How data analytics can be used to improve the network performance: some use cases

Luis Velasco
lvelasco@ac.upc.edu

Outline

Three use cases will be presented where the use of **data analytics** helps to **improve the network performance**:

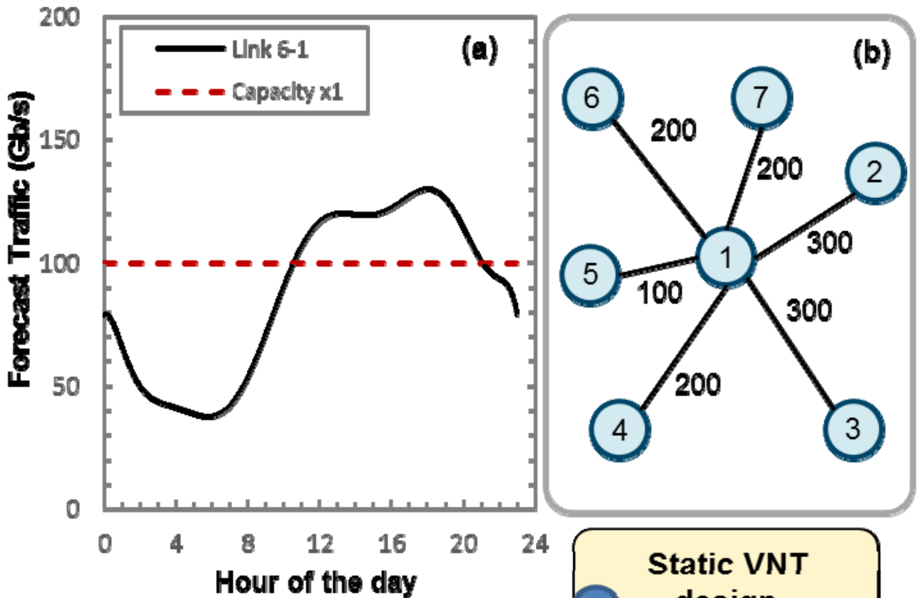
1. **Traffic prediction** that is used to reconfigure the VNT topology for the next traffic conditions.
2. Techniques to detect **traffic anomalies** in packet networks.
3. **Localization of failures** in the optical layer.

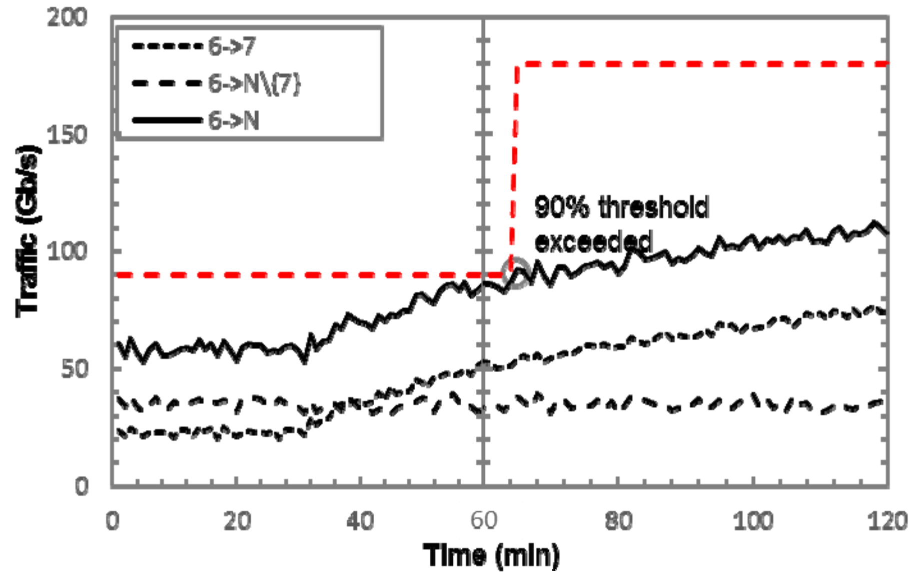
Use case 1: VNT topology reconfiguration for next traffic conditions

- Static VNT topologies are commonly designed to cope with the traffic forecast, which entails large overprovisioning.
- We propose a VNT reconfiguration algorithm based on traffic prediction.
- Origin-Destination (OD) traffic needs to be monitored in the packet nodes.

F. Morales et al., "Virtual Network Topology Adaptability based on Data Analytics for Traffic Prediction," *IEEE/OSA Journal of Optical Communications and Networking (JOCN)*, vol. 9, pp. A35-A45, 2017.

