PROJEKT

Multitenantné operačné centrum kybernetickej bezpečnosti riešené
ako otvorená cloudová služba s prvkami strojového učenia

**Previazané s balíkom KPB3 – Návrh ML algoritmov**

Previazané s výstupom V7 výsledok analýzy vhodných algoritmov

Financované
Európskou úniou
NextGenerationEU

PLÁN OBNOVY

ÚRAD VLÁDY
SLOVENSKEJ REPUBLIKY

# Obsah

**1. Metodológia ..... 3**
**2. VHODNÉ PUBLIKÁCIE ..... 4**
2.1. Enhancing intelligence SOC with big data tools ..... 4
2.2. Building a Flexible and Resource-Light Monitoring Plat form for a WLCG-Tier2 ..... 4
2.3. Observability in Microservices: An In-Depth Exploration of Frameworks, Challenges, and Deployment Paradigms ..... 5
2.4. An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity ..... 6
2.5. SUESSA: Sustainable & Ultra-Elastic Stack Security Architecture for Securing IoT Networks of Future Smart Cities ..... 7
2.6. Evaluating the Observability of Network Security Monitoring Strategies With TOMATO ..... 8
2.7. EDGE ANALYTICS AND COMPLEX EVENT PROCESSING FOR REAL TIME AIR POLLUTION MONITORING AND CONTROL ..... 8
2.8. Container-Based Honeypot Deployment for the Analysis of Malicious Activity ..... 9
2.9. Self-Healing Cloud Services in Private Multi-Clouds ..... 10
2.10. Flexible visualization of a 3rd party Intrusion Prevention (Security) tool: A use case with the ELK stack ..... 10
2.11. An ML Based Anomaly Detection System in real time data streams ..... 12
2.12. Anomaly Detection in POSTFIX mail log using Principal Component Analysis ..... 12
2.13. Cyberattacks Detection and Analysis in a Network Log System Using XGBoost with ELK Stack ..... 13
2.14. Improving State-of-the-art Compression Techniques for Log Management Tools ..... 14
2.15. SDGen: A Scalable, Reproducible and Flexible Approach to Generate Real World Cyber Security Datasets ..... 15
**Zoznam zdrojov ..... 16**

# 1. METODOLÓGIA

Literárna rešerš vhodných prípadov použitia ML/AI pri detekcii kybernetických anomálií, respektíve vyhľadávanie aplikácie AI/ML pre SOC bolo uskutočnené prostredníctvom vyhľadávania vhodných publikácií na platforme Web of Science (skrátene WoS) pre prístup k citačným údajom z vedeckých časopisov a konferenčných zborníkov. Na vyhľadávanie bol použitý reťazec kľúčových slov, respektíve kombinácia jednotlivých kľúčových slov, pričom statickým prvkom bolo použitie buď kľúčového slova „Elastic Stack / ELK Stack“ alebo „OpenSearch“, nakoľko sa jedná o otvorené dátové analytické platformy pre bezpečnosť, respektíve systémy na zber, indexovanie, vyhľadávanie vizualizáciu a analýzu logov a udalostí, ktoré podporujú detekciu kybernetických hrozieb a anomálií prostredníctvom pravidiel aj strojového učenia. Vyhľadávanie na WoS prebehlo kombináciou týchto hlavných kľúčových slov v doplnení s nasledujúcimi doplnkovými kľúčovými slovami:

A. AI
B. ML
C. SOC
D. ANOMALY DETECTION
E. CYBER SECURITY
F. SECURITY OPERATIONS
G. SECURITY MONITORING
H. ALERTING

Výsledok analýzy v prípade využitia kľúčového slova „Elastic Stack“, respektíve „ELK Stack“ je 31 publikácií (tabuľka 1).

Tabuľka 1 Počet publikácií podľa kľúčových slov v kombinácií s „Elastic Stack/ ELK Stack“

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Počet | 3 | 1 | 3 | 12 | 5 | 1 | 4 | 2 |

Najväčší počet publikácií, teda 12 publikácií, bolo možné nájsť s použitím reťazca kľúčových slov „Elastic Stack/ ELK Stack“ AND „ANOMALY DETECTION“.

Výsledok analýzy v prípade využitia kľúčového slova „OpenSearch“ v kombinácií s určenými dodatočnými kľúčovými slovami je rovný dvom publikáciám (tabuľka 2).

Tabuľka 2 Počet publikácií podľa kľúčových slov v kombinácií s „OpenSearch“

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Počet | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

V prípade vyhľadávania vhodných publikácií v kontexte analytickej platformy OpenSearch boli nájdené teda len dve publikácie, a to v doplnení kľúčového slova „SECURITY ALERTING“ a „ALERTING“-

Nájdených 33 publikácií bolo následne analyzovaných na základe dvoch kritérií, a to ich dostupnosti, teda možnosti dohľadania plného textu publikácie, a použitého jazyku publikácie (angličtina). Z 33 publikácií bolo identifikovaných 15 publikácií, ktoré spĺňajú dané kritériá, teda sú publikované v anglickom jazyku, a zároveň sú open source.

# 2. VHODNÉ PUBLIKÁCIE

Táto kapitola obsahuje stručný popis jednotlivých publikácií, respektíve ich abstrakty v pôvodnom znení (angličtine), spolu s ich slovenskými prekladmi. Folino et al. (2023) sa priamo venovali problematike detekcie kybernetických hrozieb. Kulshrestha & Durbha (2020) vytvárali systém detekcie anomálií v reálnom čase. Kyriakou & Sklavos (2018) skúmali ako zachytiť nové hrozby v oblasti kybernetickej bezpečnosti. Poat et al. (2023) sa snažili dosiahnuť automatizovanú prevenciu narušení v reálnom čase naprieč rôznymi systémami, so zameraním na okamžitý prehľad o tom, čo a odkiaľ bolo blokované, kto bol zablokovaný, a zároveň mať možnosť informovať komunitu v prípade odobratia prístupu, bez potreby vytvárania „ticketov“, ktoré môžu byť vybavené s oneskorením. Jose et al. (2021) sa zamerali na pokročilejšie prístupy systémov detekcie sieťových anomálií. S dôrazom nie len na robustný algoritmus detekcie anomálií, ale aj mechanizmus prenosu dát v reálnom čase, ktorý minimalizuje oneskorenia reakcie. Výskum sa zameriava na implementáciu streamovacieho systému v reálnom čase pre detekciu sieťových anomálií, schopného fungovať aj pri vysokom objeme dátovej prevádzky. Tran & Tran (2018) navrhli metódu na identifikáciu anomálií v dátovom súbore, ktorá dokáže detegovať anomáliu v čase, keď dôjde k útoku prostredníctvom spamových e-mailov. Yang et al. (2021) skúmali používanie umelej inteligencie a metód strojového učenia pri kybernetických útokoch, najmä so zameraním na obranné metódy proti týmto útokom ich detekciou a identifikovaním prostredníctvom pozorovania logovacích dát a analýzy, či obsahujú abnormálne správanie.

