BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF BUSINESS AND MANAGEMENT

FAKULTA PODNIKATELSKÁ

INSTITUTE OF INFORMATICS

ÚSTAV INFORMATIKY

SECURITY ENHANCEMENT DEPLOYING SIEM IN A SMALL ISP ENVIRONMENT

ZVÝŠENÍ BEZPEČNOSTI NASAZENÍM SIEM SYSTÉMU V PROSTŘEDÍ MALÉHO POSKYTOVATELE INTERNETU

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTHOR — Bc. Petr Bělousov

AUTOR PRÁCE

SUPERVISOR — Ing. Petr Sedlák

VEDOUCÍ PRÁCE

BRNO 2019

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF BUSINESS AND MANAGEMENT

Specification Master's Thesis

Department:	Institute of Informatics
Student:	**Bc. Petr Bělousov**
Study programme:	System Engineering and Informatics
Study field:	Information Management
Supervisor:	**Ing. Petr Sedlák**
Academic year:	2018/19

Pursuant to Act no. 111/1998 Coll. concerning universities as amended and pursuant to the BUT Study Rules, by the Director of the Institute, you have been assigned a Master's Thesis entitled:

Security Enhancement Deploying SIEM in a Small ISP Environment

Characteristics of thesis dilemmas:

Introduction
Aim of the Thesis
Theoretical Background
Problem Analysis and Current Situation
Proposals and Contribution of Suggested Solutions
Conclusions
References
Appendices

Objectives which should be achieve:

Comparison of available solutions to enhance security in a small ISP environment using SIEM. Selection of appropriate SIEM system for a given company. Documentation and training materials preparation. Deployment of the selected SIEM on a part of actives. Benefits evaluation.

Basic sources of information:

ČSN ISO/IEC 27001. Informační technologie – Bezpečnostní techniky – Systémy řízení bezpečnosti informací – Požadavky. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2017.

ČSN ISO/IEC 27002. Informační technologie – Bezpečnostní techniky – Soubor postupů pro opatření bezpečnosti informací. Praha: Úřad pro technickou normalizaci, metrologii a státní zkušebnictví, 2017.

CHUVAKIN Anton, Chris PHILLIPS a Kevin SCHMIDT. Logging and Log Management: The Auitative Guide to Dealing with Syslog, Audit Logs, Events, Alerts and Other It 'Noise'. Saint Louis: William Andrew, 2012. ISBN 978-15-9749-635-3.

KENT Karen a Murugiah SOUPPAYA. Guide to Computer Security Log Management. NIST SP, 2015. ISBN 978-14-9475-253-8.

Deadline for submission Master's Thesis is given by the Schedule of the Academic year 2018/19

In Brno dated 28.2.2019

L. S.

doc. RNDr. Bedřich Půža, CSc. Director of the Institute	doc. Ing. et Ing. Stanislav Škapa, Ph.D. Dean

Abstrakt

Diplomová práce se zaměřuje na zvýšení bezpečnosti v prostředí malého poskytovatele internetu nasazením SIEM systému. Dostupné systémy jsou porovnány a zhodnoceny v souladu s požadavky zadávající firmy. Projekt nasazení systému SIEM je navržen, implementován a zhodnocen v souladu s unikátním prostředím firmy.

Klíčová slova

bezpečnost, SIEM, log management, ELK, Graylog, Kyberbezpečnost, ISP

Abstract

This master's thesis is focused on improvement of security in small ISP environment by deploying SIEM system in the company. The available systems are compared and evaluated to cover the requirements. The selected SIEM system deployment is proposed, implemented and evaluated in accordance to the firm's unique characteristics.

Keywords

Security, SIEM, Log management, ELK, Graylog, Cybersecurity, ISP

Rozšířený abstrakt

Diplomová práce je zaměřena na zvýšení bezpečnosti v prostředí malého poskytovatele internetu pomocí systému SIEM. Práce se zabývá konkrétním projektem nasazení systému SIEM ve firmě Master Internet s.r.o.

Začal jsem obecným představením používané terminologie, kde SIEM systém znamená systém pro management bezpečnostních informací a událostí. Představil jsem principy systému SIEM a uvedl definice důležitých příbuzných pojmů.

Provedl jsem analýzu vnějších vlivů na firmu pomocí metody SLEPT a Porterovy analýzy pěti sil za účelem zhodnocení vnějších vlivů působících pro a proti nasazení systému SIEM. Pro zhodnocení vnitřních vlivů jsem provedl interní analýzu pomocí metody 7S. Z analýzy vyplývá, že nasazení systému SIEM zvýší konkurenceschopnost firmy v konkurenčním prostředí trhu poskytovatelů připojení k internetu. Z interní analýzy vyplývá, že firma je na nasazení systému dobře připravená.

Představil jsem firmu, její historii a zjednodušenou organizační strukturu a provedl analýzu současného stavu situace správy logů a bezpečnostních informací. Metodou asistovaného zhodnocení jsem zjistil požadavky kladené na systém zákonnými povinnostmi a jejich plnění. Během analýzy byly vidět zásadní problémy současného řešení a nemožnost naplno systém využívat.

Pro porovnání vhodnosti konkrétních řešení SIEM jsem vytvořil hodnotící tabulku, která je založená na váženém průměru jednotlivých hodnotících kritérií s váhou podle důležitosti daného kritéria pro Master Internet.

Dostupné komerční systémy SIEM je možné rozdělit dle modelu předplatného, prodejce největších komerčních systémů SIEM jsem kontaktoval a na základě příliš vysoké ceny vyřadil z dalšího výběru. Z open-source řešení dostupných zadarmo jsem dvě vybral k podrobnému zkoumání a zhodnocení.

Pro porovnání jsem zvolil ELK stack a Graylog.

Popsal jsem podrobně strukturu a složky ELK stacku skládající se ze služeb Elasticsearch, Logstash a Kibana. Představil jsem jednotlivé služby, jejich rozšíření a používání rozhraní Kibany.

Hodnocení ELK stacku shrnuje a vysvětluje hodnocení vhodnosti ELK SIEMu pro pokrytí potřeb Master Internet. Díky výhodám velmi dobré podpory, dokumentace a

kompatibility s nástroji třetích stran získal ELK SIEM celkový vážený průměr skóre 7,5 bodu.

U Graylog SIEM alternativy jsem opět představil architekturu systému, základy ovládání, jednotlivé části nastavení, které u Graylogu probíhá kompletně přes webové rozhraní.

V hodnocení se pozitivně promítla jednoduchost ovládání a výkon Graylog SIEM. Horší podpora a zálohování snížily celkový vážený průměr na hodnocení 6,9.

Na základě hodnocení a vyřazení komerčních systémů SIEM byl pro vlastní implementaci zvolen systém SIEM ELK.

V následující části práce jsem připravil projekt nasazení systému SIEM pomocí Lewinova modelu řízení změny, rozvrhl jednotlivé fáze projektu, určil jsem sponzora a agenta změny a oblasti změn. Pro celý projekt jsem následně připravil časovou osu pomocí metody PERT a zjistil, že realizace projektu bude s rezervou trvat 40 člověkodní.

Analyzoval a zhodnotil jsem rizika projektu nasazování nového systému SIEM, vytvořil jsem mapu rizik projektu a pro relevantní rizika připravil vhodná opatření pro snížení jejich hodnoty, především opatření kontrolního charakteru.

Poslední část práce popisuje nasazení systému v rámci Master Internet. Na Dell R430 server s kombinací rychlých SSD a kapacitních HDD disků jsem nainstaloval Centos 7 se standardními základními bezpečnostními opatřeními. Elasticsearch, Logstash a Kibana služby v aktuální verzi jsem nainstaloval a nakonfiguroval dle analýz a plánů uvedených v předchozích kapitolách. Všechny zmíněné informace ohledně instalace, modifikací a doporučení co dělat v případě výpadku jsou dostupné i v interní wikipedii Master Internet.

V závěrečném zhodnocení přínosů projektu jsou porovnány náklady na provedení projektu s potenciálními ztrátami způsobenými narušením provozu malého poskytovatele internetu. Velké narušení provozu způsobené bezpečnostním incidentem by bylo o přibližně 140 000 Kč dražší, než je jednorázová cena implementace projektu.

Všechny cíle diplomové práce byly splněny, systém byl úspěšně nasazen do používání ve firmě, sponzor změny je s výsledkem spokojen.

Bibliografická citace

BĚLOUSOV, Petr. *Zvýšení bezpečnosti nasazením SIEM systému v prostředí malého poskytovatele internetu*[online]. Brno, 2019[cit. 2019-05-09]. Dostupné také z: https://www.vutbr.cz/studenti/zav-prace/detail/119793. Diplomová práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce Petr Sedlák.

Čestné prohlášení
Prohlašuji, že předložená diplomová práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 9. května 2019

……
podpis autora

Poděkování

Rád bych poděkoval panu inženýrovi Sedlákovi za odborné vedení mé diplomové práce, konzultace a doporučení nejen k diplomové práci ale i do života.

Dále bych rád poděkoval firmě Master Internet, především technickému řediteli Martinu Žídkovi za umožnění spolupráce na diplomové práci; jakož i kolegům administrátorům za podporu během studia.

Na závěr patří můj neskonalý dík mé ženě a rodině za vytrvalou podporu.

Content

Introduction

The company Master Internet Inc., an internet service provider based in Brno, has many servers, switches, routers and other ICT equipment that interfaces with the internet. Due to always evolving and worsening security risks associated with cyberspace it is crucial to control and whenever possible eliminate possible threats. In order to identify and differentiate between potential incidents and false positives centralized information base from multiple sources is necessary. A SIEM system combines the capabilities of several systems into one platform to collect and categorize information, provide visualizations, aggregations and searches to allow administrators fast action.

Goals

The goal of this thesis is to analyze the requirements of the company regarding SIEM systems, choose a system most suitable to company needs, implement it in the company, prepare documentation and training of other administrators for the company and lastly to evaluate the benefits of the project.

1 Theoretical basis of the work

1.1 Definition of used terminology

SIEM

SIEM (Security Information and Event Management) encompasses a comprehensive approach to security. It combines the SIM (Security Information Management) and SEM (Security Event Management) functions into one complex system.

Fundamentally, SIEM consists of information collection, cleanup, aggregation, correlation, evaluation, reaction and storage.

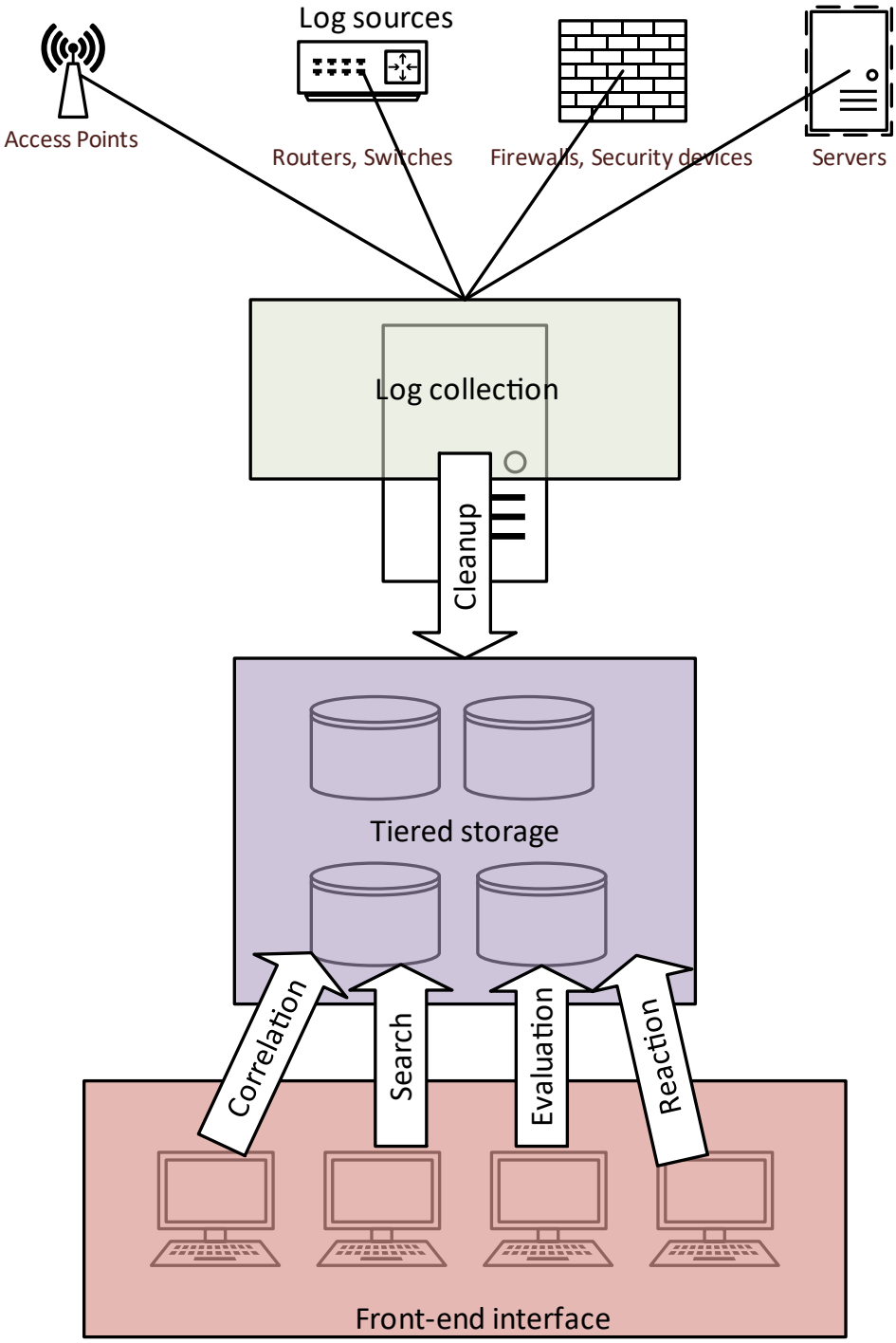

Figure 1 SIEM architecture

(Source: own creation)

A fundamental part of each SIEM is a well-functioning Log Management system.

