A SIEM Architecture for Multidimensional Anomaly Detection

K.-O. Detken · T. Klecker (DECOIT® GmbH)

C. Kleiner · T. Laue (University of Applied Sciences Hanover)

Prof. Dr. Kai-Oliver Detken
DECOIT® GmbH
Fahrenheitstraße 9
D-28359 Bremen
https://www.decoit.de
detken@decoit.de

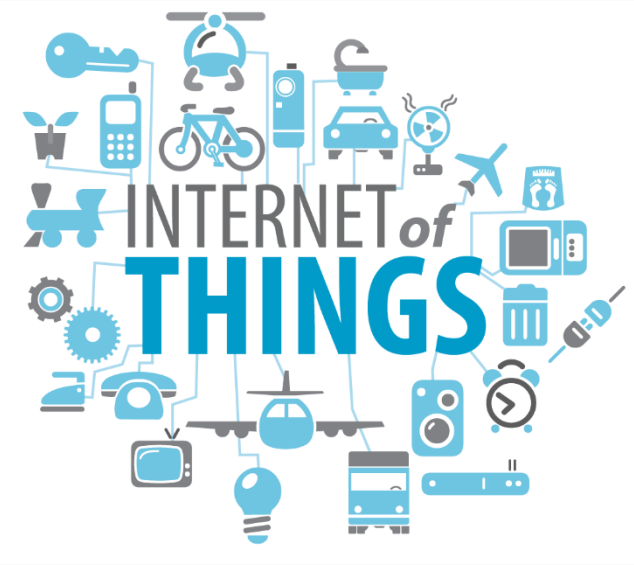
DECOIT®

Agenda

- Motivation
- SIEM definition
- GLACIER project and architecture
- Data Collection and Data Analysis
- Use Cases
- Test Network IInfrastructure
- Results of Experiments
- Conclusions

Motivation

- Growing Interfaces to the Internet
  - Internet of Things (IoT)
  - Cloud Services
  - Industry 4.0
  - Medical IT devices/equipment
  - BYOD
- Isolated islands of IT security solutions
- In most cases: IT security components of different vendors are not working together
- Real-time analysing of all security data is not available



Source: blog.surveyanalytics.com/2014/09/top-5-infographics-of-week-internet-of.html

Security Information and Event Management (SIEM)

- Main issue: overview about security status of the complete network
- Tasks:
  - Collection of security relevant information from the network
  - Assessment of the information
  - Prioritisation of the assessed information
  - Generation of messages about critical security issues
  - Provision of guidance regarding the handling of critical messages

