Estimation of correlated flows from link measurements

Mariusz Kamola*†
*NASK — Research and Academic Computer Network
Warsaw, Poland
Email: Mariusz.Kamola@nask.pl
†Institute of Control and Computation Engineering
Warsaw University of Technology
Warsaw, Poland
Email: M.Kamola@ia.pw.edu.pl

***Abstract*—The real network traffic characteristics makes it difficult to estimate origin-destination flows only from statistics for routers' interfaces. The presented approach, based on passively measured link traffic statistical properties, estimates the flows, allowing for flow interdependence and seasonality. Rescaled profile of the total daily traffic in the network is used to compensate deterministic component of flows, before the proper estimation takes place. For adequate identification of the residual flows, it is assumed that they are mixtures of a number of independent random processes, whose parameters are to be identified along with the mixing coefficients. The results obtained for real traffic in Abilene network are promising, and indicate that data preprocessing and application of correct optimization routines are of importance for further research.**

***Keywords*—correlated flows, traffic matrix estimation, origin-destination flow estimation**

I. Introduction

Identification of individual flows in a graph, given only measurements of the aggregated flow on each edge is an abstraction of many important practical problems, the functional magnetic resonance imaging [1] and road traffic origin-destination modeling [2] being the two less obvious to be recalled. The model and existing identification approaches apply equally well to traffic matrix estimation (TME) in computer networks. In the telecommunications, the knowledge of traffic demand is often a prerequisite for modern network management algorithms to operate [3]. Until now, practical applicability of TME for network planning, provisioning and routing was not appreciated much because of still inadequate accuracy of the methods, but also because networks are managed in decentralized manner in many aspects, e.g. route selection, link capacity upgrade, domain name lookup policy etc. However, the rise of software defined networks (SDN) calls for centralized management policies, which may bring the network to new levels of efficiency (in terms of e.g. quality of service - QoS, power consumption reduction, costs of upgrade) only if the network traffic is known well enough. Such demand is close to be met by appearance of modern TME approaches; our work tries to contribute in this field by proposing an alternative identification routine which takes into account the real-life traffic nature.

We propose here an alternative approach to contemporary TME routines, as follows. Let us consider a network with N nodes and L links. Each node generates traffic to all other nodes, which results in N^2-N directed flows which we can order by their source node, and then the destination node. A flow j is forwarded along its individual path that can be represented as a binary vertical vector $\mathbf{a}_j$ composed of elements a_{ij}, where $a_{ij}=1$ if flow j goes through link i, and zero otherwise. The paths form a routing matrix $\boldsymbol{A}=(\mathbf{a}_1,...,\mathbf{a}_{N^2-N})$.

The traffic measurement process happens in discrete time, with the discretization period much greater than the packet round trip time. So, if $\boldsymbol{X}(t)$ denotes the data volumes transmitted by flows in period number t, then the vector of total traffic observed in period t on every link is a vector $\boldsymbol{Y}(t)=\boldsymbol{A}\boldsymbol{X}(t)$. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ (without time index) denote all samples of traffic generated by flows over some observation period, and all flow measurements taken on links, respectively. The problem is to infer about $\boldsymbol{X}$ from $\boldsymbol{Y}$. Given the fact that only traffic aggregates are observed, this problem is inherently an undetermined one. However, with certain assumptions, one is able to figure out at least basic statistical properties of the flows, especially mean flow values, $\bar{\boldsymbol{X}}$.

With adequately strong assumptions about the generated traffic, one is able to estimate the traffic matrix using relatively simple apparatus. If we follow [4] and assume data transmitted by each flow to be independent random variables with Poisson distribution, $\boldsymbol{X}(t) \sim \mathrm{Poisson}(\boldsymbol{\lambda})$, then we can use mean observed volumes on links as consistent estimators for Poisson distribution parameters $\boldsymbol{\lambda}$. This gives us the first-order equation for $\boldsymbol{\lambda}$ estimation:

