Critical Infrastructures

(of today and tomorrow)

Simin Nadjm-Tehrani

www.ida.liu.se/~rtslab

Department of Computer & Information Science

Linköping University, Sweden

and

University of Luxembourg

Outline

- What are the challenges to today's critical infrastructures?

- Overview of the emerging infrastructures
  - A post-disaster communication network as an extreme case

- Own work in 2003-2005 in one of the first European projects on critical infrastructures

Attributes of dependability

Availability

- Readiness for use

Reliability

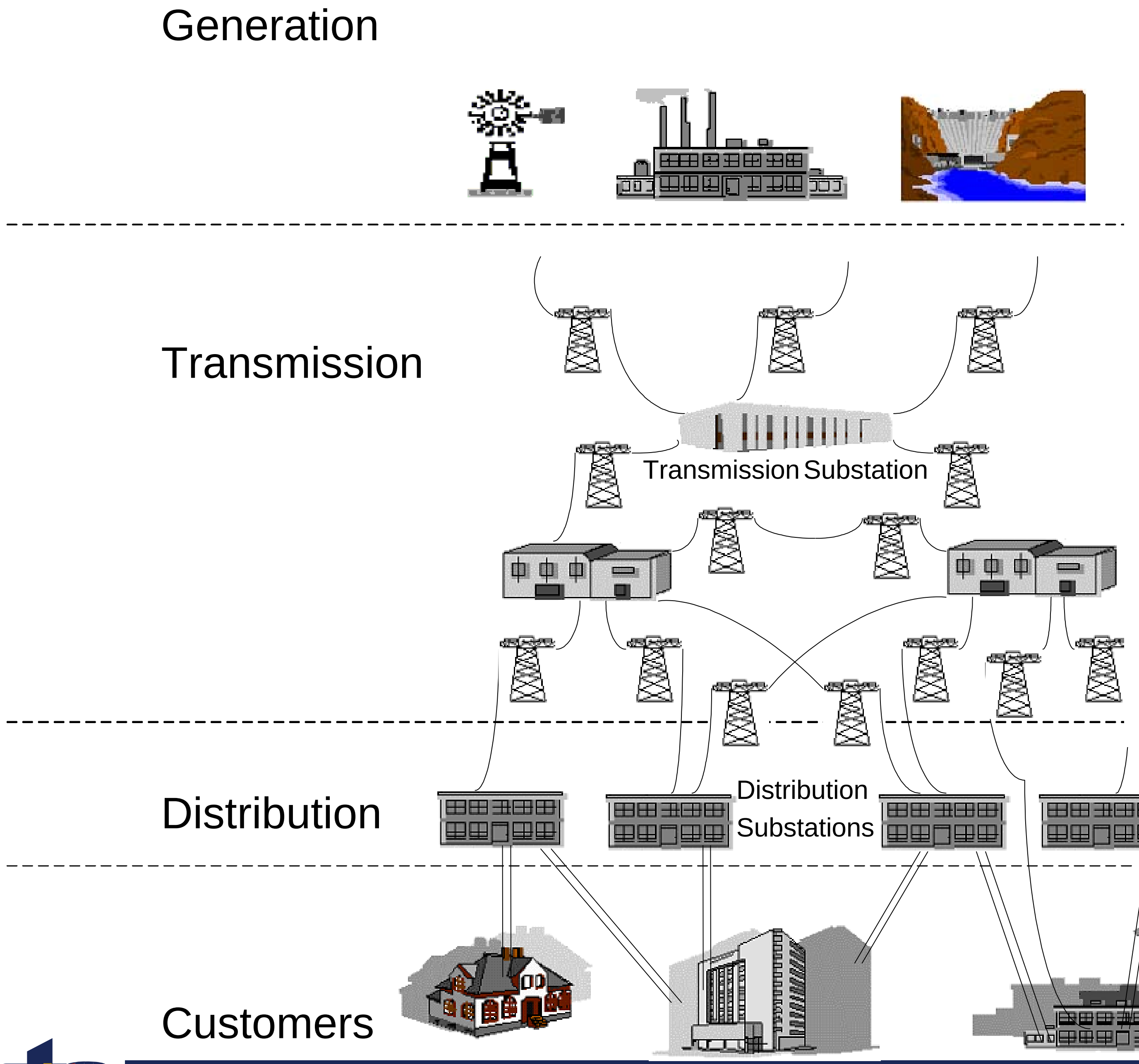
- Continuous correct service

Integrity

- No unauthorised change

Challenge 1

Complexity and interdependencies



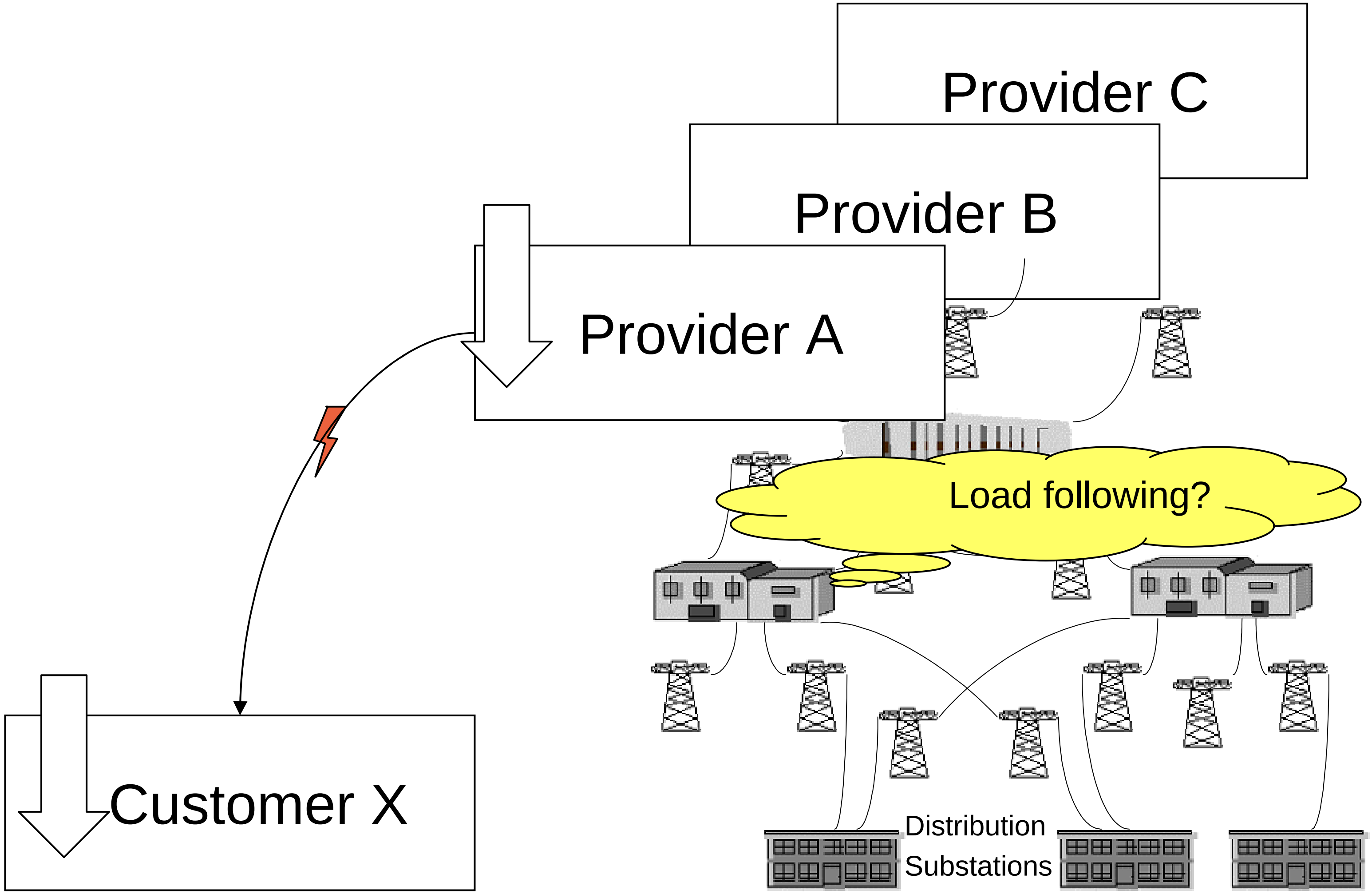
- Classic regional power grid
- One vertically integrated utility
- Grid control by frequency following
- Dedicated communication

[Source: NSTAC]

Restructuring of the Grid

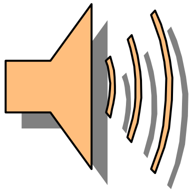
- Deregulation: organisations can enter into bilateral or multilateral power generation contracts
  - Large scale operation: from centralised to distributed control
  - Difficulty of coordination among independent service operators
- Approaching grid capacity
- New monitoring and control problems

Local impact spreads



Need for communication & trust

- Line frequency can no longer be the implicit communication channel
- Ideally contracts and capacities need to be known to everyone for cooperative control



In reality ...

- No operator wants to disclose information unless mandated by authorities
- Line frequency not enough for stablisation: one needs to know the state of equipment, detailed load profiles, pricing,…

26 August 2008

- 646 flights delayed as a direct result of a failure in a communication link that transmits flight plan data from the Georgia facility to a similar facility in Salt Lake City
- Flights from a wide swath of the United States, from Dallas and to the East Coast delayed
- The FAA: the source of the computer software malfunction was a "packet switch" that "failed due to a database mismatch."

Challenge 2

Transition from managed to unmanaged

20 August 2007

- Skype today provided a few more information pieces about the reasons behind its massive network outage last week.

- The network outage was initially caused by a "massive restart of [its] user's computers across the globe within a very short timeframe as they rebooted after receiving a routine software update."

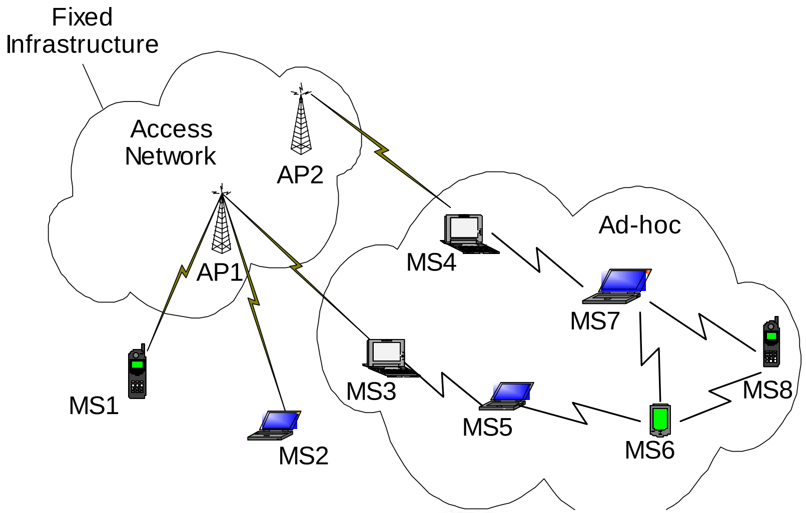
- That high number of reboots was followed by an equally high number of log-in requests, which resulted in what Skype calls a "chain reaction."

