Monitoring, Management, and Analysis of Security Aspects of IaaS Environments

Andrzej Mycek

Cracow University of Technology, Cracow, Poland

https://doi.org/10.26636/jtit.2023.4.1419

Abstract — Many companies or institutions either already have placed their resources in or plan to move them to the cloud. They do so for security reasons and are weary of the fact that by relying on cloud-based resources, they do not have to bear such extensive infrastructure-related costs. However, continuous technology advancement results not only in benefits, but also in disadvantages. The latter include the growing risk associated with IT security, forcing the individual actors to implement monitoring measures and to respond to numerous threats.
This work focuses on creating a small infrastructure setup using the publicly available Google Cloud Platform which, thanks to the monitoring systems implemented thereon, allows to rapidly respond to hardware and software faults, including those caused by external factors, such as attacks on specific components. This project may also be customized to satisfy individual needs, depending on the cloud service provider selected. The work uses public cloud provider tools as well as open-source systems available for everyone, both in the cloud and in the on-prem environment. The paper deals also with the concept of a proprietary intrusion detection system.

Keywords — cybersecurity, IaaS, monitoring, Wazuh, Zabbix

1. Introduction

Cloud computing is currently one of the most popular topics in modern computer science. Cloud services provide users with access, on an on-demand basis, to such resources as databases, networks, and software. Cloud computing is primarily characterized by its flexibility and scalability, pay-as-you-go payments, fully virtualized resources, as well as a high degree of availability and reliability.
The flexibility and scalability allow users to adjust the size of their resources in the cloud. Depending on their needs, a customer can either increase or decrease the amount of cloud resources used. The cloud environment is also distinguished by the pay-as-you-go business model. Cloud users are billed only for the resources they use, and they can roughly estimate their potential costs using tools offered by the service providers. As cloud providers ensure resource redundancy and distribution, thus ensuring high service availability, public clouds are considered the safest data storage method, due to their high availability and reliability [1].
It is worth mentioning that public clouds, such as AWS, Azure, or GCP, are not the only types of clouds available in the market. Besides publicly available clouds, other models, such as private cloud, hybrid cloud, and the so-called multicloud, are on offer too. In the case of a private cloud, we are the owners of the hardware, unlike in the case of a public cloud. This means, however, that we are responsible for the hardware data center. The hybrid cloud approach is a compromise between the public and private models. It combines a public cloud with a private system. Multicloud is the last model considered. It is a combination of multiple public clouds from different providers, joined to optimize costs [2]. The following cloud service models are available (Fig. 1):

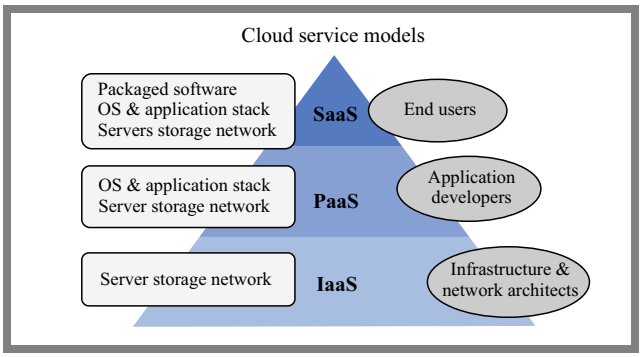

Fig. 1. SaaS vs. PaaS vs. IaaS [3].

- Infrastructure as a Service (IaaS) – a model that involves providing the end user with specific infrastructure (such as hardware, software) available via virtual machines,
- Platform as a Service (PaaS) – involves delivering, to the customer, applications which are hosted on the cloud service provider's servers,
- Software as a Service (SaaS) – in this model, the customer receives a specific solution from the provider, in the form of defined functionality of specific software,
- Function as a Service (FaaS) – here, the provider allows the customer to execute code in the form of functions.

