Machine Learning-based Detection of C&C Channels with a Focus on Locked Shields

Nicolas Känzig[(1)], Roland Meier[(1)], Luca Gambazzi[(2)],
Vincent Lenders[(2)], Laurent Vanbever[(1)]

CYCON

(1) ETH zürich

(2) Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
armasuisse

TECH \ CYBERSECURITY

UK hospitals hit with massive ransomware attack

52

Sixteen hospitals shut down as a result of the attack

By Russell Brandom | May 12, 2017, 11:36am EDT

Peter O'Conner / Flickr

A massive ransomware attack has shut down work at 16 hospitals across the United Kingdom. According to *The Guardian*, the attack began at roughly 12:30PM local time, freezing systems and encrypting files. When employees tried to access the computers, they

Ukraine power cut 'was cyber-attack'

11 January 2017 Share

Ukraine's energy grid has been attacked twice by hackers

A power cut that hit part of the Ukrainian capital, Kiev, in December has been judged a cyber-attack by researchers investigating the incident.

The blackout lasted just over an hour and started just before midnight on 17 December.

The cyber-security company Information Systems Security Partners (ISSP) has linked the incident to a **hack and blackout in 2015** that affected 225,000.

THE IRISH TIMES

Mon, May 20, 2019

Dublin 14°C

Cyber defenders fight hackers in high-tech Estonia war

Attacks on vital systems and fake news

Fri, Apr 28, 2017, 01:00

Daniel McLaughlin in Tallinn

0

The attack on the airbase began with a salvo of fake news. “A report appeared saying drones were using nerve gas,” said Lauri Luht, crisis management chief for the cyber security department of Estonia’s information system authority.

THE IRISH TIMES

Mon, May 20, 2019

Dublin 14°C

Cyber defenders fight hackers in high-tech Estonia war games

Attacks on vital systems and fake news are all part of Locked Shields exercise

Fri, Apr 28, 2017, 01:00

Daniel McLaughlin in Tallinn

0

Locked Shields, now taking place in Estonia involving 20 teams from Europe and the US, is the world's most advanced live-fire cyber defence exercise. Photograph: Daniel McLaughlin

The attack on the airbase began with a salvo of fake news. "A report appeared saying drones were using nerve gas," said Lauri Luht, crisis management chief for the cyber security department of Estonia's information system authority.

Locked Shields is the largest live-fire global cyber defense exercise

Picture: NATO CCDCOE

Locked Shields is the largest live-fire global cyber defense exercise

- Red Team vs. Blue Team exercise

Attackers	Defenders
1 Team	1 Team / country

CYCON

Picture: NATO CCDCOE

Locked Shields is the largest live-fire global cyber defense exercise

- Red Team vs. Blue Team exercise

Attackers	Defenders
1 Team	1 Team / country

- 1’200 experts from 30 nations
- 4’000 virtualized systems
- 2’500 attacks

CYCON

Picture: NATO CCDCOE

If the Swiss Blue Team had used our system at Locked Shields 2018, it would have discovered more than 80% of the C&C servers within 30 minutes.

Overview

Locked Shields:
the largest cyber defense exercise

Identifying C&C channels
in real time and
with little resources

Evaluation
on real data from Locked Shields 2017 and 2018

CYCON

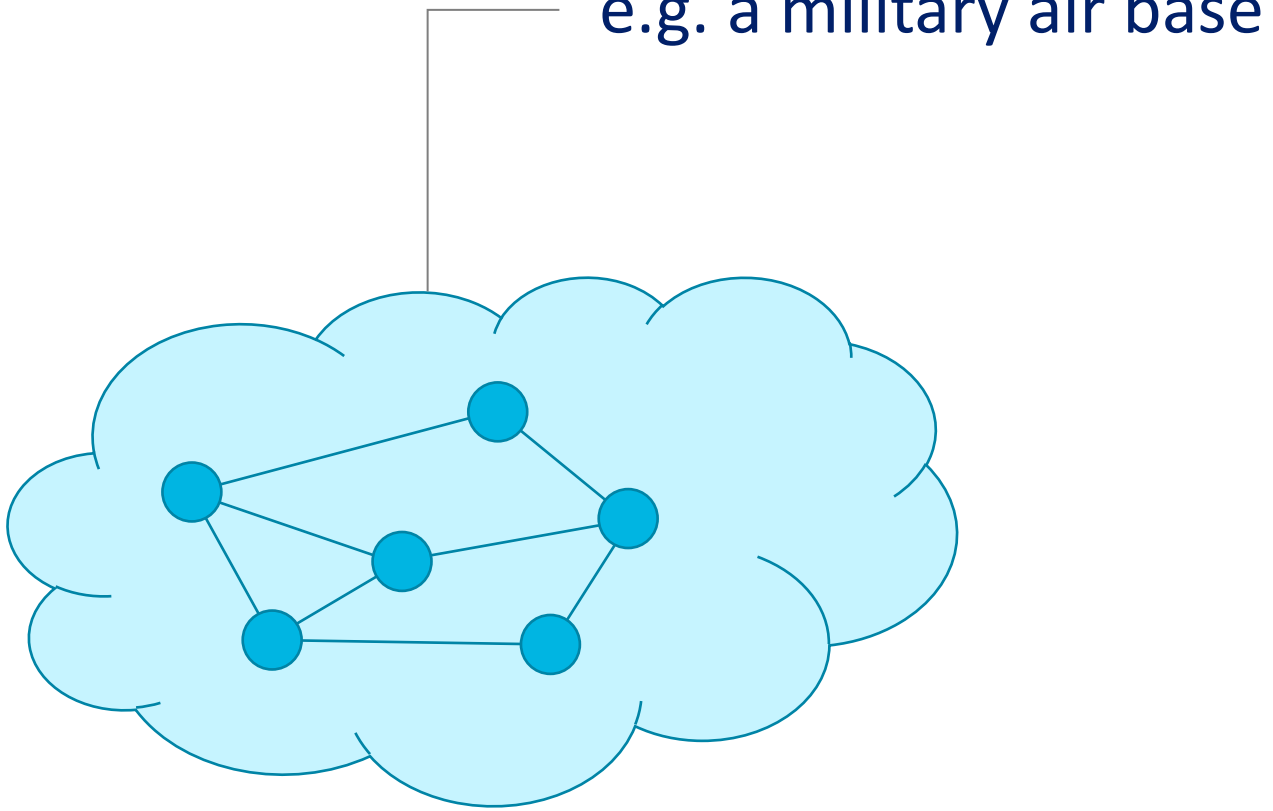
Locked Shields:
the largest cyber
defense exercise

