Stochastic Analysis of Horizontal IP Scanning

Derek Leonard, Zhongmei Yao, Xiaoming Wang, and Dmitri Loguinov

Internet Research Lab
Department of Computer Science and Engineering
Texas A&M University

March 29, 2012
Agenda

• Introduction
• Motivation
• Formalizing scanning
• Analysis of existing methods
• Stealth optimality
• Final thoughts
Introduction

- IDS (Intrusion Detection Systems) are commonly deployed to protect network assets.

- Algorithms in IDS aim to detect:
  - Malicious payload
  - Anomalous traffic patterns
  - DoS attacks
  - Scanning for open services

- To maintain scalability and adapt over time, IDS periodically expires state and performs detection using packets received only within a given time window.
Introduction 2

• To reduce false-positive rates, IDS must observe a minimum number of packets in the window before triggering an underlying estimator
  – This makes IDS oblivious to attacks that span multiple windows and never reach this threshold
  – We call such exploits stealthy

• One malicious activity whose detection is particularly sensitive to amount of IDS state is horizontal scanning
  – This entails probing of all BGP space on a given port
  – Similar techniques can be applied to vertical scanning (probing of multiple ports on a given IP)
Agenda

• Introduction
• Motivation
• Formalizing scanning
• Analysis of existing methods
• Stealth optimality
• Final thoughts
Motivation

• The only exposed technique for stealth scanning is to stretch it over several months (Staniford 2002)

• This leaves many open issues:
  – Is stealth scanning possible at faster rates?
  – For a given scan rate, with what probability will existing IDS installations notice the various types of scanners?
  – How to optimally permute the IP space during the scan?
  – How to distribute the load between multiple scanner IPs?

• We aim to address these questions through probabilistic modeling
Agenda

- Introduction
- Motivation
- Formalizing scanning
- Analysis of existing methods
- Stealth optimality
- Final thoughts
Formalizing Scanning

• Since no prior work analytically examined IDS detection rates, our first task is to develop a formalization that makes the problem tractable

• Assume $\mathcal{F} = \{0, \ldots, n - 1\}$ is the target IP space
  – For IPv4, $n = 2^{32}$ addresses, later filtered by BGP

• Attacker has access to $m$ source IPs (e.g., a botnet) from which it launches the scan
  – Not concerned with infection, only scanning
  – Thus, no new IPs are added to the botnet
Formalizing Scanning 2

• Define a scan pattern to consist of:
  - Permutation: order in which $\mathcal{F}$ is probed
  - Split: partitioning of $\mathcal{F}$ between source IPs
  - Schedule: instances when probes are transmitted

• In the literature
  - Two permutations mentioned, i.e., sequential ($\mathcal{F}$ remains intact) and uniform ($\mathcal{F}$ is randomly shuffled)
  - Split could be applied before or after permutation, but always involved contiguous chunks of space
  - Schedule amounted to constant inter-probe spacing
Formalizing Scanning 3

- Illustration of three classes of existing scan patterns
Formalizing Scanning 4

- Consider two models of IDS behavior
  - Define $\Delta$ to be window size in time units and $E$ the number of scan packets that triggers an estimator
  - Estimator is assumed to always detect the scanner

- Model IDS-A (Snort and its commercial versions)
  - Described by a separate FSM for each source IP $i$
  - FSM counts the number of unique targets probed by $i$
Formalizing Scanning 5

- Model IDS-B (Bro and certain firewalls)
  - Resets the timer each time new target is hit

- For the same pair of parameters \((\Delta, E)\), IDS-B detects all scanners that IDS-A does
  - But this comes at the expense of keeping separate timers for each source IP and longer lists of seen targets in steady-state
**Formalizing Scanning 6**

