Using extensible metadata definitions to create a vendor-independent SIEM system

Prof. Dr. Kai-Oliver Detken
(DECOIT GmbH)
*
Dr. Dirk Scheuermann
(Fraunhofer SIT)
*
Bastian Hellmann
(University of Applied Sciences and Arts of Hanover)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Agenda

- Definition of SIEM systems
- The research project SIMU
  - SIMU architecture
  - Interface for Metadata Access Points (IF-MAP)
  - Collector and Flow Controller Components
- Requirements and Strategies for Metadata Definition
  - Metadata definition
  - Requirements for the Data Model
  - Strategies for additional data definition
  - Process of metadata conception
- Data Model for Non-Proprietary SIEM Systems
  - ESUKOM data model example
  - General data model
  - Domain instances for anomaly detection
  - ESUKOM feature model example
  - SIMU data additions example
  - Vendor-independent SIEM
- Conclusions

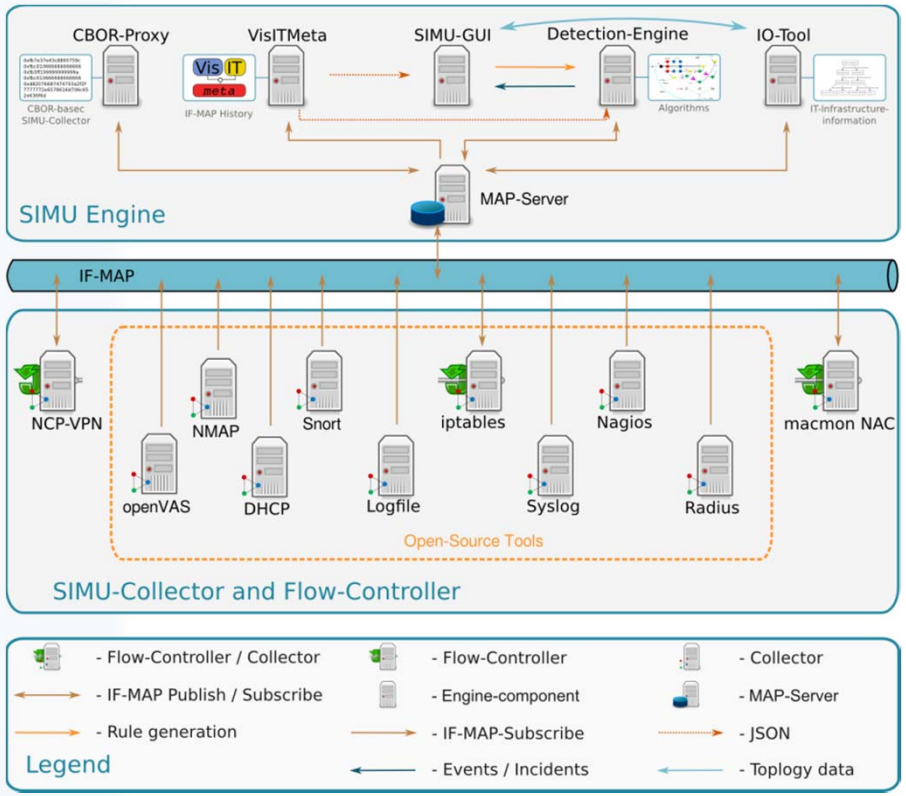
Definition of SIEM systems

- Security information and event management (SIEM) is a term for software products and services combining
  - Security Information Management (SIM) and
  - Security Event Management (SEM)
- SIEM technology provides in detail real-time analysis of security alerts, which have been generated by network hardware and applications
- SIEM is also applied to log security data and generate reports for compliance purposes

The objective of SIEM is to help companies respond faster to attacks and organize mountains of log data

The research project SIMU

SIEM architecture of SIMU



Layers of the architecture

- The *SIMU collector* and *flow-controller layer*, where the components are responsible for data collection and enforcement
- The *SIMU engine*, which includes the components for central data and knowledge storage, the data correlation, aggregation, and visualization of data, as well as interfaces to other protocols
- The layers base on the architecture of the project *ESUKOM* (www.esukom.de)

IF-MAP Specification

- IF-MAP is an open standard, client-server based protocol by the Trusted Computing Group (TCG) for sharing arbitrary metadata across arbitrary entities
- IF-MAP can provide the following benefits:
  - Integration of existing security systems by a standardized, interoperable network interface
  - Avoidance of isolated data silos within a network infrastructure
  - Extended functionality of existing security tools (e.g. automatic responses on detected intrusions, identity-based configuration of packet filters)
  - Vendor independence
- IF-MAP is the basis protocol for SIMU and previously for the project ESUKOM

