Bridging the Data Gap: *Data Related Challenges in Evaluating Large Scale Collaborative Security Systems**

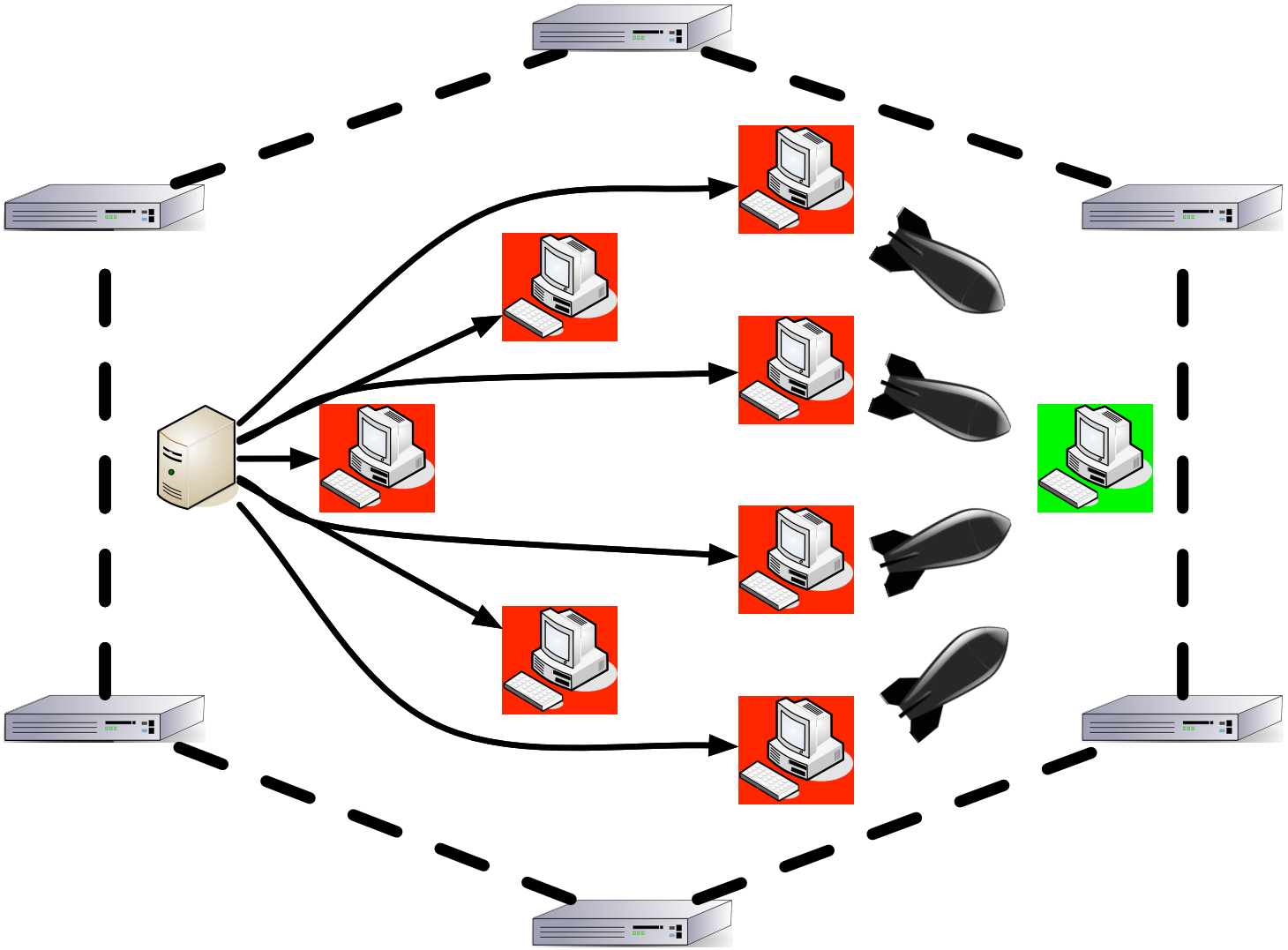
John Sonchack[1], Adam J. Aviv[2], Jonathan M. Smith[1]

[1] University of Pennsylvania

[2] USNA

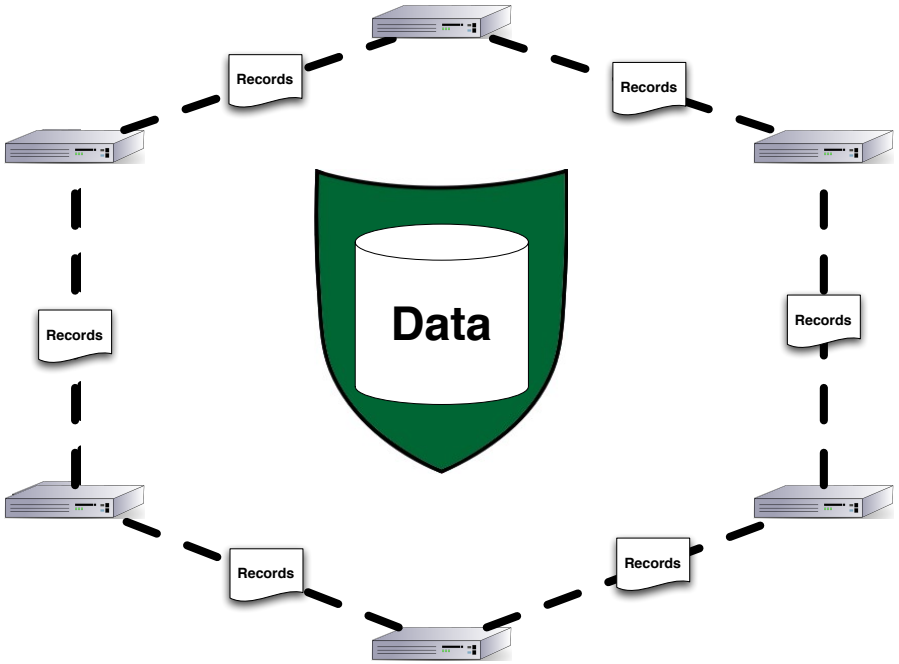
*This work was supported by ONR Grant N00014-12-1-0757