Cloud computing began to gain popularity a decade ago, but the market's interest was boosted significantly when cloud services became available in America via Amazon which offered its Amazon Web Services (AWS) solution. Later, other giants, such as Microsoft (Azure cloud) and Google (Google Cloud Platform – GCP) joined the market [4]. Cost effectiveness is an extremely important aspect that speaks in favor of the cloud approach. Cloud computing is currently relied upon in various areas, including web hosting, data processing, file storage, application development, and artificial intelligence. Additionally, the pandemic has increased the demand for cloud-based digital services. The significantly increased

popularity of remote work, digital transformation-related efforts undertaken by most companies, and the rapid growth of the healthcare sector have made cloud computing the most popular IT topic alongside AI. According to [5], every country is investing in cloud computing, and statistics show that 37% of healthcare service providers consider cloud adoption as a strategic element. 22% are in the planning stages, while 25% are in the midst of the execution phase, and expect their efforts to significantly drive their respective industries. Moreover, the 5% of companies that have already implemented cloud computing in their operations estimated that savings generated after migrating to the cloud amount to approx. 20%. This demonstrates how significant the topic of cloud computing is. Based on the above information, it is evident that this is one of the most crucial directions in the modern IT industry.

The rest of the paper is structured as follows. Section 2 reviews the topic of cloud infrastructure with an emphasis on security aspects, while Section 3 describes the most important tools used in the work. Section 4 portrays the process of deploying cloud infrastructure while relying on the best security practices. It also describes the approach developed by the author. Section 5 discusses the results achieved, while Section 6 presents the conclusions and future work.

2. Monitoring and Security of the Cloud Environment

Monitoring the systems and ensuring their proper security are highly important topics nowadays. These tasks are carried out to detect and prevent any security-related incidents. Continuous monitoring of IT infrastructures enables early detection of cyberattacks, data breaches, and other security threats. Regular monitoring allows for a swift response and the implementation of suitable measures at an early stage, to minimize potential damage. Monitoring plays a crucial role in safeguarding the confidentiality and integrity of data. In today's world, data holds immense importance, serving as an invaluable asset for both businesses and individuals alike. Monitoring allows to shield this data from unauthorized access, theft, loss, or destruction. Actions such as monitoring network traffic, managing access rights, and encrypting data aid in maintaining confidentiality and integrity of information.

Continuous monitoring significantly facilitates maintaining high availability of the system by promptly detecting operational malfunctions and allows to quickly respond to minimize system and service downtimes. High availability is particularly crucial for companies that need to ensure the continuity of their services. Monitoring also holds paramount importance in terms of potential growth planning, as it allows for the collection of data regarding the performance of systems, networks, and applications. This information can be utilized for performance analysis and optimization, identifying areas for improvement, and planning the development of IT infrastructure according to current and future needs.

3. Description of the Cloud Infrastructure

To explain security-related topics, the basic terms and commercially available services need to be described first.

Google Cloud Platform (GCP) – is a platform offered by Google, covering data processing, storage, analytics, machine learning, artificial intelligence, networking, security and many more functions. It offers flexibility and scalability, allowing users to adapt resources to their needs, manage load, and pay only for those resources that are actually used. GCP offers access to over 170 services, tools, and components that can be added or removed to he designed infrastructure – depending on your current and future needs. The services are available on GCP support, inter alia, easy launching of applications, regardless of the technologies relied upon to create them, implementation of new and modification of existing functionalities without the need for major changes on the server side, protection against attacks from the outside, identification of threats inside the organization as well as creation and management of networks, with a particular emphasis on security, availability, and performance.

Virtual Private Cloud (VPC) is a service provided by various cloud platforms, including GCP. VPC allows for the creation of a private network in the cloud, with own subnets, to manage IP addressing and define security policies. The supervisor can control which resources have access to the network, as well as what types of connections are allowed. VPC also enables routing of network traffic between different subnets and between the cloud and the LAN network.

Compute Engine is a service provided by GCP, enabling users to launch and manage virtual machines in the cloud. Compute Engine facilitates scalability of virtual machines to match application requirements. It provides a wide range of virtual machine options with various hardware configurations, including processor core count, RAM, and disk capacity. It ensures a secure environment, such as virtual machine isolation, access management capabilities, and network firewalls [6].

Bastion host is a specially configured server within a network that serves as a secure entry point to other machines or networks within the IT infrastructure. The use of a bastion host significantly reduces the risk of attacks from the inside, as it enables organizations to restrict access to their network resources.

Zabbix is an open-source tool used for monitoring IT systems, networks, and applications. With Zabbix, organizations can effectively monitor and analyze their resources, enabling swift identification and resolution of issues related to system's performance, availability, and reliability. Zabbix offers various functionalities, including monitoring and data collection, configurable alerts, real-time graphs, advanced visualization options, network maps, reports, custom graphs consolidating multiple elements into a single view, data archiving, and historical storage, automatic network device discovery, agent auto-registration, and detection of file systems, network interfaces, and SNMP identifiers [7].