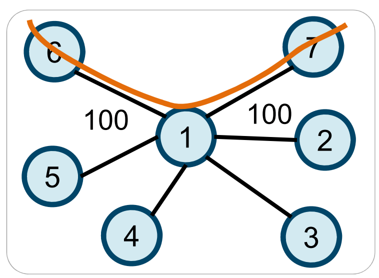
VNT Design and Reconfiguration Options

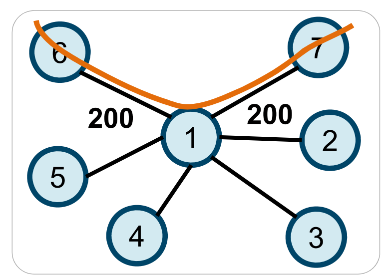




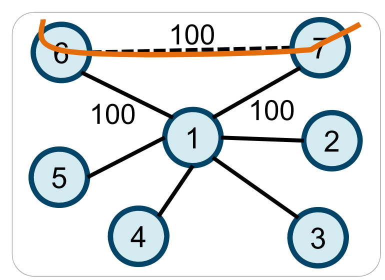
1 Static VNT design

Current VNT





2 Threshold-based reconfiguration

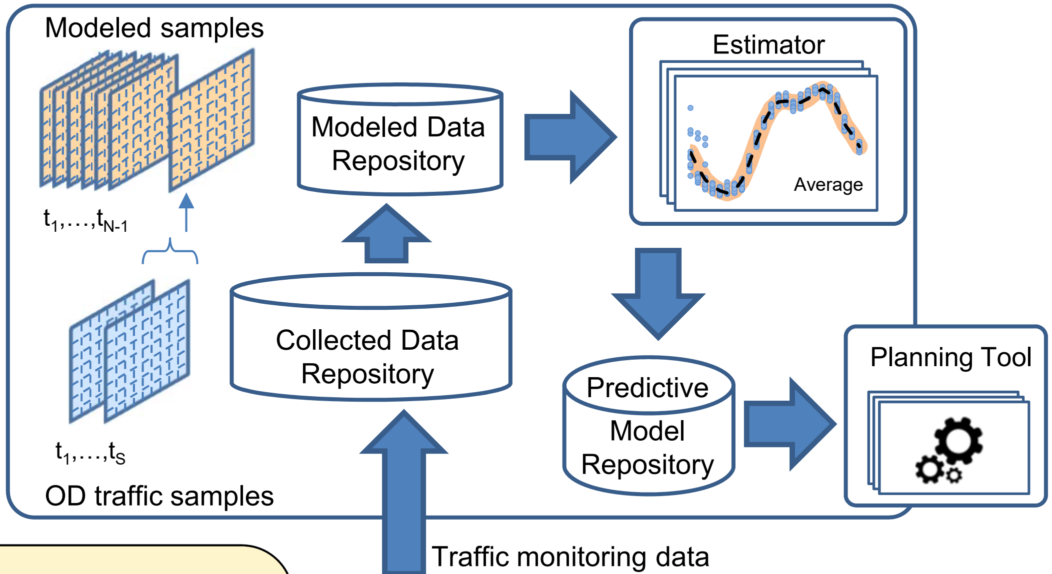


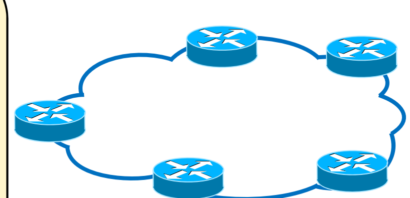
3 Reconfiguration based on OD traffic prediction

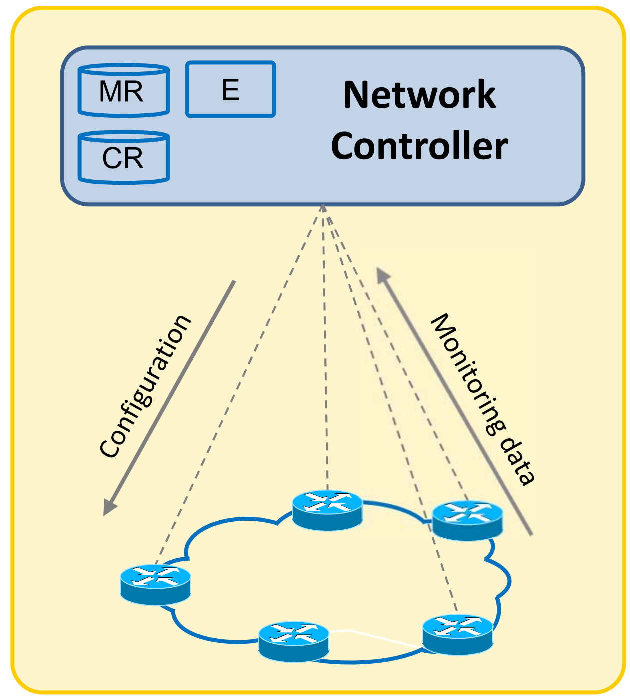
Monitoring and Data Analytics

2 Collected data is summarized by computing the average

3 Estimator: estimates the specific models (two response variables: **μ** and **σ**) for every OD pair







1 Traffic monitoring collected from the network nodes is stored in a centralized repository

The VENTURE Problem

Given:

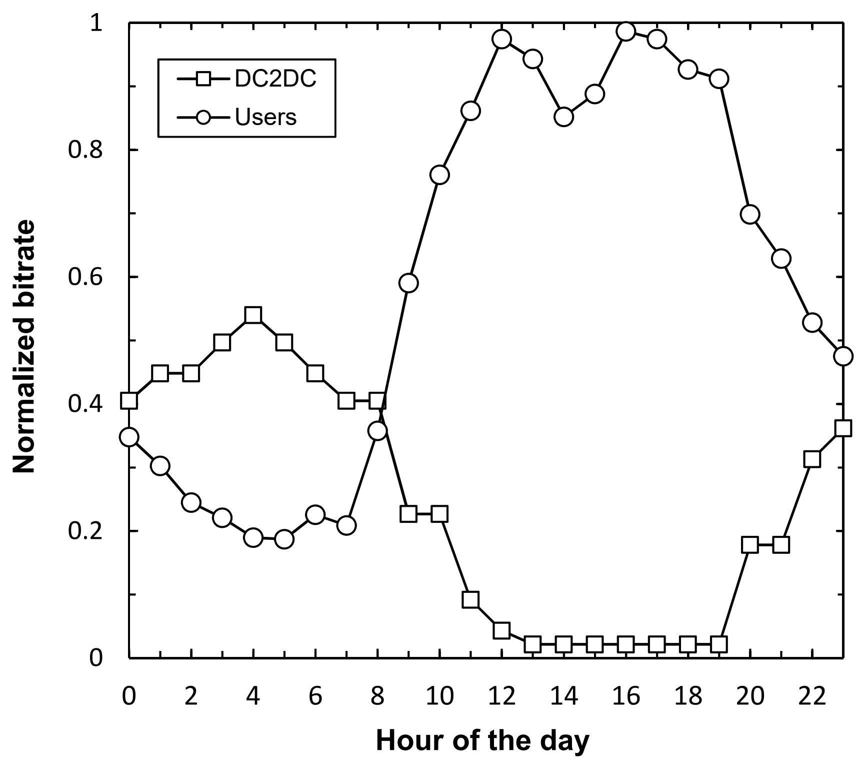
- The current VNT represented by a graph G(*N*, *E*'), being *N* the set of routers and $E' \subseteq E$ the set of current vlinks.
- The set *P* with the transponders available in the nodes; every transponder with capacity *B*.
- The current traffic matrix *D*.
- The predicted traffic matrix *OD*.

Output: The reconfigured VNT $G^*(N, E^*)$, where $E^* \subseteq E$, and the paths for the traffic on G^*.

