PROJECT

Multi-tenant cybersecurity operations centre designed
as an open cloud service with machine learning elements

Linked to work package KPB1 – Design of the virtualization solution

Linked to output
V3 – Description of the selected solution meeting the specified criteria

type – documentation

Funded by the European Union
NextGenerationEU

RECOVERY AND RESILIENCE PLAN

GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

Analysis II – Graylog

Funded by the European Union
NextGenerationEU

RECOVERY AND RESILIENCE PLAN

GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

CONTENTS

1 Graylog 4
1.1 Graylog products 4
1.2 Main functions 5
2 Installation requirements 6
2.1 Graylog stack 6
2.1.1 Ports 6
2.1.2 Recommended system requirements 6
2.1.3 Shard 7
2.1.4 Compatibility 7
3 Description of functions 8
3.1 DATA LOGGING 8
3.1.1 Logging inputs 8
3.2 Data Routing 8
3.2.1 Streams 8
3.2.2 Pipelines 8
3.2.3 Destinations 8
3.3 Data search 9
3.4 Data visualization 9
3.5 Events and alerts 9
3.5.1 Event 9
3.5.2 Alert 9
Sources 9

1 GRAYLOG

Graylog is a powerful security information and event management (SIEM) solution that offers a robust log analysis platform, simplifying the collection, search, analysis and alerting on all types of machine-generated data. It is specifically designed to capture data from various sources, which makes it possible to centralize, secure and monitor logs effectively. Graylog can perform a wide range of cybersecurity functions, such as:

- Data aggregation
- Security data analysis (reports and dashboards)
- Security event monitoring
- Forensic analysis
- Incident detection and response
- Console for responding to real-time events or alerts
- Threat intelligence
- User and entity behaviour analytics (UEBA)

1.1 GRAYLOG PRODUCTS

Graylog offers several versions to choose from. Each version offers different log management solutions. However, they require a paid licence for use, although an open-source version that is freely available is also offered.

Graylog versions:

1. **Graylog Open**: A free open-source version providing basic log management with data collection, enrichment and analysis capabilities.
2. **Graylog Enterprise**: Includes all the capabilities of Graylog Open while adding advanced features such as log archiving, user auditing, script-based notifications, scheduled reports and technical support. This package is intended for organizations with complex IT infrastructures.
3. **Graylog Security**: An add-on to the Enterprise version focused on cybersecurity. It enables the detection and investigation of, and response to, security incidents, with threat intelligence and user behaviour monitoring functionality.
4. **Graylog Cloud**: Combines the advantages of Graylog Enterprise and Security with the flexibility of a cloud environment. It is a fully managed service that allows rapid deployment, scaling and updates without the need for one's own infrastructure.
5. **Graylog API Security**: A product intended for API protection; it provides a detailed overview of API activity and identifies any misuse.
6. **Graylog Illuminate**: An add-on package for the Enterprise and Security versions that provides pre-built content (pipelines, rules, lookup tables) for processing logs for specific operations.

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

1.2 MAIN FUNCTIONS

Main functions offered by the Graylog solution:

Streams: They serve as tags for incoming messages and categorize them in real time. Rules can be used to determine which stream a message will be routed to.

Search: The Graylog Search page allows direct searching in logs using a simple Lucene-like syntax. Time ranges can be relative or absolute, and the results can be saved or displayed as aggregations that can be added to dashboards.

Dashboards: contain visualizations or summaries of information found in logs. A dashboard can be customized with widgets that visualize data (counts, averages, charts, maps), and role-based access control is available.

Alerts: They are created using event rules. Once a certain condition is met, an event is triggered, which can activate an alert.

Indices: They serve as the basic data storage units, with indices allowing retention, sharding and replication to be configured for specific data, with the option to customize their retention and rotation.

Graylog Sidecar: This agent manages logging clients that collect logs from server operating systems and send them to Graylog.

Processing pipelines: They allow rules to be applied to specific types of events, whereby messages can be modified, enriched or blocked as needed, which optimizes their processing through Graylog.

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

2 INSTALLATION REQUIREMENTS

2.1 GRAYLOG STACK

The Graylog architecture consists of three components, also referred to as the Graylog stack:

1. Graylog
2. Data node (OpenSearch)
3. MongoDB

Graylog itself is a web interface that serves as a control centre for the management of logging data. It allows users to filter and analyse logs collected from various sources efficiently.