## 2.1. Enhancing intelligence SOC with big data tools

**Abstract:** Traditional security tools are based on predetermined signatures or rules that show rigidity in the face of the dynamics of the interconnections that occur in organizations. The inclusion of analytic in security systems extends the field of vision of security analysts for proactive threat detection. This allows SOC specialists to make decisions in real time and focus on the protection of critical assets of the organization. Our contribution in this work is to analyze the applicability of the use of Big Data as a complementary tool for the detection of security events in a real CSIRT environment, validating the architecture, configuration and visualization using the ELK stack as a Big Data platform. (Andrade & Torres, 2018)

**keywords:** security operation center; cognitive security; big data; cybersecurity

**Abstrakt:** Tradičné bezpečnostné nástroje sú založené na vopred stanovených podpisoch alebo pravidlách, ktoré sú voči dynamike prepojení v organizáciách pomerne rigidné, nepružné. Zaradenie analytických nástrojov do bezpečnostných systémov rozširuje zorné pole bezpečnostných analytikov pre proaktívnu detekciu hrozieb. To umožňuje špecialistom SOC prijímať rozhodnutia v reálnom čase a sústrediť sa na ochranu kritických aktív organizácie. Prínosom v tejto práci je analýza použiteľnosti Big Data ako doplnkového nástroja na detekciu bezpečnostných udalostí v reálnom prostredí CSIRT, overenie architektúry, konfigurácie a vizualizácie pomocou ELK stacku ako platformy Big Data. (Andrade & Torres, 2018)

## 2.2. Building a Flexible and Resource-Light Monitoring Plat form for a WLCG-Tier2

**Abstract:** Software development projects at Edinburgh identified a desire to build and manage our own monitoring platform. This better allows us to support the developing and varied physics and computing interests of our Experimental Particle Physics group. This production platform enables oversight of international experimental data management, local

software development projects and active monitoring of lab facilities within our research group. Larger sites such as CERN have access to many resources to support generalpurpose centralised monitoring solutions such as MONIT. At a WLCG Tier2 we only have access to a fraction of these resources and manpower. Recycling nodes from grid storage and borrowed capacity from our Tier2 Hypervisors has enabled us to build a reliable monitoring infrastructure. This also contributes back to our Tier2 management improving our operational and security monitoring. Shared experiences from larger sites gave us a head-start in building our own service monitoring (FluentD) and multi-protocol (AMQP/STOMP/UDP datagram) messaging frameworks atop both our Elasticsearch and OpenSearch clusters. This has been built with minimal hardware and software complexity, maximising maintainability, and reducing manpower costs. A secondary design goal has also been the ability to migrate and upgrade individual components with minimal service interruption. To achieve this, we made heavy use of different layers of containerisation (Podman/Docker), virtualization and NGINX web proxies. This presentation details our experiences in developing this platform from scratch with a focus on minimal resource use. This includes lessons learnt in deploying and comparing both an Elasticsearch and OpenSearch clusters, as well as designing various levels of automation and resiliency for our monitoring framework. This has culminated in us effectively indexing, parsing and storing >200GB of logging and monitoring data per day. (Currie & Yuan, 2024)

**Abstrakt:** Softvérové vývojové projekty v Edinburghu identifikovali potrebu vybudovať a spravovať vlastnú monitorovaciu platformu. Táto iniciatíva umožňuje efektívnejšie podporovať dynamicky sa rozvíjajúce a rôznorodé záujmy v oblasti fyziky a výpočtovej techniky v rámci skupiny experimentálnej časticovej fyziky. Produkčná platforma poskytuje dohľad nad medzinárodným riadením experimentálnych dát, lokálnymi softvérovými vývojovými projektmi a zabezpečuje aktívne monitorovanie laboratórnych zariadení vo výskumnej infraštruktúre. Veľké inštitúcie, ako CERN, disponujú rozsiahlymi zdrojmi na implementáciu centralizovaných monitorovacích riešení, akým je napríklad systém MONIT. Na úrovni WLCG Tier2 však je k dispozícii iba obmedzený objem zdrojov a pracovnej kapacity. Recyklácia uzlov z grid storage a využívanie zapožičanej kapacity z Tier2 hypervízorov umožnili vybudovať robustnú a spoľahlivú monitorovaciu infraštruktúru. Tento prístup zároveň prináša pridanú hodnotu aj pre samotnú správu Tier2 prostredia zlepšením prevádzkového a bezpečnostného monitoringu. Skúsenosti zdieľané väčšími centrami poskytli významnú výhodu pri návrhu a implementácii vlastného systému monitorovania služieb (FluentD) a multiprotokolových správových rámcov (AMQP/STOMP/UDP datagram) nad klastrami Elasticsearch a OpenSearch. Celá infraštruktúra bola navrhnutá s dôrazom na minimálnu hardvérovú a softvérovú komplexnosť, čo zvyšuje jej udržiavateľnosť a znižuje náklady na pracovnú silu. Sekundárnym dizajnovým cieľom bola schopnosť migrovať a modernizovať jednotlivé komponenty s minimálnym dopadom na dostupnosť služieb. Na dosiahnutie tohto cieľa sa intenzívne využili viaceré vrstvy kontajnerizácie (Podman/Docker), virtualizácie a NGINX webové proxy servery. Táto prezentácia sumarizuje skúsenosti s budovaním platformy od základov, s dôrazom na efektívne využitie dostupných zdrojov. Obsahuje aj poznatky získané pri nasadzovaní a porovnávaní Elasticsearch a OpenSearch klastrov, ako aj pri návrhu rôznych úrovní automatizácie a odolnosti monitorovacieho rámca. Výsledkom je schopnosť efektívne indexovať, parsovať a ukladať viac než 200 GB logovacích a monitorovacích dát denne. (Currie & Yuan, 2024)

## 2.3. Observability in Microservices: An In-Depth Exploration of Frameworks, Challenges, and Deployment Paradigms

**Abstract:** The widespread adoption of microservices-based architectures in native cloud systems has amplified the need for robust observability strategies to ensure system reliability and performance. Microservices, while enabling scalability and flexibility, introduce

complexities such as distributed failure points, performance bottlenecks, and resource mismanagement. This paper provides a comprehensive review of state-of-the-art observability frameworks and tools designed for microservices architecture. It focuses on key aspects of observability at the system, service, and network levels, utilizing logs, traces, and metrics as core data sources. A thematic taxonomy is proposed, classifying the diverse approaches to monitoring microservices based on implementation paradigms, deployment platforms, and architecture patterns. The study compares existing solutions in terms of their capabilities, limitations, and effectiveness in addressing observability challenges in cloud, edge, and fog environments. Furthermore, open research issues are identified, and practical recommendations are provided to optimize observability in complex, distributed microservices ecosystems. This work aims to serve as a valuable resource for system architects, DevOps engineers, and researchers striving to enhance the reliability and performance of modern cloudnative systems. (Ummay Faseeha et al., 2025)

