Realtime Communication of MISP, Zeek, and SIEMs

Matthias Vallentin
Tenzir

Liviu Vâlsan
CERN

Intelligence in Zeek

- Architecture
  - `Intel::Item` represents intelligence
    - `Intel::Type` one of `ADDR, SUBNET, URL, SOFTWARE, EMAIL, DOMAIN, USER_NAME, CERT_HASH, PUBKEY_HASH, FILE_HASH, FILE_NAME`
- Cluster
  - Manager holds full intel data, disseminates minimal subset to workers
  - Workers report back matches to master

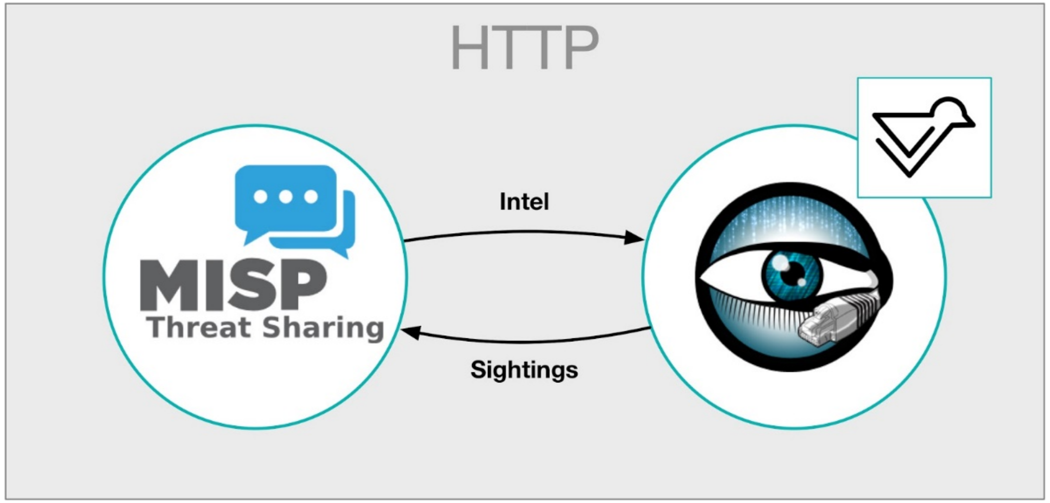
Intelligence in MISP

- MISP: Open-Source Threat Intelligence Sharing Platform
- Zeek's `Intel::Item` = MISP attribute
- Can download a snapshot of MISP intel via REST API
- ZeroMQ pub/sub for all MISP activity
  - Publisher (MISP): stream of (topic, JSON) data
  - Subscriber (User): consume and process data

