# Processing Network Logs in Elasticsearch

# Contents

1. Introduction to network log processing ..... 3

2. Elasticsearch and its role in network log processing ..... 4

3. Alternatives to Elasticsearch for network log processing ..... 5

3.1 Splunk ..... 5

3.2 Graylog ..... 6

3.3 Loggly ..... 6

3.4 Fluentd + Grafana + Loki ..... 6

4. Comparison of the advantages and disadvantages of the individual solutions ..... 6

4.1 Splunk in network log processing ..... 6

4.2 Elasticsearch in network log processing ..... 7

4.3 Graylog in network log processing ..... 7

4.4 Loggly in network log processing ..... 8

5. Properties and requirements of multi-tenant solutions ..... 9

5.1 Requirements for a multi-tenant solution ..... 9

5.2 Elasticsearch in multi-tenant mode ..... 9

5.3 Multi-tenancy in the other solutions ..... 10

6. Conclusion and recommendations ..... 10

6.1 Key findings ..... 10

6.2 Recommendations ..... 10

# 1. Introduction to network log processing

In today's digital world, continuous monitoring and analysis of network traffic is an essential part of IT infrastructure management. Every system connected to the network generates a large amount of data in the form of network logs – records of network communication, connections, requests or security-related events. These logs originate, for example, from firewalls, routers, servers, IDS/IPS systems or operational monitoring tools.

Network log processing enables:

- **Threat detection and prevention** – early identification of attacks such as DDoS, port scanning or unauthorized access;
- **Network monitoring** – monitoring the status and load of the infrastructure;
- **Audits and forensic analysis** – reconstruction of events after security incidents;
- **Compliance with legislative requirements** – activity records required for GDPR, ISO 27001 and similar standards.

Given the growing volumes of network data, efficient log processing is a challenge. Traditional databases are often insufficient due to inadequate performance when indexing and searching large amounts of data. Therefore, specialized tools such as **Elasticsearch**, which are designed for distributed, real-time log processing and analysis, are increasingly being used.

The aim of this work is to explore the possibilities of using Elasticsearch for network log processing, to describe the solution architecture, the method of data processing and visualization, as well as the advantages and limitations of this approach compared with the alternatives.

# 2. Elasticsearch and its role in network log processing

Elasticsearch is a powerful, distributed search and analytics engine based on the Apache Lucene library. In the field of network log processing, it plays a key role in storing, indexing and analysing data in real time. Within the **Elastic Stack** (ELK: Elasticsearch, Logstash, Kibana), it serves as the central storage and search engine for network events originating from firewalls, IDS/IPS systems, routers, switches, proxy servers and other devices.

**Key features of Elasticsearch in network log processing:**

- **Fast search and analysis**: Elasticsearch enables instant searching based on IP addresses, ports, protocols, country of origin and timestamps, which is important in the detection of network threats.
- **Horizontal scaling**: Thanks to its distributed architecture, it is capable of processing millions of network events per day, which is essential for larger networks and organizations.
- **Flexible data model**: It allows semi-structured data (e.g. Netflow, syslog, Suricata JSON) to be stored without the need for a fixed schema.
- **Integration with collection tools**: In conjunction with tools such as Filebeat, Logstash or Packetbeat, it can ingest diverse network logs and transform them into a usable format.
- **Visualization via Kibana**: It enables the creation of clear dashboards showing the number of connections, anomalies, the geographical distribution of attack sources and more.

**Challenges and limitations:**

- **Complexity of configuration and administration**: Managing large Elasticsearch clusters with multiple indices for different log sources requires expert knowledge.

- **Infrastructure demands**: With large data volumes (e.g. Netflow from an enterprise-wide network), substantial disk and memory resources are required.
- **Security and data isolation**: Without the use of X-Pack or OpenSearch Security, it may be necessary to handle authentication, authorization and encryption separately.

**Examples of use with network logs:**

- **Anomaly detection**: Recording a large number of connections from a single source (e.g. DDoS, port scanning).
- **Search by IP or subnet**: Rapid identification of malicious activity based on a known IP address.
- **Connection geolocation**: Identification of suspicious connections from countries that do not normally connect.
- **Forensic analysis**: Retrospective examination of events along a timeline and using a combination of multiple filters.

