

NEW APPROACH FOR NETWORK ATTACK DETECTION BASED ON IMAGE RECOGNITION

SANTIAGO HERNÁNDEZ RAMOS

[HTTPS://GITHUB.COM/SHRAMOS](https://github.com/shramos)

[@SANTIAGOHRAMOS](https://twitter.com/santiagoqramos)

NAVAJA
NEGRA

How does this start?

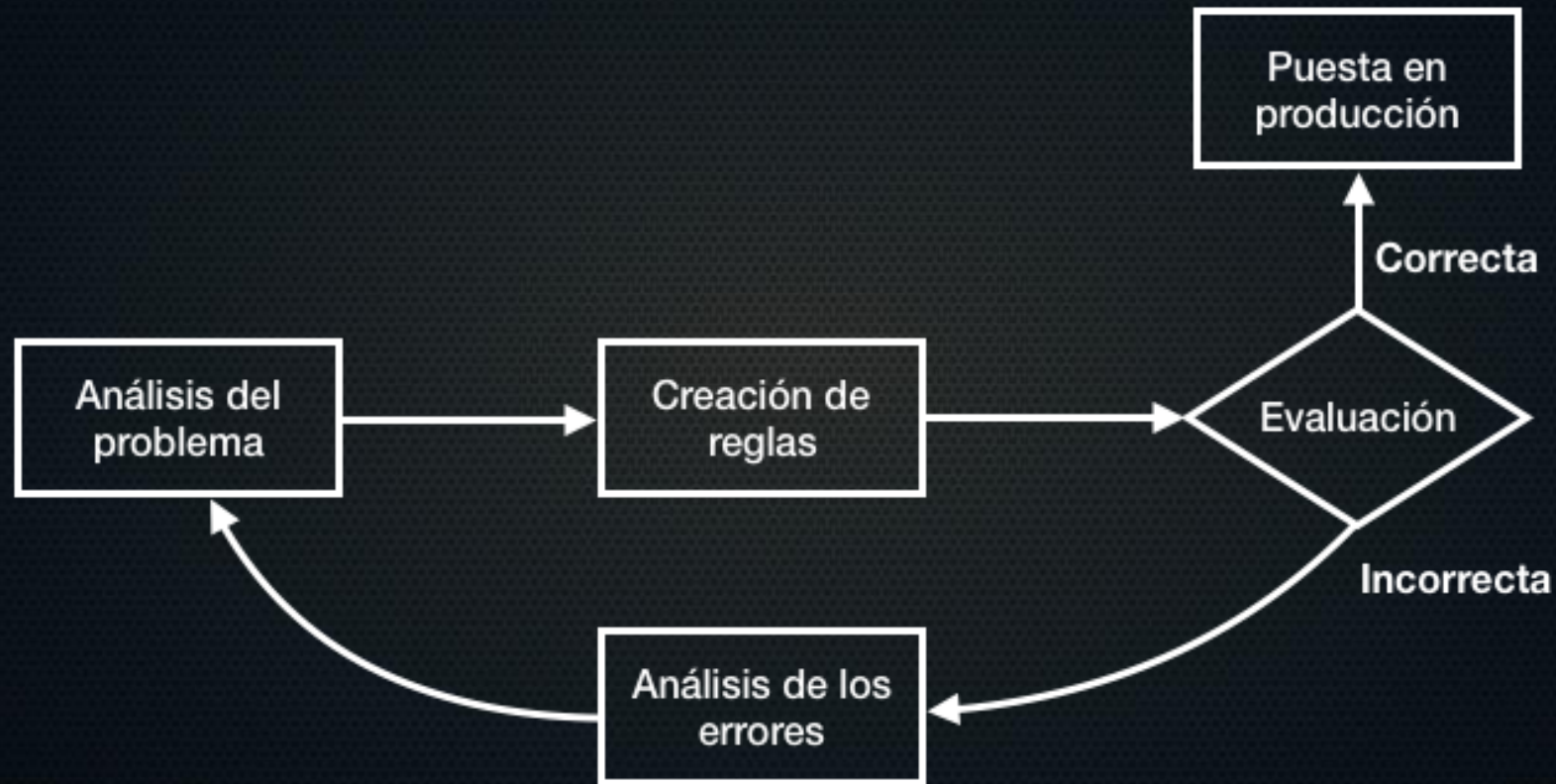
Intrusion Detection

Intrusion Detection is categorized into two types [1]:

- **Rule-based and heuristic approaches:** Produces few false positives. Detects known attacks. Does not work well for detecting new attacks.
- **Anomaly-based approaches:** Profiles normal system behavior. Can detect new attacks. Generates more false positives.

