THINK UNLIMITED

BÁO CÁO ĐỒ ÁN CUỐI KÌ
CƠ CHẾ HOẠT ĐỘNG CỦA MÃ ĐỘC

DawnGNN: Documentation augmented windows malware detection using graph neural network

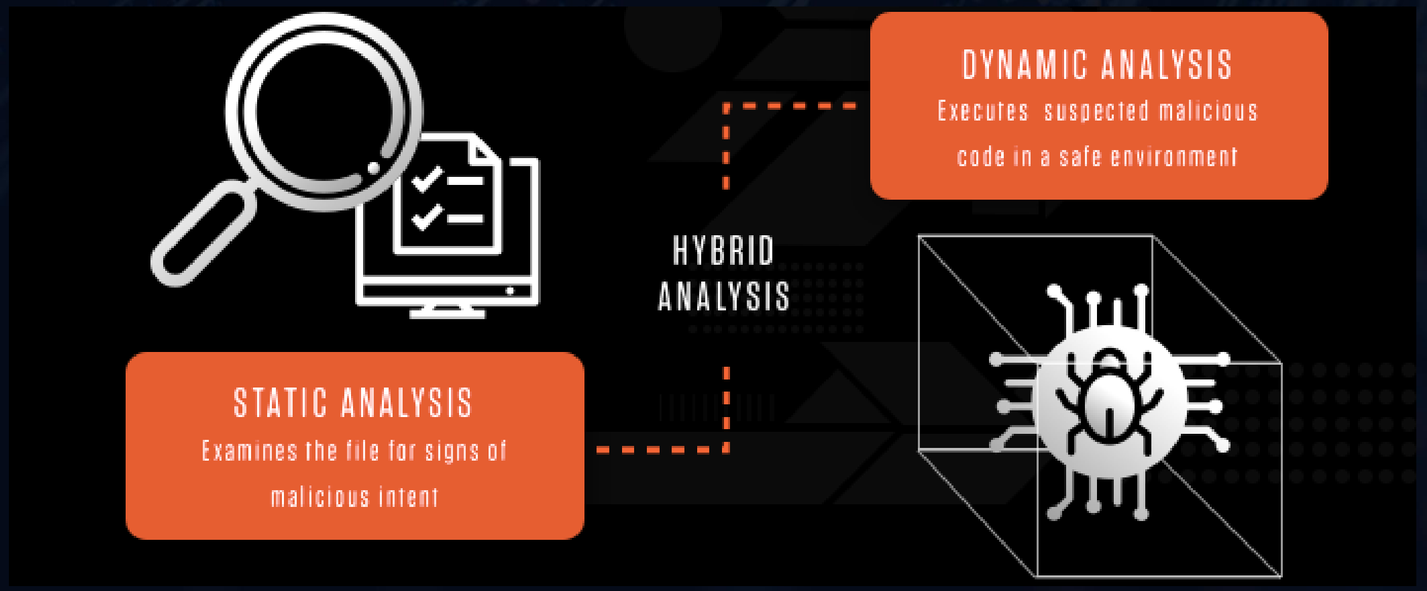
ABOUT US

23520401 - Nguyễn Ngọc Diệu Duyên
23520673 - Đoàn Việt Khải
23520938 - Nguyễn Hoàng Bảo Minh

CẤU TRÚC

MỞ ĐẦU

PAIN POINT



MỞ ĐẦU

PROPOSED METHOD

Thiếu quan tâm đến chức năng và mục đích của các hàm API.

→ **Sử dụng tài liệu API chính thức**

→ **BERT tăng cường thông tin ngữ nghĩa**

Xử lý chuỗi API theo thứ tự thời gian tuyến tính

→ **GNN-GAT khai thác thông tin từ đồ thị gọi API của chương trình.**

TỔNG QUAN MÔ HÌNH

OVERVIEW SYSTEM DESIGN

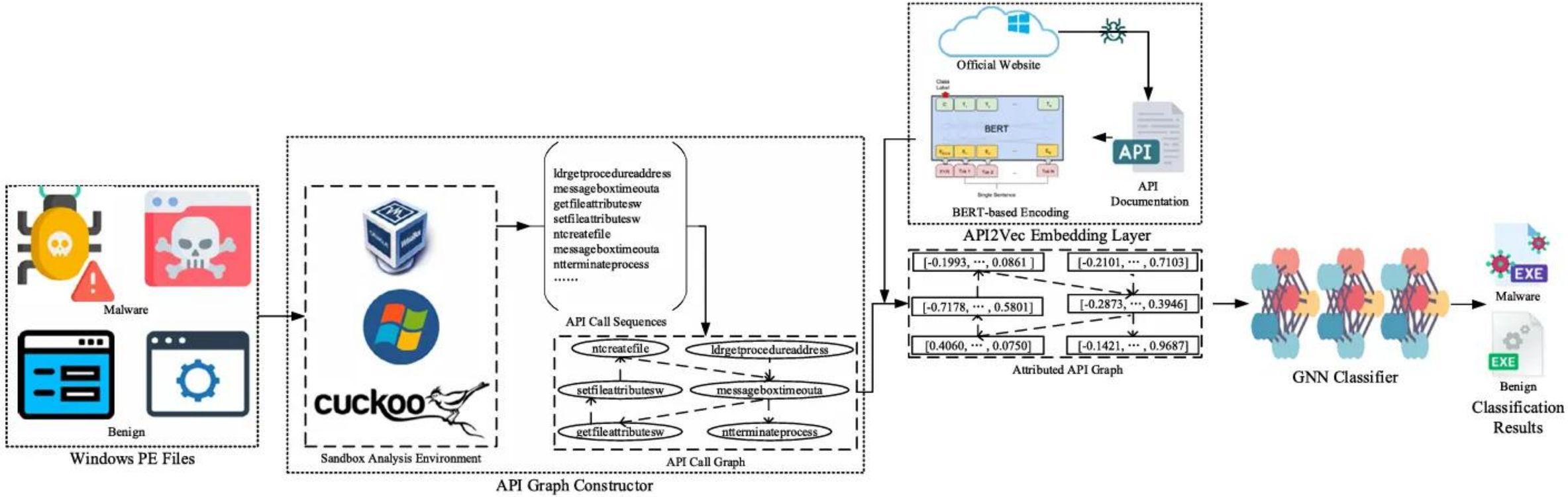
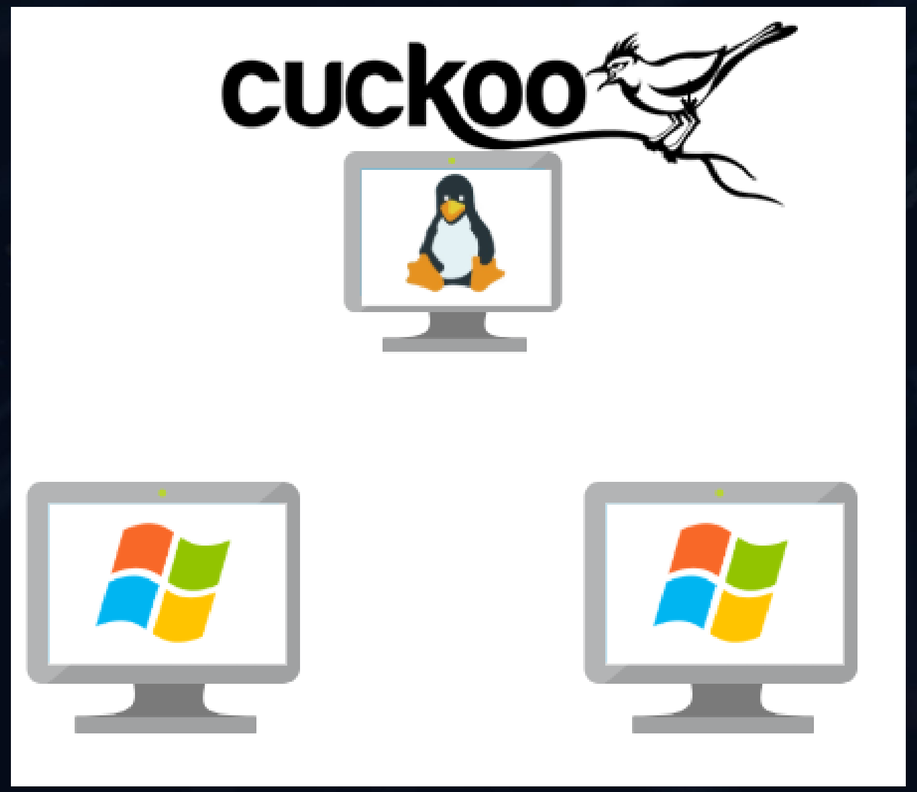


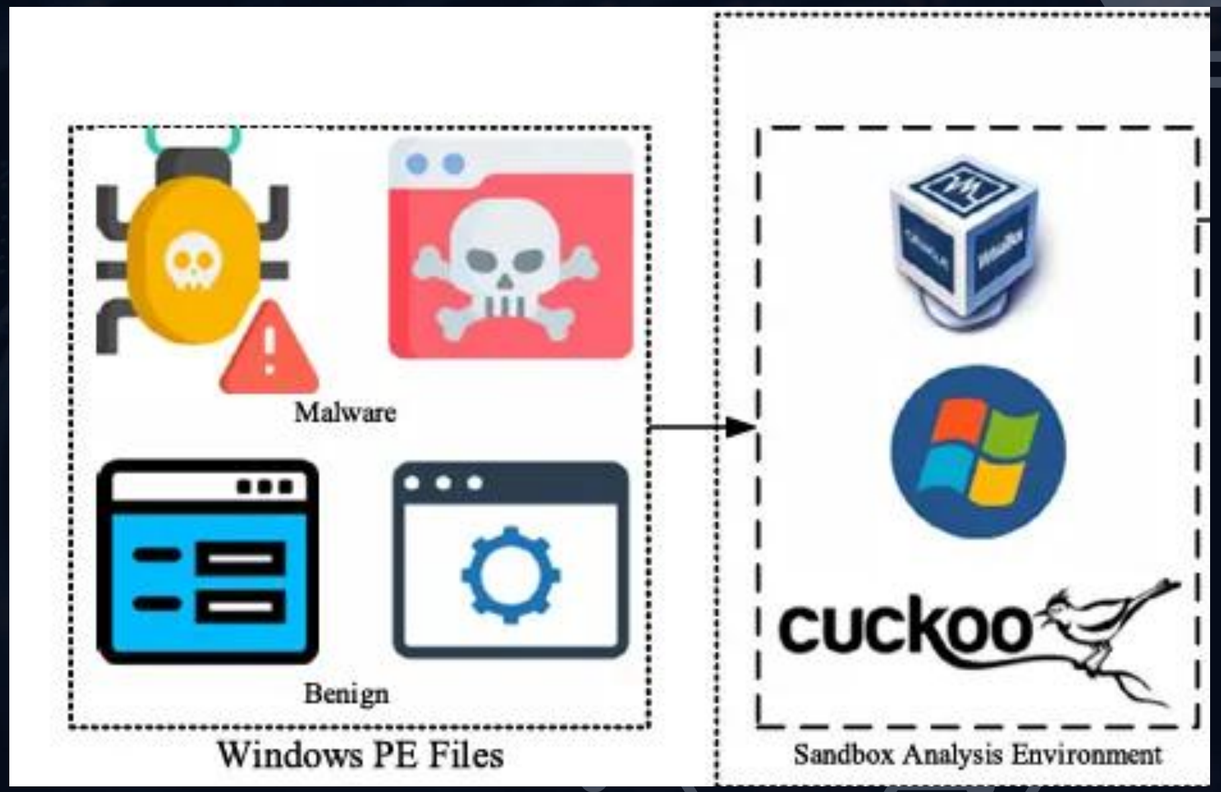
Fig. 3. The architecture of DawnGNN system.

API GRAPH CONSTRUCTOR

API SEQUENCE EXTRACTION



1 Ubuntu Host: Cuckoo sandbox
n Windows VMs: Cuckoo Agents

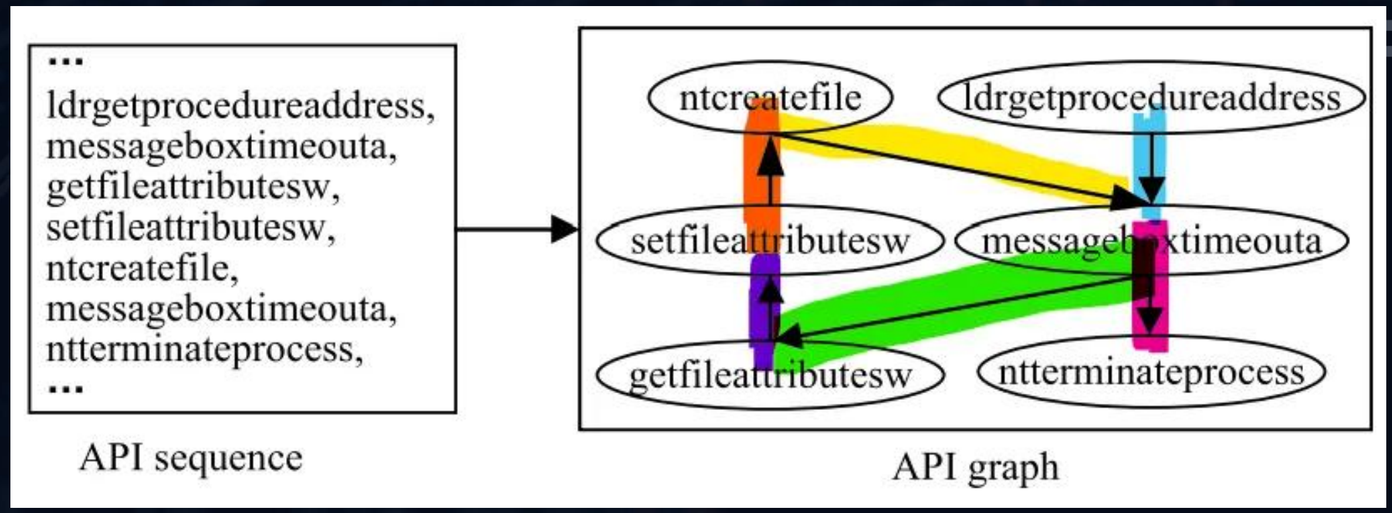


API GRAPH CONSTRUCTOR

API GRAPH CONSTRUCTOR

Chuyển đổi chuỗi API thành đồ thị có hướng g = {V, D}

Nút (V): Các API duy nhất
Cạnh (D): Trình tự thực thi



	ntcreatefile	ldrgetprocedureaddress	setfileattributesw	messageboxtimeouta	getfileattributesw	ntterminateprocess
ntcreatefile						
ldrgetprocedureaddress						
setfileattributesw						
messageboxtimeouta						
getfileattributesw						
ntterminateprocess						

