“Network Monitoring and Measurements: Challenges and Future Directions”

6th IEEE LCN Workshop on Network Measurements

colocated with

36th IEEE Conference on Local Computer Networks (LCN 2011)

4 October 2011

Antonio Pescapè

Università degli Studi di Napoli “Federico II”
Dipartimento di Informatica e Sistemistica
COMICS Research Group

Acknowledgements

- WNM Steering Committee for the invitation

Who am I

- Assistant Professor at University of Napoli Federico II (Italy)
- Member of the COMICS (***COMputers for Interaction and CommunicationS***) group at the Computer Science Department
- More than 10 years in the NM2 field
- Two small babies (to continuously monitor)… but this is another story ☺

Agenda

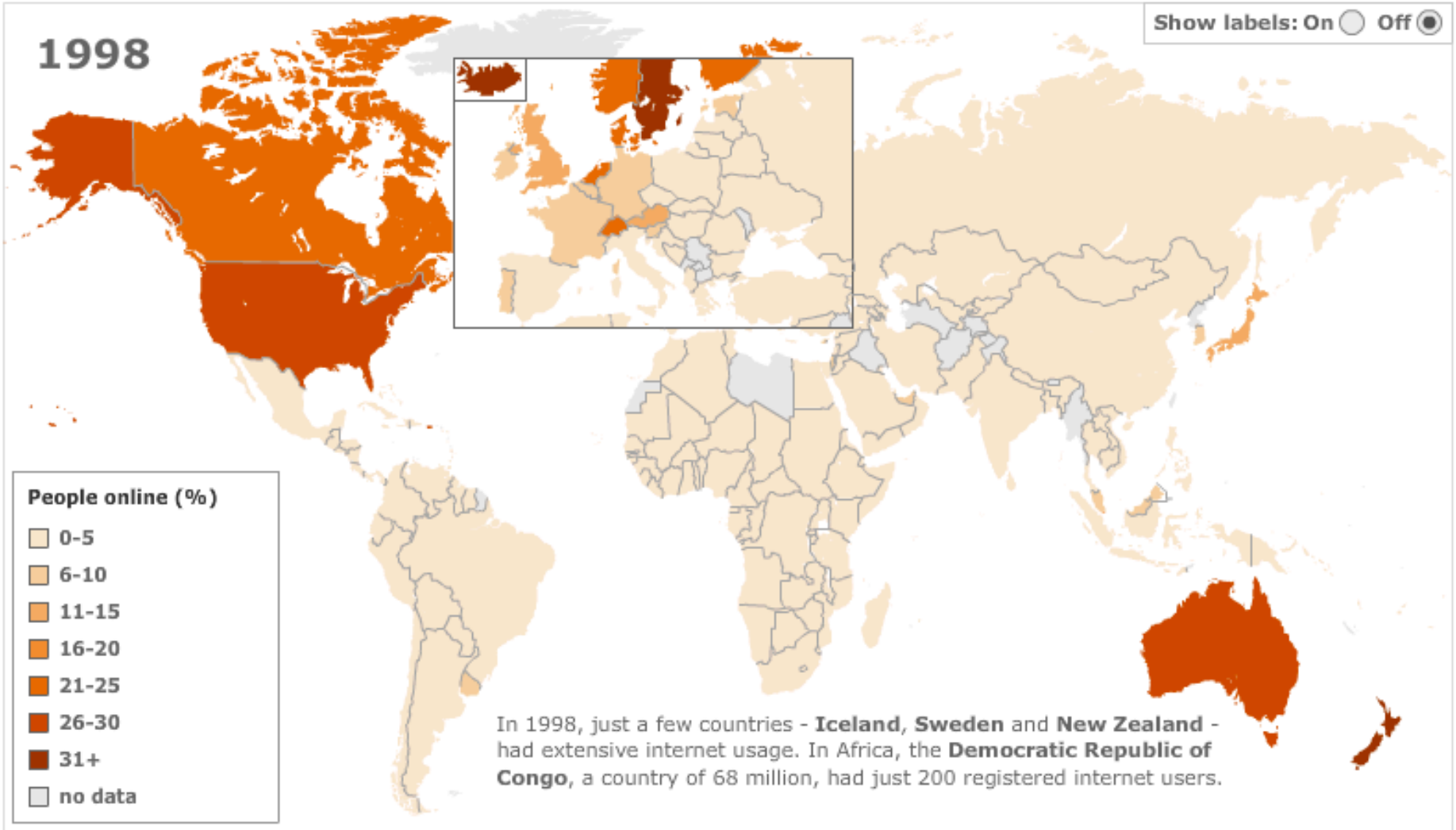
- Network Monitoring and Measurements (NM2)
  - Part 1
    - Introduction and Motivations, Hot topics, Challenges and Obstacles
  - Part 2
    - Research Contributions
      - don’t worry, no marketing, no results!!! ☺
  - Part 3
    - Conclusion

Part 1

Introduction

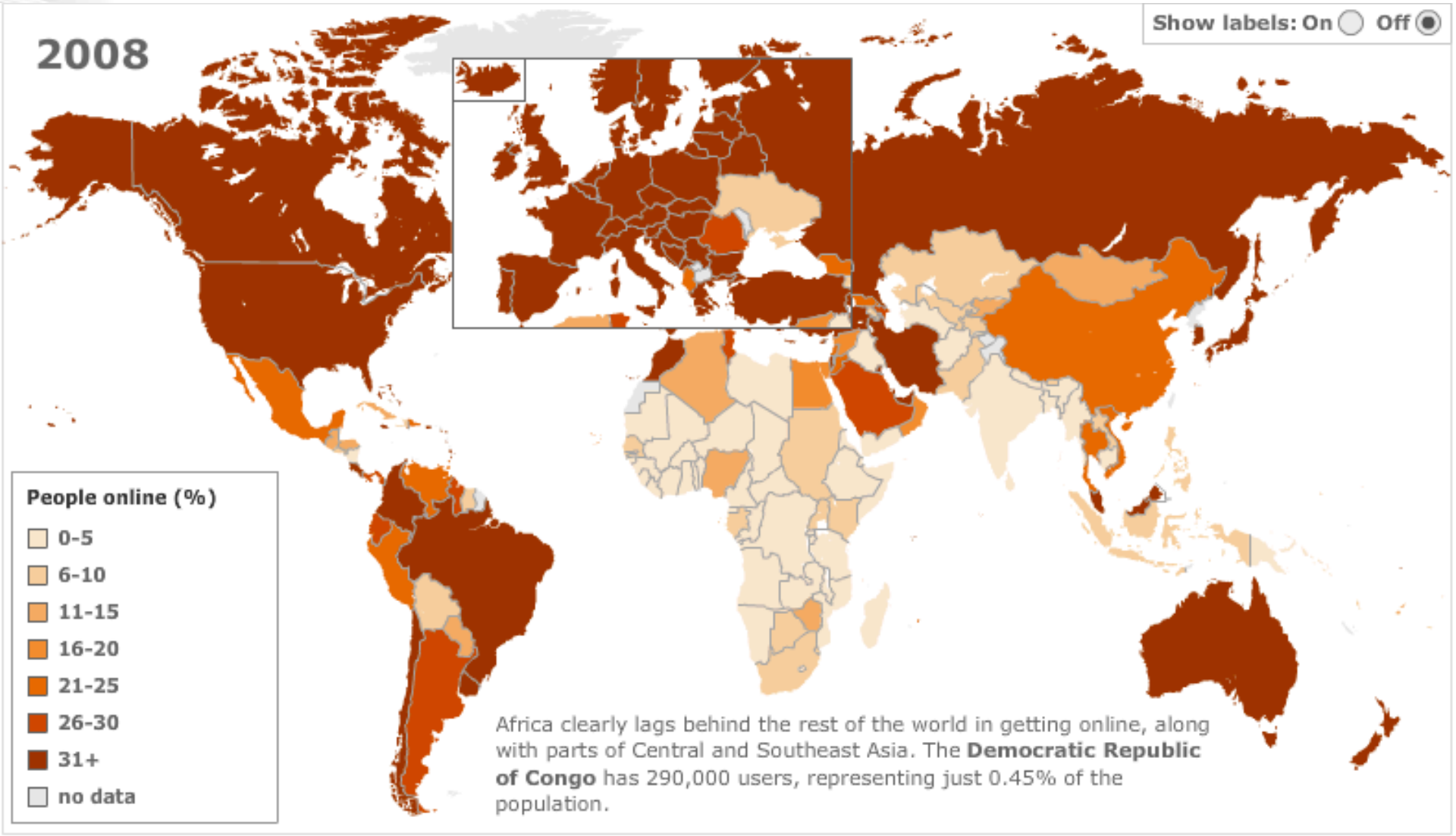
- L. Kleinrock and W. E. Naylor, “On Measured Behavior of the ARPA Network,” afips, pp.767, Proceedings of the National Computer Conference, 1974
- First network measurement publication
- I was three years old…

Internet Evolution: Internet users in 1998



Source: http://news.bbc.co.uk/2/hi/technology/8552410.stm

Internet Evolution: Internet users in 2008



Source: http://news.bbc.co.uk/2/hi/technology/8552410.stm

Internet Evolution: devices and bandwidth

- **Devices**: in 10 years factor 4 growth
  - ✓ Driven by cell phones & other new classes of consumer electronics (eBooks, tablets, Internet TV, digital picture frames …)
  - ✓ Even bigger is machine to machine (smart grids for energy management, surveillance & public safety, traffic & parking control, cars, and sensor nets …)

- **Bandwidth**: trunk speeds roughly double every 22months (driven by Moore's law)
  - ✓ …voice long ago overtaken by data,
  - ✓ moved from 75bps in 1960 to 50kbps in 1970 to 10-100Gbps singe stream today (1 billion times increase)

Motivations (1/2)

- "***Well known***" motivations (non exhaustive list)
  - ✓ Planning and design
  - ✓ Traffic engineering
  - ✓ Quality of Service
  - ✓ Billing
  - ✓ Troubleshooting and fault diagnosis
  - ✓ Security
  - ✓ Operation and Management
  - ✓ Etc.
  - ✓ Research and Intellectual Curiosity

Motivations (2/2)

- "***New comers***" motivations (non exhaustive list)
  - ✓ from Traffic/Application Monitoring to Service Monitoring
  - ✓ Customer feedback
  - ✓ Internet Outages
  - ✓ Internet Censorship
  - ✓ Overcome infrastructure limitations (eg 3G networks)
  - ✓ Etc.
  - ✓ Research and Intellectual curiosity
    - Monitoring and Measurement of one of the largest "real" complex systems
      - Network topologies
      - Overlay Networks
      - Social Networks

The role of measurement (1/2)

- The Role of Measurement in Network Research

Google Scholar search terms	Articles / %
network packet	1,060,000
network packet **system**	66%
network packet **performance**	58%
network packet **model**	39%
network packet **analysis**	34%
network packet **protocol**	30%
network packet **simulation**	29%
network packet **TCP**	16%
network packet **theory**	12%
network packet **measurement**	9%
network packet **calibration**	0.8%

Source: http://www.icir.org/vern/talks/vp-moneyshots.SIGCOMM11.pdf

The role of measurement (2/2)

- The Role of Measurement in Physics

Google Scholar **Physics**	Articles / %	Ratio to **Networking**
	5,680,000	5.4 : 1
system	44%	0.7 : 1
model	39%	1 : 1
simulation	35%	1.2 : 1
theory	22%	1.8 : 1
measurement	39%	4.3 : 1
calibration	32%	39.0 : 1

Source: http://www.icir.org/vern/talks/vp-moneyshots.SIGCOMM11.pdf

Hot Topics (1/3)

- It's really hard to list the hot topics
- Each of us is influenced by her/his research activity
- I'll show you the results of a simple exercise (no scientific ☺)
  - ✓ I have selected four workshops and conferences focusing on NM2, from 2001 to 2011
    - IMC, PAM, TMA, and (obviously ☺) WNM
  - ✓ I have (roughly) clustered using topics
    - 36 clusters (i.e., topics)
  - ✓ In this way
    - "hot topics" are defined by the community
    - we have a rough idea of their *temporal evolution*
      - 2001-2005 vs 2006-2011

