SMARTxAC: Traffic Monitoring and Analysis System for high-speed links

Advanced Broadband Communications Center (CCABA)

Technical University of Catalonia (UPC)

Pere Barlet

Josep Solé-Pareta

Jordi Domingo-Pascual

{pbarlet, pareta, jordid}@ac.upc.es

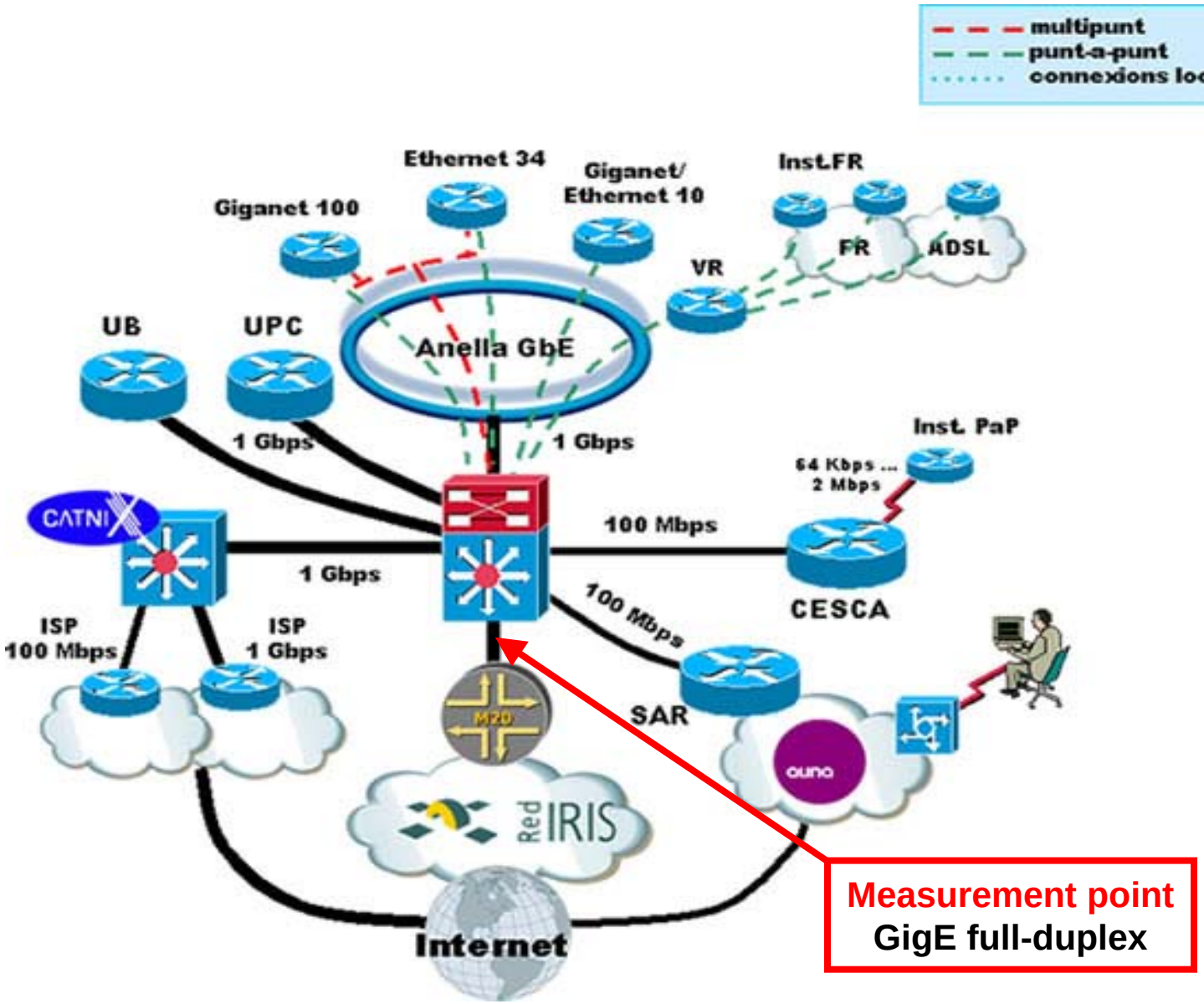
http://www.ccaba.upc.es

Acknowledgment: This work has been partially supported by CESCA (SMARTxAC agreement) and MCyT (ref. TIC2002-04531-C04-02)

SMARTxAC

- Collaboration agreement started on July 2003
  - UPC (Technical University of Catalonia)
  - CESCA (Supercomputing Center of Catalonia)
- Development and deployment of a low-cost monitoring system for the Catalan R&D network
  - Anella Cientifica (Scientific Ring)
  - Connects ~50 universities and research centers in Catalonia
- Objectives
  - Continuous monitoring and analysis of the Anella Cientifica
  - Gather knowledge about per institution usage
  - Detection of anomalies, irregular usage and attacks
- Measurement of a full-duplex GE link
  - Connection to RedIRIS (Spanish R&D) and to global Internet
  - Current traffic load ~700 Mbps / ~150 Kpps

