
Ongoing research activities

Artur Ziviani

National Laboratory for Scientific Computing (LNCC)
Petrópolis, Brazil

ziviani@lncc.br

<http://www.lncc.br/~ziviani>

Where am I since 2004?

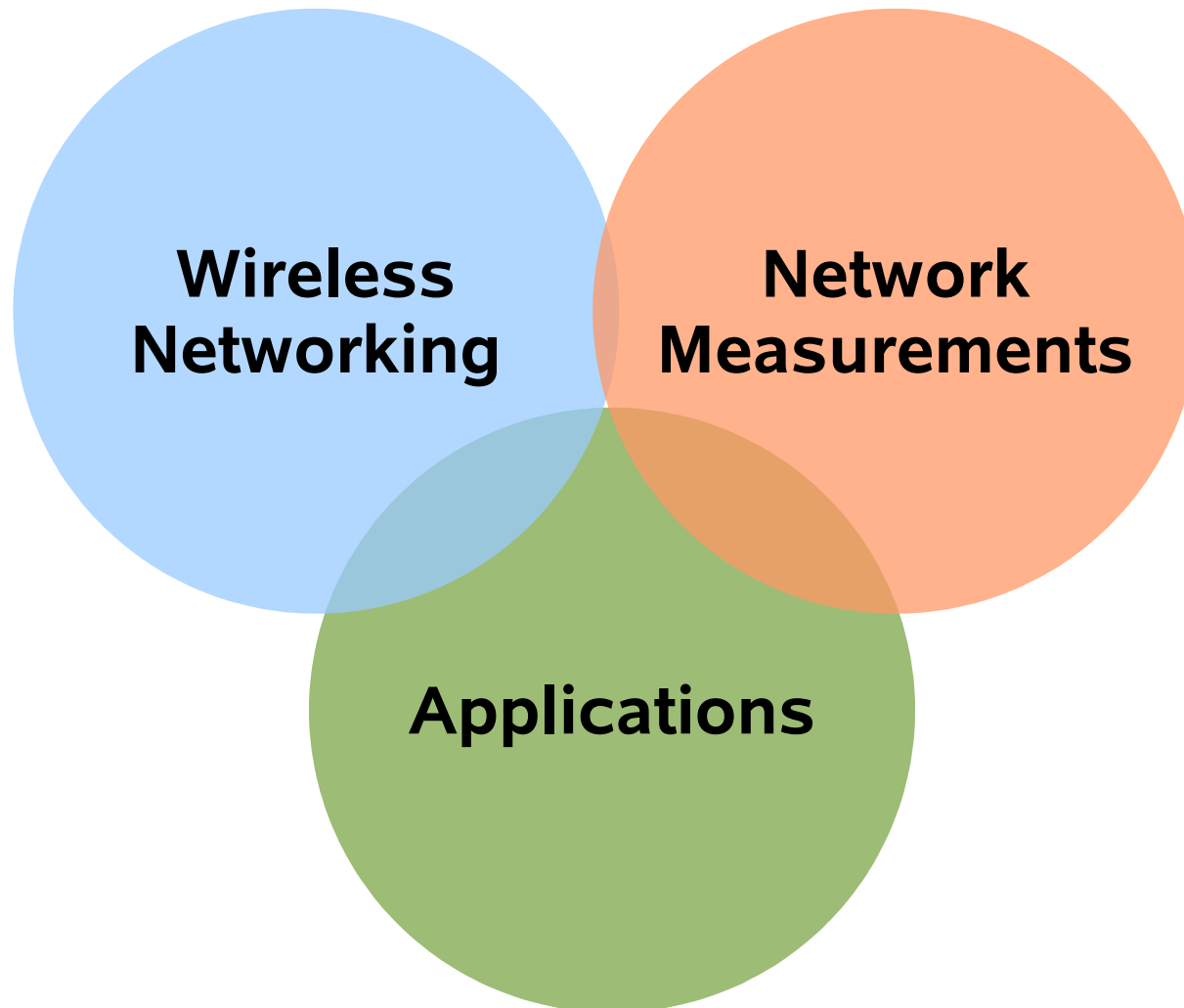


- **National Lab for Scientific Computing (LNCC)**
 - research unit of the Brazilian Ministry of Science and Technology
 - located at Petrópolis (~70 km far from Rio de Janeiro)

