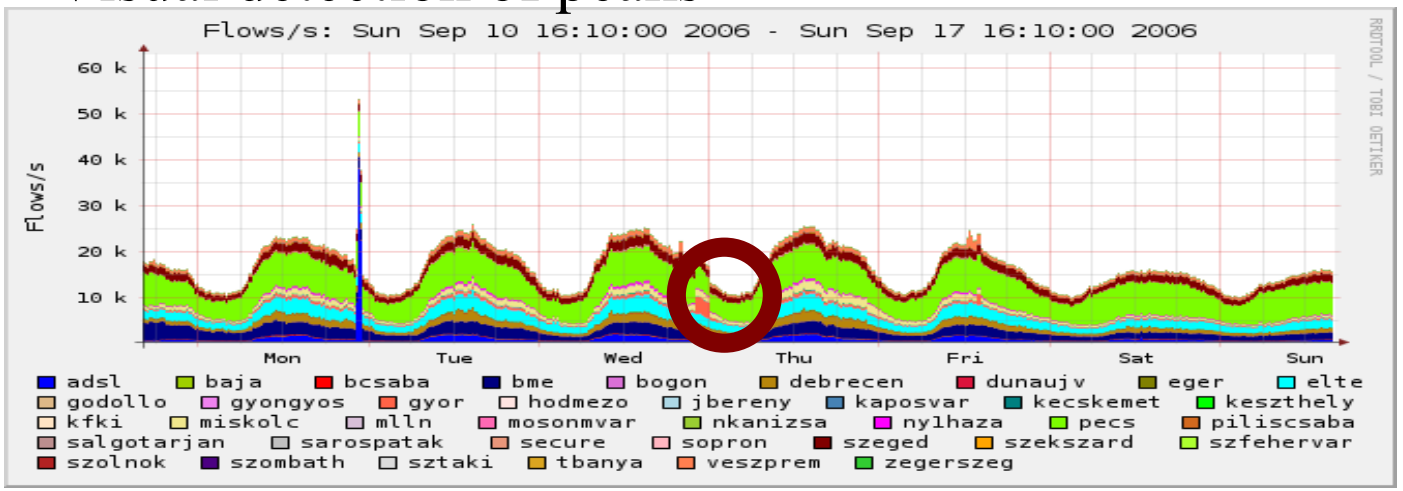
Anomaly detection for NFSen/nfdump netflow engine - with Holt-Winters algorithm