Source: www.topionetworks.com

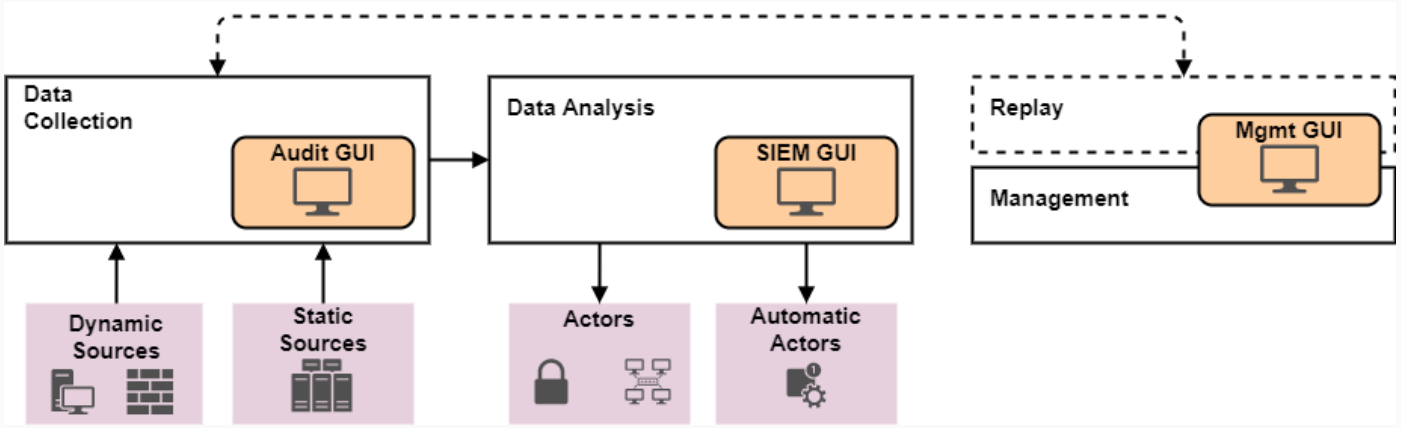
The project GLACIER

- GLACIER = Intrusion detection via multi-dimensional analysis of security data streams
- GLACIER is a cooperation project within the German BMBF with the following partners:
  - DECOIT GmbH (coordinator, developer, and SIEM specialist)
  - University of Applied Sciences of Hanover (research, deep-learning)
  - rt-solutions GmbH (trust consultant, SIEM specialist)
- Associated partner:
  - PLATE (German wholesaler for office supplies)
  - hanseWasser (sewage company from city of Bremen)
- The project has been started at May 2019 and will end at September 2021
