Attack Reducation and Anomaly Modeling in Popularly Targeted Protocols

Michael Collins, CERT/NetSA

Talk outline

The Problem

- Noise in traffic flows
- Impact on anomaly detection

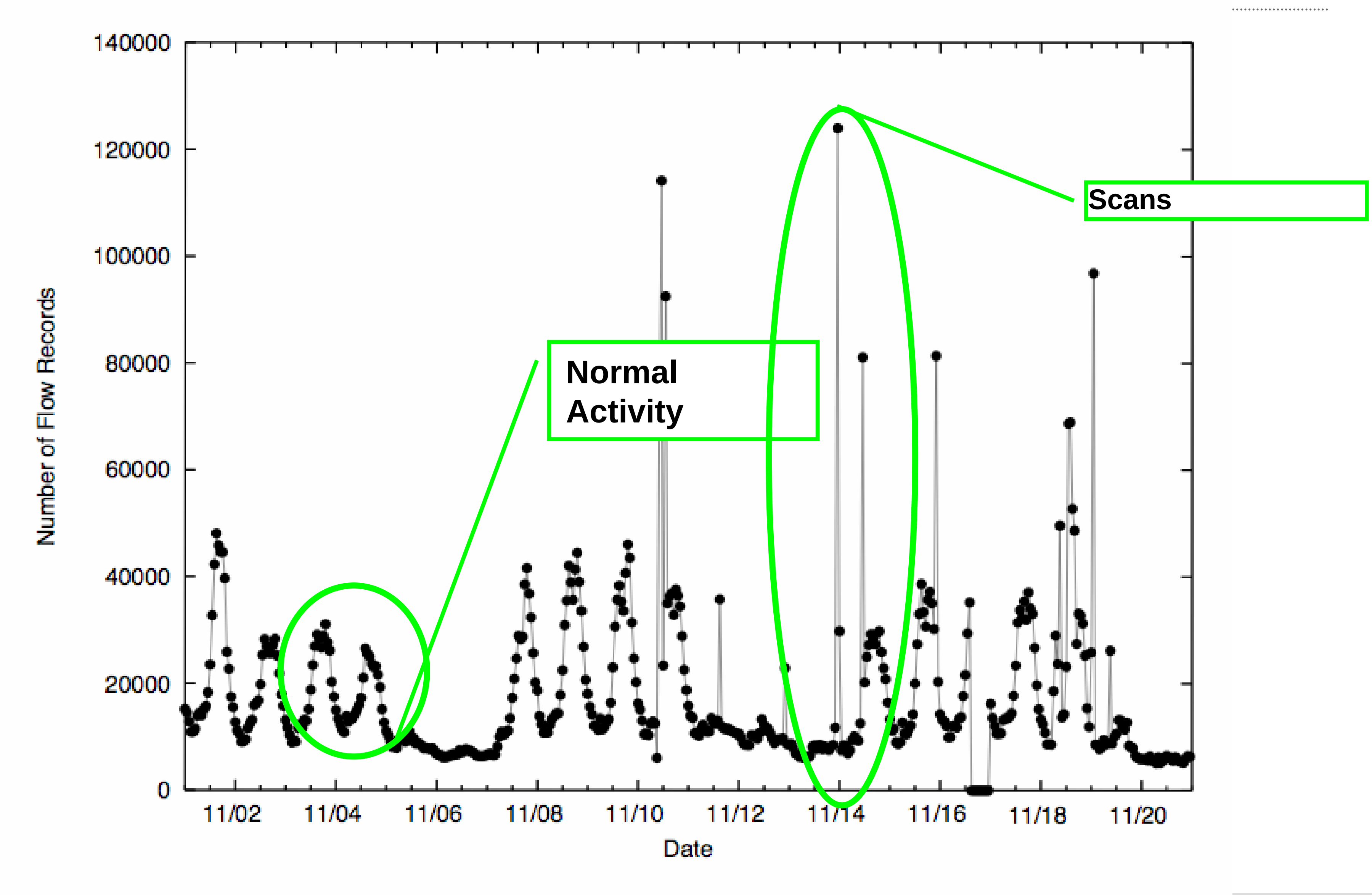
Two Stage Filtering

- Log Filtering
- State Filtering

Attack reduction