[1] Lee, W., and Stolfo, S. J. Data mining approaches for intrusion detection. In *Proc. of the 7th USENIX Security Symposium (USA, 1998)*, vol. 7, USENIX Association, pp. 79–94.

Rule-based and heuristic approaches



Anomaly Detection

- An anomaly is an **event that deviates from normal or expected behavior** and is suspicious, in this case, from a security perspective.
- Anomaly detection is the identification of rare elements, events, or observations that raise suspicions by significantly differing from most of the data [1].
- Anomaly detection was proposed for intrusion detection systems (IDS) by Dorothy Denning in 1986 [2].
- Anomalies can occur due to **two main factors** [3]:
 - Performance-related
 - Security-related

[1] https://en.wikipedia.org/wiki/Anomaly_detection

[2] D. E. Denning, P. G. Neumann, *Requirements and model for IDES—A real-time intrusion detection system*, 1985.

[3] Thottan, M., and Ji, C. *Anomaly detection in IP networks*. *IEEE Transactions on Signal Processing* 51, 8 (August 2003), 2191–2204.

Artificial Intelligence and Cybersecurity

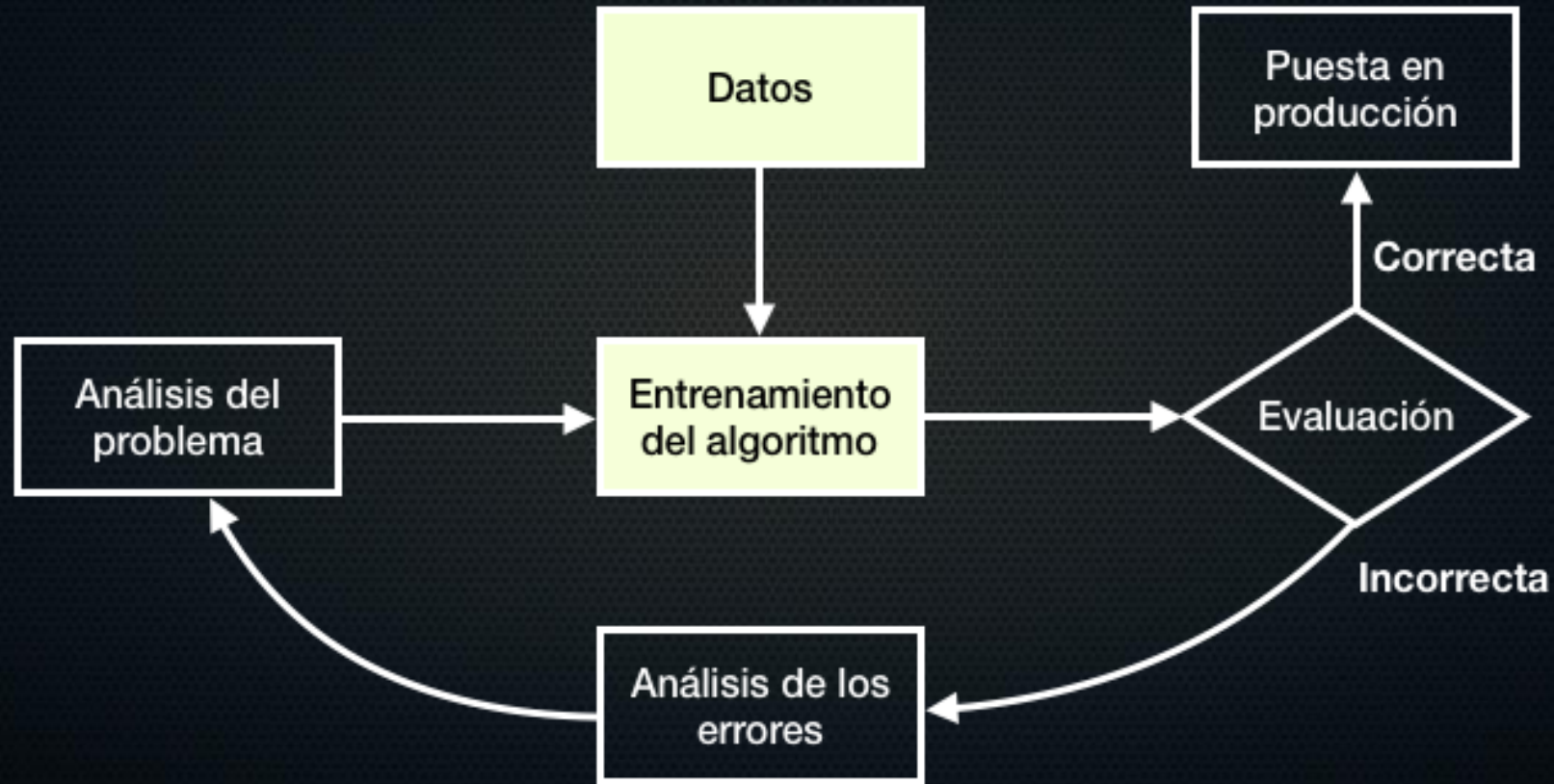
“Today’s AI has important applications in cybersecurity, and is expected to play an increasing role for both defensive and offensive cyber measures.” [1]

