Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope

Raphael Hiesgen, Marcin Nawrocki, Alistair King, Alberto Dainotti, Thomas C. Schmidt, Matthias Währisch

USENIX Security, August 12, 2022

raphael.hiesgen@haw-hamburg.de
The Share of Irregular Packets is Increasing

UCSD Network Telescope: a /9 IPv4 prefix
What is a SYN Irregularity?

- Irregular packets show one or more of:
  - High TTL ($\geq 200$)
  - No TCP options
  - Fixed IP ID (54321)

- The telescope now observes a share of roughly 75% irregular SYNs
What is a SYN Irregularity?

- Irregular packets show one or more of:
  - High TTL ($\geq 200$)
  - No TCP options
  - Fixed IP ID (54321)

- The telescope now observes a share of roughly 75% irregular SYNs
What is a SYN Irregularity?

• Irregular packets show one or more of:
  • High TTL ($\geq 200$)
  • No TCP options
  • Fixed IP ID (54321)

• The telescope now observes a share of roughly 75% irregular SYNs
What is a SYN Irregularity?

- Irregular packets show one or more of:
  - High TTL (≥200)
  - No TCP options
  - Fixed IP ID (54321)

- The telescope now observes a share of roughly 75% irregular SYNs
What is a SYN Irregularity?

Is this observation specific to the UCSD network telescope?

• The telescope now observes a share of roughly 75% irregular SYNs
A Global Phenomenon
A Global Phenomenon
A Global Phenomenon

European IXP

Share [%]

Time [D]

UCSD NT

Time [D]
A Global Phenomenon
A Global Phenomenon

Do these packets pose a threat?
Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
- Hand-crafted probes sent via raw sockets
- Recognize replies via SYN cookies

- Popularized by ZMap around 2013
- Abused by Mirai in 2016

Antonakakis et al., Understanding the Mirai Botnet, USENIX Security, 2017
Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
- Hand-crafted probes sent via raw sockets
- Recognize replies via SYN cookies

- Popularized by ZMap around 2013
- Abused by Mirai in 2016
Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
- Hand-crafted probes sent via raw sockets
- Recognize replies via SYN cookies

- Popularized by ZMap around 2013
- Abused by Mirai in 2016

Antonakakis et al., Understanding the Mirai Botnet, USENIX Security, 2017
Background: Stateless Scanning
"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
- Hand-crafted probes sent via raw sockets
- Recognize replies via SYN cookies

- Popularized by **ZMap** around 2013
- Abused by **Mirai** in 2016

---

Antonakakis et al., *Understanding the Mirai Botnet*, USENIX Security, 2017
Background: Stateless Scanning

"Scan the Internet in less than 1 hour on commodity hardware!"

- Increases scan speeds by avoiding local state
- Hand-crafted probes sent via raw sockets
- Recognize replies via SYN cookies
- Popularized by ZMap around 2013
- Abused by Mirai in 2016

How can stateless scanning be abused?

Antonakakis et al., Understanding the Mirai Botnet, USENIX Security, 2017
Two-phase Scanning

- First phase: Transport layer
  - Hand-crafted, stateless SYNs
  - *Identify responsive hosts*

- Second phase: Application layer
  - OS-level TCP handshake
  - *Deliver payloads & reconnaissance*
Two-phase Scanning

- First phase: Transport layer
  - Hand-crafted, stateless SYNs
  - Identify responsive hosts

- Second phase: Application layer
  - OS-level TCP handshake
  - Deliver payloads & reconnaissance
Spoki: Revealing Two-phase Scanners

- Spoki interacts with two-phase scanners in real time
  - Scalable system based on actors with the C++ Actor Framework (CAF)
  - Libtrace for packet ingestion, Scamper for probing
  - Collects payloads after accepting TCP connections
  - Deployed in two /24 prefixes (US, EU)

- Published source code on GitHub (https://github.com/inetrg/spoki)

Spoki rate-limits probes and uses small packets to avoid participating in DoS.
Scaling Up to 1 Million Probes Per Second

Parallel components allow Spoki to process large traffic volumes.
Share of Two-phase Sources

About 30% of sources send two-phase events each day.
Targeted Ports

Two ports are scanned exclusively in the EU.
Targeted Ports

Two ports are scanned exclusively in the EU.

TDS used by Microsoft SQL
Likely SIMATIC bug (Siemens AG)
Targeted Ports

Two ports are scanned exclusively in the EU.

- TDS used by Microsoft SQL
- Likely SIMATIC bug (Siemens AG)
- TR-069, concerns home routers
TCP Payloads

- TCP payloads are not available in traditional telescopes
- We scan payloads for *downloaders*: shell code that downloads malware

Sample names and types match known malware such as the Mozi P2P-botnet

<table>
<thead>
<tr>
<th>Event Type</th>
<th>EU</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCII</td>
<td>2,155,751</td>
<td>1,984,444</td>
</tr>
<tr>
<td>HEX</td>
<td>1,478,556</td>
<td>339,217</td>
</tr>
<tr>
<td>Downloader</td>
<td>42,303</td>
<td>143,309</td>
</tr>
</tbody>
</table>

- Spoki detected 15% of the samples earlier than VirusTotal (26% benign, 59% old)
The Maliciousness of Two-Phase Scanners

Malware distribution clearly points at malicious intent. Can we validate our findings?
The Maliciousness of Two-Phase Scanners

Malware distribution clearly points at malicious intent. Can we validate our findings?

