Toward a Search-Based Approach to Support the Design of Security Tests for Malicious Network Traffic

Davide La Gamba, **Gerardo Iuliano**, Gilberto Recupito, Giammaria Giordano,
Filomena Ferrucci, Dario Di Nucci, Fabio Palomba

University of Salerno, Italy
Department of Computer Science
Software Engineering (SeSa) Lab

geiuliano@unisa.it
GerardoIuliano.github.io
@GerardoIuliano

MIRAI, October 21, 2016

Multiple DDoS attacks occurred in DNS service using the Mirai malware.

Malware was installed on a large number of IoT devices.

sesa lab
SOFTWARE ENGINEERING
SALERNO

Security Testing

State of the art

Intensive Preprocessing of KDD Cup 99 for Network Intrusion Classification Using Machine Learning Techniques

[1]Ibrahim Obeidat, [2]Nabhan Hamadneh, [3]Mouhammd Alkasassbeh, [4]Mohammad Almseidin

[1, 2] Prince Al-Hussein Bin Abdullah II Faculty of Information Technology, Hashemi
e-mails: { imsobeidat, nabhan}@hu.edu.jo

[3]Computer Science Department, Princess Sumaya University for Technol
e-mail: m.alkasassbeh@psut.edu.jo

[4]Department of Information Technology, University of Miskolc, H-3515 Mis
e-mails: alsaudi@iit.uni-miskolc.hu

Abstract— **Network security engineers work to keep services available all the time by handling intruder attacks. Intrusion Detection System (IDS) is one of the obtainable mechanism that used to sense and classify any abnormal actions. Therefore, the IDS must be always up to date with the latest intruder attacks signatures to preserve confidentiality, integrity and availability of the services. The speed of the IDS is very important issue as well learning the new attacks. This research work illustrates how the Knowledge Discovery and Data Mining (or Knowledge Discovery in Databases) KDD dataset is very handy for testing and evaluating different Machine Learning Techniques. It mainly focuses on the KDD preprocess part in order to prepare a decent and fair experimental data set. The techniques J48, Random**

its users [1]. Remote to use
network attacks, which an
another computer or server
have permission to access as
(U2R) is a second type of
access the network resources
attempts the intruder become
is a third type of attack in
devices to determine weakn
opened ports and then use th
to personal information.
represent probing over a ne

The Journal of Systems & Software 193 (2022) 111475

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

On the use of artificial intelligence to deal with privacy in IoT systems: A systematic literature review☆

Giammaria Giordano*, Fabio Palomba, Filomena Ferrucci

SeSa Lab - Department of Computer Science, University of Salerno, Italy

ARTICLE INFO

Article history:
Received 2 December 2021
Received in revised form 29 July 2022

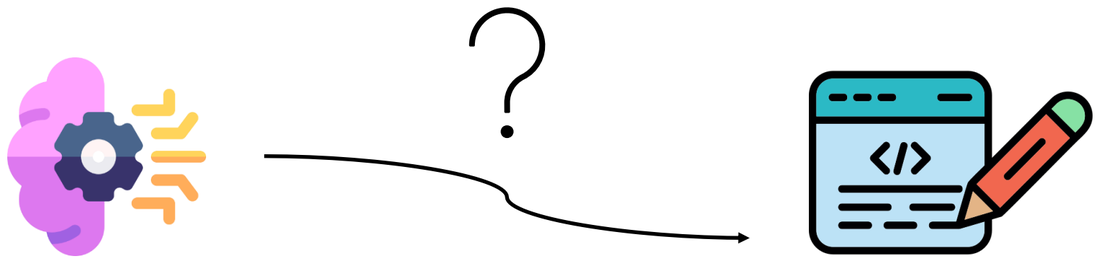
ABSTRACT

The Internet of Things (IoT) refers to a network of Internet-enabled devices that can make different operations, like sensing, communicating, and reacting to changes arising in the surrounding environ-

Limitation

ML model outputs are hard to use for test case generation.

It is difficult to understand how the model generated the results.



Our idea

Detection Rule

```
if (
duration <= 5855 AND
protocolType == icmp AND
srcBytes >= 462 AND
srvCount >= 1 AND
diffSrvRate <= 90)
then {
"attack found"}
```

Test Case

```
@Test
void testConnection(){
  Assert.IsTrue(
    duration <= 5855,
    protocolType == "icmp",
    srcBytes >= 462,
    srvCount >= 1,
    diffSrvRate <= 90,

  "DoS attack found");
}
```

Al-fuhaidi *et al.*

Al-fuhaidi et al.

DoS

Our work

DoS, Probe, U2R, R2L

Performance Evaluation of a Genetic Algorithm Based Approach to Network Intrusion Detection System

B. Abdullah*, I. Abd-alghafar**, Gouda I. Salama** and A. Abd-alhafez**

Abstract: The purpose of the work described in this paper is to provide an intrusion detection system (IDS), by applying genetic algorithm (GA) to network intrusion detection system. Parameters and evolution process for GA are discussed in detail and implemented. This approach uses information theory to filter the traffic data and thus reduce the complexity. We use a linear structure rule to classify the network behaviors into normal and abnormal behaviors. This approach applied to the KDD99 benchmark dataset and obtained high detection rate up to 99.87% as well as low false positive rate 0.003%. Finally the results of this approach compared with available machine learning techniques.

Keywords: Intrusion Detection System, Genetic Algorithm, Open Source Weka software.

1. Introduction

Internet and local area networks are expanding at an amazing rate in recent years, not just in the terms of size, but also in the terms of changing the services offered and the mobility of users that make them more vulnerable to various kinds of complex attacks. While we are benefiting from the convenience that new technology has brought us, computer systems are exposed to increasing number and complexity of security threats

Goal of the study

- To what extent genetic algorithm can be used to detect intrusion attacks.
- Provide a set of detection rules that security tester can use to build security tests.

