Rosetta:
Enabling Robust TLS Encrypted Traffic Classification in Diverse Network Environments with TCP-Aware Traffic Augmentation

Renjie Xie, Jiahao Cao, Enhuan Dong, Mingwei Xu, Kun Sun
Qi Li, Licheng Shen, Menghao Zhang

清華大學 Tsinghua University
快手
George Mason University
泉城实验室 Quan Cheng Laboratory

TLS Encryption Protocol

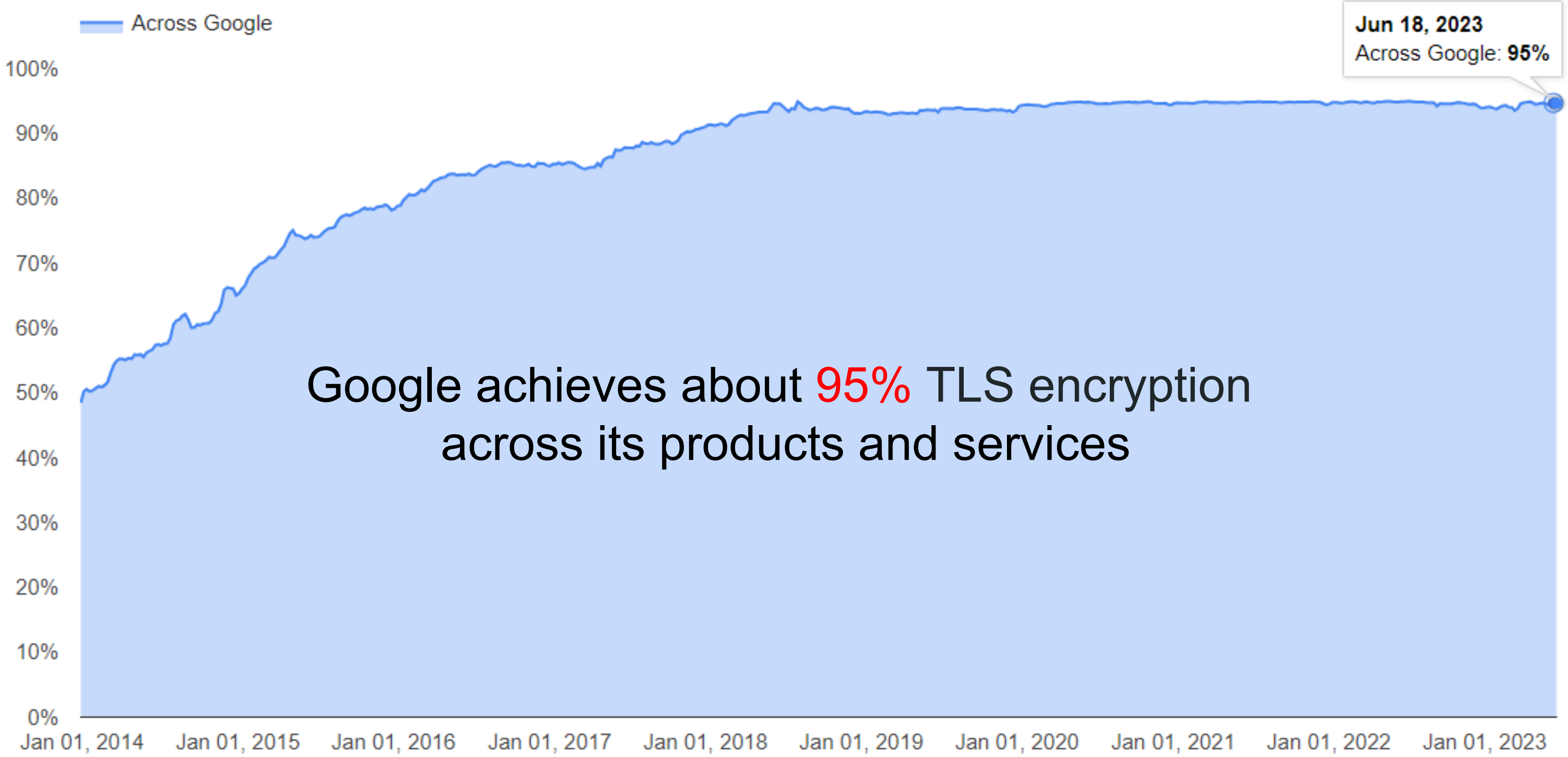
Website

Communications

Online Storage

TLS encryption is widely accepted by various applicaions

TLS Encryption Protocol



Google achieves about 95% TLS encryption across its products and services

Background

TLS encrypted traffic classification provides valuable information for

User Profiling

Intrusion Detection

Network Management

TLS Encryption

HTTP

```
0010  01 84 48 ab 40 00 40 06  b4 fc c0 a8 7b 66 22 6b   ··H·@·@· ····{f"k
0020  dd 52 87 7e 00 50 ef 1a  03 ed a3 81 39 2e 50 18   ·R·~·P·· ····9.P·
0030  fa 18 3d 43 00 00 47 45  54 20 2f 73 75 63 63 65   ··=C··GE T /succe
0040  73 73 2e 74 78 74 3f 69  70 76 34 20 48 54 54 50   ss.txt?i pv4 HTTP
0050  2f 31 2e 31 0d 0a 48 6f  73 74 3a 20 64 65 74 65   /1.1··Ho st: dete
0060  63 74 70 6f 72 74 61 6c  2e 66 69 72 65 66 6f 78   ctportal .firefox
0070  2e 63 6f 6d 0d 0a 55 73  65 72 2d 41 67 65 6e 74   .com··Us er-Agent
0080  3a 20 4d 6f 7a 69 6c 6c  61 2f 35 2e 30 20 28 58   : Mozill a/5.0 (X
0090  31 31 3b 20 55 62 75 6e  74 75 3b 20 4c 69 6e 75   11; Ubun tu; Linu
00a0  78 20 78 38 36 5f 36 34  3b 20 72 76 3a 31 30 39   x x86_64 ; rv:109
00b0  2e 30 29 20 47 65 63 6b  6f 2f 32 30 31 30 30 31   .0) Geck o/201001
00c0  30 31 20 46 69 72 65 66  6f 78 2f 31 31 35 2e 30   01 Firef ox/115.0
00d0  0d 0a 41 63 63 65 70 74  3a 20 2a 2f 2a 0d 0a 41   ··Accept : */*··A
00e0  63 63 65 70 74 2d 4c 61  6e 67 75 61 67 65 3a 20   ccept-La nguage: 
00f0  7a 68 2d 43 4e 2c 7a 68  3b 71 3d 30 2e 38 2c 7a   zh-CN,zh ;q=0.8,z
0100  68 2d 54 57 3b 71 3d 30  2e 37 2c 7a 68 2d 48 4b   h-TW;q=0 .7,zh-HK
0110  3b 71 3d 30 2e 35 2c 65  6e 2d 55 53 3b 71 3d 30   ;q=0.5,e n-US;q=0
0120  2e 33 2c 65 6e 3b 71 3d  30 2e 32 0d 0a 41 63 63   .3,en;q= 0.2··Acc
0130  65 70 74 2d 45 6e 63 6f  64 69 6e 67 3a 20 67 7a   ept-Enco ding: gz
0140  69 70 2c 20 64 65 66 6c  61 74 65 0d 0a 43 6f 6e   ip, defl ate··Con
0150  6e 65 63 74 69 6f 6e 3a  20 6b 65 65 70 2d 61 6c   nection:  keep-al
0160  69 76 65 0d 0a 50 72 61  67 6d 61 3a 20 6e 6f 2d   ive··Pra gma: no-
0170  63 61 63 68 65 0d 0a 43  61 63 68 65 2d 43 6f 6e   cache··C ache-Con
0180  74 72 6f 6c 3a 20 6e 6f  2d 63 61 63 68 65 0d 0a   trol: no -cache··
0190  0d 0a                                              ··
```