MÔ HÌNH BERT

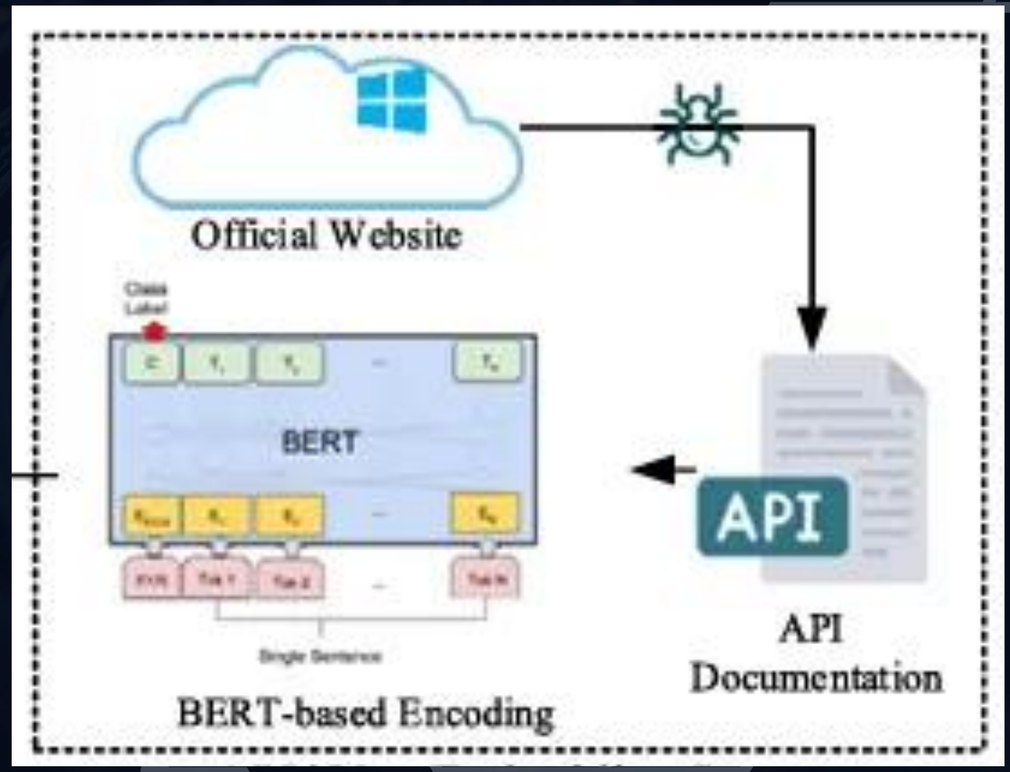
OVERVIEW

Step 01: Crawl API Definition

Step 02: Remove meaningless words of API Definition

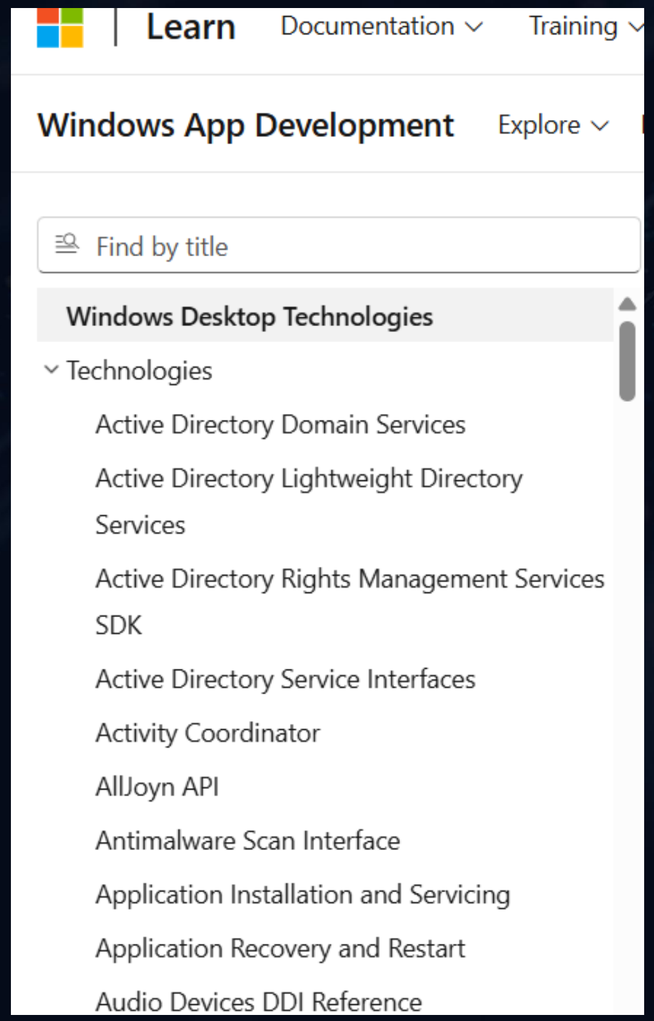
Step 03: Masked Language Model - MLM for BERT

Step 04: Feature Extraction by BERT



MÔ HÌNH BERT

API DOCUMENTATION CORPUS PREPARATION



AmsiNotifyOperation

Sends to the antimalware provider a notification of an arbitrary operation. (AmsiNotifyOperation)

AmsiOpenSession

Opens a session within which multiple scan requests can be correlated.

AmsiResultIsMalware

Determines if the result of a scan indicates that the content should be blocked.

AmsiScanBuffer

Scans a buffer-full of content for malware.

AmsiScanString

Scans a string for malware.

MÔ HÌNH BERT

API DOCUMENTATION CORPUS PREPARATION

CopyFileW

The CopyFileW (Unicode) function (winbase.h) copies an existing file to a new file.

~~The CopyFileW~~ ~~(Unicode)~~ ~~function~~ ~~(winbase.h)~~ **copies an existing file to a new file**

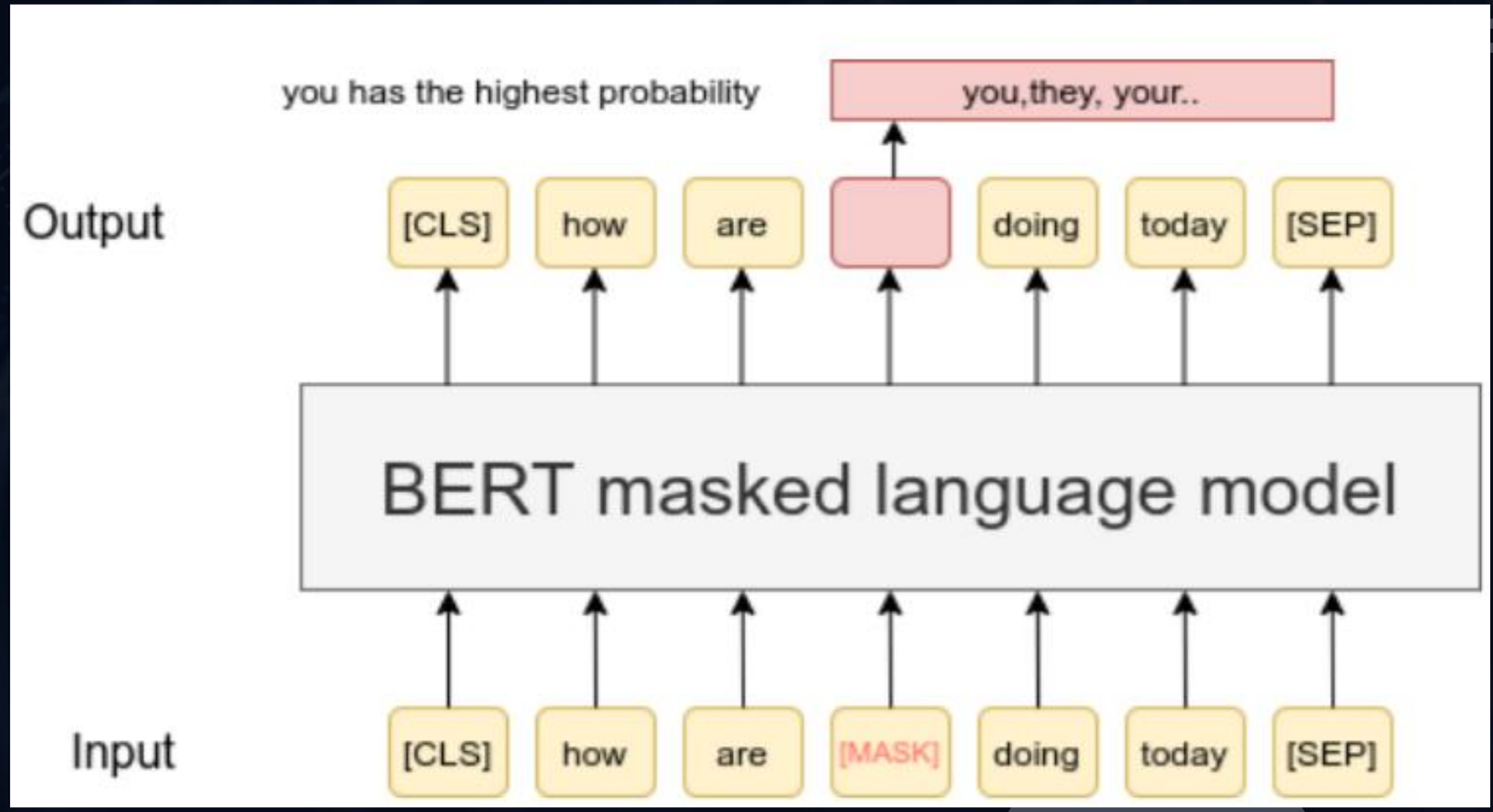
Crawled 37100 API

MÔ HÌNH BERT

MASKED LANGUAGUE MODEL TASK

Xác suất che giấu: 15%
(mlm_probability)

MASK: 80%
Unchange: 10%
Corrupted Token: 10%



MÔ HÌNH BERT

BERT - BASED SEMANTIC FEATURE EXTRACTION

acmDriverDetailsW	queries a specified ACM driver to determine its capabilities.
acmDriverID	returns the handle of an ACM driver identifier associated with a
acmDriverMessage	sends a user-defined message to a given ACM driver instance.

BERT base model (uncased)

