Predikcia situačného povedomia v kybernetickej bezpečnosti pomocou metód neurónových sietí (strojového učenia)

Pracovisko: ÚINF

Autor: Bc. Jakub Mohler

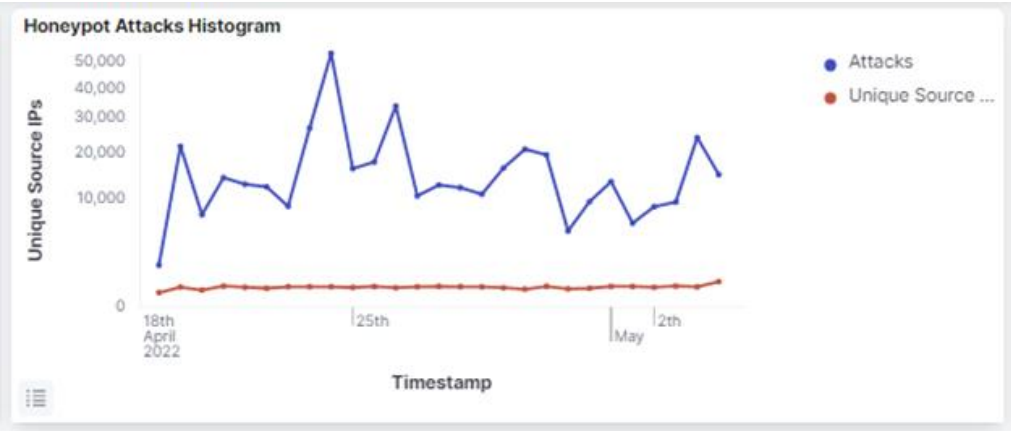
Vedúci práce: doc. RNDr. JUDr. Pavol Sokol, PhD.

