



# Intrusion Log Sharing

## University of Wisconsin-Madison

John Bethencourt (bethenco@cs.wisc.edu)

Jason Franklin (jfrankli@cs.wisc.edu)

Mary Vernon (vernon@cs.wisc.edu)



# Talk Outline

- Background: Blacklists, DOMINO System
- Open Questions
- DShield logs
- Preliminary Analysis Results
- Ongoing Work



# Background: Blacklists

Goal: deny access to hosts considered “bad” relative to the site’s policies

e.g., a host that performs a port scan

Given: a set of firewall and NIDS logs

generate: a blacklist of hosts to block

Blacklist failure: when a connection is not blocked and the source host is later on the blacklist



# Background: DOMINO System

- architecture: 20+ “axis nodes” in an overlay network
- each axis node:
  - ❑ maintains a NIDS & an active sink
  - ❑ obtains NIDS logs from “satellite nodes”
  - ❑ exchanges hourly, daily, and monthly blacklists with other axis nodes & distributes to satellite nodes
  - ❑ provides query engine with capability to trigger a new summary



# Background: DOMINO System

- Analysis of 1600 NIDS logs
  - ❑ 40-60 logs together give correct ordering of offenders & target ports
    - /24 site logs or /16 site logs
  - ❑ 1024 offenders perform 90% of the scans in a given hour
  - ❑ the similarity in 2 sites' blacklist orderings is correlated with the IP distance between the sites
  - ❑ 2 years required to scan entire subnet to avoid top 60 hourly or daily blacklist



# Background: DOMINO System

- each axis node votes for an alert if:
  - ❑ 200% increase in hourly number of scans,
  - ❑ 100% increase in number of offenders during hour, and
  - ❑ number of offenders is  $> 5$

50 access nodes & 20% consensus:

avg. reaction time for SQL Snake is close to 0



# Background: DOMINO System

- summary:
  - ❑ 20+ axis nodes, each with satellite nodes
  - ❑ sharing blacklists increases site's knowledge of global top offenders & global top target ports
  - ❑ consensus alerts for increase in scans & offenders decreases average reaction time for worms

not known: whether sharing logs decreases blacklist failure



# Open Questions

- 1) benefit of obtaining NIDS logs from other sites for blocking future attacks
  - What is the maximum increase in the percentage of future attacks that a site can block?  
  
i.e., with full blacklist from each other site compared with only the site's own blacklist
  - What percentage of future attacks can't be identified from the global blacklist?



# Open Questions

- 1) benefit of obtaining logs from a single active sink (compared to NIDS logs) for blocking future attacks
- 2) If NIDS logs are beneficial, which minimal set of peers provides the maximum benefit?
- 3) How frequently should log information be exchanged?



# Metric: Blacklist Quality

$B_i$  : blacklist at beginning of time period  $i$

$I_i$  : set of hosts that attempt connections during  $i$

- *quality* of the blacklist at time  $i$ :

$$q_i = \frac{|B_i \cap I_i \cap B_{i+1}|}{|I_i \cap B_{i+1}|}$$

- % of attackers that can't be identified:

$$1 - q_i$$



# Shared Blacklist Max. Benefit

- maximum possible benefit of shared blacklists:  
compare quality of  $B_i$  = local blacklist vs  $B_i$  = global blacklist

$$q_i = \frac{|B_i \cap I_i \cap B_{i+1}|}{|I_i \cap B_{i+1}|}$$



# Experimental Data

- DShield project: central repository of NIDS & firewall log data
- collected from 1600 providers across Internet
  - ❑ number of hosts: 1 to 65,000
  - ❑ number of log entries per second:  $10^{-5}$  to 100
- Initially: 4 consecutive days of data

