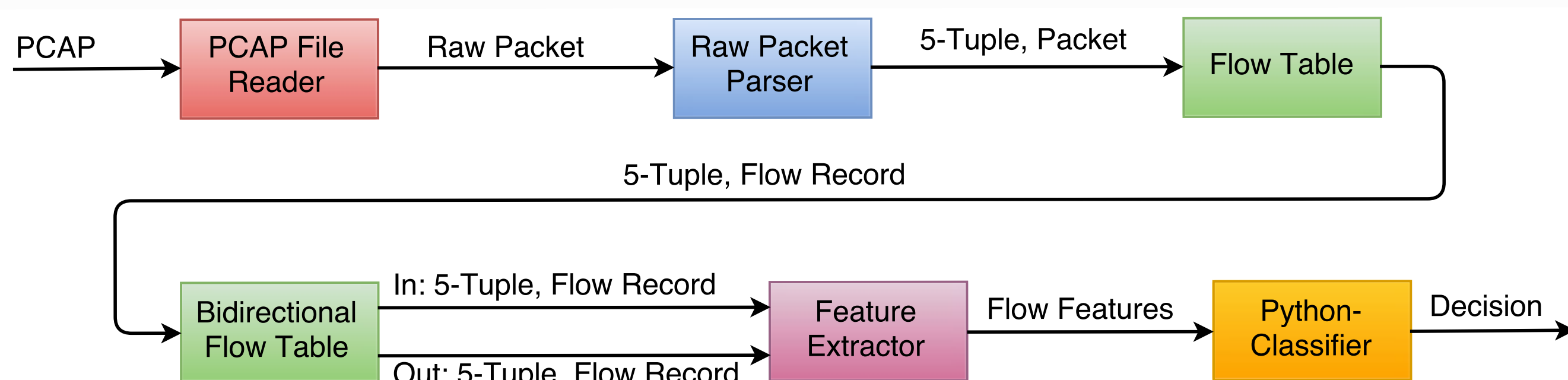
- The data extracted from a flow can be tailored to fit a user's specific application



