Unsupervised Traffic Flow Classification Using a Neural Autoencoder

TECHNISCHE UNIVERSITÄT DARMSTADT

Philipps Universität Marburg

Jonas Höchst, Lars Baumgärtner, Matthias Hollick, Bernd Freisleben

Problems and Targets

Challenges in Modern Computer Networks

- Growing popularity of smartphone and tablet usage
- Competing services on mobile devices:
  - Web browsing, Voice-over-IP, video live streaming, ...
  - Real-time, high-bandwidth applications
  - HTTP(s) as main communication channel
- Paradigm shift towards Software-Defined Networking (SDN) and Software-Defined Wireless Networking (SDWN)
- Dynamic flow configurations based on application demands

Targets

- Protocol independent traffic flow classification
- Rely on statistical flow properties, rather than port-identification or deep-packet inspection (DPI)
- Enable online-classification and reclassification
- Enable efficient on-device classification

Related Work

- Bayesian analysis into 10 fixed classes, 65% accuracy [1]
- Comparing different supervised machine learning approaches, including SVMs, up to 97,8% accuracy, using prelabeled traffic [2]
- Semi-supervised learning using K-means, subsequent cluster-labeling [3]
- Unsupervised clustering algorithm based on statistical properties and payload-based clustering [4]

Publications

[1] A. W. Moore and D. Zuev, "Internet traffic classification using Bayesian analysis techniques," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1. ACM, 2005, pp. 50–60.

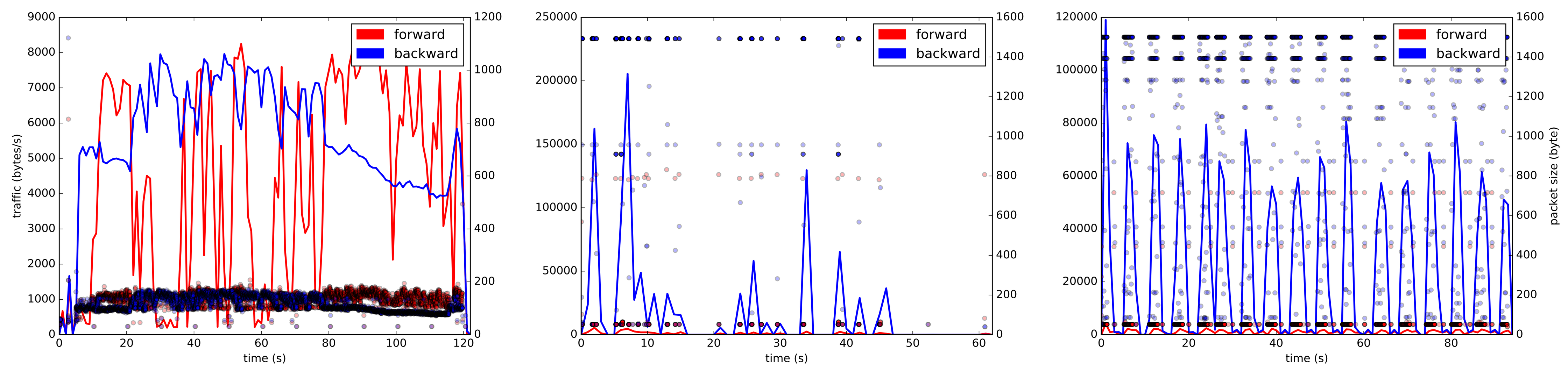
[2] H. Kim, K. C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: myths, caveats, and best practices," in *Proceedings of the 2008 ACM CoNEXT Conference*. ACM, 2008, pp. 11:1–11:12.

[3] J. Erman, A. Mahanti, M. Arlitt, I. Cohen, and C. Williamson, "Semisupervised network traffic classification," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 35, no. 1. ACM, 2007, pp. 369–370.

[4] J. Zhang, Y. Xiang, W. Zhou, and Y. Wang, "Unsupervised traffic classification using flow statistical properties and IP packet payload," *Journal of Comp. and Syst. Sciences*, vol. 79, no. 5, pp. 573–585, 2013.

Traffic Patterns

- Forward and backward traffic (lines)
- Forward and backward packets (dots)
- Typical patterns observable after a short period of time.
- Main differences observable in packet sizes, traffic shapes and inter-arrival times.