Hot Topics (2/3)

1	Measurement tools & methodologies
2	Wireless networks and Wifi performance evaluation and optimization
3	Mobile
4	Topology
5	OSN
6	Traffic analysis & simulation
7	Traffic anomalies & event detection
8	Other applications level performance & optimization
9	Overlay network
10	Addressing
11	Streaming
12	Access Networks
13	IPTV
14	Traffic sampling & manipulation
15	CDN & content distribution performance
16	Privacy
17	Traffic classification
18	Censorship
19	Cloud
20	Datacenter
21	Economics
22	Ethics and Legality
23	Green
24	Network neutrality
25	Transport layer improvements
26	Network Devices
27	Traffic generation
28	Traffic matrix
29	DNS performance
30	Tomography
31	Traffic engineering
32	Coordinate systems & Geolocationing
33	Traffic characterization and modeling
34	Bandwidth estimation
35	Routing
36	Network performance evaluation

Hot Topics (3/3)

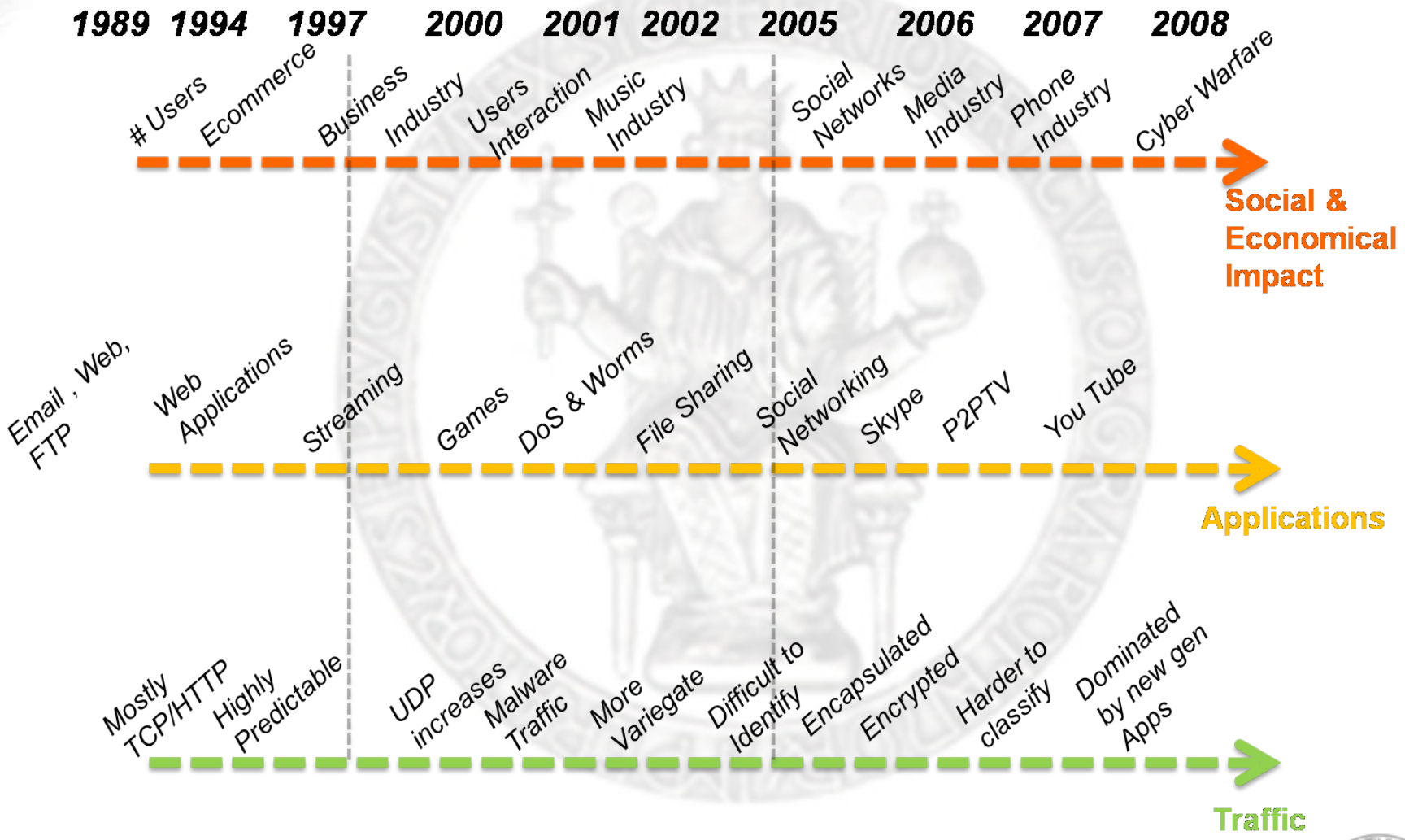
		2001-2005	2006-2011		Delta
1	Measurement tools & methodologies	32	70		38
2	Wireless networks and Wifi performance evaluation and optimization	8	43		35
3	Mobile	0	26		26
4	Topology	11	35		24
5	OSN	0	19		19
6	Traffic analysis & simulation	36	53		17
7	Traffic anomalies & event detection	19	35		16
8	Other applications level performance & optimization	9	24		15
9	Overlay network	14	28		14
10	Addressing	2	15		13
11	Streaming	8	19		11
12	Access Networks	3	12		9
13	IPTV	0	7		7
14	Traffic sampling & manipulation	27	31		4
15	CDN & content distribution performance	3	6		3
16	Privacy	0	3		3
17	Traffic classification	13	16		3
18	Censorship	0	2		2
19	Cloud	0	2		2
20	Datacenter	0	2		2
21	Economics	0	2		2
22	Ethics and Legality	0	2		2
23	Green	0	1		1
24	Network neutrality	0	1		1
25	Transport layer improvements	16	17		1
26	Network Devices	6	6		0
27	Traffic generation	2	2		0
28	Traffic matrix	5	5		0
29	DNS performance	9	8		-1
30	Tomography	10	9		-1
31	Traffic engineering	2	1		-1
32	Coordinate systems & Geolocationing	9	7		-2
33	Traffic characterization and modeling	20	17		-3
34	Bandwidth estimation	13	9		-4
35	Routing	24	17		-7
36	Network performance evaluation	24	10		-14

increasing

stable

decreasing

Topics follow/anticipate Internet evolution



Main Challenges and Obstacles in a nutshell

- Lack of available data and ground truth
  - traffic classification, topology discovery, anomaly detection, etc.
- Technology evolution (ever-increasing speed of network links, the growing use of encrypted communications, etc.)
  - traffic classification, passive monitoring, etc.
- Huge amount of network data
  - scalability of techniques and tools, mining and analysis, OSNs, etc.
- Ever-increasing need of privacy-preserving techniques
  - traffic identification, passive monitoring, etc.
- Lack of common practices
  - ✓ data management
  - ✓ rigorous evaluations and comparisons
  - ✓ repeatability
  - ✓ comparability
  - ✓ accuracy/precision

Part 2

Acknowledgements: Research Group

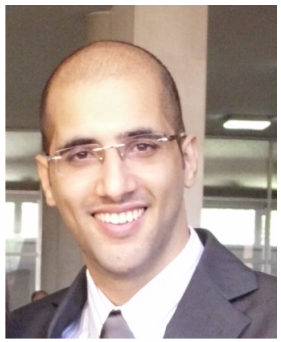
Alessio Botta

NM2

Giuseppe Aceto

Alberto Dainotti

Walter de Donato



Pietro Marchetta

Research Contributions

- Broadband Internet performance: a view from the edge (ACM Sigcomm 2011)
- Traffic Classification (IEEE Network Magazine)
- Accuracy of active probing (IEEE Com Mag 2010)
- Hybrid solutions for networks topology discovery (IEEE JSAC 2011)
- Traffic Monitoring of Mobile Broadband Networks (IEEE Com Mag 2011)

Broadband Internet performance: a view from the edge

S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, A. Pescapè, "Broadband Internet Performance: A View From the Gateway", ACM SIGCOMM 2011 proceedings, Toronto, ON, Canada, August 15-19, 2011.

W. de Donato, S. Sundaresan, N. Feamster, R. Texeira, A. Pescapè, "BISMark: A Platform for Studying Home Networks", USENIX NSDI'11, Boston, MA, USA, 30 March - 1 April, 2011.

What Affects Broadband Performance?

- Observing ISP activities is not straightforward
  - ✓ ISPs don't usually provide reliable information
  - ✓ SLAs don't consider important parameters
- Are users really getting what they pay for?


guardian.co.uk

News | Sport | Comment | Culture | Business | Money | Life & st

Money › Internet, phones & broadband

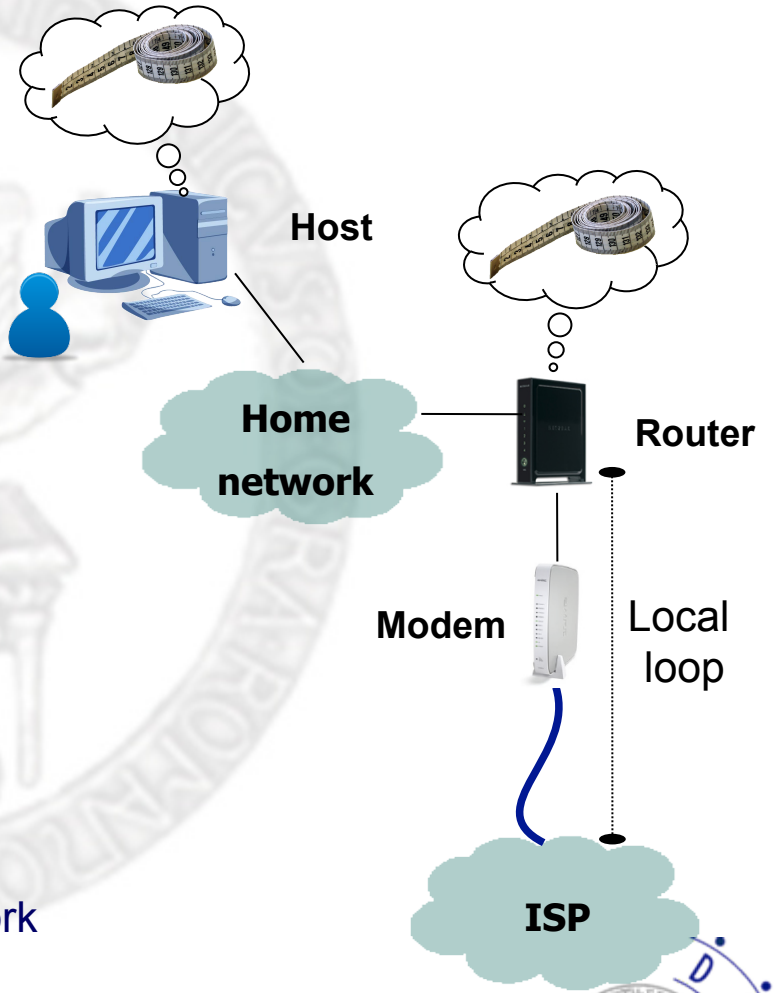
Ofcom: Broadband ISPs are pulling a fast one

- Average speed 46% below that promised by ISPs
- Mandatory code and clear penalties vital, experts say

- Important for regulators, consumers, ISPs
- Notion of performance is fuzzy
  - ✓ What metrics should we measure?
  - ✓ How to measure them?