HTTPs

```
0010  03 68 72 bf 40 00 2e 06  ba 7e d1 bc 0e 87 c0 a8   ·hr·@·.· ·~······
0020  7b 66 01 bb 9f 82 43 53  9a b8 af db 61 28 50 18   {f····CS ····a(P·
0030  f8 93 fd 21 00 00 54 5c  61 06 30 61 2d 78 d5 99   ···!··T\ a·0a-x··
0040  4e 91 02 d1 ca da a1 67  95 38 67 0a 43 14 d6 eb   N······g ·8g·C···
0050  41 4a e6 a1 93 a2 44 36  be 51 fd 45 ef cd 03 c4   AJ····D6 ·Q·E····
0060  30 f8 de dd 61 d4 b1 fa  57 11 fa 73 bd 1f 47 cf   0···a··· W··s··G·
0070  25 42 88 16 91 0b 86 3b  e5 b0 1d 29 b6 6e 0f 17   %B·····; ···)·n··
0080  75 46 4d 19 3a 82 ca b3  0a f9 56 b2 02 51 97 58   uFM·:··· ··V··Q·X
0090  00 c5 64 d9 1d cd bc e2  e8 66 1c 90 73 d7 01 39   ··d····· ·f··s··9
00a0  f3 56 cd c3 c9 0f 95 d5  03 02 6f 98 d9 e3 04 26   ·V······ ··o····&
00b0  77 73 5d e0 64 ed 44 53  35 dd 01 21 23 16 4e f1   ws]·d·DS 5··!#·N·
00c0  f1 f2 59 35 4e 39 eb c7  e4 17 b7 58 d3 b3 fc fc   ··Y5N9·· ···X····
00d0  04 e5 65 59 e8 8d 2e ea  7d dc 86 b3 4a d6 c6 05   ··eY··.· }···J···
00e0  9c 98 5e b9 0d 9b e5 a2  d9 97 43 99 84 5f 61 6b   ··^····· ··C··_ak
00f0  93 cd 10 3e 92 8d 37 31  73 af b8 a1 a0 3f c6 17   ···>··71 s····?··
0100  7b 5d f1 6d 30 dc 0c 9e  63 16 31 7b ac 3e 99 69   {]·m0··· c·1{·>·i
0110  42 3f e0 18 9d c2 92 8a  cd 25 52 55 bc fb 2c cc   B?······ ·%RU··,·
0120  f8 47 e0 f8 05 61 15 ca  62 a2 7f dd 97 f5 53 9a   ·G···a·· b·····S·
0130  8b 27 cd ea 5a 09 a7 8d  b6 9f 8d 8d ca 3d e3 fc   ·'··Z··· ·····=··
0140  e0 f1 e8 3c f7 2d bf d3  7d 35 e4 98 45 0a c6 34   ···<·-·· }5··E··4
0150  1b 13 9d b0 0f bb 99 0d  88 cc 0a d5 1d 3f 27 c4   ········ ·····?'·
0160  1d f4 b0 a2 5b 8c 5f 95  ec 90 a6 69 b6 70 bd b5   ····[·_· ···i·p··
0170  ac ec db 7b c4 13 72 e5  5d be a4 91 28 91 8e eb   ···{··r· ]···(···
0180  cd ca df fb f4 81 0b 7d  66 fd da 9a a5 66 00 4a   ·······} f····f·J
```

TLS Encryption

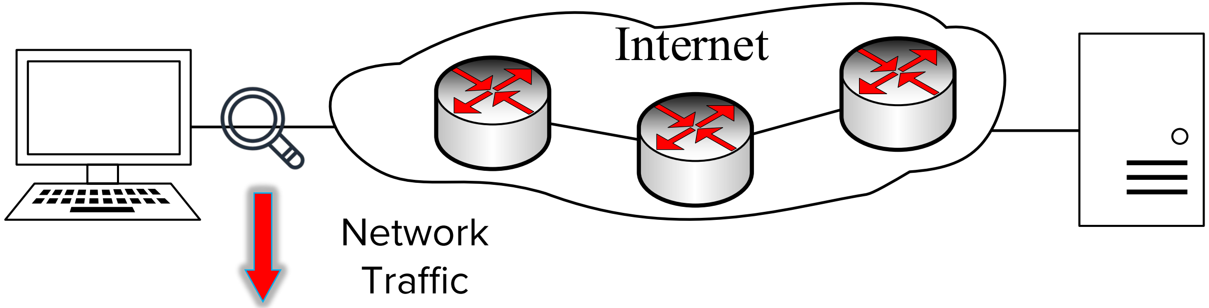
Classification based on Payload ✔

Classification based on Payload ✘

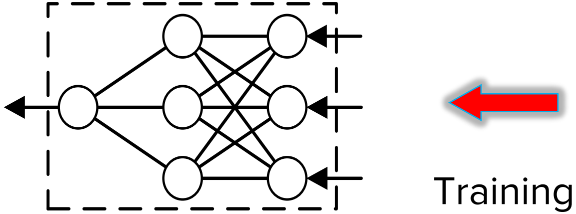
Payload become unrecognized after being encrypted by TLS

Deep Learning Models in TLS Traffic Classification

- TLS Traffic classification on **packet length sequence**



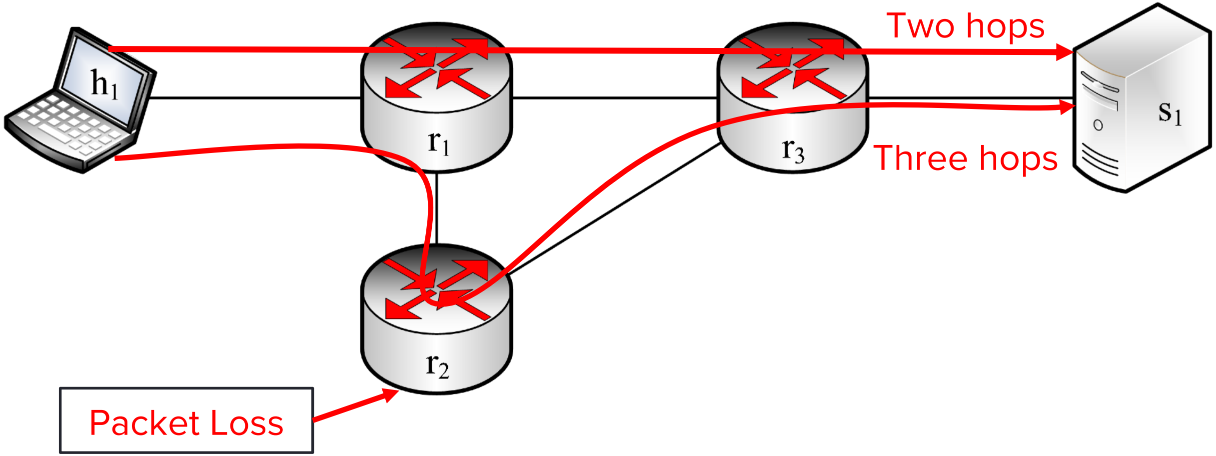
Amazon	517, 51, 163, 46, 86, 77, 168, ...
Facebook	517, 51, 163, 380, 161, 38, 46, ...
Google	190, 126, 485, 484, 485, 764, ...