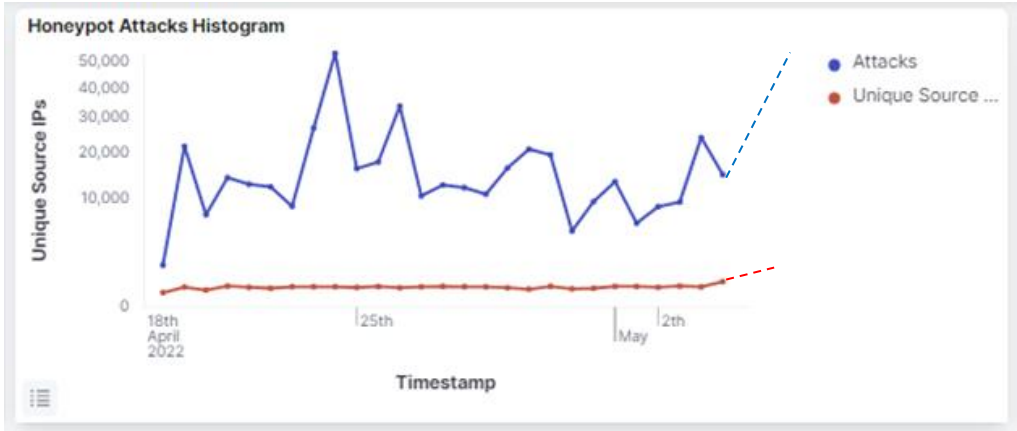
Konzultant: RNDr. Richard Staňa

Motivácia

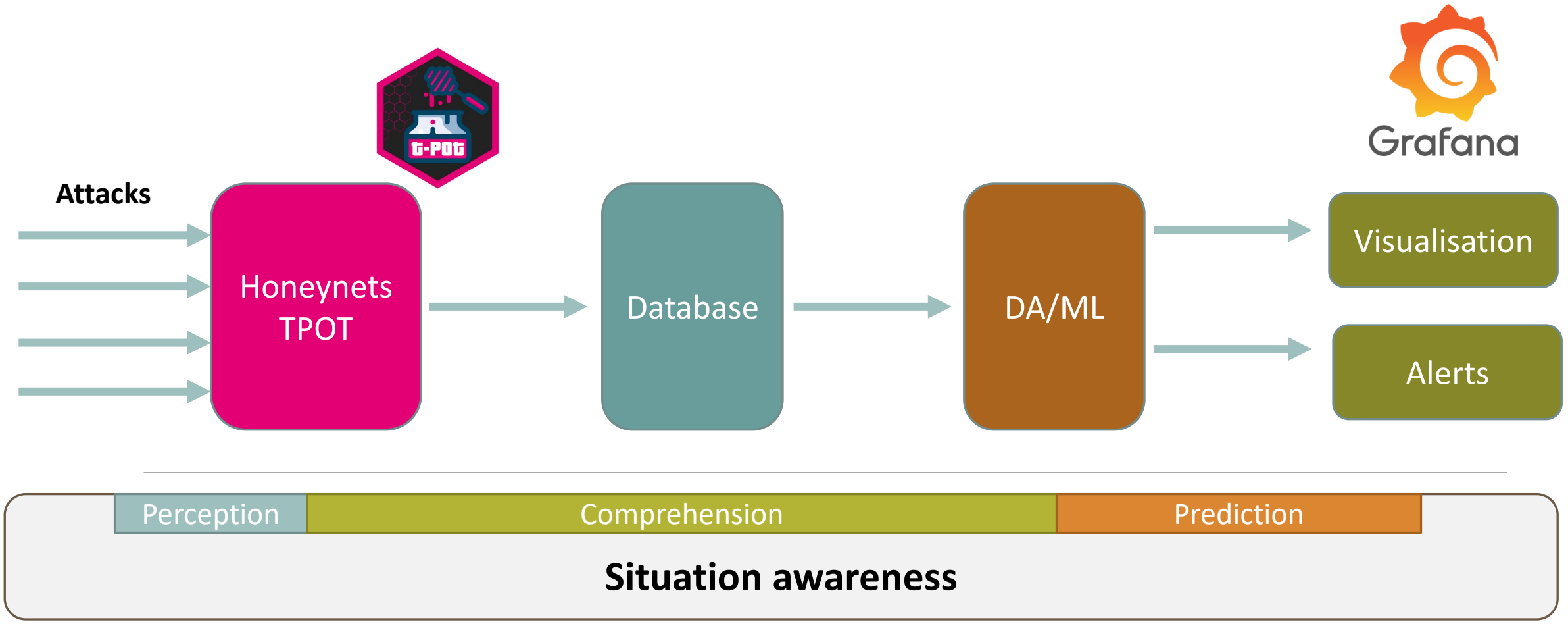

elastic
Search Elastic
Dashboard
>T-Pot
Full screen
Share
Clone
Edit
244,582 Dionaea - Attacks
159,243 Cowrie - Attacks
4,594 Adbhoney - Attacks
2,707 Tanner - Attacks
2,125 Honeytrap - Attacks
1,449 CitrixHoneypot - Attacks
848 ConPot - Attacks
823 Mailoney - Attacks
651 Heralding - Attacks
329 ElasticPot - Attacks
Honeypot Attacks Bar
Attacks
Honeypots
Honeypot Attacks Histogram
Unique Source IPs
Timestamp
Attacks
Unique Source ...
Honeypot Attack Map
Elastic Maps Service
OpenMapTiles
OpenStreetMap contributors
Attacks by Destination Port Histogram
445
22
23
80
55...
Attacks by Honeypot Histogram
Dionaea
Cowrie
Adbhoney
Tanner
Honeytr...
Attacks by Country Histogram
Vietnam
China
India
Indonesia
United Sta...
Attacker Src IP Reputation
known attacker
mass scanner
tor exit node
form spammer
bot, crawler
compromised
Attacks by Honeypot
Dionaea
Cowrie
Adbhoney
Tanner
Honeytrap
CitrixHoneypot
ConPot
Mailoney
Heralding
ElasticPot
POf OS Distribution
Linux 2.2.x-3.x
Windows 7 or 8
Linux 3.11 and newer
Linux 2.2.x-3.x (barebon...
Windows NT kernel 6.x
Windows NT kernel
Linux 3.1-3.10
Linux 2.4.x
Windows XP
Linux 2.2.x-3.x (no time...
Attacks by Country
Vietnam
China
India
Indonesia
United States
Brazil
Philippines
Taiwan
Venezuela
Russia
Attacks by Country and Port
Suricata Alert Category Histogram
Not Suspicious Traffic

Cieľ


Honeypot Attacks Histogram
Unique Source IPs
50,000
40,000
30,000
20,000
10,000
0
18th April 2022
25th
May
2th
Timestamp
Attacks
Unique Source ...

