High-speed Network and Service Monitoring

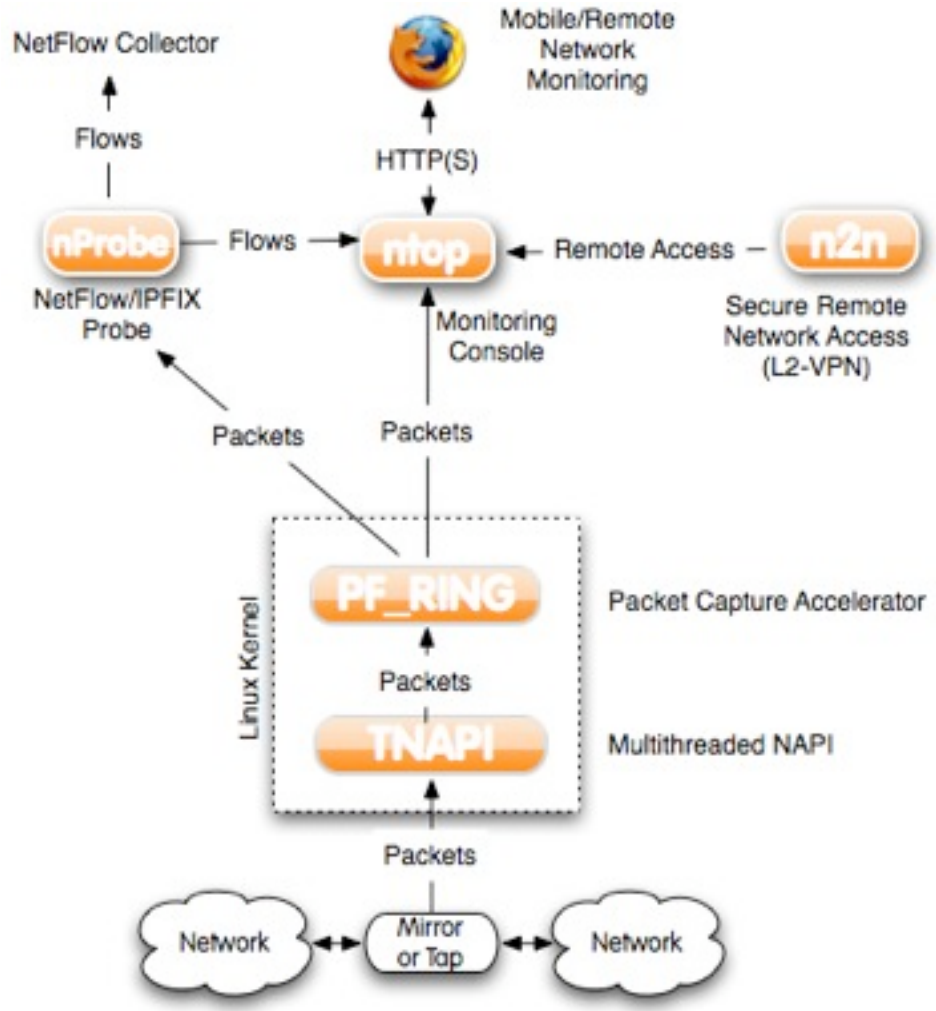
Luca Deri <deri@ntop.org>

ntop

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



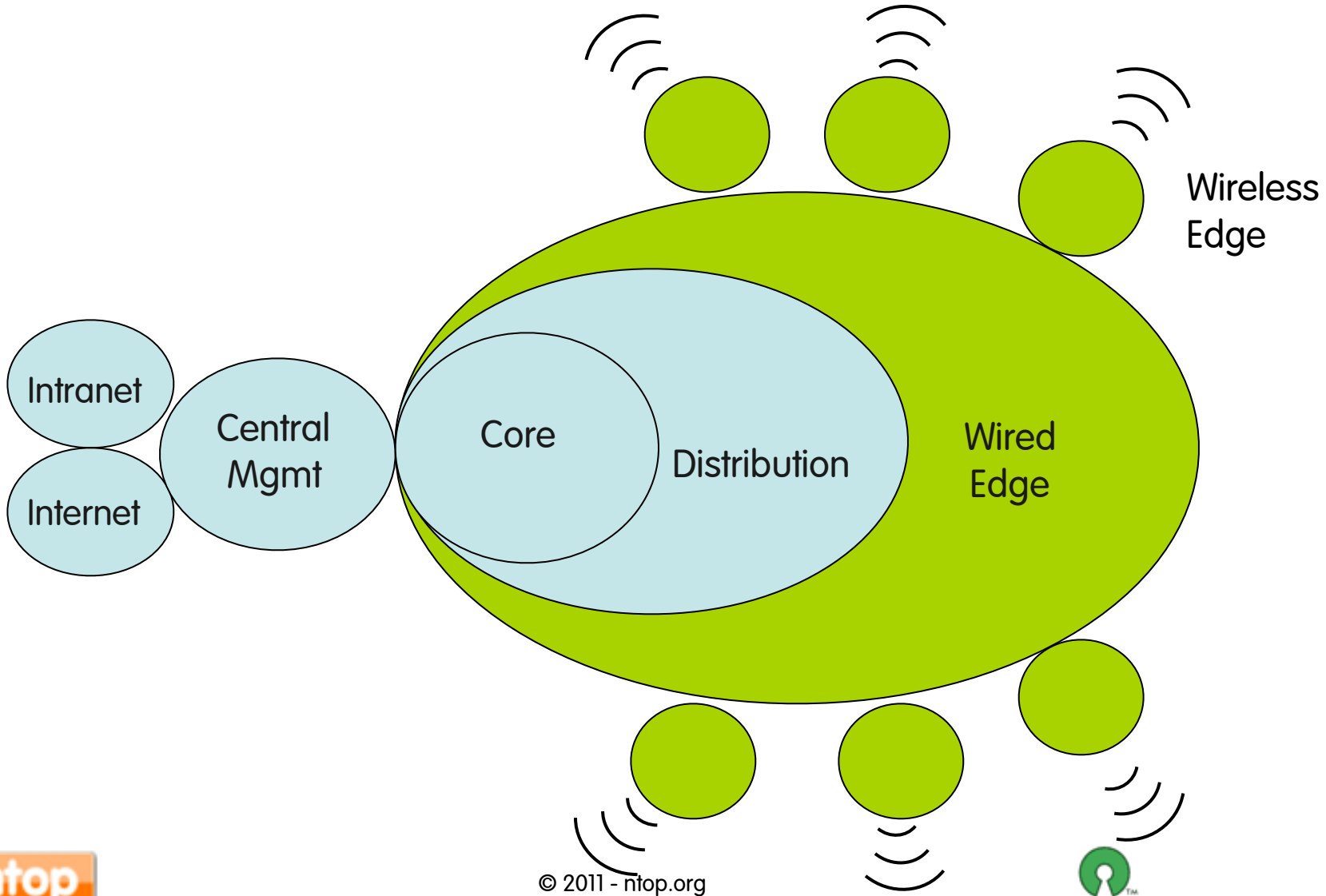
Who is Using ntop Products ?

open source

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.

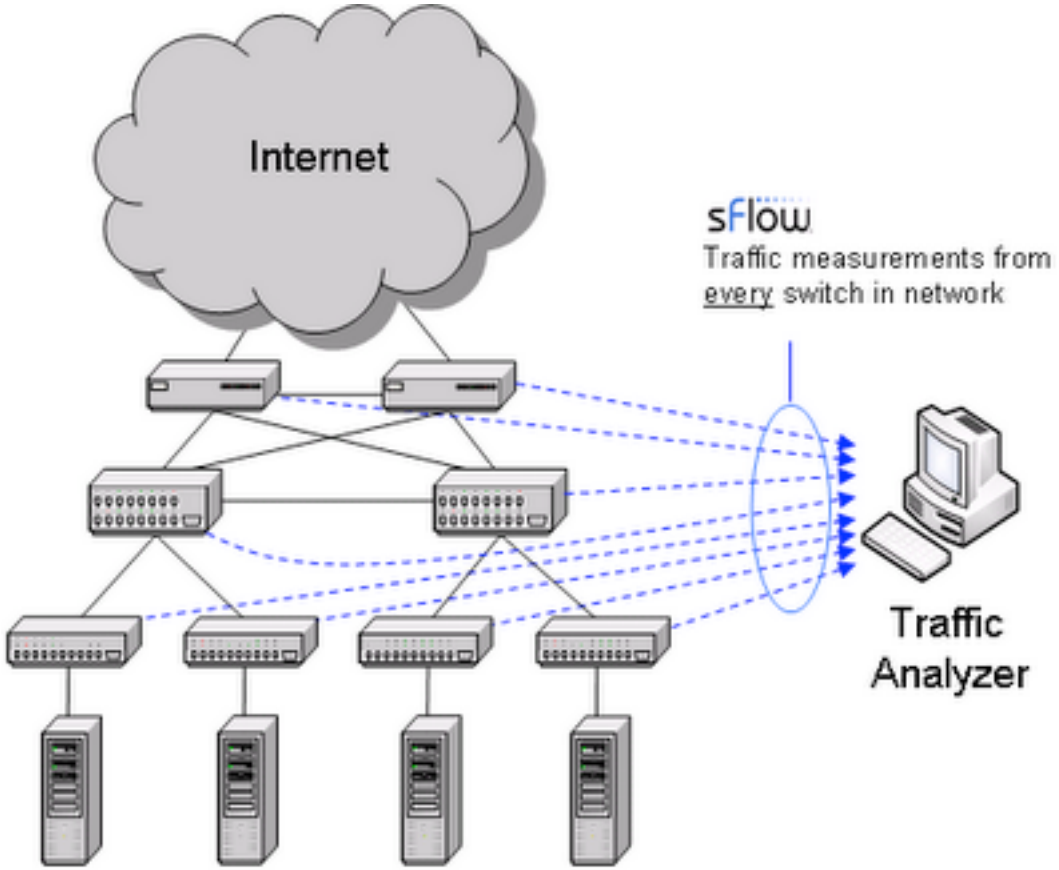
Networks are Changing… [1/2]