Deep Learning Models

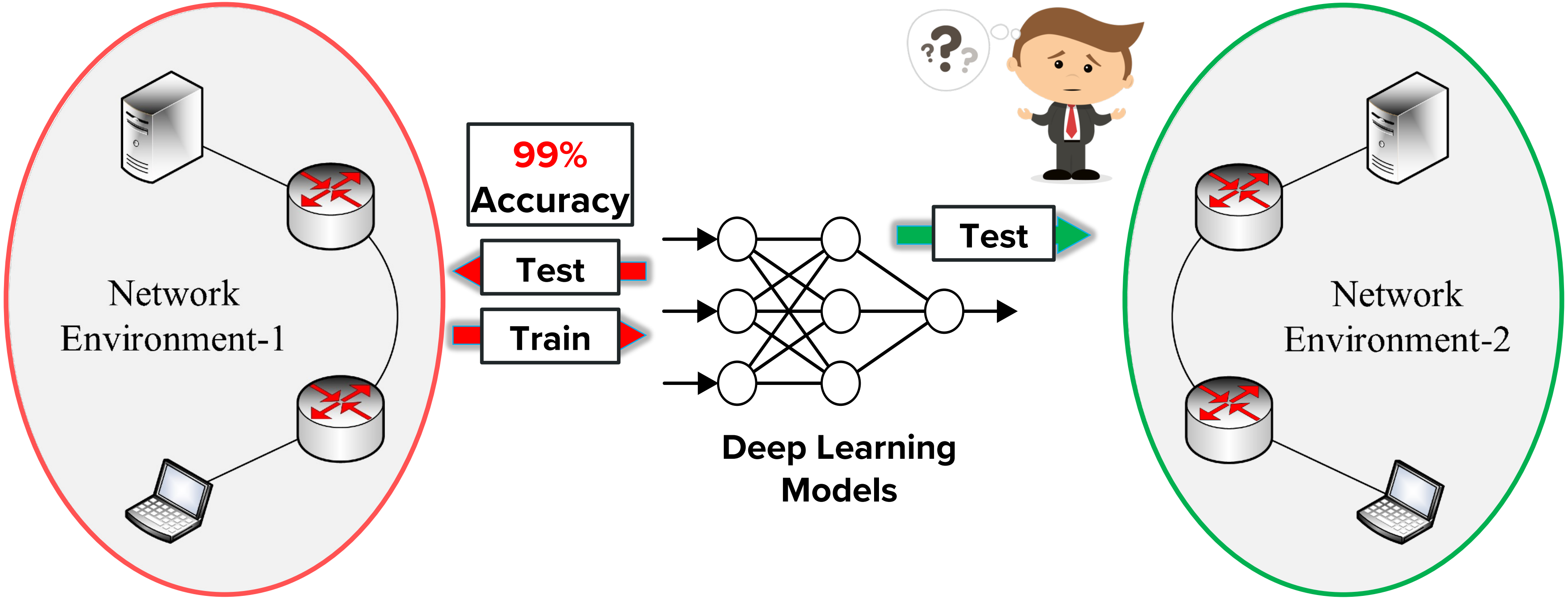
Diverse Network Environments

- Traffic may be affected by network environments in practice
  - Packet Loss
  - Routing Path



Performance in Diverse Network Environments

- DL-based classification in diverse network environments



Experimental Setup for Replayed TLS traffic

- Diverse Network Environments Construction
  - Location and Access mode
- Replayed Dataset
  - CIRA-CIC-DoHBrw-2020
- Models
  - CNN, LSTM, SDAE, DF, FS-Net, Transformer

Different Network Enviroments for **Replayed** Traffic

Network Type	Env. ID	Sender Loc.	Receiver Loc.	Access mode
Wired	θ_0	Local LAN	Local LAN	Ethernet
	θ_1	China	China	
	θ_2	Korea	China	
	θ_3	USA	China	
Wireless	θ_4	China	China	Wi-Fi
	θ_5	China	China	4G LTE
	θ_6	China	China	3G WCDMA

Evaluation on Various Deep Learning Models

- Mainstream deep learning models in Replayed Traffic (Trained in θ_0)

Model	Different Wired Network Environments								Different Wireless Access Network Environments					
	θ_0		θ_1		θ_2		θ_3		θ_4		θ_5		θ_6	
	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1
CNN	99.89%	99.84%	98.21%	98.20%	53.16%	34.91%	57.04%	36.32%	87.47%	87.03%	74.42%	71.52%	53.26%	34.96%
SDAE	95.47%	95.46%	91.47%	91.47%	56.21%	43.40%	55.75%	36.04%	88.11%	88.03%	82.11%	81.42%	55.16%	41.73%
LSTM	95.26%	95.25%	87.68%	87.47%	53.05%	35.07%	57.04%	36.57%	82.00%	81.19%	70.84%	67.34%	53.58%	36.08%
DF	99.89%	99.84%	98.42%	98.41%	53.26%	34.75%	58.03%	36.72%	88.00%	87.57%	74.95%	72.17%	53.37%	35.00%
FS-Net	92.11%	92.10%	90.74%	90.71%	61.16%	52.11%	58.10%	39.66%	88.84%	88.76%	83.68%	83.30%	56.84%	44.50%
Transformer	99.56%	99.36%	98.28%	96.00%	62.22%	54.12%	57.04%	42.00%	93.74%	91.35%	85.62%	83.12%	54.27%	47.57%
On Average	**97.03%**	**96.98%**	**94.13%**	**93.71%**	**56.51%**	**42.39%**	**57.17%**	**37.89%**	**88.03%**	**87.32%**	**78.60%**	**76.48%**	**54.41%**	**39.97%**

Baseline

avg accuracy: -39.86%

avg accuracy: -42.68%

Experimental Setup for Real TLS Traffic

- Diverse Network Environments Construction
  - Location and Access mode
- Traffic datasets
  - Website traffic dataset:
    - 1.8 million TLS flows from 12 websites
- Models:
  - CNN, LSTM, SDAE, DF, FS-Net, Transformer

Different Network Enviroments for **Real** TLS Traffic

Network Type	Env. ID	Client Loc.	Access mode
Wired	τ_1	China	Ethernet
	τ_2	Korea	
	τ_3	USA	
Wireless	τ_4	China	Wi-Fi
	τ_5	China	4G LTE
	τ_6	China	3G WCDMA

Evaluation on Various Deep Learning Models

- Mainstream deep learning models in real website traffic (Trained in τ_1)

Model	Different Wired Network Environments						Different Wireless Access Network Environments					
	τ_1		τ_2		τ_3		τ_4		τ_5		τ_6	
	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1
CNN	89.55%	89.28%	81.48%	80.88%	57.73%	52.29%	72.51%	68.51%	67.16%	60.15%	70.63%	68.73%
SDAE	82.37%	79.79%	78.13%	74.79%	70.04%	68.80%	68.04%	67.98%	64.57%	64.20%	69.94%	64.01%
LSTM	81.85%	77.39%	76.72%	74.08%	62.71%	57.26%	60.89%	60.04%	66.93%	63.60%	66.41%	61.67%
DF	91.27%	91.15%	83.95%	80.58%	83.59%	83.50%	79.90%	75.00%	70.67%	66.91%	73.03%	70.17%
FS-Net	85.81%	81.42%	73.02%	72.20%	64.42%	61.97%	70.14%	68.39%	64.84%	65.42%	67.65%	66.48%
Transformer	84.85%	82.13%	70.97%	69.57%	62.66%	58.46%	63.71%	62.14%	78.98%	75.38%	61.37%	59.74%
On Average	**85.95%**	**83.53%**	**77.38%**	**75.35%**	**66.86%**	**63.71%**	**69.20%**	**67.01%**	**68.86%**	**65.94%**	**68.17%**	**65.13%**

Baseline

avg accuracy: -19.09%

avg accuracy: -17.78%

Understanding Performance Degradation

- Three phenomenons observed in diverse network environments
  - Phenomenon-I: packet subsequence shift (caused by packet loss)

θ_0:	1 190	2 126	3 485	4 484	5 485	6 26	7 54
θ_5:	1 190	3 485	2 126	6 26	7 54	4 484	5 485

  - Phenomenon-II: packet subsequence duplication (caused by packet loss)

θ_0:	1 517	2 51	3 163	4 38	5 46		
θ_1:	1 517	2 51	3 163	4 38	5 46	3 163	4 38

  - Phenomenon-III: packet size variation (caused by delay variation)

θ_0:	1 517	2 51	3 163	4 380	5 161	6 48	7 46
θ_3:	1 517	2 514	3 175				

How to enable robust traffic classification in various environments?

