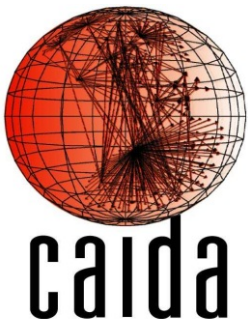# Entropy in IP Darkspace Data

Tanja Zseby

Cooperative Association for Internet Data Analysis (CAIDA)
and
Fraunhofer Institute for Open Communication Systems (FOKUS)
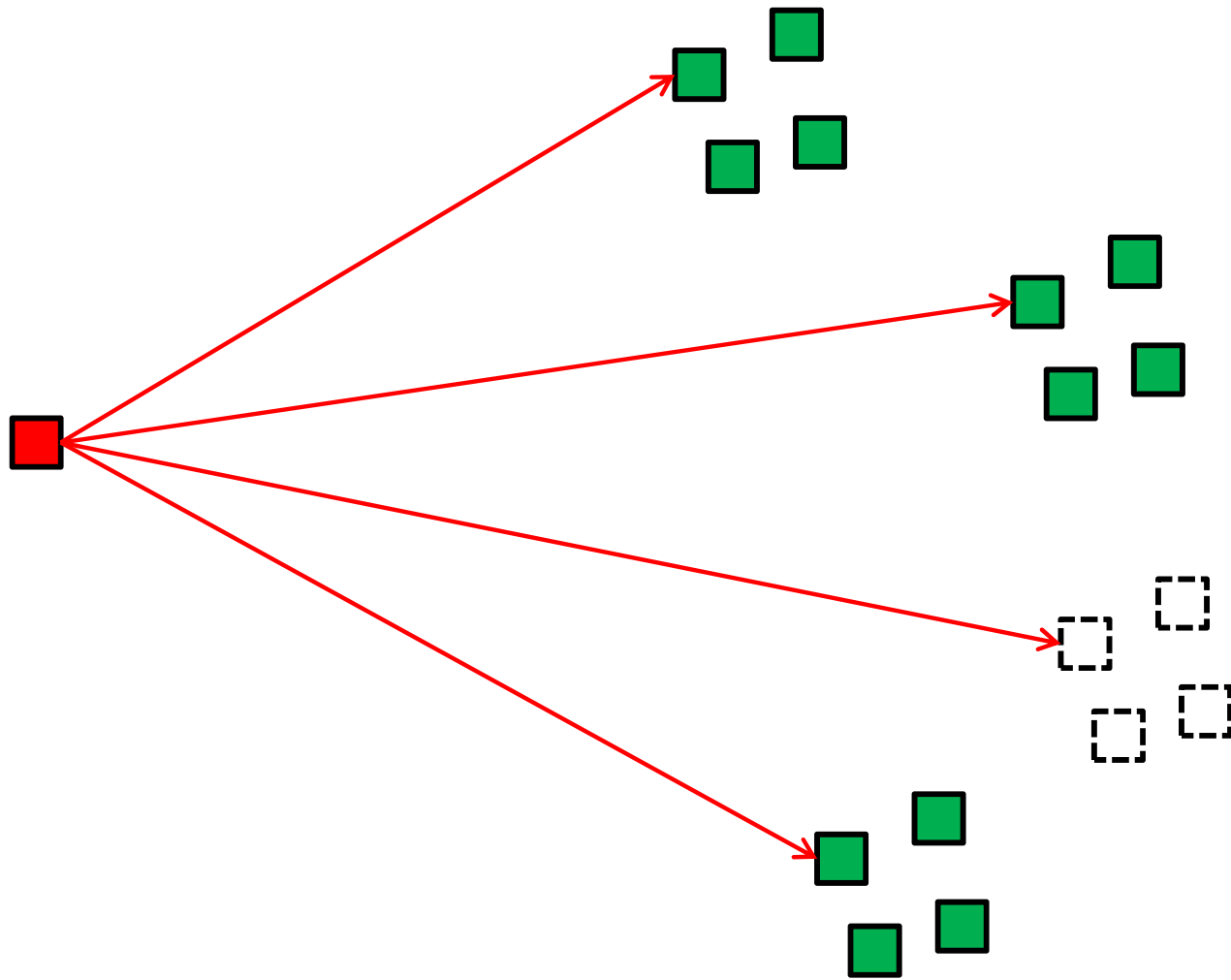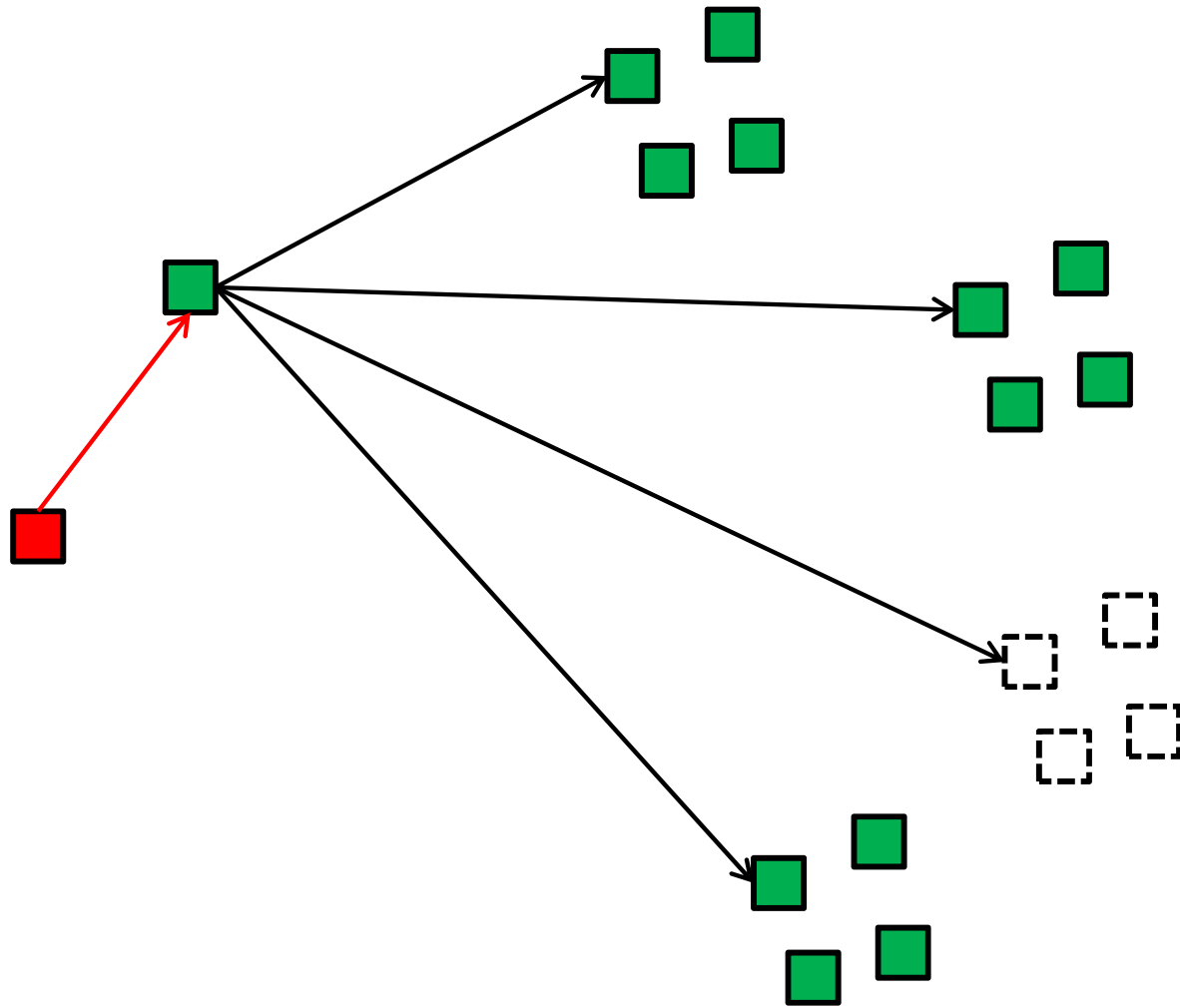
CERT FloCon, January 2012

IP Darkspace

- Global routable IP address space
  - announced by routing
  - but no hosts attached
  - ➔ all traffic destined to darkspace is unsolicited
- UCSD telescope
  - /8 darkspace
  - Used for different analysis (security, outages, etc.)
- Other IP darkspace monitors:
  - Internet Motion Sensor, Team cymru Darknet Project, iSink, …

Scanning

Backscatter

Analysis of Darkspace Data

- **Detection of incidents**
  - Scanning activities
  - Backscatter
  - Misconfigurations
  - Network outages

➔ Analysis (patterns, scope,..)

➔ Early warning

➔ „Cleaning up" address space

DSA related work

- ## General Analysis Techniques
  - Brownlee. One-way Traffic Monitoring with iatmon. To appear at PAM 2012
  - Ahmed et al. Characterising anomalous events using change - point correlation on unsolicited network traffic. In Identity and Privacy in the Internet Age, 2009.

- ## Security and Misconfigurations
  - Wustrow et al. Internet background radiation revisited. IMC 2010
  - Aben. Conficker. ISOI 2009
  - Moore et al.Code-Red: a case study on the spread and victims of an Internet worm. IMW 2002

- ## Network Outages
  - Dainotti et al. Analysis of Country-wide Internet Outages Caused by Censorship, IMC 2011

- ## Darkspace Construction
  - Janies, Collins, Darkspace Construction and Maintenance, FloCon 2011

- ## IPv6 Darkspace
  - Huston: IPv6 Background Radiation, NANOG50, 2010
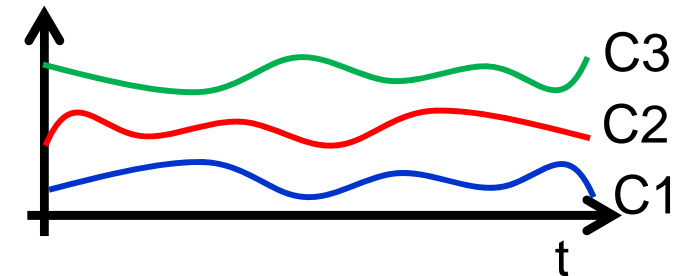 - Ford, et al. Initial Results from an IPv6 Darknet, 2006
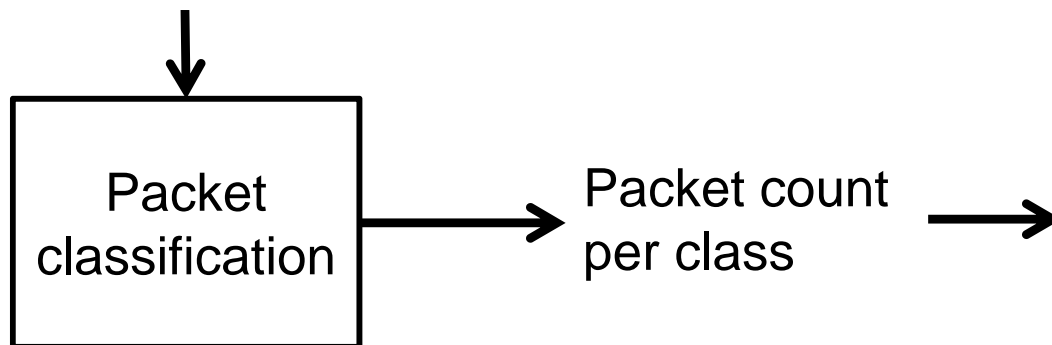
…and others.

