# MaskDroid: Robust Android Malware Detection with Masked Graph Representations

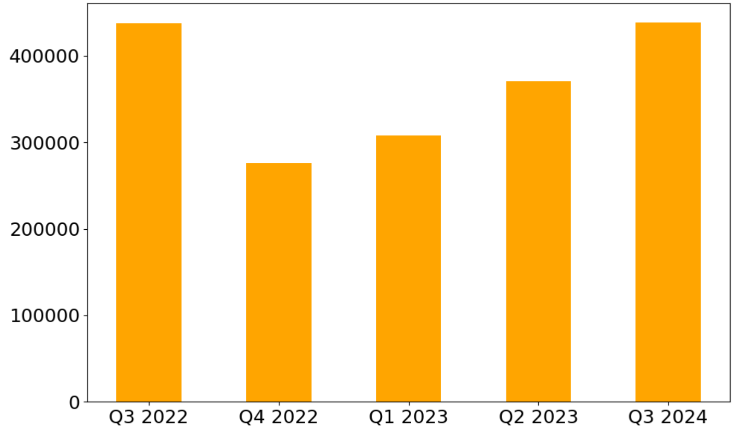
*Jingnan Zheng[*], Jiahao Liu[*], An Zhang, Jun Zeng, Ziqi Yang, Zhenkai Liang, Tat-Seng Chua*

IEEE/ACM ASE, Oct. 2024

NUS
National University of Singapore

# Security Threats and Risks of Android Malware

Android malware poses increasing security threats and risks

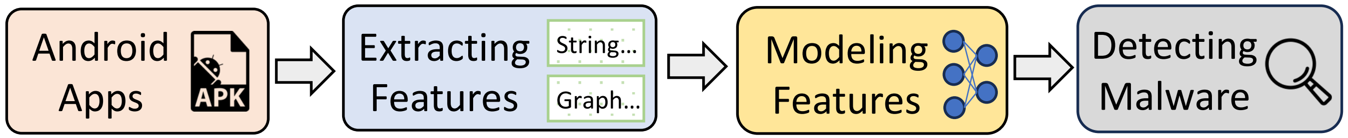



*Number of detected malicious install packages*


Detecting Android malware before installation is the key to mitigate these security threats and risks

# ML-based Android Malware Detection

Characterize apps and identify malicious patterns to distinguish malware




*General workflow of ML-based Android malware detection*

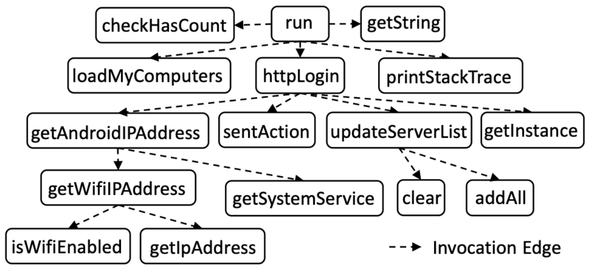
**Syntax-based methods** [Drebin @NDSS’14, XMAL @TOSEM’21, …]

- Model app behaviors with discrete features, e.g., permission, API calls

**Semantic-based methods** [Malscan @ASE’19, MsDroid @TDSC’22, …]

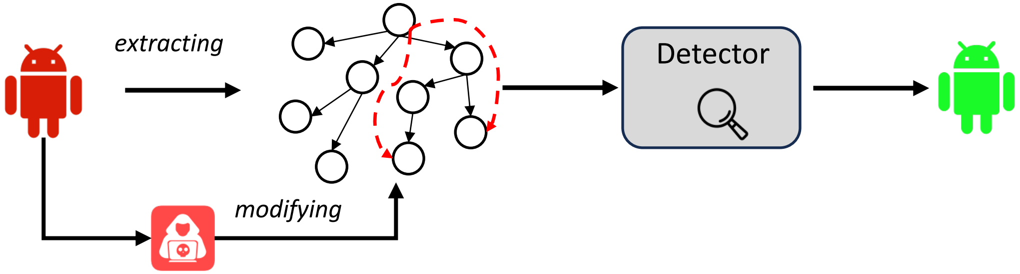
- Distill semantics from apps’ graph representations