János Mohácsi, Gábor Kiss
NIIF/HUNGARNET

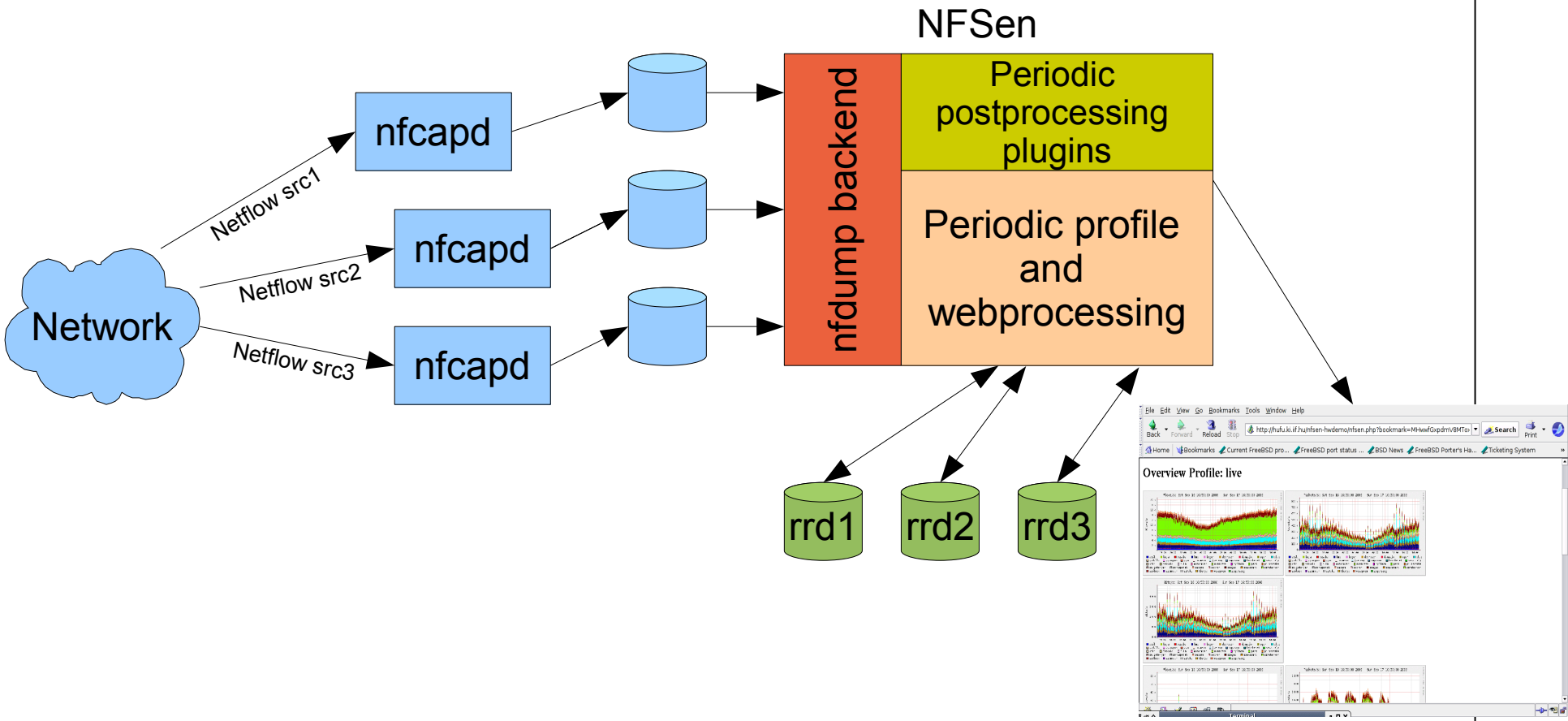
Motivation

- Usual work of CSIRT teams:
  - Find abnormal behaviour
    - Visual detection of peaks



    - Automatic abnormal detection – can you define what is abnormal
  - Report and coordinate incidents
  - Help to visually detect abnormal behaviour

Architecture of NfSen/nfdump



Anomaly detection

- Observations:
  - The traffic/flow pattern show weekly periodicity with some increasing trend in traffic volume/number of flows
  - The abnormal flow increase (DDoS, Scan) can be detected from deviation from this statistical traffic pattern
  - Processing the whole collected flow information is overwhelming – Nfsen (nfdump) already processed them and generated a digest in the RRD database
- Need: forecasting methods with ability to process data with seasonal behaviour and trends
- Alarm: when forecasted data and measured differs

Overview of Forecasting methods /1

- Single exponential smoothing – useful when there is no trends and seasonal components

$$\hat{y}_{n+1} = \sum_{i=0}^{\infty} w_i \, y_{n-i}$$

- Where w_i are the weights given to past values and sum to 1. Usually: $w_i = a*(1-a)^i$
- The formula: $\hat{y}_{n+1} = a y_n + a(1-a) y_{n-1} + a(1-a)^2 y_{n-2} + \ldots$
- Recursively: $\hat{y}_{n+1} = a y_n + (1-a)\hat{y}_n$ or $\hat{y}_{n+1} = a e_n + \hat{y}_n$

Next forecast

Current observation

Previous forecast

a: smoothing parameter,

e_n step ahead prediction error

Overview of Forecasting methods /2

- Holt-Winters Forecasting: Exponential smoothing for data with trend and/or seasonality
- Models contain estimates of trend and seasonal components
- Models "smooth", i.e. place greater weight on more recent data:
- $$\hat{y}_{n+1} = (m_n + b_n) c_{n-s+1}$$
- m_n component level, b_n component of the slope, c_{n-s+1} is the relevant seasonal component with *s* signifying the seasonal period

