# DGA Bot Detection with Time Series Decision Trees

**Anaël Bonneton**
Daniel Migault, Stephane Senecal, Nizar Kheir

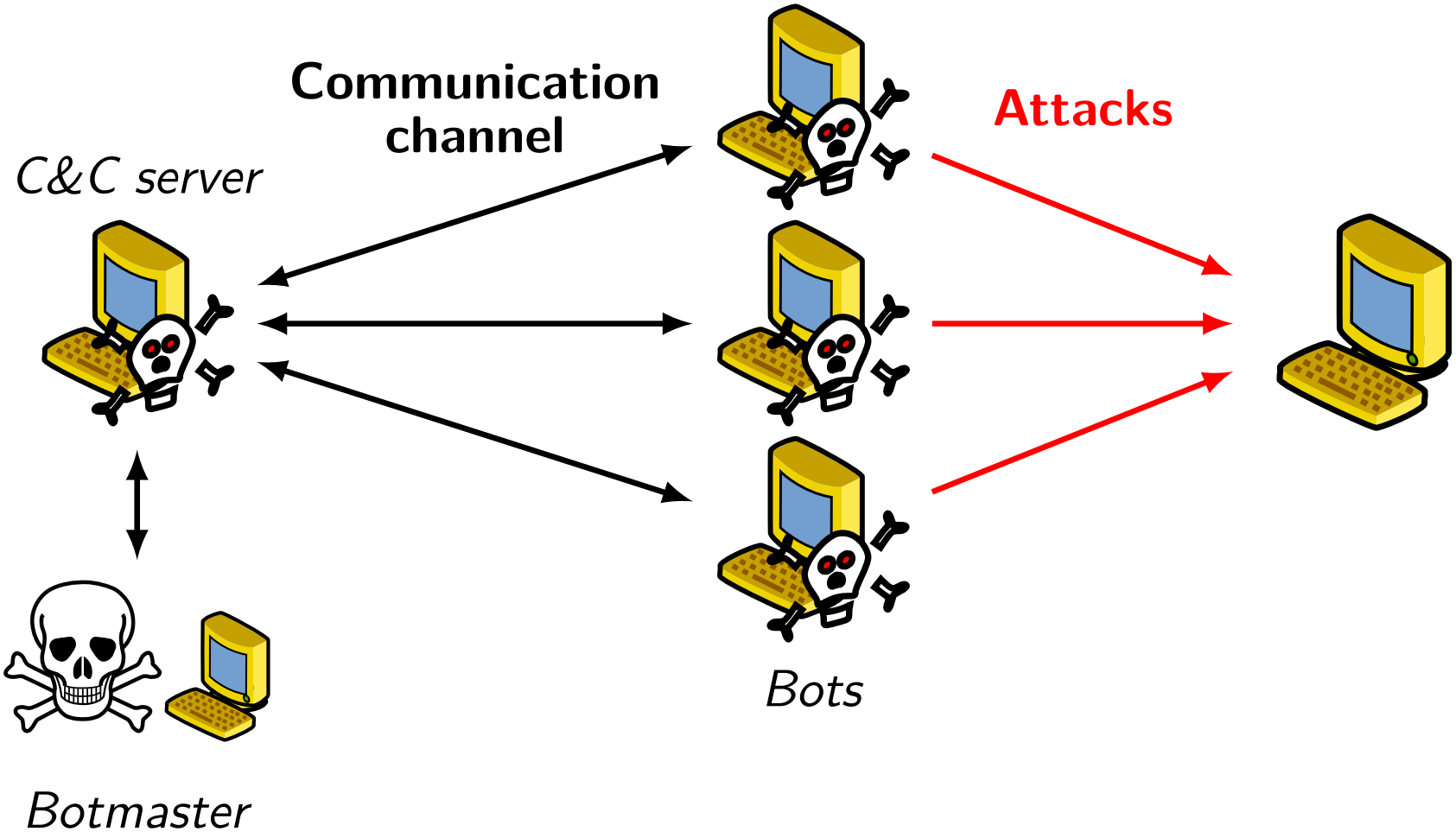
BADGERS Workshop - November, 5th 2015



Research works conducted at Orange Labs

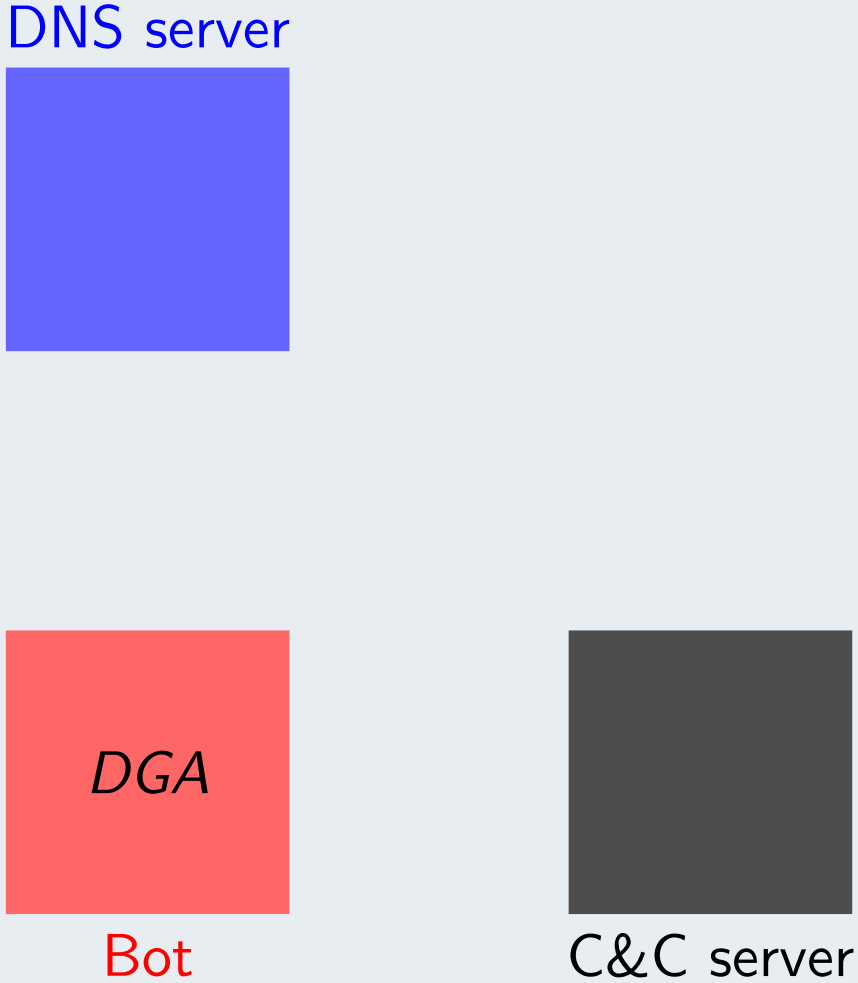
# Problem: Detecting DGA Bots

# Botnet



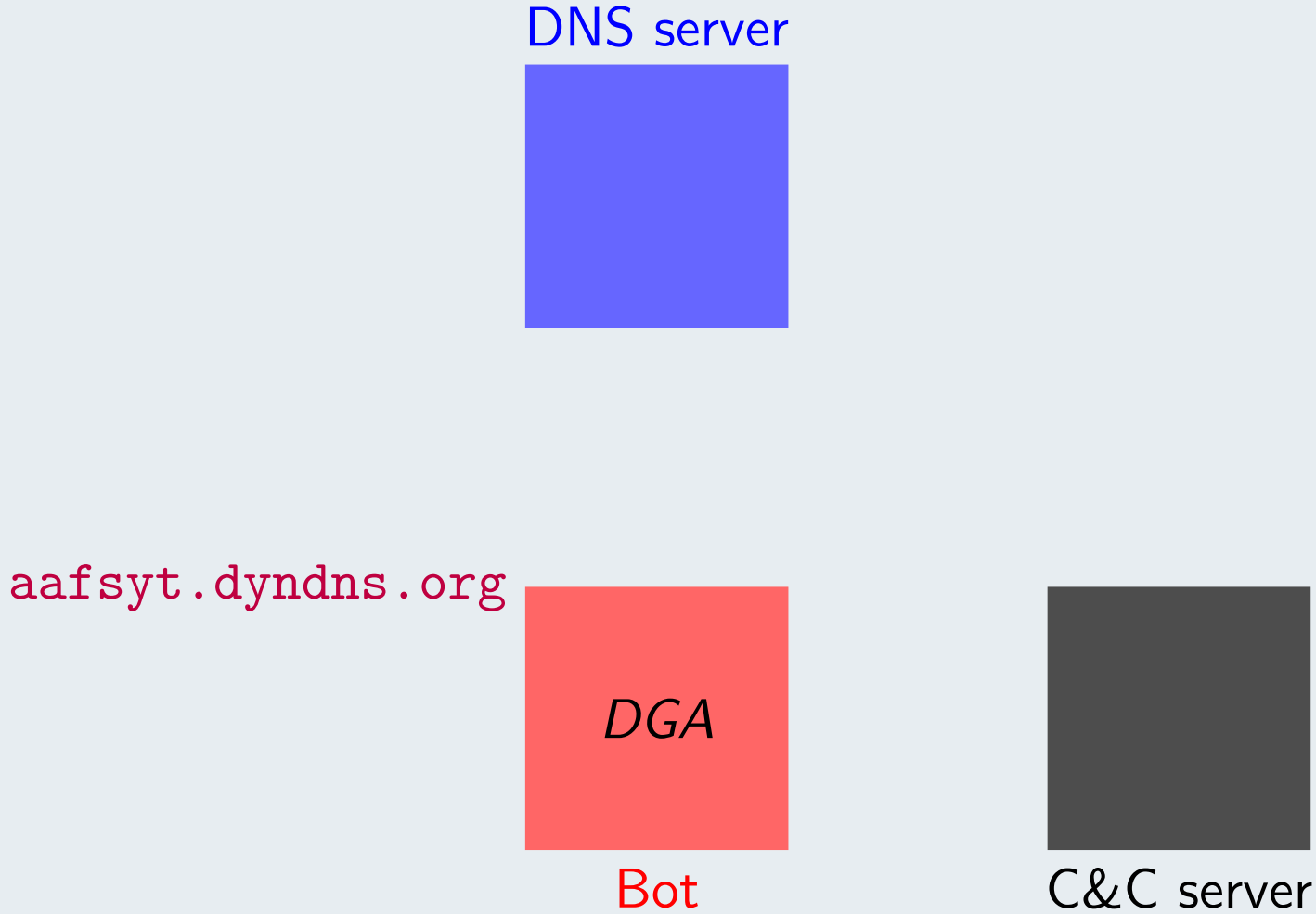
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



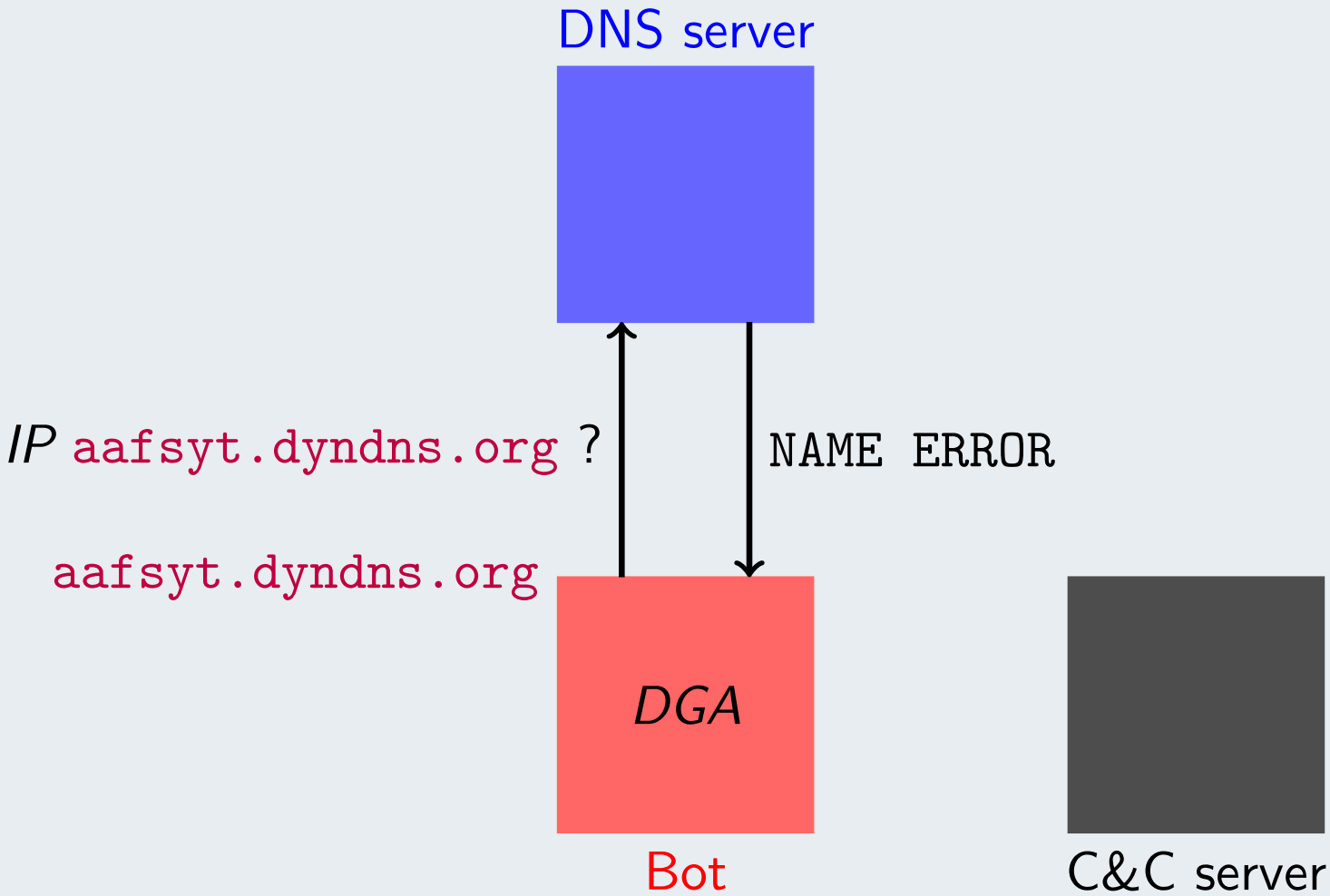
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