Challenge 3

Heterogeneity

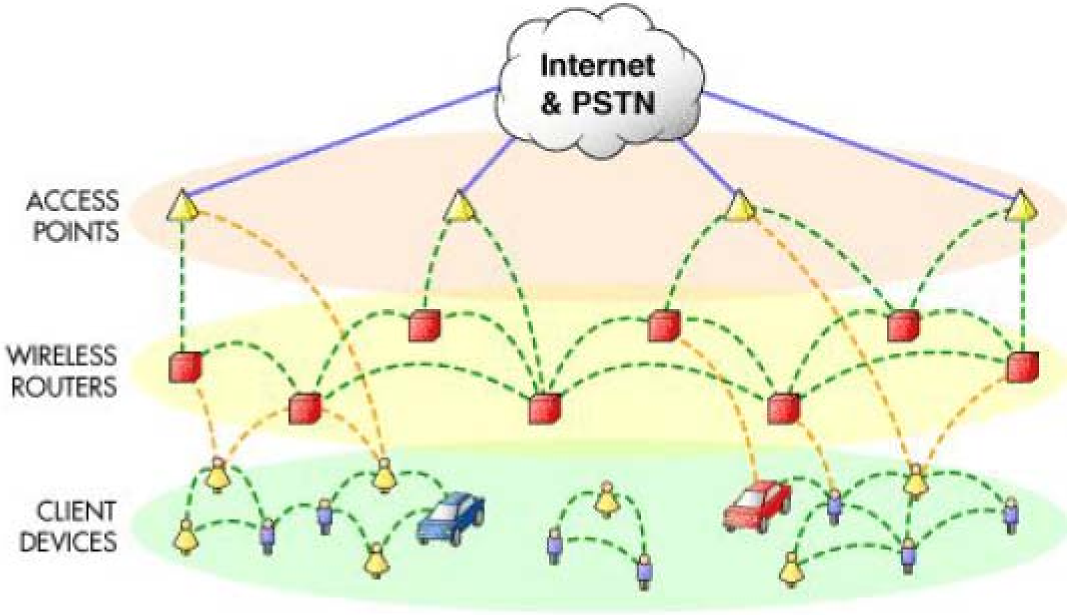
Converging networks

- From cellular …



- … to adhoc networks with no infrastructure

- … to multi-region Intermittently-connected networks



Reliance on novel technologies

- Wireless Communication
  - Almost taken for granted as part of the infrastructure today
  - GPRS, HSDPA, WiMAX, Wi-Fi, …

- Distributed cell networks
  - Local (per customer) generation of electricity
  - Dynamic energy market trading at customer level

Challenge 4

Organised threats with economic motives or adversary disruptions

Symantec Threat Report - Dec 07

- An average of 61,940 active bot-infected computers per day in the second half of 2007, an increase of 17% from the previous period.
- 499,811 new malicious code threats were reported to Symantec, a 136% increase over the first half of 2007.

Georgian govt. web attacks

- Gadi Evron, a prominent Internet security researcher and the founder of Israel's Computer Emergency Response Team, posited that the attackers are more likely nationalistic "enthusiasts" than organized criminals or Russian government operatives.

Summary

Challenge	Emerging solutions
Complexity and interdependencies	Modelling, Risk analysis, Provisioning
Transition from managed to unmanaged	P2P technologies, self-managing systems
Heterogeneity	Standardised protocols, Overlay networks, Software defined radio
Organised threat, fraud and disruptions	Hardening, Intrusion tolerance, diversity, partial rejuvenation

Outline

- What are the challenges to today’s critical infrastructures?

- Overview of the emerging infrastructures
  - A post-disaster communication network as an extreme case

- Own work in 2003-2005 in one of the first European projects on critical infrastructures

What happens in the worst case?

- Existing infrastructure collapses

Chaotic & surprising

Network: lack of resources

- Time is running out…

- Actors are spread out and mobile

- Communication culture clashes

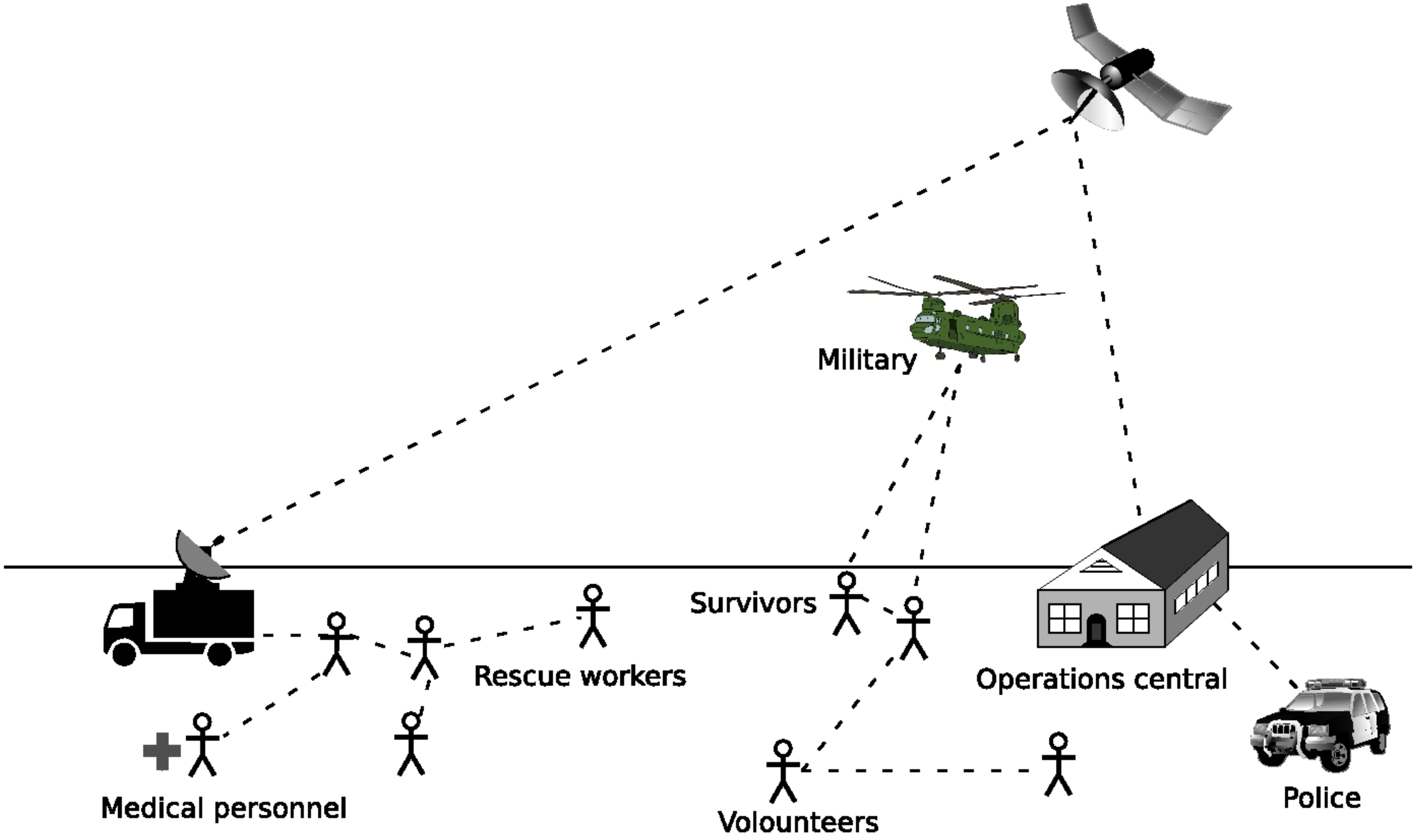
Our hypothesis

- Hastily formed networks can have a role to play

- Use commodity hardware and massively distributed software

- Have built-in mechanisms for
  - When batteries are in short supply
  - Mobility changes connectivity
  - Dealing with overload and urgency
  - Detect and respond to abuse

Multiple information owners/users



Hastily formed networks

Challenge	Emerging solutions
Disconnectivity as a norm	Store-and-forward techniques, delay-tolerant networks (DTN)
Resource constraints	QoS optimisation techniques, prioritisation
Infeasibility to centrally manage	Gossip-style distributed protocols
Heterogeneity	Overlay networks, DTN bundles
Less organised opportunistic threats, adversary disruptions	Reputation-based systems, Selfish-resistance protocols, Decentralisation

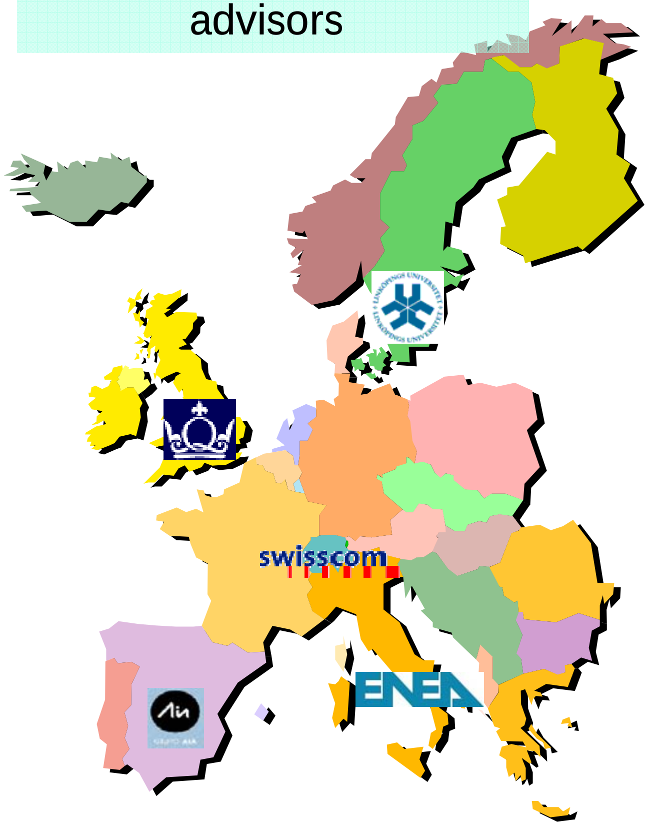
Outline

- What are the challenges to today’s critical infrastructures?
- Overview of the emerging infrastructures
  - A post-disaster communication network as an extreme case
- Own work in 2003-2005 in one of the first European projects on critical infrastructures

Project Safeguard

- Goal: to enhance survivability of Large Complex Critical Infrastructures (LCCIs)
- Electricity and telecommunications networks as practical examples
- Granted pre 9/11!
- Ended in 2004

plus a panel of senior government and industry advisors



Challenges

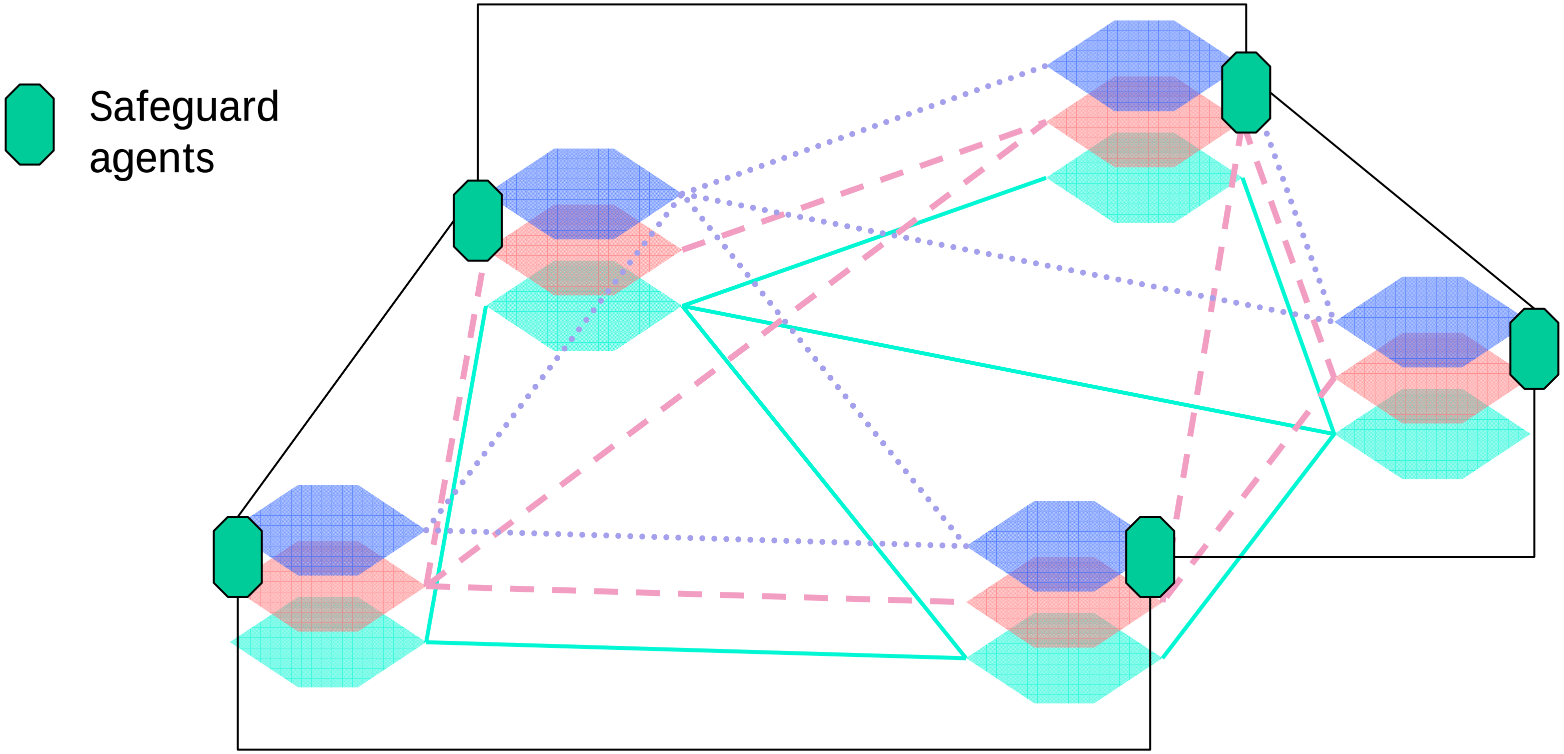
General:

- Increase information quality for administrator
- Recognise unknown attacks
- Predict future overloads

Telecom specific:

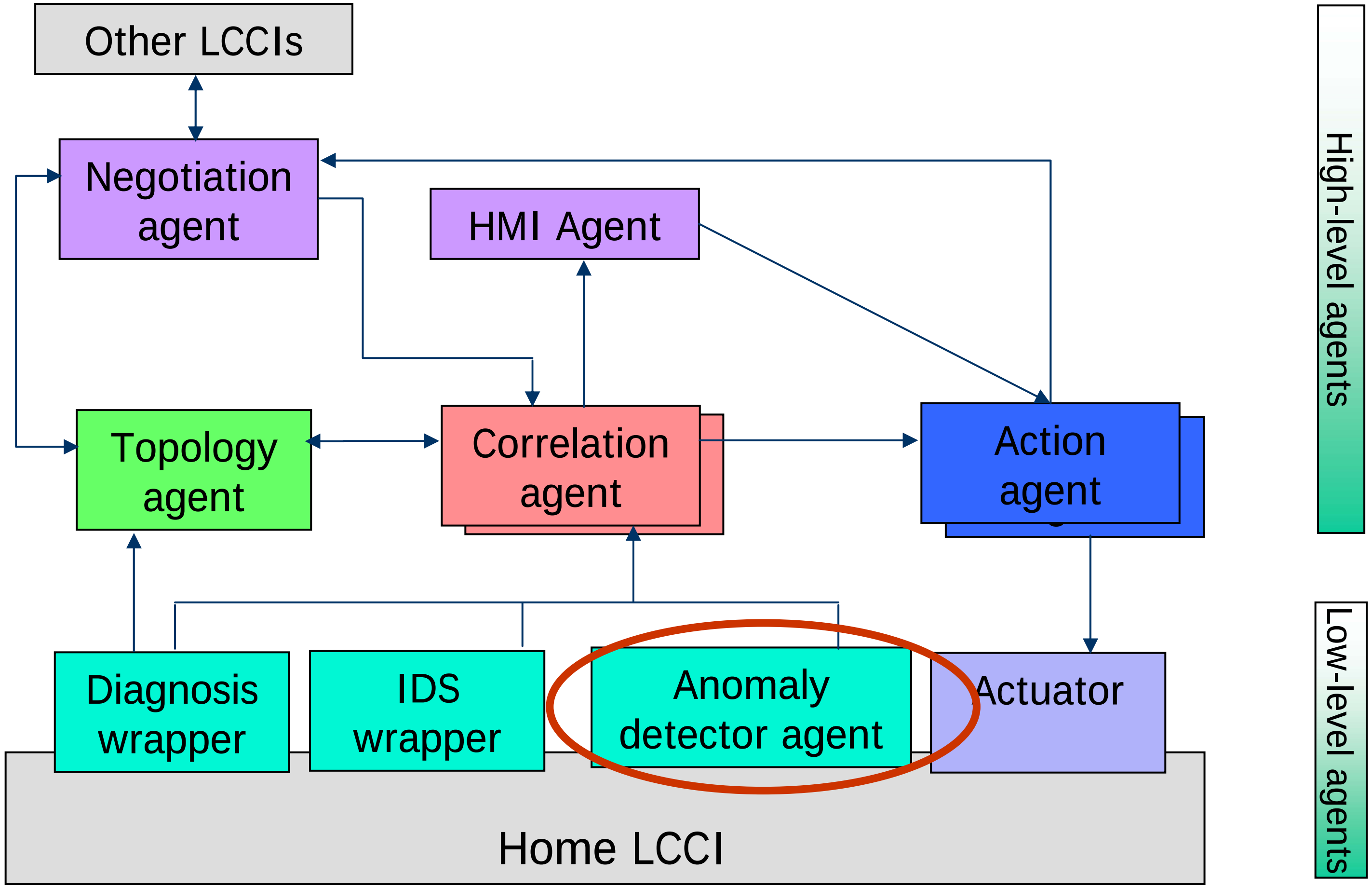
- Decrease no. of alarms
- Decrease false positives (higher availability)

The Safeguard approach



Safeguard agents

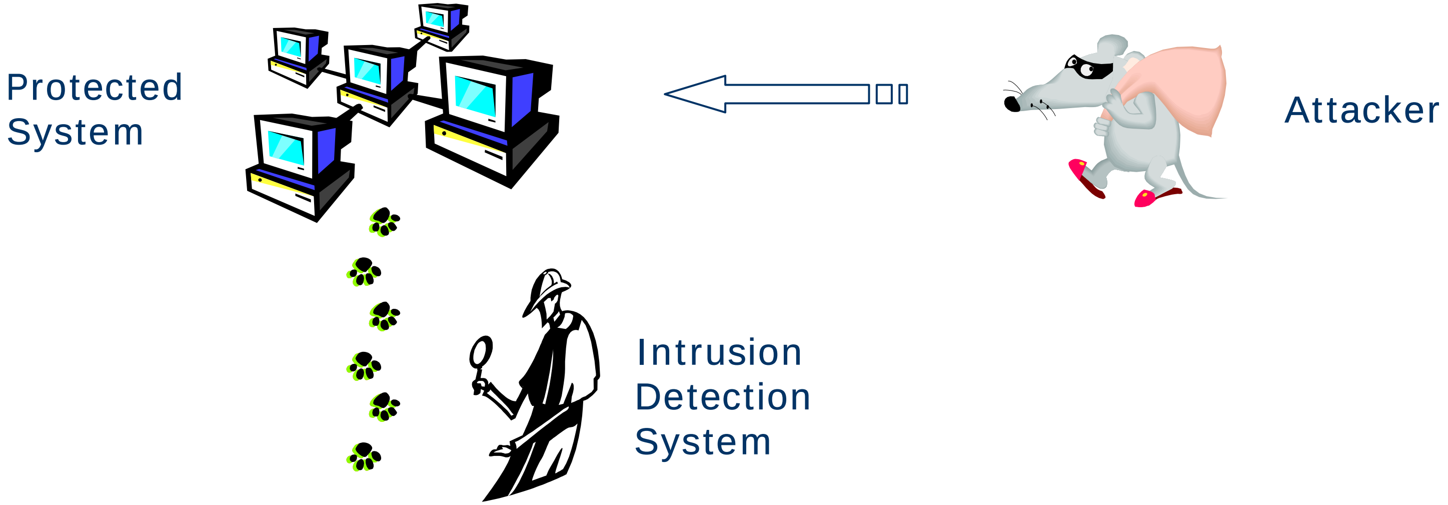
Safeguard architecture



Anomaly Detection

- ADWICE: Anomaly Detection With fast Incremental ClustEring
- Joint work with Kalle Burbeck
- Not a silver bullet: part of the larger Safeguard context

Intrusion detection

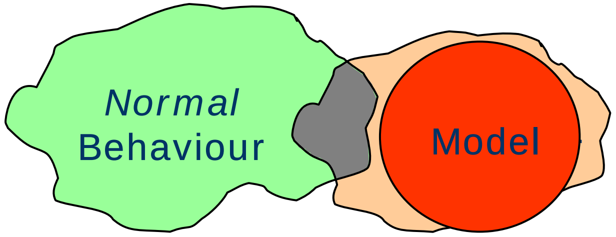


Protected System

Attacker

Intrusion Detection System

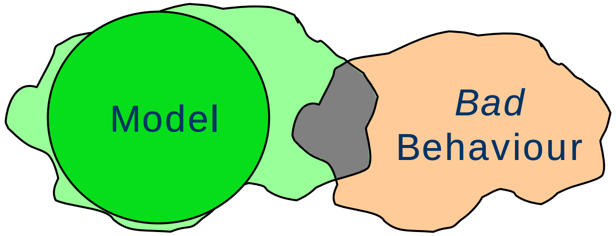
Misuse Detection



Normal Behaviour

Model

Anomaly Detection

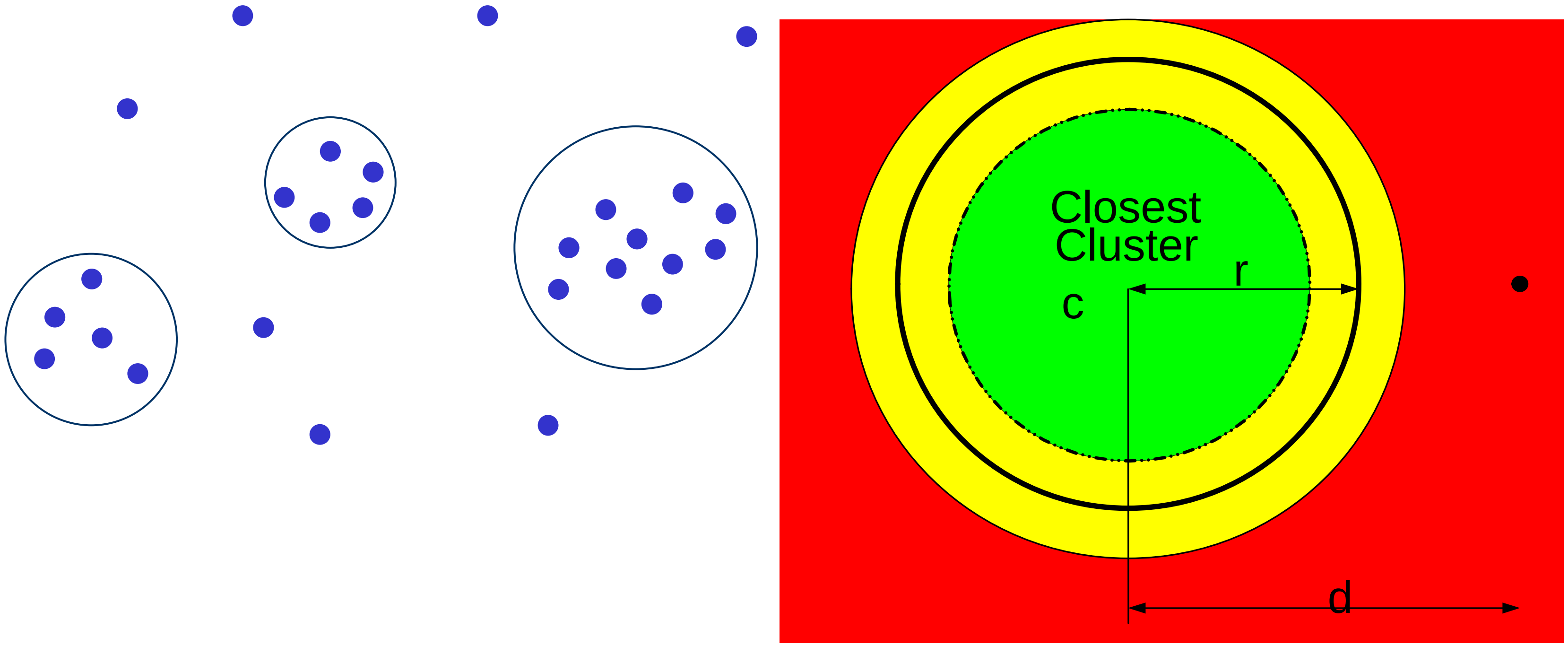


Model

Bad Behaviour

Clustering

- ADWICE uses clusters to represent normality
- Adaptation of an existing data mining algorithm (BIRCH)



What is a data point?

