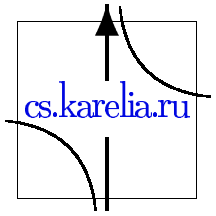


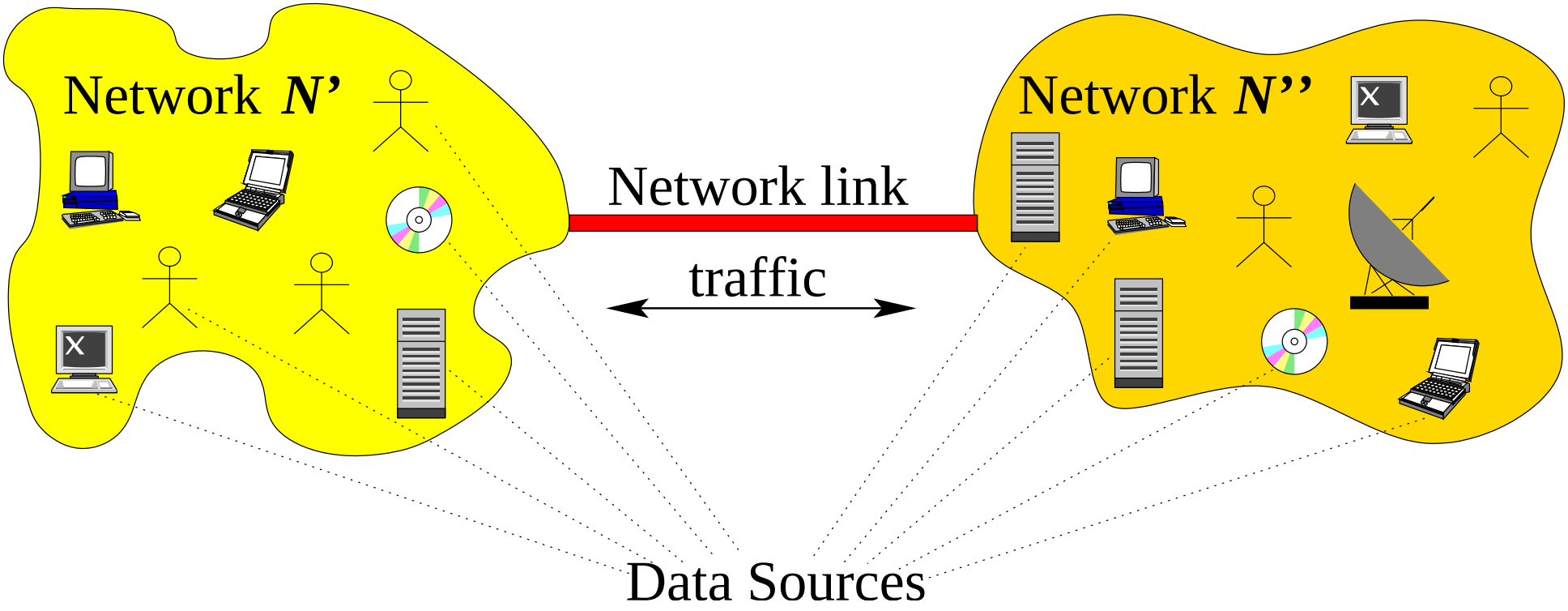
On the Problem of Modeling and Identification for Traffic Structure of a Network Link

Dr. Yury A. Bogoyavlenskiy, Dr. Dmitry G. Korzun

{ybgv,dkorzun}@cs.karelia.ru

June 26, 2003

Preliminaries



Data ***sources*** send data to the ***link***.

All data are multiplexed and form ***traffic*** of the link.