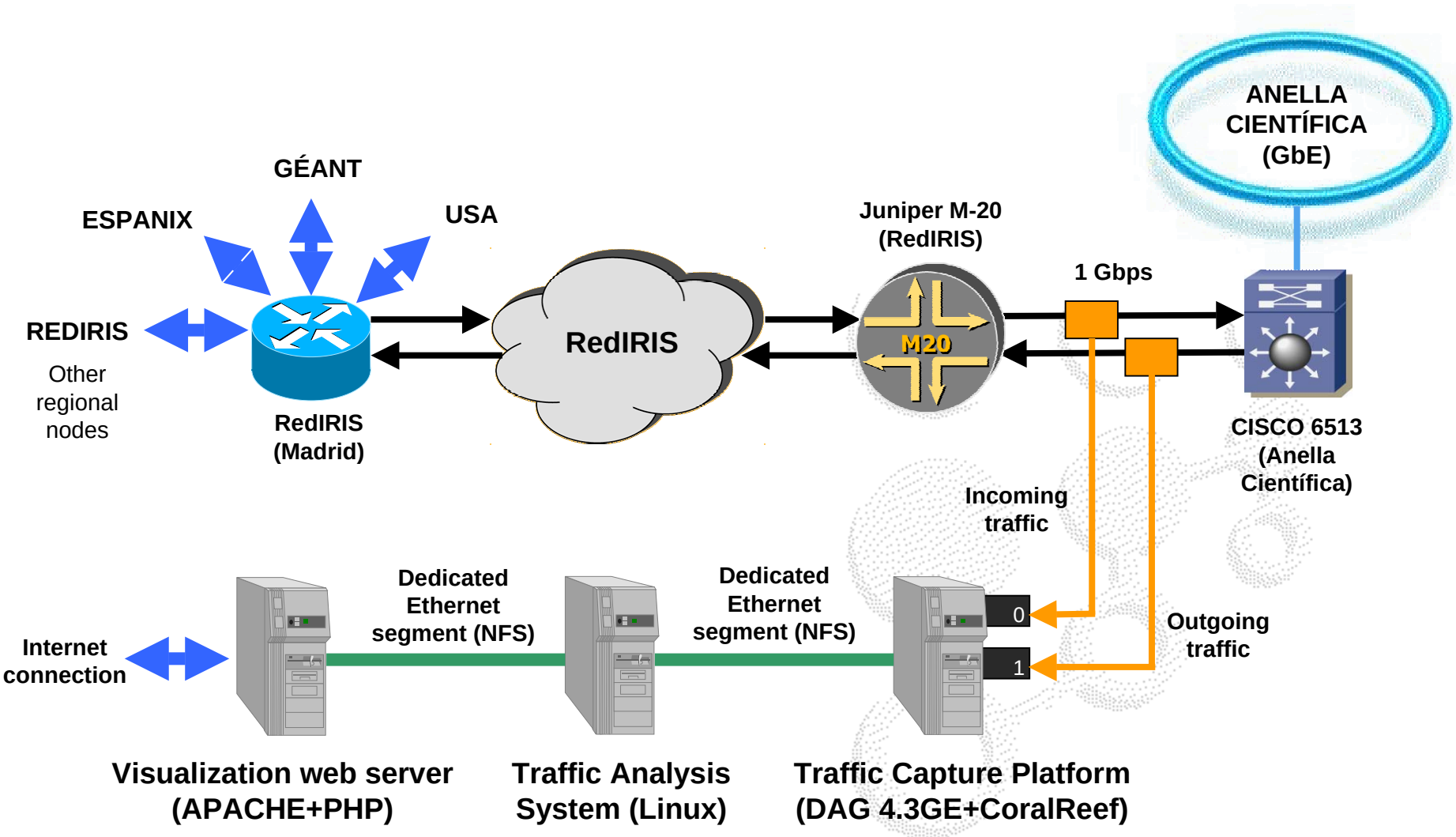
Anella Cientifica



Main characteristics

- Passive measurement
  - Full capture (no sampling)
  - Equipment: DAG 4.3GE + optical splitters
  - Precise timestamping using GPS (Trimble Acutime 2000)
  - CAIDA CoralReef: Packet capture + flow aggregation
    - *CoralReef is going to be replaced shortly by SMART*
- Traffic analysis
  - Analysis of all traffic at full-line rate
  - Header-only analysis due to:
    - *Performance reasons*
    - *Encryption techniques*
    - *Legal restrictions*
  - Permanent storage of all analysis results
- Web-based graphical interface
  - On-demand visualisation of all computed statistics

Measurement scenario



Flow classification approach

- Traditional 5-tuple flows are translated into 3-tuple “classified flows” (<inst, dst, app>)
  - IP addresses → Institution and destination network
    - *Longest prefix match algorithm using a BGP dump*
  - Ports + protocol → Application
  - Bidirectional aggregation
- Unknown traffic and relevant traffic to detect anomalies is logged with more details
- Aggregation into accounting periods
  - Daily, weekly and monthly data-aggregation
- Higher aggregation than traditional flow classification
  - E.g. disk occupation in “Anella Cientifica” reduced by >99%
    - *5-tuple flows: ~25 GB/day*
    - *Classified flows: ~20 MB/day*

Traffic analysis statistics

- Traffic statistics per institution
  - Application time-series plot (bytes/sec, pkts/sec, flows/sec)
  - Institution breakdown
  - Destination network breakdown
  - Application breakdown
  - Protocol breakdown
  - Destination per applications breakdown
  - Destination per institution breakdown
  - Unknown ports log
  - Unknown IP addresses log
  - Unknown protocols log
  - Top-N ports and applications
  - Top-N IP addresses
  - Top-N protocols
  - Threshold-based anomaly detection graphs
  - Self-similarity estimation
  - Packet size distribution

SMART

- SMART is currently under development
  - Specifically designed to perform at gigabit speeds
  - Integration of capture engine and analysis system
    - *Only one measurement box runs capture and analysis processes*
    - *CoralReef is no longer needed and will not be used*
  - New statistics (e.g. ASxAS matrices)
  - Data anonymization
  - Anomaly detection capabilities
  - IPv6 support
- Collaboration between UPC and NLANR/PMA
  - Tested in one of NLANR/PMA OC192MON's located on SDSC's TeraGrid Cluster
    - *http://pma.nlanr.net/Special/tera2.html*
  - IP trace from “Anella Cientifica” was collected for NLANR/PMA
    - *http://pma.nlanr.net/Special/cesc1.html*

Anomaly detection

- Detect anomalies based on changes in institution traffic patterns has several difficulties
  - Define an “ordinary” (expected) traffic profile per institution
  - Rule to decide which deviations are considered as an anomaly
  - Inherent variations of traffic by itself
    - *E.g. burstiness, day/night, workday/weekend, etc.*
  - Minimise number of false alarms

- Anomaly detection (only header analysis) based on:
  - Simple thresholds per institution (packets, bytes, flows, etc.)
    - *Already implemented and working*
  - Adaptive traffic prediction
    - *“Ordinary” traffic profile has not to be defined explicitly*
    - *First tests using “adaptive normalized least mean square error linear predictor” were very successful*
  - Combination of both methods to avoid limitations
    - *E.g. use of thresholds can mitigate prediction limitation when constant traffic changes occur*

Integration of SMARTxAC and CoMo

- Collaboration Intel Research Cambridge - UPC
- Objective: Integration of SMARTxAC with CoMo
  - CoMo has been designed as an open monitoring infrastructure
  - Migrate statistics computed by SMARTxAC as CoMo modules
  - Migrate SMARTxAC graphical interface into CoMo
  - Collaborate in the design and development of CoMo to:
    - *Identify limitations which can difficult such integration*
    - *Facilitate the development and integration of custom modules*
- Participation also in the design and development of CoMo core
  - Capture, export and query processes

