POLITECNICO
MILANO 1863

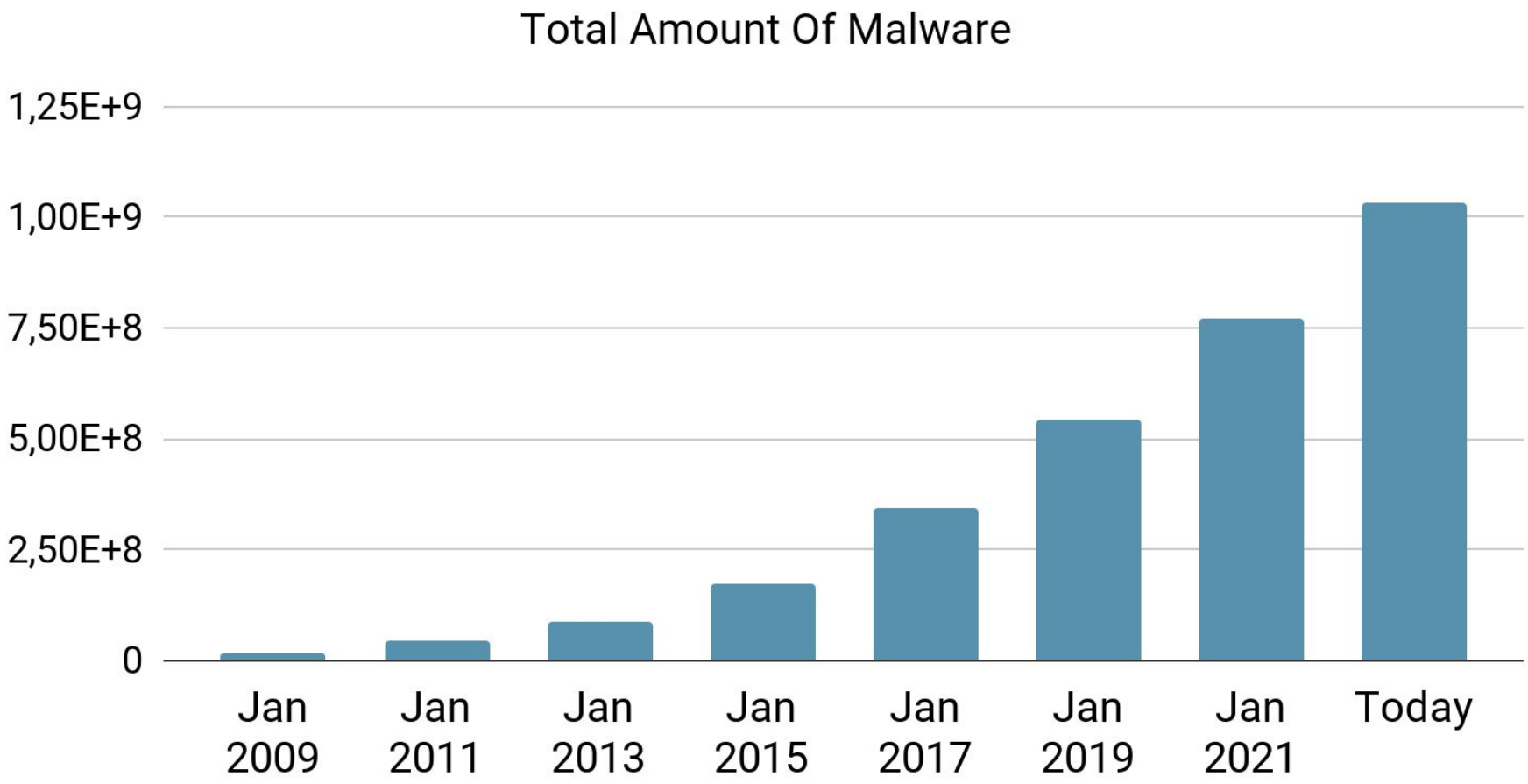
Tarallo: An End-to-End Framework for Malware Behavior Obfuscation

Authors: Gabriele Digregorio, Salvatore Maccarrone

Advisor: Prof. Michele Carminati
Co-advisor: Mario D'Onghia

Academic Year: 2021-22

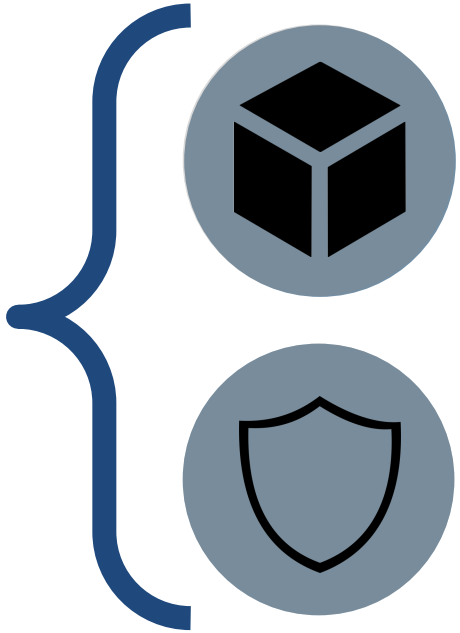
The Malware Phenomenon



Thesis Goal

Evade API call sequence-based machine learning malware classifiers.

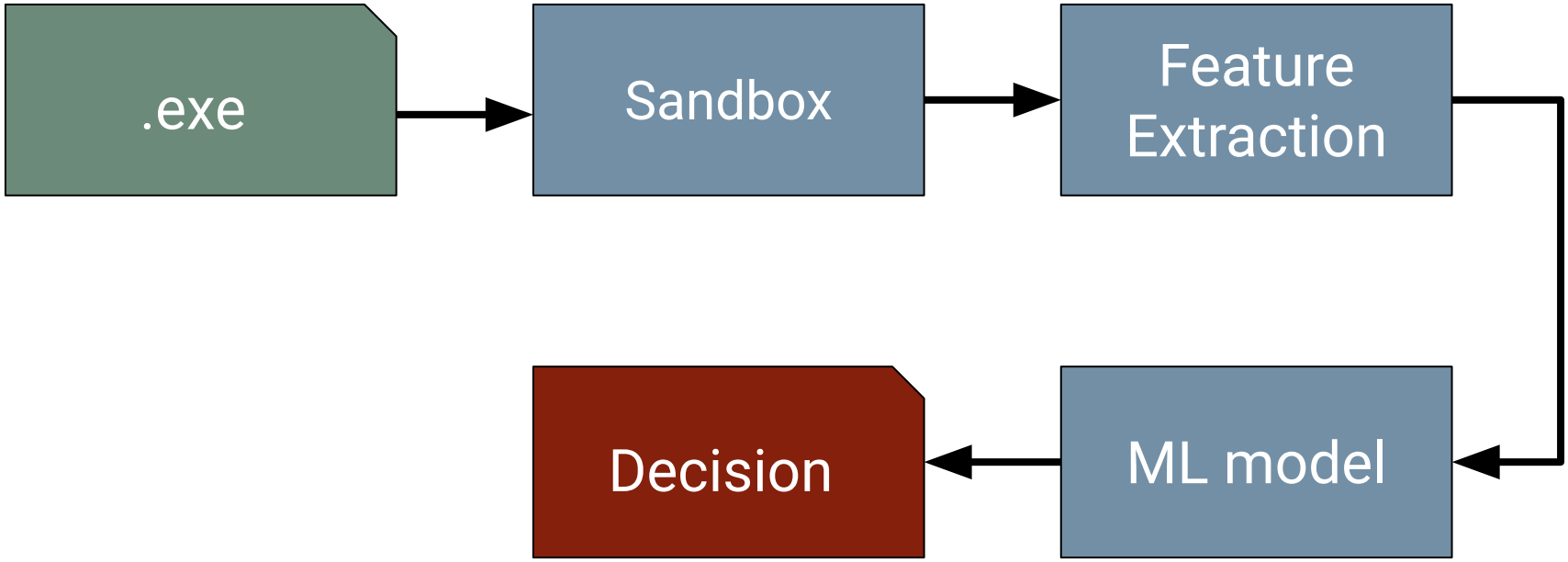
Motivations:



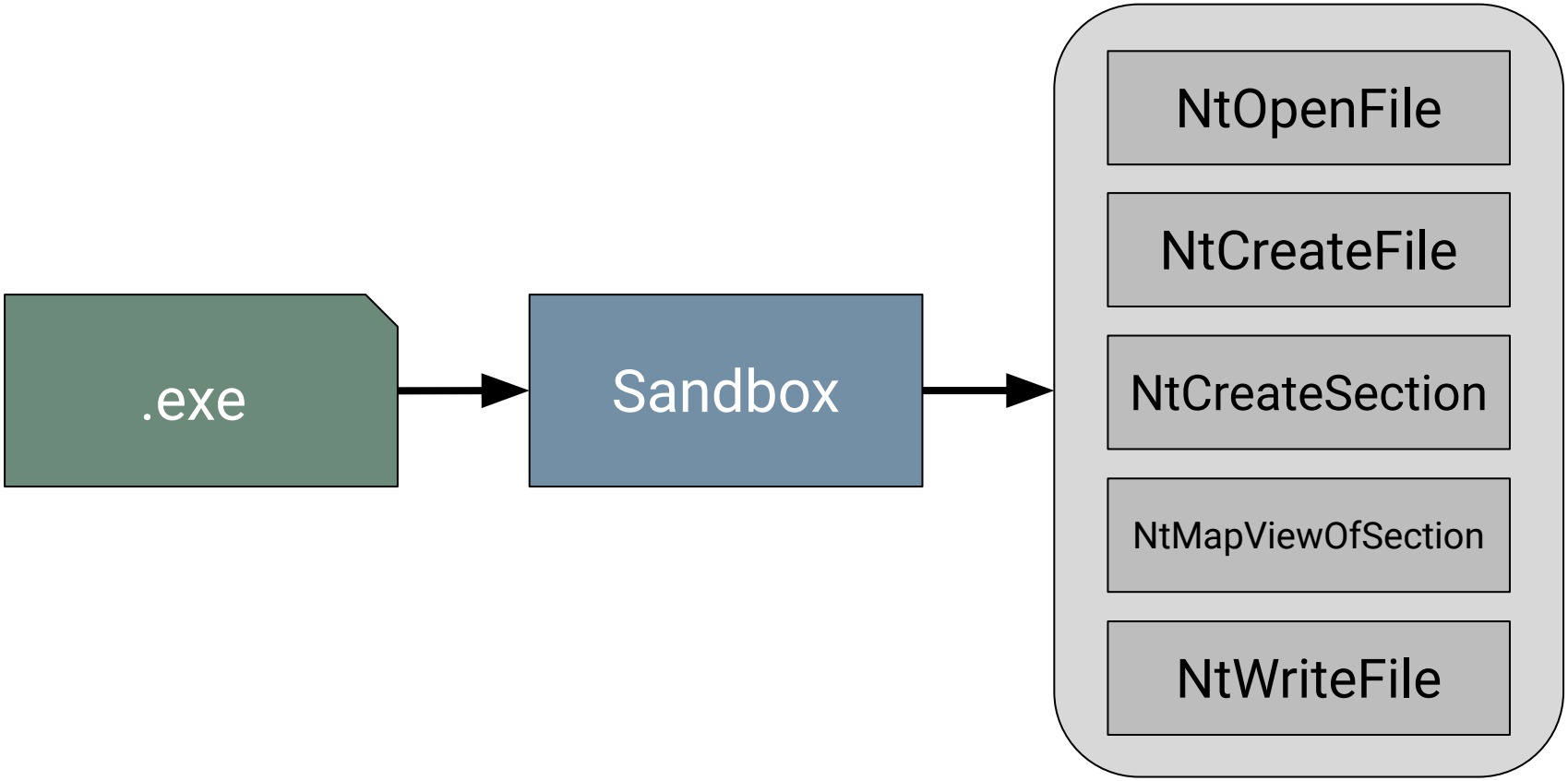
Next generation anti-malware are based on these classifiers

Improve defense mechanisms

ML-based Detection System



API Call Sequences

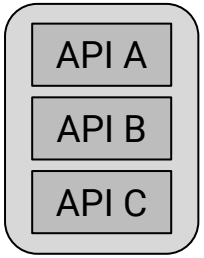


Adversarial Machine Learning

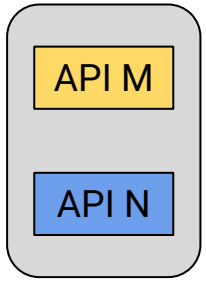
Original Image
Panda

Perturbation

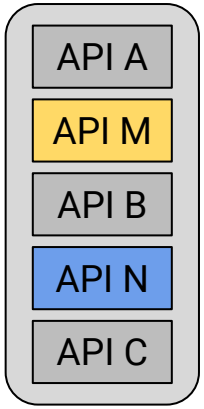
Perturbed Image
Gibbon



Original PE File
Malware



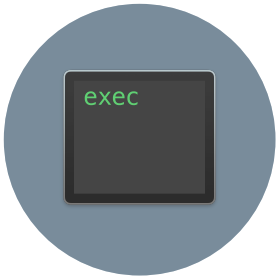
Perturbation



Perturbed PE File
Goodware

Thesis Challenges

Adversarial machine learning techniques in the malware domain



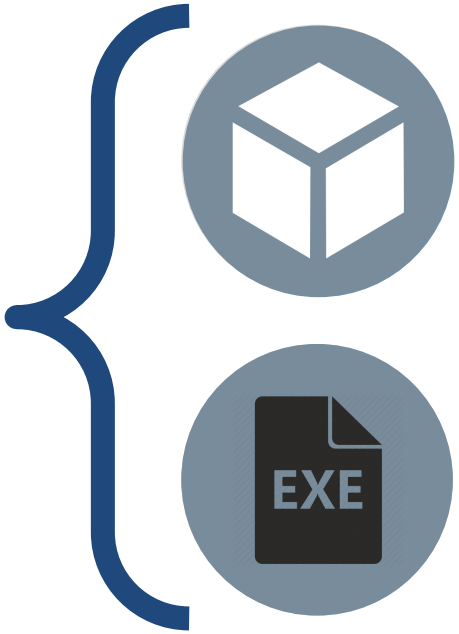
From feature representation to executable domain

