LAMDA: A Longitudinal Android Malware Benchmark for Concept Drift Analysis

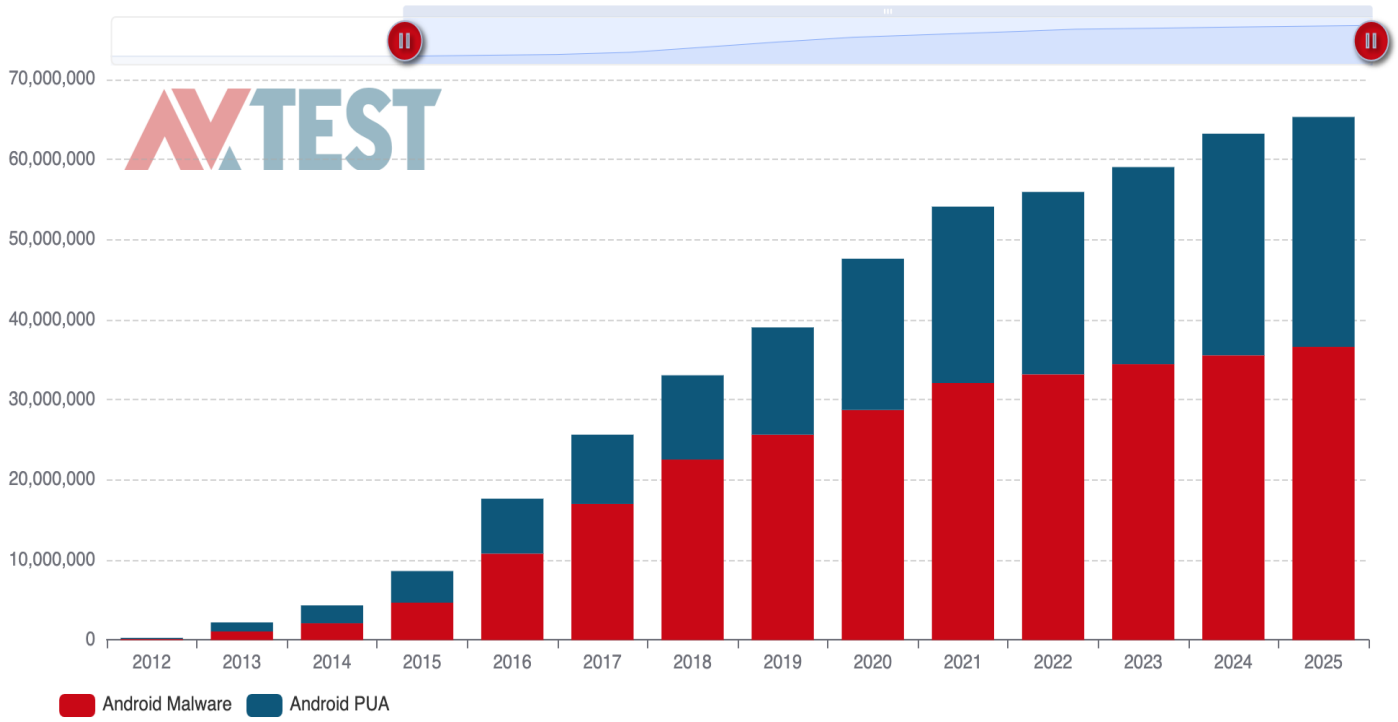
Md Ahsanul Haque, Ismail Hossain, Md Mahmuduzzaman Kamol, Md Jahangir Alam, Suresh kumar Amalapuram, Sajedul Talukder, Mohammad Saidur Rahman

Presenter: **Md Ahsanul Haque**
IQSeC Lab
Department of Computer Science
University of Texas at El Paso