Research Questions

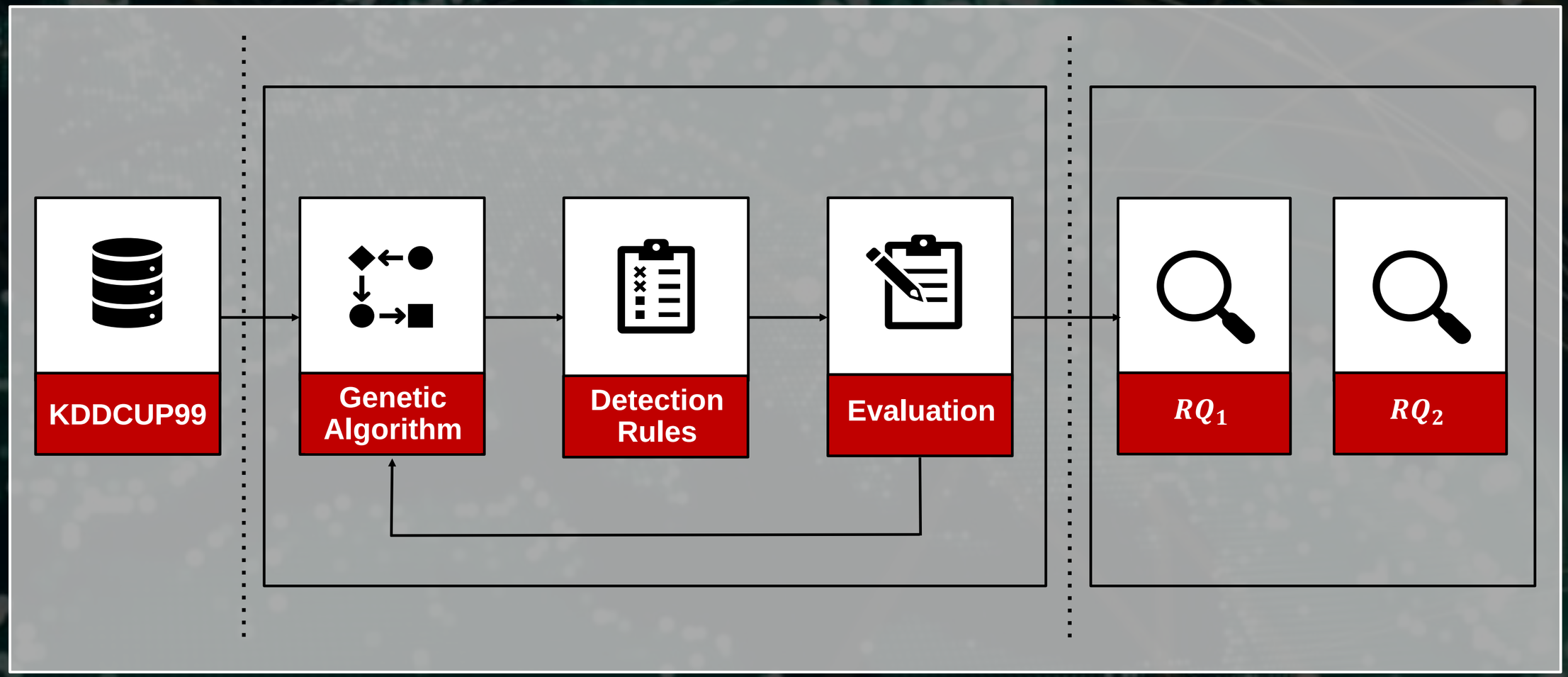
RQ_1

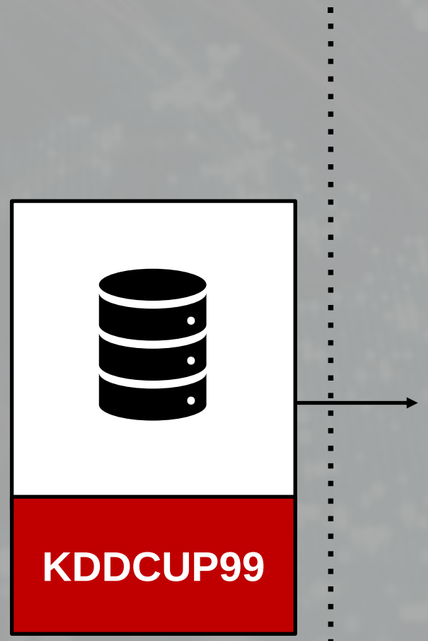
To what extent can genetic algorithms generate detection rules to identify DoS attacks?

Research Questions

RQ_1	To what extent can genetic algorithms generate detection rules to identify DoS attacks?
RQ_2	What are the performance of GAs in detecting intrusion attacks?

Research Method





Attack Type

4

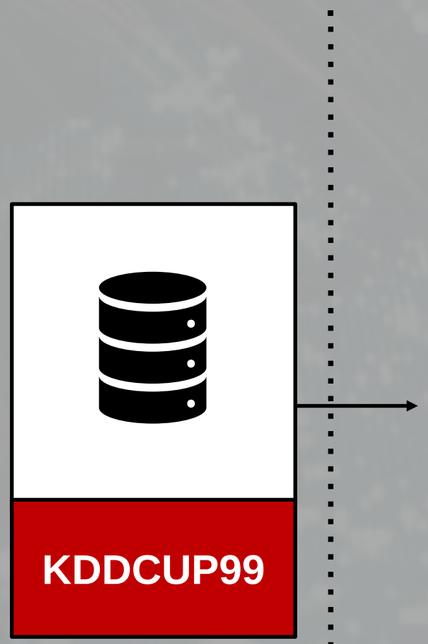
Denial of Service Attacks (**DoS**)

Probing Attacks (**Probe**)

User To Root Attacks (**U2R**)

Remote To Local Attacks (**R2L**)