LNCC

- **About 50 PhDs in different fields**
 - mathematics, physics, biology, engineering, computer science
- **Multidisciplinary approaches on**
 - grid computing
 - support to both infrastructure and distributed applications
 - bioinformatics
 - computer-aided medicine (e.g. telemedicine)
 - mathematical modeling of Amazonian problems

Ongoing R&D activities



Wireless networking

- **Sensor networks**

- topology control

- with LIP6 and FOKUS Berlin

- data dissemination to support mobile sink nodes

- with LIP6 and INRIA-INSA Lyon

- **Service discovery within ad hoc networks**

- with PUC-Rio and INRIA-INSA Lyon

- **Telemedicine for emergency assistance**

- with Medical School of UFRJ

Network measurements

- **Network support for**
 - collaborative virtual environments (CVEs)
 - large-scale grid computing
- **Measurement-based methods for**
 - geolocation of Internet hosts
 - different parts with UFRJ, LIP6, Univ. Boston, and Univ. Catholique de Louvain
 - performance evaluation and capacity planning of web search systems
 - with UFMG, Google Engineering, and Yahoo! Research
 - network anomaly detection



Network Anomaly Detection using Nonextensive Entropy

**A. Ziviani, M.L. Monsores,
P.S.S. Rodrigues, A.T.A. Gomes**

National Laboratory for Scientific Computing (LNCC)
Petrópolis, Brazil

ziviani@lncc.br

<http://www.lncc.br/~ziviani>

Introduction

- **Network traffic anomaly**

“significant and unusual changes on the traffic patterns in one or multiple links, be them intentional or not”

- **Possible causes within an AS**

- ongoing distributed denial of service attacks (DDoS)
- changes in IP forwarding
 - failures/mistakes in routers, changes in BGP policy
- ...

Network anomaly diagnosis^[Lakhina04]

Three components

- **Detection**
 - *when* an network anomaly is happen?
- **Identification**
 - *what* is causing the network anomaly?
- **Quantification**
 - *How* critical is the network anomaly?

Network anomaly diagnosis

Three components

- **Detection**

- *when* an network anomaly is happen?

**our
focus**

- **Identification**

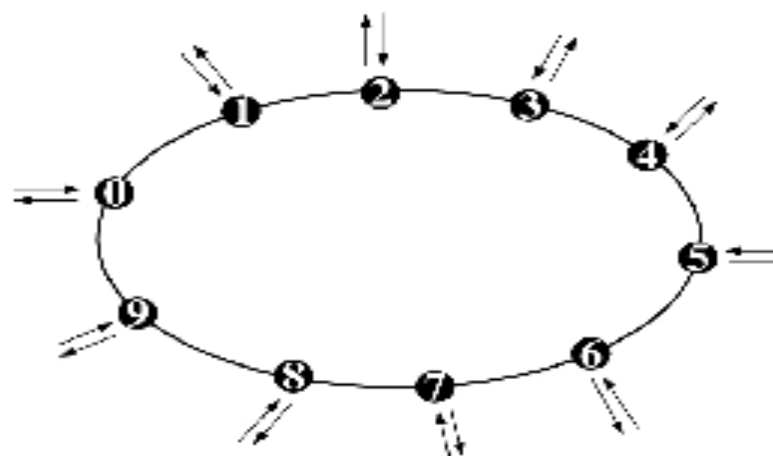
- *what* is causing the network anomaly?

- **Quantification**

- *How* critical is the network anomaly?