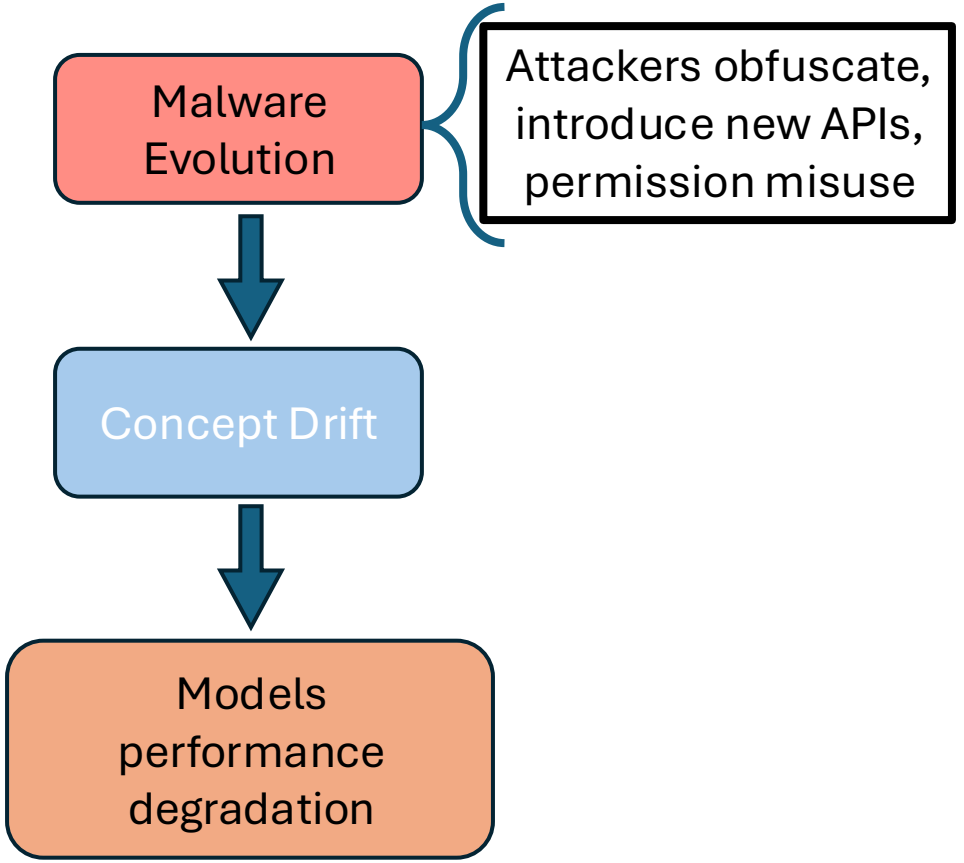
ICLR

The Fourteenth International Conference on Learning Representations - 2026

Growth of Malicious Software (Malware)

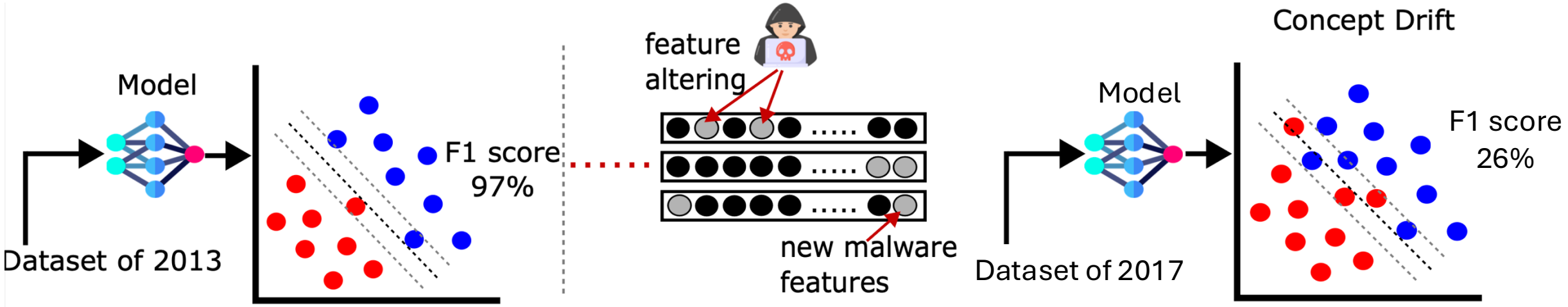


Around 450K new malware emerge each day [1].



[1] https://www.av-test.org/en/statistics/malware/

Concept Drift in Malware



Example: Android trojan SoumniBot obfuscates its manifest file to evade detection [2].

[2] https://thehackernews.com/2024/04/new-android-trojan-soumnibot-evades.html

Why Yet another Benchmark: LAMDA

- Existing works do not reflect real-world malware evolution
  - Limited temporal coverage
  - Reduced family diversity and singleton sample
- Need detailed analysis for real-world and longitudinal drift analysis [3].
- Need label drift analysis
- Need a benchmark to evaluate concept drift adaptation methods.

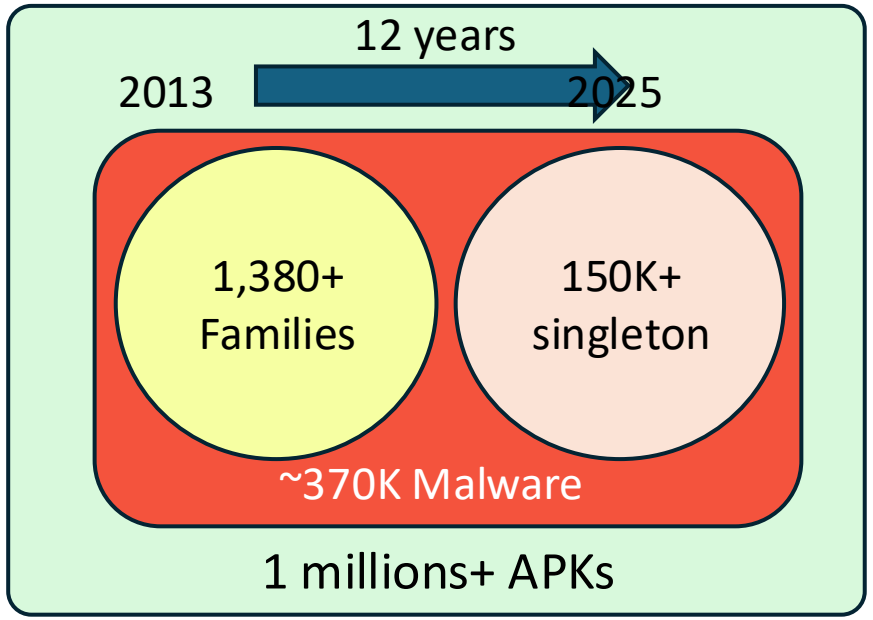
Comparison against Previous Benchmark

	APIGraph	LAMDA (Ours)
Total Samples	~290,000	**1,008,381**
Malware Samples	~32,000	**~370,000**
Malware Families	1,120	1,380
Singleton Samples	~370	**~150,000**
Temporal Span	7 years	**12 years**