Normal Connections

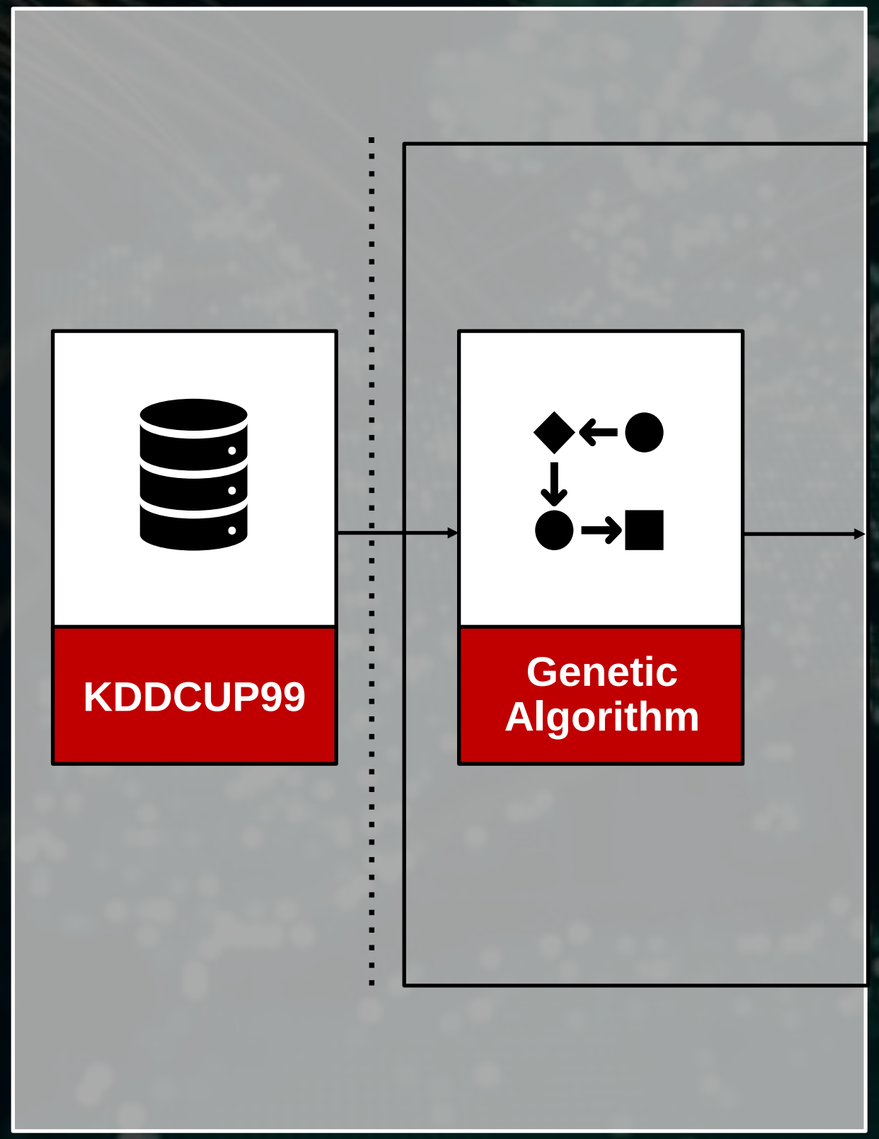


Features

31

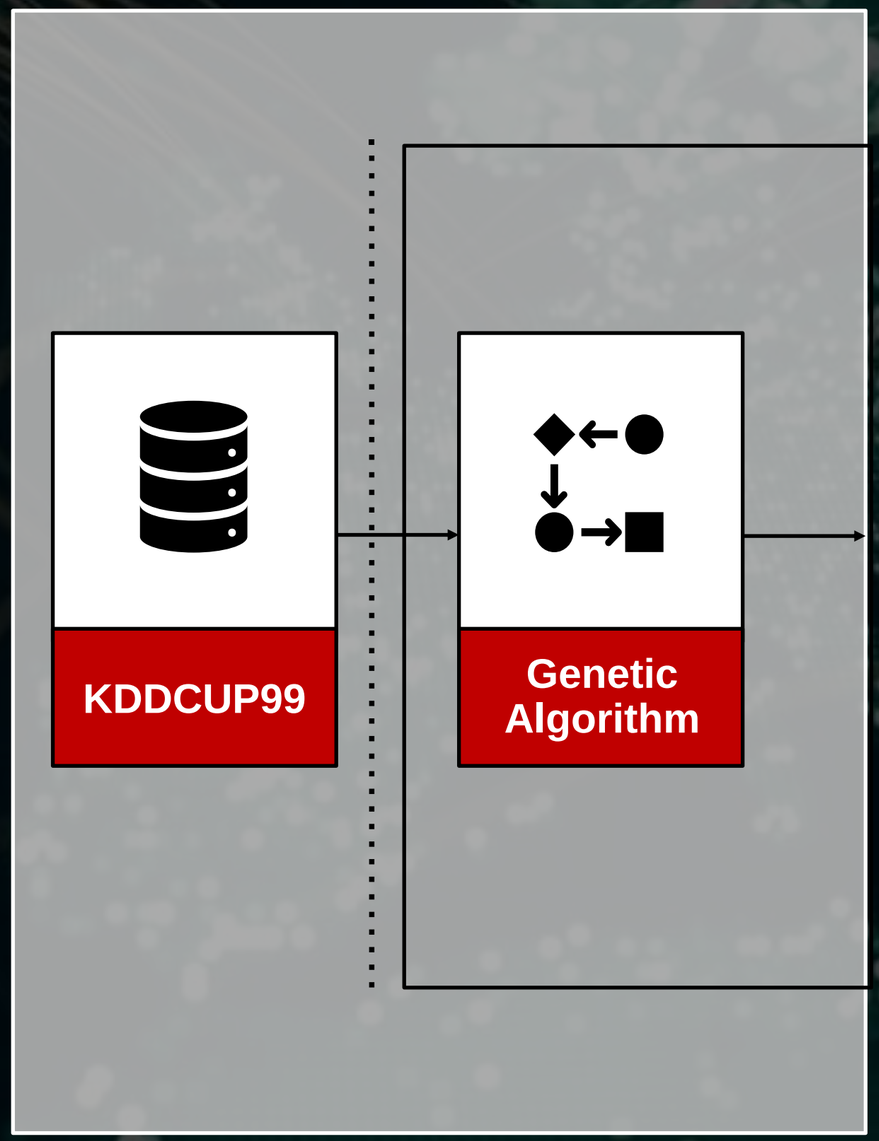
Basic features of TCP connections

Derived features



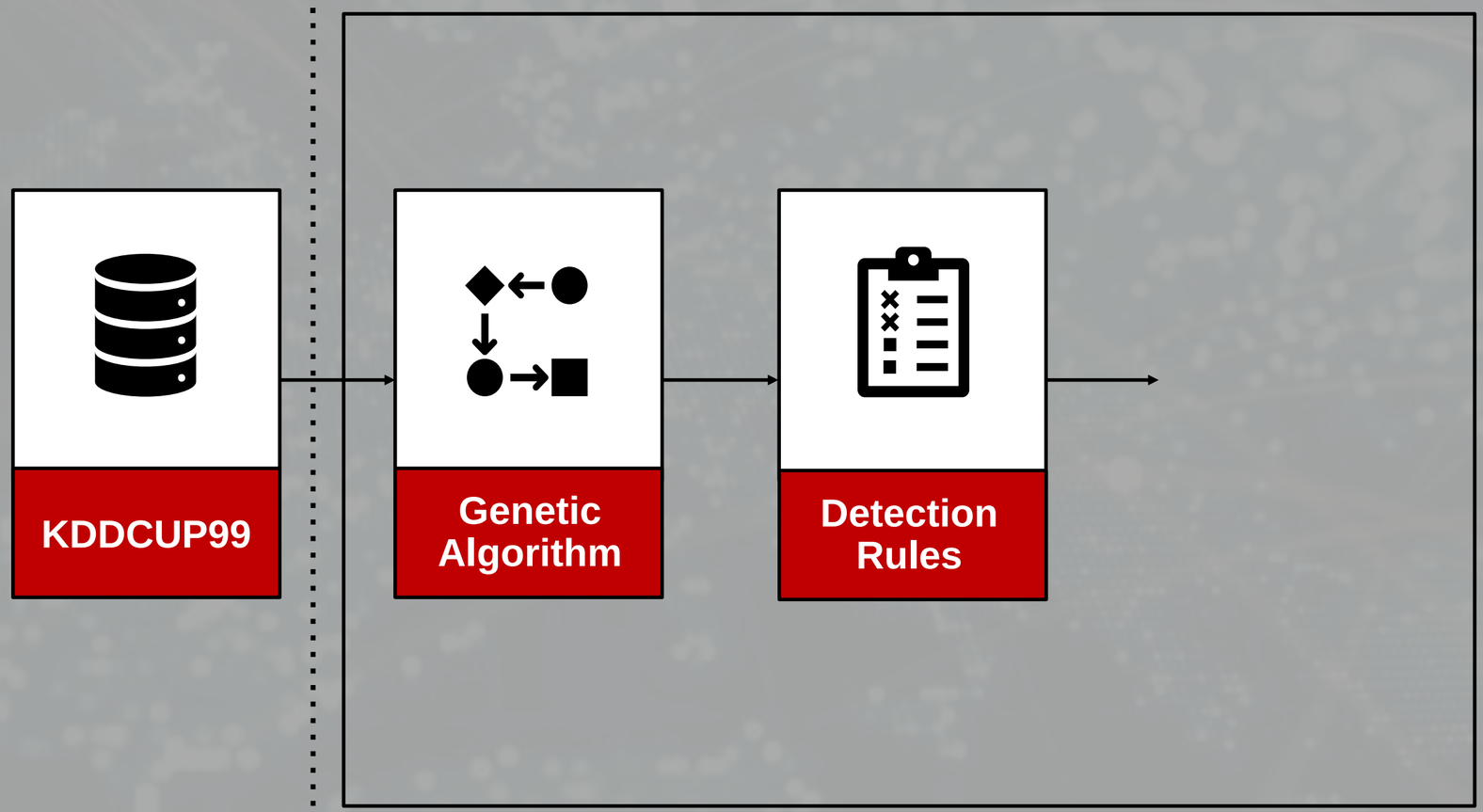