Permissions: *INTERNET …*
API calls: *getIpAddress …*



# Vulnerable to Adversarial Attacks

Adversarial attack purposely modifies the graph structure to bypass the detection




**A Comprehensive Study of Learning-based Android Malware Detectors under Challenging Environments**

**Black-box Adversarial Example Attack towards FCG Based Android Malware Detection under Incomplete Feature Information**

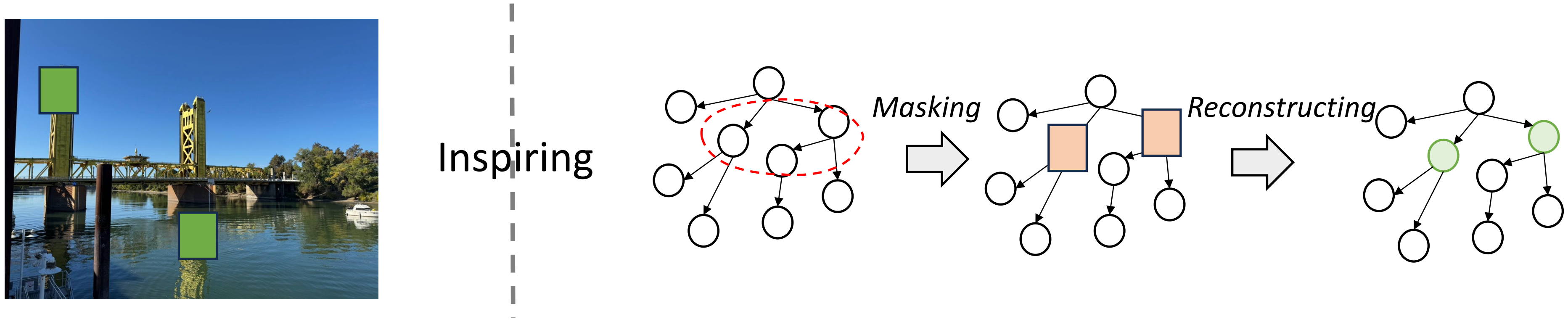
Heng Li[†], Zhang Cheng[‡,†], Bang Wu[†], Liheng Yuan[†], Cuiying Gao[†], Wei Yuan[†] [*], Xiapu Luo[*]
[†] *Huazhong University of Science and Technology*
[*] *The Hong Kong Polytechnic University*
[‡] *NSFOCUS Technologies Group Co., Ltd.*
*{liheng,wubangm,ylh,gaocy,yuanwei}@hust.edu.cn*
*chengzhang@nsfocus.com,csxluo@comp.polyu.edu.hk*

**Research Problem**: Given the graph representation, could we design a robust and effective Android malware detector?

# Our Insights

Masking and reconstructing mechanisms are effective for robust learning

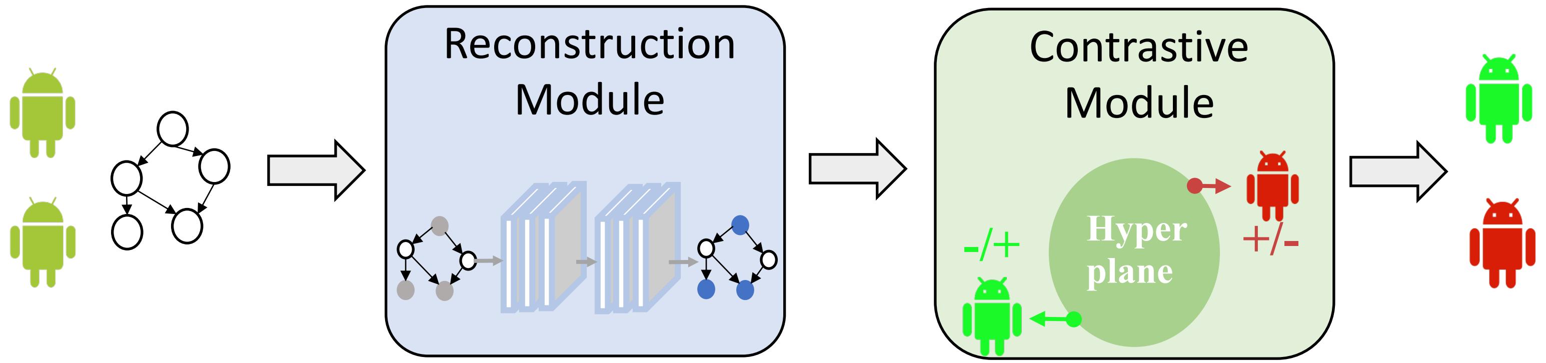
- Encourage models to capture overall information even if some features are purposely changed