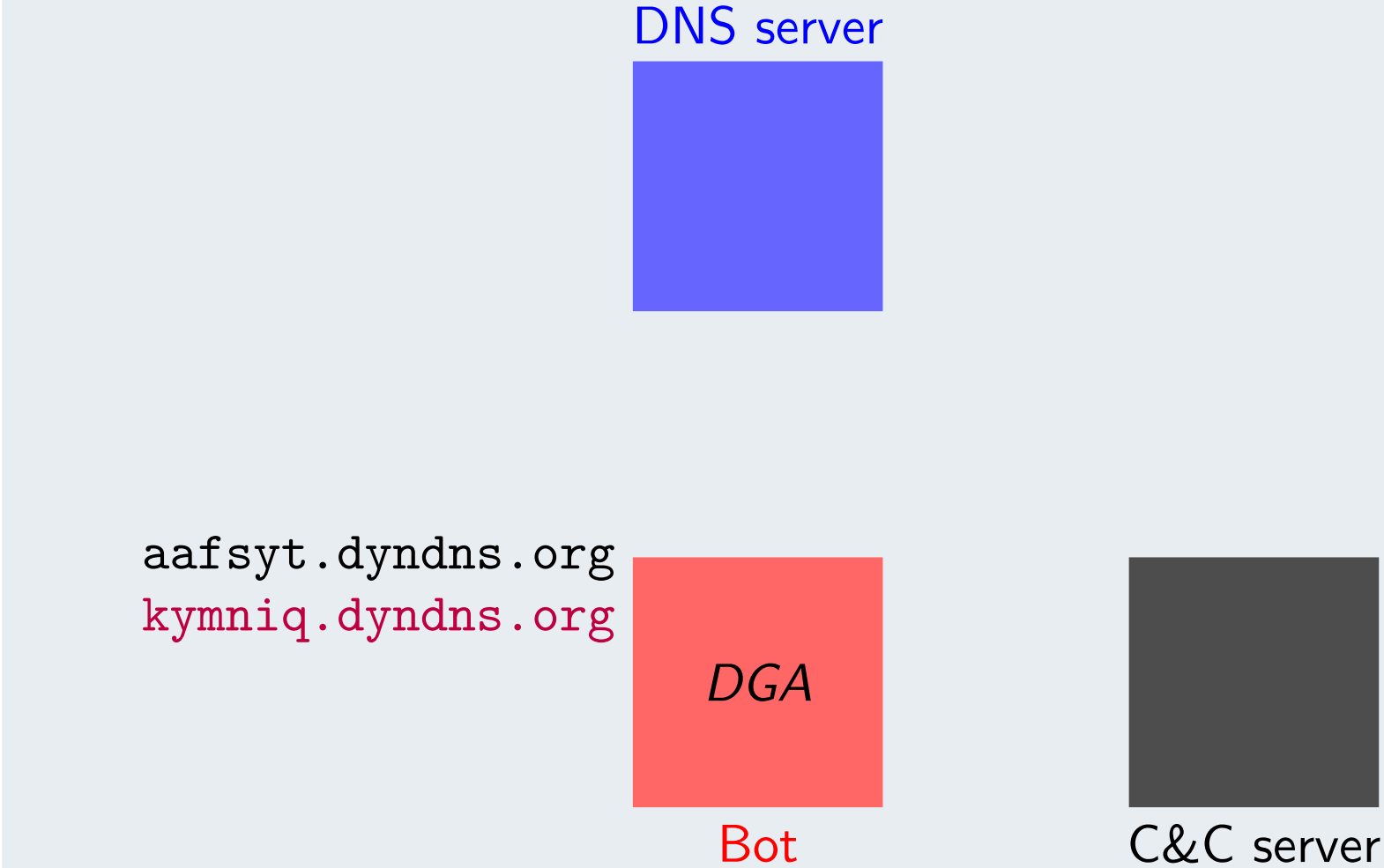
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



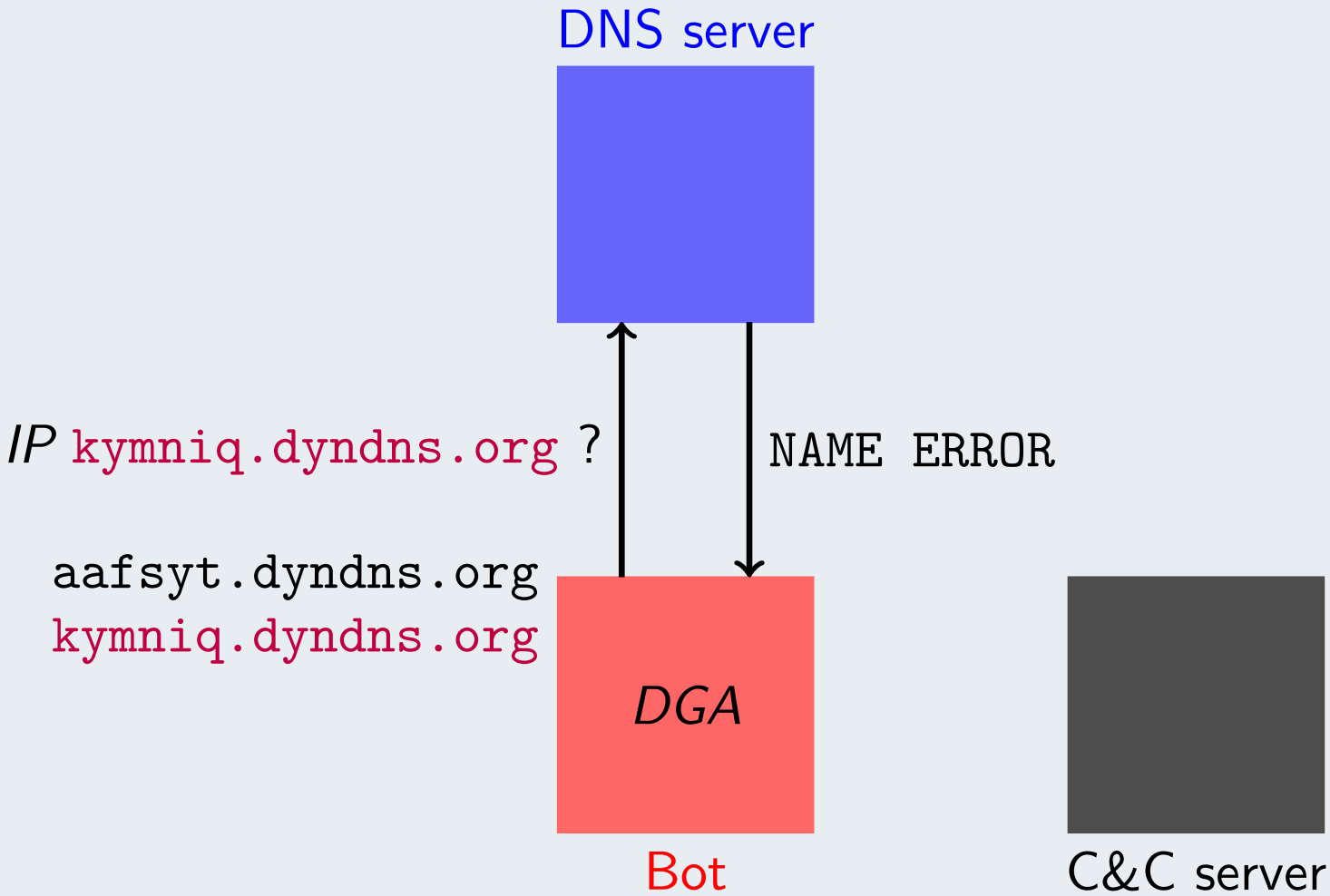
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



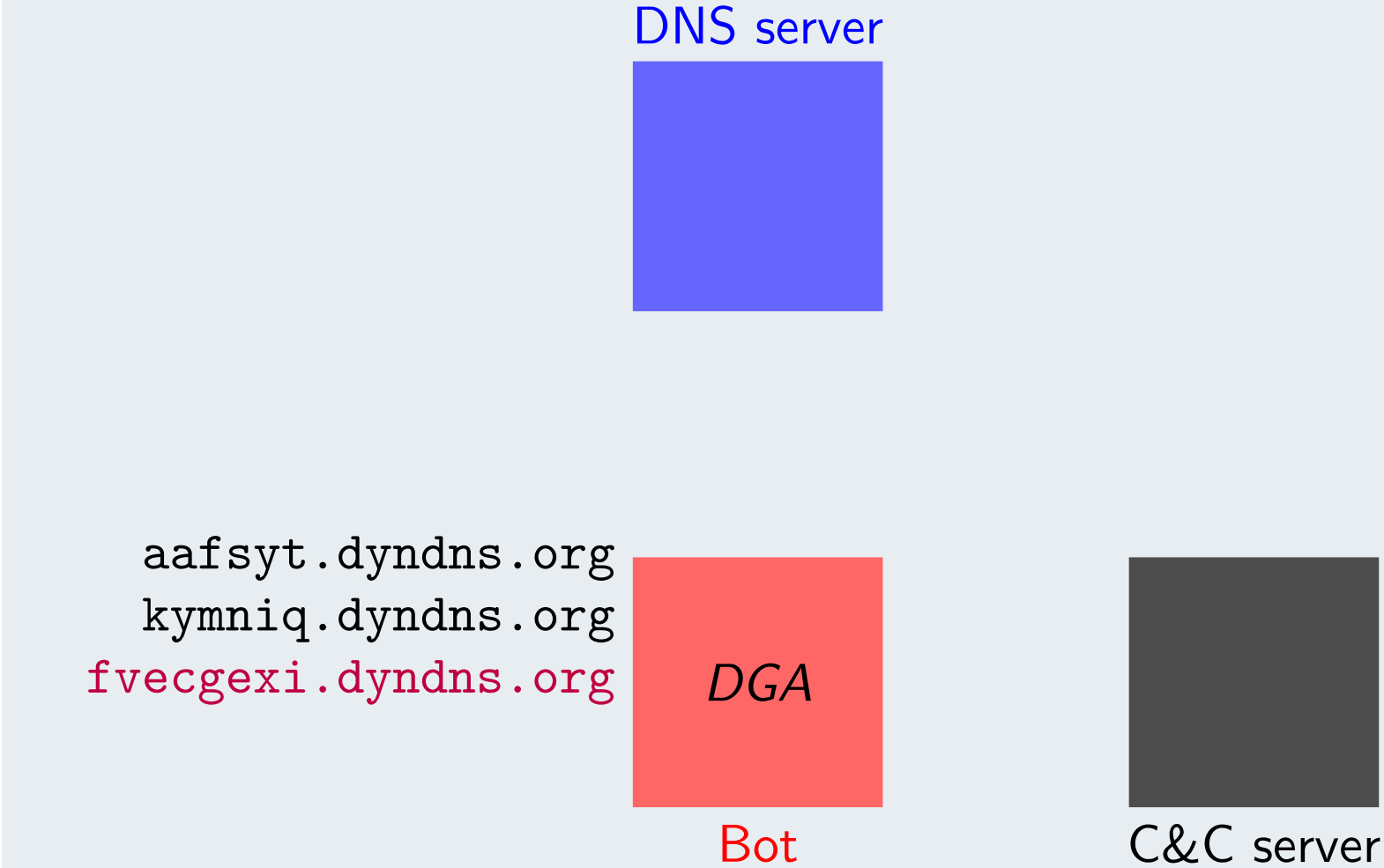
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