Collector and Flow Controller

IF-MAP environment

- Flow controller and collectors are typical security components and services in a network infrastructure
- They collect information or manage the network behavior
- Several clients have been adopted to support integration into an IF-MAP environment
- SIMU components work on basis of ESUKOM components

IF-MAP-Clients

- Collectors
  - DHCP
  - RADIUS
  - syslog
  - Nagios / Icinga REST
  - Snort
  - Nmap
  - OpenVAS
  - Logfile
  - LDAP
  - Android
- Flow controller
  - iptables
  - macmon NAC
  - NCP-VPN
  - OpenVPN

Metadata Definition

- Metadata plays an important role for securing network applications
- The TCG already established specifications providing large amounts of standardized metadata and identifiers useful for network security
- But, the existing standard data schemes often do not provide appropriate types for all **data objects** relevant for a certain application like a desired SIEM system
- Additionally, the design of new applications (e.g. Android App) often requires the definition of additional and domain-specific metadata
- That means the existing **data model** has to be extended

Requirements for a Data Model

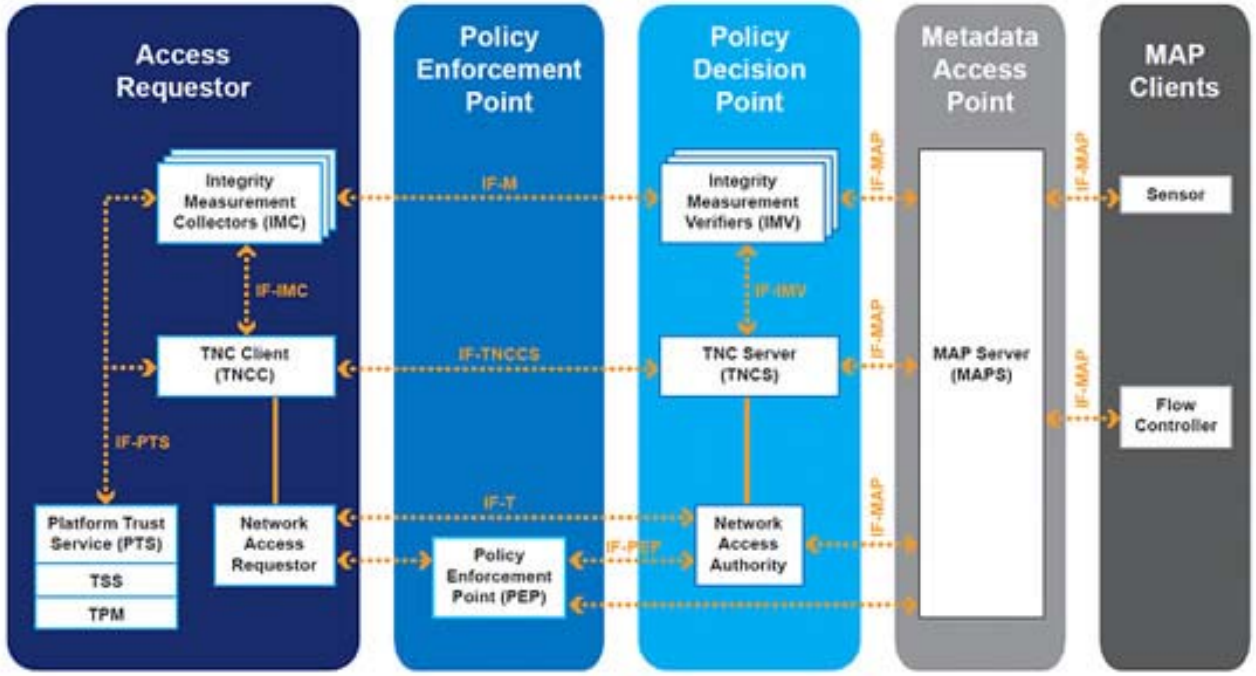
- **Integration of arbitrary metadata:** Metadata from multiple different domains must be used within the new data model. As different components have a specific view onto a network, the data model has to be flexible and non-restrictive in terms of which values can be expressed.
- **Technology independence:** The model itself has to be independent of any concrete technology. An implementation of the data model has to ensure that all needed components can exchange the data in a platform on independent way.
- **Allowing enlargements:** To allow the use of our model in future use cases and scenarios, the model itself has to be extensible. Thus, the definition of data has to be done in a flexible way.
- **Covering all intended use cases:** All previously identified concepts and key features need to be represented by the data model. The model has to be able to include all metadata that is needed to solve the key features.

