High-speed Network and Service Monitoring

Luca Deri <deri@ntop.org>
Who’s ntop.org?

• Started in 1998 as open-source monitoring project for developing an easy to use passive monitoring application.

• Several project spin-off
  – Accelerated packet capture
  – 1 and 10 Gbit packet capture
  – NetFlow/sFlow probes
  – Peer-to-Peer VPN (n2n)
ntop.org at a Glance
Who is Using ntop Products?
Some Challenges

• SNMP is good for element management (e.g. router and server monitoring) but poor for traffic measurement.
• Not all routers/switches speak NetFlow/sFlow: we need to deploy soft probes.
• 1 and 10 Gbit networks can produce a lot of monitoring data: our monitoring apps must be able to handle all this traffic.
Networks are Changing… [1/2]
Networks are Changing… [2/2]

• Without edge control there’s no real network control.

• Central traffic monitoring isn’t enough anymore: not all traffic flows through the center.

• Edge equipment is often very basic and it means that there’s no visibility at the edge: think about this before purchasing your network equipment.
Typical Monitoring Deployment: LAN
Typical Monitoring Deployment: Internet Traffic
Typical Monitoring Deployment: Cloud and Intra-VM Monitoring
Some Lessons Learnt

• In order to monitoring the traffic we need to deploy a probe where the traffic is flowing.

• We need to make sure we can handle both NetFlow and sFlow if we want to have complete network visibility.

• Cloud computing and server virtualization push us to monitor in-VM virtual networks.
How can ntop.org help me?

• ntop can act as central network monitoring console.
• nProbe can act as sFlow and NetFlow proxy/probe/collector.
• 10 Gbit packet capture acceleration and filtering, in host and VMs, using PF_RING.
• Ability to query billion of flows with sub-second response time.
A Web-based Monitoring Console
What is ntop ?
### Local Hosts Characterization

<table>
<thead>
<tr>
<th>Host</th>
<th>Unhealthy Host</th>
<th>L2 Switch</th>
<th>Gateway</th>
<th>Printer</th>
<th>NTP/DNS Server</th>
<th>SMTP/POP/MAP Server</th>
<th>Directory/FTP/HTTP Server</th>
<th>DHCP/WINS Server</th>
<th>DHCP Client</th>
<th>P2P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host059-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host062-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host053-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host003-160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host005-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host029-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host028-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dns03.ablia.net</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dns02.ablia.net</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dns01.ablia.net</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host119-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host118-160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host117-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host074-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host073-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host065-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host063-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>host062-160</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Traffic Trends

Current Day Throughput

Max: 179.4M  Avg: 112.3M  Last: 69.5M

Last Month Throughput

Max: 173.8M  Avg: 102.7M  Last: 53.6M
### Host Health

<table>
<thead>
<tr>
<th>Data Rcvd Stats</th>
<th>0 %</th>
<th>100 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP vs. Non-IP Rcvd</td>
<td>IP 100 %</td>
<td>Non-IP 0 %</td>
</tr>
<tr>
<td>Sent vs. Rcvd Pkts</td>
<td>Sent 51.8 %</td>
<td>Rcvd 48.2 %</td>
</tr>
<tr>
<td>Sent vs. Rcvd Data</td>
<td>Sent 33.2 %</td>
<td>Rcvd 66.8 %</td>
</tr>
<tr>
<td>Host Type</td>
<td>Name Server</td>
<td></td>
</tr>
<tr>
<td>Historical Data</td>
<td>[ ]</td>
<td></td>
</tr>
<tr>
<td>Host Healthness (Risk Flags)</td>
<td>![Risk Flags]</td>
<td></td>
</tr>
</tbody>
</table>

1. **Unexpected packets (e.g. traffic to closed port or connection reset):**
   - [Rcvd: rejected]
   - [Rcvd: port unreac]
   - [Rcvd: hostnet unreac]