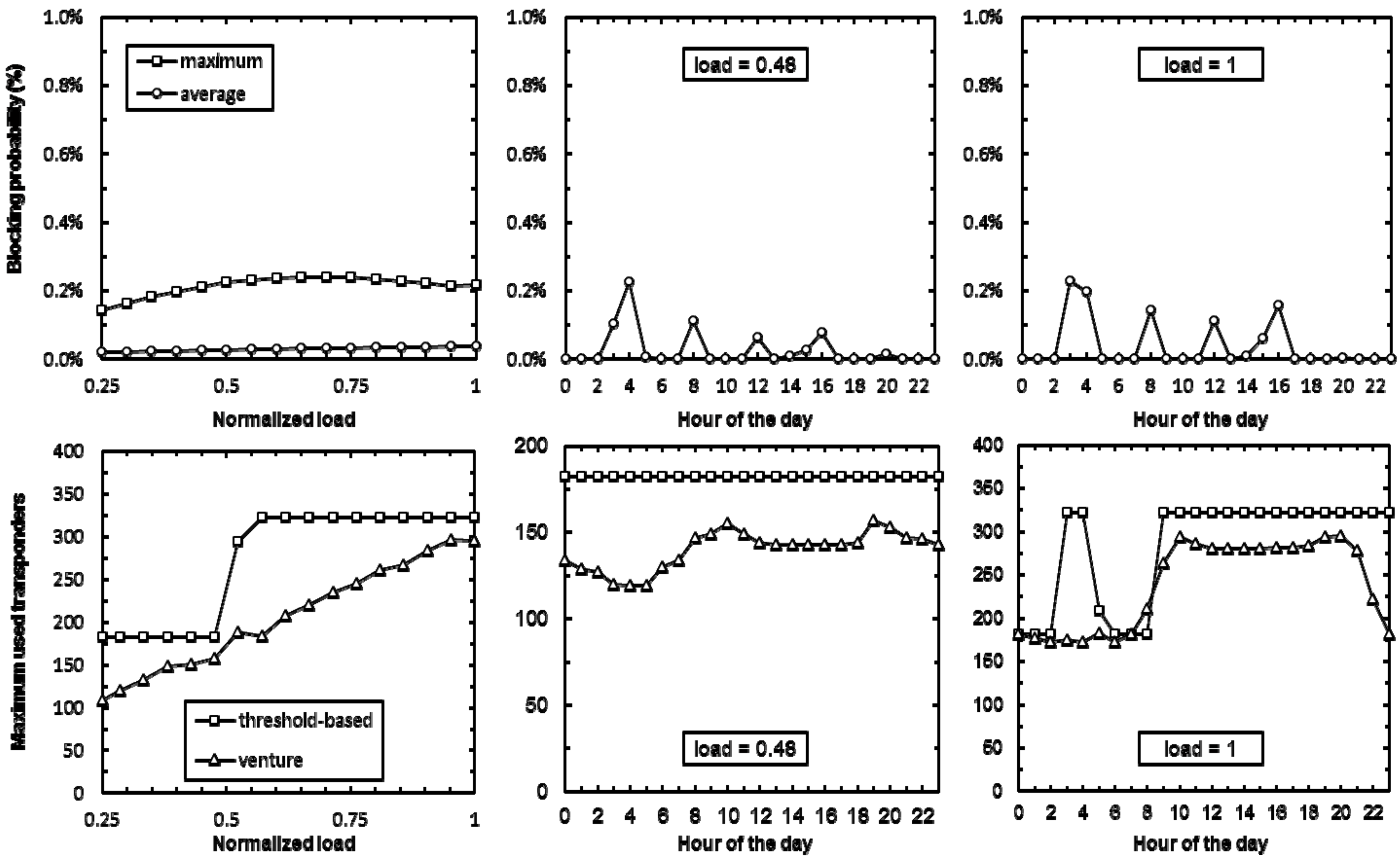
Objective: Maximize current and predicted served traffic matrices, whilst minimizing the total number of transponders used.

Simulation

- Comparison:
  - VENTURE.
  - Threshold-based VNT reconfiguration (full-mesh VNT).
- Scenario: 14-node VNT with 26 TPs in every node.
- Dynamic traffic profiles and loads considered.
- Traffic changes in **intensity** and in **direction**.



Results

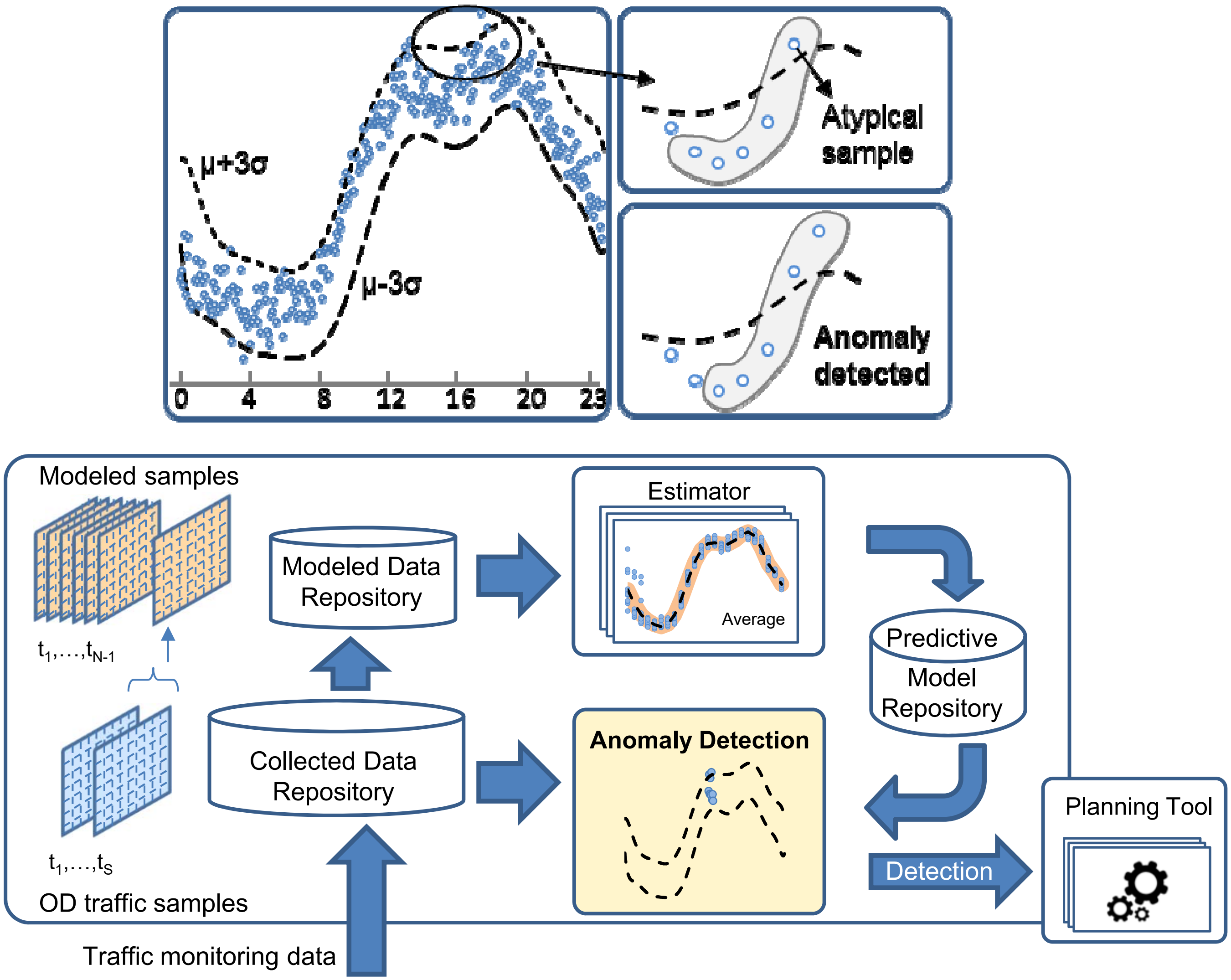

Blocking probability (%)
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
maximum
average
0.25
0.5
0.75
1
Normalized load
load = 0.48
load = 1
0 2 4 6 8 10 12 14 16 18 20 22
Hour of the day
Maximum used transponders
400
350
300
250
200
150
100
50
0
threshold-based
venture
Normalized load
load = 0.48
load = 1
Hour of the day

