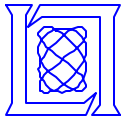
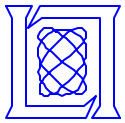


Results of the DARPA 1998 Offline Intrusion Detection Evaluation

Richard P. Lippmann, Robert K. Cunningham, David J. Fried,
Isaac Graf, Kris R. Kendall, Seth E. Webster, Marc A. Zissman
rpl@sst.ll.mit.edu

MIT Lincoln Laboratory
Room S4-121
244 Wood Street
Lexington, MA 02173-0073

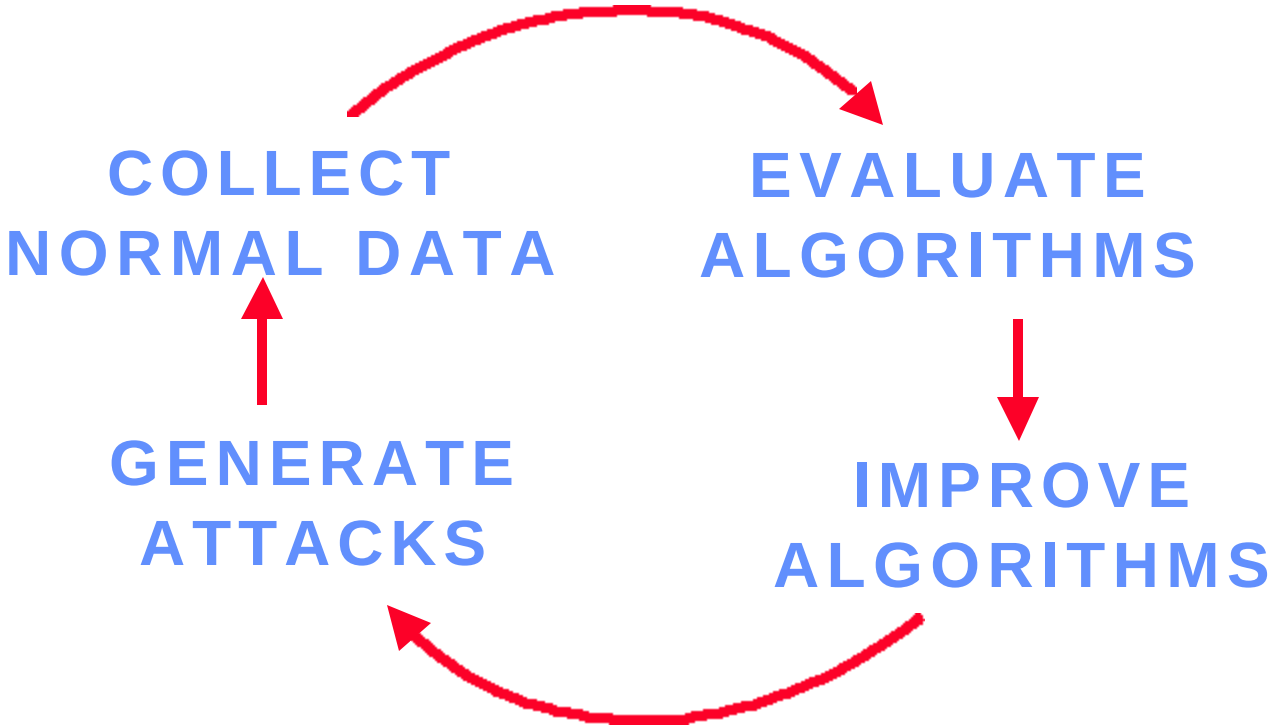
Presented at the Recent Advances in Intrusion Detection, RAID 99 Conference,
7-9 September
West Lafayette, Indiana, USA



Outline

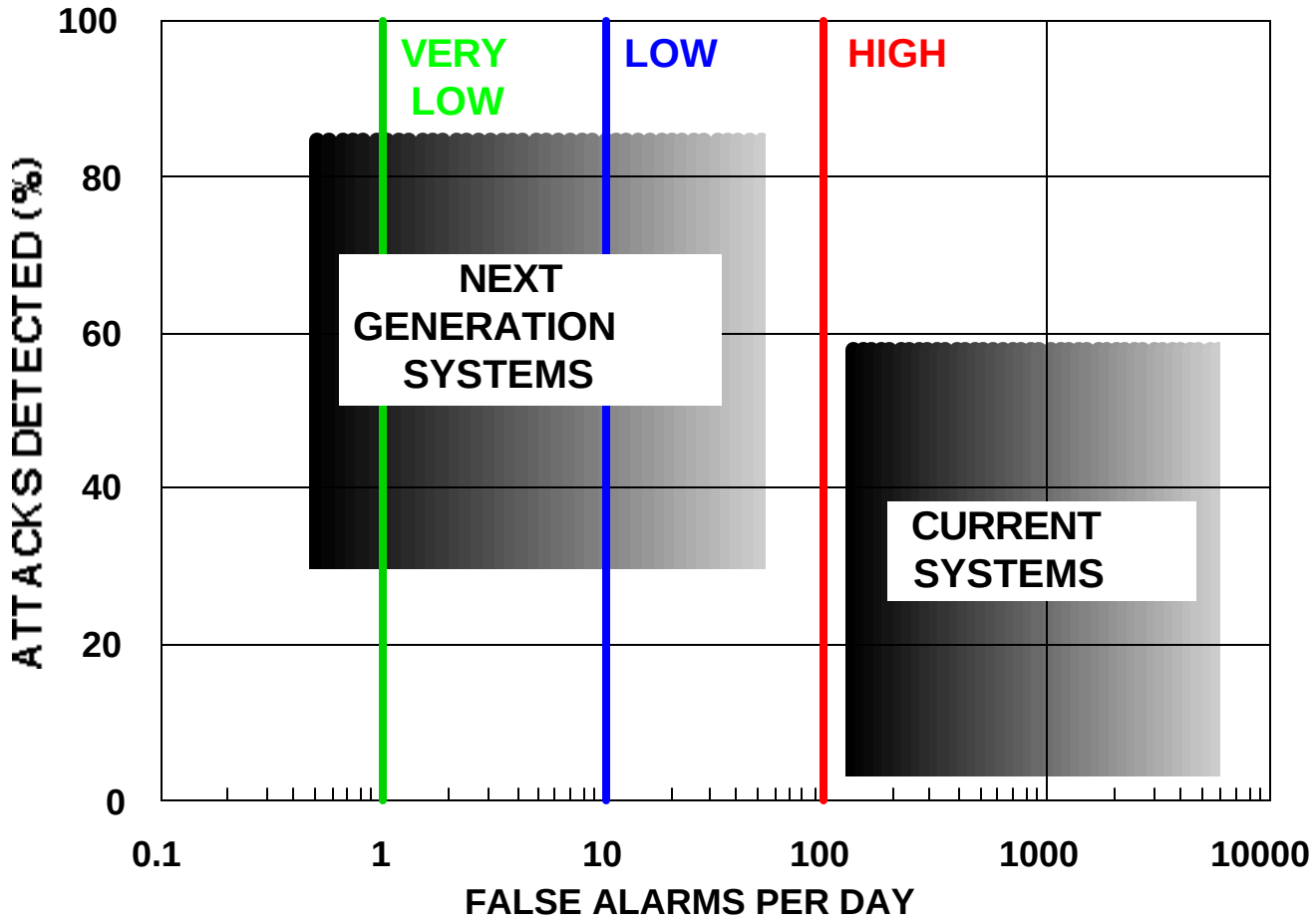
- Background and Introduction
- Analysis/Synthesis Approach to Generate Normal Background Traffic
- Attacks
- Results
- Summary and Conclusions