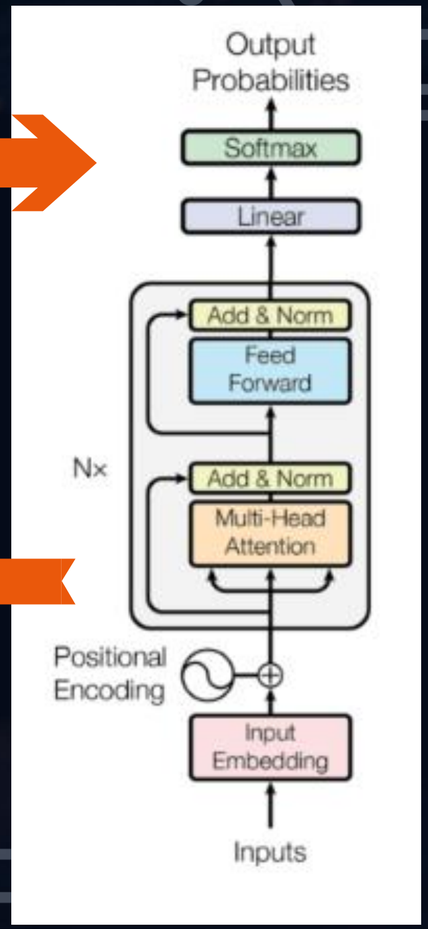
{ API Name : [API Features] }

1 x 768

BERT Small

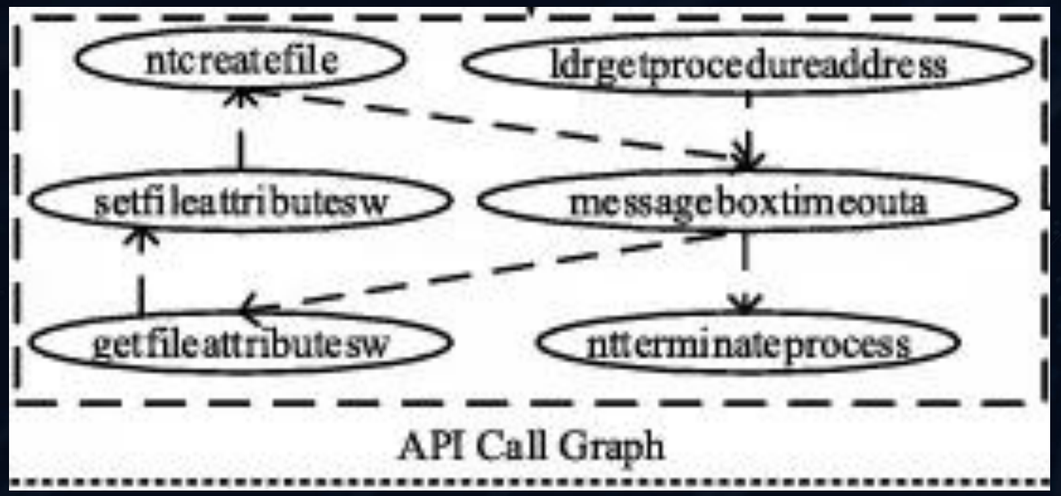
{ API Name : [API Features] }

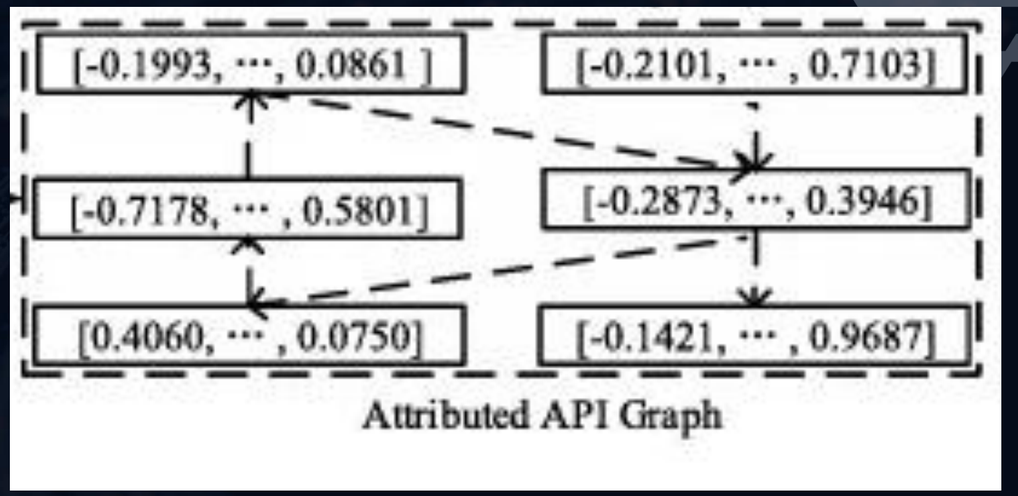
1 x 512

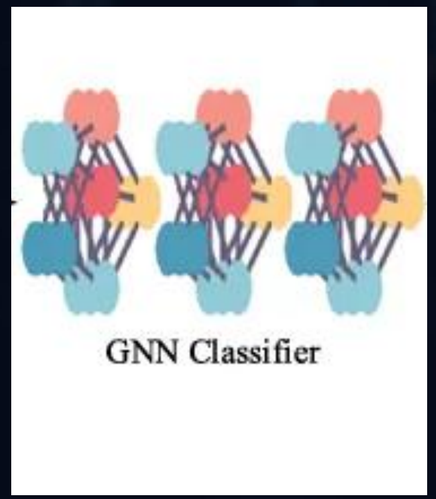


MÔ HÌNH GNN

OVERVIEW



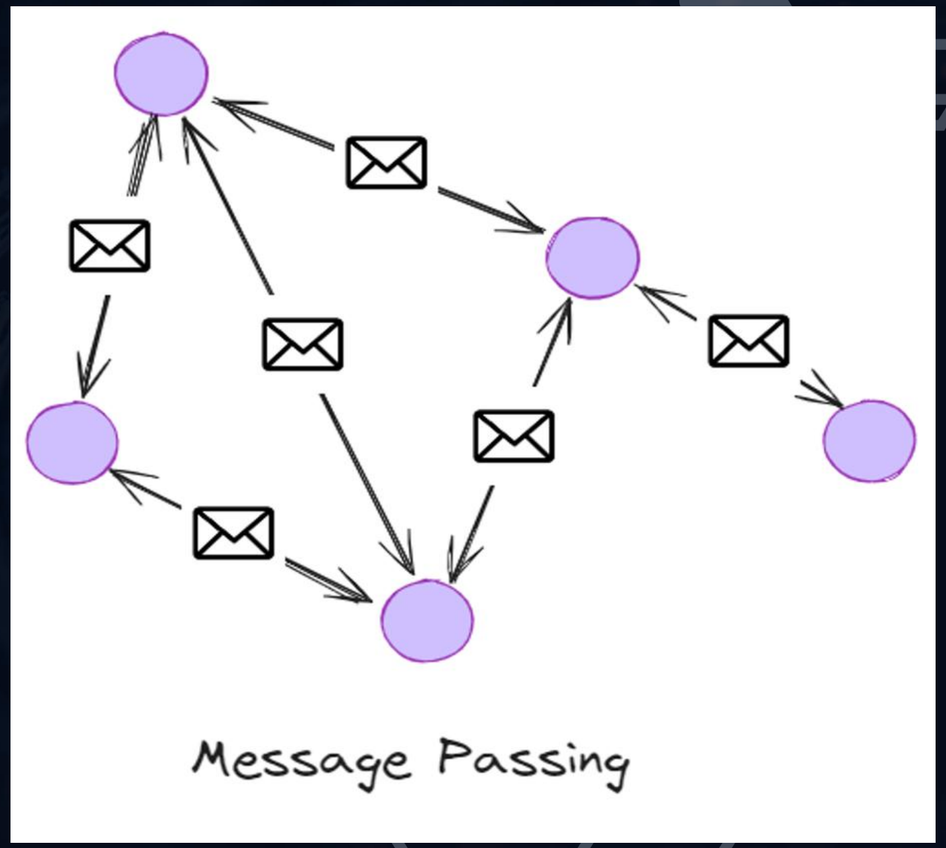




MÔ HÌNH GNN

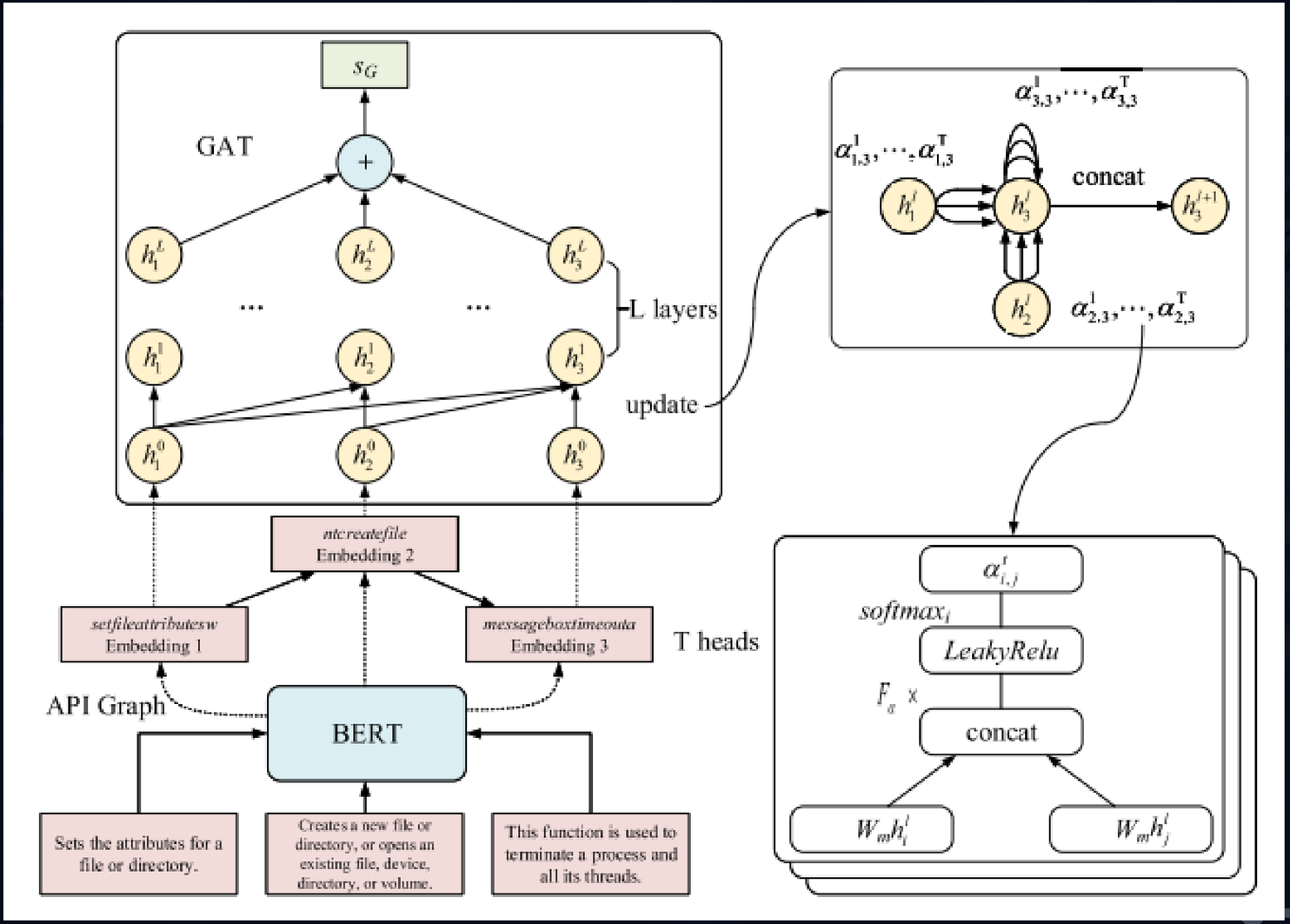
GRAPH NEURAL NETWORK - GNN

Bước 01: Truyền thông điệp - Message Passing đến các nút lân cận.
Bước 02: Tổng hợp thông điệp.
Bước 03: Cập nhật trạng thái.
Bước 04: Lặp lại K lần.



MÔ HÌNH GNN

GRAPH ATTENTION NETWORKS ARCHITECTURE - GAT

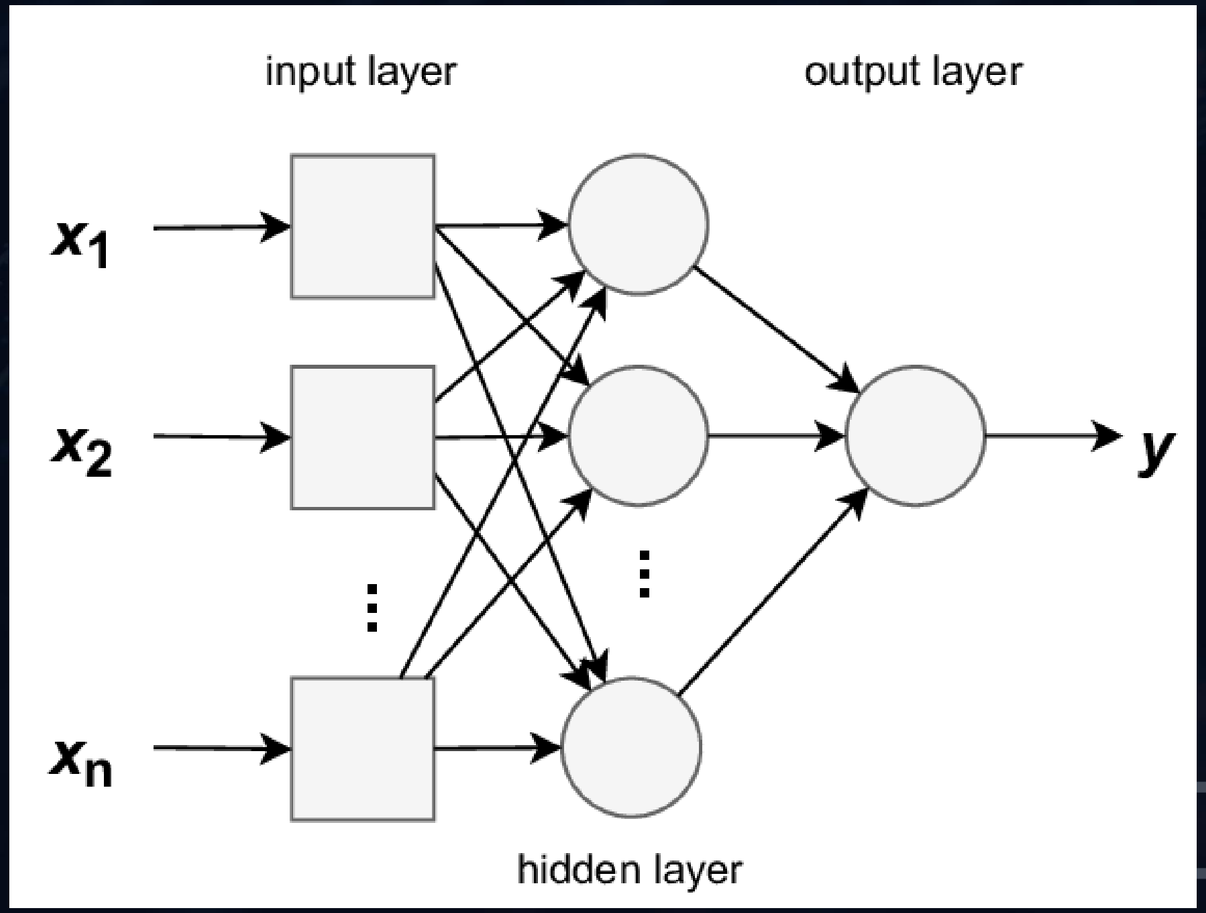


Tính toán mức độ quan trọng của các node lận cận.

Multi - head attention

MÔ HÌNH GNN

MLP



THỰC NGHIỆM & KẾT QUẢ

OVERVIEW & ENVIRONMENT SET UP

	Model	Dataset	Evaluation Metrics	Goals
Experimental Setup 01	BERT based uncased BERT small	MalBehavD-V1 PE_APICALLS APIMDS	Accuracy, Precision F1 Score, Recall	Hiện thực Flow chính của tác giả.
Experimental Setup 02	GCN, GIN WORD2VEC + GAT BERTbase + LSTM	MalBehavD-V1	Accuracy, Precision F1 Score, Recall	Chứng minh sự ưu việt của DawnGNN.
Experimental Setup 03	BERT based uncased BERT small	API Call Sequences Malware	Accuracy, Precision F1 Score, Recall	Kiểm tra tính hiệu quả của mô hình trên Dataset mới.

THỰC NGHIỆM & KẾT QUẢ

DATASET 01: WINDOWS_PE_APICALLS

Malware	APIcalls
Backdoor	SetUnhandledExceptionFilter, GetSystemTimeAsFileTime, NtDelayExecution, GetSystemTimeAsFileTime, NtDelayExecution, GetSystemTimeAsFileTime,...
benign	LdrGetDllHandle, NtTerminateProcess, NtClose, LdrGetDllHandle, NtClose, NtClose, NtClose, NtClose, NtClose,...
Trojan	NtOpenFile, NtCreateSection, NtClose, LdrLoadDll, LdrGetProcedureAddress, LdrLoadDll, LdrGetProcedureAddress, LdrGetDllHandle,...
Virus	LdrLoadDll, LdrGetProcedureAddress, LdrGetProcedureAddress, LdrGetProcedureAddress, LdrGetProcedureAddress, LdrGetProcedureAddress,...
Worm	NtProtectVirtualMemory, SHGetFolderPathW, NtQueryAttributesFile, GetSystemTimeAsFileTime, NtDelayExecution, MoveFileWithProgressW,...
Backdoor	GetSystemTimeAsFileTime, NtDelayExecution, GetSystemTimeAsFileTime, NtDelayExecution, GetSystemTimeAsFileTime, NtDelayExecution,...
Worm	NtProtectVirtualMemory, SHGetFolderPathW, NtQueryAttributesFile, GetSystemTimeAsFileTime, NtDelayExecution, MoveFileWithProgressW,...