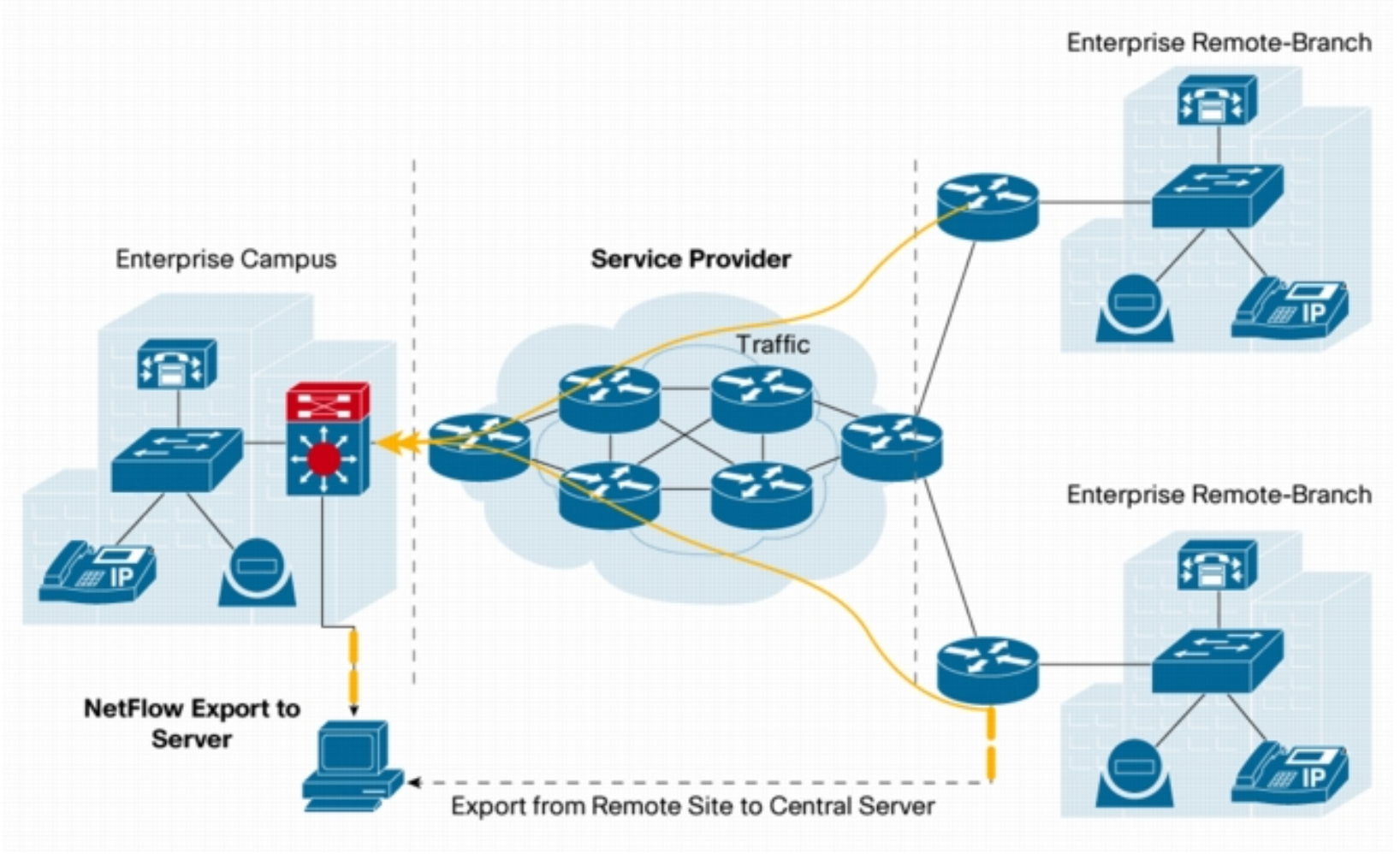
Networks are Changing… [2/2]

- 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.

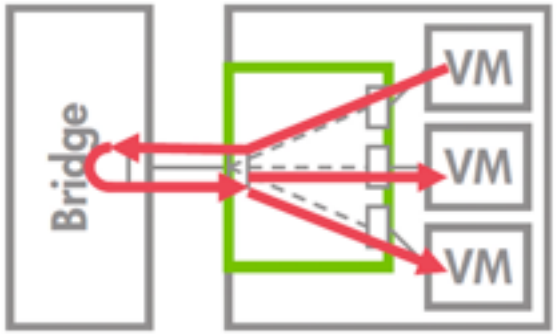
Typical Monitoring Deployment: LAN

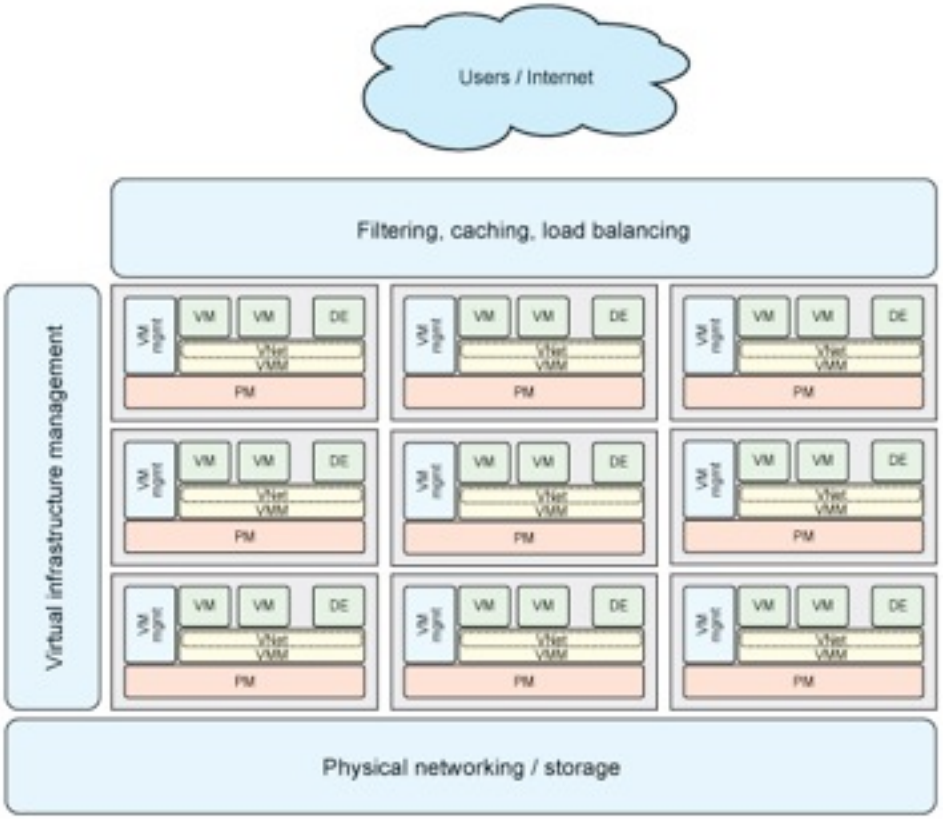


Typical Monitoring Deployment: Internet Traffic



Typical Monitoring Deployment: Cloud and Intra-VM Monitoring





Some Lessons Learnt

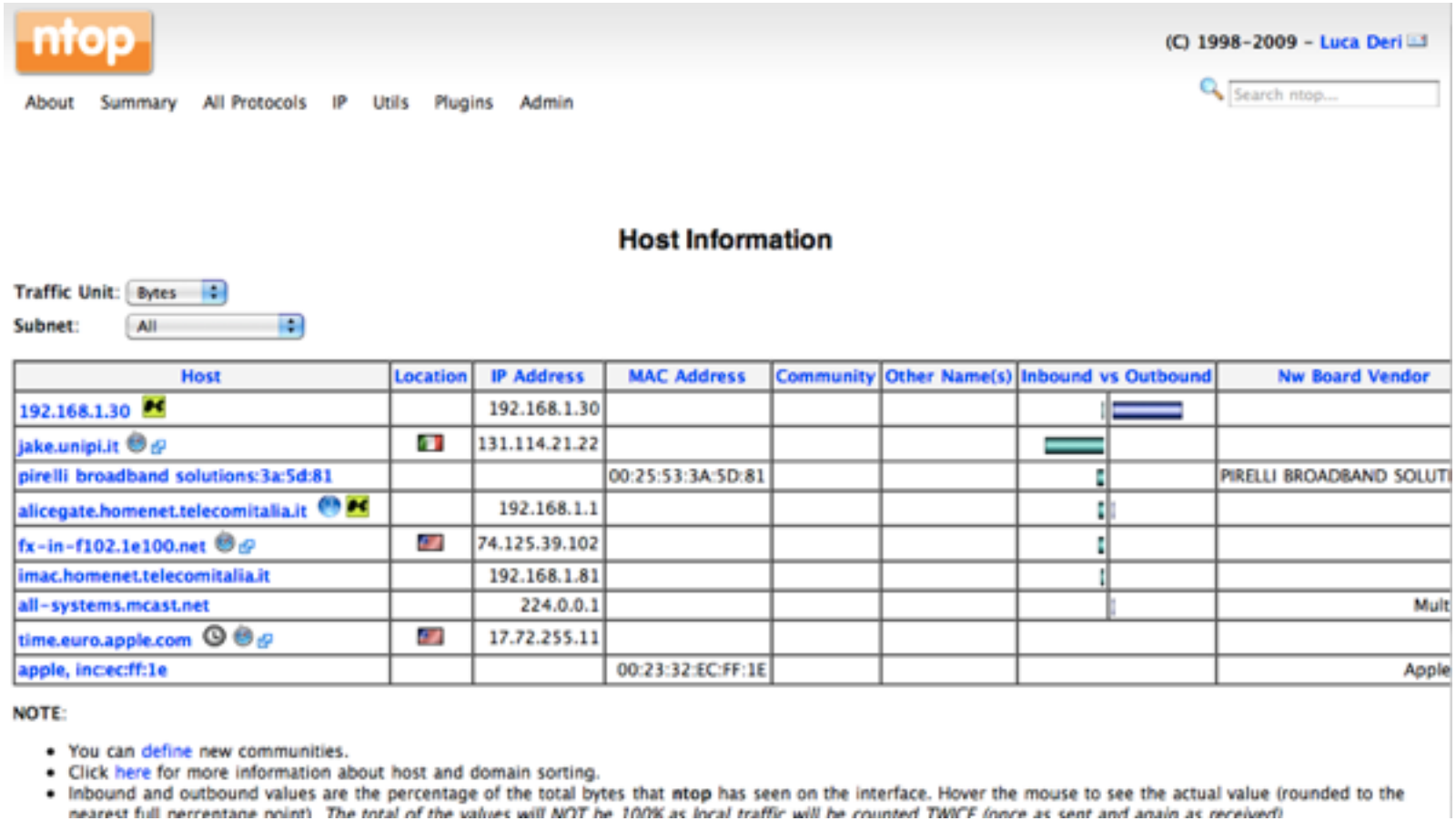
- In order to monitoring the traffic we need to deploy a probe where the traffic is flowing.
- We need to make sure we can handle both NetFlow and sFlow if we want to have complete network visibility.
- Cloud computing and server virtualization push us to monitor in-VM virtual networks.

How can ntop.org help me?