### Host Traffic Stats

<table>
<thead>
<tr>
<th>Time</th>
<th>Tot. Traffic Sent</th>
<th>% Traffic Sent</th>
<th>Tot. Traffic Rcvd</th>
<th>% Traffic Rcvd</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 AM</td>
<td>13.4 MB</td>
<td>74.7 %</td>
<td>26.6 MB</td>
<td>74.0 %</td>
</tr>
<tr>
<td>10 AM</td>
<td>4.5 MB</td>
<td>25.3 %</td>
<td>9.3 MB</td>
<td>26.0 %</td>
</tr>
<tr>
<td>9 AM</td>
<td>0</td>
<td>0.0 %</td>
<td>0</td>
<td>0.0 %</td>
</tr>
<tr>
<td>8 AM</td>
<td>0</td>
<td>0.0 %</td>
<td>0</td>
<td>0.0 %</td>
</tr>
</tbody>
</table>
# VoIP Support

<table>
<thead>
<tr>
<th>Client</th>
<th>Server</th>
<th>Data Sent</th>
<th>Data Rcvd</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>130.192.225.34</td>
<td>130.192.225.44</td>
<td>58.6 KB</td>
<td>70.3 KB</td>
<td>valter called livio</td>
</tr>
<tr>
<td>130.192.225.34</td>
<td>130.192.225.44</td>
<td>224</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>stun01.sipphone.com</td>
<td>130.192.225.34</td>
<td>216</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>130.192.225.34</td>
<td>bill.ipv6.polito.it</td>
<td>2.8 KB</td>
<td>2.3 KB</td>
<td>valter called livio</td>
</tr>
<tr>
<td>130.192.225.44</td>
<td>bill.ipv6.polito.it</td>
<td>4.5 KB</td>
<td>5.0 KB</td>
<td>valter called livio</td>
</tr>
<tr>
<td>130.192.225.44</td>
<td>130.192.225.34</td>
<td>462</td>
<td>361</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Host Type</th>
<th>VoIP Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Users</td>
<td>stefano &lt;101&gt;</td>
</tr>
</tbody>
</table>
ntop Scripting using Python

http://ntop.local:3000/python/hello.py

handlePythonHTTPRequest(...)

<html>
  ...
  </body>
</html>
Where is my traffic going to?
Flow-based Traffic Monitoring: nProbe
nProbe: IPFIX/NetFlow Probe

Packet Capture → nProbe → Data Dump → Raw Files / MySQL / SQLite / FastBit

sFlow → nProbe → Flow Export
nProbe: Main Features

• Ability to keep up with Gbit speeds on Ethernet networks handling thousand of packets per second without packet sampling on commodity hardware.

• Support for major OS including Unix, Windows and MacOS X.

• Full NetFlow v9/IPFIX and sFlow (no probe) support

• V9 extensions: payload, network/application latency, VoIP, RTP.

• Ability to extend the probe with user-written plugins.

• BGP Peering with the router for full AS monitoring.
Problem Statement [1/2]

• NetFlow and sFlow are the current state-of-the-art standard for network traffic monitoring.
• As the number of generated flows can be quite high, operators often use sampling in order to reduce their number.
• Sampling leads to inaccuracy so it cannot always be used in production networks.
• Thus network operators have to face the problem of collecting and analyzing a large number of flow records.
Problem Statement [2/2]

Where to store collected flows?

- Relational Databases
  - Pros: Expressiveness of SQL for data search.
  - Cons: Sacrifice flow collection speed and query response time.

- Raw Disk Archives
  - Pros: Efficient flow-to-disk collection speed (> 250K flow/s).
  - Cons: Limited query facilities as well search time proportional to the amount of collected data (i.e. no indexing is used).
What are we looking for?

• Ability to execute multidimensional queries on arbitrary large amounts of data with response time in the order of seconds (in many cases, milliseconds).

• Efficient yet simple flow record storage architecture in terms of disk space, query response time, and data collection duration.