Use case 2: Traffic anomalies detection

- Traffic anomalies are hard to detect, but its prompt detection helps to avoid traffic congestion by preparing the network:
  - by modifying routing tables, or
  - by reconfiguring the VNT.
- ITU-T specifies 15 min monitoring period → high anomaly detection times.
- Study the performance of OD traffic anomaly detection methods and find the best architectural approach.

A. P. Vela et al., "Bringing Data Analytics to the Network Nodes," in Proc. *European Conference on Optical Communication (ECOC)*, 2016.

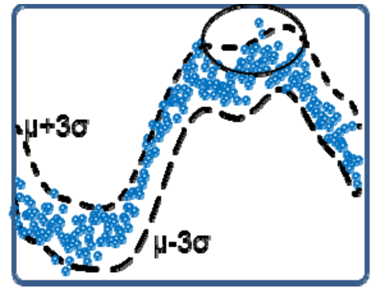
OD Traffic Anomaly Detection



Methods for Traffic Anomaly Detection

Threshold-based

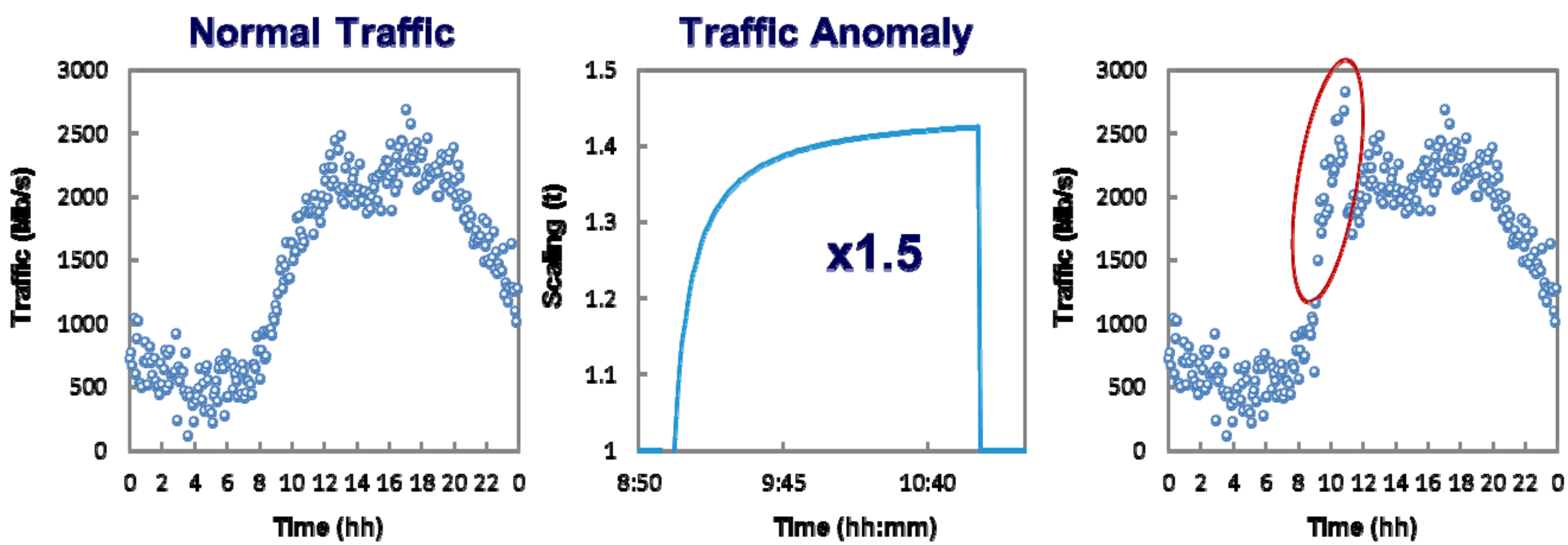
- Detecting anomalies after receiving a certain number of out-of-bound traffic samples with respect to: μ ±3σ