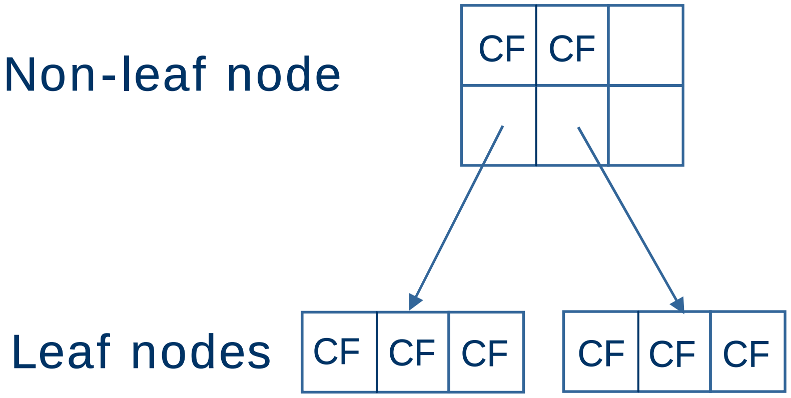
- General: A set of numeric values
  - E.g. measurements from sensors

- What about IP packets?
  - A vector of alphanumeric values in header of an IP packet
  - Transformed into vector of numeric values
  - In our tests: 41 dimensions

- Need efficient storage and search among summaries of collections of data points

Basic ADWICE concepts

- CF (Cluster Feature)
  - Summary of cluster
  - [No, Sum, Sum of sq]
- Index: CF Tree



Non-leaf node

CF CF

Leaf nodes

CF CF CF

CF CF CF

- Maximal number of clusters (M)
- Threshold requirement (TR)
- Branching factor (B)

Efficient operations

- We have: CF = $\langle \ n, \sum v_i, \sum v_i^{\ 2} \ \rangle$
- Can compute the Centroid v_0 :

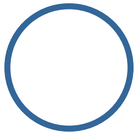
$$\sum v_i / n$$

- Can compute the Radius:

$$\sqrt{\sum (v_0 - v_i)^2 / n}$$

ADWICE training

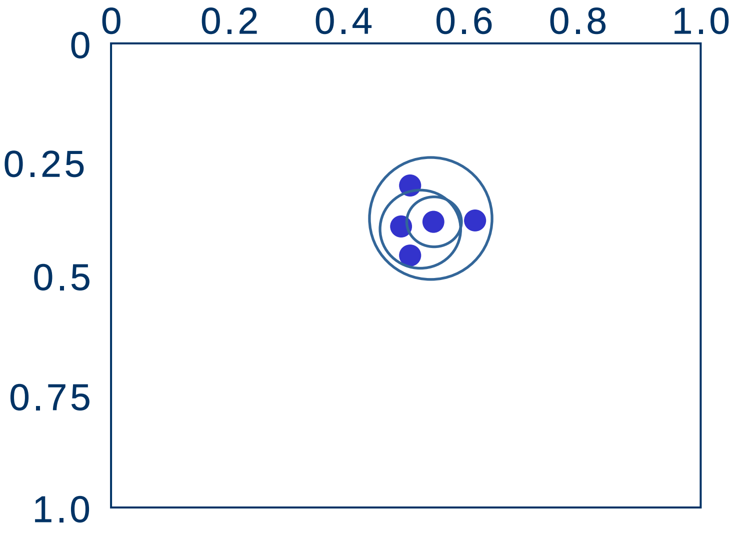
Threshold:



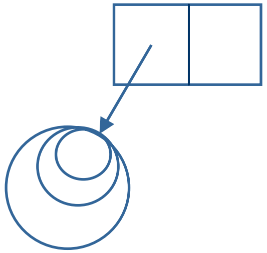
Max Number of Clusters: 3

Branching factor: 2

Data Space

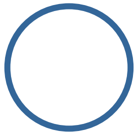


CF Tree



ADWICE training

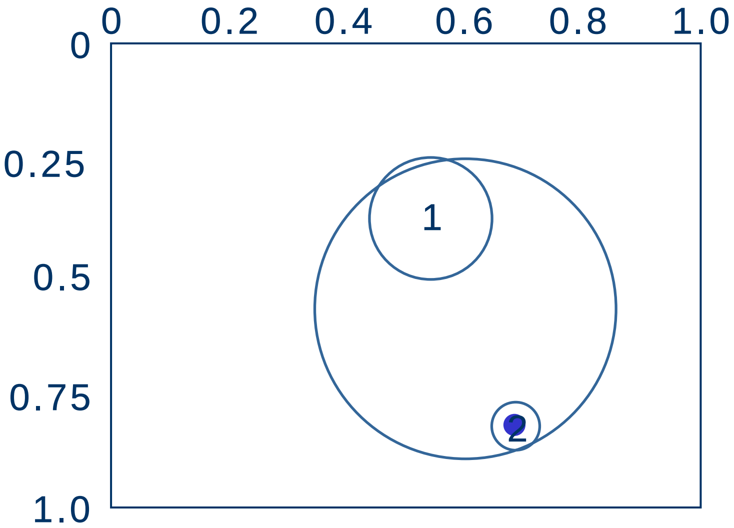
Threshold:



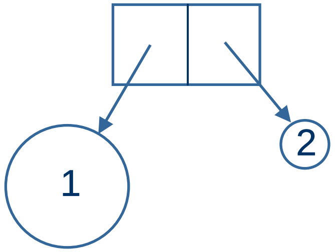
Max Number of Clusters: 3

Branching factor: 2

Data Space

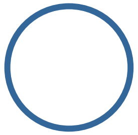


CF Tree



ADWICE training

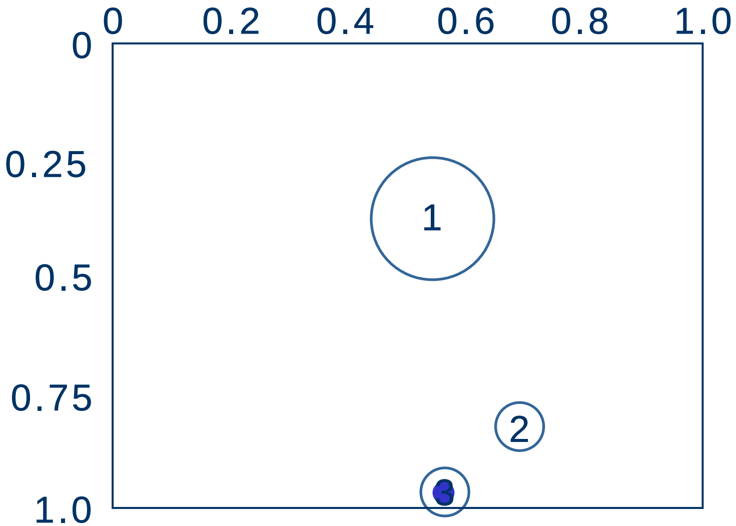
Threshold:



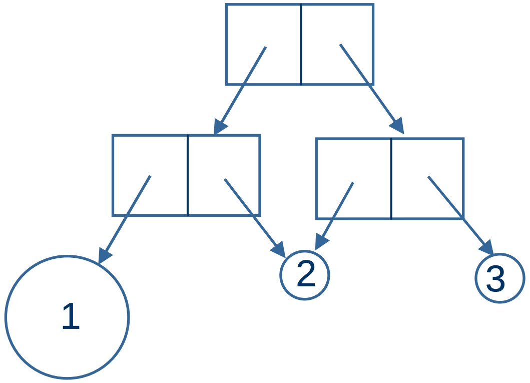
Max Number of Clusters: 3

Branching factor: 2

Data Space

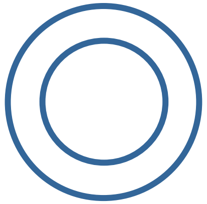


CF Tree



ADWICE training

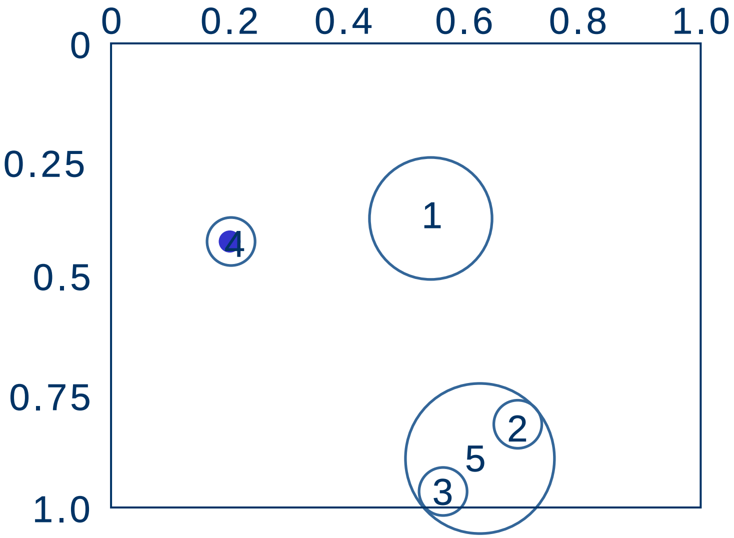
Threshold:



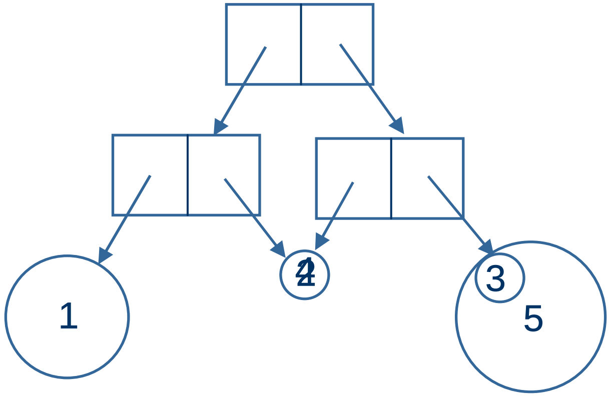
Max Number of Clusters: 3

Branching factor: 2

Data Space

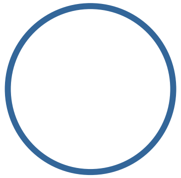


CF Tree



ADWICE detection

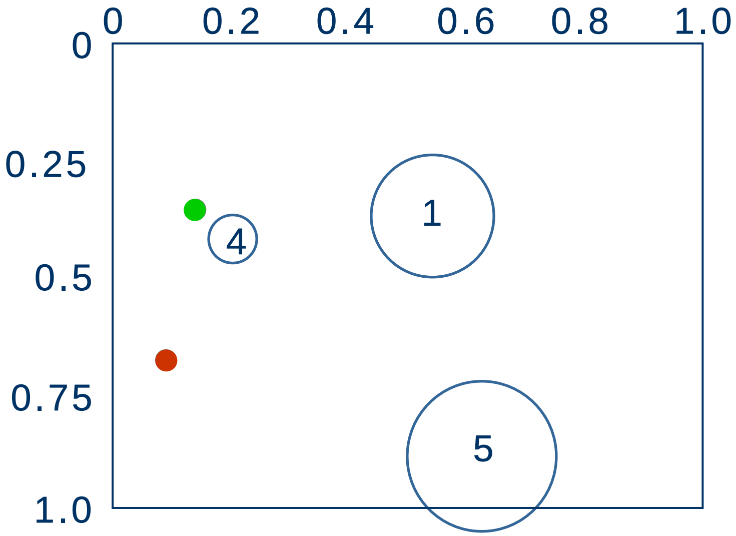
Threshold:



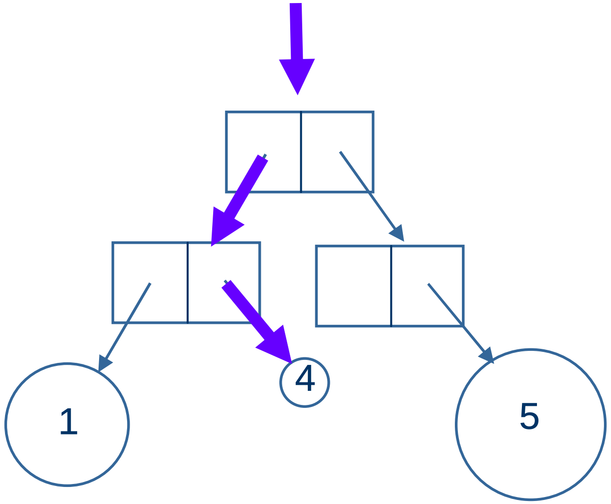
Max Number of Clusters: 3

Branching factor: 2

Data Space



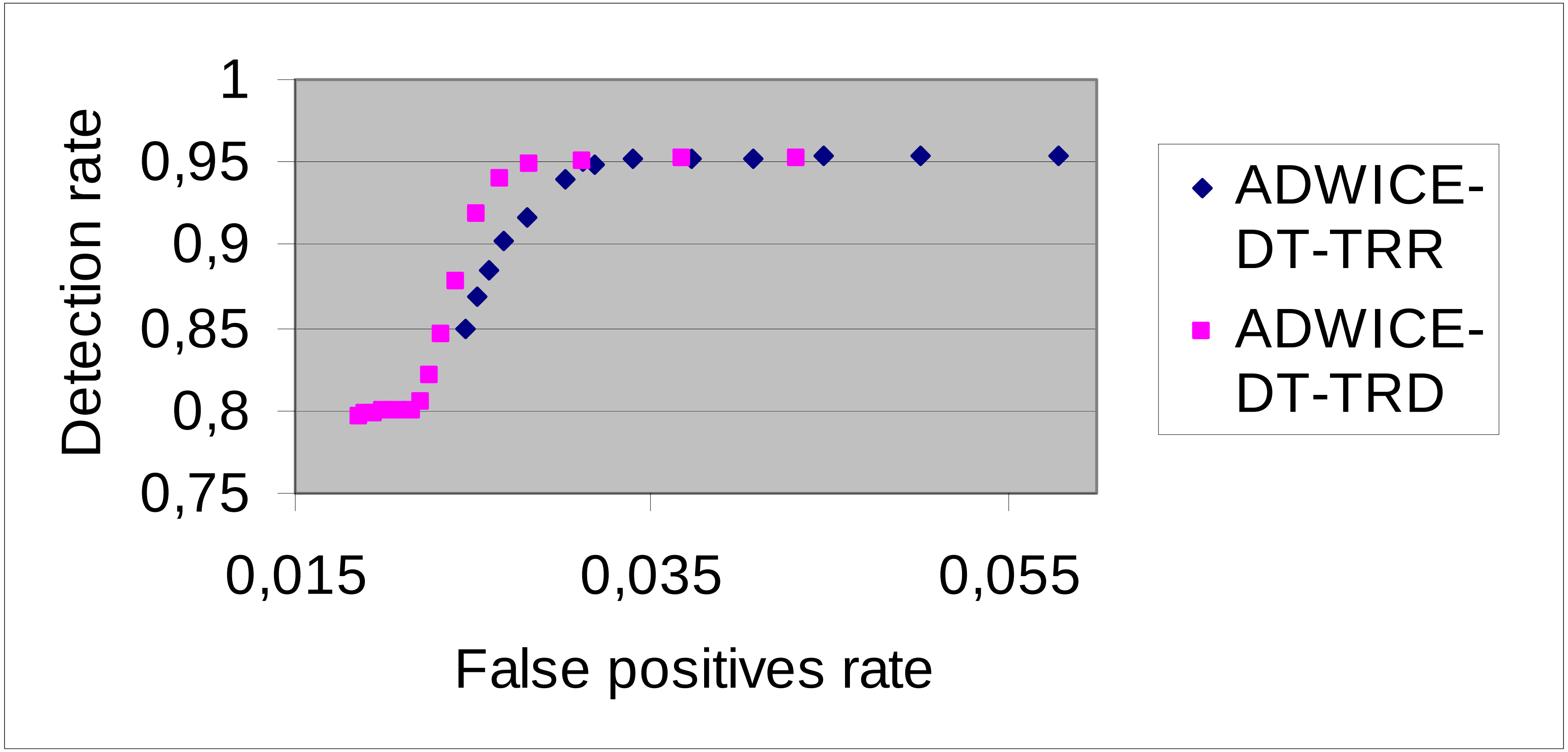
CF Tree



Evaluation

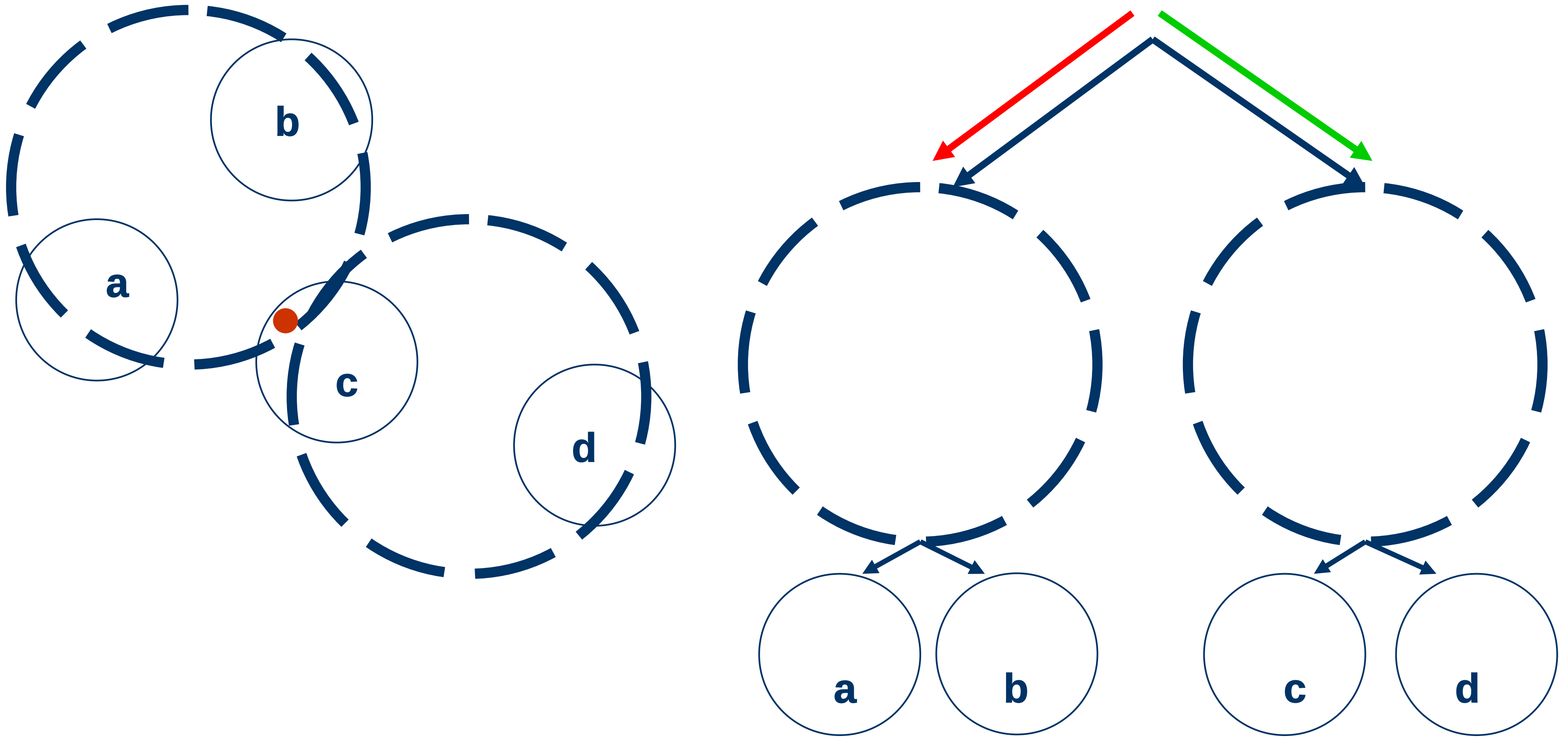
- KDD99 Data
- General properties
  - Session records (TCP/UDP summaries)
  - 41 features (flags, service, traffic stats ...)
- Training data
  - 4 898 431 session records
  - 972 781 normal, the rest (attacks) not used
- Testing data
  - 311029 session records
  - normal data and 37 different attack types

Detection rate vs. false positives



Index errors

- Some false positives are due to index errors

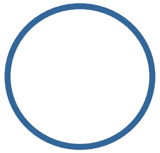


ADWICE-Grid

- A new version of the algorithm: separates cluster formation and index updates
- How does ADWICE- Grid work?

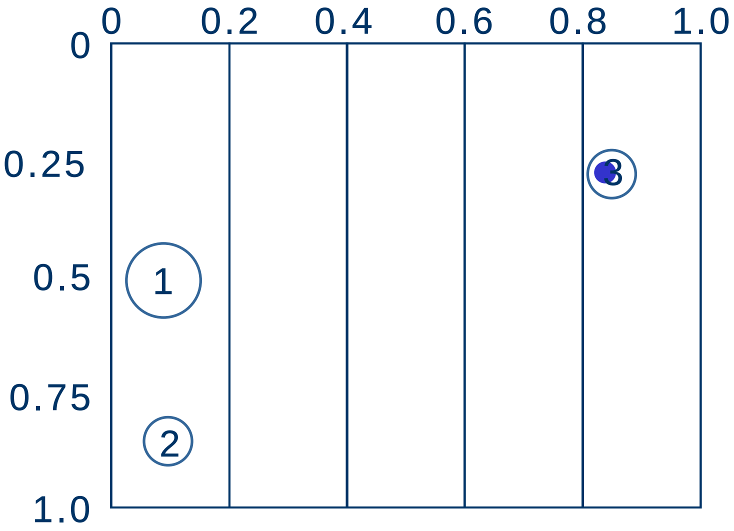
ADWICE-Grid: Training

Threshold:

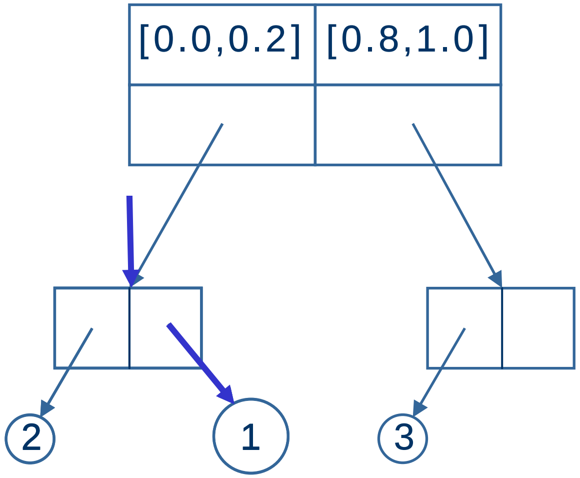


Max clusters in Leaf: 2

Data Space



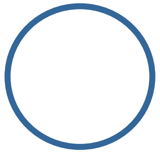
CF Tree



ADWICE-Grid: Training

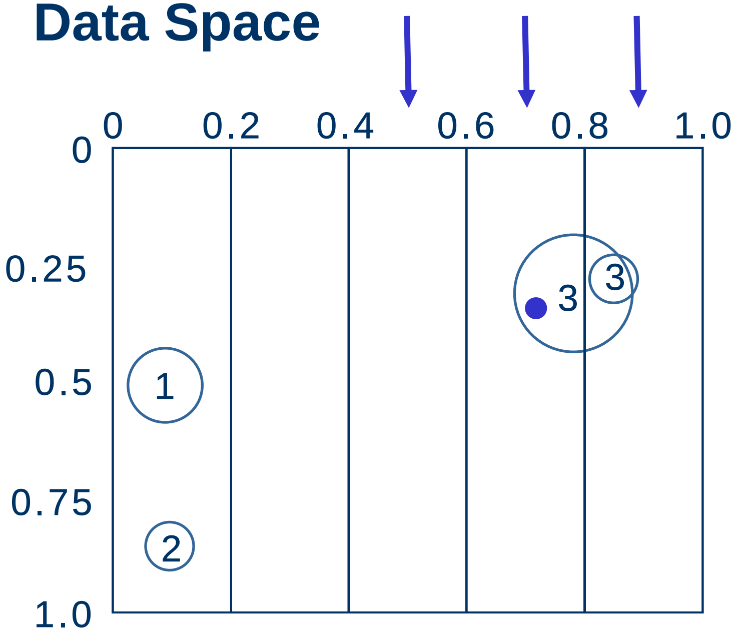
Threshold:

(Search width)