**keywords:** Microservices observability; Cloud-based microservices; System reliability; Containerized computing environments; Real-time performance monitoring

**Abstrakt:** Rozšírené prijatie architektúr založených na mikroslužbách v cloudových systémoch natívnych pre prostredie cloudu zvýraznilo potrebu robustných stratégií pozorovateľnosti (observability strategies) na zabezpečenie spoľahlivosti a výkonu systémov. Mikroslužby síce umožňujú škálovateľnosť a flexibilitu, avšak zároveň prinášajú komplexitu v podobe distribuovaných miest zlyhania, výkonnostných úzkych profilov a neefektívneho riadenia zdrojov. Táto práca poskytuje komplexný prehľad najmodernejších observability rámcov a nástrojov navrhnutých pre architektúru mikroslužieb. Zameriava sa na kľúčové aspekty observability na úrovni systému, služby a siete, pričom využíva logy, trasovania a metriky ako základné dátové zdroje. Je navrhnutá tematická taxonómia, ktorá klasifikuje rôzne prístupy k monitorovaniu mikroslužieb na základe implementačných paradigmat, nasadzovacích platforiem a architektonických vzorov. Štúdia porovnáva existujúce riešenia z hľadiska ich možností, obmedzení a efektívnosti pri riešení výziev observability v prostrediach cloudu, edge computingu a fog computingu. Okrem toho identifikuje otvorené výskumné problémy a prináša praktické odporúčania na optimalizáciu observability v komplexných distribuovaných ekosystémoch mikroslužieb. Cieľom tejto práce je slúžiť ako hodnotný zdroj pre systémových architektov, DevOps inžinierov a výskumníkov usilujúcich sa o zvýšenie spoľahlivosti a výkonu moderných cloud-native systémov. (Ummay Faseeha et al., 2025)

## 2.4. An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity

**Abstract:** Nowadays, the speed of the user and application logs is so quick that it is almost impossible to analyse them in real time without using high-performance systems and platforms. In cybersecurity, human behaviour is responsible directly or indirectly for the most common attacks (i.e. ransomware and phishing). To monitor user behaviour, it is necessary to process fast user logs coming from different and heterogeneous sources, having part of the data or some entire sources missing. A framework based on the elastic stack (ELK) to process and store log data in real time from different users and applications is proposed for this aim. This system generates an ensemble of models to classify user behaviour and detect anomalies in real time, exploiting the advantages of the ELK-based software architecture and of the Kubernetes platform. In addition, a distributed evolutionary algorithm is used to classify the users by exploiting their digital footprints derived from many data sources. Experiments conducted on two real-life data sets verify the approach's goodness in detecting anomalies in user behaviour, coping with missing data and lowering the number of false alarms. (Folino et al., 2023)

**keywords:** Cybersecurity · Ensemble learning · Anomaly detection · Behavior analytics

**Abstrakt:** V súčasnosti je rýchlosť generovania používateľských a aplikačných logov taká vysoká, že ich analýza v reálnom čase je prakticky nemožná bez využitia vysokovýkonných systémov a platforiem. V oblasti kybernetickej bezpečnosti je ľudské správanie priamo alebo nepriamo zodpovedné za najčastejšie typy útokov (napr. ransomware a phishing). Na monitorovanie používateľského správania je nevyhnutné spracovávať rýchlo prichádzajúce logy z rôznych heterogénnych zdrojov, pričom časť dát alebo dokonca celé zdroje môžu chýbať. Na tento účel je navrhnutý rámec založený na Elastic stacku (ELK), ktorý umožňuje spracovávať a ukladať logovacie dáta v reálnom čase z rôznych používateľských a aplikačných zdrojov. Systém generuje súbor modelov na klasifikáciu používateľského správania a detekciu anomálií v reálnom čase, pričom využíva výhody softvérovej architektúry založenej na ELK a platformy Kubernetes. Okrem toho sa využíva distribuovaný evolučný algoritmus na klasifikáciu používateľov na základe ich digitálnych stôp odvodených z viacerých dátových zdrojov. Experimenty uskutočnené na dvoch reálnych dátových súboroch potvrdzujú efektívnosť navrhovaného prístupu pri detekcii anomálií v používateľskom správaní, pri práci s neúplnými dátami a pri znižovaní počtu falošných poplachov. (Folino et al., 2023)

## 2.5. SUESSA: Sustainable & Ultra-Elastic Stack Security Architecture for Securing IoT Networks of Future Smart Cities

**Abstract:** Recently, the concept of a super smart city has received special attention and has become the most popular agenda among city planners, policy makers and IT industries around the globe. In this paper, we proposed a novel Sustainable Ultra-Elastic Stack Security Architecture (SUESSA) which can address various future security challenges of artificial intelligence (AI), big data and communication networks. Our architecture employs the holistic approach based on sustainability of network security which is different than that of the prevailing security practices. We propose to apply the sustainable approach in the Network security architecture through augmentation of cloud, networks and end-point level, our approach provides a robust and ultra-elastic mechanism of security that surpasses the prevailing architecture. We evaluated the feasibility and performance of the security stack containing a machine learning component, Elastic search engine, Kibana and Logstash as the building blocks of the architecture. Based on our survey, SUESSA architecture overcomes the disadvantages of current architecture specifically solving the problem of sustainability strongly demanded in security architecture of the future smart cities. (Gautam & Norio, 2020)

**keywords:** Sustainable & Elastic Security Architecture, Holistic Approach, Super Smart City, Future Cities, Artificial Intelligence

**Abstrakt:** V poslednom období získal koncept super inteligentného mesta osobitnú pozornosť a stal sa jednou z najpopulárnejších tém medzi mestskými plánovačmi, tvorcami politík a IT priemyslom na celom svete. V tomto článku je navrhnutá nová architektúra Sustainable Ultra-Elastic Stack Security Architecture (SUESSA), ktorá dokáže riešiť rôzne budúce bezpečnostné výzvy spojené s umelou inteligenciou (AI), big data a komunikačnými sieťami. Táto architektúra využíva holistický prístup založený na udržateľnosti sieťovej bezpečnosti, ktorý sa odlišuje od tradičných bezpečnostných praktík. Navrhuje sa aplikovať princípy udržateľnosti do architektúry sieťovej bezpečnosti prostredníctvom rozšírenia na úrovni cloudu, sietí a koncových bodov. Tento prístup poskytuje robustný a ultra-elastický bezpečnostný mechanizmus, ktorý prekonáva existujúce architektúry. Overila sa realizovateľnosť a výkon bezpečnostného stacku, ktorého základné komponenty tvoria strojové učenie, vyhľadávací engine Elasticsearch, vizualizačný nástroj Kibana a spracovateľ logov Logstash. Na základe tohto výskumu architektúra SUESSA prekonáva nevýhody

súčasných riešení, pričom konkrétne rieši problém udržateľnosti, ktorý je kriticky požadovaný v bezpečnostných architektúrach budúcich inteligentných miest. (Gautam & Norio, 2020)