Log management

"Log management can benefit an organization in many ways. It helps to ensure that computer security records are stored in sufficient detail for an appropriate period of time. Routine log reviews and analysis are beneficial for identifying security incidents, policy violations, fraudulent activity, and operational problems shortly after they have occurred, and for providing information useful for resolving such problems. Logs can also be useful for performing auditing and forensic analysis, supporting the organization's internal investigations, establishing baselines, and identifying operational trends and long-term problems." (1)

Log management needs to deal with a large amount of computer-generated log messages, their collection, aggregation, storage, analysis (real-time and historical), normalization, searching and reporting.

To collect the logs LM utilizes two approaches agentless collection and agent driven.

Linux/Unix based systems support an older syslog protocol service, designed to run on the port 514 which can be utilized for agentless log shipment.

Agent driven log collection utilizes specialized, low overhead log collecting, buffering, transmitting and rotating (the process of maintaining logs for the selected period of time, organizing by date and deleting older than required logs) programs that connect to the central logging server. The logs should be secured in transit to prevent accidental information leak.

Log management considerations

Typical organization does not contain a single type of device with a single log format. That poses several problems. A few of them as well as their respective solutions follow.

- Device connection to logging server
  Small ISP has an advantage in this regard, as the network components between the log management server and the log generating device are controlled and managed by the company itself, thus decreasing the possibility of leak or loss of

information during transit. Transport layer encryption and VLAN separation of the logging messages transport is still recommended to minimize risks.

- Inconsistent log content

  Different log sources deem different information redundant and don't include them in the log. Another problem is the syntax of information (IP address 127.0.0.1 can be represented by dot separated number format or a localhost or LOCALHOST format) Correlation of information is therefore more difficult. Normalization and proper verification are required to eliminate missing information due to different representation.

- Inconsistent timestamps

  Each host generates logs with internal time information. To minimize differences in time all hosts are configured with a unified source of accurate time, the ntp servers of master internet, specifically ntp.master.cz and ntp2.master.cz.

- Inconsistent log format

  Each source of logs may utilize a different format to present data. The most popular formats are CSV (Comma Separated Values), tab-separated text files, SNMP (Simple Network Management Protocol), XML (Extensible Markup Language), syslog, YAML (YAML Ain't Markup Language) or binary format.

  The only solution is to use a log ingest pipeline capable of understanding and normalizing every format of input data.

Security Information Management (SIM)

Once logs are collected from endpoints, normalized, cleaned up and stored, the next step in security related terms comes into play. Security Information Management is a term used to describe long term storage, analysis and reporting based on relations between separate information. Information correlation, false positive elimination, and knowledge derivation is only possible with enough underlying data.

Security Event Management (SEM)

Once a security event occurs, it should be dealt with as soon as possible. As the security of any system can be represented by an onion model where each layer of security defense makes it harder for the attacker to succeed.

Security Event Management is concerned with real-time monitoring of possible events, analysis, reporting, alerting and swift event resolution support.

SIEM should merge the SIM and SEM systems into one cohesive system. The underlying hardware as well as software must be stable and fast enough to enable real time monitoring, alerting and reporting duties adopted from SEM, as well as cost effective and spacious enough to allow long term storage for future reference. (2)

Security event

An event is any observable occurrence in a system or network. Events include a user connecting to a file share, a server receiving a request for a web page, a user sending email, and a firewall blocking a connection attempt. Adverse events are events with a negative consequence, such as system crashes, packet floods, unauthorized use of system privileges, unauthorized access to sensitive data, and execution of malware that destroys data. (3)

Computer security incident

A computer security incident is a violation or imminent threat of violation of computer security policies, acceptable use policies, or standard security practices. Examples of incidents are:

- An attacker commands a botnet to send high volumes of connection requests to a web server, causing it to crash.
- Users are tricked into opening a “quarterly report” sent via email that is actually malware; running the tool has infected their computers and established connections with an external host.
- An attacker obtains sensitive data and threatens that the details will be released publicly if the organization does not pay a designated sum of money.
- A user provides or exposes sensitive information to others through peer-to-peer file sharing services. (3)

1.2 Business environment analysis

1.2.1 External influence analysis

SLEPT

SLEPT analysis examines the effects of external influences on the company, considering sociology, legislation, economics, politics and technology.

Social factors

Factors influencing the company from the social point of view are, above all, the growing demands for the constant availability of offered services, the pressure on the ecological operation of the datacenter and security. At the same time, more emphasis is placed on the gender balance of employees and on the integration of people with disabilities, especially in IT environment.

The availability of potential IT staff in the Czech Republic is inadequate, therefore it is important to work on retention of existing employees and their training.

Legislative factors

Legislative adjustments in the area of IT security are at the forefront of discussions due to large data leaks mainly from social networks within the European Union (and thus the Czech Republic).

The harmonization of European Union general data protection regulation (GDPR) introduced last year sets large fines in case of loss of control of user data.

In general, every Internet provider in the Czech Republic must follow a Cyber Decree, setting reporting obligations, deploy measures to eliminate the risk of a cyber incident inside the company, and toward other entities in cyberspace.

Another frequent problem is the Copyright Act, especially the problematic copyright infringement by customers in the hosting part of ISP business. The Digital Millennium Copyright Act (DMCA) warrants many copyright owners' requests to remove unauthorized distribution of their intellectual property.

From the upcoming regulations articles 11 and 17 of the Copyright Act reform should be well watched, as their impact on the liability for illegal content on Internet portals directly

affects all public cloud providers. Their enforcement and subsequent use in legal disputes may result in fines for platform providers instead of end users, which could be disastrous for cloud and internet providers.

Economic factors

The Czech economy is still in a moderate but slowing growth, with some companies preparing for a possible crisis. In the IT area, salary growth has slowed down, most companies use employee bonuses as flexible working hours, 5 or more weeks of vacation, working from home, employer-paid breakfasts, corporate events and other as a motivation.

Political factors

The political scene in the Czech Republic is stable, the current government supports the development of technology companies. At the same time, however, security is not included in most campaigns, therefore no direct financial contribution or tax breaks can be expected from the government.

Technological factors

The acquisition of more powerful and faster servers is supported by the European Funds for High-Speed Internet Development.

The increased competitiveness of AMD's processors results in a significant shift in available solutions, not only in terms of performance, but also in price.

The speed of moral obsolescence of server technologies has accelerated over the last 2 years from the previous approximately 10 years to today's 5 years, but the technology cycle is expected to prolong to more than 10 years due to the limits of available technology at material level.

Security enhancements are available at a good technological level.

Porter's five forces analysis

I analyze the external environment of the company from the perspective of five threat factors

Threat of existing competitors

In the Czech Republic and especially in Brno, it is possible to divide the competition in the area of interest of the company into several areas. From the perspective of the Internet provider, the main competitors are small providers of high-speed connections to large peering networks with the possibility of interconnection into transit, in Brno mainly the company Faster.cz

In terms of server housing and server hosting, these are again small providers such as Faster.cz or coolhousing.net, but there is also a threat from multinational public cloud operators, especially Amazon (AWS), Google (Google Cloud) and Microsoft (Azure).

As far as security and logistics issues are concerned, the big providers are in the role of impartial provider of virtual hardware as a service and the security solution is left to the user.

Faster.cz actively deployed a SIEM solution and offers it to its customers as a service, other providers could follow suit.

Threat of new entrants

The competition in the local market is challenging, it is necessary to place a suitable place to build a datacenter, as well as considerable resources and willing cooperation of other Internet providers, therefore there is no need to worry about new competitors from small providers. The greater risk is the expansion of one of the large cloud providers to the Czech Republic, where it could influence the prices of the services offered. One of the major providers that has already begun to expand into the Czech Republic is the Chinese Aliyun (Alibaba Cloud), which does not offer security solutions, and conversely, Chinese ownership questions the security of data in this cloud.

Threat of substitutes

I was unable to identify substitution competition between existing and announced technologies, customer requirements for server technologies are very conservative, and

backward compatibility, stability, and long-term reliability is very important. Theoretically, serverless application runtime technology can be considered as a substitution option, but long-term development is more likely to strengthen than server failures.

Threat of bargaining power of suppliers

The supply chain is a vital part of server hosting business, and the negotiating power of suppliers is generally quite large. On the other hand, by standardizing and modularizing server solutions, this threat is not too large. Overall, it is necessary to consider the development of the fight between supplying companies, both as a threat and as an opportunity.

Threat of bargaining power of customers

We can divide our customers into several categories.

Big Business

Large-scale IT facilities mostly rent space, connectivity and energy in the datacenter. Migrating several racks of technology to another data center is logistically challenging and inefficient for many companies. This group of customers can strive to reduce prices, but they are mostly satisfied, and most of them expand the number of technologies in the data center, due to ever increasing demand on ICT technologies.

Medium businesses

The combination of several physical and multiple virtual servers is a usual mid-size company server stack, their bargaining position is stronger than that of large corporations due to larger flexibility. The loss of this type of customer is very simple and the chances of being dissatisfied or worried about stability and security are high. This group also prefers prepared solutions and managed solutions as the cost of services is lower than hiring a dedicated specialist inhouse.

Small businesses and individuals

The group of small businesses and individuals with a small share of total traded volume represents potentially more profitable customers in the future. Sensitive to the cost of the service, they are easy to lose. Unfortunately, plenty of illegitimate customers rent cheap virtual servers for illegal activities such as targeted attacks, spamming, or sharing

copyright-protected content. They represent a potentially interesting customer or potential security and legal problem.

1.2.2 7S analysis of internal factors

Strategy

The company's long-term strategy is to provide services at a high level of quality, at the internet service provider side of business as well as in hosting services. The company's advantage over large providers such as Amazon or Google is local support, flexibility of solutions, know-how in server administration and local specifics and highly available solutions designed according to customer needs.

The company focuses not only on developing cooperation with existing customers, but also on expanding the portfolio of services on offer for new customers.

The company's long-term intentions are based on the high-quality education of its employees, which the company supports through regular training and participation in international conferences and lectures.

Structure

The organizational structure of the company is combined; in this work I focus on the part of the company where direct management with line manager is deployed, as can be seen in Figure 3 Organizational structure. The Chief Technical Officer is directly responsible for managing data center and administrators and the security of the offered services.

From the perspective of the whole company, the management is linearly staffed with project structure, overlapping into the divisional structure (it is advantageous for some projects to create working groups across departments with specialized staff responsible for part of the project)

Systems

An internally developed information system is the main support system available to all employees to varying degrees. Processes are formalized, employees have access to an internal Wikipedia with guidelines and recommended workflows. The current security event management system is outdated and inefficient.

Style of management

Direct management of the datacenter team of administrators and technicians combined with individual responsibility for quarterly projects on internally used services as well as customer facing services. The high level of autonomy of the administrators in solving individual projects and problems together with the management which listens to the recommendations of employees regarding future projects means that the company utilizes a combination of democratic and Laissez-Faire styles of management.

Staff

Employees are IT professionals with specializations in networking, virtualization, server management with both Windows and Linux, security, design of highly available solutions, automation and other capabilities. The company takes care of the professional development of its employees by regular training. Among the workers there a friendly atmosphere is supported by corporate events officially organized and initiated by employees. Collaboration on issues related to specialization among colleagues is often used.

Skills

At the same time, the high level of specialization is an advantage in solving highly specific problems, but it also poses a problem in taking leave / substitutability. The skills specifically related to the security project discussed in this work are not common, documentation and internal training is one of the reasons that management wants to implement this project. Management hopes to increase all workers ability to work with the SIEM system.

Shared values

Quality, safe and sustainable solutions are made possible through shared values among employees. They represent the level of quality of solutions below which employees do not stoop, despite customer requirements, as these requirements could jeopardize the quality of service offered and the safety of not only the customer but potentially all customers.

2 Current state of security and log management

2.1 Introduction of the company

In order to understand the requirements in the context of the target company I will briefly introduce it. Master Internet is a small ISP (Internet Service Provider) and datacenter operator at the same time.

Figure 2 Master Internet logo (4)

2.1.1 History of the company

"*Master Internet has been in the server hosting market for more than 15 years. After the first year of operation with servers located in the United States, they moved them to Brno, where Master Internet continue their business activities to this day. Then as now, they focused on server leasing, but with their own network they could also provide internet access. In 2003, Master Internet was the first provider on the Czech market to offer the virtual server product, and since then they have been focusing on virtualized solutions, and cloud solutions in particular.*" (5)

2.1.2 Organizational structure

The technical part of business is managed by two teams, Administrators and 24/7 technical support. Other parts of the company (sales, accountant, HR etc.) are not important for the project described in this thesis. The simplified organizational structure can be found in Figure 3.