Collaborative Cyber Security Systems



Proposed Collaborative Security Systems

- Highly Predictive Blacklisting
- Autograph
- Internet Scale Anomaly Detection



Obtaining Evaluation Data

- Collect live traffic
- Challenges: privacy, legality, ethics

Policy

- Repositories and public data sets
- Challenges: anonymization, filtering

Policy

MIT Lincoln Laboratory

PREDICT
Protected Repository for the Defense of Infrastructure Against Cyber Threats

caida

- Simulation tools
- Challenges: configuration, validation, realism

Design

emulab
total network testbed

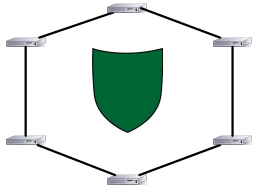
Mininet

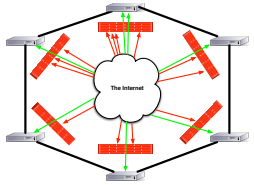
vm

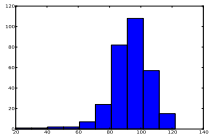
ns-3
NETWORK SIMULATOR

Outline

- Introduction
- Data and Experimental Ideals
- Case Studies
- The promise of simulation
- Conclusion



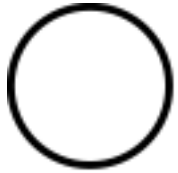
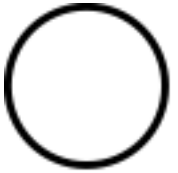
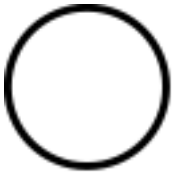