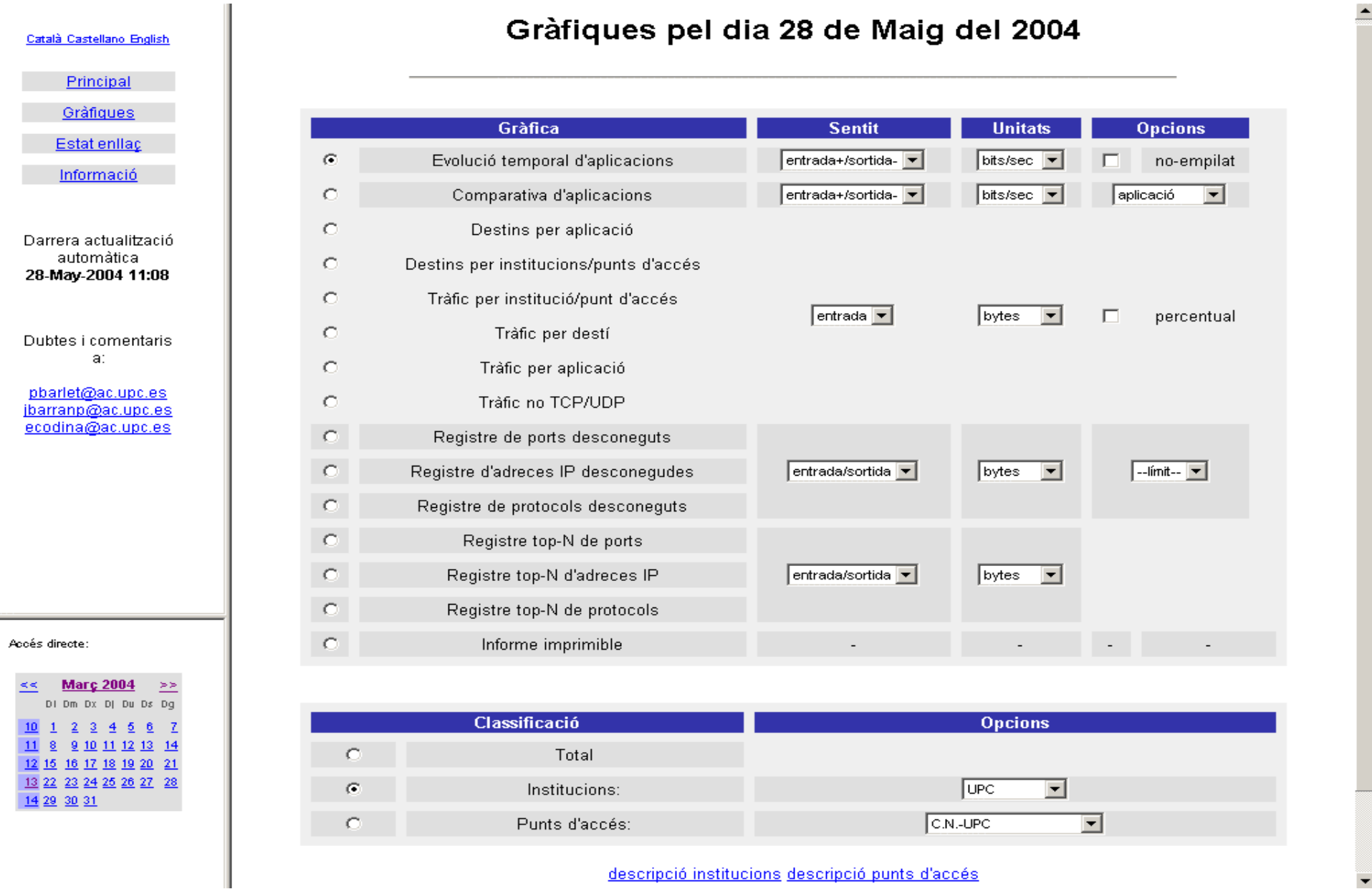
Online demo

http://smartxac.ccaba.upc.es

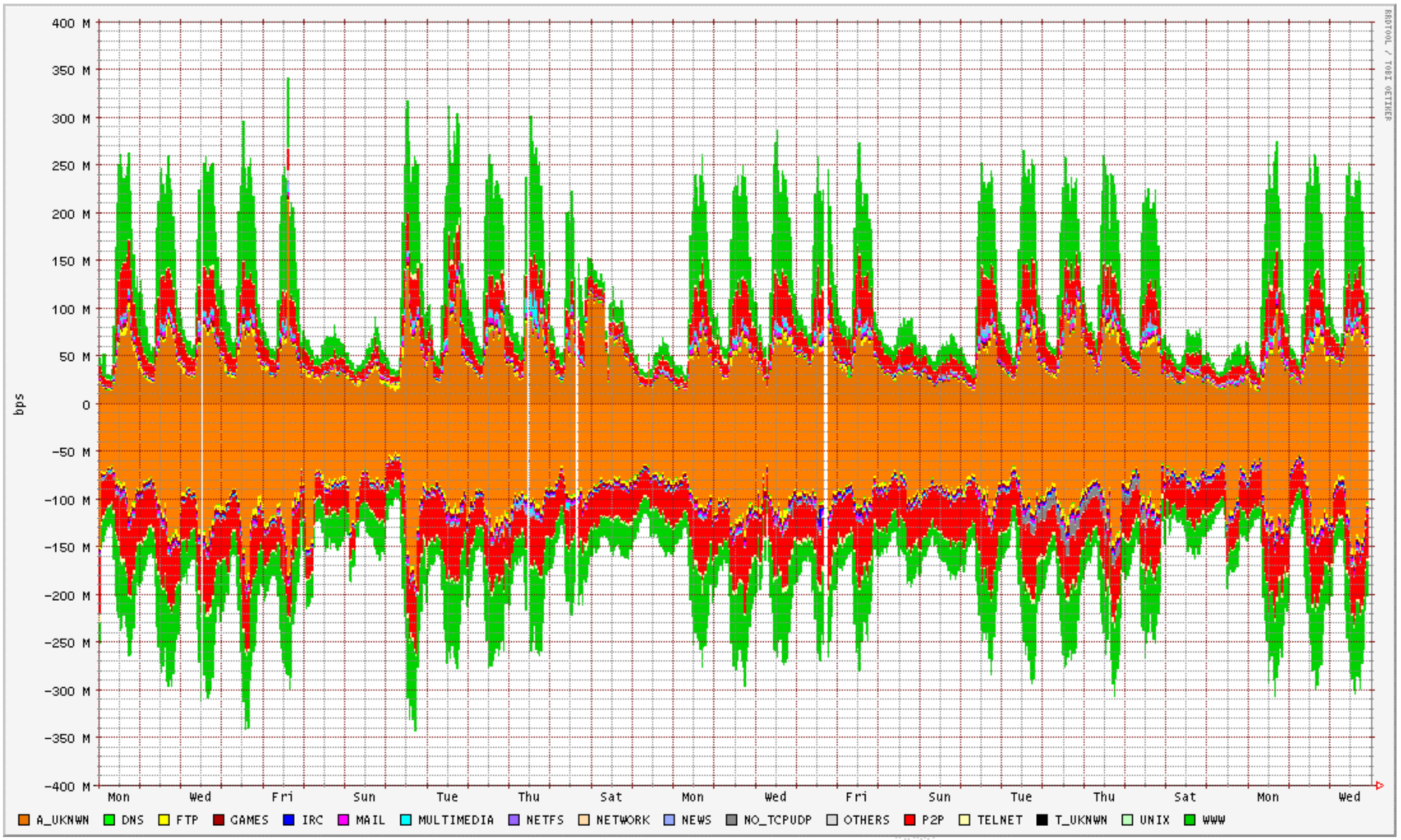
(access to this site is restricted due to data confidentiality)