Contrastive strategy can better investigate the relationships among samples

- samples within the same class are drawn closer together, while those from different classes are pushed further apart

# MaskDroid: Overview

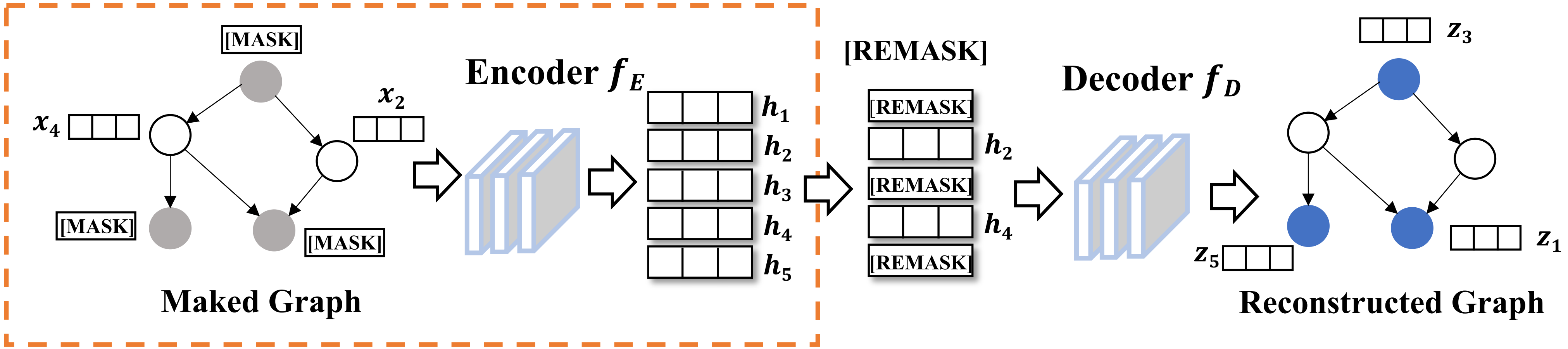



*Given an Android app, output the probability of being malicious*

- Mask and reconstruct graphs to learn robust representations
- Separate malware with a contrastive strategy

# Reconstruction Module

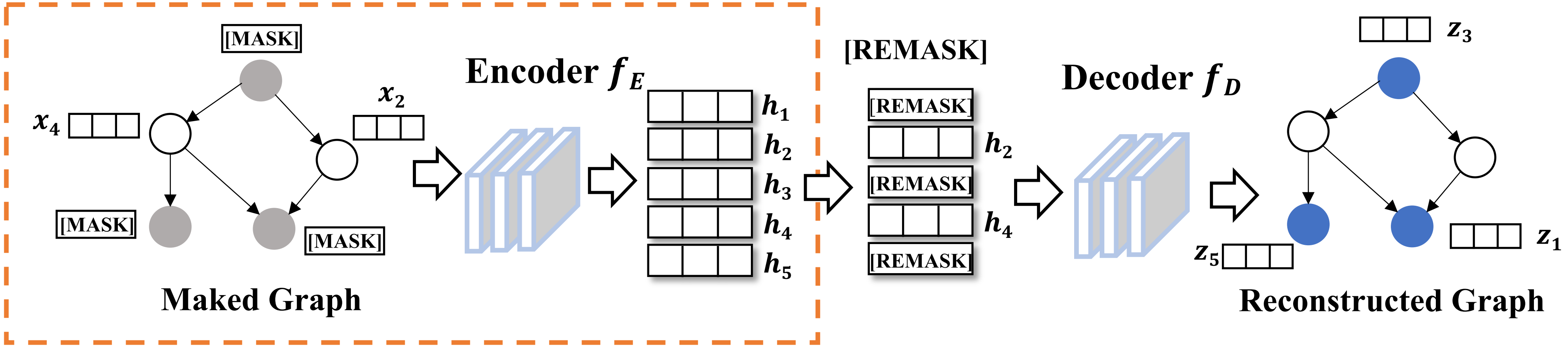
**Key Idea:** masking and reconstructing the graph structure to learn a robust representation of malicious behavior