• A system that operates on raw flow records without first reducing or summarizing them.

• The reduction of the time needed to explore a large dataset and the possibility to display query results in real-time, making the exploration process truly interactive.
nProbe + FastBit

- FastBit is not a database but a C++ library that implements efficient bitmap indexing methods.
- Data is represented as tables with rows and columns.
- A large table may be partitioned into many data partitions and each of them is stored on a distinct directory, with each column stored as a separated file in raw binary form.
- nProbe natively integrates FastBit support and it automatically creates the DB schema according to the flow records template.
- When a partition is fully dumped, columns to be indexed are first sorted then indexed.
## Handling Billion of Flows

### nProbe+FastBit vs MySQL

<table>
<thead>
<tr>
<th>Query</th>
<th>MySQL</th>
<th>nProbe + FastBit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>22.6</td>
<td>5.6</td>
</tr>
<tr>
<td>Q2</td>
<td>69</td>
<td>0.5</td>
</tr>
<tr>
<td>Q3</td>
<td>971</td>
<td>12.5</td>
</tr>
<tr>
<td>Q4</td>
<td>1341</td>
<td>48.2</td>
</tr>
<tr>
<td>Q5</td>
<td>2257</td>
<td>30.7</td>
</tr>
</tbody>
</table>

### nProbe+FastBit vs nfdump

<table>
<thead>
<tr>
<th></th>
<th>nProbe+FastBit</th>
<th>nfdump</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45</td>
<td>1500</td>
</tr>
</tbody>
</table>

All measurements are in seconds
How to Add Geolocation Data [1/2]

• Routers are unable to export any geolocation information.

• NetFlow/IPFIX flows do not contain any information about geolocation into standard flow formats.

• Solution:
  – Let the collector add geolocation information to flows received by routers
  – Let nProbe export this information to collectors.
How to Add Geolocation Data [2/2]

• nProbe takes advantage of GeoIP library (GPL) to:
  – Add geolocation information to flows
  – Map IP addresses to ASN (Autonomous System Numbers) for adding ASN awareness.
  – GeoIPASNum.dat (ASN)
  – GeoLiteCity.dat (GeoLocation)
# Constructor
$update = Net::BGP::Update->new(
    NLRI => [ qw( 10/8 172.168/16 ) ],
    Withdraw => [ qw( 192.168.1/24 172.10/16 192.168.2.1/32 ) ],
    # For Net::BGP::NLRI
    Aggregator => [ 64512, '10.0.0.1' ],
    AsPath => [ 64512, 64513, 64514 ],
    AtomicAggregate => 1,
    Communities => [ qw( 64512:10000 64512:10001 ) ],
    LocalPref => 100,
    MED => 200,
    NextHop => '10.0.0.1',
    Origin => INCOMPLETE,
);
### HTTP Traffic Analysis

**Plugin HTTP Protocol Dissector templates:**

- `[NFv9 57652][IPFIX 35632.180] %HTTP_URL` HTTP URL
- `[NFv9 57653][IPFIX 35632.181] %HTTP_RET_CODE` HTTP return code (e.g. 200, 304...)
- `[NFv9 57654][IPFIX 35632.182] %HTTP_REFERER` HTTP Referer
- `[NFv9 57655][IPFIX 35632.183] %HTTP_UA` HTTP User Agent
- `[NFv9 57656][IPFIX 35632.184] %HTTP_MIME` HTTP Mime Type

