Labeled Full Packet/Flow Level Data Capture

Towards A Framework For Instrumenting Cyber Warfare Exercises

DoD Disclaimer

This document was prepared as a service to the DoD community. Neither the United States Government nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Reference herein to any specific commercial products, process, or service by trade name, trademark manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government.
The opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government, and shall not be used for advertising or product endorsement purposes.

Outline

- Objective
- Previous work
- Approach
- Framework
- Future work
- Acknowledgements

Objective

- Design a framework that:
  - Accurately auto-labels and captures full packet / flow level data sets from different classes of cyber defense exercises
  - Produces data sets that are of value to the intrusion detection and or security community

Previous Work
DARPA IDEVAL DATA SETS

- Pros
  - Well-known, publicly-available data set for repeatable intrusion detection experimentation
  - Fairly rich emulation of live network traffic
  - Well-documented and well-controlled data sets
- Cons
  - The included threats and protocols are showing their age
  - Modest traffic rate and network size, discrete attack events
  - Statistical artifacts exploitable by anomaly detection
- Note
  - This work is not intended to replace the IDEVAL data sets, but to explore techniques to provide additional resources for researchers and developers

Approach

- Survey opportunity
  - What aspects of a data sets are of value to researchers
- Design framework
  - Objectives
  - Requirements
- Implement framework
    - USMA Internal Cadet Cyber Defense Exercise (March)
    - Inter-Service Academy CDX 2009 (April)
- Package and share data
- Analyze data
- Gather feedback
- Repeat

Variety of Cyber Defense/Offense Exercises

- The framework must be robust enough, and generic enough, to ensure that it can be applied to most CND/CNO exercise
  - Inter-military service Cyber Defense Exercise (Defend only)
  - DEFCON Capture The Flag (Attack Only)
  - Colligate Cyber Defense Exercise (Defend Only)
  - Intra-USMA Cadet Cyber Warfare Exercise (Mixed)

Design Framework

- Objectives (What do we want out of our data captures)
  - Recorded traffic from multiple sensor locations
  - Attack event labeling
  - Network, host and service status information

  - Exercise reports
  - Network and host configuration change log
  - Host-base LOGS
  - Exercise IDS Alert DB or logs

Design Framework

- Requirements (what do we need in order to meet our objectives)
  - Complete recordings of traffic at sensor locations
  - Accurate description, source, destination and timing of Red Cell attack events
  - Recording of network, host and service status