[3] Yang, Limin, et al. "CADE: Detecting and Explaining Concept Drift Samples for Security Applications." USENIX Security 2021.

Contributions in this Work

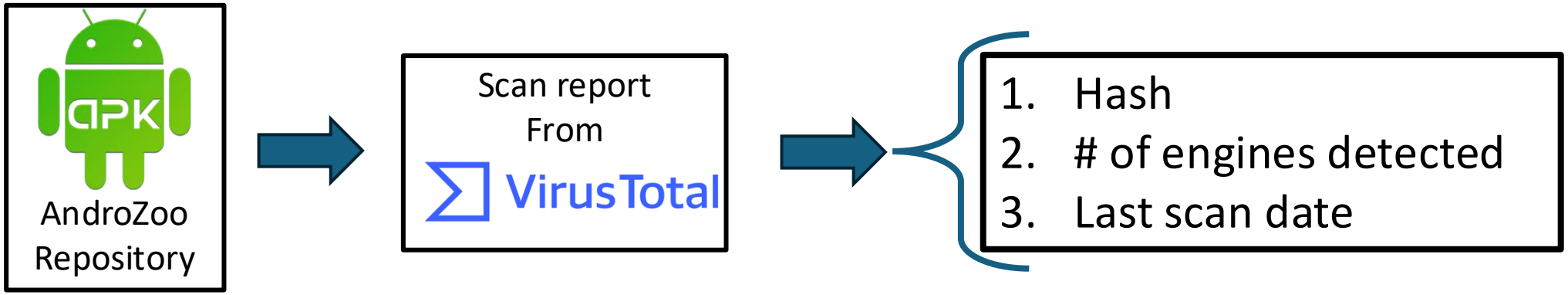
- LAMDA, a large-scale Android malware benchmark.
- Provide LAMDA creation pipeline
- Perform a detail concept drift detection.
- Conduct comprehensive drift analysis.
- Evaluate existing drift adaptation methods.
- Compare LAMDA with previous benchmark.
- Evaluate a broader application of LAMDA.



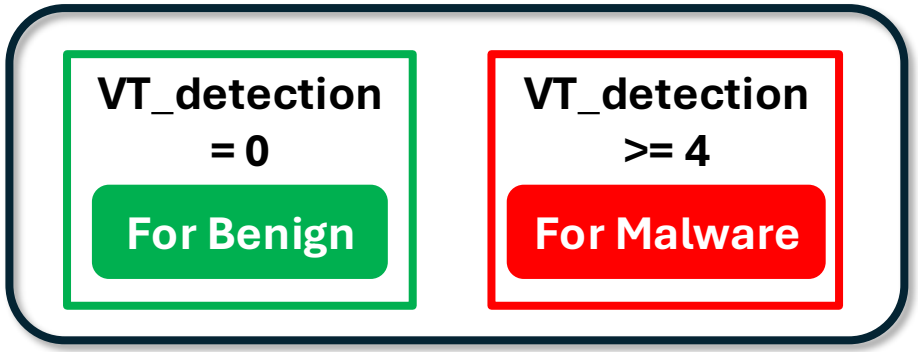
LAMDA

The LAMDA Creation

Collection Strategy and Label Assignment:

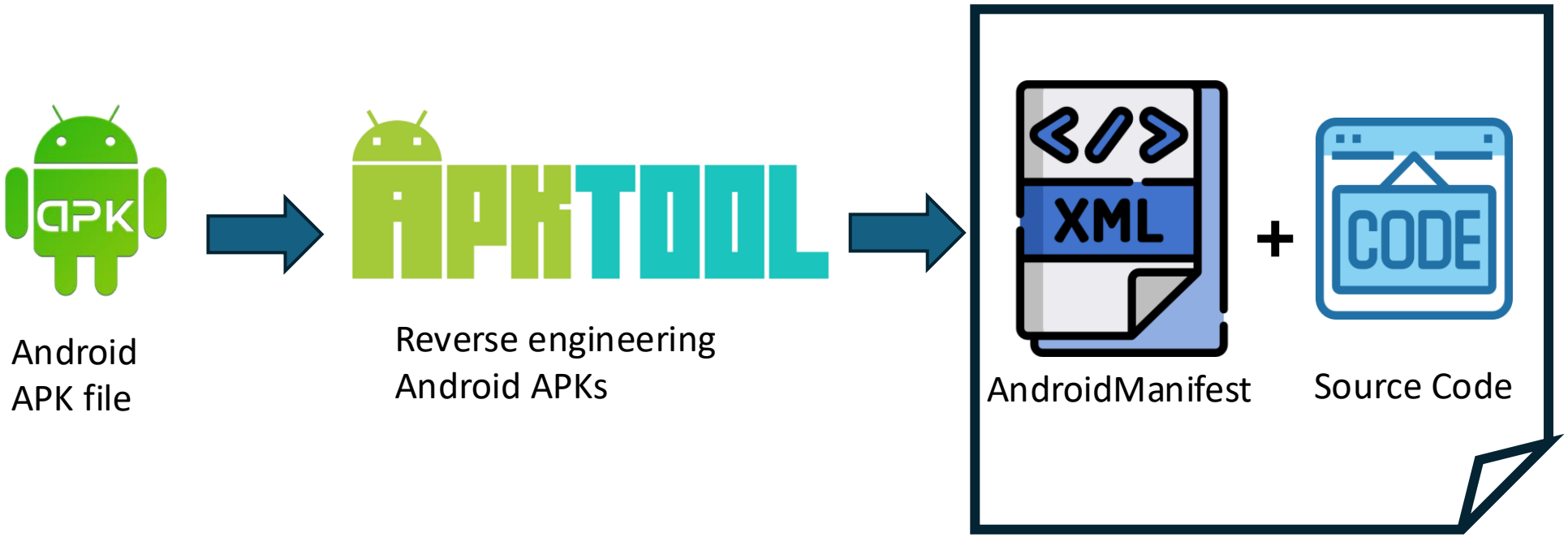


Labeling Strategy [4]



[4] Pendlebury, Feargus, et al. "TESSERACT: Eliminating experimental bias in malware classification across space and time." USENIX Security 2019.

Decompilation and Static Feature Extraction



APK Collection and Train Test Split

Year	Malware Samples	Benign Samples	Total Samples
2013	44,383	42,048	86,431
2014	45,756	55,427	101,183
2016	45,134	64,059	109,193
2017	21,359	77,785	99,144
2018	39,350	64,942	104,292
2019	41,585	49,465	91,050
2020	46,355	55,718	102,073
2021	35,627	45,528	81,155
2022	41,648	44,768	86,416
2023	7,892	46,462	54,354
2024	794	47,633	48,427
2025	23	44,640	44,663
Total	369,906	638,475	1,008,381

Train Split → 80% from each year

Test Split → 20% from each year

Vectorization and Temporal Feature Alignment

All Extracted Train APKs

~9.69 millions

Global Vocabulary Construction

Vectorization

Sample	Feat_1	Feat_2	Feat_3	……	Feat_9,690,000
APK_1	0	1	0	……	1
APK_2	0	0	1	……	0
……	…….	……	…..	……	…..
APK_n	0	0	0	……	1

Variance Threshold

Baseline
Threshold 0.001
4561 features

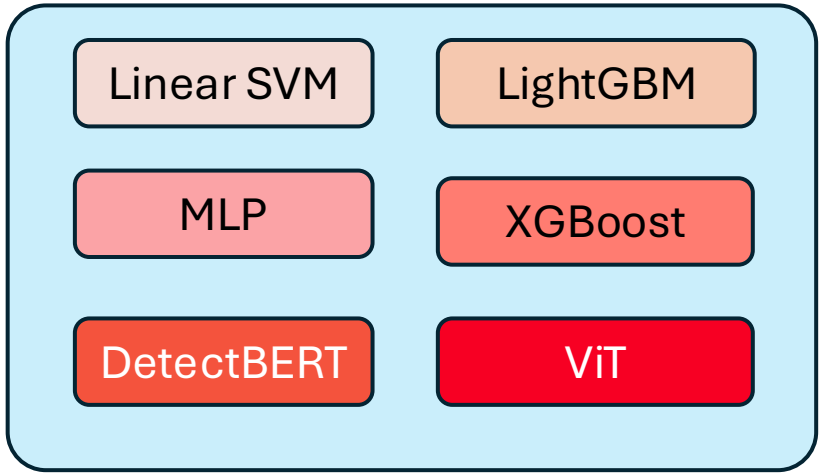
V2 Threshold
0.0001
25460 features

V3 Threshold
0.01
925 features

LAMDA Variants

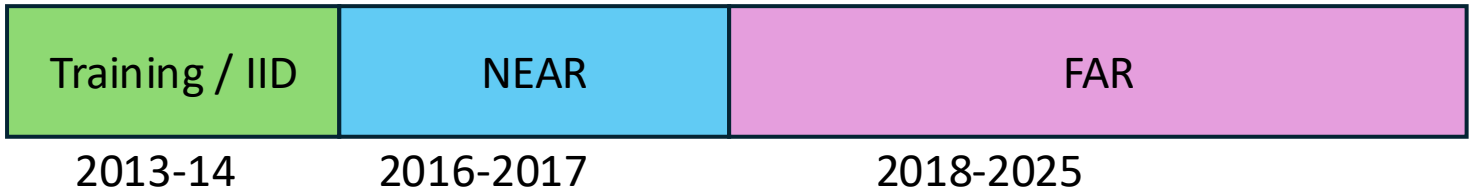
Concept Drift Detection with Supervised Learning

Models Used:



Temporal Split of the data [5]:

- IID - Independent and Identically Distributed with Training Distribution.
- NEAR – Smaller temporal distance from training.
- FAR – Larger temporal distance from training.