## 2.6. Evaluating the Observability of Network Security Monitoring Strategies With TOMATO

**Abstract:** Monitoring systems for malicious behavior increasingly requires aggregating and analyzing data from various sources, such as network flows, host logs, and end-point monitoring platforms. However, there's currently a lack of metrics and methodologies to compute the observability and efficiency of a security monitoring strategy. This manuscript introduces TOMATO (Threat Observability & Monitoring Assessment Tool), which is a platform to evaluate the effectiveness of a security monitoring strategy by exploring both the number of known adversarial techniques that can be detected within a network, along with evaluating the number of false-positives produced by the monitoring strategy. The output produces both an observability score and efficiency score of a set of deployed monitoring techniques, which are evaluated based on the data from the environment, and simulated attacks generated from MITRE ATT&CK. The proposed approach is then integrated into an ELK stack and evaluated on real SCADA devices within the WSU Smart City Testbed. (Halvorsen et al., 2019)

**keywords:** Computer security, network security, monitoring.

**Abstrakt:** Monitorovacie systémy zamerané na detekciu škodlivého správania čoraz viac vyžadujú agregáciu a analýzu dát z rôznych zdrojov, ako sú sieťové toky, hostiteľské logy a platformy na monitorovanie koncových bodov. V súčasnosti však chýbajú metriky a metodológie na vyhodnocovanie pozorovateľnosti (observability) a efektívnosti bezpečnostnej monitorovacej stratégie. Tento článok predstavuje TOMATO (Threat Observability & Monitoring Assessment Tool), platformu určenú na hodnotenie účinnosti bezpečnostnej monitorovacej stratégie. Nástroj vyhodnocuje jednak počet známych adversariálnych techník, ktoré je možné detegovať v sieti, ako aj počet falošne pozitívnych nálezov generovaných monitorovacou stratégiou. Výstupom sú dve metriky, observability score a efficiency score, pre súbor nasadených monitorovacích techník. Tieto metriky sú vypočítané na základe dát zo skúmaného prostredia a simulovaných útokov vytvorených podľa rámca MITRE ATT&CK. Navrhovaný prístup je následne integrovaný do ELK stacku a vyhodnotený na reálnych zariadeniach SCADA v rámci WSU Smart City Testbed. (Halvorsen et al., 2019)

## 2.7. EDGE ANALYTICS AND COMPLEX EVENT PROCESSING FOR REAL TIME AIR POLLUTION MONITORING AND CONTROL

**Abstract:** The advent of the Internet of Things (IoT) has led to the generation of tremendous amounts of data from various sources. Cloud based systems are effective in storage and application of machine learning algorithms on such datasets. However, in some cases it is important to enable real time processing for making immediate decisions. There are many applications which require instantaneous analysis of the generated data for remedies in event of an anomaly. Data associated with such use cases remains significant only for a short duration of time. Various electronic sensors, e.g. Temperature, Moisture, Air Quality, Pressure, Wind Velocity etc. present in a Wireless Sensor Network generate streams of values. It can be processed using pipelines which provide prompt and quick analysis for decision making. Stream Processing Systems can be helpful in such cases as they analyse data streams within a few milliseconds to a few seconds. In this paper, we discuss an event based processing of streaming data from air pollution sensors to create a real time anomaly detection system. To reduce the delays associated with the generation of alarms in our pipeline, Apache Foundation's Stream Processing Tools, Kafka and Flink were used for operations on our

streams. To further accelerate the process, all the analysis is conducted on an embedded edge computing gateway device rather than sending data to the cloud for batch processing. The results are obtained in the form of a geographical map visualization using ELK stack. The map highlights the coordinates of the location with an unhealthy air quality index in real time. (Kulshrestha & Durbha, 2020)

**keywords:** Internet of Things, Complex Event Processing, Air Pollution Sensors, Stream Processing, LoRaWAN, Geospatial Big Data Analysis

**Abstrakt:** Nástup Internetu vecí (IoT) viedol k masívnej generácii dát z rôznych zdrojov. Cloudové systémy sú efektívne pri ukladaní a aplikácii algoritmov strojového učenia na takéto dátové súbory. V niektorých prípadoch je však dôležité umožniť spracovanie dát v reálnom čase, aby bolo možné okamžite prijímať rozhodnutia. Existuje mnoho aplikácií, ktoré vyžadujú okamžitú analýzu generovaných dát na prijatie opatrení pri výskyte anomálií. Dáta spojené s takýmito prípadmi majú význam iba počas krátkeho časového obdobia. Rôzne elektronické senzory, napríklad senzory teploty, vlhkosti, kvality ovzdušia, tlaku, rýchlosti vetra a pod., ktoré sú súčasťou bezdrôtovej senzorovej siete, generujú prúdy hodnôt. Tie je možné spracovávať pomocou dátových pipeline, ktoré poskytujú rýchlu a okamžitú analýzu na podporu rozhodovania. Stream Processing Systems sú v takýchto prípadoch užitočné, pretože dokážu analyzovať dátové toky v časovom horizonte od niekoľkých milisekúnd po niekoľko sekúnd. V tomto článku sa diskutuje o udalostne orientovanom spracovaní dátových tokov zo senzorov znečistenia ovzdušia s cieľom vytvoriť systém detekcie anomálií v reálnom čase. Na zníženie oneskorení pri generovaní alarmov v pipeline boli použité nástroje Apache Foundation pre streamové spracovanie dát, Kafka a Flink. Na ďalšie zrýchlenie procesu je celá analýza vykonávaná na zabudovanom edge computing gateway zariadení, namiesto odosielania dát do cloudu na batch spracovanie. Výsledky sú následne prezentované vo forme geografickej mapovej vizualizácie využívajúcej ELK stack, pričom mapa v reálnom čase zvýrazňuje súradnice lokalít s nezdravým indexom kvality ovzdušia. (Kulshrestha & Durbha, 2020)

## 2.8. Container-Based Honeypot Deployment for the Analysis of Malicious Activity

**Abstract:** In today's world, the field of cyber security is a fastpaced changing environment. New threats are continuously emerging, and the ability to capture and effectively analyze them is paramount. In our work, we are deploying multiple honeypot sensors in order to monitor and study the actions of the attackers. The selected honeypots are Cowrie, Dionaea and Glastopf, presented as a Linux host, a Windows host and a Web application respectively. This enables us to have a diverse and broad environment that can attract attackers aiming at different attack surfaces. The sensors are running on a containerization platform, Docker and in this way, they are lightweight, resilient and could be easily deployed and managed. Our goal is the creation of a single dashboard that can present the captured data effectively in real-time and both in macroscopic and microscopic levels. Thus, we are utilizing the Elastic Stack and we are enriching our data sources using Virus Total's analysis engine. The proposed system ran for a three-month period and provided numerous data points, from which instantaneous useful conclusions were drawn for the behavior and nature of the malicious users. (Kyriakou & Sklavos, 2018)