Strategies for additional data definition

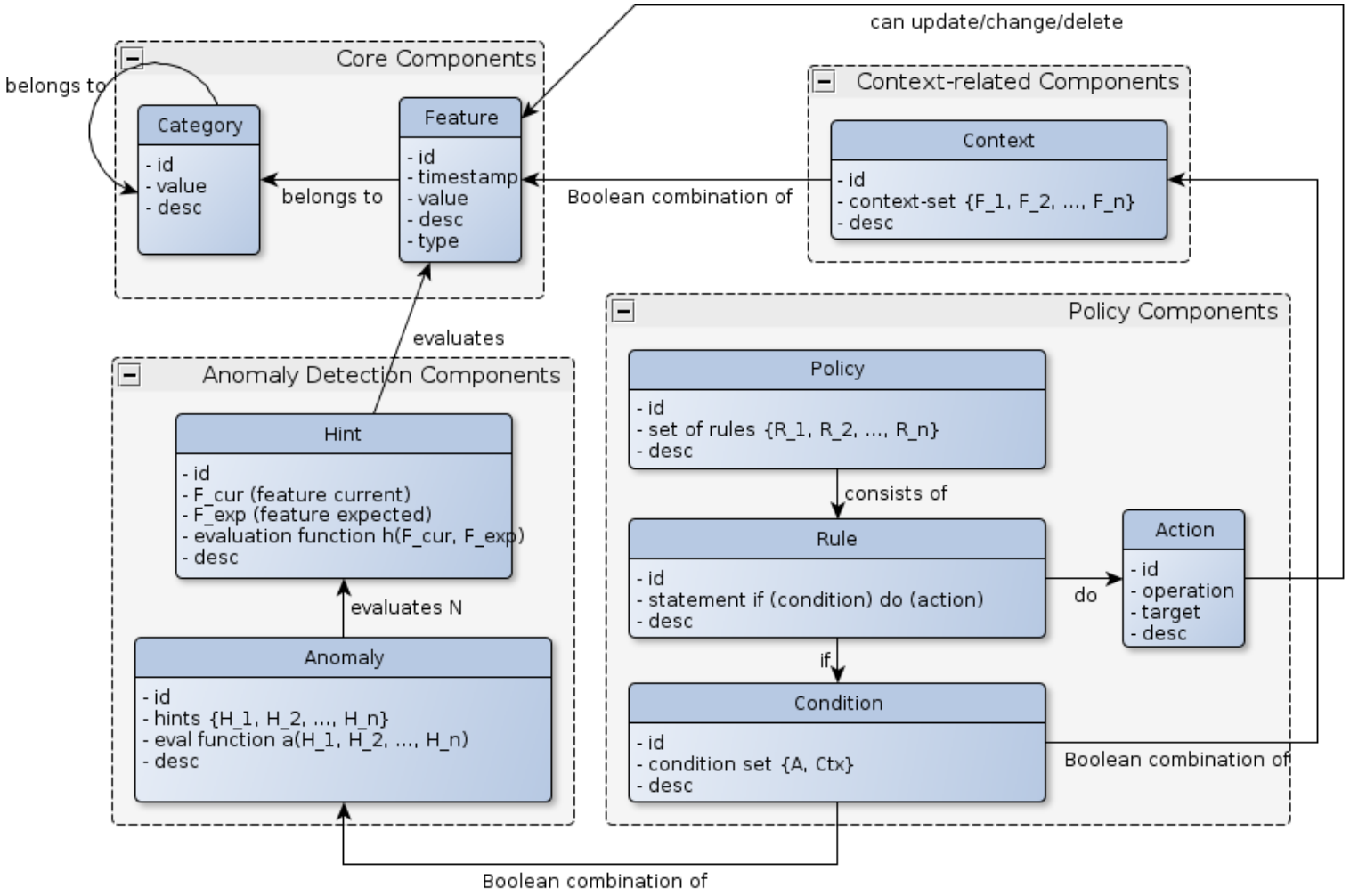
- **Additional vendor specific data:** use already existing functionality for extension within a specification. Although this would leave the specification as it is, it will not be appropriate for future standard applications. Interoperability with non-SIMU components would be a problem.
- **Enlargement of specification:** The original specification can be extended by the types and attributes of the data model. This process usually takes a long time, as changing a specification involves multiple rounds of review by the corresponding working group, and has the disadvantage that it cannot always be done.
- **Define a standardized way to enhance metadata specifications:** A compromise of the two previous suggestions is to encourage the specification of the working group to adjust their own policies for enlargement so that additions like the ones of the SIMU project can be easily added to the specification.

Process of metadata conception

- **Definition of a generic model:** according to the requirements of the actual problem domain, a generic data model that holds all required information has to be created.
- **Mapping onto IF-MAP:** this generic model then has to be mapped onto IF-MAP, that is onto new identifiers and metadata types.