Metrics and Techniques

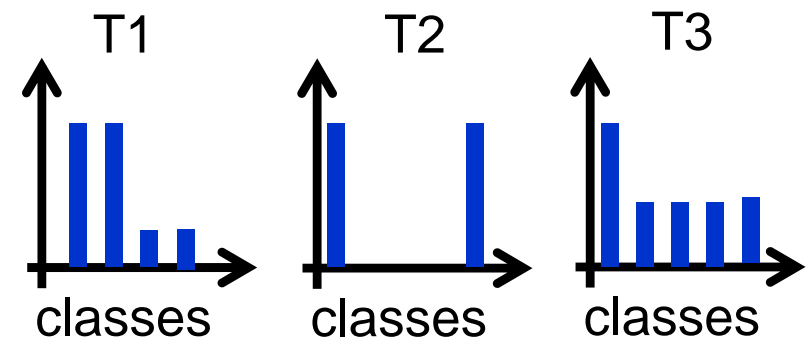Classification rules
(feature combinations)

Packet
classification

Packet count
per class
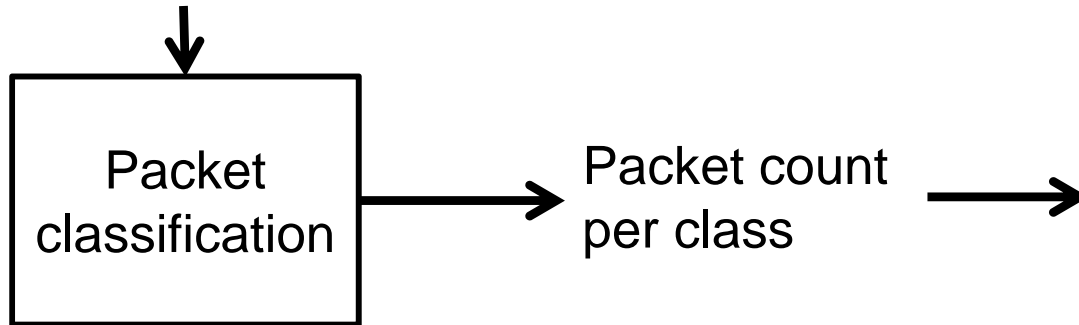
Time series of packet counts
for selected feature combinations

C3
C2
C1
t

Classification rules
(selected features)

Packet
classification

Packet count
per class

Distributions for selected
features

T1   T2   T3

classes   classes   classes

Example Metrics

- ## Time series of packet counts
  - Overall packet count
  - Packets to a specific port
  - Packets with specific TCP flags

- ## Source groups based on source behavior
  - Packet features (e.g. SYNs to specific port)
  - Inter Arrival Times (IATs)

- ## Distributions
  - IP addresses, port numbers

Challenges

- High amount of data
  - Many repetitions/boring events (TCP-SYNs,…)
  - whole distributions ➔ huge amount of data

- Selection of suitable classification rules
  - Separate known events from new/interesting packets
  - Feature selection difficult
  - Features of interest may change
  - High analysis effort
  - Detection of different events requires various metrics

Problem Statement

- **Goal:** detect and classify „events of interest"
  - New vulnerabilities (increased scanning)
  - New victims of attacks (increased backscatter)
  - Misconfigurations
  - Network outages
- Ideal: Comprehensive metric
  - capture **all** events of interest
- Conditions
  - Keep storage requirements low

Characteristics of DS Events

- Hostscans (new vulnerability)
  - Many new sources (attackers) send to a specific destination port

- Backscatter (from DoS attacks with spoofed addresses)
  - Several sources (victims) send a lot of data to many destination addresses using a specific source port

- Misconfiguration (configuration of wrong destination IP)
  - Several sources send to a specific destination IP and specific destination port

- Outages
  - Source IPs from outage region are missing ➔ fewer source IPs

- DDoS (to a destination IP in darkspace)
  - Many new sources (bots or spoofed) send to a specific destination IP and specific destination port

- Portscan
  - One or several hosts send to a specific destination IP and many destination port

Expected Effects on Distributions

	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random (attackers)	specific (victims)	specific	specific (some missing)	random (attackers)	specific (attackers)
dIP	random	random	specific	depends	specific	specific
sPort	random*	specific	depends	depends	random*	random*
dPort	specific	random*	specifc	depends	specific	random

Distinction of specific/random ➔ **entropy** !

*assuming random sPort selection by attack tools

Sample Entropy

*"You should call it **entropy**, […]*
…no one really knows what entropy really is, so in a debate you will always
have the advantage."
John von Neumann's suggestion to Claude Shannon according to Max Jammer "*Dictionary of*
the History of Ideas: Entropy"

Sample Entropy

Definition from [LaCD05]:

Histogram $\qquad X = \{n_i, i = 1, ..., N\}$

Total number of observations $\quad S = \sum_{i=1}^{N} n_i$

$$H(X) = -\sum_{i=1}^{N} \left(\frac{n_i}{S}\right) \log_2 \left(\frac{n_i}{S}\right)$$

[LaCD05] Lakhina, Crovella, Diot: Mining Anomalies Using Traffic Feature Distributions. *SIGCOMM* 2005

Related Work

Entropy-based anomaly detection:

- Lee/Xiang 2001
  - Information Theoretic Measures for Anomaly Detection

- Feinstein/Schnackenberg 2003
  - Detection of DDoS attacks based on source IP entropy

- Lakhina et al.2005
  - Detection of scanning, DDoS, outages based on combinations of entropy from addresses and ports

Entropy Example

freq
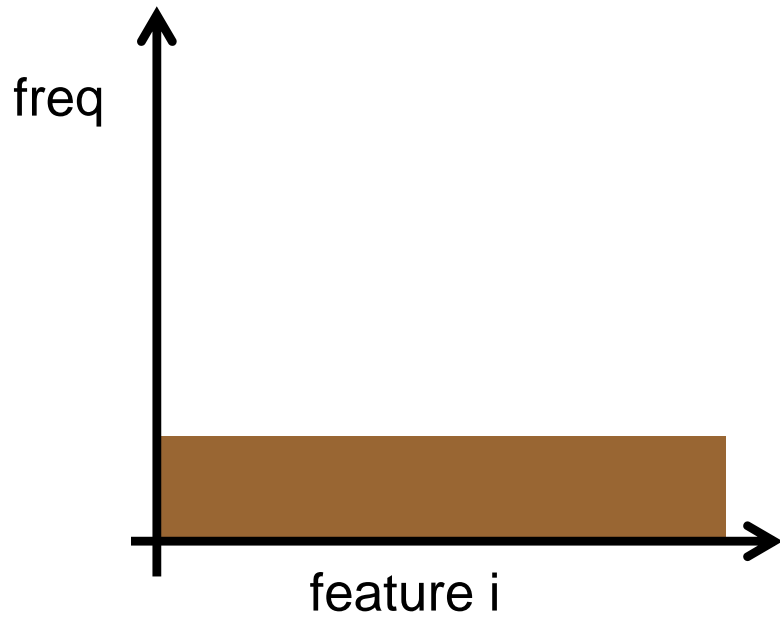
Each packet different

➔ Entropy= max
$H(X) = \log_2 N$

feature i

H(x)=max

freq
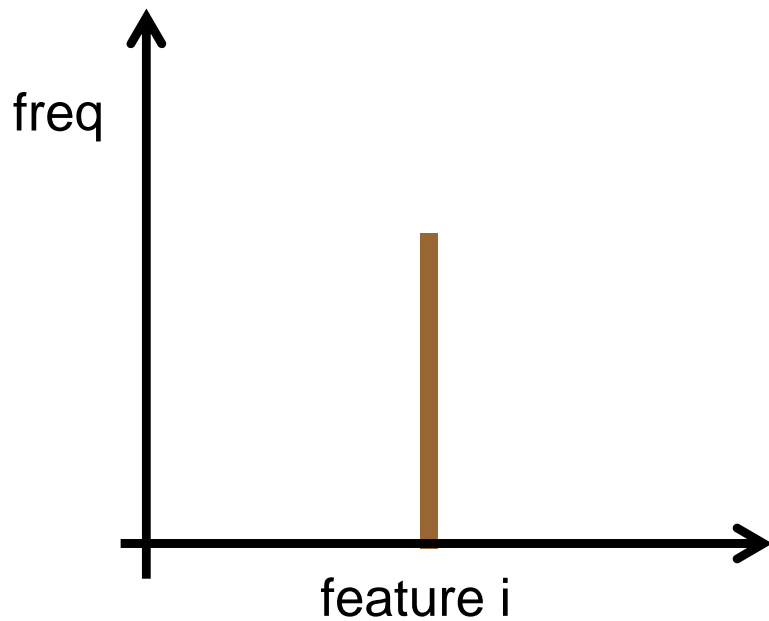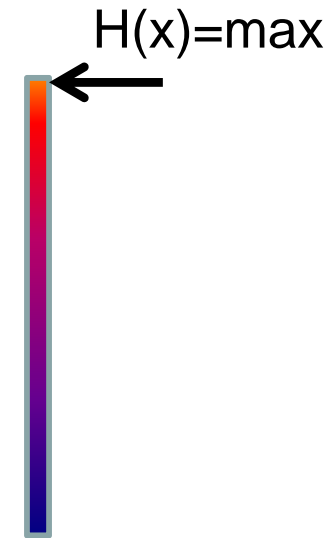
All packets equal

➔ Entropy = min
$H(X) = 0$

feature i

H(x)=min
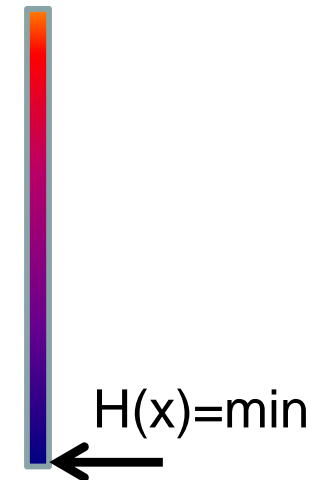
Expected Entropy Patterns

	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random ↑	specific ↓	specific ↓	specific ↓	random ↑	specific ↓
dIP	random** ↑	random** ↑	specific ↓	depends	specific ↓	specific ↓
sPort	random* ↑	specific ↓	depends	depends	random* ↑	random* ↑
dPort	specific ↓	random* ↑	specific ↓	depends	specific ↓	random ↑