Individual Encoding

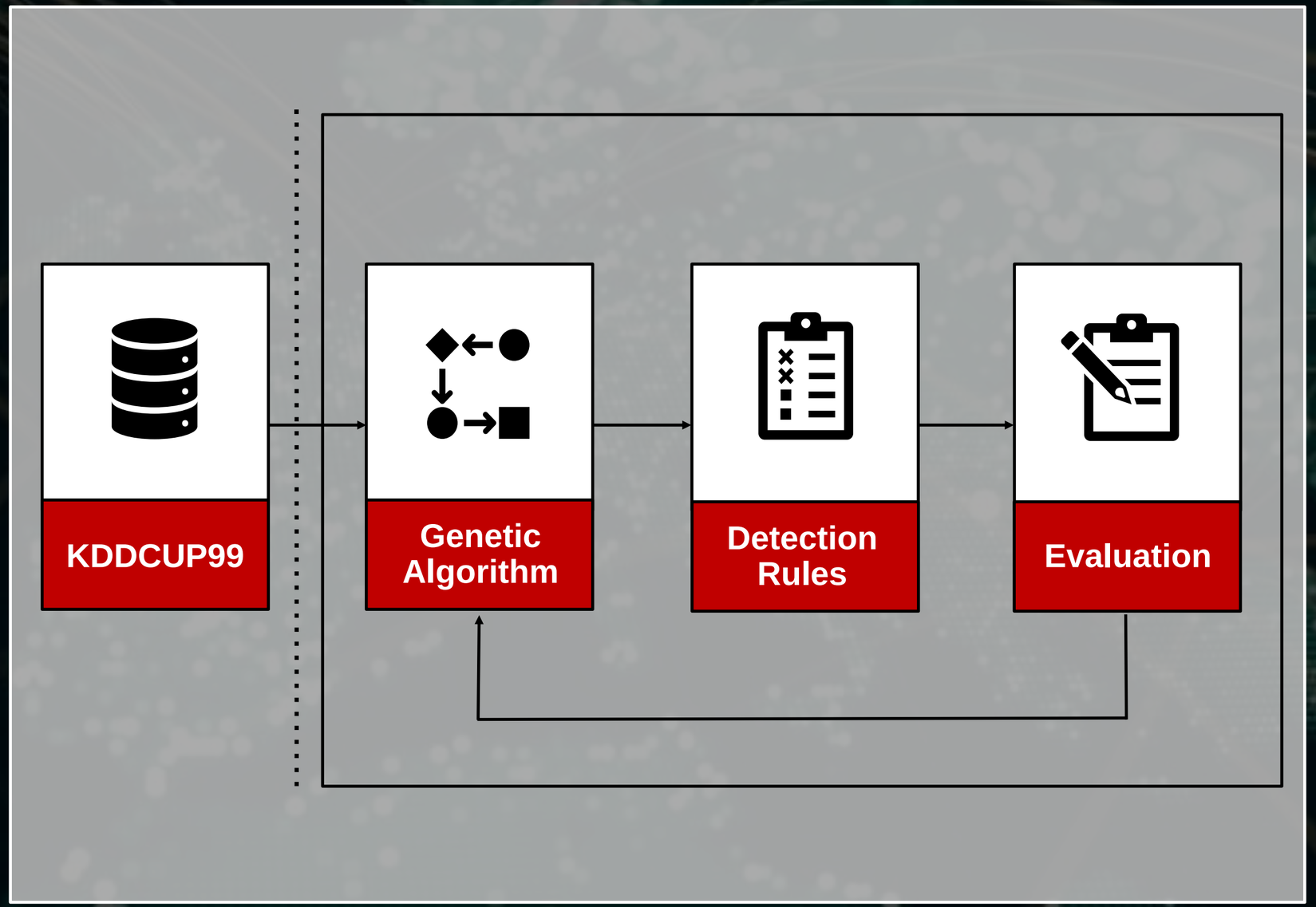
```
1. if (
2. duration <= 5855 AND
3. protocolType == icmp AND
4. srcBytes >= 462 AND
5. srvCount >= 1 AND
6. diffSrvRate <= 90)
7. then {
8. attack found}
```