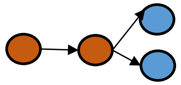

- ***Mask*** the graph to construct incomplete graph representation
    - Concatenate opcode and permission to represent graph node features
    - Apply uniform random sampling to choose a subset of nodes and mask their representations

# Reconstruction Module

**Key Idea:** masking and reconstructing the graph structure to learn a robust representation of malicious behavior

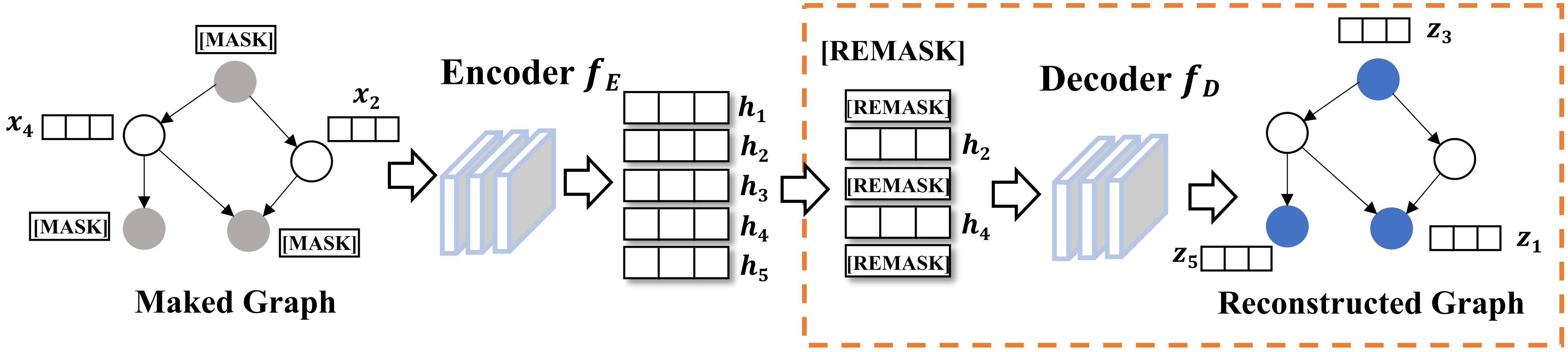



- ***Encode*** the graph representation with GNN Encoder
  - Propagate and aggregate node information across edges, enabling the model capture both local and global graph dependencies
  - getAndroidAddress -> getWifiIpAddress -> isWifiEnabled | getIpAddress