Trước

```
print("Original dataset size:", df.shape)
```

✓ 0.1s

```
Original dataset size: (552, 2)
```

Sau

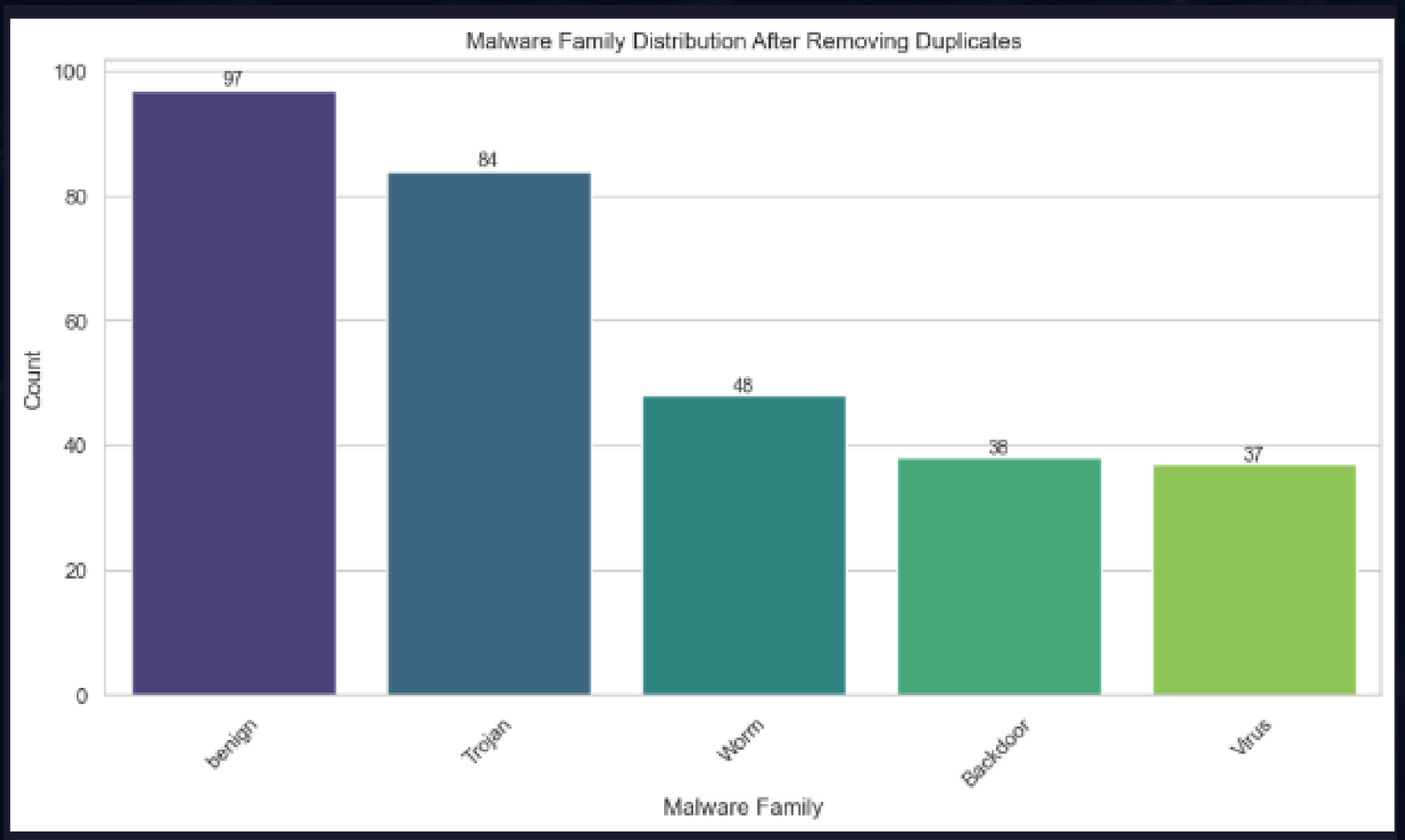
```
df.drop_duplicates(inplace=True)
df.shape
```

✓ 0.0s

```
(304, 2)
```

THỰC NGHIỆM & KẾT QUẢ

DATASET 01: WINDOWS_PE_APICALLS



THỰC NGHIỆM & KẾT QUẢ

DATASET 02: MalBehavD-V1

sha256	labels	0	1	2	....	Unnamed 173	Unnamed 174
5c18291c481a192ed5...	0	LdrUnloadDll	RegCloseKey	NtOpenSection	.....	NaN	NaN
4683faf3da550ffb594...	0	NtOpenMutant	NtOpenSection	CoUninitialize	.....	NaN	NaN
9a0aea1c7290031d7c...	0	GetForegroundWindow	LoadStringW	GetFileType	.....	NaN	NaN
.....	.....	.....	.....	.....	.....	.....	.....
e0f3e4d5f50afd9c31e...	1	CreateToolhelp32Snapshot	NtOpenSection	CreateThread	.....	NaN	NaN
ec2b6d29992f13e740...	1	CreateToolhelp32Snapshot	NtOpenSection	CreateThread	.....	NaN	NaN

Trước

```
print("Original Dataset size:", df.shape)
```

0.0s

```
Original Dataset size: (2570, 177)
```

Sau

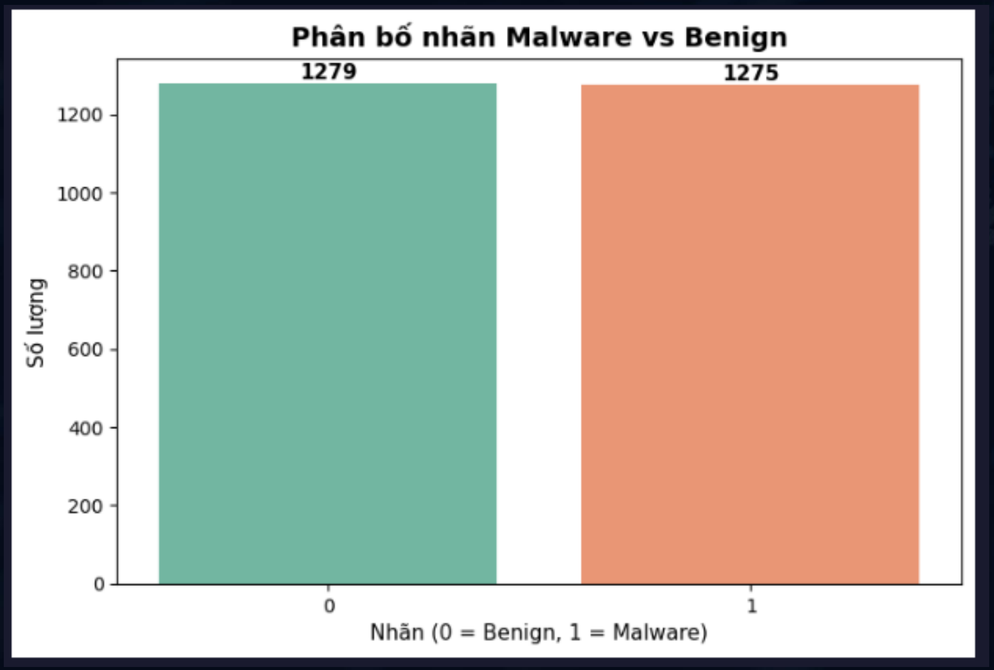
```
df.drop_duplicates(inplace=True)
print("Dataset size after removing duplicates:", df.shape)
```

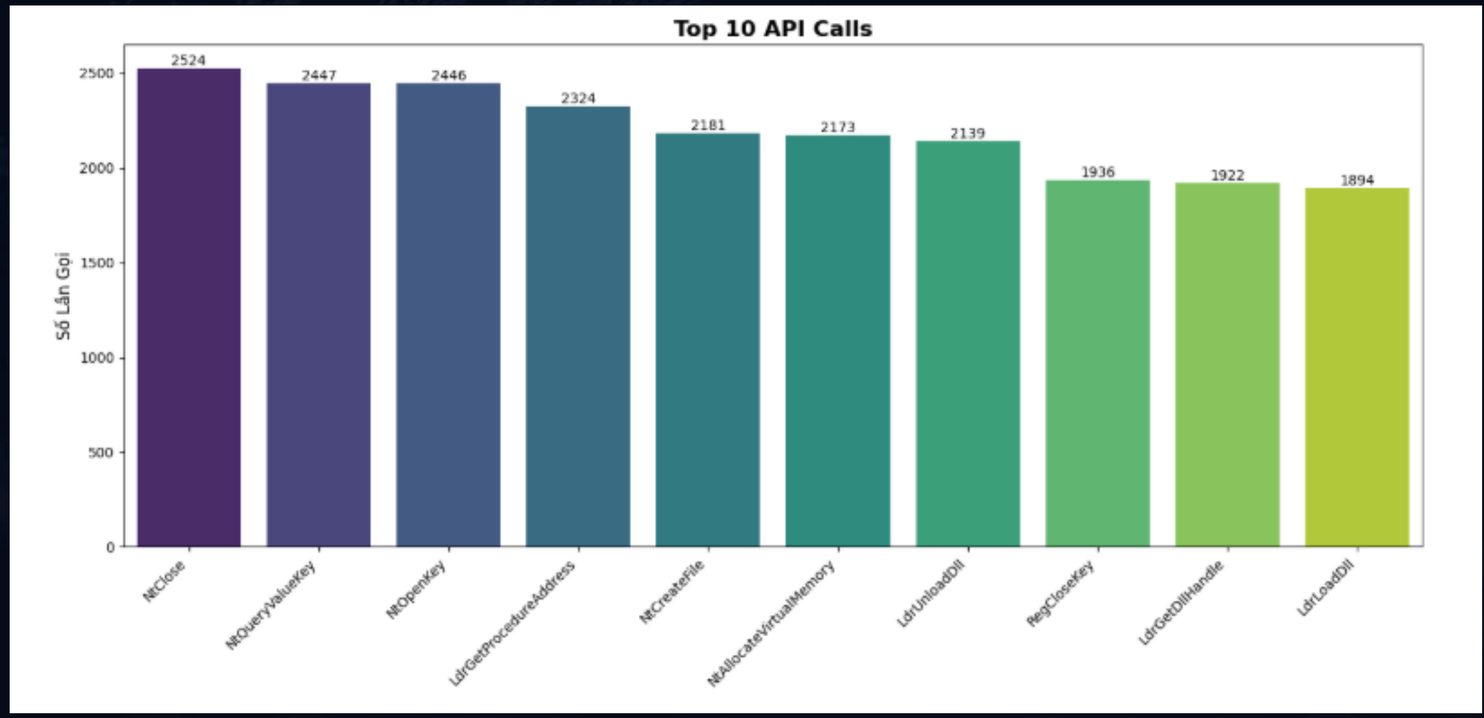
0.0s

```
Dataset size after removing duplicates: (2554, 177)
```

THỰC NGHIỆM & KẾT QUẢ

DATASET 02: MalBehavD-V1





THỰC NGHIỆM & KẾT QUẢ

DATASET 03: APIMDS

Malware.exe
Benign.exe → Trích xuất API ✗

Search → Repos ✓

```
print("Original Dataset size:", df.shape)
```

✓ 1.2s

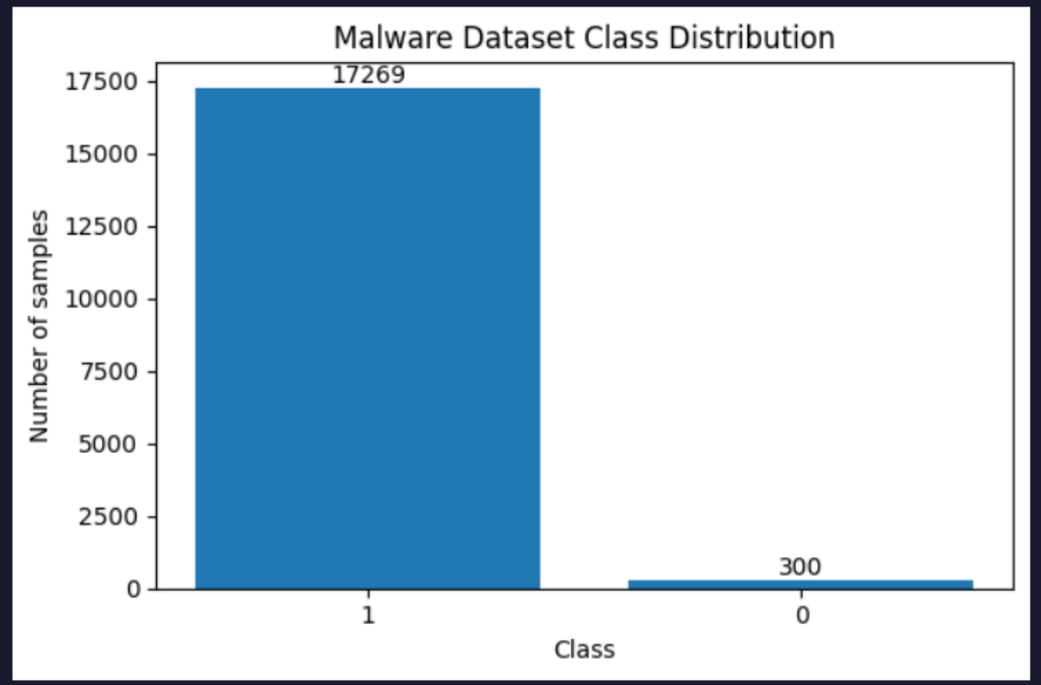
```
Original Dataset size: (17569, 597)
```

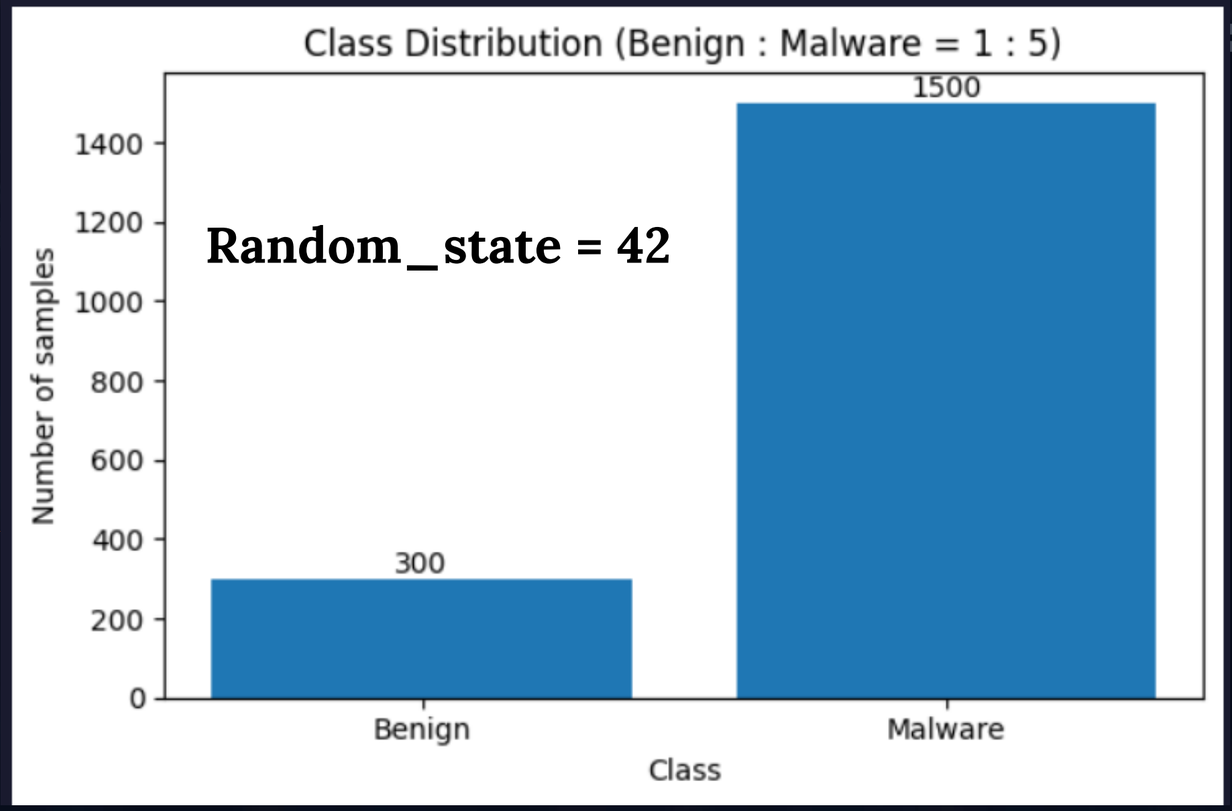
label	hash	0	1	...
1	034691b9a0558842....	CreateEventW	GetCommandLineA	...
1	0f74250f63874840...	CreateEventW	DisableThreadLibraryCalls	...
...	...	...	...	...
0	1772d592e6115cc84f....	RegCloseKey	RegQueryValueExW	...