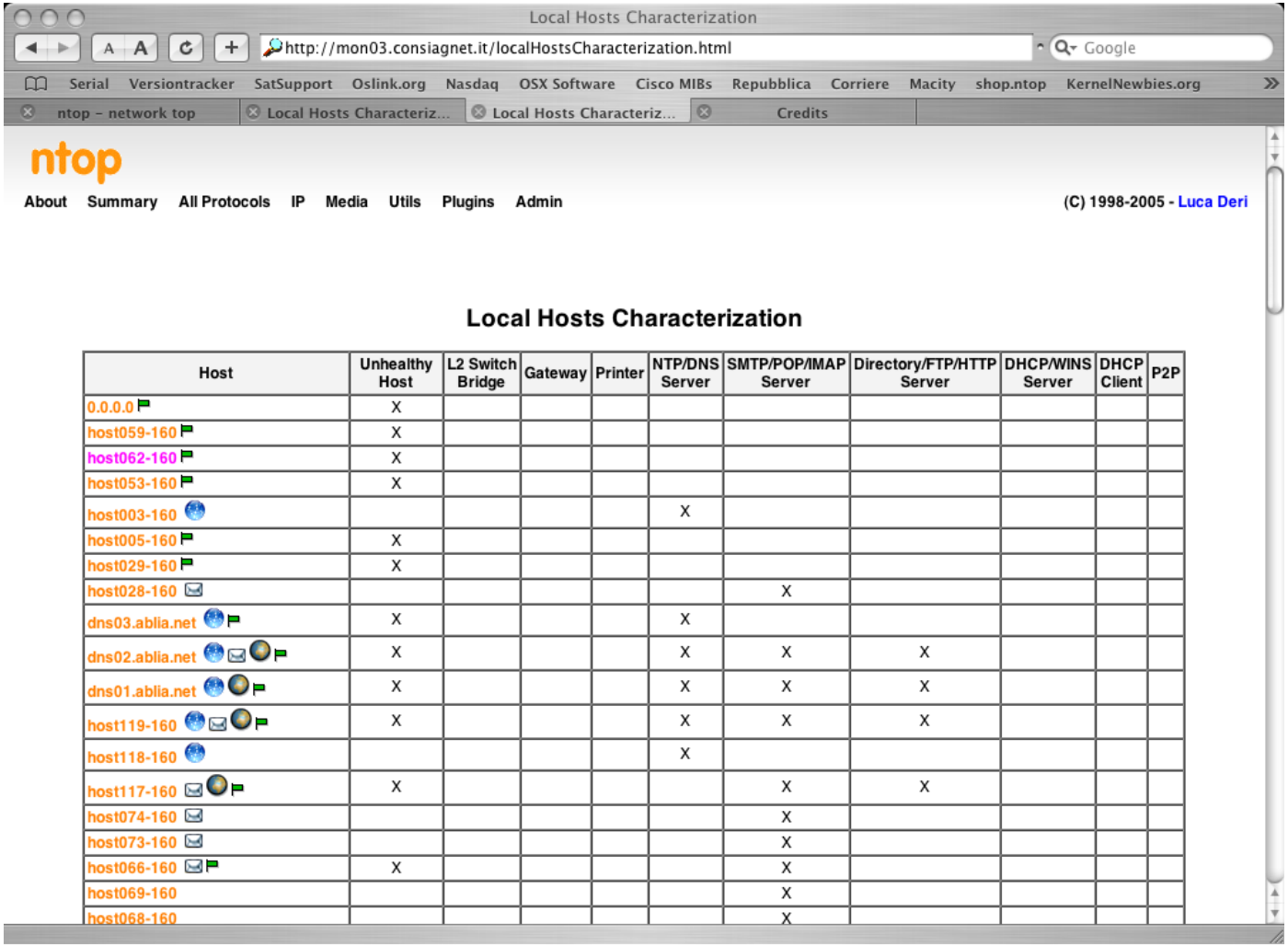
- ntop can act as central network monitoring console.
- nProbe can act as sFlow and NetFlow proxy/probe/collector.
- 10 Gbit packet capture acceleration and filtering, in host and VMs, using PF_RING.
- Ability to query billion of flows with sub-second response time.

ntop:
A Web-based Monitoring Console

What is ntop ?



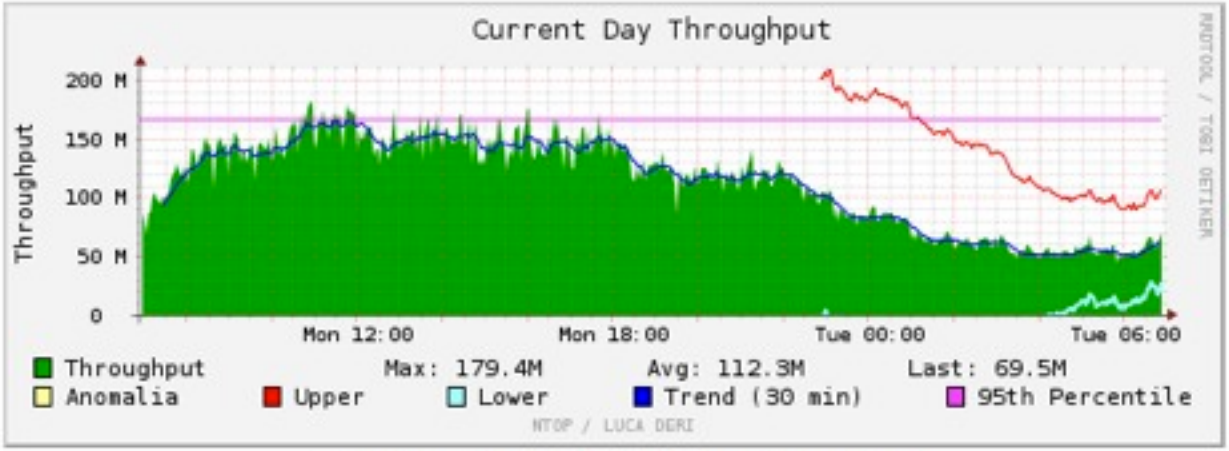
Network Inventory



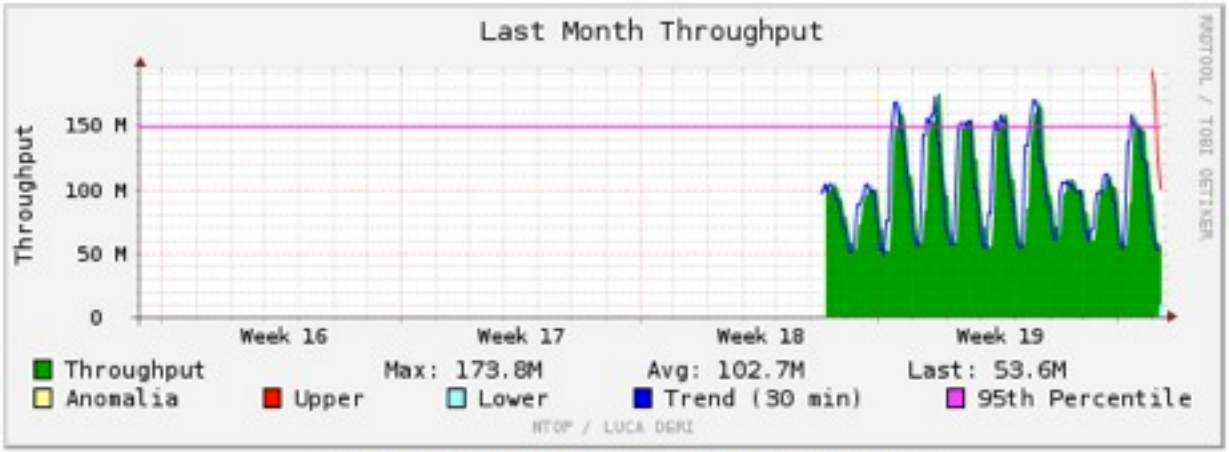
Local Hosts Characterization

Host	Unhealthy Host	L2 Switch Bridge	Gateway	Printer	NTP/DNS Server	SMTP/POP/IMAP Server	Directory/FTP/HTTP Server	DHCP/WINS Server	DHCP Client	P2P
0.0.0.0	X									
host059-160	X									
host062-160	X									
host053-160	X									
host003-160					X					
host005-160	X									
host029-160	X									
host028-160						X				
dns03.ablia.net	X				X					
dns02.ablia.net	X				X	X	X			
dns01.ablia.net	X				X	X	X			
host119-160	X				X	X	X			
host118-160					X					
host117-160	X					X	X			
host074-160						X				
host073-160						X				
host066-160	X					X				
host069-160						X				
host068-160						X				

Traffic Trends



Time [Mon May 17 06:52:02 2010 through now]



Time [Sun Apr 18 06:52:02 2010 through now]

Host Health

Data Rcvd Stats	0 % Rem 100 %
IP vs. Non-IP Rcvd	IP 100 % Non-IP 0 %
Sent vs. Rcvd Pkts	Sent 51.8 % Rcvd 48.2 %
Sent vs. Rcvd Data	Sent 33.2 % Rcvd 66.8 %
Host Type	Name Server
Historical Data	[ ]
Host Healthness (Risk Flags)	1. Unexpected packets (e.g. traffic to closed port or connection reset): [Rcvd: rejected] [Rcvd: port unreac] [Rcvd: hostnet unreac]

Host Traffic Stats

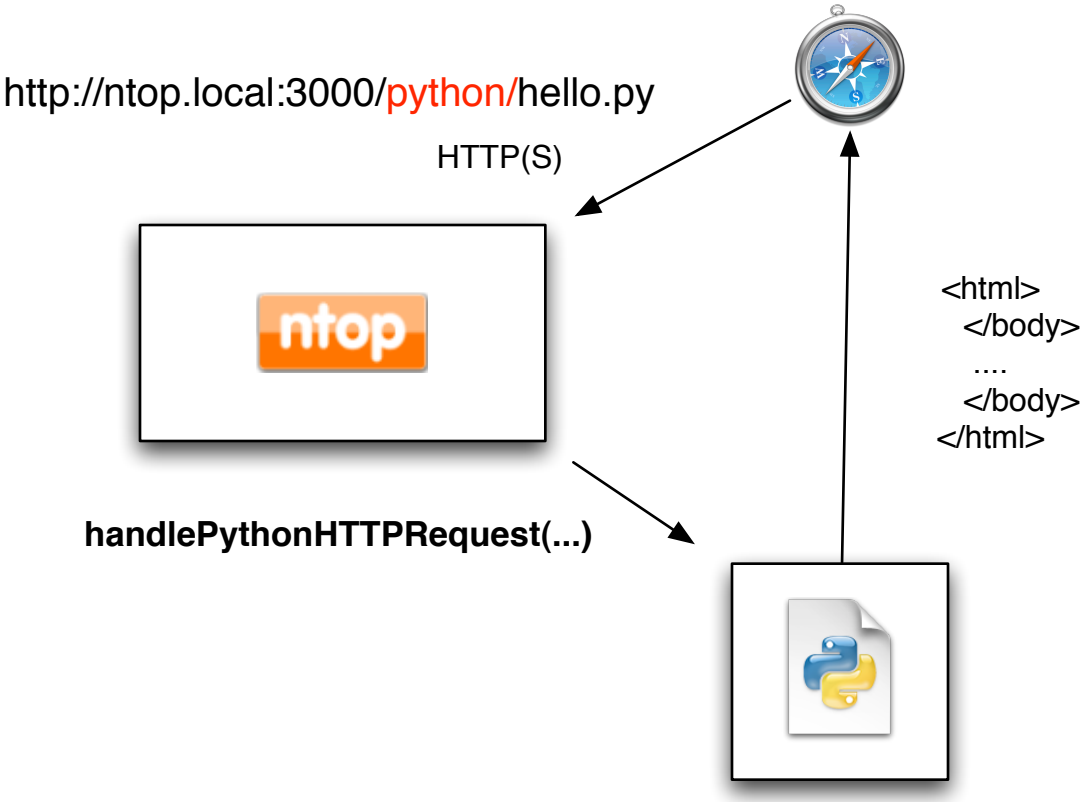
Time	Tot. Traffic Sent	% Traffic Sent	Tot. Traffic Rcvd	% Traffic Rcvd
11 AM	13.4 MB	74.7 %	26.6 MB	74.0 %
10 AM	4.5 MB	25.3 %	9.3 MB	26.0 %
9 AM	0	0.0 %	0	0.0 %
8 AM	0	0.0 %	0	0.0 %