General information and sample graphs can be found at:
http://www.ccaba.upc.es/smartxac

Web-based graphical interface

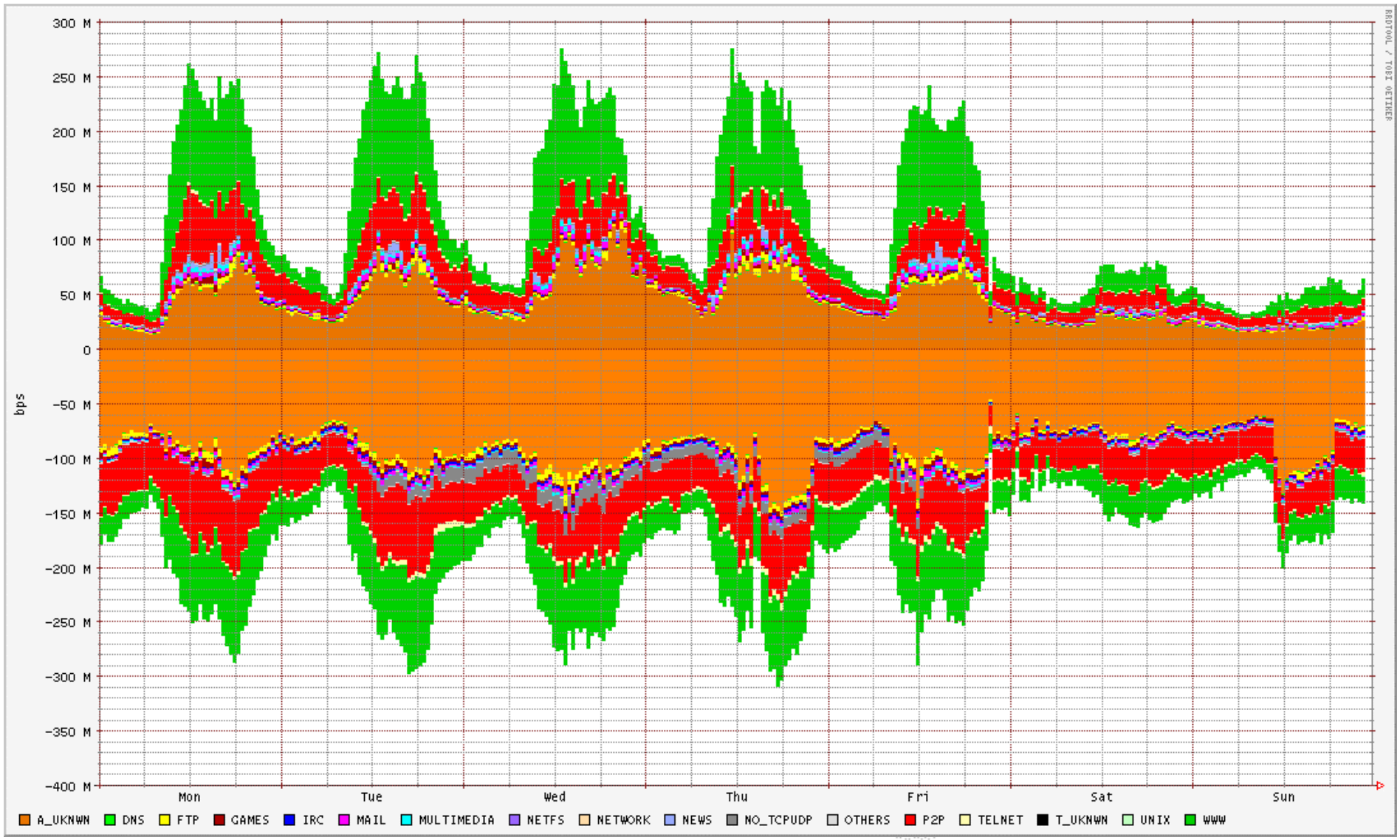


Monthly traffic per application

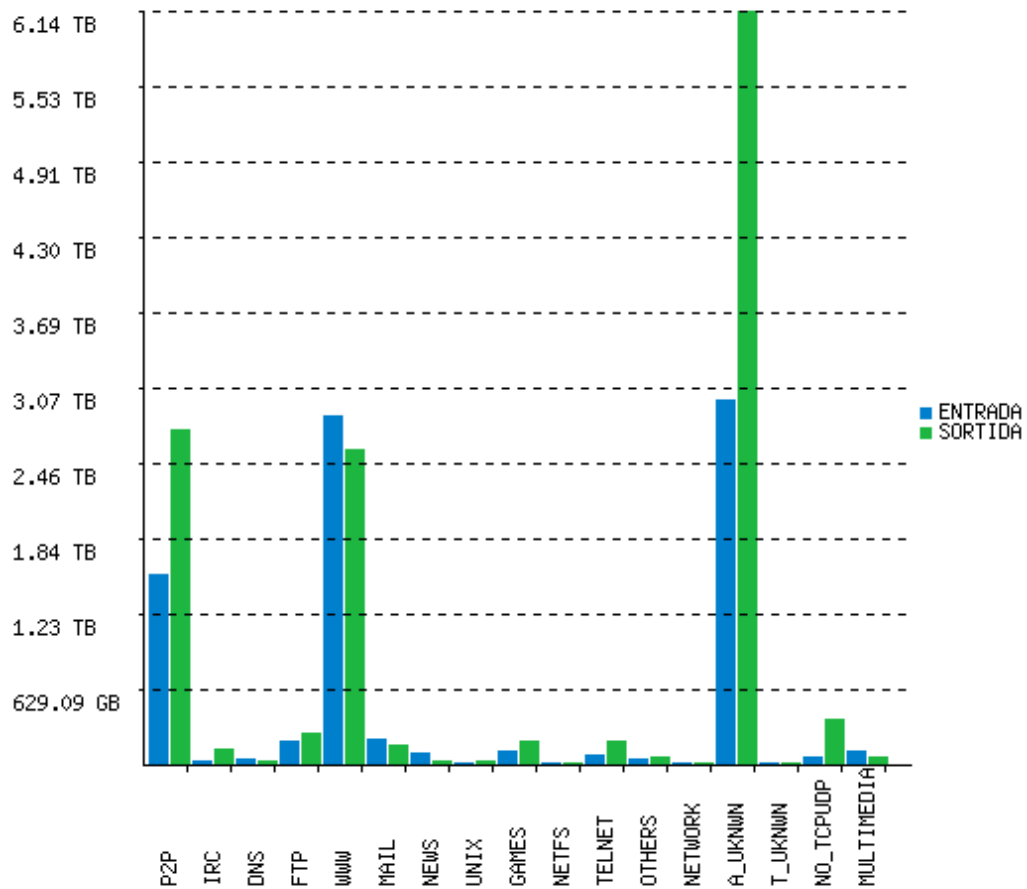


Institution graphs are not shown in order to preserve institution confidentiality