**keywords:** honeypot, Docker, virtualization, intrusion detection, malware, Cowrie, Dionaea, Glastopf, Elastic Stack

**Abstrakt:** V súčasnom svete je oblasť kybernetickej bezpečnosti dynamicky sa meniace prostredie. Neustále vznikajú nové hrozby a schopnosť ich zachytiť a efektívne analyzovať je kľúčová. V rámci tejto práce sa nasadzuje viacero honeypot senzorov s cieľom monitorovať a skúmať správanie útočníkov. Vybrané honeypoty sú Cowrie, Dionaea a

Glastopf, reprezentujúce Linux hostiteľa, Windows hostiteľa a webovú aplikáciu. Tento prístup umožňuje vytvoriť rôznorodé a rozsiahle prostredie, ktoré dokáže prilákať útočníkov zameraných na rôzne útočné plochy. Senzory sú prevádzkované na platforme Docker, čím sú ľahké, odolné a jednoducho nasaditeľné a spravovateľné. Cieľom je vytvorenie jednotného dashboardu, ktorý dokáže efektívne prezentovať zachytené dáta v reálnom čase, a to na makroskopickej aj mikroskopickej úrovni. Na tento účel sa využíva Elastic Stack a dátové zdroje sa obohacujú pomocou analytického enginu VirusTotal. Navrhovaný systém bežal počas trojmesačného obdobia a poskytol množstvo dátových bodov, z ktorých bolo možné okamžite vyvodiť užitočné závery o správaní a povahe škodlivých používateľov. (Kyriakou & Sklavos, 2018)

## 2.9. Self-Healing Cloud Services in Private Multi-Clouds

**Abstract:** As clouds continue to dominate as the de facto means of consuming shared computing resources, cloud providers are required to ensure that their services always meet or exceed service level agreements in terms of availability, reliability, scalability, performance etc every hour, day and year. These challenging requirements are even exacerbated when it comes to multi-cloud environments. In the era of automation and continuous cost reductions, using humans to manually test and trigger corrective actions when needed in order to fulfill these service level agreements is considered cumbersome and too expensive. In order to thrive, cloud service providers need a solution which is efficient, cheaper to deploy and easy to operate. This paper introduces a distributed container based solution for testing and self-healing of cloud resources. The proposed solution is based on a containerized multi-agent architecture which leverages Elasticsearch, Logstash, Kibana (ELK) stack and uses a rule-based self-healing algorithm implemented using Drools rule engine and Spring Boot framework. Preliminary tests show that the solution meets its target objectives in terms of maintaining service level agreements, lowering deployment costs, high performance and operability. (Mfula & Nurminen, 2018)

**keywords:** self-healing; multi-cloud; cloud computing

**Abstrakt:** Keďže cloudové riešenia naďalej dominujú ako de facto spôsob využívania zdieľaných výpočtových zdrojov, poskytovatelia cloudu musia zabezpečiť, aby ich služby vždy spĺňali alebo presahovali úroveň service level agreements (SLA) z hľadiska dostupnosti, spoľahlivosti, škálovateľnosti, výkonu a pod., a to v každej hodine, dni aj roku. Tieto náročné požiadavky sa ešte viac zvýrazňujú v prostredí multi-cloud infraštruktúr. V ére automatizácie a kontinuálneho znižovania nákladov je využívanie ľudskej práce na manuálne testovanie a spúšťanie korekčných akcií pre plnenie SLA považované za zdĺhavé a finančne príliš náročné. Pre úspešné fungovanie potrebujú poskytovatelia cloudových služieb riešenie, ktoré je efektívne, lacnejšie na nasadenie a jednoduché na prevádzku. Tento článok predstavuje distribuované kontajnerové riešenie pre testovanie a samoopravné mechanizmy cloudových zdrojov. Navrhované riešenie je založené na kontajnerizovanej multi-agentovej architektúre, ktorá využíva Elasticsearch, Logstash a Kibana (ELK) stack a aplikuje pravidlovo riadený samoopravný algoritmus, implementovaný pomocou Drools rule engine a Spring Boot frameworku. Predbežné testy ukazujú, že riešenie spĺňa stanovené ciele z hľadiska dodržiavania SLA, znižovania nákladov na nasadenie, dosahovania vysokého výkonu a jednoduchej prevádzkovateľnosti. (Mfula & Nurminen, 2018)

## 2.10. Flexible visualization of a 3rd party Intrusion Prevention (Security) tool: A use case with the ELK stack

**Abstract:** A difficult aspect of cyber security is the ability to achieve automated real time intrusion prevention across various sets of systems. To this extent, several companies are

offering comprehensive solutions that leverage an “accuracy of scale” and moving much of the intelligence and detection on the Cloud, relying on an ever-growing set of data and analytics to increase decision accuracy. Often, they provide tools to visualize the decision workflows in attack prevention (as well as tune the algorithm) but those solutions are not always practical as companies see the problem as “global” that is, from a unified Cyber-security standpoint. However, a key to a successful Cyber-security program is transparency and trust: from an experimental team viewpoint, this specifically means having the ability to immediately see what and from where, who has been blocked and being able to inform the community in case of a revoked access without the need for filing a “ticket” (that may eventually be answered) – in other words, rapid response to their user-base is essential but solutions targeting “sub-groups” in an organization are not often available. We have come up with a versatile solution leveraging the ELK stack (Elasticsearch, Logstash, & Kibana) and an IPS (Intrusion Prevention System) based WAF (Web Application Firewall) from Signal Sciences. Signal Science allows the streaming of detailed logs in a Logstash format suitable for custom solutions for visualization. By combining these two tools, we have strengthened our security posture and enabled individual experiments to monitor their own traffic. Specifically, the IPS WAF provides unique data such as country of origin, protocol, response code, source IP, and paths accessed. In this contribution, we will show how we engineered a visualization solution so experiment groups could access a dashboard with predefined graphs but also, where they can create individual customizable dashboards used to display blocked traffic and troubleshoot latency issues. We will discuss the details and procedures for developing and configuring these tools and how it benefits cyber security postures across our scientific based environment. (Poat et al., 2023)