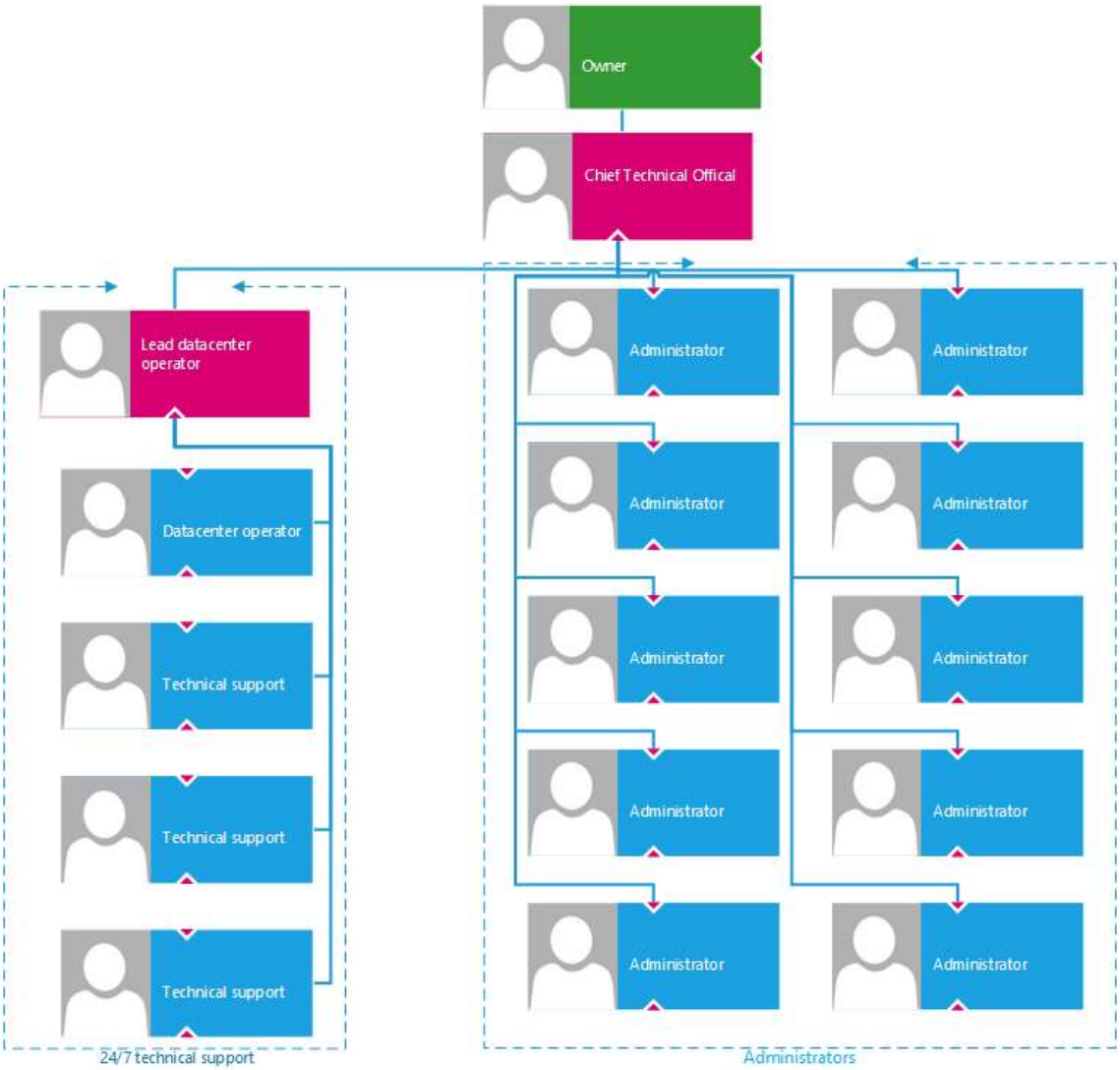

Figure 3 Organizational structure

(Source: own creation)

2.2 Log management currently

Currently a rudimentary subset of SIEM is deployed in the company only for a subset of internal devices.

Log management is handled by an old version of Logstash, an open source free to use log collector. It is running bare metal on a single instance. Collection is handled on UDP port 514 (traditional unix syslog port). Incoming requests are allowed for the whole internal network and select IP addresses from public IP ranges. Allowed source IP addresses are defined in the firewall configuration.

Logstash filters the logs, adding tags according to the type of log sent, setting up categorization and reverse resolving sources. To make resolving faster a small local lightweight DNS cache server in the form of dnsmasq is utilized.

Output of Logstash is then directed both to Elasticsearch for storage and ad hoc requests and Riemann for pattern recognition and email reporting.

Kibana acts as the GUI for browsing the logs and ad hoc search request creation.

Unfortunately, older, no longer supported versions of service are utilized, thus lowering the effectiveness, usability, utilization and security of the whole solution.

2.2.1 Log sources

To fully utilize SIEM systems it is necessary to gather information from all sources that are crucial in problem resolution and all sources where a potentially dangerous event can occur.

Current log sources are mostly from managed networking equipment

- Routers with their firewalls, SNMP, SSH, access control
- Central firewalls with main DHCP server, SNMP, SSH, access control
- Managed L2 switches, SNMP and SSH
- Managed L3 switches, SNMP and SSH
- TACACS server
- Wifi access points
- Hypervisor hosts

However, as I discovered during the implementation phase of the project, not all sources were correctly managed. Only 67% of all sources contained all the necessary information in the right parts of log management. 22% of all sources were managed by the system, but due to configuration errors weren't parsed correctly, meaning that orientation in logs coming from these sources was more challenging for the administrators. 11% of all sources were not present in the old log management at all.

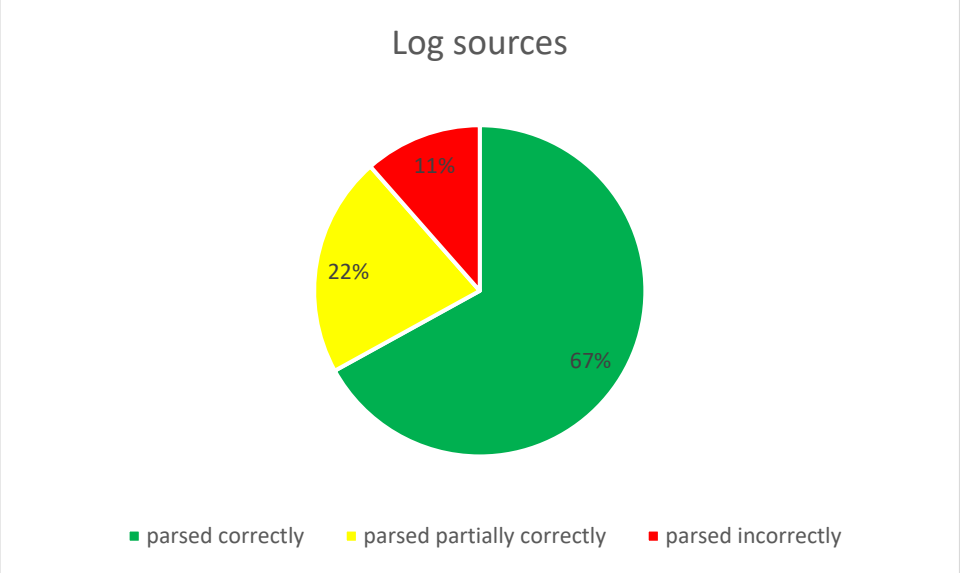

Figure 4 Log sources originally

(Source: own creation)

Potential internal log sources include (but are not limited to)

- Active Directory
- Services servers, possible sources of information include fail2ban, iptables, apache
- Nagios
- Wiki
- Ticketing system
- IS
- Certificate management server
- Flowmon netflow
- Radware Vision
- dashboard
- Biometric access authenticator

- Perimeter alarm
- Development servers of application team
- Managed customer servers

To make collected data more relevant it might be beneficial to include third party sources of information (from reputable sources only and with mostly informational value)

Possible external sources

- Botnet IP lists
- Spam blacklists
- IP location services
- National cybersecurity alerts

The process of using information from logs to identify false positives or true threats is reliant on the experience of administrators and the combination of information from the log management, service monitoring system and information received from third parties such as national cybersecurity team, customers, phishing monitoring services etc. To shorten the decision-making process several log enriching processes are to be configured such as Ip address to DNS name resolution,

2.2.2 Assisted evaluation

To analyze the current state, I utilized a technique called assisted evaluation, developed by the Czech national cyber security team NÚKIB in cooperation with doc. Petr Sedlák. It consists of a questionnaire in the form of a matrix of open questions which the evaluating company can use as a starting point. The questionnaire is targeted mainly at CII (Critical Information Infrastructure) providers, although it also contains relevant recommendations for all security conscious enterprises.

It is recommended to select all parts of assisted evaluation that apply to your individual company, as the methodology and the questionnaire are meant for a broad range of (mostly government) subjects. I therefore created a modified questionnaire with a subset of points to cover.

Table 1 Assisted evaluation

(Source: own creation based on (6))

Description of requirements	How fulfilled	Codified/ described in
Network and information system access control. Physical and logical access to networks and information systems is authorized and limited according to security requirements.	Security locks and surveillance on all access points to the premises as well as throughout the datacenter. Logical access control with centralized identity management system, separation of networks into dedicated zones.	Internal directive in the form of a set of wiki pages
Physical and environmental security, covering failure of the system, failure of the administrator, willful interventions and environment induced failures	Fire extinguishers, redundant backups, configuration check mechanisms as well as monitoring, logging, updates and upgrades of software and hardware	Code of conduct, backup plan, logging directive

Maintaining and testing of procedures and processes of anomaly detection in order to maintain early and adequate information	Regularly run anomaly detection program, cooperation with national CSIRT for early external threat notices	Internal directive for anomaly detection
procedures and policies concerning incident reporting and weakness identification and reporting	Code of conduct, procedures in wiki	internal directives for reporting, wiki
Reaction in accordance with defined procedures and reporting on results of deployed measures	Procedures as described in internal wiki, reporting in the ticketing system each issue has its own ticket	Internal wiki, ticketing system
Incident severity appraisal, incident analysis documentation, collection of evidence and continuous improvement process support	Evaluation scale, wiki	Evaluation scale, wiki, log management system
Crisis plans proposal and application, based on the analysis of impact on business continuity, regularly reassessed and tested.	Crisis plan	Crisis plan
Ability to restore operation after an exceptional event, regularly assessed and tested	Backup plan	Backup plan
Carry out series of measurements or observations in order to determine, if	Monitoring system with defined thresholds, regular maintenance	Monitoring system, code of conduct

networks and information systems function as intended		

3 Proposals and contribution

3.1 SIEM comparison and selection

In this chapter I create an evaluation table with regards to the requirements of the company on the SIEM system, compare the available solutions and choose the most suitable SIEM system to implement.

3.1.1 Requirements compliance evaluation table

The basis for any cyber security defense enhancements should be properly implemented and adhered to ISMS (Information Security Management System) in the company.

To compare the available solutions, I have prepared a scoring Table 2 that reflects the requirements on the SIEM system by a small ISP. The resulting score represents a weighted average according to the importance of each scoring component. Scoring represents how well the requirement is fulfilled by the evaluated system on the scale of 0 to 10, where 0 means cannot fulfill and 10 fully satisfies requirements.

Table 2 Evaluation table

(Source: own creation)

Evaluation requirement	Scoring	Additional information	Weight
Ease of use			3
Compatibility with 3rd party services			2
Scalability			3
Storage tiering capability			1
Security			3
Documentation, community			2
Configuration backup			1
Backup and restore of data			1
Support			2
Performance			1
Total sum score		Total weighted average	

3.1.2 Commercial SIEMs

Commercial SIEMs are provided by either one of large organizations that are developing SIEM tools for internal purposes and license them as well to offset part of the development costs or a company specializing on data management applications in general or SIEMs in particular.

Commercial SIEM systems are expensive, the pricing of these systems can be divided into several groups.

1. One-time upfront payment

   Usually either a basic version of the system with either limitations in performance or scalability, lacks features for centralized user management, can be more cost effective depending on companies current and future needs, not scalable. Future upgrades and support beyond warranty represent additional costs.

2. Time based subscription model

   No limitation on the amount of ingested data, priced per month with usually a 1 to 2-year commitment. Can be advantageous in larger deployments, for small to medium sized companies usually represents a large investment during deployment. Upgrades and support included (depending on the price tier).

3. Traffic based subscription model

   Pricing per ingested gigabyte per month, shipping side optimization recommended to lower the expenditure. Increasing the number of monitored nodes increases the costs, depends heavily on proper projection of future event volume and rate of occurrence.

4. Node based subscription model

   Pricing per monitored node, better than traffic-based model in deployments where a smaller number of sources produce large amounts of data. Difficult to scale in the future.

3.1.3 OSSIEM

Open Source SIEM is a type of SIEM with open source code available.

Many commercial SIEM developers are maintaining a stripped-down version of their proprietary tools to boost development of modules available for their SIEMs, as a free to use version to entice customers to try their solution and when they require more functionality out of the system to migrate them onto their premium platform. The benefits of open source projects are that multiple interested companies can support the development, security and feature enhancements than a single company could. Auditability of the underlying code for potential backdoor access (intentional or malicious) is also possible when source code is available, enabling researchers to analyze, find and report for correction anything potentially dangerous.

The open nature of OSSIEMs also usually results in an easier to extend and modify interfaces available, enabling both third party module support and standard APIs for in-house connections. (Elasticsearch for example has an API for all queries that can be made from the Kibana interface, enabling custom connections to other information systems, connecting status board, internal IS, customer IS to show relevant information).

The open nature of OSSIEMs means that community support and tips are widely available, enabling easier creation of highly specialized use cases. The downside of open nature is that initial configuration tends to be more complicated and the learning curve of the system tends to be steeper.