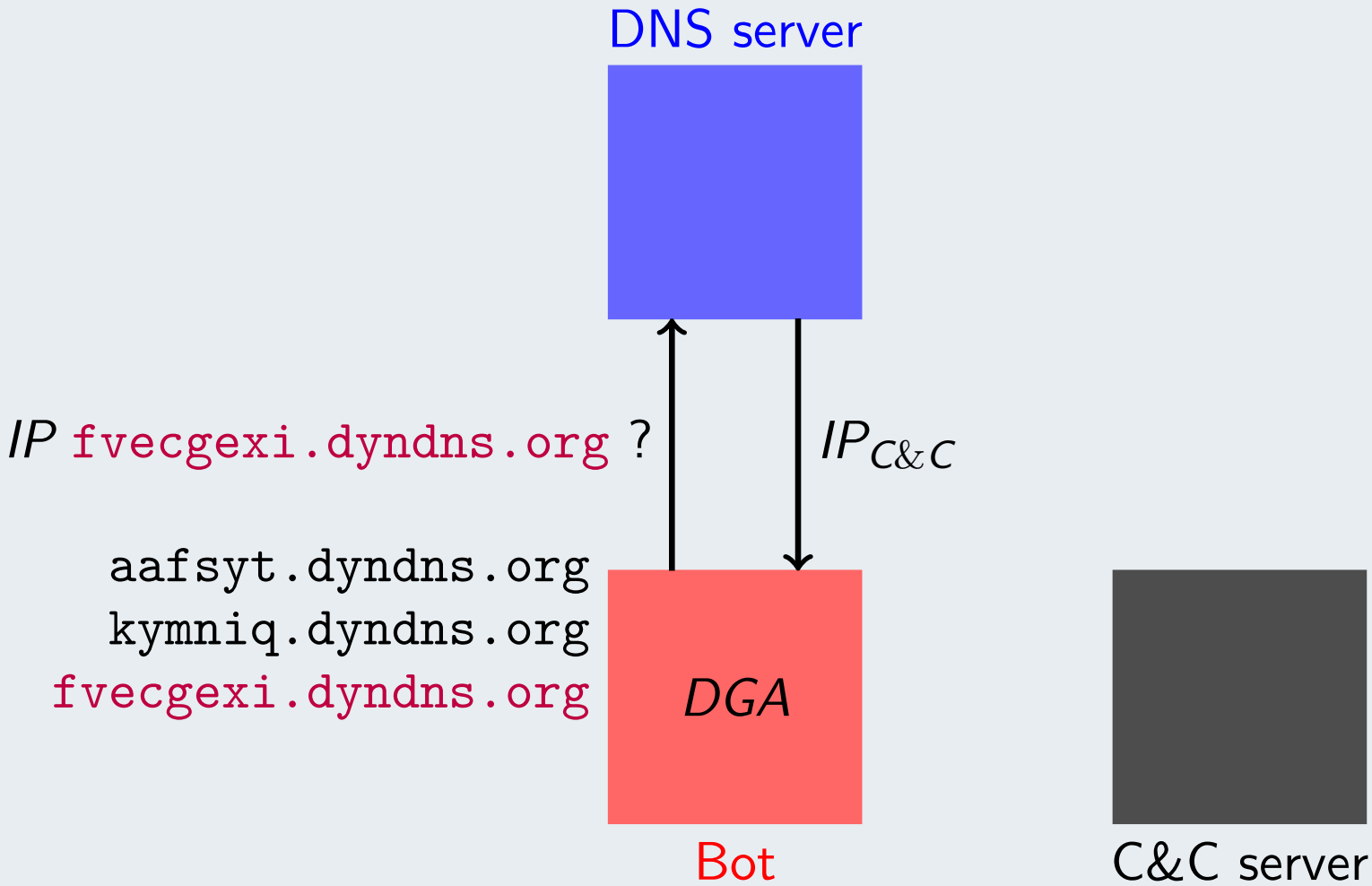
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



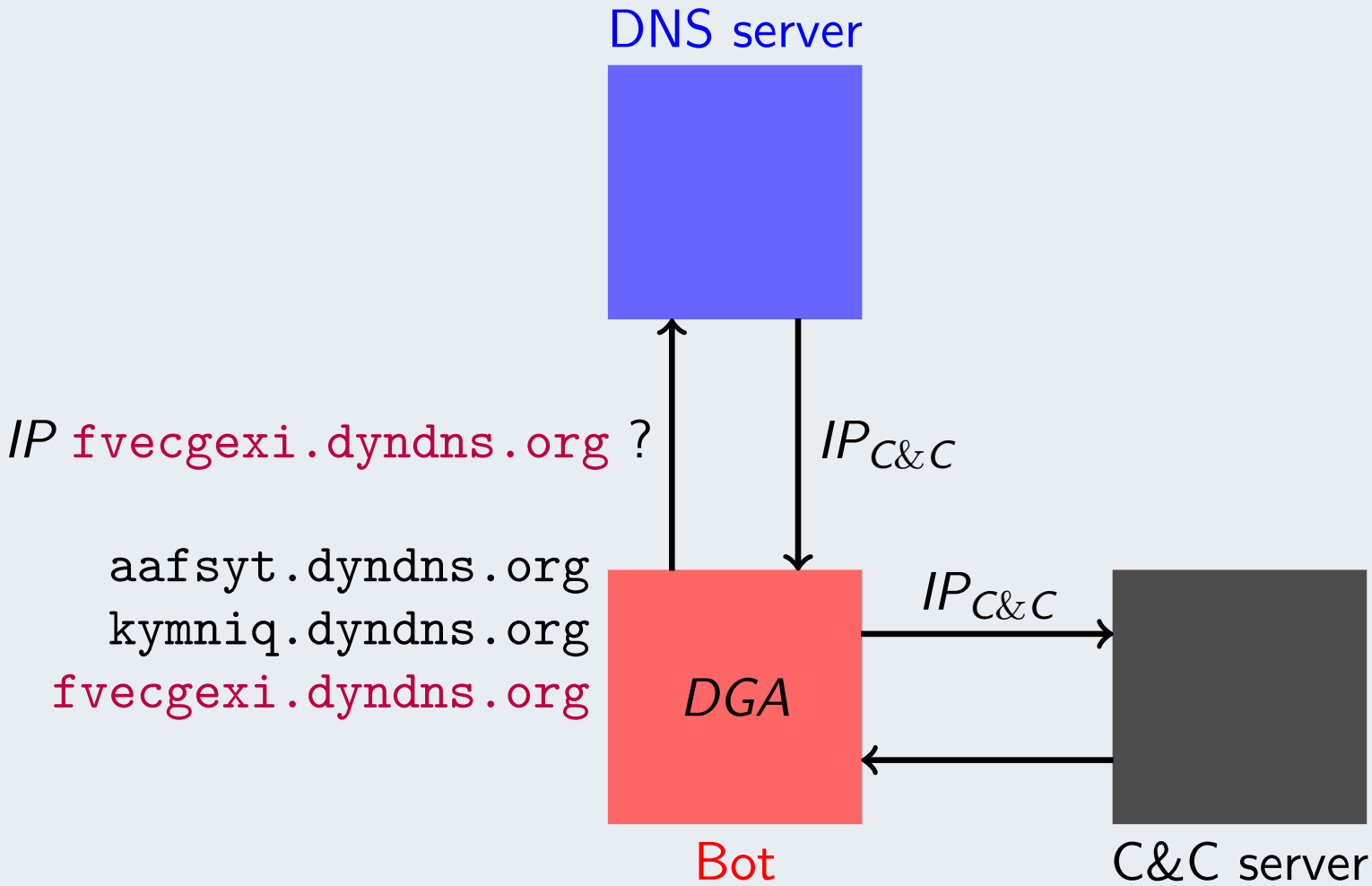
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm



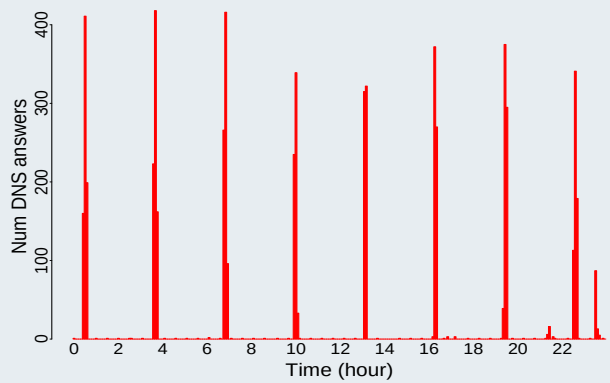
# Communication between bots and the C&C server