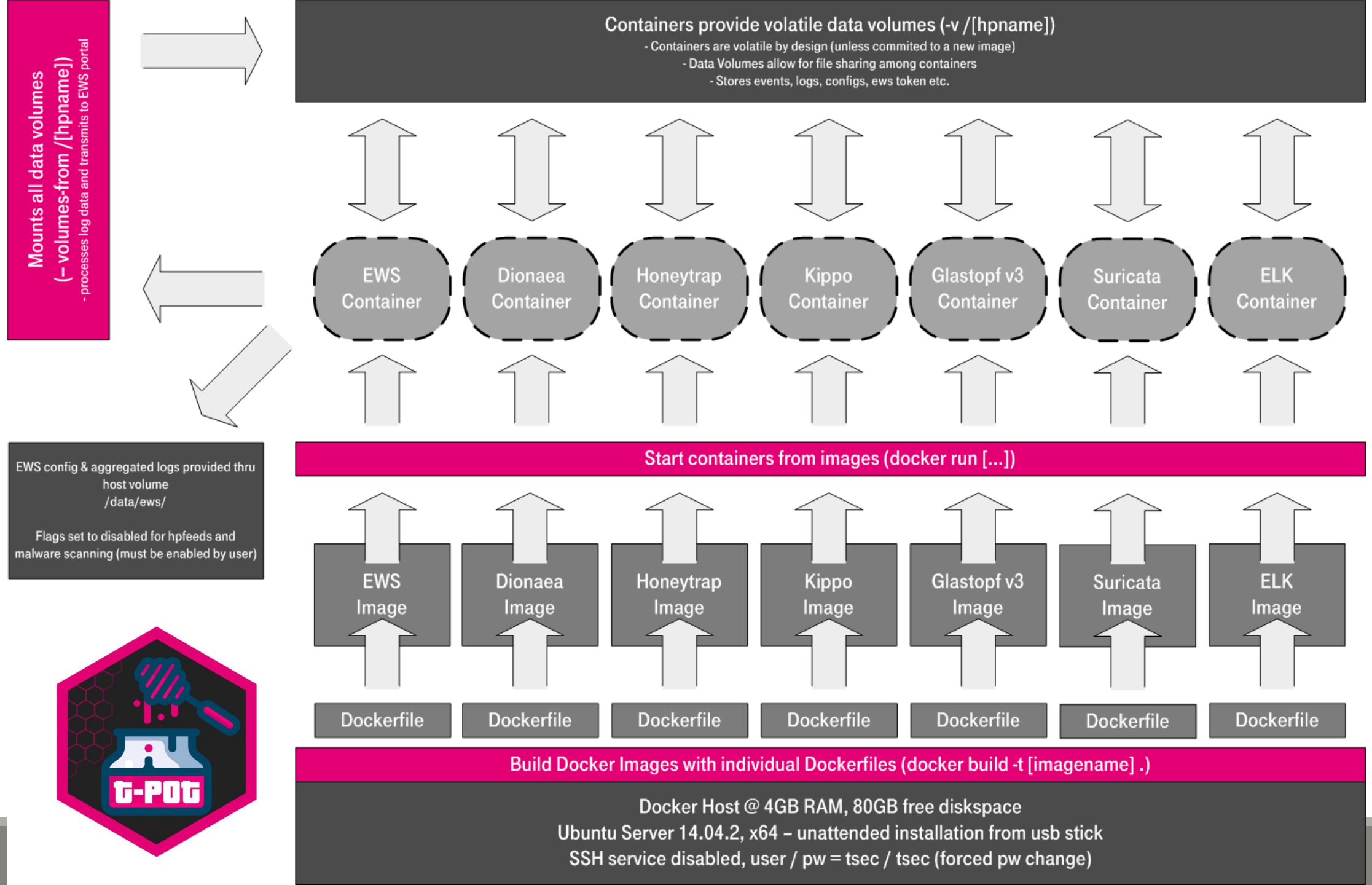

Honeypot Attacks Histogram
Unique Source IPs
50,000
40,000
30,000
20,000
10,000
0
18th April 2022
25th
May
2th
Timestamp
Attacks
Unique Source ...

Návrh


Attacks
Honeynets
TPOT
Database
DA/ML
Visualisation
Alerts
Grafana
Perception
Comprehension
Prediction
Situation awareness

Honeynets

- Zber dát
- Rôzny honeypoty = rôzne dáta
- Časové rady
- Extrakcia dát z T-Pot
- Dataset (existujúci + vytváranie nového)
- Rozšírenie siete honeynetov


Mounts all data volumes
(– volumes-from /[hpname])
- processes log data and transmits to EWS portal
Containers provide volatile data volumes (-v /[hpname])
- Containers are volatile by design (unless commited to a new image)
- Data Volumes allow for file sharing among containers
- Stores events, logs, configs, ews token etc.
EWS Container
Dionaea Container
Honeytrap Container
Kippo Container
Glastopf v3 Container
Suricata Container
ELK Container
EWS config & aggregated logs provided thru host volume
/data/ews/
Flags set to disabled for hpfeeds and malware scanning (must be enabled by user)
Start containers from images (docker run [...])
EWS Image
Dionaea Image
Honeytrap Image
Kippo Image
Glastopf v3 Image
Suricata Image
ELK Image
Dockerfile
Dockerfile
Dockerfile
Dockerfile
Dockerfile
Dockerfile
Dockerfile
Build Docker Images with individual Dockerfiles (docker build -t [imagename] .)
Docker Host @ 4GB RAM, 80GB free diskspace
Ubuntu Server 14.04.2, x64 – unattended installation from usb stick
SSH service disabled, user / pw = tsec / tsec (forced pw change)
T-POT