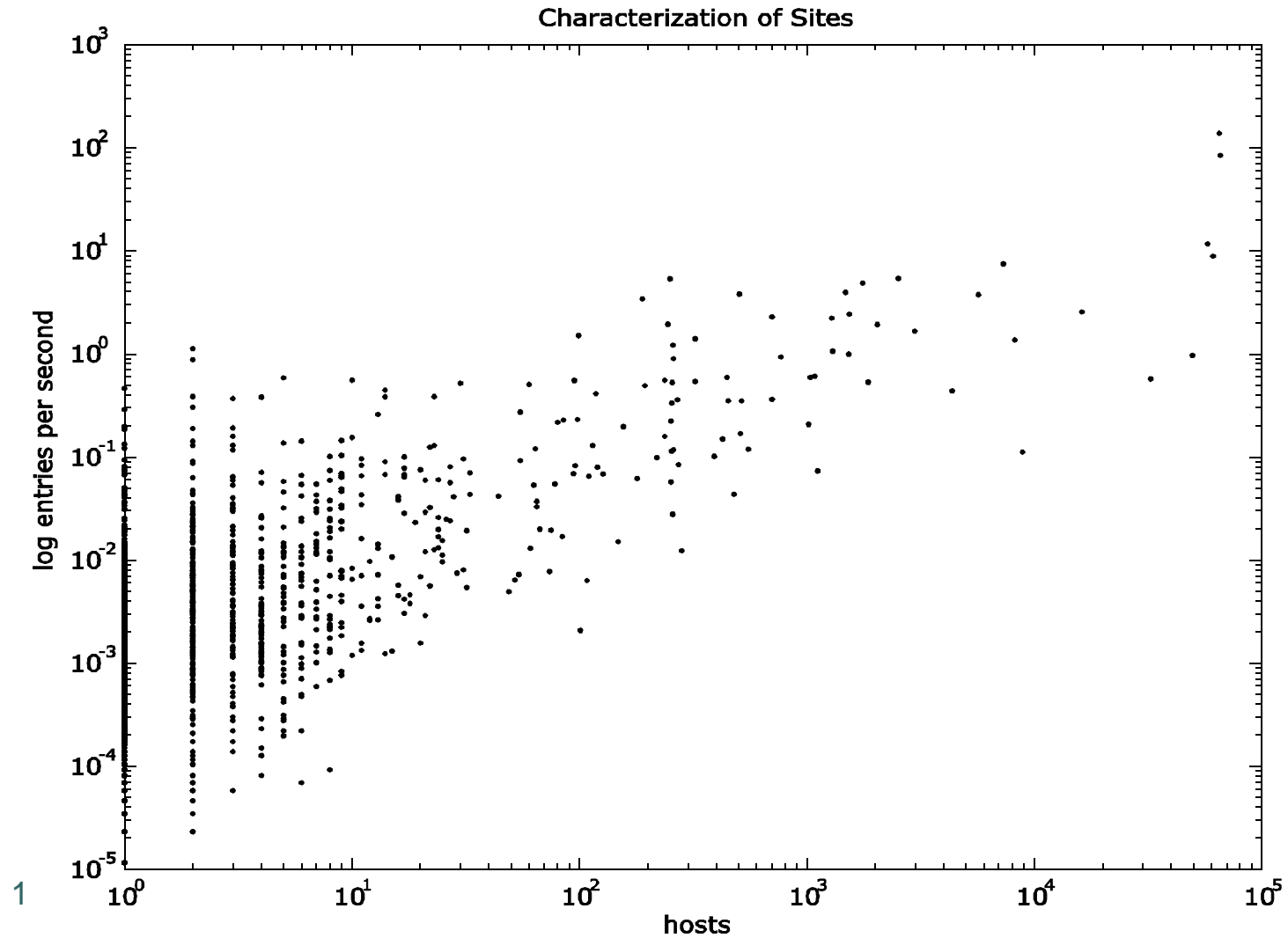


# What is DShield?

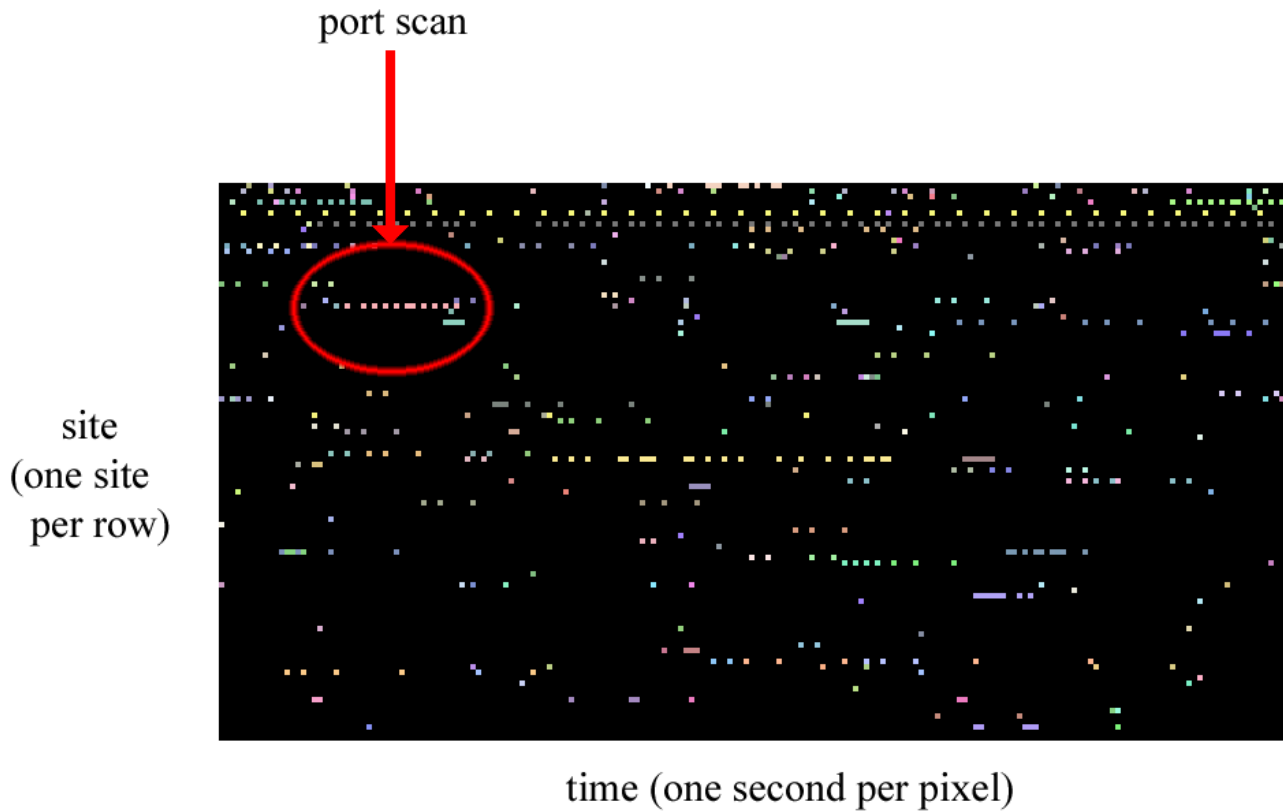
- ~30 million log entries collected per day
- Log entries are of the form
  - <time, date, submitter, source IP, source port, dest. IP, dest. port>
- Live statistics published at [www.dshield.org](http://www.dshield.org)



# DShield Sites: Size & Log Entry Rate



# Visualization of Log Data



- Each pixel, colored by source IP address, represents a failed connection
- Port scans show up as horizontal sequences of identically colored pixels