Data Node is a component of the Graylog architecture that is responsible for managing OpenSearch. This function allows Graylog to manage the search backend so that it is not necessary to install and manage OpenSearch separately. It improves the security of the data layer by implementing certificates. It also ensures the correct version of OpenSearch and its extensions.

MongoDB is a NoSQL database responsible for storing metadata and configuration data for Graylog. It is also used to store data for OpenSearch nodes, which ensures data redundancy and access. MongoDB is also scalable, which makes it ideal for large amounts of data. It requires only a small amount of system resources and storage space to run.

2.1.1 Ports

For smooth operation, Graylog requires access to the following ports:

- :9000 – The Graylog API runs on this port, which must be accessible to all Graylog nodes
- :443 – HTTPS port
- :9515 – ChromeDriver for report generation
- :27017 – MongoDB
- :514 – Syslog input
- :1514 – Alternative syslog port
- :4739 – NetFlow input
- :5044 – Beats input
- :5555 – Graylog Extended Log Format (GELF)

2.1.2 Recommended system requirements

System requirements may vary depending on the amount of data expected. The recommended requirements for processing 1–10 GB of data per day are:

- 8-core CPU
- 16 GB RAM

For the data node, the following is recommended:

- 8-core CPU
- 24 GB RAM

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

2.1.3 Shard

A shard is essentially a part of the data in a database or search engine. Shards are key to the performance and scaling of the Graylog system and distribute data across multiple data nodes.

2.1.4 Compatibility

Graylog can be installed on almost any Linux distribution.

When installing Graylog, compatibility with other software must be maintained. Graylog version 6.1 has the following requirements:

- Minimum MongoDB version: 5.0.7
- Maximum MongoDB version: 7.x
- Minimum OpenSearch version: 1.1.x
- Maximum OpenSearch version: 2.15.x

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

3 DESCRIPTION OF FUNCTIONS

3.1 DATA LOGGING

3.1.1 Logging inputs

Logging inputs are used to accept log messages in Graylog. Some basic message types are supported by Graylog; others need to be added by installing plugins. Inputs are further divided into listener and pull inputs.

Listener inputs listen on a port for a given application and send data to Greylog.

Pull inputs communicate with end devices and retrieve logs using an API or other methods.

Common inputs include:

- Ingest syslog
- Ingest journald
- Ingest Windows eventlog
- Ingest CEF
- Ingest Raw/Plaintext
- Ingest GELF
- Ingest from files
- Ingest JSON path from API
- Ingest Application Data

For secure data transmission, TSL encryption can be enabled for individual inputs

3.2 DATA ROUTING

Once logs have been received by Graylog, the next step is to filter, modify and further forward these logs to target destinations.

3.2.1 Streams

Once logs have been received in Graylog, they can be assigned to specific streams, which are specific processes used to filter and route logs through Graylog and beyond it. Rules can be assigned to streams that determine which messages are routed to which stream.

3.2.2 Pipelines

Using pipelines, we can transform and modify messages after they have been sent to a stream. In essence, it is a sequence of processes through which a message passes. These may be operations such as filtering, transformations or forwarding.

3.2.3 Destinations

After processing, the data are sent as output to:

- Index sets – a storage space for logs, where they are stored according to certain properties or needs.
- Outputs – sending data to external systems or destinations.

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC

3.3 DATA SEARCH

Data can be searched in the Graylog web interface. The data found are visualized through various widgets. The logs found can be saved and exported.

3.4 DATA VISUALIZATION

Data are visualized through widgets. Using widgets, we can display data from various perspectives. Data can be displayed in a chart or a table. Aggregation widgets or message table widgets can be used to visualize data.

Aggregation widgets are data shaped so as to clearly answer a specific question. They display only the required fields, functions or metrics. These may be, for example, messages from a certain IP address over the last two days.

Message table widgets display messages and their fields. They provide a more detailed examination of individual messages.

3.5 EVENTS AND ALERTS

3.5.1 Event

An event is a change in the environment that differs from the expected behaviour. It may be, for example, a change in firewall rules or a failed login attempt from a blocked address. In Graylog, parameters can be set for events and, if the data meet the parameters, Graylog raises an alert.

3.5.2 Alert

An alert is a notification that can be created for an event.

SOURCES

https://go2docs.graylog.org/current/what_is_graylog/what_is_graylog.htm

Funded by the European Union NextGenerationEU
RECOVERY AND RESILIENCE PLAN
GOVERNMENT OFFICE OF THE SLOVAK REPUBLIC