THỰC NGHIỆM & KẾT QUẢ

DATASET 03: APIMDS

Imbalance data ???





Chia subset cho malware

THỰC NGHIỆM & KẾT QUẢ

DATASET 04: ANALYSIS DATASETS: API CALL SEQUENCES MALWARE

```
data = pd.read_csv("dynamic_api_call_sequence_per_malware_100_0_306.csv")
print("Original data size:", data.shape)
```

✓ 0.3s

```
Original data size: (43876, 102)
```

hash	t0	t1	...	t99	Malware
071e8c3f8922e18...	112	274	...	35	1
33f8e6d08a6aae9...	82	208	...	112	1
....	...	....	...	...	...
654139d715abcf7...	82	280	...	141	1
078c9d4e7be4819...	112	274	...	71	1

Mapping lại

hash	t0	t1	...	t99	Malware
071e8c3f8922e18...	RegOpenKeyExA	NtOpenKey		GetSystemMetrics	1
33f8e6d08a6aae9...	GetSystemTimeAsFileTime	NtAllocateVirtualMemory		RegOpenKeyExA	1
....	...	....	...	...	...

API được đánh thành số

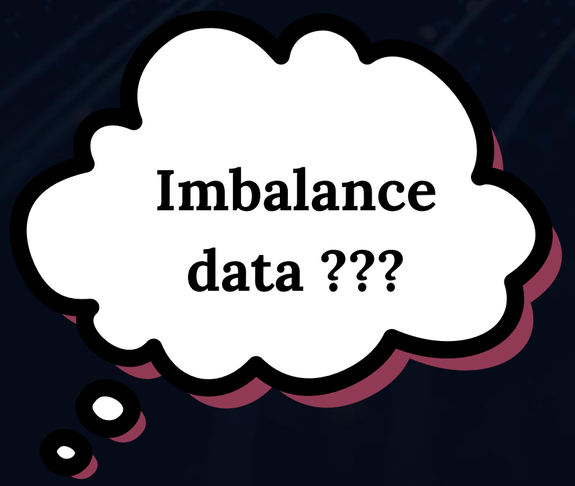
INSTRUCTIONS: ieee-dataport.org

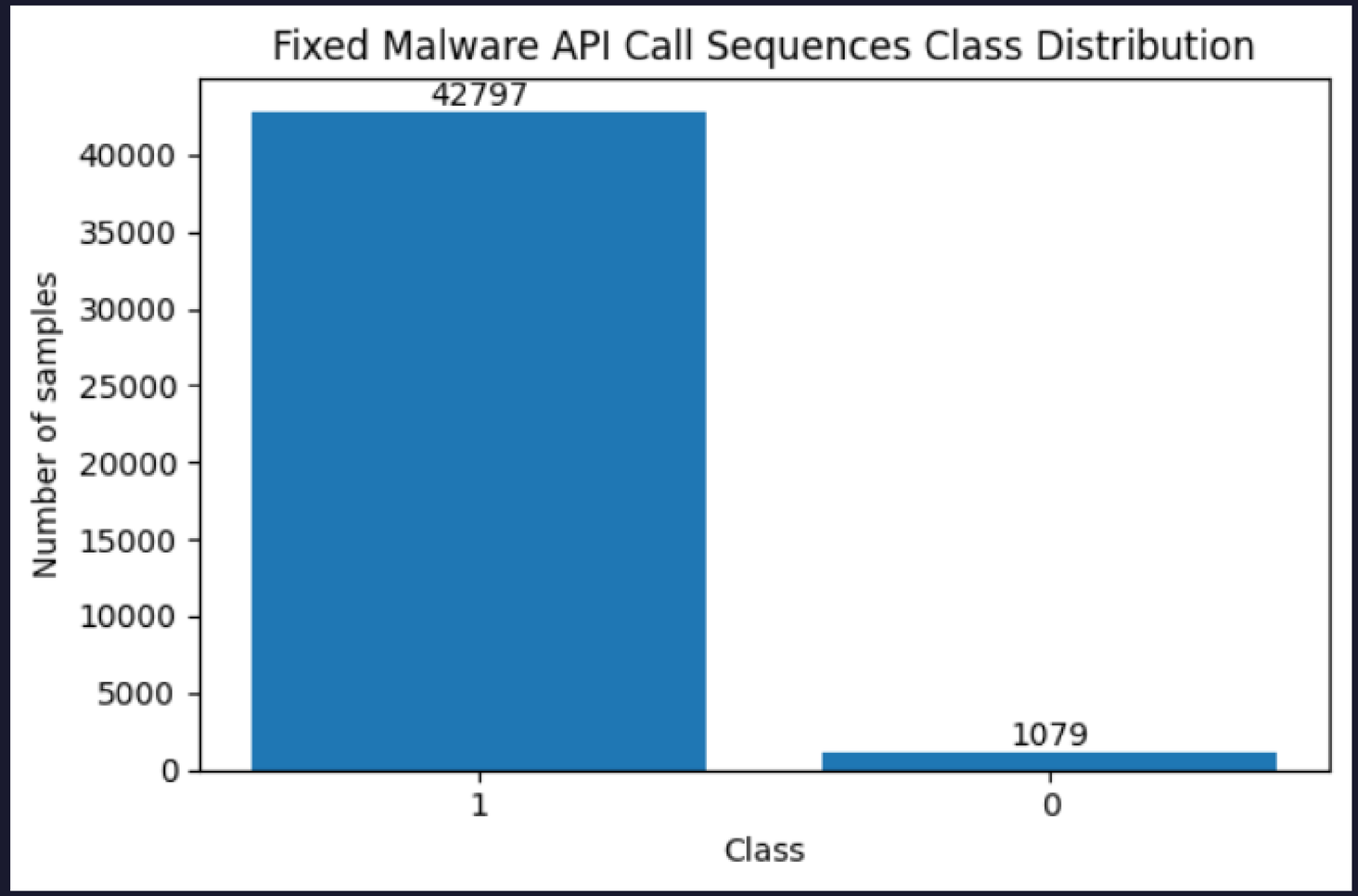
* FEATURES * Column name: hash Description: MD5 hash of the example Type: 32 bytes string Column name: t_0 ... t_99 Description: API call Type: Integer (0-306) Column name: malware Description: Class Type: Integer: 0 (Goodware) or 1 (Malware) API Calls: ['NtOpenThread', 'ExitWindowsEx', 'FindResourceW', 'CryptExportKey', 'CreateRemoteThreadEx', 'MessageBoxTimeoutW', 'InternetCrackUrlW', 'StartServiceW', 'GetFileSize', 'GetVolumeNameForVolumeMountPointW', 'GetFileInformationByHandle', 'CryptAcquireContextW', 'RtlDecompressBuffer', 'SetWindowsHookExA', 'RegSetValueExW',

THỰC NGHIỆM & KẾT QUẢ

DATASET 04: ANALYSIS DATASETS: API CALL SEQUENCES MALWARE

Phân phối dữ liệu





THỰC NGHIỆM & KẾT QUẢ

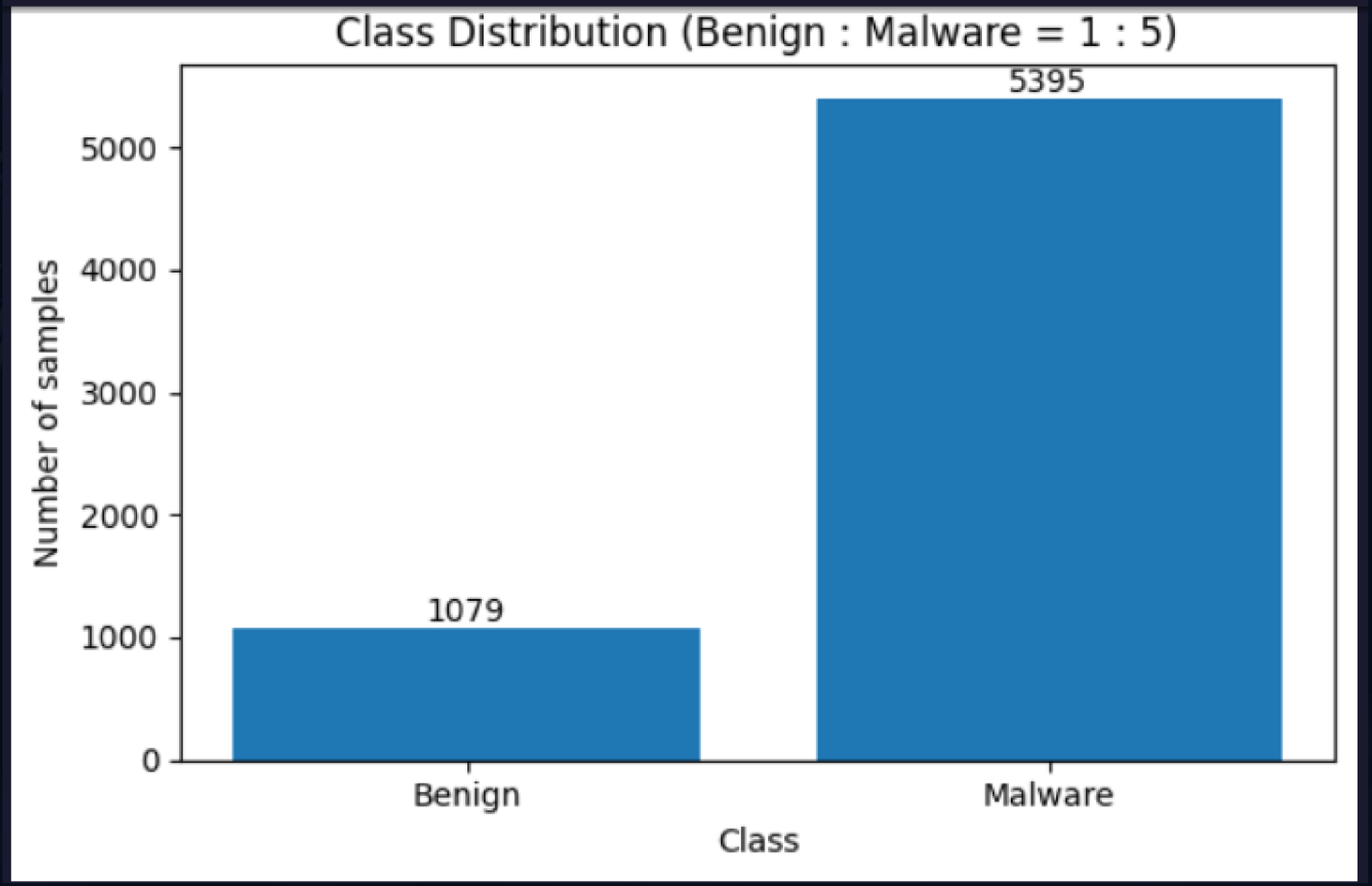
DATASET 04: ANALYSIS DATASETS: API CALL SEQUENCES MALWARE

Xử lí imbalance

Chia subset cho malware

Benign : Malware = 1 : 5

Random_state = 42



THỰC NGHIỆM & KẾT QUẢ

EVALUATION METRICS

ACTUAL VALUES	POSITIVE	NEGATIVE
POSITIVE	TP	FN
NEGATIVE	FP	TN

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 01: PERFORMANCE OF MALWARE DETECTION