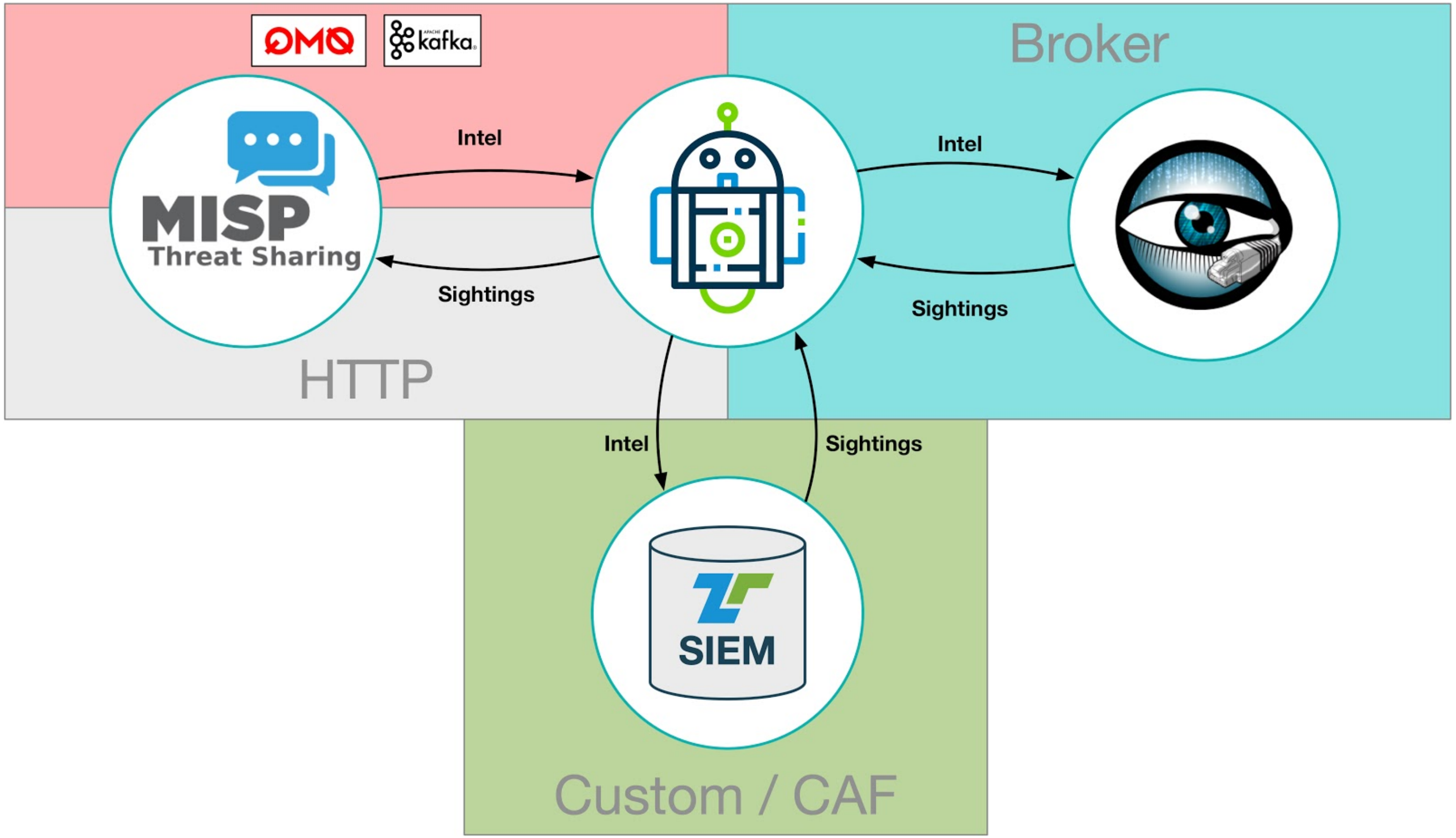
MISP Threat Sharing

Related Work: **dovehawk**

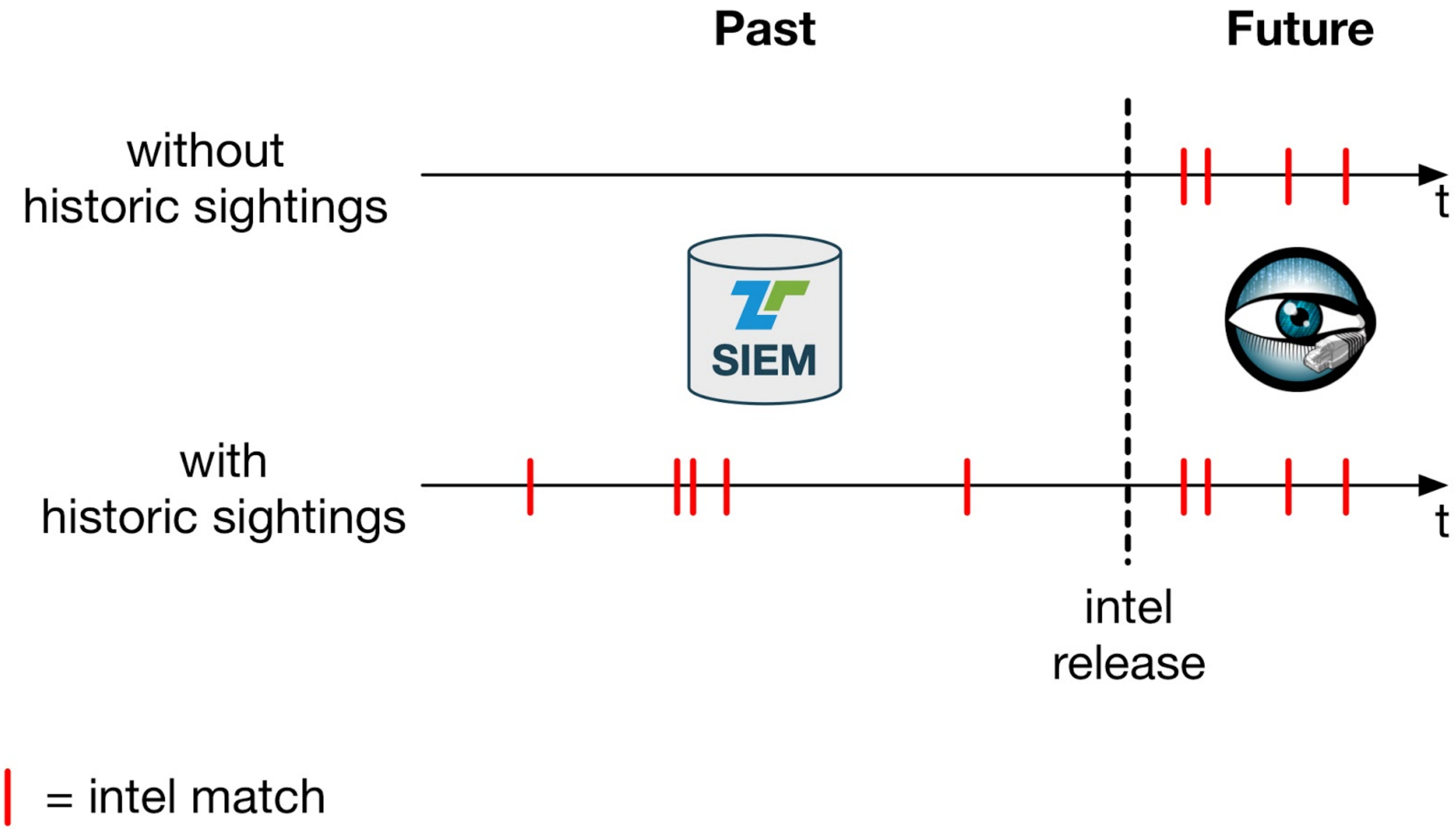
- https://github.com/tylabs/dovehawk
- Direct use of MISP's REST API
- Can report Zeek intel matches
  - Back to MISP as sightings
  - To Slack channel
- Implementation
  - Via Zeek's `ActiveHTTP` framework
  - Periodic download of intel snapshot
  - Intel framework weeds out duplicates
- Limitations
  - Snapshot-based
  - No real-time feed of deltas



A new approach: The Robo Investigator