ESUKOM data model example



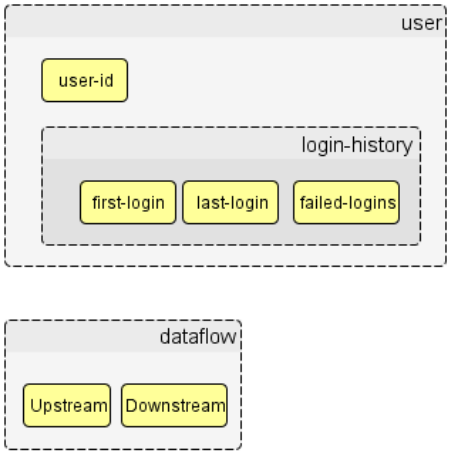
ESUKOM Data Model Components

General Data Model

- **Specification of data objects:** Identifiers shall be defined for instances publishing and subscribing data as well as formats for the exchanged data (metadata).
- **Specification of anomalies:** Abnormal system states shall be represented by combinations of certain data values.
- **Specification of policies:** Actions shall be taken if certain system states, including anomalies, are detected.

Domain instances for anomaly detection

- Domain instances including categories, features, signatures, policies etc. with assigned values
- Example: domain instances that define a smartphone with its operating system version, its apps, the apps permissions, etc. can be used to detect anomalies by the behavior of the device



ESUKOM feature model example

- **Definition of a unique metadata type „feature“**
  - Contains a **name attribute** for a specific feature (metadata type)
  - Contains a **type attribute** to distinguish between quantitative (concrete metered value), qualitative (enumeration) or arbitrary (any string)
  - Contains a **value attribute** for metadata value
  - Contains an **attribute group** ContextParameters
- May be applied to any new datatype needed and not present in current IF-MAP specification!
- Data type to be proposed for enlargement of IF-MAP specification!

SIMU data additions example

- **Definition of a identifier set for services:**
  - **Service identifier:** a definition of a generic service, like SSH server
  - **Vulnerability identifier:** defines a specific vulnerability (e.g. CVE-2014-1692)
  - **Implementation identifier:** defines an actual implementation of a service (e.g. openssh v1.5.9)
- Helps to represent and detect attacks on actual vulnerable services in a network

Vendor-independent SIEM

- With the possibility to both generate own metadata (using the *feature definition*) and IF-MAP in general as a data and communication protocol, a SIEM-like system integrating arbitrary components can be designed
- Additional components can easily be added by using new, specific metadata or reusing already existing definitions
- Homogenization of data is implicitly done by using IF-MAP

Conclusions

- A standard specification was extended by two strategies:
  - Feature model approach (from the project ESUKOM)
  - Service Identifier (from the project SIMU)
- This *feature model* approach makes it possible to create a flexible proprietary data model if needed
- But it is recommended to include the *service identifier* into future versions of the TCG specifications, given that it is as basic as other standard identifier types, such as *ip-address* or *device*, when describing the state of a network with IF-MAP
- As a final conclusion, it is recommended to use the existing specification description of the TCG regarding interoperability with other IF-MAP components
- If an extension is needed it would be useful to integrate this definition also into the standard specification
- If that is not possible the *feature model* solution can be used
- But that includes interoperability lacks between different IF-MAP components and will work only in proprietary environment

SIMU

Thank you

…for your attention!

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

References

- SIMU project website, http://www.simu-project.de
- Federal Ministry of Education and Research, http://www.bmbf.de/en/index.php
- Jamil, A.: The difference between SEM, SIM and SIEM. 29th July (2009)
- Williams, A.: The Future of SIEM – The market will begin to diverge. 1st January (2007)
- TCG: TNC IF-MAP Metadata for Network Security. Trusted Network Connect, Specification Version 1.1, Revision 8, Trusted Computing Group (2012)
- TCG: TNC IF-MAP Binding for SOAP. Trusted Network Connect, Specification Version 2.2, Revision 9, Trusted Computing Group (2014)
- Birkholz, H., Sieverdingbeck, I., Sohr, K., Bormann, C.: IO: An interconnected asset ontology in support of risk management processes. IEEE Seventh International Conference on Availability, Reliability and Security, Page 534-541 (2012)
- M. Shahd, M. Fliehe: Fast ein Drittel der Unternehmen verzeichnen Cyberangriffe. BITKOM news release from 11th of March 2014, CeBIT, Hanover (2014)
- ESUKOM project website, http://www.esukom.de
- Ahlers, V., Heine, F., Hellmann, B., Kleiner, C., Renners, L., Rossow, T., Steuerwald, R.: Replicable security monitoring: Visualizing time-variant graphs of network metadata. Joint Proceedings of the Fourth International Workshop on Euler Diagrams (ED 2014) and the First International Workshop on Graph Visualization in Practice (GViP 2014) co-located with Diagrams 2014, number 1244 in CEUR Workshop Proceedings, pages 32-41 (2014)

Copyright 2013-2015

Das dem Projekt zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen „16KIS0041K“ gefördert. Die Verantwortung für den Inhalt liegt bei den Autoren.