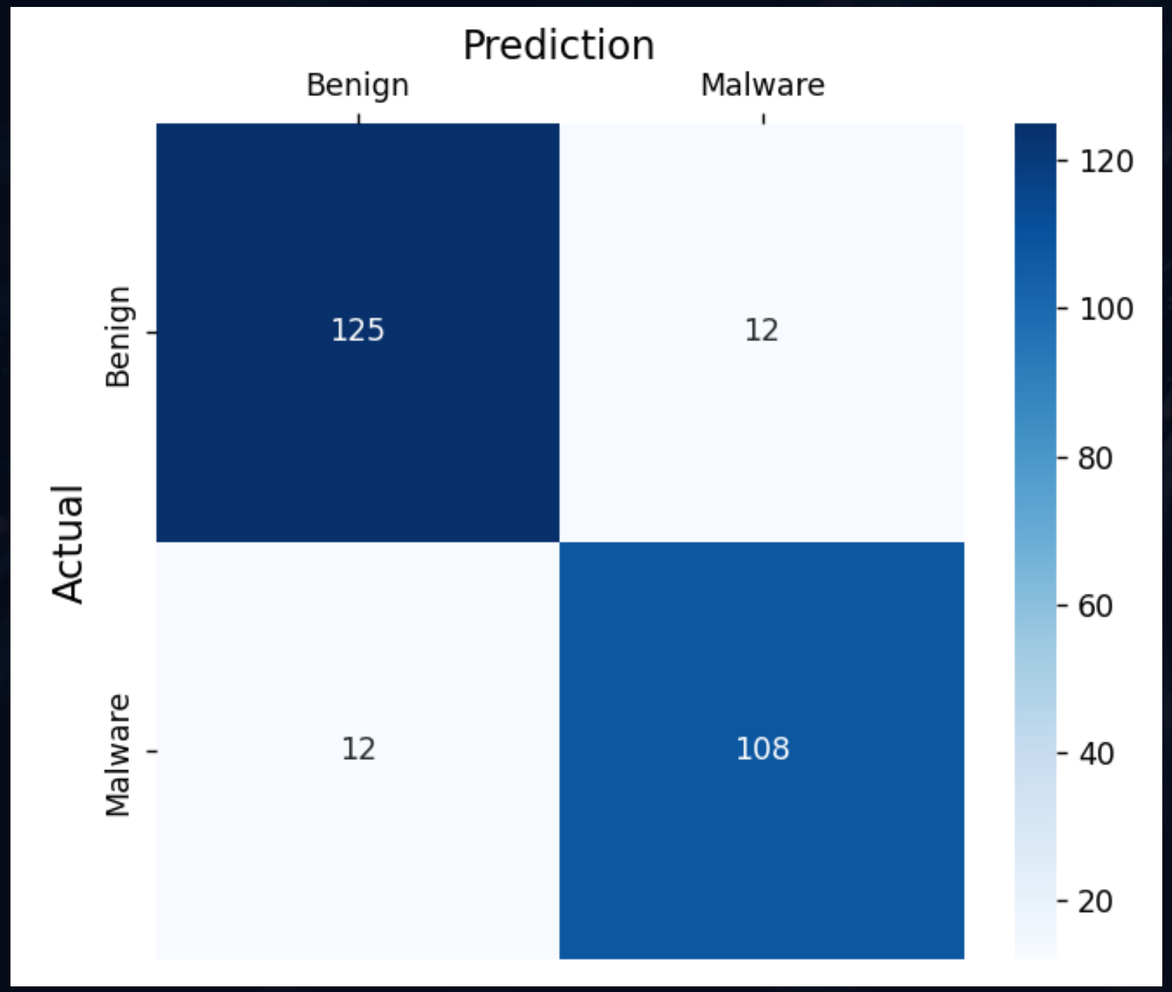
Thực nghiệm của Tác giả

Thực nghiệm của Nhóm

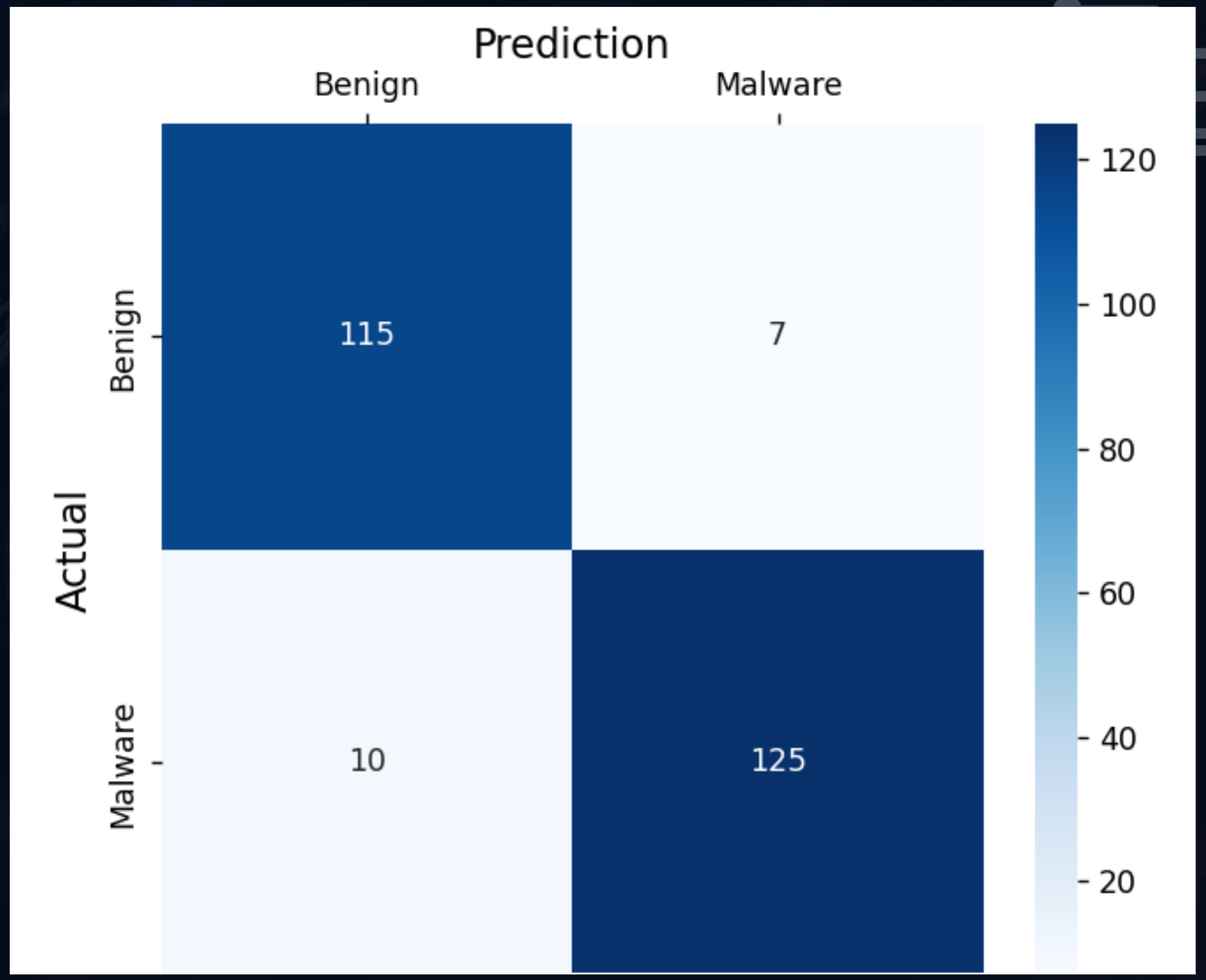
Model	Precision	Recall	F1 - Score	Accuracy	Dataset
BERTsmall + GAT	0.9667	0.9756	0.9683	0.9607	MalBehavD-V1
BERTsmall + GAT	0.9066	0.9000	0.9000	0.9000	MalBehavD-V1
BERTbase + GAT	0.9697	0.9788	0.9711	0.9638	MalBehavD-V1
BERTbase + GAT	0.9470	0.9259	0.9363	0.9339	MalBehavD-V1

THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 01: PERFORMANCE OF MALWARE DETECTION



$BERT_{small}$ + GAT



$BERT_{base}$ + GAT

EXPERIMENTAL 01: PERFORMANCE OF MALWARE DETECTION →

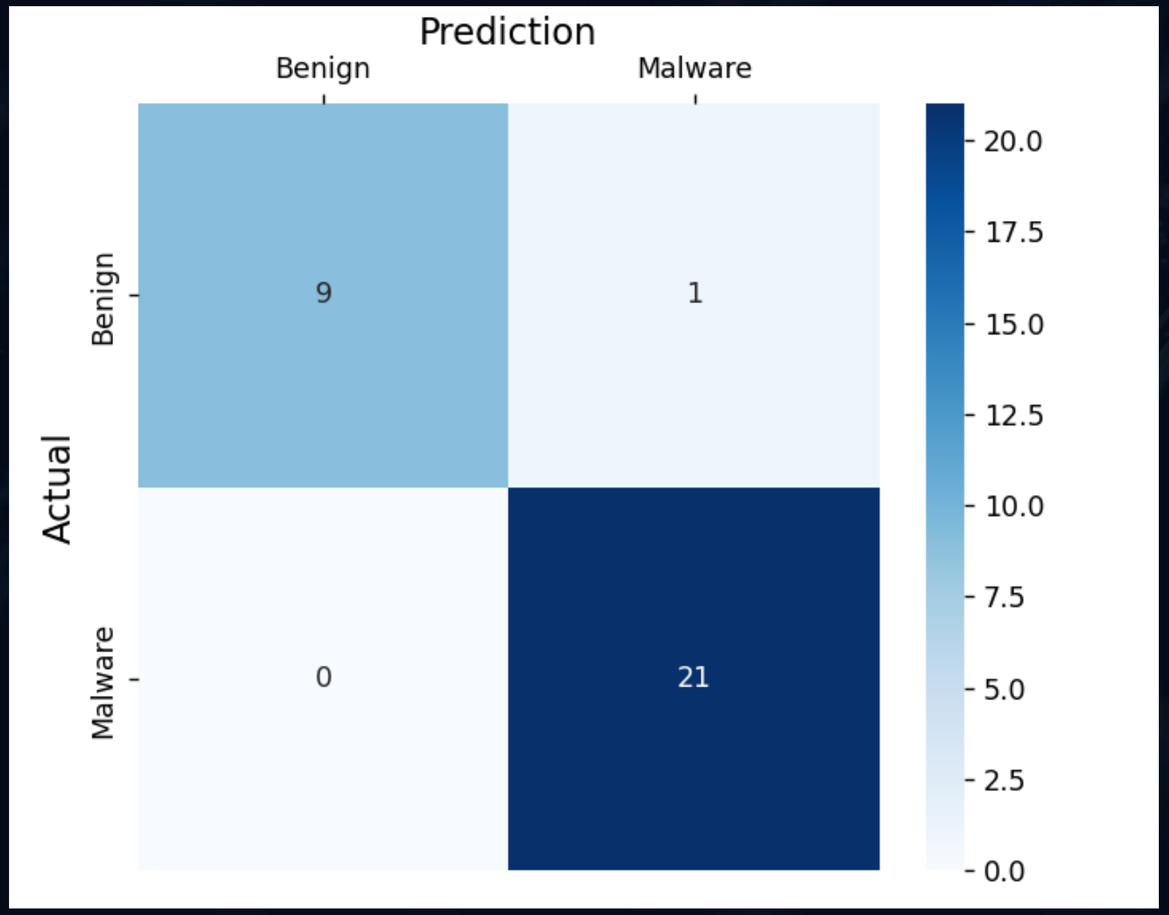
Thực nghiệm của Tác giả

Thực nghiệm của Nhóm

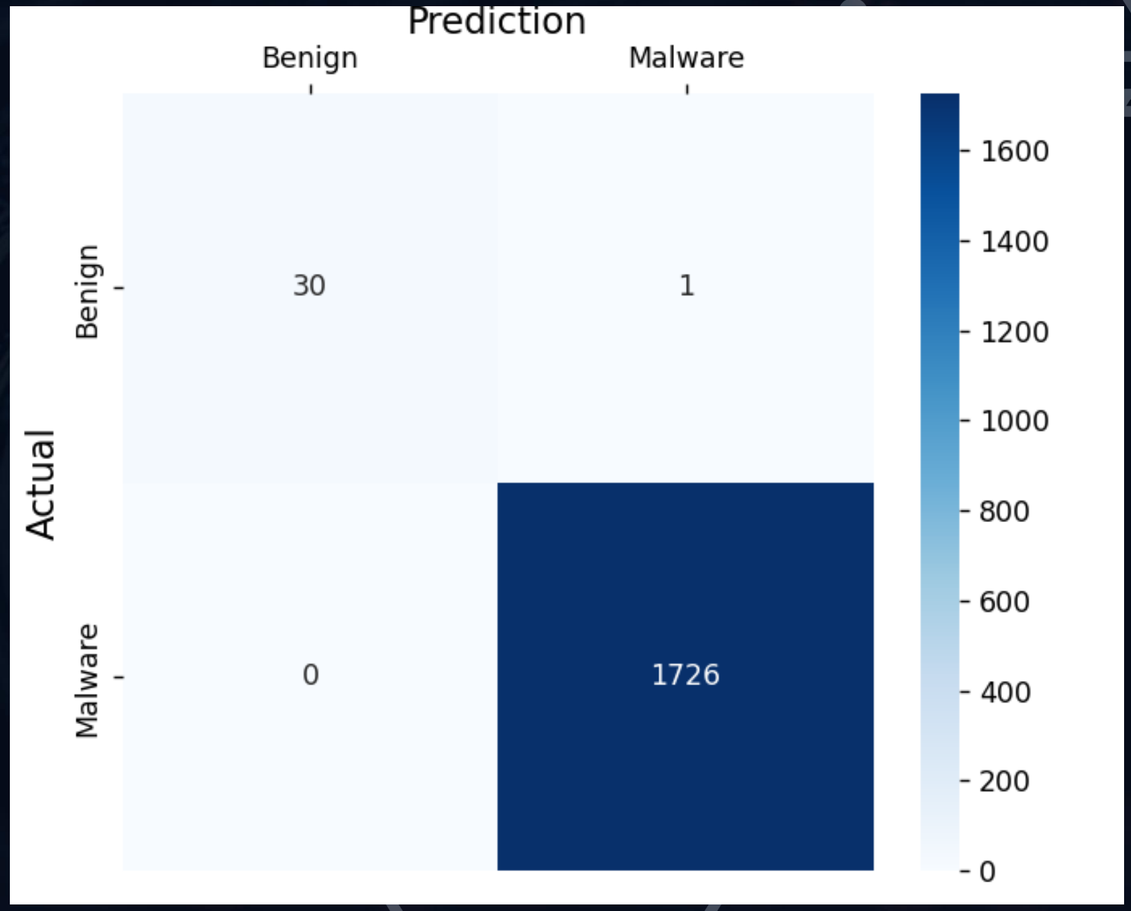
Model	Precision	Recall	F1 - Score	Accuracy	Dataset
BERTbase + GAT	0.9722	1.0000	0.9855	0.9762	PE_APICALLS
BERTbase + GAT	0.9545	1.0000	0.9767	0.9677	PE_APICALLS
BERTbase + GAT	0.9969	1.0000	0.9984	0.9975	APIMDS
BERTbase + GAT (Full Dataset)	0.9994	1.0000	0.9997	0.9994	APIMDS
BERTbase + GAT	1.0000	1.0000	1.0000	1.0000	APIMDS

THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 01: PERFORMANCE OF MALWARE DETECTION