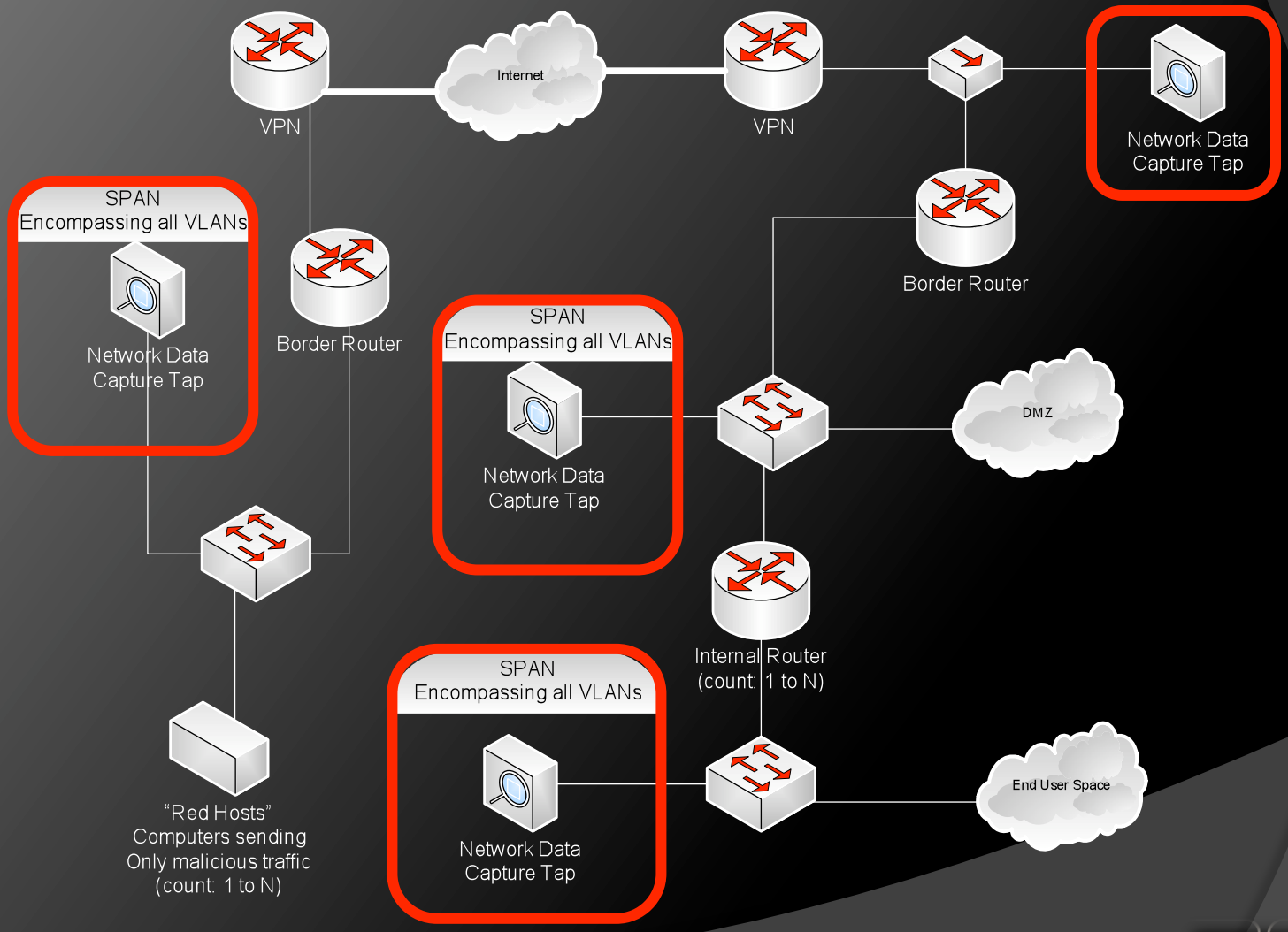
Abstract Network Design

- Placement of network ‘taps’ is critical when evaluating the type and amount of data captured
  - Are we only interested in the attacks?
  - Are we interested in the attacks and resulting traffic?
  - Are we interested in flow patterns without regard to type of traffic?