Network anomaly detection

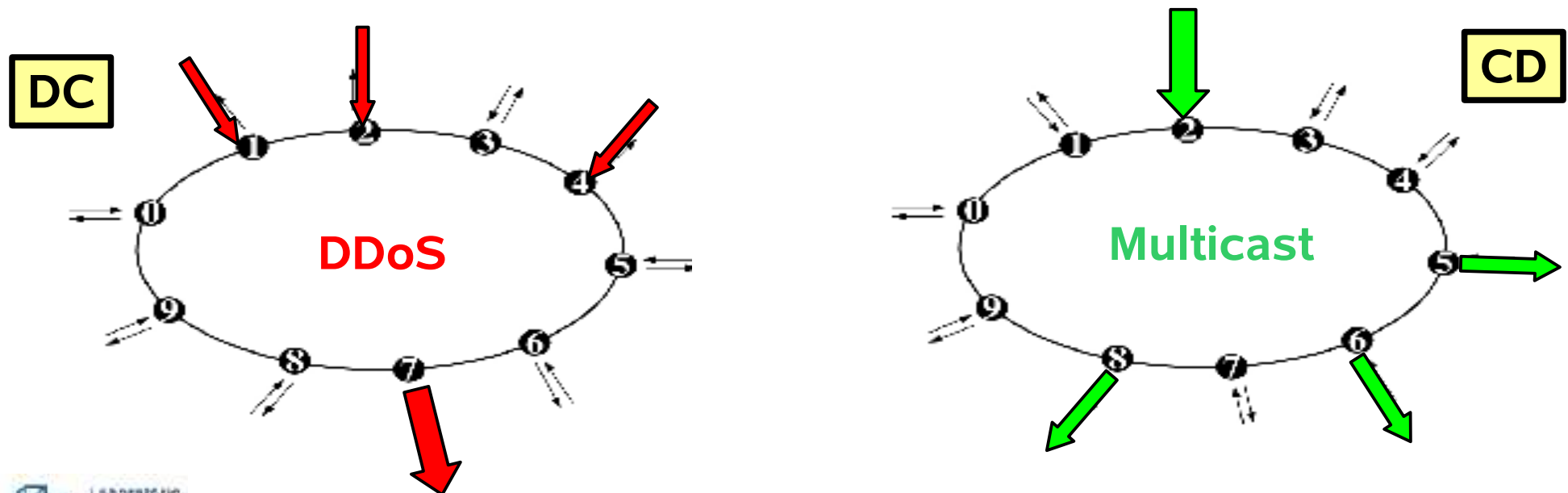
- **Using Shannon entropy** [Lakhina 05]
 - characterization of traffic pattern within an AS
 - concentrated (C) ou dispersed (D)
 - with respect to the traffic origin and destination



Dispersed-Dispersed (DD)

Network anomaly detection

- **Using Shannon entropy** [Lakhina 05]
 - characterization of traffic pattern within an AS
 - concentrated (C) ou dispersed (D)
 - with respect to the traffic origin and destination



Shannon entropy

$$H_S = - \sum_{i=1}^N p_i \log_2 p_i$$

- **Minimal entropy**

$$H_S^{min} = 0$$

- maximum concentration

$$p = \{0, 0, 0, 1\}, N = 4$$

- **Maximum entropy**

$$H_S^{max} = \log_2 N$$

- maximum dispersion $p = \{0.25, 0.25, 0.25, 0.25\}, N = 4$

Using Tsallis entropy

- **Nonextensive (Tsallis) entropy** [Tsallis 88]
 - conceived in physics for systems presenting **long-range dependency** (LRD)
 - Tsallis entropy is a **one-parameter generalization** of Shannon entropy