<table>
<thead>
<tr>
<th>#</th>
<th>Client</th>
<th>Server</th>
<th>Protocol</th>
<th>Method</th>
<th>URL</th>
<th>HTTPReturnCode</th>
<th>Location</th>
<th>Referrer</th>
<th>UserAgent</th>
<th>ContentType</th>
<th>Bytes</th>
<th>BeginTime</th>
<th>EndTime</th>
<th>Flow</th>
<th>Hash</th>
<th>Cookie</th>
<th>Terminator</th>
<th>ApplLatency</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.0.200</td>
<td>api.leoslyrics.com</td>
<td>http</td>
<td>GET</td>
<td>/api_search.php?auth=mindquirk_harmonic&amp;artist=Franco+Battiato&amp;songtitle=Povera+Patria</td>
<td>200</td>
<td>curl/7.13.1 (powerpc-apple-darwin8.0) libcurl/7.13.1 OpenSSL/0.9.7i zlib/1.2.3 text/xml</td>
<td>10244</td>
<td>1133966831.996</td>
<td>1133966832.910</td>
<td>2423982224</td>
<td>0</td>
<td>C</td>
<td>0.152</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.168.0.200</td>
<td>elyrics.net</td>
<td>http</td>
<td>GET</td>
<td>/go/f/Franco-Battiato-lyrics/Povera-Patria-lyrics/</td>
<td>302</td>
<td>ควลที่7.13.1 (powerpc-apple-darwin8.0) libcurl/7.13.1 OpenSSL/0.9.7i zlib/1.2.3 text/html</td>
<td>1186</td>
<td>1133966832.527</td>
<td>1133966832.908</td>
<td>2413138730</td>
<td>0</td>
<td>S</td>
<td>0.114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.168.0.200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>http</td>
<td>GET</td>
<td>/images/filewave01.gif</td>
<td>200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>27750</td>
<td>1133966828.928</td>
<td>1133966830.606</td>
<td>26992029</td>
<td>0</td>
<td>S</td>
<td>0.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.168.0.200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>http</td>
<td>GET</td>
<td>/images/iwas01b.gif</td>
<td>200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>12469</td>
<td>1133966828.574</td>
<td>1133966829.932</td>
<td>26992028</td>
<td>0</td>
<td>S</td>
<td>0.369</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.168.0.200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>http</td>
<td>GET</td>
<td>/images/filewave02.gif</td>
<td>200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>25505</td>
<td>1133966827.681</td>
<td>1133966829.196</td>
<td>26992027</td>
<td>0</td>
<td>S</td>
<td>0.387</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192.168.0.200</td>
<td><a href="http://www.macintouch.com">www.macintouch.com</a></td>
<td>http</td>
<td>GET</td>
<td>/</td>
<td>200</td>
<td>Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13 image/gif</td>
<td>52474</td>
<td>1133966827.127</td>
<td>1133966829.339</td>
<td>26992026</td>
<td>0</td>
<td>S</td>
<td>0.308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2011 - ntop.org
nProbe: VoIP Traffic Analysis

Plugin RTP templates:

[NFv9 57622][IPFIX 35632.150] %RTP_FIRST_SSRC  First flow RTP Sync Source ID
[NFv9 57623][IPFIX 35632.151] %RTP_FIRST_TS  First flow RTP timestamp
[NFv9 57624][IPFIX 35632.152] %RTP_LAST_SSRC  Last flow RTP Sync Source ID
[NFv9 57625][IPFIX 35632.153] %RTP_LAST_TS  Last flow RTP timestamp
[NFv9 57626][IPFIX 35632.154] %RTP_IN_JITTER  RTP Jitter (ms * 1000)
[NFv9 57627][IPFIX 35632.155] %RTP_OUT_JITTER  RTP Jitter (ms * 1000)
[NFv9 57628][IPFIX 35632.156] %RTP_IN_PKT_LOST  Packet lost in stream
[NFv9 57629][IPFIX 35632.157] %RTP_OUT_PKT_LOST  Packet lost in stream
[NFv9 57630][IPFIX 35632.158] %RTP_OUT_PAYLOAD_TYPE  RTP payload type
[NFv9 57631][IPFIX 35632.159] %RTP_IN_MAX_DELTA  Max delta (ms*100) between consecutive pkts
[NFv9 57632][IPFIX 35632.160] %RTP_OUT_MAX_DELTA  Max delta (ms*100) between consecutive pkts