*assuming random sPort selection by attack tools

**dIP has already high entropy in "normal" operation

Analysis

- Time periods
  - Nov 2008
  - Jan/Feb 2011
  - Oct 2011

- Calculation of Sample Entropy
  - sIP, dIP, sPort, dPort
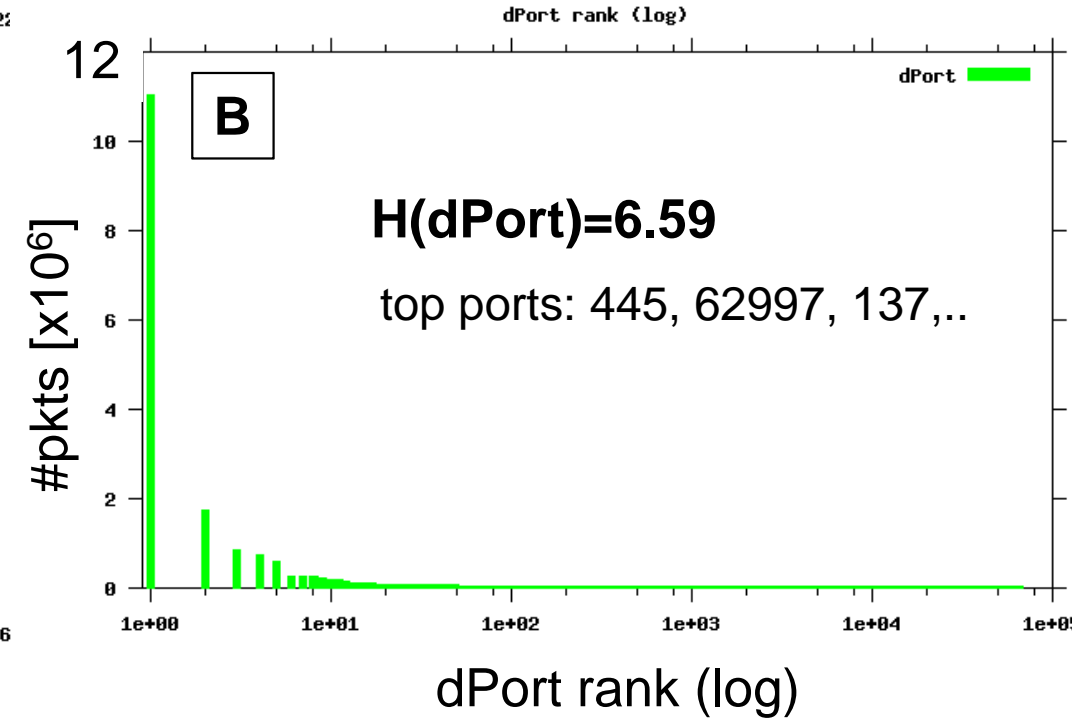 - Time intervals: 1 hour
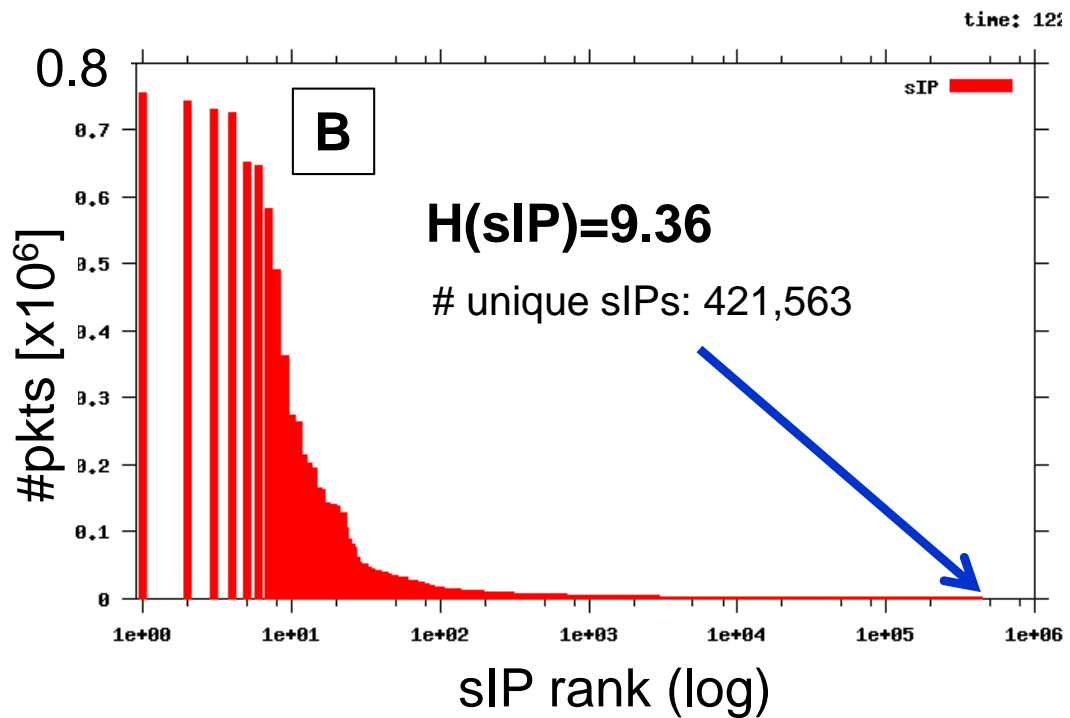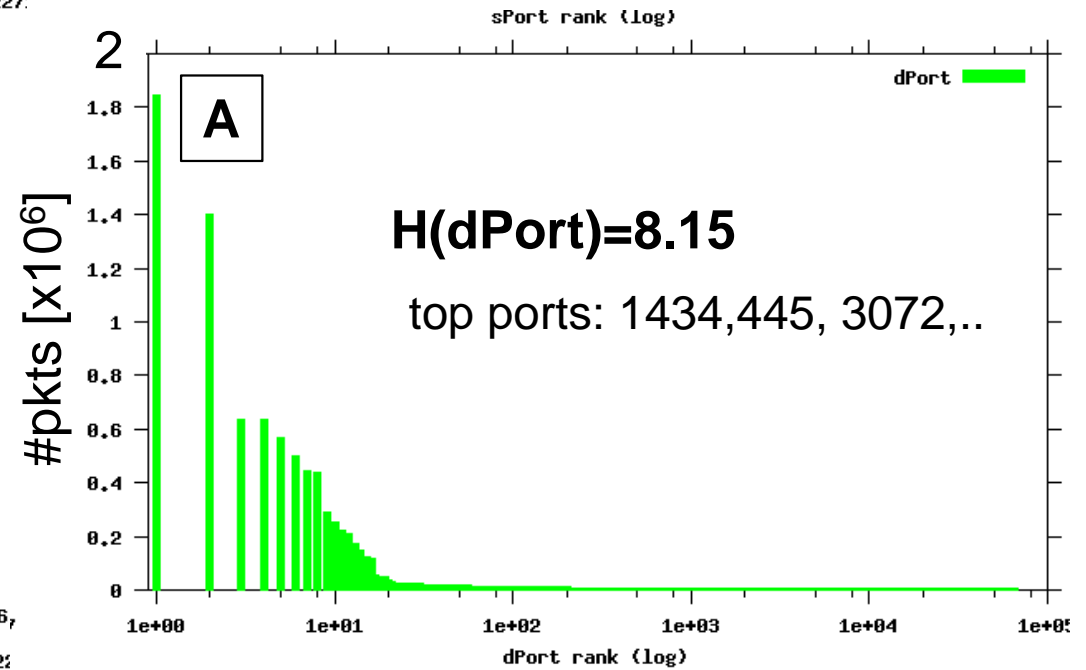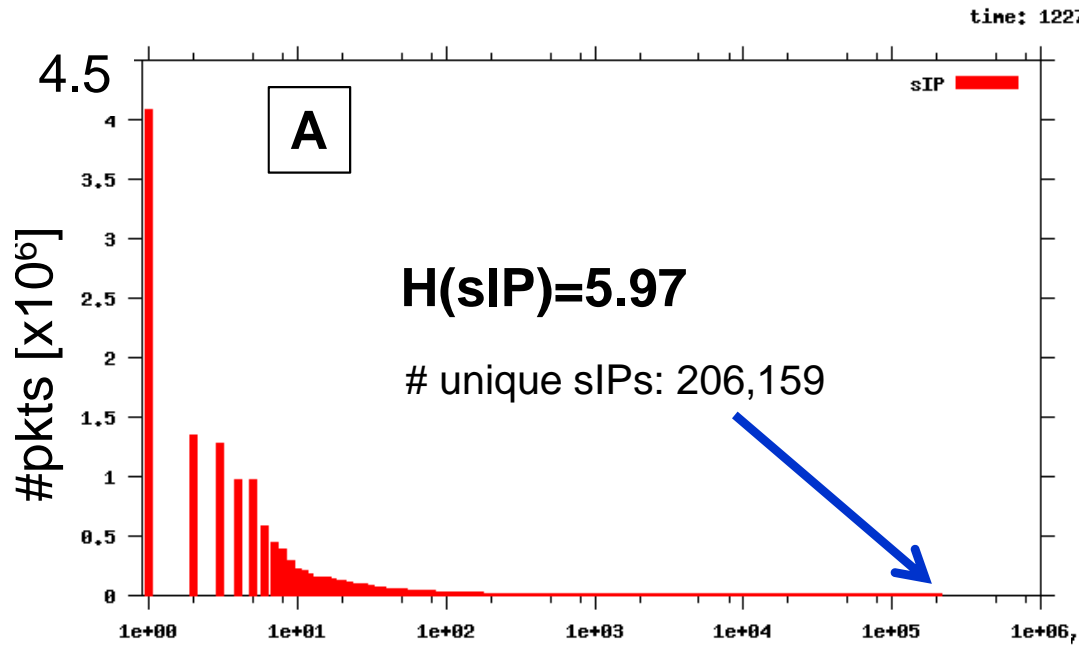
- Tools: SiLK, R