Identifying C&C channels
in real time and
with little resources

Evaluation
on real data from Locked
Shields 2017 and 2018

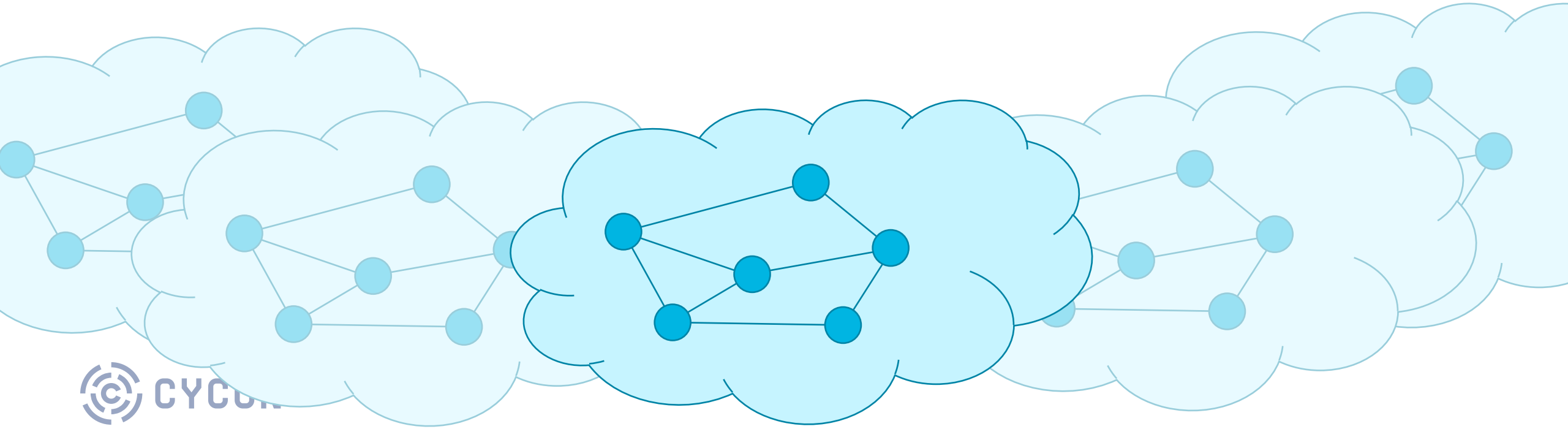
CYCON

Each Blue Team has its own network (“Gamenet”) to defend

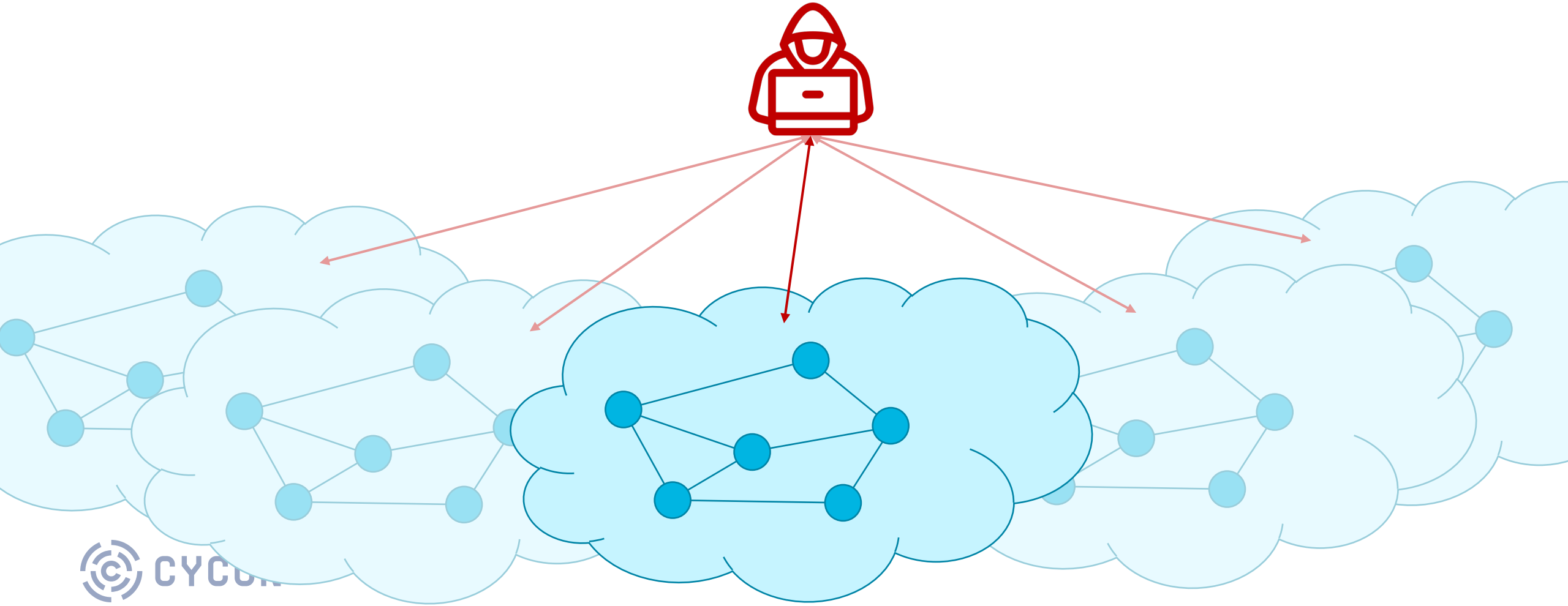


CYCON

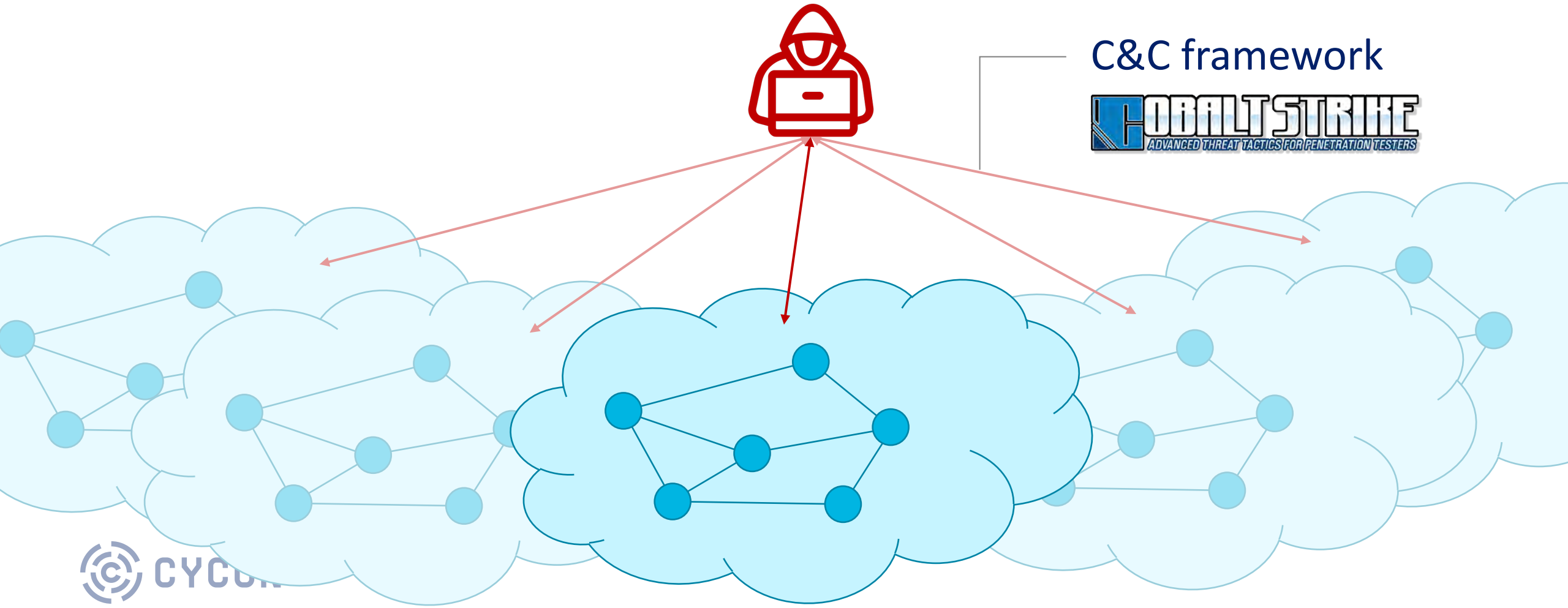
Each Blue Team has its own network (“Gamenet”) to defend



One Red Team attacks all Gamenets



One Red Team attacks all Gamenets (partially) using C&C infrastructure

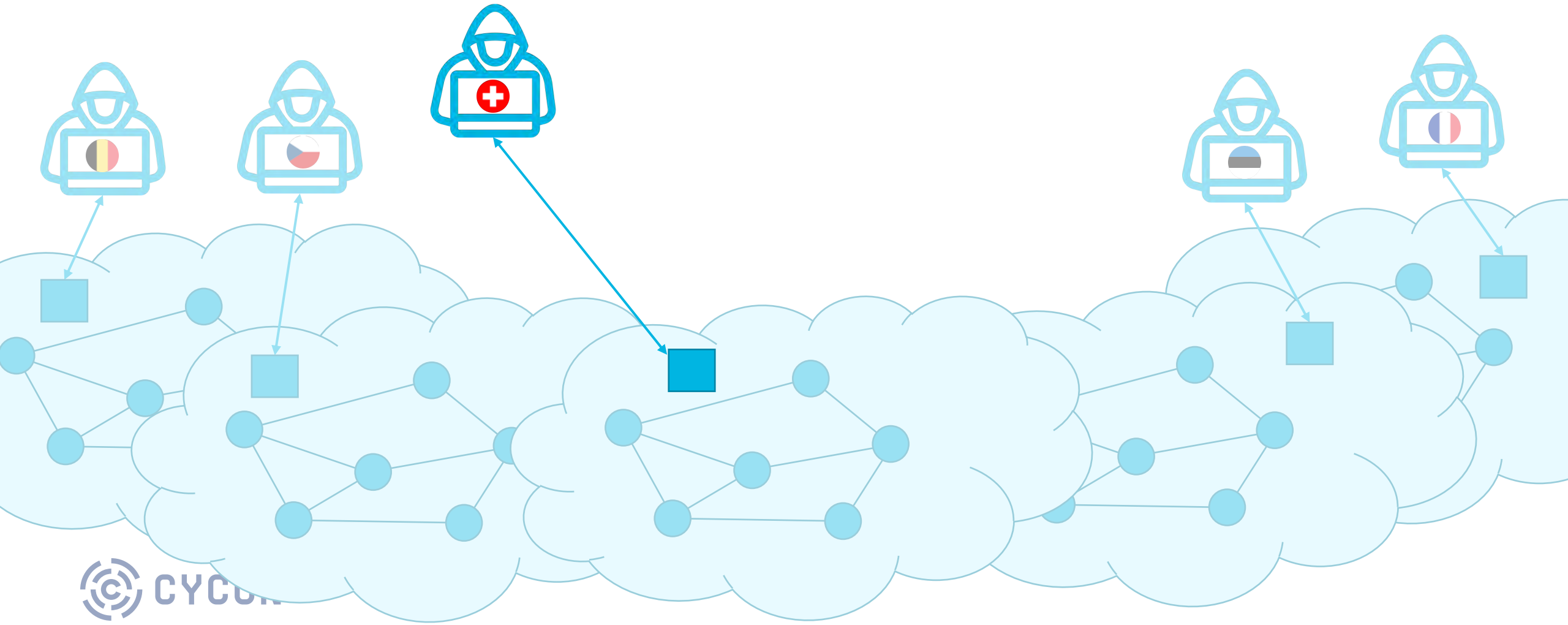


The setting for the Red Team during Locked Shields resembles the one for real attackers