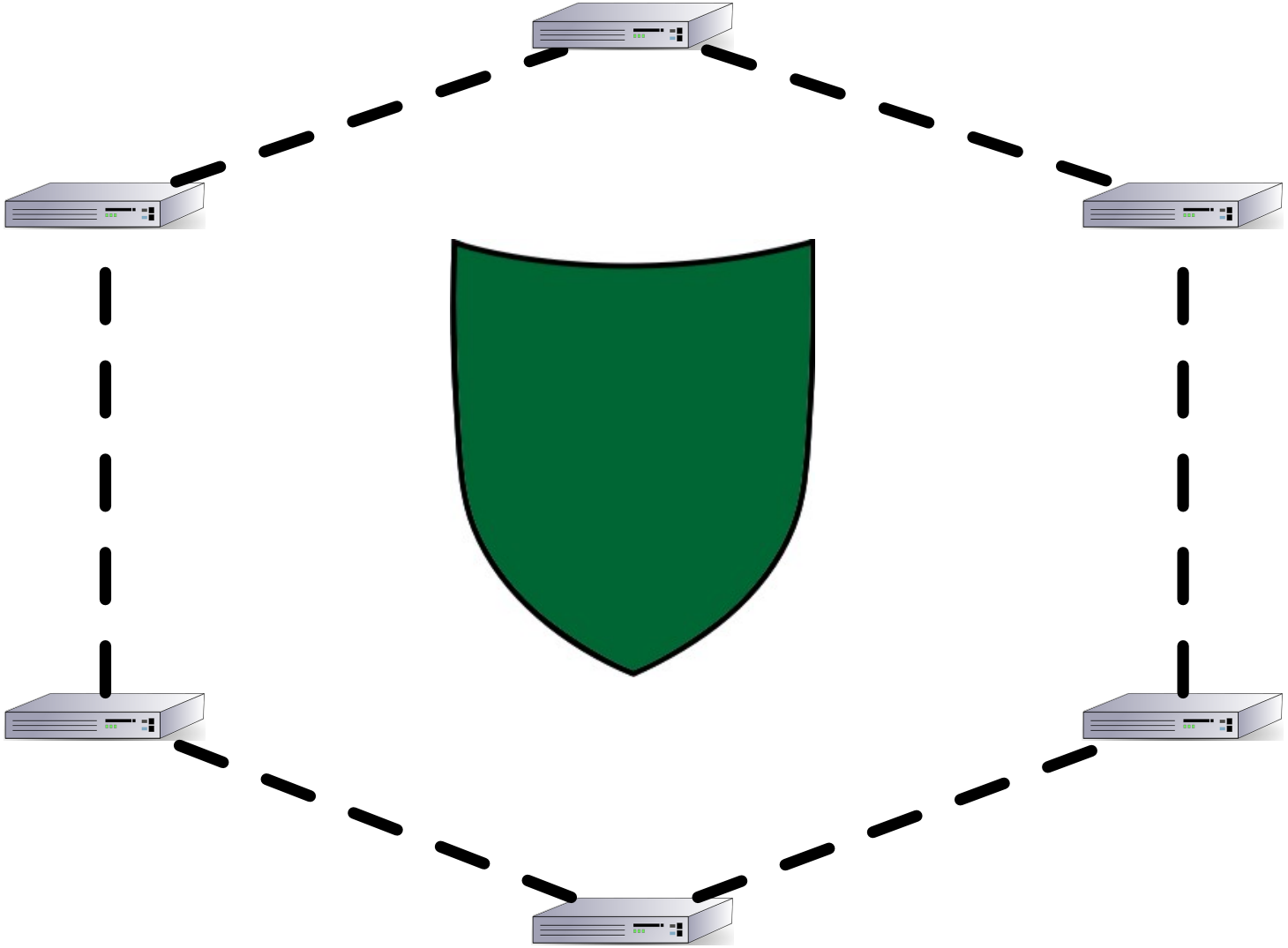
Data and Experimental Ideals

- Reproducibility
  - Build on previous work
- Experimental Control
  - Evaluate the effects of factors
- Ground Truth
  - Measure accuracy
- Evaluation at Scale
  - Large scale systems

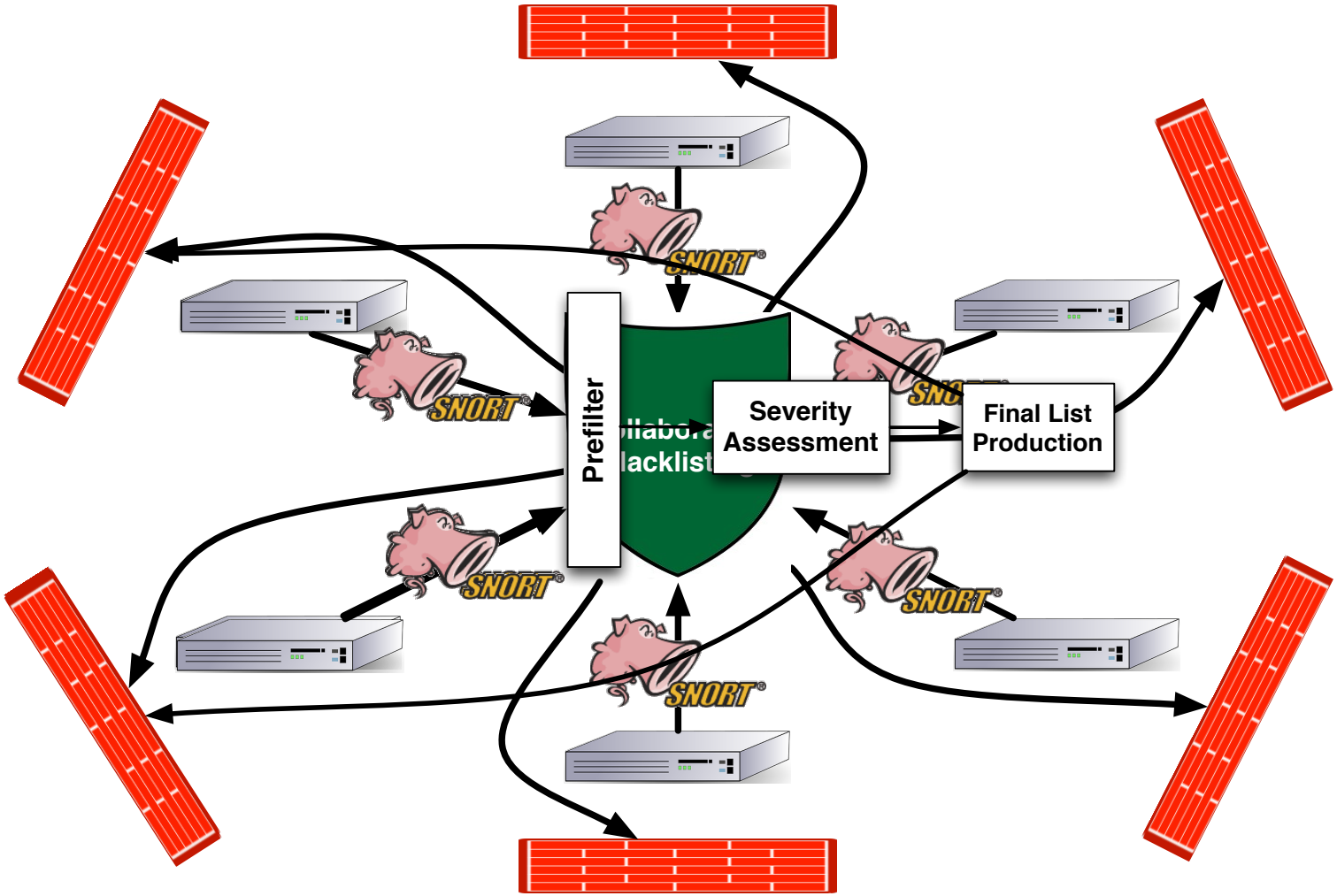
Data and Experimental Ideals

	Reproducibility	Experimental Control	Ground Truth	Evaluation at Scale
Live Data				
Repositories				
Simulation				

