# ADAPTIVE ANOMALY DETECTION FOR SECURITY AND PERFORMANCE MONITORING

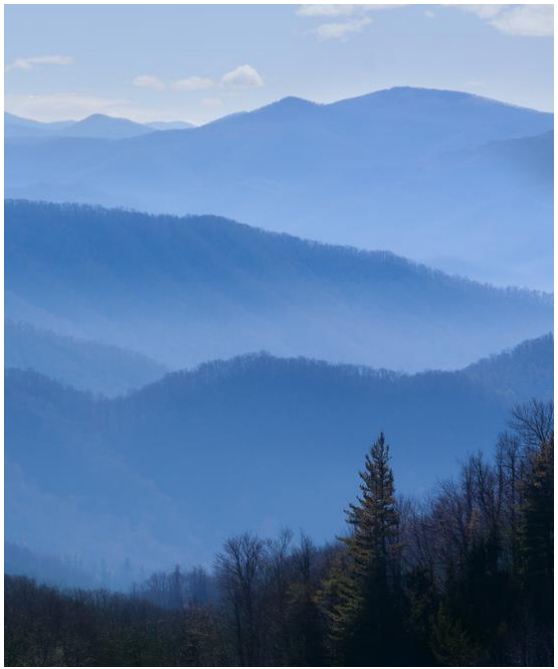
## AKA “SMOKY MOUNTAINS”

Pieter Venemans

TNO innovation for life

# ABOUT THE PRESENTATION

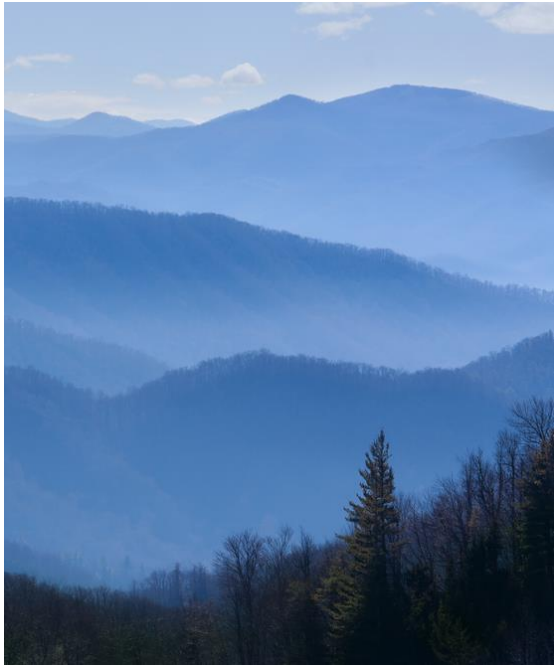
- A good example of how TNO could help a customer to innovate and improve their products with our scientific (mathematical) knowledge and experience in the field of security monitoring and detection
- Algorithms work with data that was already available at Netdialog and were adapted to the customer's needs. Implementation of future work may require modifications in the Netdialog software.
- After the introduction by TNO, NetDialog will tell about their experiences with the results.



NET DIALOG

# MOTIVATION

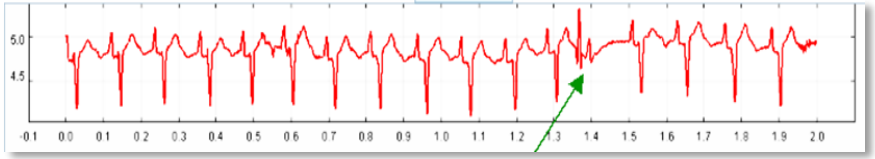
- Many security and performance problems can be detected in an early stage by monitoring the right parameters
  - Traffic volumes, CPU loads, error counts, number of firewall / IDS alerts, number of login attempts etc
- Practical problems:
  - Amount of data is too much to handle manually / by eye
  - Data patterns are often too complex for simple thresholds
- Wish: a robust monitoring and detection algorithm that mimics the human eye:
  - Recognize repeating patterns (day/night, workday/weekend)
  - Understands the normal amounts of uncertainty
  - Produce alerts for unexpected observations
  - Does not require much tuning and tweaking



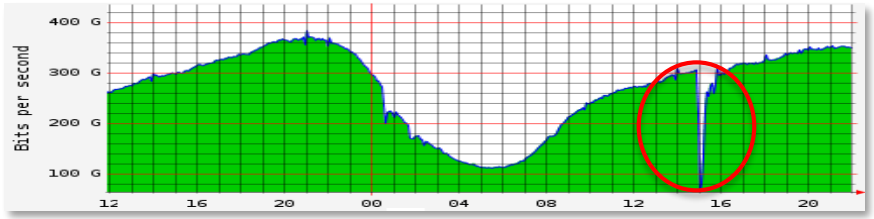
# ANOMALY DETECTION IN TIME SERIES

- Observe quantities over time, construct some model and detect observations that do not match with our expectations
- Deviations could be caused by:
  - equipment failures,
  - overload,
  - operating errors,
  - malware,
  - fraud,
  - cyber attacks etc.