Databáza(y)

JSON Databáza

- Typ databázy
- Návrh databázy
- Ukladanie JSON z T-Pot

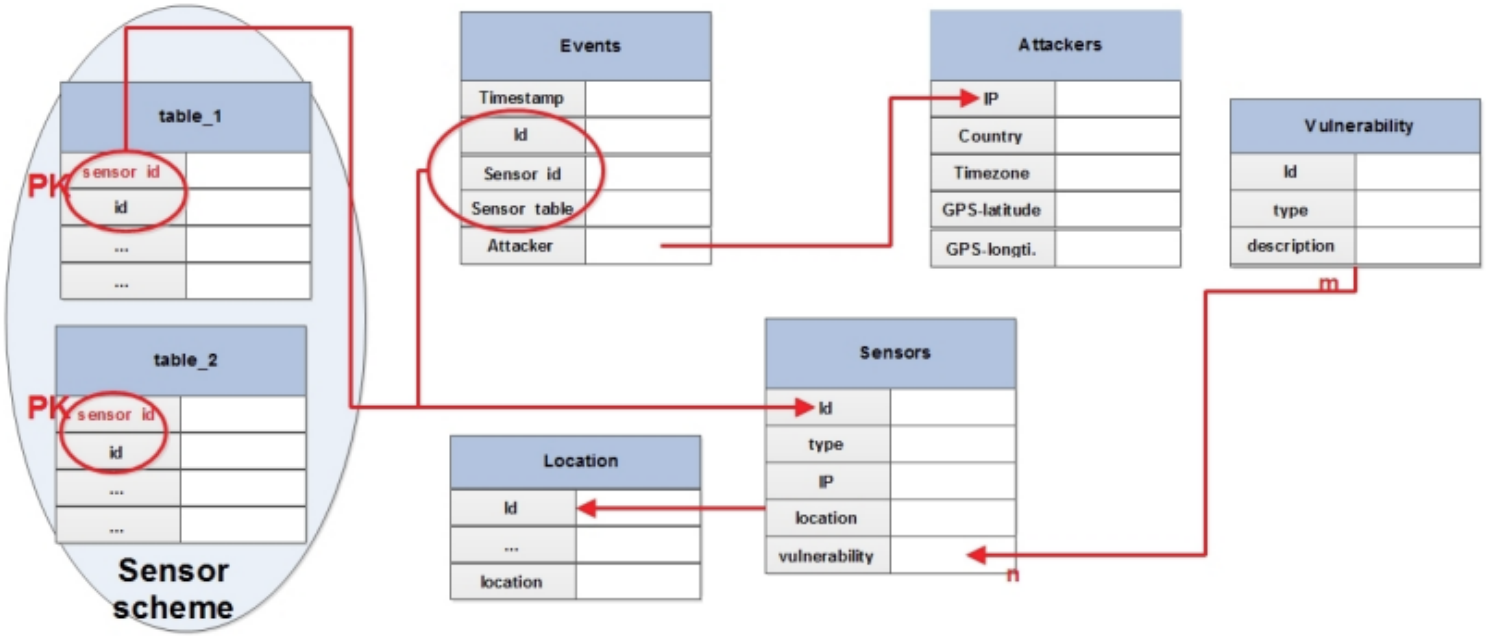


Figure 4: Scheme of central database.

Zdroj: https://web.archive.org/web/20180421062733id_/http://spi.unob.cz/papers/2015/2015-19.pdf