[1] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

AI Algorithms Classification

- Supervised learning.
- Semi-supervised learning.
- Unsupervised learning.
- Reinforcement learning.

Based on Machine Learning



Problems in the application of Machine Learning techniques

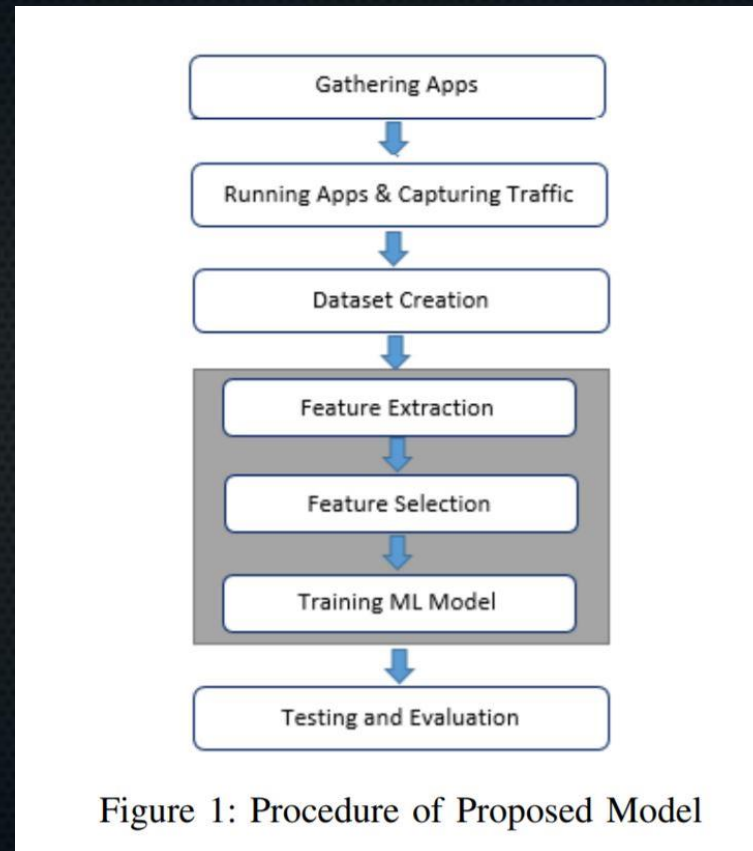


Figure 1: Procedure of Proposed Model

Towards a Network-Based Framework for Android Malware Detection and Characterization

Problems in the application of Machine Learning techniques

Out[4]:

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_count	dst_host_srv_count	dst_host_same_
95141	0	tcp	http	SF	214	14939	0	0	0	0	...	52	255	
37486	0	tcp	private	S0	0	0	0	0	0	0	...	255	2	
34926	0	tcp	http	REJ	0	0	0	0	0	0	...	255	8	
34589	0	tcp	http	SF	257	259	0	0	0	0	...	255	255	
11420	0	udp	other	SF	516	4	0	0	0	0	...	255	255	
46955	0	tcp	private	S0	0	0	0	0	0	0	...	255	20	
32661	0	tcp	smtp	RSTO	0	0	0	0	0	0	...	255	1	
21066	0	icmp	eco_i	SF	8	0	0	0	0	0	...	2	129	
22128	0	tcp	private	S0	0	0	0	0	0	0	...	255	17	
21455	0	udp	private	SF	1	0	0	0	0	0	...	255	110	