Goal of DARPA 1998 Intrusion Detection Evaluation



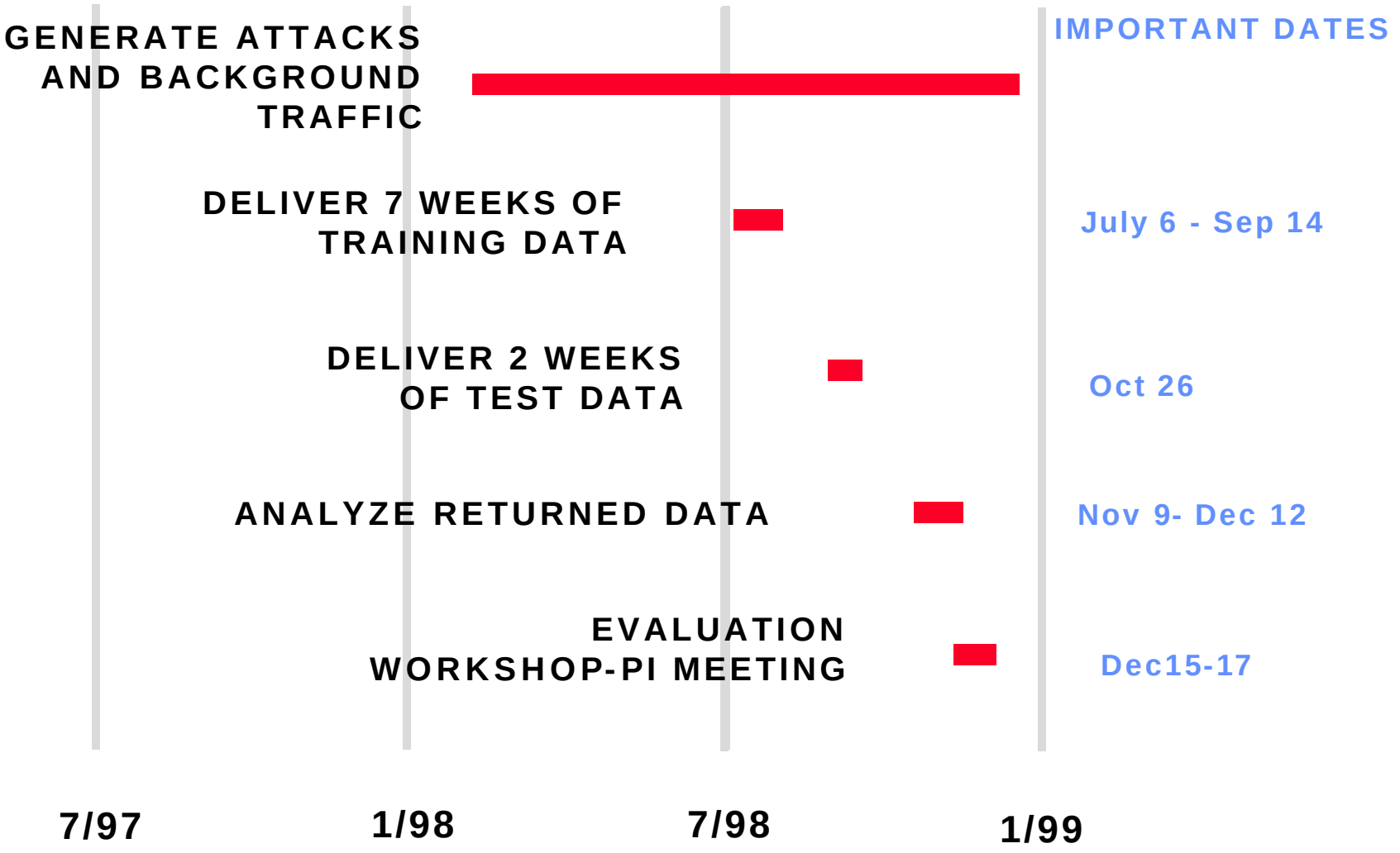
- **Evaluations Required to Determine Current System Capabilities**
- **Lead to Iterative Performance Improvements**
- **Difficult Because No Standard Comparison Metrics, No Existing Attack or Background Traffic Collections, Privacy/Security Restrictions**

Desired Receiver Operating Characteristic Curve (ROC) Performance

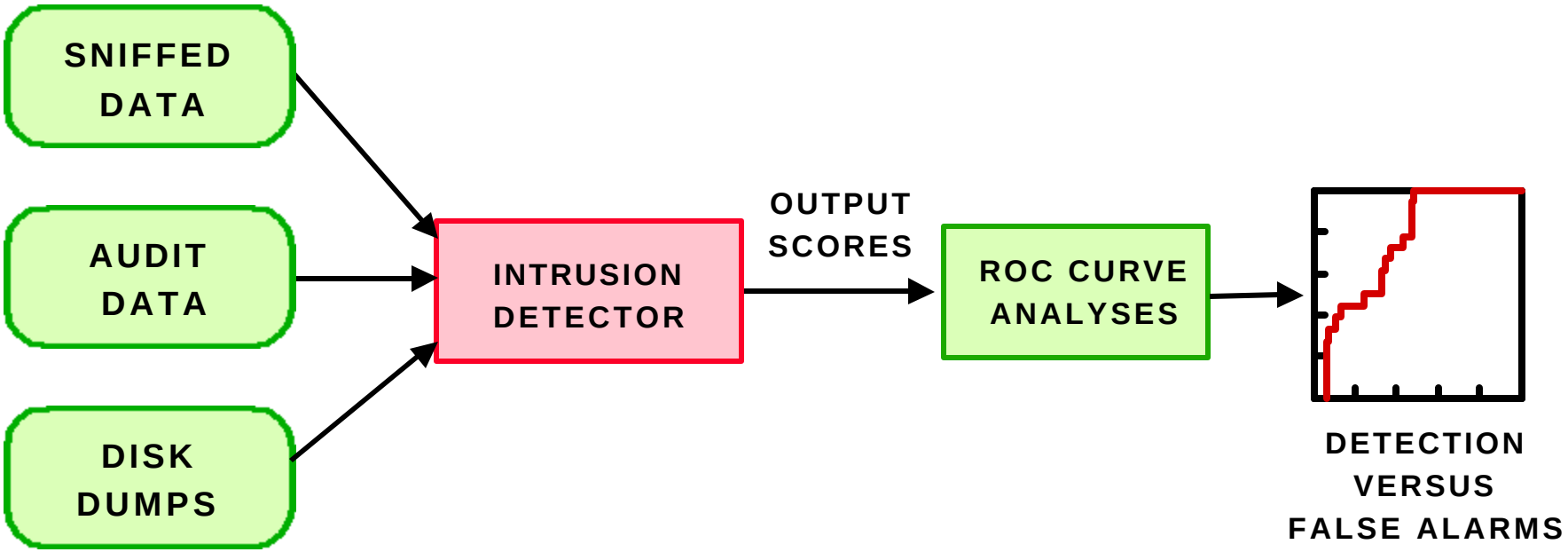


- Goal is to Reduce False Alarm Rates by Two to Three Orders of Magnitude and Improve Attack Detection Accuracy

Major Tasks and Timeline



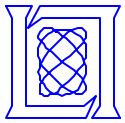
Data Types and Evaluation Overview



- Focus on UNIX, Outsider Attacks
- Generate More than Two Months of Data with Attacks
  - Network Sniffing Data (All Packets In/Out of Simulated Base)
  - Host Audit Data (Solaris Host BSM Audit Records)
  - Host File System Dumps (Solaris)
- Analyze and Compare False Alarm and Detection Rates

Outline

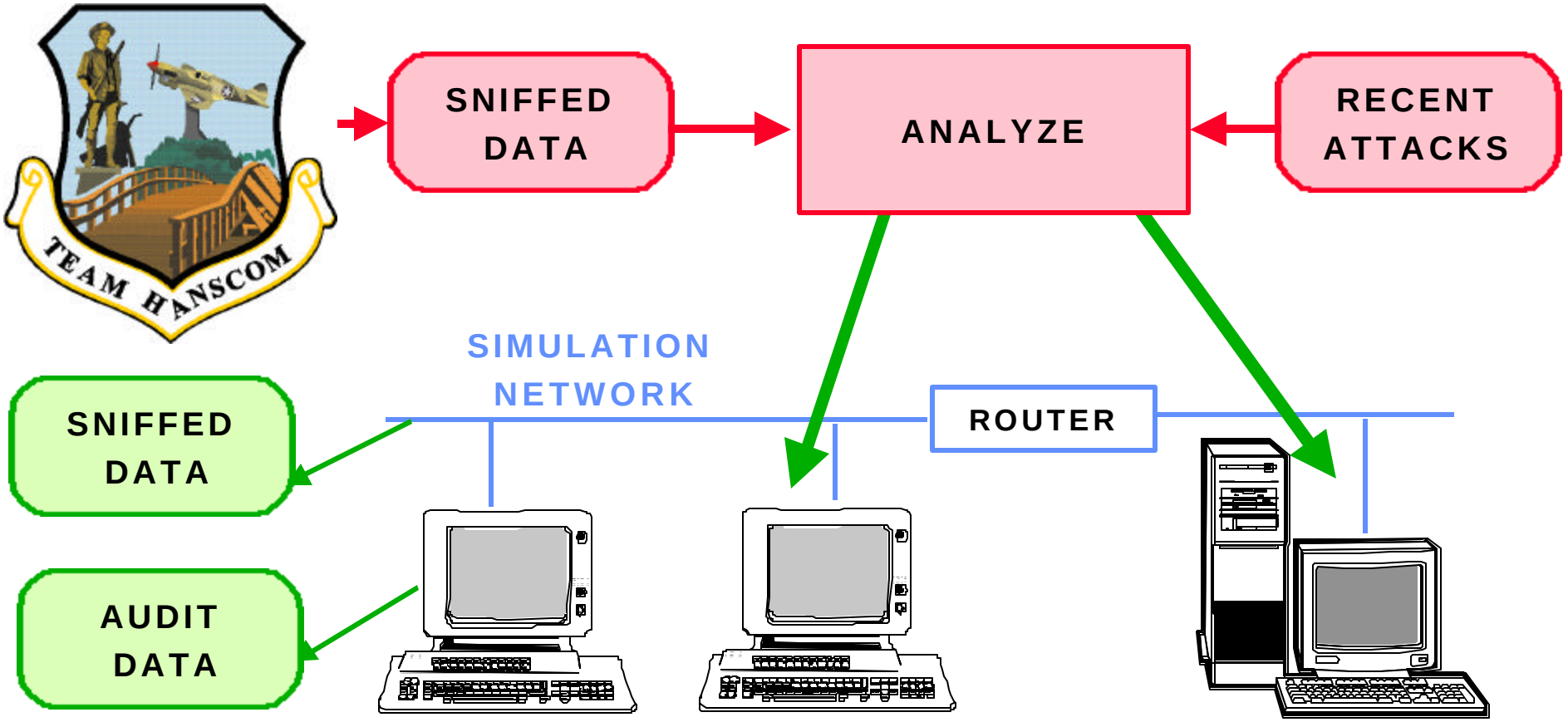
- Background and Introduction
- Analysis/Synthesis Approach to Generate Normal Background Traffic
- Attacks
- Results
- Summary and Conclusions



