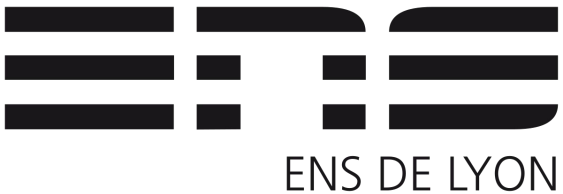
Towards Designing Robust and Efficient Classifiers for Encrypted Traffic in the Modern Internet

Xi Jiang, Shinan Liu, Saloua Naama, Francesco Bronzino, Paul Schmitt, Nick Feamster



Network Traffic Classification v.s. Increasing Encryption

Network Traffic Classification:

- Inference of Internet services and applications.
- Useful for capacity and resource planning, QoS monitoring, traffic prioritization, etc.

Conventional Approach:

- Expert knowledge-based.

ML Approach:

- Shallow and Deep learning.

Why are current encrypted traffic classifiers not enough?

Existing classifiers experience inadequate efficiency due to lack of attention to feature space:

- Increasing encryption alters the classification feature space:
  - Reducing the usefulness of affected features (e.g. DNS encryption)
  - Shifting the feature importance distribution (e.g. TLS v.s. VPN)
- DL models do not address the issue
  - High system overhead
  - Low inference speed
  - Low interpretability

Why are current encrypted traffic classifiers not enough?

Classifiers evaluated using closed-world datasets do not guarantee model transferability.

- Low robustness given newer network traffic across
  - Domain
  - Time
  - etc.
- Prominent datasets:
  - ISCX VPN-NonVPN
  - UNIBS-2009

What are some plausible research directions?

Utilize interpretable models to reduce feature space to improve efficiency.

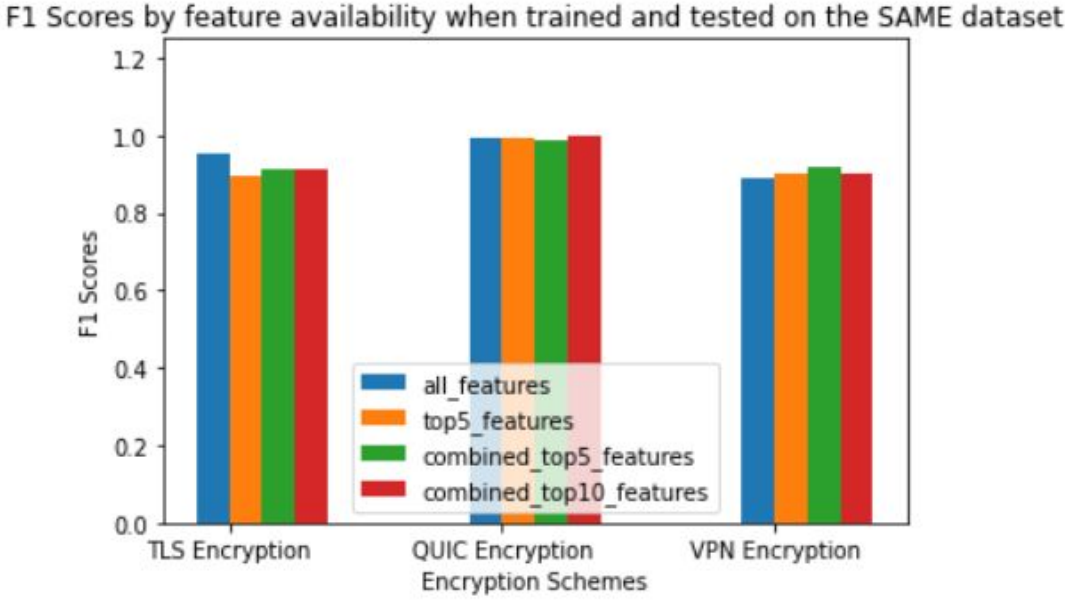
- Reduce the number of features to consider while obtaining reasonably good classification results.
- Simple techniques can be used:
  - SHAP
  - Recursive feature removal based on p-value
  - Etc.
- Lower system cost and higher inference speed

What are some plausible research directions?

Utilize interpretable models to reduce feature space to improve efficiency.