## DGA : Domain Generation Algorithm

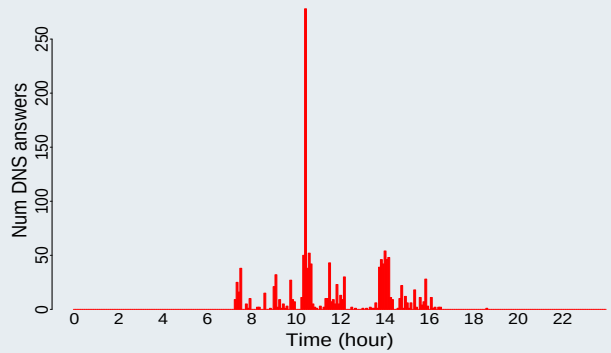


# Discriminating DNS Temporal Profiles

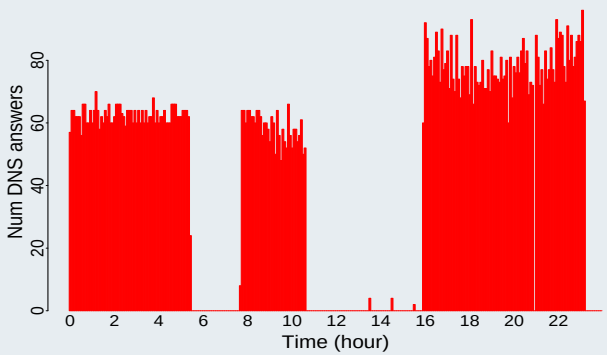
## Conficker A




## Conficker B

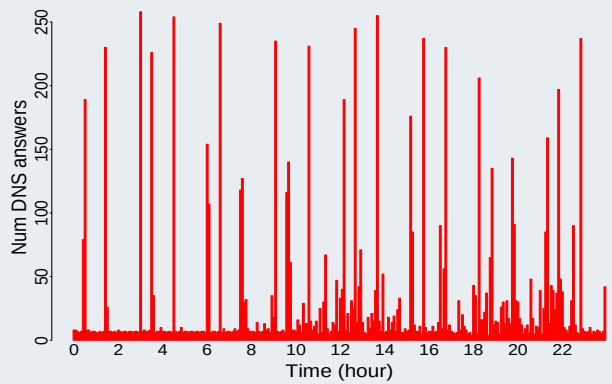



## Kraken

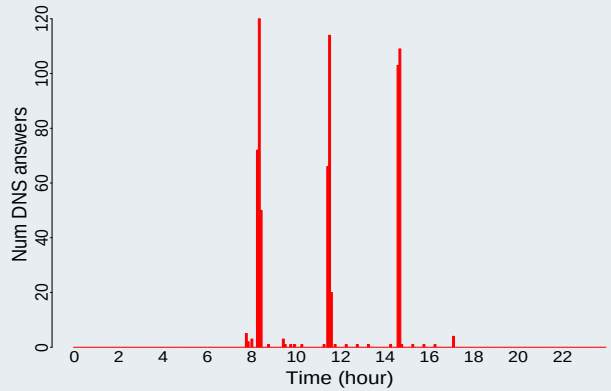


# Temporal Profiles for Conficker A

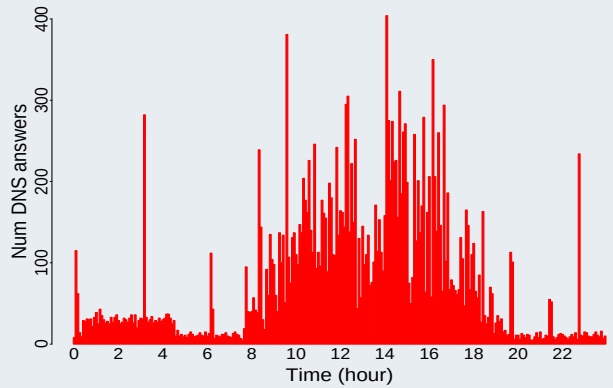
## Several Infected Devices




## Switched off Device

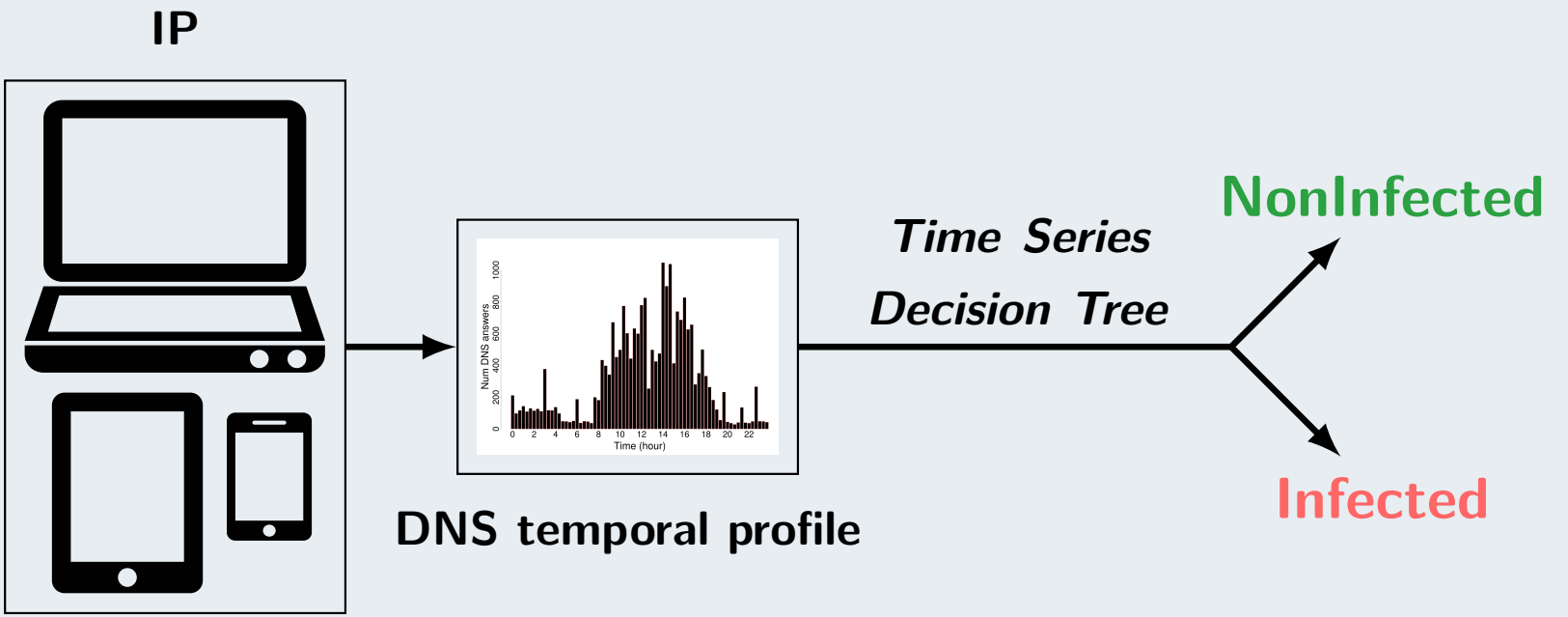



## Noisy Time Series



# Problem Statement

## Detecting Infected IPs



# Behavioral Detection Model: Time Series Decision Trees

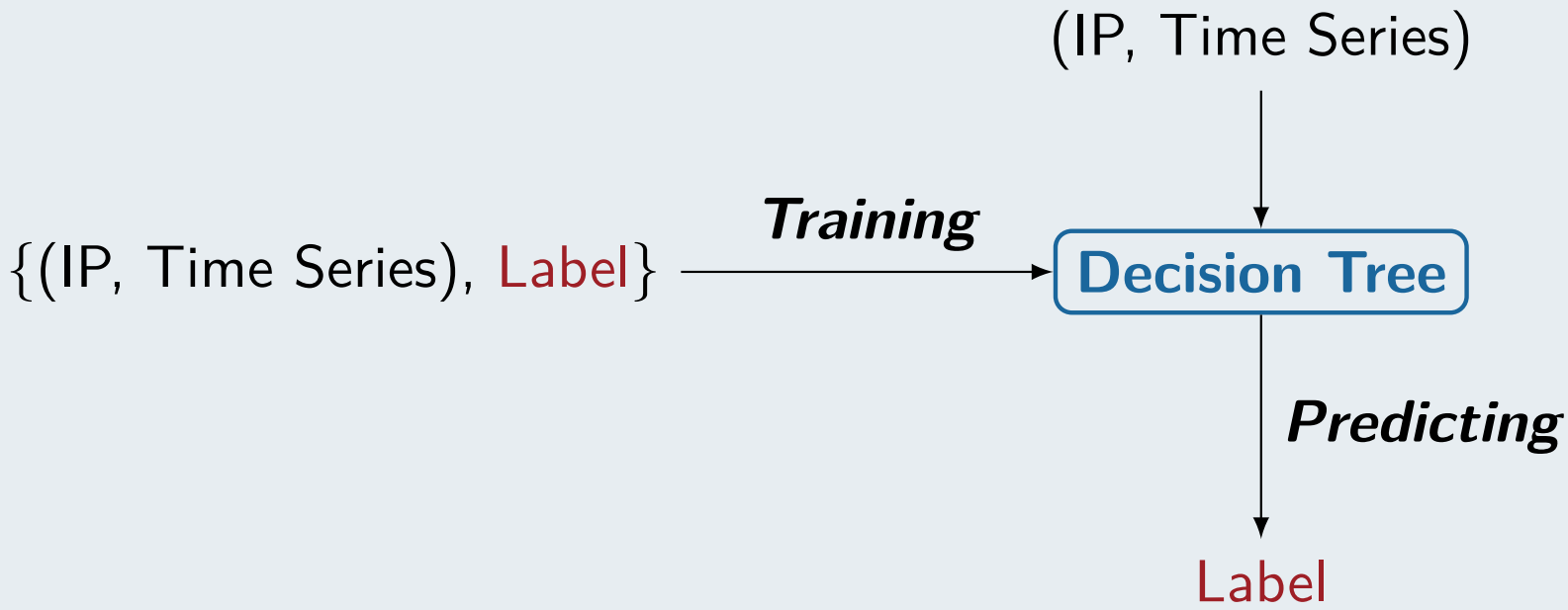
# Supervised Learning

## Training and Predicting

{(IP, Time Series), Label} —***Training***→ **Decision Tree**

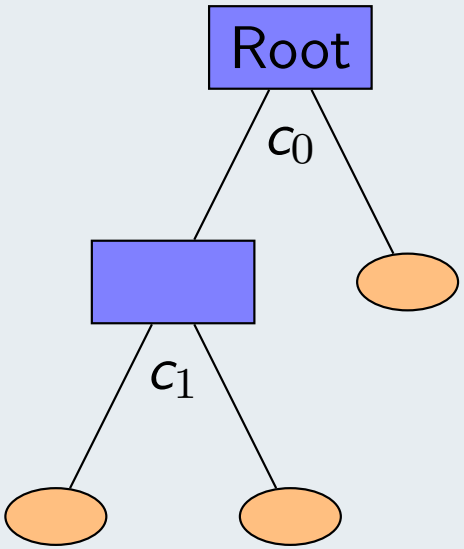
# Supervised Learning

## Training and Predicting



# Decision Tree

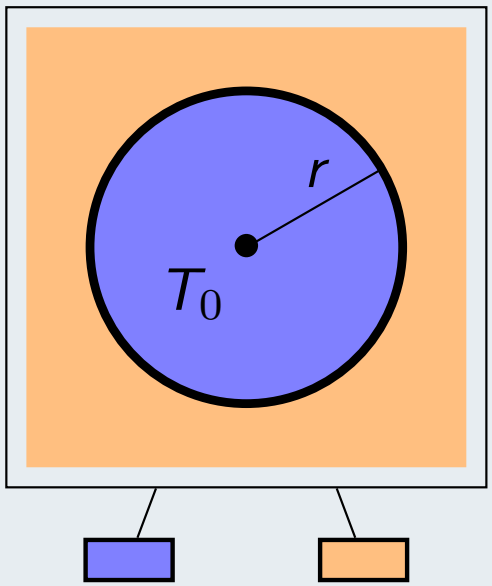
## Classifier based on a Binary Tree




- Recursive partition
  - Root = all the training data
  - Root = ∪ ∪
- Split conditions = decision rules

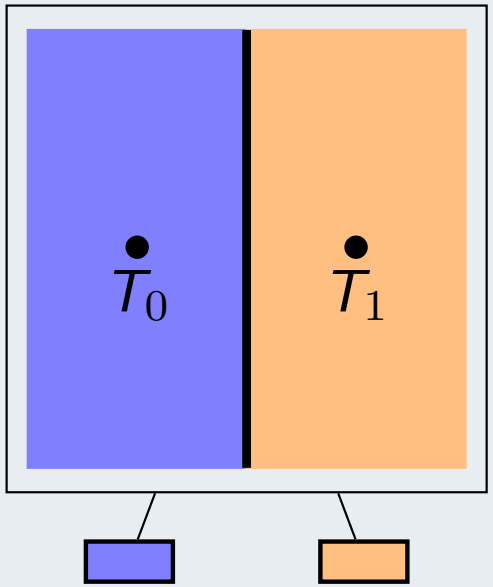
# Splits for Time Series

## Standard Split




$T_0$ a time series, $r \in \mathbb{R}_+^\star$

## Cluster Split

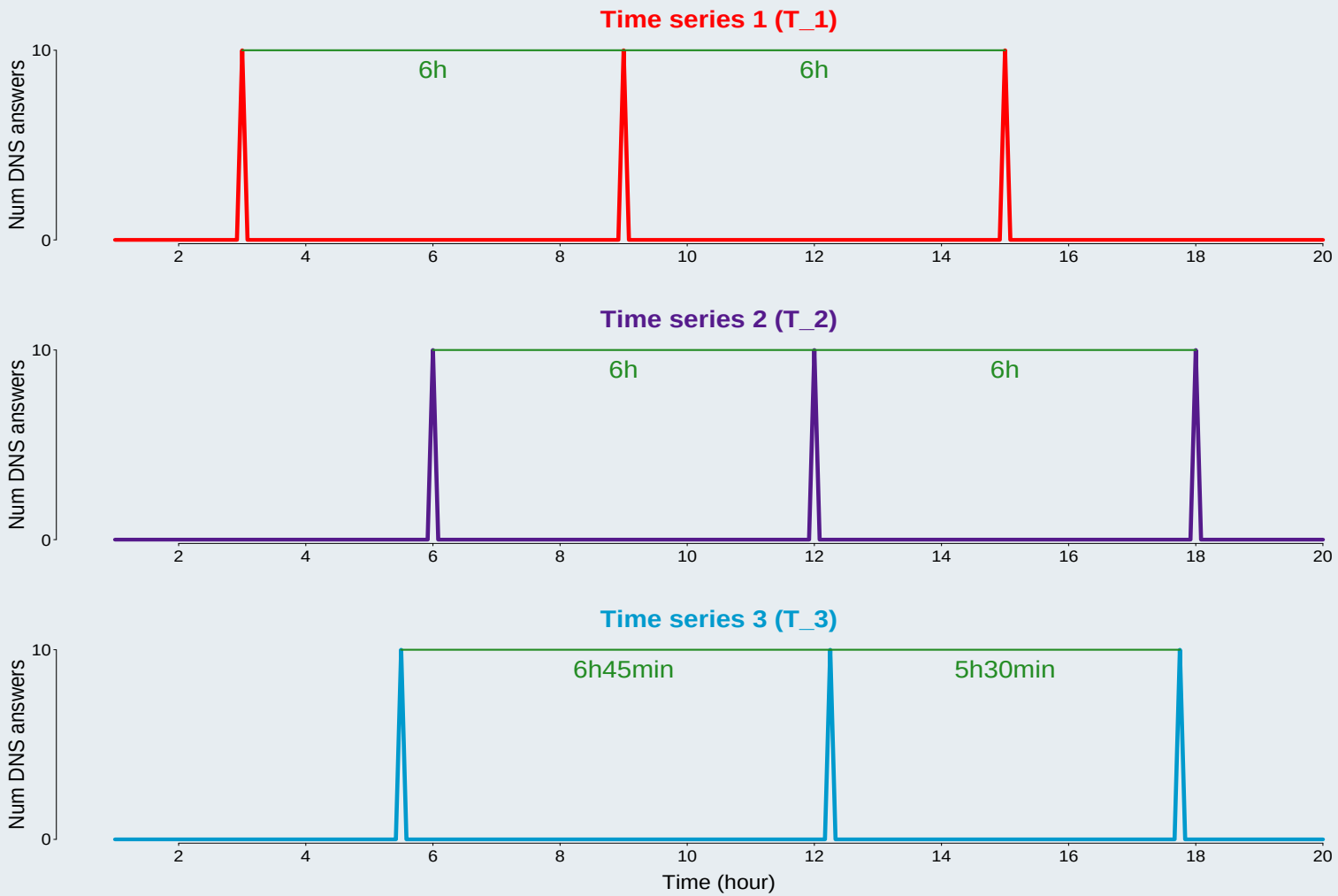



$T_0$ and $T_1$, 2 time series

Ref : Y. Yamada et al. "Decision-tree induction from time-series data based on a standard-example split test", in ICML 2003