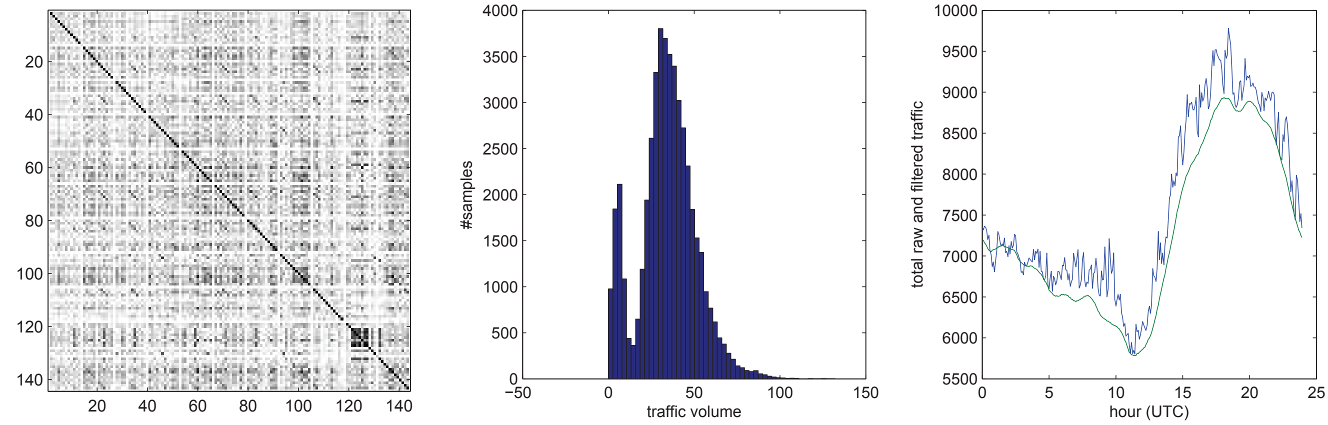
$$\bar{\boldsymbol{Y}}=\boldsymbol{A}\boldsymbol{\lambda}\,, \tag{1}$$

where $\bar{\boldsymbol{Y}}$ is the sample mean for each link. Considering that variance equals the expected value for Poisson distribution, it was observed that the sample covariance for any two links i,i' is equal to the total of variances of all flows that pass this link pair: $\mathrm{cov}(Y_i,Y_{i'})=\sum_j a_{ij}a_{i'j}\lambda_j$ — with a big enough sample number. Such second-order equation can be rewritten for all links in matrix notation:

$$\mathrm{cov}\,\boldsymbol{Y}=\boldsymbol{A}(\boldsymbol{A}^T\circ\boldsymbol{\lambda})\,, \tag{2}$$

where '$\circ$' stands for elementwise multiplication, and $\boldsymbol{\Lambda}$ is a matrix made of $\boldsymbol{\lambda}$ vector replicated L times:

$$\boldsymbol{\Lambda}\stackrel{\mathrm{df}}{=}[\overbrace{\boldsymbol{\lambda},\boldsymbol{\lambda},\ldots,\boldsymbol{\lambda}}^{L}]\,.$$



a) Absolute values of correlation coefficients of flows (white is 0, black is 1).

b) Exemplary histogram of volume samples.

c) Daily total traffic profile: raw (ragged, blue) and filtered (smooth, green).

Fig. 1: Properties of original Abilene traffic.

Equations (1) and (2) form an overdetermined system, usually solved by some optimization routine for minimization of weighed square norm for RHS-LHS. Applying a gradient-based optimization algorithm greatly improves the search time. Unfortunately, the assumption about independent and Poissonian traffic nature is completely unrealistic, thus adding up to the ill-defined nature of (1, 2). We present related work that handles these difficulties in Sec. II — then we propose our approach in Sec. III. Numerical results are given and discussed in Sec. IV; and we conclude in Sec. V.