$$H_q = \frac{1 - \sum_{i=1}^N p_i^q}{q-1}$$

$q < 1$: events with **lower** probability contribute more to the entropy value

$q \rightarrow 1$: **Shannon entropy**

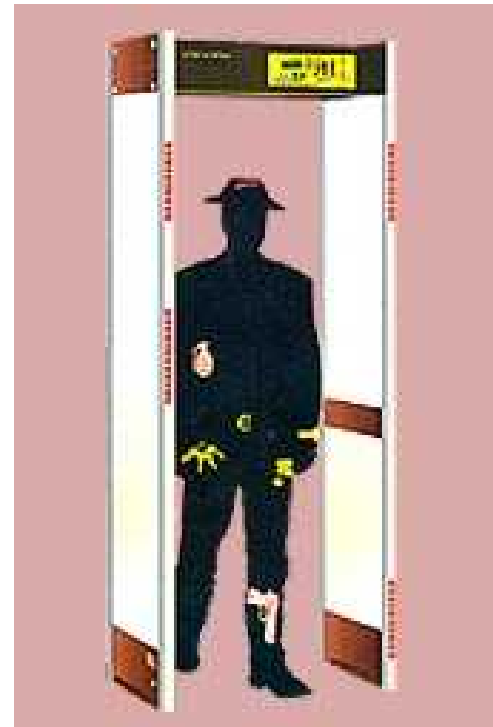
$q > 1$: events with **higher** probability contribute more to the entropy value



Detection sensitivity

- **Variation of the q parameter**
 - fine-tuning the sensitivity of the detection system