Plugin SIP templates:

[NFv9 57602][IPFIX 35632.130] %SIP_CALL_ID  SIP call-id
[NFv9 57603][IPFIX 35632.131] %SIP_CALLING_PARTY  SIP Call initiator
[NFv9 57604][IPFIX 35632.132] %SIP_CALLED_PARTY  SIP Called party
[NFv9 57605][IPFIX 35632.133] %SIP_RTP_CODECS  SIP RTP codecs
[NFv9 57606][IPFIX 35632.134] %SIP_INVITE_TIME  SIP SysUptime (msec) of INVITE
[NFv9 57607][IPFIX 35632.135] %SIP_TRYING_TIME  SIP SysUptime (msec) of Trying
[NFv9 57608][IPFIX 35632.136] %SIP_RINGING_TIME  SIP SysUptime (msec) of RINGING
[NFv9 57609][IPFIX 35632.137] %SIP_OK_TIME  SIP SysUptime (msec) of OK
[NFv9 57610][IPFIX 35632.138] %SIP_BYE_TIME  SIP SysUptime (msec) of BYE
[NFv9 57611][IPFIX 35632.139] %SIP_RTP_SRC_IP  SIP RTP stream source IP
[NFv9 57612][IPFIX 35632.140] %SIP_RTP_SRC_PORT  SIP RTP stream source port
[NFv9 57613][IPFIX 35632.141] %SIP_RTP_DST_IP  SIP RTP stream dest IP
[NFv9 57614][IPFIX 35632.142] %SIP_RTP_DST_PORT  SIP RTP stream dest port
nProbe: Further Traffic Analysis

Plugin DNS Protocol Dissector templates:
[NFv9 57677][IPFIX 35632.205] %DNS_QUERY
[NFv9 57678][IPFIX 35632.206] %DNS_QUERY_ID
[NFv9 57679][IPFIX 35632.207] %DNS_QUERY_TYPE
[NFv9 57680][IPFIX 35632.208] %DNS_RET_CODE
[NFv9 57681][IPFIX 35632.209] %DNS_NUM_ANSWER

Plugin MySQL Plugin templates:
[NFv9 57667][IPFIX 35632.195] %MYSQL_SERVER_VERSION
[NFv9 57668][IPFIX 35632.196] %MYSQL_USERNAME
[NFv9 57669][IPFIX 35632.197] %MYSQL_DB
[NFv9 57670][IPFIX 35632.198] %MYSQL_QUERY
[NFv9 57671][IPFIX 35632.199] %MYSQL_RESPONSE
nProbe: Network Performance and Response Time

© 2011 - ntop.org
# nProbe: Network Awareness

<table>
<thead>
<tr>
<th>SysUpTime</th>
<th>HTTP RET CODE</th>
<th>HTTP URL</th>
<th>ingressInterface</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>200</td>
<td><a href="http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/">http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/</a></td>
<td>35</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td><a href="http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/">http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/</a></td>
<td>35</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td><a href="http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL">http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL</a></td>
<td>35</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td><a href="http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL">http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL</a></td>
<td>35</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td><a href="http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=KPSM">http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=KPSM</a></td>
<td>35</td>
</tr>
</tbody>
</table>
How can I Improve my Internet Presence?