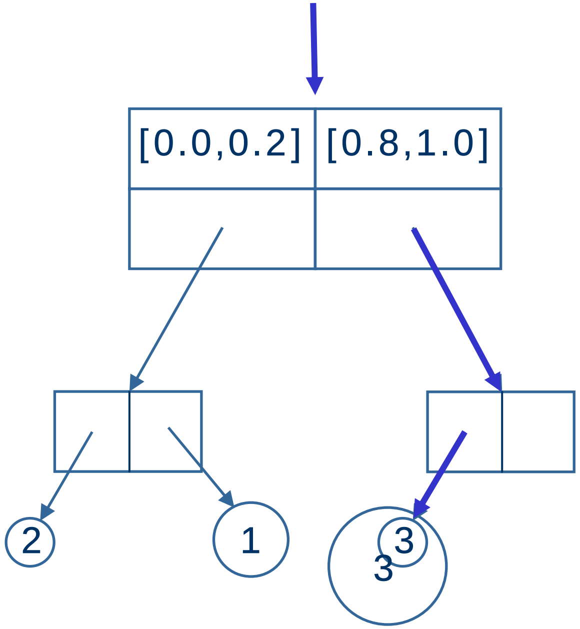


Max clusters in Leaf: 2

Data Space

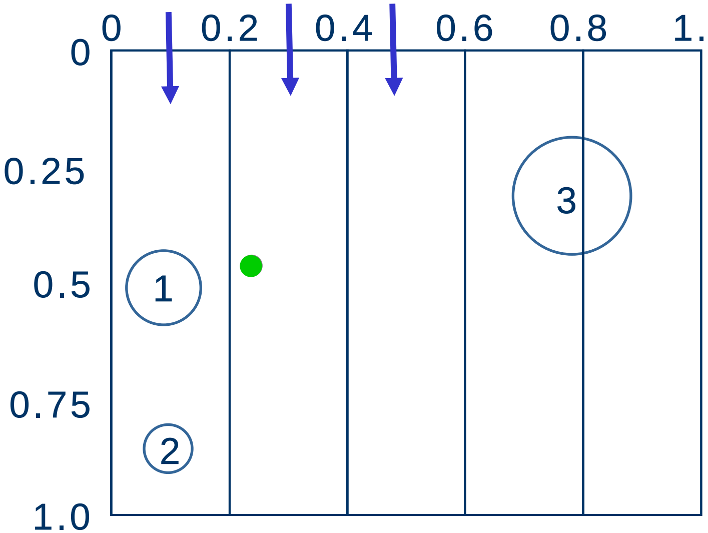


CF Tree

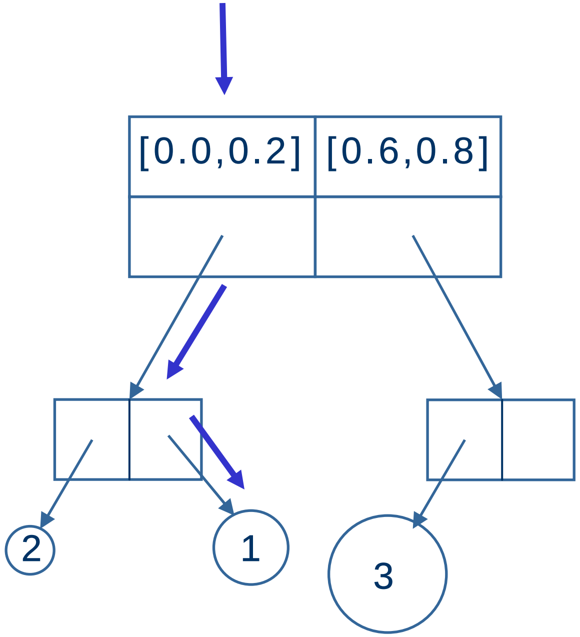


ADWICE-Grid: Detection (1)

Data Space

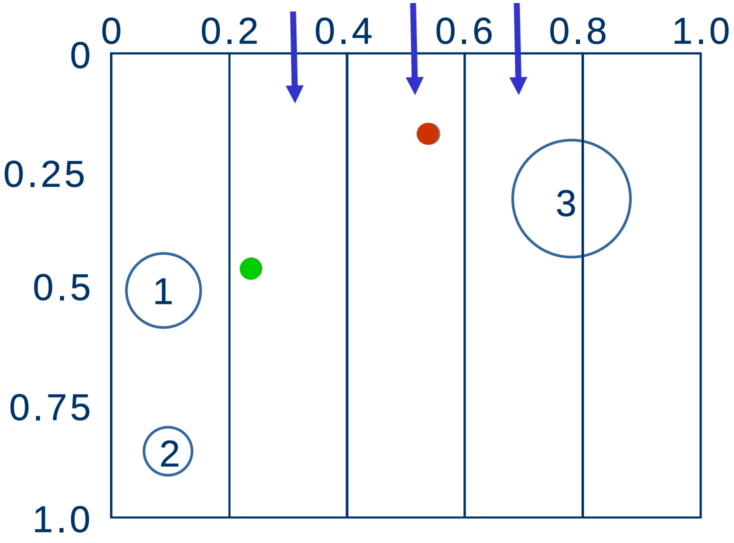


CF Tree

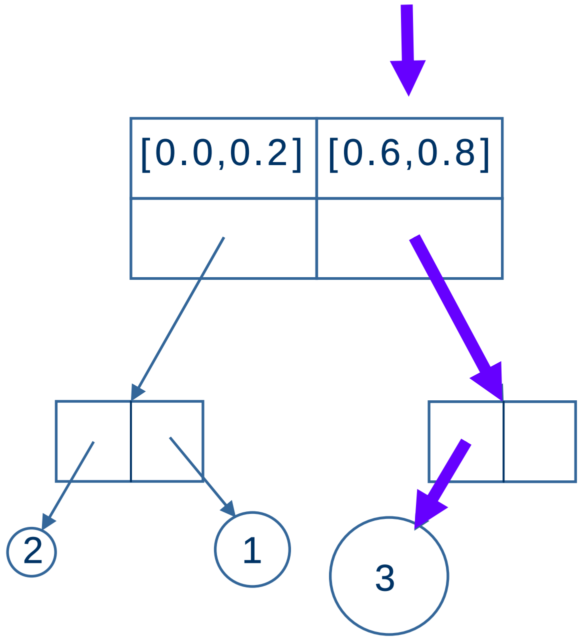


ADWICE-Grid: Detection (2)

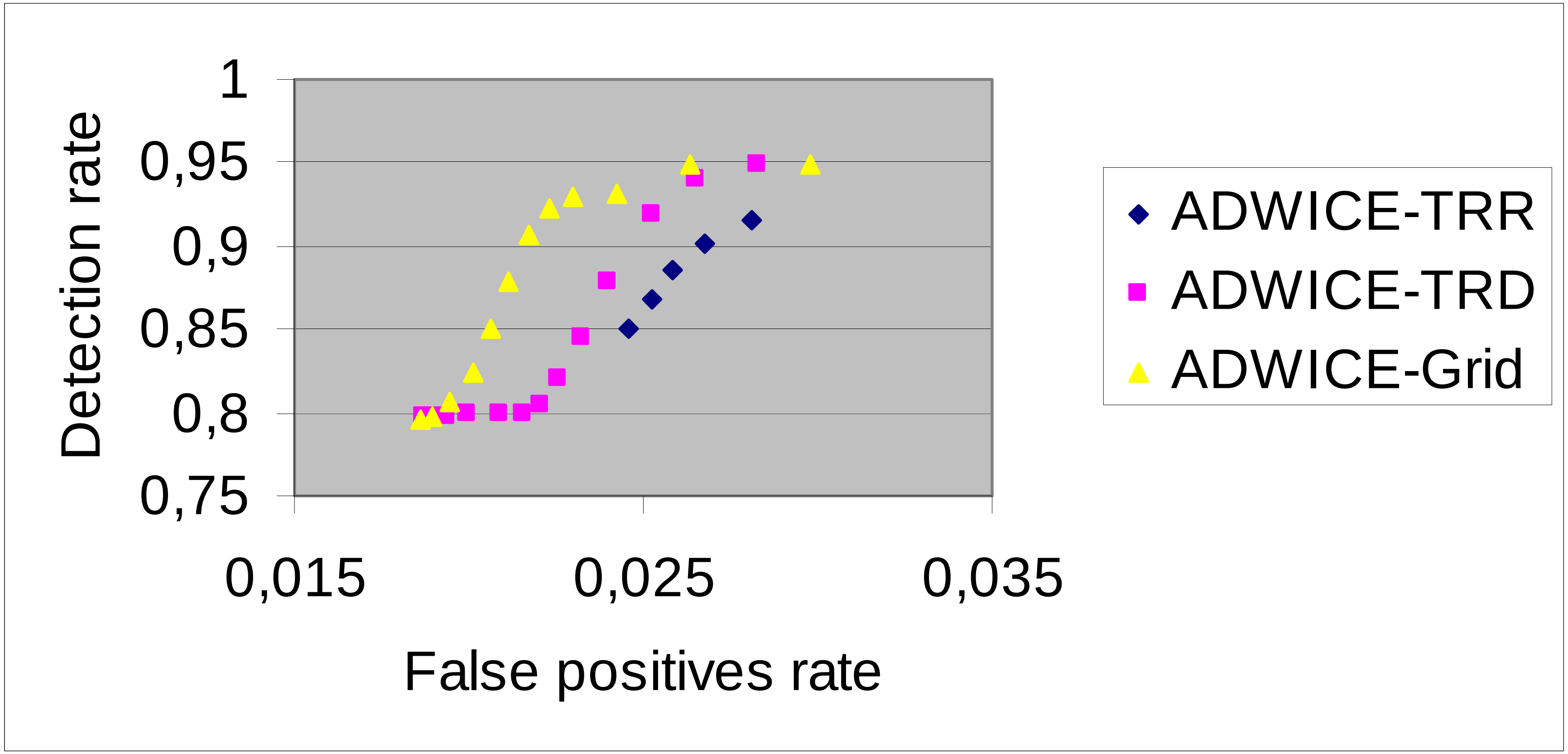
Data Space



CF Tree



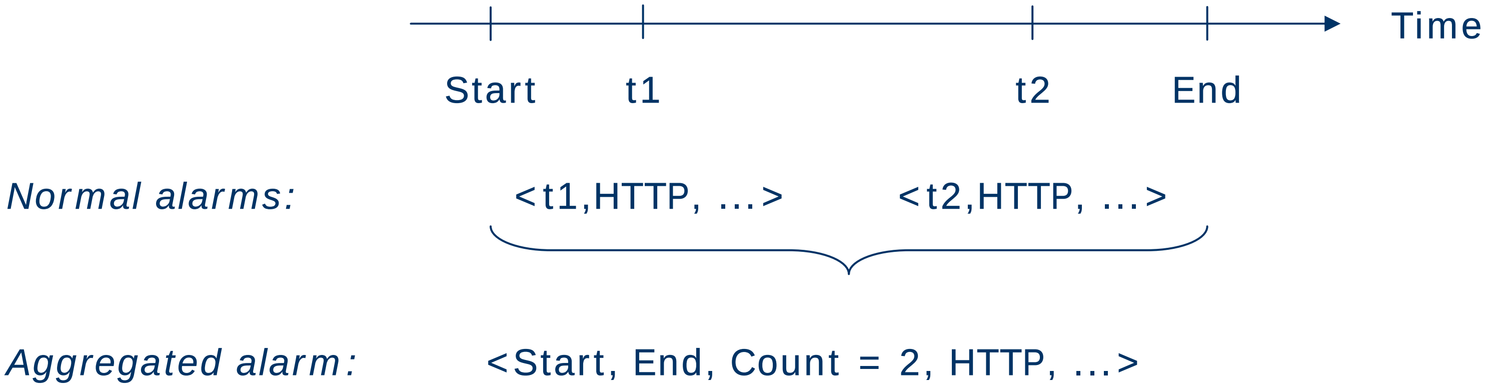
Detection rate vs. false positives



Source: [Burbeck & Nadjm-Tehrani 04,07]

Alarm aggregation

- Anomaly detection may produce many similar alarms (e.g. DoS, Probes, False positives)
- Similar alarms can be aggregated without losing accuracy