**Abstrakt:** Jedným z náročných aspektov kybernetickej bezpečnosti je schopnosť dosiahnuť automatizovanú prevenciu narušení v reálnom čase naprieč rôznymi systémami. V tejto súvislosti viaceré spoločnosti ponúkajú komplexné riešenia, ktoré využívajú tzv. „accuracy of scale“ a presúvajú veľkú časť inteligencie a detekcie do cloudu, pričom sa spoliehajú na neustále rastúci objem dát a analytických nástrojov na zvýšenie presnosti rozhodovania. Často poskytujú aj nástroje na vizualizáciu rozhodovacích workflow pri prevencii útokov (ako aj možnosti ladenia algoritmov), no tieto riešenia nebývajú vždy praktické, keďže firmy vnímajú problém z globálnej perspektívy jednotnej kybernetickej bezpečnosti. Kľúčom k úspešnému programu kybernetickej bezpečnosti je však transparentnosť a dôvera. Z pohľadu experimentálneho tímu to znamená mať okamžitý prehľad o tom, čo a odkiaľ bolo blokované, kto bol zablokovaný, a zároveň mať možnosť informovať komunitu v prípade odobratia prístupu, bez potreby vytvárania „ticketov“, ktoré môžu byť vybavené s oneskorením. Inými slovami, rýchla reakcia voči používateľskej základni je zásadná, no riešenia zamerané na sub-skupiny v rámci organizácie nie sú často dostupné. Navrhnuté bolo preto univerzálne riešenie založené na ELK stacku (Elasticsearch, Logstash, Kibana) a IPS (Intrusion Prevention System) typu WAF (Web Application Firewall) od spoločnosti Signal Sciences. Signal Sciences umožňuje streamovanie detailných logov vo formáte Logstash, ktoré sú vhodné pre tvorbu vlastných vizualizačných riešení. Kombináciou týchto dvoch nástrojov sa posilnila bezpečnostná architektúr a umožnila sa jednotlivým experimentom monitorovať vlastná prevádzka. Konkrétne IPS WAF poskytuje unikátne údaje, ako sú krajina pôvodu, protokol, návratový kód, zdrojová IP adresa a prístupované cesty. V tomto príspevku je ukázané, ako sa navrhlo vizualizačné riešenie, ktoré umožňuje experimentálnym skupinám pristupovať k dashboardu s preddefinovanými grafmi, ale aj vytvárať individuálne prispôsobené dashboardy na zobrazovanie blokovanej prevádzky a riešenie problémov s latenciou. Podrobne sú predstavené postupy pri vývoji a konfigurácii týchto nástrojov a ukázané, ako prispievajú k posilneniu kybernetickej bezpečnosti v prostredí orientovanom na vedecký výskum. (Poat et al., 2023)

## 2.11. An ML Based Anomaly Detection System in real time data streams

**Abstract:** Due to the advancements in machine learning and artificial intelligence applied fields, network anomaly detection systems have experienced an evolution from traditional signaturebased methods for intrusion detection. Nonetheless, as security measures evolve, more sophisticated attacks are also constantly being developed by hackers. Not only a robust anomaly detection algorithm is needed, but also a real-time data feeding mechanism for minimizing the reaction-time impact is required. Moreover, DDoS attacks can flood the network data channels with more than thousands of packets per second with the latent effect of overloading most traditional monitoring systems that rely on data storage. Due to this, the research presented in this paper focuses its efforts on implementing a real-time data streaming system for network anomaly detection that can operate during a high volume of traffic data. The solution includes the deployment of a flow collector platform connected to Apache Kafka for receiving NetFlow data from network switches. Also, real-time big data processing techniques are applied through Apache Spark, where the ML anomaly detection is triggered. The detection of anomalies is performed by a combination of the unsupervised learning clustering algorithm k-means and the supervised learning classifier KNN (k- nearest neighbors). Finally, a monitoring system consisting of an ELK stack collects historical data for further evolution of the ML algorithms. (Jose et al., 2021)

**keywords:** Anomaly Detection, Machine Learning, Big Data, NetFlow, Data Streams

**Abstrakt:** Vďaka pokrokom v oblasti strojového učenia a aplikovanej umelej inteligencie prešli systémy detekcie sieťových anomálií vývojom od tradičných signatúrových metód detekcie narušení k pokročilejším prístupom. Napriek tomu, ako sa bezpečnostné opatrenia zdokonaľujú, útočníci neustále vyvíjajú sofistikovanejšie techniky útokov. Preto je nevyhnutné nielen nasadiť robustný algoritmus detekcie anomálií, ale aj vytvoriť mechanizmus prenosu dát v reálnom čase, ktorý minimalizuje oneskorenia reakcie. Navyše, DDoS útoky môžu zahltiť dátové kanály siete tisíckami paketov za sekundu, čo vedie k preťaženiu väčšiny tradičných monitorovacích systémov, ktoré sa spoliehajú na ukladanie dát. Z tohto dôvodu sa výskum prezentovaný v tomto článku zameriava na implementáciu streamovacieho systému v reálnom čase pre detekciu sieťových anomálií, schopného fungovať aj pri vysokom objeme dátovej prevádzky. Navrhované riešenie zahŕňa nasadenie flow collector platformy prepojenej s Apache Kafka na príjem NetFlow dát zo sieťových prepínačov. Následne sa aplikujú techniky spracovania big data v reálnom čase prostredníctvom Apache Spark, kde sa spúšťa ML detekcia anomálií. Detekcia je realizovaná kombináciou neučenej (unsupervised) metódy klastrovania k-means a učeného (supervised) klasifikátora KNN (k-nearest neighbors). Nakoniec je súčasťou riešenia aj monitorovací systém na báze ELK stacku, ktorý zhromažďuje historické dáta na ďalší rozvoj ML algoritmov. (Jose et al., 2021)

## 2.12. Anomaly Detection in POSTFIX mail log using Principal Component Analysis

**Abstract:** Anomalies in a dataset are the abnormal observations of the data, and they often represent the problems or defectives coming from some changes in the system operation. Detecting the anomalies in a data set is the first step from which further analysis can be done to locate the actual root cause in the data. In this paper, we propose a general method to pinpoint the anomalies from a set of Postfix mail log. This method utilizes the Elastic Stack to collect and process log data, then applies Principal Component Analysis to separate the abnormal data from the normal one. To demonstrate the feasibility of the solution, we test it against a collection of a million Postfix log messages. The result is promising under

experimentation as it can detect an anomaly at the time a spam mail attack occurs. (Tran & Tran, 2018)

**keywords:** anomaly, elasticsearch, elastic stack, postfix, kibana, logstash, watcher

**Abstrakt:** Anomálie v dátovom súbore predstavujú abnormálne pozorovania dát, ktoré často signalizujú problémy alebo chyby spôsobené zmenami v prevádzke systému. Detekcia anomálií v dátovom súbore je prvým krokom, na ktorý nadväzuje ďalšia analýza s cieľom určiť skutočnú príčinu v dátach. V tomto článku je navrhnutá všeobecná metóda na identifikáciu anomálií zo súboru logov Postfix mail servera. Navrhovaná metóda využíva Elastic Stack na zber a spracovanie logovacích dát a následne aplikuje analýzu hlavných komponentov (PCA – Principal Component Analysis) na oddelenie abnormálnych dát od normálnych. Na overenie uskutočniteľnosti riešenia sa otestoval na súbore obsahujúcom milión logovacích správ Postfixu. Výsledky experimentov sú sľubné, keďže metóda dokáže detegovať anomáliu v čase, keď dôjde k útoku prostredníctvom spamových e-mailov. (Tran & Tran, 2018)