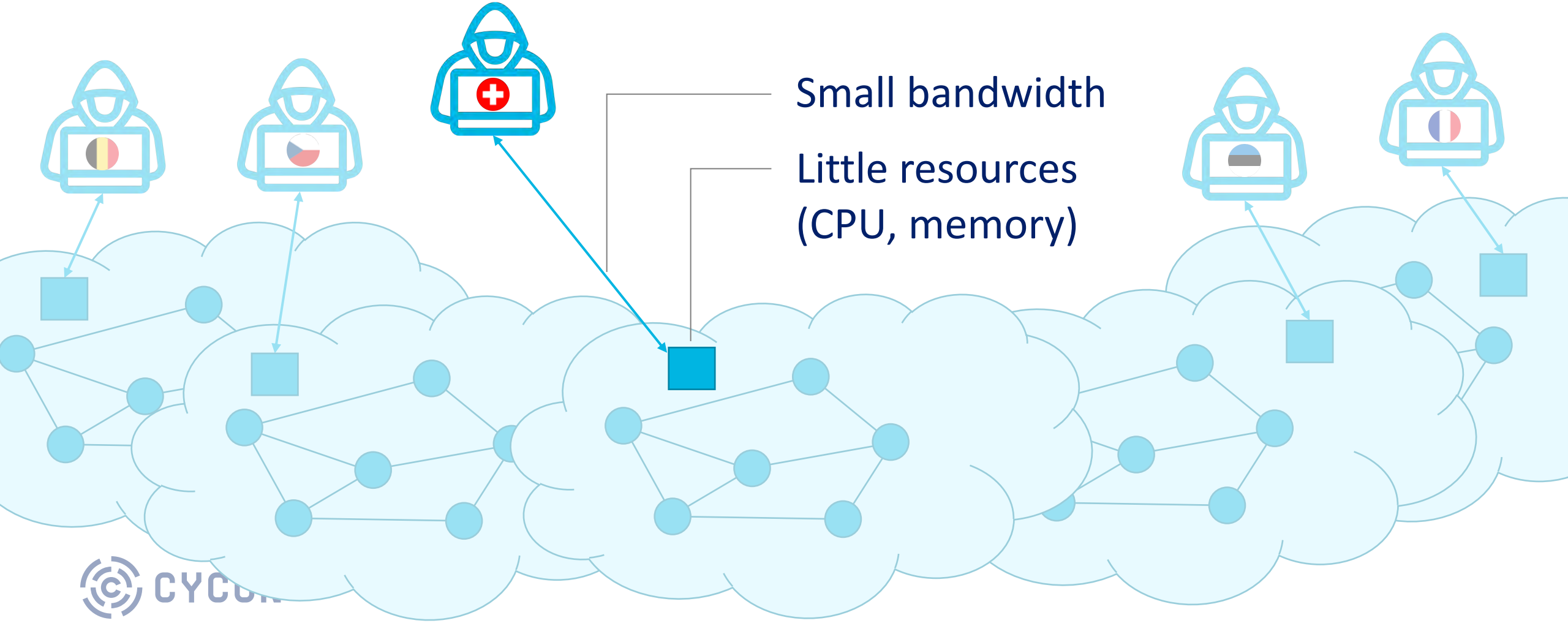
- Skilled attackers
- Exploit various weaknesses of a network
- Run attacks (partially) via a C&C infrastructure

CYCON

Blue Teams can run tools and sniff traffic inside the Gamenet



Blue Teams can run tools and sniff traffic inside the Gamenet *with some constraints*



The setting for Blue Teams during Locked Shields resembles the one for real defenders

- Arrive when the network is already under attack
- Systems are already compromised and poorly documented
- Users demand availability of the infrastructure

CYCON

Locked Shields:
the largest cyber defense exercise

Identifying C&C channels in real time and with little resources

Evaluation
on real data from Locked Shields 2017 and 2018

CYCON

CYCON

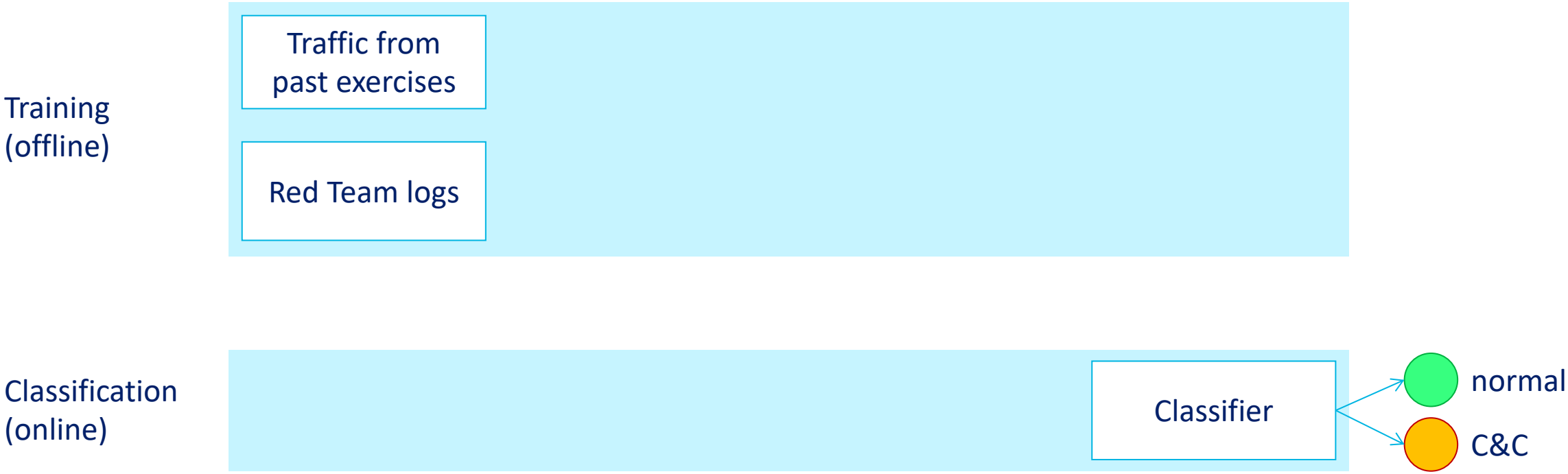
We aim at a classifier that classifies between normal and C&C traffic

CYCON

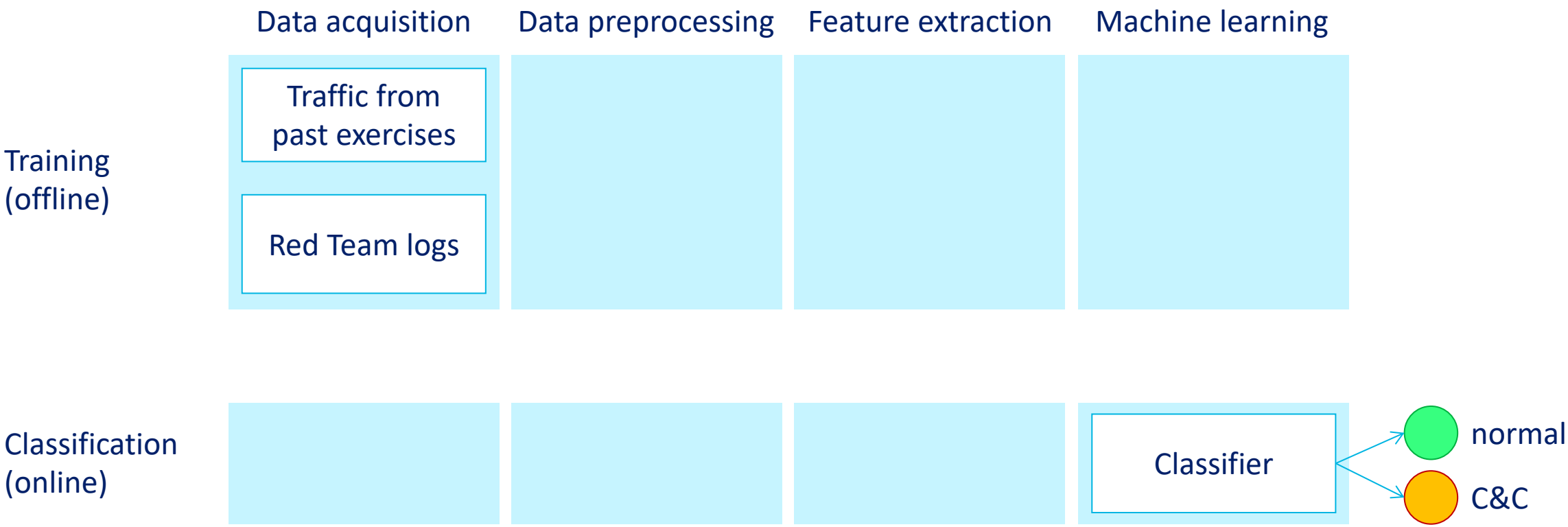
We aim at a classifier that classifies between normal and C&C traffic