# Reconstruction Module

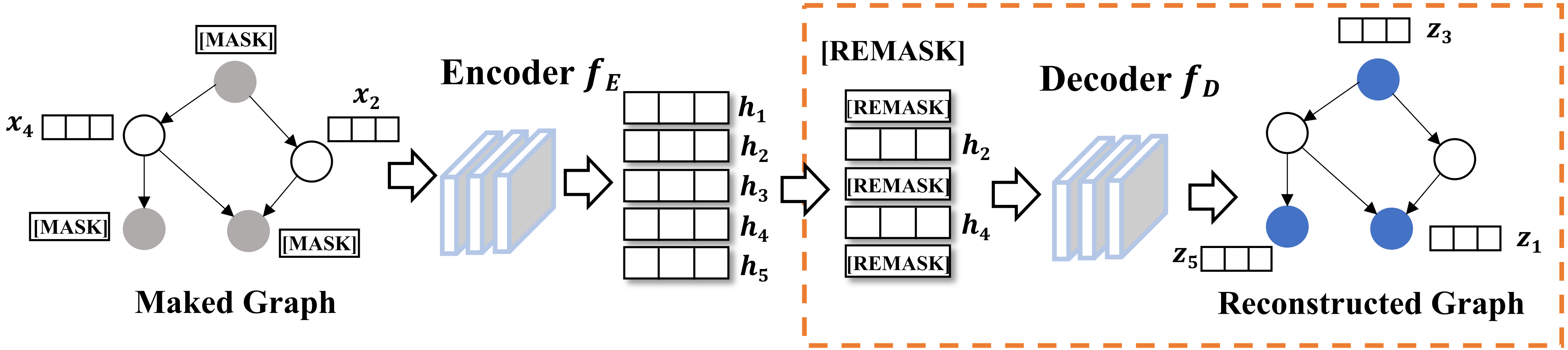
**Key Idea:** masking and reconstructing the graph structure to learn a robust representation of malicious behavior




- ***Reconstruct*** the masked nodes to allow the model can infer overall information from partial nodes and edges
  - Ensure the model recover masked nodes based solely on their surrounding nodes -> remask
  - Another GNN to capture structural information and recover nodes

# Reconstruction Module

**Key Idea:** masking and reconstructing the graph structure to learn a robust representation of malicious behavior

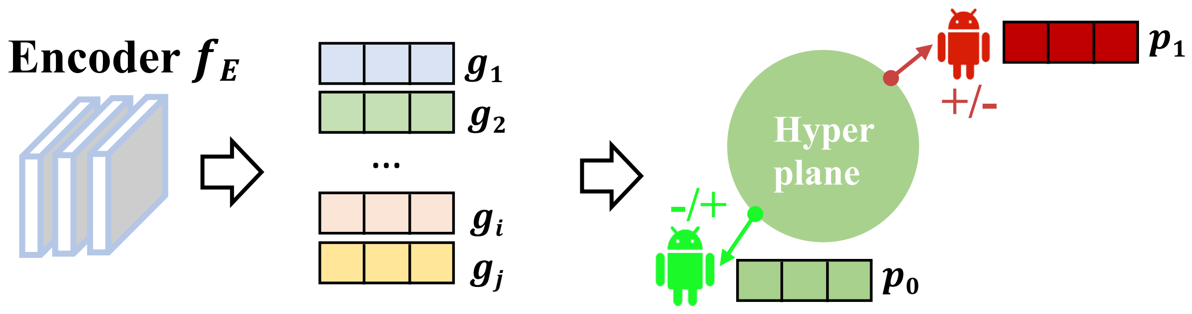



- ***Objective:*** measure the distance between the original node features and the reconstruction features Reconstruction Loss $(z_i, x_i)$

With the reconstruction module, MaskDroid can recover the overall graph information, even if the graph is partly corrupted

# Contrastive Module

**Key Idea:** Apps within the same class should be closer to each other, while apps from different classes should be more distant

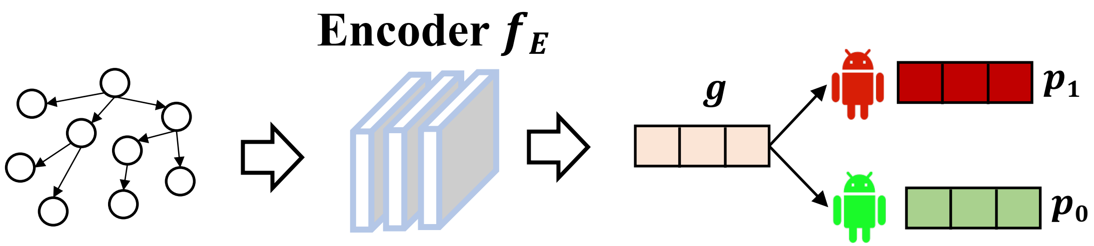



- ***Define*** two proxies for benign and malicious classes to guide the contrastive learning
  - Each instance is pulled closer to the proxy of its own class while being pushed further from the other proxy Proxy-Based Contrastive Loss $(g_i, p_0, p_1, y_i)$