Corpus Generation Options

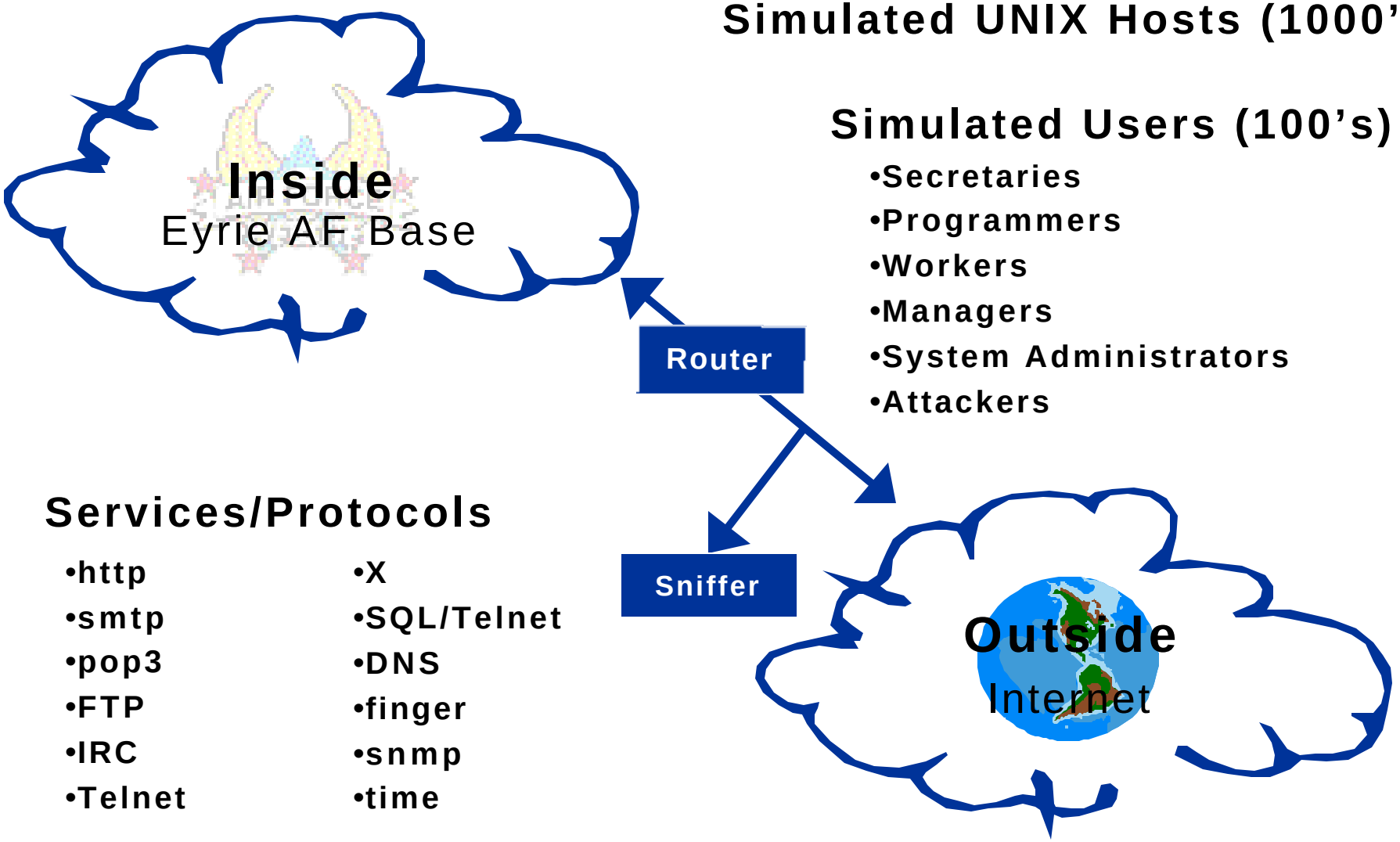
- **Option I: Sniff/Audit Real Operational Data and Attack Base**
  - **Real-World, but Can't Attack Operational Base and Can't Release Private Email, Passwords, Userid's, ...**
- **Option II: Sanitize Operational Data, Mix in Attacks**
  - **Too Difficult to Sanitize All Data Types, Mixing in Attacks Would Introduce Artifacts**
- **Option III - Synthesize Both Normal and Attack Sessions on a Private Network**
  - **Generate Non-Sensitive Traffic Similar to That Seen on a Base Using Public Domain and Randomly Generated Data Sources**
  - **Automate Normal Traffic Generation and Attacks Using Same Network Software (e.g. sendmail, ftp, telnet) Used on Base**
  - **Distribute Sniffing and Audit Data for Training and Testing Without Security or Privacy Concerns**

Analysis/Synthesis Approach



- Examine 4 Months of Data From Hanscom Air Force Base and More than 50 Other Bases, and Add Attacks
- Recreate Traffic on Simulation Network

Simulation Network Overview



Simulated UNIX Hosts (1000's)

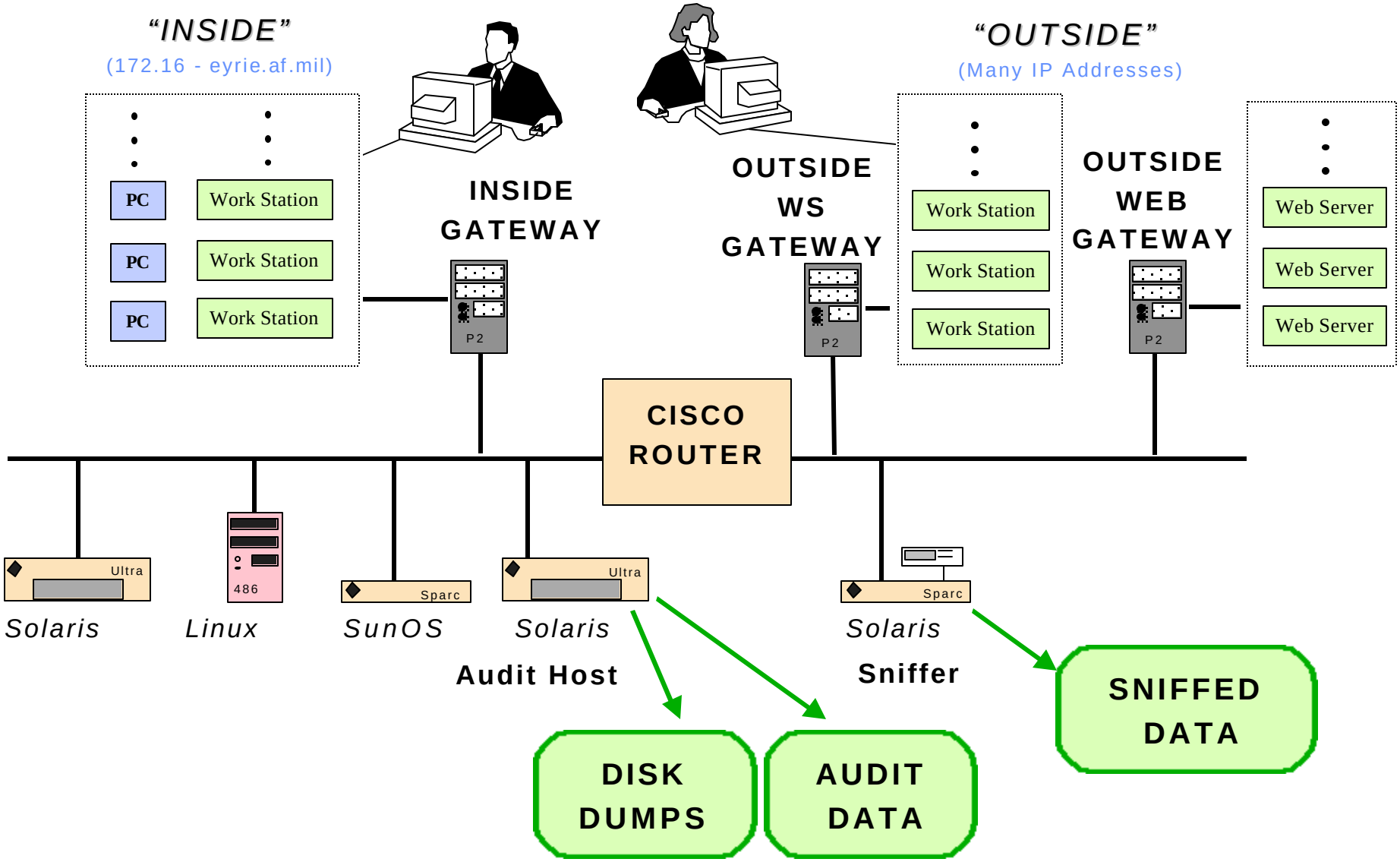
Simulated Users (100's)