Human heart beat (ECG)

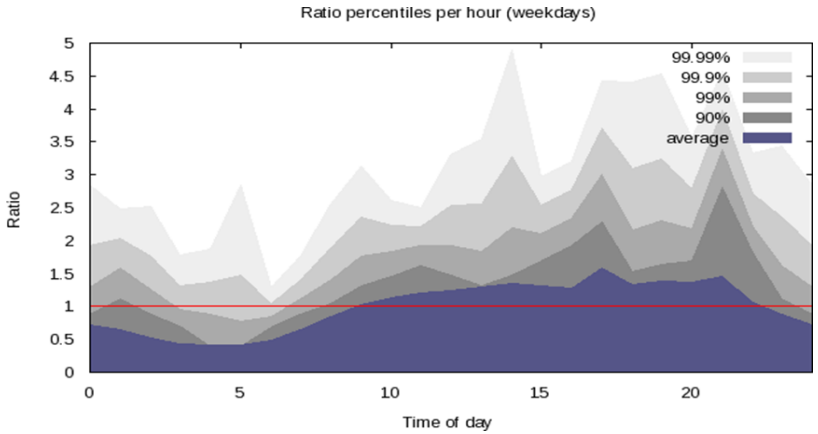



AMSIX traffic



# STOCHASTIC MODELS FOR TIME SERIES

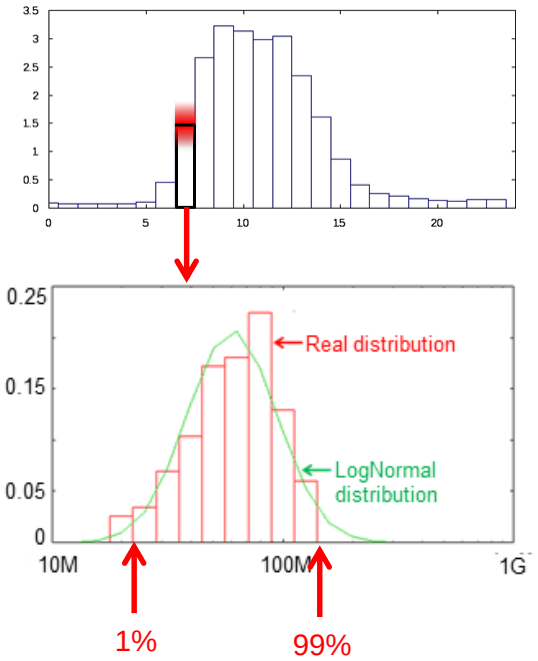
- Network traffic (and other time series) are often non-deterministic (unpredictable)
- But certain statistical properties of the traffic can often be modelled with time-invariant model parameters.
- Distribution of traffic values is often “long-tailed”.
- A good model allows us to do forecasting as well as anomaly detection: data that doesn’t match with the prediction may be a symptom of a problem.