Functionality preservation

Approach

Approaching The Problem

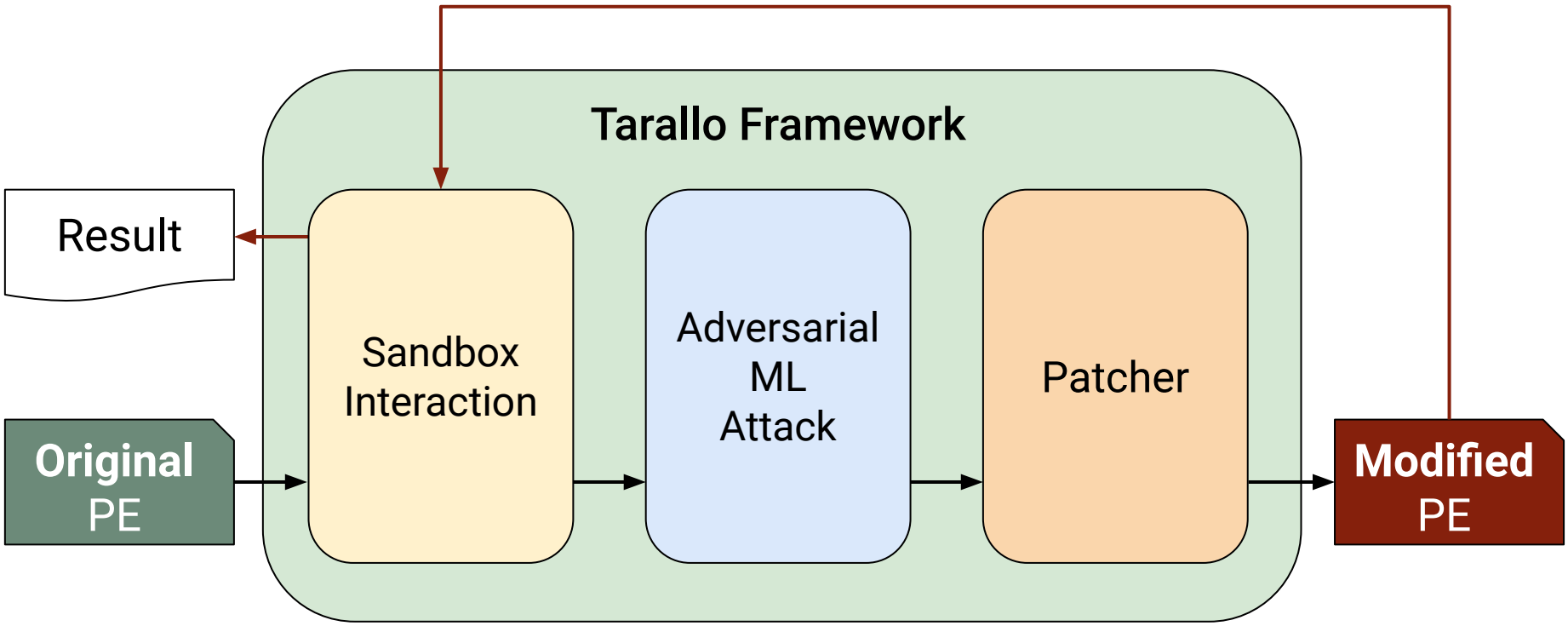
Two sub problems:



Evading Detection

Change dynamic apparent behavior

Tarallo Framework

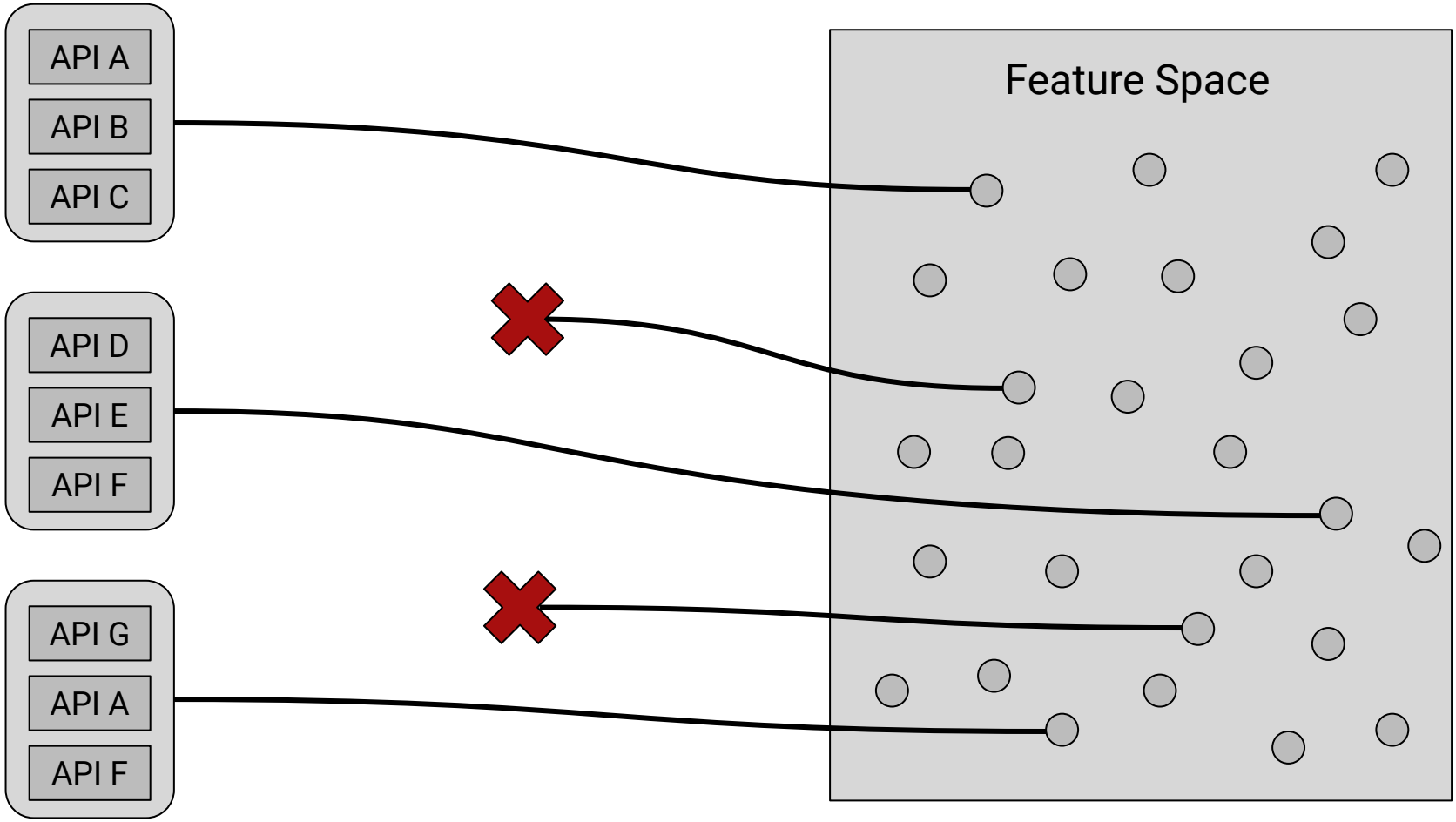


Evading Detection

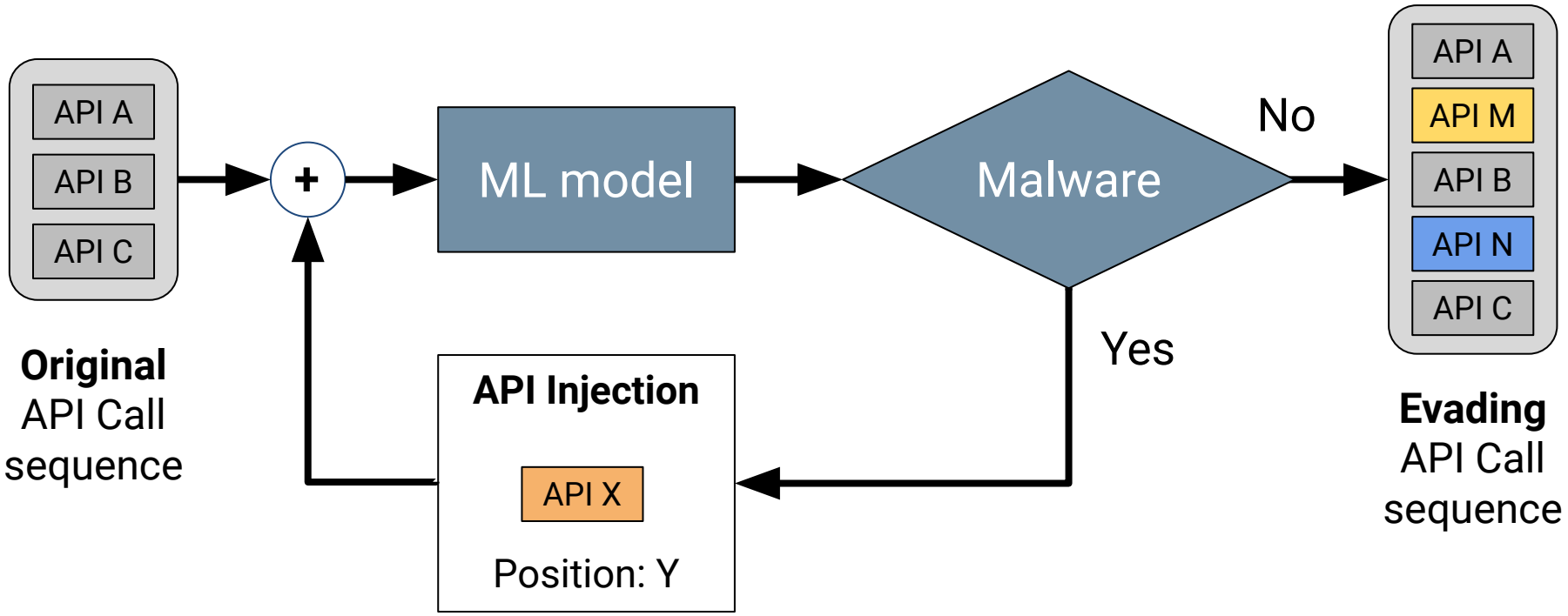
Gabriele Digregorio, Salvatore Maccarrone