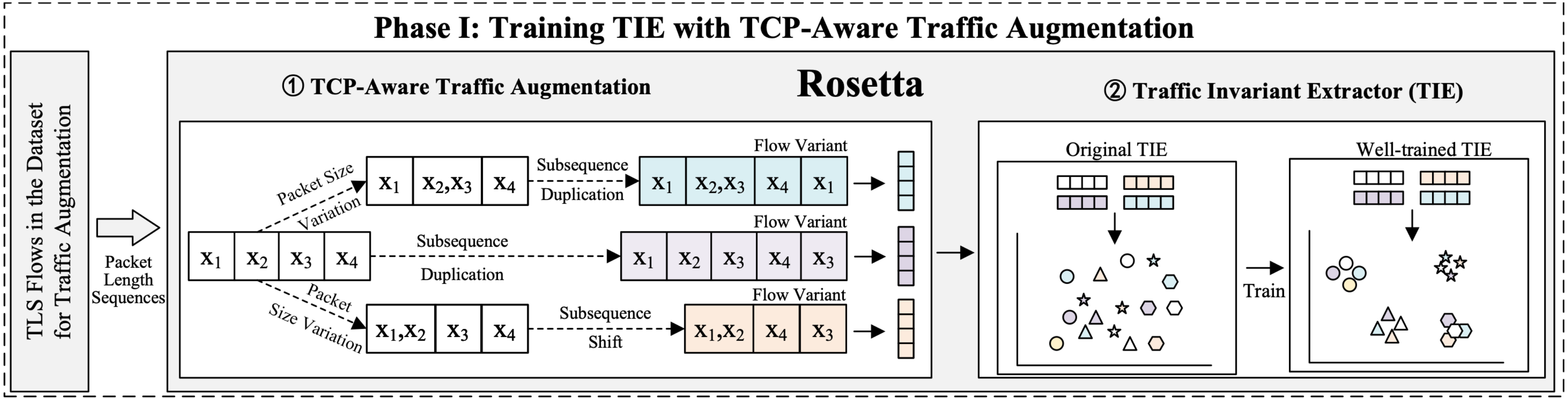
- Contribution I:

Make deep learning models aware of these regular packet sequence changes with TCP semantics.

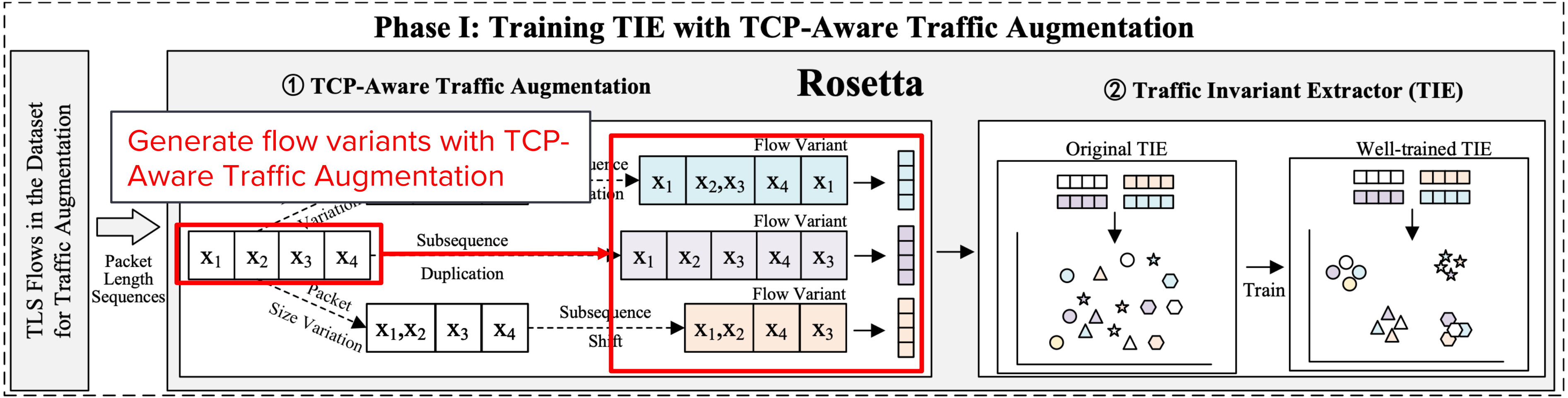
- Contribution II:

Extract robust features from flows for traffic classification.