With the contrastive module, MaskDroid can learn a compact representation for each class and forms clear boundaries between different classes

# Detecting Android Malware

Transforms the input graph into a graph-level representation using the encoder and calculate the distance with benign and malicious proxies




*Detecting Android malware by calculating the distance*

# Evaluation

## Experiment Setup:

- Around 114k apps collected from AndroZoo – a continuously expanding repository of Android apps sourced from platforms such as Google Play, Appchina, and Anzhi
    - Covering a wide range of apps (5 years)
    - Mirroring real-world malware distribution (the ratio of goodware to malware is set as 9:1)
    - Filtering out grayware with positive anti-virus alerts from VirusTotal

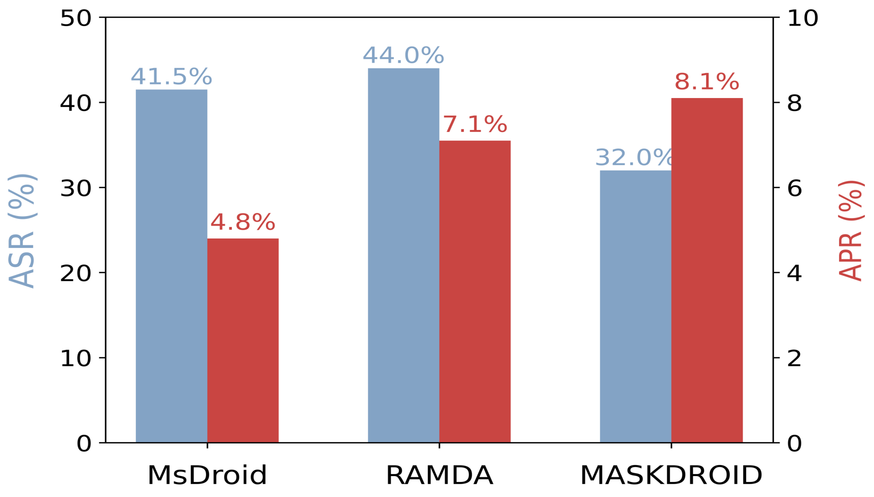
## Evaluation aspects:

- Does MaskDroid improve the robustness against different adversarial attacks?
- Does MaskDroid sacrifice detection effectiveness to enhance its robustness?
- To what extent do different design choices affect MaskDroid’s performance?

# Robustness Enhancement

Investigate whether MaskDroid can enhance its robustness against adversarial attacks compared to existing solutions (white-box and black-box)

- Attack Success Rate (ASR), Average Perturbation Ratio (APR)




*White-box Attack*

| Detectors | Malscan | MamaDroid | Drebin |
|---|---|---|---|
| ASR | 98.5% | 69.0% | 100% |
| APR | - | - | - |
| **Detectors** | **MsDroid** | **RAMDA** | **MaskDroid** |
| ASR | 13.2% | 19.2% | 19.1% |
| APR | 0.5% | 1.5% | 10.1% |

*black-box Attack*

Compared with state-of-the-art solutions, MaskDroid enhances robustness against adversarial attack, especially in white-box attack.

# Effectiveness Comparison

Investigate whether the improved robustness comes at the expense of detection performance (F1-score)