II. Related work

To address the problem of inadequate number of measurements for solving (1, 2), extra assumptions about the traffic nature must be made. In early works, they were based on: prior traffic matrix estimates, routing weights or destination preferences for sources. The gravity model [5], falling into the last category, has gained much attention. It assumes that a traffic source distributes its traffic across destinations proportionally to the ratios of the totals for ingoing and outgoing network-wide traffic. The concept gets developed in [6], and termed *choice model*, as it is assumed that the outgoing traffic distribution reflects destinations' appeal (e.g. in the sense of content, QoS). Others combine gravity model with entropy penalization [7].

TME based on second-order characteristics (2) has gained critical reviews [8] as being inferior to a Bayesian approach, which consists in calculation of conditional probability distributions of flows, given link measurements. However, others [9], [10] have tried to adapt Vardi's original approach to the non-Poissonian nature of real traffic, respectively, by replacing (2) with nonlinear formula and applying a non-stationary model — then, by projecting the problem onto a subspace defined by (1), thus accelerating numerical search for the solution.

The next generation of TME methods started with the observation [11] that the actual traffic can be represented as mixture of a relatively small number of common components (called *eigenflows*) of three types: deterministic, spike and noise. This reduces problem dimension, simultaneously capturing major traffic properties: seasonality, correlation and long tails. Proper construction of the matrices that map eigenflows into real-life traffic flows is in fact principal component analysis (PCA). This technique gets further developed e.g. in [12] to handle common phenomena like traffic anomalies. Alternatively, [13], [14] try to address inherent traffic variability by using neural networks. The initial learning phase requires, however, true flow information to be available. Similarly, [15], [16] postulate to apply partial flow measurements — this time in order to reduce computational complexity of TME in large networks.

III. Proposed approach

Hereby we propose a procedure for traffic matrix identification that takes into account most adverse real-life traffic properties: daily profile, flow correlation and typical far-from-Poisson flow stochastic nature. The procedure has been developed and tested on Abilene network data [17], similarly to a number of works cited above. Abilene flow data were collected for 167 days in 5' periods: they are correlated, cf. Fig. 1a, not similar to Poisson distribution (cf. Fig. 1b), and show a clear daily pattern (cf. Fig. 1c, ragged blue graph).

High correlation of flows results definitely from the fact that they all follow, to various extent, the same daily profile; but not only: higher-layer operations (like VPN) may also be the reason. To make the matters worse, the raw data may be considered dirty, in the sense that unusual long-lasting traffic surges or idle periods can be easily found for some links, indicating errors in measurement or data processing phases, or abnormal activities, e.g. denial-of-service attacks. This issue has already been addressed in [12].

To deal with the daily flow variability, we first need to calculate the typical daily traffic profile, υ, common for all the network. As the processed Abilene data cover a short period, we see no reason to model the trend, i.e. long-term tendency for traffic growth/shrinking. However, to work safely with the daily profile, we had to exclude weekend samples from further analysis, which still gives us 119 days of observations.

Furthermore, days with total of the traffic lying outside the interval defined by standard deviation were treated as outliers, and excluded. This reduced the number of valid days to 103. The daily 5' traffic profile was calculated by averaging corresponding time-of-day samples smoothed over a 1-hour window (cf. Fig. 1c, green smooth line). Usefulness of such profile was verified by applying it, rescaled individually, to known flows so that the total covariance of residual traffic was as small as possible. The residual total traffic was then analyzed, and still showed a fat tail distribution, meaning there were still samples that would not be handled by any finite-variance traffic model. Once again, outliers were hunted, but on much smaller timescale: all five-minute samples outside of 95% confidence interval were replaced by the corresponding data from profile, and normalized. Such operation in fact reshapes the measurement data and limits the modeling scope, but on other hand it curbs traffic variability to limits that are predictable and manageable by the kind of model proposed. Those samples were decided to be dropped off from further analysis, and the daily profile was recalculated, giving the final traffic profile, $\upsilon(t)$, used consistently for all links.