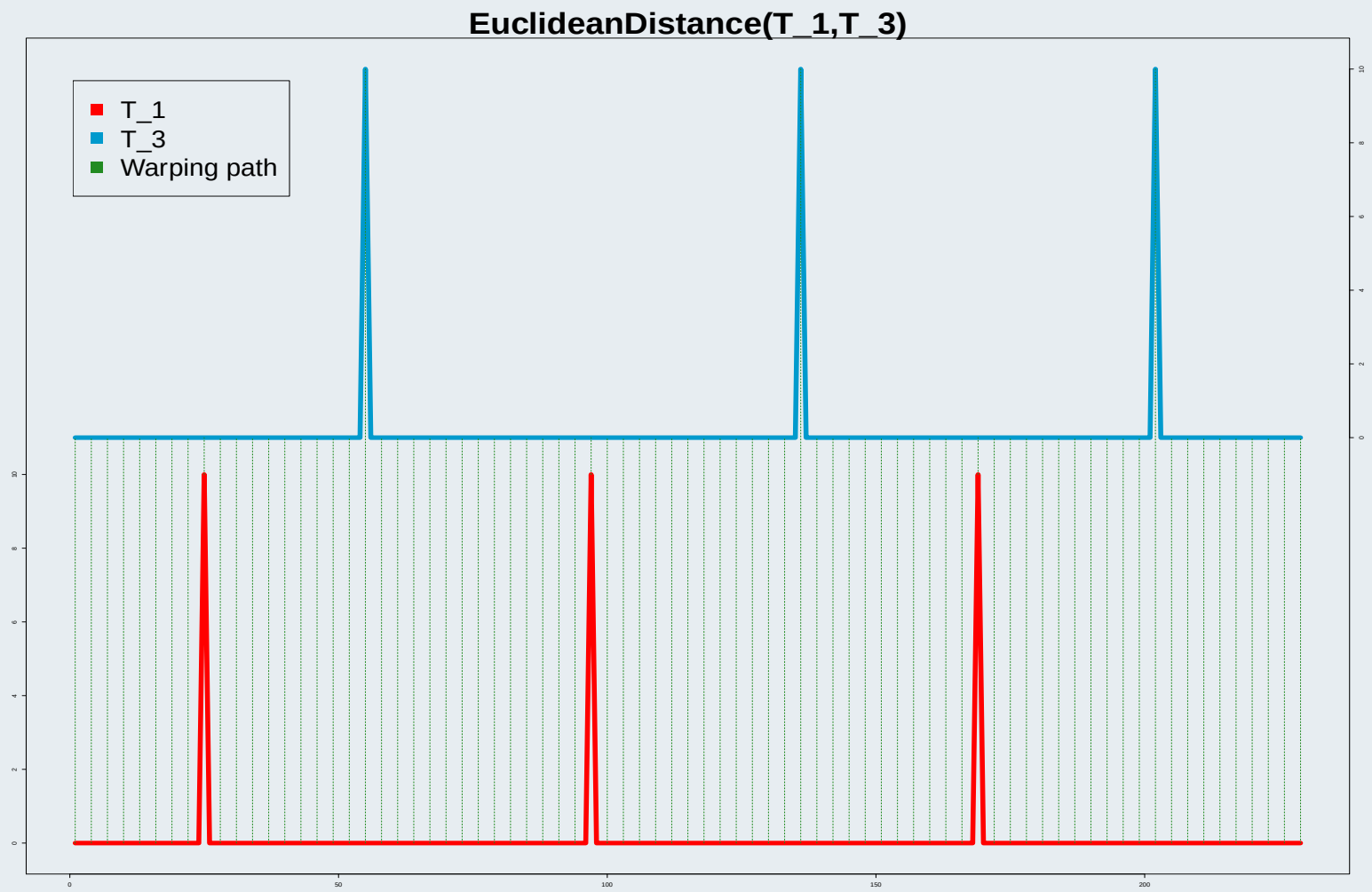
# Distance between two Time Series

## Similar Time Series



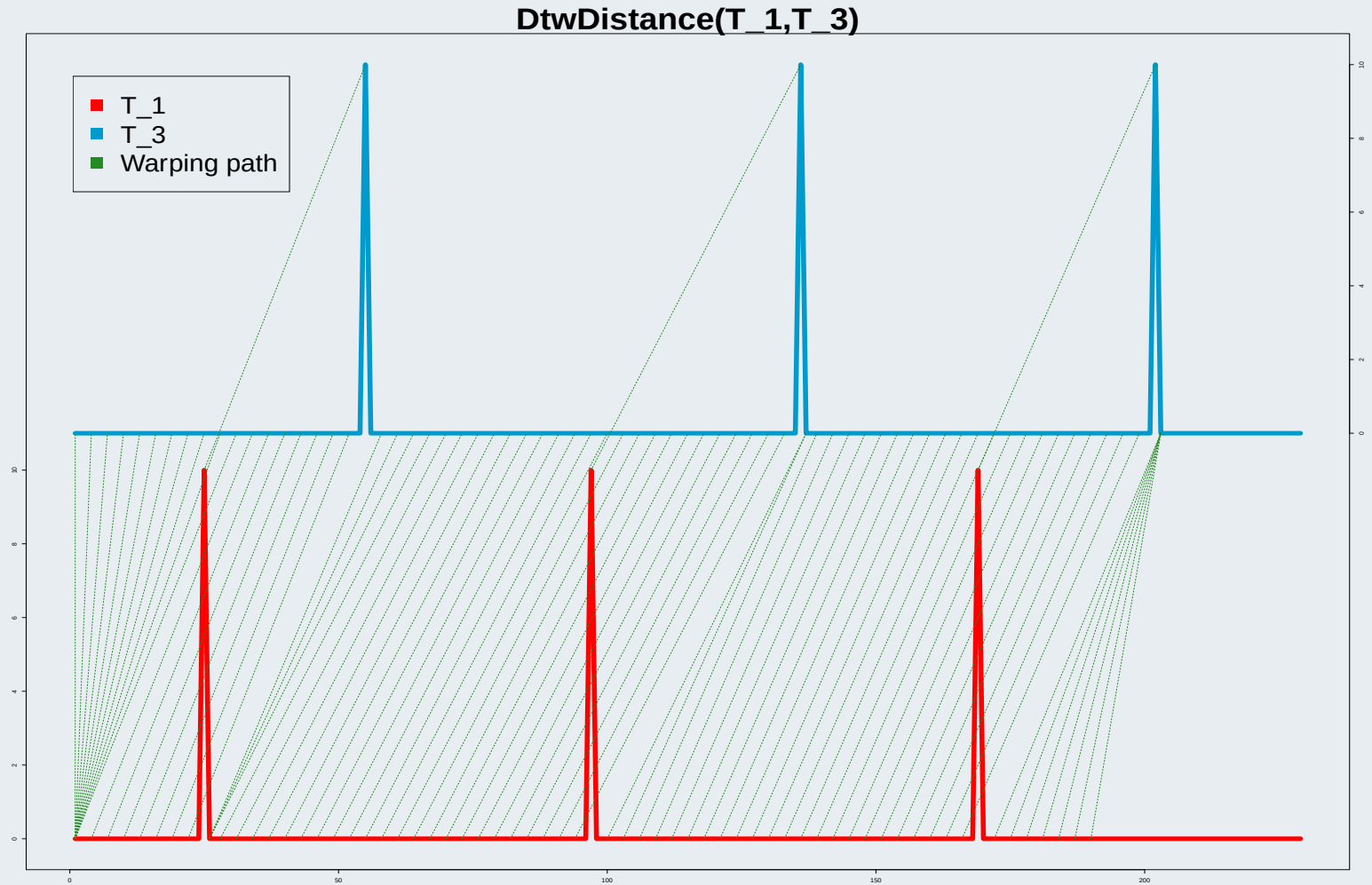
# Euclidean Distance

## Not Suitable



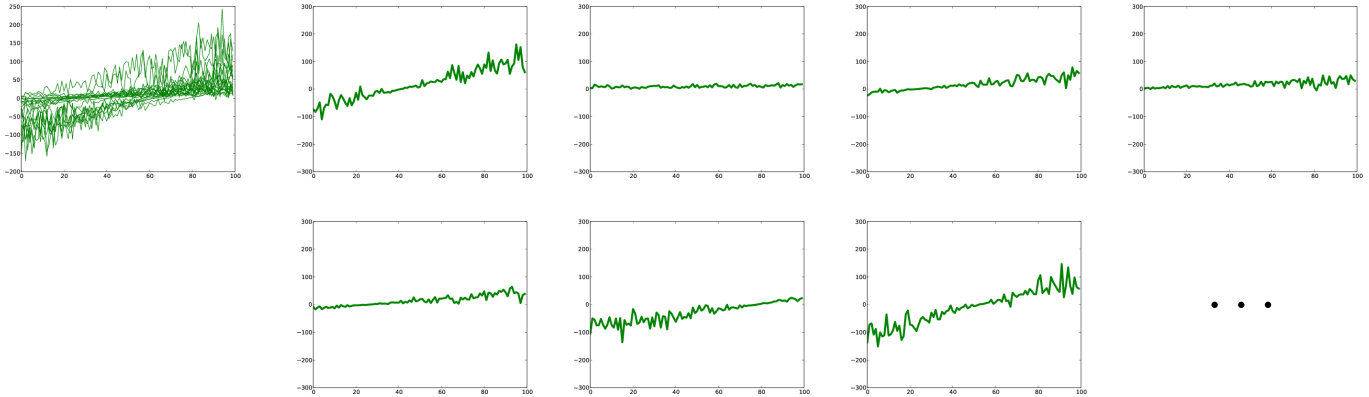
# Dynamic Time Warping (DTW)

## Suitable

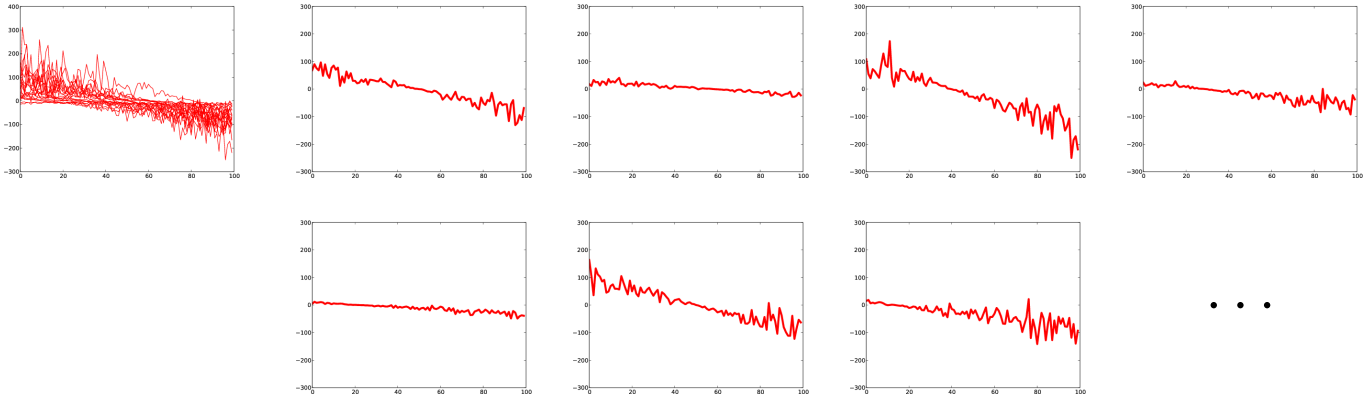


# Example: Increasing vs Decreasing Time Series

Increasing

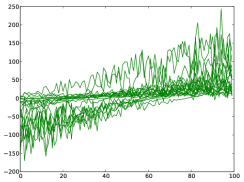


Decreasing

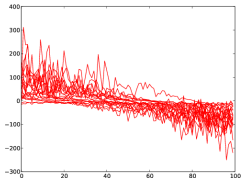


# Example: Building a Time Series Decision Tree

Increasing

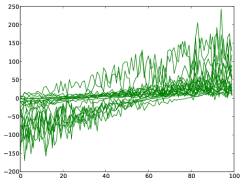


Decreasing

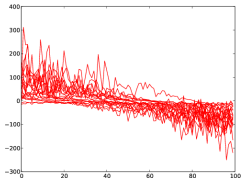


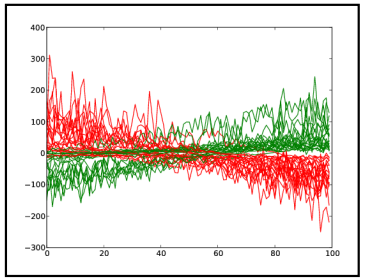
# Example: Building a Time Series Decision Tree

Increasing



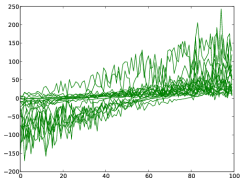
Decreasing



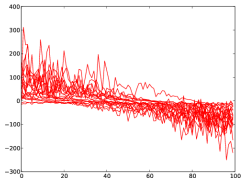


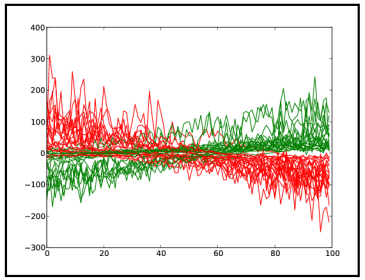
# Example: Building a Time Series Decision Tree

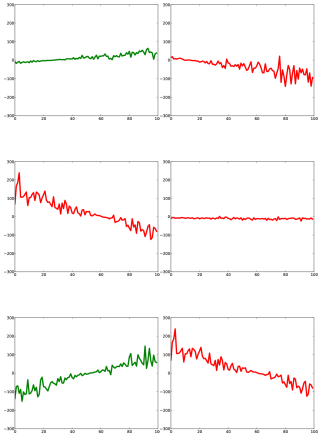
Increasing



Decreasing



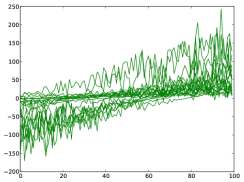




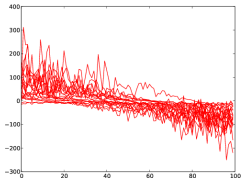
. . .