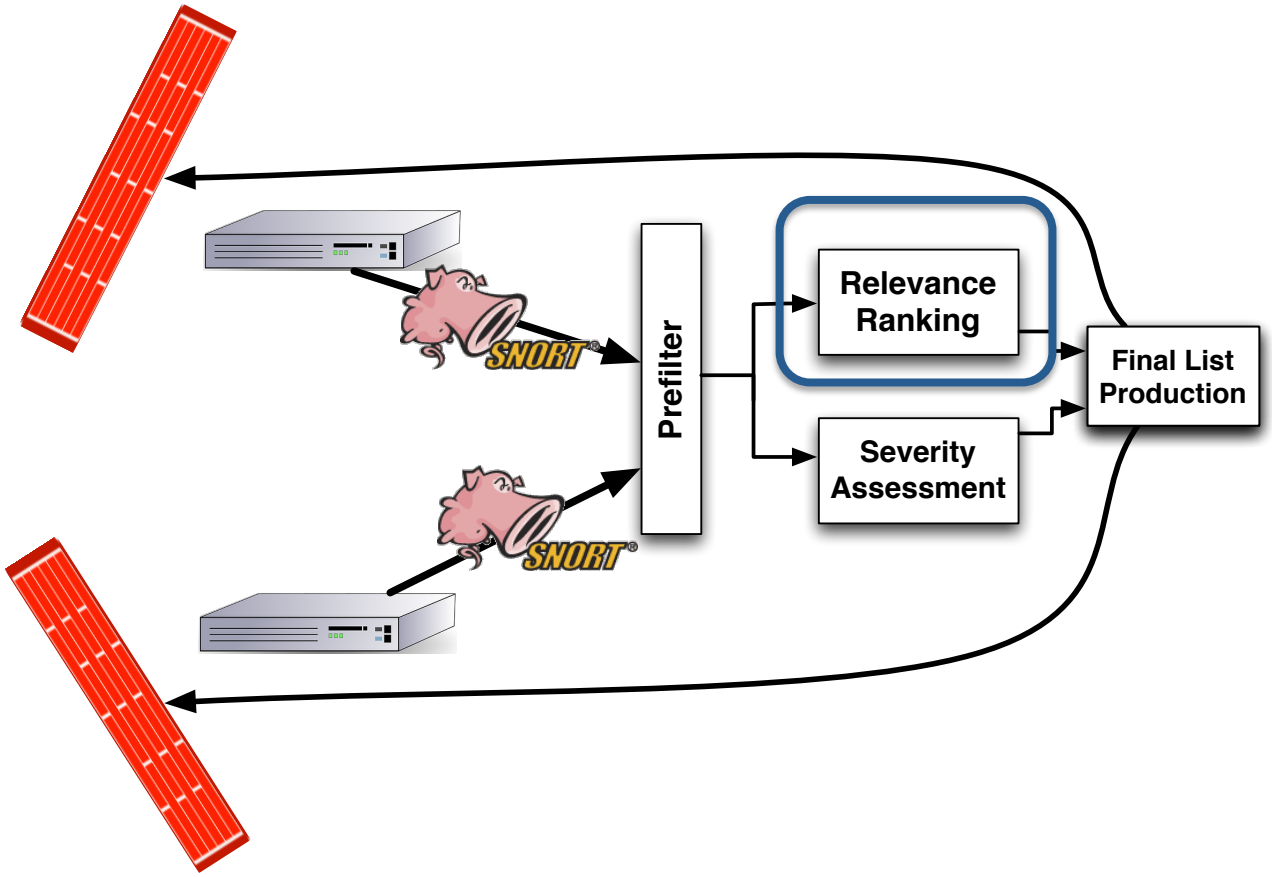
Collaborative Security Case Studies



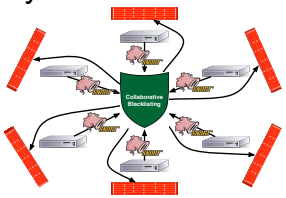
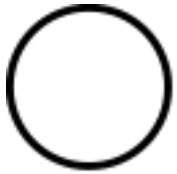
Highly Predictive Blacklisting

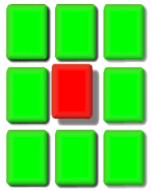


Highly Predictive Blacklisting



Highly Predictive Blacklisting

Evaluation	Reproducibility	Experimental Control	Ground Truth	Evaluation at Scale
Highly Predictive Blacklisting 				

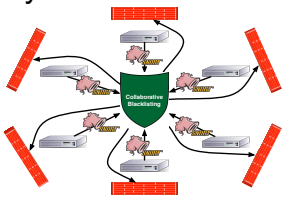
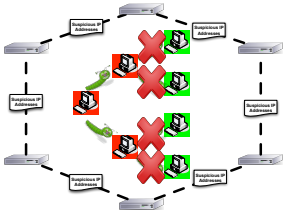
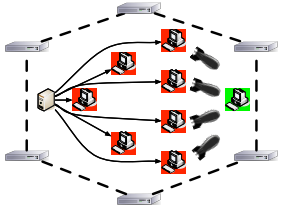