- Core assumptions
- Method for data removal
- Impact

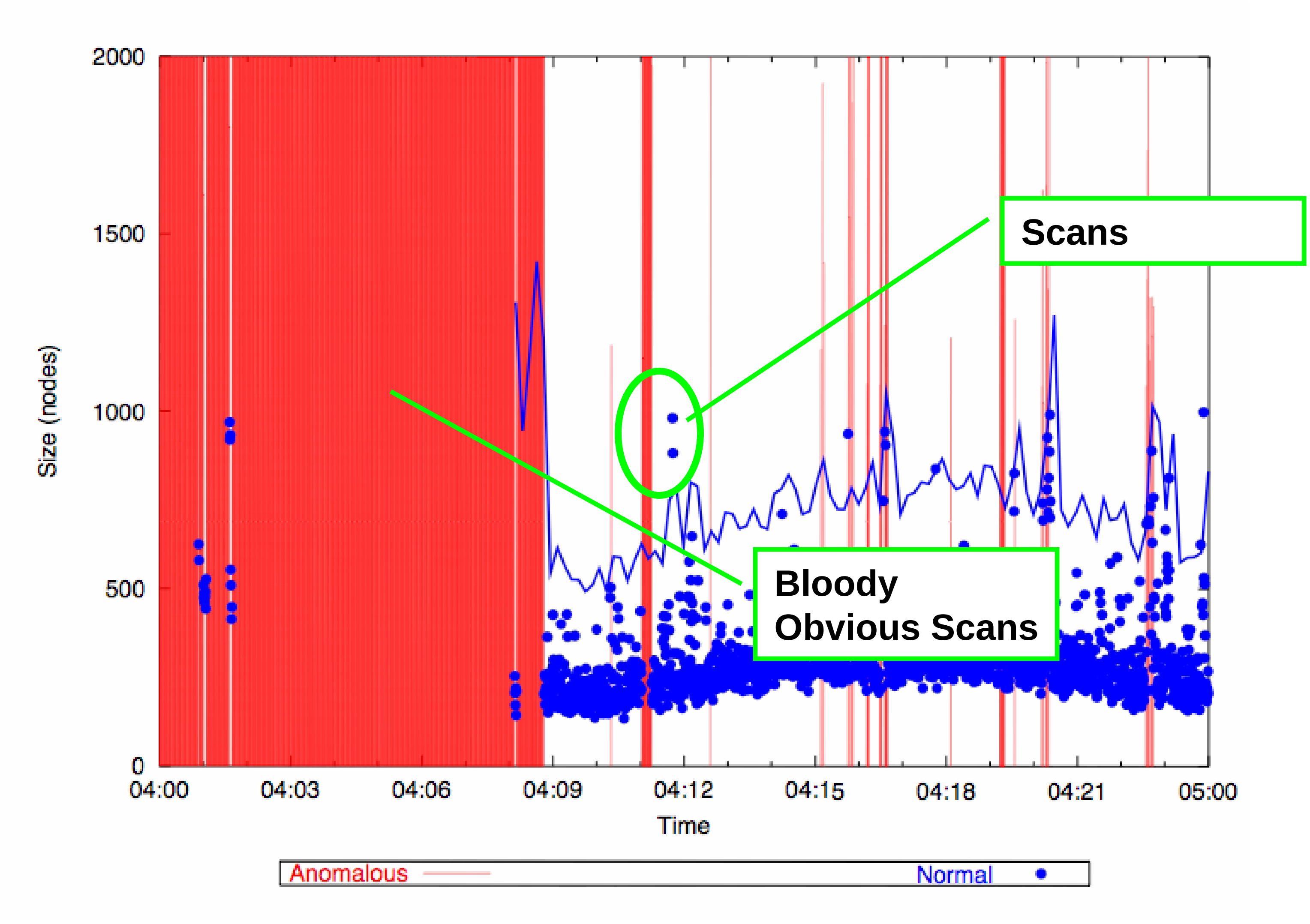
Innocuous Attacks



Normal Activity

Scans

Normal SSH Activity



Scans

Bloody Obvious Scans

Raw SSH Data

A Hypothesis

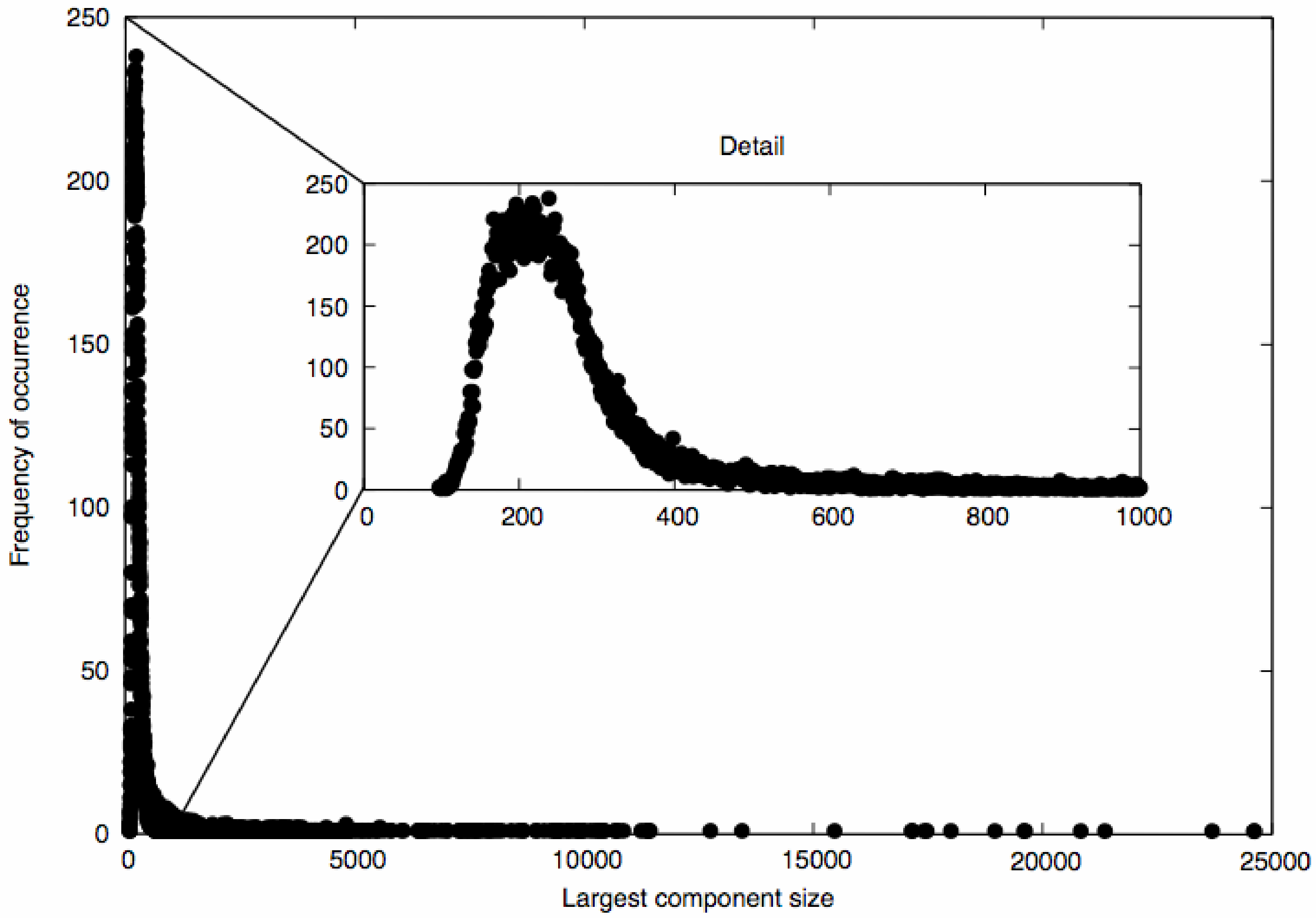
We see two populations:

- Normal users, who know where they’re going
- Attackers, primarily scanners, who have no idea about the network’s structure

The majority of attackers are clumsy

- Low success rates
- Picking targets effectively at random
- Pick many more targets than there are actual targets
  - >350,000 per 30s period, vs. ~ 10,000 real targets