CYCON

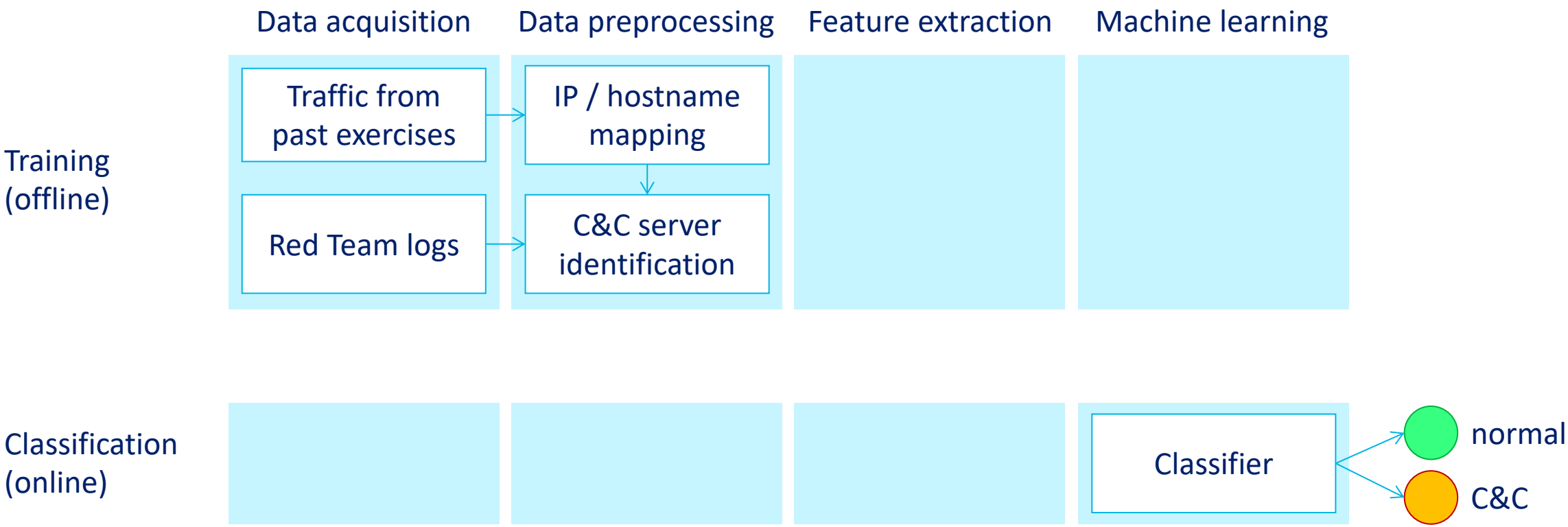
Training happens offline, classification happens online



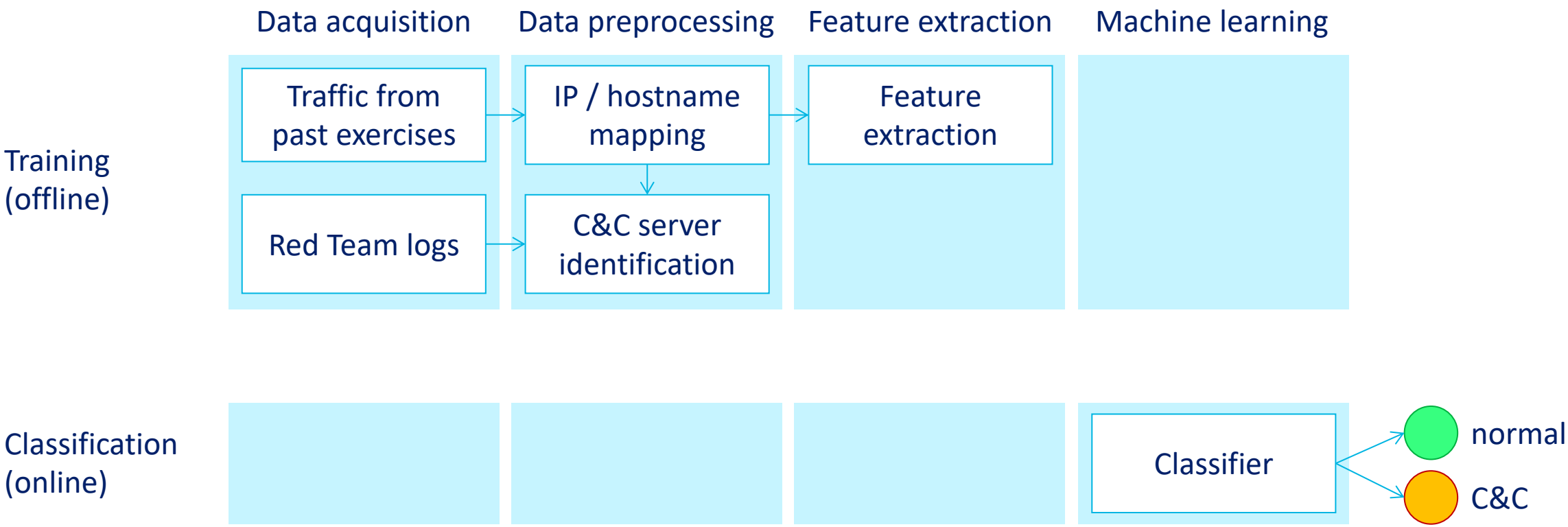
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
Classification (online)
Classifier
normal
C&C

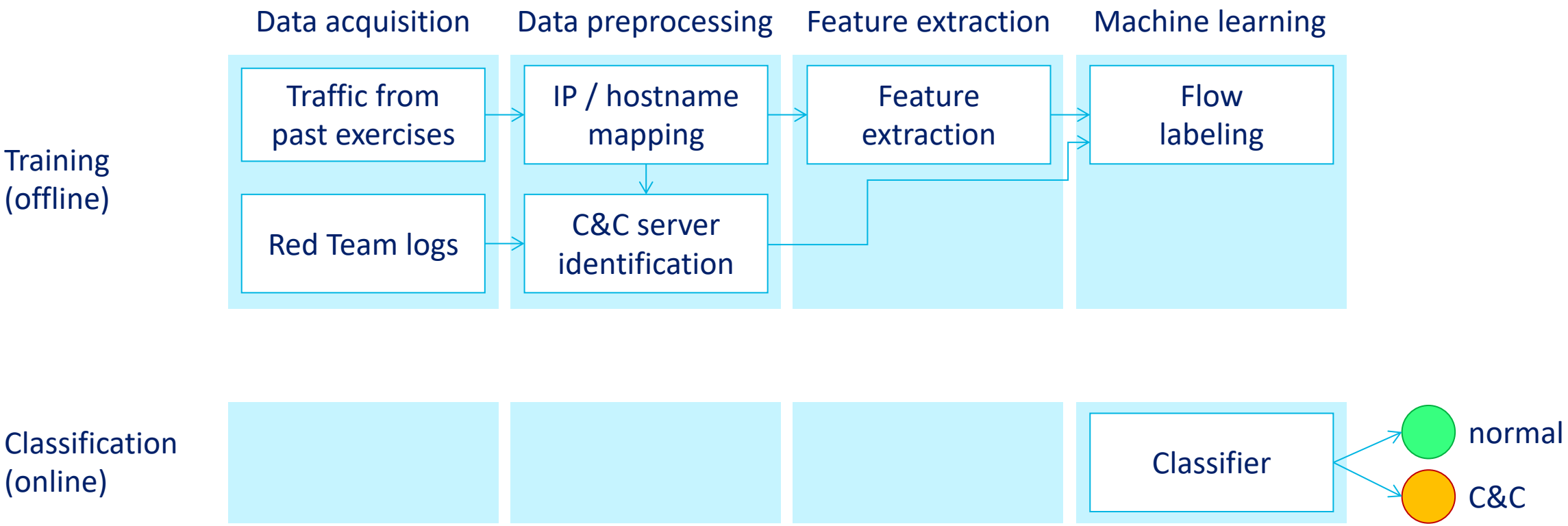
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Classification (online)
Classifier
normal
C&C

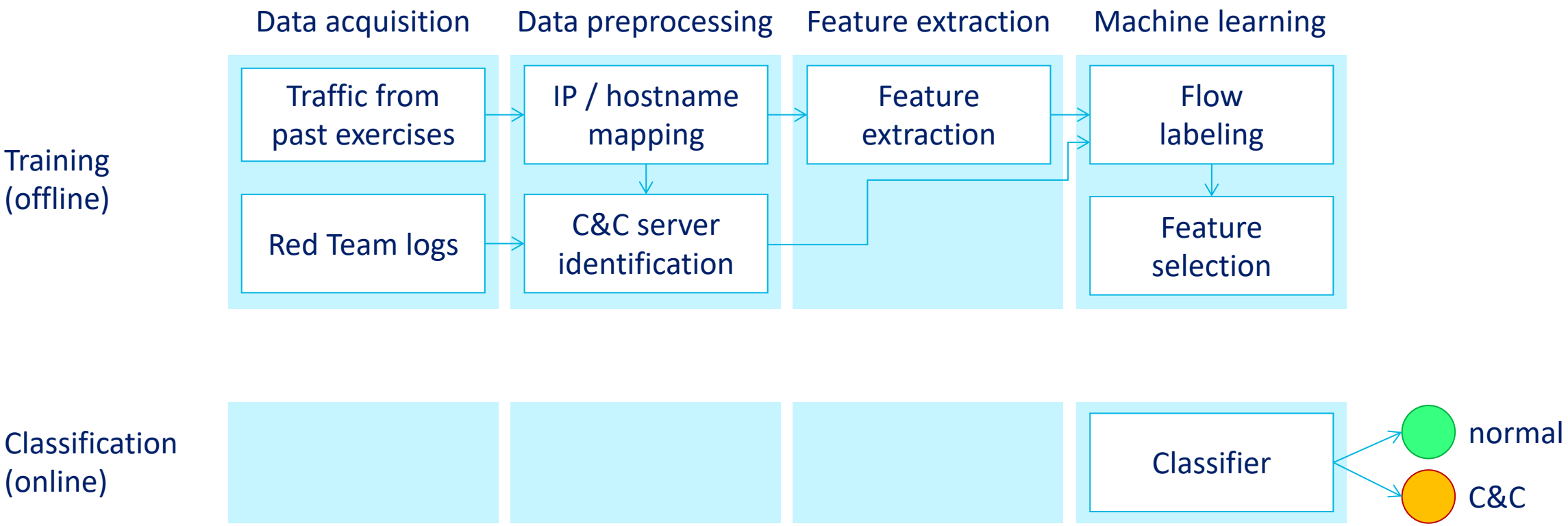
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Classification (online)
Classifier
normal
C&C