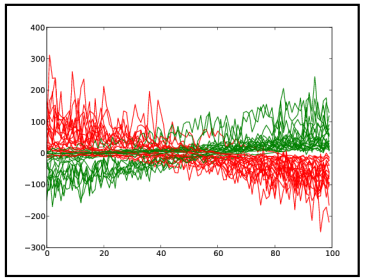
# Example: Building a Time Series Decision Tree

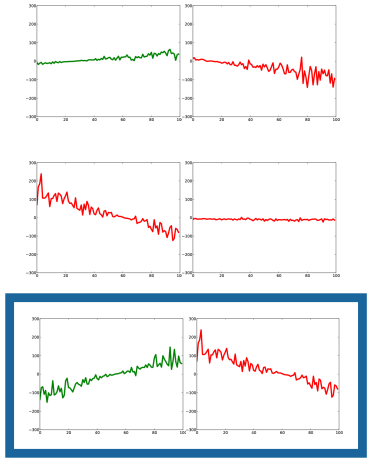
Increasing



Decreasing



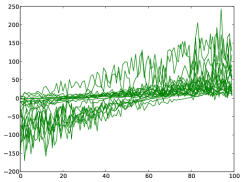




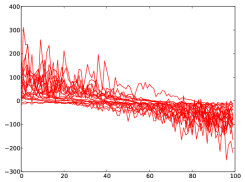
. . .

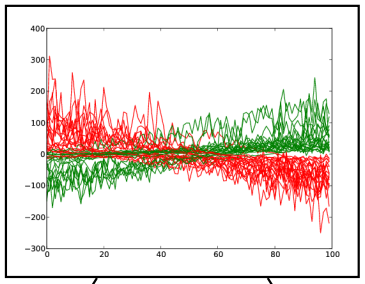
# Example: Building a Time Series Decision Tree

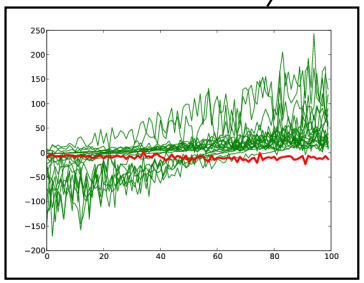
Increasing

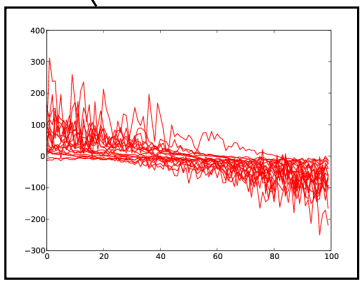


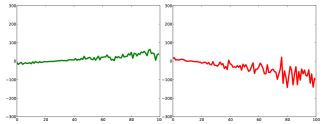
Decreasing

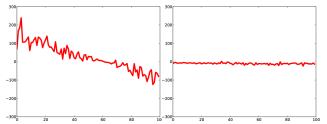


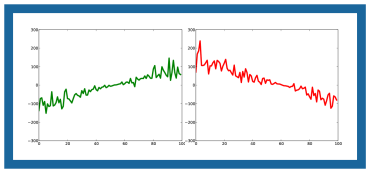






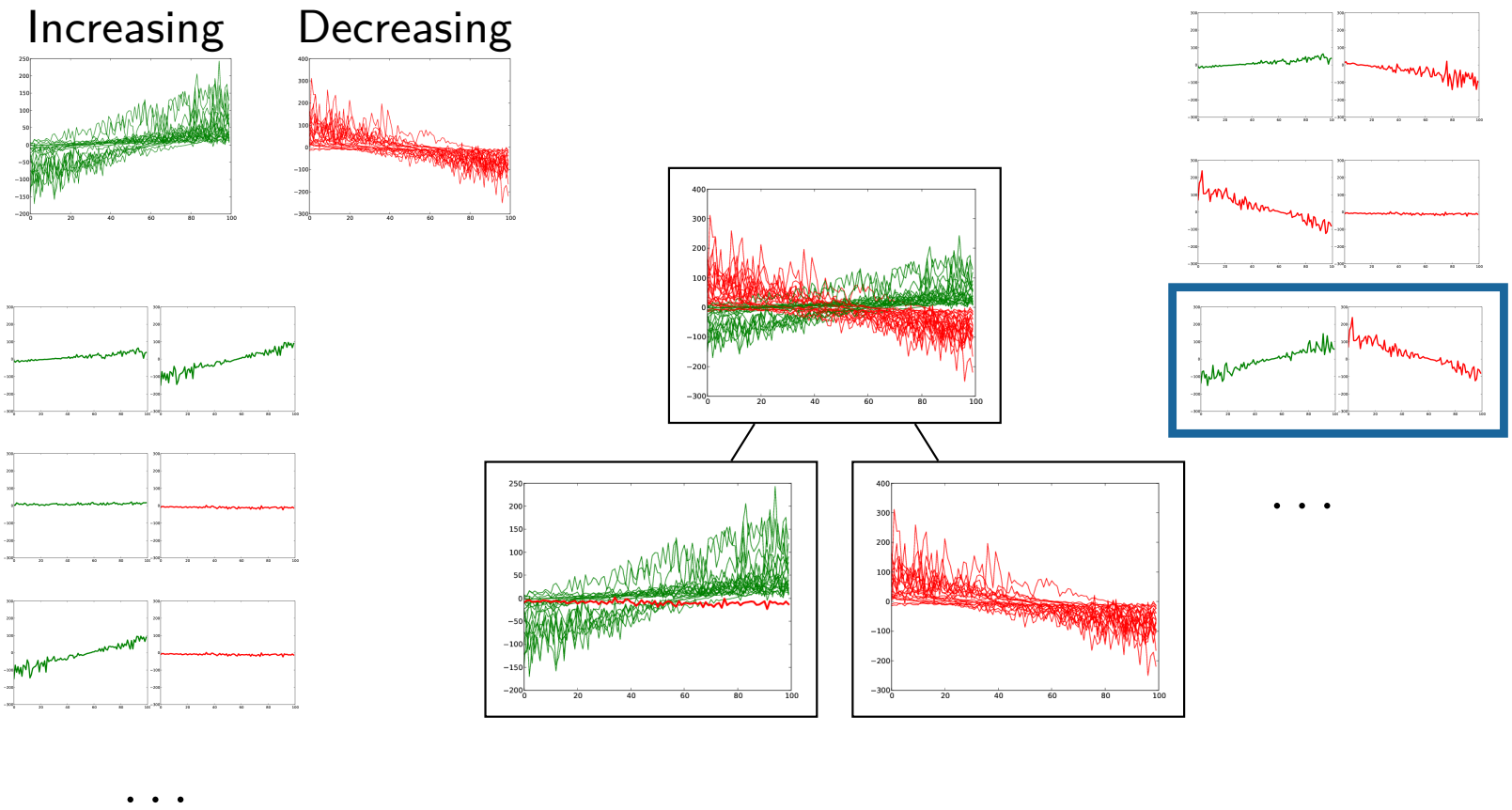




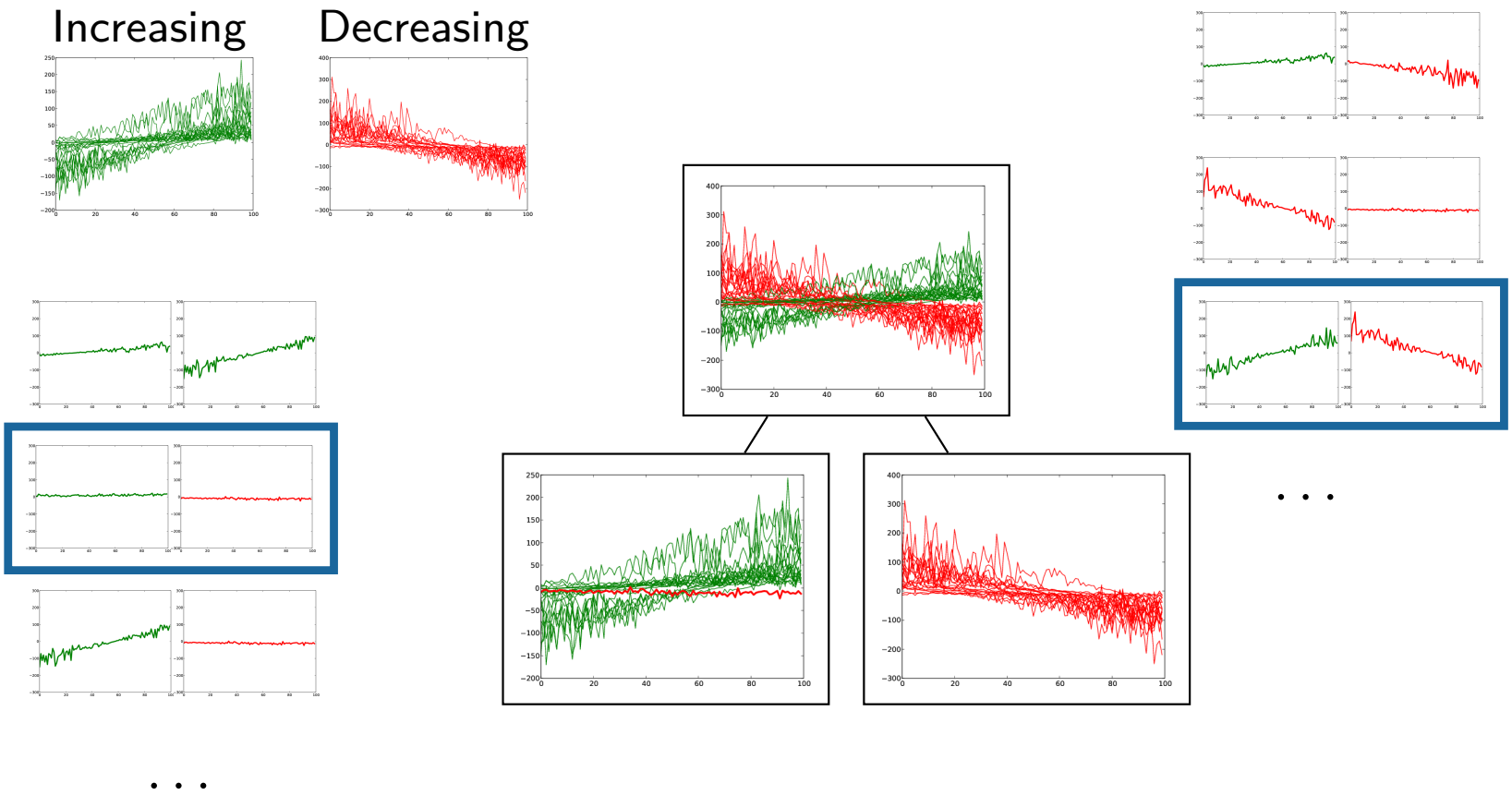


. . .