3.1.4 Available commercial SIEMs

Obtaining approximate prices of commercial SIEM solutions is not easy, very few companies offer an easy-to-compare upfront cost. Contacting the sales departments and going through company websites yielded following prices (all prices were converted to USD for better comparison, solutions with different pricing model than single payment have the pricing model closest to the requirements of small ISP selected)

Table 3 SIEM pricing

(Source: own creation)

AlienVault Unified Security Management	$12900
Splunk Enterprise Security per GB	$2076
LogRhythm NextGen SIEM	$28000
SolarWinds SIEM	$4272
Fortinet FortiSIEM perpetual license	$21179
Fortinet FortiSIEM 50 node license	$8700
SIEMonster	$15000
The SMB Edition 150 node license	$5000
IBM QRadar	$20000
X-pack per control node	$7206

Commercial SIEMs are fairly expensive as can be seen in the Table 3 SIEM pricing. As an alternative, open source solutions might be a more cost-effective solution. In ISP environment it is not unreasonable to expect a capable team of administrators able to provision a self-hosted solution. The hardware costs, operating costs can be easily covered, the required manpower is also already available. As the prices of commercial SIEM systems are too high for Master Internet to accept, I focused on the open-source options, the two most suitable options being ELK stack and Graylog.

3.1.5 Comparison of available free solutions

3.1.6 ELK stack

ELK stack is a combination of opensource projects Elasticsearch, Logstash and Kibana in a single service. As the ELK stack has been evolving to support also a lightweight log shipper called Beats it is sometimes referred to as Elastic stack rather than ELK stack, but as in this thesis I am not using Beats component I will use the ELK stack name. The currently utilized structure of ELK stack can be found in Figure 5 ELK stack visualization.

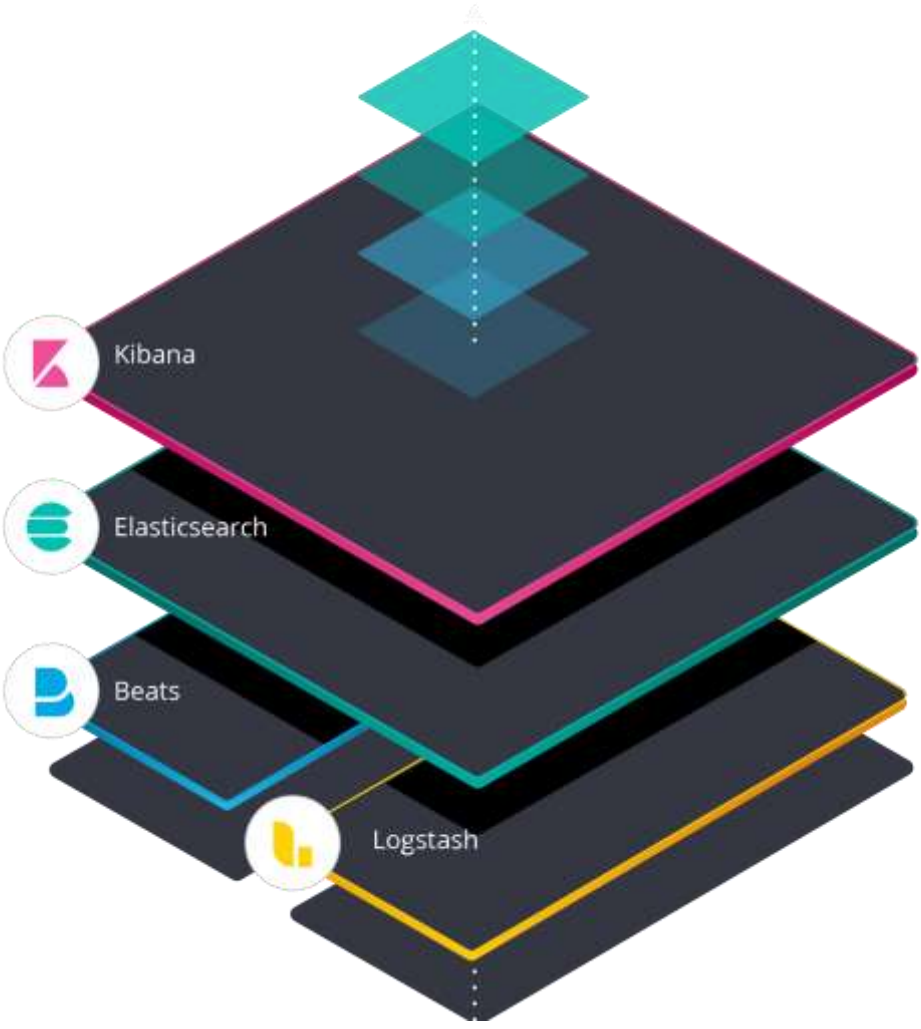

Figure 5 ELK stack visualization (7)

Elasticsearch is a NoSQL search and analytics database designed to handle large volumes of data. To store and search for values it utilizes several hierarchically ordered levels of abstraction.

Node structure

Elasticsearch structure is split into nodes with different functions. To coordinate the cluster a master node is voted from within the Elasticsearch cluster master-eligible nodes. Once a master node has been selected it coordinates the initialization of data nodes and ingest nodes. A data node is responsible for storage of data as well as search and indexing capabilities. An index is a collection of documents that have somewhat similar characteristics. An index is identified by a name (that must be all lowercase) and this name is used to refer to the index when performing indexing, search, update, and delete operations against the documents in it. A document is a basic unit of information that can be indexed. Documents in Elasticsearch are expressed in JSON (JavaScript Object Notation), a ubiquitous internet data interchange format. An index can contain large number of documents, limited only by available hardware. To avoid hardware limitations of a single server Elasticsearch supports a distributed storage of indices by subdividing a single index into shards. A shard is fully self-contained index, which can be hosted on any storage node on the cluster. To avoid loss of data or unavailability of shard during partial cluster downtime Elasticsearch contains replication. It is possible to create replicas of primary shards which are exact copies of original shard on one or more Elasticsearch nodes. It is possible to allocate as many replicas as necessary, in smaller environments with replicated storage a single shard deployment is possible. As your needs scale it is also possible to add replicas after the fact, increasing the security of the data as well as increasing the possible search performance as the search operations can be parallelized.

Logstash is a server-side real-time data processing pipeline designed to ingest data from multiple sources, transform the data and send it to a database. Logstash is not limited to outputting into Elasticsearch, current output plugins officially supported by Elastic can be found at https://www.elastic.co/guide/en/logstash/current/output-plugins.html The compatible outputs include, but are not limited to graphite, kafka, loggly, mongodb,

rabbitmq, redis and others. Logstash features a customizable pipeline to unify data from disparate sources as well as cleanse unnecessary information from the stream. Logstash pipeline configuration is divided into three main stages: Input, Filter and Output

The input part of pipeline configures the way data will be ingested into Logstash. The most common options include

file: reads data from a file on the filesystem, similar to UNIX tail -0F

syslog: listens on well-known port 514 for syslog formatted messages and parses them according to the RFC3164 format

beats: processes events sent by the Elastic Beats log shipping agent

The filter part is a configurable intermediary processing device in the pipeline. Rules based application of filters as well as combination of filters is supported. Most useful filters include

grok: parse and structure arbitrary text. Grok is currently the best way in Logstash to parse unstructured log data into something structured and queryable. 120 patterns are available built-in to Logstash with custom pattern creation possible.

mutate: perform general transformations on event fields. You can rename, remove, replace, and modify fields in your events.

drop: drop an event completely, for example, debug events.

clone: make a copy of an event, possibly adding or removing fields.

geoip: add information about geographical location of IP addresses, the basis for geolocation maps in Kibana.

The output part represents the final phase of Logstash processing. An event can pass through multiple outputs, after processing through the last of them an event has finished execution. Commonly used outputs include

elasticsearch: send event data to Elasticsearch.

file: write event data to a file on disk.

graphite: send event data to graphite, a popular open source tool for storing and graphing metrics.

riemann: send metrics to Riemann service

Kibana is the easy to use visualization and control frontend. It allows creation of visualization charts and graphs as well as general queries.

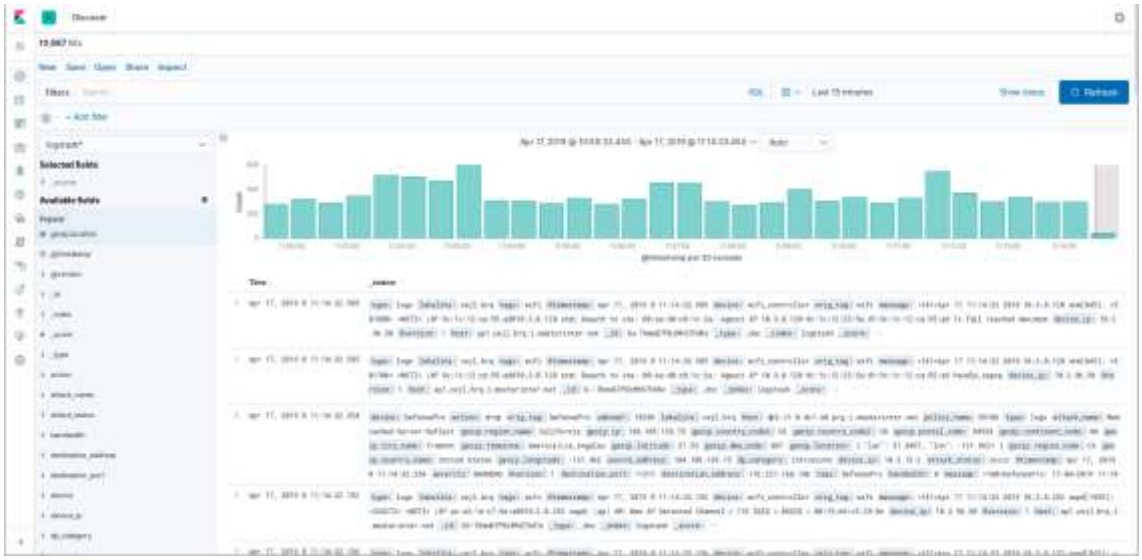

Figure 6 Main window with menu

(Source: own creation)

Every ingested log contains a timestamp upon which a time series is created, thus allowing for time-based searches with frequently used options available as a direct shortcut. It is possible to select either a sliding time window (relative time to current) or an absolute time range.

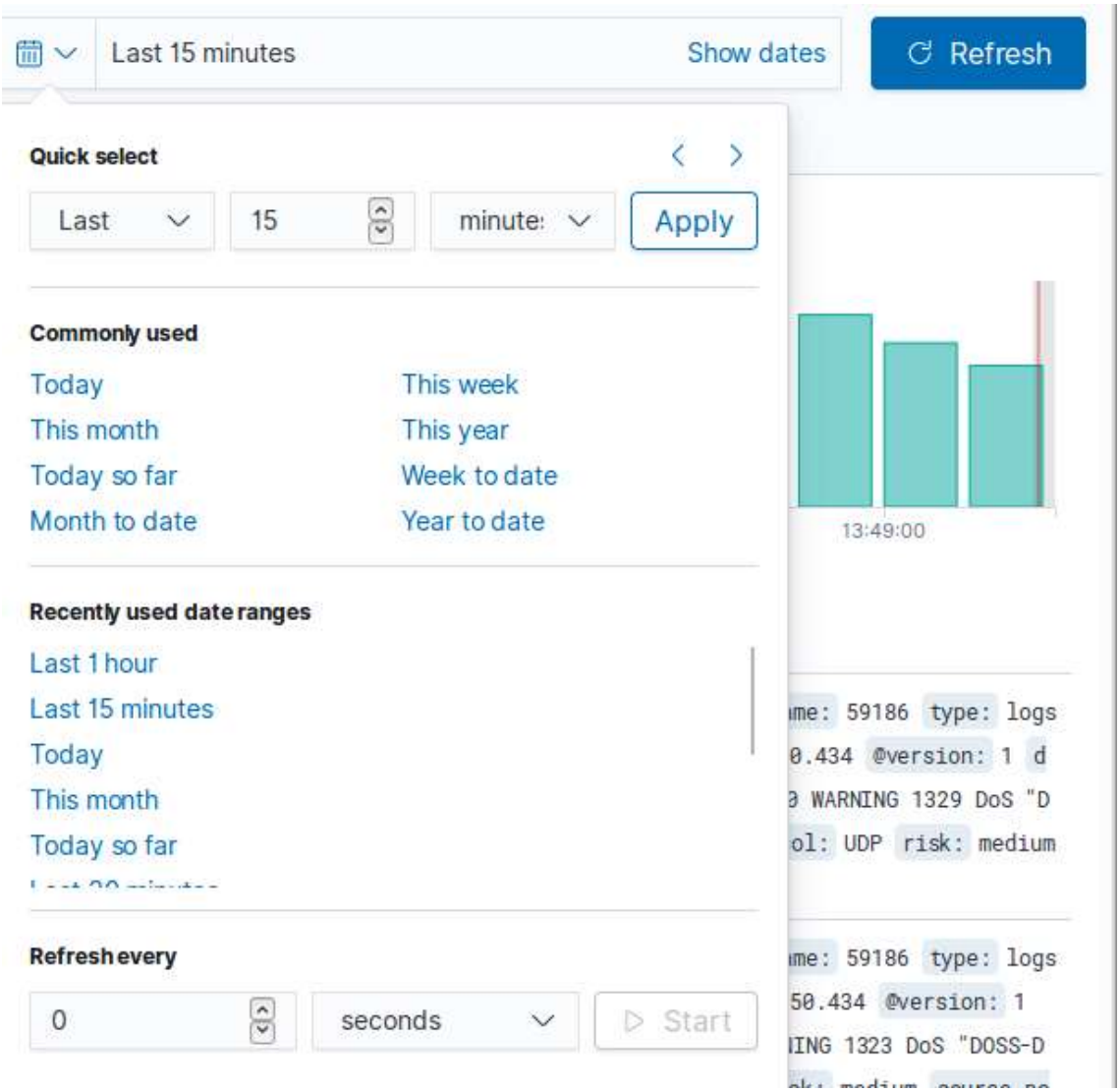

Figure 7 Time selection shortcuts

(Source: own creation)

Kibana also has an interface for manual Elasticsearch GET, PUT and POST queries with syntax checks built in. It provides an easy way of developing custom queries not available through the Kibana interface.