NOV 2008

Nov 2008
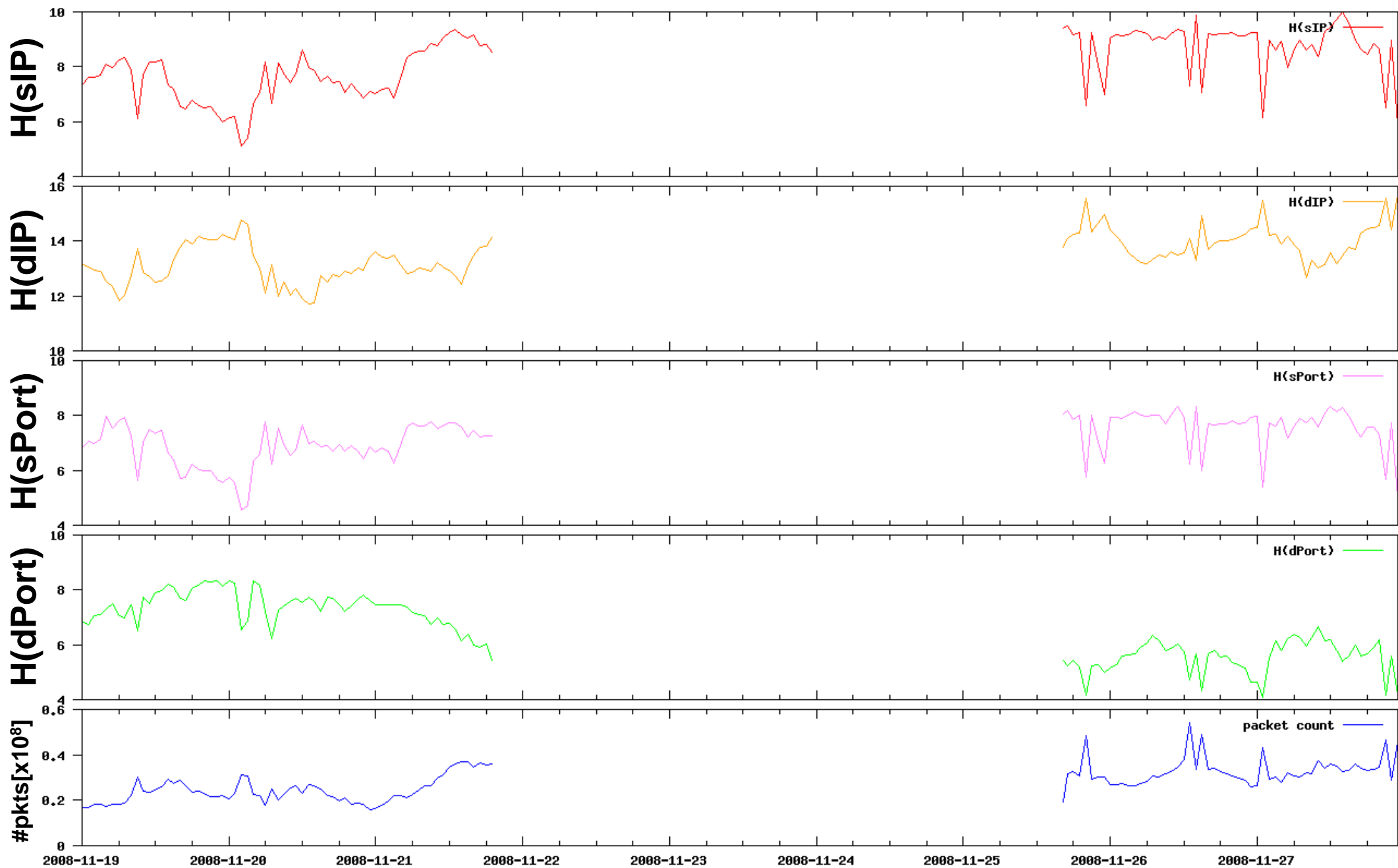
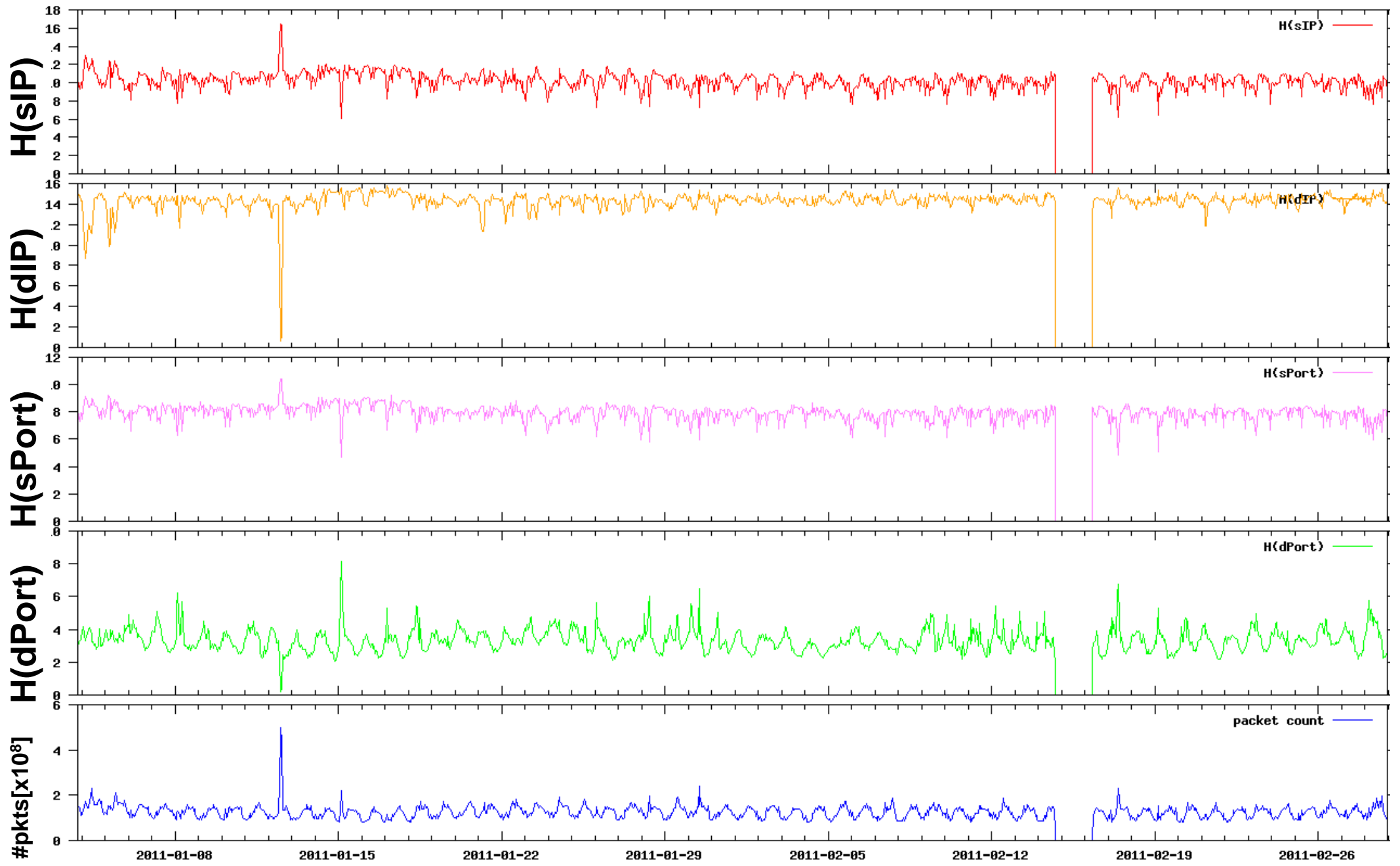Nov 2008

Classification of Event B

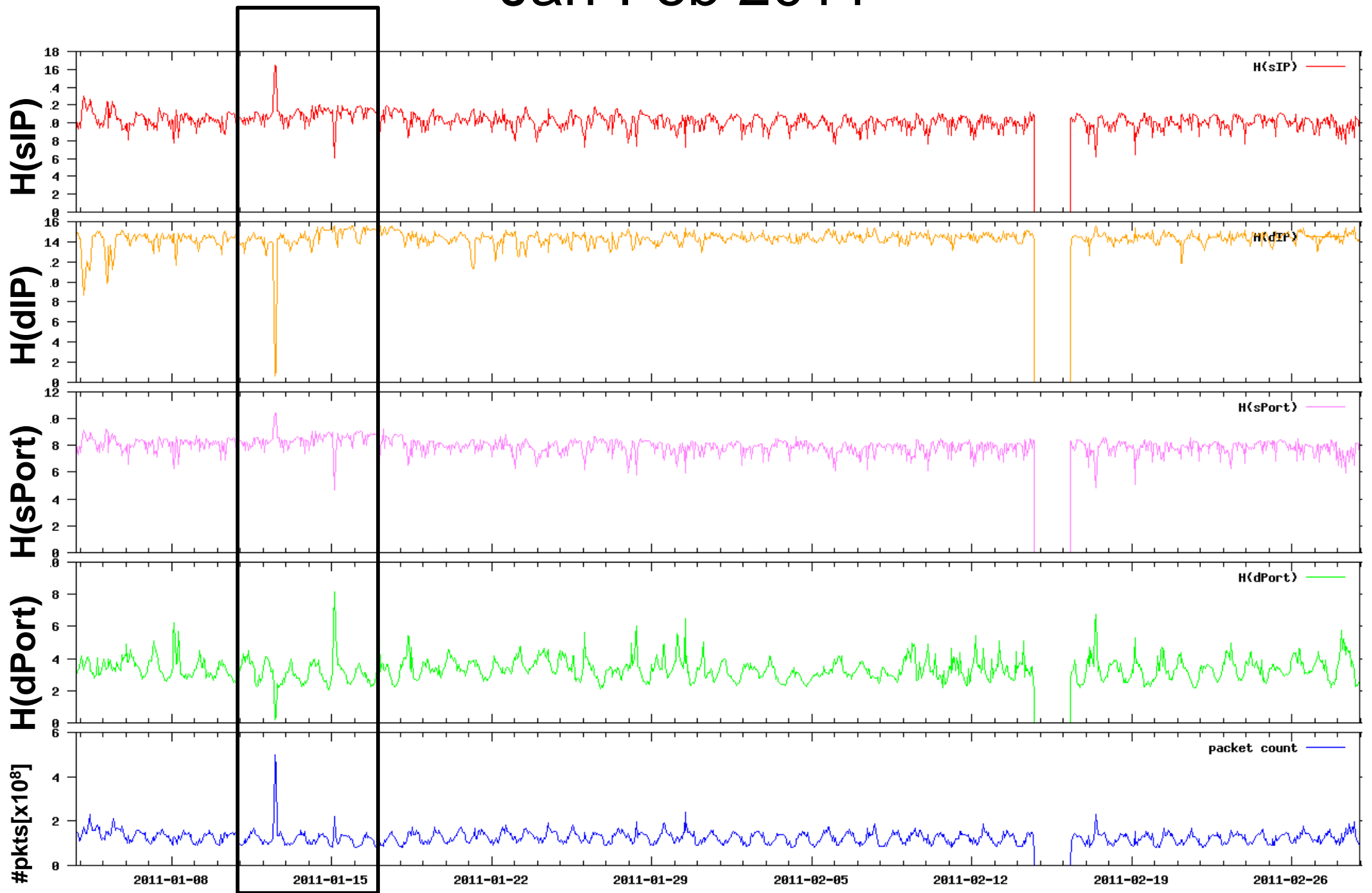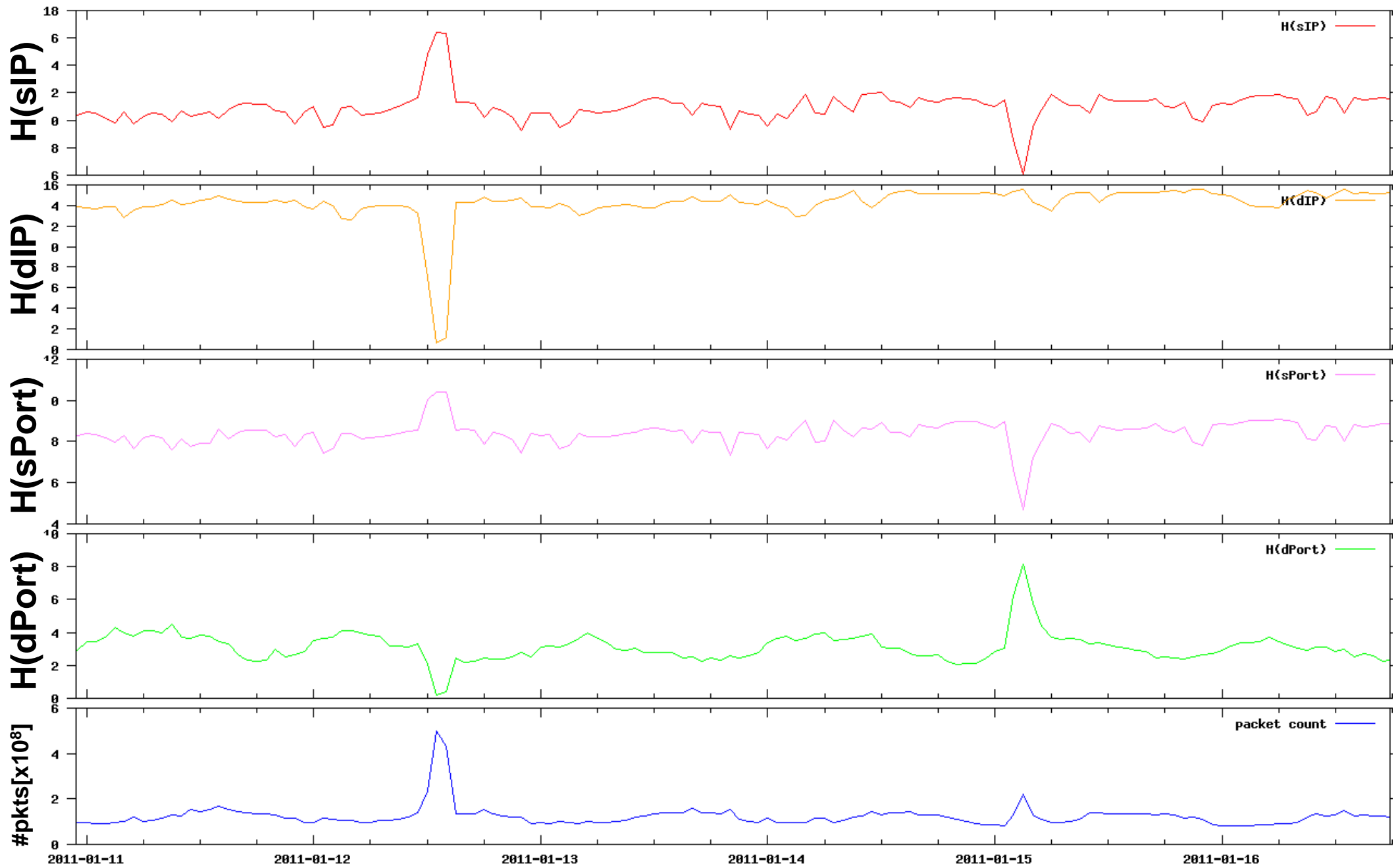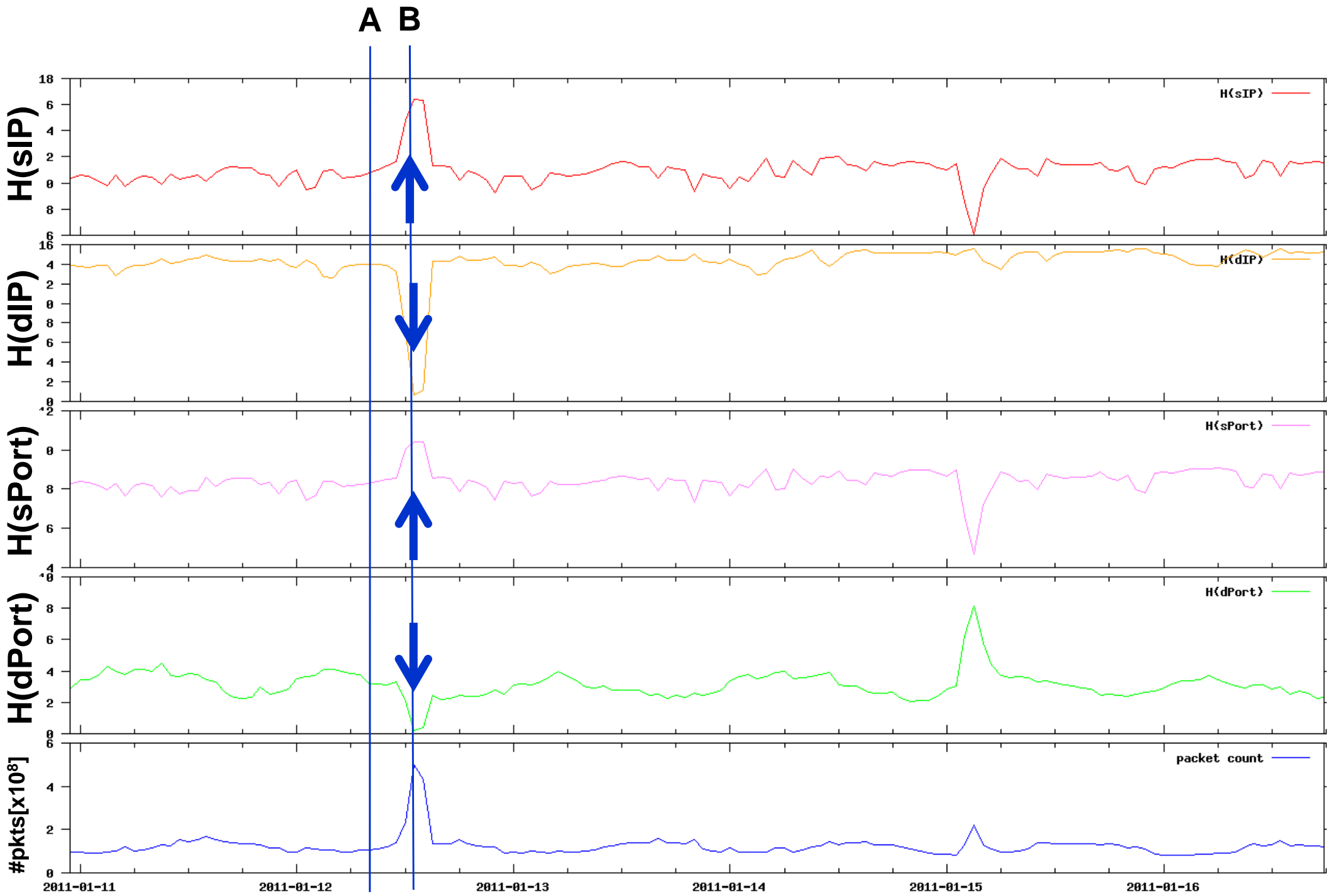|	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random ↑	specific ↓	specific ↓	specific ↓	random ↑	specific ↓
dIP	random** ↑	random** ↑	specific ↓	depends	specific ↓	specific ↓
sPort	random* ↑	specific ↓	depends	depends	random* ↑	random* ↑
dPort	specific ↓	random* ↑	specific ↓	depends	specific ↓	random ↑