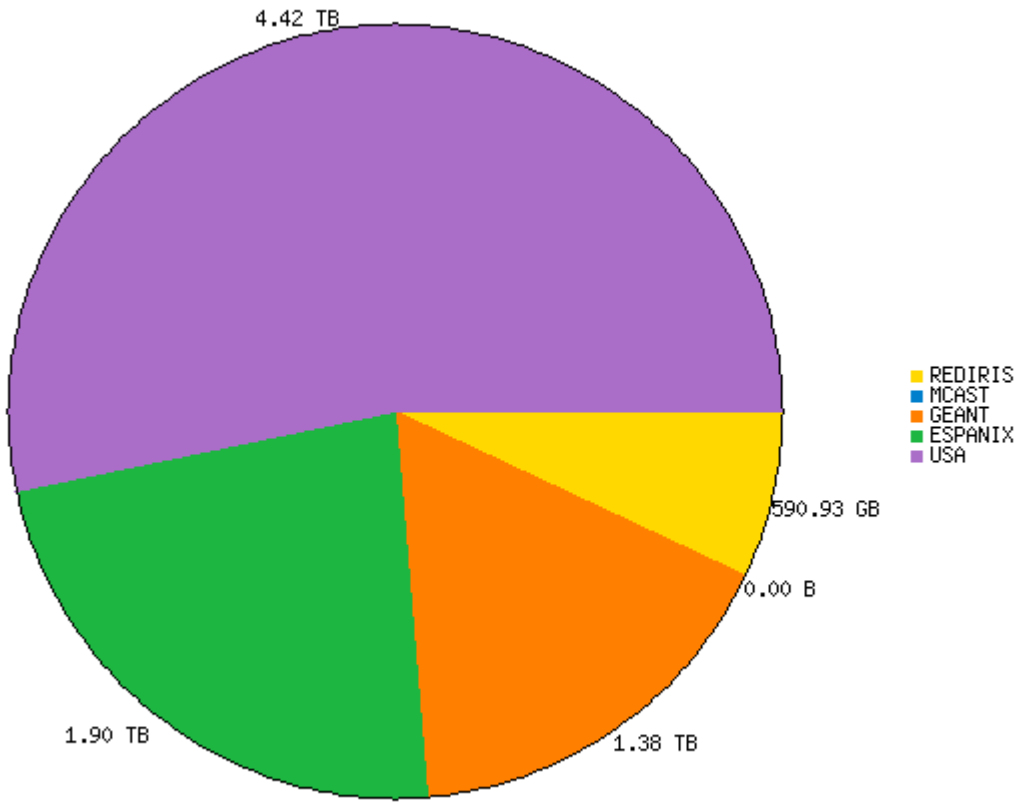
Weekly traffic per application



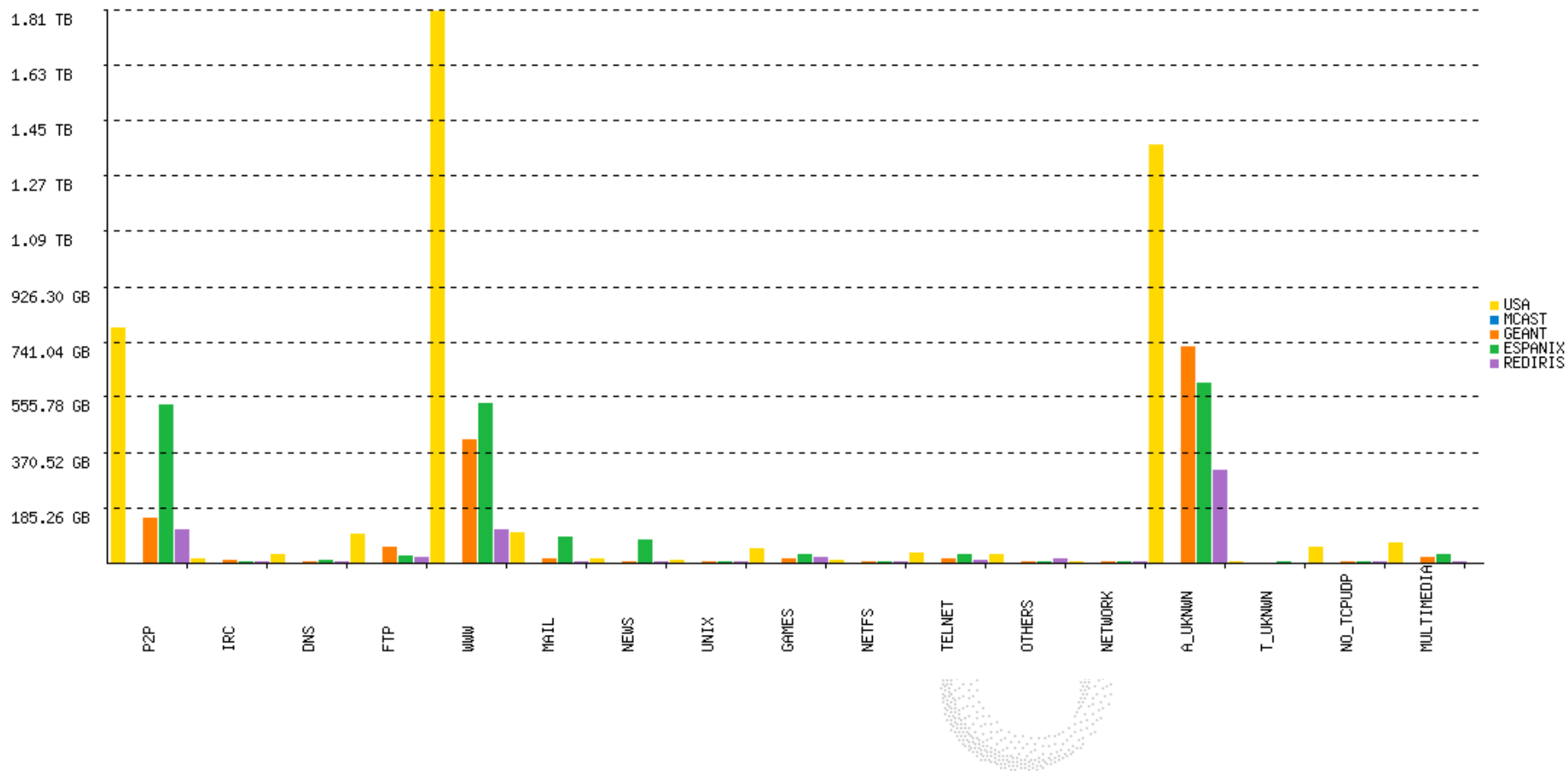
Application breakdown