Figure 8 Console for queries

(Source: own creation)

Kibana built in monitoring of the whole cluster stats. Monitoring is divided into the Elasticsearch monitoring and Kibana monitoring.

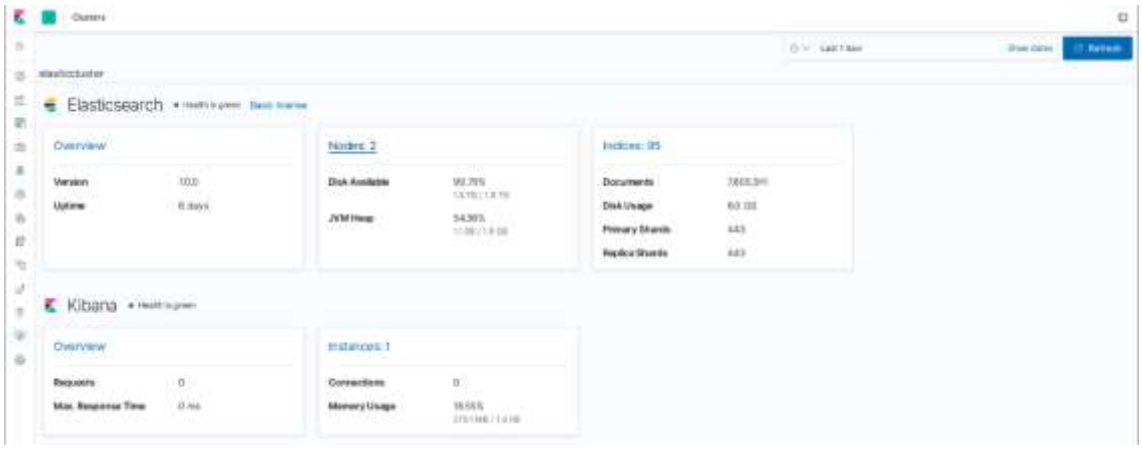

Figure 9 Monitoring interface

(Source: own creation)

Elasticsearch monitoring allows administrators to monitor the general state of Elasticsearch cluster, the amount of memory used by Elasticsearch, number of shards, documents and data stored, as well as charts regarding the search rate and latency and indexing rate and latency.

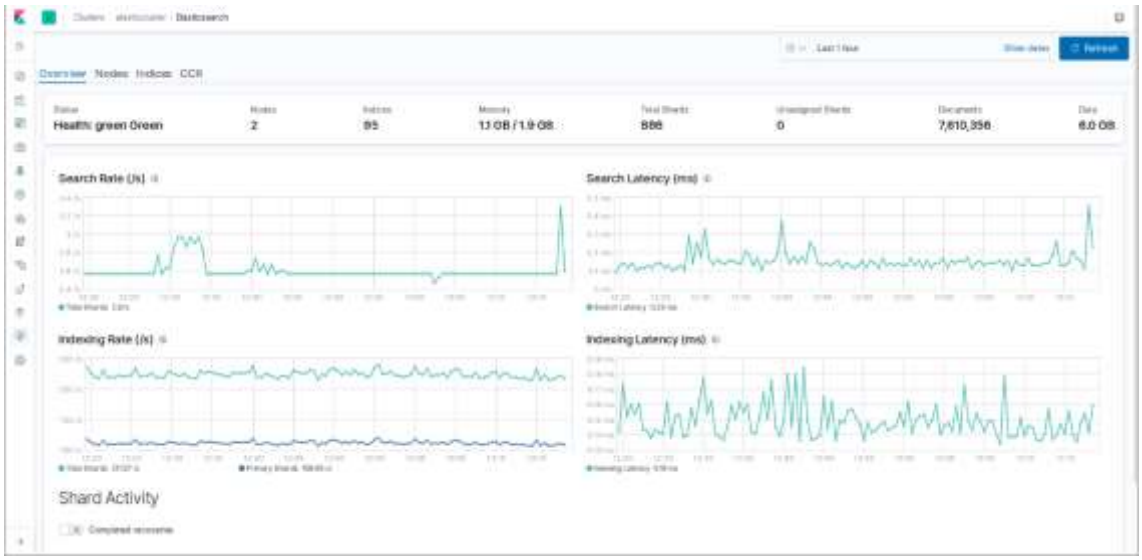

Figure 10 Elasticsearch monitoring graphs

(Source: own creation)

Individual node statistics are also available with detailed CPU utilization, memory, system load or latency graph of current utilization.

Figure 11 Node load graphs

(Source: own creation)

Kibana monitoring contains Kibana load graphs as well as information on the number of requests, memory usage and number of http connections.

Figure 12 Kibana monitoring

(Source: own creation)

Logstash monitoring is not available through Kibana, although a simple heartbeat input plugin configured to write to Elasticsearch every 10 seconds alleviates the need for monitoring.

Lifecycle management

To automate the lifecycle of data in ELK there are two ways to manage the transition of data from current to long term storage to deletion. Starting from ELK version 7 the built-in index lifecycle management settings allow creating lifecycle policies from Kibana or traditionally through the API. The policies determine the maximum size and/or the maximum number of documents and/or the maximum number of days an index can contain before a rollover to a new index is triggered. Rollover can also trigger the optional next phase, the warm phase. It is possible to specify a node denomination that determines the hierarchy of data.

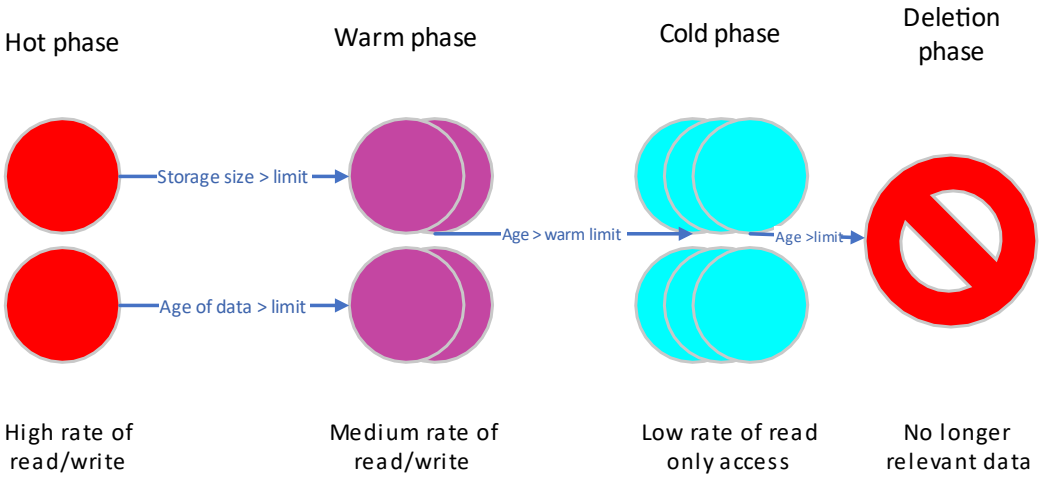

Figure 13 Hot warm cold architecture

(Source: own creation)

Elasticsearch lifecycle management allows 4 phases of data availability hot, warm, cold and deletion. As the ILM (Index Lifecycle Management) has been integrated only recently and is not compatible with older-style daily Logstash rotation of indices I didn't choose this method.

The older but still functional and more lightweight solution to lifecycle management is through the Curator program. Originally developed by a third-party developer Elastic saw the usefulness of the tool and hired the original developer Jordan Sissel and continued to support the tool. Curator offers an easy to use CLI interface for automation of shard allocation based on time. Unlike ILM it isn't necessary to modify the index templates (that allows daily rotation of Logstash index as in past versions) and works on even older versions of Elasticsearch reliably.

Elasticsearch evaluation table

Table 4 Elasticsearch evaluation table

(Source: own creation)

Evaluation requirement	Scoring	Additional information	Weight
Ease of use	8	Intuitive GUI, Logstash configuration not in GUI	3
Compatibility with 3rd party services	9	High number of built in and 3rd party plugins	2
Scalability	8	Elasticsearch engine scales well	3
Storage tiering capability	7	Tiered storage supported, difficult setup	1
Security	5	Lacking OOB user management	3
Documentation, community	10	Excellent documentation and community support	2
Configuration backup	7	Easy configuration backup	1
Backup and restore of data	6	Snapshot support, bad cross-version import support	1
Support	8	Frequent and well-maintained updates	2
Performance	6	Higher recommended specifications	1
Total sum score	74	Total weighted average	7,5

Evaluation summary

The ease of use of ELK stack is given by a modern, responsive design of the interface with good deal of nice to have features such as query completion suggestions, most used filters, quick toggles with understandable icons as well as helpful labels. The initial configuration is mostly file-based, lowering the score a little, with no configuration of Logstash directly from Kibana adding to a total score of 8.

Compatibility and availability of third party addons is a strong point of ELK, with most major services compatible either OOB or via an official or community plugin.

Elasticsearch is highly horizontally scalable, although cluster management is initially slightly confusing, thus only a score of 8 points.

Tiering of the storage for fast access to recent data is possible, but more difficult than need be.

Security of the ELK stack is dependent on proper configuration; initial deployment is geared toward local testing environment and needs security hardening before deployment into production environment. User management OOB is provided on the paid tier, a recommended setup via a secured apache or nginx proxy alleviates this problem.

Elastic documentation is excellent with in-depth configuration options explained with examples. Community around Elasticsearch is also very good both online and offline, Elastic officially supports an Elastic User Group meetup in Brno every few weeks. (8)

Backup and recovery of ELK configuration is conveniently file-based, making recovery to a new host in the event of catastrophic failure easy.

Backup and restore of data inside ELK stack include support for snapshots, recovery of indices created in older versions of Elasticsearch is not supported directly (It is possible to recover indices created in 1 last major version so Elasticsearch 7 can import indices from version 6 directly, but indices from version 5 have to be reindexed in a version 6 cluster first)

Updates to ELK stack are timely, with mostly meaningful additions. Version 7 broke a small number of legacy features.

Performance is generally good, although highly memory intensive. Logstash indexing and transformation is CPU performance dependent, lowering the score of this section.

3.1.7 Graylog

Figure 14 Graylog logo (9)

Graylog is an alternative to ELK stack, also with open source version available. It also utilizes Elasticsearch as its database, as it is considered the most mature, fast and scalable NoSQL database, most suitable for log storage applications. However, unlike ELK stack it utilizes MongoDB for metadata storage and its own Graylog service for both data ingest and as a server for web client control.

The main interface is divided into the top navigation menu and the search window. The interface shows the histogram of results, individual messages that satisfy the search conditions, query input box and result filtering interface. The controls are very similar to older version of Kibana, missing some of the more interactive features of ELK 7 such as click and drag selection of timeframe.

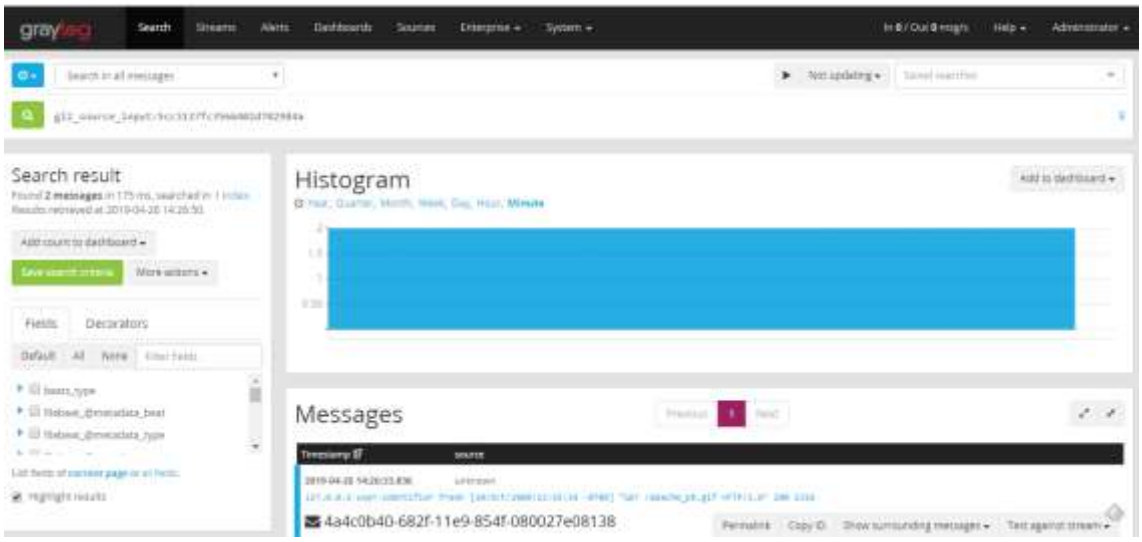

Figure 15 Graylog search interface

(Source: own creation)

The time selection interface of Graylog is barebones compared to ELK, lacking frequently used timeframes. It also lacks finer control.