M. Tavallae, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

END-TO-END LEARNING

We delegate the feature extraction function to the algorithm: **Deep Learning**

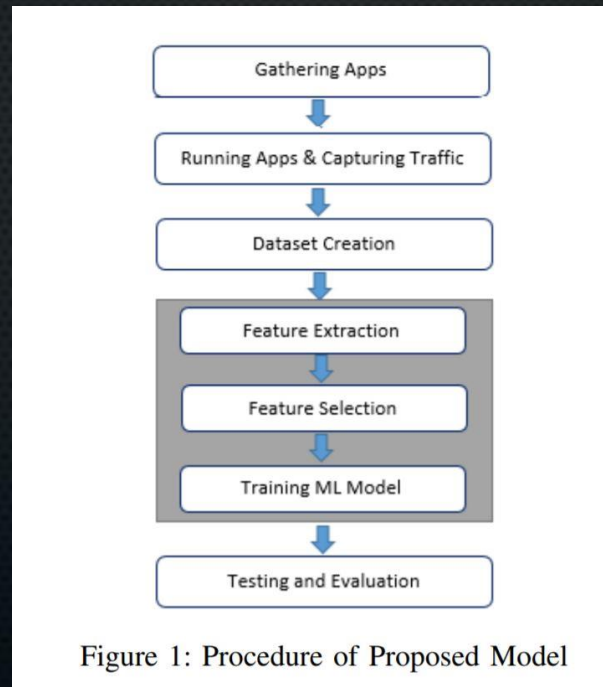


Figure 1: Procedure of Proposed Model

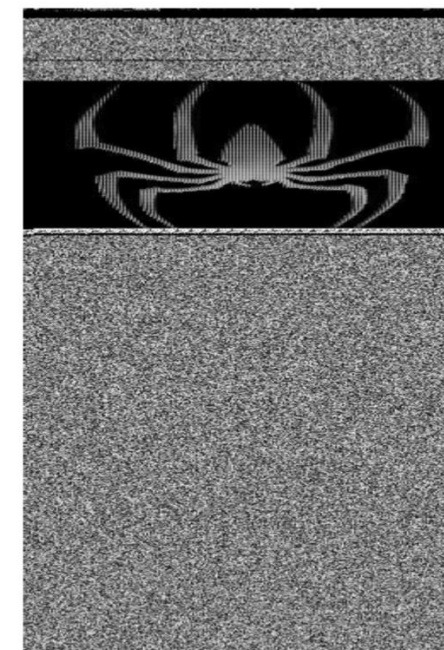
Malware Images: Visualization and Automatic Classification