Parameter Configuration

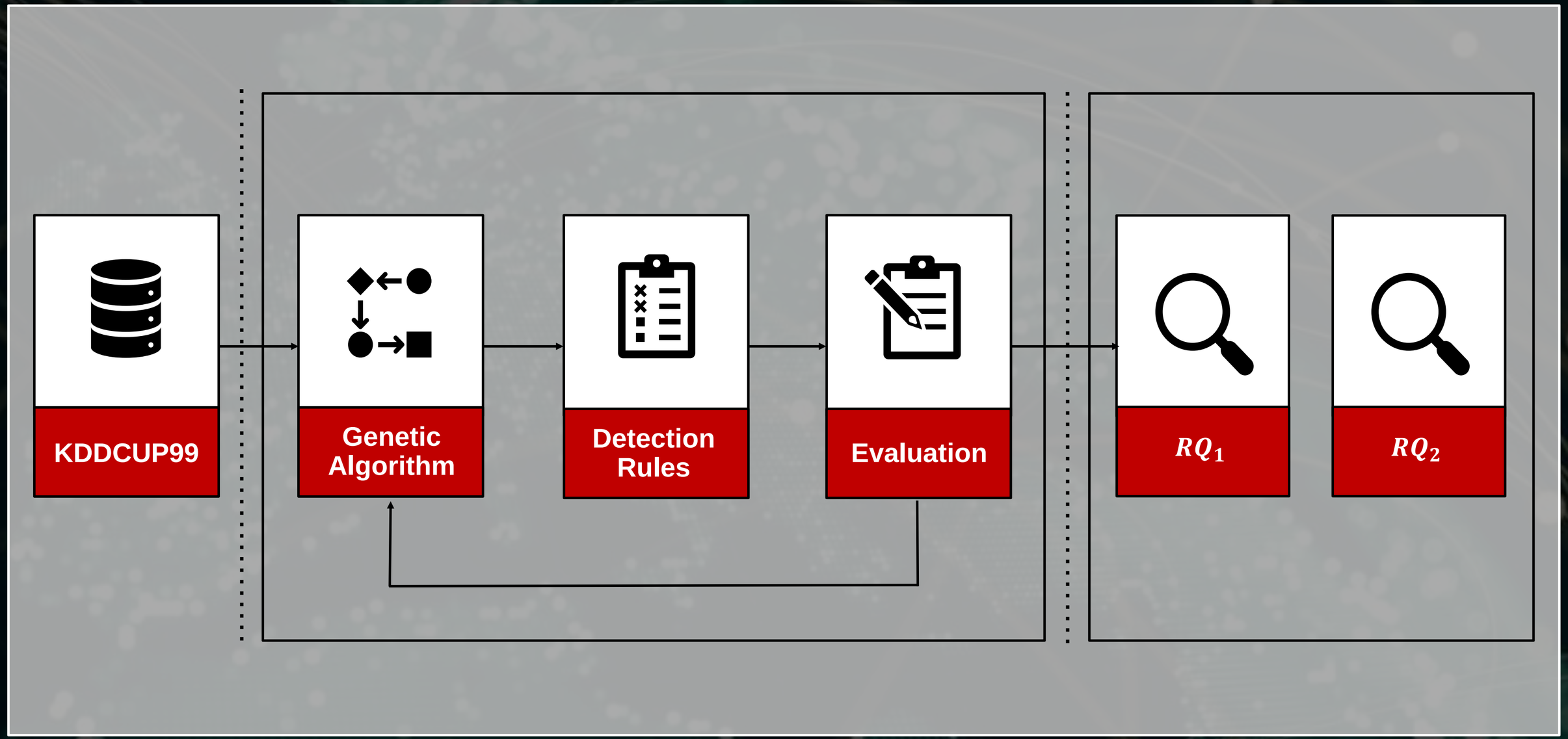
Parameters	Values
Initial Population	***1'000*** individuals
Stopping Criteria	***1'000*** generations
Mutation	***10%*** probability
Crossover	***90%*** probability
Selection Operator	***Elite*** selector