VoIP Support

Client	Server	Data Sent	Data Rcvd	Note
130.192.225.34 :8000	130.192.225.44 :32854	58.6 KB	70.3 KB	valter called livio
130.192.225.34 :8001	130.192.225.44 :32855	224	146	
stun01.sipphone.com :3478	130.192.225.34 :47575	216	0	
130.192.225.34 :5060	bill.ipv6.polito.it :5060	2.8 KB	2.3 KB	valter called livio
130.192.225.44 :5060	bill.ipv6.polito.it :5060	4.5 KB	5.0 KB	valter called livio
130.192.225.44 :5060	130.192.225.34 :5060	462	361	

Host Type	VoIP Host
Known Users	stefano <101> [VoIP]

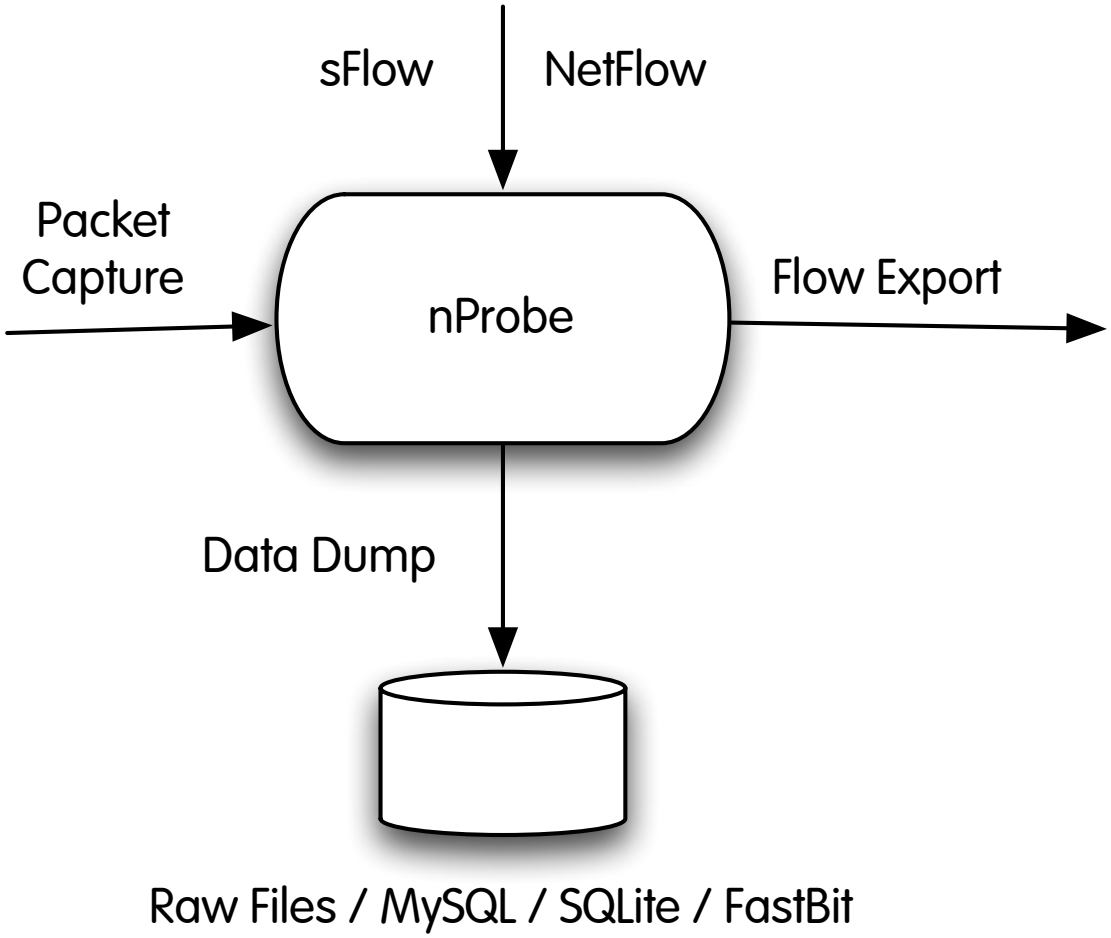
ntop Scripting using Python



Where is my traffic going to ?

Flow-based Traffic Monitoring: nProbe

nProbe: IPFIX/NetFlow Probe



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

Query	MySQL	nProbe + FastBit
Q1	22.6	5.6
Q2	69	0.5
Q3	971	12.5
Q4	1341	48.2
Q5	2257	30.7

nProbe+FastBit vs nfdump

nProbe+FastBit	45
nfdump	1500

All measurements are in seconds

How to Add Geolocation Data [1/2]

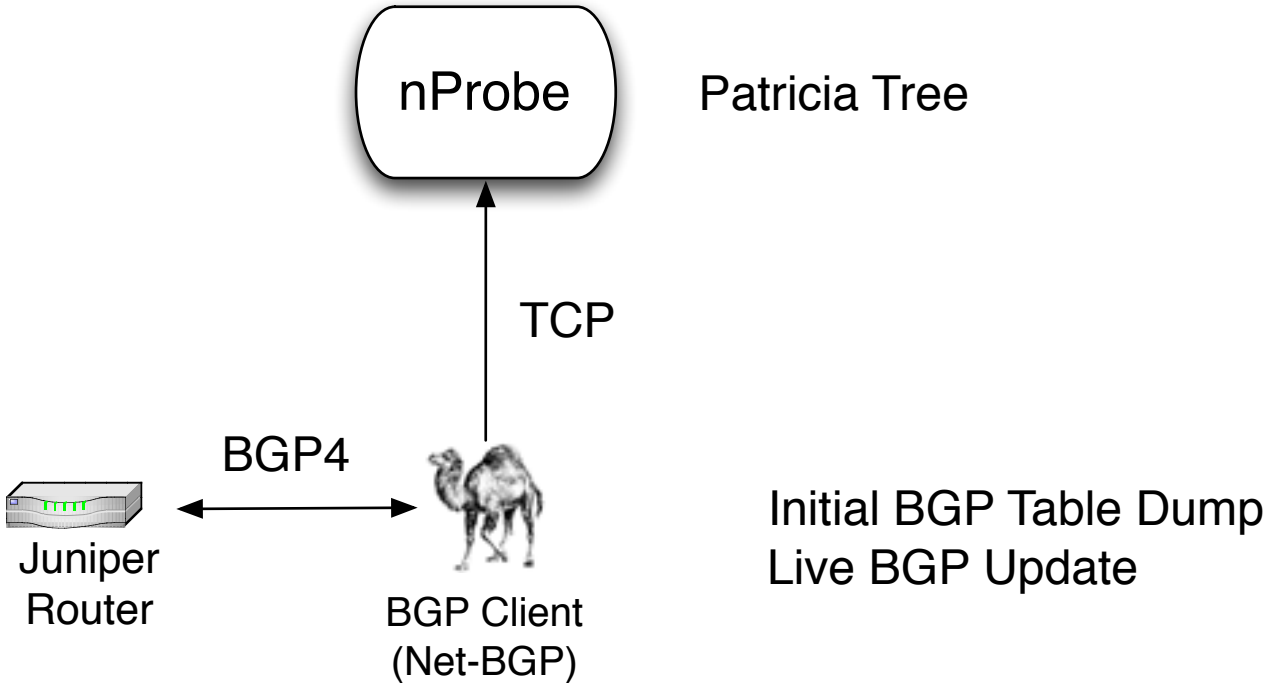
- Routers are unable to export any geolocation information.
- NetFlow/IPFIX flows do not contain any information about geolocation into standard flow formats.
- Solution:
  - Let the collector add geolocation information to flows received by routers
  - Let nProbe export this information to collectors.

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)

BGP Data Integration [1/2]



BGP Data Integration [2/2]

```
# 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,
);
```

nProbe: 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
```

```
#
# Client        Server  Protocol        Method  URL     HTTPReturnCode  Location        Referer UserAgent
ContentType     Bytes   BeginTime       EndTime Flow Hash       Cookie  Terminator      ApplLatency
#
192.168.0.200   api.leoslyrics.com      http    GET     /api_search.php?auth=mindquirk_harmonic&artist=Franco
+Battiato&songtitle=Povera+Patria  200                      curl/7.13.1 (powerpc-apple-darwin8.0) libcurl/7.13.1
OpenSSL/0.9.7i zlib/1.2.3  text/xml        10244   1133966831.996  1133966832.910  2423982224      0       C
0.152
192.168.0.200   elyrics.net     http    GET     /go/f/Franco-Battiato-lyrics/Povera-Patria-lyrics/      302
www.elyrics.net/inc/404.html            curl/7.13.1 (powerpc-apple-darwin8.0) libcurl/7.13.1 OpenSSL/0.9.7i zlib/
1.2.3  text/html       1186    1133966832.527  1133966832.908  2413138730      0       S       0.114
192.168.0.200   www.macintouch.com      http    GET     /images/filewave01.gif  200             www.macintouch.com
Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13       image/gif
27750   1133966828.928  1133966830.606  26992029        0       S       0.261
192.168.0.200   www.macintouch.com      http    GET     /images/iwas01b.gif     200             www.macintouch.com
Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13       image/gif
12469   1133966828.574  1133966829.932  26992028        0       S       0.369
192.168.0.200   www.macintouch.com      http    GET     /images/filewave02.gif  200             www.macintouch.com
Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13       image/gif
25505   1133966827.681  1133966829.196  26992027        0       S       0.387
192.168.0.200   www.macintouch.com      http    GET     /       200                     Mozilla/5.0 (Macintosh; U;
PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13       text/html       52474   1133966827.127
1133966829.339  26992026        0       S
       0.308
```

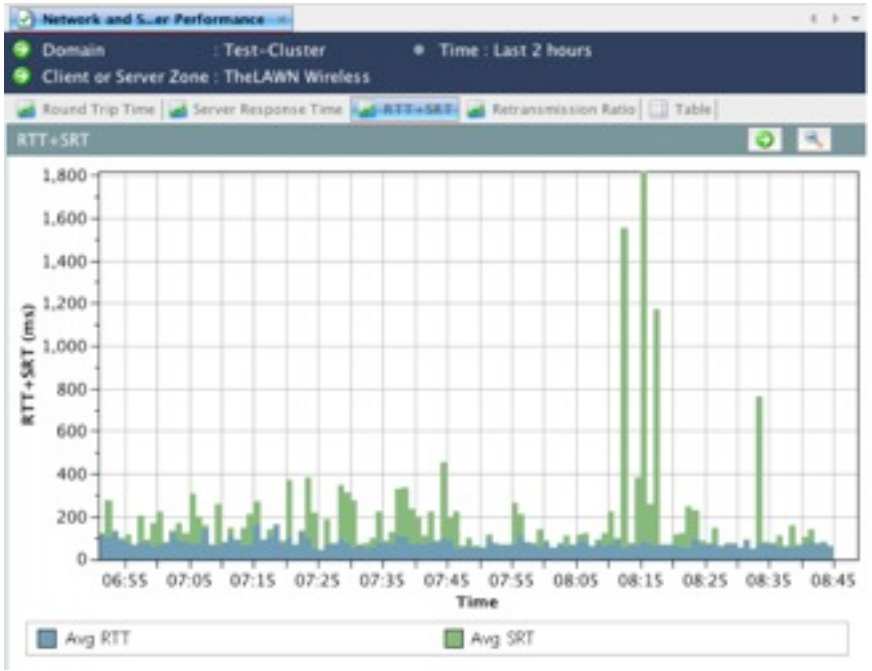
nProbe: VoIP Traffic Analysis