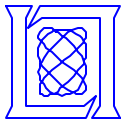
- Secretaries
- Programmers
- Workers
- Managers
- System Administrators
- Attackers

Services/Protocols

- http
- smtp
- pop3
- FTP
- IRC
- Telnet
- X
- SQL/Telnet
- DNS
- finger
- snmp
- time

Simulation Network Details

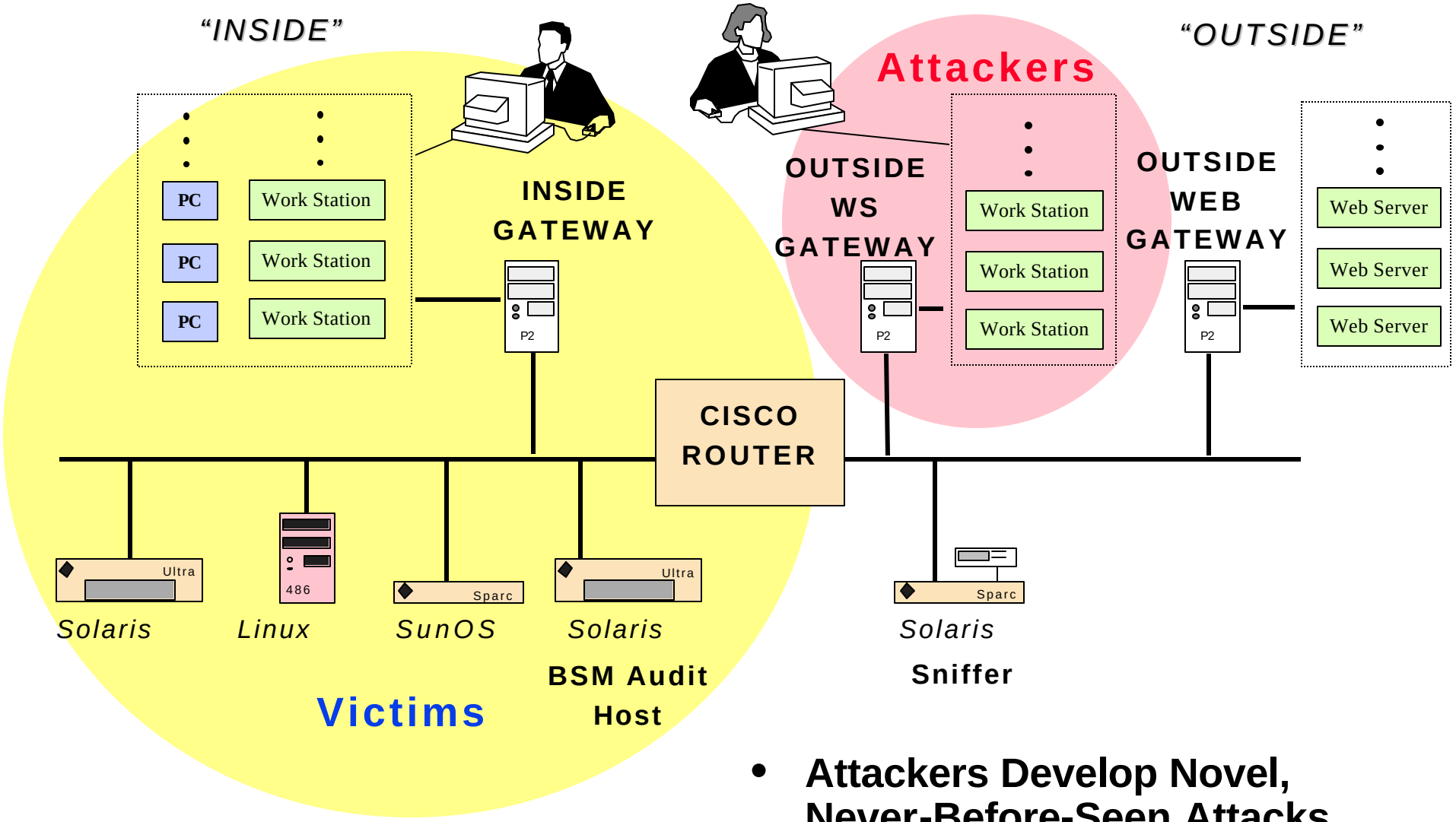




Outline

- Background and Introduction
- Analysis/Synthesis Approach to Generate Normal Background Traffic
- Attacks
- Results
- Summary and Conclusions

Attackers and Victims in Simulation



- **Attackers Develop Novel, Never-Before-Seen Attacks**

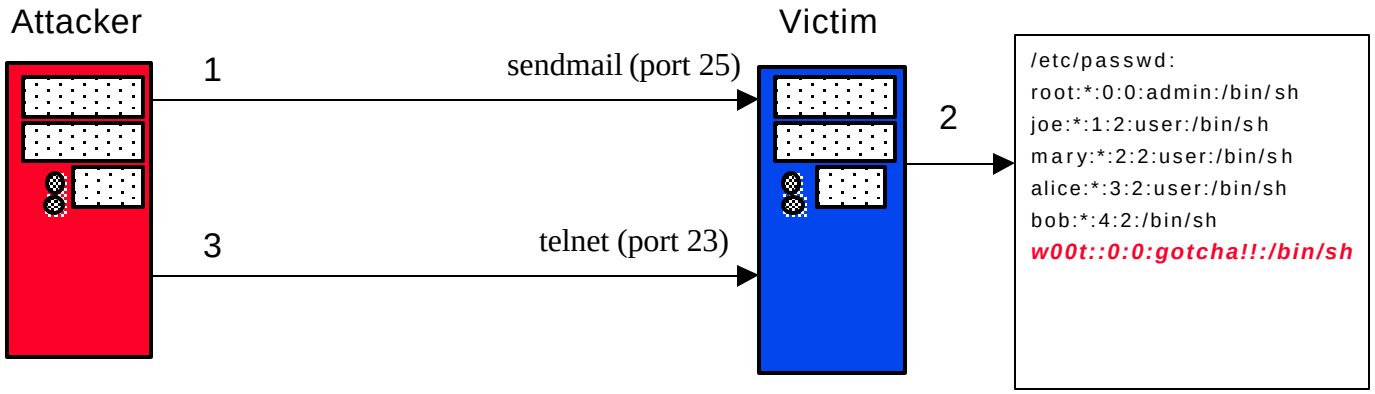
38 Attack Types in 1998 Test Data

	Solaris Server (audited)	SunOS internal	Linux internal	Cisco Router
DENIAL OF SERVICE (11 Types, 43 Instances)	•back •Neptune •Ping of death •Smurf •Syslogd •land •Apache2 •Mailbomb •Process Table •UDP Storm	•back •Neptune •Ping of death •Smurf •land •Apache2 •Mailbomb •Process Table •UDP Storm	•back •Neptune •Ping of death •Smurf •Teardrop •land •Apache2 •Mailbomb •Process Table •UDP Storm	•snmpgetattack
REMOTE TO LOCAL (14 Types, 17 Instances)	•dictionary •ftp-write •guest •phf •httptunnel •xlock •xsnoop	•dictionary •ftp-write •guest •phf •httptunnel •xlock •xsnoop	•dictionary •ftp-write •guest •imap •phf •httptunnel •named •sendmail •xlock •xsnoop	
USER TO ROOT (7 Types, 38 Instances)	•eject •ffbconfig •fdformat •ps	•loadmodule •ps	•perl •xterm	
SURVEILLANCE /PROBE (6 Types, 22 Instances)	•ip sweep •nmap •port sweep •satan •mscan •saint	•ip sweep •nmap •port sweep •satan •mscan •saint	•ip sweep •nmap •port sweep •satan •mscan •saint	•ip sweep •nmap •port sweep •satan •mscan •saint