Web- vs Client - vs Router-based

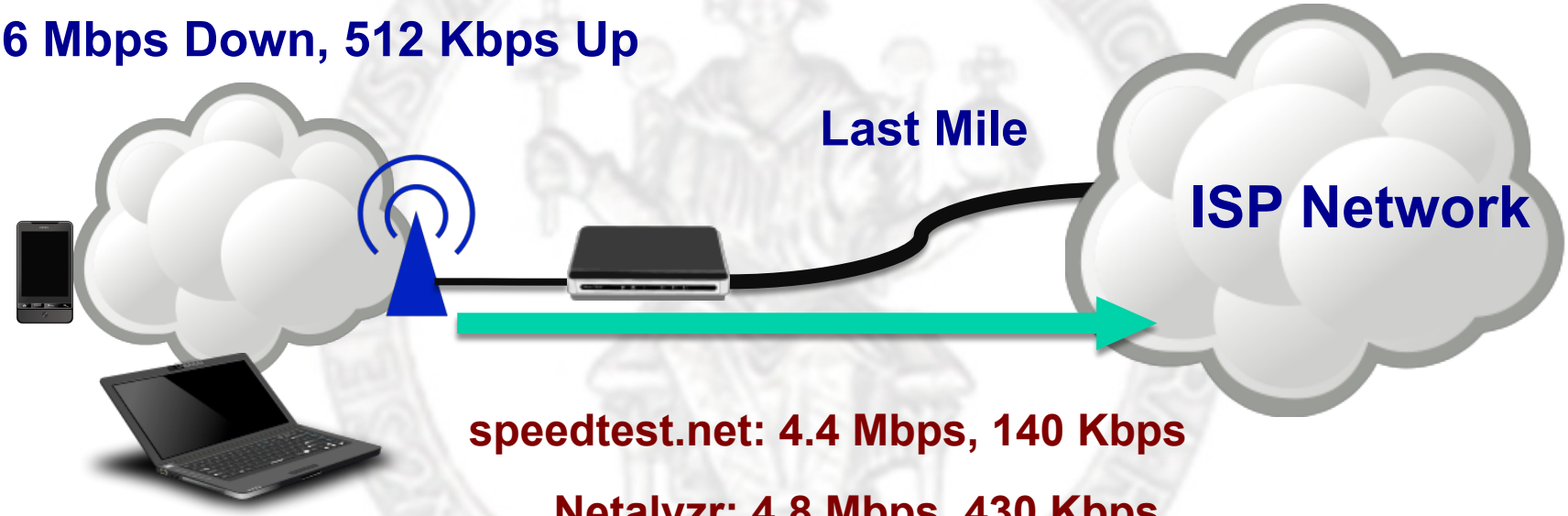
- Measuring from the edge → Independent point of view
- Different approaches
  - **Web-based** (*Speedtest.net*, *Netalizr*, ...)
    - easy to use
    - one-shot measure
    - affected by interferences
  - **Client-based** (*Grenouille*, *Isposure*, ***HoBBIT***, ...)
    - repeated/periodical measures
    - easy large scale deployments
    - active only when the PC is turned on
    - unable to account for interferences
  - **Router-based** (*SamKnows*, ***BISMark***)
    - continous periodical measures
    - observes all traffic passing through network
    - can take into account interfereces
    - difficult to obtain large scale deployments



The Case For the Gateway

Home Network: AT&T DSL

6 Mbps Down, 512 Kbps Up



speedtest.net: 4.4 Mbps, 140 Kbps

Netalyzr: 4.8 Mbps, 430 Kbps

Gateway: 5.6 Mbps, 460 Kbps

Gateway enables periodic measurements, and can account for confounding factors

Measuring from the gateway: the deployments

- Breadth: The FCC/SamKnows study
  - ✓ 4,000 gateways, 16 ISPs, multiple service plans
- Depth: The BISMark study
  - ✓ 16 gateways in Atlanta, on-demand measurements
  - ✓ http://projectbismark.net/
- Duration: Dec 2010 – Jan 2011 (ongoing project)

Currently Supported devices

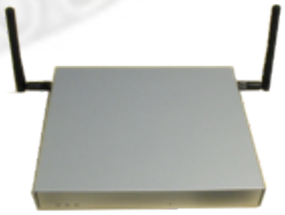
Netgear WNDR3700

680Mhz MIPS CPU

64 MB RAM

8MB Flash

Custom OpenWrt OS



NOX Box

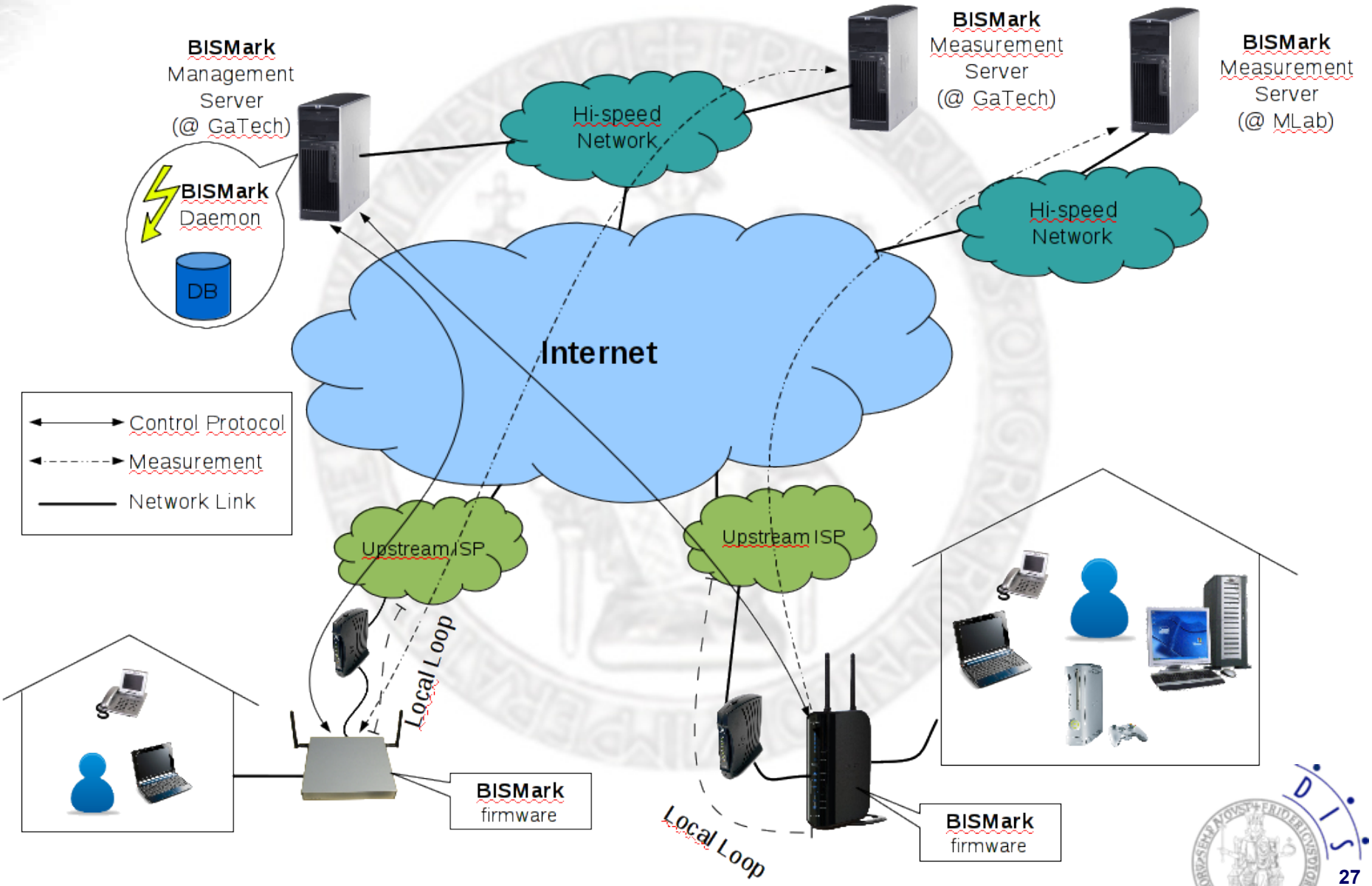
500Mhz Geode CPU

256 MB RAM

2GB Flash

Custom Debian OS

BISMark (1/2)


BISMark
Management
Server
(@ GaTech)
BISMark
Daemon
DB
Hi-speed
Network
BISMark
Measurement
Server
(@ GaTech)
BISMark
Measurement
Server
(@ MLab)
Hi-speed
Network
Internet
Control Protocol
Measurement
Network Link
Upstream ISP
Upstream ISP
Local Loop
Local Loop
BISMark
firmware
BISMark
firmware

BISMark (2/2)

Current worldwide deployment status (updated June 2011)

- 2 management servers
- more than 70 BISMark routers
- more than 50 measurement servers (Universities, MLab)

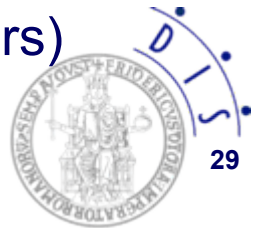
What do we measure?

- Actively

Parameters	Type	L4 Proto	Frequency	Tool
Latency	End-to-end Last-mile Upstream load Downstream load DNS	ICMP, UDP ICMP ICMP ICMP UDP	5 min 5 min 30 min 30 min 5 min	ping, D-ITG ping ping ping nslookup
Packet loss	End-to-end	UDP	15 min	D-ITG
Jitter	End-to-end	UDP	15 min	D-ITG
Downstream throughput	Single-thread Passive Capacity	TCP N/A UDP	30 min 30 min 12 hrs, 60 min	Curl (GET), D-ITG /proc/net/dev shaperprobe, D-ITG
Upstream throughput	Single-thread Passive Capacity	TCP N/A UDP	30 min 30 min 12 hrs, 60 min	Curl (PUT), D-ITG /proc/net/dev shaperprobe, D-ITG

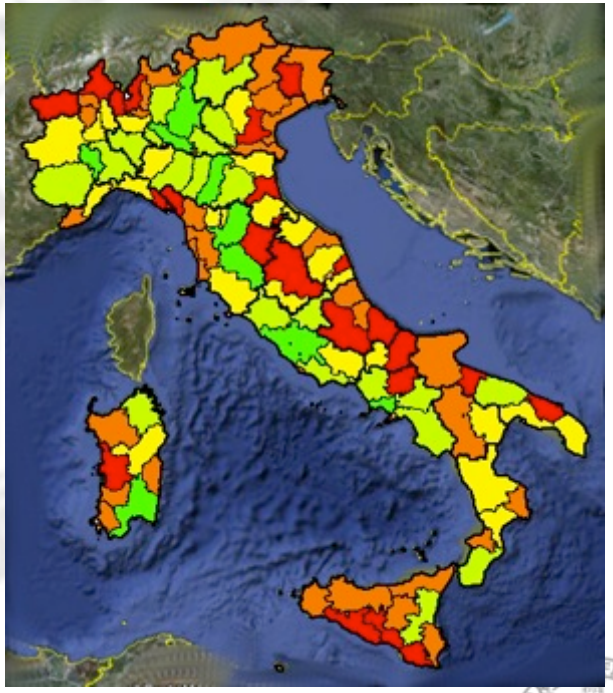
We are also collecting per-application throughput (currently just port-based)

- Passively (with anonymization)
  - Flows (application labels, packet-level stats, HTTP headers)
  - Wireless activity, DHCP events, ARP associations