<table>
<thead>
<tr>
<th>Distance: 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASN</strong></td>
<td><strong>AS Name</strong></td>
<td><strong>Traffic</strong></td>
</tr>
<tr>
<td>2597</td>
<td>REGISTRO CTRLD IT</td>
<td>738.8 MB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distance: 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASN</strong></td>
<td><strong>AS Name</strong></td>
<td><strong>Traffic</strong></td>
</tr>
<tr>
<td>3356</td>
<td>Level 3 Communications, LLC</td>
<td>392.0 MB</td>
</tr>
<tr>
<td>12637</td>
<td>Seewebe Srl</td>
<td>3.1 MB</td>
</tr>
<tr>
<td>137</td>
<td>GARR Italian academic and research network</td>
<td>15 MB</td>
</tr>
<tr>
<td>21309</td>
<td>ACANTHO SPA</td>
<td>420.7 KB</td>
</tr>
<tr>
<td>64862</td>
<td>??</td>
<td>190.1 KB</td>
</tr>
<tr>
<td>21056</td>
<td>Welcome Italia S.p.A.</td>
<td>30.5 KB</td>
</tr>
<tr>
<td>16004</td>
<td>MIX S.r.l.</td>
<td>9.8 KB</td>
</tr>
<tr>
<td>15469</td>
<td>Warinet NOC AS</td>
<td>356.0 bytes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distance: 3</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASN</strong></td>
<td><strong>AS Name</strong></td>
<td><strong>Traffic</strong></td>
</tr>
<tr>
<td>9020</td>
<td>INTECOnet</td>
<td>139.0 MB</td>
</tr>
<tr>
<td>9035</td>
<td>Wind Telecomunicazioni spa</td>
<td>17.4 MB</td>
</tr>
<tr>
<td>6453</td>
<td>Teleglobe Inc.</td>
<td>16.0 MB</td>
</tr>
<tr>
<td>1273</td>
<td>Cable &amp; Wireless Deutschland GmbH</td>
<td>15.4 MB</td>
</tr>
<tr>
<td>3549</td>
<td>Global Crossing</td>
<td>13.4 MB</td>
</tr>
<tr>
<td>702</td>
<td>UUNET - Commercial IP service provider in Europe</td>
<td>10.7 MB</td>
</tr>
<tr>
<td>24940</td>
<td>Hetzner Online AG RZ-Nuernberg</td>
<td>9.2 MB</td>
</tr>
<tr>
<td>6762</td>
<td>Telecom Italia International high speed</td>
<td>9.0 MB</td>
</tr>
<tr>
<td>8218</td>
<td>Neo Telecoms</td>
<td>7.2 MB</td>
</tr>
<tr>
<td>286</td>
<td>KPNQwest Backbone AS</td>
<td>6.8 MB</td>
</tr>
<tr>
<td>7473</td>
<td>Singapore Telecom</td>
<td>6.0 MB</td>
</tr>
<tr>
<td>1299</td>
<td>TeTiaNet Global Network</td>
<td>4.6 MB</td>
</tr>
<tr>
<td>1239</td>
<td>Sprint</td>
<td>4.1 MB</td>
</tr>
<tr>
<td>10310</td>
<td>Yahoo!</td>
<td>4.0 MB</td>
</tr>
</tbody>
</table>
Interactive Data Search
Packet Capture Acceleration: PF_RING
Socket Packet Ring (PF_RING)

Application A

mmap() → Outgoing Packets

Read Index

Socket (ring)

Write Index

Application Z

Socket (ring)

Network Adapter

Incoming Packets

Userspace

Kernel

© 2011 - ntop.org
PF_RING: Benefits

• It creates a straight path for incoming packets in order to make them first-class citizens.
• No need to use custom network cards: any card is supported.
• Transparent to applications: legacy applications need to be recompiled in order to use it.
• No kernel or low-level programming is required.
• Developers familiar with network applications can immediately take advantage of it without having to learn new APIs.
PF_RING Packet Journey [1/2]

Packet Received → Parse Packet (up to layer 4)

Defragment packet (optional)

Same as above for PF_RING socket clusters

Added the packet to PF_RING sockets that potentially match it (packet and socket device match)

Return control to the kernel
PF_RING Packet Journey [2/2]