POLITECNICO MILANO 1863

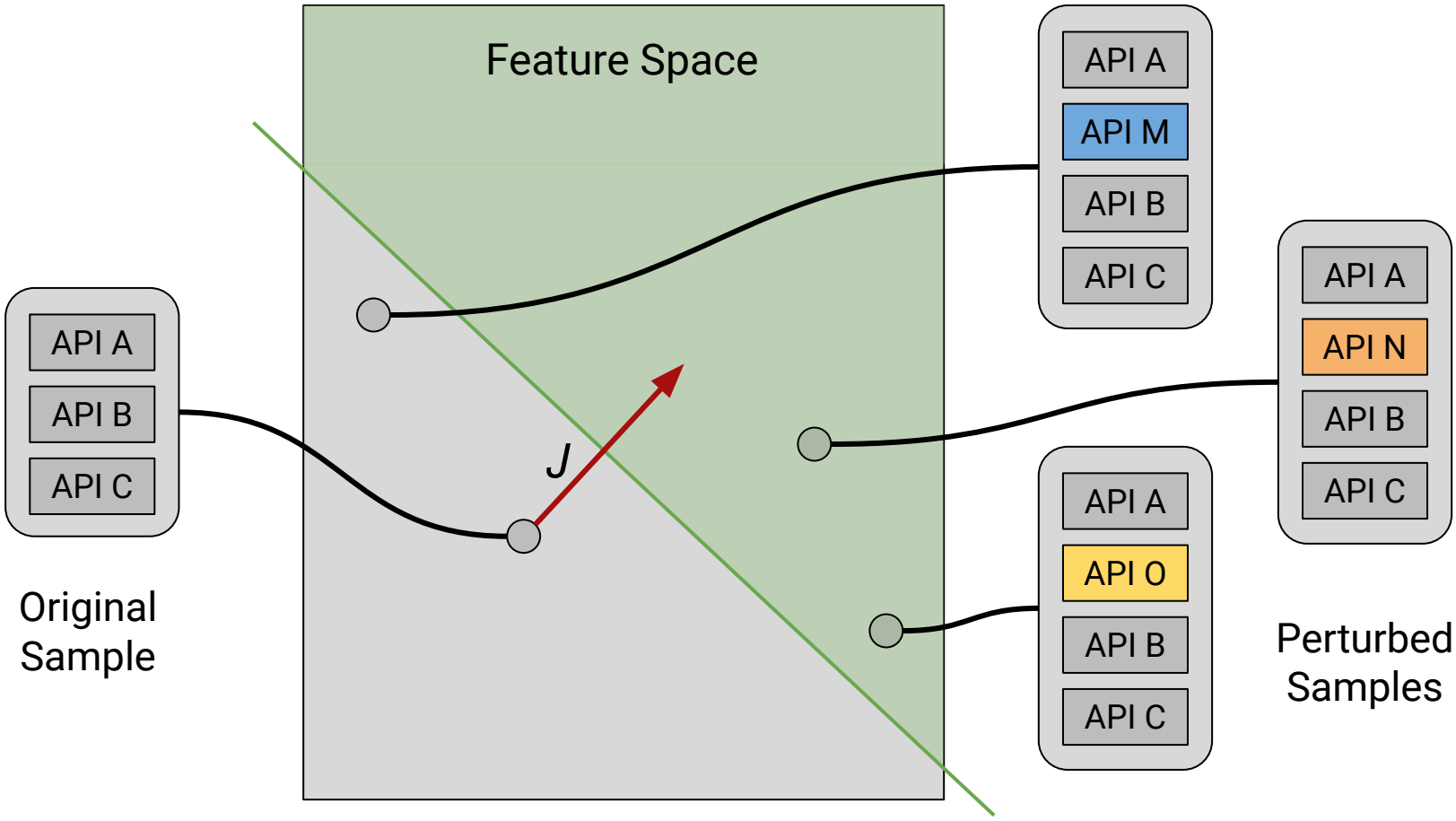
Feature Space Mapping



Adversarial Machine Learning Strategy

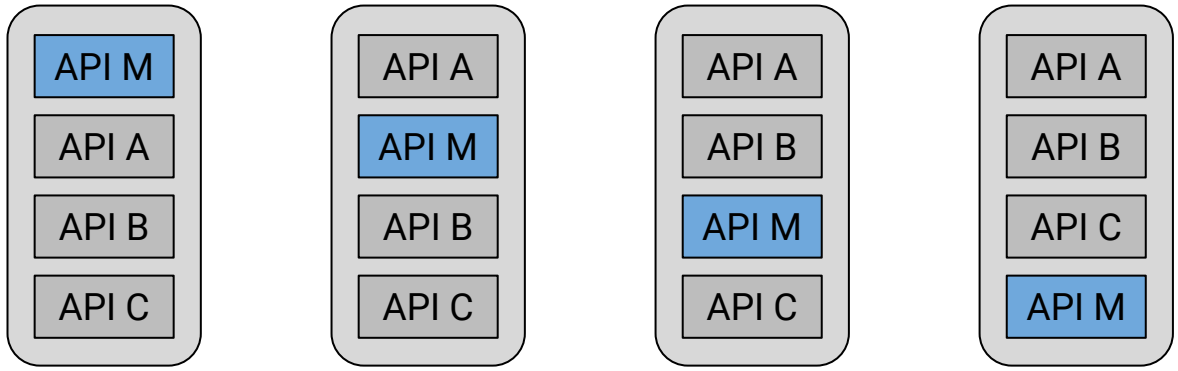


Jacobian Discrete Approximation



Computing Best Position