(a) Audio call (b) Website interaction (c) Buffered videostream

Methods and Approaches

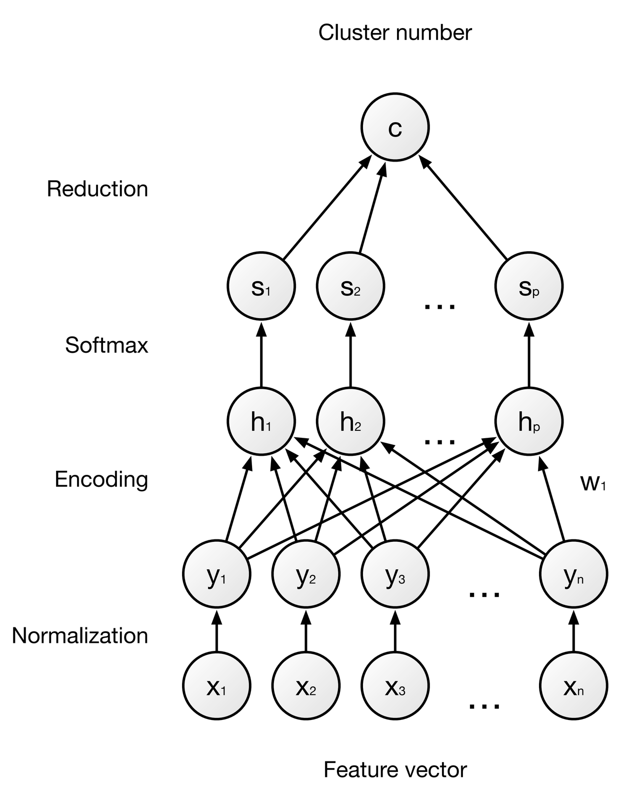
Feature Vector Construction

- Low number of statistical features to reduce computational amounts and memory usage:
  - Number of packets & bytes, avg./stdev./sum of packet sizes, mean DSCP
- Feature computation in forward (client to server) and backward direction
- Snapshots of statistical features after exponentially growing intervals, after 1, 2, 4, 8, 16, … seconds.

Data Clustering using a Neural Autoencoder

1. Feature normalization using standard score
2. Data encoding using the trained autoencoder
3. Apply softmax to raise output contrast
4. Reduction by choosing the index of the greatest element.

Training: Summed squared error combined with the Adaptive Moment Estimator (Adam).



Autolabeling Clusters

1. Clustering flows of equally sized sets per traffic class
2. Assign cluster labels by choosing the label with the highest occurrence in the cluster.

Clustering using massive amounts on unlabeled data.

Classification using a small amount of labeled data.

Class	Principal feature	Example mobile application
Browsing	ephemeral	Wikipedia, Spiegel, Heise
Interactive	long lasting	Online Games, Facebook, Twitter
Download	large downstream	Updates, Dropbox
Livestream	constant bitrate	Streaming, iTunes Webradio
Videostream	periodic buffering	Youtube, Vimeo, Facebook, Twitch
Call	low iat, symmetric	Skype, Apple FaceTime, Google Hangouts, WhatsApp
Upload	large upstream	YouTube, Facebook, WhatsApp

Experimental Evaluation

Aggregation Method

- Using statistical non-cumulative features is 15% better than using cumulative values.

Number of Clusters

- Sweet spot when using 60 clusters – no further improvement using more clusters.

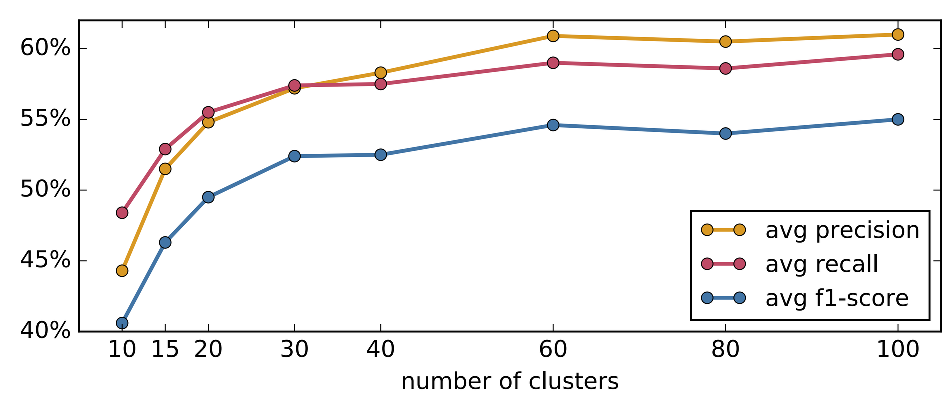


Fig. 2: Classification quality vs. number of clusters

Scaler

- Using a scaler improves the precision and recall to an average of around 60%.

Classification Quality

- Best result: 100 clusters, 30 epochs, standard scaler, full dataset (including UDP and TCP flows):
  - Average precision of **80%**
  - average recall of **75%**
  - F1 Score of **0.76**.

TABLE II: Classification quality

	precision	*recall*	*F1 score*
videostream	0.47	0.80	0.59
upload	1.00	0.85	0.92
livestream	0.86	0.67	0.75
browsing	0.91	0.50	0.65
download	0.80	0.80	0.80
call	0.87	1.00	0.93
interactive	0.71	0.60	0.65
avg/total	**0.80**	**0.75**	**0.76**

Conclusion

- Novel time interval based feature vector and semi-automatic cluster labeling method.
- Clustering independent independent of known traffic classes, classification using limited set of example flows.
- Future Work: a) using deep and stacked Autoencoding, b) improving the SoftMax function to improve Clusters, c) real-time classification.

SFB 1053 MAKI

LOEWE-Schwerpunkt NICER