Add Packet to PF_RING → Packet Filtering

Sampling Rate Check

Queue Packet on PF_RING

PF_RING Reflector Check

Back to PF_RING
PF_RING in VMs: vPF_RING

KVM Hypervisor

VM Guest

KVM Kernel Modules

PF_RING

© 2011 - ntop.org
PF_RING Packet-to-Disk: n2disk

- n2disk Controller
  - Commands
  - Responses
  - Notifications

- Userspace
  - Kernel

- Packet Buffer
  - Kernel Thread
  - VFS

<table>
<thead>
<tr>
<th>n2disk (only userland)</th>
<th>750 Kpps</th>
<th>Quad-Xeon with 8 RAID disks</th>
</tr>
</thead>
<tbody>
<tr>
<td>n2disk (kernel+userland)</td>
<td>650 Kpps</td>
<td>Atom with single SATA Disk</td>
</tr>
</tbody>
</table>

© 2011 - ntop.org
Towards 10 Gbit Monitoring
Modern Networking Architectures

- Optimized TCP/IP stack
- Affinitized network data flow for balanced computing across multiple CPUs
- Chipset
- Memory
- Enhanced direct memory access (DMA) for more efficient memory copies
- Network Data Stream
Scaling to 10 Gbit: Divide et Impera

- CPU manufacturers are scaling with multicore.

- **Multicore equations:**
  - more cores = more total CPU power
  - more cores = less single core power

- Software scales with multicore only if it can exploit it:
  - multiprocess or multithread

- A “simply faster” 10 Gbit NIC is not enough:
  - one 10G card means that several threads need to compete for packets hence that a lot of time will be wasted on semaphores
Multicore+Networking Design Flaws

![Diagram showing Multicore Networking Design Flaws](image)

- **Thread**: Connections between Userland and Linux Device Driver
- **RX Queue**: Single Resource Competition, Merge & Split
- **Linux Device Driver**: Sequential Queue Polling
- **MSI-X**: RSS (Resource Side Scaling)
- **RSS (Resource Side Scaling)**
- **10 Gbit PHY**

© 2011 - ntop.org
PF_RING+TNAPI

Thread
Thread
Thread
Thread

RX Queue
RX Queue
RX Queue
RX Queue

PF_RING

RSS (Resource Side Scaling)
[Hardware per-flow Balancing]

1 Gbit / 10 Gbit NIC

Userland

No Mutex Needed

Virtual PF_RING Ethernet Queue

Threaded Polling

© 2011 - ntop.org
Intel 82599 Ethernet Controller [1/2]

- Latest generation of Intel 10 Gbit Ethernet Controller.
- Ability do define up to 32,000 perfect rules per port (unlimited hashing rules).
- Commodity adapter (<350 USD/port).
- Hardware support for virtualization (i.e. in-NIC L2 Switch) and multi RX/TX queues.
- Limitation: OSs exploits only basic NIC capabilities.
Intel 82599 Ethernet Controller [2/2]

- In 82599 packet filtering is performed in hardware at wire rate.
- Filtering is necessary to decide to which RX queue a packet must be assigned.
- Assigning a packet to a non-existing RX queue (\(\leq\) number of available CPU cores) drops the packet.
Using Hardware Filters in Real Life

- Signaling-based realtime multimedia (e.g. VoIP, IPTV) monitoring.
- Network Troubleshooting: Wireshark.
- Traffic Classification and Balancing.
- Lawful Interception of IP Traffic.
- 10 Gbit Firewalling.
Towards 10 Gbit Snorting

User space

Kernel space

PF_RING

RX 0 Queue

RX 1 Queue

RX 2 Queue

RX 3 Queue

RSS (Resource Side Scaling)

1-10 Gbit PHY

© 2011 - ntop.org
Divide et Impera

Network Monitoring Servers

10 Gbit Traffic Splitter / Switch

Selective Packet Drop on 82599

10 Gbit Ingress Stream
References

• Home Page:
  http://www.ntop.org/

• Platforms:
  Win32 (except PF_RING) and Unix.

• License:
  Gnu Public License (GPL) and Commercial.