As the flows are currently unknown, we propose to use υ to bring down the interdependence of observed link traffic. However, the flows comprising that observed traffic are unknown yet. So we propose to assume temporarily that each flow is of the form:

$$\psi_i(t) = \alpha_i \upsilon(t) \ , \tag{3}$$

that is, a common profile scaled by individual coefficient. After a close practical examination of subsequent flows, we found that adding a time lag parameter to (3) did not improve the fitting significantly, while doubling the size of model parameter space. By cloning $\psi_i(t)$ onto all days we get a temporary matrix of "samples" $\boldsymbol{\Psi}$ (parametrized by $\boldsymbol{\alpha}$). Now let us find $\boldsymbol{\alpha}$ (i.e., "generate traffic") such that, when subtracted from observed link traffic, leaves residual link traffic correlated as little as possible. Formally, one looks for

$$\boldsymbol{\alpha}^{\star} = \arg\min_{\boldsymbol{\alpha}} || \operatorname{cov}(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{\Psi})||_1 \ . \tag{4}$$

The above procedure does its best to eliminate the deterministic component from unknown flows, taking into account the routing matrix. The search for optimum must be constrained to avoid unreasonable solutions, and should use gradient formula rather than estimation to ensure convergence.

Consequently, whatever the real flows are, the remaining residual traffic observed on links is

$$\tilde{\boldsymbol{Y}} \stackrel{\text{df}}{=} \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{\alpha}^{\star}\upsilon \ . \tag{5}$$

Although somewhat deprived of seasonality, $\tilde{\boldsymbol{Y}}$ is all the same the effect of mixing correlated *and* non-Poissonian flows. To address the latter issue, let us peek at sample distributions of the original flows with daily profile removed:

$$\tilde{\boldsymbol{U}} \stackrel{\text{df}}{=} \boldsymbol{X} - \boldsymbol{\alpha}^{\star}\upsilon \ . \tag{6}$$

Representative histograms of some residual flows and residual flow variances are shown in Fig. 2. They certainly do not resemble Poisson or Gauss distributions, and flow means prevalently are negative. We propose to assume the residential flows to have distribution similar to extreme value (EV) distribution, for practical reasons. EV has two parameters: shift μ and scale σ — which is still twice as many parameters as for Poisson, but we hope it to be a reasonable choice, rather than using more general 3-parameter distributions. EV is widely used to model natural phenomena as rainfall or sea level [18], [19]; in those cases with purpose to estimate respective maximum values. When examining residual flows, we found empirical distributions clearly asymmetric, and we chose to use EV to find values of μ (responsible for low limit of generated traffic) and σ (responsible for variance-to-mean ratio) that modeled the individual flow best. Fig. 3 presents ratios of variance to mean for all nonzero flows. Unlike for Poison process, the ratios are generally much bigger than one, and also highly variable. This motivates usage of EV, however it does not leave a clue about their typical values that would help accelerating the identification.

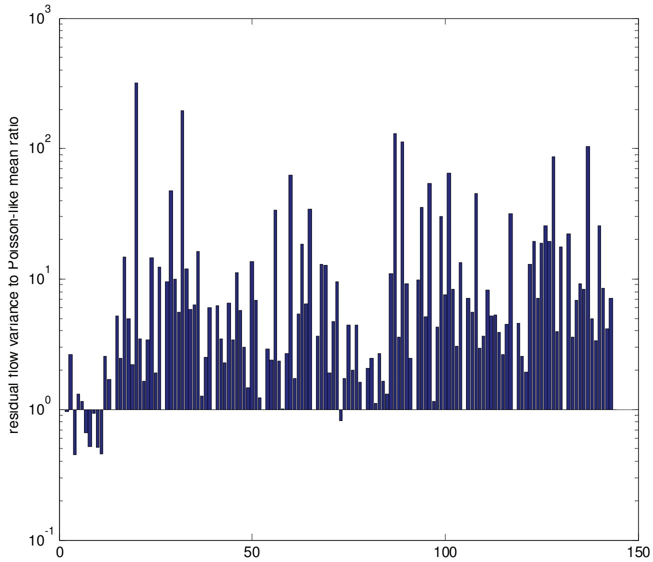


Fig. 3: Ratios of residual flows variance to the flow means.