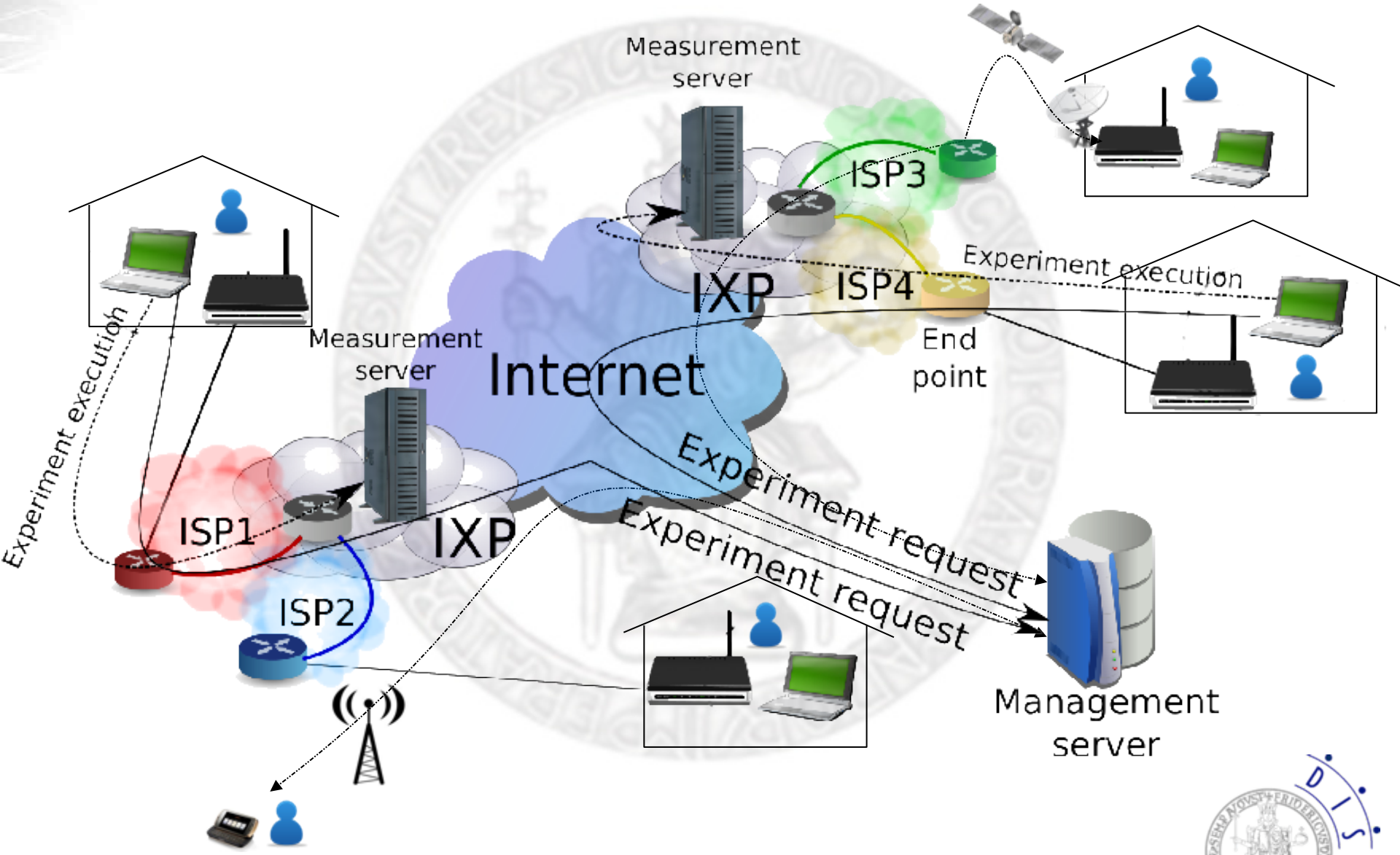


HobbIT: (1/2)

- ***Host Based Broadband Internet Telemetry***
- Network measurements taken from the users' PC
  - ✓large scale deployments
  - ✓National (Italy) project
- Active measurements using standard tools
  - ✓extensible measurement framework
  - ✓data geolocation and mapping
  - ✓fine-grained resource management
- Main features
  - ✓multi-platform client
  - ✓automatic updates
  - ✓per-application measurements
- Users can
  - ✓monitor their Internet connection
  - ✓compare results with others in the same location
- ✓ *http://hobbit.comics.unina.it/*



HobbIT (2/2)


Measurement server
ISP3
IXP
ISP4
End point
Experiment execution
Measurement server
Internet
Experiment execution
ISP1
IXP
Experiment request
Experiment request
ISP2
Management server

What we are doing and contributions

- Study of access networks from both the gateway and the client
  - ✓ gateway provides unique insight into home network
- Insight into how to benchmark performance
  - ✓ Throughput measurement technique depends on usage scenario and measurements are affected by measurement techniques and shaping
  - ✓ Traffic shaping is highly variable across users
  - ✓ Effect of access link properties, home equipment
    - Access link characteristics affect performance
    - Modem buffers induce high latency
- Insight into broadband performance
  - ✓ Consistency of performance
  - ✓ Loss, jitter, time-of-day effects
- ISPs generally deliver consistent throughput, with some time-of-day effects
- Ongoing
  - ✓ Comparison among client- and router-based approaches
  - ✓ How to mitigate buffering effects
  - ✓ Home networks analysis

Traffic Classification

A. Dainotti, A. Pescapé, K. C. Claffy, “Issues and Future Directions in Traffic Classification", IEEE Network, 2011, to appear

G. Aceto, A. Dainotti, W. de Donato, A. Pescapè, "PortLoad: taking the best of two worlds in traffic classification", IEEE INFOCOM 2010 - WIP Track - March 2010, San Diego (CA, USA)

A. Dainotti, W. De Donato, A. Pescapè TIE: a Community-Oriented Traffic Classification Platform", International Workshop on Traffic Monitoring and Analysis (TMA'09) @ IFIP Networking 2009 - May 2009, Aachen (Germany)

Traffic Classification

- Associating traffic flows to network applications that generate them
- Approaches
  - ✓ **Port-based**
  - ✓ **Payload inspection**
  - ✓ **Pattern Recognition**

Traffic Classification

- Recent interest of Research & Industry
  - ✓Ports are not reliable anymore
  - ✓Payload-based approaches have issues
  - ✓New applications
  - ✓New scenarios (3G networks+Smartphones+Video)
  - ✓No perfect solution up to today

Challenges (1/2)

- Traffic Evolution
  - ✓ New applications (often) with undisclosed proprietary protocols (e.g. *Skype*)
    - New applications emerge continuously and it is difficult to investigate each of them in order to update approaches and/or signatures
  - ✓ Protocol encapsulation
    - E.g. over HTTP (*MSN, Kazaa,* …)
  - ✓ Encryption
    - Application payload
    - Application protocol encapsulation (SSL, SSH, …)
    - Network level (IPSec Tunnels, …)
- Link speed: we often need to do classification online
  - ✓ Speed / computational complexity of algorithms
    - Payload inspection (complexity)
    - Other approaches (how much data do we need?)
  - ✓ Storage

Challenges (2/2)

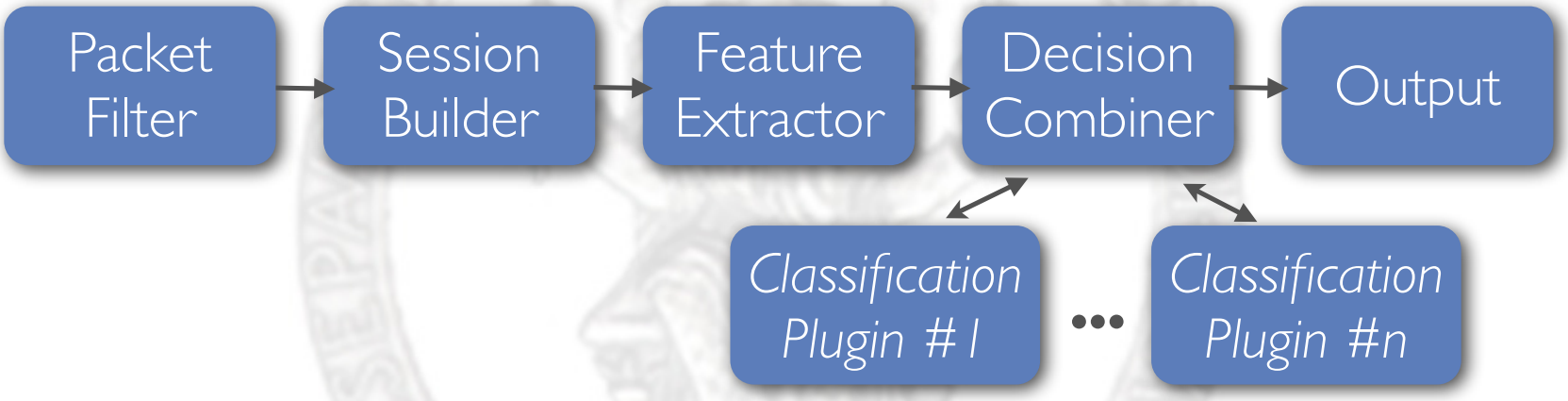
- Privacy
  - ✓ How invading a technique is?
  - ✓ Access to full payload may be not allowed
  - ✓ Storage may be not allowed
  - ✓ Trace anonymization (issues)
- Ground truth
  - ✓ Payload-based (eg L7filter)
  - ✓ Heuristics
  - ✓ Manual Inspection
  - ✓ Alternative techniques requiring user collaboration
- Available data
  - Traffic traces from operational networks
- Consistent Evaluation and Comparison Methods
  - ✓ Rigorous evaluation and comparison of techniques requires standard testing and validation procedures and benchmarking metrics

What we are doing

- Developing and sharing a community platform for supporting the research and experimentation in the field of traffic classification
- Developing novel traffic/application/service classification techniques
- Traffic Classification against new comers applications and services
- Integration and combination of different approaches (multi-classification)

TIE: Traffic Identification Engine (1/3)

- ✓ TIE is an open-source software platform to allow the research community to work with *shared tools and data*
- ✓ *http://tie.comics.unina.it*



TIE: Traffic Identification Engine (2/3)

- Support for multiple approaches and techniques
- Allow the comparison of different techniques
- Able to act as a multi-classifier
- Target online classification
- Three available operating modes: *Offline*, *Realtime* and *Cyclic* mode.
- Written in C, runs on Linux and FreeBSD platforms

TIE: Traffic Identification Engine (3/3)

- Support for different definitions of *objects*
  - ✓ *Flows, Biflows, TCP connections, Hosts*
- Support for different definitions of *classes*
  - ✓ *Application IDs, Sub-IDs, Group IDs*
- *Easy to add*: classification techniques, classification features, combination strategies
- API for *Classification Plugins*
- *Defined format* of Output & Input Tables
- *Tools* for numerical and graphical analysis and comparison
  - ✓ Several common *metrics*: Accuracy, Byte-Accuracy, Precision, F-Measure, Recall
  - ✓ Confusion Matrix

Novel Classification Technique

- Technique is called ***PortLoad****
  - ✓ **Port-based** is *fast* and *privacy-friendly* because:
    - It needs the *1st* packet only
    - It uses fixed fields (protocol and port)
    - It uses few data

    It can be considered as a special case of ***packet-classification*** techniques developed for routers, flow-monitors, etc.
  - ✓ **Payload-based** is *accurate* because relies on application-level headers and other information from the payload
    - Payload-based signatures
- ***Port + Payload = PortLoad***

- G. Aceto, A. Dainotti, W. de Donato, A. Pescapè, "PortLoad: taking the best of two worlds in traffic classification", IEEE INFOCOM 2010 - WIP Track - March 2010, San Diego (CA, USA)

* *Patent N.*: NA2010AOOOO11

Conclusions

- Traffic Classification is important for understanding and controlling the Internet traffic.
- Despite the large quantity of research works there are still several open issues.
- Because of the continuously evolving scenario and the emergence of new applications, research in this field will probably keep being very active in the future.
- Multi-Classification can achieve higher accuracy than any single classifier, and are more robust to changes in the sample population, including the nature and mix of applications ("concept drift").
- Common tools and techniques are needed.