```
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              DNS QUERY
[NFv9 57678][IPFIX 35632.206] %DNS_QUERY_ID           DNS query transaction Id
[NFv9 57679][IPFIX 35632.207] %DNS_QUERY_TYPE         DNS query type (e.g. 1=A,2=NS)
[NFv9 57680][IPFIX 35632.208] %DNS_RET_CODE           DNS return code
[NFv9 57681][IPFIX 35632.209] %DNS_NUM_ANSWER         DNS # of returned answers

Plugin MySQL Plugin templates:
[NFv9 57667][IPFIX 35632.195] %MYSQL_SERVER_VERSION   MySQL server version
[NFv9 57668][IPFIX 35632.196] %MYSQL_USERNAME         MySQL username
[NFv9 57669][IPFIX 35632.197] %MYSQL_DB               MySQL database in use
[NFv9 57670][IPFIX 35632.198] %MYSQL_QUERY            MySQL Query
[NFv9 57671][IPFIX 35632.199] %MYSQL_RESPONSE         MySQL server response
```

nProbe: Network Performance and Response Time



flowDirection	APPL_LATENCY_SEC	APPL_LATENCY_USEC	applicationId	CLIENT_NW_DELAY_SEC	CLIENT_NW_DELAY_USEC
INGRESS*	0	25396	domain (dns) (53 UDP)	0	0
INGRESS*	0	99368	domain (dns) (53 UDP)	0	0