**Just like a metal
detector door frame!**



Some preliminary results

- **Experimental data**

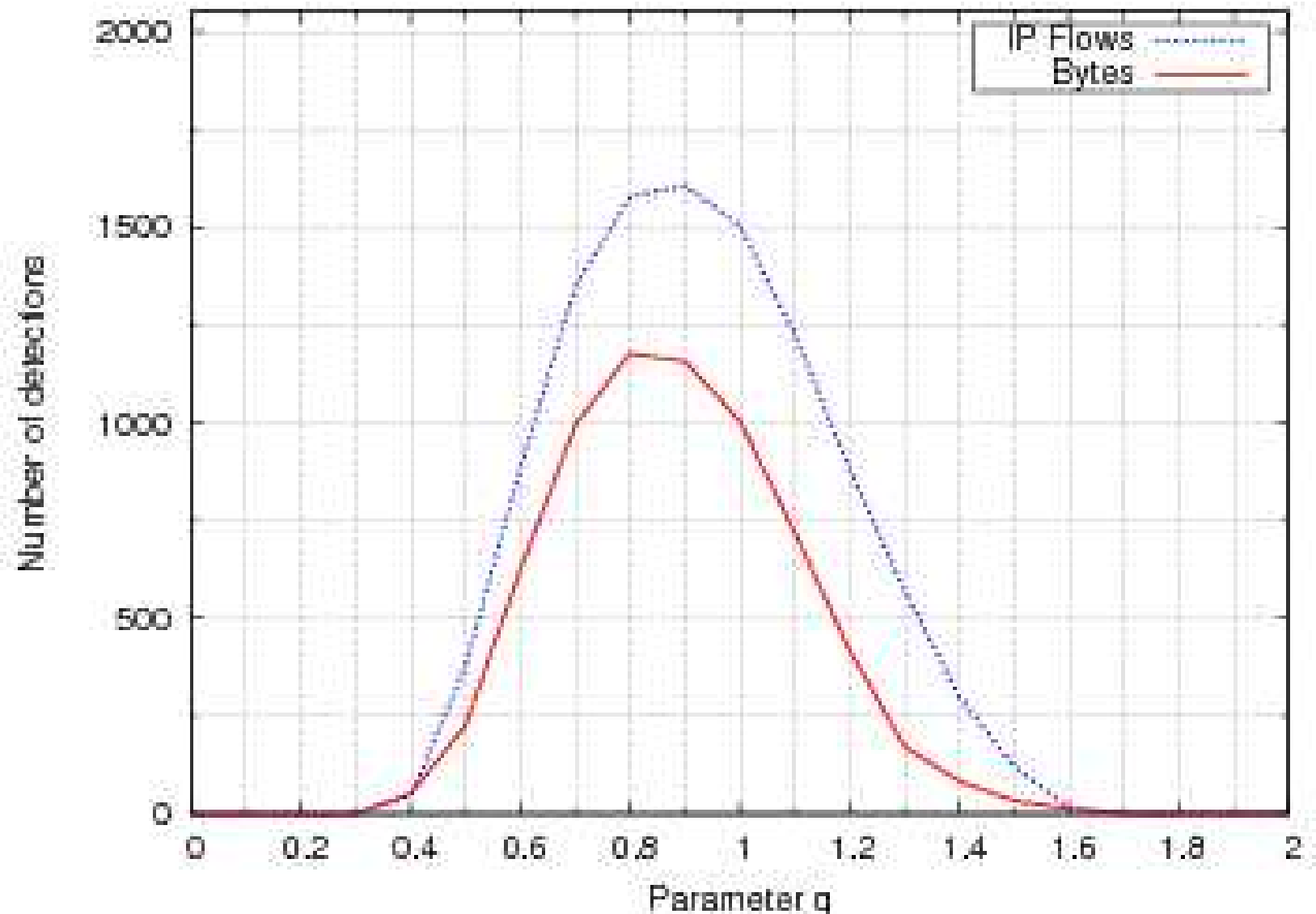
- **Abilene traces (Internet2) - USA**

- metrics: IP flows and bytes
 - one week: April 7-11 2003 (Monday to Sunday)
 - traffic volume in 5-min intervals (2016 time intervals) of the metric distribution among origin and destination nodes

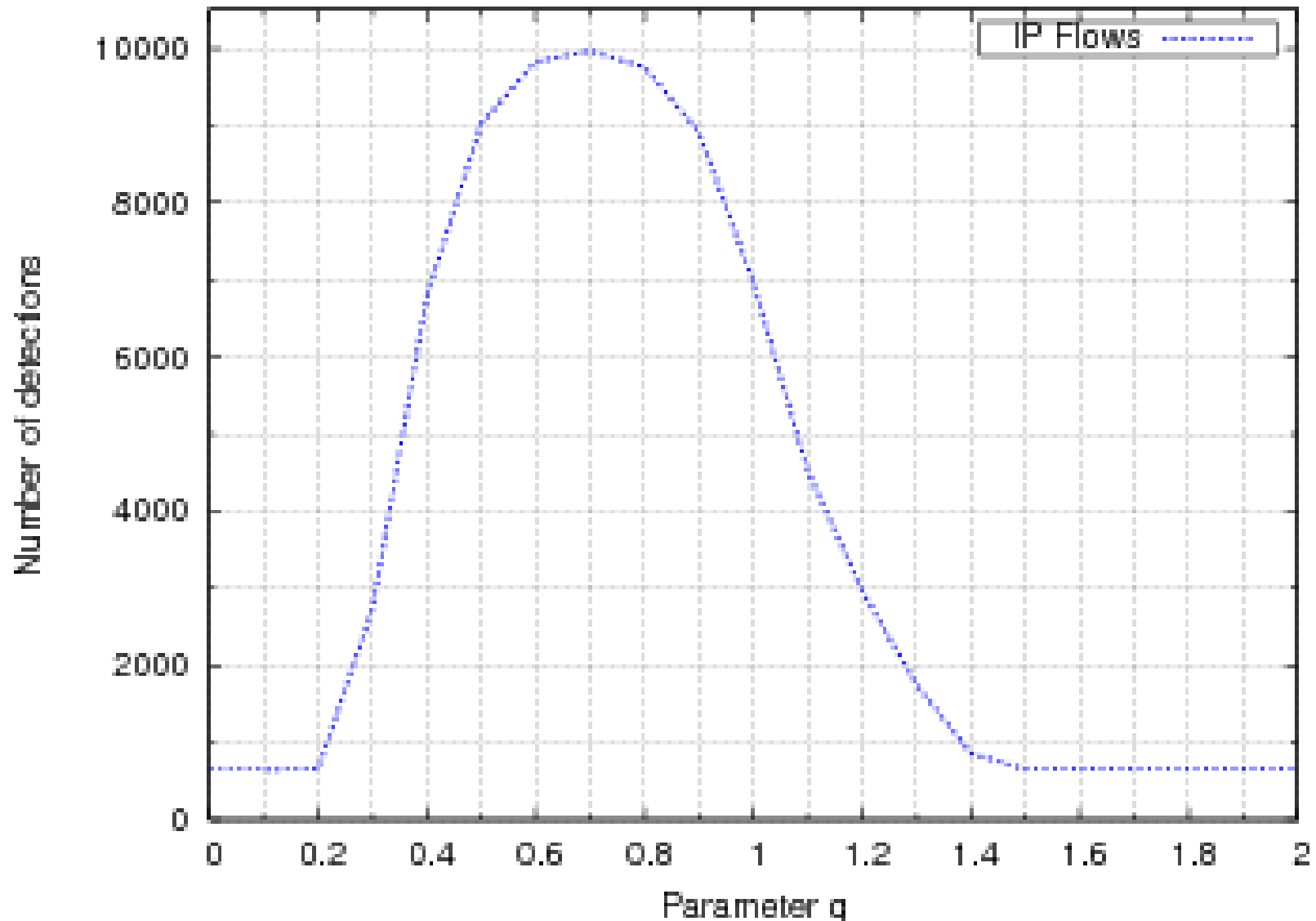
- **direct comparison with the use of Shannon entropy**