Assuming flows to be EV-distributed still does not address the issue of flows correlation. Since it was already observed [11] that the network traffic can be accurately represented in practice by a small number of stochastic processes, we propose to do likewise. Let us assume that the network traffic on any link can be expressed as a weighted sum of a small number of truly independent EV random variables, represented by a random vector Φ:

$$\tilde{\boldsymbol{U}} = \boldsymbol{C}\Phi(\boldsymbol{\mu}, \boldsymbol{\sigma}) \ . \tag{7}$$

Matrix $\boldsymbol{C}$ serves to model the phenomenon of traffic flow interdependence, by assigning a fraction c_{ij} of truly independent random j-th EV variable to flow i. Assuming all elements of $\boldsymbol{C}$ to be identified from link samples is unrealistic, one has to decide a priori about $\boldsymbol{C}$ structure and, consequently, the number of nonzero $\boldsymbol{C}$ elements and, consequently, Φ parameters to be identified. While designing its structure and initializing it for optimization, we want to:

- keep the product $\boldsymbol{AC}$ nonsingular — to keep first-moment equations solvable;

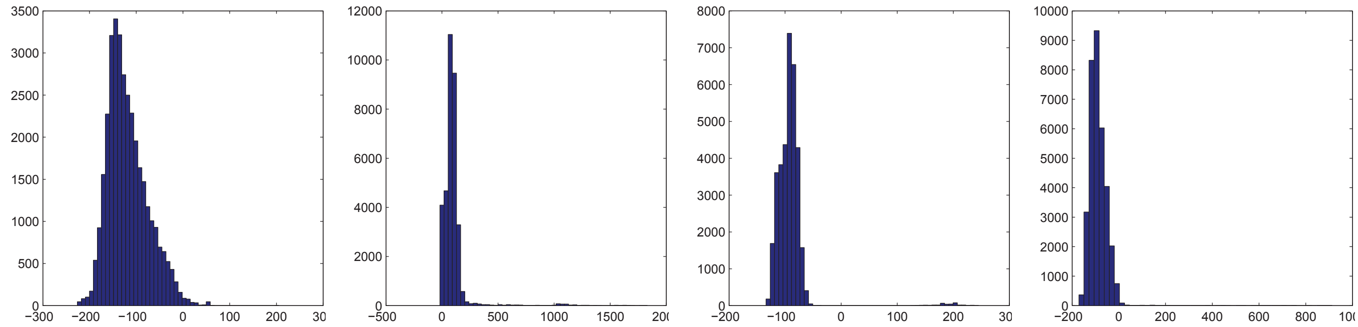


Fig. 2: Sample distributions for biggest residual flows #63, 87, 94, 99.

- keep the number of truly independent flows that compose a real flow small — because superposition of many flows would make the real flow look Gauss-like, which was not observed in reality;
- find reasonable initial values for $\boldsymbol{C}$ — this may greatly accelerate the optimization process.

Let us assume the number of truly independent EV processes to be K, a number slightly bigger than the number of links. In our case, $K = 35$. This is obviously much less than the total number of flows — but we believe that just so few processes determine most of the network traffic, as other authors already observed.

To decide which $\boldsymbol{C}$ elements are to become decision variables, we apply the routine presented in Algorithm 1. The algorithms guarantees assignment of truly independent flows to almost all flows, and good scatter of nonzero coefficients of C (marked with dots in Fig. 4).

After completing the procedure of qualification of $\boldsymbol{C}$ elements as decision variables, we can compile the decision variable vector $\boldsymbol{z} = [\boldsymbol{\sigma}, \boldsymbol{\mu}, c_{i,j:(i,j)\in W}]$, where W stands for indices of nonzero $\boldsymbol{C}$ elements. Analogously to Vardi's approach (1,2) we want to find $\boldsymbol{z}^\star$ such that theoretical 1st and 2nd order properties of the generated and mixed EV traffic correspond to the observed link traffic properties. As for 1st order properties, we propose to match theoretical distribution of a sum of EV variables to the sample distribution,