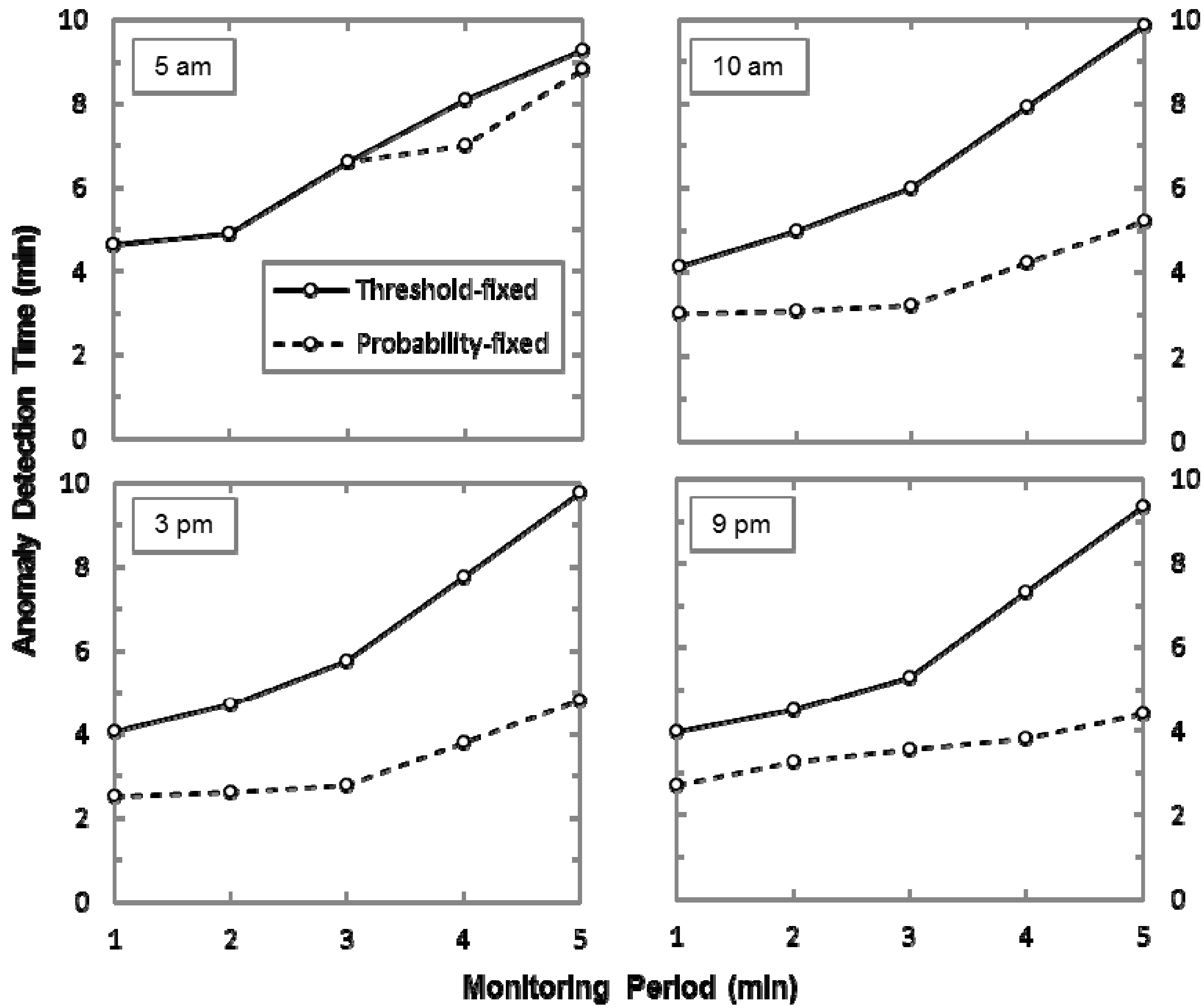
Probability-based

- Self-learning classifier with two labels for the response: *Normal/Anomaly*
- *w*(*H*): distance function → how likely is that feature vector *H* belongs to normal class
- w_{thr}: threshold distance that if *H* is normal never exceeds