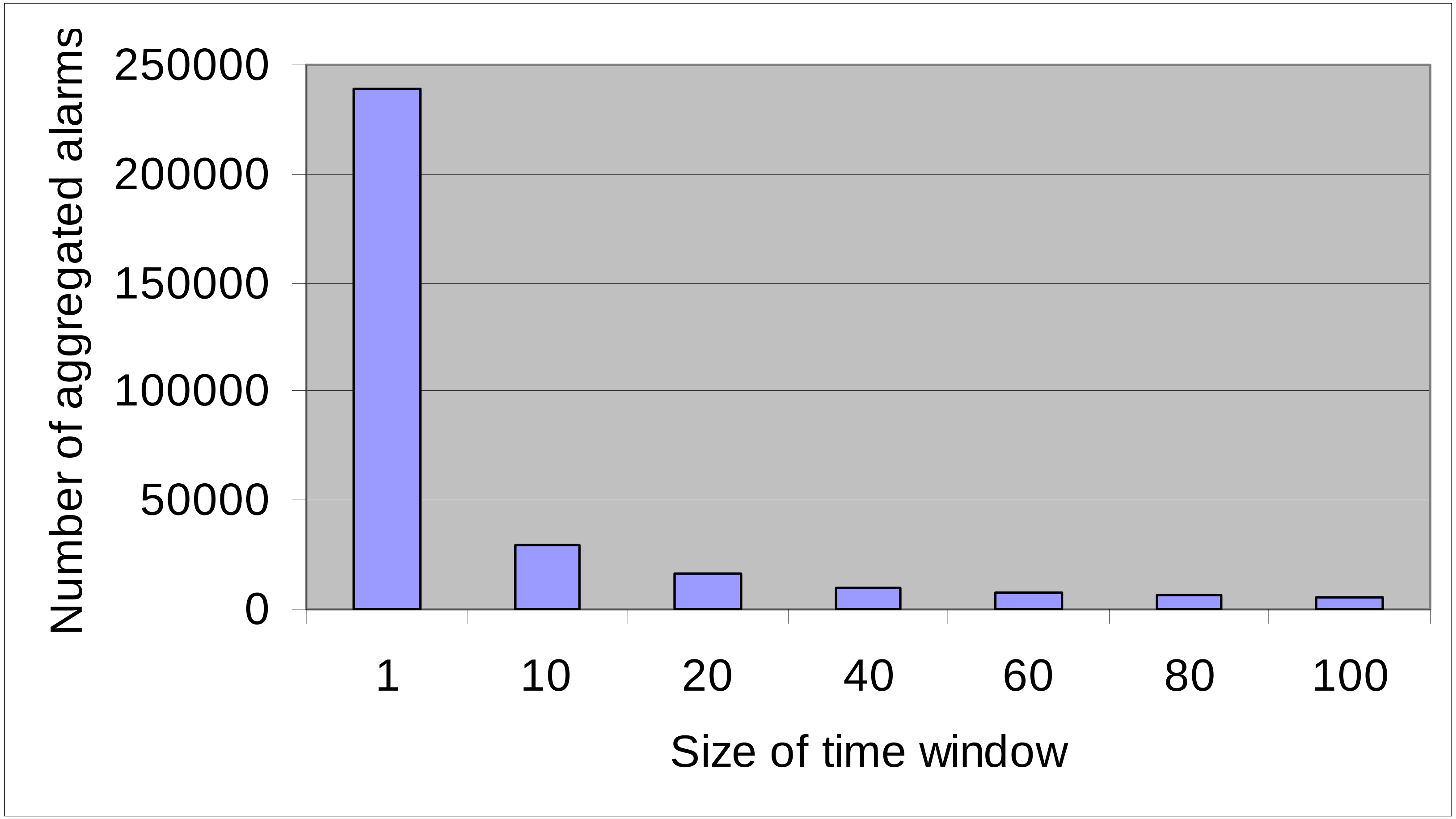
Time

Start t1 t2 End

Normal alarms: <t1,HTTP, ...> <t2,HTTP, ...>

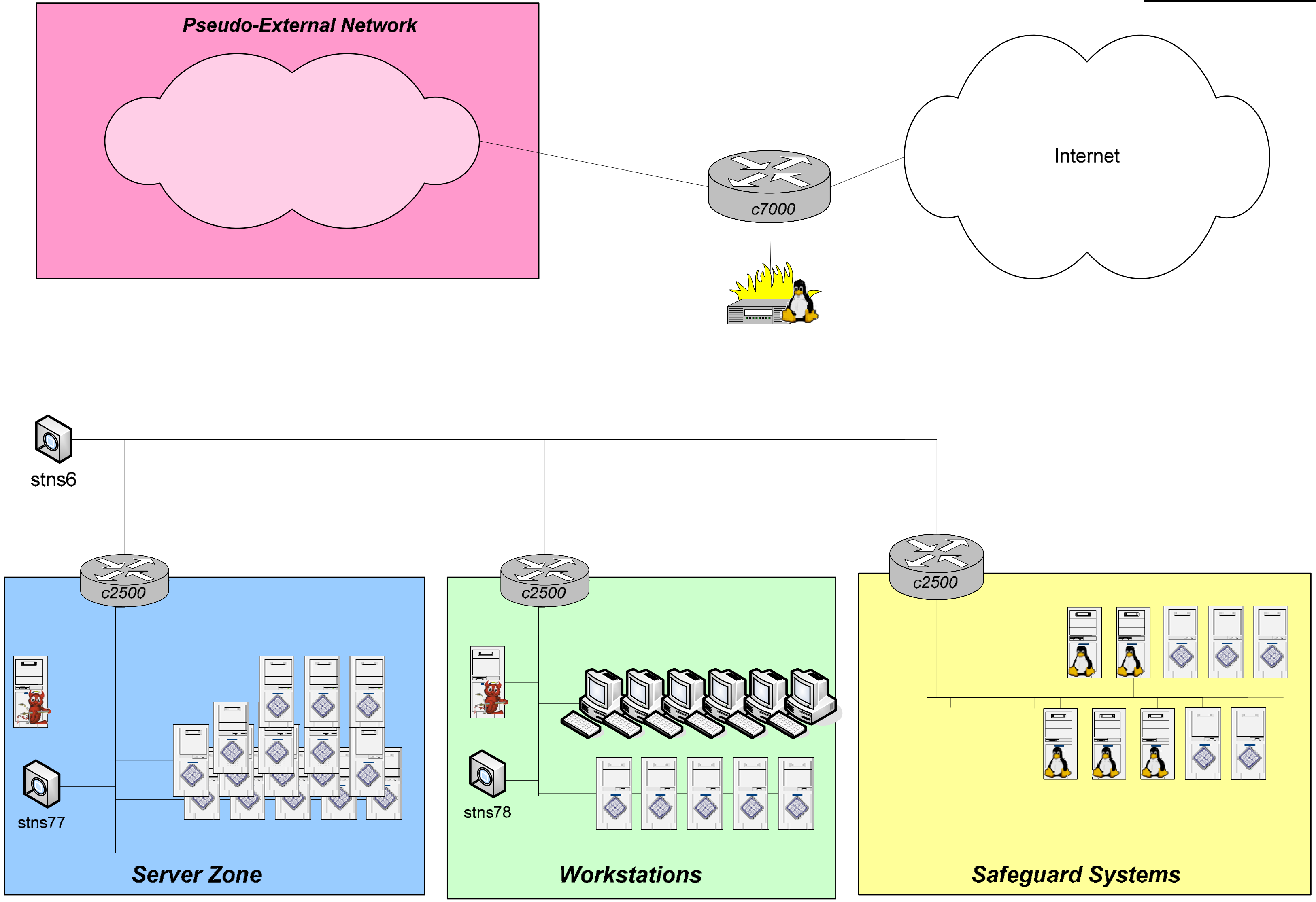
Aggregated alarm: <Start, End, Count = 2, HTTP, ...>