Distributions: sIP, dPort
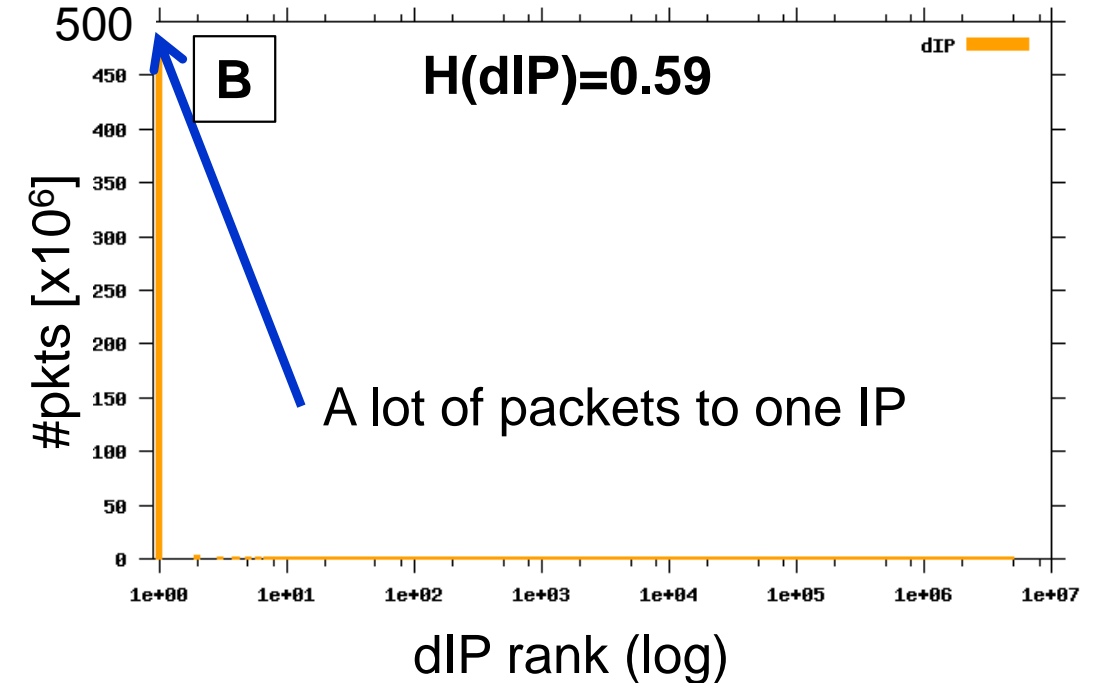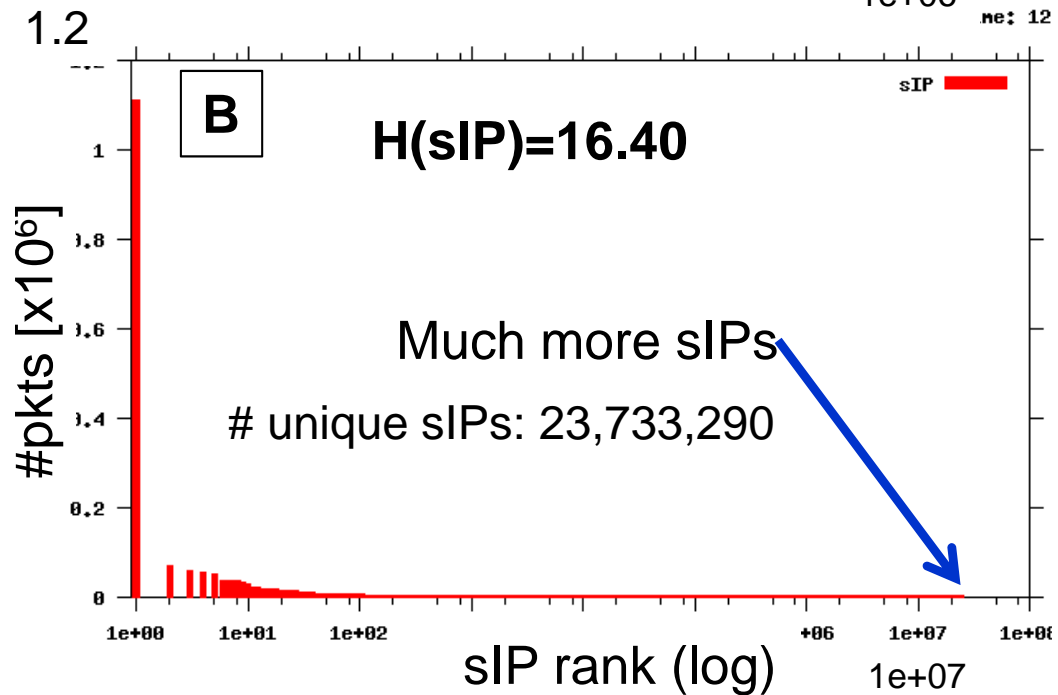
JAN/FEB 2011