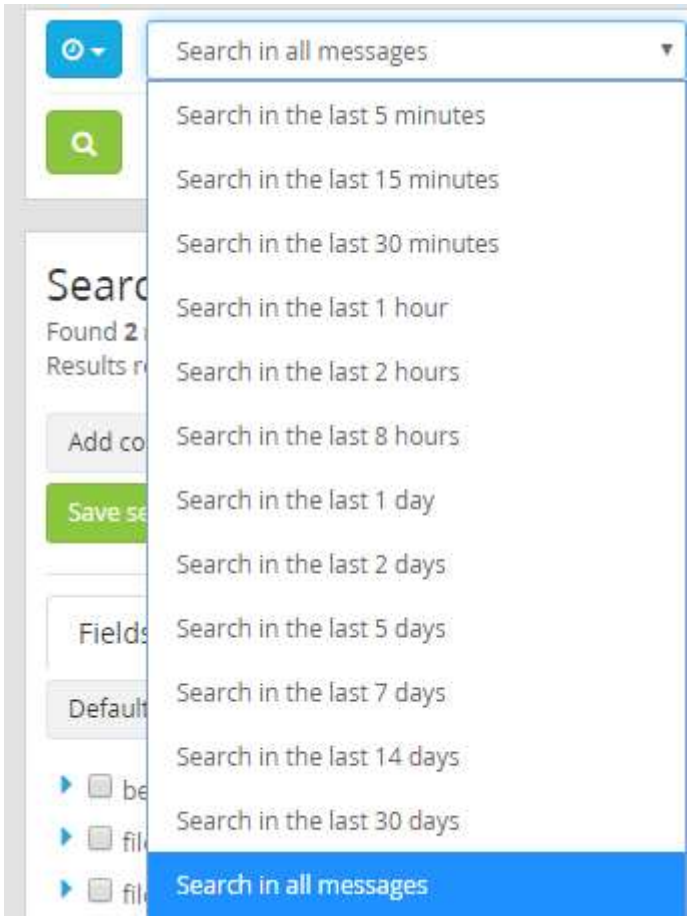

Figure 16 Graylog time selection

(Source: own creation)

Graylog offers built in alerts right from the web interface in the free version. Alert system is based on monitoring streams of data for a passing condition. Alert condition types include:

- Message count

  This condition triggers whenever the stream contains more than the specified number of messages. Suitable for environments where a small number of errors is acceptable and alert after exceeding the number.

- Field aggregation

  Computes statistical aggregate of numerical messages in the stream and triggers whenever the aggregate is lower or higher than the specified amount. Suitable for

monitoring performance problems, such as when the standard deviation of response time is higher than X in the last Y minutes.

- Field content

  Monitors the stream for a message with a filed set to a given value. Suitable for type-based alert filtering of messages.

Different notification types can be applied to alerts, allowing Graylog to send notifications to external systems. Free version contains the email notification and the HTTP notification. HTTP notification allows the configuration of an endpoint in other services such as ticketing system, Slack, Mattermost, IS or other to accept, parse and further modify the incoming message. Standard JSON message syntax is used.

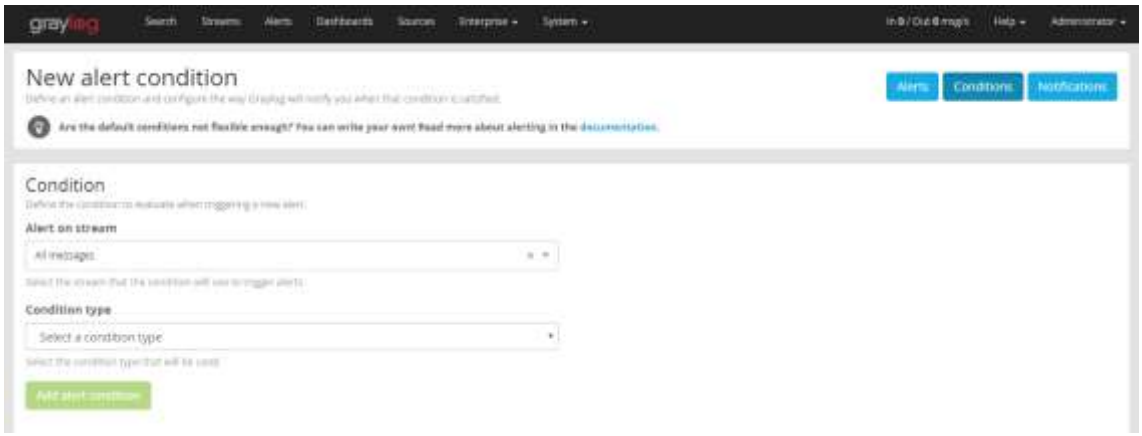

Figure 17 Graylog alert creation

(Source: own creation)

Graylog does not utilize Logstash for filtering, normalization and transformation of incoming data. Streams in combination with processing pipeline are used instead, offering the control over the ingestion pipeline through the Graylog web interface. Incoming messages are compared against a set of rules stored in the MongoDB database, ID of each matched stream is added to an array within the message, which is then sent to an output router, which determines where will the message be sent to. This is similar to the Logstash output plugins.

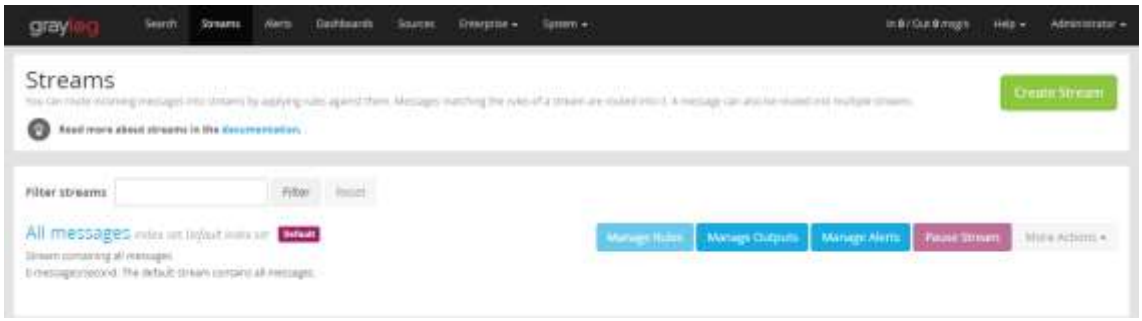

Figure 18 Graylog Streams

(Source: own creation)

The processing pipeline operates similarly to the filter plugin of Logstash. Pipelines must be connected to streams to process messages. Rules enable conditional enrichment of messages. Example rule:

```
rule "from firewall subnet"
when
    cidr_match("10.10.10.0/24", to_ip($message.gl2_remote_ip))
then
end
```

(10)

Simple node monitoring with information on the load and details about the active nodes in the cluster, similar to ELK.

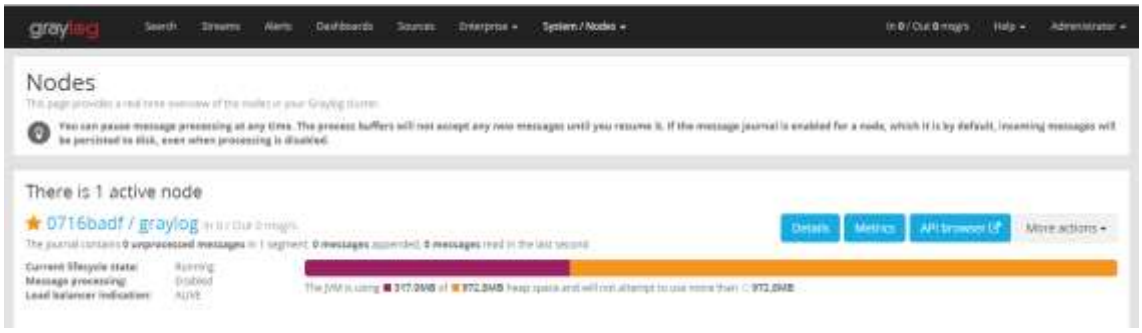

Figure 19 Graylog node statistics

(Source: own creation)

Graylog lifecycle management offers both the Elasticsearch lifecycle management methods described in chapter 3.1.6 on page 42, as well as custom Graylog rotation of indices. Unfortunately, using both can sometimes lead to unexpected results if both Graylog and Elasticsearch start a rollover at the same time.

Graylog evaluation table

Table 5 Graylog evaluation table

(Source: own creation)

Evaluation requirement	Scoring	Additional information	Weight
Ease of use	9	Fully controllable through GUI	3
Compatibility with 3rd party services	6	Lacking some 3rd party services support	2
Scalability	8	Elasticsearch engine scales well	3
Storage tiering capability	7	Same underlying storage engine	1
Security	6	Better user management OOB	3
Documentation, community	6	Smaller userbase, less mature documentation	2
Configuration backup	6	MongoDB is more difficult than files to recover	1
Backup and restore of data	6	Snapshot support, bad cross-version import support	1
Support	6	Longer update schedule	2
Performance	8	Recommended specifications lower	1
Total sum score	68	Total weighted average	6,9

Evaluation summary

Graylog is well designed to be fully controllable through the web GUI interface. Initial setup is easy, the interface is well designed, missing only minor quality of life features.

Third party compatibility is not a strong point of Graylog, although major systems are supported, 3rd party plugin development is less developer-friendly with a smaller community.

Scalability of storage is the same as in ELK due to the use of Elasticsearch storage engine.

Tiering of the storage for fast access to recent data is possible, same limitations as in ELK apply for Graylog as well.

Initial setup is fairly secure, initial setup password-protects the system OOB, user management is available.

Documentation is well designed, with all important information accessible. Configuration examples are not comprehensive enough and community around Graylog is much smaller with less help available.

Configuration is stored in MongoDB, making configuration backup and restore more difficult.

Backup and restore of data feature the same limitations and features as ELK due to the use of Elasticsearch engine with added need for compatibility with Graylog.

Graylog operates a longer update schedule than Elastic, cross-version compatibility is generally good.

Lower recommended specifications as well as the streaming design of Graylog transformation result in a good performance.

3.1.8 SIEM selection

As the requirements and available technologies of a small ISP are specific, the selection of a SIEM solution will vastly differ from larger companies. Thankfully, apart from costly enterprise solutions open-sourced technologies offer similar features for much lower, even free upfront cost. The requirement of free opensource SIEMs to be self-hosted is an advantage in ISP environment, as keeping potentially sensitive log data inside the company is one of the requirements.

After careful consideration of the available solutions and the requirements from the CTO and administrators I have selected the ELK stack as the most suitable solution for company needs. The weighted average of evaluated requirements has resulted in favor of ELK stack. Also, the interface and usability during evaluation resulted in a better overall experience than competition.

3.2 Lewin’s change management model

3.2.1 Implementation of change

Unfreezing phase

In the first phase, we will analyze which company assets need to be secured using the SIEM system, i.e. create an up-to-date list of information sources for the SIEM system.

Based on this list, it is then necessary to prepare the SIEM system itself, divide all the assets on the list according to the type of information provided and create corresponding rules for each type in the SIEM system. The rules will be based on previous rules with necessary conversion of tags, adding missing assets and a better logical categorization.

The phase of change

Switching the endpoint systems from the original logging system to the new SIEM system will be done in the change phase. During the transfer, any compatibility issues not detected during the defrost phase need to be resolved and the functionality of the new system verified. The functionality of all components, including notifications must be verified.

The change also includes the preparation of documentation and training of employees, but the completion of the documentation will occur at the next stage.

Freezing phase

After proper verification of the functionality, the SIEM system will be put into routine operation and the old unsuitable system will be terminated and cleaned. Based on the experience gained during the change phase, the documentation will be completed and training of the administrators in proper use of the newly deployed SIEM system will be scheduled and performed.

3.2.2 Change Agent

Personally, as the responsible administrator, I am the change agent and I am preparing and executing the entire process of changing the SIEM system.

3.2.3 Sponsor of change

The sponsor of change is the Technical Director of Master Internet. The sponsor has dedicated part of the working hours to the implementation of the project and allocated the required hardware for use in this project. At the same time, funds are earmarked for the operation of the SIEM server system and training to increase knowledge of working with it.

3.2.4 Intervention areas

The change indirectly affects the whole company - the level of security is defined by the weakest link in the defense chain. Direct changes can be identified especially in technology - increasing the company's technological equipment, at the same time the change is a potential service for the company's customers (installation and management of the SIEM system is not trivial, but it may be desirable for a larger number of managed servers of costumers). The change to communication flows (simplifies traceability of key information) and processes (extending the scope of operation of all security administrators) have a major impact.