- **120 Attacks in 2 Weeks of Test Data**

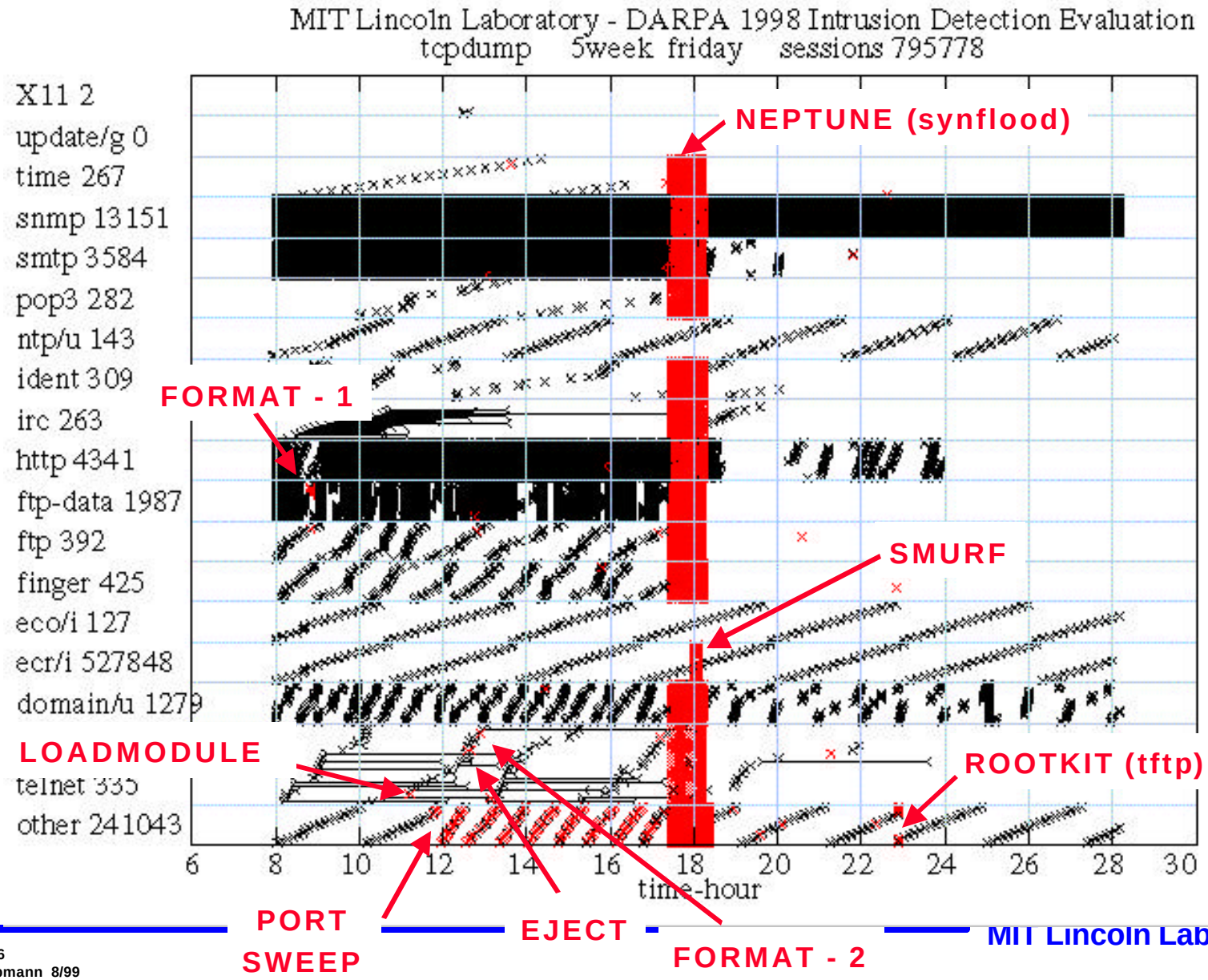
≡ test only

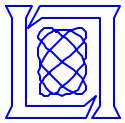
Novel Sendmail Remote to User Attack



- Novel Attack Code Developed for this Evaluation
- To Our Knowledge No One Else has Attack Code that Exploits this Vulnerability
- An Attacker Sends One Email message to the Victim with a MIME header field that Causes a Buffer Overflow and Modifies the Password File
- After this the Attacker Has Free Access to the Victim Machine as Root using Telnet

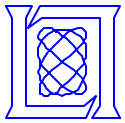
Training Data Traffic, Week 5, Friday





Outline

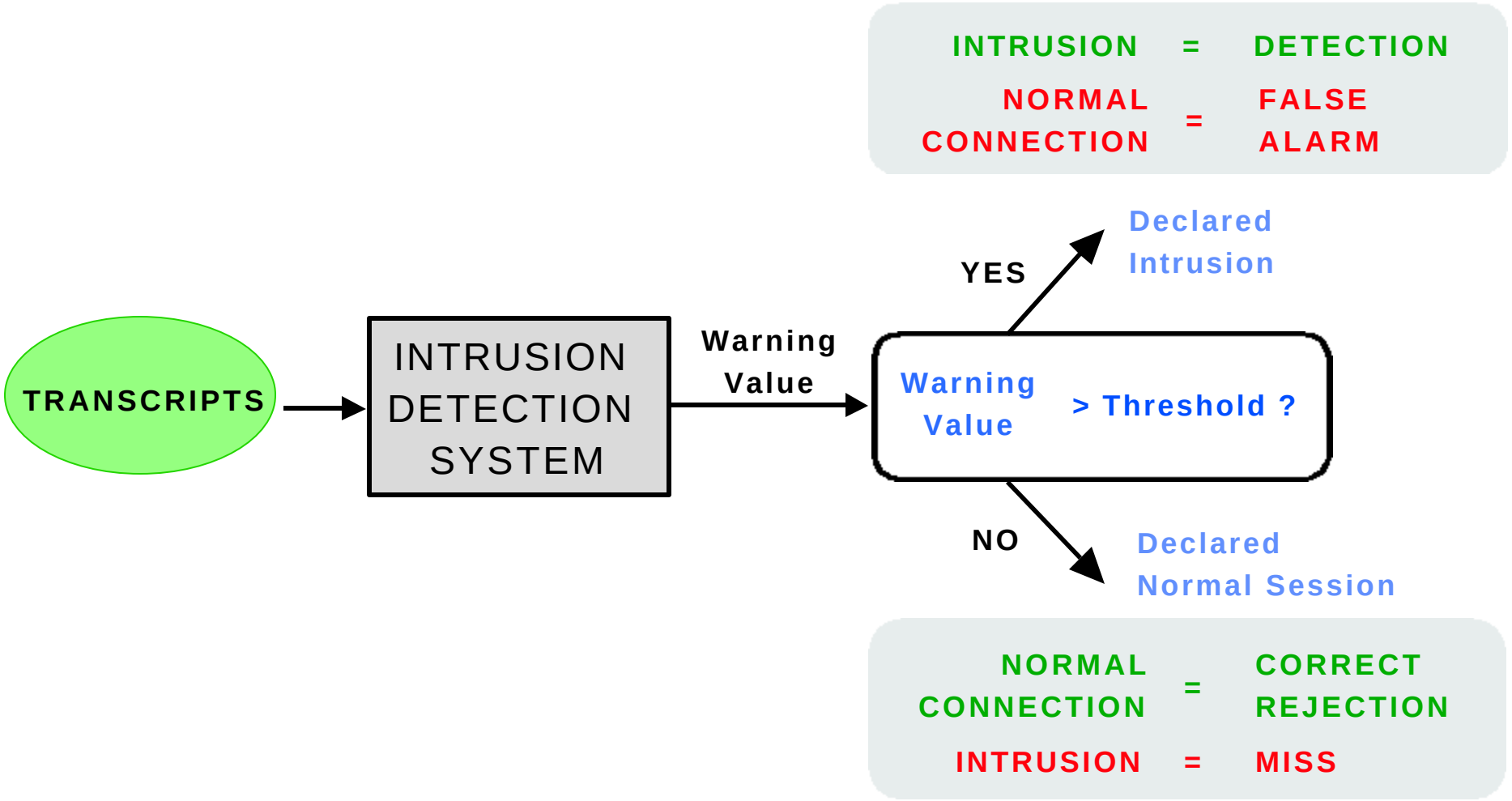
- Background and Introduction
- Analysis/Synthesis Approach to Generate Normal Background Traffic
- Attacks
- Results
  - Participants
  - Generating Receiver Operating Characteristic (ROC) Curves
  - Overall ROC of Best Composite System
  - ROCs With Network Sniffing Data for Four Attack Categories (Denial of Service, Probes, User to Root, Remote to Local)
  - ROC with Host Audit Data for User to Root Attacks
- Summary and Conclusions