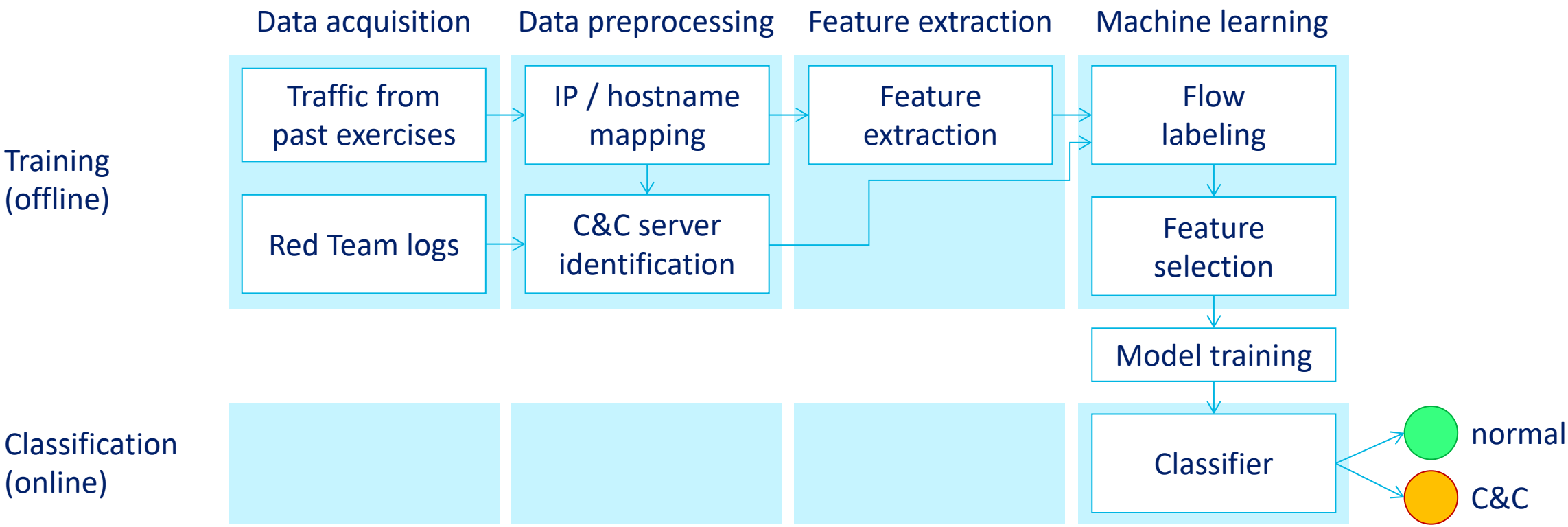
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Classification (online)
Classifier
normal
C&C

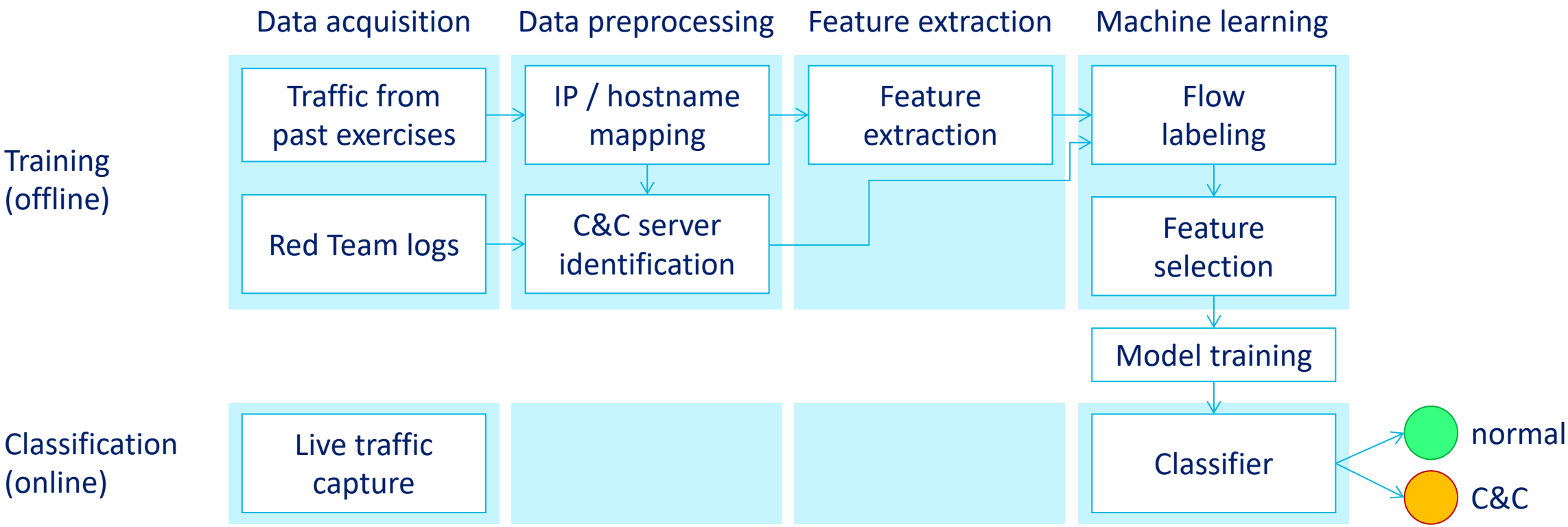
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Classification (online)
Classifier
normal
C&C

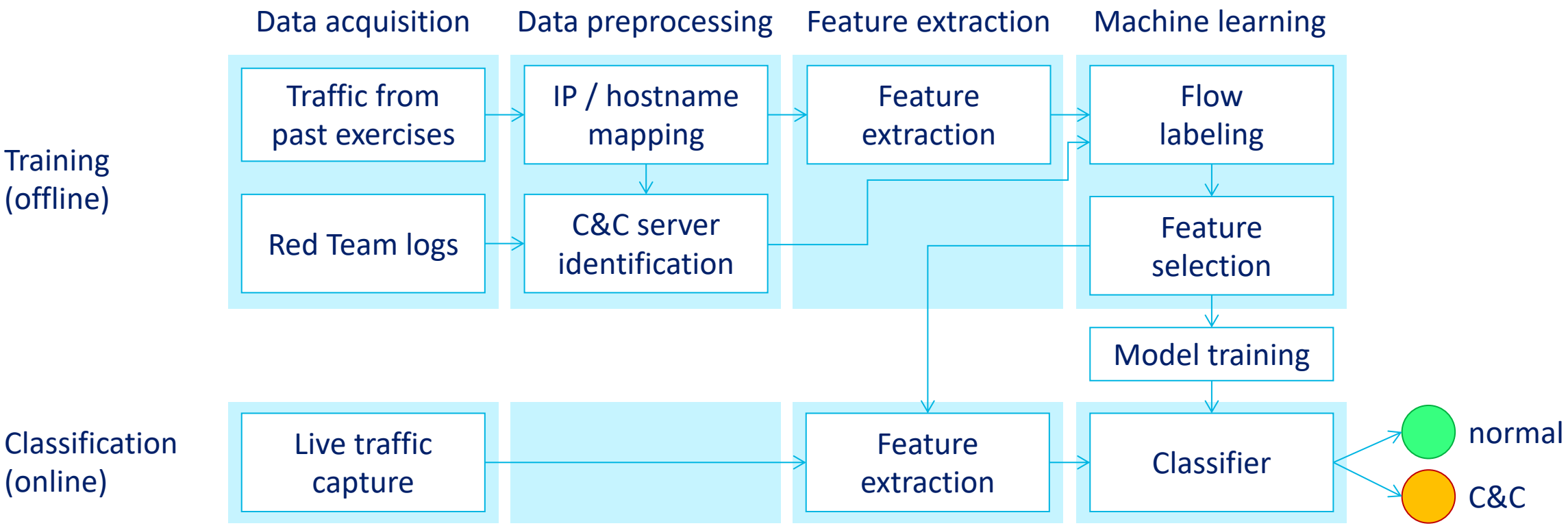
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Classifier
normal
C&C