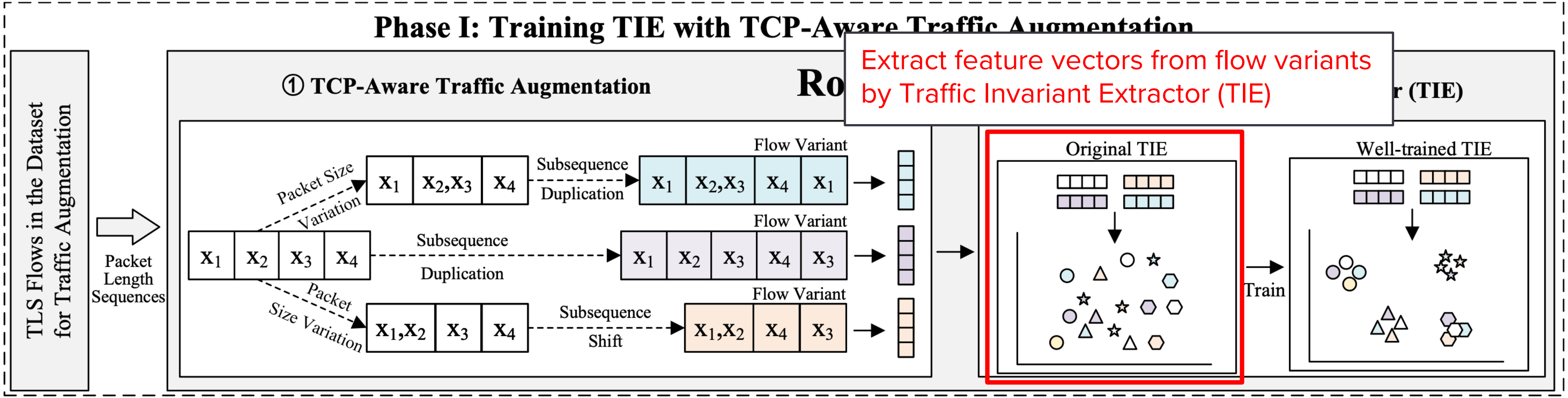
Rosetta Overview



Rosetta Overview

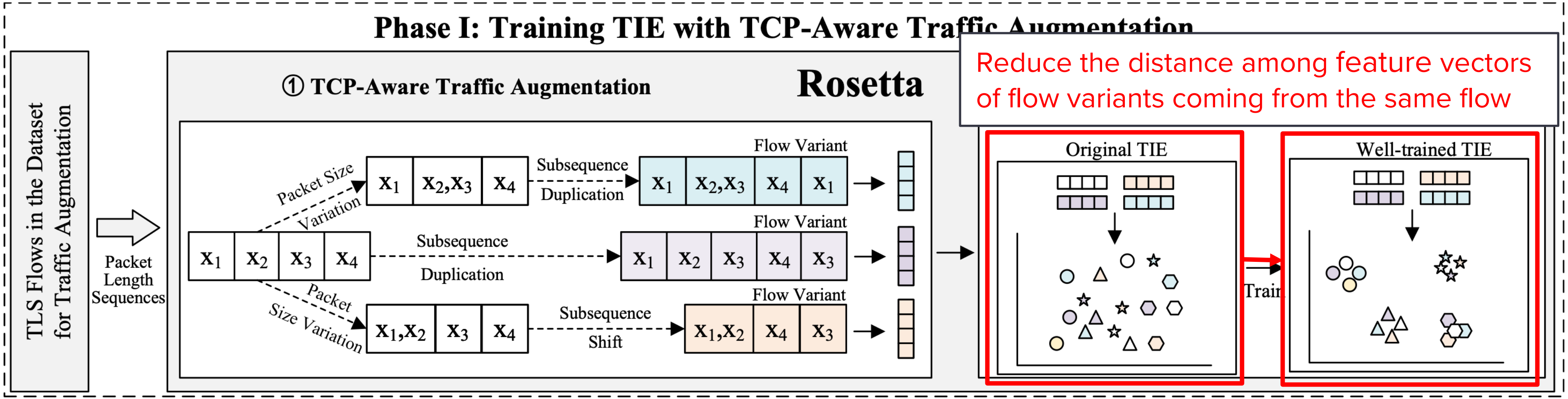


Rosetta Overview

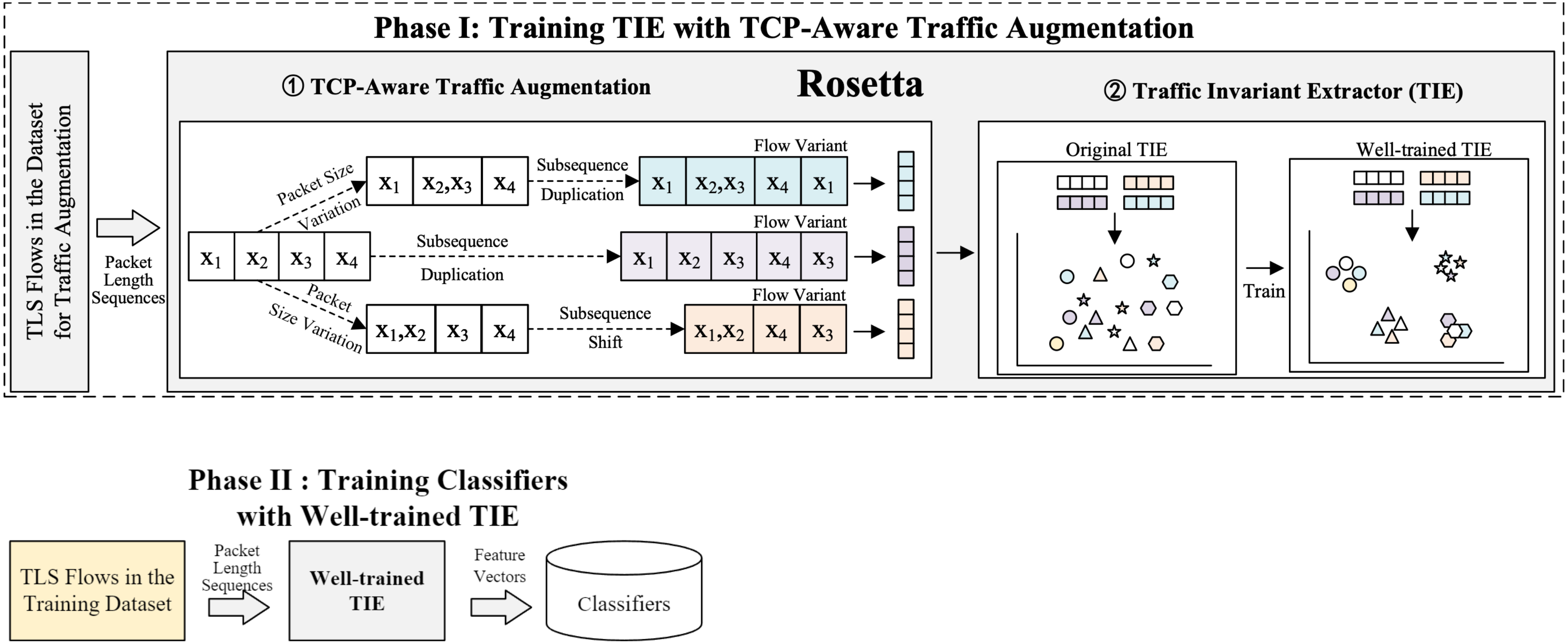


Extract feature vectors from flow variants by Traffic Invariant Extractor (TIE)