Databáza časových radov

elastic

influxdb

Prometheus

Grafana

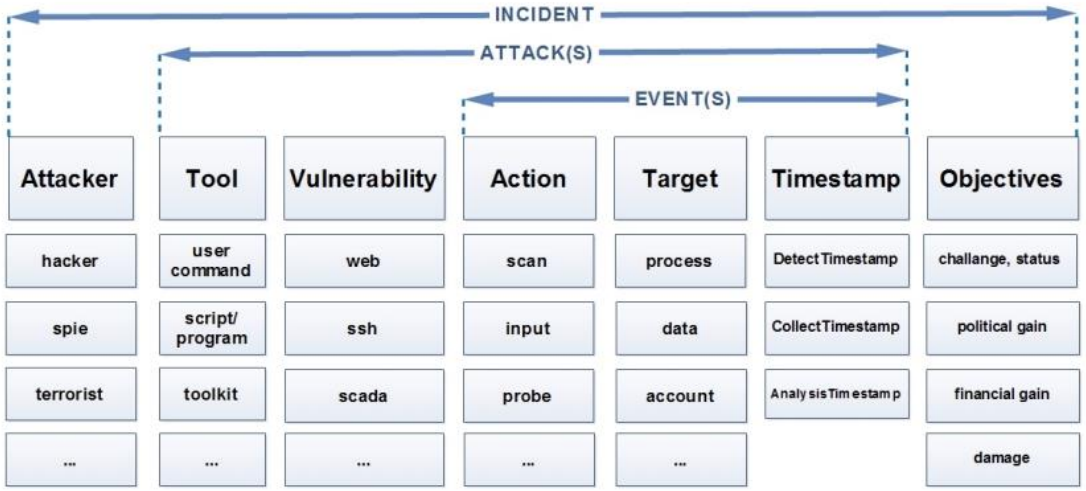
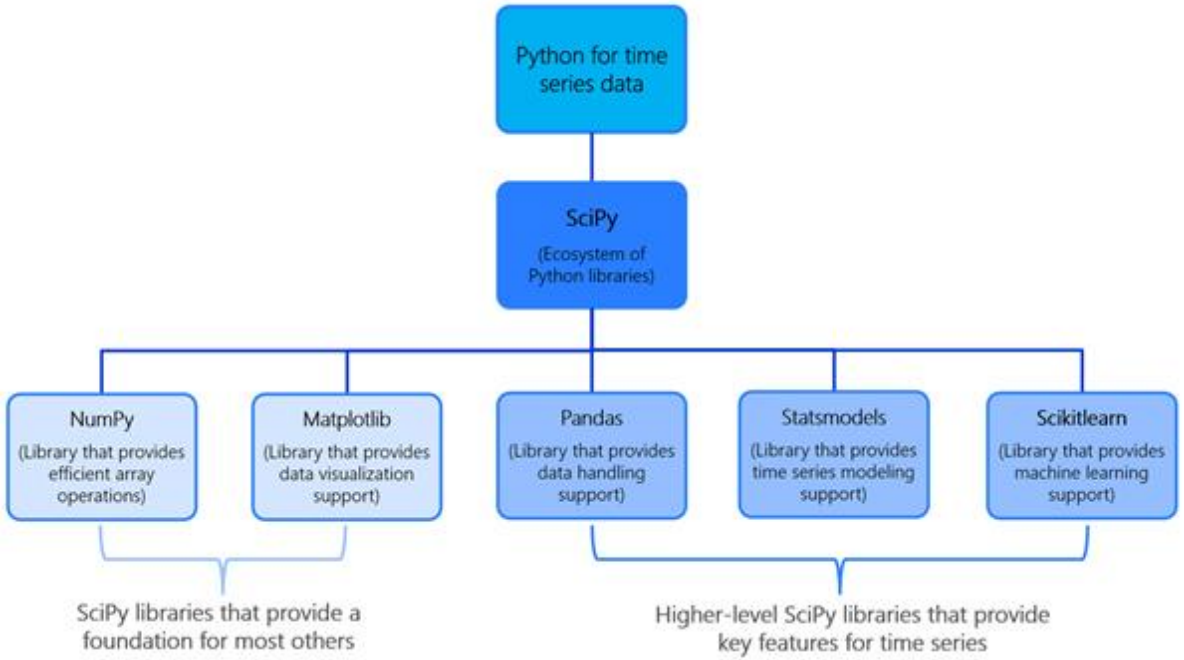


Figure 7: Incident taxonomy based on honeypots' and honeynets' data.

Dátová analýza / Strojové učenie

- Články, práce kolegov
- ML v Elasticu je drahý
- Vytvorenie predikcie
- Použitie vhodných knižníc



Verifikácia (I.)

- Výber metrík (MASE, MAE, MSE)
- Diebol Marino test
  - Reálne hodnoty
  - Predikcia 1. modelu
  - Predikcia 2. modelu

  - Porovnanie modelov

Zdroje: https://gist.github.com/bshishov/5dc237f59f019b26145648e2124ca1c9
https://github.com/johntwk/Di-ebold-Mariano-Test

Verifikácia (II.)