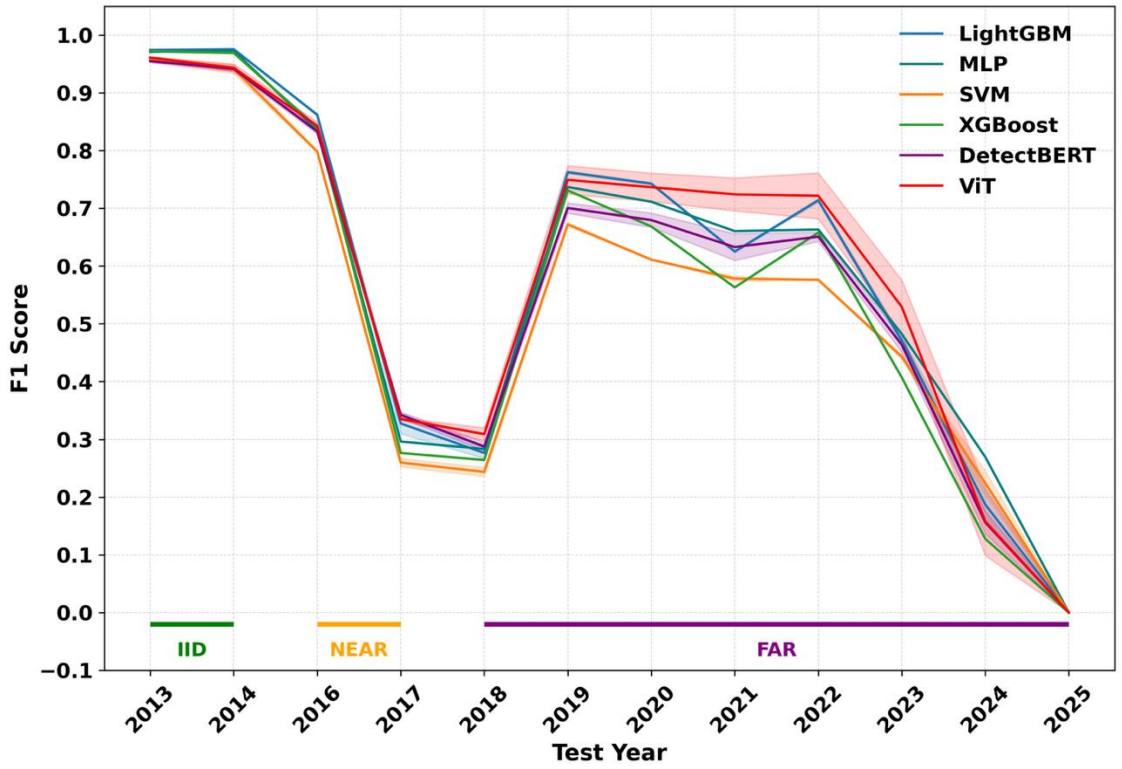


[5] Dragoi, Marius, et al. "Anoshift: A distribution shift benchmark for unsupervised anomaly detection." *NeurIPS* 2022.