KDDCUP99
Genetic Algorithm
Detection Rules

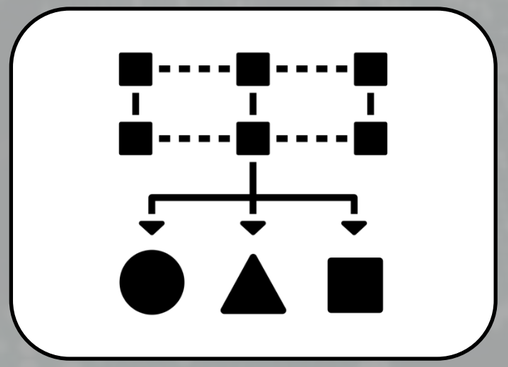


Metrics

- Precision
- Recall
- F-Measure
- MCC
- Accuracy
- Specificity

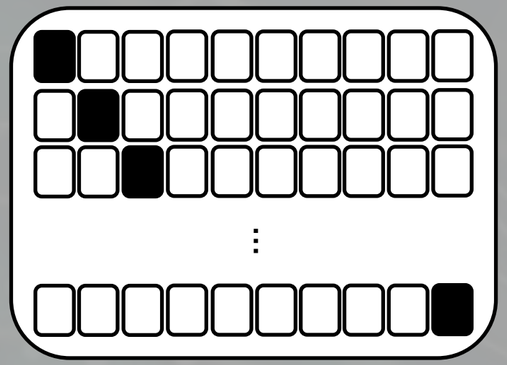

KDDCUP99
Genetic Algorithm
Detection Rules
Evaluation
RQ_1
RQ_2

Validation



Multiclass Classification

No Data Balancing



10-Fold Cross-Validation

Results

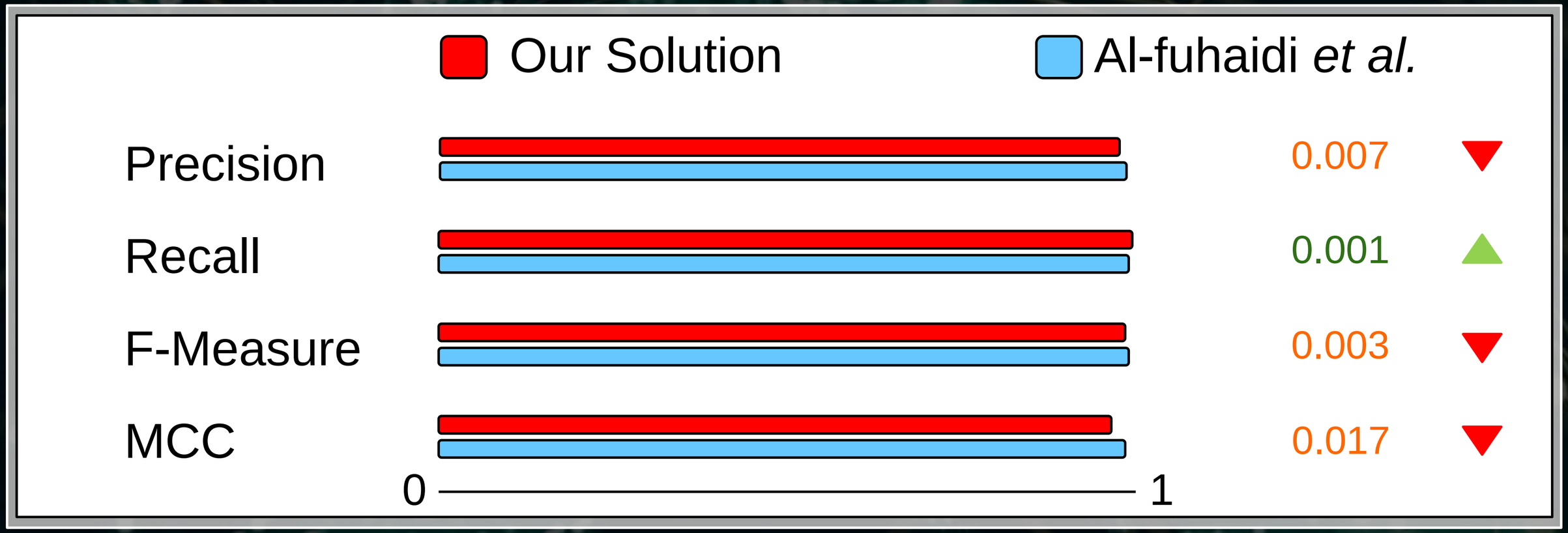
17 Detection Rules

11
DoS Attacks

4
Probe Attacks

2
U2R Attacks

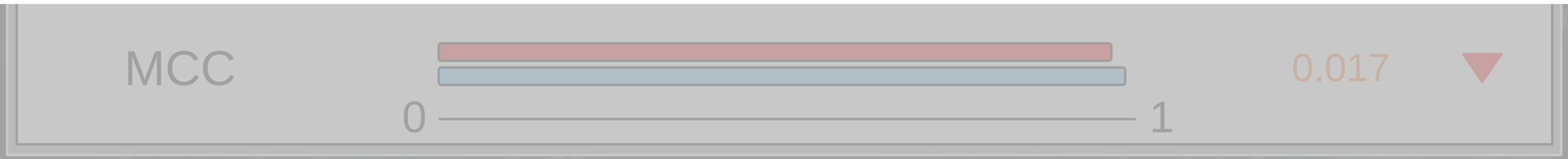
Results RQ_1



RQ_1 To what extent can genetic algorithms generate detection rules to identify DoS attacks?

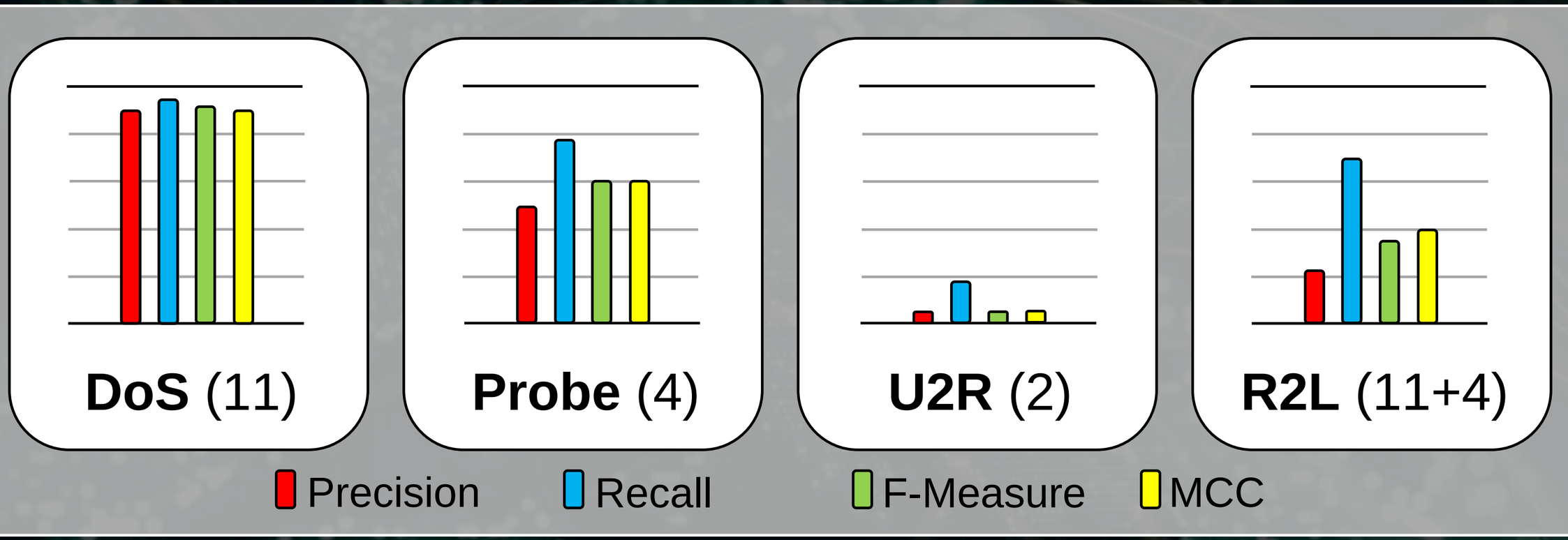
Results RQ_1

State-Of-The-Art has been **confirmed**.



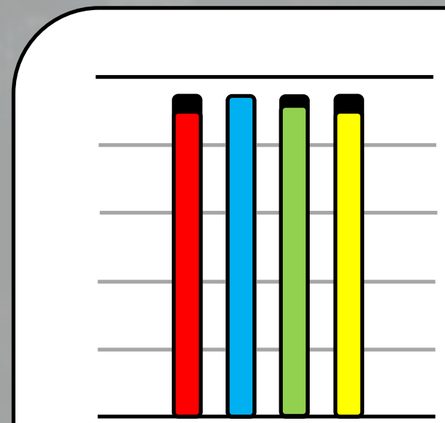
RQ_1 To what extent can genetic algorithms generate detection rules to identify DoS attacks?