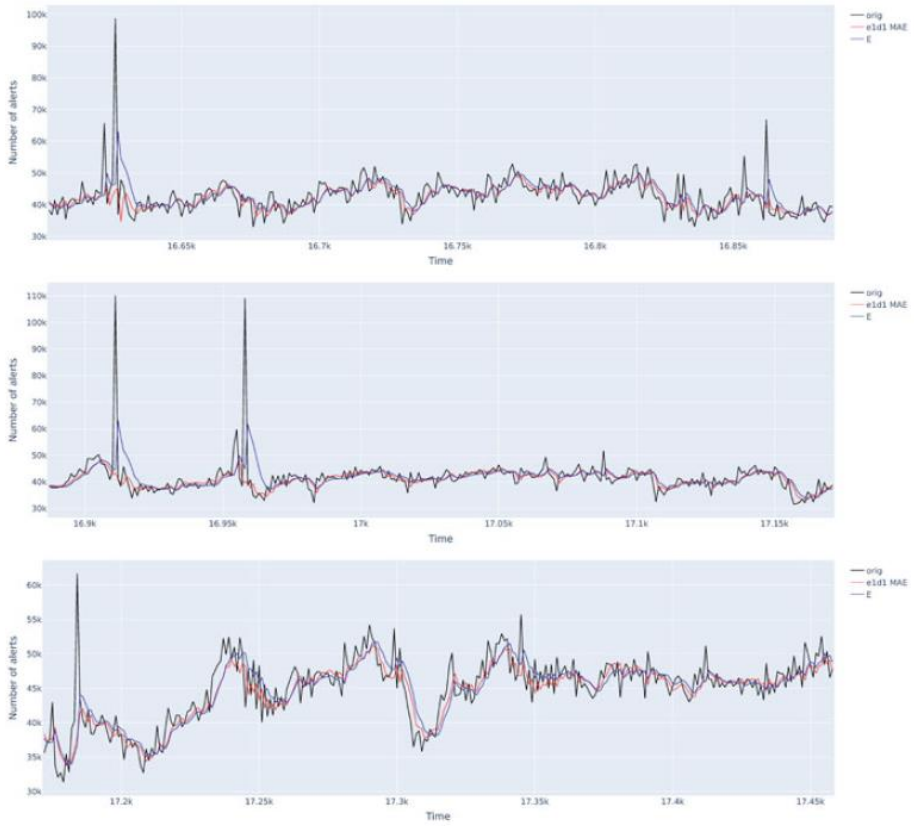


Fig. 1 Graphical comparison of best models based on neural networks (e1d1 with MAE loss) and statistical models (exponential smoothing) for the total number of alerts