- For each source $i$, IDS can be modeled as a discrete-state stochastic process (counter) $C_i(t)$
  - Define $\tau_i(t)$ to be the **first hitting time** of $C_i(t)$ on the absorbing state $E$ after the first packet arrives from $i$
    \[ \tau_i(t) = \inf\{ t > 0 : C_i(t) = E | C_i(0) = 1 \} \]
- Assume $T$ is the fixed duration of the scan
  - Then, the number of detected scanner IPs is given by random variable $D$:
    \[ D = \sum_{i=1}^{m} 1\{\tau_i(t) < T\} \]
  - and the IDS succeeds at detecting the scan with probability $\rho(T) = P(D \geq 1)$
Formalizing Scanning 7

- Define **stealth-cover time** (SCT) to be the duration of the scan that keeps detection probability \( \rho(T) \) below some threshold \( \epsilon \)

\[
\delta = \inf \{ T > 0 : \rho(T) \leq \epsilon \}
\]

- **Main objectives:**
  - Derive \( \delta \) for existing methods (sequential, uniform) and analyze how \( m \) and pre/post-permutation splits affect it
  - Investigate the existence of **optimal** scan patterns that minimize \( \delta \) under both IDS-A and IDS-B
  - Compare the various scan techniques to each other

- Only a portion of this is covered today
Agenda

- Introduction
- Motivation
- Formalizing scanning
- Analysis of existing methods
- Stealth optimality
- Final thoughts
Analysis of Existing Methods

• Sequential scanning is very simple to analyze
  – SCT is computed for $\epsilon = 0$ (no detection):
    \[
    \delta = \Delta \frac{n}{m \zeta}, \quad \text{where} \quad \zeta = \begin{cases} 
    E - 1 & \text{IDS-A} \\
    1 & \text{IDS-B}
    \end{cases}
    \]

• Observations:
  – IDS-B requires a factor of $(E - 1)$ longer scan durations than IDS-A
  – Scan time reduces linearly with botnet size $m$

• Scan rate at all networks is constant $n / (mT)$
  – For $T = 24$ hrs and $m = 1$, this is 49.7 thousand pps
  – Clearly noticeable and intrusive
Analysis of Existing Methods 2

• Uniform scanning is more interesting
  – The paper develops a single unifying model to handle pre/post permutation splits and different botnet sizes $m$

• With certain approximations, IDS-A is tractable
  – Probability of noticing a scan at subnet $s$:
    \[
    \rho(T) \approx 1 - \left( \sum_{j=0}^{E-1} \binom{|s|}{j} q^j (1 - q)|s|-j \right)^{1/q}
    \]
  – where
    \[
    q = \frac{\Delta}{\omega T} \quad \text{and} \quad \omega = \begin{cases} 1 & \text{pre-permutation} \\ m & \text{post-permutation} \end{cases}
    \]
Analysis of Existing Methods

• Model accurate across all input parameters
Analysis of Existing Methods 5

• IDS-B is more challenging
  – Larger threshold $E$ creates non-trivial memory of previous observations of scanner probes

• Only asymptotic results are possible
  – Using the Chen-Stein theorem for sums of dependent Bernoulli variables, we have:

$$\rho(T) \approx 1 - e^{-(|s|-E+1)(1-\chi)\chi^{E-1}}$$

  – where

$$\chi = 1 - (1 - q)^{|s|}$$

  – as long as

$$\frac{(|s| - E)(1 - \chi)}{m} \gg 1$$
Analysis of Existing Methods

- Even for small subnets ($|s| = 2^8$), model is quite accurate, except when threshold $E$ is large.
Analysis of Existing Methods

• We invert both IDS-A/B models to obtain stealth cover time (SCT) $\delta$
  - After simplifications and approximations for $\epsilon \to 0$:
    $$\delta \approx \frac{|s|^{1+c}\Delta}{\omega\gamma\epsilon^c}$$
  - where
    $$c = \frac{1}{E-1}$$
    and
    $$\gamma = \begin{cases} 
      (E!)^c & \text{IDS-A} \\
      1 & \text{IDS-B}
    \end{cases}$$