Input API Call Sequence Heatmap



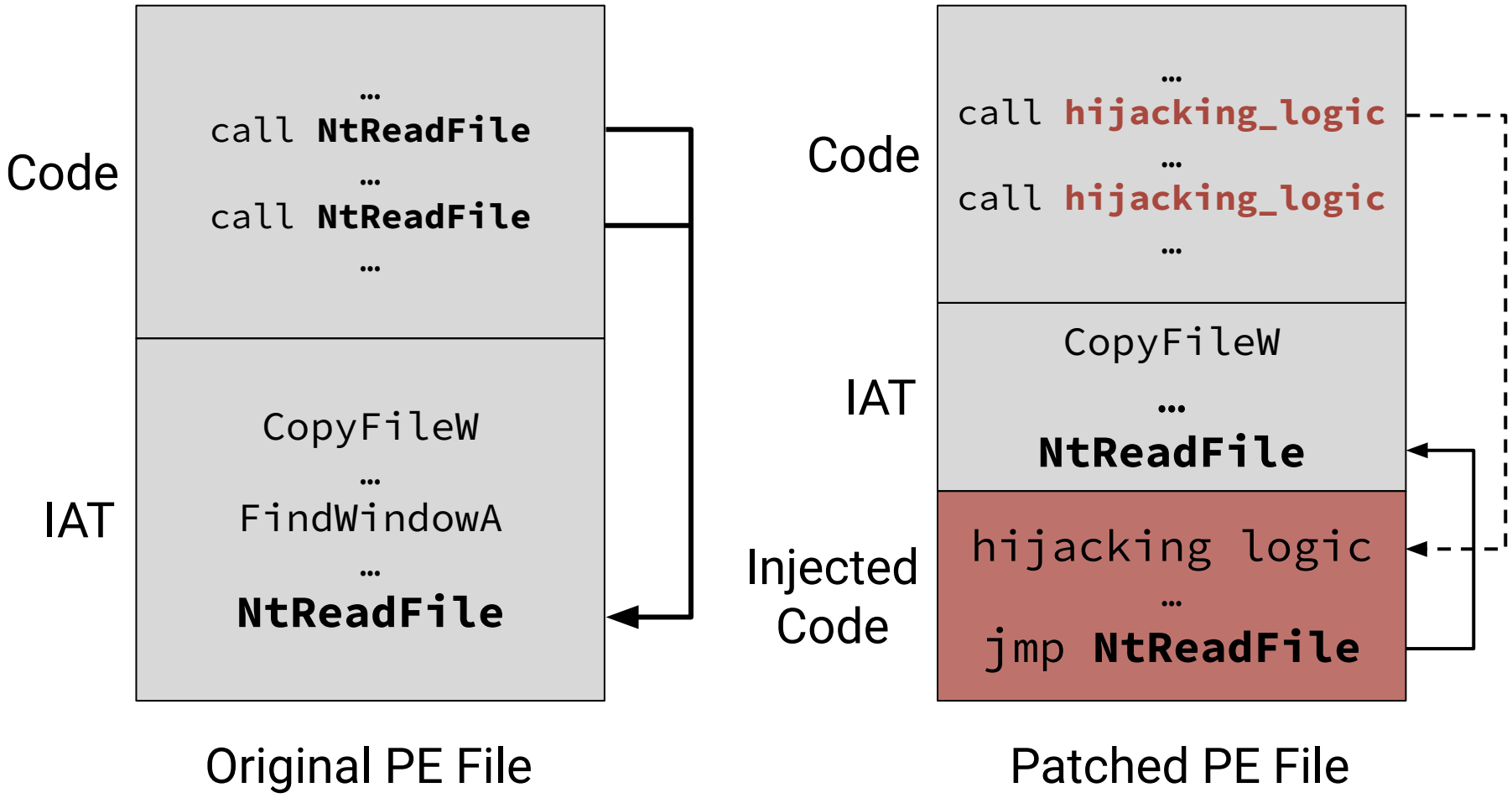
Prediction score: 0.32 0.57 0.45 0.63

Modifying Dynamic Apparent Behavior

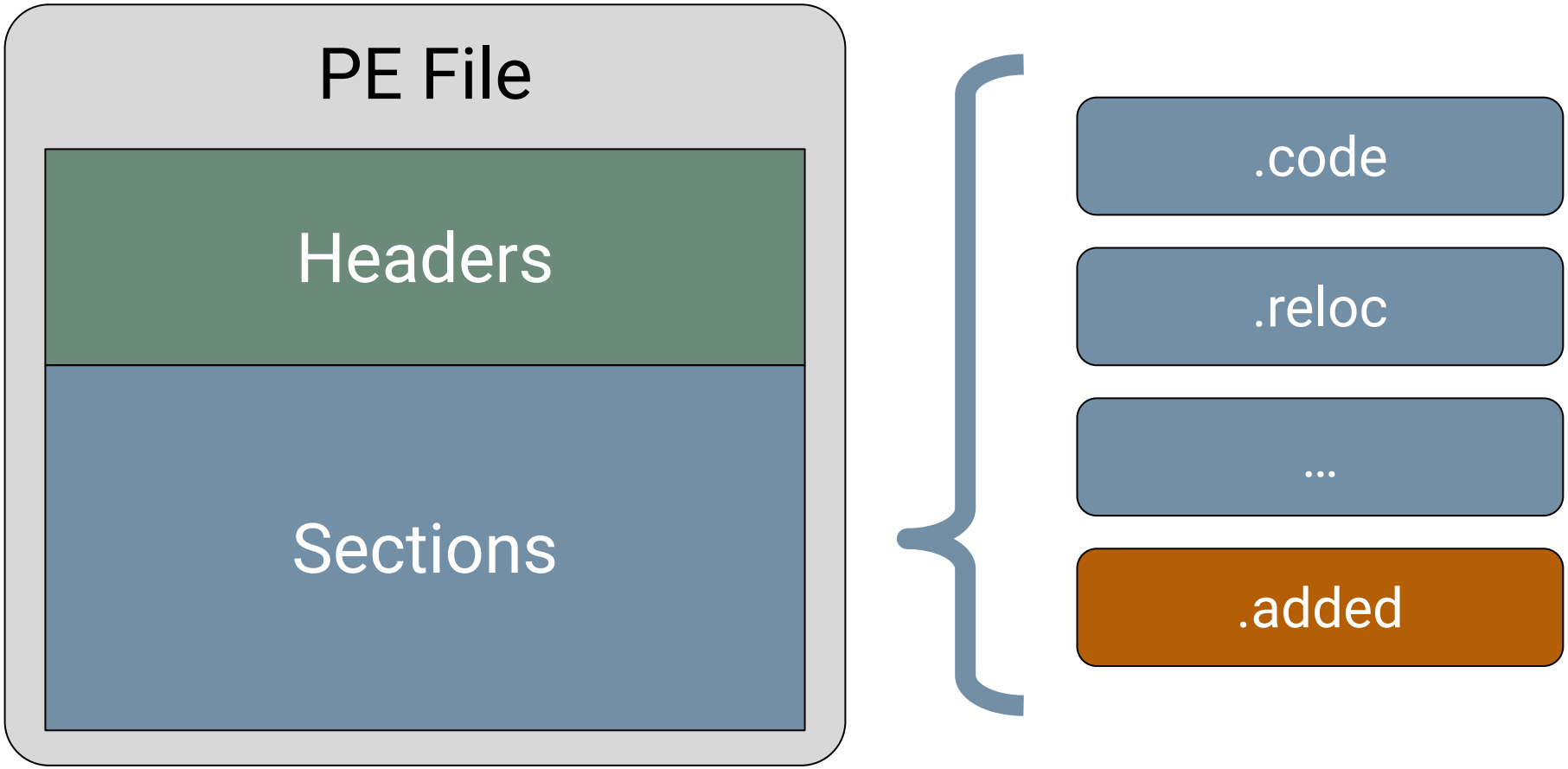
Gabriele Digregorio, Salvatore Maccarrone

