An Input-Agnostic Hierarchical Deep Learning Framework for Traffic Fingerprinting

Jian Qu  
Xi’an Jiaotong University

Xiapu Luo  
Hong Kong Polytechnic University

Zhenhua Li  
Tsinghua University

Xiaobo Ma  
Xi’an Jiaotong University

Lei Xue  
Sun Yat-sen University

Li Feng  
Southwest Jiaotong University

Jianfeng Li  
Xi’an Jiaotong University

Junjie Zhang  
Wright State University

Xiaohong Guan  
Xi’an Jiaotong University

August 2023
Outline

- Background and Problem Description
- System Design
- Evaluation
- Conclusions
Background and Problem Description

Related work

Website Fingerprinting
Shen et al. [ACM ARES, 2019]
Di Martino et al. [IEEE ICC, 2019]

Application Fingerprinting
App-Net [INFOCOM WKSHPS, 2020]
FOAP [USENIX Security, 2022]

Internet of Things Fingerprinting
Ma et al. [IEEE INFOCOM, 2020]
IoTFinder [EuroS&P, 2020]

......
Motivation

Related work

Feature-based traffic fingerprinting
k-fingerprinting [USENIX Security, 2016]
Shafiq et al. [The Journal of Supercomputing, 2019]

Deep learning-based traffic fingerprinting
Deep fingerprinting [ACM CCS, 2018]
Var-cnn [PETS, 2019]
SHAME [ACM WPES, 2021]

……
Outline

- Background and Problem Description
- System Design
- Evaluation
- Conclusions
System Design
System Design
System Design

Module 0 (M0) - Packet Vectorization
Module 1 (M1) - Packet-to-flow Mapping
Module 2 (M2) - Flow-to-trace Mapping
Module 3 (M3) - Trace-to-label Classification

- Flow 1
  - Packets
  - Packet Vectorization

- Flow 2
  - Heterogeneous input
  - Packet Vectors

- Flow 3

- Flow 4

- Trace X
  - Trace Vector

- Flow Layer
  - Flow Vectors
  - Flow Vector 1
  - Flow Vector 2
  - Flow Vector 3
  - Flow Vector 4

- Trace Layer
  - Trace Vector

- Fully Connected Layer
  - Output Layer
  - Softmax
  - P(X=label-1)
  - P(X=label-2)
  - P(X=label-3)
  - P(X=label-n)
System Design

Diagram showing the system design with modules and layers for packet vectorization, packet-to-flow mapping, flow-to-trace mapping, and trace-to-label classification.
System Design

NN Structure in Packet-to-flow (M1) Mapping and Flow-to-trace (M2) Mapping

(1) Chain-structured
(2) Tree-structured
(3) Attention-structured
(4) Hybrid (uses multiple neural network structures)
System Design

Use CNN Compression to Speed up Training

(a) Without CNN Compression

(b) With CNN Compression
System Design

Techniques to Handle Overfitting

- Early Stopping
- Weight Decay
- Dropout
- Batch Normalization
- Auxiliary Loss
- Data Enhancement
System Design

Techniques to Handle Overfitting

- Early Stopping
- Weight Decay
- Dropout
- Batch Normalization
- Auxiliary Loss
- Data Enhancement

\[
\text{Loss} = \text{CrossEntropy}(P, \mathcal{L}) + \frac{1}{N} \sum_{i=1}^{N} \text{CrossEntropy}(P_i', \mathcal{L})
\]

Trace Classification loss

Flow Classification loss
System Design

Techniques to Handle Overfitting

- Early Stopping
- Weight Decay
- Dropout
- Batch Normalization
- Auxiliary Loss
- Data Enhancement

Cropping

Noising

Dropping
System Design

Techniques to Handle Overfitting

- Early Stopping
- Weight Decay
- Dropout
- Batch Normalization
- Auxiliary Loss
- Data Enhancement
- Hybrid
Outline

- Background and Problem Description
- System Design
- Evaluation
- Conclusions
### Evaluation

Performance comparison with the SOTA methods.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Our F1-score</th>
<th>SOTA method</th>
<th>SOTA F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>UAV [4]</td>
<td>0.928 (±0.023)</td>
<td>K-Means +RF [4]</td>
<td>0.957</td>
</tr>
<tr>
<td>SWF</td>
<td><strong>0.996 (±0.003)</strong></td>
<td>RF+LCS [43]</td>
<td>0.982</td>
</tr>
<tr>
<td>KWF</td>
<td><strong>0.977 (±0.009)</strong></td>
<td>PSC+ET [44]</td>
<td>0.974</td>
</tr>
<tr>
<td>IDI [30]</td>
<td><strong>0.940 (±0.040)</strong></td>
<td>RF [45]</td>
<td>0.91</td>
</tr>
<tr>
<td>ISD [31]</td>
<td><strong>0.984 (±0.022)</strong></td>
<td>CCR-ELM [46]</td>
<td>0.961</td>
</tr>
</tbody>
</table>

◆ Our method effectively fingerprint traffic across multiple tasks.