Jan-Feb 2011

Jan-Feb 2011

Jan 2011

Classification of Event B

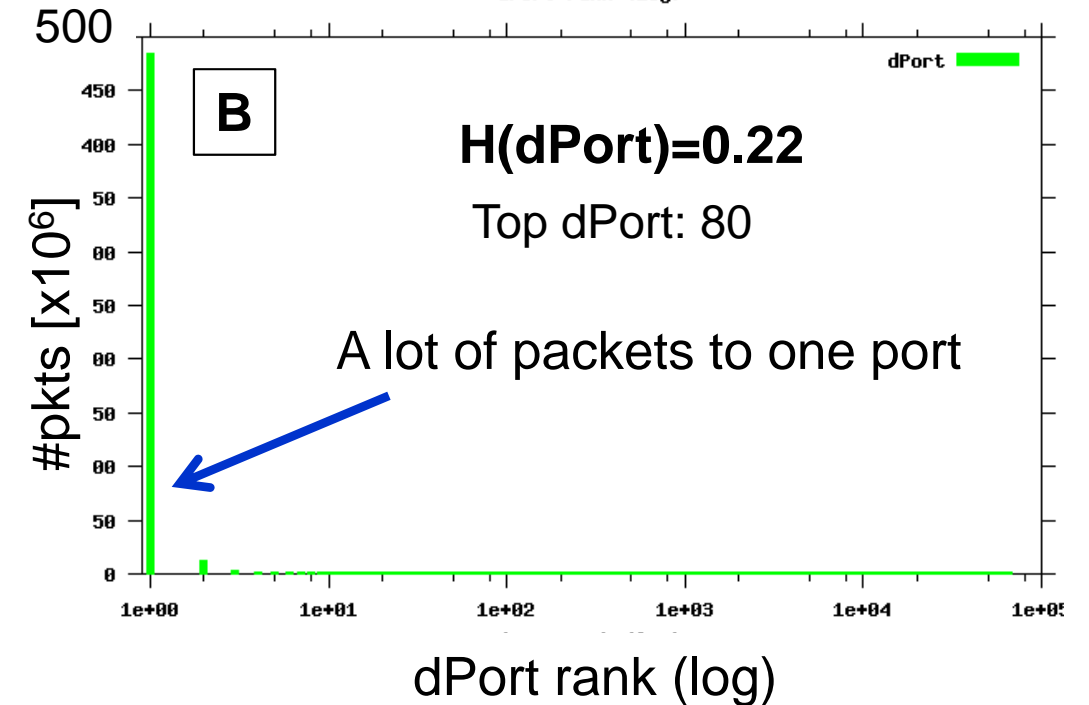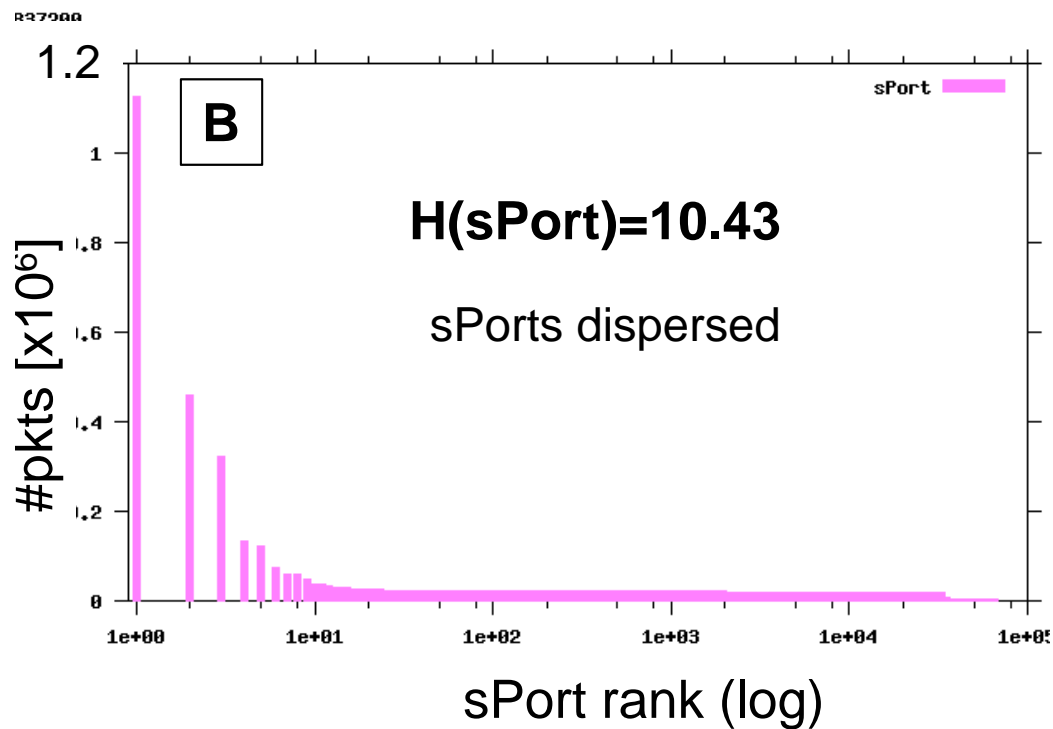|	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random ↑	specific ↓	specific ↓	specific ↓	random ↑	specific ↓
dIP	random** ↑	random** ↑	specific ↓	depends	specific ↓	specific ↓
sPort	random* ↑	specific ↓	depends	depends	random* ↑	random* ↑
dPort	specific ↓	random* ↑	specific ↓	depends	specific ↓	random ↑

Distributions: sIP, dIP

Distributions: sPort, dPort

A H(sPort)= 8.42

A H(dPort)=3.19

Top dPort: 445

B H(sPort)=10.43

sPorts dispersed

B H(dPort)=0.22

Top dPort: 80
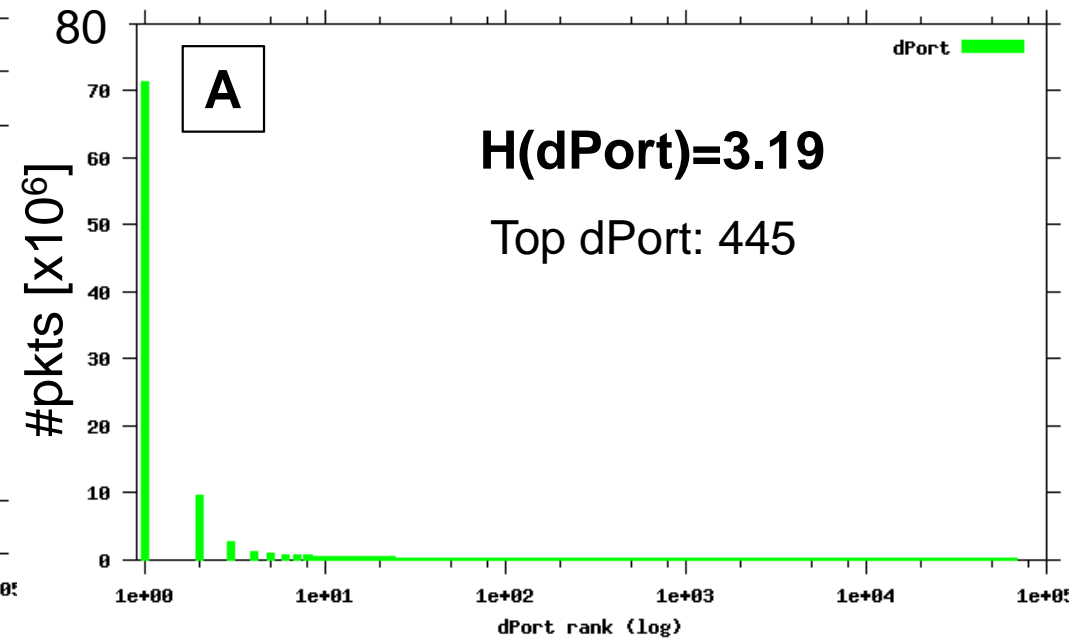
A lot of packets to one port

sPort rank (log)

dPort rank (log)

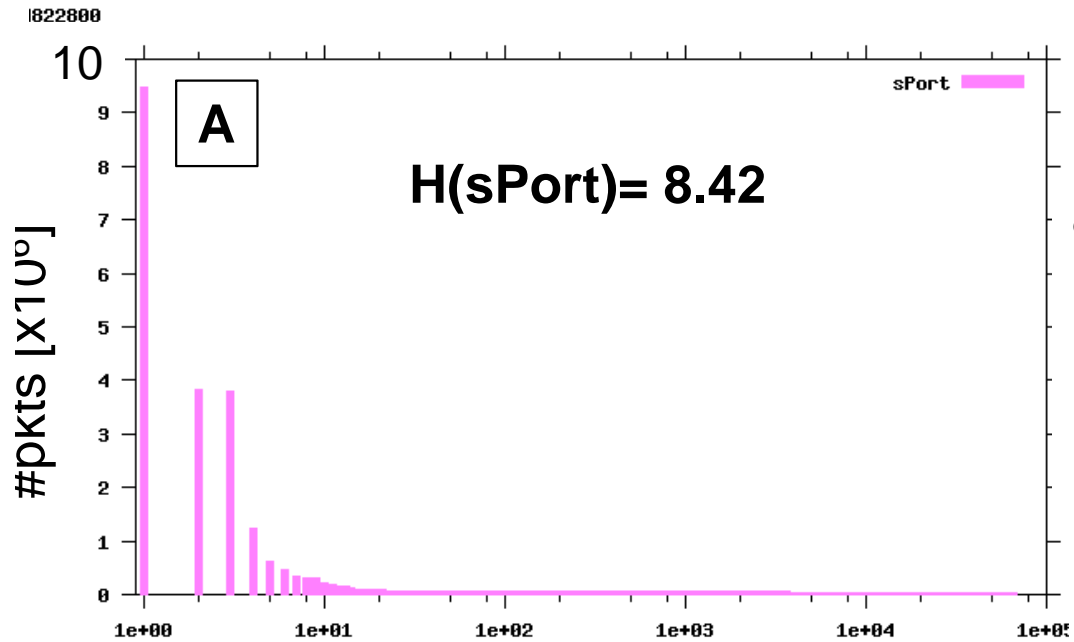# Classification of Event C

	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random ↑	specific ↓	specific ↓	specific ↓	random ↑	specific ↓
dIP	random** ↑	random** ↑	specific ↓	depends	specific ↕	specific ↓
sPort	random* ↑	specific ↓	depends	depends	random* ↑	random* ↑
dPort	specific ↕	random* ↑	specific ↓	depends	specific ↓	random ↑