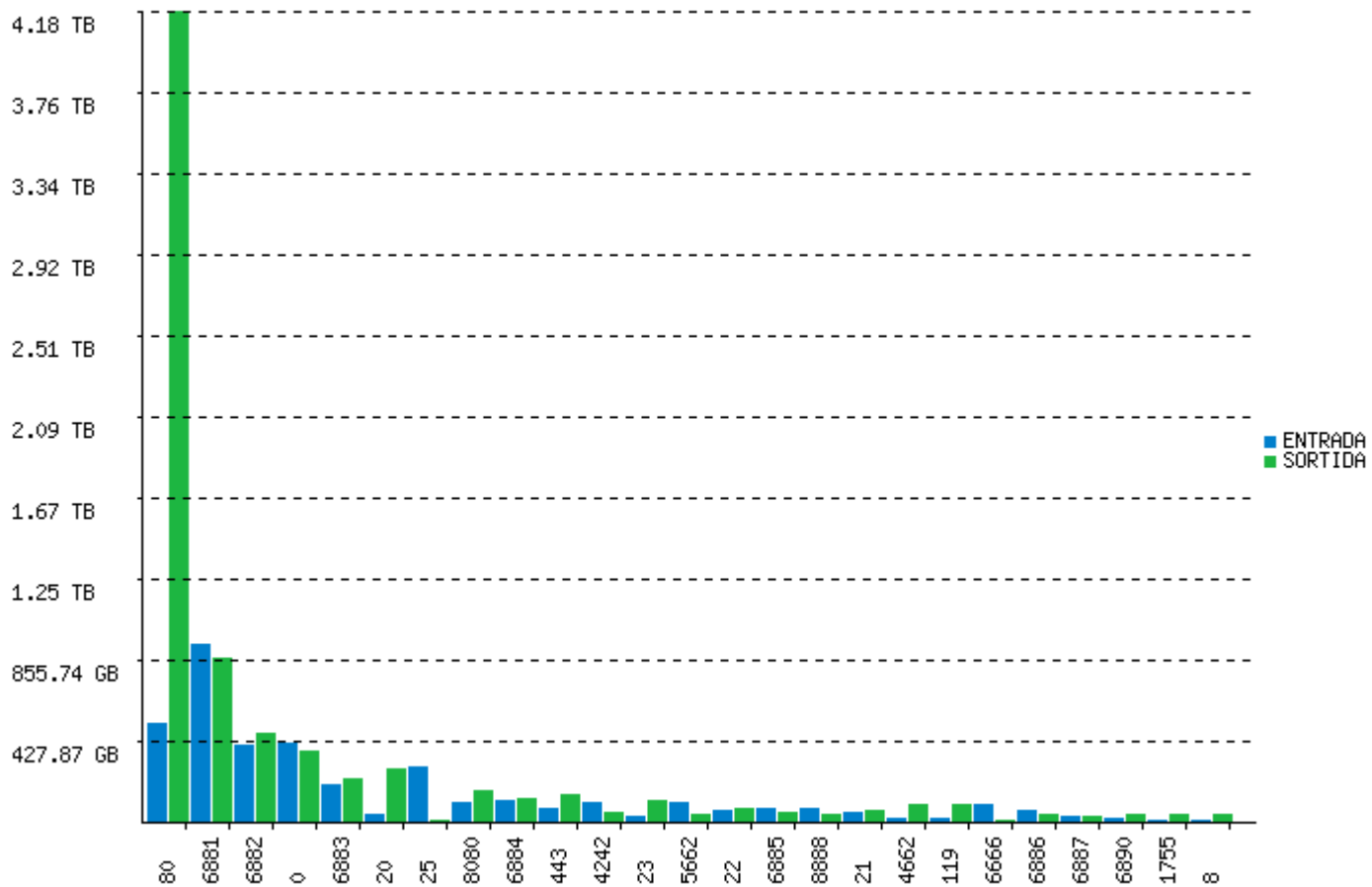
Traffic per AS group



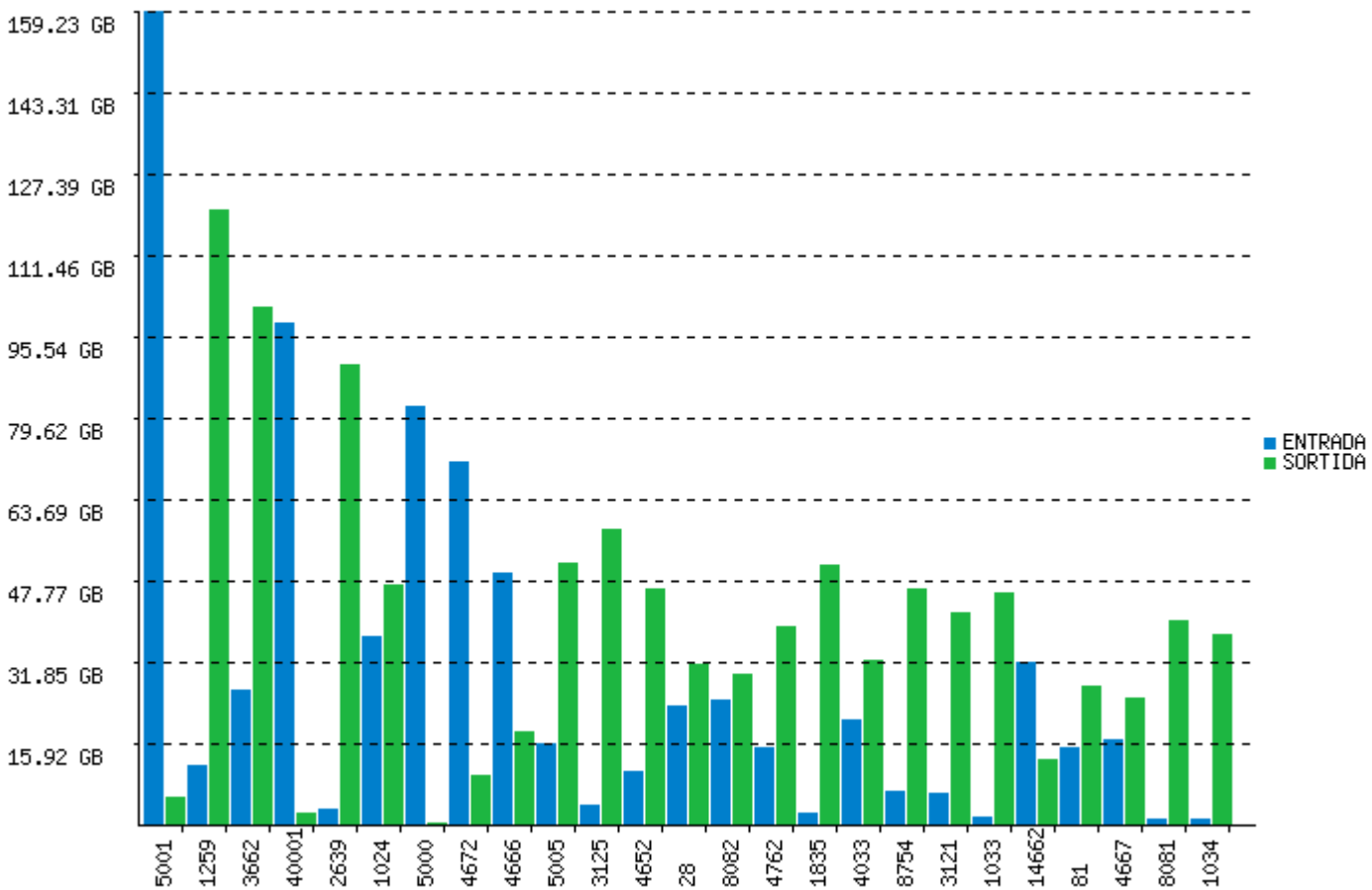
Traffic per AS group and application