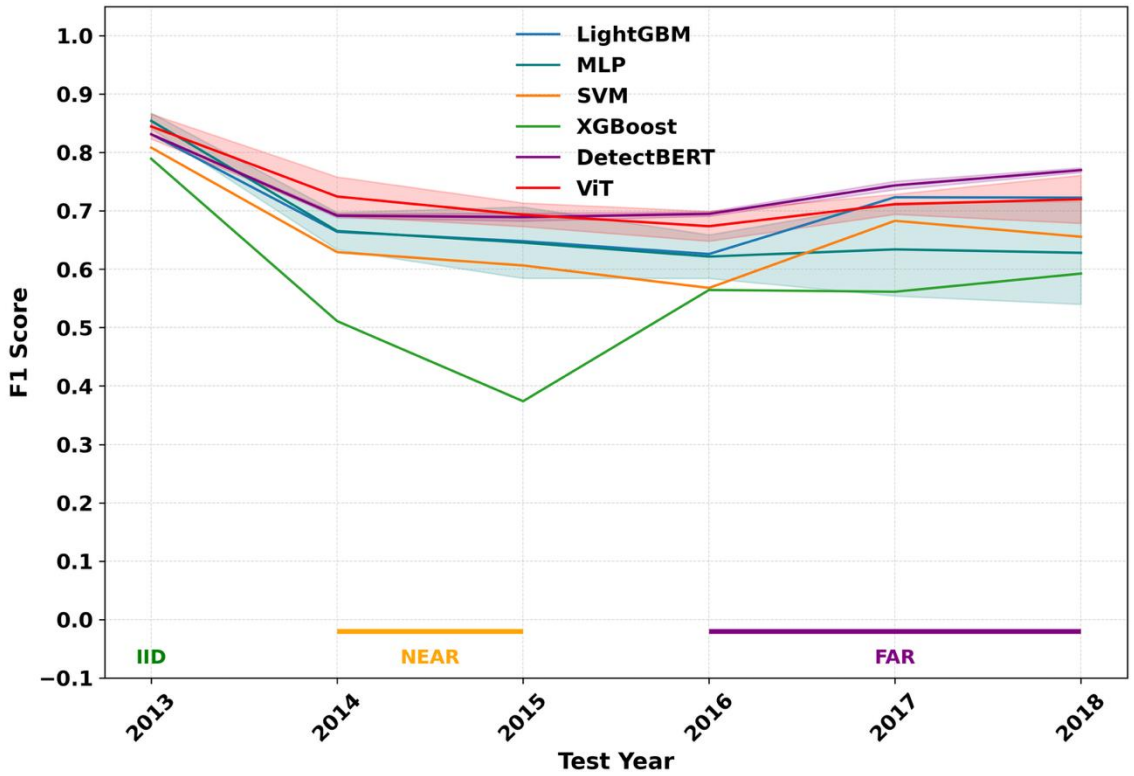
Results Comparison

LAMDA	IID = 97.4%, on NEAR = 59.4% and on FAR = 47.2%
APIGraph	IID = 85.9%, on NEAR = 66.7% and on FAR = 68.2%

LAMDA Performance



APIGraph Performance



Temporal Drift Analysis on Common Malware Families

Jaccard similarity → measures the total number of common features [6]