Overview of Forecasting methods /3

$$m_t = \alpha \frac{y_t}{c_{t-s}} + (1-\alpha)(m_{t-1} + b_{t-1})$$

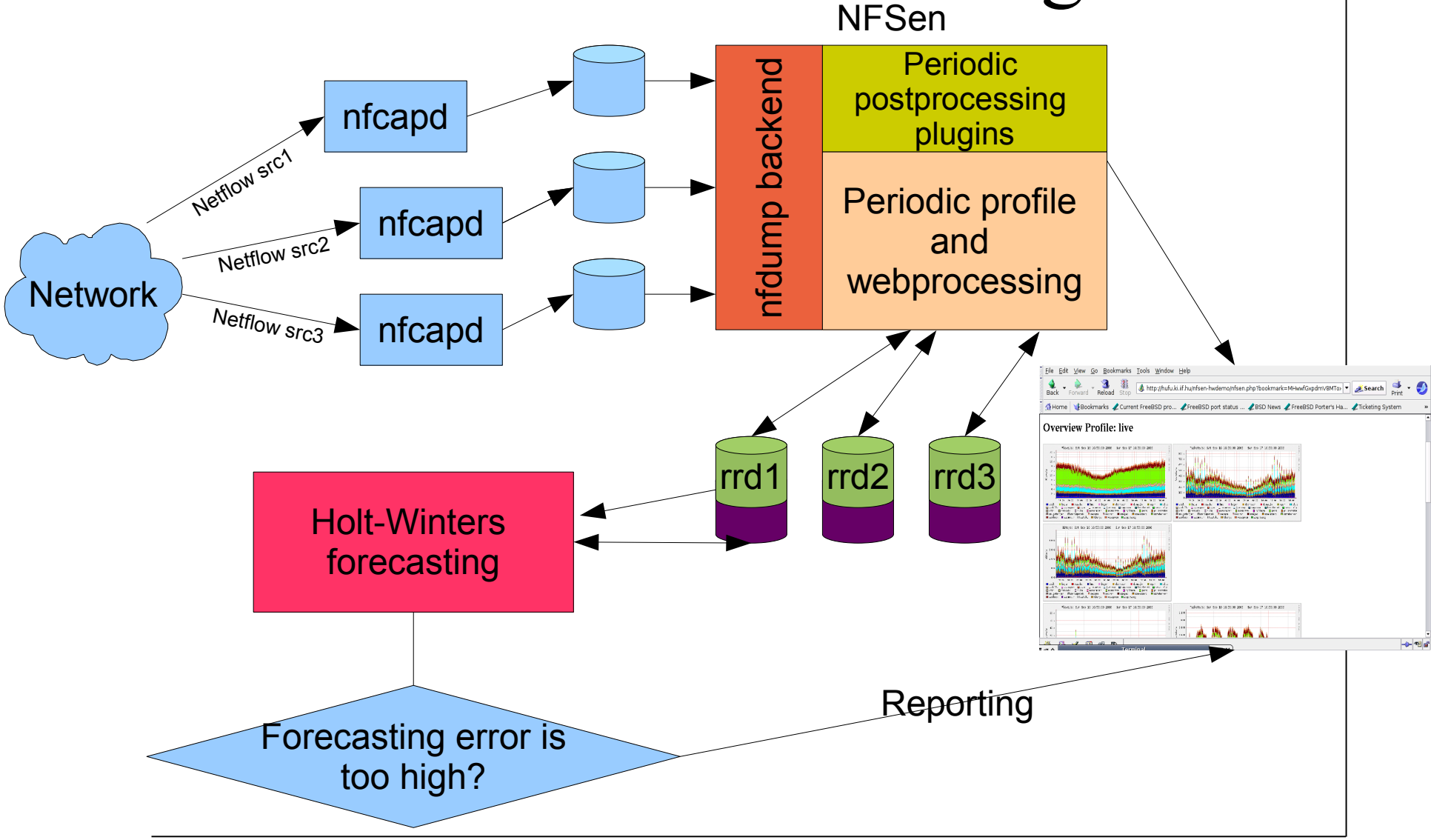
$$b_t = \beta (m_t - m_{t-1}) + (1-\beta) b_{t-1}$$

$$c_t = \gamma \frac{y_t}{m_t} + (1-\gamma) c_{t-s}$$

- α level parameter (between 0 and 1)
- β slope parameter (between 0 and 1
- γ seasonal factor parameter (between 0 and 1)