Wazuh is a tool designed for collecting, aggregating, indexing, and analyzing security data. By monitoring logs, system

events, network traffic, and other data, Wazuh enables organizations to swiftly respond to intrusion attempts, attacks, and other potential threats.

Cloud IDS is a service provided by GCP, aimed at detecting and monitoring unauthorized activity and threats within the cloud environment.

Every major IT company in the cloud computing market has its own IDS, i.e., Google Cloud IDS or Microsoft Sentinel, available in the Microsoft Azure cloud environment – a solution that is gaining significant popularity. Microsoft Sentinel is a solution that provides security information and event management (SIEM), as well as security orchestration, automation, and response (SOAR) mechanisms. The fundamental difference between a SIEM system and IDS is that SIEM tools take proactive measures, whereas an IDS system focuses on detecting and reporting identified events. Thanks to Microsoft Sentinel, we have access – while working on this paper – to a comprehensive solution that allows us to detect attacks and respond to threats. It also has a rich database that enables an in-depth analysis of all past events. The ability to collect vast amounts of data from all applications and devices, infrastructure components, and additionally from various clouds or on-premises environments is another advantage of this particular solution [8].

However, just like any other tool, Microsoft Sentinel comes with its own drawbacks. These include, for example, relatively complex implementation. In the IT world, it is simply not possible to be an expert in everything. Considering that every major cloud provider has its own tools, organizations require engineers specializing in various technologies. This significantly extends the deployment time of the solution within the organization and generates additional costs in terms of recruiting IT engineers, hiring them, and integrating them into their specific tasks. Another issue is adapting the platform to case-specific needs. Since each provider has its own solution, when migrating from one cloud provider to another, user-specific implementations become useless and need to be rebuilt. In the case of the aforementioned Sentinel tool, for instance, we must be aware that it is dependent on the Microsoft ecosystem and is offered as a commercial license. As far as having one's own cloud infrastructure is concerned, data storage costs and licensing fees are an important factor. Therefore, in this paper, the concept of a system providing significantly lower costs is presented, as it essentially relies solely on free software tools, nevertheless satisfying the majority of needs.

Legal considerations are also important. For example, many institutions require their data to be stored only in data centers located within the European Union. In the case of such tools, there is often uncertainty, as tech giants may not provide clear information on this matter. Consequently, at the stage of choosing a cloud provider, the location of data storage may lead to rejecting a specific provider.

Another SIEM tool offering greater flexibility is the IBM Security QRadar solution. This software can be deployed in several ways: as a software, hardware, or virtual solution. The architecture of IBM QRadar includes flow processors tasked with collecting data from the fourth layer of the OSI model, QFlow processors for conducting deep inspection and analysis of packets in the seventh layer of the OSI model, i.e., the application layer, and a centralized console that is highly useful for security operations center (SOC) personnel. Thanks to the QRadar security intelligence module, data is, of course, collected in real-time. Another interesting and highly useful functionality is the IBM QRadar vulnerability manager module which serves as a vulnerability scanner and utilizes additional data from other scanners, such as Rapid7 or Nessus. However, the drawbacks of this solution include product updates, hardware requirements, and costs. In the case of QRadar, licensing costs and the resources required for deployment and maintenance need to be taken into consideration as well [9].

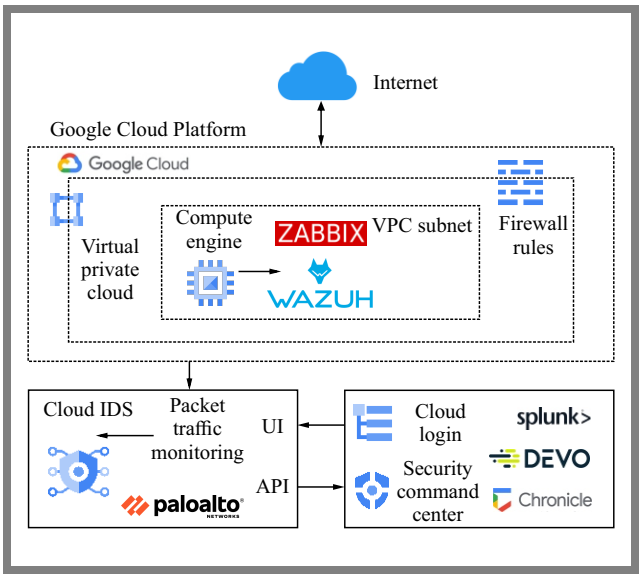

Fig. 2. Environment architecture scheme.