- Source: DShield.org repository
- Alerts: >15 million

- Networks: >1000

- IP addresses: >10,000

Data and Experimental Ideals *in practice*

Evaluation	Reproducibility	Experimental Control	Ground Truth	Evaluation at Scale
Highly Predictive Blacklisting 	●	◑	○	●
Autograph 	◑	◑	◑	◑
Internet Scale Detection 	○	◑	●	◑

The Promise of Simulation

Simulation

emulab total network testbed

Mininet vm

ns-3 NETWORK SIMULATOR

Existing Simulation Tools

Traffic Generators

D-ITG

MGEN

Swing

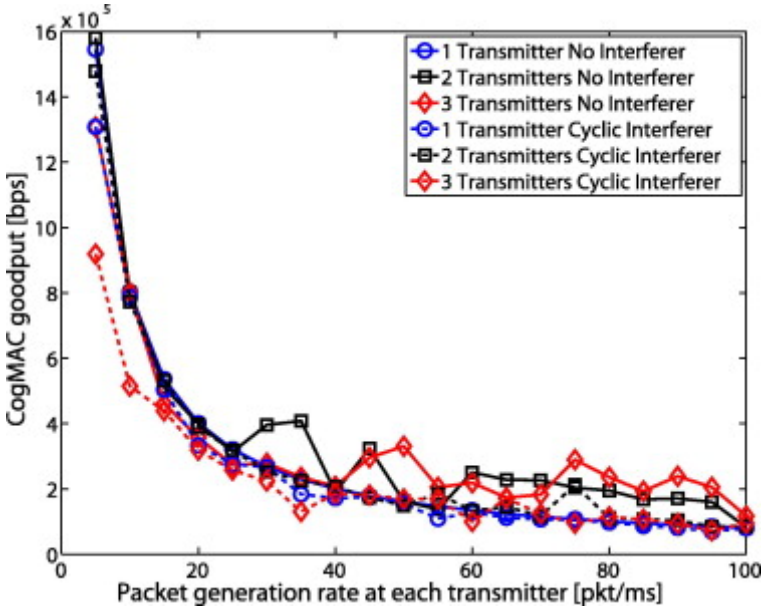
Virtual Testbeds

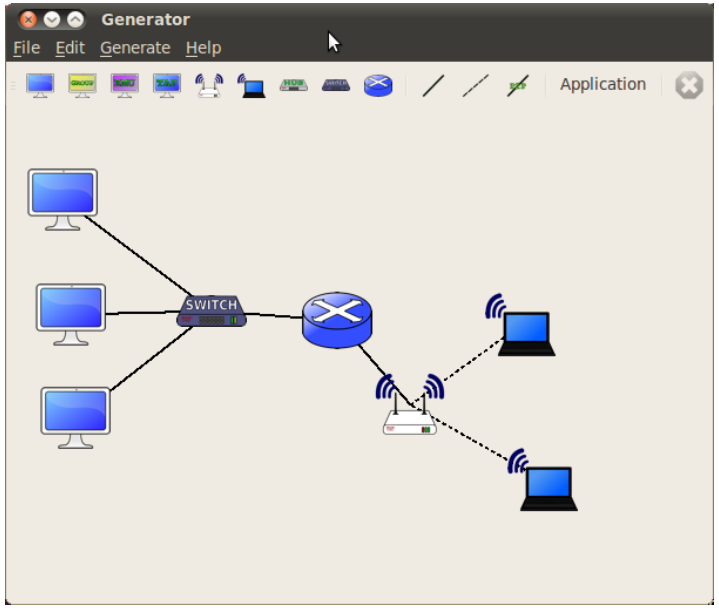
emulab
total network testbed

ns-3