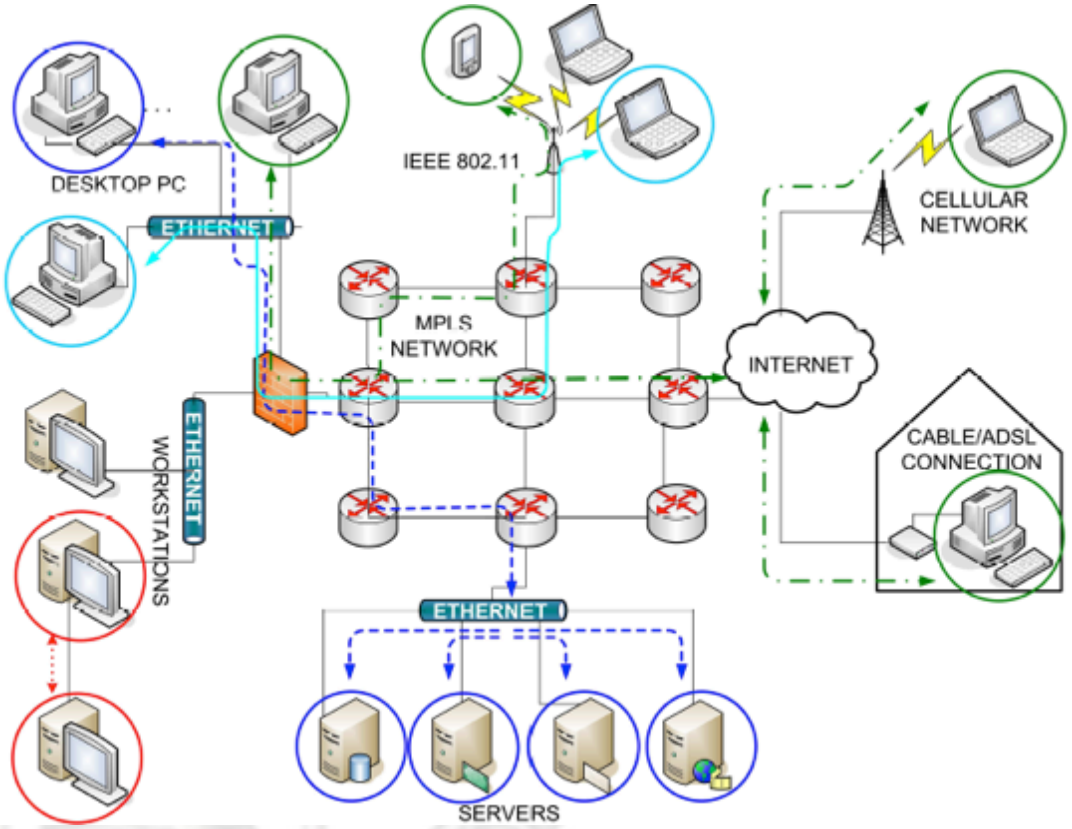
Accuracy of Active Probing

A. Botta, A. Dainotti, A. Pescape', "Do You Trust Your Software-based Traffic Generator?", IEEE Communications Magazine, vol.48, no.9, pp.158-165, Sept. 2010.

A. Botta, A. Dainotti, A. Pescapè, "Do you know what you are generating?", Poster at Co-Next 2007 Student Workshop. 2-pages abstract to be published in Co-Next '07 Proceedings

Active Probing and Application Traffic Generation

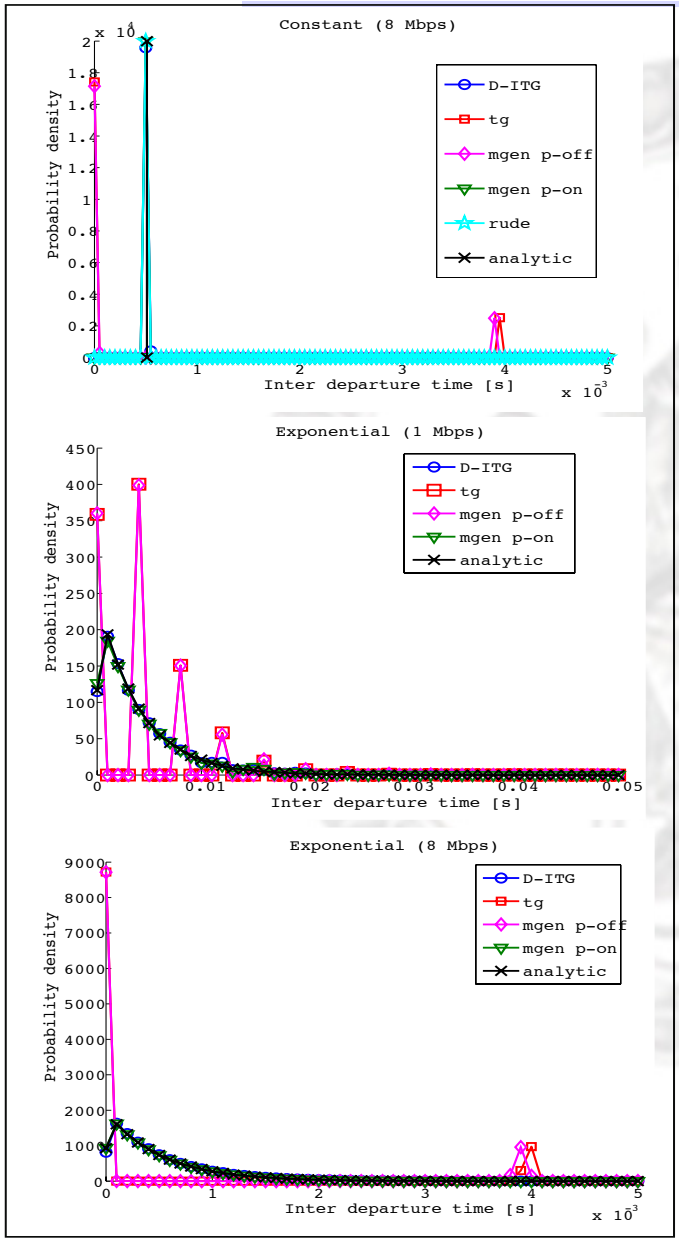
- **Why?**
  - Network Performance
  - Testing/benchmarking
    - Network Infrastructures
    - Device capabilities
    - Quality of Service (QoS) architectures
    - Queuing disciplines
    - Traffic shapers
    - Etc.



Traffic generation scenario

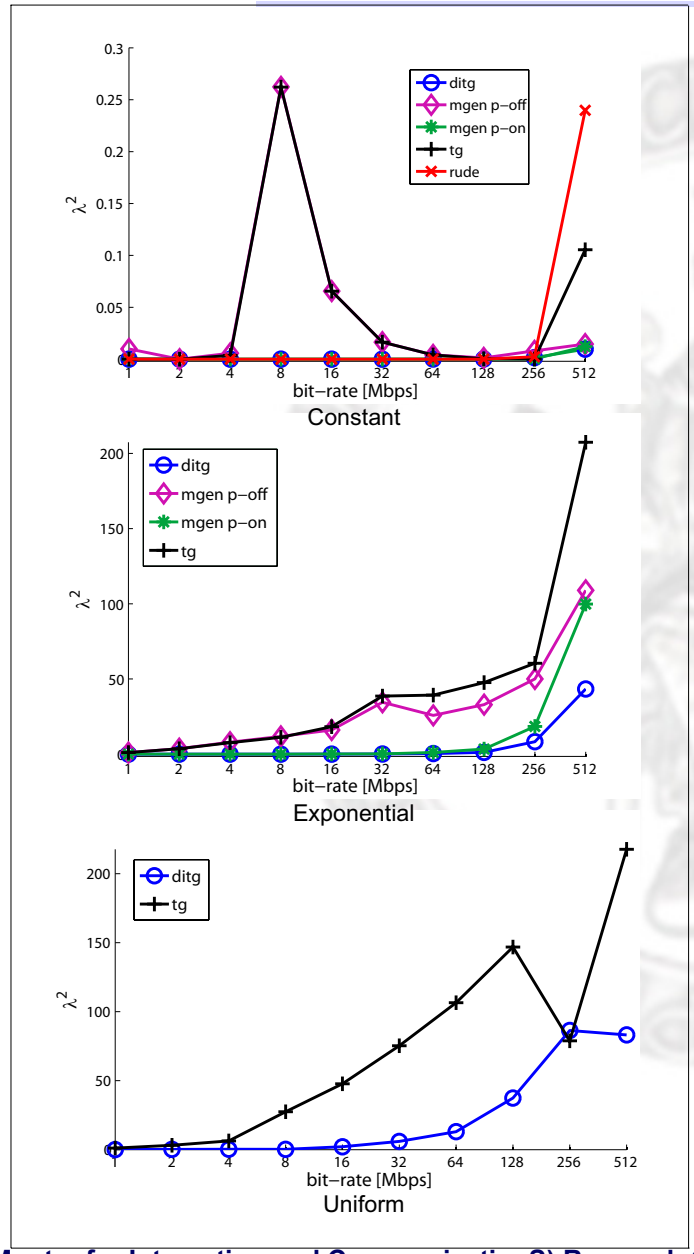
- **What?** Generation of **realistic traffic**
  - replicating as accurately as possible real applications
  - collecting information on how the single packets have been processed by the SUT

Accurate packet timings



- Constant and Exponential distribution at 1 and 8 Mbps (low rates)
- IPT generation process is often poisoned even starting from such low rates (i.e., within the stable working range of the generator)