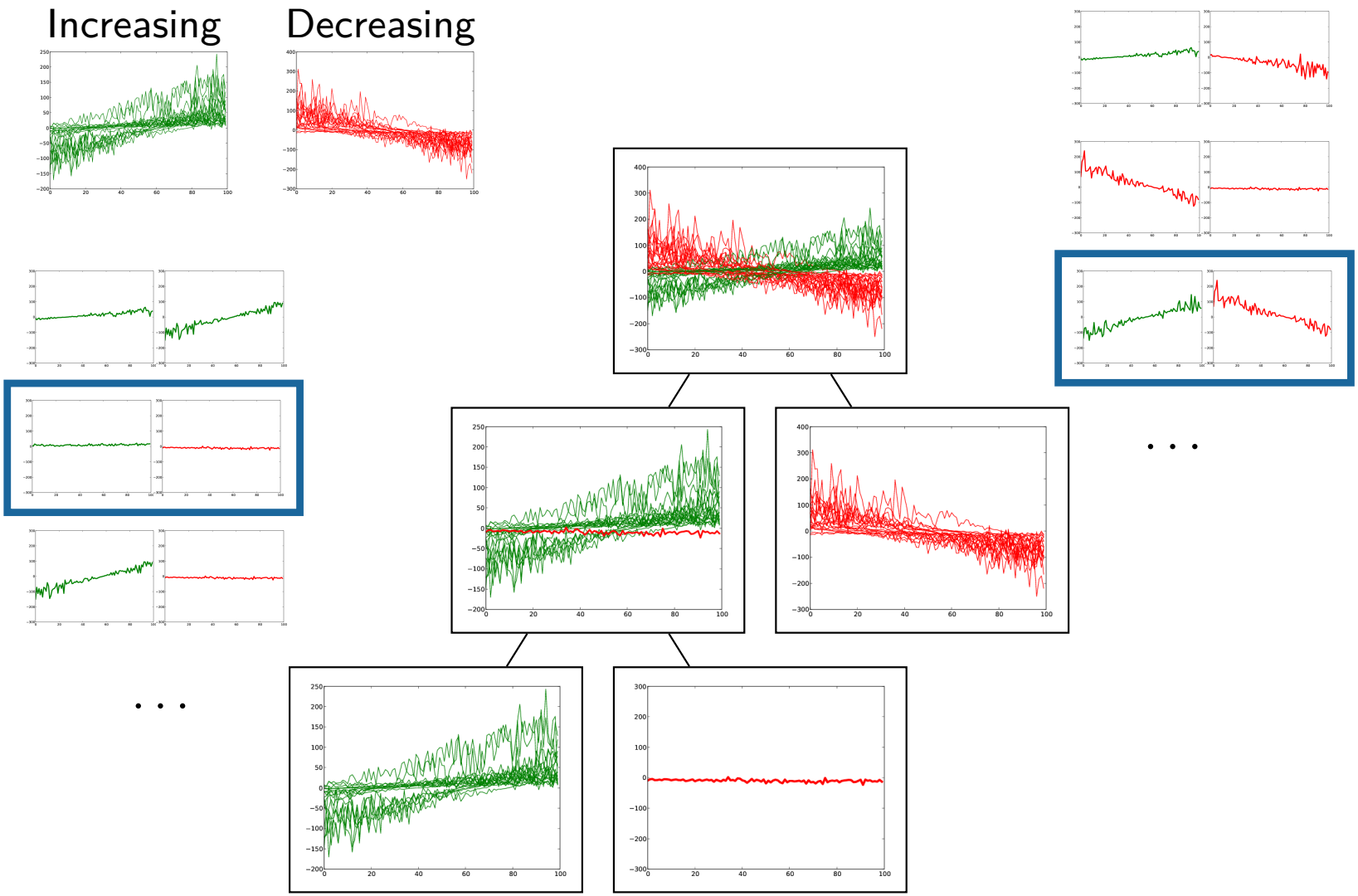
# Example: Building a Time Series Decision Tree



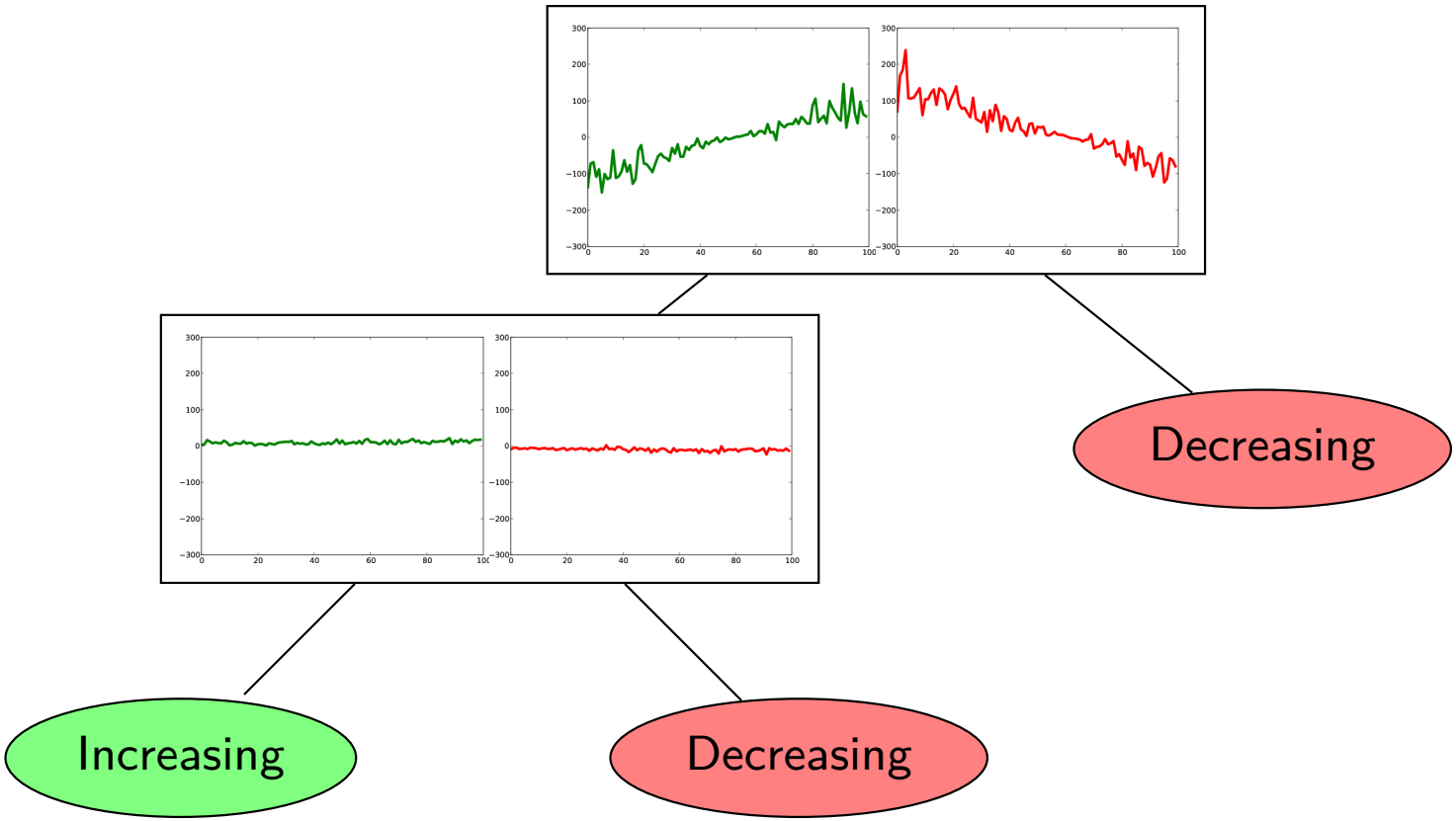
# Example: Building a Time Series Decision Tree



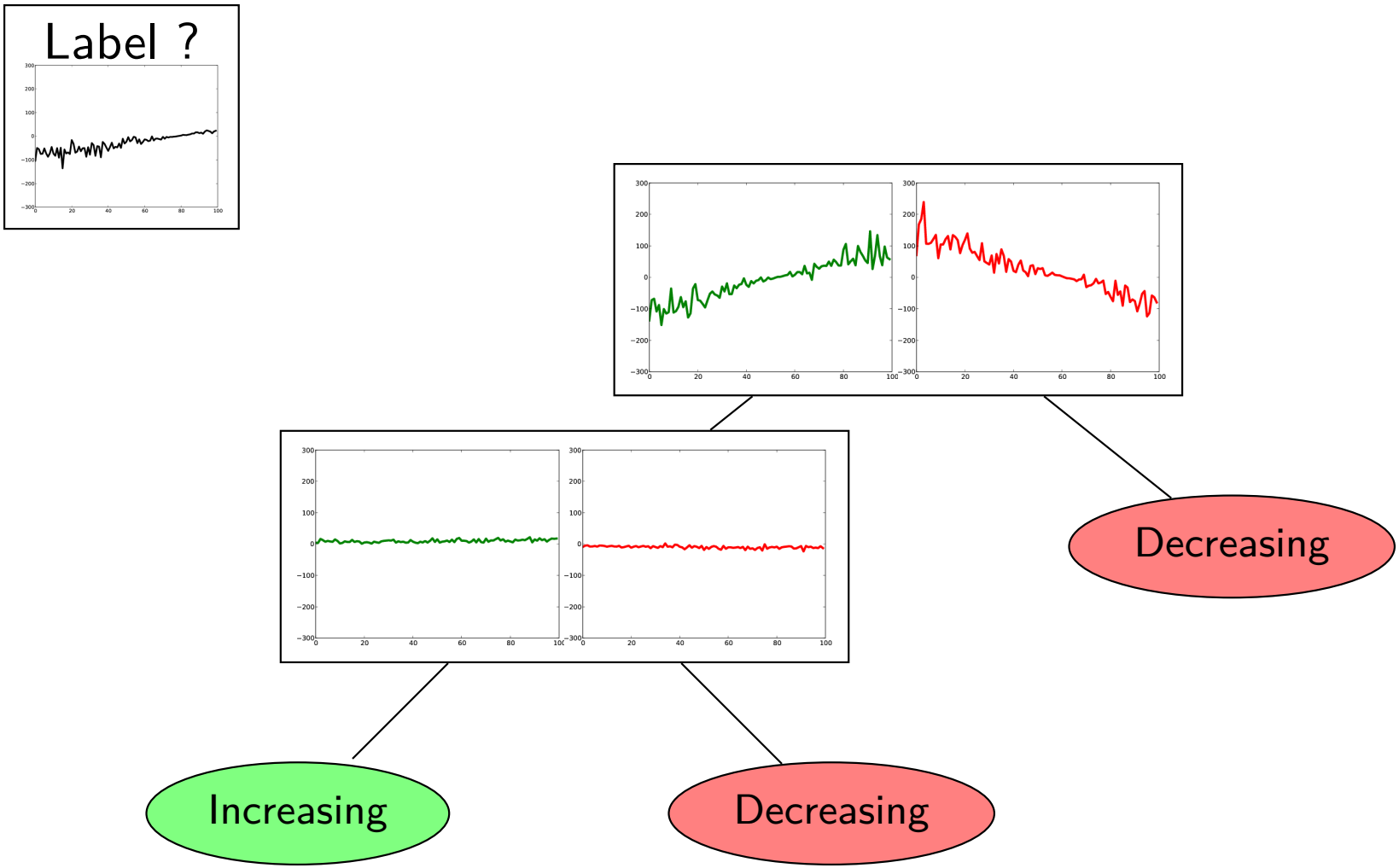
# Example: Building a Time Series Decision Tree



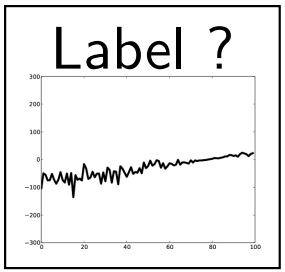
# Example: Predicting with a Time Series Decision Tree

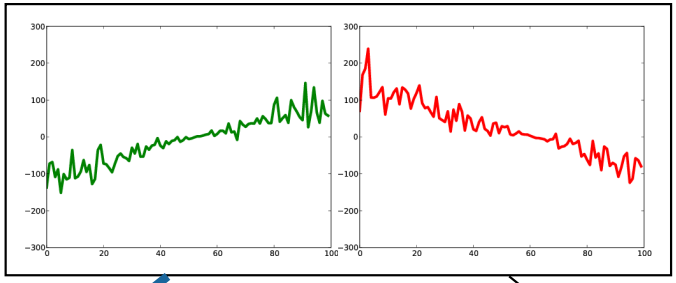


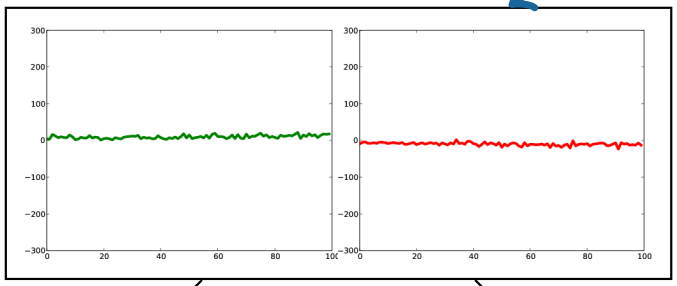
# Example: Predicting with a Time Series Decision Tree