- Project website: http://www.glacier-project.de

Focus of GLACIER

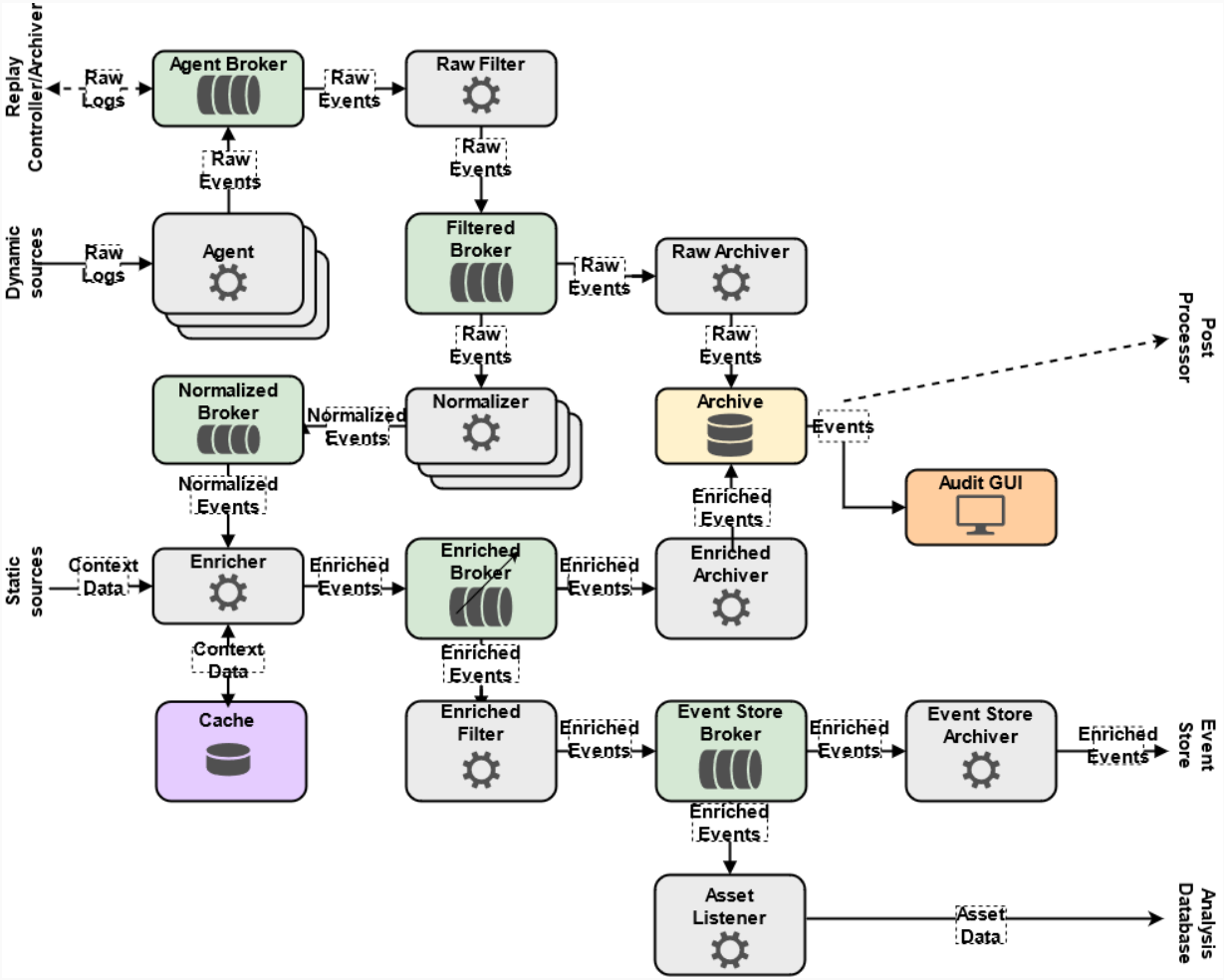
- Develop of advanced concepts for automatic aggregation and analysis of network data related to information security
- Cover all possible data structures to detect a variety of anomalies, automatic aggregation directly yields the view of the data that best displays anomalies
- As the aggregations are generated automatically, the configuration of the system has to be simplified
- Further features are:
  - Efficiently of the main concept
  - Horizontal scalability
  - Presentation of results in a structured format
  - Visualization of all relevant information