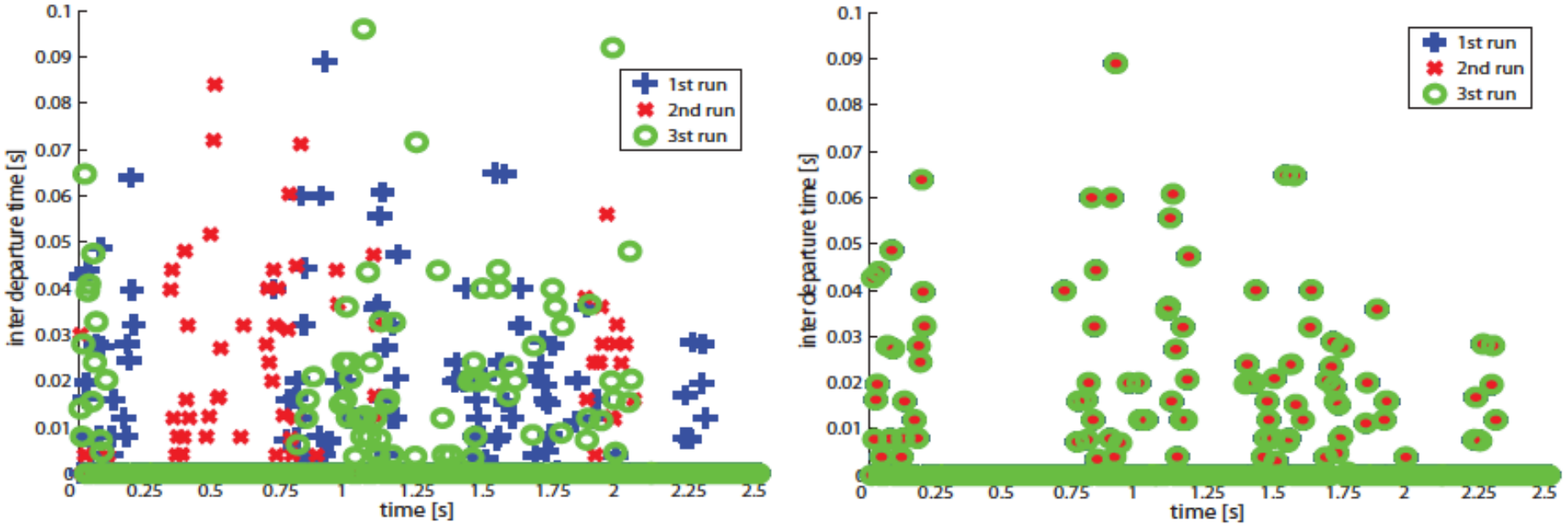
Accurate packet timings



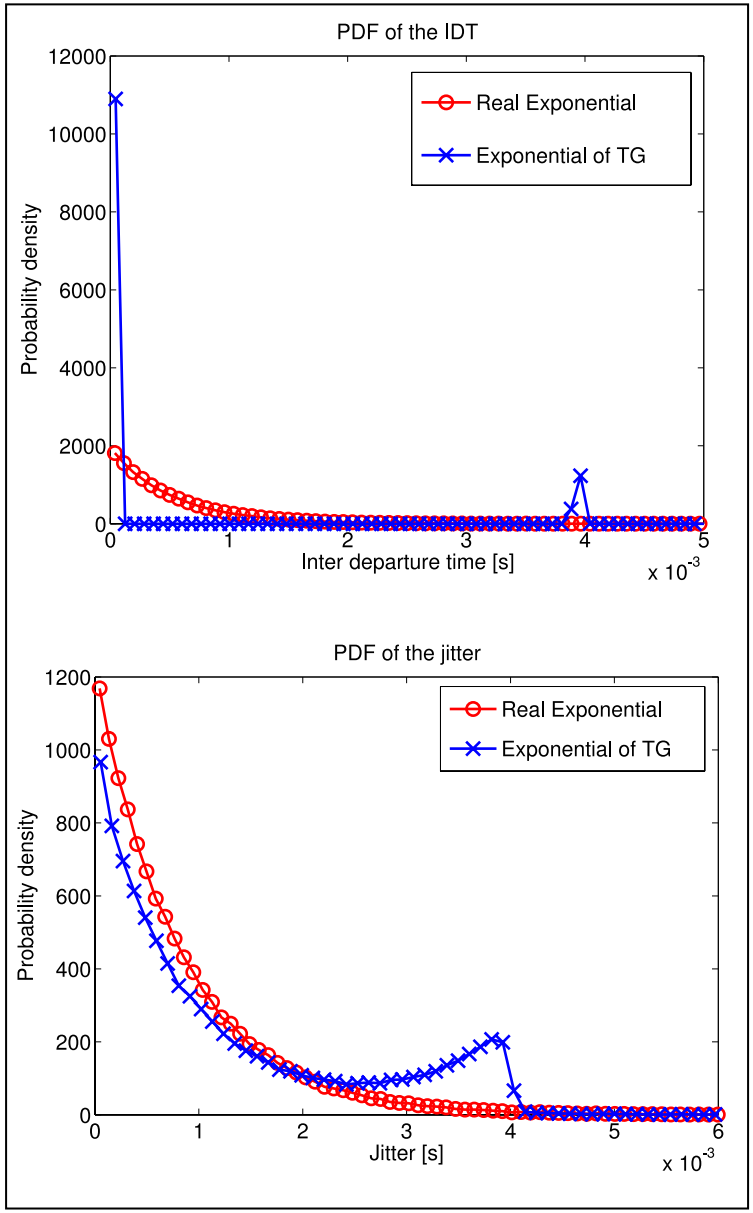
- As the requested rate becomes higher all generators increasingly deviate from the expected distribution.
- We conclude that we cannot take for granted distributions that could be very different from those imposed.

Accurate packet timings

- Setting the seed for random-based generation of statistical profiles



Accurate packet timings



- The impact of inaccurate traffic profiles can significantly alter the results of the measurement
- Ill-behaved generators can affect the results of an experiment.
- The two PDFs of the jitter are quite different.

What we are doing

- We are developing a community platform for supporting the research and experimentation in the field of active probing and application traffic generation
  - ✓ *http://www.grid.unina.it/software/ITG/*
- We are improving the platform to adhere to the *realistic generation* feature
  - ✓ Protocols
  - ✓ Applications
  - ✓ Metrics
  - ✓ Etc.
- We are working on the *metrological review* of software-based active probing architectures
  - ✓ to overcome inaccuracies due to SW/OS (Polling, IDT-recovery, Buffered binary logging, etc.)

Traffic Monitoring over Mobile Broadband Networks

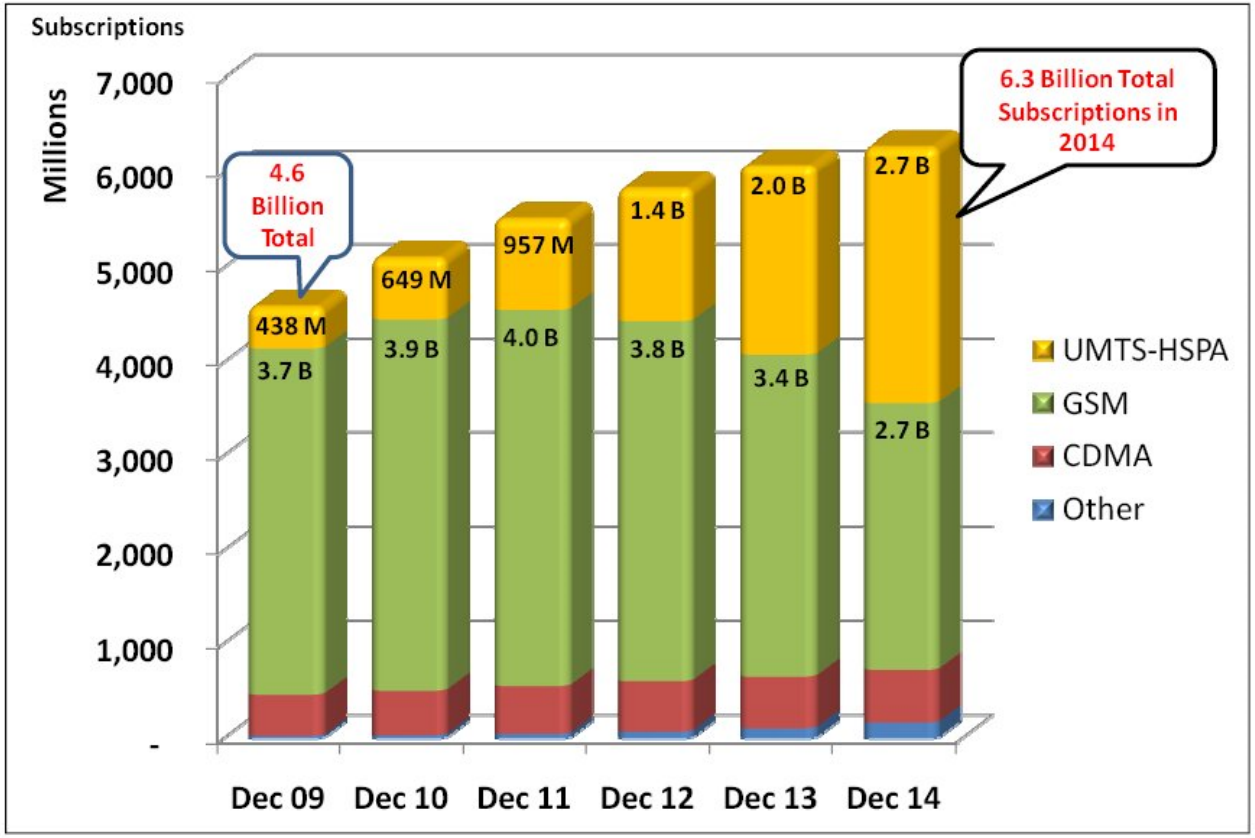
A. Botta, A. Pescapè, C. Guerrini, M. Mangri, A Customer Service Assurance Platform for Mobile Broadband Networks, IEEE Communications Magazine, October 2011

A. Botta, A. Pescapè, "Monitoring and measuring wireless network performance in the presence of middleboxes", The 8th International Conference on Wireless On-demand Network Systems and Services (WONS), Bardonecchia (TO), Italy, January 2010

A. Botta, A. Pescapè, G. Ventre, E. Biersack, S. Rugel, "Performance footprints of heavy-users in 3G networks via empirical measurement", 6th International workshop on Wireless Network Measurements (Winmee) 2010, Avignon, France, May 2010

Wireless Networks Evolution

- Users: Wireless Subscribers Forecast 2014

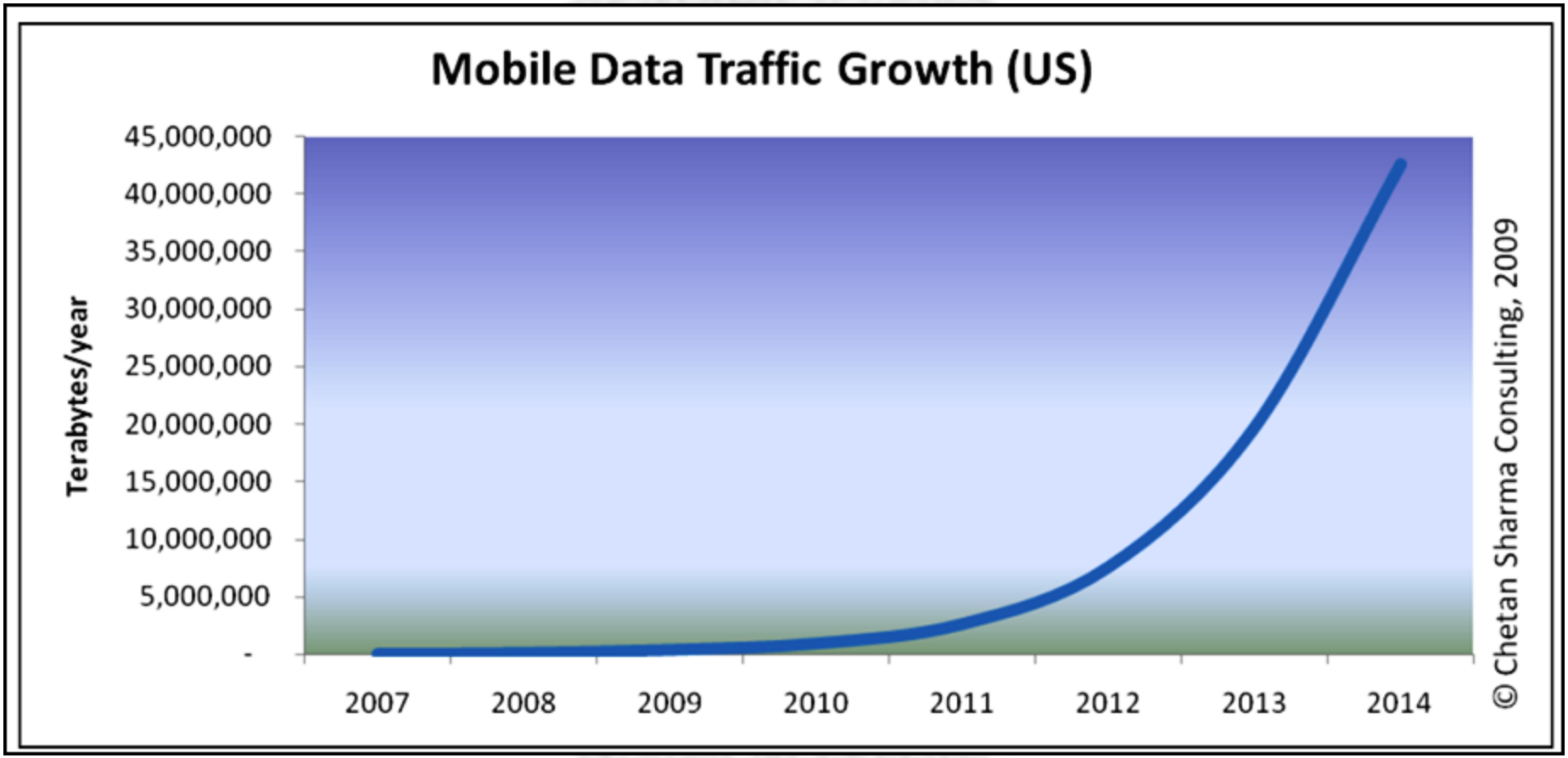


	2008	2009	2010	2011	2012	2013	2014
WiMAX	0.5	2.8	7.5	16.7	37.1	82.1	
LTE	0.0	0.0	0.5	3.5	13.1	44.5	131.5
HSPA	304	438	649	957	1400	2000	2700