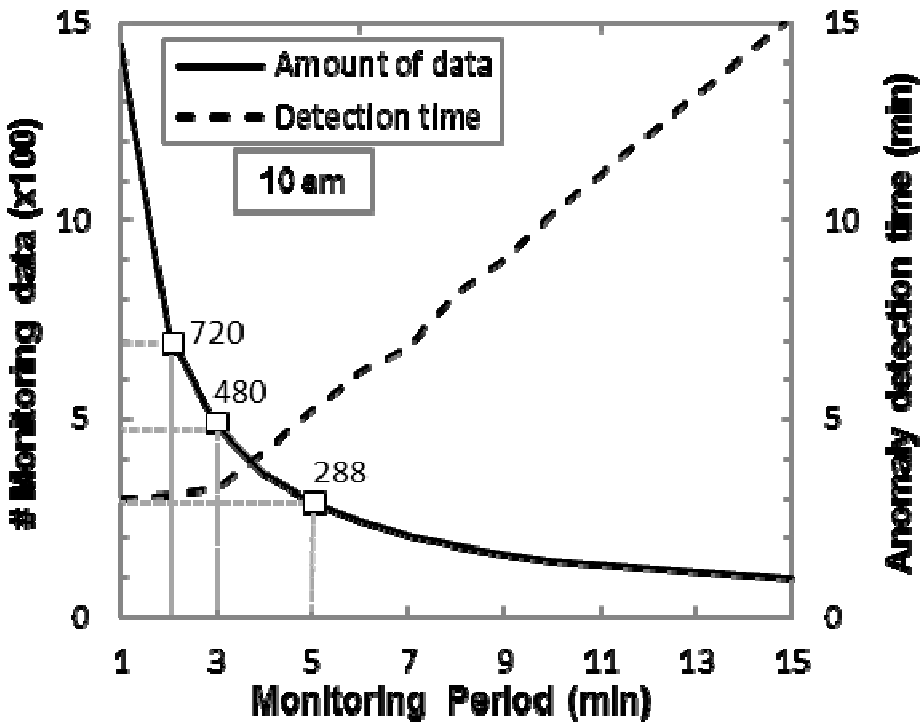
Traffic and anomaly generation



Anomaly detection time vs Monitoring period

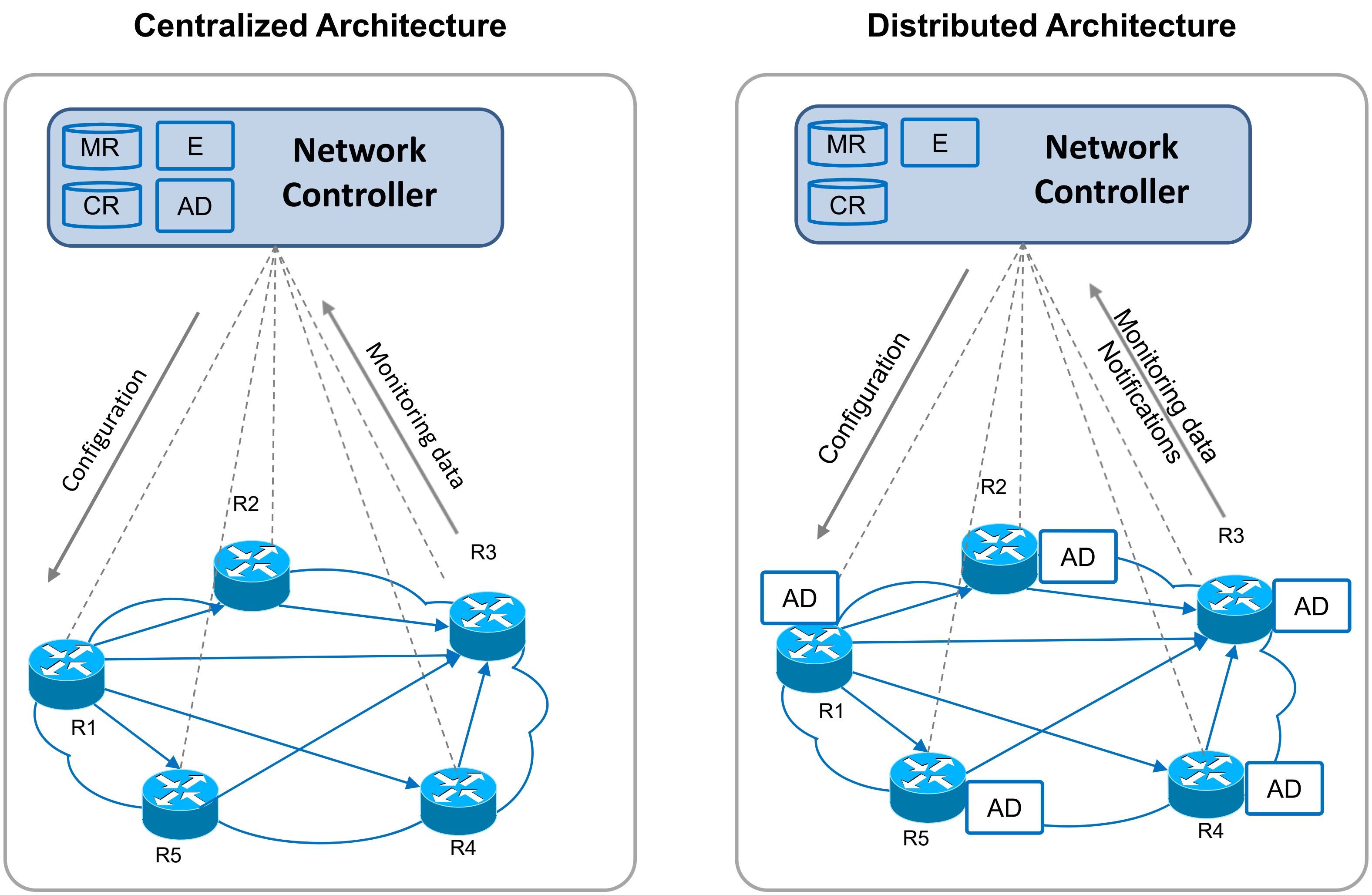


Detection time and monitoring data vs Monitoring period



- Reducing the monitoring period exponentially increases the amount of collected data:
  - E.g., 2 min period entails 720 samples per OD (|*N*|*(|*N*|-1) ODs) and day to be collected and stored.