Comparing the two populations



Impact on anomaly detection

Almost every anomaly detection system requires advance knowledge

- Mean, standard deviations
- Map of known servers

This information may not be easily acquired

- Inventory is nontrivial
- Going by the data can lead to false positives from attackers

We need to train the system while acknowledging the hostility

Filtering: Log Filtering

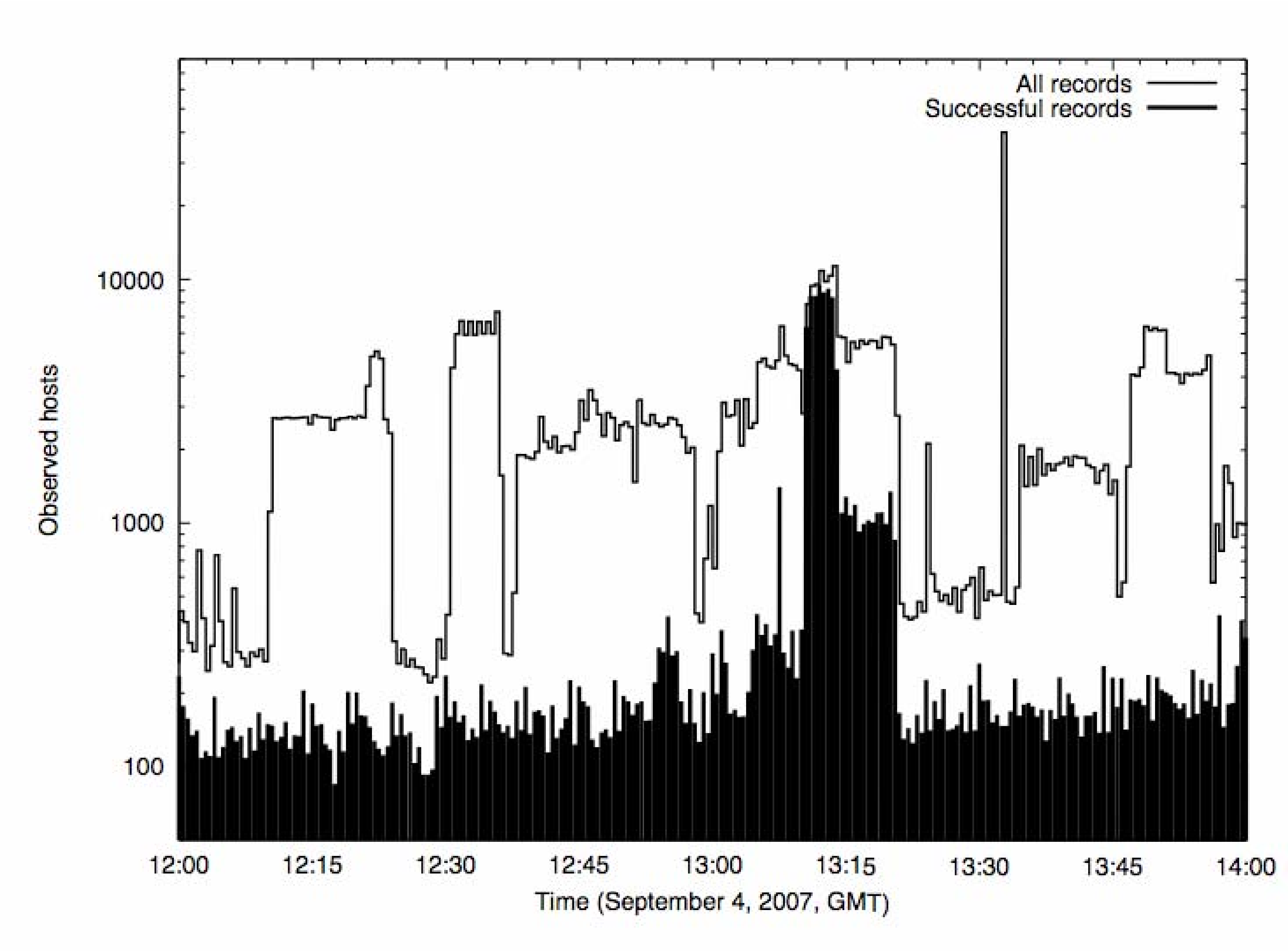
At least with TCP data, we can rely with the state machine