NETWORK SIMULATOR

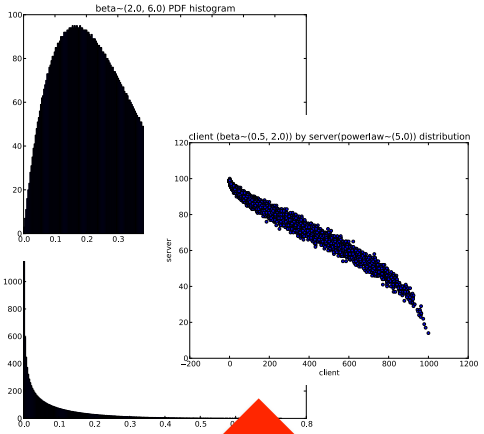
Mininet

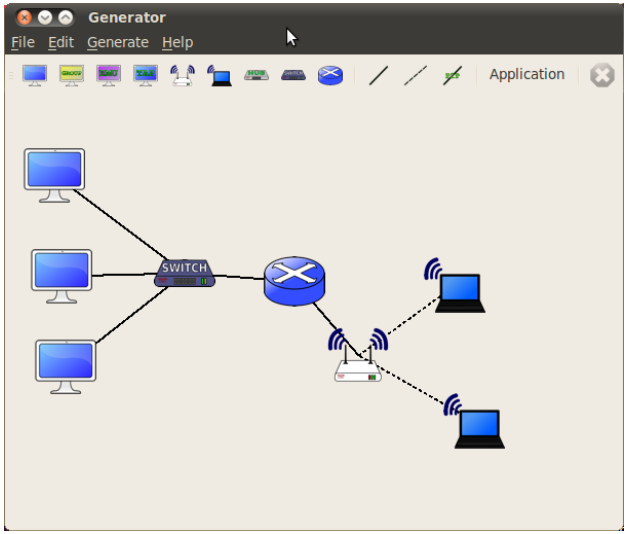
vm

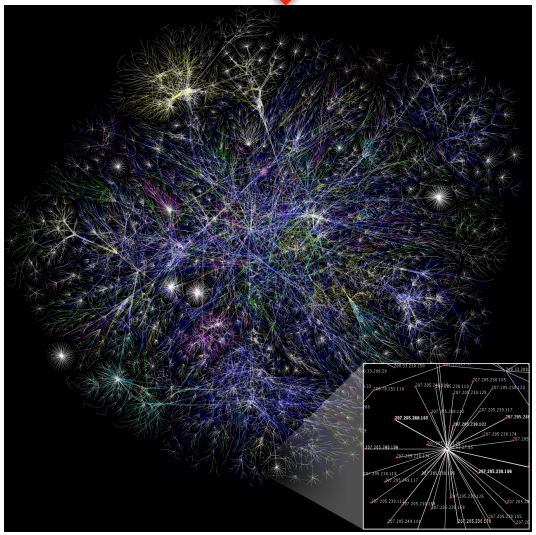




Designing Simulators for Collaborative Security Systems

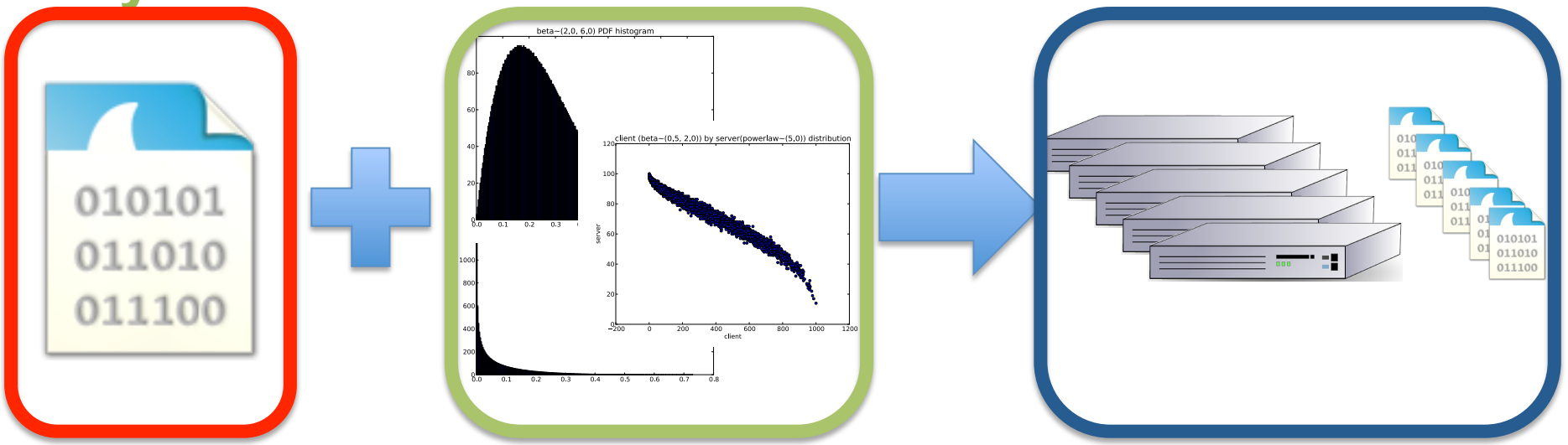




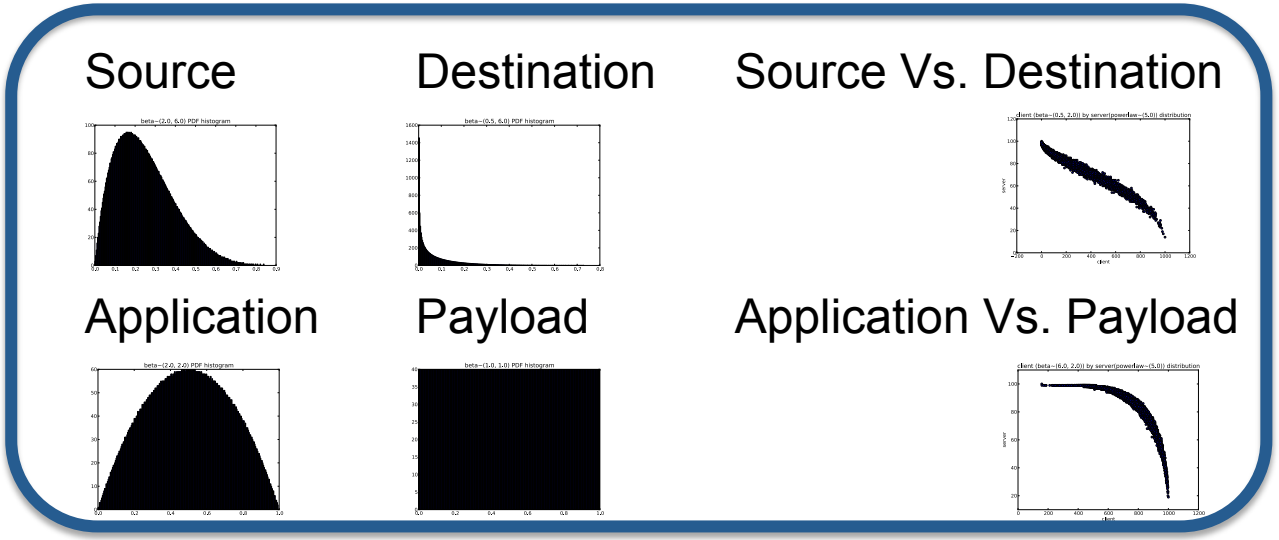


Parameterized Trace Scaling

Parameterized Trace Scaling extracts flow payloads from a **small scale input trace**, and then replays the flows between simulated hosts in an event driven **large scale simulation**, where the **events are generated using statistical models of network factors relevant to collaborative security systems**.