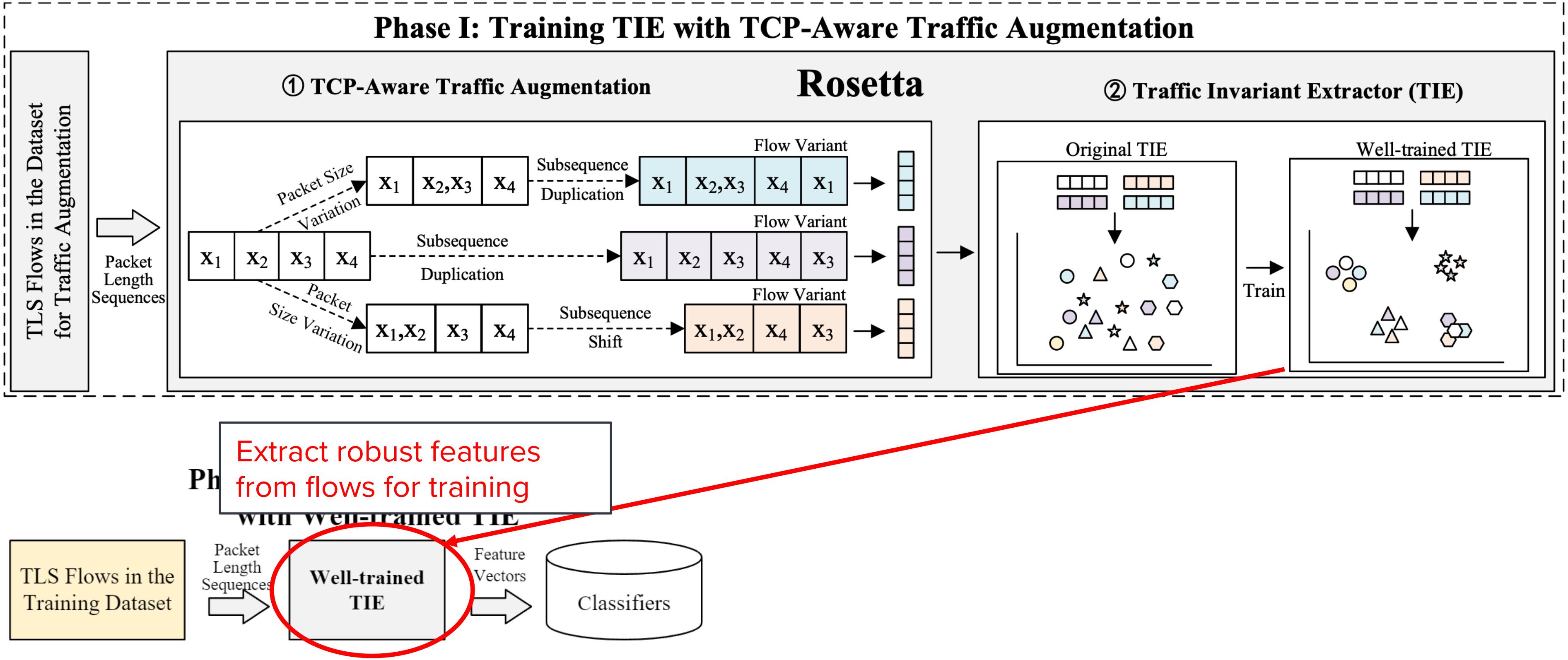
Rosetta Overview



Rosetta Overview

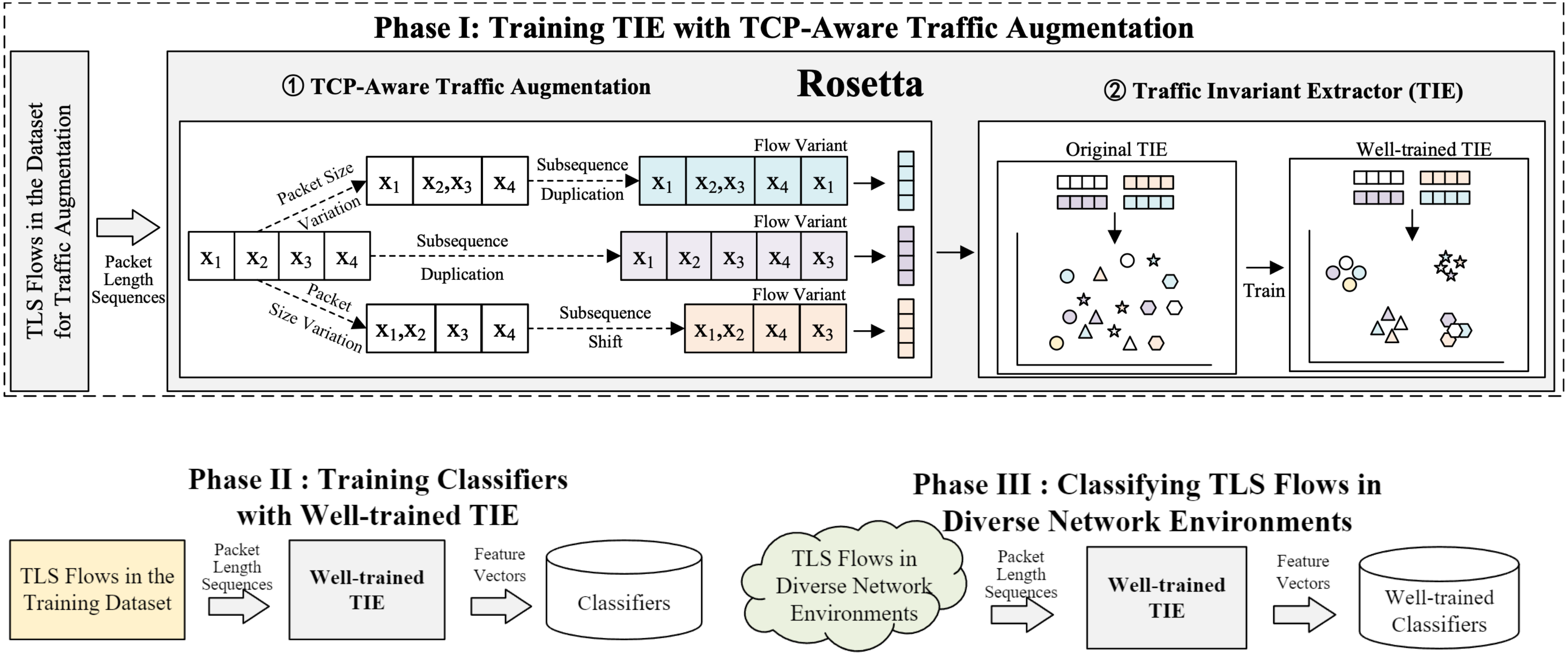


Rosetta Overview

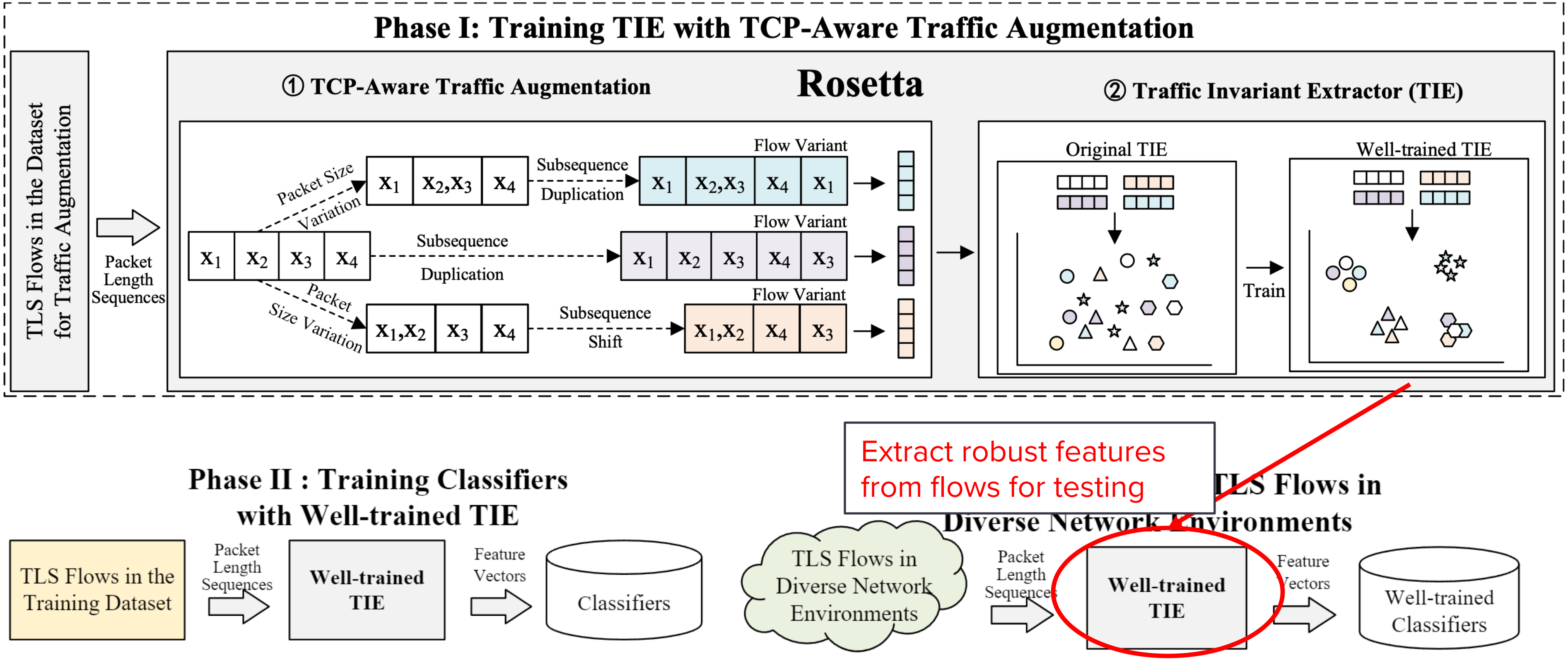


Extract robust features from flows for training

Rosetta Overview



Rosetta Overview



TCP-Aware Traffic Augmentation

- Three types of traffic augmentation algorithms
  - Packet Subsequence Duplication Augmentation
    - Fast retransmit and RTO
  - Packet Subsequence Shift Augmentation
    - Fast retransmit and RTO
  - Packet Size Variation Augmentation

TCP-Aware Traffic Augmentation

An example of Packet Subsequence Duplication Augmentation via Fast Retransmit

Original sequence:	100	200	300	400	500	600	
Augmented sequence:	100	200	300	200	400	500	600

TCP-Aware Traffic Augmentation

An example of Packet Subsequence Duplication Augmentation via RTO

Original sequence:	100	200	300	400	500	600		
Augmented sequence:	100	200	300	400	200	300	500	600

TCP-Aware Traffic Augmentation

An example of Packet Subsequence Shift Augmentation via Fast Retransmit

Original sequence:	100	200	300	400	500	600
Augmented sequence:	100	300	200	500	600	400

TCP-Aware Traffic Augmentation

An example of Packet Subsequence Shift Augmentation via RTO

Original sequence:	100	200	300	400	500	600
Augmented sequence:	100	400	200	300	500	600