## 2.13. Cyberattacks Detection and Analysis in a Network Log System Using XGBoost with ELK Stack

**Abstract:** The usage of artificial intelligence and machine learning methods on cyberattacks increasing significantly recently. For the defense method of cyberattacks, it is possible to detect and identify the attack event by observing the log data and analyzing whether it has abnormal behavior or not. This paper implemented the ELK Stack network log system (NetFlow Log) to visually analyze log data and present several network attack behavior characteristics for further analysis. Additionally, this system evaluated the extreme gradient enhancement (XGBoost), Recurrent Neural Network (RNN), and Deep Neural Network (DNN) model for machine learning methods. Keras was used as a deep learning framework for building a model to detect the attack event. From the experiments, it can be confirmed that the XGBoost model has an accuracy rate of 96.01% for potential threats. The full attack data set can achieve 96.26% accuracy, which is better than RNN and DNN models. (Yang et al., 2021)

**keywords:** Cyber Security , Machine Learning , ELK Stack , XGBoost , NetFlow Log

**Abstrakt:** Používanie umelej inteligencie (AI) a metód strojového učenia (ML) pri kybernetických útokoch v poslednom období významne narastá. Z hľadiska obranných metód proti týmto útokom je možné detegovať a identifikovať útok prostredníctvom pozorovania logovacích dát a analýzy, či obsahujú abnormálne správanie. V tomto článku bol implementovaný systém ELK Stack pre sieťové logy (NetFlow Log) na vizuálnu analýzu logovacích dát a prezentáciu vybraných charakteristík správania sieťových útokov pre ďalšiu analýzu. Okrem toho systém vyhodnotil metódy strojového učenia: Extreme Gradient Boosting (XGBoost), Recurrent Neural Network (RNN) a Deep Neural Network (DNN). Na vytvorenie modelu pre detekciu útokov bol použitý framework Keras pre hlboké učenie. Výsledky experimentov potvrdzujú, že model XGBoost dosahuje presnosť 96,01 % pri identifikácii potenciálnych hrozieb. Pri práci s úplnou množinou dát útokov je dosiahnutá presnosť 96,26 %, čo je lepší výsledok v porovnaní s modelmi RNN a DNN. (Yang et al., 2021)

## 2.14. Improving State-of-the-art Compression Techniques for Log Management Tools

**Abstract:** Log data records important runtime information about the running of a software system for different purposes including performance assurance, capacity planning, and anomaly detection. Log management tools such as ELK Stack and Splunk are widely adopted to manage and leverage log data in order to assist DevOps in real-time log analytics and decision making. To enable fast queries and to save storage space, such tools split log data into small blocks (e.g., 16KB), then index and compress each block separately. Previous log compression studies focus on improving the compression of either large-sized log files or log streams, without considering improving the compression of small log blocks (the actual compression need by modern log management tools). The evaluation of four state-of-the-art compression approaches (e.g., Logzip, a variation of Logzip by pre-extracting log templates named Logzip-E, LogArchive and Cowic) indicates that these approaches do not perform well on small log blocks. In fact, the compressed blocks that are preprocessed using Logzip, Logzip-E, LogArchive or Cowic are even larger (on median 1.3 times, 1.5 times, 0.2 times or 6.6 times) than the compressed blocks without any preprocessing. Hence, we propose an approach named LogBlock to preprocess small log blocks before compressing them with a general compressor such as gzip, deflate and lz4, which are widely adopted by log management tools. LogBlock reduces the repetitiveness of logs by preprocessing the log headers and rearranging the log content leading to an improved compression ratio for a log file. Our evaluation on 16 log files shows that, for 16KB to 128KB block sizes, the compressed blocks by LogBlock are on median 5% to 21% smaller than the same compressed blocks without preprocessing (outperforming the state-of-the-art compression approaches). LogBlock achieves both a higher compression ratio (a median of 1.7 to 8.4 times, 1.9 to 10.0 times, 1.3 to 1.9 times and 6.2 to 11.4 times) and a faster compression speed (a median of 30.8 to 49.7 times, 42.6 to 53.8 times, 4.5 to 6.0 times and 2.5 to 4.0 times) than Logzip, Logzip-E, LogArchive and Cowic. LogBlock can help improve the storage efficiency of log management tools. (Yao et al., 2021)

**keywords:** Software log compression, Software logging, Log management tools

**Abstrakt:** Logovacie dáta zaznamenávajú dôležité informácie o behu softvérového systému, ktoré sa využívajú na rôzne účely, vrátane zabezpečenia výkonu, plánovania kapacity a detekcie anomálií. Nástroje na správu logov, ako ELK Stack a Splunk, sú široko používané na správu a využívanie logovacích dát s cieľom podporiť DevOps pri analýze logov v reálnom čase a pri rozhodovaní. Aby bolo možné vykonávať rýchle dopyty a šetriť úložný priestor, takéto nástroje rozdeľujú logovacie dáta na malé bloky (napr. 16 KB), ktoré následne indexujú a komprimujú samostatne. Predchádzajúce štúdie v oblasti kompresie logov sa zameriavali na zlepšenie kompresie buď veľkých logovacích súborov, alebo logovacích tokov, pričom nebrali do úvahy kompresiu malých logovacích blokov, čo je pritom skutočná potreba moderných nástrojov na správu logov. Vyhodnotenie štyroch najmodernejších prístupov ku kompresii, Logzip, jeho variácie Logzip-E (s predextrahovanými šablónami logov), LogArchive a Cowic, ukazuje, že tieto prístupy nedosahujú dobré výsledky pri malých logovacích blokoch. V skutočnosti sa ukazuje, že bloky spracované pomocou týchto prístupov sú po kompresii dokonca väčšie (v mediáne 1,3×, 1,5×, 0,2× alebo 6,6×) než bloky komprimované bez akejkoľvek predspracovateľskej úpravy. Preto je navrhutý nový prístup s názvom LogBlock, ktorý umožňuje predspracovanie malých logovacích blokov pred ich kompresiou všeobecnými kompresormi, ako sú gzip, deflate alebo lz4, ktoré sú široko používané v nástrojoch na správu logov. LogBlock znižuje redundanciu logov predspracovaním hlavičiek logov a preusporiadaním obsahu, čo vedie k lepšiemu kompresnému pomeru pri logovacích súboroch. Evaluácia na 16 logovacích súboroch ukazuje, že pri veľkosti blokov 16 KB až 128 KB sú bloky

komprimované s LogBlock-om v mediáne o 5 % až 21 % menšie ako tie isté bloky komprimované bez predspracovania (a zároveň prekonávajú existujúce prístupy). LogBlock dosahuje nielen vyšší kompresný pomer (v mediáne 1,7–8,4×, 1,9–10,0×, 1,3–1,9× a 6,2–11,4×) ale aj vyššiu rýchlosť kompresie (v mediáne 30,8–49,7×, 42,6–53,8×, 4,5–6,0× a 2,5–4,0×) než Logzip, Logzip-E, LogArchive a Cowic. LogBlock tak predstavuje efektívny spôsob, ako zlepšiť úložnú efektivitu nástrojov na správu logov. (Yao et al., 2021)

