

Tony Bartoletti, Computer Incident Advisory Capability (CIAC) - Cyber Security Research

Paul D'Avilar, Johns Hopkins University (JHUI)  
Chris Mueller, University of California at Davis (UC Davis)

## Introduction

Every day, millions of hostile probes bombard the outer defenses of Livermore's network. A significant portion of these probes constitute network scans. During a network scan, the attacker sends a connection request to every possible network address and listens for replies indicating the presence of a (possibly vulnerable) computer. The methodology resembles that of a telemarketer calling every possible phone number for a given area code, from (925)000-0000 to (925)999-9999. Since a network scan often serves as a precursor to an attack, reliable identification of scanners can significantly enhance cyber-security. Furthermore, the ability to map adversary hierarchies and correlate attacks with events in the real world contributes to counterintelligence work. The patterns in packet arrival timing may hold the key to consistent identification of Internet adversaries.

## Scope and Objectives

### Traffic Characterization

Identify an ensemble of methods optimized to distinguish voluminous hostile activity according to attackers' techniques, tools employed, platforms employed, and routing interference (true network location).

### Confidence Level

Develop confidence-level metrics for hypothesis resolution

### Operational Capability

Demonstrate operational capability to isolate and characterize hostile activity under real-world load conditions.

### Reporting

Provide report on methodology and results, with expectation for publication in appropriate journals

## Tasking Outline

### Data Collection

Redeploy TIPS scan sensor at LLNL to collect refined hostile scan data

### Control Tests

Conduct controlled studies using probe tools such as Nmap, Nessus, etc. under varied system loads and similar constraints  
Expand controlled studies to perform "hostile" scans from varied remote locations.

Independently varying network location and platform activity is critical to resolving the import of issues in the derived feature space

### Signature Database

Refine signature database and efficient look-up capability

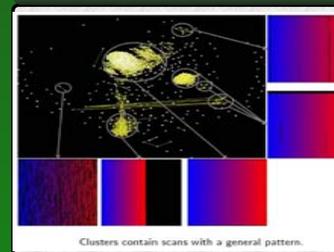
## Abstract

Network scanners constantly probe the Livermore network looking for vulnerabilities. Analyzing packet arrival timing data reveals highly distinctive patterns that may correlate with the attacker's choice of tools, physical platform and/or network location. Consistent identification will improve network security and aid counterintelligence efforts. We have developed tools to pre-process scan data, using wavelet techniques to achieve over 1,000x compression ratio while still preserving the essential features. Initial experiments indicate our methods consistently identify patterns in the data.

## Probe Tools – Control Tests

Scan Tools	OS		Randomization				Time	Source Hiding		Packet Content	
	Unix	Linux	Src Port	Src IP	Dst Port	Dst IP	Adjust Rate	Stealth	Spoofing	Adjustable TCP Flags	Adjustable Date
Nmap	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Nessus	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Netcat	Y	Y	Y	Y	Y	Y	Y	Y*	Y*	Y*	Y
Strobe		Y	Y			Y**	Y		Y*		
Hping2		Y	Y	Y		Y	Y		Y	Y	Y
Superscan	Y				Y	Y	Y				
Scansite	Y			Y	Y	Y					

## ScanVis – UC Davis



## Data Analysis

We can visualize packet arrival data by plotting the target address space:



Even after several days, scans consistently display highly distinct patterns:



Identical patterns in one data mode become distinct in another mode:



## Wavelet Summarization

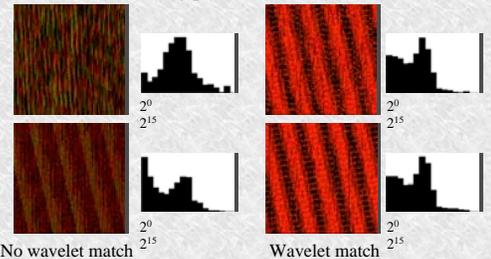
Wavelets summarize data at multiple resolutions

Given a series of  $N=2^n$  items,  $d_1, \dots, d_N$ , we can calculate:

$$d'_{i,2} = \frac{d_i + d_{i+1}}{2}, \quad s_{i,2} = \frac{d_i - d_{i+1}}{2}, \quad \sigma_{i,2} = \frac{s_{i,2}}{2^{n-1}}, \quad i = 0, 2, 4, \dots, 2^n - 2$$

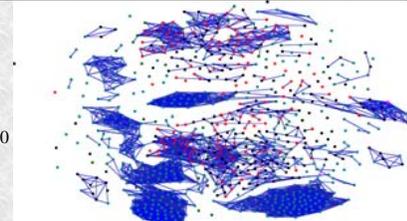
Repeated recursively, the sigma values estimate the variance at each resolution

Wavelets effectively preserve features



## Wavelet Clustering

- Port 80
- Port 139
- Port 445
- Port 1433
- Port 17300
- All other



Evidence of effectiveness of this technique is illustrated by this figure. The nodes in this graph are clustered by applying wavelets 29 on the arrival time deviation of probe data, which clearly reveal strong port base clustering/alignment. We surmise this is due to worms or other processes that execute with consistent behavior on specific platforms.

## Status – Current Activities

### Data Collection

- Controlled probe data is being collected to help
  - Characterize the tools
  - Characterize the host hardware
  - Characterize the host OS
  - Ignoring the effects of routing

### Control Tests

- Using control tests to verify and understand wavelets and data modes abilities to identify similar scan features