INGRESS*	0	0	domain (dns) (53 UDP)	0	0
INGRESS*	1	239407	domain (dns) (53 UDP)	0	0
INGRESS*	0	0	domain (dns) (53 TCP)	0	354
INGRESS*	0	374	domain (dns) (53 TCP)	0	354
INGRESS*	0	0	domain (dns) (53 TCP)	0	517
INGRESS*	0	365	domain (dns) (53 TCP)	0	517

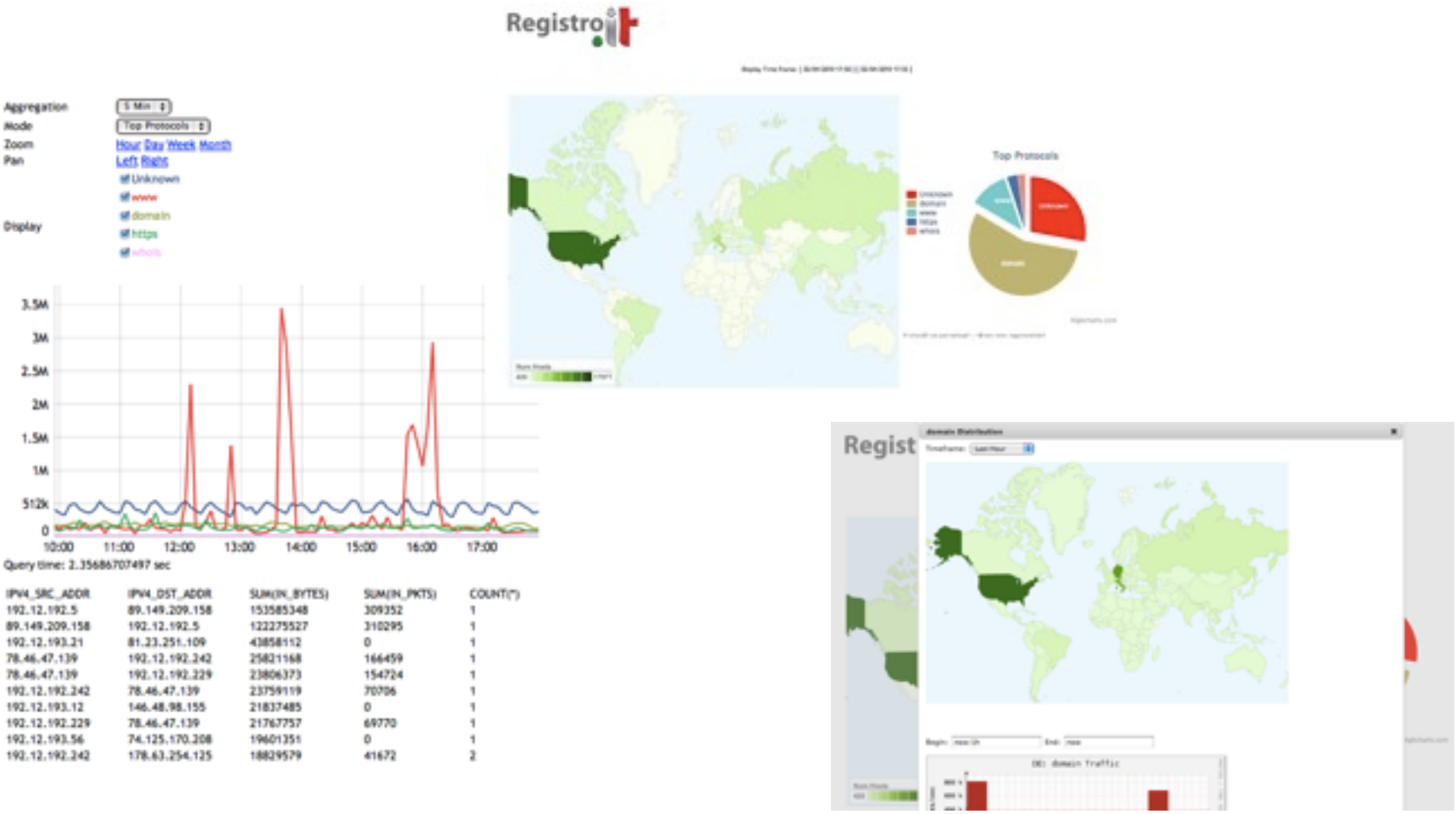
nProbe: Network Awareness

SysUpTime	HTTP_RET_CODE	HTTP_URL ↓	ingressInterface
	200	http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/	35
	200	http://x4.last.fm/user/23519048/73afdc0b19d9a93d9c4718dab0240e21/10016/	35
	200	http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL	35
	200	http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=RPLL	35
	200	http://weather.noaa.gov/cgi-bin/mgetmetar.pl?cccc=KPSM	35

How can I Improve my Internet Presence ?

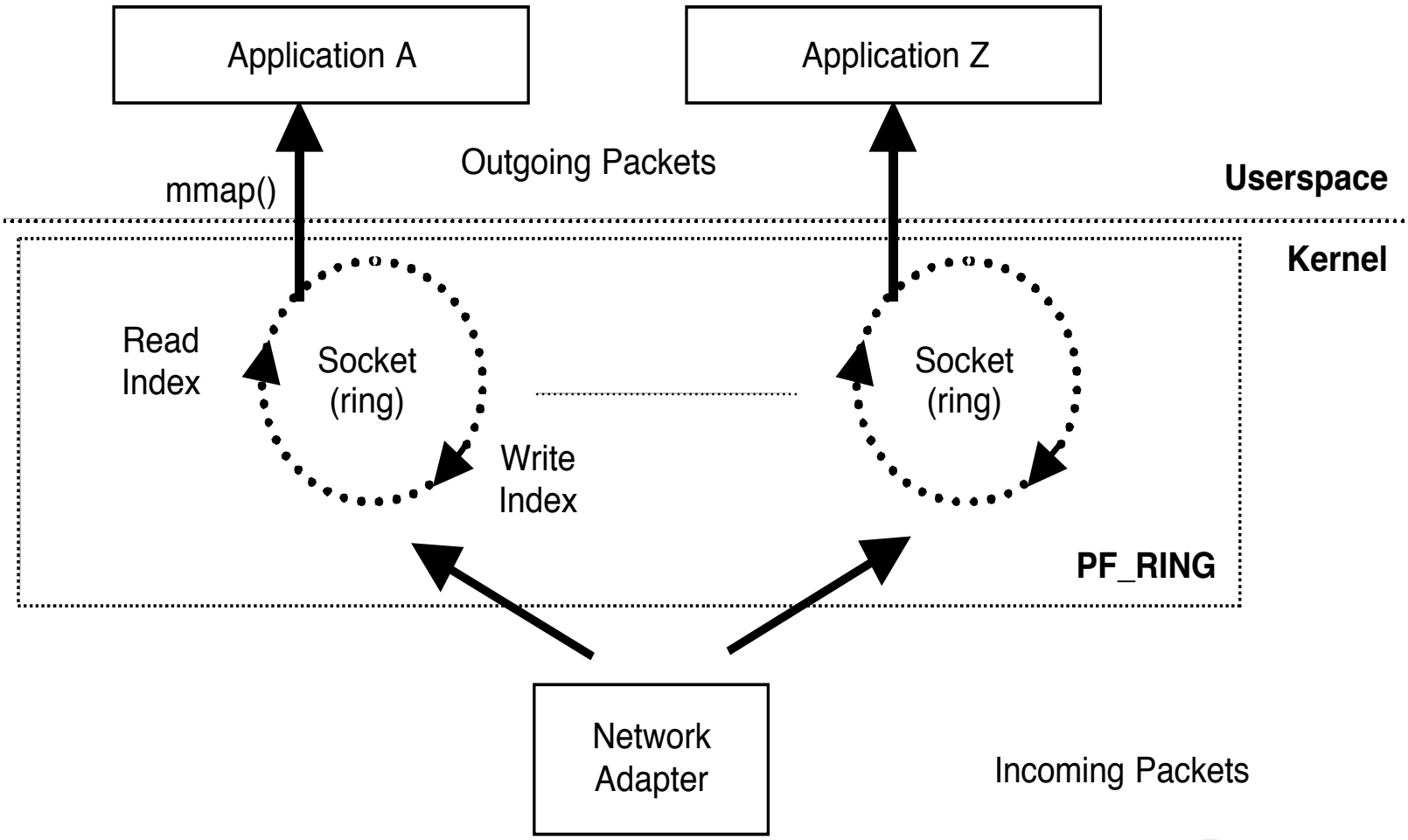
Distance: 1		
ASN	AS Name	Traffic
2597	REGISTRO CCTLD IT	738.8 MB
Distance: 2		
ASN	AS Name	Traffic
3356	Level 3 Communications, LLC	392.0 MB
12637	Seeweb Srl	3.1 MB
137	GARR Italian academic and research network	1.5 MB
21309	ACANTHO SPA	420.7 KB
64862	??	190.1 KB