$$\tilde{\boldsymbol{Y}} = \boldsymbol{AC}\Phi\ , \tag{8}$$

using least squares, weighted by link traffic volume. Technically, it requires convolving selected, rescaled elements of Φ contributing to link traffic. As for 2nd order properties, we minimize LHS-RHS of

$$\operatorname{cov}\tilde{\boldsymbol{Y}} = \boldsymbol{D}(\boldsymbol{D}^T \circ \boldsymbol{\Delta})\ , \tag{9}$$

where $\boldsymbol{D} = \boldsymbol{AC}$ and with $\boldsymbol{\delta}$ being a vector of theoretical variances of independent flows, $\delta_i = \sigma_i^2\pi/6$,

$$\boldsymbol{\Delta} \overset{\mathrm{df}}{=} \overbrace{[\boldsymbol{\delta}, \boldsymbol{\delta}, \ldots, \boldsymbol{\delta}]}^{L}\ .$$

Algorithm 1 Construction and initialization of $\boldsymbol{C}$

1: Initialize $\boldsymbol{C}$ with zeros; initialize index of truly independent flow q with 1.
2: Find unprocessed link pair (i,j) with biggest $||\operatorname{cov}(\tilde{\boldsymbol{Y}}_i, \tilde{\boldsymbol{Y}}_j)||$.
3: Find set of flows F that pass either i or j but not both.
4: **for** every flow $f \in F$ **do**
5:   **if** number of nonzero $\boldsymbol{C}$ elements in row corresponding to f is less than 3 **then**
6:     Initialize those elements with $\frac{1}{n}$ each, n is the number of flows using link i or j, whichever bigger.
7:   **end if**
8: **end for**
9: **if** $\operatorname{rank}(\boldsymbol{AC})$ has not increased since long **then**
10:   Roll back to step 2 and continue from there
11: **end if**
12: **if** all flows are assigned to truly independent flows or $q = K$ **then**
13:   exit
14: **end if**
15: Increment q and proceed to step 2.

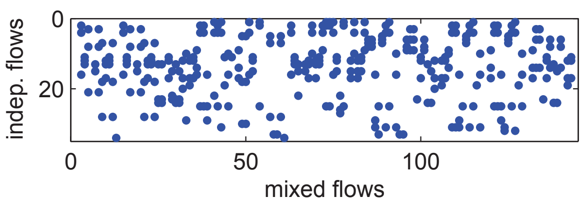


Fig. 4: Location of nonzero elements of traffic mixing matrix $\boldsymbol{C}$.

IV. Results

To find $\boldsymbol{z}^\star$ via minimization of mean square errors in equations (8,9), one must decide upon proportions between

TABLE I: MRE for original and filtered traffic data.

	MRE		
	95%	90%	50%
filtered traffic	1.02	0.87	0.64
original traffic	1.04	0.89	0.67

RHS-LHS imbalance in both systems. We assumed that the sum of mean square errors in (8) counts as is, while analogous sum in (9) is scaled by ratio γ and then added to the goal function. Optimization was performed for a number of γ's, and the results of flow estimation were evaluated using mean $\boldsymbol{X}$ error (MRE) measure [8]. MRE_n (for the n-th percentile of traffic) is the mean relative error for biggest flows that amount to n percent of the total traffic in the network. We found that for a range of $\gamma > 1$ optimization results were practically the same. Table I compares estimation errors for original and filtered traffic data, measured for various percentiles. The errors, in the sense of mean flow values are still quite big, although the luckier half of flows get estimated with error of 60%. This result, being still too high for e.g. energy-efficient network management, can be already useful for rough dimensioning.

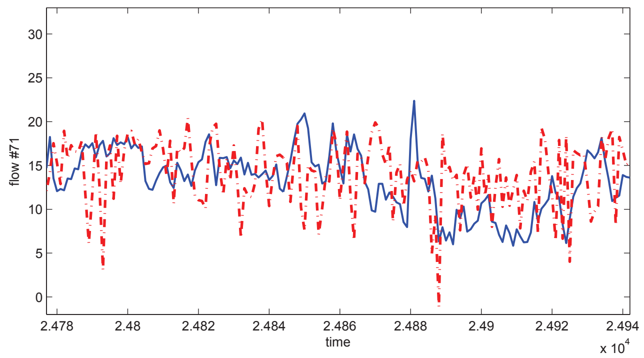


Fig. 5: Real (blue, solid) and estimated (red, dash-dotted) traffic for flow 71.