Overview of component groups



- *Data Collection:* heterogeneous data is gathered from dynamic or static sources and consolidated as necessary for security analysis (e.g. hosts, firewalls, OT).
- *Data Analysis:* Enriched data are forwarded and analyzed for anomalies. These can be visualized in the SIEM GUI.
- *Management GUI:* All GLACIER components are configured and supervised by the management area. Replay functionality for replaying events to recreate previously encountered situations.

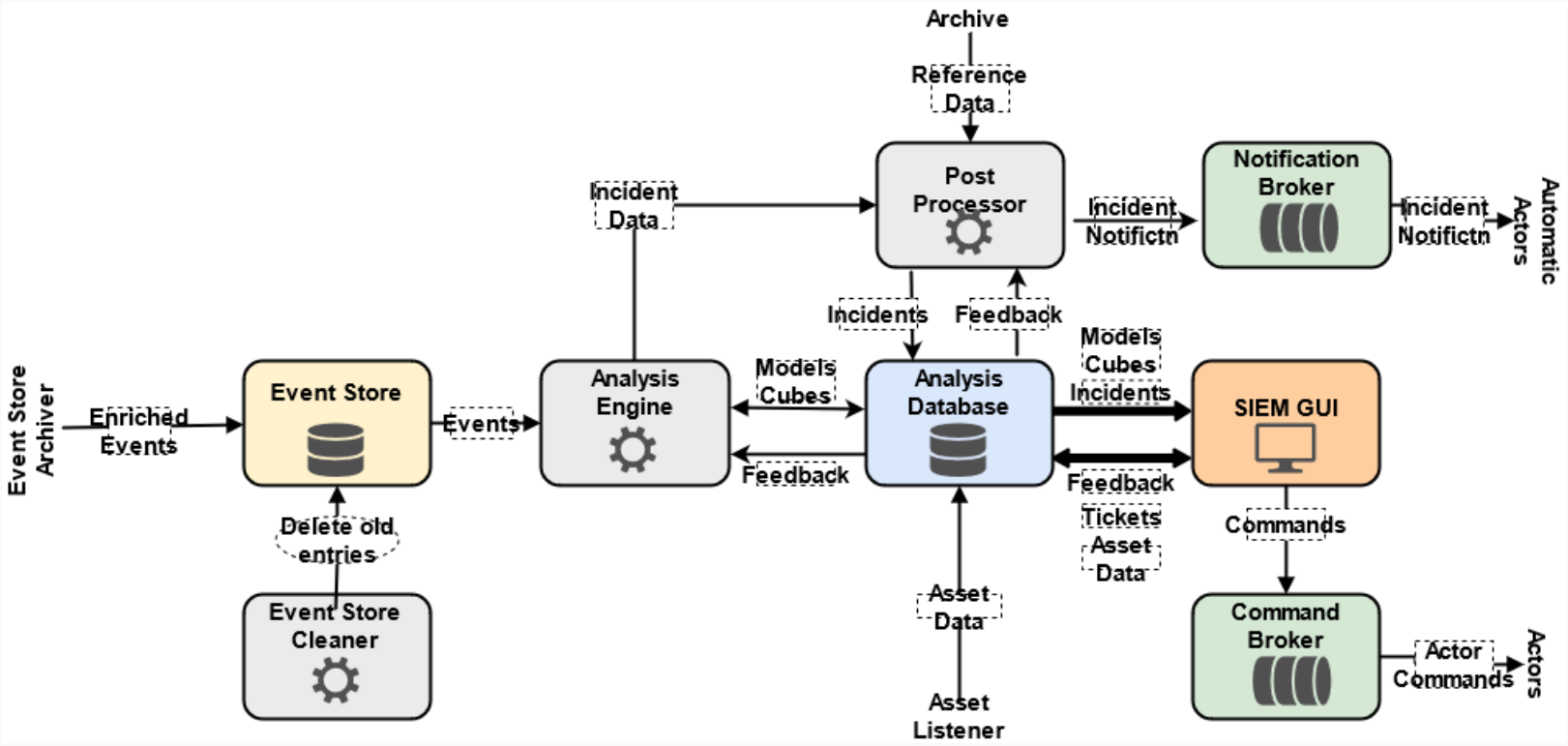
Data Collection Components



Data Collection

- These components collect event data from different points in the network, pre-process and archive it, and finally pass it on to data analysis:
  - *Brokers* are used to buffer events messages in order to decouple different stages of data processing, thus enabling horizontal scalability (RabbitMQ with protocol AMQP)
  - *Archivers* are components that insert data into databases
  - Each *Agent* has a *Normalizer* tasked with transforming its JSON output into a common format to integrate data from all sources
  - *Elastic Common Schema (ECS)* has been used as common format
  - The *Enricher* pick up context information (i.e. user related data from LDAP)
  - *Raw Filter* collects the *Raw Events* produced by the Agents, giving each event an ID which uniquely identifies it across the remainder of the system