Participants and Systems

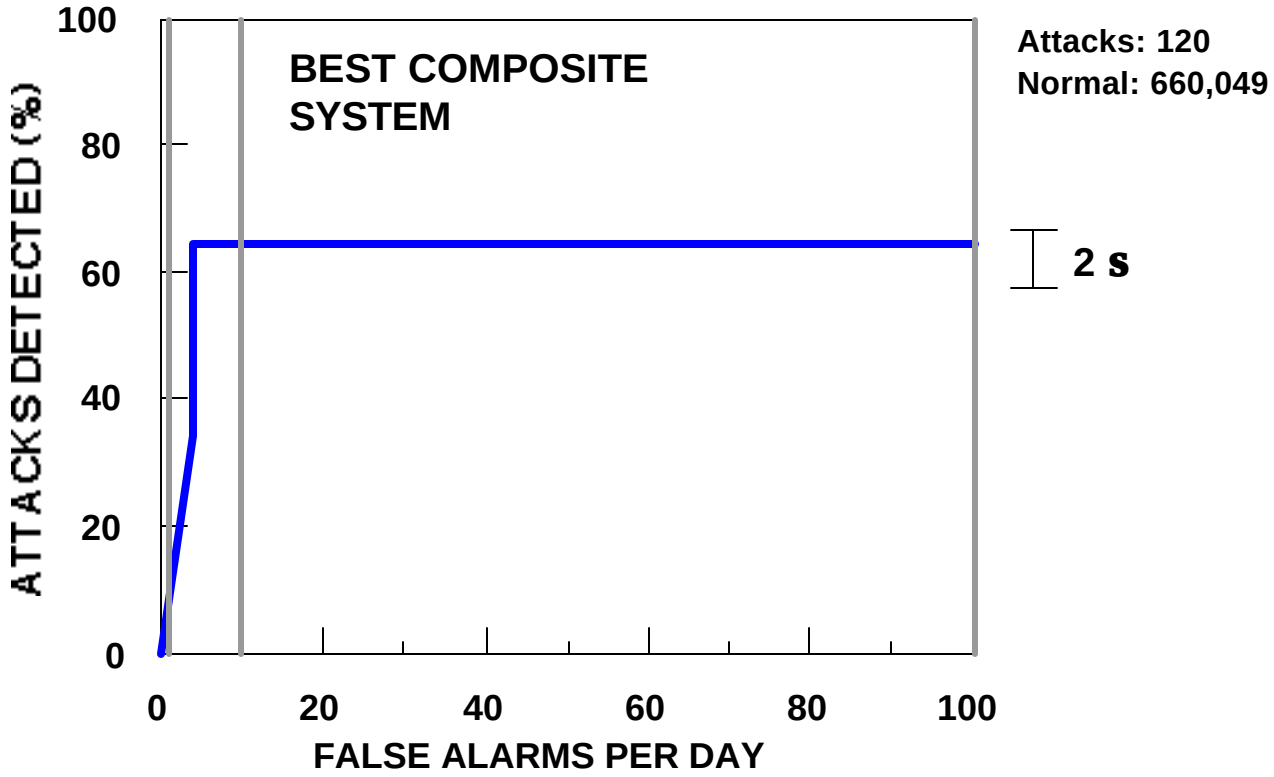
- Six Participants Submitted Seven Systems
  - Network Sniffer Inputs Only (3)
  - Host Audit BSM Inputs Only (2)
  - Both Host Audit and Sniffer Inputs (1)
  - File System Dumps (1)
- All Participants Followed the Blind Test Procedures
- System Types
  - Finite-State Machine or Rule-Based Signature Detection
  - Expert Systems
  - Pattern Classification/Data Mining Trained System

Generating A Receiver Operating Characteristic (ROC) Curve



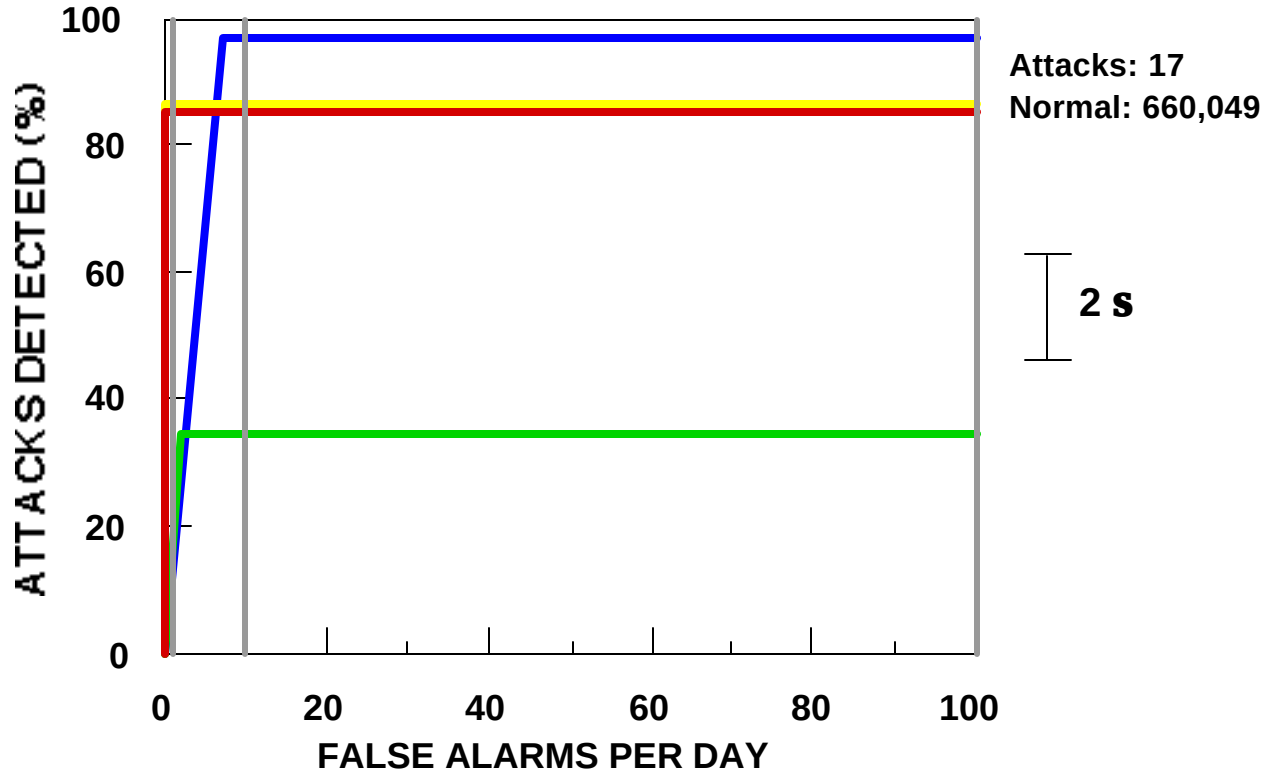
- **Vary Threshold to Obtain Different False Alarm and Miss Values and Trace out ROC Curve**

Best Composite ROC Across All Systems for All Attacks



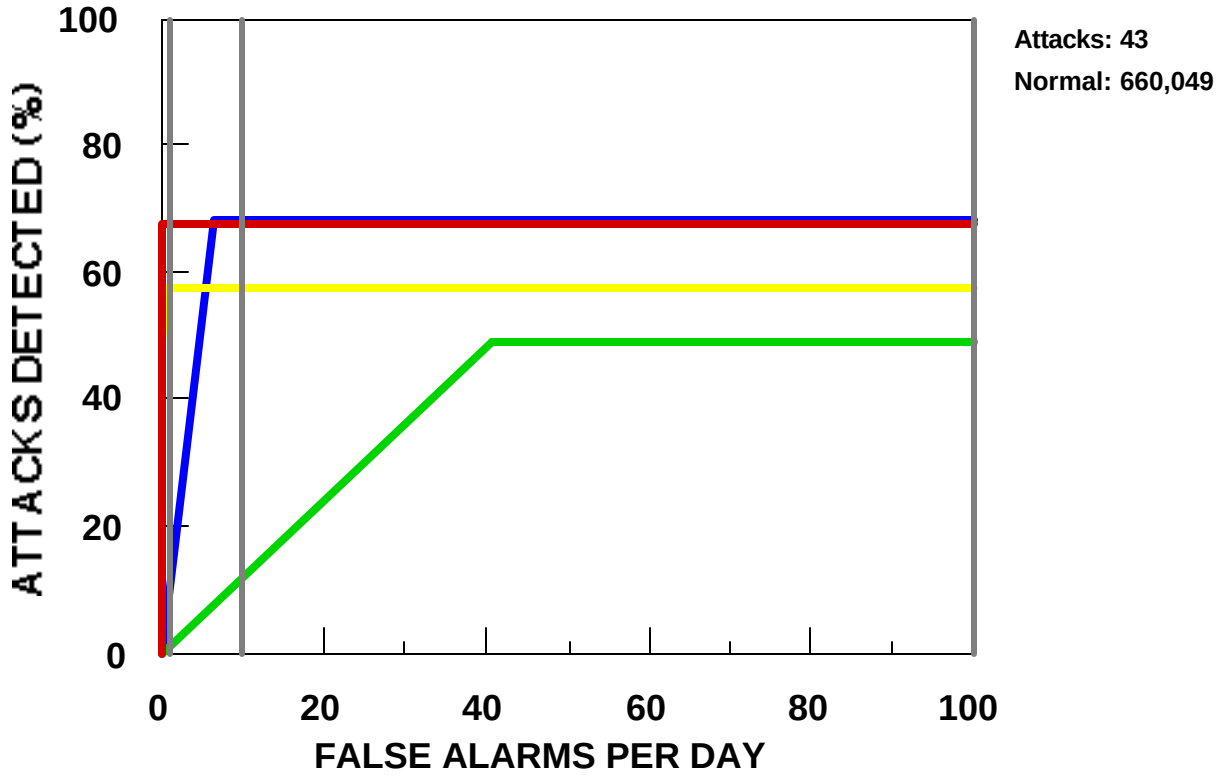
- Roughly 65% Detection at 5 False Alarms Per Day
- Low False Alarm Rate, But Poor Detection Accuracy
- Most Systems Miss New and Novel Attacks