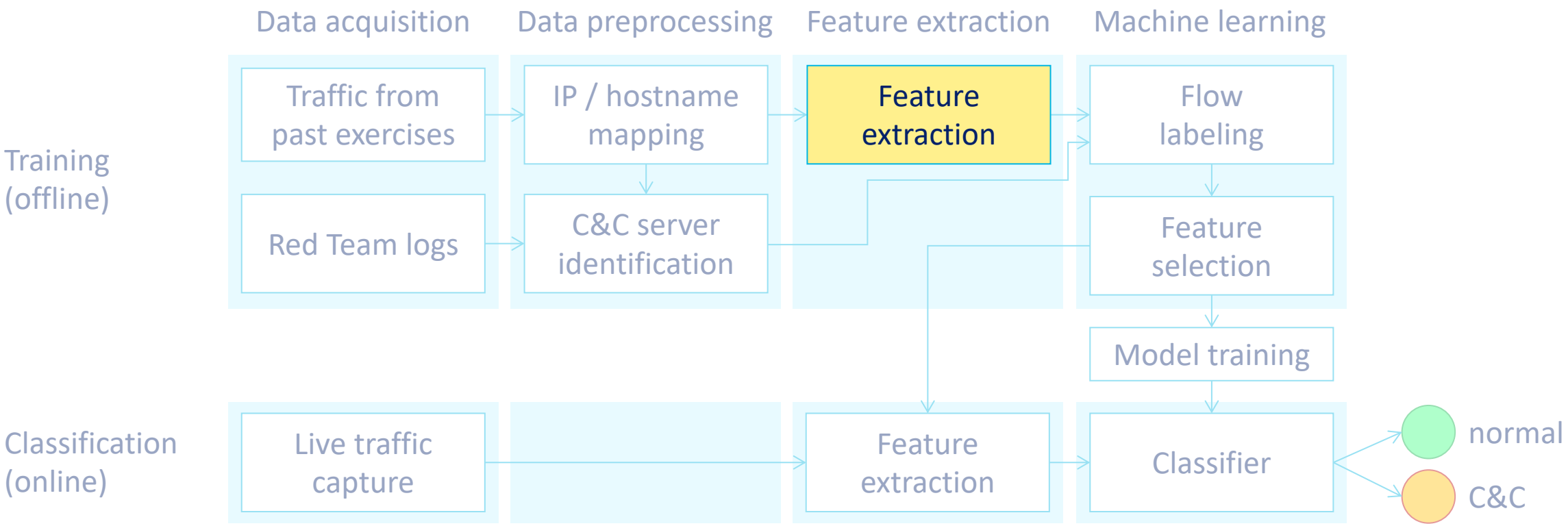
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Classifier
normal
C&C

CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

CYCON

We consider 77 widely used network traffic features

Metadata

- Flow direction
- L3/L4 protocol
- Internal / external
- …

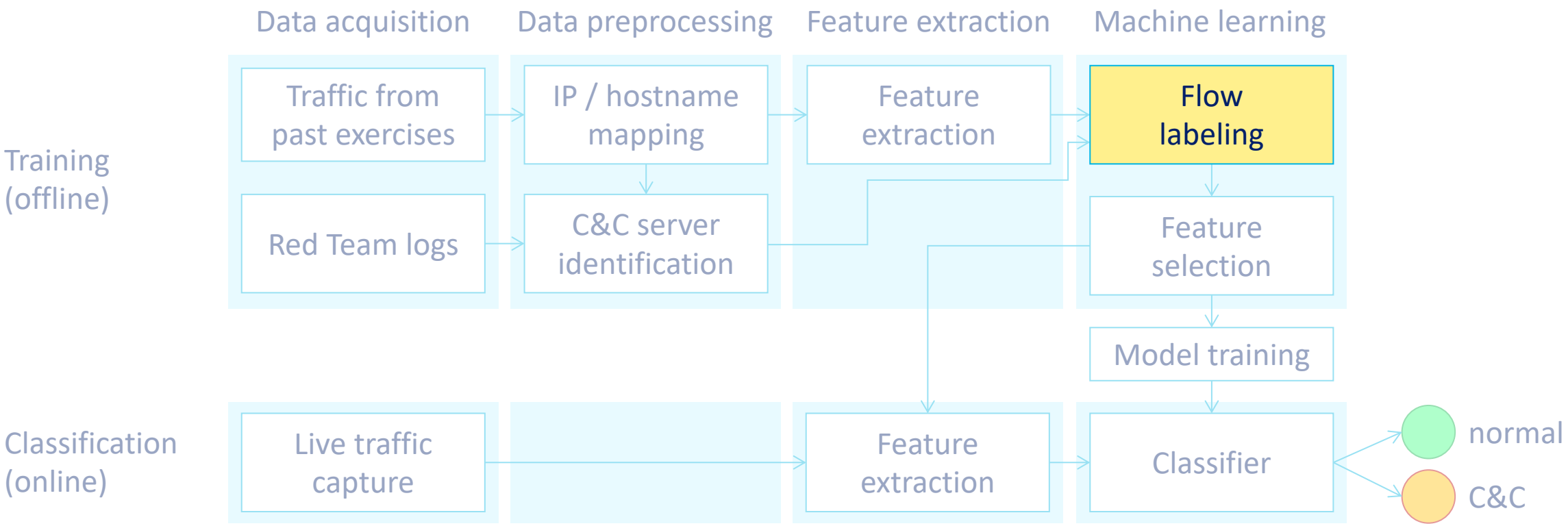
Time-related

- Flow duration
- Packets / s
- Inter arrival time
- …

Volume-related

- Number of packets
- Bytes / s
- Packet size
- …

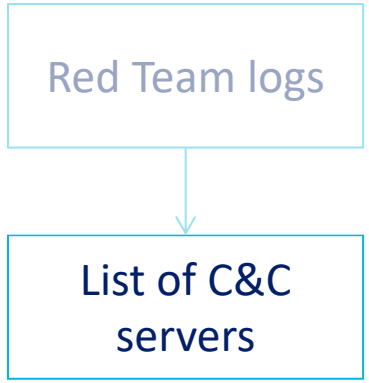
CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

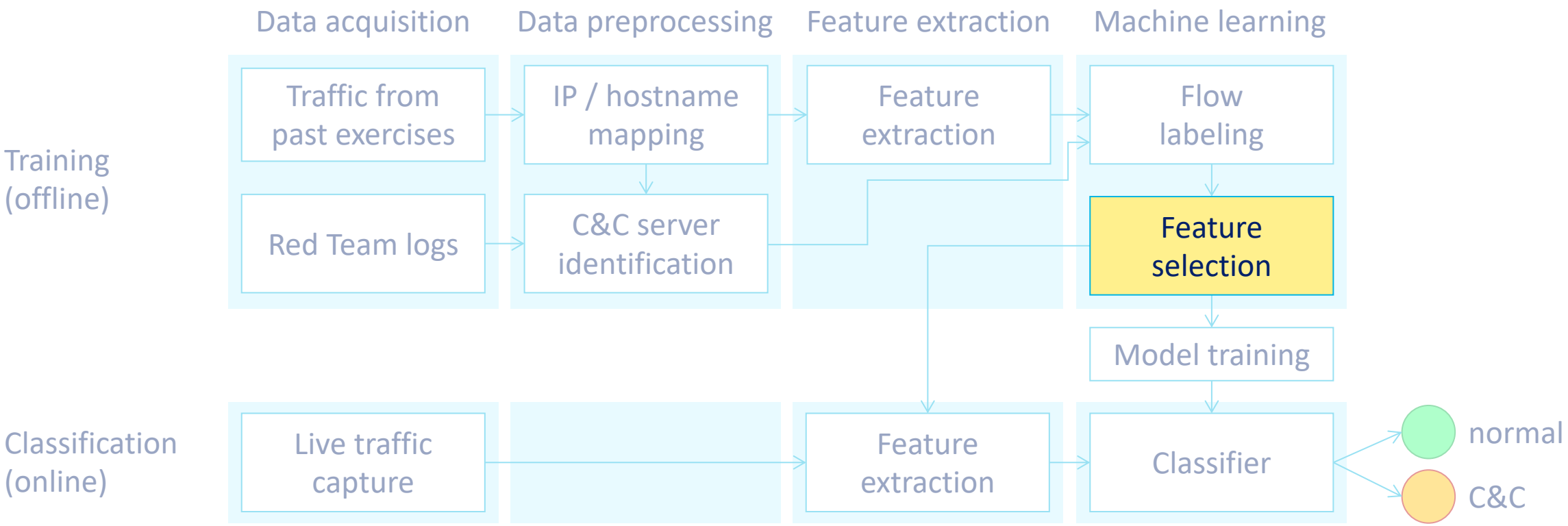
CYCON

We label all flows from or to a C&C server as C&C traffic

```
For each flow:
    If (source or destination ∈ List of C&C servers)
      Then: flow is C&C
      Else: flow is normal
    End If
```



CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

CYCON

We select the most useful network traffic features

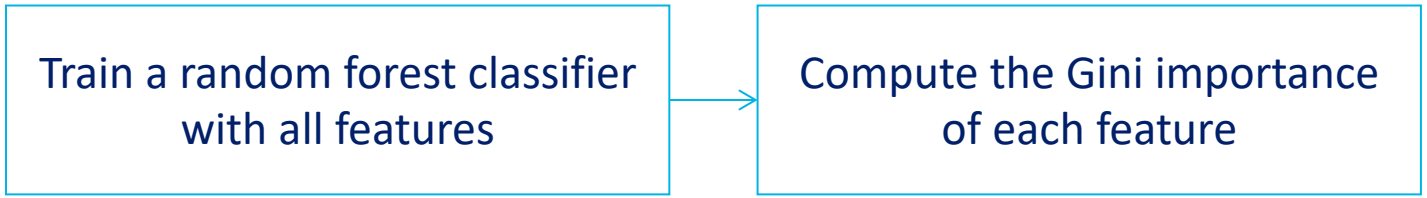
CYCON

We select the most useful network traffic features

Train a random forest classifier with all features

CYCON

We select the most useful network traffic features

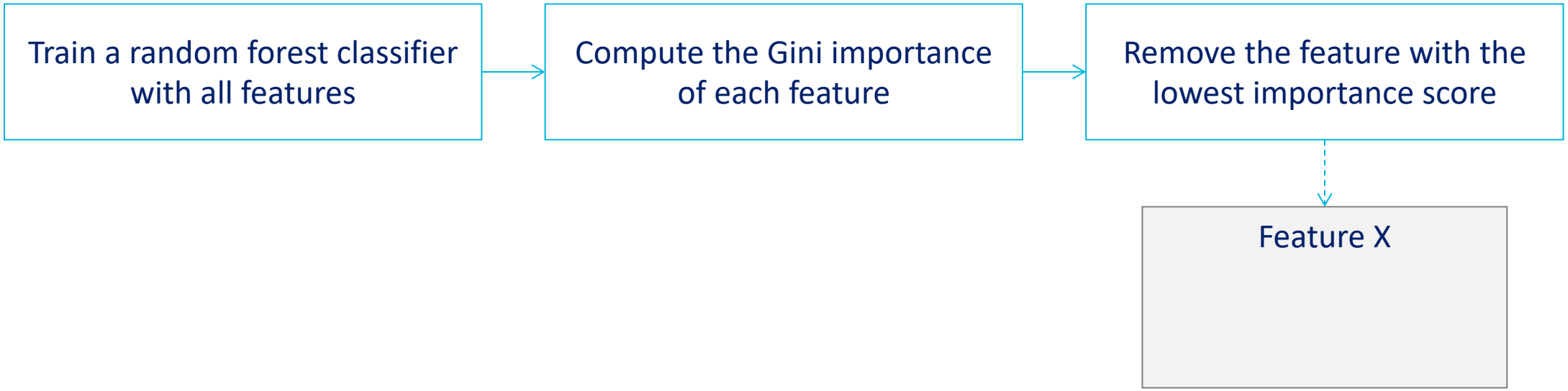


CYCON

We select the most useful network traffic features

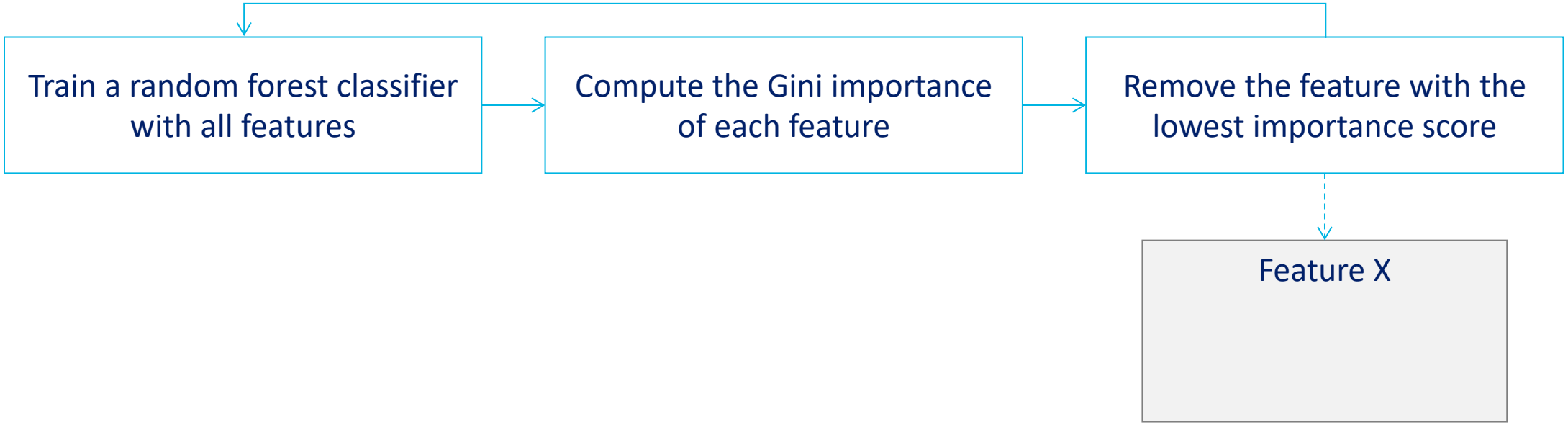
CYCON

We select the most useful network traffic features



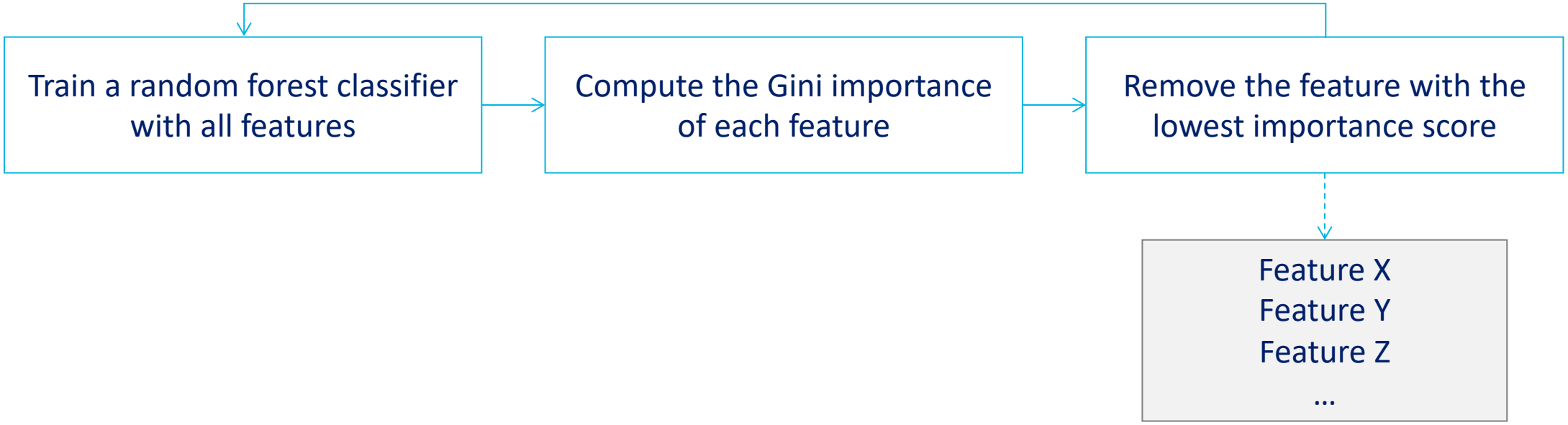
CYCON

We select the most useful network traffic features



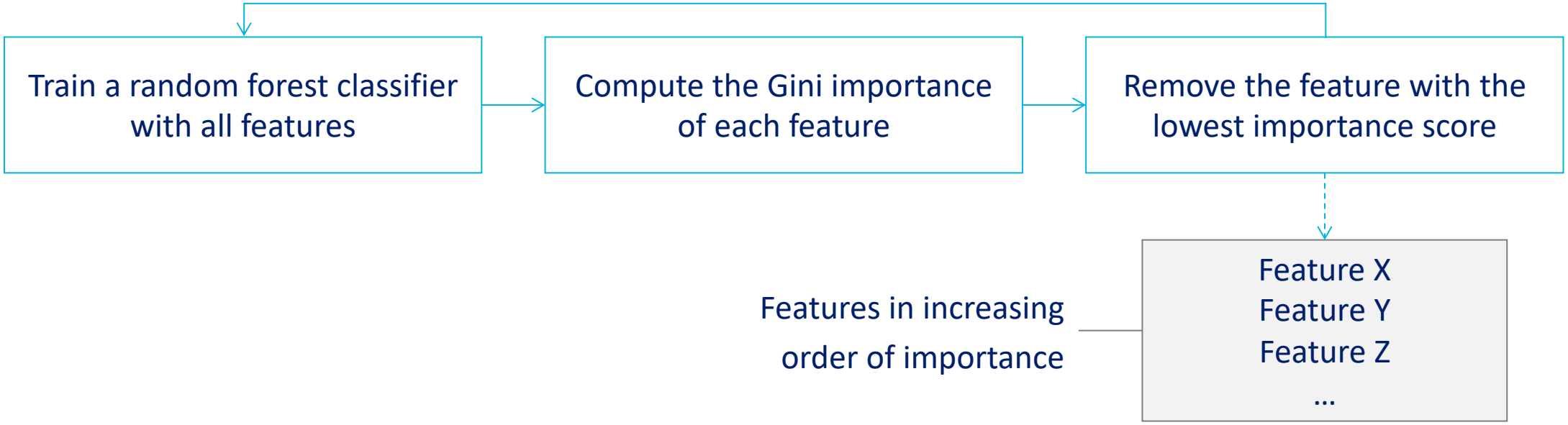
CYCON

We select the most useful network traffic features

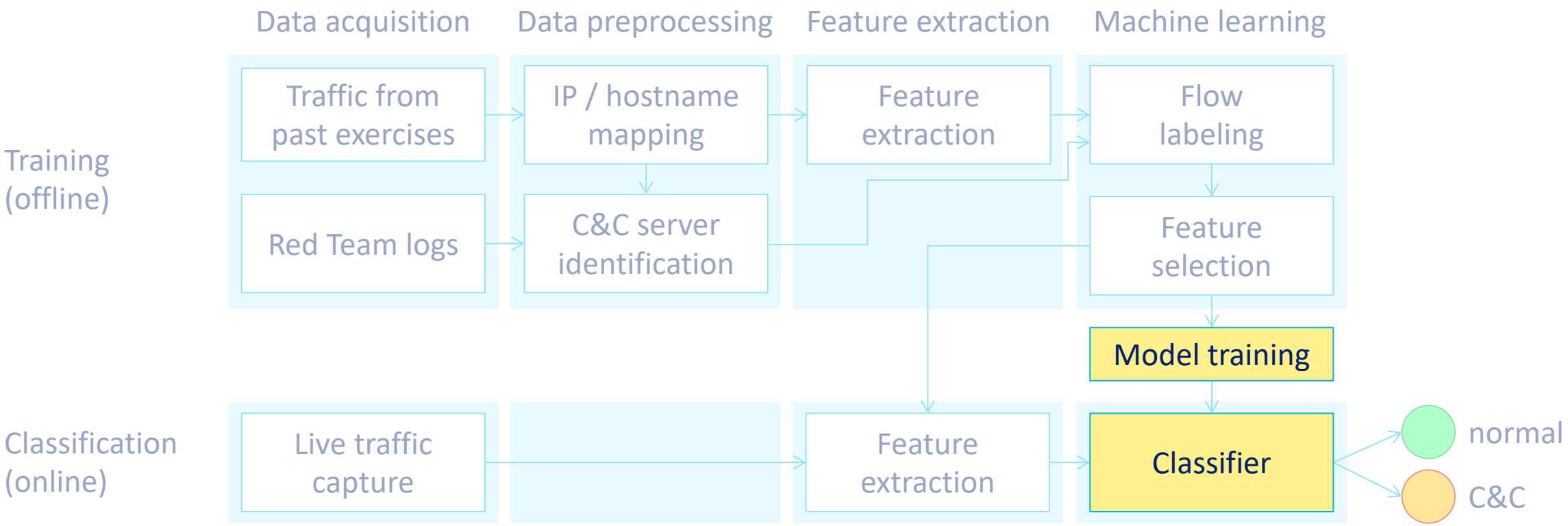


CYCON

We select the most useful network traffic features



CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

CYCON

Random Forest models achieve good results in an efficient matter and in little time

Challenges:

- “normal” traffic is different in each year
- Resources (CPU, Memory) are constrained
- Want (near) real-time classification

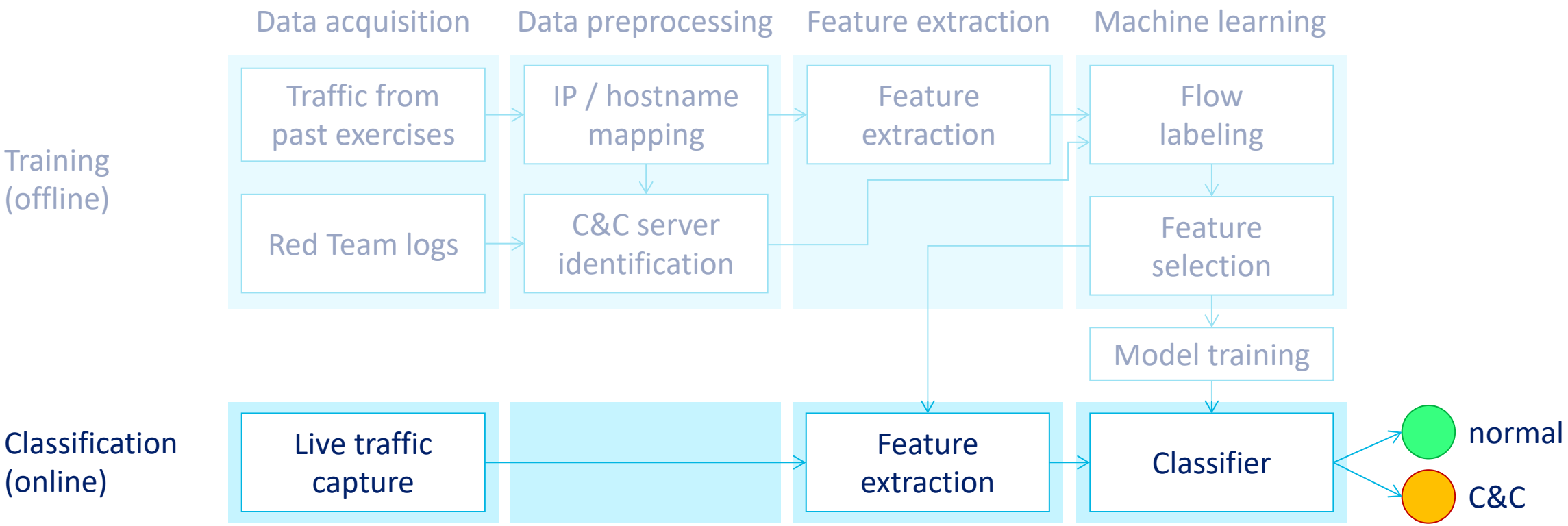
CYCON

Random Forest models achieve good results in an efficient matter and in little time

Challenges:

- “normal” traffic is different in each year
- Resources (CPU, Memory) are constrained
- Want (near) real-time classification
- We found that random forest models performed best while satisfying all constraints

CYCON


Data acquisition
Data preprocessing
Feature extraction
Machine learning
Training (offline)
Traffic from past exercises
Red Team logs
IP / hostname mapping
C&C server identification
Feature extraction
Flow labeling
Feature selection
Model training
Classification (online)
Live traffic capture
Feature extraction
Classifier
normal
C&C

CYCON

Locked Shields:
the largest cyber defense exercise

Identifying C&C channels
in real time and
with little resources

Evaluation
on real data from Locked Shields 2017 and 2018

CYCON

We evaluate 2 models

	Training	Testing
LS17 model	LS17	LS18
LS18 model	LS18	LS17

CYCON

We evaluate 2 models

	Training	Testing	# Trees	Max. depth	Features