Data Analysis Components



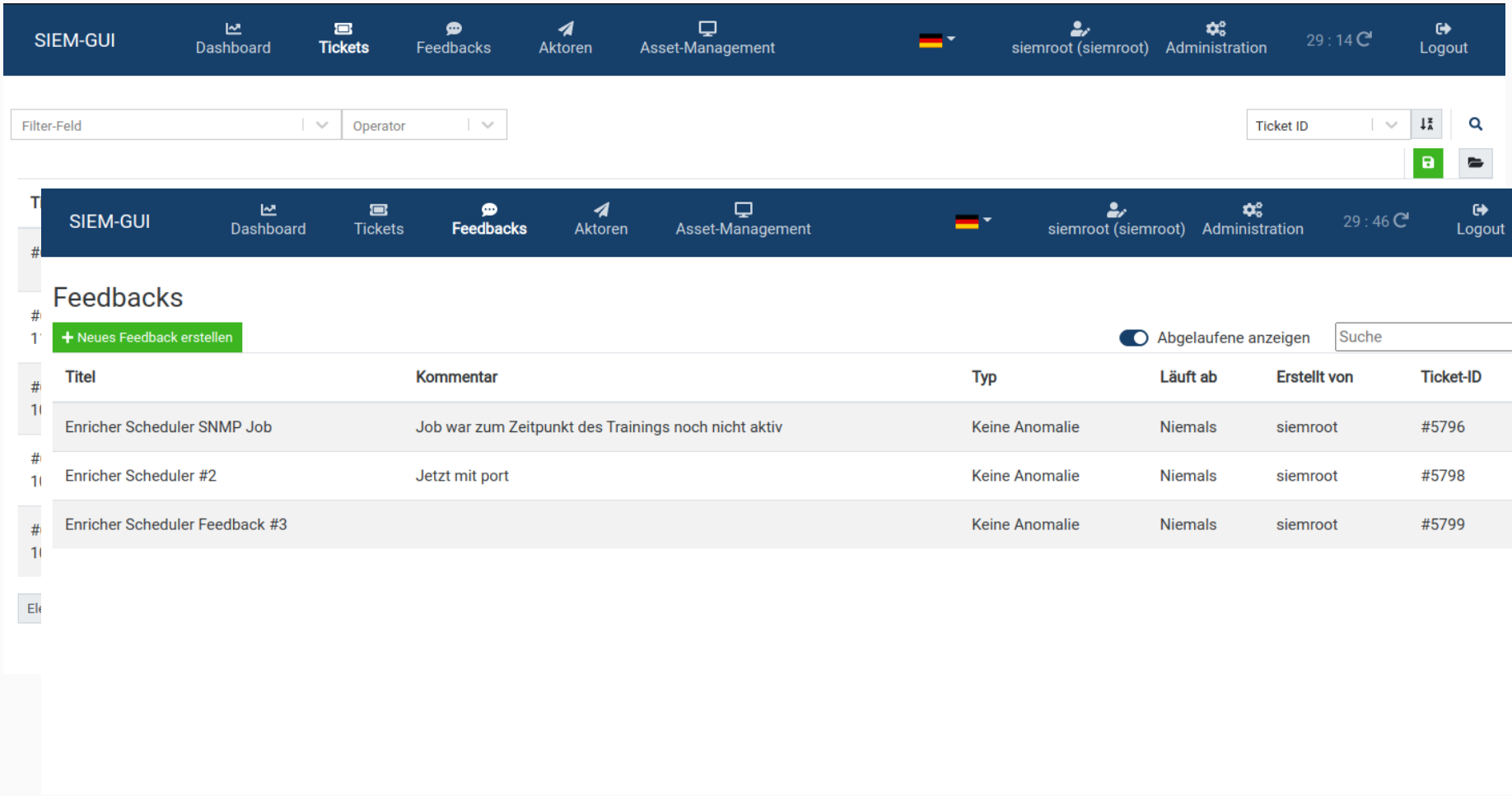
Data Analysis

- These components of the data analysis are tasked with analyzing the gathered events for anomalies and presenting them to users:
  - *Event Store Database:* Events enter the data analysis chain through it
  - This database contains a relatively short history of fully enriched events and offers high-bandwidth access for near real-time analysis (Elasticsearch)
  - *Analysis Engine*: responsible for finding anomalies in the event stream by the use of anomaly detection (deep learning) algorithms (own development and/or Elastic SIEM)
  - *Analysis Database:* Any data analysis results are stored in this database (PostgreSQL)
  - Machine learning models, cubed training, inference data, and user feedback are stored in the analysis database and retrieved if needed

SIEM-GUI visualisation

- Visualisation of incidents with a description
  - It is possible to access related incidents
  - This can be cross-referenced with a visualization of network assets
- SIEM-GUI gives users recommendations for reacting of incidents and access to actors which carry out the reactions (e.g. by moving a host to quarantine zone)
- Users can also give feedback to the analysis engine on each incident (e.g. adjusting the anomaly score)
- The history of actions users merge with one incident are stored as a ticket in the analysis database