# Example: Predicting with a Time Series Decision Tree



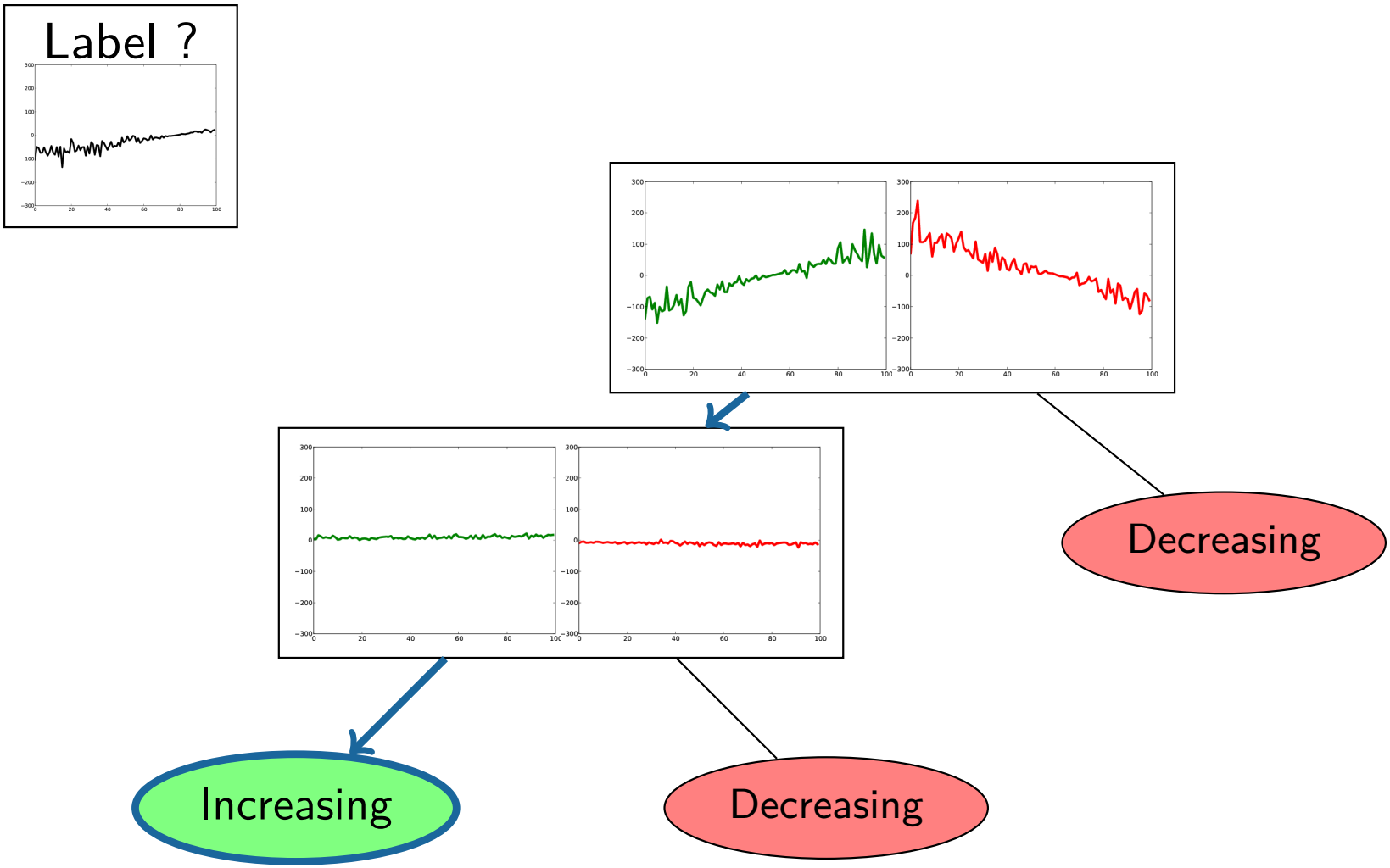


Decreasing

Increasing

Decreasing

# Example: Predicting with a Time Series Decision Tree



# Experimental Results

# Interpretation and Selection of the Parameters

## Parameters of the Supervised Learning

- Time Series
  - *Sampling interval* : 30min, 10min, 5min, 3min, 2min
- Decision trees
  - *Kind of splits* : `Cluster`, `Standard`, `Cluster/Standard`
- DTW distance
  - *Local distance* : from $\mathcal{L}_1$ to $\mathcal{L}_{10}$ , $\mathcal{L}_p(x, y) = (\sum(x_i - y_i)^p)^{\frac{1}{p}}$
  - *Distortion window* : from 0h to 24h

## Methodology

- Independent selection of the best parameters
- F-score

# Results for Conficker A

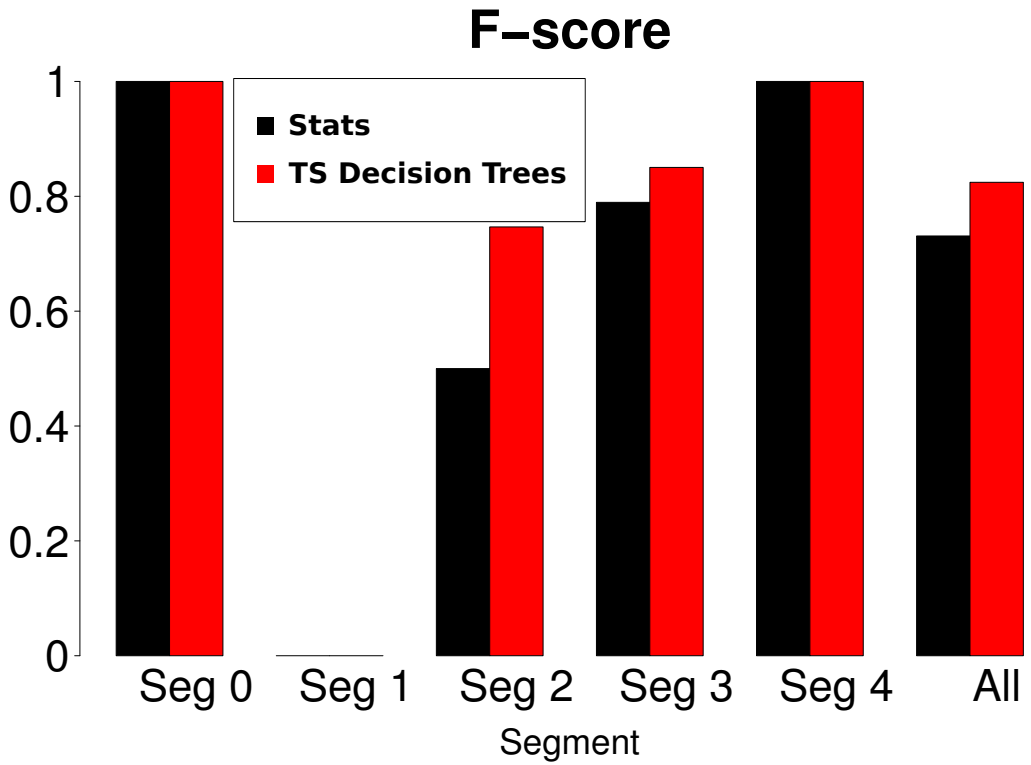
## Best Parameters

| | |
|---|---|
| **Sampling Interval** | 5min |
| **Split** | Stand. |
| **Local Distance** | $\mathcal{L}_7$ |
| **Distortion Window** | 14h |

## Best Model Performance

| | |
|---|---|
| **F-score** | 82.44% |
| **Accuracy** | 90.40% |
| **False Alarm Rate** | 7.38% |
| **Detection Rate** | 84.33% |

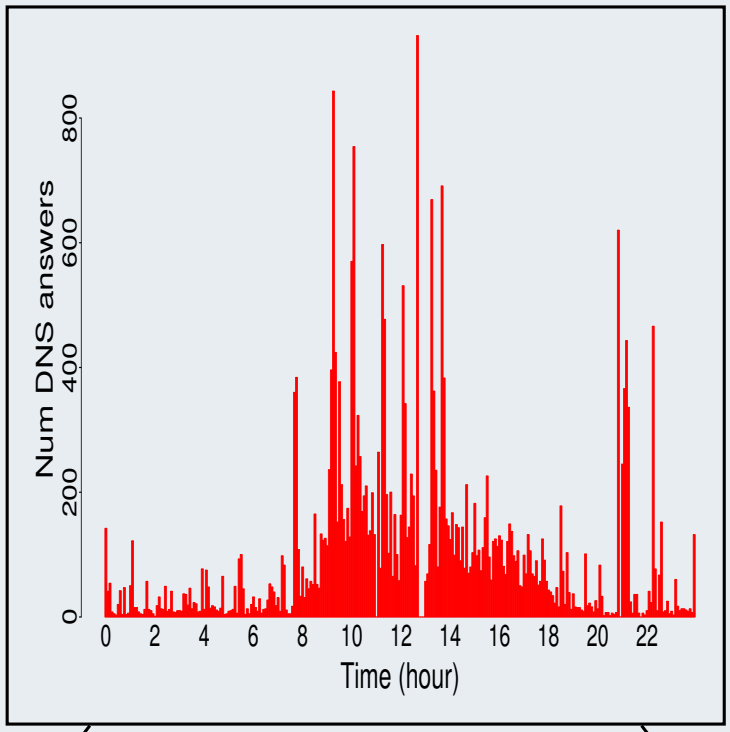
# Better Results with Whole Time Series




**Segments**

| Seg. | # DNS ans. |
|---|---|
| 0 | 1 to 10 |
| 1 | 11 to 100 |
| 2 | ... |

# Standard Split Decision Tree

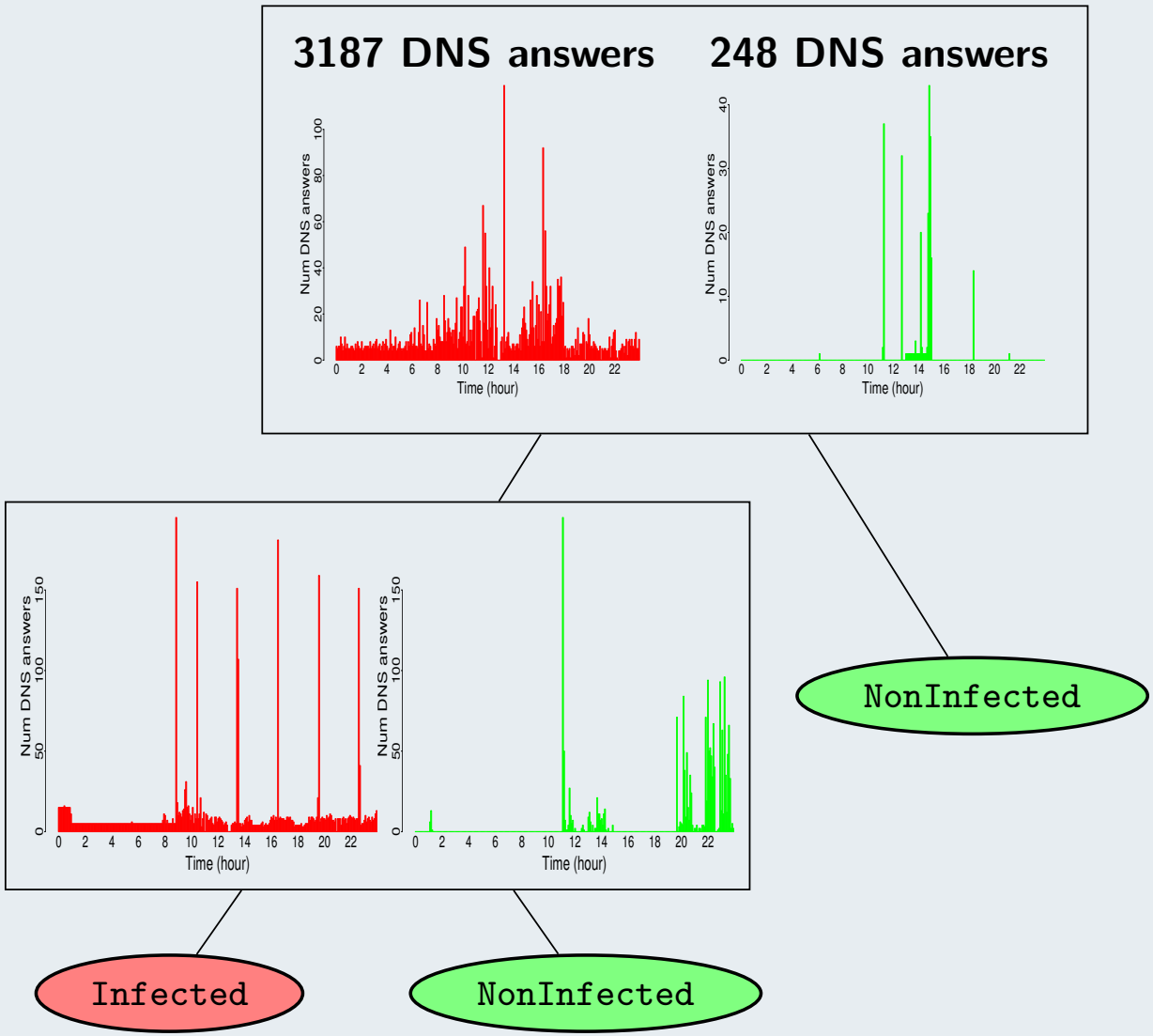



$r = 795.91$

Infected

NonInfected

# Cluster Split Decision Tree



# Conclusion

# Conclusion

## Behavioral Detection Model

- Classifier easy to interpret
- Whole time series as input
- Promising results

## From a Behavioral Detection Model to a Detection System

- More features
- Random forests
- Multi-class decision trees