Alarm aggregation results

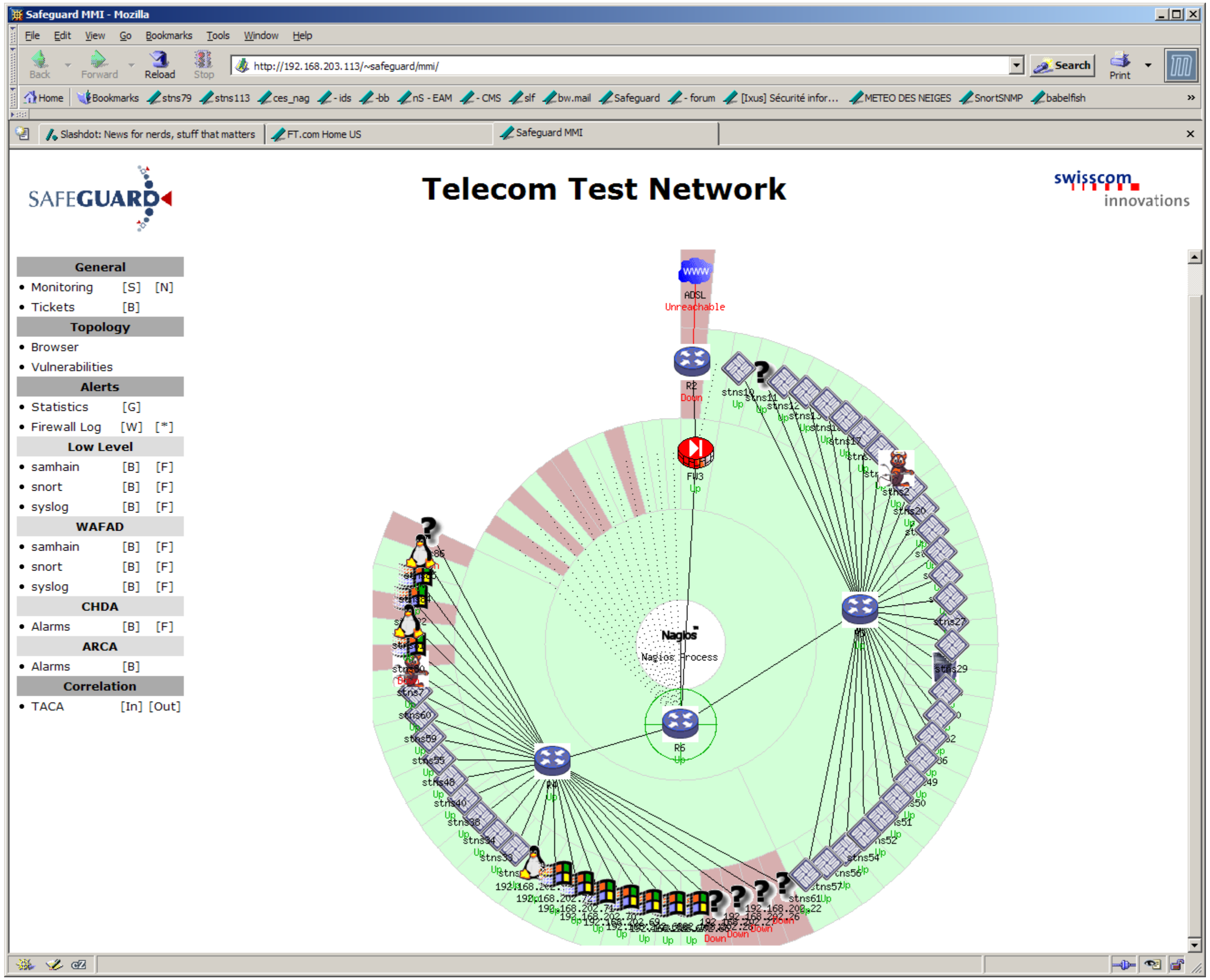


Safeguard 100+ test network

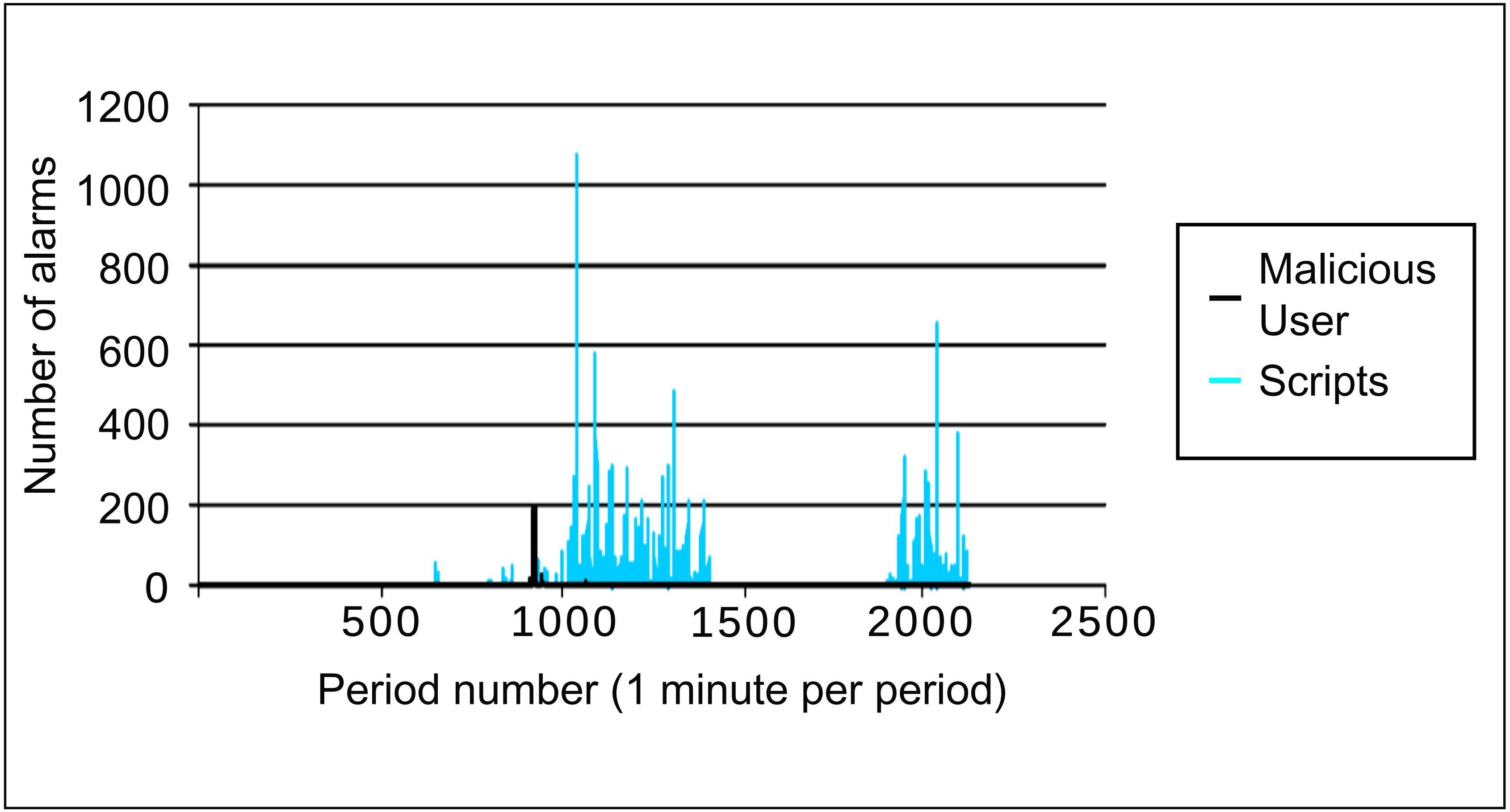
Safeguard



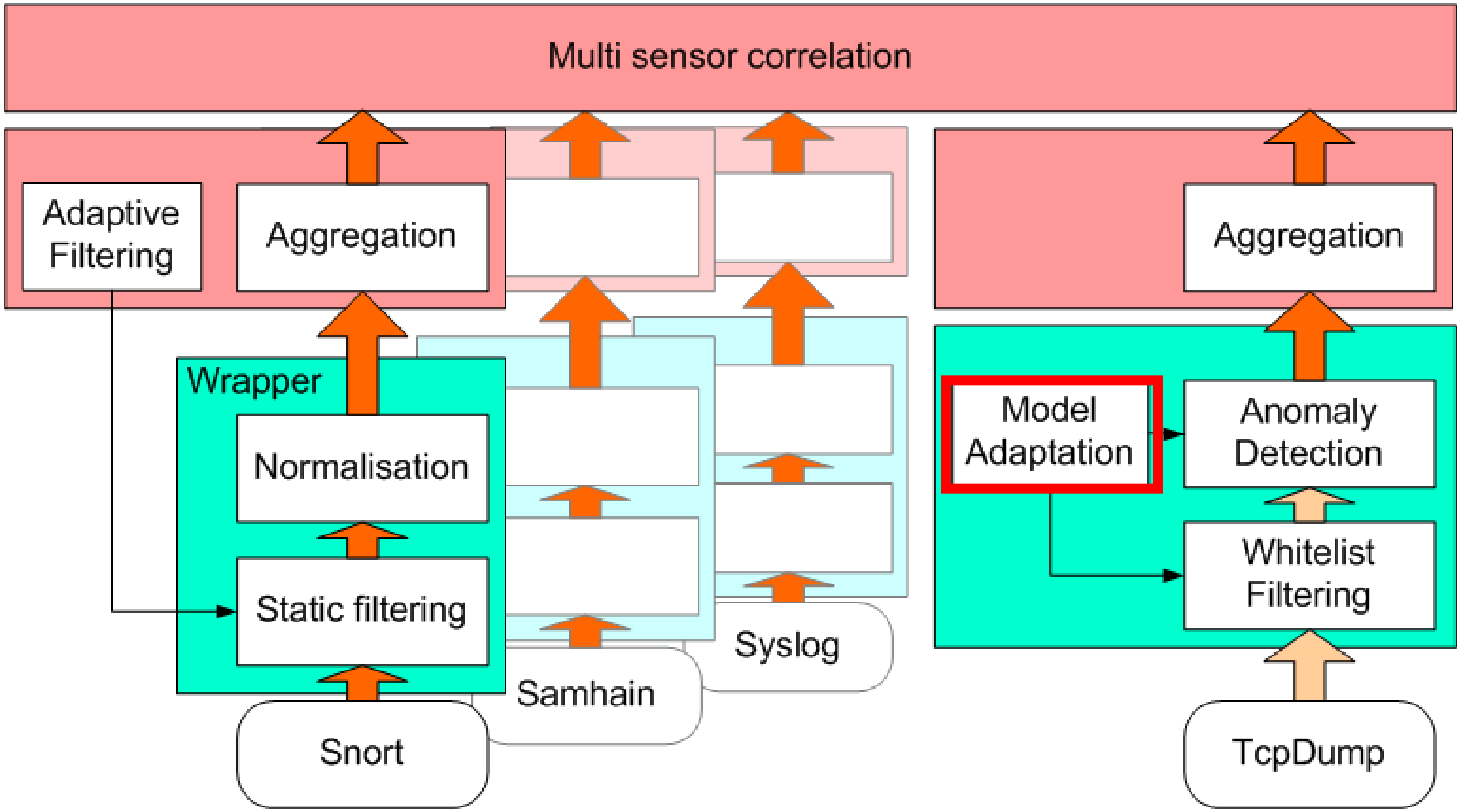
One HMI agent interface



A Safeguard scenario

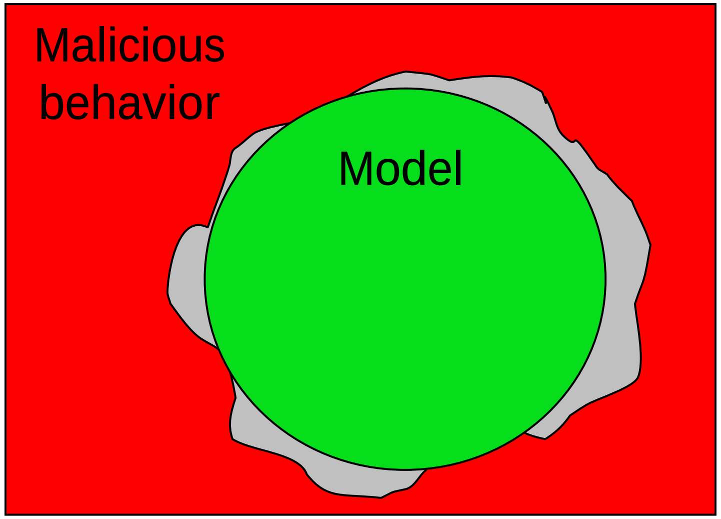


Correlating alarms



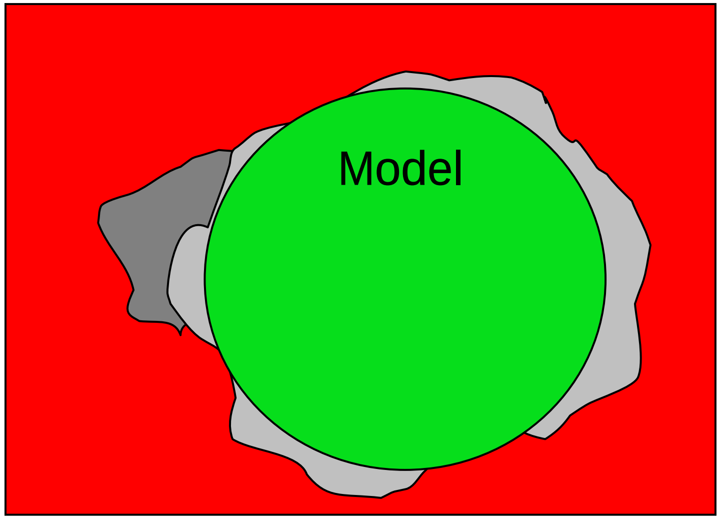
Need for normality adaptation

- Normality is not static!



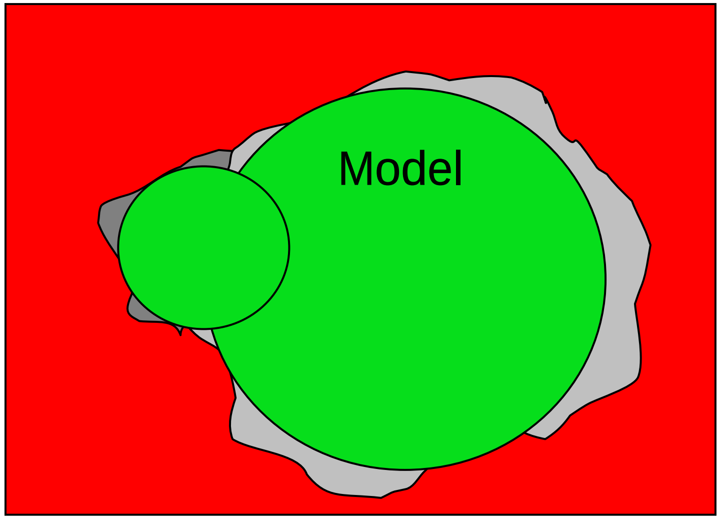
New cases of normality

- Normality changes
  - New type of normal behaviour
- Old model incomplete
  - Evaluation using KDD data gives ~300 false positives for new normality

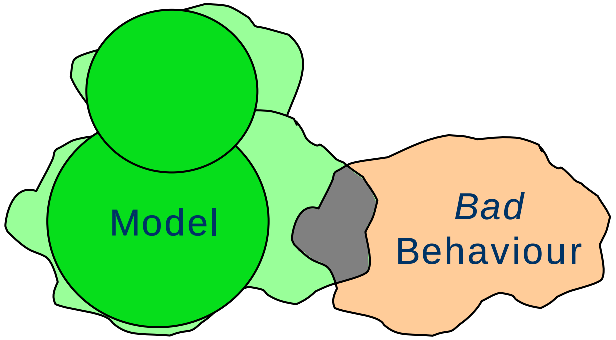


Evaluation of normality adaptation

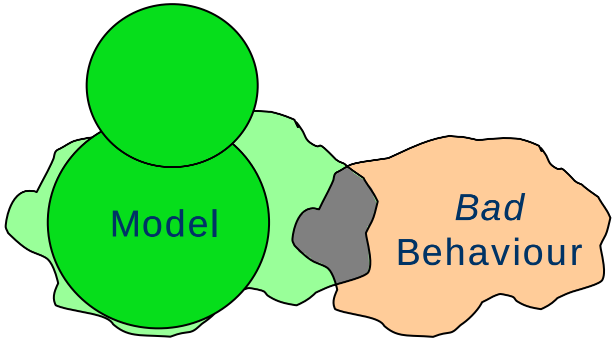
- Admin or system reacts
  - Recognize new false positives
  - Tells ADWICE to learn this behaviour
- Normality model adapted
  - From 300 to 3 false positives!



Forgetting



- System keeps track of model usage
  - If time since last usage is very long for subset of clusters
  - Decrease size (influence) of those clusters and finally remove them if not used



Lessons Learnt

Safeguarding critical infrastructures needs:

- Adaptive elements
- Incremental and scalable algorithms
- High performance for large volume of data

- Demonstration on realistic test beds
  - Research on open data sets :-)

- Understanding and mitigating interdependencies

Current track

- Application of ADWICE in anomaly detection for water management systems
  - Cooperation with Environment Protection Agency (EPA), USA
  - Time series data from simulated water system over an interval of one week

- Talk to me if interested to join!