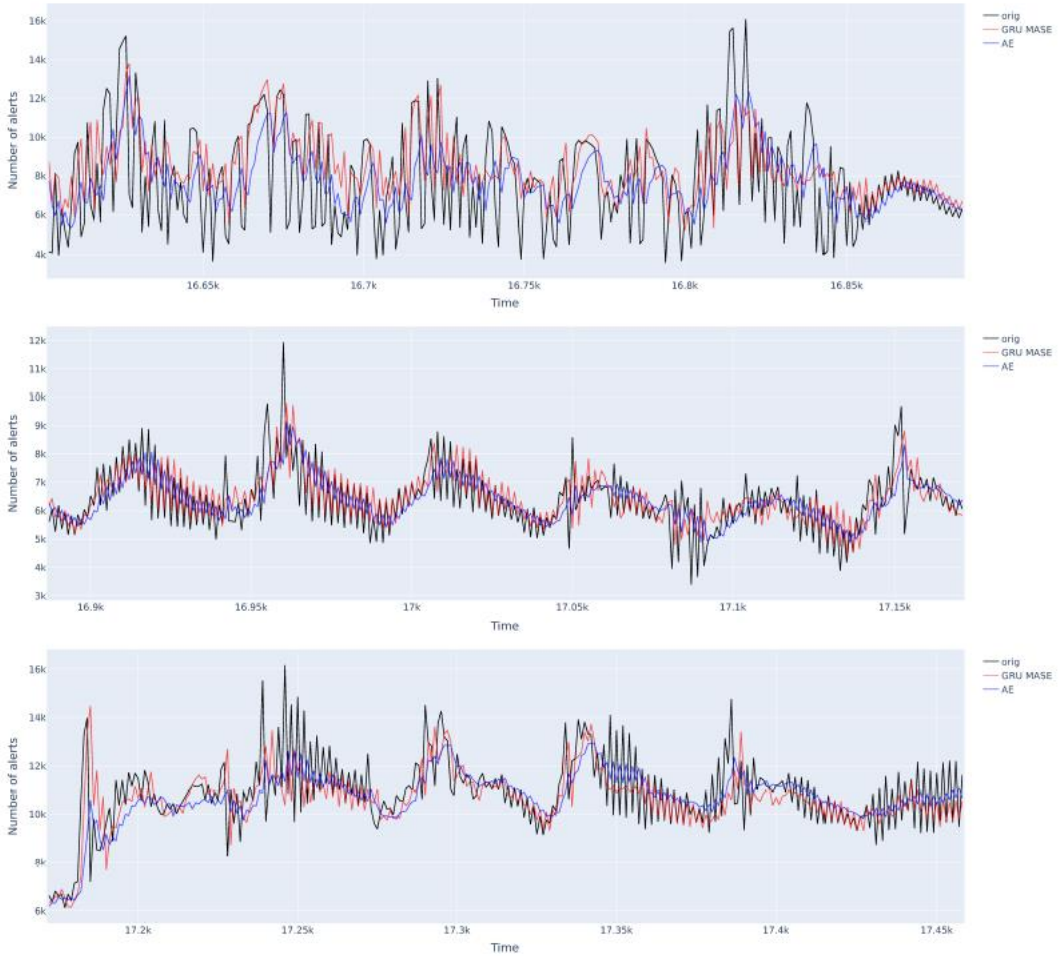


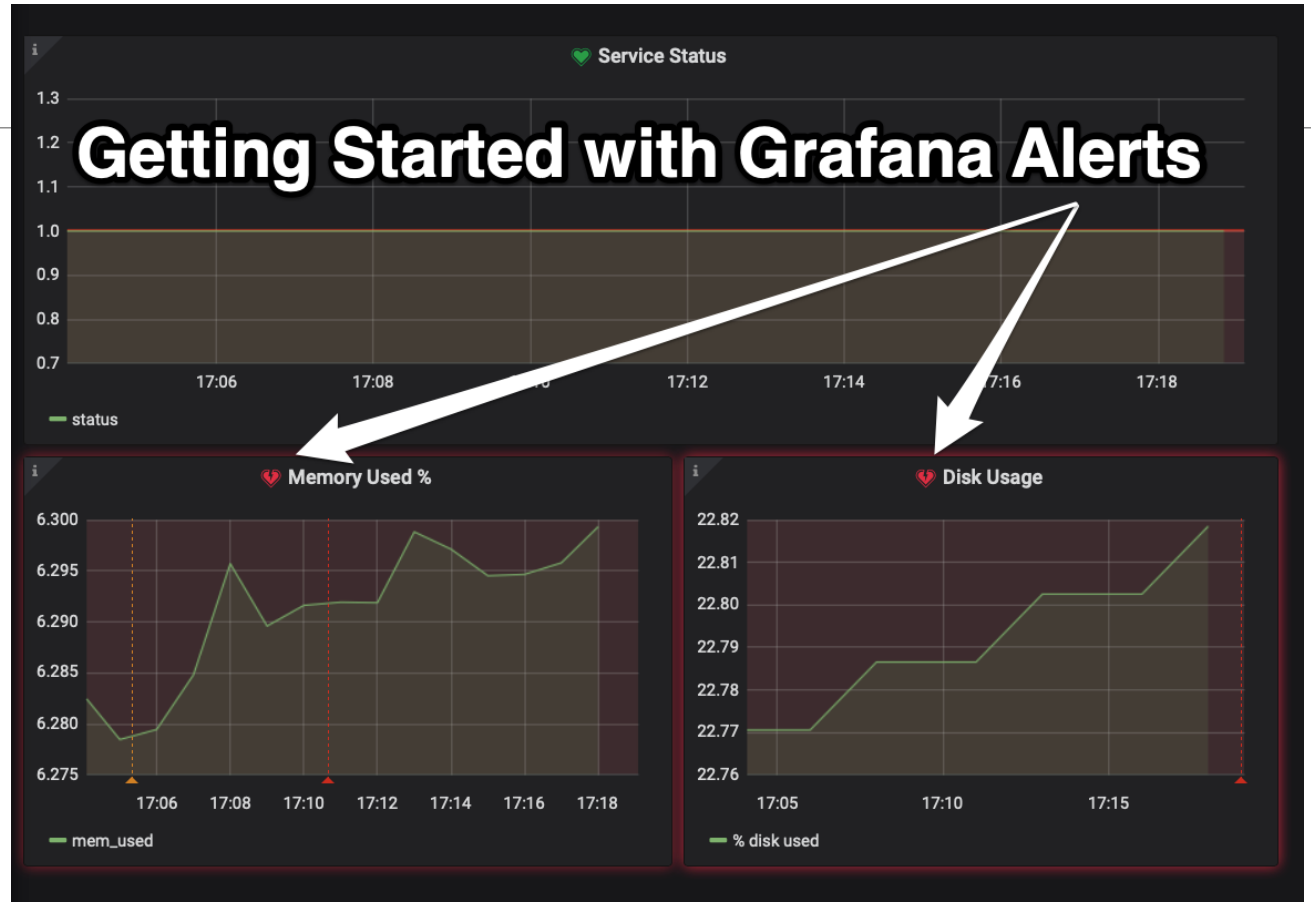
Fig. 2. Graphical comparison of best models based on neural networks (Gru with MASE loss) and statistical models (combination of ARIMA and Exponential smoothing) for port 445/TCP.