Malware Binary

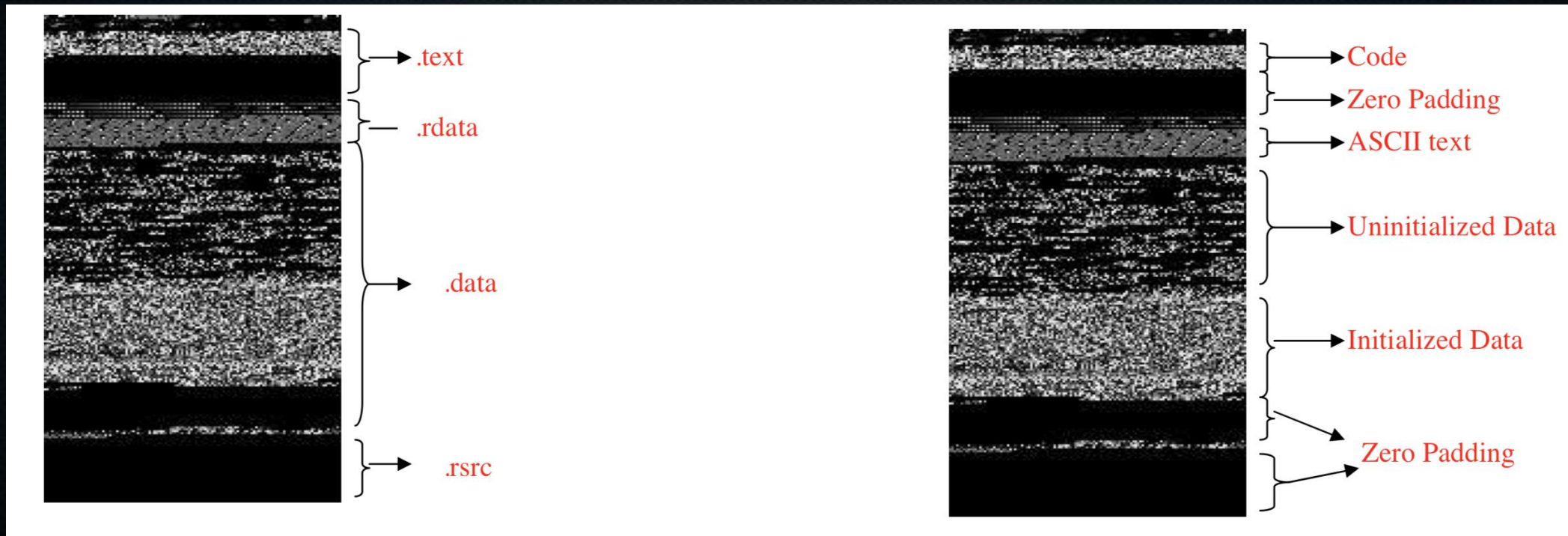
```
011100110101  
100101011010  
10100001..
```

