Synthesizing Intrusion Detection System Test Data

Jared Chandler
jared.chandler@tufts.edu
Tufts University

Adam Wick
awick@fastly.com
Fastly

LangSec 2023
Intrusion Detection Systems (IDS)

• Recognize traffic using Rules.
• Rules describe Malicious Packets

• Shared Rules allow us to recognize threats we’ve never encountered.

• We don’t share the malicious traffic to trigger those rules because it contains sensitive data.

• Without test data it’s hard to know if your IDS is working correctly.
Example Suricata Rule

alert tcp $EXTERNAL_NET any -> $HOME_NET any
(msg:"ET TROJAN Possible Metasploit Payload Common Construct Bind_API (from server)";
flow:from_server,established; content:"|60 89 e5 31|"; content:"|64 8b|"; distance:1; within:2;
content:"|30 8b|"; distance:1; within:2;
content:"|0c 8b 52 14 8b 72 28 0f b7 4a 26 31 ff|";
distance:1; within:13; content:"|ac 3c 61 7c 02 2c 20 c1 cf 0d 01 c7 e2|"; within:15; content:"|52 57 8b 52 10|";
distance:1; within:5; metadata:
former_category TROJAN; classtype:trojan-activity;
sid:2025644; rev:1; metadata:affected_product Any,
attack_target Client_and_Server, deployment
Perimeter, deployment Internet, deployment Internal,
deployment Datacenter, tag Metasploit,
signature_severity Critical, created_at 2016_05_16,
updated_at 2018_07_09;)

Research Goal

```
alert tcp $EXTERNAL_NET any -> $HOME_NET 21
(msg:"ET SCAN PRO Search Crawler Probe";
flow:to_server,established; content:"PASS ";
nocase; depth:5; content:"crawler"; nocase;
within:30; pcre:"/^PASS\s+PRO(- |\s)*search\s+Crawler/smi";
reference:url,sourceforge.net/project/showfiles.php?group_id=149797;
reference:url,doc.emergingthreats.net/2008179;
classtype:not-suspicious; sid:2008179; rev:3;
metadata:created_at 2010_07_30, updated_at 2010_07_30;)
```
Prior Work

• *Deceptive Self-Attack for Cyber Defense, HICSS 2023*
• Originally Developed to Blind an IDS / Distract an Adversary
• Bombard a network with spurious attacks
1. Signature is parsed.

2. Content generated for regular expressions.

3. Generated content and positioning constraints encoded as SMT problem.

4. Generated content positioned according to SMT solution.
Survey of Rules In Practice

- Proof Point Emerging Threats Dataset
- Open Source Corpus of over 30,000 rules
- Contributions spanning more than a decade
Survey: Results

• What we found
• Protocols people use
• Distribution over time
• What features do people use

![Bar chart showing protocol usage]

- HTTP: 18321
- DNS: 6143
- TCP: 3012
- TLS: 2431
- UDP: 388
- OTHER: 415
<table>
<thead>
<tr>
<th>Protocol</th>
<th>Named Buffer</th>
<th>Qty of Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP</td>
<td>http.uri</td>
<td>10426</td>
</tr>
<tr>
<td></td>
<td>http.method</td>
<td>6449</td>
</tr>
<tr>
<td></td>
<td>http.request_body</td>
<td>2916</td>
</tr>
<tr>
<td></td>
<td>http.host</td>
<td>2668</td>
</tr>
<tr>
<td></td>
<td>http.header_names</td>
<td>2550</td>
</tr>
<tr>
<td></td>
<td>http.user_agent</td>
<td>2354</td>
</tr>
<tr>
<td></td>
<td>http.header</td>
<td>1251</td>
</tr>
<tr>
<td></td>
<td>http.stat_code</td>
<td>813</td>
</tr>
<tr>
<td></td>
<td>http.content_type</td>
<td>605</td>
</tr>
<tr>
<td></td>
<td>http.request_line</td>
<td>326</td>
</tr>
<tr>
<td></td>
<td>http.cookie</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td>http.accept</td>
<td>126</td>
</tr>
</tbody>
</table>
Expanded Approach

• Handle named buffers

```plaintext
alert http any any - > [$HOME_NET,$HTTP_SERVERS] any (msg:"ET EXPLOIT eMerge E3 Command Injection Inbound (CVE-2019-7256)"; flow:established,to_server; 
http.method; content:"GET";
http.uri; content:"/card_scan"; startswith;
fast_pattern; content:".php"; distance:0;
within:15; content:"=|60|"; reference:cve,2019-7256; classtype:attempted-admin; sid:2033757;
rev:1; metadata:created_at 2021_08_22, cve CVE_2019_7256, former_category EXPLOIT,
updated_at 2021_08_22;)
```
http.method; content:"get";
http.uri; content:"json"; distance:4; within:10;
http.protocol; content:"3.0";
http.hostname; content "abc.com"
Synthesize Each Named Buffer Individually
Insert Into Template

Example Signature Fragment:

- `http.method;content:"get";
- `http.uri; content:"json"; distance:4; within:10;
- `http.protocol; content:"3.0";
- `http.hostname; content "abc.com"

Buffer Start Locations:

1. GET
2. http.uri buffer
3. HTTP /
4. http.protocol buffer
5. HOST:
6. http.hostname buffer

Content Copied Directly From Buffer Description:

- Relative Distance 4 bytes
- Relative Within 10 bytes

Insert Into Template
GET /xy.JSON HTTP/3.0 HOST: ABC.COM
GET /AB.JSON HTTP/13.0 HOST: FOOABC.COM
GET /WXYJSONABC HTTP/ABC123.0 HOST: ABC.COM
GET /ABCD/JSON HTTP/13.0.XYZ HOST: FOOABC.COM
Evaluation

6689 rules
Evaluation

- 6689 rules
- 6145 rules
- 92% Success rate
Interesting Failures

http.uri; content:"Flash Player.exe";
Interesting Failures

http.uri; content:"Flash Player.exe";
Interesting Failures

http.uri; content:"Flash Player.exe";
PCRE Introduced Constraints

http.uri; content:“index.html”; offset:0;

http.uri; pcre:“^index.html”;
Overlap for Performance

```plaintext
content: "foo"; offset: 0;
pcre: "fooba+r"; offset: 0;
```
Rule Debugging

content: "a"; content: "b";

content: "a"; depth: 1, content: "b"; depth: 1;

content: "a"; depth: 1, content: "cb"; depth: 2;
Next Steps

• Expand to handle more named buffers.
• More complex relationships: Move byte-string synthesis within Z3.
• Help rule authors debug rules.
• Synthesize examples which are usefully dissimilar.
Thank You! Questions?

Jared Chandler
jared.chandler@tufts.edu