## Flow Records and Processing

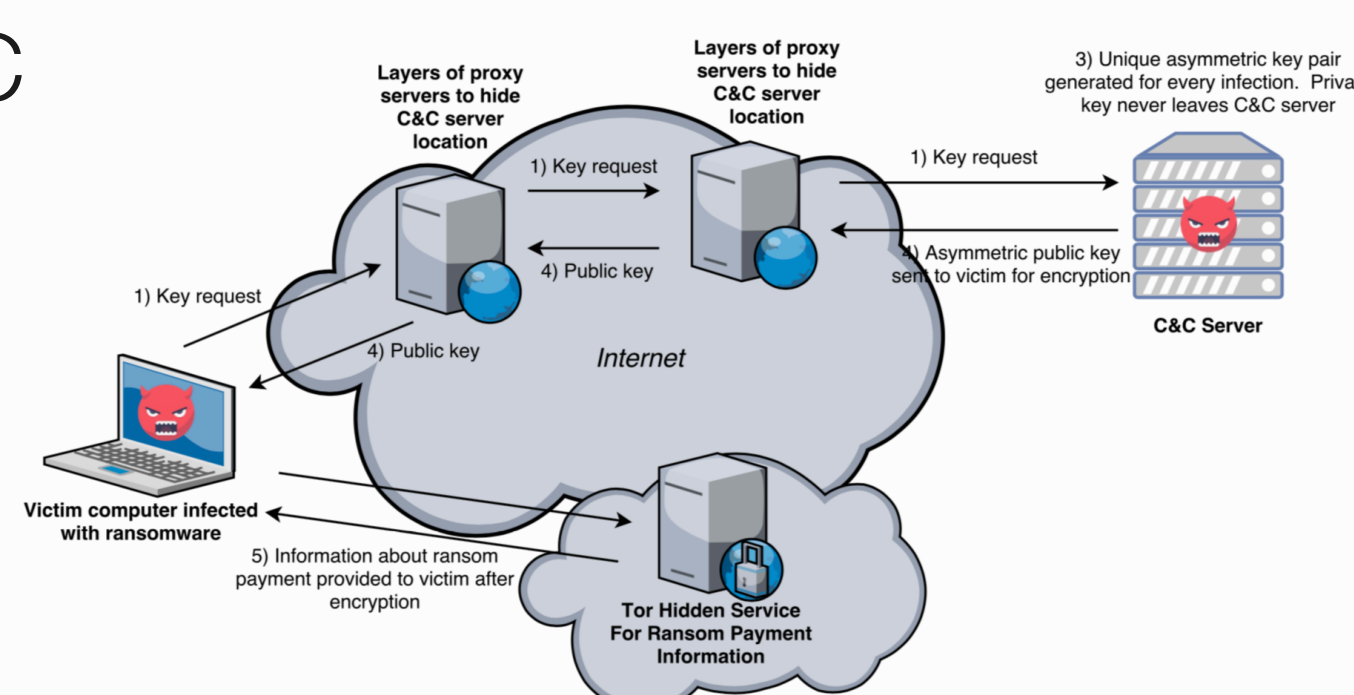
### Stream processor

- 5 kernel stream processor
- Simulates PFE-generated compact, rich flow records
- Extracts vital features from flow records
- Feeds into Python classifier



## Ransomware Overview

- Victim makes initial key request to C&C server
- C&C server returns encryption key
- Tor hidden service communicates method of payment



[Cabaj, Mazurczyk. CoRR '16]

## Ransomware Classification Results

### Goal

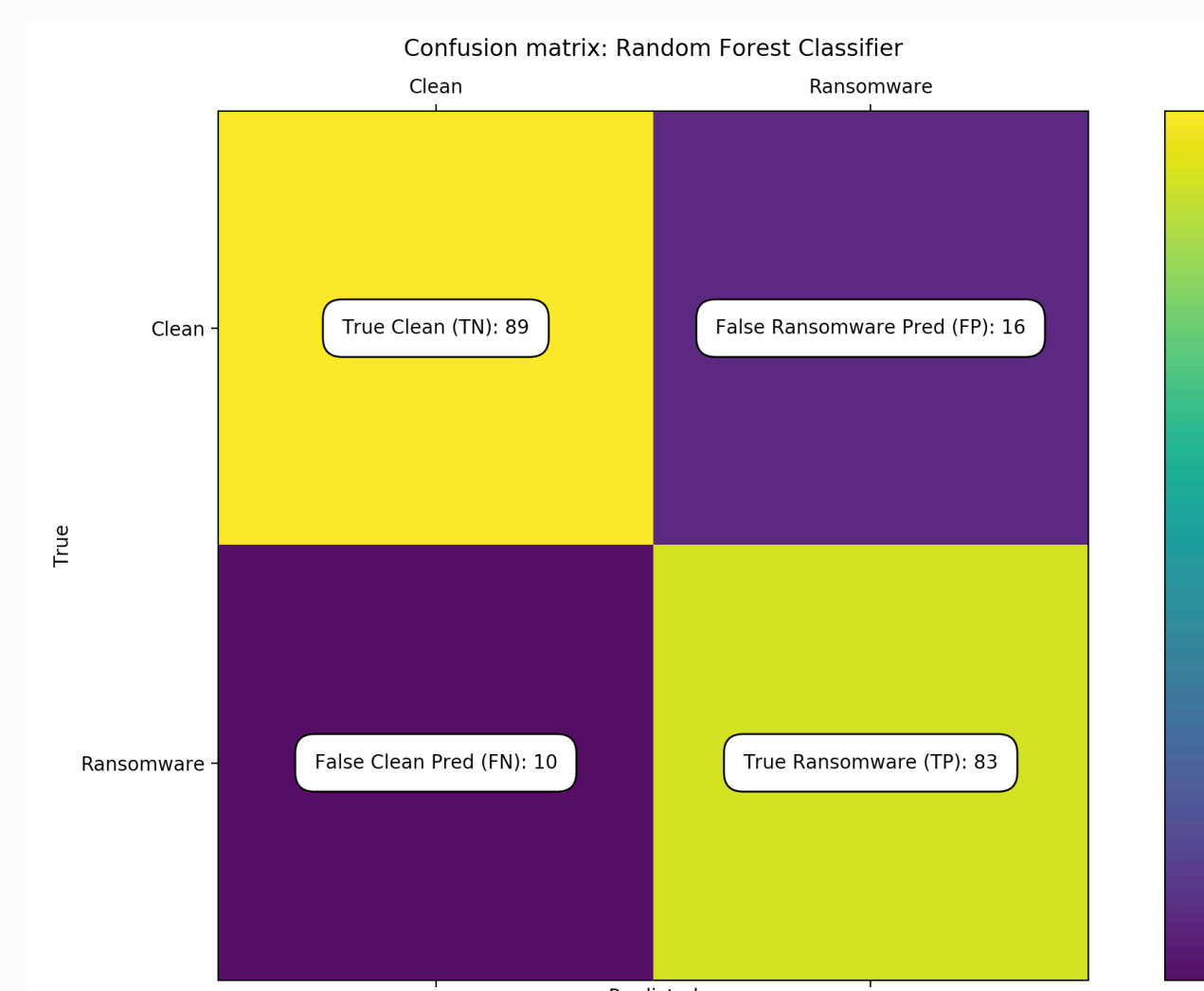
- Minimize false negatives
- Balance FNR and FPR