Historical Intel Matching



Past

Future

without
historic sightings

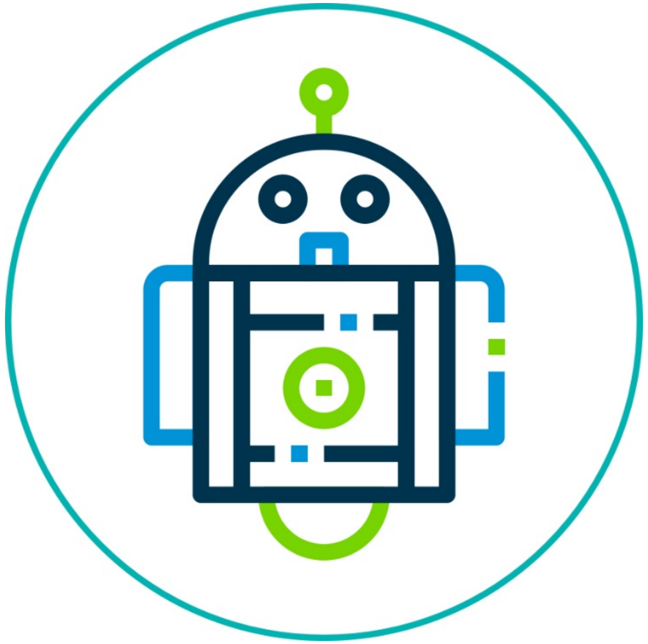
with
historic sightings

intel
release

| = intel match

Robo Investigator - Architecture

- Pluggable **producer / consumer architecture**:
  - Producers: MISP (candidates: IntelMQ, STIX, passive DNS)
  - Consumers: Zeek, VAST/Tenzir (candidates: Sigma)
- **Bidirectional communication** channels
- Written in Python 3
  - `pymisp, broker, confluent_kafka, pyzmq`
  - `asyncio` for coroutine-based concurrency