4. Creating Infrastructure in the Cloud

At the infrastructure creation stage, planning is paramount to cope, from the outset, with specific issues, primarily related to costs or the various infrastructure development methods. Currently, Infrastructure as Code (IaC) is a highly popular approach. IaC makes it possible to manage the entire infrastructure, e.g., virtual machines, load balancers, virtual networks, security policies, and other services without any manual work performed to create a specific environment, as the entire infrastructure is stored in the repository, in the form of code. IaC also reduces the possibility of making mistakes that are unavoidable in the case of infrastructures created manually creation [10].

Therefore, an in-depth business analysis is the first activity that should be undertaken. At this stage, it is necessary to analyze the organization's needs and choose optimized services that will meet the specific business goals. Implementation and maintenance costs are another important aspect, especially in the context of cloud computing. It is important to keep in mind that in the cloud, we rely on the pay-as-you-go model [1]. Not all services may benefit the organization when the need-to-cost ratio is taken into consideration, and some

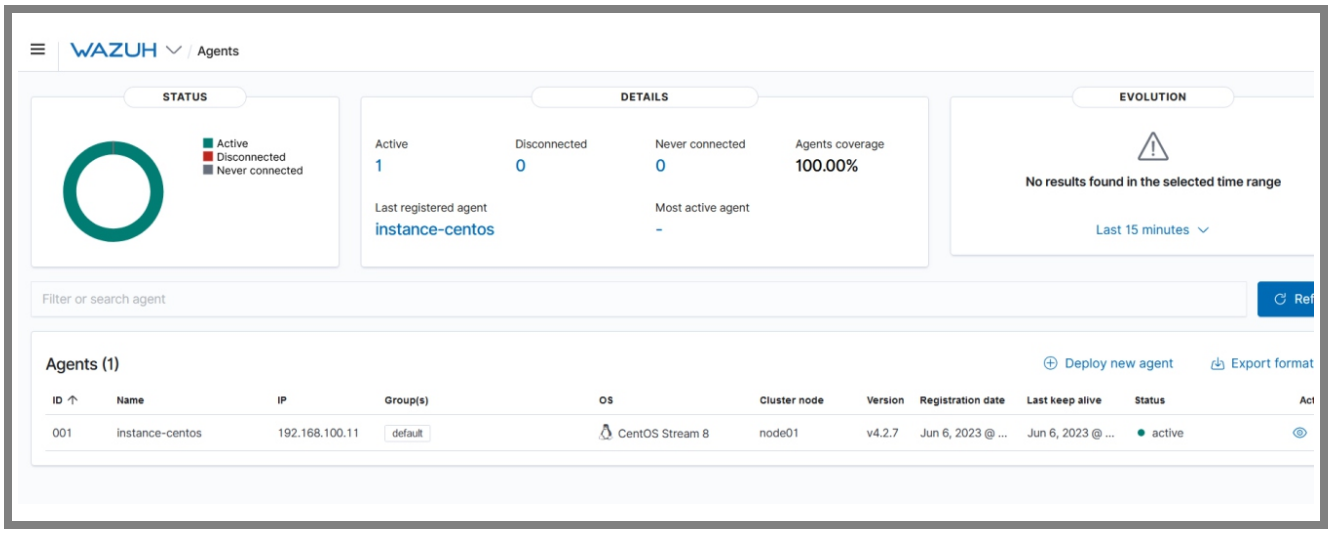

Fig. 3. Adding a virtual machine to Wazuh.

services hosted in an in-house data center might be more cost-effective. Therefore, we often encounter the so-called hybrid cloud model, i.e., a combination of a public and a private cloud system. Depending on the needs or requirements related to data storage, services relying on both public and private clouds are taken advantage of.