*Are there invariants—**traffic structure**?*

Two-Component Description of Traffic

Model: traffic = structure $\mathcal{S}$ + stochastic $\mathcal{P}$

Structure: invariant rules of traffic forming

Stochastic: changeability of traffic

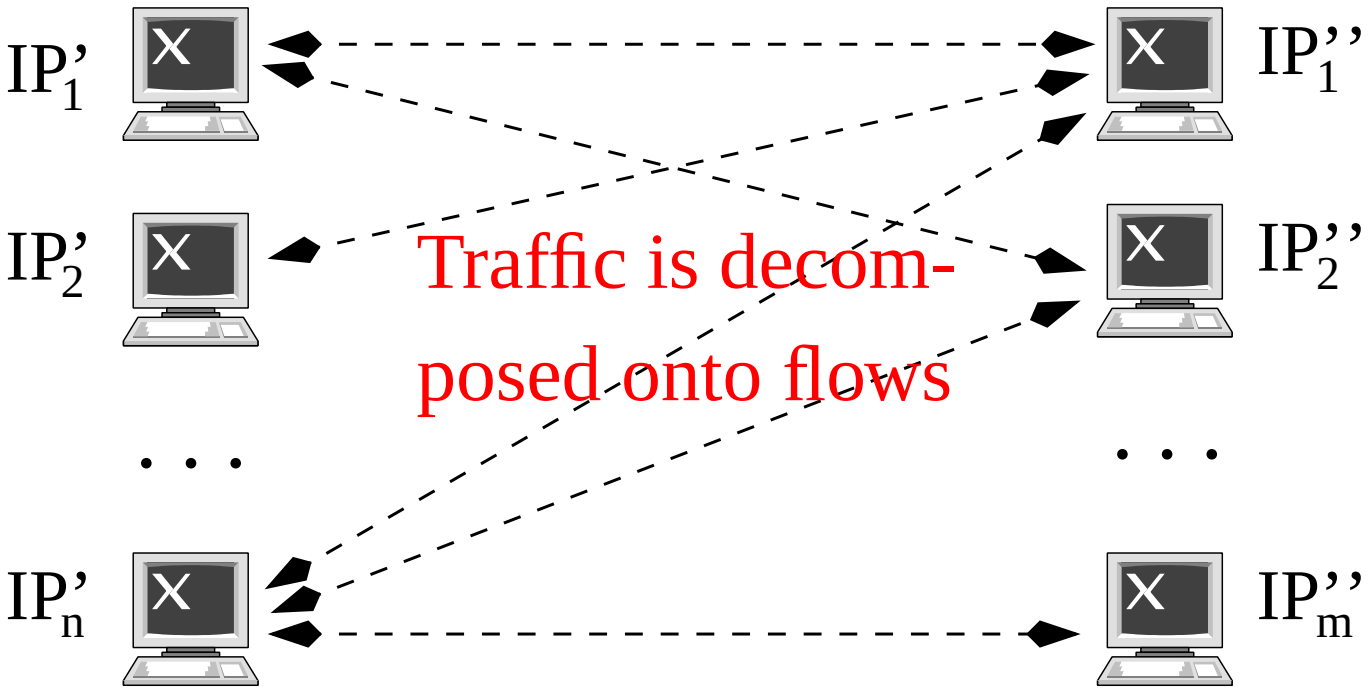
Structural model of a traffic characteristic (e.g. throughput**):**

$$V(\Delta) = F_{\mathcal{S}}\left(x_1(\mathcal{P}), \ldots, x_m(\mathcal{P})\right)$$

$V(\Delta)$ is modeled variable on time interval Δ

x_i are input (measured) parameters (e.g. data volumes)

Flows (TCP connections, user sessions, etc.)



Network endpoints form the structure.

Observations:

5 month,

$2.5 \cdot 10^6$ IPs,

75.9 Mb/day,

Total 1.8 Tb,

$9 \cdot 10^8$ flows,

$6.7 \cdot 10^9$ packets

$$\text{Troughput}(\Delta) = \frac{1}{\Delta} \sum_{(\text{IP}'_i, \text{IP}''_j)} v(\text{IP}'_i, \text{IP}''_j)$$

Aggregated Data Sources

- The number of real sources is too large
- Heterogenous sources should be classified
- Model should be simple and compact

$$\text{Troughput}(\Delta) = \sum_{s \in \mathcal{H}} \nu_s(\Delta) v_s$$

$\nu_s(\Delta)$ — activity of the aggregated source s

v_s — permanent characteristic of the aggregated source s

How real sources can be transformed to aggregated ones?

Model Identification Problem

$x = \big(x_1(\Delta), x_2(\Delta), \ldots, x_m(\Delta)\big)$ is classified data volumes

Hypothesis: there exists a linear Diophantine system $\forall \Delta$

$$Ax = \mathbb{O}\,, \qquad A \in \mathbb{Z}^{n\times m}\,, \quad x \in \mathbb{Z}_+^m \tag{1}$$

Model: $\text{Troughput}(\Delta) = \sum_{s\in\mathcal{H}} \nu_s(\Delta) h^{(s)}$,

where $\mathcal{H} = \{h^{(1)}, \ldots, h^{(q)}\}$ is Hilbert basis of (1)

Problems:

- equality in (1) is approximate and scalable
- identification of the matrix A
- searching $\mathcal{H}$ and $\nu_s(\Delta)$

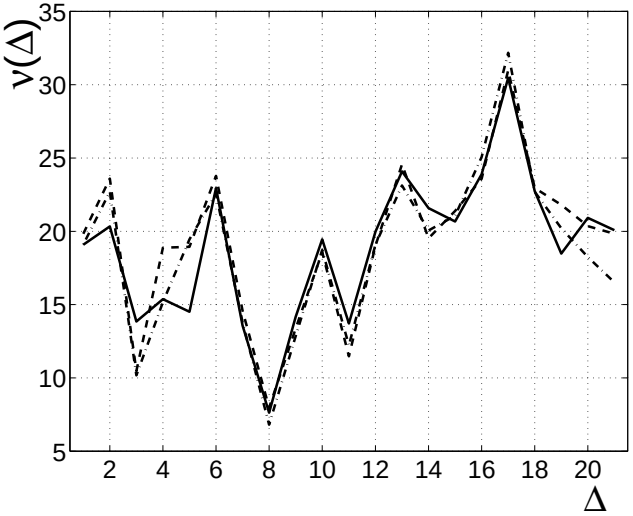
Experiment confirmation

Classifi cation: 6 groups by volume of flows

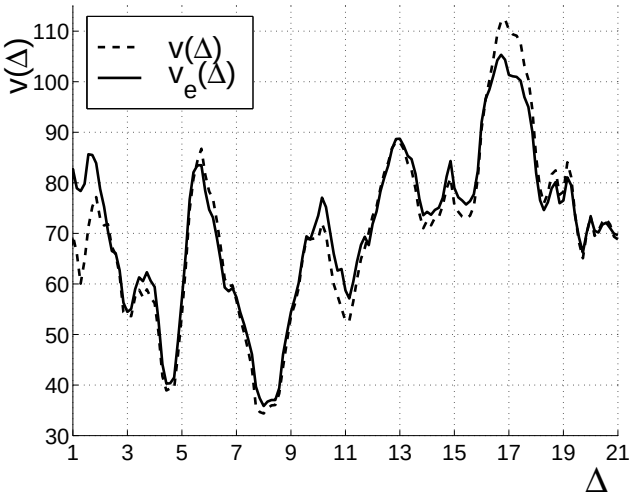
Gr.1	Gr.2	Gr.3	Gr.4	Gr.5	Gr.6
$[0, 100)$,	$[100, 500)$,	$[500, 10^3)$,	$[10^3, 5\cdot10^3)$,	$[5\cdot10^3, 10^4)$,	$[10^4, +\infty)$

Principal aggregated sources: $h^{(\mathrm{v})} = (0, 0, 0, 2, 1, 1)$,

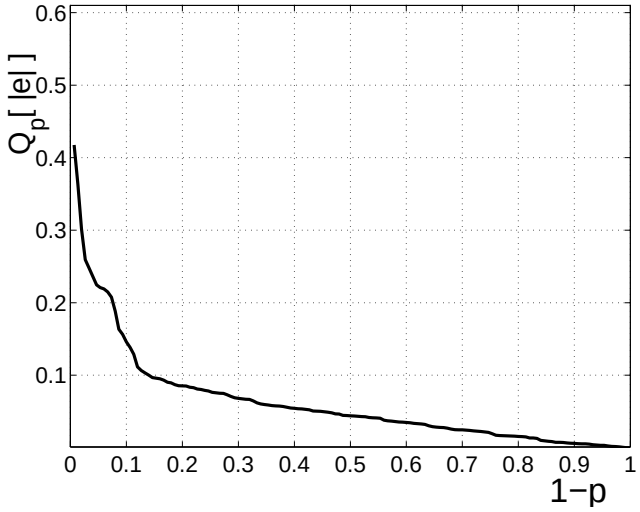
$h^{(\mathrm{m})} = (0, 0, 1, 2, 0, 0)$, $h^{(\mathrm{s})} = (1, 1, 1, 0, 0, 0)$



Activity $\nu(\Delta)$ of h^{s}
act/sec



Throughput $V(\Delta)$
fbws/sec



Percentile of abs.err.

Conclusion

1. Discrete model of traffic structure is proposed
2. The model identification is based on a linear Diophantine system construction and Hilbert basis searching
3. Our experiments confirm the applicability of the model (existence of stable aggregated sources, compactness, scalability, representativeness)