Robo Investigator - Benefits

- **Real-time processing** of new / changing intel
  - No need to wait for next snapshot
  - Only delta requires processing: constant-time work -> finally scales!
- New Kafka interface from CERN enables **reliable intel delivery**
- Integration of **SIEM context**
  - Historic sightings reconstruct full picture of incident
- **Decoupled components** improves flexibility and maintainability
  - Can add different intel providers
  - Zeek scripts are agnostic to intel format

Zeek Consumer

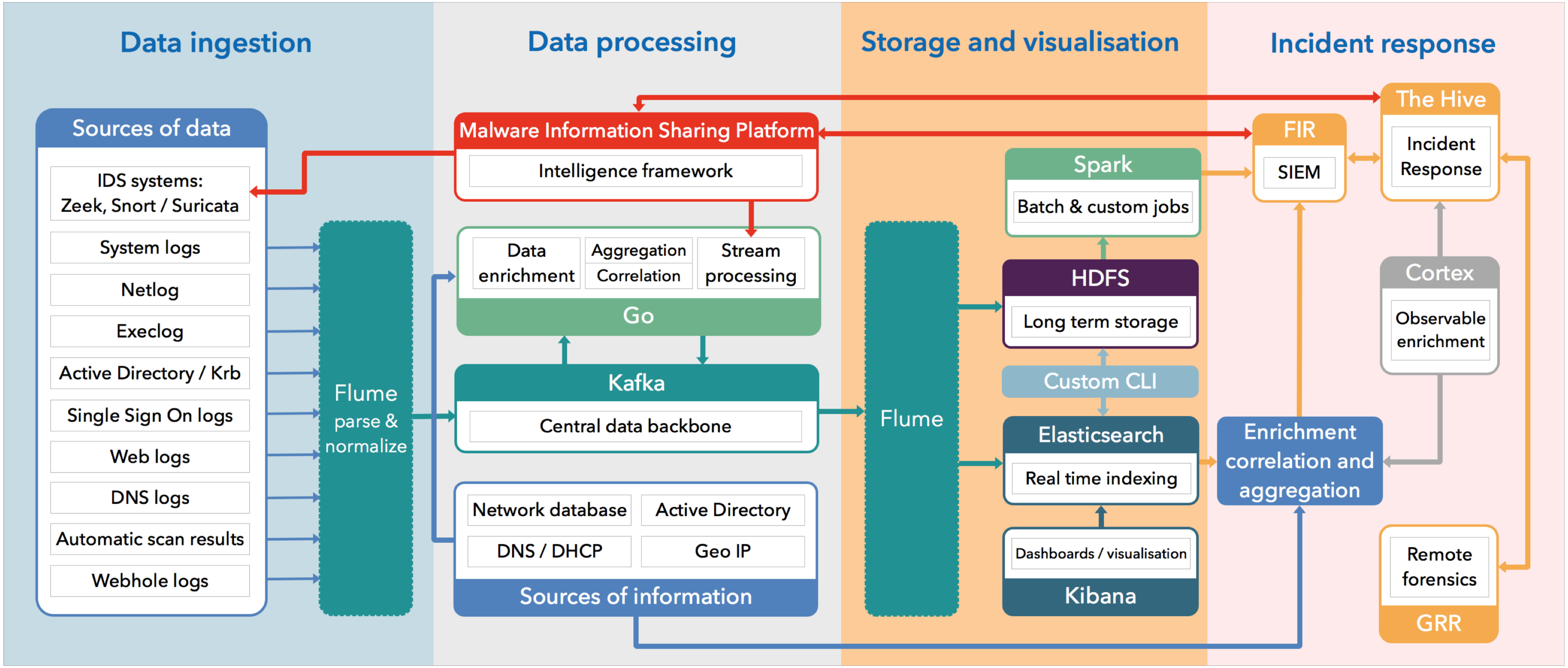
- **Broker-based** communication
- Supports **standalone and cluster** mode
- Can ask for **intel snapshot at startup**
- Noisy intel feature:
  - Handling matches of heavy hitters causes high CPU load
  - Zeek sends special event if intel matches exceed a certain rate
  - Zeek then removes intel locally (high CPU load otherwise)
  - Robo sends a proposal to remove IDS flag from corresponding MISP attribute

VAST / Tenzir Consumer

- Example of SIEM integration
- Translates intel into historical queries
- Extracts timestamps from results
- Reports sighting times for given intel
- Efficient control and data channel*
  - Fast communication via CAF
  - Zero-copy data sharing via Apache Arrow

