

# IPSOFACTO: A Visual Correlation Tool for Aggregate Network Traffic Data

Flip Korn  
AT&T Labs-Research

S. Muthukrishnan  
AT&T Labs and Rutgers Univ.

Yunyue Zhu<sup>\*</sup>  
New York University

## ABSTRACT

IP network operators collect aggregate traffic statistics on network interfaces via the Simple Network Management Protocol (SNMP). This is part of routine network operations for most ISPs; it involves a large infrastructure with multiple network management stations polling information from all the network elements and collating a real time data feed. This demo will present a tool that manages the live SNMP data feed on a fully operational large ISP at industry scale. The tool primarily serves to study correlations in the network traffic, by providing a rich mix of ad-hoc querying based on a user-friendly correlation interface and as well as canned queries, based on the expertise of the network operators with field experience. The tool is called IPSOFACTO for *IP Stream-Oriented FAsT Correlation TOol*.

## 1. IPSOFACTO

In the academic literature and in practice, the need for understanding correlations in time series data has long been known. It is our experience that the correlations in network phenomena are often complex, and root causes for events are often subtle, but a flexible, browsing tool with the ability to pose sophisticated correlation queries is valuable in the hands of a network analyst. IPSOFACTO is designed to meet this expectation.

### 1.1 Visualization

IPSOFACTO has a graphical user interface for selecting network elements and links, filtering them on desired aspects, and displaying traffic time series and correlation plots. Links are bichromatic to indicate the link's endpoints. Once a link-edge is clicked on, the display confirms this by highlighting the link and identifying the routers at the endpoints. Collocated links are represented by a single edge, but a drop-down menu can be used to select from a list of links when a representative edge is clicked on. Once (an endpoint of) a single link is selected and the "get traffic" button is clicked on, IPSOFACTO then displays a plot of the traffic rate for the selected interface on the specified date. The user can then compute autocorrelations or correlations with other links as described next.

<sup>\*</sup>This work was done while the author was a DIMACS visitor at AT&T Labs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD 2003, June 9-12, 2003, San Diego, CA.

Copyright 2003 ACM 1-58113-634-X/03/06 ...\$5.00.

## 1.2 Correlations

Finding correlations automatically is a formidable task with time-varying effects, and dynamics we do not completely understand. IPSOFACTO allows users to express complex, ad-hoc correlations in a principled way. These queries can be developed and refined interactively. There are three query modes: specific-pair (given a pair of individual links, find their correlation); all-elements (given a single link, find all links that correlate); and all-pairs (find all pairs of links that correlate). Each mode can either correlate pairs of links over the same time period or autocorrelate a link with itself at two different times.

Given some pair  $(x, y)$  of time series vectors to correlate, our interface allows the user to specify four components of the correlation function  $corr(x, y)$  that will be computed on these vectors: **preprocessing** (scaling, smoothing, aggregation, etc.); **temporal band** (determines at which timestamps  $t_i$  and  $t_j$  to "join" the elements  $x_i$  and  $y_j$  of respective vectors  $x$  and  $y$ ); **scalar operator** (to be applied on these pairs  $(x_i, y_j)$ ); and **summary** (how to aggregate the output from the scalar operations into an output stream).

## 2. DEMO

The demo will highlight the flexibility and effectiveness of the visualization interface and the power of performing correlation analysis using the functionality provided by IPSOFACTO. Some sample ad-hoc queries are as follows:

- What is the correlation between a specified link  $x$  and a specified adjacent link  $y$ ?
- What is the autocorrelation on link  $x$  of traffic from this Wednesday compared with that of last Wednesday?
- Given a link  $x$  and the devices at its endpoints, which other links between the same device-pair exhibit a similar burst at roughly the same time?
- Find all links for which the difference in outbound traffic at one endpoint versus the inbound traffic at the other endpoint (summed over a day) is greater than  $\epsilon$ .

We will also demonstrate the ability of the tool to be interactively used to follow an "event" in the network. For example, say there is a spike of traffic at a router. The tool can be used to look at the neighboring routers and zoom in on links which correlate most significantly with this spike. Our experience has been that this functionality is very helpful for a network analyst to develop viable hypotheses for causes and extent of "events" in the network and do historic correlation searches to test such hypotheses.