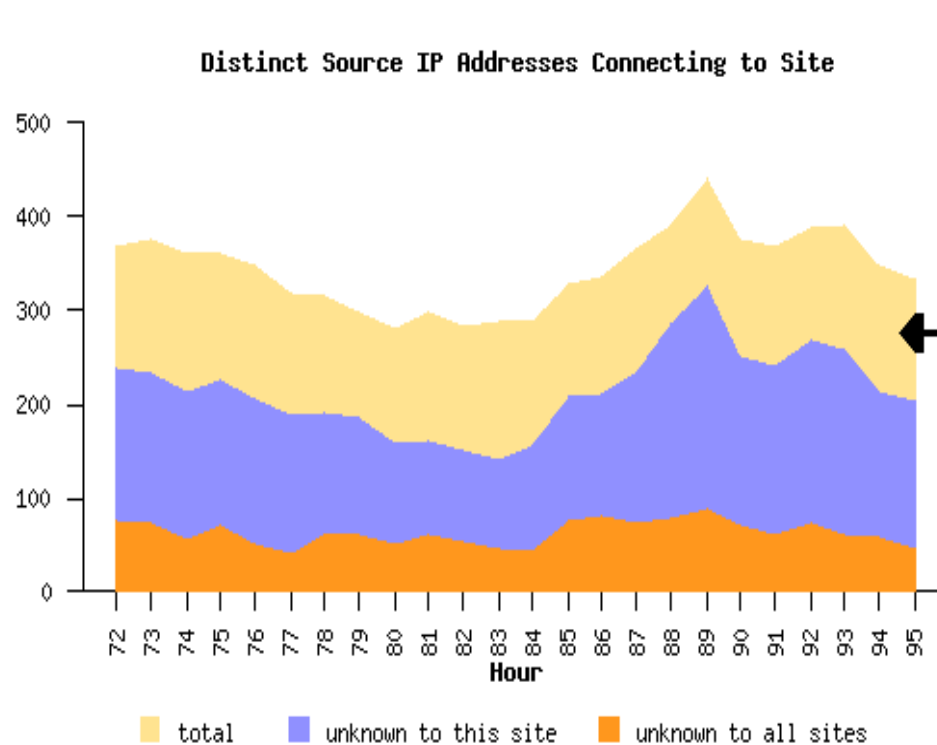


# Experimental Setup

- Network intrusion logs from 1600 sites obtained from DShield project
  - ❑ 4 consecutive days of complete log data
  - ❑ Compute for each hour:
    - Number of attack sources not in site's blacklist
    - Number of attack sources not in global blacklist