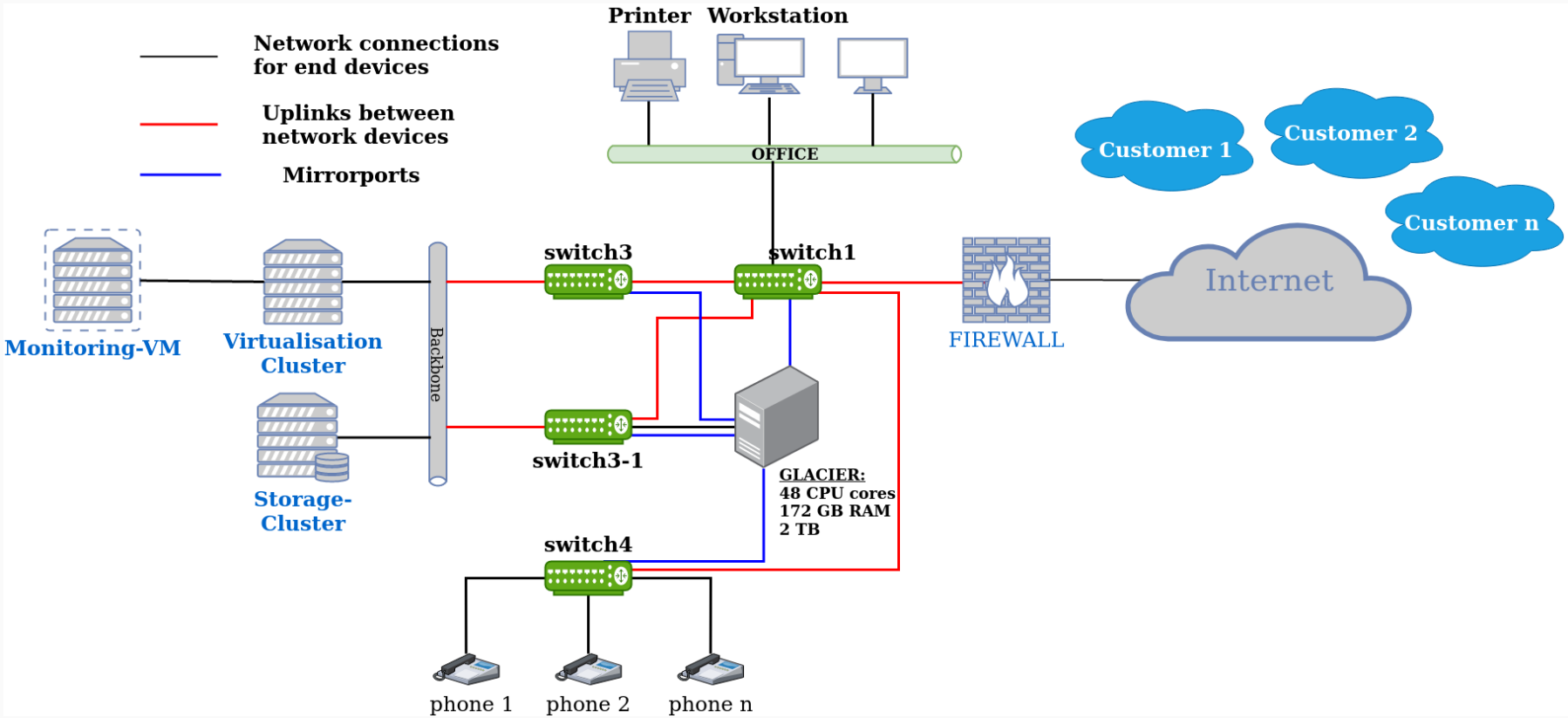
SIEM-GUI surface



Defined Use Cases

- *Use Case 1:* Definition of communication rules (hosts, networks, time restrictions) for detecting violations
- *Use Case 2:* Analysis of logs (Windows event logs and syslog)
- *Use Case 3:* File integrity monitoring (access via SSH or agents on the corresponding system)
- *Use Case 4:* Detection of failed SSH logins
- *Use Case 5:* Vulnerability Scan
- *Use Case 6:* Malware detection in network communication
- *Use Case 7:* Login attempts on Windows server systems
- *Use Case 8:* Detecting new network connections
- *Use Case 9:* Detecting new protocols within the network

Test Network Structure



Test topology of a real company environment

Test of GLACIER architecture (1)

- The tests were performed according to a specific scheme in order to obtain unambiguous results:
  - Running the installation script
  - Creating networks for asset management
  - Starting an vulnerability scan
  - Monitoring the components of availability
  - Collecting data using intrusion detection components (learning phase)
  - Starting the analysis engine (one week later)
  - Adjusting settings for the analysis algorithm and restarting with new training data from an entire week
  - Improving the analysis and restarting with training data from four weeks
  - Test of static analysis (define rules and create rule violations)

Test of GLACIER architecture (2)

- The results can be sorted into different categories:
  - *Analysis Engine:* the statements of most tickets referred to unusually high or low data traffic at certain times. Currently, there are still too many anomalies. This can be corrected with more training data and other settings.
  - *Static Analysis:* generates too many duplicate incidents (tickets). Instead of generating only one incident, 70 tickets were written during the tests.
  - *Vulnerability Scan:* the vulnerability analysis generated tickets that point to CVE vulnerabilities. All CVE vulnerabilities that were known were found. All of them were non-critical.
  - *Asset Management:* successfully captured the assets in the network. The ports and protocols used were shown per asset. Tickets referenced the assets found and could be used to track anomalies.

Test results

- Overall, the tests showed that the SIEM architecture worked with the self-developed analysis engine well enough.
- However, too many tickets were still generated, which is an area for improvement.
- Also, duplicate tickets should be avoided in the future by means of adding a duplicate detection mechanism early on.
- For a efficient anomaly detection it is necessary to use a big amount of data in the implemented databases.
- Grow of data traffic with the usage of more advanced hardware components (e.g. 10 Gbps for Ethernet) is predictable for the future.

Conclusions

- The GLACIER architecture will provide the required features for security-based anomaly detection in IT and OT environments
- More sensors have to be added to the system to improve anomaly detection results for further options for describing the normal system state and in consequence analyze potential deviations
- In the data analysis component group the focus in future developments will be on improving the analysis engine at its core
- More realistic and comprehensive training data sets will also be essential for improving the analysis component
- By integrating information from a wide range of input sensors and by using multidimensional anomaly detection algorithms the system is able to detect modifications that could not have been detected previously

Thank you for your attention!

DECOIT GmbH
Fahrenheitstraße 9
D-28359 Bremen

https://www.decoit.de
info@decoit.de