Results RQ_2. GA Performance

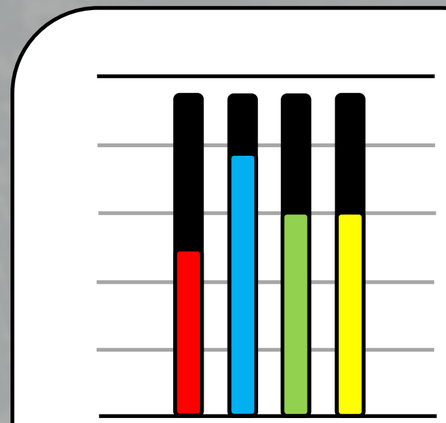


RQ_2 What are the performance of GAs in detecting intrusion attacks?

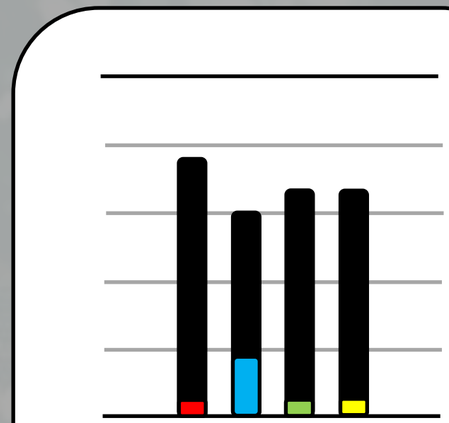
Results RQ_2. GA vs ML



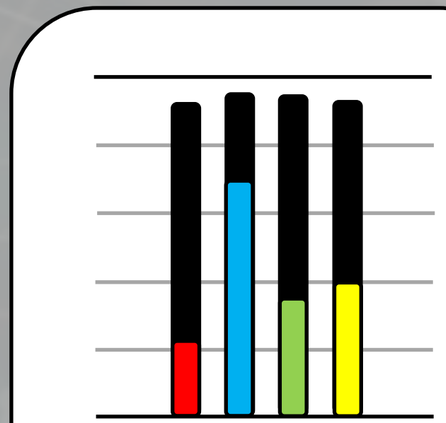
DoS (11)



Probe (4)



U2R (2)



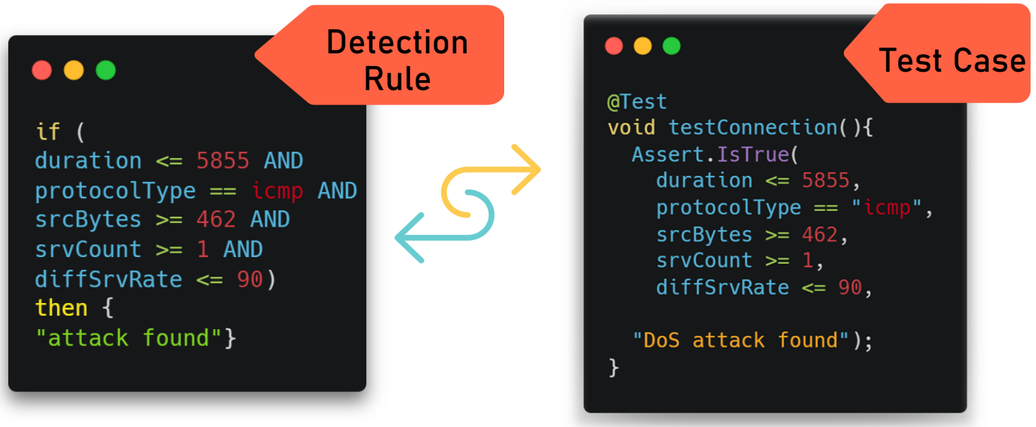
R2L (11+4)

Precision Recall F-Measure MCC

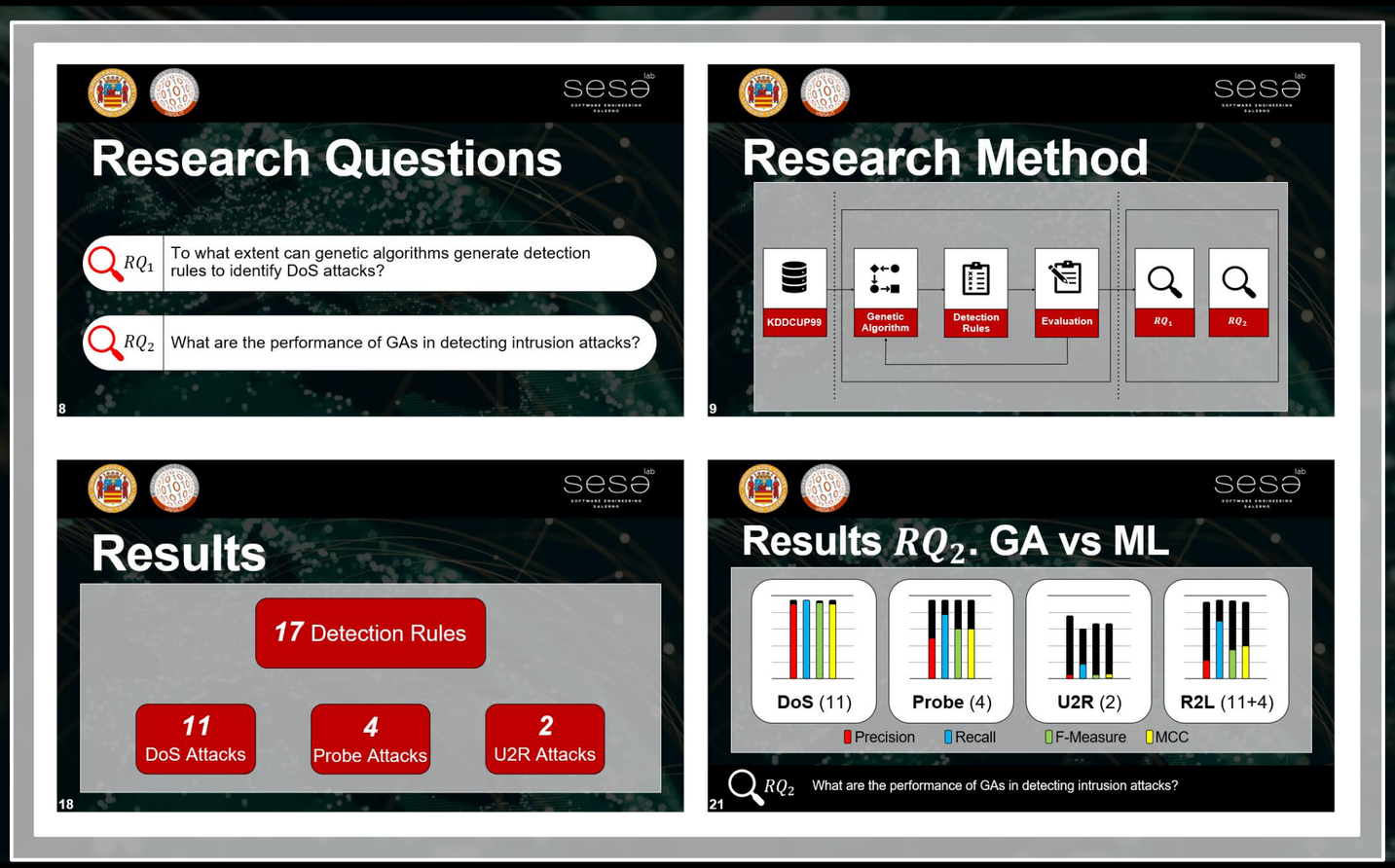
RQ_2 What are the performance of GAs in detecting intrusion attacks?