Bringing Data Analytics to the Network Nodes

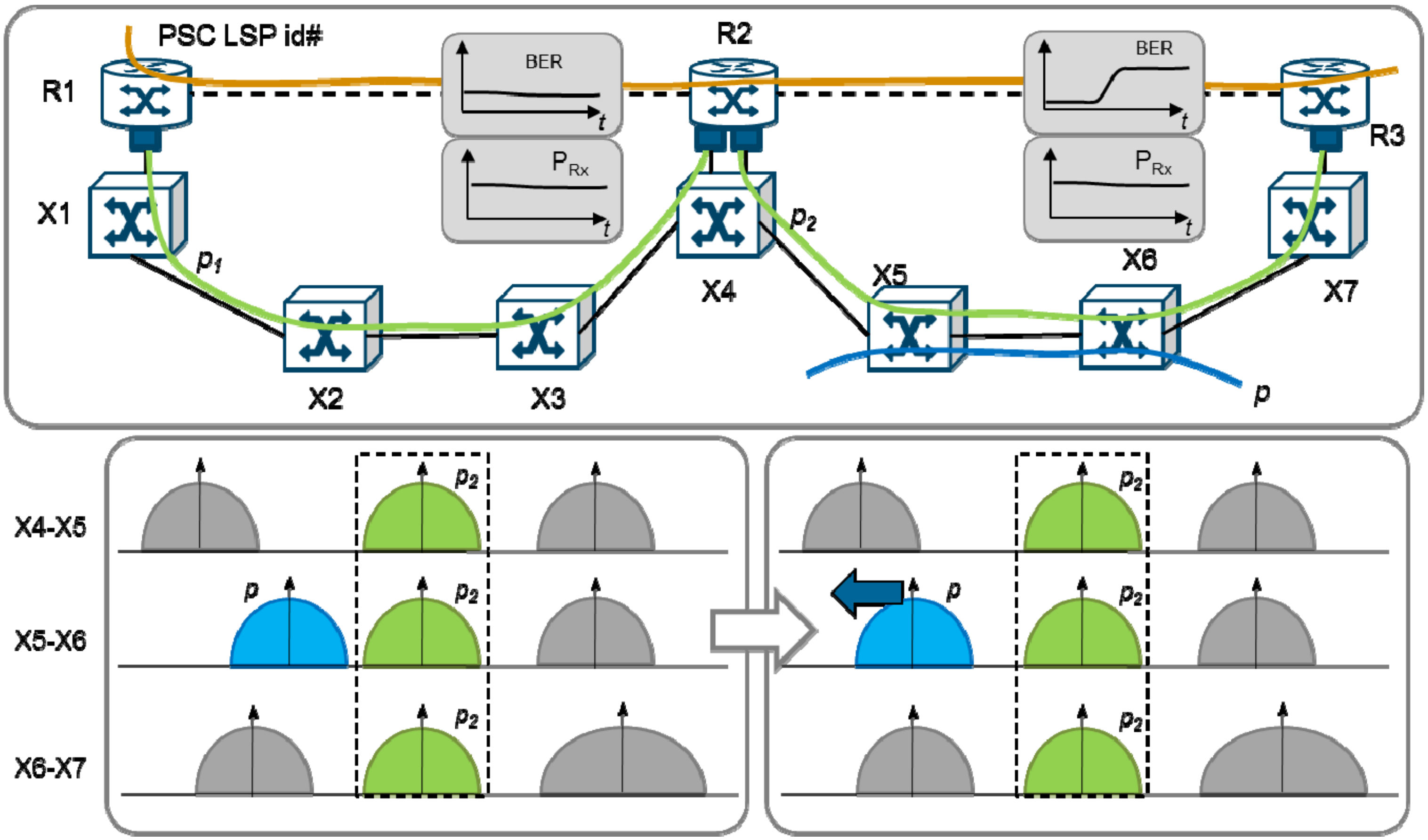


Use case 3: Failure Localization/Identification in the Optical Layer

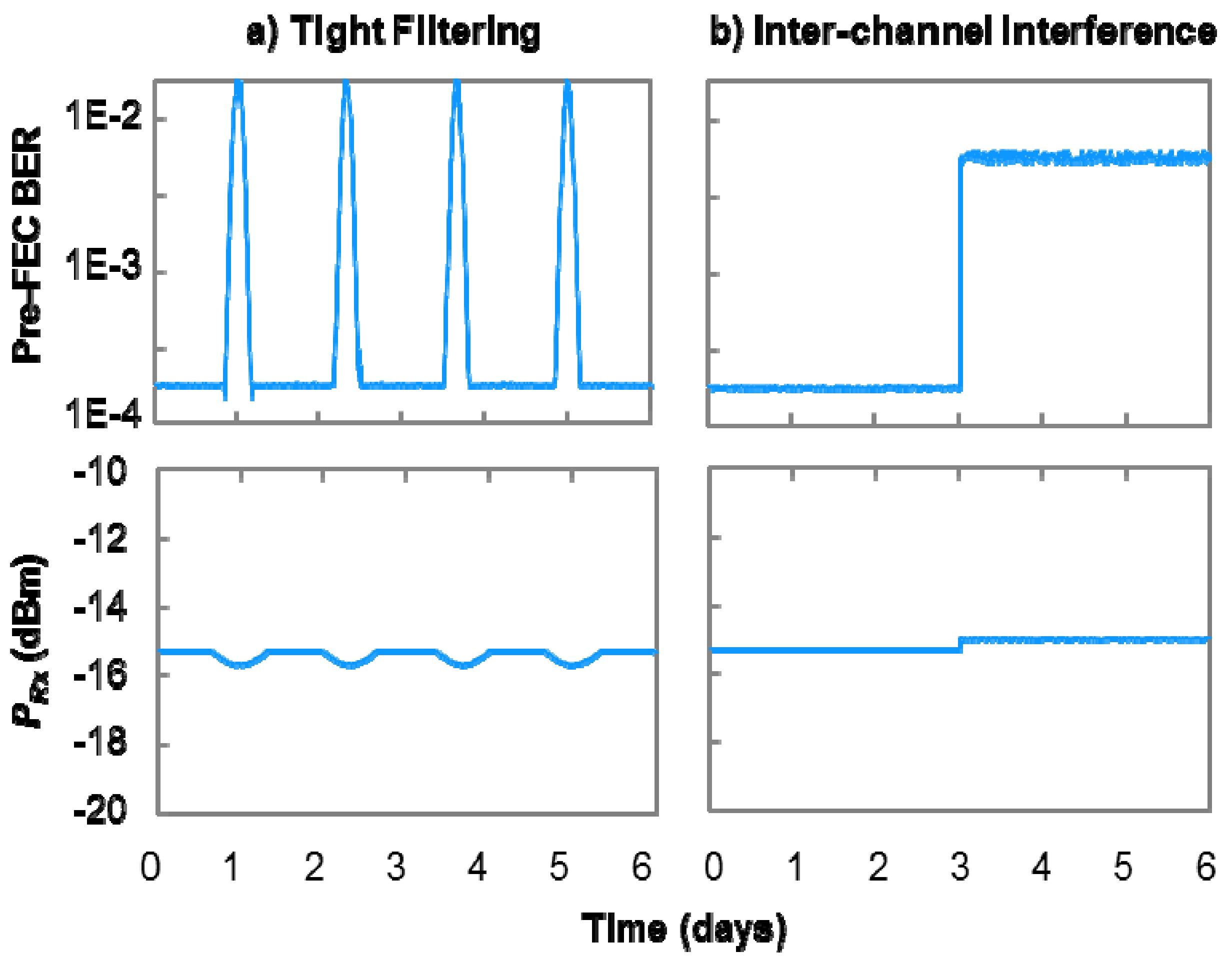
- *vlinks* in VNTs are ultimately supported by lightpaths.
- Errors in lightpaths translate into errors in service flows that might cause packet losses and retransmissions leading to unacceptable QoS.
- Monitoring the physical layer is key to verify SLA fulfillment and to localize the case of faults or degradations.
- We study the effects on QoT monitoring parameters of two failures in the optical layer and propose an effective ML algorithm.

M. Ruiz et al., "Service-triggered failure identification/localization through monitoring of multiple parameters," in Proc. *European Conference on Optical Communication (ECOC)*, 2016.