Combined Overall Feature Importance

	header	field	importance
11	ipv4	ttl	0.785778
21	tcp	opt	0.197444
10	ipv4	tos	0.076111
31	udp	len	0.076000
9	ipv4	tl	0.071778
1	ipv4	dfbit	0.061889
29	tcp	wsize	0.052222
17	tcp	doff	0.034889
0	ipv4	cksum	0.020111
19	tcp	fin	0.017778
30	udp	cksum	0.009667
13	tcp	ackf	0.008111
12	ipv4	ver	0.002889
15	tcp	cksum	0.002778



What are some plausible research directions?

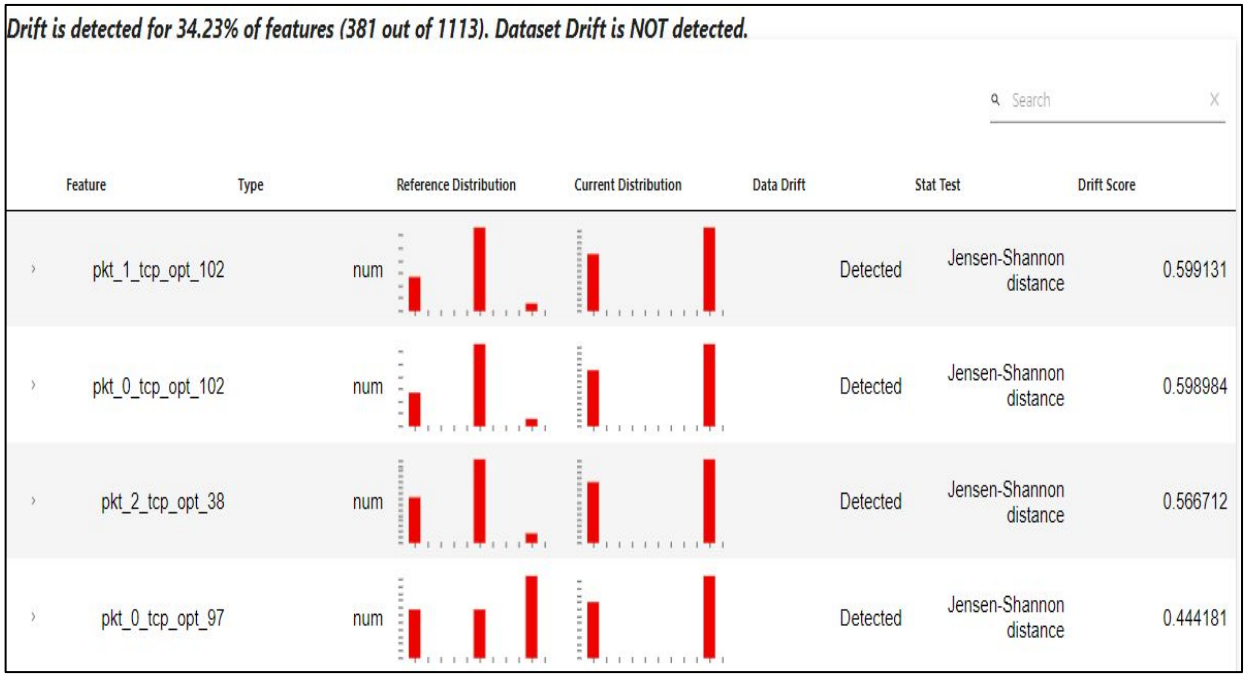
Perform statistical analysis on multiple datasets to locate features robust to improve model transferability.

- Robust features:
  - can achieve similar performance when tested on a new dataset that it has never seen before, i.e., non-environment specific features.
- How do we locate these features?
  - Statistical analysis across datasets.

What are some plausible research directions?

Perform statistical analysis on multiple datasets to locate features robust to improve model transferability.

Bit level drift score calculation:



Fields with MEAN drift scores lower than 0.1 (aggregated and averaged)

	header	field	drift_score
1	ipv4	dfbit	0.048140
3	ipv4	foff	0.006982
4	ipv4	hl	0.006982
5	ipv4	id	0.032307
6	ipv4	mfbit	0.006982
7	ipv4	opt	0.000000
8	ipv4	proto	0.054713
9	ipv4	rbit	0.006982
11	ipv4	tl	0.036799
12	ipv4	tos	0.006982
14	ipv4	ver	0.006982
24	tcp	opt	0.054071

Conclusion

- There exists room for improvement for network traffic classifiers due to
  - Lack of efficiency for practical deployment
  - Low model transferability
- Re-examination of the space is required to design methods that are
  - Robust against encryption
  - Efficient/accurate in classification

Open Question and Comments?