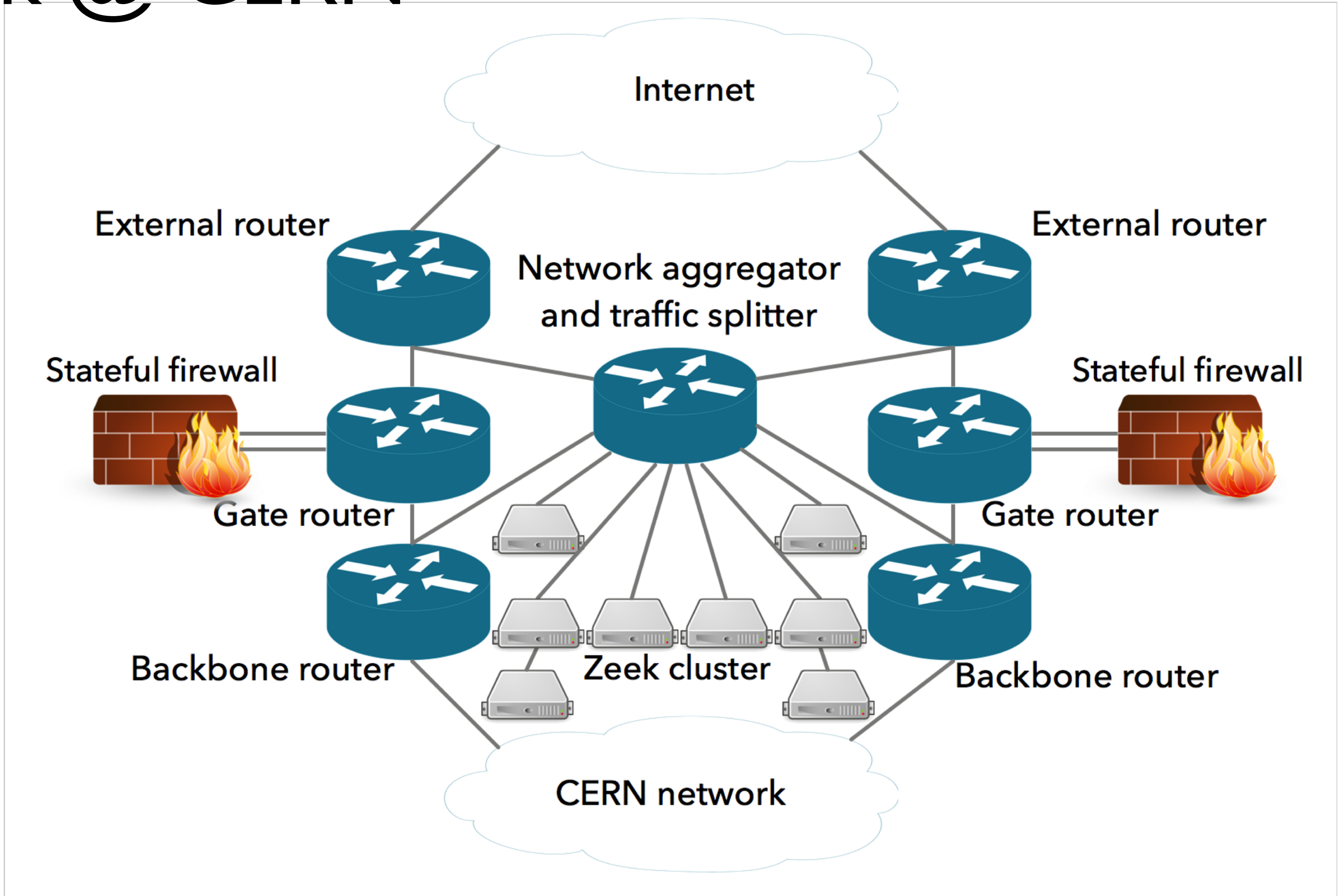
*under development

CERN SOC



Data ingestion
Data processing
Storage and visualisation
Incident response
Sources of data
IDS systems: Zeek, Snort / Suricata
System logs
Netlog
Execlog
Active Directory / Krb
Single Sign On logs
Web logs
DNS logs
Automatic scan results
Webhole logs
Flume parse & normalize
Malware Information Sharing Platform
Intelligence framework
Data enrichment
Aggregation Correlation
Stream processing
Go
Kafka
Central data backbone
Network database
Active Directory
DNS / DHCP
Geo IP
Sources of information
Flume
Spark
Batch & custom jobs
HDFS
Long term storage
Custom CLI
Elasticsearch
Real time indexing
Dashboards / visualisation
Kibana
FIR
SIEM
The Hive
Incident Response
Cortex
Observable enrichment
Enrichment correlation and aggregation
Remote forensics
GRR