Distributions: sIP, dIP
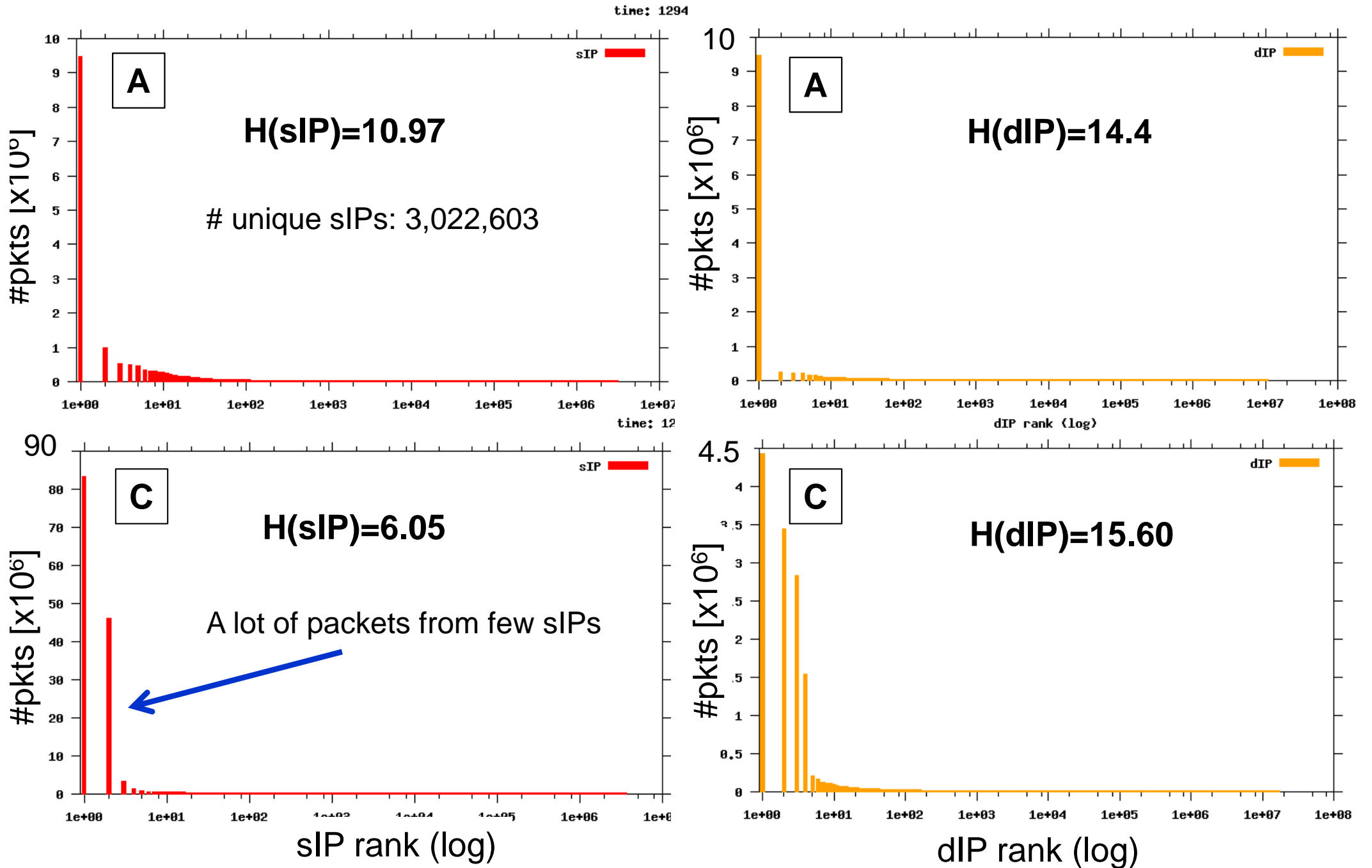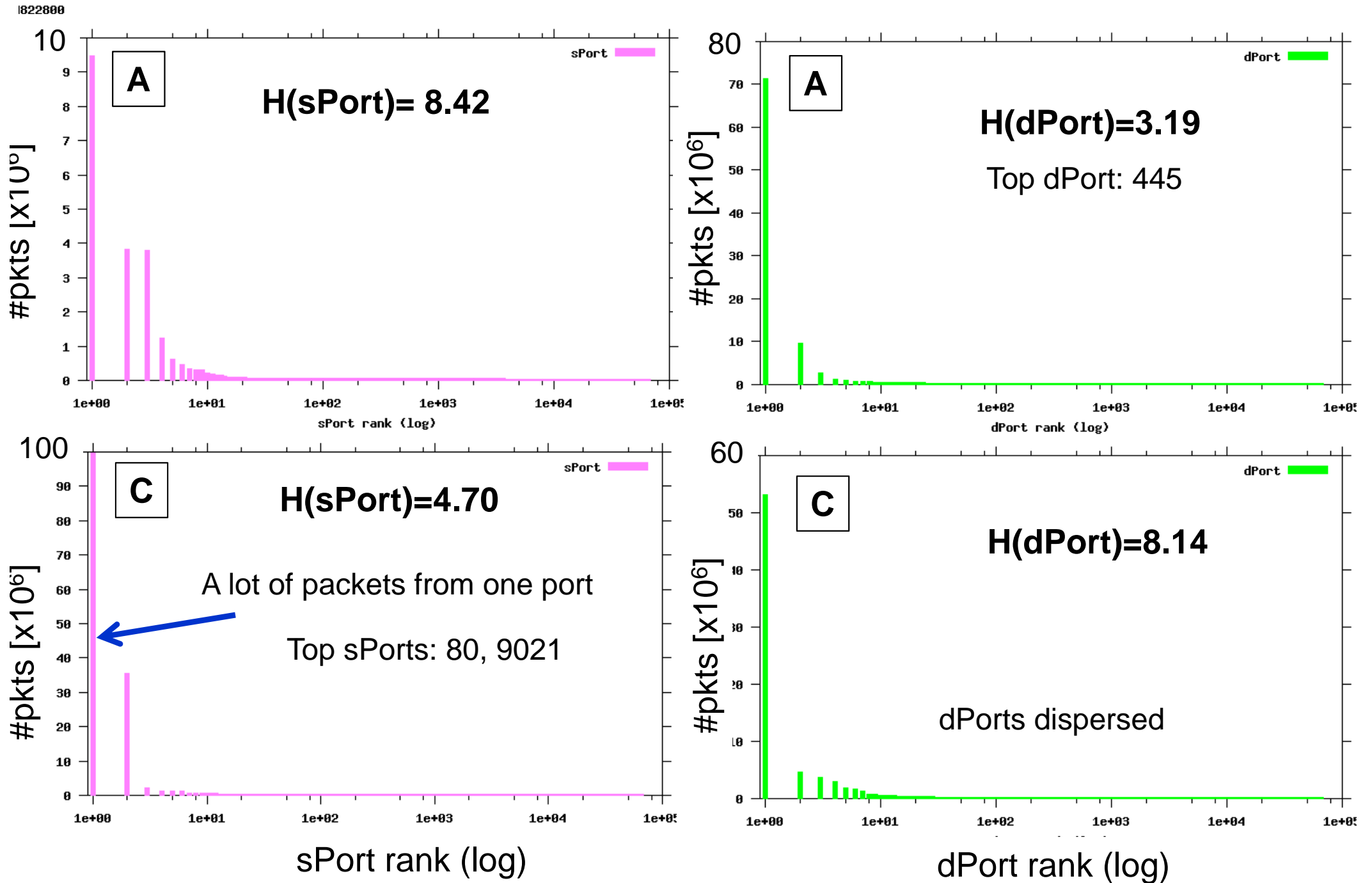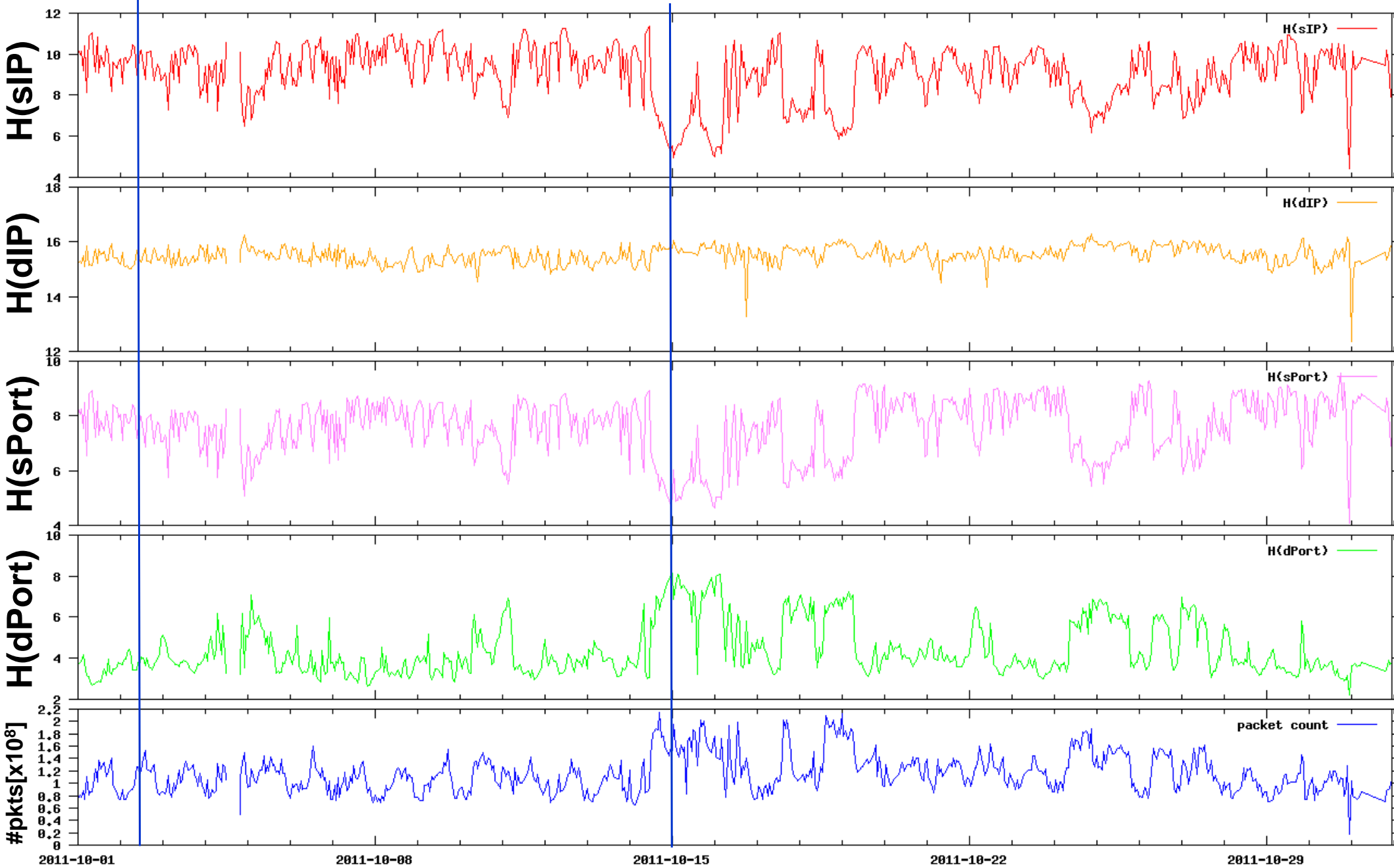
Distributions: sPort, dPort