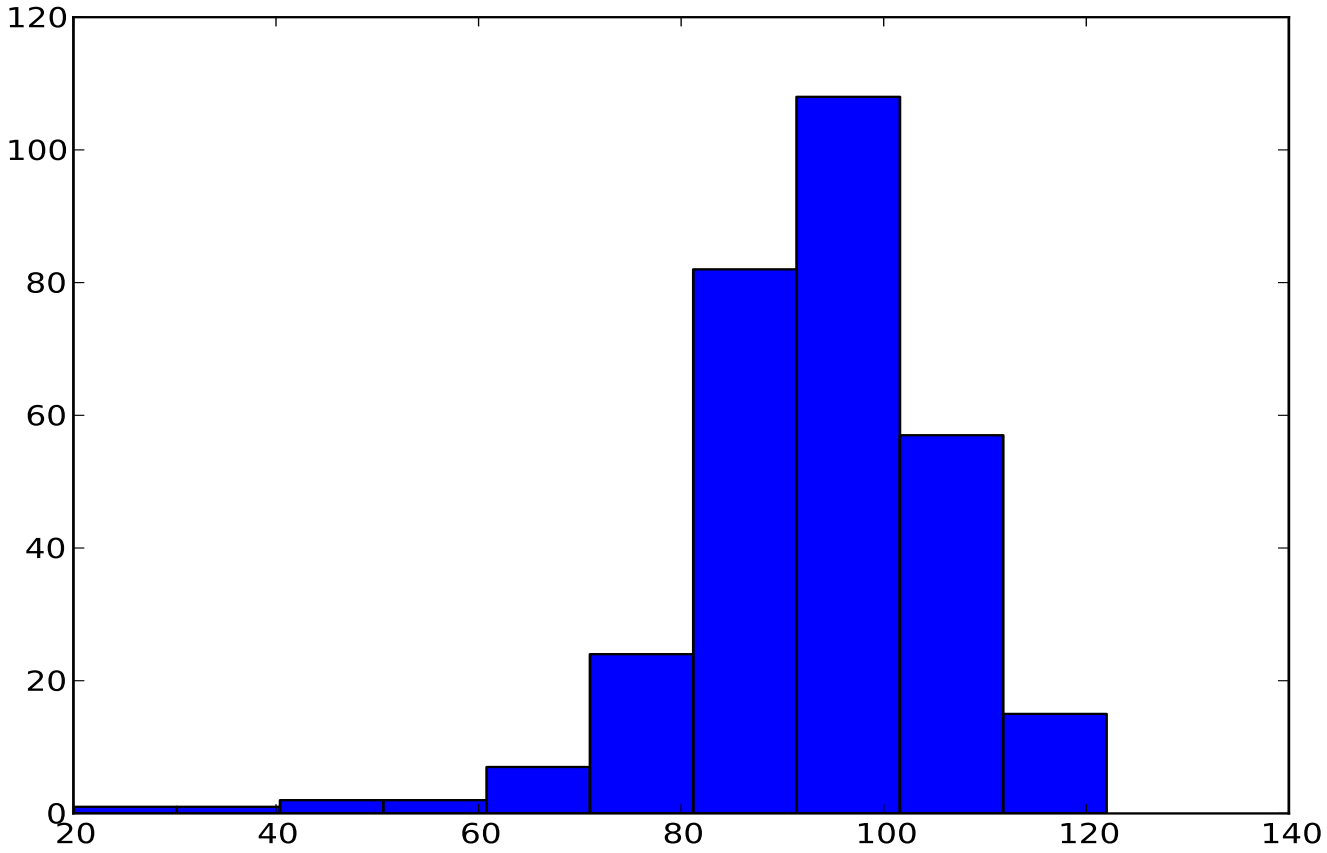


Input Distributions and Parameters



Parameterized Trace Scaling

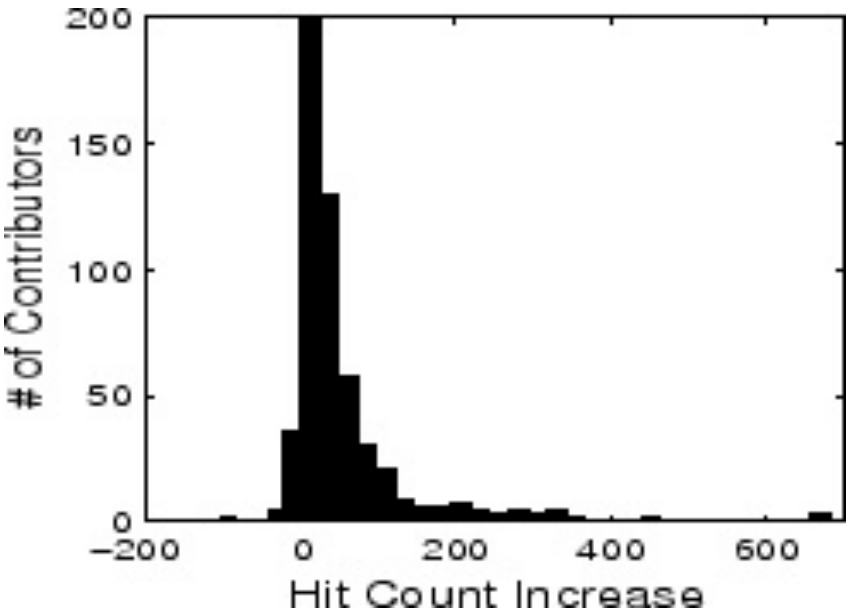
Preliminary Results (**Not** in the paper)



Preliminary Results

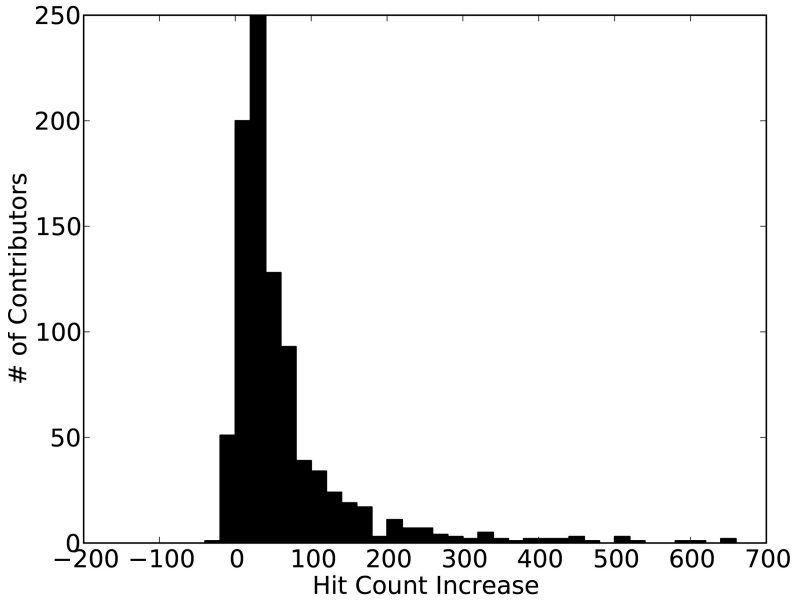
Predictive Blacklisting Evaluation

- 15 million alert logs collected from DShield



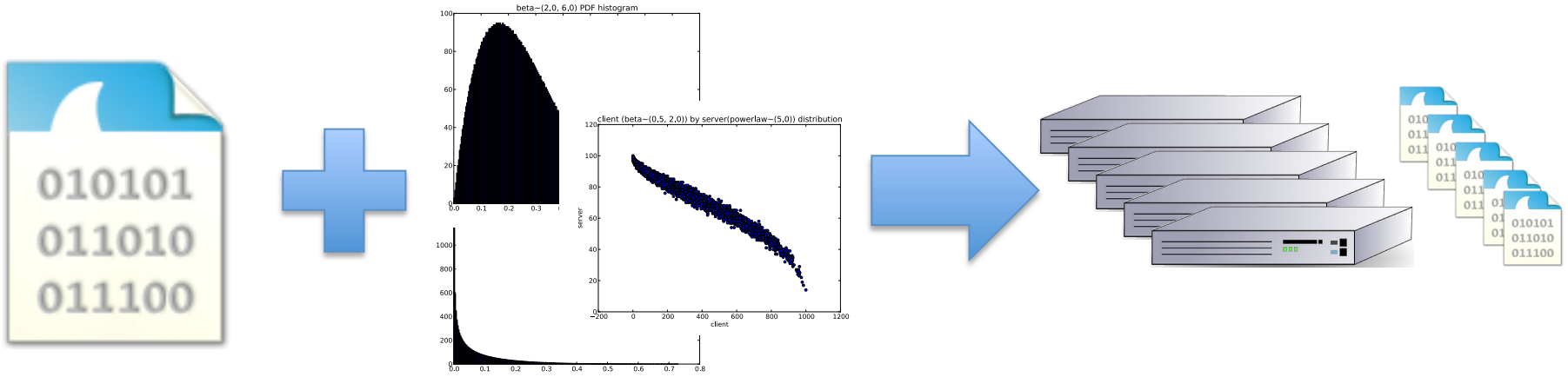
Parameterized Trace Scaling Re-evaluation

- single network, openly available trace (scaled up)



Parameterized Trace Scaling and Overcoming Data Challenges

Original Evaluations	Reproducibility	Experimental Control	Ground Truth	Evaluation at Scale
Highly Predictive Blacklisting	●	◐	○	●
Autograph	◐	◐	◐	◐
Internet Scale Detection	○	◐	●	◐
Parameterized Trace Scaling	●	●	●	◐



Conclusion

- Collaborative Techniques
  - Great potential for Internet Security
  - Many questions
- Data Sources
  - present challenges
  - affect experimentation
- Policy Vs. Design
- Parameterized Trace Scaling
  - A simulator designed for large scale collaborative security evaluations
- Questions?