TOWARDS INSTRUMENTING NETWORK WARFARE COMPETITIONS TO GENERATE LABELED DATASETS

Ben Sangster
TJ OConnor

Agenda

- Motivation
- What is the Cyber Defense Exercise?
- CDX ’09 Network
- Example Captured Attack Vector
- Logs from CDX ‘09
- Benefits to our Approach
- The Data
- Shortcomings to our Approach
- The Road Ahead
- Questions

Motivation

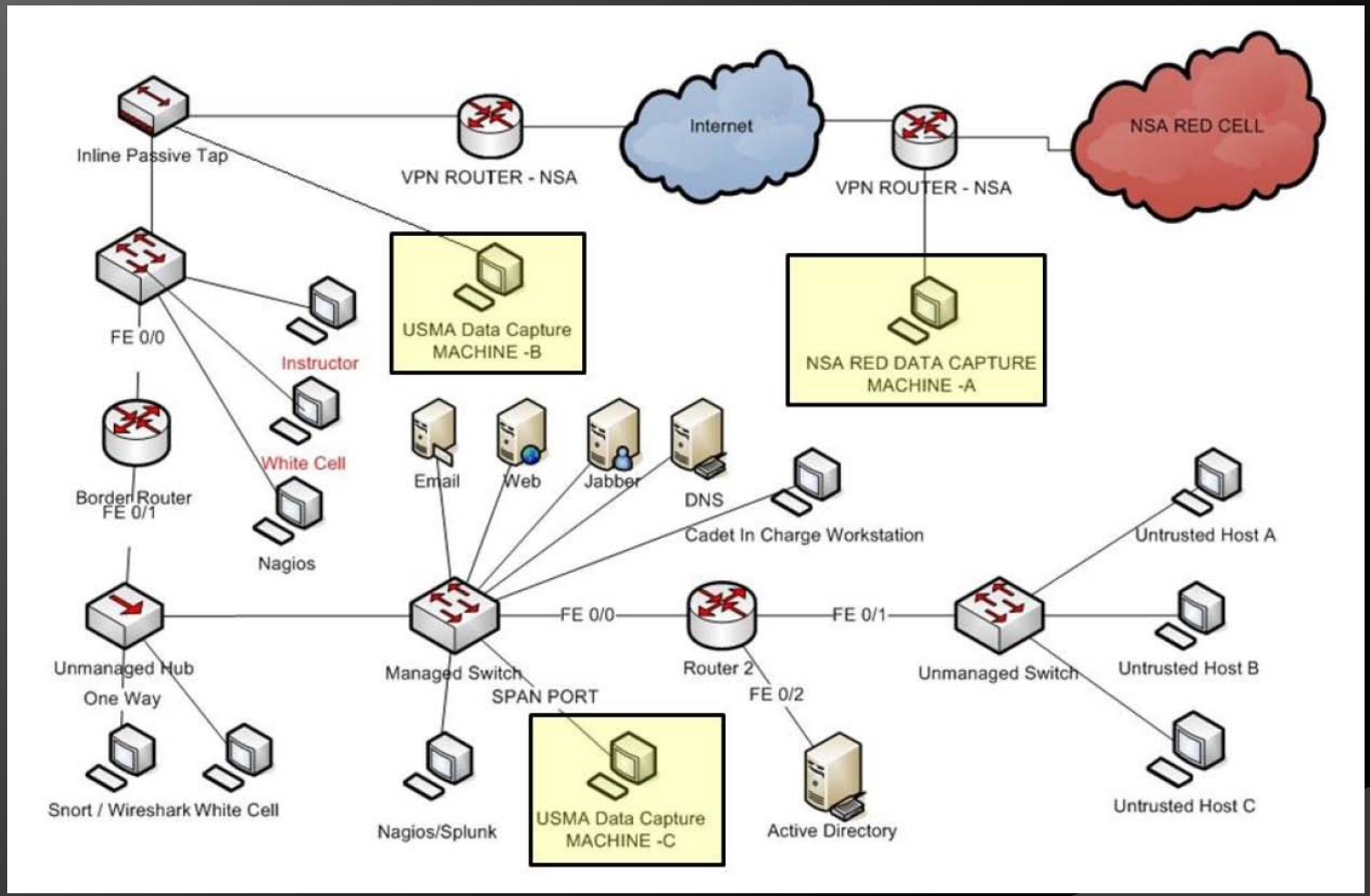
- Most commonly used datasets:
  - dated attacks
  - artificial in nature
  - trivial artifacts
  - don’t represent the true nature of an attacker
- Network warfare games:
  - zero day attacks
  - potential for scalability
  - human driven interaction

What is the Cyber Defense Exercise?

- Annual network warfare competition
- Numerous service academies (‘09 participants):
  - United State’s Military Academy at West Point
  - Naval Post Graduate School
  - AFIT (2 teams)
  - United States Naval Academy
  - United States Air Force Academy
  - United States Merchant Marine Academy
  - United States Coast Guard Academy
  - Royal Canadian Military Academy
- Students must **defend** computer networks from constant attack
- Attackers form the Red Cell
- Red Cell comprised of National Security Agency and Department of Defense experts
- Competition spans four days
- Red Cell offensive operations authorized 9:00 am to 4:00 pm

ITOC

CDX '09 Network



Example Captured Attack Vector

```
[**] [119:18:1] (http_inspect) WEBROOT DIRECTORY TRAVERSAL
     [**]
11/08-13:05:07.678919 10.2.190.254:33208 -> 154.241.88.201:80
TCP TTL:61 TOS:0x0 ID:46182 IpLen:20 DgmLen:1200 DF
***A**** Seq: 0xE257391E  Ack: 0x7D615C25  Win: 0xB7  TcpLen: 32
TCP Options (3) => NOP NOP TS: 953733 77976283
```

Types of Attacks (not exclusive):

- Web Application Attacks
- DNS Spoofing
- Source Routing
- Reverse Shells

ITOC

Logs from CDX ‘09

- Intrusion Detection Alert Database
- DNS Service and Message Logs
- Web Server Access and Error Logs
- Splunk Logserver Aggregate Logs

Benefits to our Approach

- Reduced artificiality
  - 30 Red Cell personnel
  - 20 White Cell personnel
- Scale of network
  - 30 person team using entire class C network
- Aggregated logs from West Point competition network

The Data

- Website with data and logs
  - http://www.itoc.usma.edu/research/dataset/

Shortcomings to our Approach

- IDS Researchers—mixture of cover traffic with malicious inhibits clearly labeled red traffic
- Lack of volume and diversity of traffic normally seen in production networks
- Length of CDX potential point of concern for anomaly detection that requires training period

ITOC

The Road Ahead

- West Point:
  - Increase number of capture sensors
  - Potential to provide actual virtual machines used during the exercise
  - Potential for observing worm behavior on exercise network
- Potential for data capture on other network warfare games in near future

ITOC

Questions