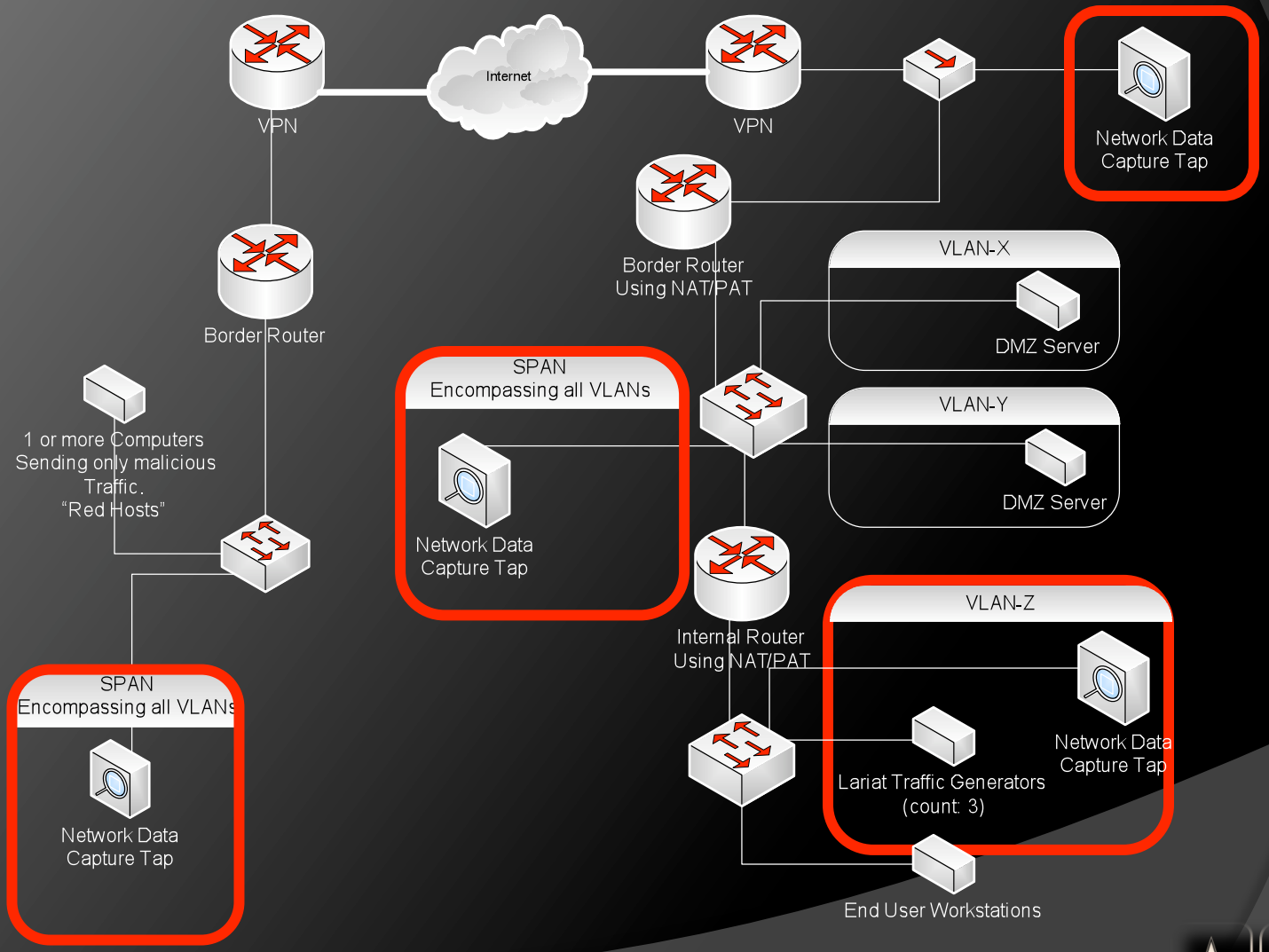
Limitations and Concerns

- One of the biggest concerns is to not impact the exercise/game in any way
  - Need to have the monitoring boxes and network config completely invisible to the participants
  - Limited number of monitoring boxes means that there may data loss from unmonitored portions of the network
  - Human recording is needed to know exactly what malicious events were occurring at what times
    - We cannot guarantee that every event will be recorded

Abstract Network Design



Network Design – CDX 2008 Example



Software Implementation

- Network taps (capturing boxes) will be running FreeBSD 7.1
  - Will be using tcpdump with the capturing interface in promiscuous mode to dump traffic directly to 1 Gigabyte files
  - Will capture during the entire exercise

Hardware Implementation

- 4Data Capture Nodes:
  - Dual-CPU (Intel Xeon 2.33 GHz)
  - 24 GB RAM (DDR2-667)
  - 1 .5 TB RAID-5
- Will be at the listening end of a one-way Ethernet cable
- That Ethernet cable will be plugged into a spanned port on a Cisco switch to allow monitoring of all traffic in all VLANs

Future Work

- Package and share data
- Community analysis
- Gather feedback
- Adjust Framework according to feedback
- Repeat

Conclusion

- Need for revitalized IDEVAL data sets
- A number of existing opportunities to capture data
- Proposed framework to capture and label relevant Data
- Infuse into existing IDEVAL data sets