3.2.5 Project timeline

The time schedule was prepared using the PERT method, this project is unique and therefore it is not possible to estimate the time values based on previous experience, so I assigned 3 durations to the individual activities, optimistic, realistic and pessimistic estimate. To calculate a single resulting time, the data must be converted using the weighted average according to the formula $t = \frac{a+4m+b}{6}$

Table 6 PERT processes

(Source: own creation)

ID	Activity	Follower	a	m	b	t	σ	σ^2
1	Create a list of all monitored devices	2,3,4	5	6	7	6	0,333	0,111
2	Sort devices to categories	7	5	6	7	6	0,333	0,111
3	Prepare rules for each category	7	3	5	7	5	0,667	0,444
4	Prepare the server (physically)	5	0,1	0,2	0,3	0,2	0,033	0,001
5	Prepare the server (OS, basic SW)	6	0,5	1	1,5	1	0,167	0,028
6	Prepare SIEM services	7	0,5	1	1,5	1	0,167	0,028
7	Application and validation of prepared rules	8,10	2	3	4	3	0,333	0,111
8	Test traffic	9	0,3	1	3	1,22	0,450	0,203
9	Function validation	11	4	5	9	5,5	0,833	0,694
10	Worker education	13	2	3	5	3,17	0,500	0,250
11	Transfer all traffic	12	3	5	9	5,33	1,000	1,000
12	Validate traffic	13	7	9	11	9	0,667	0,444
13	End of project		0,1	0,1	0,1	0,1	0,000	0,000

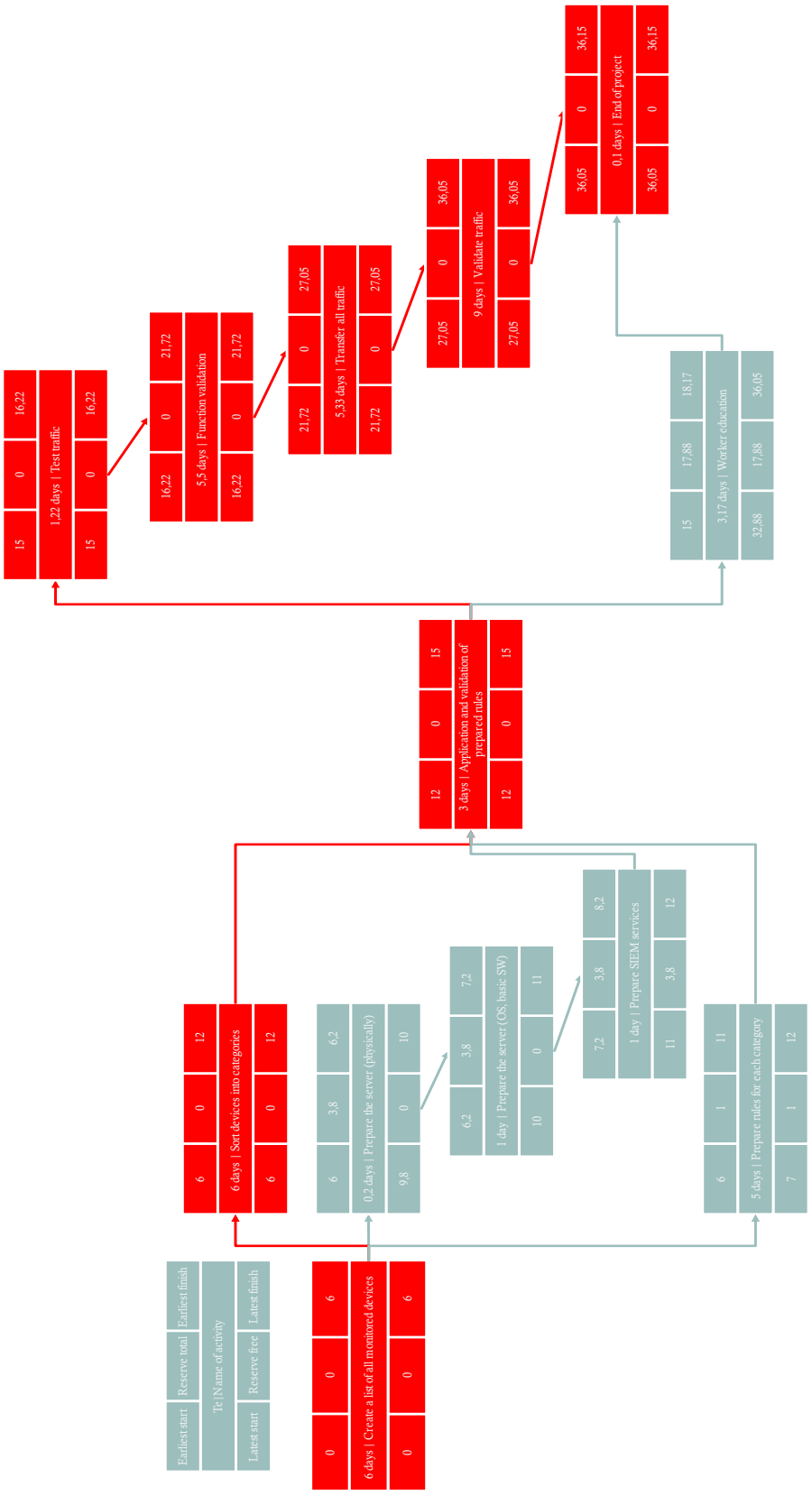

Figure 20 PERT *(Source: own creation)*

PERT commentary

The processes in the PERT graph are serialized, low level of parallelization is given by the dependencies in the flow of consecutive tasks. A critical path is highlighted in red. In order to achieve the planned length of the project it is necessary to pay attention to the processes on the critical path, as delays in these processes would prolong the whole project. The change duration will be 36,15 days, 40 accounting for a reasonable project buffer.

3.3 Risk analysis

Risk analysis is a subset of requirements analysis in which I will assess the risks involved in proposed change. I shall use both quantitative as well as qualitative analysis, determining the potential risks, their probability, their impact on the company business and the vulnerability of involved assets.

Probability and impact assessment table

Table 7 Risk analysis scoring

(Source: own creation)

Values	Probability	Impact
0-2	Very low	Minimal
3-5	Low	Low
5-7	Medium	Medium
8-10	High	High

3.3.1 Risk assessment

The table contains the identified risks that may arise during deployment of the SIEM system (either in the planning phase, the change phase, or in the post-project use). Each threat is assigned a corresponding scenario. The threat value is evaluated by assessing the

likelihood and impact of the threat. High value threats are highlighted in deeper shades of red. The highest rated threats should be treated appropriately, see below.

Table 8 Risk assessment table

(Source: own creation)

ID of threat	Threat description	Probability	Impact	Threat value
1	Delay in one of the processes on the critical path	9	2	18
2	Unsuitable chosen system	2	8	16
3	System incompatibility	2	7	14
4	Incorrect configuration	4	7	28
5	Omission of actives	5	3	15
6	Loss incurred due to system failure	2	5	10
7	Loss of data	3	7	21

3.3.2 Risk map of the project

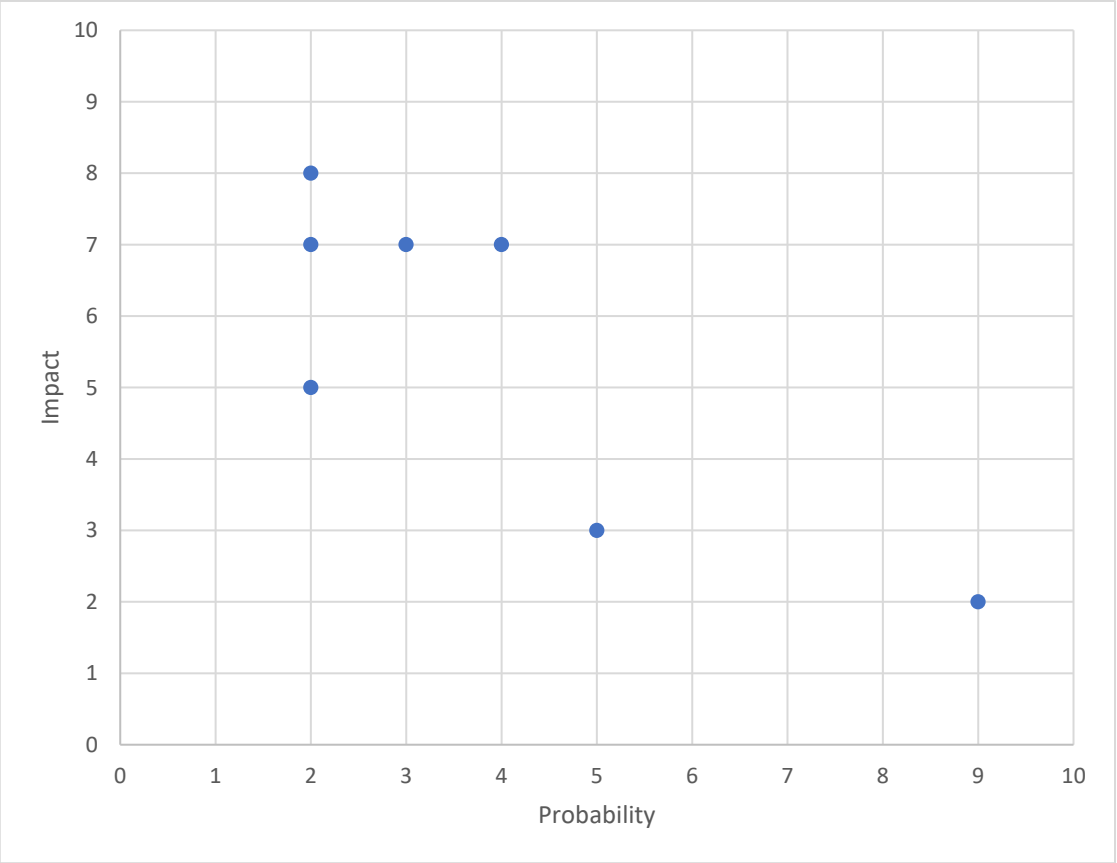

Figure 21 Risk map of the project

(Source: own creation)

Standout threats that should be dealt with are especially threads Incorrect configuration ID 4 with the highest value of threat. Threats with impact higher than 5 should be lowered.

Precautions to lower or eliminate the threats

Table 9 Threat precautions

(Source: own creation)

ID of the threat	Precaution	Type of precaution
1	Adequate project buffer	Lower probability
2	Use a test version of the system	Lower probability
3	Test version verification	Lower probability
4	Standard templates used as basis, second administrator verification where necessary	Lower probability, transfer
5	Thorough preparation phase, double verification	Lower probability
6	Monitoring in place, short disruption acceptable	Acceptance
7	Backup plan	Elimination

4 Deployment on a selection of assets, training and documentation

4.1 Deployment

4.1.1 Hardware

In this part I described the hardware of the server used for the project, as well as creating the required denomination in the IS (with regards to IP address allocation, recovery passwords, hardware configuration utilized etc.)

In order to satisfy the need for faster searches in recent logs and for longer term storage I selected the DELL R430 server with quad core Intel processor and 32GB of RAM and hardware raid controller. The data will be split into two groups hot and warm. Hot data will be stored on RAID1 (mirrored) connected Intel DC S4500 480GB SSDs for fast access, while the older data will be stored on RAID1 connected WD Gold 2TB HDDs.

As the incoming logs will be transferred through network interface the server selected is equipped with a 10Gb network card.

4.1.2 Installation

This part describes the utilized standard distribution of OS used as well as steps taken to prepare the SIEM software itself.

I utilized current release of Centos 7 as the underlying OS in the version 7.6 with added security measures for bruteforce attacks (fail2ban) as well as custom firewall rules to enable connections only from select servers as well as administrator IP range.

Installation of ELK stack

The ELK stack can be installed using standard package distribution method (in case of CentOS 7 the yum package manager) in the form of standard repository available directly from Elastic, the maintainer of ELK stack. To utilize the repository simply add the public signing key of the repository

sudo rpm --import https://packages.elastic.co/GPG-KEY-elasticsearch

and create the repository with a suitable name in your /etc/yum.repos.d/ directory with the following lines

[elasticsearch-7.x]

name=Elasticsearch repository for 7.x packages

baseurl=https://artifacts.elastic.co/packages/7.x/yum

gpgcheck=1

gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch

enabled=1

autorefresh=1

type=rpm-md

After doing so the services Elasticsearch Logstash and Kibana can be installed using

sudo yum install elasticsearch logstash kibana

and updated as necessary using

sudo yum update elasticsearch logstash kibana

In order to output to nonstandard outputs or ingest data from nonstandard inputs, it is necessary to also install the necessary input or output plugins. Plugins can be installed by calling (example using the Riemann plugin)

/usr/share/logstash/bin/logstash-plugin install logstash-output-riemann (11)

4.1.3 Usage

In this part of the documentation I list the ways you can check each ELK stack component in case of service malfunction or after a breaking change in an update.

Each part of the ELK stack is in this configuration installed through and updated through the rpm repository. Major version updates are distributed in separate repositories. Accidental update to a new major version is therefore unlikely. Current rpm repository can be found in https://www.elastic.co/guide/en/elasticsearch/reference/current/rpm.html When a new major version is available an update assistant in the Kibana Management tab

can be utilized to check if any settings of the current configuration are deprecated and need to be changed prior to updating.

Elasticsearch

Elasticsearch service operates via two primary communication ports. An http port (default TCP port 9200) is used for GET and PUT requests on Elasticsearch. Any action that can be called through the Kibana interface is also available as a http request on this port. Basic information about the service can be obtained by calling a GET request on the server.

curl -XGET http://localhost:9200/?pretty

The main points regarding Elasticsearch that should be monitored is whether the service is running and if the heartbeat input from Logstash is correctly stored. For this I set up our monitoring service Nagios to monitor the service via NRPE.

All other actions required by administrators regarding Elasticsearch can be accomplished through Kibana.

Logstash

The ingest pipeline adjustments after adding more monitored devices have a clearly defined place and format in the Logstash configuration files.

Kibana

The Kibana service should be running on the server and the configured apache http proxy, responsible for LDAP connection and authorization of users should be running as well.

The connection to Kibana is restricted on the firewall as well, allowing connections only from the IP addresses of administrators' office.