POLITECNICO MILANO 1863

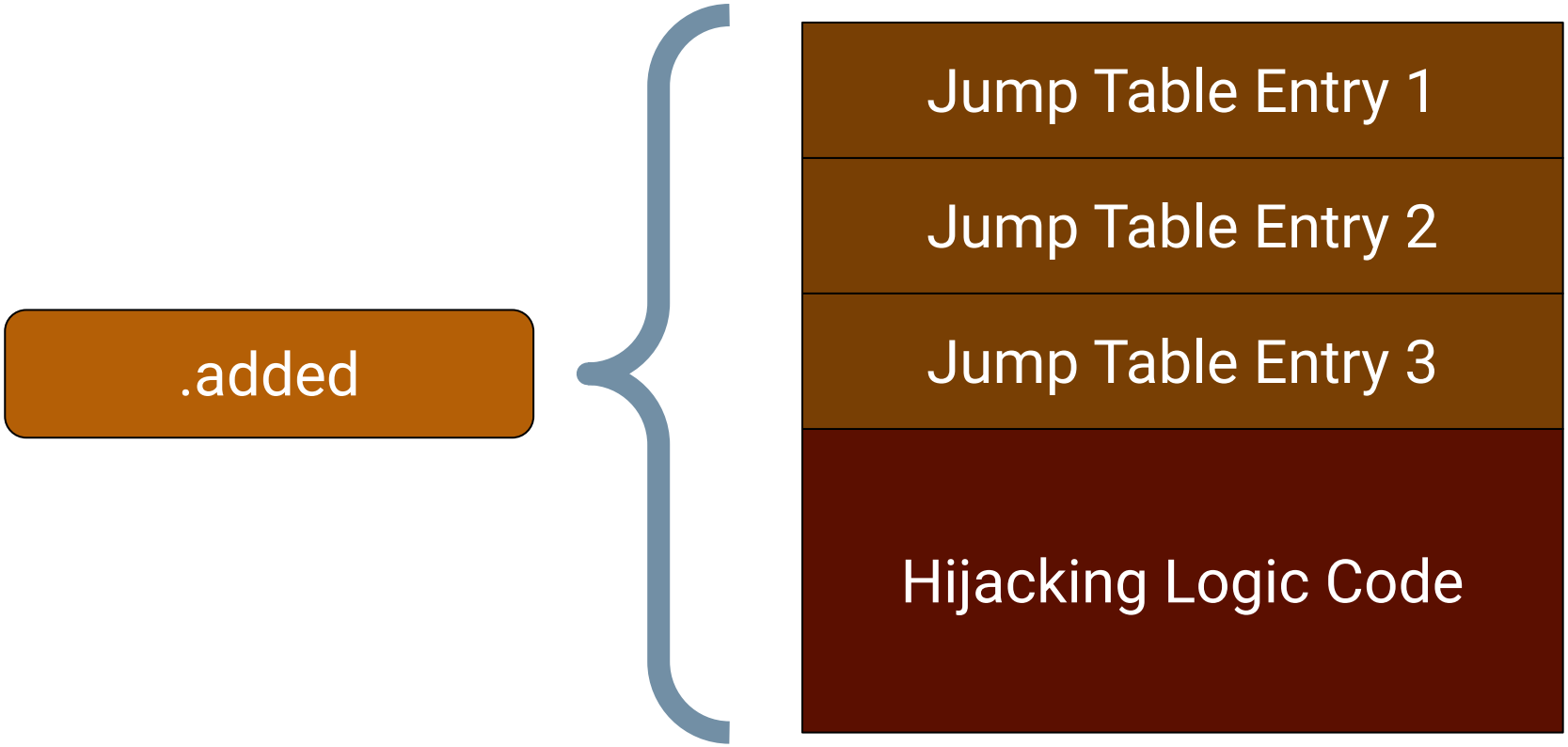
Patching PE Files



Hijacking Logic Location



Hijack Logic



Experiments

Gabriele Digregorio, Salvatore Maccarrone

Experiments - Goals

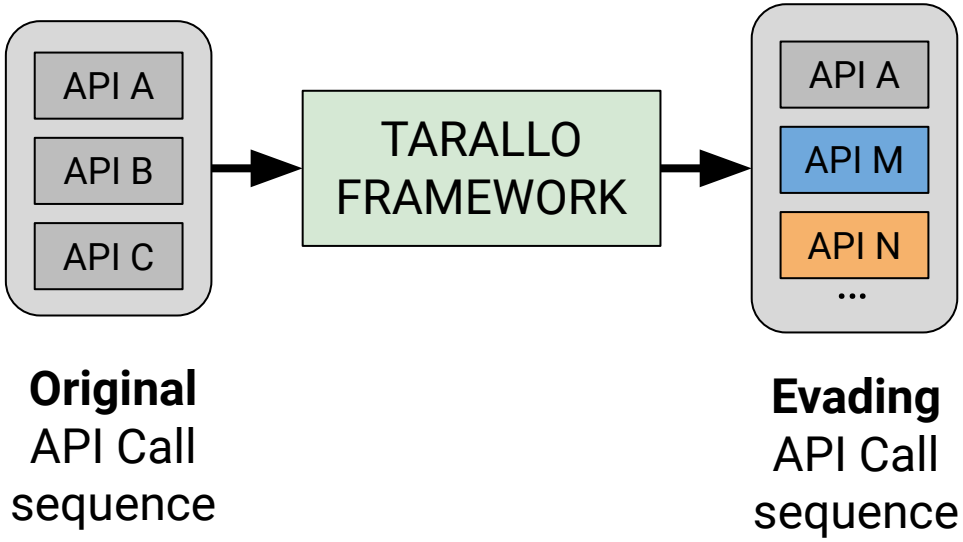
Evasion effectiveness

Overhead Comparison

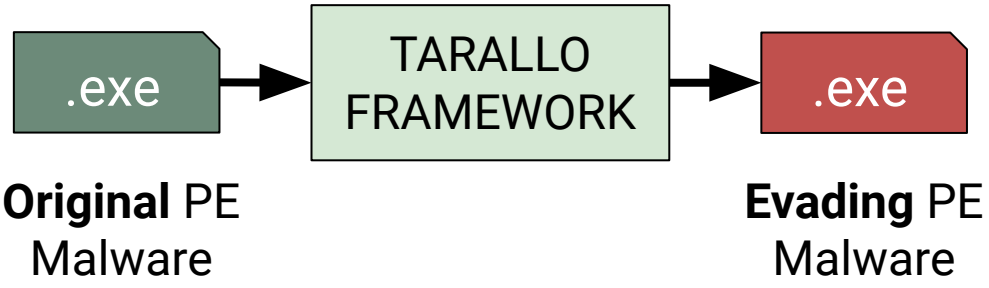
Functionality preservation

Experiments - Evasion Effectiveness

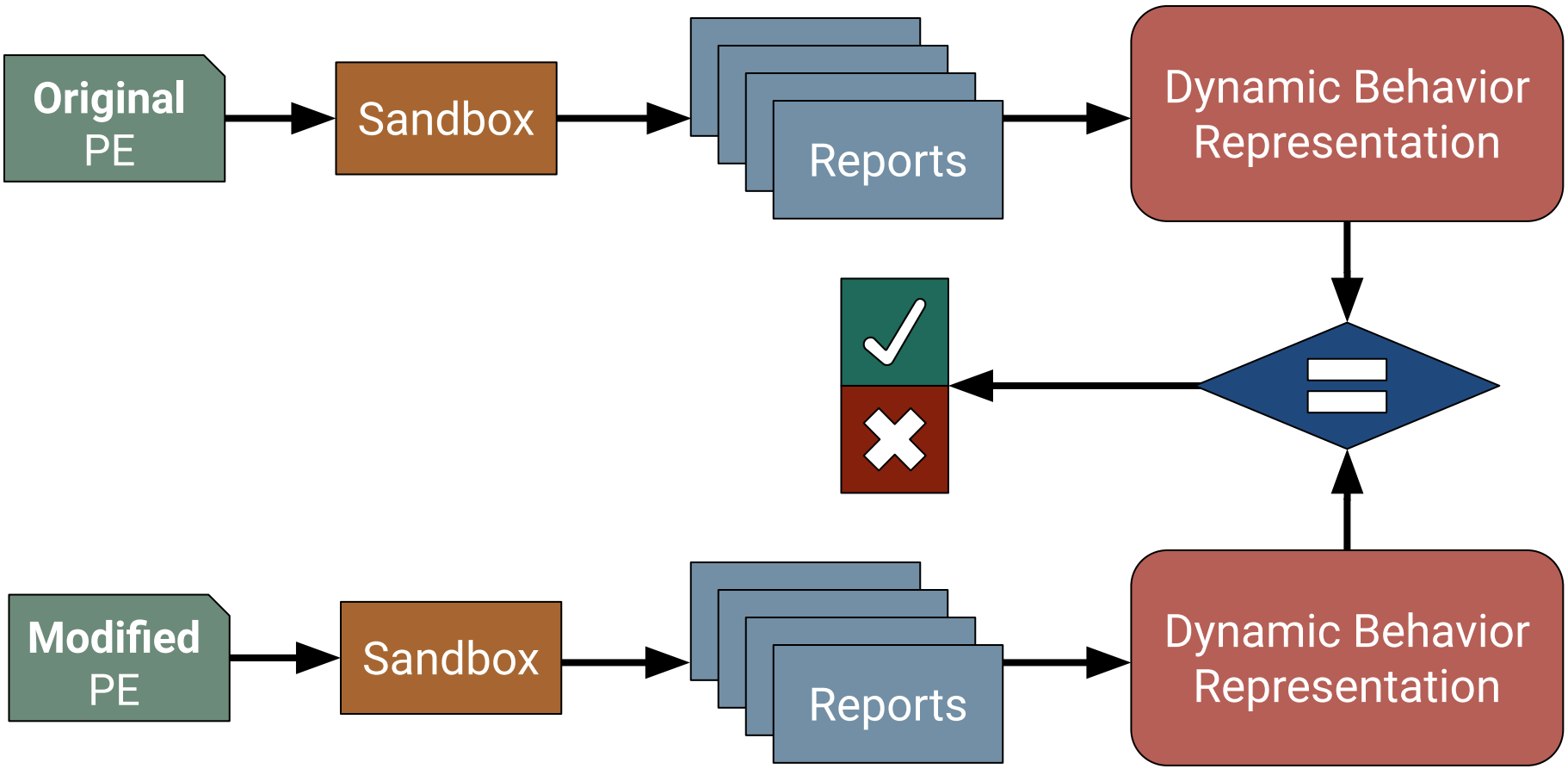
- **Feature** level attack



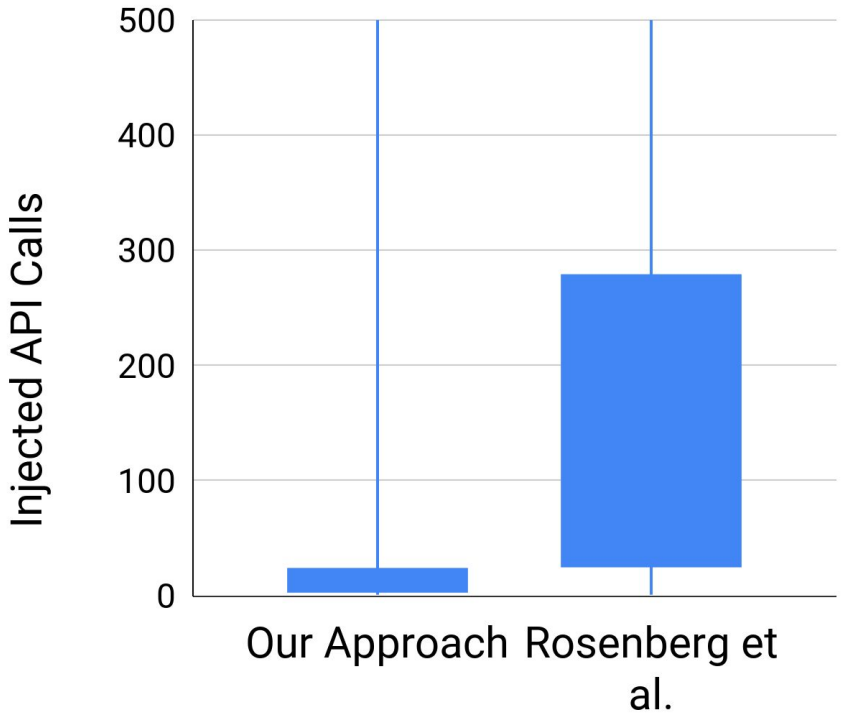
- **End-to-end** attack



Experiments - Functionality Preservation



Feature Level Attack Results



Overhead limit	20%	50%	90%
Our Approach	**0.9742**	**0.9841**	**0.9904**
Rosenberg et al.	0.267	0.3674	0.9535

End-To-End Attack Results