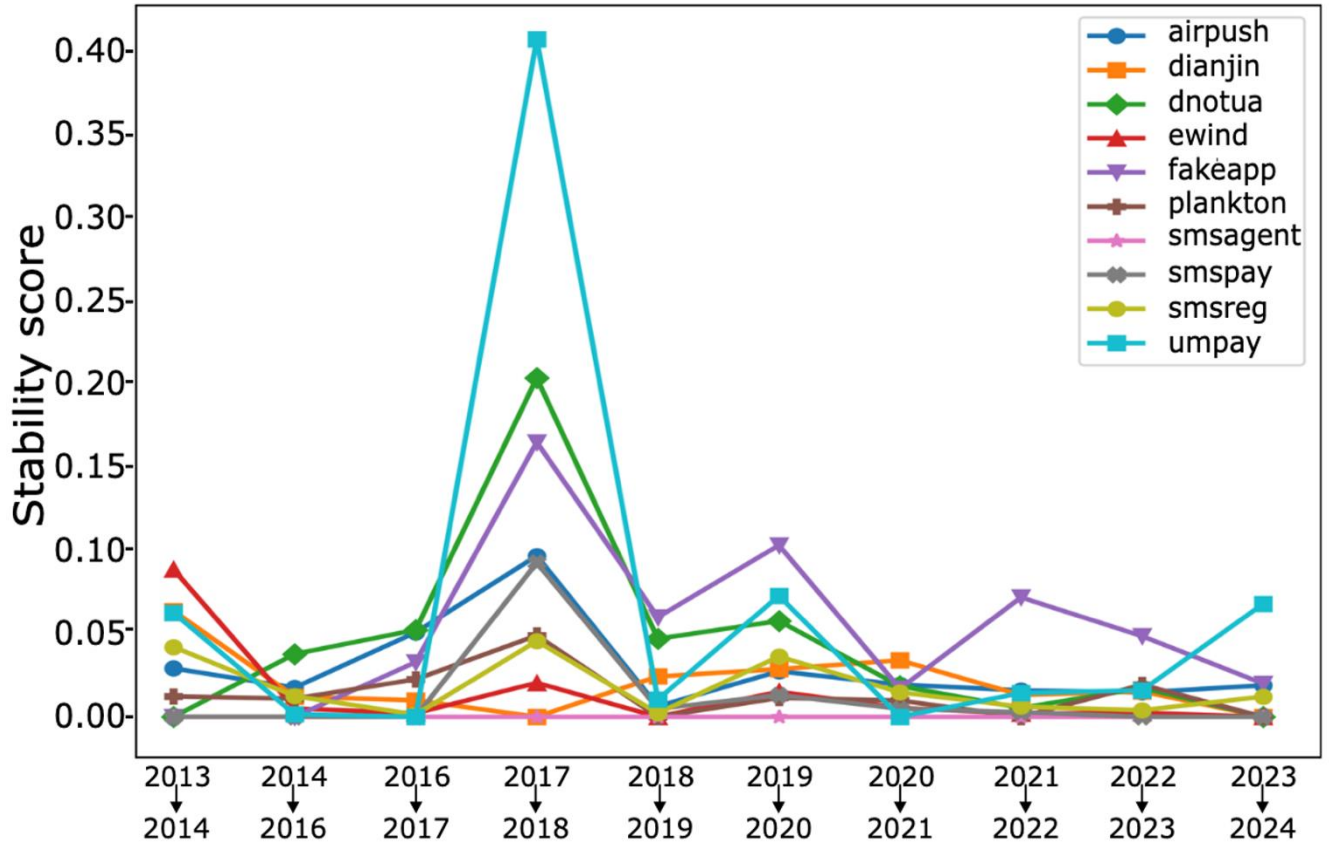
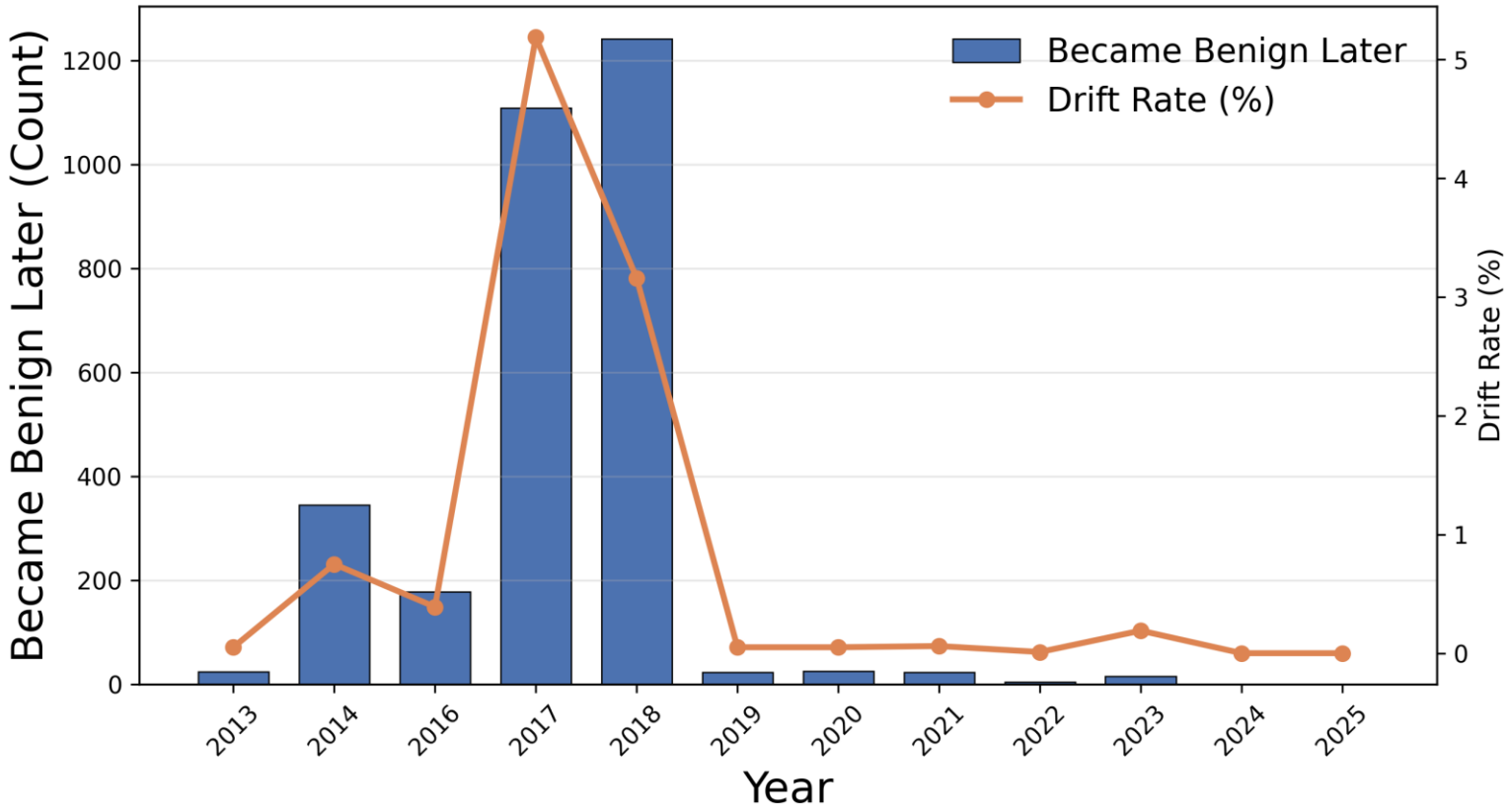


Fig: Jaccard similarity-based stability score [7]

[6] Masud, et al. “Classification and Adaptive Novel Class Detection of Feature-Evolving Data Streams.” IEEE TKDE, 2011.