Surprisingly small difference of errors between filtered and raw data can be explained by comparing time series of estimated vs. real traffic. Many real flows still exhibit traffic surges that passed into identification phase, without being correctly suppressed by the filtering procedure. The proposed model cannot handle them well; the resulting estimated traffic is matched roughly w.r.t. the mean value, which results in permanent overestimation. However, there is a number of flows that get identified quite well in the stochastic sense, cf. e.g. Fig. 5.

Alternatively, we may present the results using cumulative distribution of the spatial relative errors for flows (SRE – cf. [20]). The distributions are given in Fig. 6. The graph has the same shape as in [16], but with bigger errors. However, in this case we can compare only shapes because the mentioned authors carried out identification with detailed measurements of selected flows.

Looking in more detail at estimation errors for flows (cf. Fig. 7), we see that many of the big flows have been identified with ratio well below 1; while this might be considered still an unimpressive result, it is hoped that there still exists possibility for further improvement.

The proposed approach was also applied to another popular test case, the European GÉANT network [21], frequently referred to in the literature. However, the obtained results turned out to be rather disappointing: the error rates went considerably higher than in the Abilene case. Closer examination of the original flows showed their huge variability in

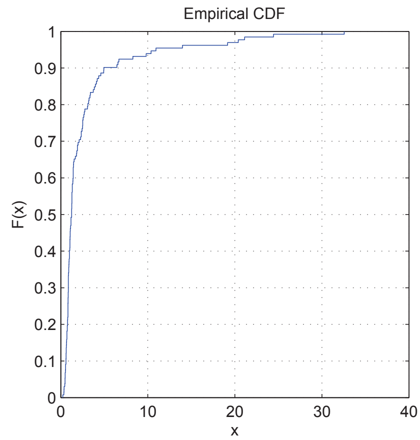


Fig. 6: CDF for spatial relative errors.

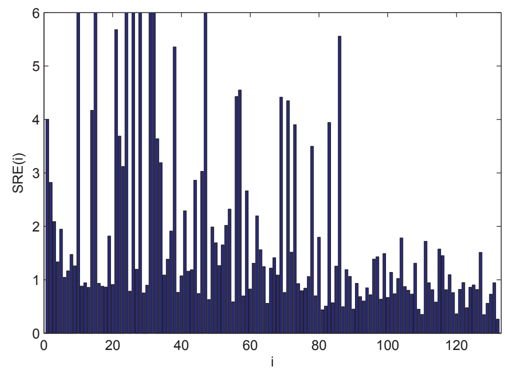


Fig. 7: Relative errors (SRE) for flows (in ascending order of flow volumes).

time; sometimes in the order of magnitude. Consequently, the de-profiling phase did not work correctly, and the subsequent phases failed accordingly. The approach proposed here simply does not handle scenarios with so varying traffic.

V. Conclusions

In work presented here we tried to manage adverse properties of the network traffic (correlation, fat tails, seasonality), while extending as little as possible the original apparatus for first and second moments-based identification of flows. Although the results still qualify as preliminary, as compared to third-generation of methods, we have reasons to claim our approach still to be subject for improvement. Firstly, in the numerical layer: it turned out a number of times that only a change of optimization algorithm or analytical gradient calculation made progress possible. The problem presented here definitely demands a research using versatile optimization algorithms. Secondly, further data preprocessing or taking into account evident anomalies may improve prediction (as emphasized in [12]). Finally, a number of other distribution types may be considered for modeling the original independent flows: in our view, EV does not handle multiple-peak and fat-tail real distributions well. The above remarks may serve as indications for further research.

References

[1] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," DTIC Document, Tech. Rep., 1997.