### Random Forest Classifier

- 40 decision trees with depth 15

### Performance

- Precision: 0.89
- Recall: 0.83
- F1 Score: 0.87
- AUC of ROC: 0.93



### Key Flow Features

- Packet interarrival times
- Inflow to outflow packet ratios
- Burst lengths
- Flow duration

## Shadowsocks Classification Results

### Goal

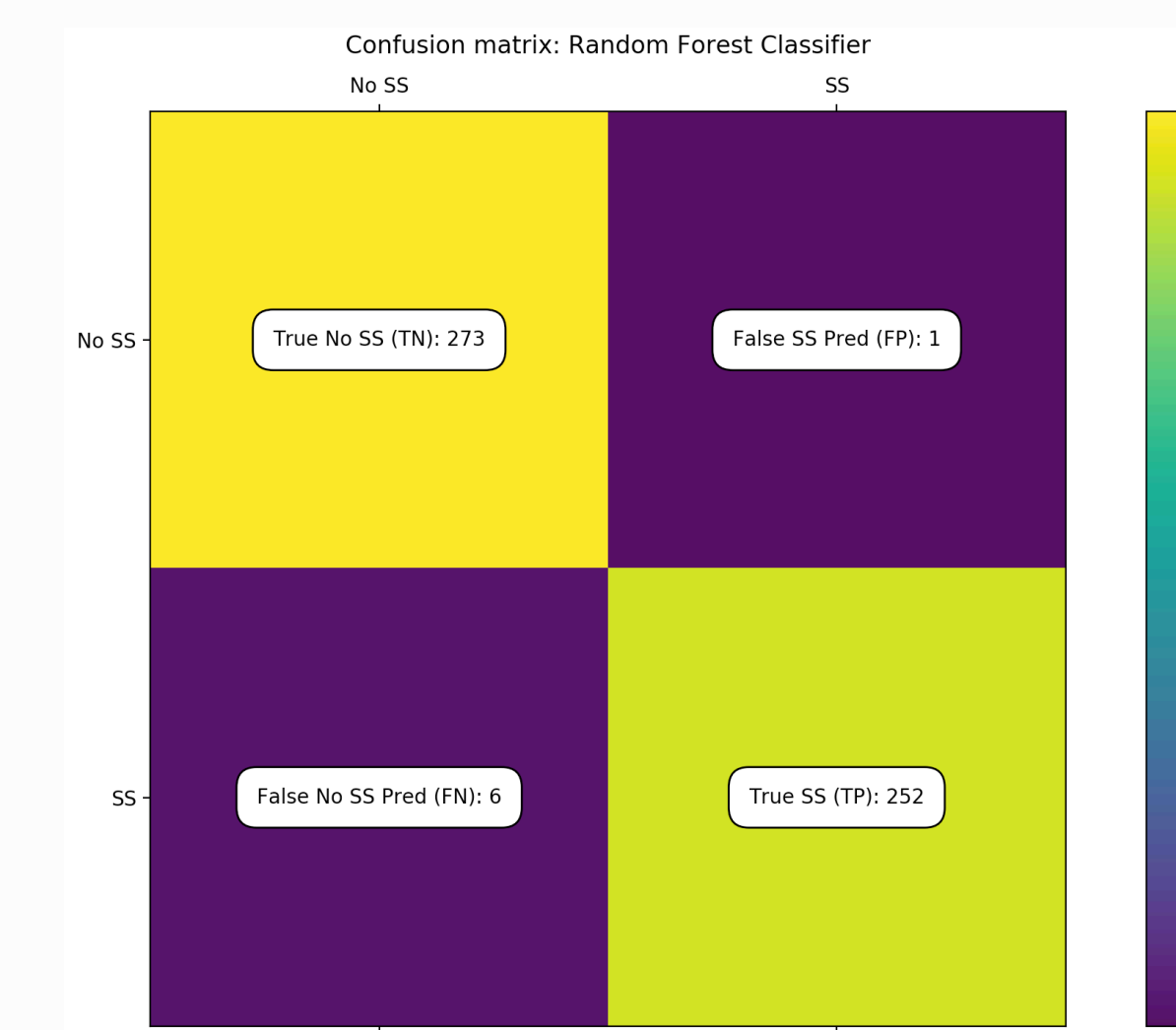
- Minimize false positives

### Random Forest Classifier

- 10 decision trees with depth 10

### Performance

- Precision: 0.996
- Recall: 0.977
- F1 Score: 0.987
- AUC of ROC: 0.999



### Key Flow Features

- Packet interarrival times
- Traffic latency