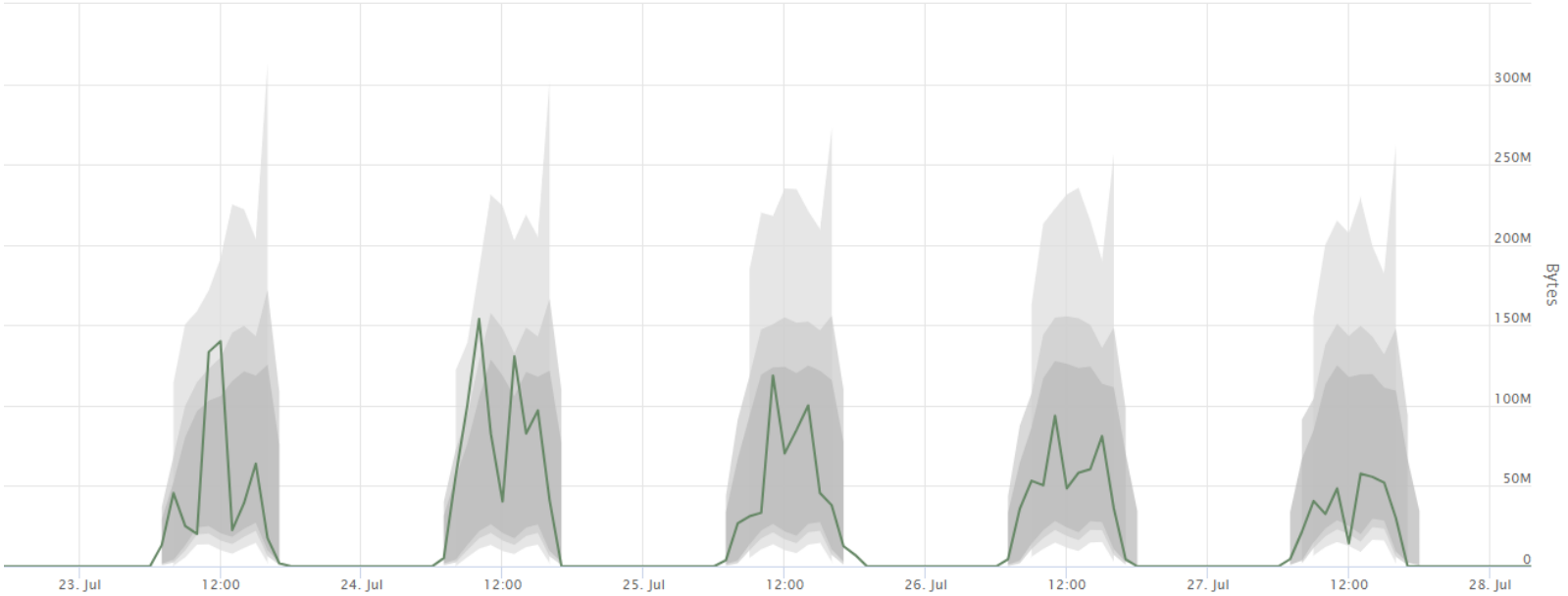

*Traffic during a day, including (extrapolated) upper percentiles of a Log-Normal (long-tailed) distribution of peak values*

# SMOKY MOUNTAINS ALGORITHM

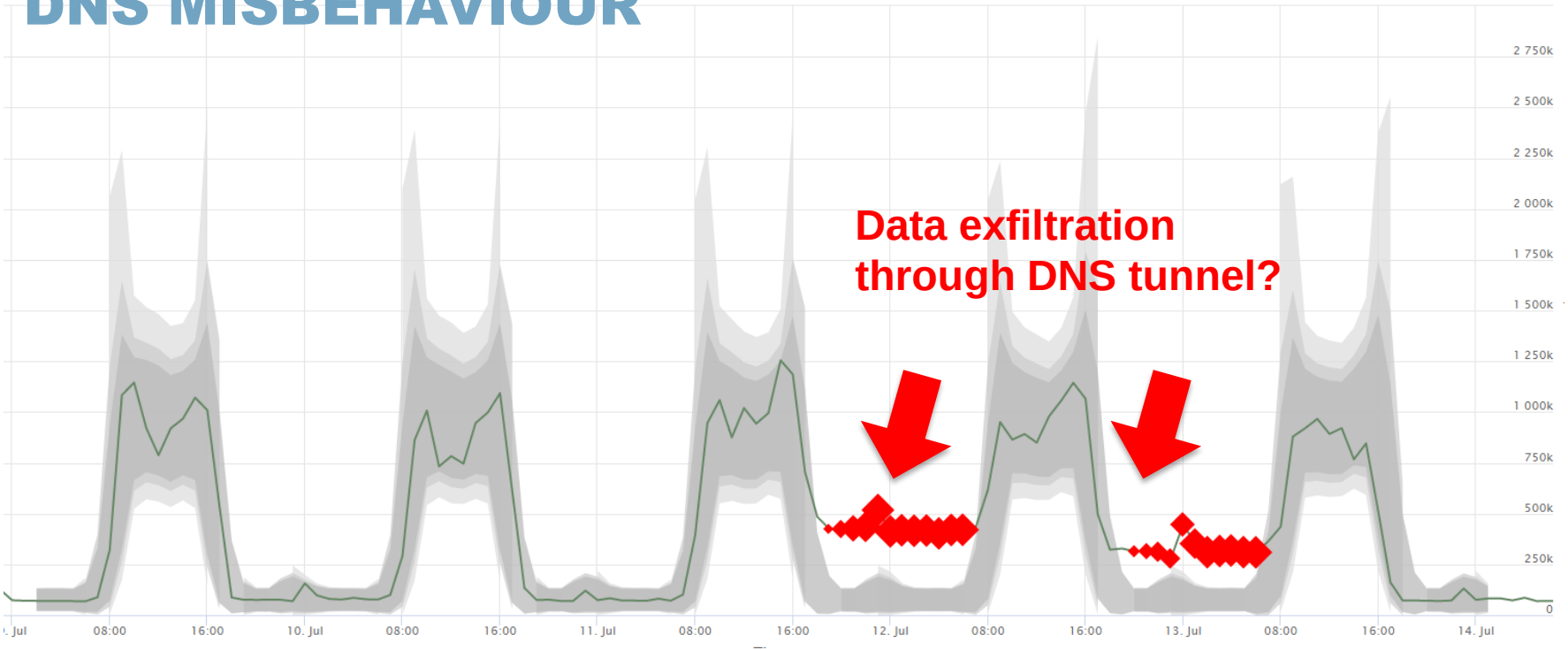
- Models a parameter by a set of 24 statistical models, one for each hour of the (working or weekend) day
- Optional: statistical models for the correlation between adjacent hours
- Produces dynamic upper and lower bounds for “normal” behaviour of the parameter and generates alerts when bound are exceeded
- Self-learning, using a negative exponential sliding window (typically 4 weeks)
- Keeps track of the model accuracy / confidence: disables alerts if confidence is too low