Zeek @ CERN


Internet
External router
External router
Network aggregator and traffic splitter
Stateful firewall
Stateful firewall
Gate router
Gate router
Backbone router
Zeek cluster
Backbone router
CERN network

Zeek @ CERN

- **Zeek** as the **primary** Intrusion Detection System
- **Monitoring all traffic** passing at the borders:
  - Between CERN and the public Internet
  - Guest WiFi network
  - Between specific CERN network domains
  - 200 Gbps total bandwidth
- **16 Zeek servers** in total
  - 10 production active nodes
  - 2 production backups
  - 4 QA

MISP @ CERN

- **MISP** as the **sole threat intelligence platform**
- A total of 4 MISP instances
- All instances behind Single Sign On
- Main instance
  - > **1.3 million IoCs** (MISP attributes)
  - > 400 contributing organizations
- Most intel coming from other MISP instances
- Importing of special purpose, private intel feeds

MISP
Threat Sharing

MISP & Zeek @ CERN

- Periodic export from MISP into Zeek intel framework format
  - IoCs from events (re)published in the past 30 days
  - IoCs from events with specific tags
- On average **100 000 IoCs** being **actively used**
- Issues:
  - Full export every time
  - High load of the small VM hosting MISP
  - Delay before intel gets added / remove from Zeek
  - Intel used only for realtime detection
  - Sightings are not reported back to MISP

Extending MISP

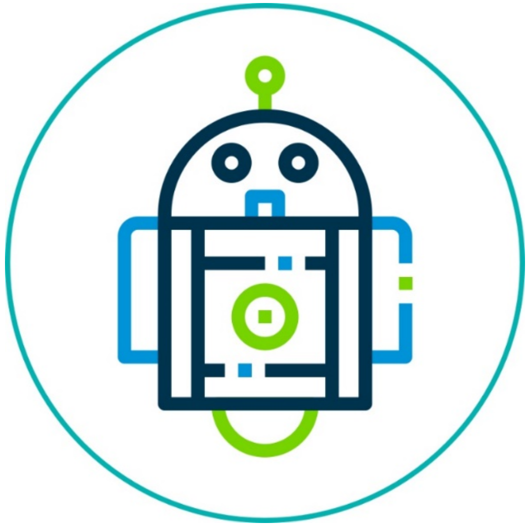
- MISP support for ZeroMQ publishing since 2015 (MISP v2.3.87)
- ZeroMQ implementation does not fit into our setup
- Attributes published as soon as they are added to MISP
- CERN contributed **Kafka support** in **MISP**
  - Available starting from MISP v2.4.104
  - Feature equivalent to ZeroMQ support
  - Kafka topic for (re)published MISP events

Deployment of Robo Investigator @ CERN

- Deployment on one of the QA Zeek node
  - Receiving an **exact copy of the traffic** going to a **production** Zeek instance
- Connected to our development MISP instance
- Successfully validated core functionality:
  - Real-time ingestion / removal of intel items
  - Dump of all intel items from Zeek
  - Removal of noisy intel items
    - Proposal added to MISP for removing the IDS flag
  - Intel sightings from Zeek to MISP

Next steps for Robo Investigator @ CERN

- Perform exhaustive loading of intel database
- Trigger historical searches for newly added IoCs
    - Add new intel consumer
- **Transition into production** deployment



Summary

- Intelligence is a key driver for threat hunting and incident response
- For maximum efficacy: feed intel to **detection** and **forensics tools**
- Demonstrated an integrated solution to do this **in real time**
  - MISP + Zeek + SIEM
  - Key benefit: reduced time to detect critical intel
- Operational validation at CERN SOC
  - Core features add value
  - Next step: transition into production deployment

Questions?

Matthias Vallentin
matthias@tenzir.com

Liviu Vâlsan
liviu.valsan@cern.ch