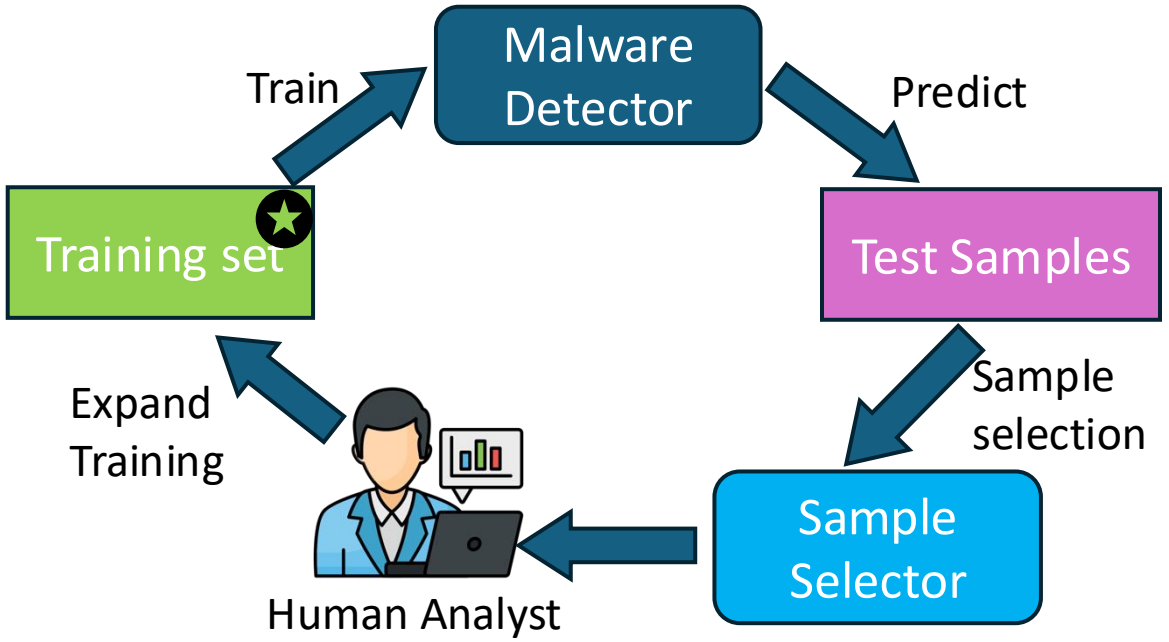
Label Drift Analysis

LAMDA Android Malware Samples → **Rescan from VirusTotal** → **New report**

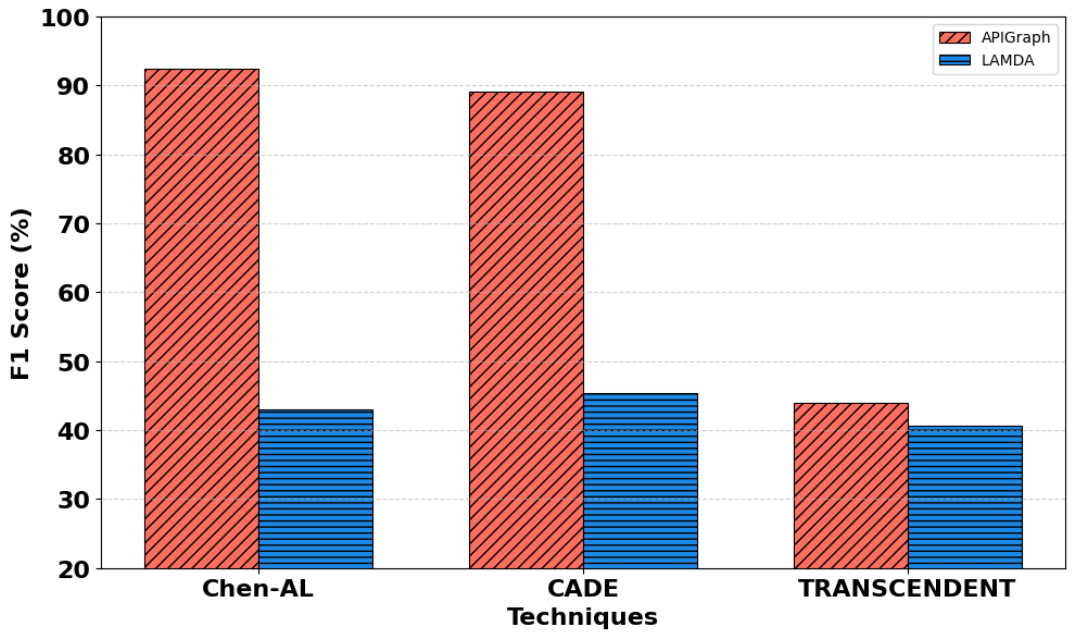


Concept Drift Adaptation

We utilize the concept drift adaptation framework using active learning following Chen-AL [7]



Concept Drift Adaptation Techniques comparison on 400 Monthly Budgets

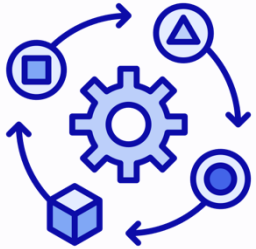


[7] Chen, Yizheng, et al. "Continuous Learning for Android Malware Detection." USENIX Security 2023.

Broader Applications of LAMDA

Evaluation of generalization

Study Malware Evolution

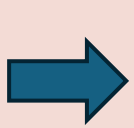


Development of Adaptive Models

Supports Scalability

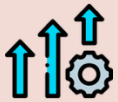
Continual Learning

Continual Learning



Domain Incremental Learning (Domain-IL)

Task Incremental Learning (Task-IL)



Class Incremental Learning (Class-IL)

Conclusion and Takeaways

- Benchmark dataset for concept drift detection and adaptation.
- Long-term evaluation of malware detection systems.
- Analysis for detecting and analyzing concept drift.
- Future research directions.

Thank you!
Questions?

https://github.com/IQSeC-Lab/LAMDA

Computer Science