# 3. Alternatives to Elasticsearch for network log processing

Although Elasticsearch is one of the most widely used tools for log processing and analysis, there are several alternative solutions that provide similar functionality – some with a higher degree of simplicity, others with extended security capabilities or integrated analytical functions.

When choosing a technology for network log processing, it is important to consider factors such as:

- the ability to work with formats such as **Syslog**, **Netflow**, **Suricata** or **Zeek**,
- performance when processing large volumes of logs in real time,
- the level of visualization and event correlation,
- support for a multi-tenant or distributed environment.

The main alternatives to Elasticsearch include in particular:

## 3.1 Splunk

Splunk is a commercial tool focused on processing and analysing large amounts of data, including network logs. It supports log ingestion from various sources, threat detection,

visualization and the creation of correlation rules. It has a powerful query language (SPL – Search Processing Language).

### 3.2 Graylog

Graylog is an open-source solution designed for centralized log management. It enables log-based collection, normalization, analysis and alerting. It is suitable for network logs thanks to native support for the Syslog format and integration with tools such as Suricata or Zeek.

### 3.3 Loggly

Loggly (from SolarWinds) is a cloud-based log processing service. It supports a wide range of inputs, including network sources, simple searching and the creation of visualizations. It is particularly suitable for smaller teams and rapid deployment.

### 3.4 Fluentd + Grafana + Loki

An alternative to the classic ELK Stack is the combination of **Fluentd** (for log collection), **Grafana Loki** (for indexing and storage) and **Grafana** (for visualization). This approach provides a more easily deployable and less resource-intensive solution for cloud-native infrastructures.

## 4. Comparison of the advantages and disadvantages of the individual solutions

In this chapter, we focus on comparing the advantages and disadvantages of four main solutions for transaction log processing: **Splunk**, **Elasticsearch**, **Graylog** and **Loggly**.

### 4.1 Splunk in network log processing

**Splunk** is a robust commercial solution designed for the analysis of logging and telemetry data. In the field of network security and network log processing, it is used mainly for:

- attack detection,
- event management (SIEM),
- correlation of data from multiple sources (firewalls, IDS/IPS, proxy, VPN, etc.),
- the creation of customized detection rules.

**Advantages:**

- A powerful query language (SPL) with the ability to correlate multiple types of data.
- An intuitive user interface with advanced dashboards.
- A high degree of customizability – alerts, detections, Machine Learning add-ons.
- Support for large environments, integration with the cloud, AD, syslog, Netflow, etc.

**Disadvantages:**

- Licensing based on data volume – can be very costly for large networks.
- Higher hardware requirements with a large amount of indexed logs.
- Closed ecosystem – limited openness compared with open-source solutions.

## 4.2 Elasticsearch in network log processing

**Elasticsearch** is often the choice for open-source network log processing solutions. It is used in combination with **Logstash**, **Filebeat**, **Packetbeat** and **Kibana**. It is suitable for:

- centralized log collection (syslog, firewall, proxy, IDS...),
- real-time attack analysis,
- visualization of data flows and suspicious activities.

**Advantages:**

- Fast indexing and searching of large volumes of logs.
- Horizontally scalable, suitable for distributed systems.
- Open and modular – numerous plugins, Beats, integrations.
- Suitable for real-time analysis and visualization using Kibana.

**Disadvantages:**

- Requires advanced configuration and expertise in larger deployments.
- Less straightforward correlation between logs of different types without custom development.
- Without X-Pack or OpenSearch Security, it has limited built-in security.

## 4.3 Graylog in network log processing

**Graylog** is an open-source solution for centralized log management that is popular mainly for its simplicity, extensibility and good support for network logs. It allows data to be received via

**Syslog**, **Beats**, **REST API** and other protocols. It is suitable for analysing logs from firewalls, routers, VPN gateways, IDS/IPS or proxy servers.

**Advantages:**

- Simple deployment and administration – suitable for medium-sized and smaller organizations.
- Web UI with search, filtering and dashboard capabilities.
- Support for **stream-based alerting** – incoming logs are evaluated in real time.
- Possibility of log enrichment (e.g. with geolocation data, DNS resolution).