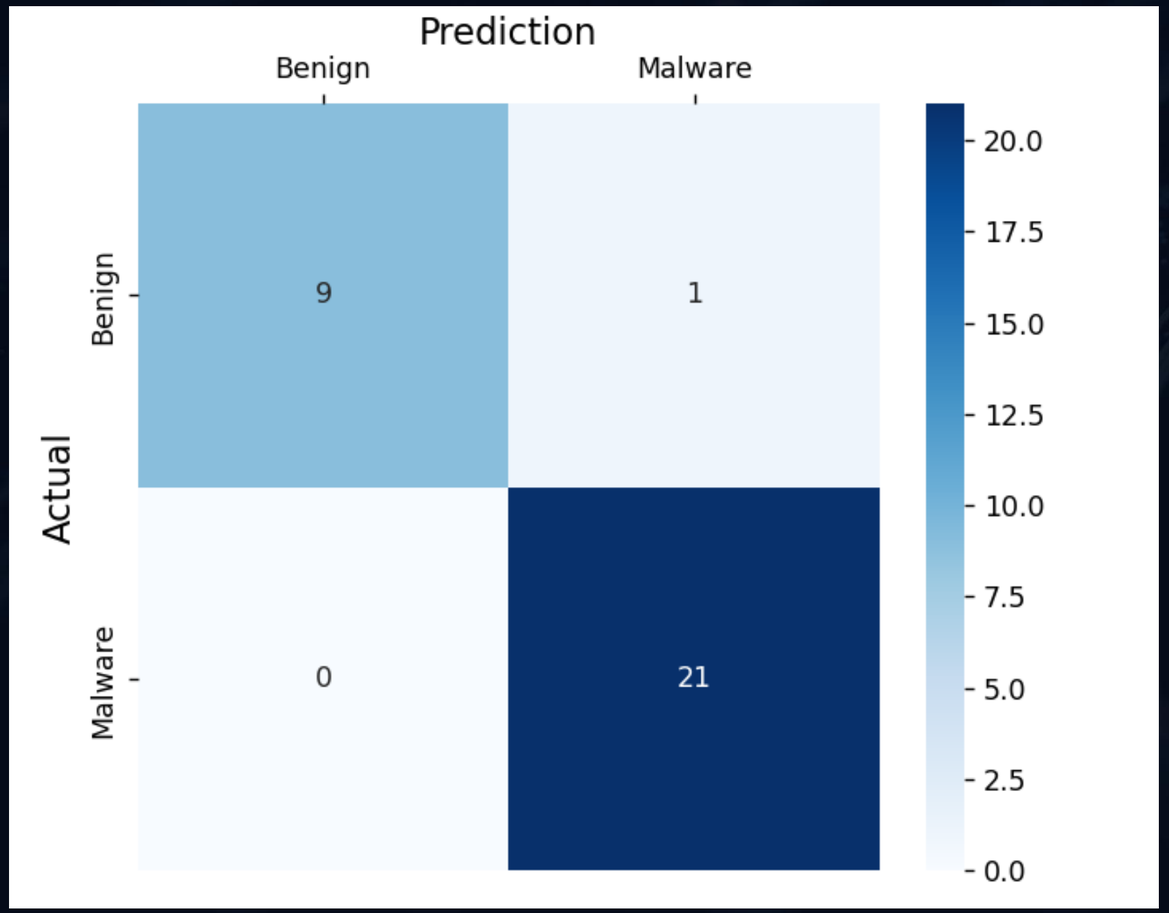
PE_APICALLS



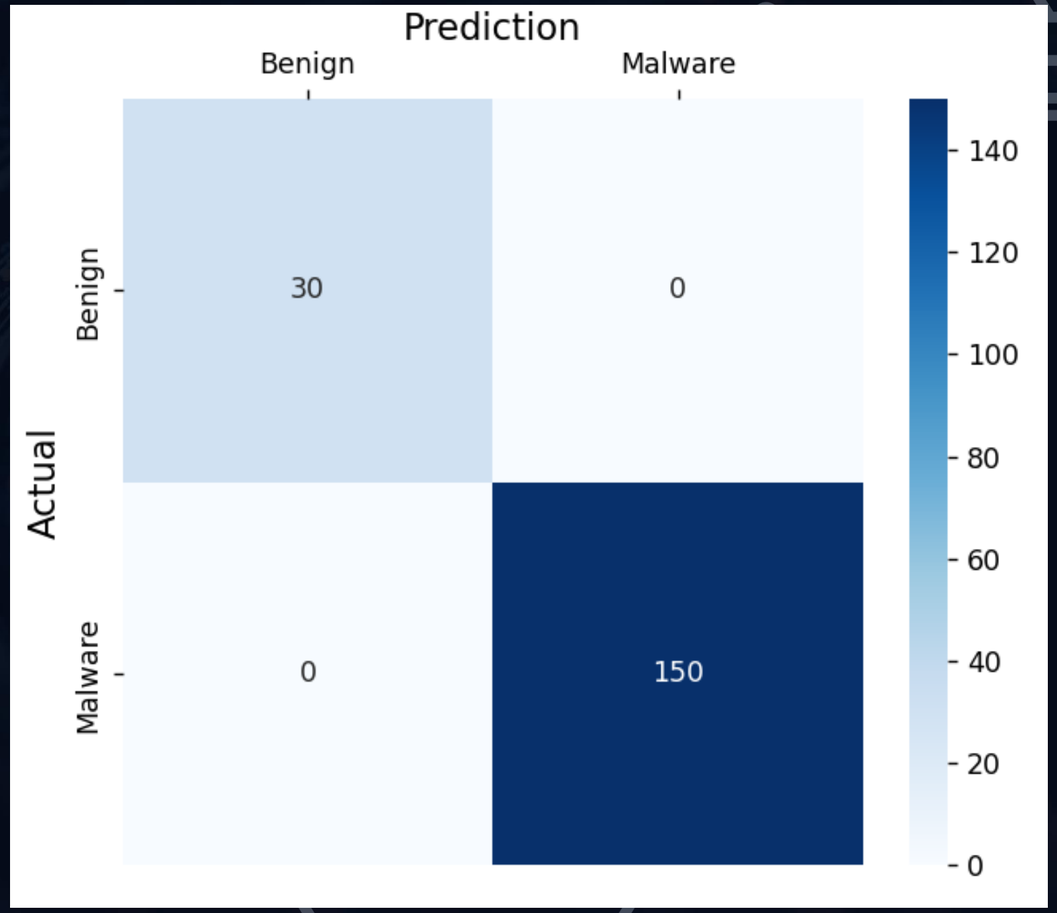
APIMDS full Dataset

THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 01: PERFORMANCE OF MALWARE DETECTION



PE_APICALLS



APIMDS (Subset)

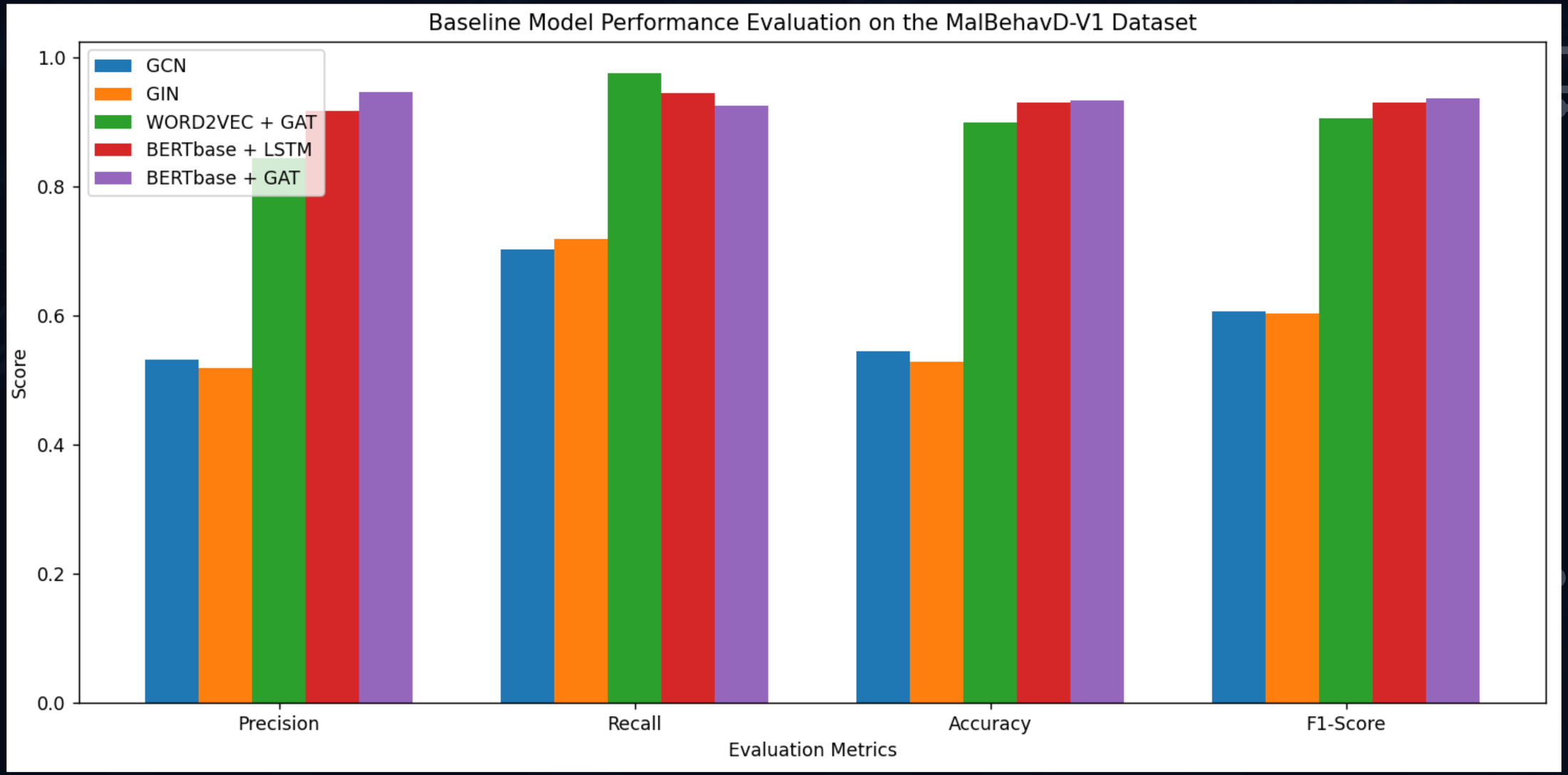
THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 02: COMPARISON WITH BASELINE MODELS

Model	Precision	Recall	F1 - Score	Accuracy	Dataset
GCN	0.5325	0.7031	0.6061	0.5447	MalBehavD-V1
GIN	0.5198	0.7188	0.6032	0.5292	MalBehavD-V1
WORD2VEC + GAT	0.8446	0.9765	0.9058	0.8988	MalBehavD-V1
BERTbase + LSTM	0.9167	0.9453	0.9307	0.9300	MalBehavD-V1
BERTbase + GAT	0.9470	0.9259	0.9363	0.9339	MalBehavD-V1

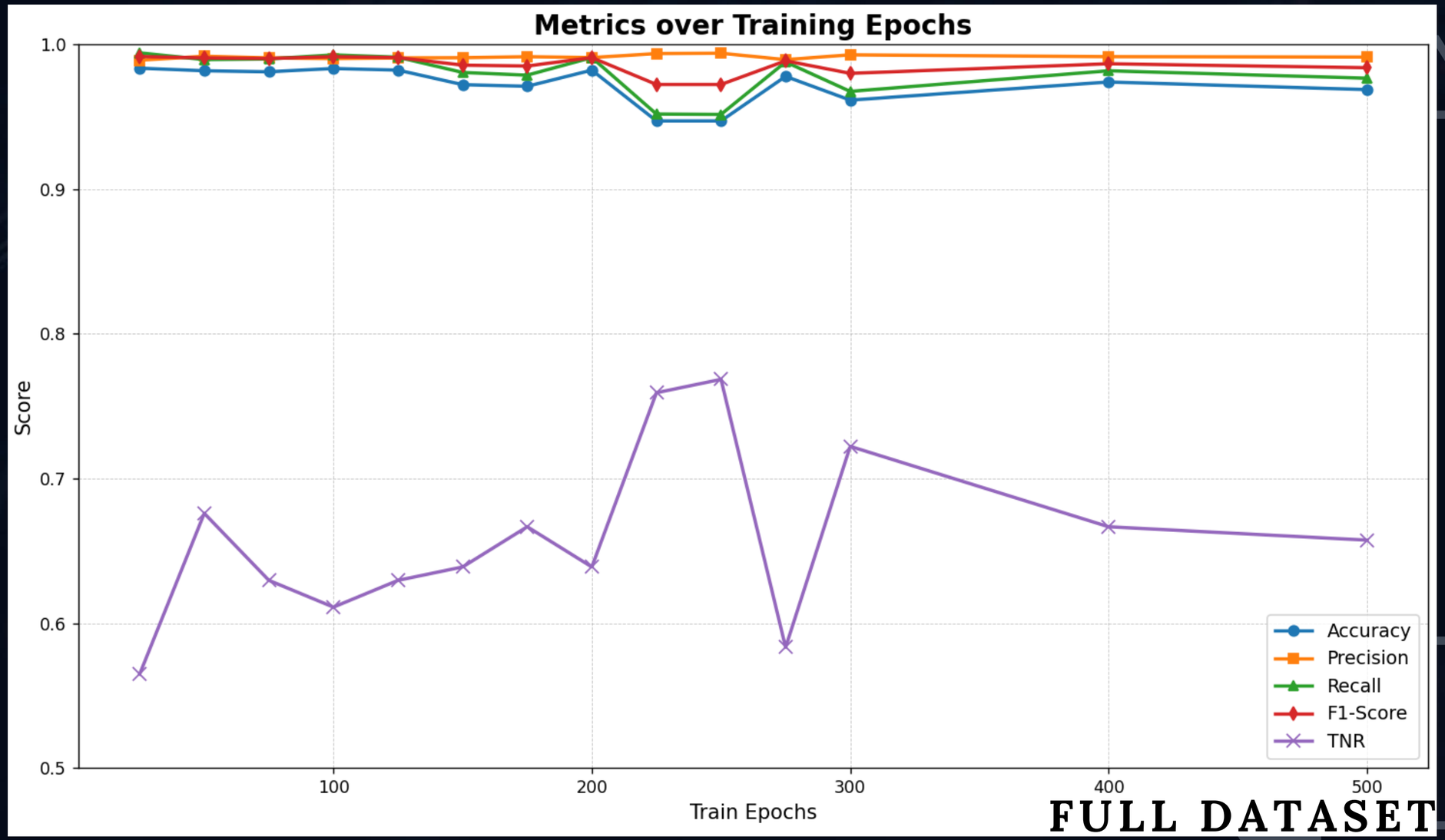
THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 02: COMPARISON WITH BASELINE MODELS



THỰC NGHIỆM & KẾT QUẢ

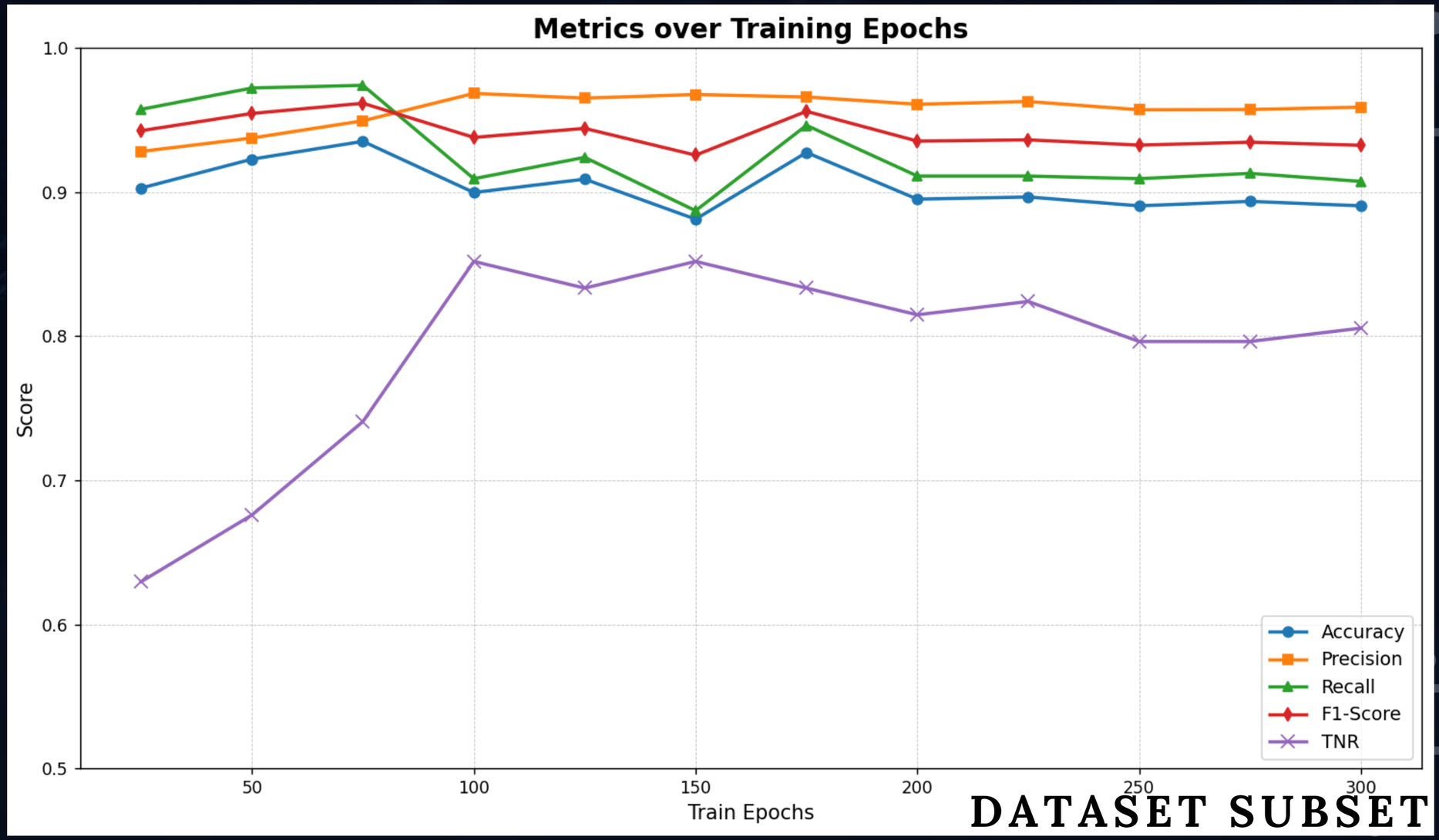
EXPERIMENTAL 03: EXTERNAL VALIDATION & ROBUSTNESS