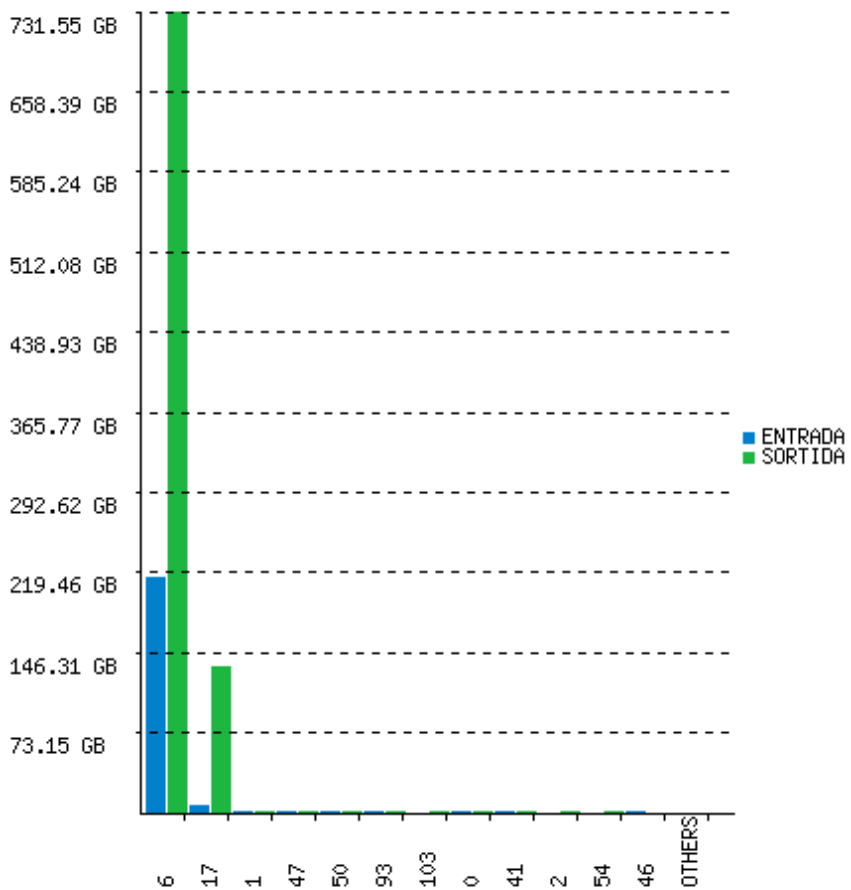
Top-N known ports



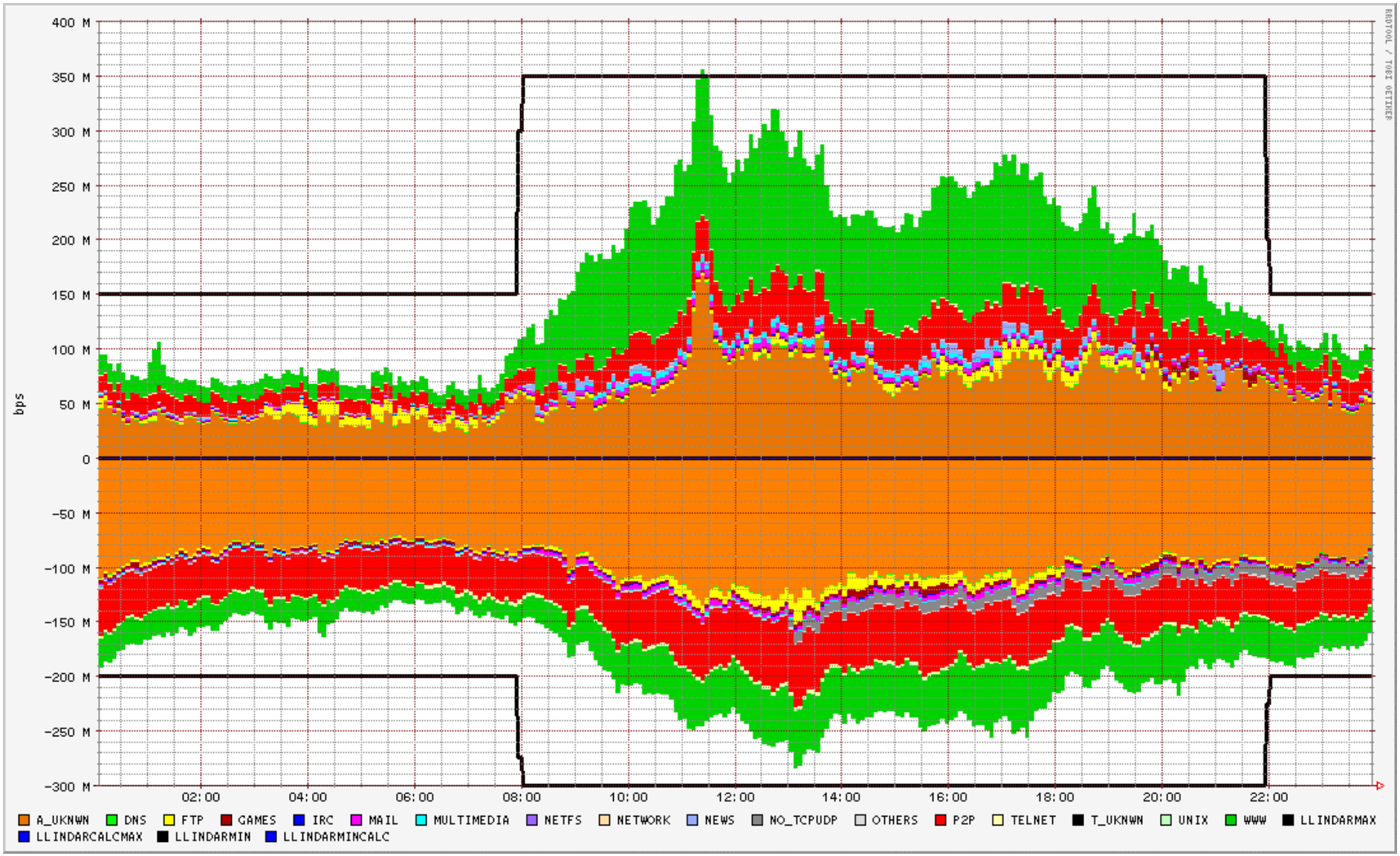
Top-N unknown ports