Binary to
8 bit
vector

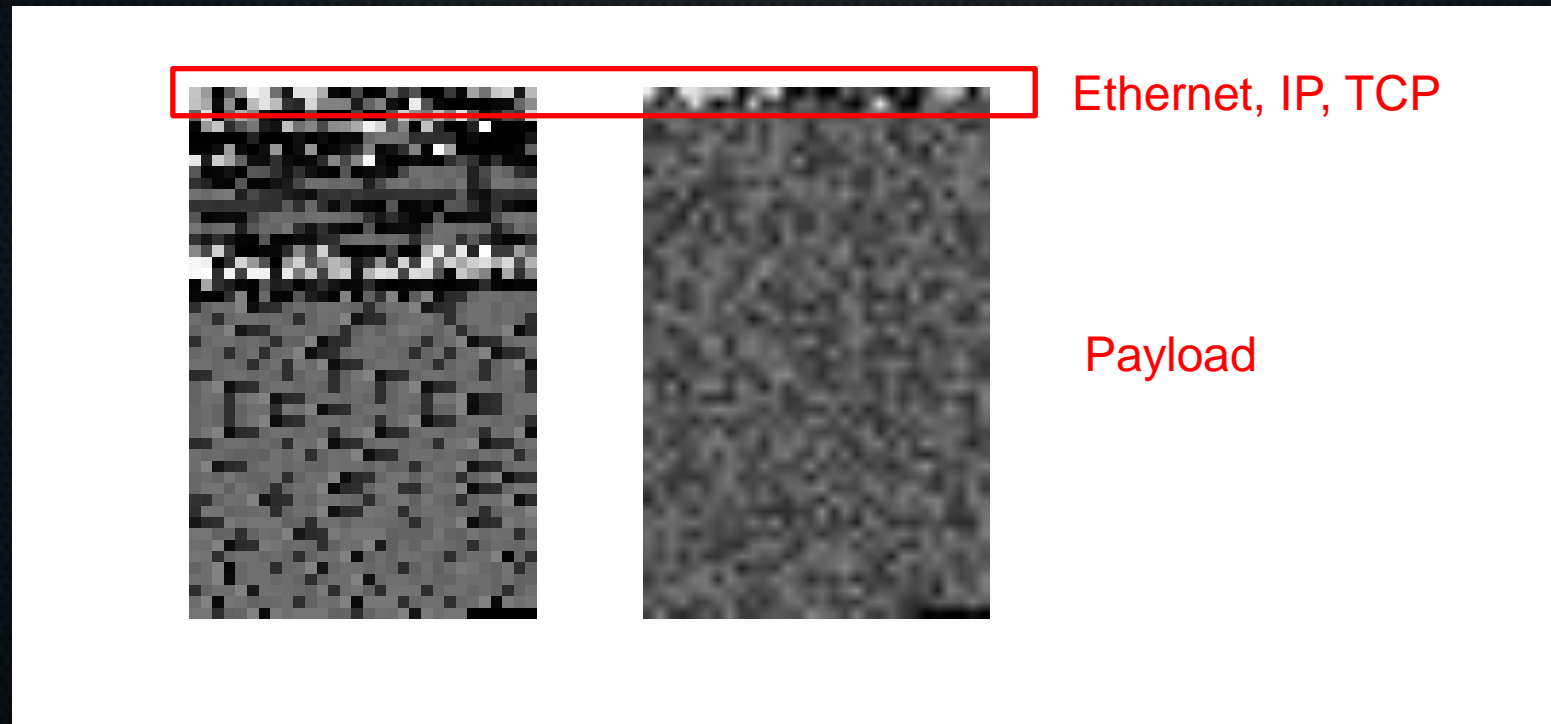
8 Bit vector to
Grayscale
Image



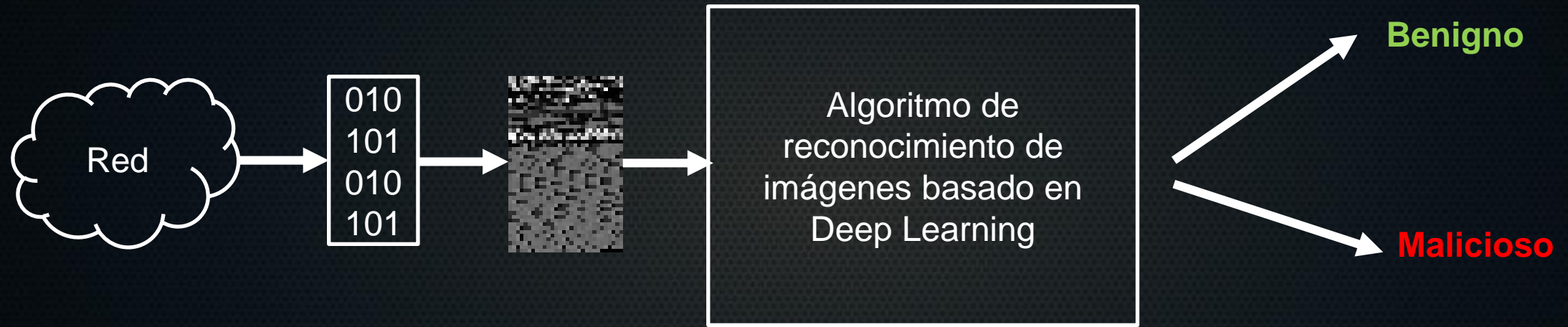
Malware Images: Visualization and Automatic Classification



Malware Images: Visualization and Automatic Classification



NAVAJA NEGRA



NAVAJA
NEGRA

Where do we start?

Where do we start?

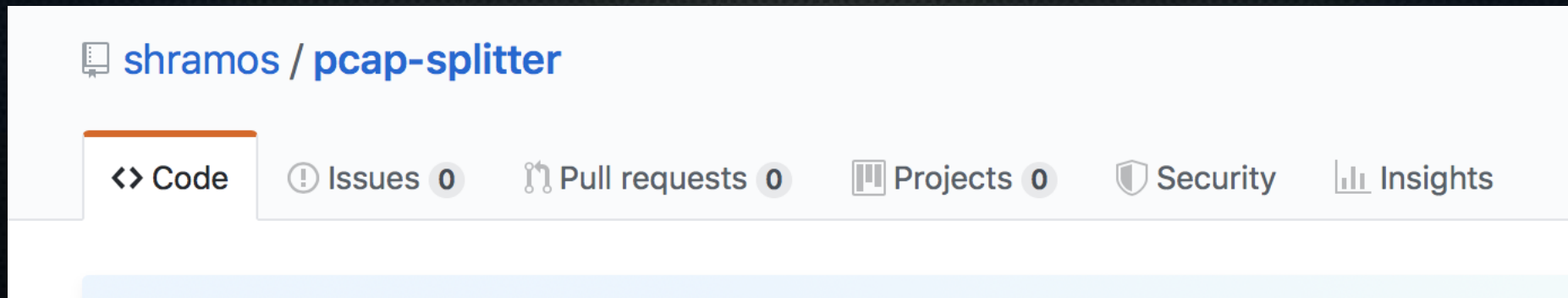
1. Convert network traffic into images.
2. Select and train a Deep Learning algorithm.
3. Offline evaluation.
4. Save the model.
5. Deploy the model in real-time.
6. Online evaluation.

