nDPI: what’s new, what’s next and what is it useful for?

Ivan Nardi
Who am I?

• Ivan Nardi, @ AI2M
  – lawful interception, investigation analysis, big data retention
  – voice/IP metadata collection, processing and reporting
  – network probes and DPI

• ivan@ai2m.eu
nDPI: what is it?

• An open-source library providing:
  – deep packet inspection engine for network visibility: protocol classification, metadata extraction, flow risks computation
  • basic blocks for a cyber-security application
  – algorithms for data analysis:
    • data forecasting and anomaly detection
    • clustering and similarity evaluation
    • (sub-)string searching and IP matching
    • probabilistic data structures: bloom filters, cardinality estimation
nDPI: some statistics

- In the last year (i.e. from ntopconf22)
  - 24 contributors
  - ~600 commits
nDPI: what is it useful for?

• Protocol classification and statistics
  - small ISP: statistics about Youtube/Netflix traffic to optimize peering

• Firewall: block/allow some kind of traffic
  - with subscriber info, block/allow rules per user
  - zero-rating applications (by ISPs)

• Cybersecurity applications

• Bandwidth control & QoS
nDPI: what is it useful for?

- Block of (only) BitTorrent traffic on a VPN free plan
- Traffic control on enterprise VPNs
- Active honeypot
- Algorithms: detect performance regressions via statistical anomaly detection
nDPI: what’s new

• Usual, boring stuff:
  – be sure that old stuff is still working…
  – new flow risks and ~50 new protocols: some VPNs, some games, few streaming services…
  – better performances, less resources, better testing
  – better internal algorithms
  – add some general statistics: Patricia tree, automa, LRU cache…

• Some documentation skeletons
nDPI: what’s new

• Better identification of VoIP/RTP traffic, even when it is explicitly blocked:
  – Zoom classification
  – RTP stream type (audio, video, screen sharing)

• More algorithms: bloom filter, count-min Sketch, popcount
nDPI: what’s new

• Better custom rules (support for custom BPF protocol definition using nBPF)

• Add an heuristic to detect fully encrypted flows

• Preliminary work to handle ECH

• Detection of (illegal) gambling sites
nDPI: fuzzing support

• What is fuzzing? This bash one liner but fancier:
  - while 1; do ./a.out < /dev/urandom &; done

• oss-fuzz integration started 4 years ago
### nDPI: fuzzing support

#### Fuzzing Introspection of OSS-Fuzz projects

<table>
<thead>
<tr>
<th>Project name</th>
<th>Language</th>
<th>Fuzz target count</th>
<th>Runtime code coverage</th>
<th>Total lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>go-containerregistry</td>
<td>go</td>
<td>1</td>
<td>98.75</td>
<td>160</td>
</tr>
<tr>
<td>evo-inflectortarget</td>
<td>java</td>
<td>1</td>
<td>98.50</td>
<td>133</td>
</tr>
<tr>
<td>faed2</td>
<td>c</td>
<td>5</td>
<td>97.49</td>
<td>12185</td>
</tr>
<tr>
<td>md4c</td>
<td>c</td>
<td>1</td>
<td>97.39</td>
<td>4290</td>
</tr>
<tr>
<td>tailscale</td>
<td>go</td>
<td>1</td>
<td>96.73</td>
<td>153</td>
</tr>
<tr>
<td>uint256</td>
<td>go</td>
<td>2</td>
<td>96.62</td>
<td>1888</td>
</tr>
<tr>
<td>jsonparser</td>
<td>go</td>
<td>14</td>
<td>96.38</td>
<td>773</td>
</tr>
<tr>
<td>cgi</td>
<td>c</td>
<td>1</td>
<td>96.19</td>
<td>919</td>
</tr>
<tr>
<td>gonids</td>
<td>go</td>
<td>1</td>
<td>96.08</td>
<td>1021</td>
</tr>
<tr>
<td>nats</td>
<td>go</td>
<td>2</td>
<td>95.91</td>
<td>709</td>
</tr>
<tr>
<td>cppiter tools</td>
<td>c++</td>
<td>1</td>
<td>95.51</td>
<td>401</td>
</tr>
<tr>
<td>h3</td>
<td>c</td>
<td>20</td>
<td>94.88</td>
<td>3868</td>
</tr>
<tr>
<td>ndpi</td>
<td>c++</td>
<td>27</td>
<td>94.85</td>
<td>39480</td>
</tr>
<tr>
<td>jpeg2dec</td>
<td>c++</td>
<td>1</td>
<td>93.48</td>
<td>5657</td>
</tr>
<tr>
<td>pcre2</td>
<td>c++</td>
<td>1</td>
<td>93.48</td>
<td>16202</td>
</tr>
</tbody>
</table>
nDPI: what’s’ next

• Usual, boring stuff: more protocols, better performances, more flow risks, more configuration options, better multi-core support...

• Significant improvements on BitTorrent, STUN (i.e. VoIP apps) and VPN traffic

• A new algorithm to detect DGA domains?

• Better handling of asymmetric traffic