Source: Informa Telecoms & Media, WCIS+, June 2009

Wireless Networks Evolution

- Data: Mobile Data Growth in the United States



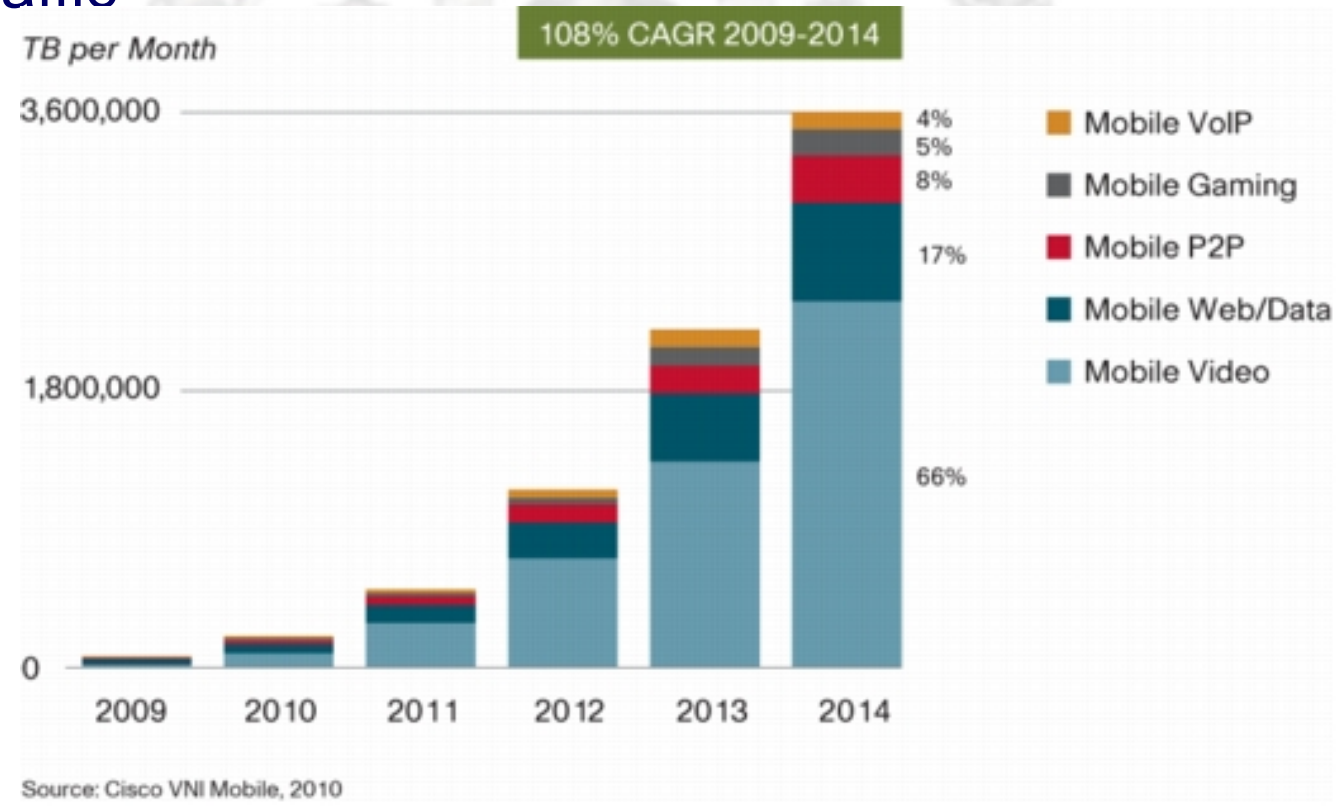
Source: Managing Growth and Profits in the Yottabyte Era
Chetan Sharma, July 2009.

Wireless Networks Evolution

- Shifts in Internet Services and Applications
  - ✓ Many content providers and consumers
  - ✓ User Generated Content (UGC)
  - ✓ Services become ubiquitous and more interactive
  - ✓ Large scale communities
  - ✓ Fastly spreading new *killer applications*
  - ✓ But, most of all:
    - Video
    - Video
    - Video
    - Video
    - VideoA

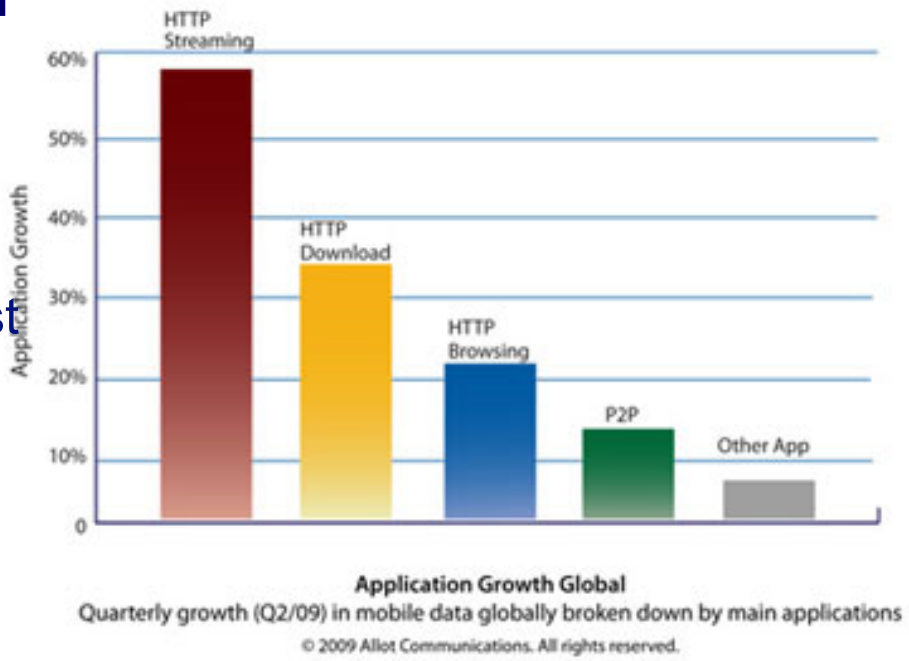
Wireless Networks Evolution

- Overall mobile data traffic is expected to grow to 3.6 exabytes per month by 2014
- Over 2.3 of expected 3.6 exabytes are due to mobile video traffic



Wireless Networks Evolution

- The "***YouTube Effect***" (2009)
  - ✓ Streaming surpassing P2P in 3G networks.
  - ✓ According to data collected by Allot Communications
    - HTTP streaming was the fastest growing application in terms of mobile bandwidth usage in the second quarter and accounted for nearly a quarter of world's 3G network traffic.
    - streaming video and audio is now equal to or greater than P2P mobile traffic in all regions of the world and definitely growing at a faster clip.

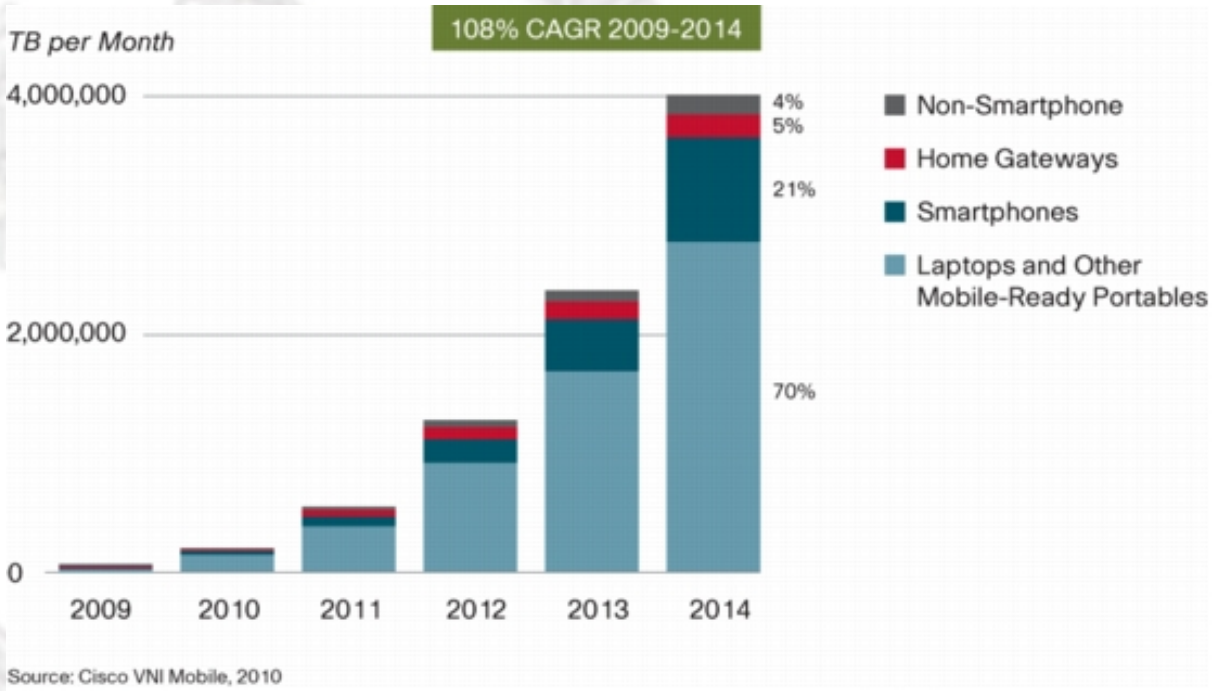


Application Growth Global

Quarterly growth (Q2/09) in mobile data globally broken down by main applications

Wireless Networks Evolution

- Smartphones and portables will account for 91 percent of all mobile data traffic by 2014.
- The average smartphone user generates 10 times the amount of traffic generated by the average non-smartphone user.
- Handset traffic is highest in regions with the highest smartphone penetration (Italy).



What we are doing

- Impact of new bandwidth-hungry applications over broadband mobile networks
- Mobile video consumption trends
- Traffic Management (eg effect of shapers and other middleboxes operated by ISPs and Telcos)
- How device type is driving data traffic
- How workload is changing
- Performance issues
  - ✓ TCP connections limited by the network and TCP connections limited by end-user devices
- Impact of traffic migration from fixed to mobile networks
  - ✓ In many countries in Europe, mobile operators are offering mobile broadband services at prices and speeds comparable to fixed broadband.

Hybrid solutions for networks topology discovery

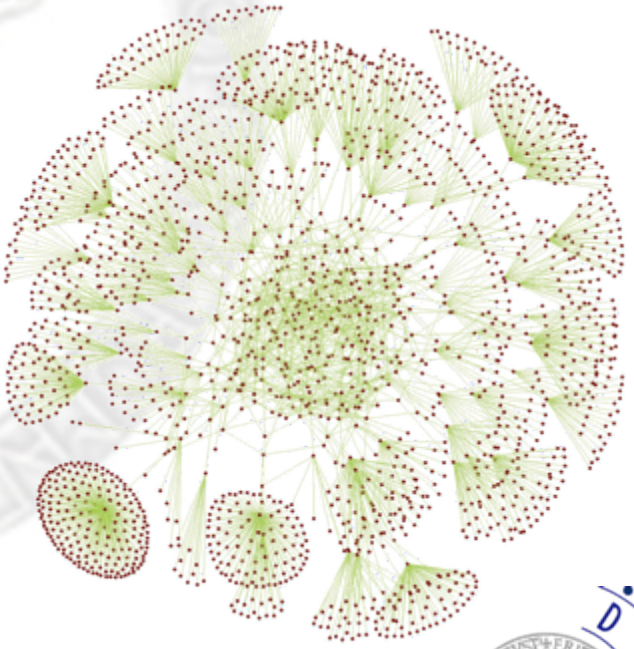
Pietro Marchetta, Pascal Mérindol, Benoit Donnet, Antonio Pescapé and Jean-Jacques Pansiot. "Topology Discovery at the Router Level: A New Hybrid Tool Targeting ISP Networks". IEEE Journal on Selected Areas in Communication (JSAC), Special Issue on Measurement of Internet Topologies, 2011, October

NM^2: Network Mapping (1/3)

- **Why?**
  - ✓ Network control and management
    - Fault isolation, performance analysis, service locations, etc.
  - ✓ Network simulations
    - It is difficult to generate realistic topologies
  - ✓ Network aware applications
    - E.g. to improve the performance
- **What?**
  - ✓ Automatic discovery of network maps in terms of: routers, links, subnets, layer-2 devices, etc.
  - ✓ Achieving
    - Completeness (i.e. discover the entire topology)
    - Accuracy (i.e. make no mistakes)
    - Low intrusiveness (i.e. reduce both the discovery duration and the traffic overhead)
  - Integration with Network Inventory solutions

NM^2: Network Mapping (2/3)

- **How?**
  - ✓ Combining multiple passive/active methodologies and techniques
  - ✓ Hybrid approaches
  - ✓ Novel techniques based on: IGMP, ParisTraceroute, IP Options, ...
  - ✓ Hynetd (single vantage point)
    - *http://www.grid.unina.it/software/TD*
  - ✓ MERLIN (multiple vantage points)
    - *http://svnet.u-strasbg.fr/merlin*