**Disadvantages:**

- Less powerful with very large log volumes compared with Elasticsearch or Splunk.
- Some advanced features are available only in the enterprise version.
- Less extensive documentation and community compared with the Elastic Stack.

## 4.4 Loggly in network log processing

**Loggly** is a cloud-based solution for centralized log collection and analysis that is part of the SolarWinds portfolio. It is oriented mainly towards ease of use, rapid deployment and integration with applications and infrastructure. For network logs, it supports receiving data via the Syslog protocol or REST API.

**Advantages:**

- Rapid configuration without the need for one's own infrastructure.
- Intuitive UI with basic charts, filters and search.
- Support for integrations with various platforms (AWS, Heroku, Docker, etc.).

**Disadvantages:**

- Limited functionality for advanced analysis or data correlation.
- Considerably less customizable than Elasticsearch or Splunk.
- The price depends on the volume of logs received and the required retention

# 5. Properties and requirements of multi-tenant solutions

Multi-tenant log processing means that multiple independent users (tenants) – for example departments, branches, customers or projects – share a single infrastructure for the collection, processing and analysis of logging data. In the context of **network logs**, multi-tenant solutions are important mainly for:

- **managed security service providers (MSSPs)**,
- **organizations with separate security zones or infrastructures**,
- **SaaS providers** that need to separate access to the logs of individual clients.

## 5.1 Requirements for a multi-tenant solution

Effective multi-tenant log processing must ensure:

- **Data isolation** – each tenant has access only to its own logs;
- **Sharing of computing resources** – central management of the infrastructure for all tenants;
- **Auditability** – the ability to check retrospectively who has worked with the data;
- **Secure access** – fine-grained management of permissions (e.g. using RBAC/DLS);
- **Customizable visualizations** – each tenant has its own dashboards, alerts and filters.

## 5.2 Elasticsearch in multi-tenant mode

Elasticsearch does not natively support multi-tenancy as an isolated container system, but it allows several approaches:

- **Separate indices for each tenant** – each tenant has its own logical index or index pattern (e.g. customer1-*, deptA-*).
- **Shared index with a label (tenant_id)** – all data in a single index, but access is filtered using **Document Level Security (DLS)**.
- **Integration with OpenSearch Security / Elastic Security** – these layers make it possible to define permissions down to the level of a field or record type.

The challenge in the case of Elasticsearch is managing a large number of indices and ensuring scalability without affecting performance.

## 5.3 Multi-tenancy in the other solutions

- **Splunk**: supports multi-tenant access by means of so-called "apps" and "roles", with access to data being restricted at the level of the sourcetype, the index or a specific application.
- **Graylog**: provides multiple "streams", and each user can be granted access only to selected streams (filtered according to rules).
- **Loggly**: offers limited multi-tenancy – better suited to a single client or simple data separation.

# 6. Conclusion and recommendations

In this work, we focused on comparing the options for processing **network logs**, with an emphasis on the use of the **Elasticsearch** tool and its alternatives. Network logs originating from firewalls, IDS/IPS systems, routers, proxy servers or other security devices constitute a key source of information for infrastructure security, incident analysis and regulatory compliance.

## 6.1 Key findings

- **Elasticsearch** proves to be a powerful and flexible solution for real-time network log processing, with support for horizontal scaling, fast searching and integration with visualization tools such as Kibana.
- **Alternatives such as Splunk, Graylog or Loggly** offer different levels of functionality – from enterprise-grade security analytics to simple cloud-based log management systems.
- **Multi-tenant solutions** are increasingly important for large organizations and service providers. Elasticsearch allows several approaches to the logical separation of data, but requires a well-thought-out design of permissions and architecture.

## 6.2 Recommendations

- **For smaller teams or an easily scalable solution**, we recommend Elasticsearch in combination with Filebeat/Logstash and Kibana as an open and well-documented foundation.

- **For enterprise environments with complex audit and compliance requirements**, Splunk or Graylog Enterprise may be more suitable, especially if SIEM functionality is required.
- **When designing a multi-tenant environment**, it is important to define in advance the method of data isolation, index planning, the data lifecycle and the permission system (e.g. using DLS/RBAC).
- **Regardless of the tool used**, network log processing should be part of the organization's broader security strategy – including incident detection, forensic analysis and threat response.