Injectable APIs	1%	10%	20%
Available samples	1611	670	198
Evading samples	1016	498	150
Ratio	**0.64**	**0.74**	**0.76**

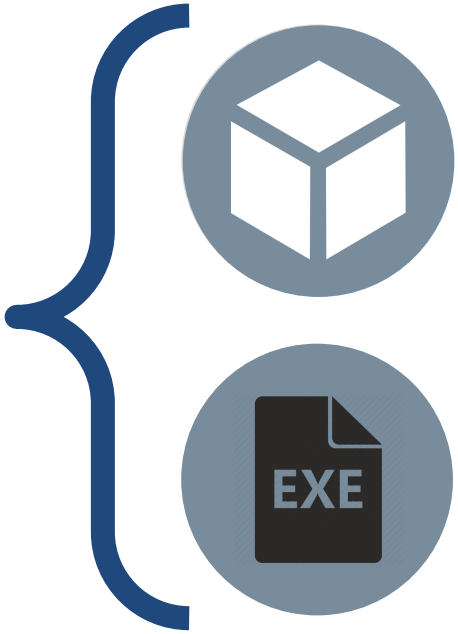
Malware samples with preserved functionalities:
89%

Conclusions

Conclusions And Future Works

Evade API call sequence-based machine learning malware classifiers.

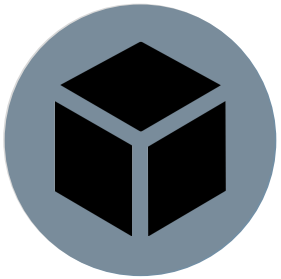
Contributions:



Novel machine learning attack

Novel patching strategy

Future Works



Black-Box Adversarial ML Attack

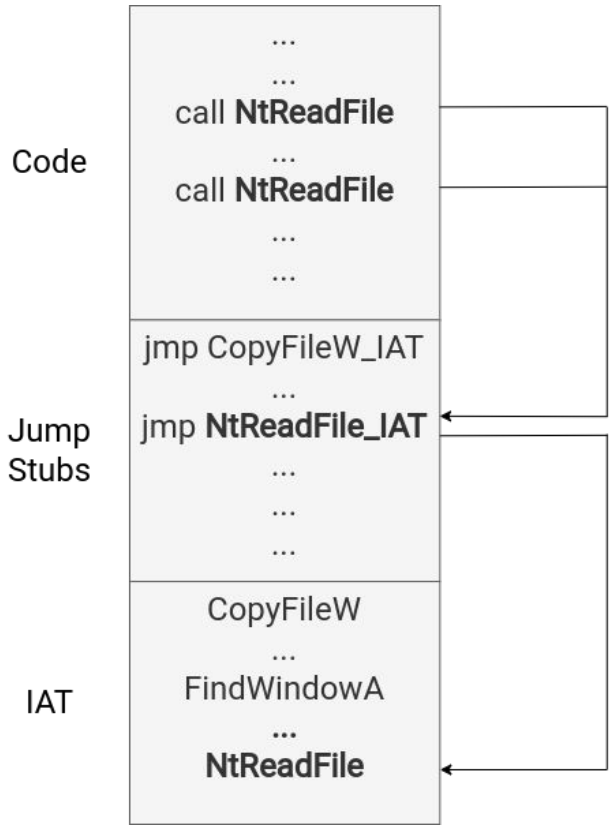
New Countermeasures

Thank You For Your Attention

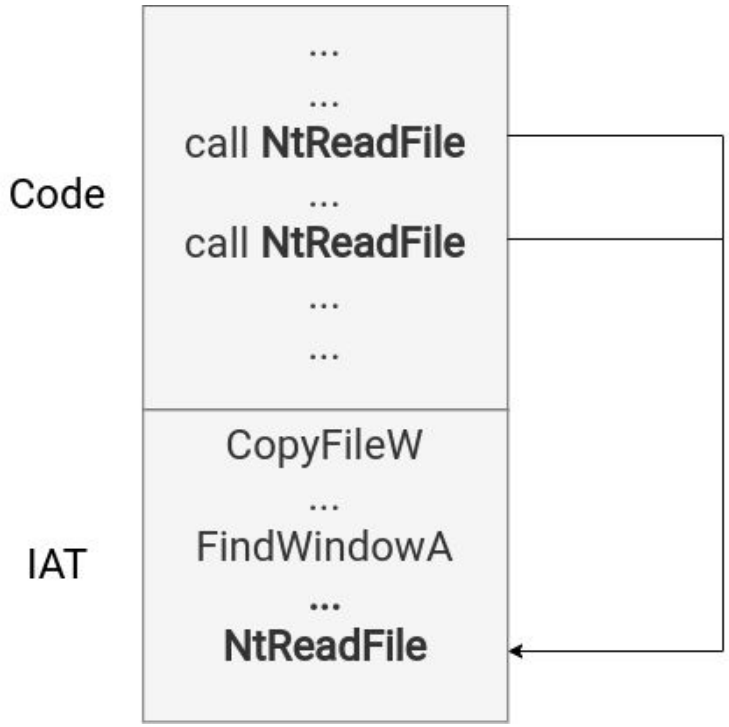
Gabriele Digregorio, Salvatore Maccarrone

POLITECNICO MILANO 1863

PE Dynamic Linking



Jump Stub



Indirect Call