COST-TMA

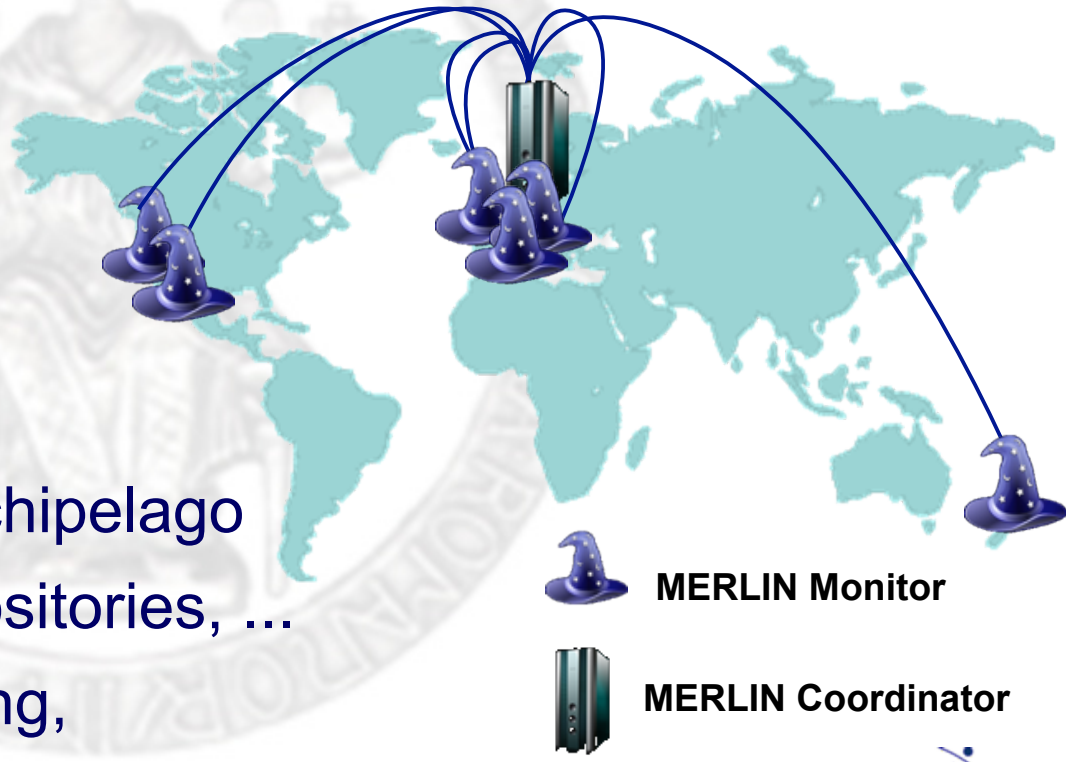
LSiiT

ip networking lab
uci, louvain-la-neuve, belgium

NM²: Network Mapping (3/3)

MERLIN: MEasure the Router Level of the Internet

- Target a specific Autonomous System network
- Multiple techniques integrated and optimized
  - Improved IGMP probing
  - ✓ Paris traceroute
  - ✓ Alias resolution
- Several input sources
  - BGP dumps, CAIDA Archipelago datasets, MaxMind repositories, ...
- Geo-Location, DNS mapping, IPtoAS mapping, ...



PART 3
Conclusion Remarks

The actors and the community

- A lot of folks are (still) measuring the Internet
  - ✓ Organizations
    - CAIDA
    - SLAC
    - RIPE
    - WIDE

  - ✓ Research groups all over the world
  - ✓ *Well Known* workshops and conferences (IMC, PAM, etc)
  - ✓ *New comers* workshops at flagship conferences (eg, W-MUST at SIGCOMM 2011)

Research Projects

- (recent) Research Projects
  - ✓EU

    - Cost ACTION **IC0703 TMA**, Traffic Monitoring and Analysis
    - **DEMONS**, DEcentralized, cooperative, and privacy-preserving MONitoring for trustworthinesS
    - **PRISM**, PRIvacy-Aware Secure Monitoring
    - **MOMENT**, Monitoring and Measurement in the Next generation Technologies
  - ✓WorldWide

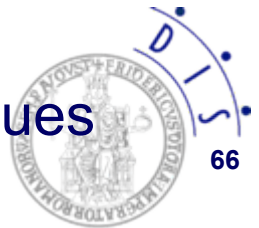
MLAB

    - Measurement Lab
    - PlanetLab supports a lot of works related to NM2

Summary

1	Measurement tools & methodologies
2	Wireless networks and Wifi performance evaluation and optimization
3	Mobile
4	Topology
5	OSN
6	Traffic analysis & simulation
7	Traffic anomalies & event detection
8	Other applications level performance & optimization
9	Overlay network
10	Addressing
11	Streaming
12	Access Networks
13	IPTV
14	Traffic sampling & manipulation
15	CDN & content distribution performance
16	Privacy
17	Traffic classification
18	Censorship
19	Cloud
20	Datacenter
21	Economics
22	Ethics and Legality
23	Green
24	Network neutrality
25	Transport layer improvements
26	Network Devices
27	Traffic generation
28	Traffic matrix
29	DNS performance
30	Tomography
31	Traffic engineering
32	Coordinate systems & Geolocationing
33	Traffic characterization and modeling
34	Bandwidth estimation
35	Routing
36	Network performance evaluation

- Network Monitoring and Measurements
  - ✓ is a really active research field
  - ✓ is becoming a (mature) discipline…
  - ✓ but…still needs common and rigorous approaches when producing results (*see next*)
  - ✓ with a number of interesting emerging topics (datacenters, virtualization, neutrality, censorship, privacy, user experience, access networks, wireless, etc.)
  - ✓ facing with scaling issues



Strategies for Sound Internet Measurement

- Summary of Strategies:

Strategy #1:	*maintain meta-data*
Strategy #2:	*run your intended methodology by colleagues*
Strategy #3a:	*examine outliers and spikes*
Strategy #3b:	*employ self-consistency checks*
Strategy #3c:	*compare multiple measurements/computations*
Strategy #4:	*structure for reproducible analysis*
Strategy #5:	*periodically analyze ongoing measurements*
Strategy #6:	*package analysis for "data reduction requests"*
Strategy #7:	*subsample large datasets, assess variability*

V. Paxson, Strategies for Sound Internet Measurement, Proc. ACM IMC, October 2004.
Source: http://www.icir.org/vern/talks/vp-strategies-imc04.pdf

Rules on how to manipulate "data"

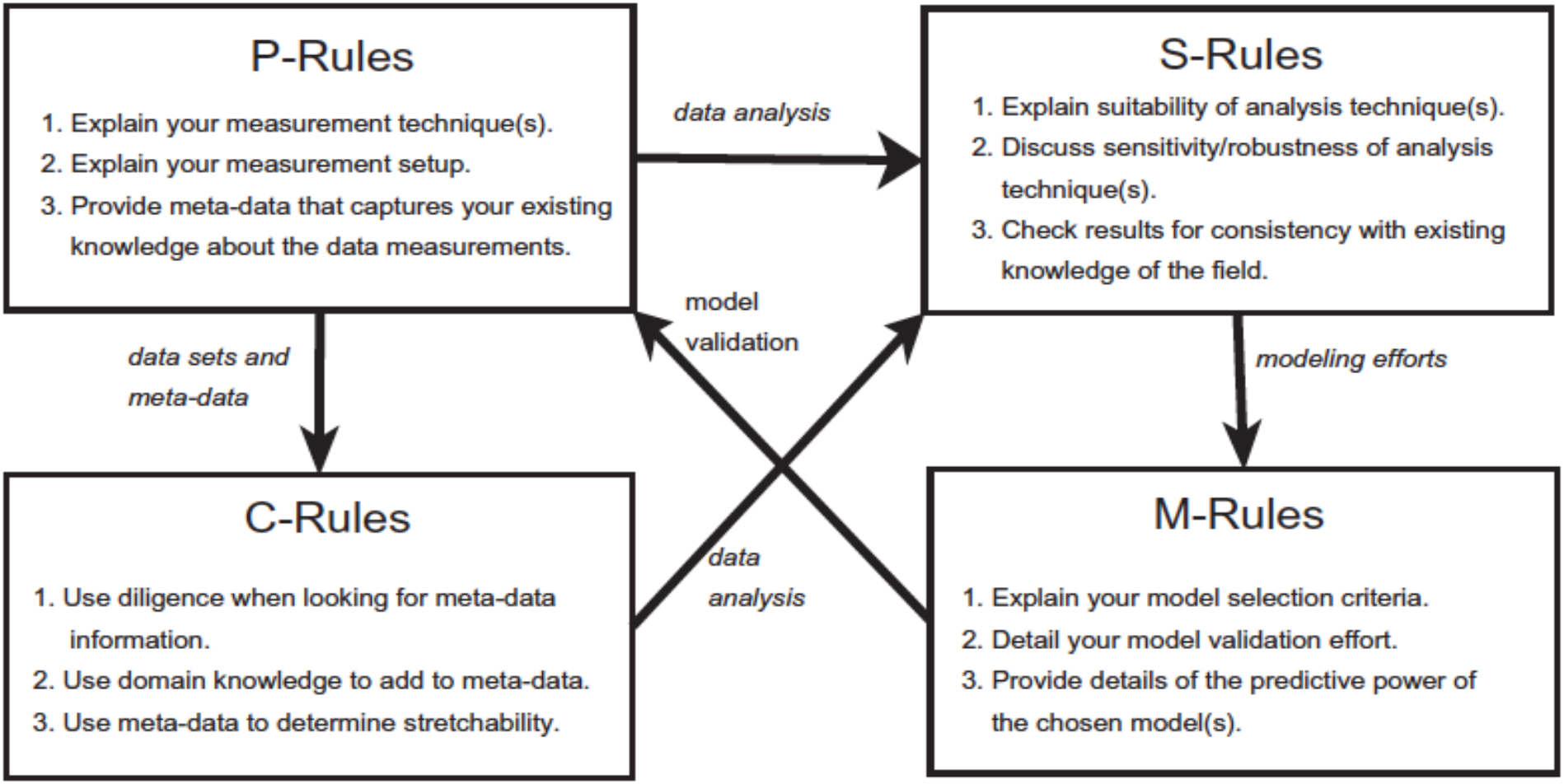


Fig. 1. The Socratic approach in a nutshell.

Source: *"A Socratic method for validation of measurement-based networking research", from Bala Krishnamurthy, Walter Willinger et al. , Computer Communications 2011*

Conclusion

- In the past, no significant difference between *operations* a *research* (due to size, users, data, etc)
- Nowadays, an increasing difference between operational and research measurement
  - ✓Results in research could be wrong in operation
  - ✓In-vitro experimentations could led to wrong assumptions due to scaling limitations

Recommendations

- Sharing and publication of open-source monitoring and measurement implementations and datasets
  - *Code to the data?*
- Integration of real/operational networks and large scale testbeds
- Increase the collaboration among researchers and operators
  - Cross-comparisons of both real tools and real data are necessary to advance the entire field over both operational networks and research testbeds

Thanks for the attention!!!

pescape@unina.it

http://wpage.unina.it/pescape

Credits

- Les Cottrell, Alberto Dainotti, Pietro Marchetta, Vern Paxson, Ernst Biersack, Srikanth Sundaresan, Walter de Donato, Nick Feamster, Renata Teixeira, Sam Crawford, Giuseppe Aceto, Alessio Botta, and all my co-authors.

Contacts

Antonio Pescape'

Dipartimento di Informatica e Sistemistica

University of Napoli "Federico II"

Via Claudio, 21 - 80125, Napoli (Italy) [Room n. 3.10]

tel. +39 081 7683856

fax +39 081 7683816

e-mail: pescape@unina.it (or pescape@ieee.org)

Personal web-page: http://wpage.unina.it/pescape

Teaching web-site (in Italian): http://www.docenti.unina.it/antonio.pescape