21056	Welcome Italia S.p.A.	30.5 KB
16004	MIX S.r.L.	9.8 KB
15469	Warinet NOC AS	356.0 bytes
Distance: 3		
ASN	AS Name	Traffic
9031	INTICOnet	139.0 MB
9035	Wind Telecomunicazioni spa	17.4 MB
6453	Teleglobe Inc.	16.0 MB
1273	Cable & Wireless Deutschland GmbH	15.4 MB
3549	Global Crossing	13.4 MB
702	UUNET - Commercial IP service provider in Europe	10.7 MB
24940	Hetzner Online AG RZ-Nuernberg	9.2 MB
6762	Telecom Italia international high speed,	9.0 MB
8218	Neo Telecoms	7.2 MB
286	KPNQwest Backbone AS	6.8 MB
7473	Singapore Telecom	6.0 MB
1299	TeliaNet Global Network	4.6 MB
1239	Sprint	4.1 MB
10310	Yahoo!	4.0 MB

Interactive Data Search



Packet Capture Acceleration: PF_RING

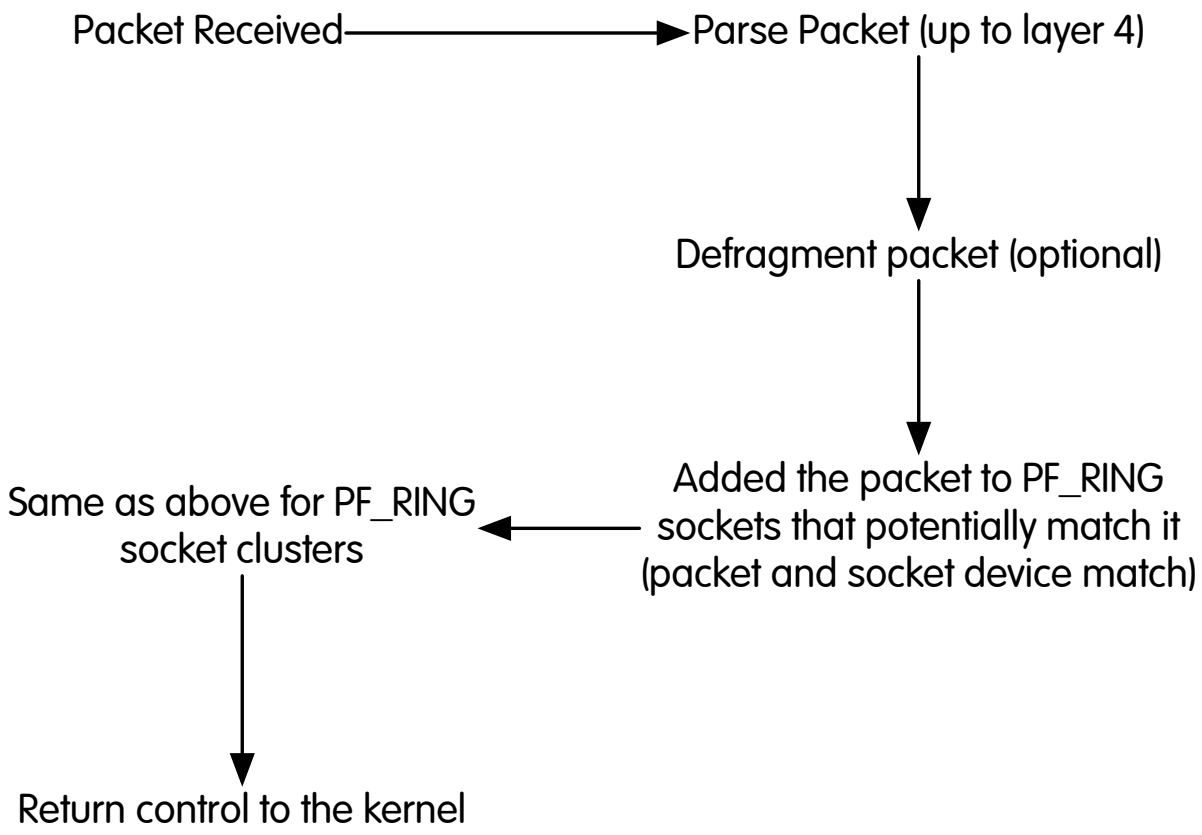
Socket Packet Ring (PF_RING)



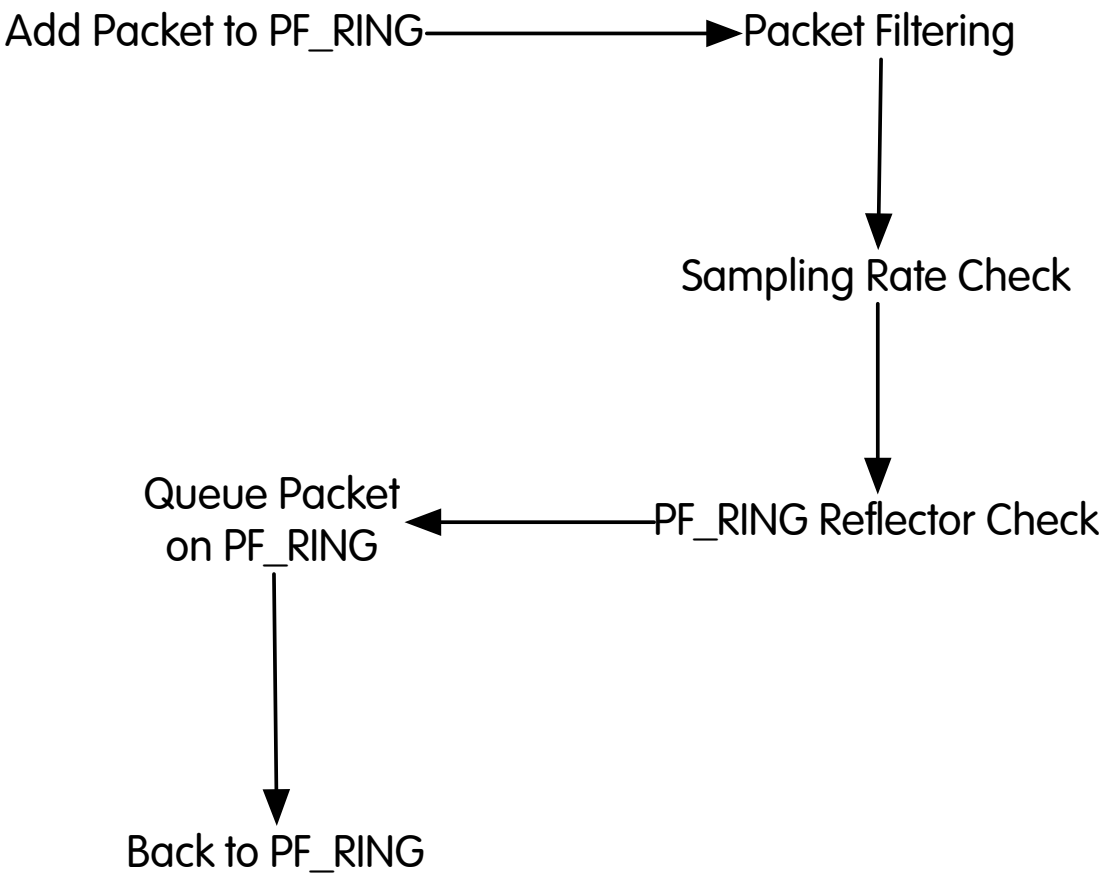
PF_RING: Benefits

- 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.

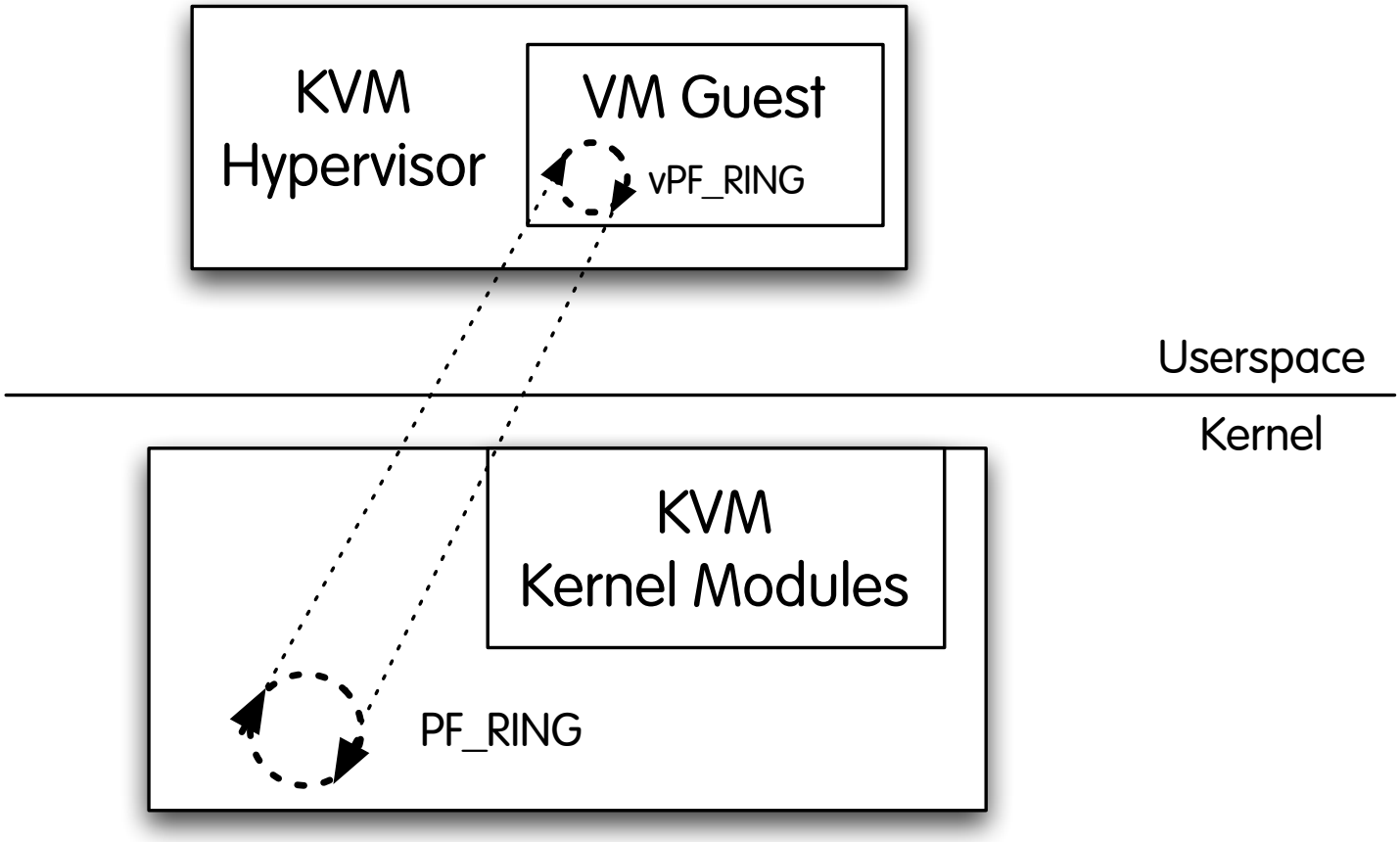
PF_RING Packet Journey [1/2]



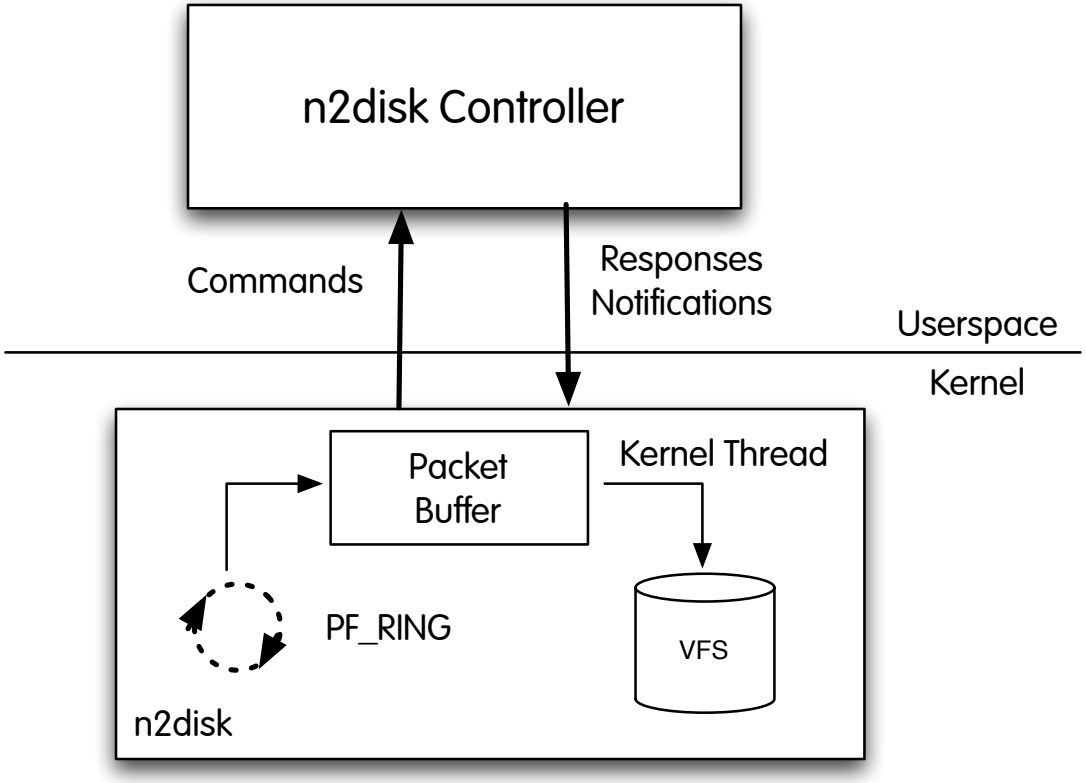
PF_RING Packet Journey [2/2]



PF_RING in VMs: vPF_RING