Conclusion

- The results denote **good performance** in detecting intrusion attacks when the dataset used has an **adequate number of malicious instances**.
- In other cases, **genetic algorithms can complement machine learning** models by providing support to the security test in writing the test case.



geiuliano@unisa.it
GerardoIuliano.github.io
@GerardoIuliano



Toward a Search-Based Approach to Support the Design of Security Tests for Malicious Network Traffic

sesa lab
SOFTWARE ENGINEERING SALERNO

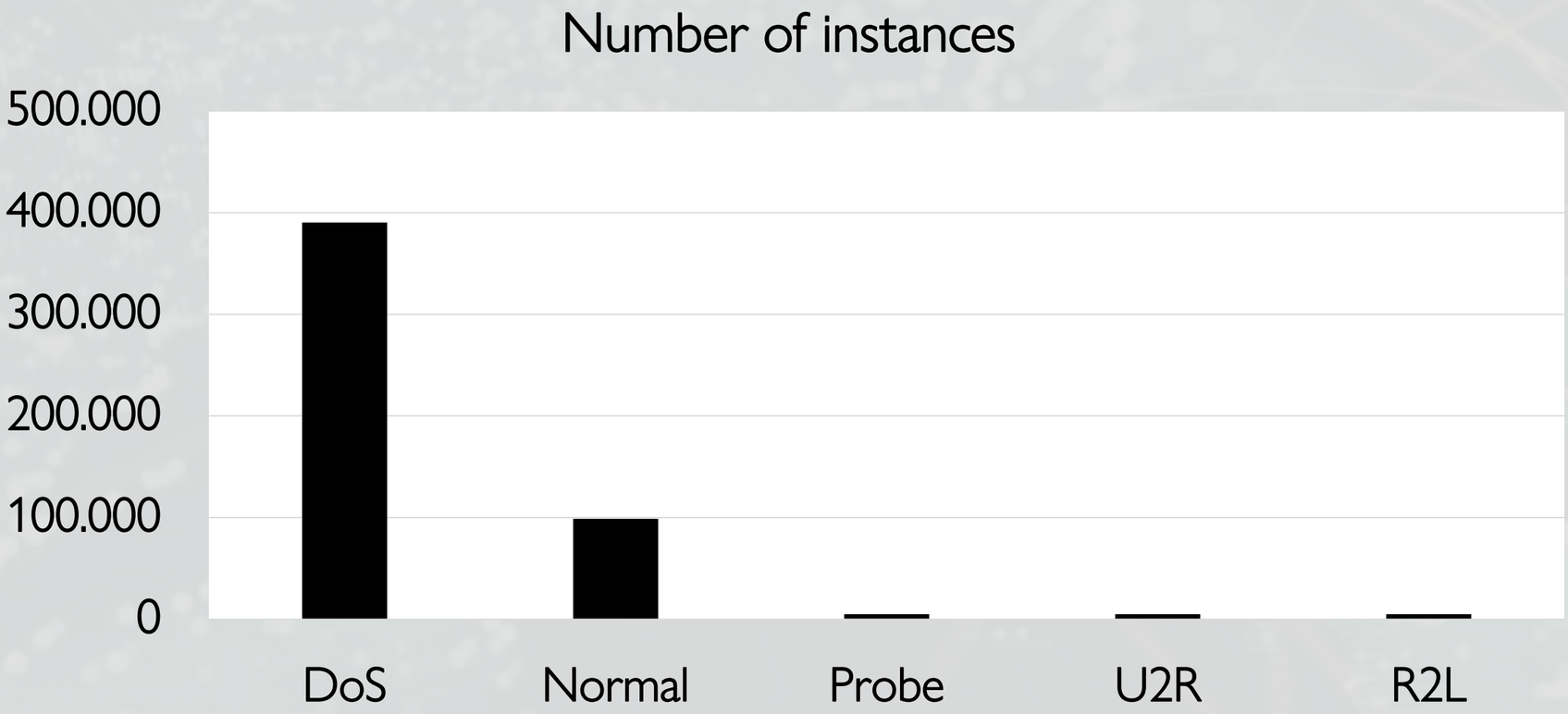
geiuliano@unisa.it

Gerardoluliano.github.io

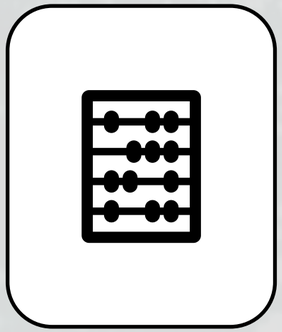
@GerardoIuliano


Backup Slides

KDDCUP99 (10% version)



Future Work



We will apply and validate these features for test case generation in industrial contexts.

We will investigate the characteristics of R2L attacks, finding correlations with other types of attacks based on the features that we selected.

Threats To Validity

- We are conscious that the dataset selection may influence the results.
- The imbalanced dataset may affect the generalizability of our results.
- Another critical concern is the risk of overfitting GAs.

To mitigate this, we introduced multiple detection rules for each attack and diversified conditions for intrusion detection.

Fitness Function

$$Fitness = \frac{Detected\ Attacks}{Total\ Attacks} - \frac{False\ Attacks}{Total\ Connetions}$$

Higher values indicate rules that are better able to detect attacks in the dataset, so the fitness measure should be maximized.

Difference with SOTA

Our Work	Al-fuhaidi *et al*.
Detection Rules	
Denial of Service	*Denial of Service*
Probe	*X*
User To Root	*X*
Remote To Local	*X*
Validation	
Bayes Network	*Bayes Network*
Decision Tree	*Decision Tree*
Support Vector Machine	*Support Vector Machine*
Decision Table	*X*
Naive Bayes	*X*
Random Forest	*X*