[2] Y. Chan, E. J. Regan, and W. Pan, "Inferring an origin-destination matrix directly from network flow sampling," *Transportation Planning and Technology*, vol. 11, no. 1, pp. 27–46, 1986. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03081068608717328

[3] E. Niewiadomska-Szynkiewicz, A. Sikora, P. Arabas, and J. Kołodziej, "Control system for reducing energy consumption in backbone computer network," *Concurrency and Computation: Practice and Experience*, vol. 25, no. 12, pp. 1738–1754, 2013.

[4] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 365–377, 1996. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476697

[5] S. Erlander and N. F. Stewart, *The gravity model in transportation analysis: theory and extensions*. Vsp, 1990, vol. 3.

[6] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot, "Traffic matrix estimation: existing techniques and new directions," *SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 4, pp. 161–174, Aug. 2002. [Online]. Available: http://doi.acm.org/10.1145/964725.633041

[7] J. Ni, S. Tatikonda, and E. M. Yeh, "Qrp06-5: A large-scale distributed traffic matrix estimation algorithm," in *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*. IEEE, 2006, pp. 1–5.

[8] A. Gunnar, M. Johansson, and T. Telkamp, "Traffic matrix estimation on a large ip backbone: a comparison on real data," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 149–160. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028807

[9] J. Cao, D. Davis, S. Vander Wiel, and B. Yu, "Time-varying network tomography: Router link data," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1063–1075, 2000. [Online]. Available: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.2000.10474303

[10] I. Juva, S. Vaton, and J. Virtamo, "Quick traffic matrix estimation based on link count covariances," in *Communications, 2006. ICC'06. IEEE International Conference on*, vol. 2. IEEE, 2006, pp. 603–608.

[11] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, "Structural analysis of network traffic flows," in *Proceedings of the joint international conference on Measurement and modeling of computer systems*, ser. SIGMETRICS '04/Performance '04. New York, NY, USA: ACM, 2004, pp. 61–72. [Online]. Available: http://doi.acm.org/10.1145/1005686.1005697

[12] Z. Wang, K. Hu, K. Xu, B. Yin, and X. Dong, "Structural analysis of network traffic matrix via relaxed principal component pursuit," *Computer Networks*, vol. 56, no. 7, pp. 2049–2067, 2012.

[13] D. Jiang, Z. Xu, H. Xu, Y. Han, Z. Chen, and Z. Yuan, "An approximation method of origin–destination flow traffic from link load counts," *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 1106 – 1121, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045790611001029

[14] D. Jiang, Z. Xu, L. Nie *et al.*, "An approximate approach to end-to-end traffic in communication networks," *Chinese Journal of Electronics*, vol. 21, no. 4, pp. 705–710, 2012.

[15] H. Tian, M. Roughan, Y. Sang, and H. Shen, "Diffusion wavelets-based analysis on traffic matrices," in *Parallel and Distributed Computing, Applications and Technologies (PDCAT), 2011 12th International Conference on*. IEEE, 2011, pp. 116–121.

[16] L. Nie, D. Jiang, and L. Guo, "A power laws-based reconstruction approach to end-to-end network traffic," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 898–907, 2013.

[17] SndLib. Demand matrices from measurements in the U.S. Abilene network. [Online]. Available: http://sndlib.zib.de/download/abilene-readme.txt

[18] S. Caires, "Extreme value analysis: Still water level," *JCOMM Technical Report No. 58*.

[19] K. K. Jose and B. Abraham, "A counting process with Gumbel inter-arrival times for modeling climate data," *Journal of Environmental Statistics*, vol. 4, no. 5, 4 2013. [Online]. Available: http://jes.stat.ucla.edu/v04/i05

[20] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot, "Traffic matrices: balancing measurements, inference and modeling," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1. ACM, 2005, pp. 362–373.

[21] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon, "Providing public intradomain traffic matrices to the research community," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 83–86, Jan. 2006. [Online]. Available: http://doi.acm.org/10.1145/1111322.1111341