**Approach 1: Semi-Manual Analysis**

- Reveals malicious payloads such as:

<table>
<thead>
<tr>
<th>Port</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>1433</td>
<td>TDS, SQL, SIMATIC</td>
</tr>
<tr>
<td>7545</td>
<td>TR-069, routers</td>
</tr>
<tr>
<td>5555</td>
<td>ADB crypto miner</td>
</tr>
<tr>
<td>9530, 4567</td>
<td>Embedded devices</td>
</tr>
<tr>
<td>5432</td>
<td>Realtek UPnP</td>
</tr>
</tbody>
</table>
Malware distribution clearly points at malicious intent. Can we validate our findings?

**Approach 1: Semi-Manual Analysis**

- Reveals malicious payloads such as:

<table>
<thead>
<tr>
<th>Port</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>1433</td>
<td>TDS, SQL, SIMATIC</td>
</tr>
<tr>
<td>7545</td>
<td>TR-069, routers</td>
</tr>
<tr>
<td>5555</td>
<td>ADB crypto miner</td>
</tr>
<tr>
<td>9530, 4567</td>
<td>Embedded devices</td>
</tr>
<tr>
<td>5432</td>
<td>Realtek UPnP</td>
</tr>
</tbody>
</table>

**Approach 2: Query GreyNoise**

- Classifies IPs into: *malicious*, *benign*, and *unknown*
- Two-phase events have a higher share of malicious sources:

![Graph showing malicious share](image)
Geographical Scanning Locality

- Scanners focus on different ports in **Europe** and the **USA**
- Different vendors and deployments attracts different attacks

<table>
<thead>
<tr>
<th>Payload Prefix</th>
<th>EU</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Share</td>
<td>Ports</td>
</tr>
<tr>
<td>TDS7 Pre-login</td>
<td>74.52%</td>
<td>1433</td>
</tr>
<tr>
<td>TLS Client Hello</td>
<td>4.55%</td>
<td>443, 8443</td>
</tr>
<tr>
<td>ADB Connect</td>
<td>4.97%</td>
<td>5555</td>
</tr>
<tr>
<td>SMB Negotiate</td>
<td>11.04%</td>
<td>445</td>
</tr>
<tr>
<td>PSQL/UPnP</td>
<td>0.35%</td>
<td>5432</td>
</tr>
<tr>
<td>TSAP</td>
<td>0.45%</td>
<td>102</td>
</tr>
<tr>
<td>MongoDB</td>
<td>0.27%</td>
<td>27017</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.16%</td>
<td>28967</td>
</tr>
</tbody>
</table>

**Targets non-ASCII payloads**

TDS: Tabular Data Stream used by Microsoft SQL
ADB: Android Debug Bridge
TSAP: Transport Service Access Point protocol port, used for x.400, X.500; vulnerabilities in a variety of SIMATIC devices
Topological Scanning Locality

- Six of the top-ten source prefixes in the EU share a /16 with our /24 vantage point
  - This scanning behavior is associated with botnets
  - A similar locality cannot be observed in the US

- Crosscheck (sampled) traffic at a European IXP
  - Local, irregular SYN packets in 370 prefixes (150 packets per host)
  - Very focused: 96% target 23, 7547, 8291 (multiple sources identified as MiktoTik routers)

- No correlation of /16 local, irregular SYN packets at an Asian ISP
Takeaways

• Spoki makes two-phase scanners visible
• Irregular SYNs dominate SYN on the Internet: ~75%
• Two-phase scans
  • ... act as a catalyst
  • ... are used for malicious activities
  • ... follow locality patterns
  • ... have detectable signatures
Takeaways

• Spoki makes two-phase scanners visible
• Irregular SYN dominate SYN on the Internet: ~75%
• Two-phase scans
  • ... act as a catalyst
  • ... are used for malicious activities
  • ... follow locality patterns
  • ... have detectable signatures

⇒ Short update cycles needed
Takeaways

• Spoki makes two-phase scanners visible
• Irregular SYN\textsc{s} dominate SYN\textsc{s} on the Internet: \(\sim 75\%\)
• Two-phase scans
  • ... act as a catalyst
  • ... are used for malicious activities
  • ... follow locality patterns
  • ... have detectable signatures

\(\rightarrow\) Short update cycles needed
\(\rightarrow\) Deliver a variety of malware
Takeaways

• Spoki makes two-phase scanners visible
• Irregular SYNs dominate SYNs on the Internet: ~75%
• Two-phase scans
  • ... act as a catalyst ➔ Short update cycles needed
  • ... are used for malicious activities ➔ Deliver a variety of malware
  • ... follow locality patterns ➔ Ensure your data fits your deployment
  • ... have detectable signatures
Takeaways

- Spoki makes two-phase scanners visible
- Irregular SYNs dominate SYNs on the Internet: ~75%
- Two-phase scans
  - ... act as a catalyst
  - ... are used for malicious activities
  - ... follow locality patterns
  - ... have detectable signatures
  - Short update cycles needed
  - Deliver a variety of malware
  - Ensure your data fits your deployment
  - Can be tracked and their packets filtered

Ensure your data fits your deployment
Thank you for your attention!

Find the paper, code, and artifacts at:
https://spoki.secnow.net

Contact: raphael.hiesgen@haw-hamburg.de