### Datasets

- User Activities (UAV)
- IoT Device Identification (IDI)
- Intrusion Detection (ISD)
- Keyword Searching (KWS)
- Shadowsocks Website Fingerprinting (SWF)
Evaluation

Macro F1-scores using different neural network structures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>0.925 (±0.019)</td>
<td>0.995 (±0.004)</td>
<td>0.977 (±0.010)</td>
<td>0.968 (±0.024)</td>
<td>0.990 (±0.010)</td>
</tr>
<tr>
<td>Chain</td>
<td>0.922 (±0.029)</td>
<td>0.992 (±0.006)</td>
<td>0.979 (±0.014)</td>
<td>0.954 (±0.033)</td>
<td>0.995 (±0.008)</td>
</tr>
<tr>
<td>Tree</td>
<td>0.923 (±0.023)</td>
<td>0.993 (±0.001)</td>
<td>0.976 (±0.010)</td>
<td>0.937 (±0.037)</td>
<td>0.986 (±0.014)</td>
</tr>
<tr>
<td>Hybrid</td>
<td>0.920 (±0.025)</td>
<td>0.993 (±0.003)</td>
<td>0.974 (±0.020)</td>
<td>0.940 (±0.040)</td>
<td>0.997 (±0.005)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>0.912 (±0.031)</td>
<td>0.991 (±0.003)</td>
<td>0.979 (± 0.009)</td>
<td>0.924 (±0.055)</td>
<td>0.992 (±0.008)</td>
</tr>
<tr>
<td>Chain</td>
<td>0.920 (±0.027)</td>
<td>0.994 (±0.004)</td>
<td>0.979 (± 0.009)</td>
<td>0.963 (±0.030)</td>
<td>0.982 (±0.022)</td>
</tr>
<tr>
<td>Tree</td>
<td>0.907 (±0.027)</td>
<td>0.987 (±0.006)</td>
<td>0.950 (±0.014)</td>
<td>0.848 (±0.041)</td>
<td>0.992 (±0.012)</td>
</tr>
<tr>
<td>Hybrid</td>
<td>0.920 (±0.025)</td>
<td>0.993 (±0.003)</td>
<td>0.974 (±0.020)</td>
<td>0.940 (±0.040)</td>
<td>0.997 (±0.005)</td>
</tr>
</tbody>
</table>