The Kibana interface is divided into two main parts. The navigation menu on the left side of the screen and the main window. The content of the main window changes based on the open app from within the navigation menu.

The most important part is the discover menu. In the discover menu you can create queries based on the ingested logs. Filters are also available with most frequent terms recommended for each term. It is possible to create filters directly with the query window, or through the recommended values of terms.

Kibana 7.0 replaced the original Lucene language syntax available in the older versions of Kibana with a new modified language called KQL (Kibana Query Language). The older syntax relied on spaces as a separator between search tokens, whereas the new syntax requires explicit Boolean operators, such as and, or, not.

Terms without a designated field are searched in all available fields.

Every ingested log contains a timestamp upon which a time series is created, thus allowing for time-based searches with frequently used options available as a direct shortcut. It is possible to select either a sliding time window (relative time to current) or an absolute time range.

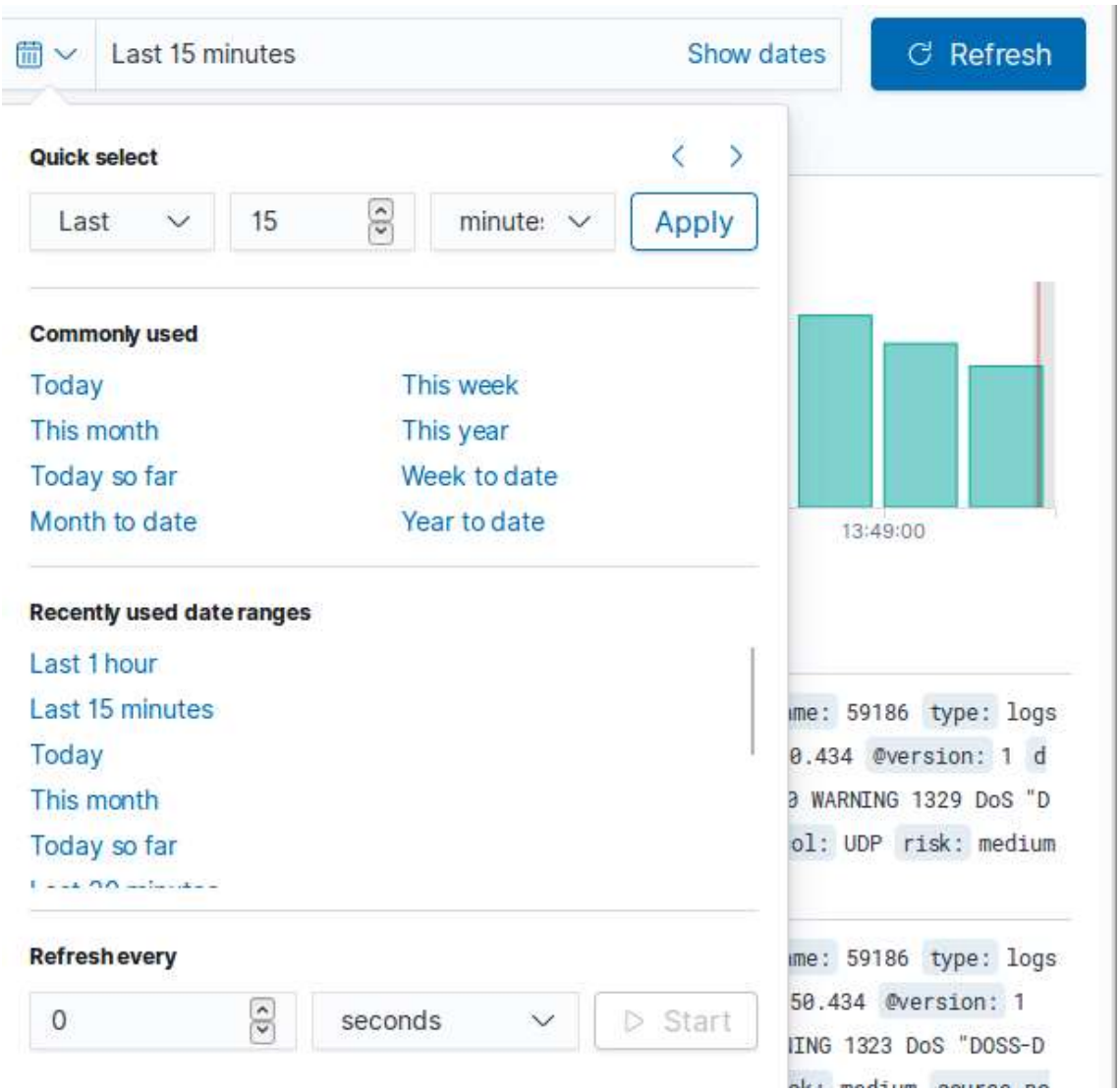

Figure 22 Time selection shortcuts

(Source: own creation)

Kibana also has an interface for manual Elasticsearch GET, PUT and POST queries with syntax checks built in. It provides an easy way of developing custom queries not available through the Kibana interface.

Figure 23 Console for queries

(Source: own creation)

4.1.4 Modifications

As Master Internet uses an internal DNS server with certain services available only on the local network and to speed up the process of DNS resolution, I prepared a local DNS caching service using the dnsmaq tool. It is configured to query internal DNS server and cache results to speed up Logstash transformation pipeline.

4.1.5 Troubleshooting recommendations

As all services of the ELK stack are connected and the function of the server and services running on it are monitored by internal Nagios monitoring server any SIEM service disruption will trigger an alert and the 24/7 technical support will be notified. First steps in troubleshooting should include checking the state of Elasticsearch, Logstash, Kibana and riemann services, depending on the type of alert received. As both Logstash and Kibana depend on Elasticsearch in order to function properly the restoration of Elasticsearch service should have a priority. Elastic utilizes java virtual machine in order to maintain a single codebase regardless of the platform it is deployed on. Java memory

garbage collection can sometimes cause issues and render the JVM inoperable. A restart of the disrupted service can force the garbage collection to start from scratch and resolve the issue.

Hardware errors are to be handled according to internal regulations with a component swap and testing of the offending component in an isolated environment and warranty return where applicable.

Networking errors can be solved using OOB (Out Of Band) management interface connected to a separate segment of the network.

4.2 Training

To properly and fully utilize available systems and processing power it is crucial to educate employees responsible for daily operation in the proper control and setting of the system. SIEM systems are no different from any other asset and uneducated administrator can pose a greater risk than an outside attacker. An internal seminar was conducted to introduce ELK to other administrators, how to operate both internal system and customer-deployed managed ELK instances.

4.3 Documentation

All documentation is made available in the form of internal Wikipedia in Master internet. Most relevant information for quick recovery of service is also directly on the SIEM server. Documentation includes information provided in chapter 4.1 and its subchapters.

4.4 Benefit analysis

To analyze the benefits of implemented project I compare the financial cost of the project to a potential major service disruption cost.

The cost of project can be split into a one-time cost of implementation and ongoing yearly maintenance costs. The hardware required to run the SIEM system and the time required to implement the system add up to the one-time cost. Electricity, cooling costs, updates and small modifications of the SIEM yield the maintenance costs. As the datacenter cooling and electricity costs must be spent anyway, we can omit them in the final calculation. The server hardware would be used for internal administrator services regardless, negating the server costs as well. The only real one-time cost of the project is therefore the time spent on the project, which I calculated as 40 MD (man-days) of administrator time, if we calculate with an hourly rate of 500 Kč the cost of project is 160 000 Kč. The yearly maintenance cost is expected to be approximately 10 MD of administrator time to add new logged services, maintain updates etc. The maintenance cost will be approximately 5000 Kč yearly. The project should provide a means to shorten or eliminate a security related service disruption, and as in ISP environment even a short disruption is very costly, I can calculate the financial benefit of the project. The cost of a single day disruption of service would represent more than 1/365th of a yearly income, as ISP SLAs (Service Level Agreements) contain penalty clauses depending on the length of service unavailability.

$$\frac{Yearly\ income\ of\ the\ company}{days\ in\ a\ year}\ (12) = \frac{110439000}{365} \doteq 302573\ Kč$$

Given that a single major disruption would cost approximately 140 000 Kč more than the project the financial benefit is obvious.

4.4.1 Future recommendations

Security is a never-ending process. New threats emerge every day and old threats change form to escape detection. For security measures to stay relevant and effective in mitigating the threats it is necessary to keep evolving the defensive measures. As the project of deploying a SIEM system is not trivial, there are still many possibilities to increase the breadth of monitored endpoints. In the process of deployment of SIEM system I recognized that the system would greatly benefit from adding correlation of IP addresses with lists of potentially dangerous source addresses provided by major spam listing companies, known malware URLs, brute force attackers and other bad actors. For example, https://threatfeeds.io/ is a free open-source aggregator of known or suspected bad actors from large security-focused projects such as AlienVault, abuse.ch or blocklist.de.

As the number of logged devices will increase it may eventually be necessary to add storage nodes to hold all the logs as well as satisfy the speed requirements.

Conclusion

The importance of security in the context of an ISP is increasing. Higher volumes of attacks as well as new attack vectors force companies to increase security. I therefore analyzed the requirements of Master Internet on the SIEM system, prepared an evaluation methodology of the systems and compared the most relevant available solutions. As most commercial solutions are very costly, I chose to implement an open-source SIEM solution using ELK stack.

I analyzed threats to the project of SIEM implementation, chose necessary countermeasures and prepared a timeline of the project.

In the implementation phase I followed the projected steps to prepare the necessary services, templates and security of the SIEM system itself. After a testing period with a subset of sources I proceeded to replace the old inadequate solution, finished the documentation on general system usage with tips for troubleshooting potential problems. I lead internal training of other administrators to fully utilize the potential of the system.

In the end I evaluated the benefits of the project and summarized potential future improvements.

The SIEM system for Master Internet has been successfully deployed into usage, all goals of the diploma thesis were achieved, and the author has fulfilled the objective of this paper.

References

1. *SP 800-92, Guide to Computer Security Log Management.* **(NIST), Author: Karen Kent and (NIST), Author: Murugiah Souppaya.** NIST SP 800-92.

2. **Swift, David.** A Practical Application of SIM/SEM/SIEM, Automating Threat Identification. [Online] [Cited: April 22, 2019.] http://www.sans.org/reading-room/whitepapers/logging/practical-application-sim-sem-siem-automating-threat-identification-1781.

3. **Cichonski, Paul R., et al.** Computer Security Incident Handling Guide. [Online] 2004. [Cited: April 23, 2019.] https://nist.gov/publications/computer-security-incident-handling-guide.

4. **Master Internet.** Logo of Master Internet. *Master.cz.* [Online] [Cited: April 25, 2019.] https://www.master.cz/images/logo.png.

5. **Master Internet inc.** Master Internet History. *Master Internet.* [Online] [Cited: April 21, 2019.] https://www.masterdc.com/history/.

6. **NÚKIB.** NÚKIB Auditní checklist. *NÚKIB materiály ke stažení.* [Online] [Cited: November 12, 2018.] https://nukib.cz/download/kii-vis/container-nodeid-580/vkbchecklistfinalv21rev.pdf?fbclid=IwAR3aYYEcGzmwjAVgpLvtZn1zYYdM3ry0WIrn3yHUcgq8McbNb5FJBFHBc-o.

7. **Elastic.** Elastick Stack. *Elastic.* [Online] [Cited: April 20, 2019.] https://www.elastic.co/elk-stack.

8. **Elastic User Group meetup. *Elastic User Group CZ meetup.* [Online] [Cited: February 13, 2019.] https://www.meetup.com/CZ-Elastic-Fantastics/.**

9. Graylog. Graylog homepage. *Web Graylog.* [Online] [Cited: April 23, 2019.] https://www.graylog.org/.

10. —. Graylog Rules. [Online] [Cited: April 25, 2019.] http://docs.graylog.org/en/latest/pages/pipelines/rules.html.

11. Elastic. Elastic rpm repository setup. *Elastic.com.* [Online] [Cited: January 1, 2019.] https://www.elastic.co/guide/en/beats/filebeat/current/setup-repositories.html.

12. Ministerstvo spravedlnosti České republiky. Sbírka listin Master Internet s.r.o. *Veřejný rejstřík a sbírka listin.* [Online] [Cited: April 28, 2019.] https://or.justice.cz/ias/ui/vypis-sl-detail?dokument=54435432&subjektId=236361&spis=724383.

13. Anton, Chuvakin, Phillips, Chris and Schmidt, Kevin. *Logging and Log Management: The Auitative.* Saint Louis : William Andrew, 2012. ISBN 978-15-9749-635-3.

List of figures

List of tables

List of abbreviations

SIEM Security Information and Event Management
SIM Security Information Management
SEM Security Event Management
LM Log Management
ISP Internet Service Provider
VLAN Virtual Local Area Network
CSV Comma Separated Values
SNMP Simple Network Management Protocol
XML Extensible Markup Language
YAML YAML Ain’t Markup Language
DMCA Digital Millennium Copyright Act
ICT Information and Communication Technologies
DNS Domain Name System
GUI Graphical User Interface
NÚKIB Národní Úřad pro Kybernetickou a Informační Bezpečnost
CII Critical Information Infrastructure
ISMS Information Security Management System
OSSIEM Open Source SIEM
API Application Programming Interface
ELK Elasticsearch Logstash Kibana
JSON JavaScript Object Notation
ILM Index Lifecycle Management
CLI Command Line Interface
RAID Redundant Array of Independent Disks
SSD Solid State Drive
HDD Hard Disk Drive
OS Operating System
OOB Out Of Band