FULL DATASET

THỰC NGHIỆM & KẾT QUẢ

EXPERIMENTAL 03: EXTERNAL VALIDATION & ROBUSTNESS



DATASET SUBSET

KẾT LUẬN

ƯU ĐIỂM

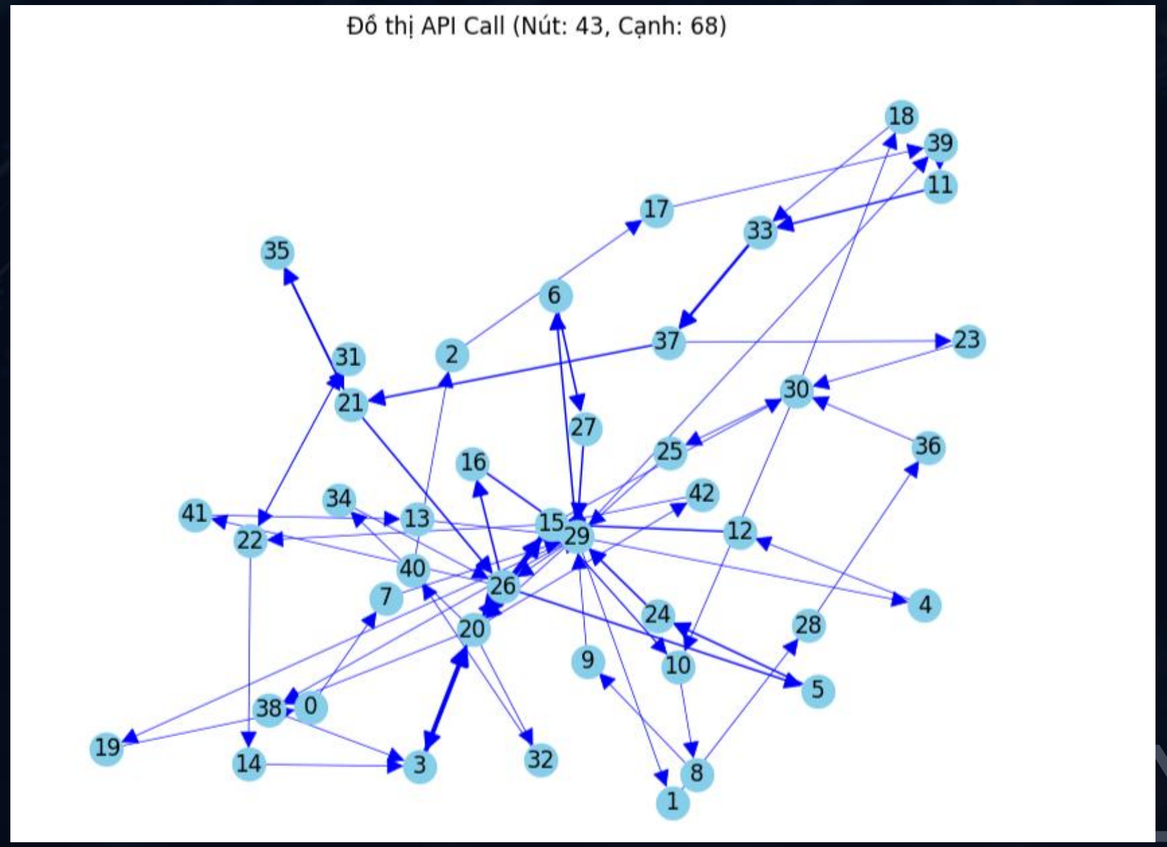
- Dữ liệu từ Microsoft liên tục được cập nhật
- Không gánh nặng tính toán khi dự đoán
- Nhìn ra mối quan hệ và sự tác động qua lại của API

NHƯỢC ĐIỂM

- Malware được cài chế độ Anti Sandbox - Anti Debugger
  → Không trích xuất được API độc hại
- Attacker đổi tên hoặc sử dụng API không có tài liệu chính thống
- Cần bypass được Website Microsoft (Bảo trì script crawl data)
- Gọi các API rác
  → Ảnh hưởng đến khả năng phân loại của GAT

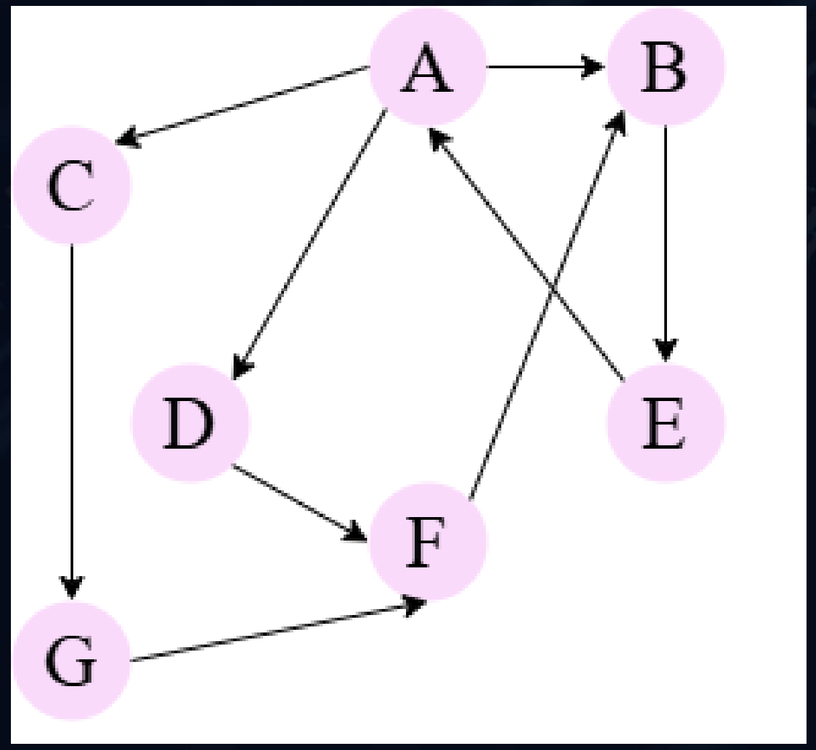
APPENDIX

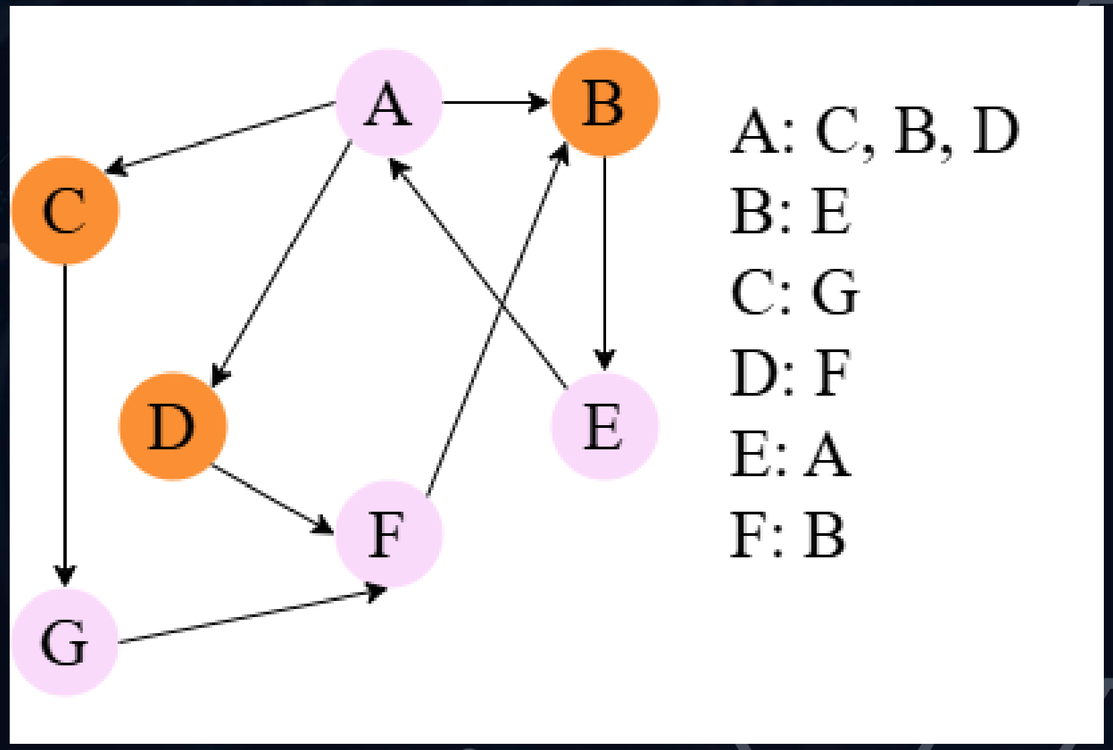
API SEQUENCE TO GRAPH



APPENDIX

GNN STIMULATION

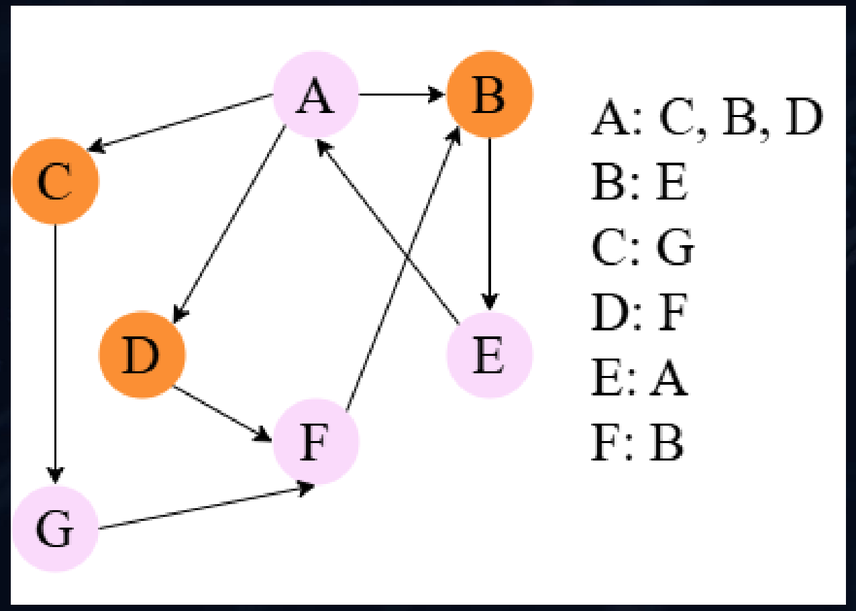




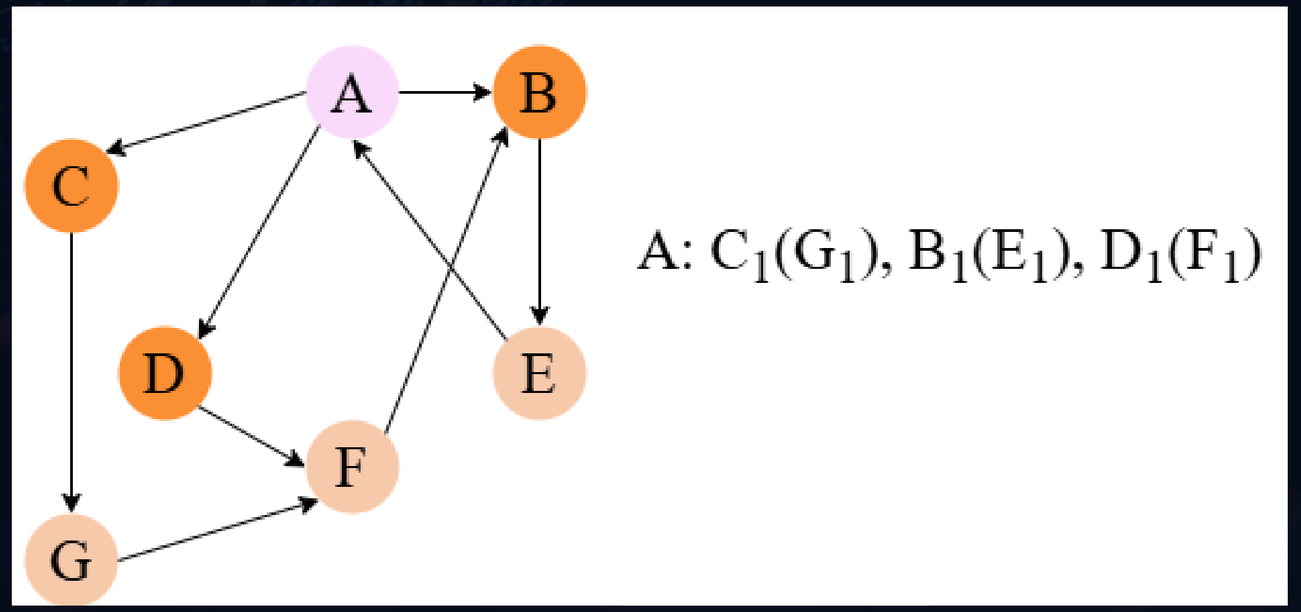
K = 1

APPENDIX

GNN STIMULATION



K = 1



K = 2