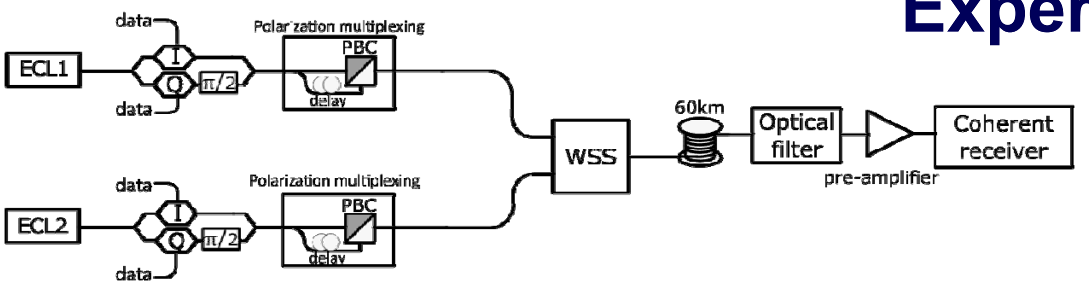
Failure localization algorithm

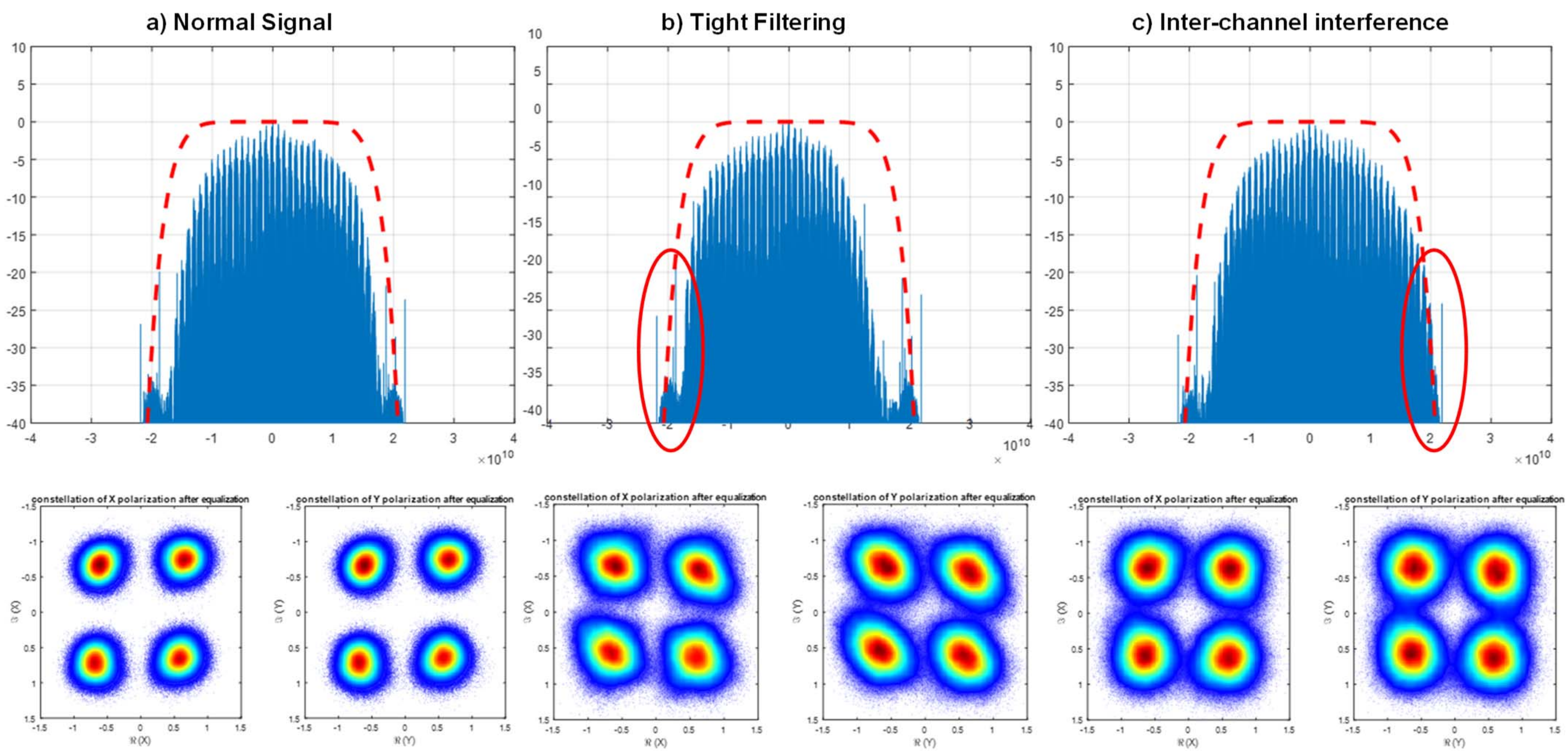


Considered failures

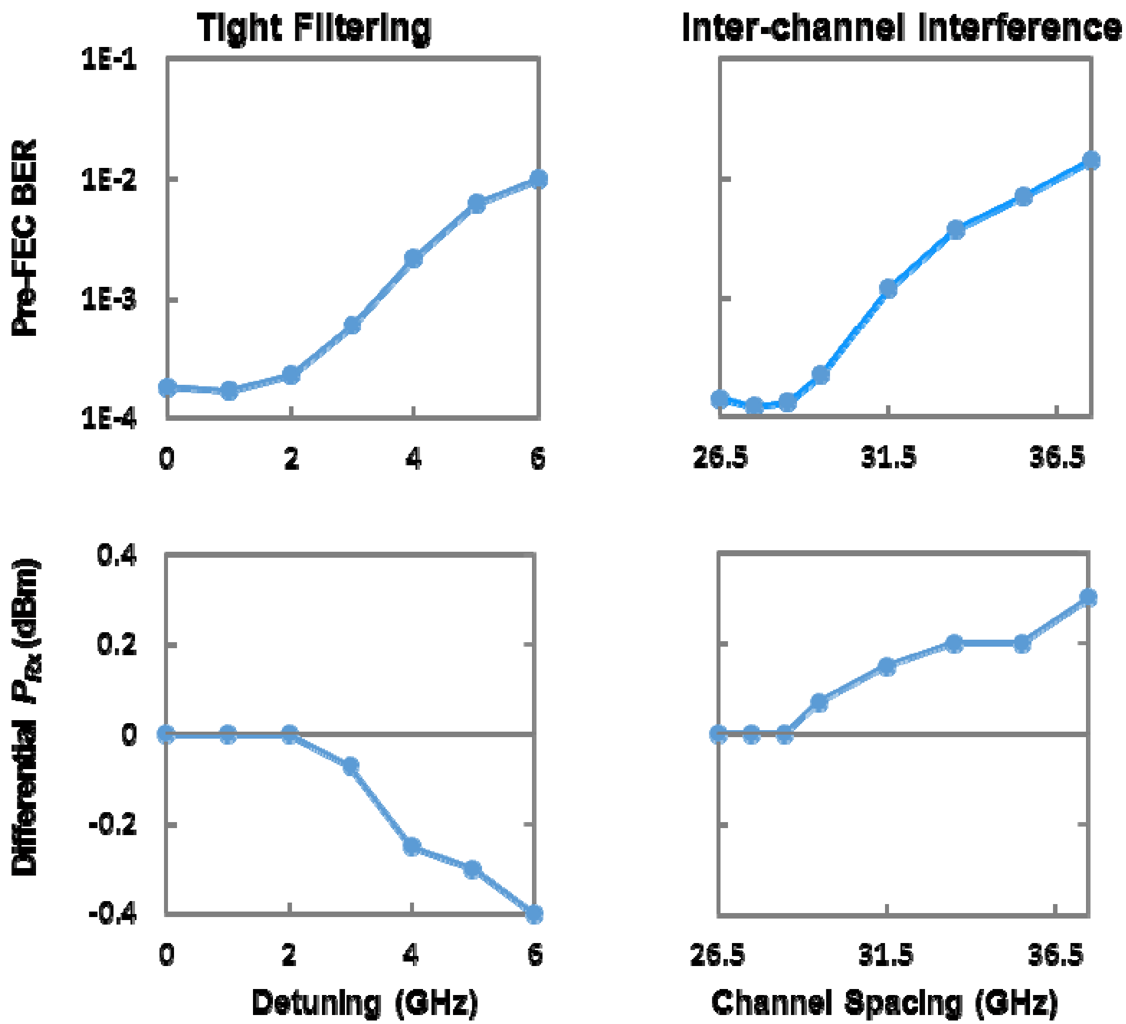


Experimental results (CNIT)





Experimental results (CNIT)



Experimental results for normal and degraded conditions

		Real		
		Normal	Filtering	Interference
Pre-diction	Normal	**99.2%**	0.8%	0%
	Filtering	0%	**100%**	0%
	Interference	0%	0%	**100%**

Goodness-of-fit computed as the probability that the BN predicts the actual failure cause as the first option

- Only 0,8% error observed in some tests where a normal signal was predicted instead of a tight filtering failure

Conclusions

Three use cases where the use of data analytics helped to improve the network performance have been presented:

1. An approach to **adapt** the current **VNT** to **future traffic**.
   - Data analytics are applied to OD monitored traffic data.
   - Traffic prediction is used as input to a VNT reconfiguration algorithm.
2. Different strategies for **OD traffic anomaly detection**.
   - 1 min monitoring period entails large amount of data being collected and stored in the centralized architecture.
   - We proposed bringing data analytics to the network nodes.
3. **Failure localization/identification** in the optical layer
   - Two failure causes have been characterized.
   - The obtained measurements were used to generate time series to train a BN to localize and identify the most probable cause of the errors.

Thank you for your attention !

How data analytics can be used to improve the network performance: Some use cases

Luis Velasco
lvelasco@ac.upc.edu