## 2.15. SDGen: A Scalable, Reproducible and Flexible Approach to Generate Real World Cyber Security Datasets

**Abstract:** Real world cyber security datasets are essential for developing and evaluating new techniques to counter cyber attacks. Ideally, these datasets should represent modern network infrastructures with upto-date cyber attacks. However, existing datasets commonly used by researchers are either synthetic, unscalable or easily outdated due to the dynamic network infrastructure and evolving nature of cyber attacks. In this paper, we introduce a security dataset generator (SDGen) which focuses on a scalable, reproducible and flexible approach to generate real world datasets for detection and response against cyber attacks. We implement SDGen within a virtual environment using DetectionLab, ELK (Elasticsearch, Logstash, Kibana) stack with Beats and AttackIQ (a security control validation platform). This implementation in fact provides a proof-of-concept (POC) of SDGen to demonstrate the dataset generation of an organisation being compromised by several types of Ransomware. We showcase that our proposed dataset generator, SDGen, provides scalability, reproducibility and flexibility in generating cyber security datasets by modifying the configurations in DetectionLab, Vagrant-Files and launching different types of attacks in AttackIQ. (Abigail et al., 2022)

**keywords:** Dataset generation; MITRE ATT&CK framework; Application security

**Abstrakt:** Reálne kyberbezpečnostné dátové súbory sú nevyhnutné pre vývoj a hodnotenie nových techník na boj proti kybernetickým útokom. Ideálne by mali tieto dátové súbory reprezentovať moderné sieťové infraštruktúry s aktuálnymi typmi útokov. Existujúce dátové súbory, ktoré výskumníci bežne používajú, sú však často syntetické, ťažko škálovateľné alebo rýchlo zastarávajú v dôsledku dynamiky sieťovej infraštruktúry a neustále sa vyvíjajúcej povahy kybernetických útokov. V tomto článku je predstavený Security Dataset Generator (SDGen), ktorý sa zameriava na škálovateľný, reprodukovateľný a flexibilný prístup k tvorbe reálnych dátových súborov určených na detekciu a reakciu proti kybernetickým útokom. SDGen je implementovaný vo virtuálnom prostredí pomocou DetectionLab, ELK stacku (Elasticsearch, Logstash, Kibana) s Beats a platformy AttackIQ (platforma na validáciu bezpečnostných kontrol). Táto implementácia predstavuje proof-of-concept (POC) nástroja SDGen, ktorý demonštruje generovanie datasetu organizácie kompromitovanej rôznymi typmi ransomvéru. Ukázané je, že navrhnutý generátor datasetov SDGen prináša škálovateľnosť, reprodukovateľnosť a flexibilitu pri tvorbe kyberbezpečnostných dátových súborov úpravou konfigurácií v DetectionLab, Vagrant-Files a spúšťaním rôznych typov útokov v AttackIQ. (Abigail et al., 2022)

# ZOZNAM ZDROJOV

[1] Abigail, Xie, M., Ryan, Sterner, C., Choi, T., & Dong, N. (2022). SDGen: A Scalable, Reproducible and Flexible Approach to Generate Real World Cyber Security Datasets. *Communications in Computer and Information Science*, 102–115. https://doi.org/10.1007/978-981-19-0468-4_8.

[2] Andrade, R., & Torres, J. (2018). *Enhancing intelligence SOC with big data tools*. https://doi.org/10.1109/iemcon.2018.8614779.

[3] Currie, R., & Yuan, W. (2024). Building a Flexible and Resource-Light Monitoring Platform for a WLCG-Tier2. *EPJ Web of Conferences*, 295, 07008. https://doi.org/10.1051/epjconf/202429507008.

[4] Folino, G., Otranto Godano, C., & Pisani, F. S. (2023). An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity. *The Journal of Supercomputing.* https://doi.org/10.1007/s11227-023-05049-x.

[5] Gautam, B. P., & Norio, S. (2020). SUESSA: Sustainable & Ultra-Elastic Stack Security Architecture for Securing IoT Networks of Future Smart Cities. *2020 Eighth International Symposium on Computing and Networking Workshops (CANDARW)*, 387–390. https://doi.org/10.1109/candarw51189.2020.00079.

[6] Halvorsen, J., Waite, J., & Hahn, A. (2019). Evaluating the Observability of Network Security Monitoring Strategies With TOMATO. *IEEE Access*, *7*, 108304–108315. https://doi.org/10.1109/ACCESS.2019.2933415.

[7] Jose, J., Talha Ahmed Khan, Akbar, W., Muhammad Afaq, & Song, W.-C. (2021). An ML Based Anomaly Detection System in real-time data streams. *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. https://doi.org/10.1109/csci54926.2021.00270.

[8] Kulshrestha, U., & Durbha, S. (2020, September 1). *Edge Analytics and Complex Event Processing for Real Time Air Pollution Monitoring and Control*. IEEE Xplore. https://doi.org/10.1109/IGARSS39084.2020.9323584.

[9] Kyriakou, A., & Sklavos, N. (2018, October 1). *Container-Based Honeypot Deployment for the Analysis of Malicious Activity*. IEEE Xplore. https://doi.org/10.1109/GIIS.2018.8635778.

[10] Mfula, H., & Nurminen, J. K. (2018). Self-Healing Cloud Services in Private Multi-Clouds. *2018 International Conference on High Performance Computing & Simulation (HPCS)*. https://doi.org/10.1109/hpcs.2018.00041.

[11] Poat, M. D., Lauret, J., & Fedele, D. (2023). Flexible visualization of a 3rd party Intrusion Prevention (Security) tool: A use case with the ELK stack. *Journal of Physics. Conference Series*, *2438*(1), 012040–012040. https://doi.org/10.1088/1742-6596/2438/1/012040.

[12] Tran, C.-P., & Tran, D.-K. (2018). Anomaly Detection in POSTFIX mail log using Principal Component Analysis. *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, 107–112. https://doi.org/10.1109/kse.2018.8573410.

[13] Ummay Faseeha, Syed, H. J., Samad, F., Sehar Zehra, & Ahmed, H. (2025). Observability in Microservices: An In-Depth Exploration of Frameworks, Challenges, and Deployment Paradigms. *IEEE Access*, 1–1. https://doi.org/10.1109/access.2025.3562125.

[14] Yang, C.-T., Chan, Y.-W., Liu, J.-C. L., Endah Kristiani, & Lai, C.-H. (2021, August 23). *Cyberattacks Detection and Analysis in a Network Log System Using XGBoost with ELK Stack*. https://doi.org/10.21203/rs.3.rs-838650/v1.

[15] Yao, K., Sayagh, M., Shang, W., & Hassan, A. E. (2021). Improving State-of-the-art Compression Techniques for Log Management Tools. *IEEE Transactions on Software Engineering*, 1–1. https://doi.org/10.1109/tse.2021.3069958.