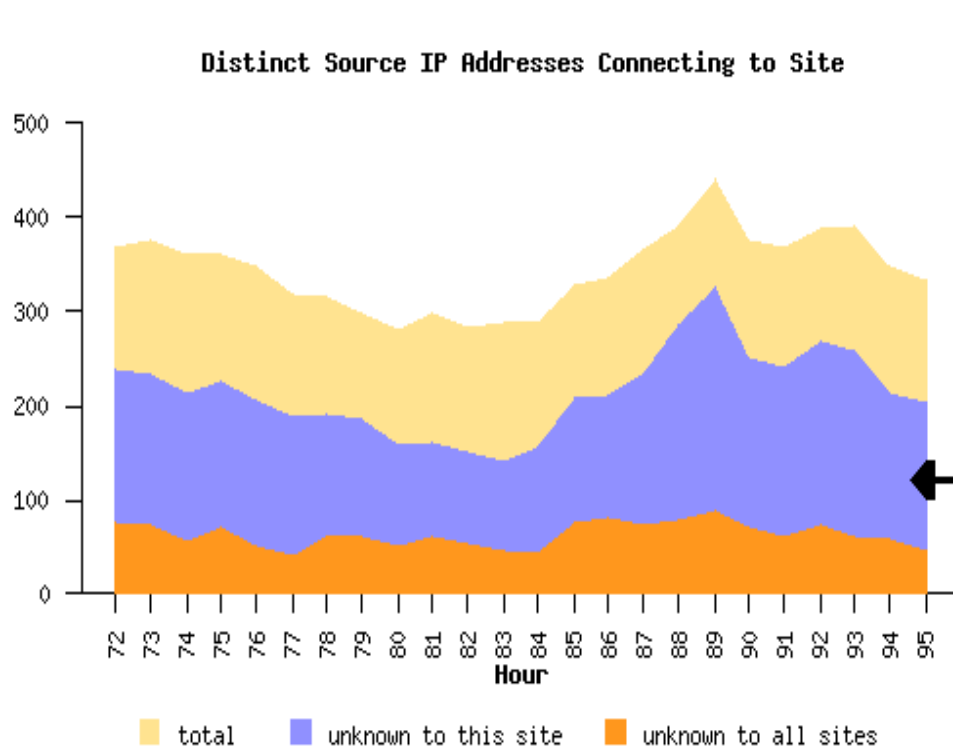
# Preliminary Results

- Connections this site can block using only its own experience



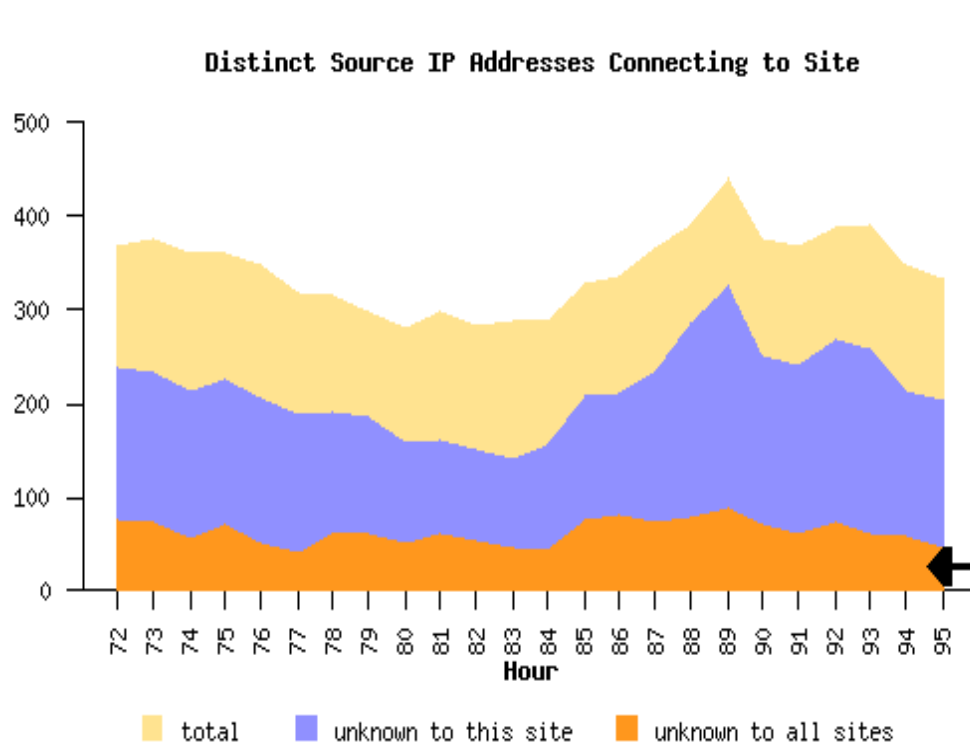
# Preliminary Results

- Connections this site can block with the help of other sites



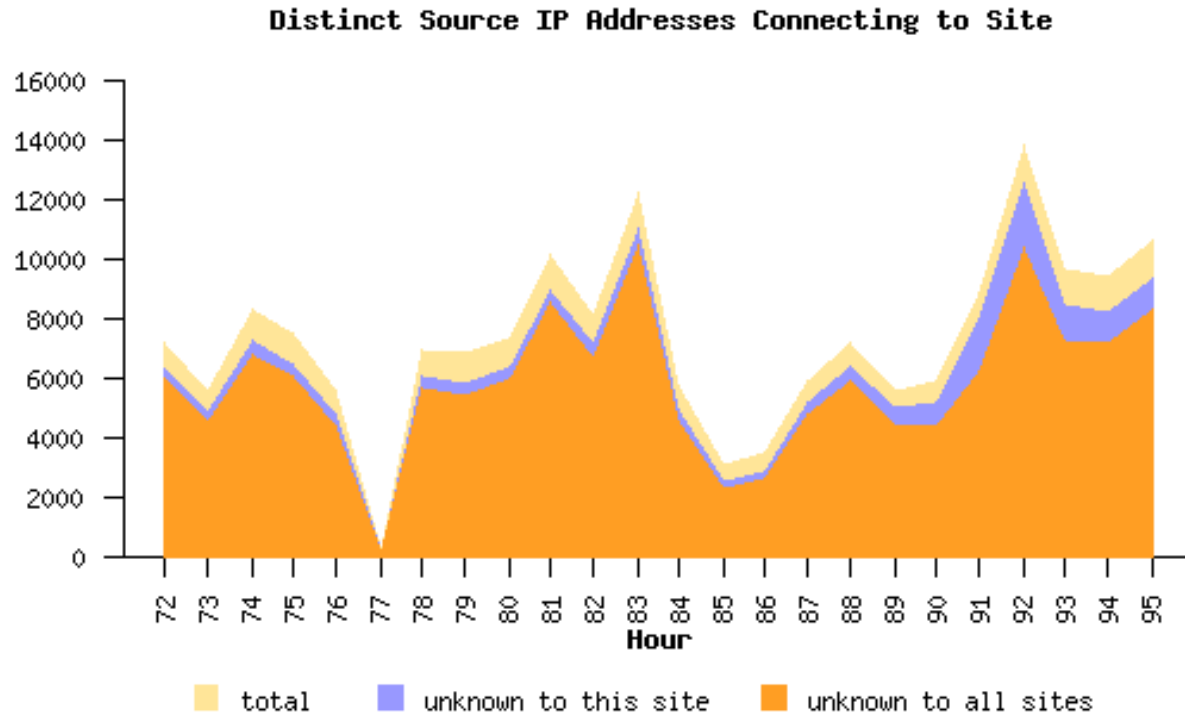
# Preliminary Results

- Connections this site cannot block with any data



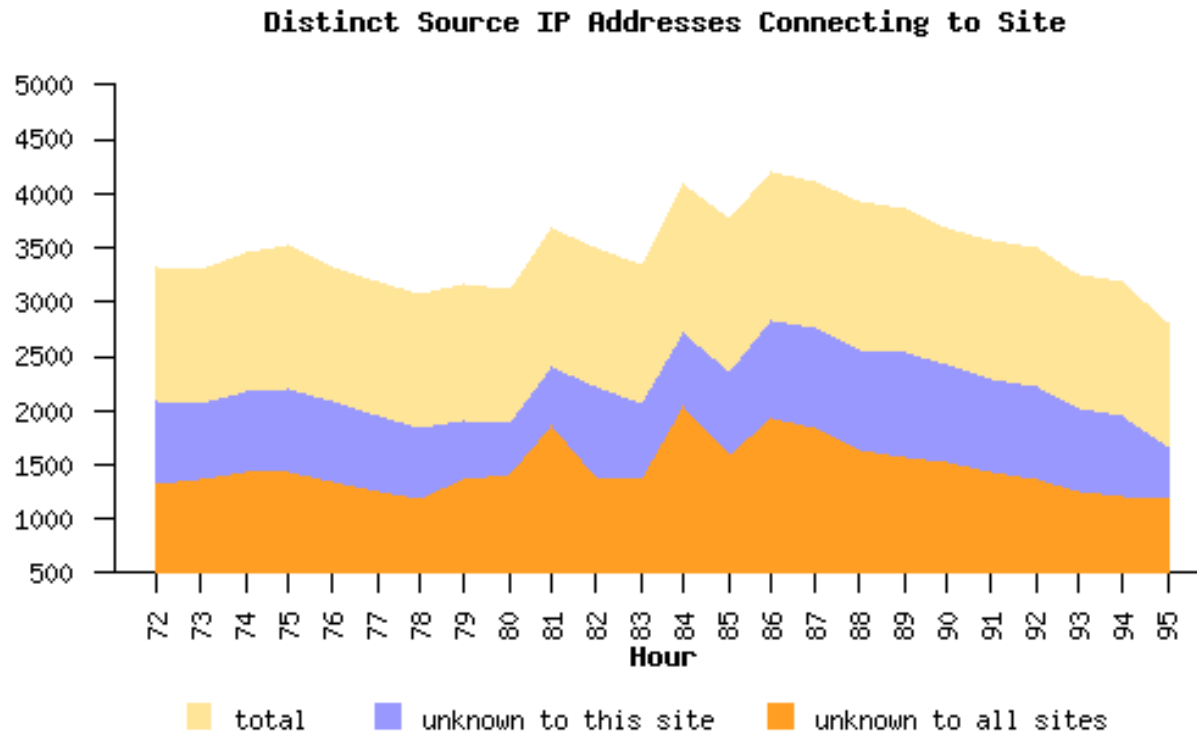
# Preliminary Results

- Site with max attack sources, max not in site blacklist & max not in global blacklist



# Preliminary Results

- Site with max absolute gain from log sharing





# On-going Work

- **Active Sink Benefit**  
maximum increase in percentage of future attacks that can be blocked using blacklist from a single active sink vs size of the sink



# Ongoing Work

- Optimal Peering Problem
  - ❑ which minimal set of peers provides maximum benefit for blocking future attacks?
  - ❑ is the optimal set of peers time-varying?
  - ❑ static or adaptive algorithm for determining the set of peers



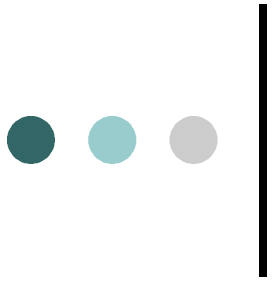
# Ongoing Work

- Essential Data Selection problem
  - ❑ What minimum subset of data exchange achieves maximum benefit for blocking future attacks?
  - ❑ Log file summary creation
  - ❑ Incremental log transmission



# Questions

- Questions?
- Comments?





# Intrusion Log Sharing

## University of Wisconsin-Madison

John Bethencourt (bethenco@cs.wisc.edu)

Jason Franklin (jfrankli@cs.wisc.edu)

Mary Vernon (vernon@cs.wisc.edu)