Zdroje: Sokol, P., Staňa, R., Gajdoš, A. and Pekarčík, P., 2022. Network security situation awareness forecasting based on statistical approach and neural networks. *Logic Journal of the IGPL*., Husák M, Bartoš V, Sokol P, Gajdoš A. Predictive methods in cyber defense: Current experience and research challenges. Future Generation Computer Systems. 2021 Feb 1;115:517-30

Vizualizácia

Grafana

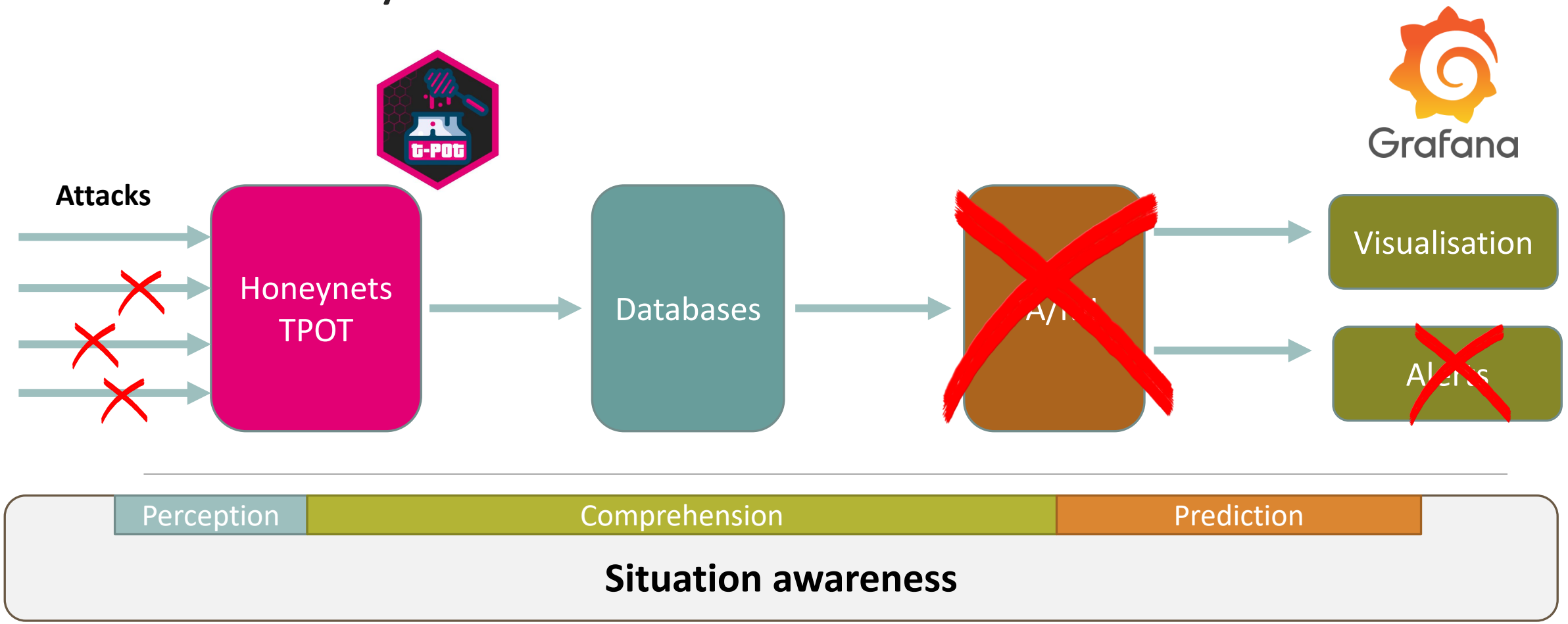
Alerty


Service Status
1.3
1.2
1.1
1.0
0.9
0.8
0.7
17:06
17:08
17:12
17:14
17:16
17:18
status
Getting Started with Grafana Alerts
Memory Used %
6.300
6.295
6.290
6.285
6.280
6.275
17:06
17:08
17:10
17:12
17:14
17:16
17:18
mem_used
Disk Usage
22.82
22.81
22.80
22.79
22.78
22.77
22.76
17:05
17:10
17:15
% disk used

Nasledujúce kroky

- Preskúmať spomínané nástroje
- Navrhnúť databázové modely
- Vytvoriť odľahčenú formu podľa návrhu
- Postupne to rozširovať, pridávať ďalšie požiadavky
  - Spojazdniť to “takmer real-time“

Odľahčený návrh



Ďakujem za pozornosť.