ELK Lab — Log Collection & Detection Exercise

Prepared by: Lokesh

Executive Summary

This ELK Lab exercise demonstrates core Security Operations Center (SOC) capabilities: centralized log collection, normalization, visualization, and basic detection. Using a controlled lab environment, FTP activity from a vulnerable host was exercised to generate relevant security events. These events were ingested into an ELK (Elasticsearch, Logstash, Kibana) pipeline and analyzed through purpose-built dashboards. The exercise validates the end-to-end pipeline from event generation through ingestion, visualization, and initial triage, and highlights practical considerations for detection tuning and enrichment in a production SOC.

Objective

The primary objective of this exercise was to validate the ELK stack's capability to:

- Receive and reliably store host and service logs from a target system.
- Normalize and index log data into a searchable format suitable for correlation and analytics.
- Provide timely visualizations that support detection of suspicious authentication activity and anomalous source IP behavior.
- Produce artifacts and evidence suitable for incident triage and reporting.

Background

Centralized logging and timely analysis are foundational to SOC operations. Modern SOCs depend on a SIEM or search-and-analytics platform to convert disparate telemetry into actionable intelligence. The ELK stack is commonly used in academic and operational settings for log aggregation (Elasticsearch), lightweight ingestion and transformation (Logstash), and visualization (Kibana). This lab focuses on FTP authentication events as a representative example of authentication abuse and credential-based attacks, which are frequently observed in real-world intrusions.

Methodology (High Level)

A controlled credential-stress scenario was performed in an isolated lab environment to produce authentication related log entries on the target service. Logs containing connection attempts, authentication successes and failures, and session metadata were collected and exported from the target host. Ingestion and normalization were performed within the ELK pipeline so that event fields (timestamp, source IP, username, result) were parse able and indexed for analytics. Dashboards were created to visualize key indicators: top source IPs by event count, failed authentication trends over time. These visualizations were used to assess whether the generated activity produced distinguishable signals that could be surfaced by rule-based alerts or manual triage.

Findings

1. **Successful Ingestion and Indexing:** The ELK pipeline reliably consumed exported FTP logs, and indexed key fields enabling fast full-text search, aggregation, and time-series analysis. Normalization produced consistent field names that facilitated dashboard creation.

2. **Distinct Event Patterns:** Authentication failures clustered around a small set of source IPs during the test window. These clusters were visually prominent in the “Top Source IPs” and “Failed Logins Over Time” panels, demonstrating that repeated credential attempts are easily identifiable with appropriate visualizations.

Analysis

The exercise confirms that ELK is effective for surface-level detection of credential-based attacks when logs are properly normalized and visualizations are thoughtfully constructed. The visual patterns observed (repeated attempts from one or a few IPs within a short timeframe) align with canonical indicators of brute-force or credential-stuffing campaigns. However, the analysis also highlights common operational challenges: the need to tune alert thresholds to the environment, the importance of log enrichment to reduce false positives, and the necessity of integrating contextual data (asset importance, business hours, known-good actors) into the triage pipeline.

Key analytical observations:

- **Temporal signature:** Rapid sequences of failed logins are a high-fidelity indicator if the baseline noise level is low; however, in environments with legitimate automation or transient errors, relying solely on counts will increase false positives.
- **Source reputation matters:** Mapping source IPs to reputation or geolocation aids in prioritization. An internal IP producing multiple attempts requires a different investigation path than a foreign, low-reputation IP.
- **Evidence chain:** The collection of raw server logs, ingestion records, and dashboard screenshots provides a defensible evidence trail for follow-up investigation and reporting.

Lessons Learned

- **Ingestion consistency is critical:** Even small variations in log format across services can complicate normalization. Standardized parsing rules and field naming conventions reduce analysis overhead.
- **Start visual, then rule:** Building dashboards first helps design effective rule conditions. Visual inspection of event distributions should inform alert thresholds.
- **Enrichment reduces noise:** Adding contextual data (asset criticality, geo-IP, threat intelligence) significantly improves triage accuracy and prioritization.
- **Document reproducible steps:** Maintaining reproducible lab documentation and artifacts (logs, screenshots, short procedural summaries) accelerates post-exercise reporting and knowledge transfer.

Conclusion

This ELK lab exercise demonstrates that an ELK-based pipeline can effectively support fundamental SOC functions: log collection, normalization, visualization, and initial detection. The controlled credential-stress scenario produced clear visual signals that are readily translatable into detection rules, while also highlighting practical operational needs—enrichment, tuning, and documented playbooks—necessary for reliable production deployments.