• Observations
  - Compared to IDS-A, scans against IDS-B must be slower by a factor of $(E!)^c$ (rather than $E-1$ as for sequential) for the same probability of detection
Analysis of Existing Methods 8

- Pre-permutation split ($\omega = 1$) does not improve scan time with botnet size $m$; post-permutation benefits linearly

- SCT scales super-linearly $\sim |s|^{1+c}$ with subnet size
  - In fact, for $E = 2$ ($c = 1$), this rate is quadratic
  - This means that sometimes sequential is less detectable than uniform for the same scan rate!
  - Specifically, sequential is more stealthy in subnets of size

  $$|s| > \left( \frac{n\gamma \epsilon^c}{\zeta} \right)^{\frac{E-1}{E}}$$

  for $E = 2$ and $\epsilon = 10^{-3}$, this is means all /20 and larger networks

- Uniform has optimal average scanning rate
  - But on small timescales, it can be bursty
Agenda

• Introduction
• Motivation
• Formalizing scanning
• Analysis of existing methods
• Stealth optimality
• Final thoughts
Stealth Optimality

• Analysis above begs a few questions
  – Can lower SCT be achieved?
  – What is the stealthiest possible scan pattern?
  – Can both IDS-A and IDS-B be scanned with equal detection rates?

• Our solution is a new scan method we call STealth-OPtimal (STOP) that consists of 3 elements
  – A new permutation that delivers packets to all subnets maximally spaced apart (see paper)
  – A novel split that guarantees optimal spacing across multiple botnet IPs (see paper)
  – A new schedule that makes evading IDS-B as easy as IDS-A (briefly covered next)
Stealth Optimality 2

- STOP pattern seen at each subnet
  - Raises counter to $E - 1$, then delays the next burst by $\Delta$

\[ m = 2, \ E = 4 \quad \quad \quad \quad \quad m = 3, \ E = 5 \]

- Instead of one packet per $\Delta$ window, STOP can scan IDS-B (and similarly IDS-A) with $E - 1$ packets per window without detection
Stealth Optimality 3

• Requires knowledge of some lower bound $\beta$ on $E$
  - For example, no mainstream IDS utilizes $E$ less than 4
  - Some have $E$ between 20-200 (Bro, NIKSUN, Juniper)
  - The larger this lower bound $\beta$, the better STOP’s performance compared to prior methods

• STOP provably achieves the lowest possible SCT against both IDS-A and IDS-B:
  $$\delta = \frac{|s|\Delta}{m(\beta - 1)}$$
  - Linear in all parameters $m$, $|s|$, $\beta - 1$, $\Delta$

• How does this compare to existing methods?
Stealth Optimality 4

• Compared to sequential (/16 subnets, $\beta=4$)
  – STOP can scan 64K times faster against IDS-A and 196K times faster against IDS-B
  – This translates into a reduction of total scan duration $T$ from 1 year to 8 and 2.6 minutes, respectively

• Compared to uniform (/16 subnets, $\beta=4$, $\epsilon = 10^{-3}$)
  – STOP is 419 times faster against IDS-A and 1209 times faster against IDS-B
  – Reduction in $T$ from 1 year to 21 and 7 hours, respectively

• Many more results and comparisons in the paper
Agenda

• Introduction
• Motivation
• Formalizing scanning
• Analysis of existing methods
• Stealth optimality
• Final thoughts
Final Thoughts

• Linear increase in stealth with $m$ is quite peculiar
  – Suggests that hijacking unused IPs on the subnet can significantly benefit viruses
  – Aliasing $k$ IPs to the same NIC allows the host to become $k$ times stealthier in terms of SCT
  – Extra steps needed are detection of NAT and DHCP conflicts with existing hosts, but both are doable

• Methods to improve IDS?
  – While tweaking $E$ and $\Delta$ is possible, this may lead to increased false-positive rates
  – Future work will address design of new algorithms for better IDS window maintenance