ROC's for Probe Attacks Using Network Sniffing Data



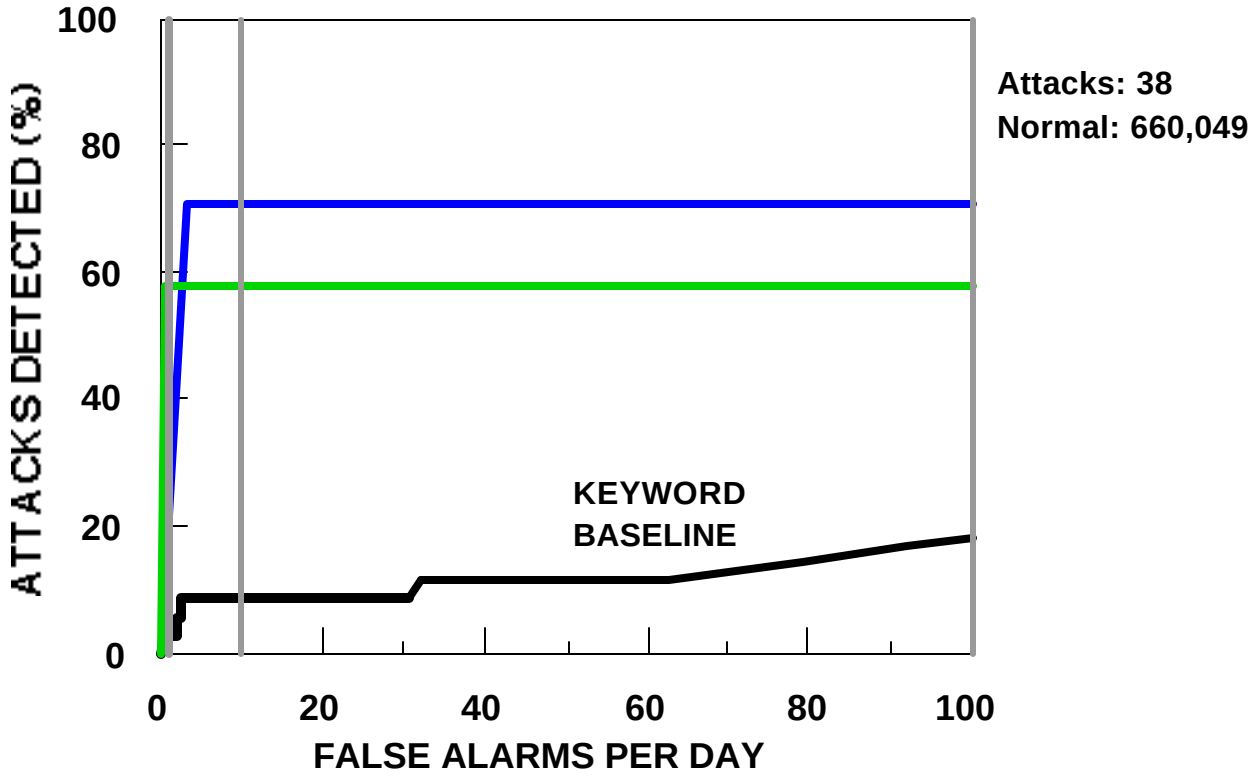
- Good Performance for Old and New Probes
- Some Research Systems Find Almost all Probe Attacks at Low (1 False Alarm Per Day) False Alarm Rates
- Old and New Probes are Similar (Satan, IP Sweeps, NMAP)

ROC's for Denial of Service (DoS) Attacks Using Network Sniffing Data



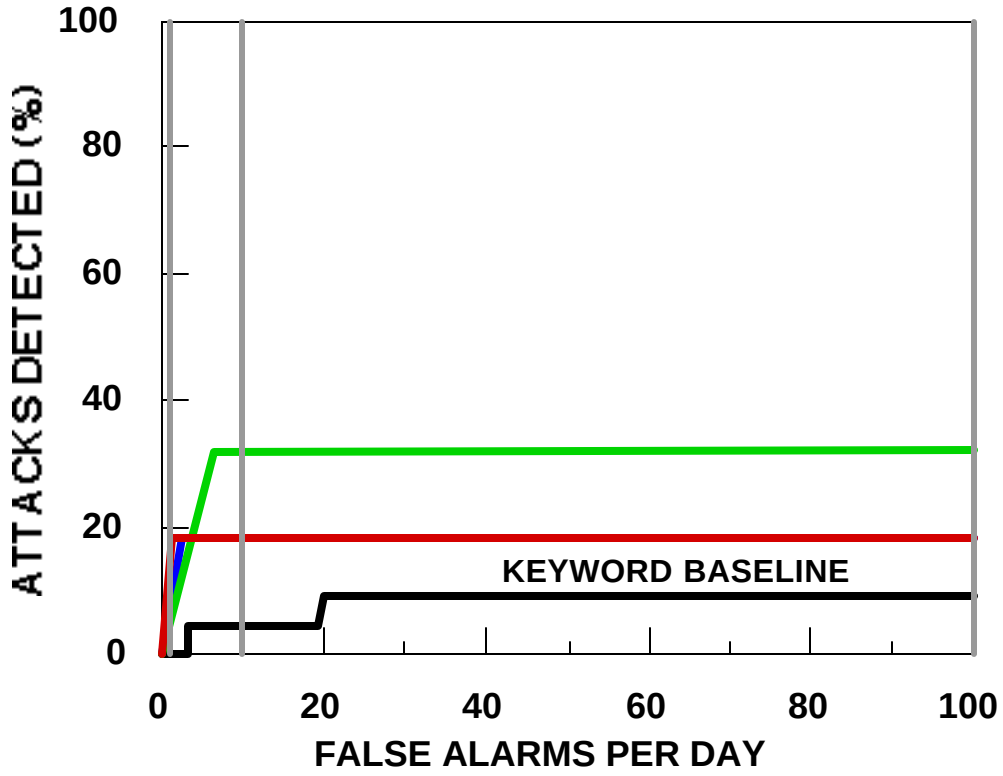
- Research Systems Don't Find all DoS Attacks
- Systems Find Old Attacks but Miss New Attacks (Process Table Exhaustion, Mail Bomb, Chargen/Echo Storm)

ROC's for User to Root (u2r) Attacks Using Network Sniffing Data



- Research Systems Don't Find all User to Root Attacks
- Research Systems Perform Substantially Better than Baseline Keyword Reference System Which is Similar to Many Commercial and Government Systems

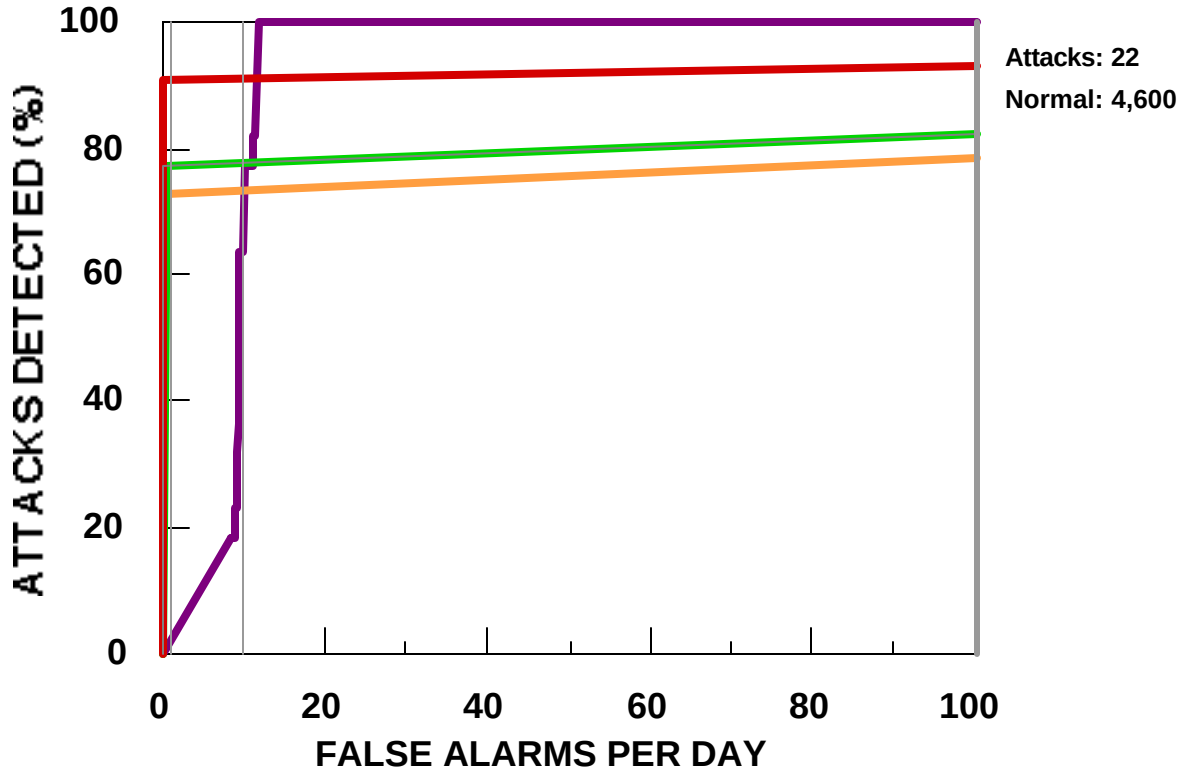
ROC's for Remote to Local (r2l) Attacks Using Network Sniffing Data



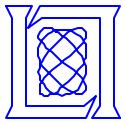
Attacks: 22
Normal: 660,049

- All Systems Have Low Detection Rates
- Many New Attacks, Highly Varied Attack Mechanisms (imap, dictionary, http tunnel, named, sendmail, xlock, phf, ftp-write)