◆ Hybrid structures should be adopted for stable Macro F1-scores.
Evaluation

Macro F1-scores using different solutions to handle overfitting. H-* removes method * from the hybrid solution.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Handling</td>
<td>0.918 (±0.027)</td>
<td>0.956 (±0.025)</td>
<td>0.204 (±0.275)</td>
<td>0.721 (±0.137)</td>
<td>0.998 (±0.005)</td>
</tr>
<tr>
<td>Pure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>0.913 (±0.034)</td>
<td>0.953 (±0.024)</td>
<td>0.203 (±0.275)</td>
<td>0.723 (±0.134)</td>
<td>0.995 (±0.010)</td>
</tr>
<tr>
<td>WD</td>
<td>0.922 (±0.022)</td>
<td>0.945 (±0.027)</td>
<td>0.391 (±0.391)</td>
<td>0.720 (±0.066)</td>
<td>0.998 (±0.005)</td>
</tr>
<tr>
<td>DO</td>
<td>0.922 (±0.024)</td>
<td>0.736 (±0.161)</td>
<td>0.052 (±0.132)</td>
<td>0.277 (±0.105)</td>
<td>0.994 (±0.008)</td>
</tr>
<tr>
<td>BN</td>
<td>0.922 (±0.021)</td>
<td>0.994 (±0.003)</td>
<td>0.970 (±0.011)</td>
<td>0.859 (±0.211)</td>
<td>0.992 (±0.011)</td>
</tr>
<tr>
<td>AL</td>
<td>0.919 (±0.023)</td>
<td>0.990 (±0.005)</td>
<td>0.869 (±0.010)</td>
<td>0.833 (±0.070)</td>
<td>0.992 (±0.008)</td>
</tr>
<tr>
<td>DE</td>
<td>0.921 (±0.022)</td>
<td>0.980 (±0.016)</td>
<td>0.140 (±0.261)</td>
<td>0.789 (±0.071)</td>
<td>0.995 (±0.010)</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-ES</td>
<td><strong>0.932</strong> (±0.023)</td>
<td>0.996 (±0.002)</td>
<td>0.820 (±0.076)</td>
<td>0.940 (±0.049)</td>
<td>0.998 (±0.005)</td>
</tr>
<tr>
<td>H-WD</td>
<td>0.918 (±0.026)</td>
<td>0.995 (±0.004)</td>
<td>0.974 (±0.007)</td>
<td>0.933 (±0.037)</td>
<td>0.989 (±0.016)</td>
</tr>
<tr>
<td>H-BN</td>
<td>0.917 (±0.024)</td>
<td>0.991 (±0.008)</td>
<td>0.872 (±0.033)</td>
<td>0.848 (±0.080)</td>
<td>0.994 (±0.008)</td>
</tr>
<tr>
<td>H-AL</td>
<td>0.916 (±0.021)</td>
<td>0.993 (±0.003)</td>
<td>0.970 (±0.015)</td>
<td>0.944 (±0.040)</td>
<td>0.995 (±0.007)</td>
</tr>
<tr>
<td>H-DE</td>
<td>0.924 (±0.024)</td>
<td>0.996 (±0.004)</td>
<td>0.973 (±0.005)</td>
<td><strong>0.958</strong> (±0.026)</td>
<td>0.990 (±0.011)</td>
</tr>
<tr>
<td>H</td>
<td>0.928 (±0.023)</td>
<td><strong>0.996</strong> (±0.003)</td>
<td><strong>0.977</strong> (±0.009)</td>
<td>0.935 (±0.041)</td>
<td><strong>0.998</strong> (±0.004)</td>
</tr>
</tbody>
</table>

◆ Hybrid solutions should be adopted for high Macro F1-scores.
## Evaluation

### Macro F1-scores when confronted with hierarchy unawareness deep learning methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HA-1.1</td>
<td>0.835 ($\pm$ 0.024)</td>
<td>0.964 ($\pm$ 0.012)</td>
<td>0.388 ($\pm$ 0.365)</td>
<td>0.816 ($\pm$ 0.072)</td>
<td>0.811 ($\pm$ 0.057)</td>
</tr>
<tr>
<td>HA-1.2</td>
<td>0.906 ($\pm$ 0.022)</td>
<td>0.966 ($\pm$ 0.013)</td>
<td>0.927 ($\pm$ 0.021)</td>
<td>0.840 ($\pm$ 0.112)</td>
<td>0.914 ($\pm$ 0.063)</td>
</tr>
<tr>
<td>HA-2</td>
<td>0.556 ($\pm$ 0.025)</td>
<td>0.800 ($\pm$ 0.022)</td>
<td>0.204 ($\pm$ 0.012)</td>
<td>0.749 ($\pm$ 0.018)</td>
<td>0.872 ($\pm$ 0.029)</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>0.928</strong> ($\pm$ 0.023)</td>
<td><strong>0.996</strong> ($\pm$ 0.003)</td>
<td><strong>0.977</strong> ($\pm$ 0.009)</td>
<td><strong>0.940</strong> ($\pm$ 0.040)</td>
<td><strong>0.997</strong> ($\pm$ 0.005)</td>
</tr>
</tbody>
</table>

- HA-1.1: Treat a trace consisting of multiple flows as a sample, **without** distinguishing between flows.
- HA-1.2: Treat a trace consisting of multiple flows as a sample, **with** distinguishing between flows.
- HA-2: Treat each flow of a trace as a sample, and classifying it into different trace labels.

◆ Hierarchy awareness is important.
Outline

- Background and Problem Description
- System Design
- Evaluation
- Conclusions
Conclusions

◆ We take the first step to designing an input-agnostic hierarchical deep learning framework to seamlessly land deep learning onto traffic fingerprinting.

◆ Our framework successfully applies in various fingerprinting tasks where SOTA methods rely on handcrafted features and deep learning is not easily applicable.

◆ We proposed techniques to handle overfitting and analyzed real-world factors that affect performance.

◆ Code available at https://github.com/shashadehuajiang/trace_classifier
Thank you!

Feel free to contact with any questions:

qj904154277@stu.xjtu.edu.cn