- subset of the data used by [Lakhina05]
 - thanks to A. Lakhina, M. Crovella, and C. Diot for sharing their experimental data for a direct comparison
 - further validation with Géant data
 - thanks to the Totem Project (Steve Uhlig)

Number of CC detections (Abilene)



Number of CC detections (Géant)



Performance comparison

Table 1: Performance comparison considering the number of detected anomalies.

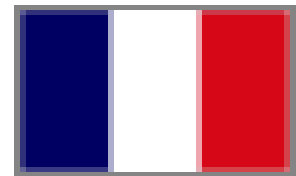
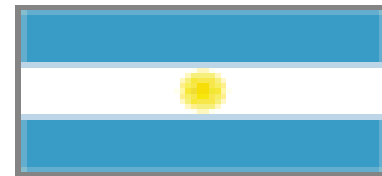
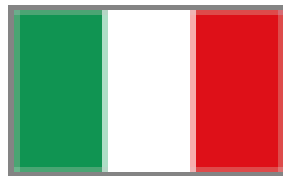
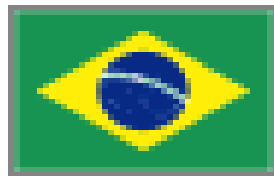
Dataset	Metric	Traffic Pattern	Number of detected anomalies		Improvement (%)
			H_S	$H_{q_{optimal}}$	
Abilene	IP Flows	CC	1505	1606	6%
	IP Flows	CD	206	444	115%
	IP Flows	DC	201	450	124%
Abilene	Bytes	CC	1005	1177	17%
	Bytes	CD	186	226	22%
	Bytes	DC	539	793	47%
Géant	IP Flows	CC	6970	9964	43%
	IP Flows	CD	1838	5364	192%
	IP Flows	DC	1090	1582	45%

Final remarks

- **Fine-tuning the detection sensitivity**
- **Ongoing work**
 - evaluation of false positives
 - a more sensitive system provides more detections, but how many of these are false positives?
 - sensitivity analysis and modeling
 - promising results with $\gamma(q) = \frac{e^q}{q}$
 - further uses for Tsallis entropy in measurements?
 - *potentially* any useful application using Shannon entropy may have better results using Tsallis entropy

Et enfin...

Bienvenus à **la cour des grands...**



1958
1962
1970
1994
2002

1934
1938
1982
2006

1954
1974
1990

1978
1986

1998



1950
1998

1970
1994

1966
1982
1986
2002

1930
1990

2006

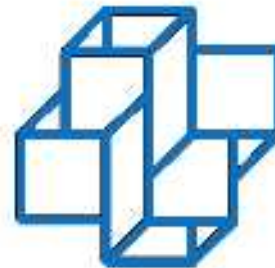
pour ceux qui savent

gagner mais aussi perdre des finales



Merci! Obrigado!

- **Questions?**
- **Contact info:**
 - Artur Ziviani (LNCC/MCT)
 - ziviani@lncc.br
 - <http://www.lncc.br/~ziviani>



Laboratório
Nacional de
Computação
Científica