TCP-Aware Traffic Augmentation

An example of Packet Size Variation Augmentation

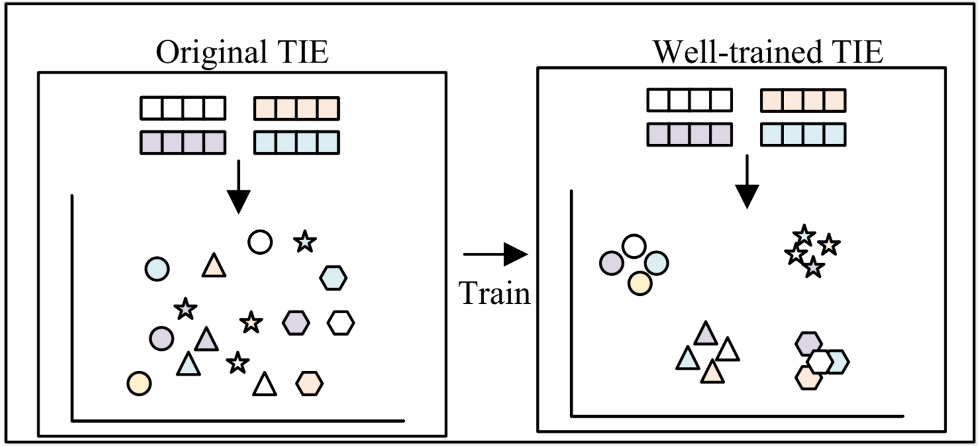
Original sequence:	100	200	300	400	500	600
Augmented sequence:	100	500	400	500	600	

Traffic Invariant Extractor (TIE)

- Loss Function of TIE

$$\mathcal{L}_{\alpha,\zeta} = \| \overline{p_\alpha}(m_\alpha) - \overline{m'_\zeta} \|_2^2 = 2 - 2 \cdot \frac{< p_\alpha(m_\alpha), m'_\zeta >}{\| p_\alpha(m_\alpha) \|_2 \cdot \| m'_\zeta \|_2}$$

- Robust Feature Extraction



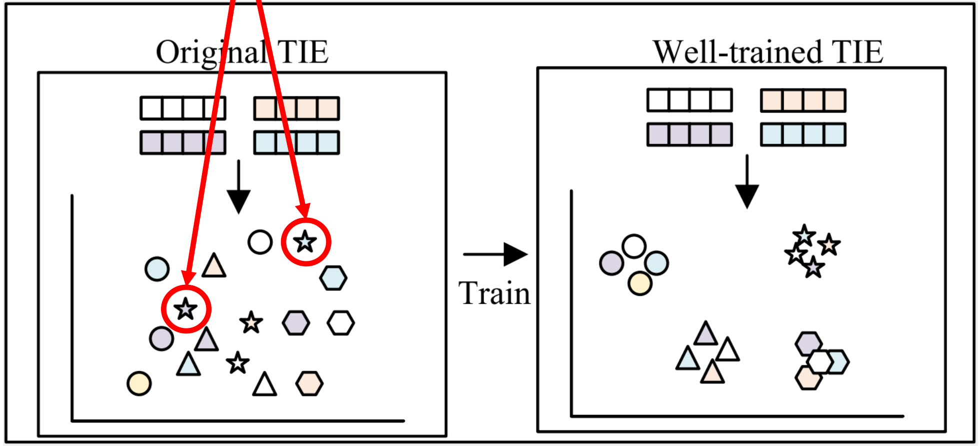
Traffic Invariant Extractor (TIE)

- Loss Function of TIE

$$\frac{< p_\alpha(m_\alpha), m'_\zeta >}{\cdots \| m'_\zeta \|_2}$$

Large distance among feature vectors of flow variants coming from the same flow

- Robust Feature Extraction



Traffic Invariant Extractor (TIE)

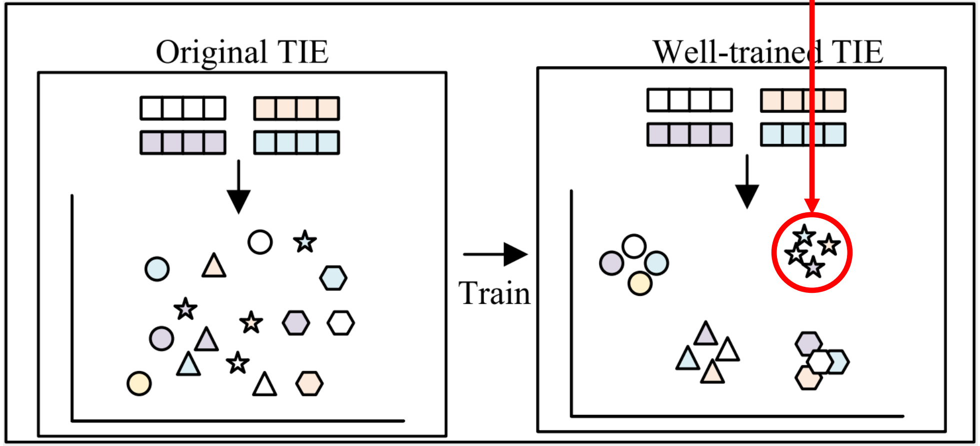
- Loss Function of TIE

$$\mathcal{L}_{\alpha,\zeta} = \| \overline{p_\alpha}(m_\alpha) - \overline{m} \ldots$$

$$< p_\alpha(m_\alpha), m'_\zeta >$$

Small distance among feature vectors of flow variants coming from the same flow

- Robust Feature Extraction



Evaluation with Rosetta

- Improvement on replayed traffic

Model	Different Wired Network Environments					
	θ_1		θ_2		θ_3	
	AC	F1	AC	F1	AC	F1
CNN + Rosetta	93.05% (↓5.16%)	93.03%(↓5.17%)	82.00% (↑28.84%)	81.78%(↑46.87%)	83.72% (↑26.68%)	82.85%(↑46.53%)
SDAE + Rosetta	91.89% (↑0.42%)	91.77%(↑0.30%)	86.63% (↑30.42%)	86.63%(↑43.23%)	84.17% (↑28.42%)	83.69%(↑47.65%)
LSTM + Rosetta	86.63% (↓1.05%)	84.03%(↓3.44%)	79.89% (↑26.84%)	78.32%(↑43.25%)	82.00% (↑24.96%)	78.98%(↑42.41%)
DF + Rosetta	94.42% (↓4.00%)	94.39%(↓4.02%)	86.63% (↑33.37%)	86.63%(↑51.88%)	86.01% (↑27.98%)	85.83%(↑49.11%)
FS-Net + Rosetta	89.26% (↓1.48%)	89.12%(↓1.59%)	84.63% (↑23.47%)	84.47%(↑32.37%)	84.17% (↑26.07%)	83.50%(↑43.84%)
Transformer + Rosetta	94.11%(↓4.17%)	93.74%(↓2.26%)	84.11% (↑21.89%)	83.60%(↑29.48%)	83.37% (↑26.33%)	80.38%(↑38.38%)
On Average	**91.56%(↓2.57%)**	**91.01%(↓2.70%)**	**83.98%(↑27.47%)**	**83.57%(↑41.18%)**	**83.91%(↑26.74%)**	**82.54%(↑44.65%)**