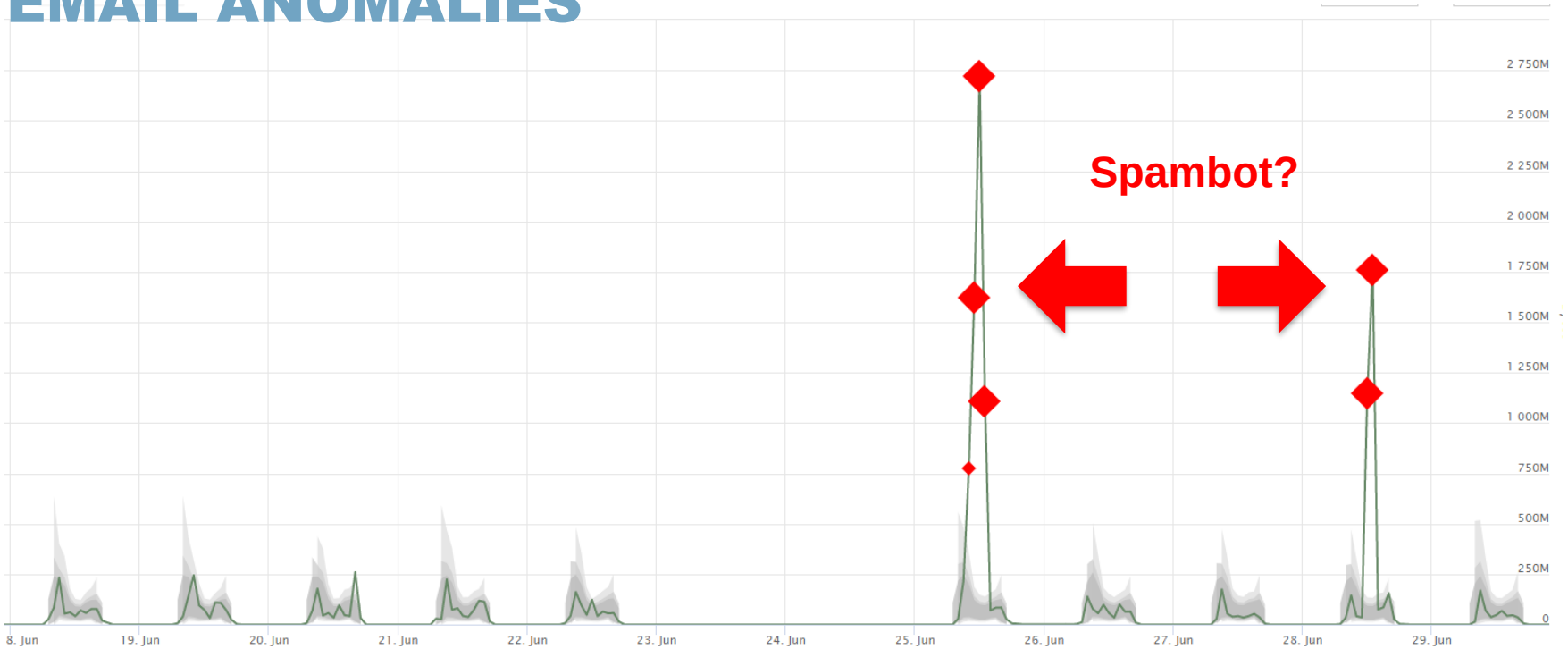
# EXAMPLE OF “NORMAL” OFFICE TRAFFIC



# DNS MISBEHAVIOUR



# EMAIL ANOMALIES



# NIGHTLY ACTIVE DIRECTORY ACTIVITY