ROC's for User to Root (u2r) Attacks Using Host Audit Data



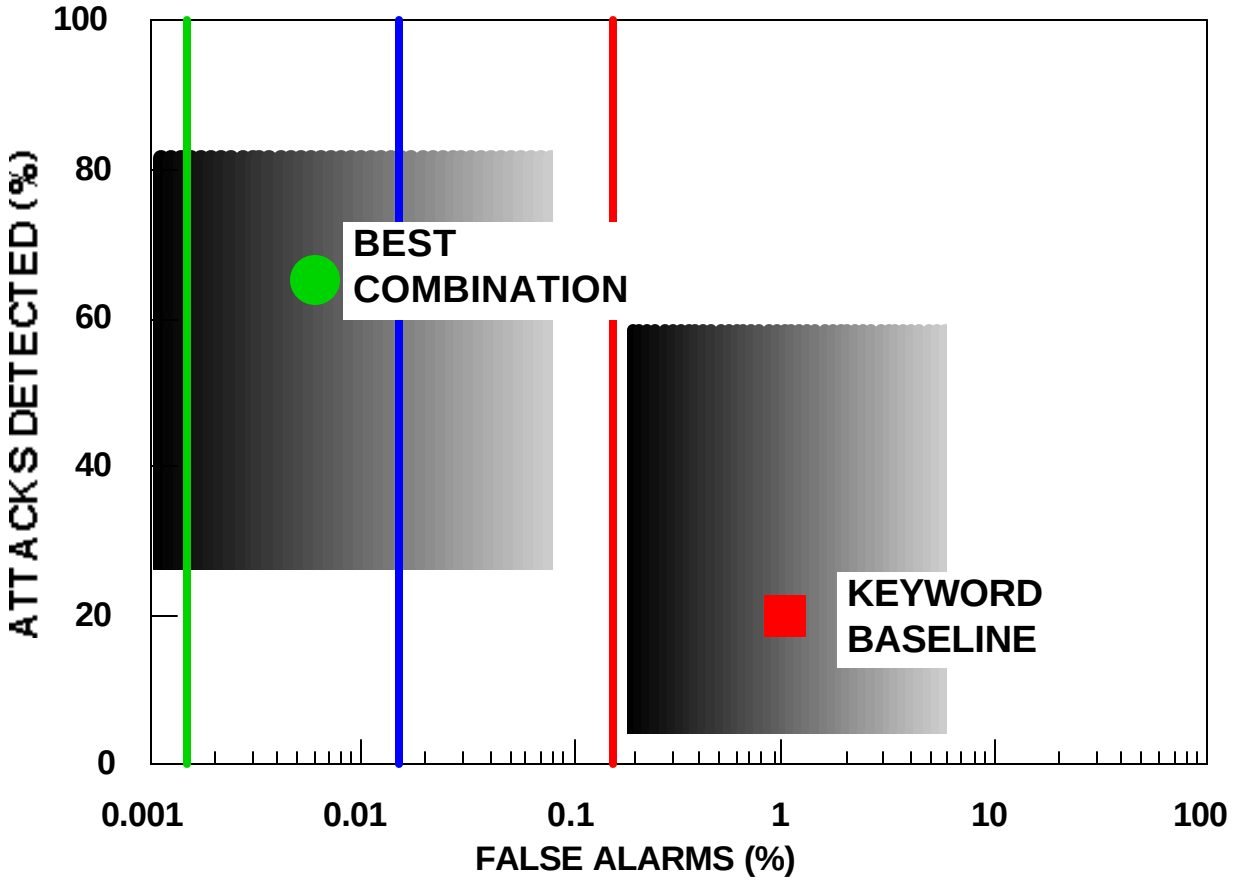
- Excellent Performance Using Host Auditing to Detect Local Users Illegally Becoming Root
- But This Requires Auditing on Each Host and is Only for User to Root Attacks



Outline

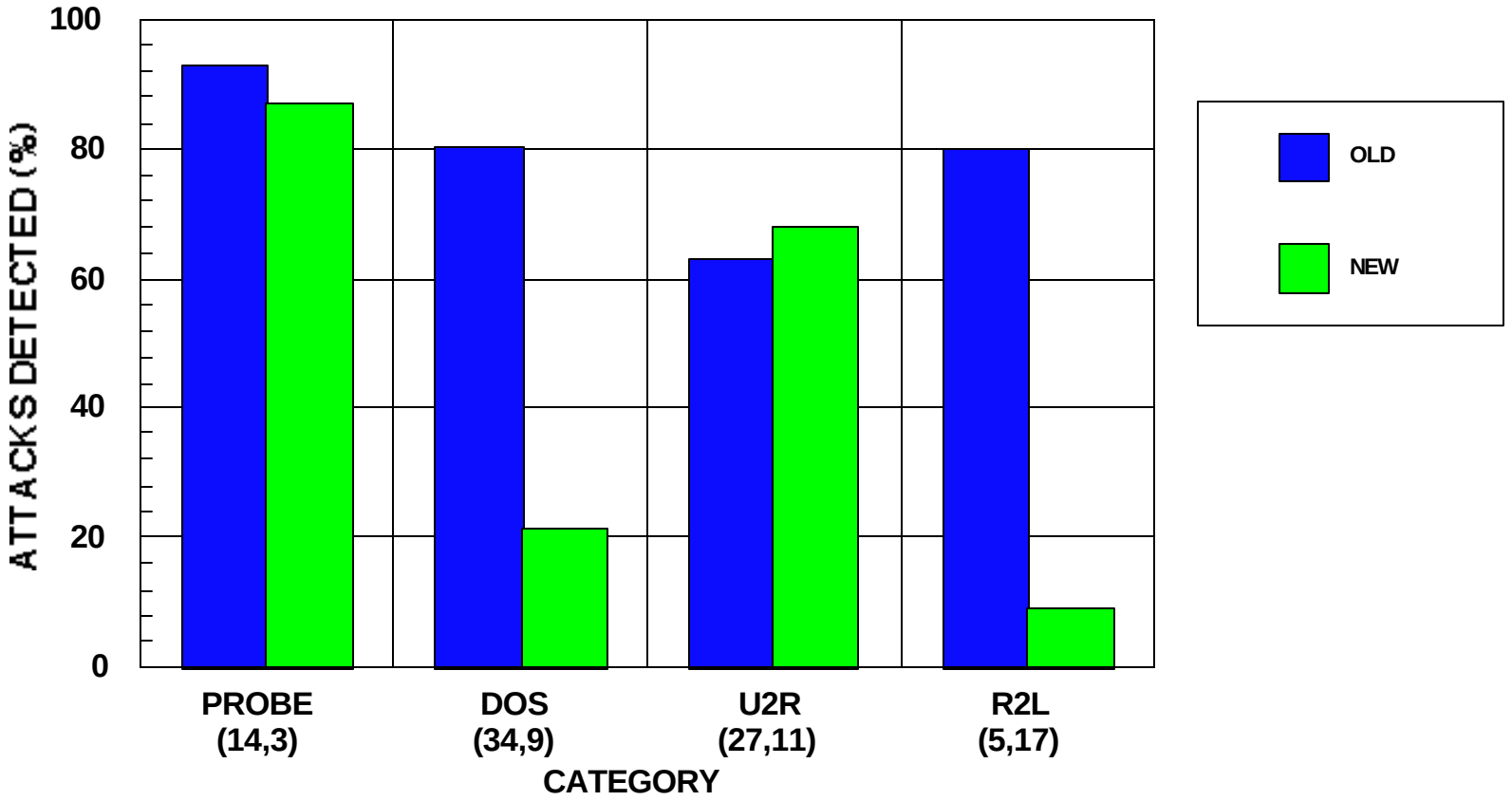
- Background and Introduction
- Analysis/Synthesis Approach to Generate Normal Background Traffic
- Attacks
- Results
- Summary and Conclusions

Best Combination System from This Evaluation Compared to Keyword Baseline

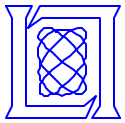


- False Alarm Rate Is More Than 100 Times Lower
- Detection Rate Is Significantly Better
- Keyword Baseline Performance Similar to Commercial and Government Keyword-based Systems

Best Systems in This Evaluation Don't Accurately Detect New Attacks



- Systems Generalize Well to New Probe and User to Root Attacks, but Miss New Denial of Service and Remote to Local Attacks
- Basic Detection Accuracy for Old Attacks Must Also Improve



Summary and Future Plans

- **We Have Developed an Intrusion Detection Test Network Which Simulates a Typical Air Force Base**
  - **Generate Realistic Background Traffic With 1000's of Simulated Hosts and 100's of Simulated Users**
  - **Insert More Than 35 Types of Automated Attacks**
  - **Measure Both Detection and False Alarm Rates**
- **The 1998 DARPA Evaluation Successfully Demonstrated**
  - **1) Research Intrusion Detection Systems Improve Dramatically Over Existing Keyword Systems**
  - **2) Research Systems, However, Miss New Denial-of-service and Remote-to-local Attacks and Do Not Perfectly Detect Old Attacks**
- **The 1999 DARPA Evaluation Will Add Windows NT Hosts and Many New Attacks**
  - **Focus in on Detecting New Attacks and Maintaining Low False Alarm Rates**