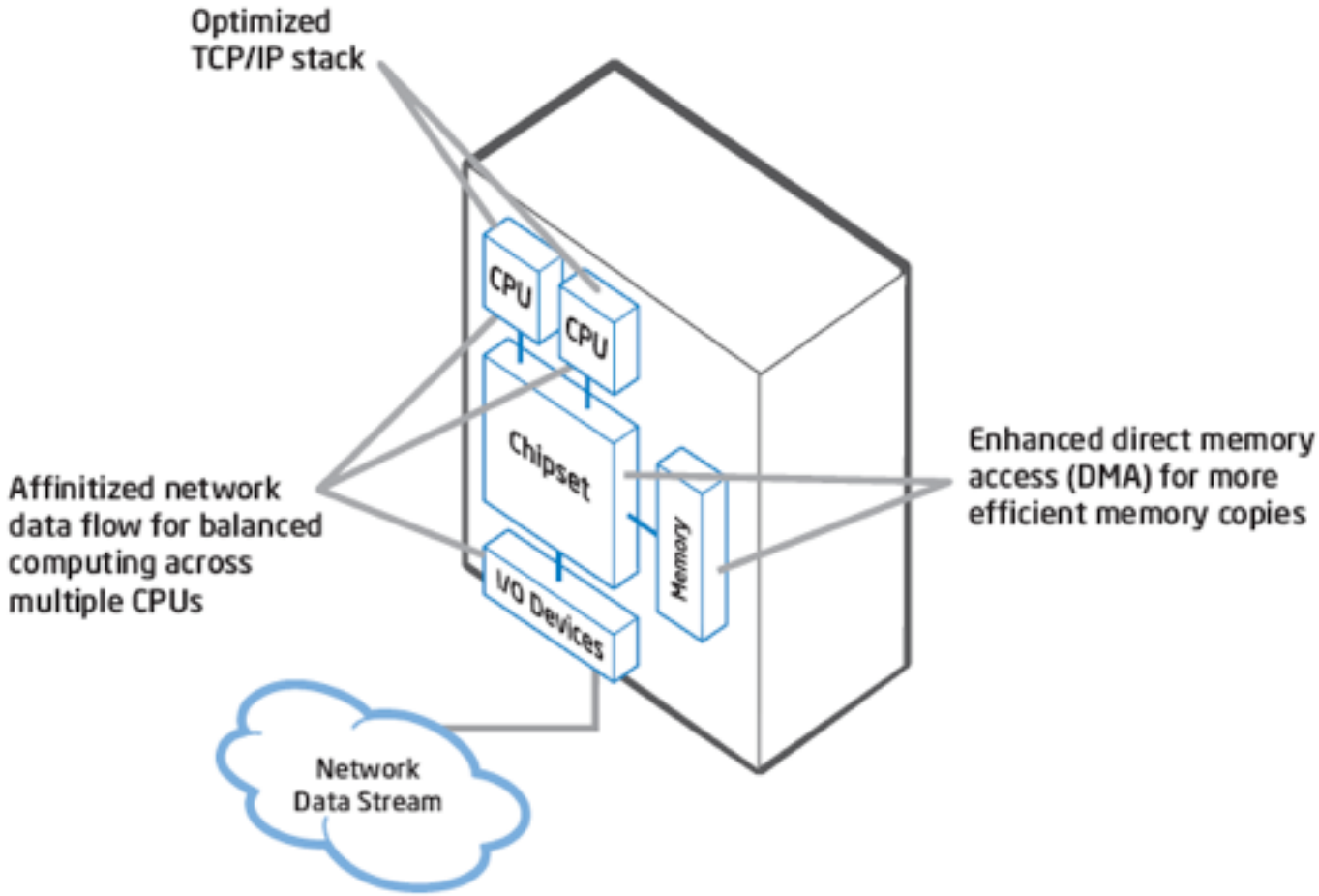
PF_RING Packet-to-Disk: n2disk



n2disk (only userland)	750 Kpps	Quad-Xeon with 8 RAID disks
n2disk (kernel+userland)	650 Kpps	Atom with single SATA Disk

Towards 10 Gbit Monitoring

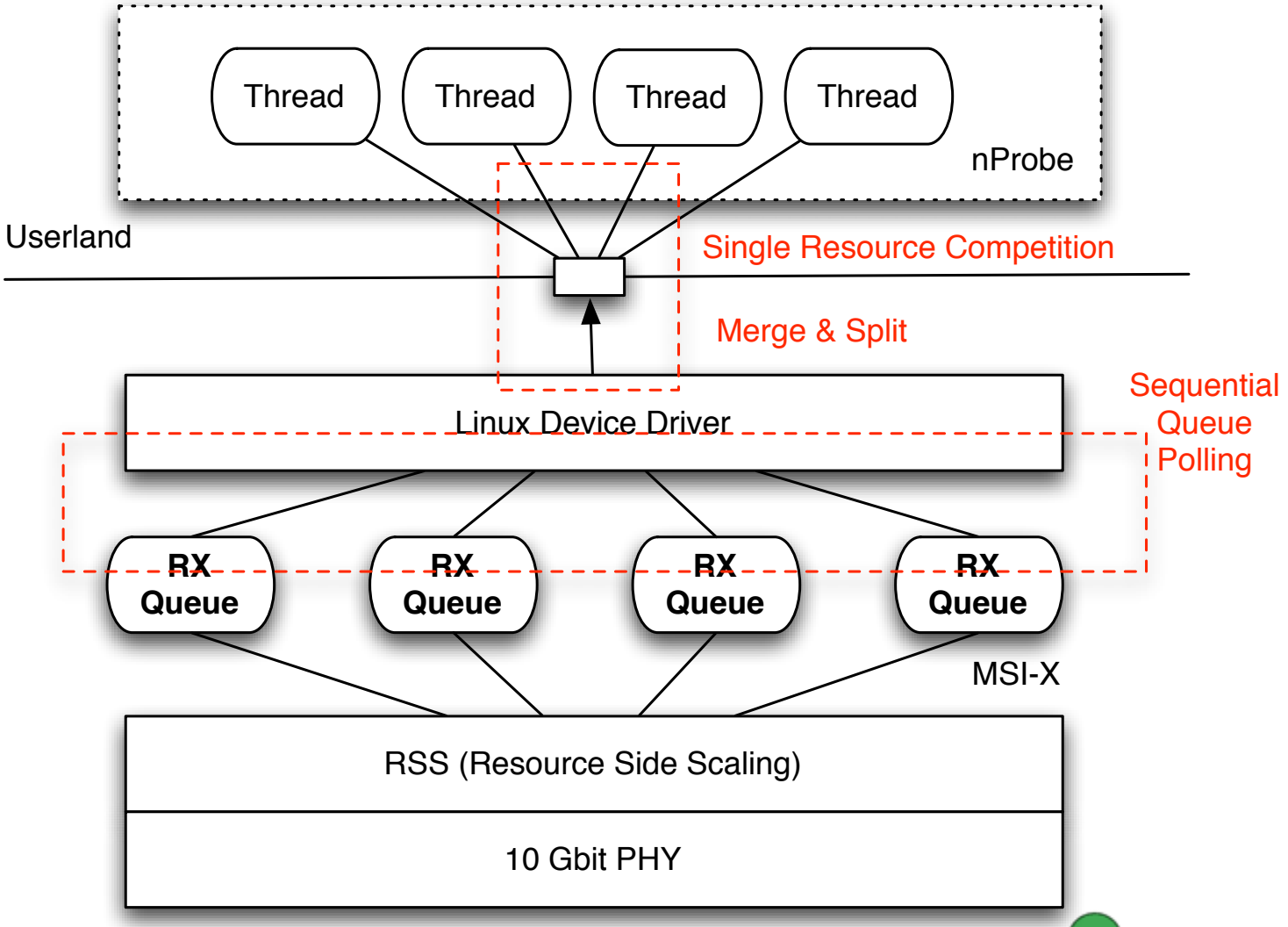
Modern Networking Architectures



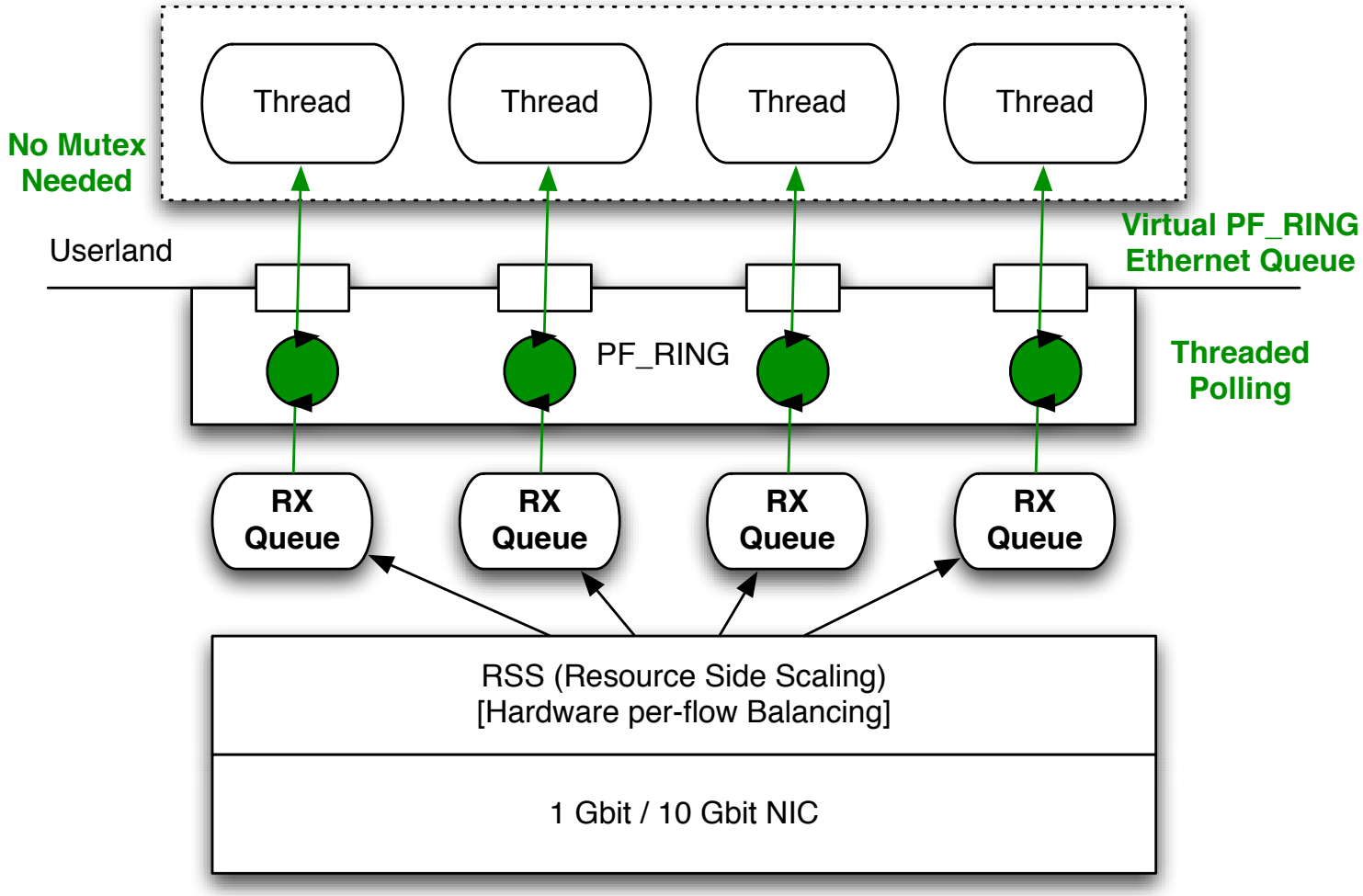
Scaling to 10 Gbit: Divide et Impera

- 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



PF_RING+TNAPI

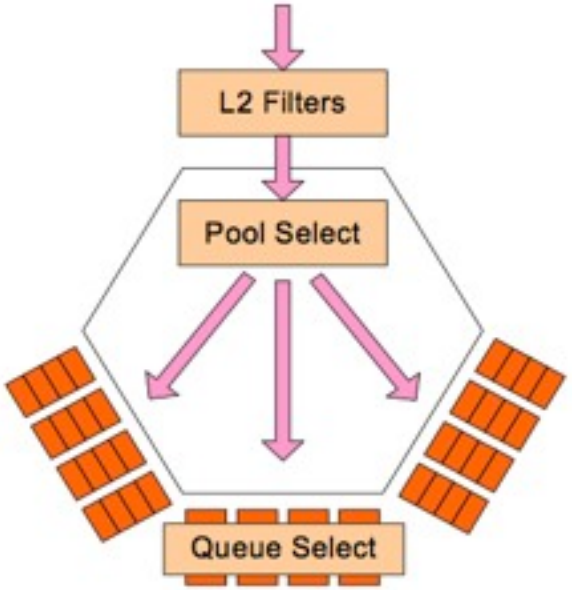


Intel 82599 Ethernet Controller [1/2]

- 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.

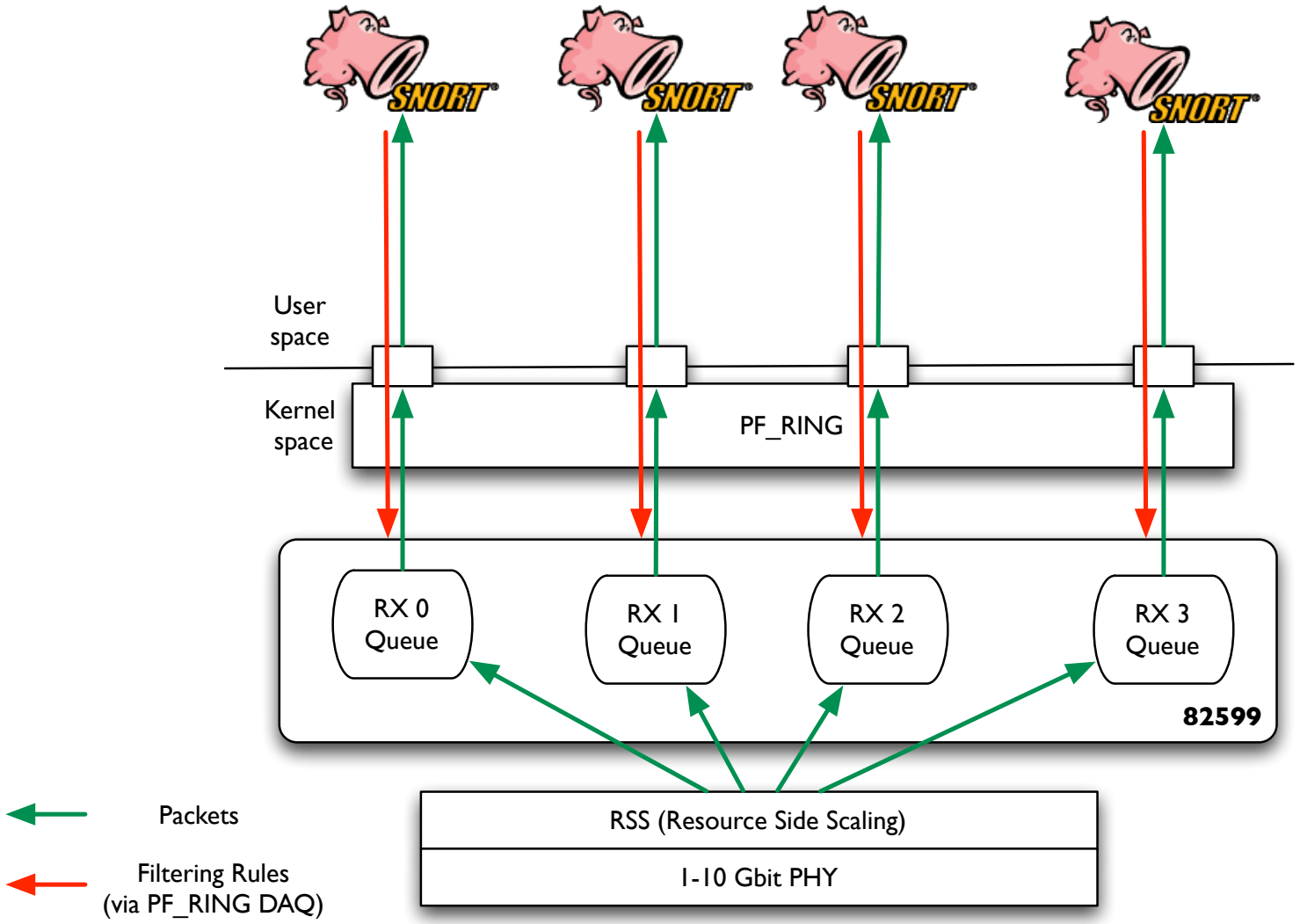


- Assigning a packet to a non-existing RX queue (<= number of available CPU cores) drops the packet.

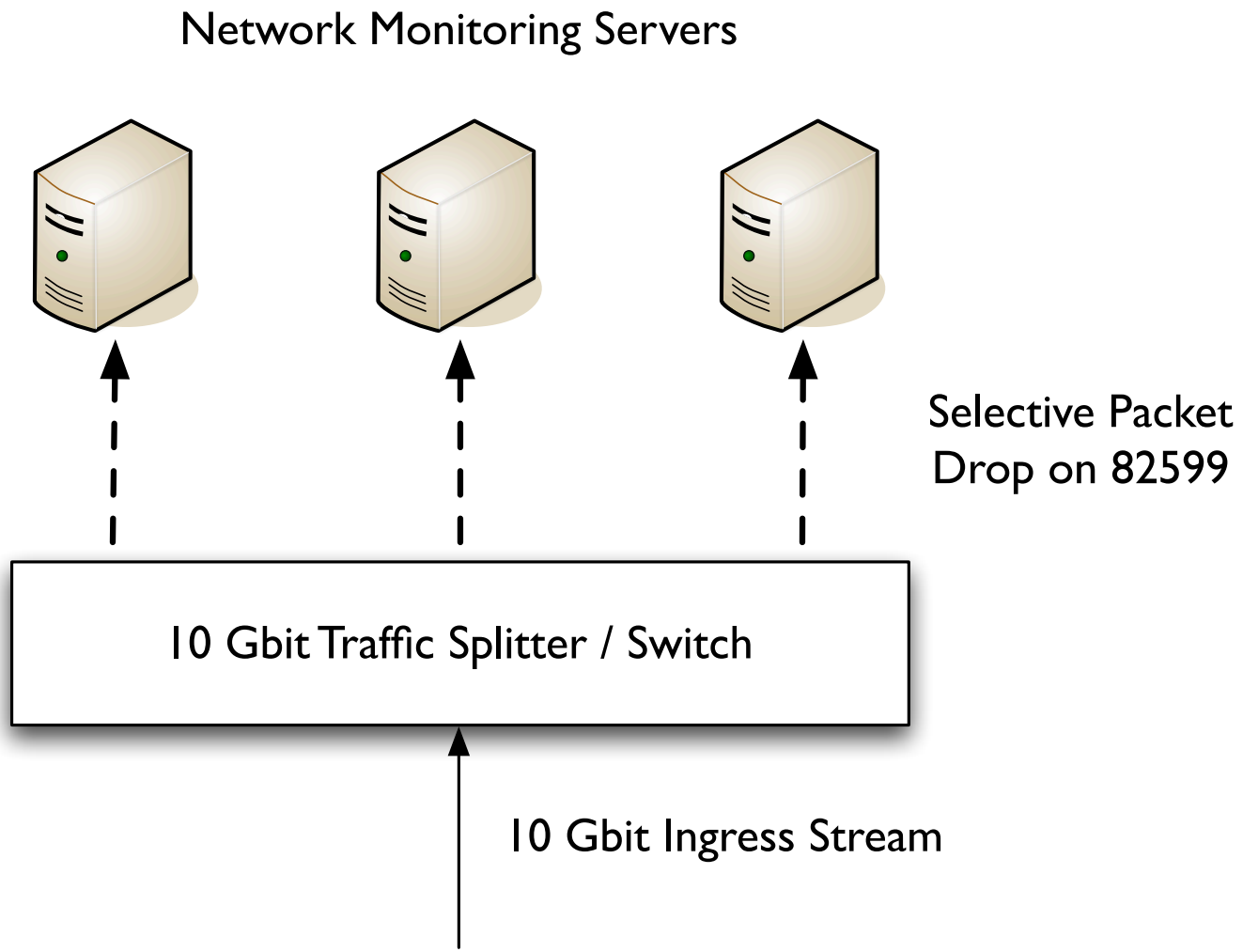
Using Hardware Filters in Real Life

- 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



Divide et Impera



References

- Home Page:
  http://www.ntop.org/

- Platforms:
  Win32 (except PF_RING) and Unix.

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