- Payload entropy

## Discussion

### Takeaway

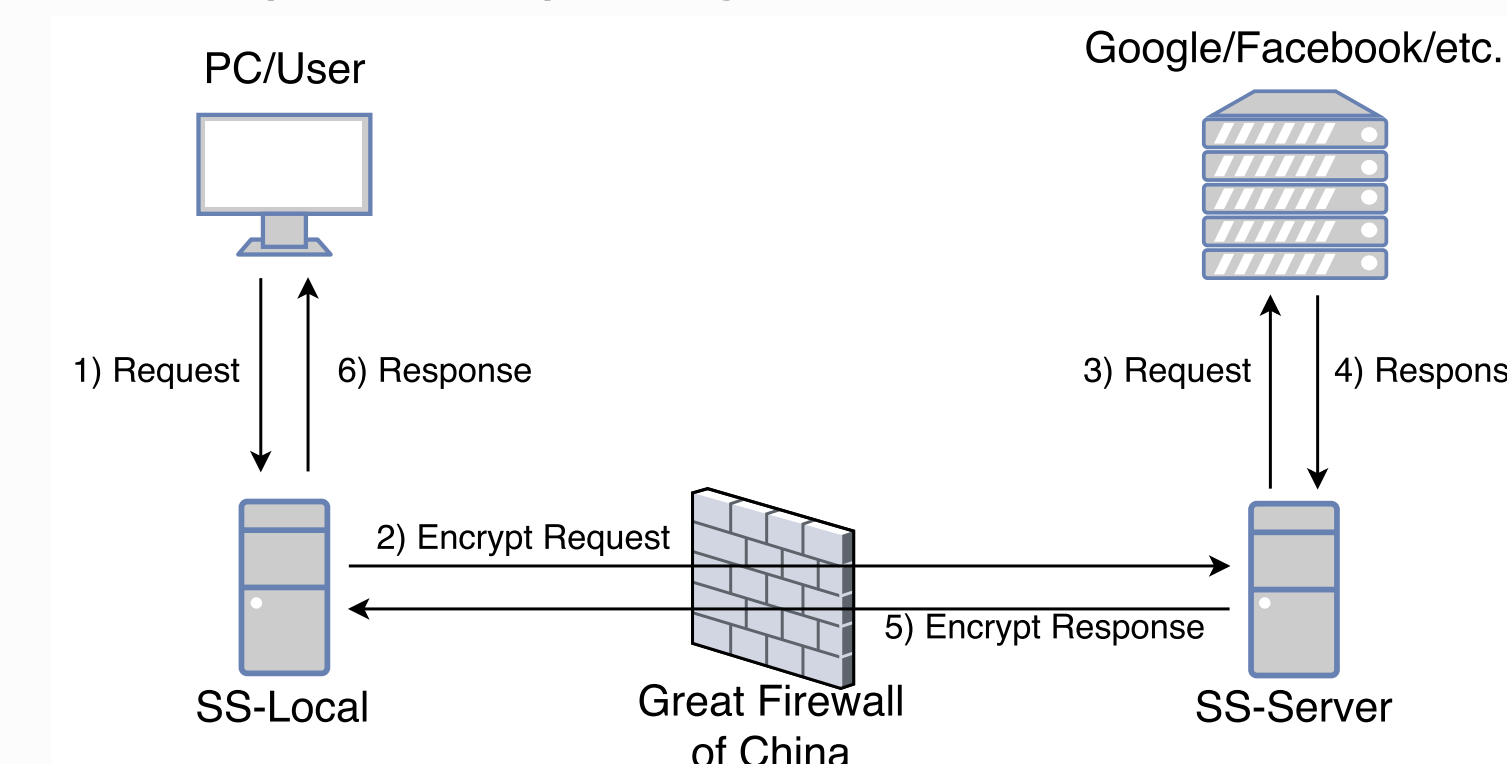
- Preliminary results show efficacy of utilizing high-rate PFE-generated, rich flow records to fingerprint different types of web traffic

### Future Work

- Write classifiers in C++ for line rate classification
- Continue to explore other classification techniques

## Shadowsocks (SS) Overview

- Censorship circumvention tool
- User sets up own proxy server outside GFW domain



## Acknowledgements

This work was supported in part by the NSF grants 1652698 (CAREER) and 1406192 (SaTC), and by the NSF and VMware grant 1700527 (SDI-CSCS).