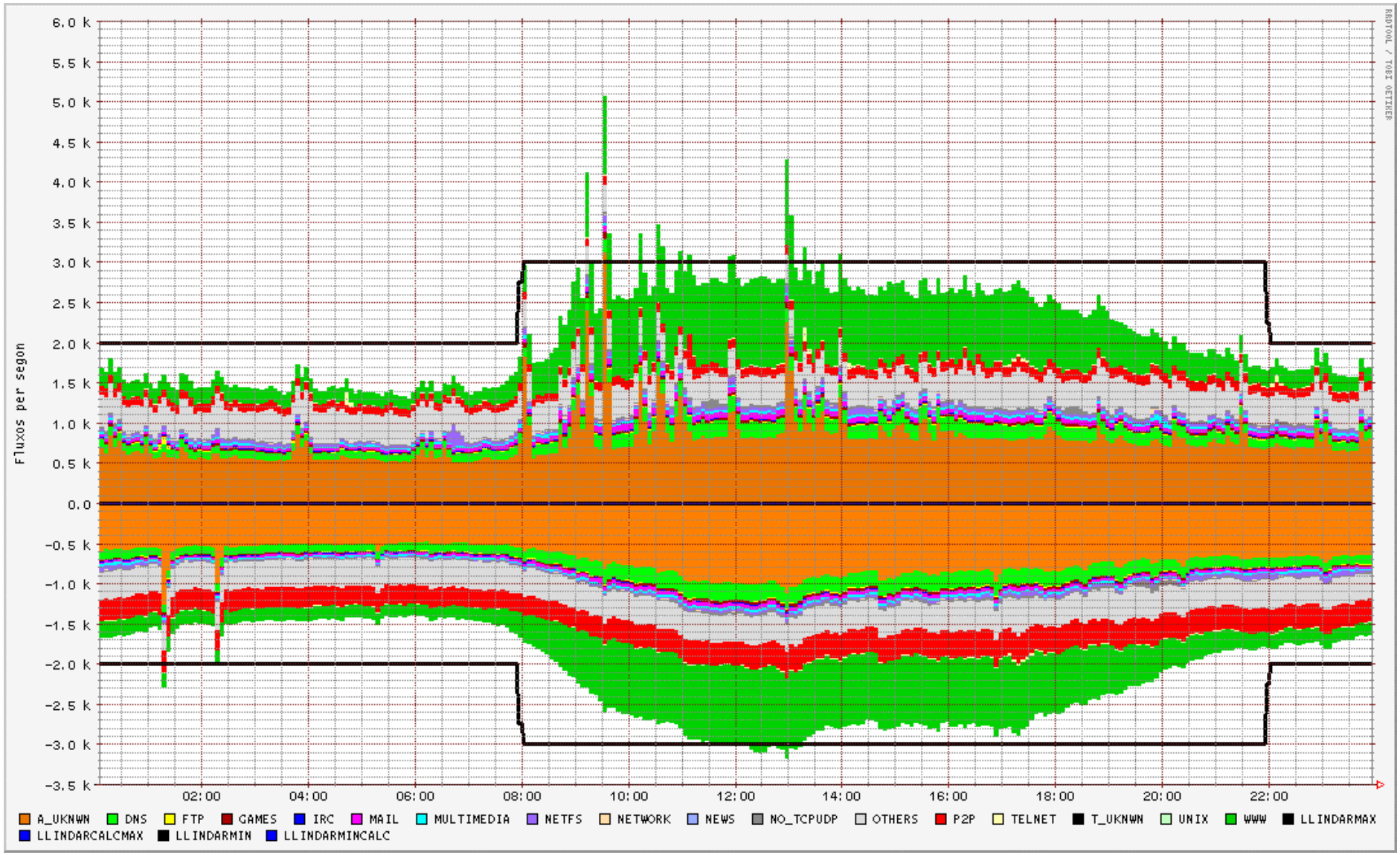
Top-N protocols



Threshold-based anomaly detection (bps)



Threshold-based anomaly detection (flows)



Packet size CDF