Convert network traffic into images

01005e000001f88e85ded79208004600002000004000
0102432dc0a80101e000000194040000110aeef50000
0000803f5d5c071b8c8590bb3db9080045000047b43b
0000ff1183cac0a8012cc0a80123db3c003500336bf6bd
57010000010000000000000000c70726f6473746f7261676
53208636c6f7564617070036e6574000001000101005
e7ffffab853ac6d1ce2080045000099522a00000111b56
0c0a80127effffffad23c076c008545e84d2d5345415243
48202a20485454502f312e310d0a484f53543a2032333
92e3235352e3235352e323

Convert network traffic into images

Divide by packets, **sessions**, connections, flows, size, number of packets, etc.



Train a Deep Learning algorithm

IMGENET Large Scale Visual Recognition Challenge (ILSVRC)

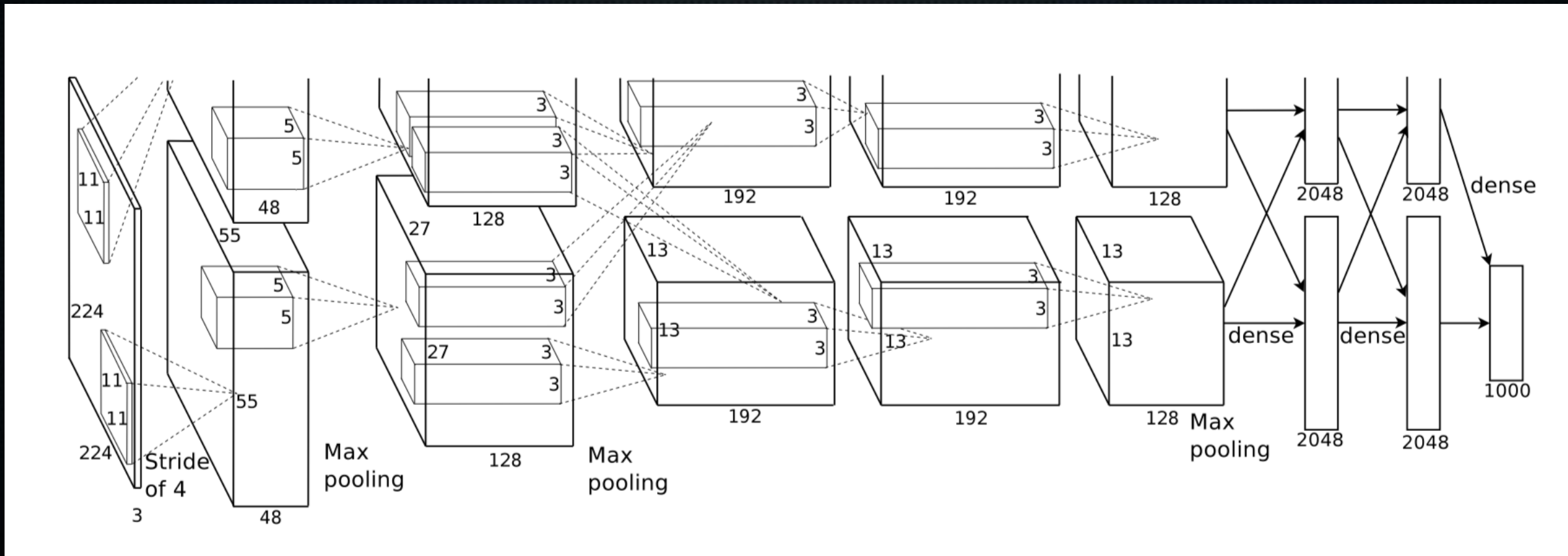
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Train a Deep Learning algorithm



Offline evaluation

CICIDS2017: *Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018*

CICInvesAndMal2019: *Laya Taheri, Andi Fitriah Abdulkadir, Arash Habibi Lashkari; Extensible Android Malware Detection and Family Classification Using Network-Flows and API-Calls, The IEEE (53rd) International Carnahan Conference on Security Technology, India, 2019*