- ≤ 3 packets implies it is most likely a scan
- > 3 packets may be legitimate

In a two week ssh dataset:

- ≤ 3 packets make up 87% of the flows
- ≤3 packets make up 1% of total bandwidth

Log Filtering is Insufficient



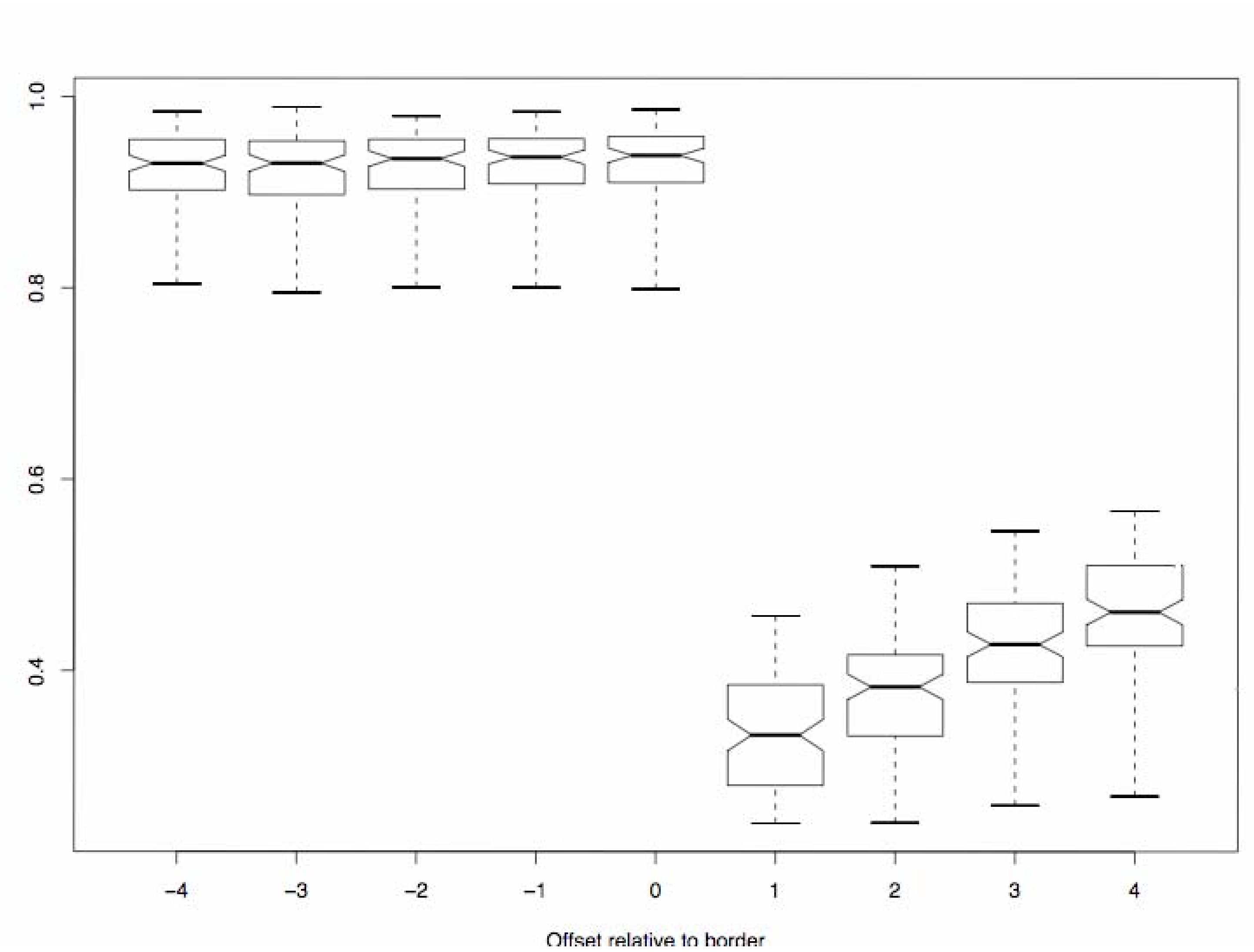
State Filtering

If we assume activity is Gaussian, then we can identify and eliminate outliers

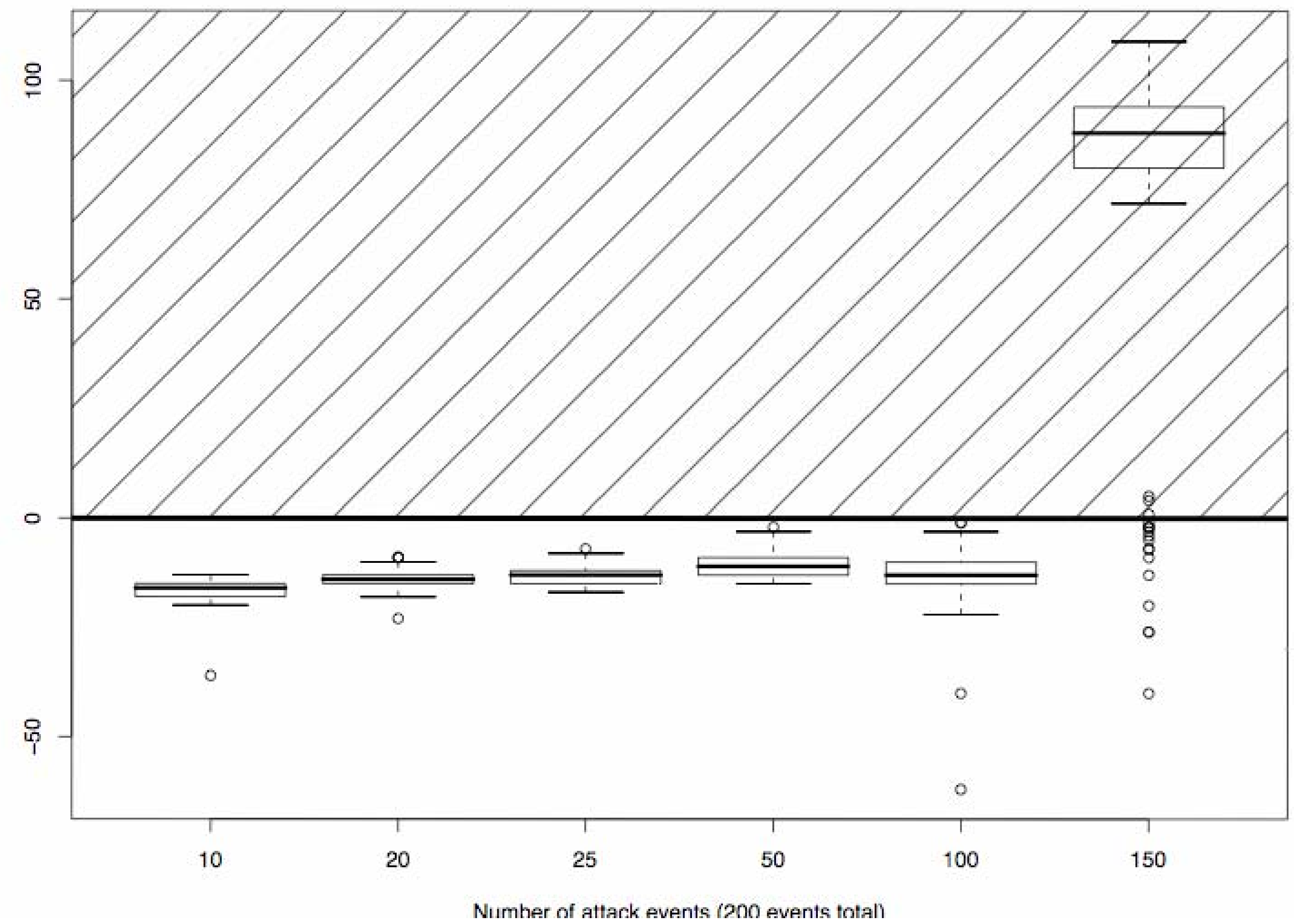
Simple test: Shapiro-Wilk test for normalcy

- God for 25-2000 samples
- Doesn’t require an estimate of mean or standard deviation

Very coarse…



How many attacks can we stand?



Conclusions

Constant noise is managable

- But it requires integrating multiple filtering mechanisms
- It also means assuming a certain mode of behavior
  - This method assumes gaussian, other tests are available

Open questions:

- What do we do with scans once we know they're there?