- All parameters are computable from observations of previous period

Architecture of NfSen/nfdump with Holt-Winters forecasting

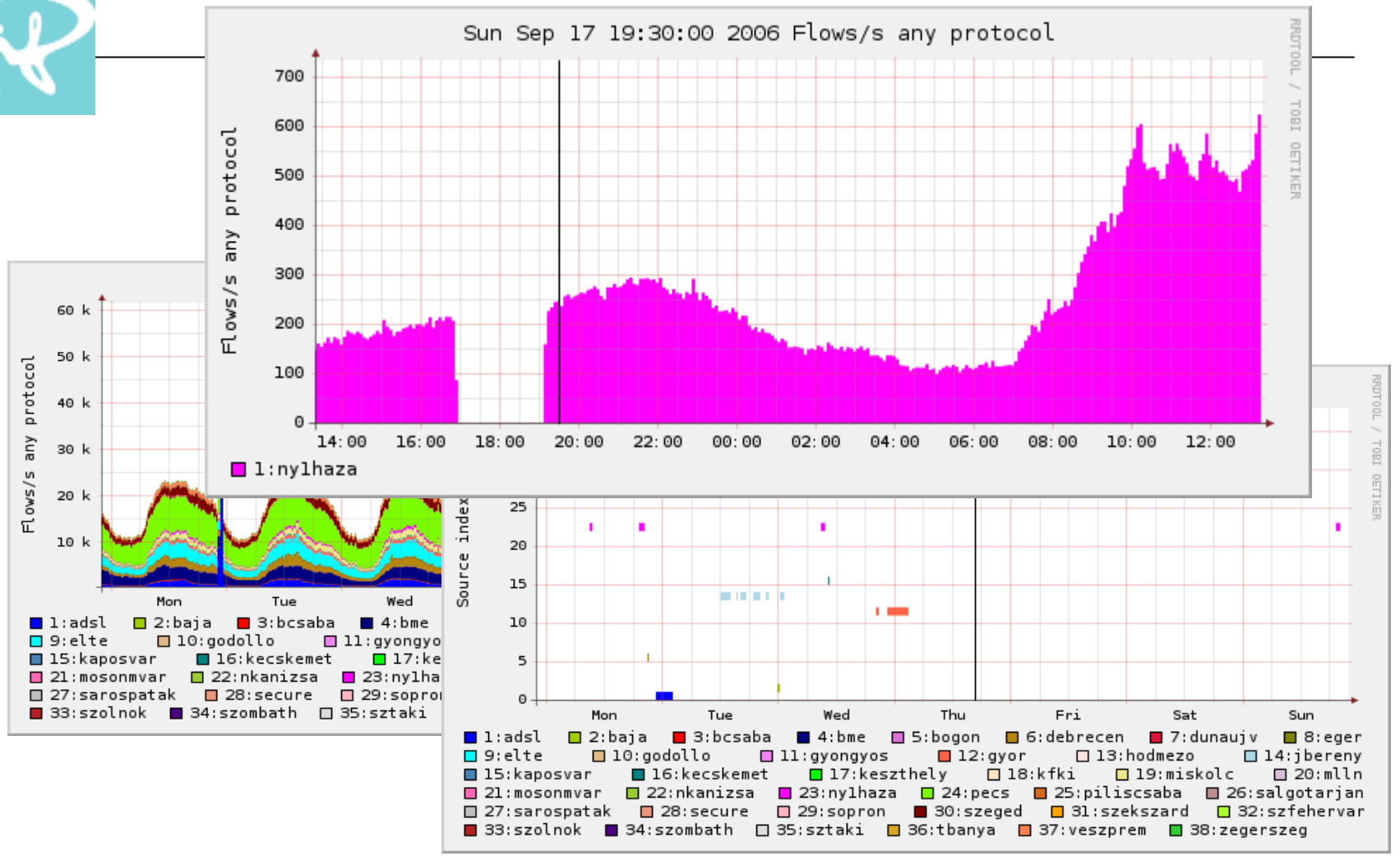


Holt-Winters implementation of RRDTool

- All the computational parameters are stored in RRA:
  - HWPREDICT – forecast computed by Holt-Winters algorithm – for each data points
  - SEASONAL – seasonal coefficient – for each datapoint
  - DEVPREDICT – array of deviation parameters
  - DEVSEASONAL – array of seasonal deviations
  - FAILURES – boolean indicators

Holt-Winters implementation of RRDTool/2

- New parametes for rrdtool create - Simplified interface to create:
  - RRA:HWPREDICT:<array length>:<alpha>:<beta>:<period>
- New parameters for rrdtool tune:
  - Specifying alpha, beta, gamma, positive/negative confidence interval, failure treshold



- Separate RRD updating – for testing only
- More intuitive interface – separate graph – 2 click listings of abnormal behaviour
- Detects abnormality even if you cannot see them by sights

Future

- Problems – needed further investigation
  - Actual settings of α, β and γ Some statistician volunteering?
    - RRA:HWPREDICT:1440:0.1:0.035:288
    - Deltapos 5, deltaneg 10 – to avoid traffic decrease alarm
  - Selection of periods – week is OK?
  - Requires at least one week data before prediction... - and abnormal behavior detection.
  - What did we detect? - are there any other important abnormal behavior that should be detected?

Further information

- Available to test:

  http://bakacsin.ki.iif.hu/~kissg/project/nfsen-hw
- About mathematics:

  http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc435.htm
- RRDTool implementation (J. Brutlag):

  http://cricket.sourceforge.net/aberrant/rrd_hw.htm

  http://www.usenix.org/events/lisa2000/full_papers/brutlag/brutlag_html/