Significant improvement

Model	Different Wireless Access Network Environments					
	θ_4		θ_5		θ_6	
	AC	F1	AC	F1	AC	F1
CNN + Rosetta	89.05% (↑1.58%)	88.93%(↑1.90%)	85.37% (↑10.95%)	85.08%(↑13.55%)	80.42% (↑27.16%)	80.37%(↑45.41%)
SDAE + Rosetta	89.89% (↑1.78%)	89.74%(↑1.71%)	83.47% (↑1.36%)	82.95%(↑1.52%)	81.89% (↑26.73%)	81.88%(↑40.15%)
LSTM + Rosetta	85.37% (↑3.37%)	82.34%(↑1.15%)	82.53% (↑11.69%)	78.22%(↑10.87%)	76.53% (↑22.95%)	73.42%(↑37.33%)
DF + Rosetta	86.84% (↓1.16%)	86.53%(↓1.05%)	82.11% (↑7.16%)	81.31%(↑9.14%	82.63% (↑29.26%)	82.57%(↑47.57%)
FS-Net + Rosetta	85.58% (↓3.26%)	85.16%(↓3.60%)	84.42% (↑0.74%)	83.89%(↑0.60%)	77.26% (↑20.42%)	76.97%(↑32.47%)
Transformer + Rosetta	90.74% (↓3.00%)	89.81%(↓1.54%)	89.16% (↑3.54%)	88.20%(↑5.08%)	81.47% (↑27.20%)	79.63%(↑32.06%)
On Average	**87.91%(↓0.12%)**	**87.09%(↓0.24%)**	**84.51%(↑5.91%)**	**83.27%(↑6.79%)**	**80.03%(↑25.62%)**	**79.14%(↑39.17%)**

Significant improvement

Evaluation with Rosetta

- Improvement on real website traffic

Model	Different Wired Network Environments					
	τ_1		τ_2		τ_3	
	AC	F1	AC	F1	AC	F1
CNN + Rosetta	86.63%(↓2.92%)	86.06%(↑4.19%)	84.83%(↑3.35%)	81.33%(↑0.45%)	91.04%(↑33.31%)	91.04%(↑38.75%)
SDAE + Rosetta	84.67%(↑2.30%)	81.54%(↑12.50%)	85.54%(↑7.41%)	84.49%(↑9.70%)	89.47%(↑19.43%)	85.87%(↑17.07%)
LSTM + Rosetta	84.17%(↑2.32%)	82.07%(↑5.48%)	76.01%(↓0.71%)	74.13%(↑0.05%)	88.52%(↑25.81%)	88.14%(↑30.89%)
DF + Rosetta	90.37%(↓0.90%)	90.10%(↑5.43%)	85.19%(↑1.24%)	81.00%(↑0.42%)	90.15%(↑6.56%)	90.14%(↑6.64%)
FS-Net + Rosetta	86.99%(↑1.18%)	86.47%(↑6.66%)	84.83%(↑11.81%)	76.24%(↑4.04%)	88.41%(↑23.99%)	88.40%(↑26.43%)
Transformer + Rosetta	90.02%(↑5.17%)	87.93%(↑1.74%)	85.36%(↑14.39%)	81.37%(↑11.80%)	89.70%(↑27.04%)	89.69%(↑31.23%)
On Average	**87.14%(↑1.19%)**	**85.69%(↑2.17%)**	**83.63%(↑6.25%)**	**79.76%(↑4.41%)**	**89.55%(↑22.69%)**	**88.88%(↑25.17%)**

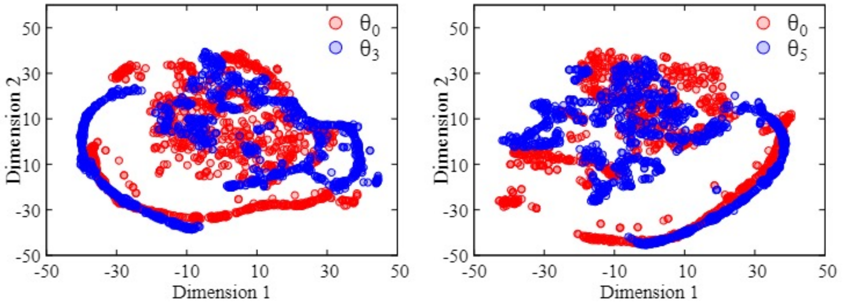
Significant improvement

Improved in all the wired networks

Model	Different Wireless Access Network Environments					
	τ_4		τ_5		τ_6	
	AC	F1	AC	F1	AC	F1
CNN + Rosetta	77.24%(↑4.73%)	75.02%(↑6.51%)	83.58%(↑16.42%)	82.38%(↑22.23%)	75.92%(↑5.29%)	70.66%(↑1.93%)
SDAE + Rosetta	79.10%(↑11.06%)	77.54%(↑9.56%)	74.31%(↑9.74%)	71.18%(↑6.98%)	71.95%(↑2.01%)	65.80%(↑1.79%)
LSTM + Rosetta	69.28%(↑8.39%)	79.53%(↑19.49%)	75.16%(↑8.23%)	74.37%(↑10.77%)	69.59%(↑3.18%)	62.84%(↑1.17%)
DF + Rosetta	84.13%(↑4.23%)	81.67%(↑6.67%)	84.19%(↑13.52%)	80.48%(↑13.56%)	77.58%(↑4.55%)	76.10%(↑5.93%)
FS-Net + Rosetta	77.95%(↑7.81%)	74.79%(↑6.40%)	75.95%(↑11.11%)	72.87%(↑7.45%)	72.83%(↑5.18%)	67.44%(↑0.96%)
Transformer + Rosetta	76.84%(↑13.13%)	74.80%(↑12.66%)	75.66%(↓3.32%)	72.60%(↓2.78%)	76.60%(↑15.23%)	70.23%(↑10.49%)
On Average	**77.42%(↑8.23%)**	**77.22%(↑10.21%)**	**78.14%(↑9.28%)**	**75.64%(↑9.70%)**	**74.08%(↑5.91%)**	**68.85%(↑3.71%)**

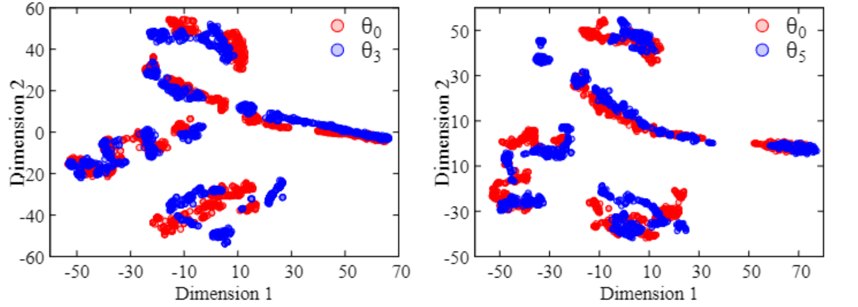
Improved in all the wireless networks

Feature visualization in 2D space



(a) Distribution between θ_0 and θ_3. (b) Distribution between θ_0 and θ_5.

Feature vectors extracted by DF



(a) Distribution between θ_0 and θ_3. (b) Distribution between θ_0 and θ_5.

Feature vectors extracted by Rosetta

Evaluation on Traffic Augmentation Algorithms

- Compare with other data augmentation methods
  - Random Mask (RM) and Random Shift (RS) in NLP
  - Model: DF

Data Aug.	Different Wired Network Environments								Different Wireless Access Network Environments						On Average	
	θ_0		θ_1		θ_2		θ_3		θ_4		θ_5		θ_6			
	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1	AC	F1
RM [17]	97.89%	97.80%	89.47%	88.12%	53.26%	11.56%	58.03%	16.47%	78.00%	71.72%	61.58%	34.00%	52.84%	14.44%	70.15%	47.73%
RS [60]	99.79%	99.77%	86.42%	83.09%	56.26%	16.16%	56.13%	21.84%	77.47%	68.53%	58.53%	20.88%	53.16%	16.74%	69.68%	46.72%
Ours	95.16%	95.14%	94.42%	94.39%	86.63%	86.63%	86.01%	85.83%	86.84%	86.53%	82.11%	81.31%	82.63%	82.57%	**87.69%**	**87.49%**

Better than other data augmentation methods in most networks

Conclusion

- Mainstream DL models cannot robustly classify TLS encrypted traffic in different network environments.

- Rosetta enables robust TLS encrypted traffic classification by
  - TCP-aware traffic augmentation
  - Traffic invariant extractor

- We improve the encrypted traffic classification performance of existing DL models for replayed and real network traffic.

Thank you and Questions?

Contact:

xrj21@mails.tsinghua.edu.cn