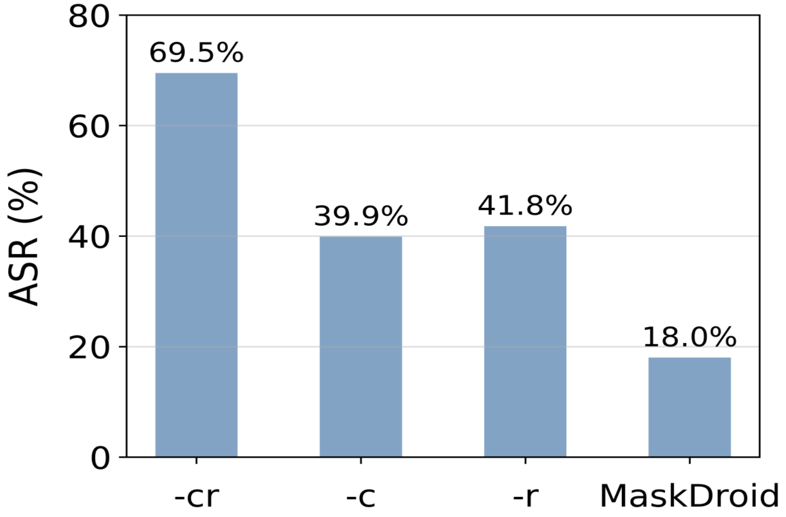
- Training and testing data within the same period
- Temporal split: Training with previous, testing with later data

| Time Span | Malscan | Mamadroid | Drebin | MsDroid | RAMDA | MaskDroid |
|---|---|---|---|---|---|---|
| **Same time** | 0.808 | 0.838 | 0.736 | 0.598 | 0.811 | **0.783** |
| **Time bias** | 0.473 | 0.421 | 0.573 | 0.317 | 0.546 | **0.582** |

MaskDroid achieves detection effectiveness comparable to existing Android malware detectors in both same-time and temporal bias scenarios.

# Evaluating the Design of MaskDroid

Investigate the effects of reconstruction and contrastive modules on the robustness and effectiveness of MaskDroid




*Robustness of various models*

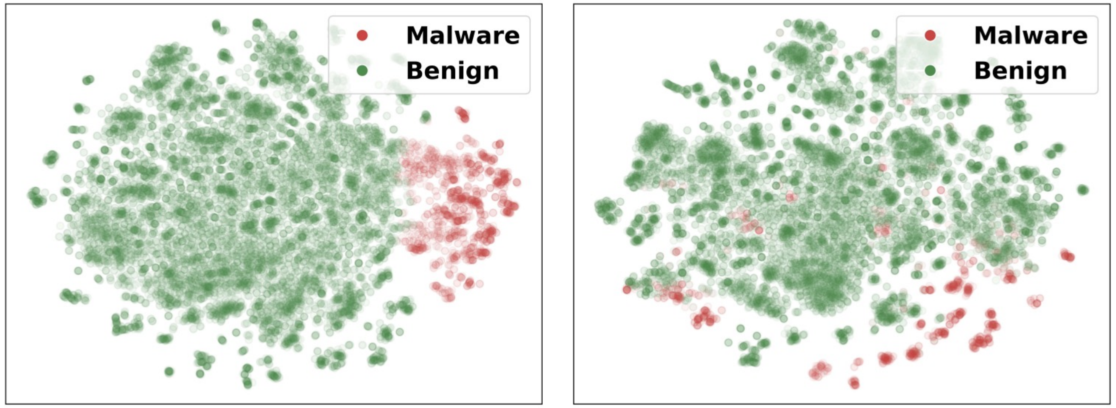
| Models | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| MaskDroid-cr | 0.918 | 0.730 | 0.813 | 0.965 |
| MaskDroid-c | 0.886 | 0.688 | 0.774 | 0.958 |
| MaskDroid-r | 0.896 | 0.720 | 0.799 | 0.962 |
| MaskDroid | 0.772 | 0.883 | 0.824 | 0.961 |

*Effectiveness of various models*

Both reconstruction and contrastive modules contribute to the performance of MaskDroid.

# Evaluating the Design of MaskDroid

Visualize the embeddings of MaskDroid and its variant with the contrastive and reconstruction modules disabled




(a) MaskDroid (b) MaskDroid-cr

MaskDroid more effectively separates malware from benign apps, providing a compact representation and clear boundaries.

# Conclusion

We propose MaskDroid:

- Learn robust graph representation encoding malicious behaviors with graph masking and reconstruction
- Incorporate proxy-based contrastive learning to better separate benign and malicious Android apps
- Release code and data at: https://github.com/SophieZheng998/MaskDroid