H(sPort)= 8.42
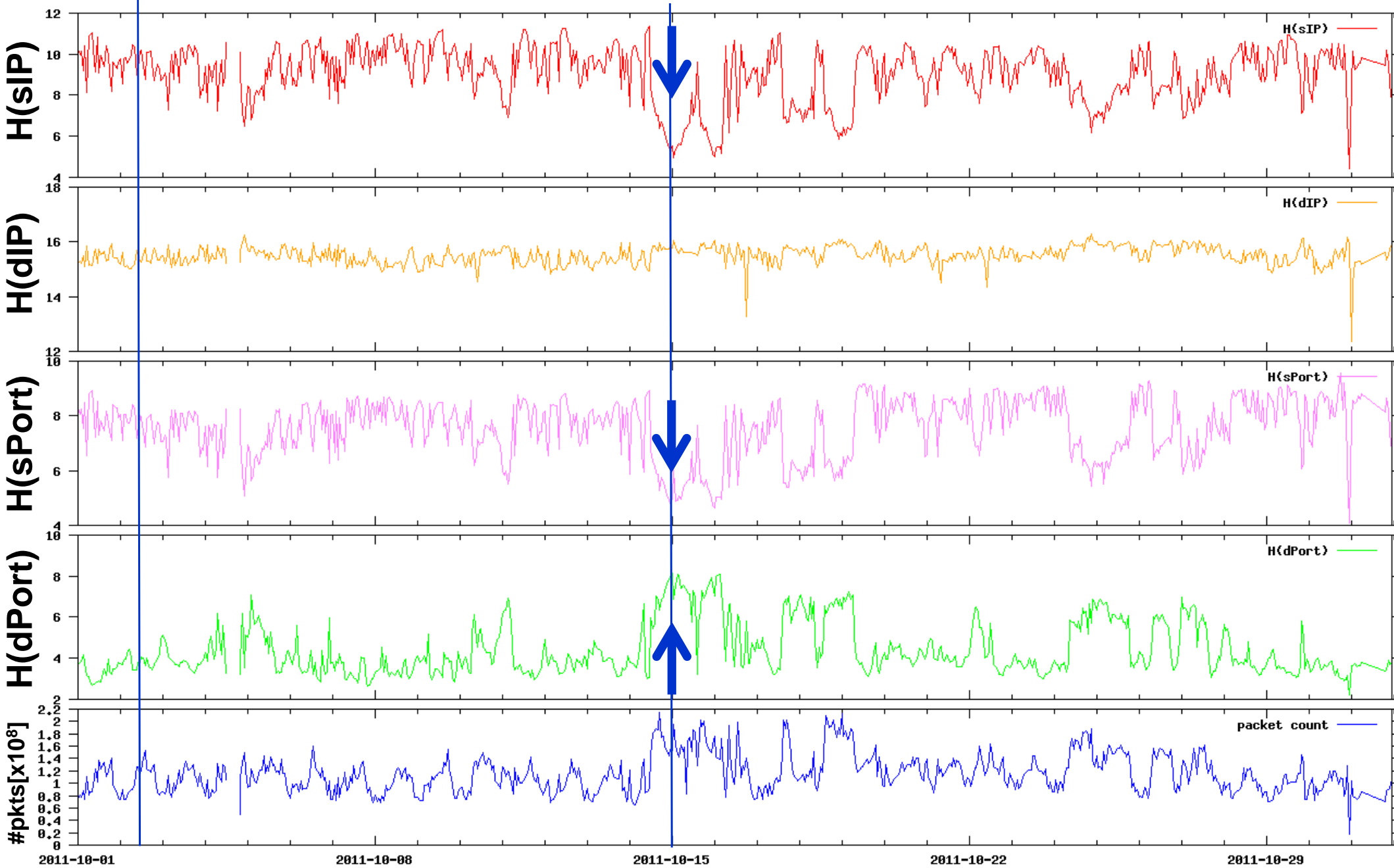
H(dPort)=3.19

Top dPort: 445

H(sPort)=4.70

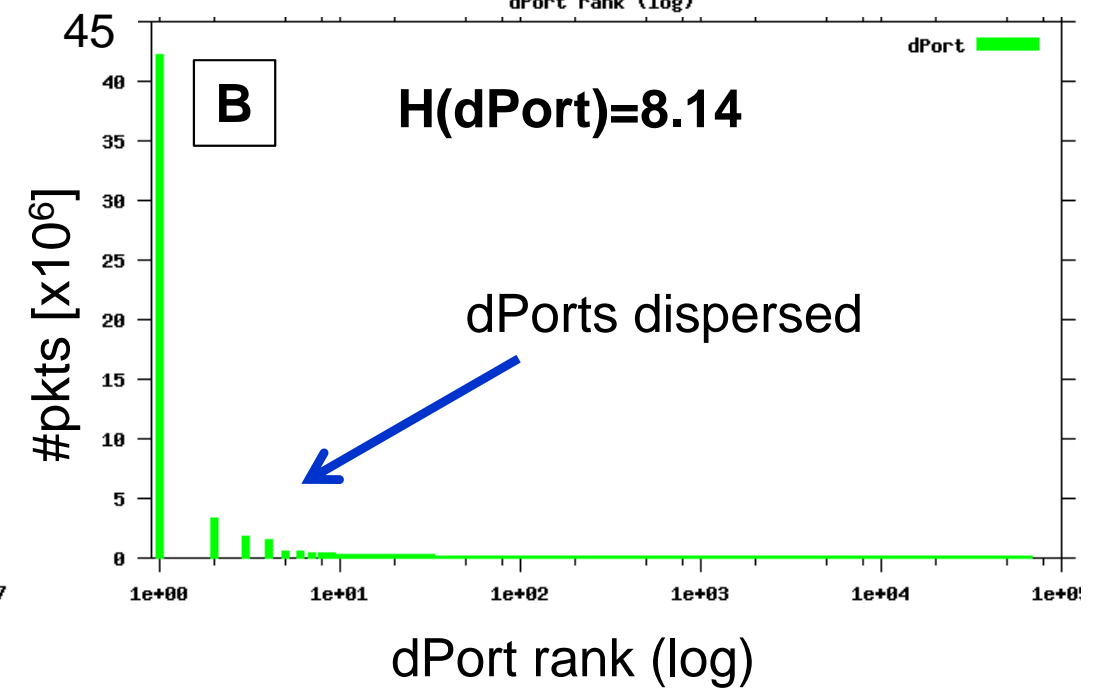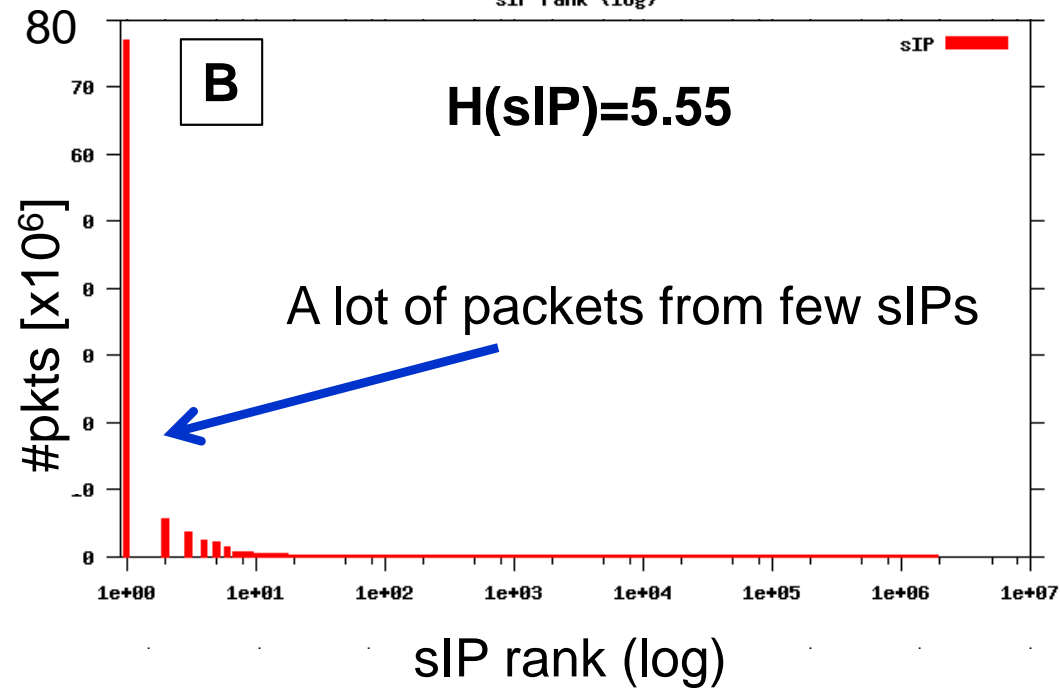A lot of packets from one port

Top sPorts: 80, 9021

H(dPort)=8.14

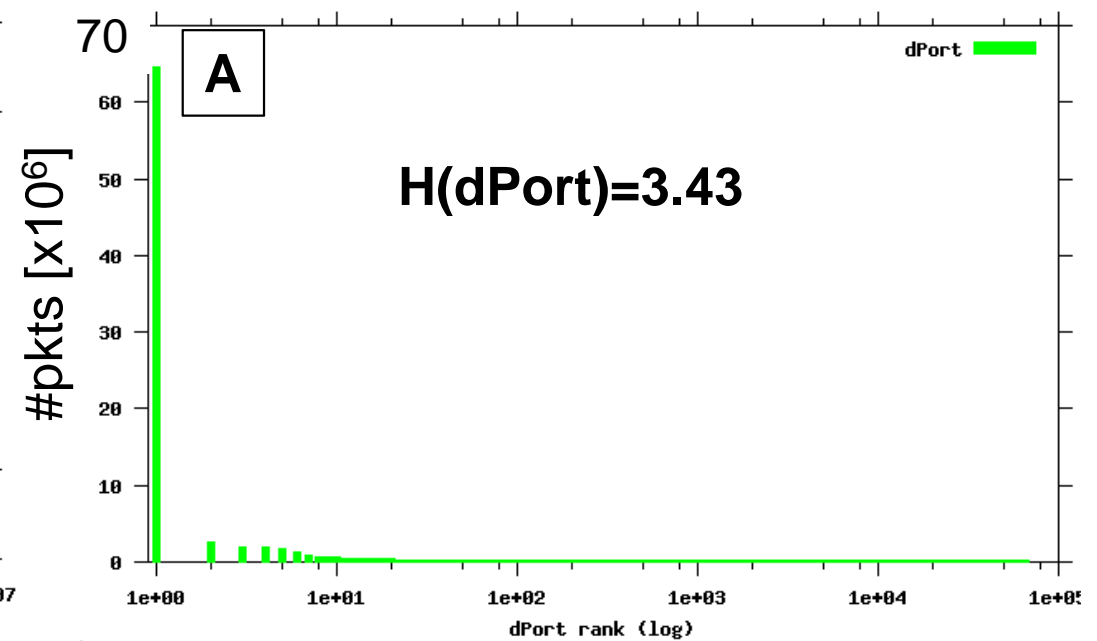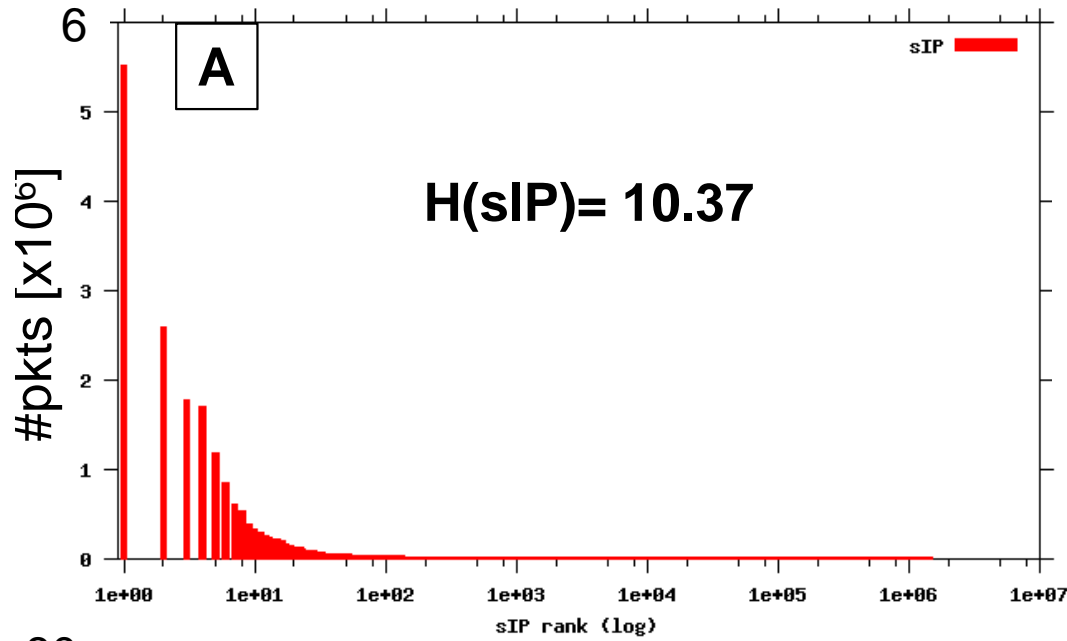dPorts dispersed

OCT 2011

Oct 2011

Classification of Event B

	Hostscan	Backscatter	Misconfig	Outage	DDoS (rare)	Portscan (rare)
sIP	random ↑	specific ↓	specific ↓	specific ↓	random ↑	specific ↓
dIP	random** ↑	random** ↑	specific ↓	depends	specific ↓	specific ↓
sPort	random* ↑	specific ↓	depends	depends	random* ↑	random* ↑
dPort	specific ↓	random* ↑	specific ↓	depends	specific ↓	random ↑

Distributions: sIP, dPort

Oct 2011

Discussion

- Entropy
  - Good indicator for new incidents in darkspace
  - Comprehensive metric to detect and classify different incidents

- Future considerations:
  - Detection of slow and small changes
    - Outages were not visible with current time interval
    - Stealth scanning
    - ➔ check fine grained time intervals
  - Time interval vs. calculation effort
  - Entropy calculation effort compared to other methods
  - Problems with nested events
  - Combination with other metrics (geolocation, source groups,…)
  - Combination with other DS monitors

CAIDA Workshop on Darkspace Analysis

- May 2012, San Diego
- Objectives
  - Bring community together
  - Share experiences
  - Share data, results
  - Establish global distributed DS network

- Participation by invitation
  - If interested ➔ contact me

Thank You!

Contact: **tanja@caida.org**