LS17 model	LS17	LS18	128	10	20
LS18 model	LS18	LS17	128	10	20

CYCON

High precision and recall for identifying C&C channels

	Training	Testing	Precision	Recall
LS17 model	LS17	LS18		
LS18 model	LS18	LS17		

CYCON

High precision and recall for identifying C&C channels

	Precision	Recall
	# correctly classified C&C flows / # C&C flows classified as C&C	# correctly classified C&C flows / # existing C&C flows

CYCON

High precision and recall for identifying C&C channels

	Training	Testing	Precision	Recall
LS17 model	LS17	LS18	99 %	98 %
LS18 model	LS18	LS17	99 %	90 %

CYCON

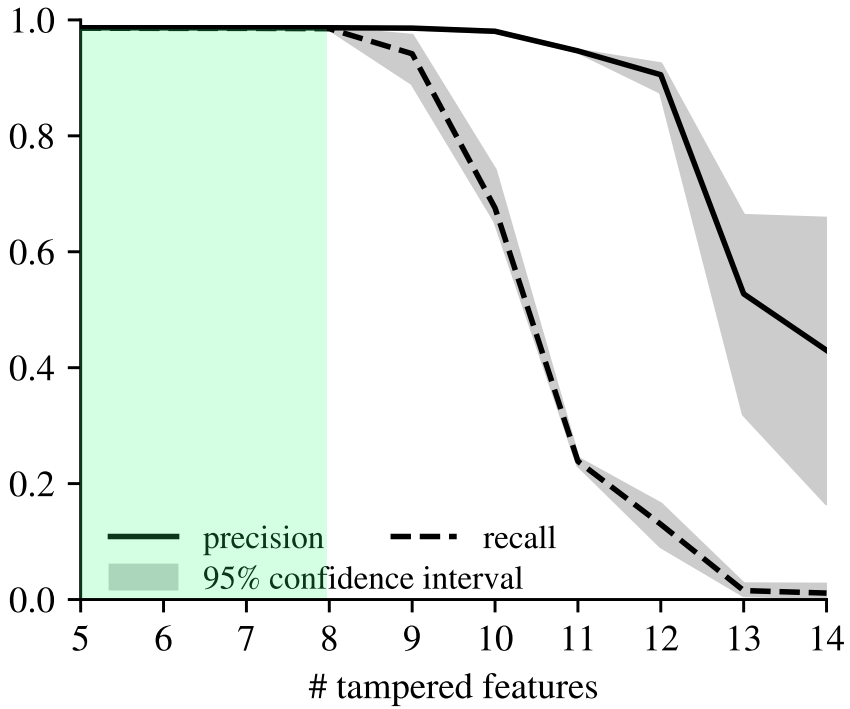
Classification of one flow only takes microseconds

	Training	Testing	Classification
LS17 model	LS17	LS18	3.1 µs
LS18 model	LS18	LS17	3.3 µs

Per flow

CYCON

Our model is robust against tampering with up to 8 features



CYCON

The Swiss Blue Team did use our system at Locked Shields 2019, and it helped them discovering C&C channels which they would not have seen otherwise.

CYCON

The Swiss Blue Team used our system successfully in Locked Shields 2019

Changes:

- Models from LS17 and LS18
- Same features (but different feature extraction tool)

CYCON

The Swiss Blue Team used our system successfully in Locked Shields 2019

Changes:

- Models from LS17 and LS18
- Same features (but different feature extraction tool)

Observations:

- Very high true positive rate
- Detected unknown C&C servers
- Confirmed known C&C servers

CYCON

Locked Shields:
the largest cyber defense exercise

Identifying C&C channels
in real time and
with little resources

Evaluation
on real data from Locked Shields 2017 and 2018

CYCON

Machine Learning-based Detection of C&C Channels with a Focus on Locked Shields

Locked Shields:
the largest cyber defense exercise

Identifying C&C channels
in real time and
with little resources

Evaluation
on real data from Locked Shields 2017 and 2018

Nicolas Känzig*, Roland Meier*, Luca Gambazzi*, Vincent Lenders*, Laurent Vanbever
(* attending CyCon)

We thank the Swiss Blue Team for sharing their data and expertise with us
and for their constant support throughout this project.

CYCON