CSE-CIC-IDS2018: *Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018*

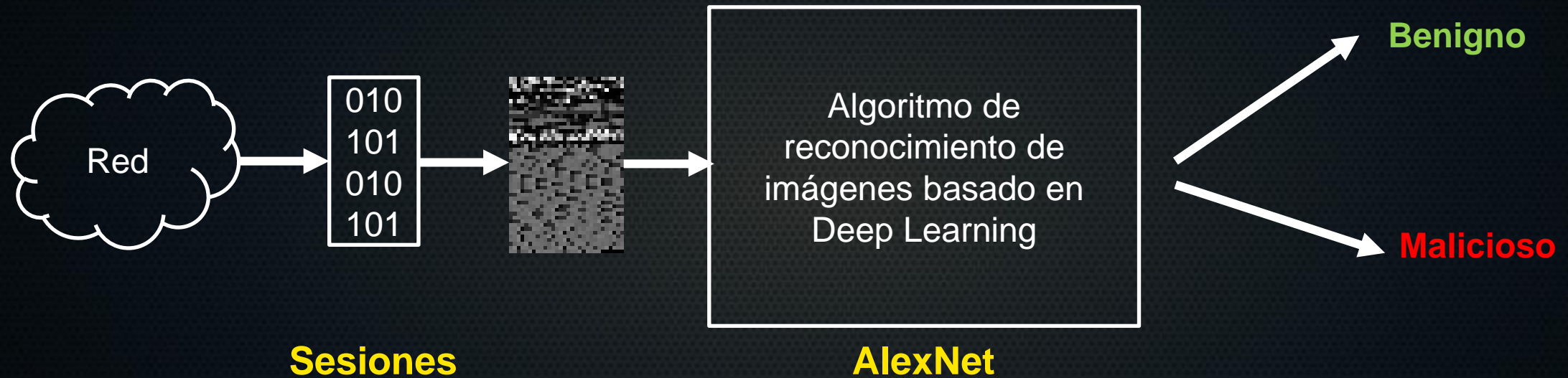
A total of more than 100GB of network traffic

Offline evaluation

Algorithm	Pr	Rc	F1	Execution (Sec.)
KNN	0.96	0.96	0.96	1908.23
RF	0.98	0.97	0.97	74.39
ID3	0.98	0.98	0.98	235.02
Adaboost	0.77	0.84	0.77	1126.24
MLP	0.77	0.83	0.76	575.73
Naive-Bayes	0.88	0.04	0.04	14.77
QDA	0.97	0.88	0.92	18.79

Dataset:	Training (10-fold cross validation)								
Scenario:	A (Malware Binary)			B (Malware Category)			C (Malware Families)		
Algorithm:	RF	KNN	DT	RF	KNN	DT	RF	KNN	DT
Precision (%):	84.00	83.60	85.10	46.50	45.70	46.50	22	21.50	21.00
Recall (%):	87.50	87.30	88.00	45.50	44.80	44.70	21.50	21.60	21.40

Deployment of the model in real-time



NAVAJA
NEGRA

Offline evaluation - DEMO

Advantages

- Does not require feature extraction or selection.
- Increases efficiency and enables real-time detection.
- Allows the creation of customized models for specific environments.
- Does not require an analyst's expertise for creating and maintaining detection rules.

Disadvantages

It **requires a large dataset** to function correctly. While benign traffic is easy to obtain, malicious traffic is not.

NAVAJA
NEGRA

THANK YOU VERY MUCH!

@SANTIAGOHRAMOS

[HTTPS://GITHUB.COM/SHRAMOS](https://github.com/shramos)



shramos

[✉ Sign in to view email](#)

[Block or report user](#)

Overview

Repositories **8**

Projects **0**

Stars **6**

Followers **38**

Following **1**

Popular repositories

[polymorph](#)

Polymorph is a real-time network packet manipulation framework with support for almost all existing protocols

● Python ★ 325 🍴 47

[Awesome-Cybersecurity-Datasets](#)

A curated list of amazingly awesome Cybersecurity datasets

★ 110 🍴 44

[winregmitm](#)

Perform MiTM attack and remove encryption on Windows Remote Registry Protocol.

● Python ★ 17 🍴 <https://t.me/learningnets>

[pyc-cfg](#)

Pyc-cfg is a pure python control flow graph builder for almost all Ansi C programming language.

● Python ★ 14 🍴 3