Security is another important topic that is, unfortunately, underestimated on many occasions. It is necessary to review the tools available in the cloud and ensure appropriate security of data and applications stored in the cloud, taking into account authentication, access control, and encryption mechanisms. Monitoring is essential and should not be overlooked (this includes monitoring of the performance of applications and resources, as well as security monitoring). Additionally, factors like redundancy required to ensure high availability of services, compliance with legal regulations, developing contingency plans and offering staff training, should not be disregarded as well. In this work, all of these issues have been addressed and implemented. For the purposes of this project, the infrastructure has been created using GCP. The core infrastructure was built using the HashiCorp's Terraform tool, while relying on the IaC approach. Terraform allows the user to create the cloud infrastructure and resources using the HashiCorp configuration language (HCL). HCL is a declarative and high-level language, where each code block defines a resource. Syntax readability and ability to interact with other tools are the primary characteristics of HCL. A declarative language (unlike the imperative variety) describes what you want to achieve, and not how you intend to do that.

The project was created in such a way that it can be adapted, without complications, while switching to another public cloud provider (e.g., AWS or Microsoft Azure), or while relying on a private cloud e.g., OpenStack. The process of creating resources with Terraform consists of three steps: writing the resources to be created in HCL, generating a plan, and running the tool. The first stage involves gathering requirements and writing the source code. Then, the cloud provider and the services we want the infrastructure to address need to be defined. The definitions of the provider and the resources we intend to create are stored in a .tf file. The next step is to generate a plan. At this stage, Terraform searches local directories for configuration files, checks the current state and displays the changes that will be made. The final step is to apply the changes. Before Terraform takes action, it asks for confirmation of the plan, and then it proceeds with the action. IaC is currently the main approach in the IT market when it comes to automated deployment of environments. It allows for easy replication of the entire infrastructure in different conditions and environments, also enabling quick implementation of any changes, version control, and ensuring that the infrastructure may be easily scaled depending on current needs. [10].

In this project, after writing the infrastructure in the HCL language and using the Terraform tool, the infrastructure was created in seconds. In contrast, creating the entire core infrastructure manually would have taken from minutes to hours. In addition, once the infrastructure is written, it can be duplicated and adapted to specific needs with almost no effort. In the case of a failure, the supervisor can easily restore and recreate the infrastructure, making sure that it is identical to that used in the initial production deployment [10].

As part of the project, both security and monitoring solutions are provided by the vendor (in our case, Google Cloud IDS). Those that can be installed in our own environment, such as Zabbix and Wazuh, are available as well (Fig. 2).

The first step in the cloud infrastructure implementation phase is to establish the permission structure. The management of users in GCP is significantly more similar to the approach adopted in Microsoft Azure than in Amazon Web Services. GCP employs the model of the so-called federated accounts (Google accounts). In the case of AWS, a special user is created. The absence of a master account, such as Subscription, is the factor differentiating this solution from Azure. In GCP, we have two main types of objects: organizations and projects. Organizations are associated with settings that pertain to the entire organization, such as GDPR compliance or global secu-

rity policies. In GCP, an organization is automatically created for each user. It manages multiple projects and the relationship itself is referred to as project ownership [11]. In the case of projects, user roles are assigned to them (e.g., billing operator, administrator, etc.). After identifying (authentication) a specific user using their cloud identity, the next step is to determine what may be done in Google Cloud (authorization). Control of access to resources is precisely managed, in GCP, through cloud identity and access management (IAM) policies. The permission rules that control access to Google Cloud resources have the form of sets of IAM bindings. IAM roles group interrelated granular permissions.

There are three main types of roles: standard, custom, and predefined. When using IAM, IAM policies are mapped to service identities using groups. In accordance with the best cloud security practices from GCP, the recommendation is to use groups, minimize points where IAM policies are applied using folders, and optionally enforce domain membership using organizational policies, such as *iam.allowedPolicyMemberDomains*. Once the "core security infrastructure" has been planned, the next step is to implement the environment using the IaC approach with Terraform. After creating the security policies, the portion responsible for IaaS is created. The VPC network and subnets were created in the Europe-West 3 (Frankfurt, Main) region, and the appropriate inbound and outbound rules were added to the firewall, including these allowing outbound traffic from the network and incoming traffic using port 22 of the TCP protocol (SSH). Additionally, a rule was added, inter alia, permitting outbound traffic via port 10050 TCP for Zabbix agents. This was necessary because, for monitoring purposes, each endpoint had a Zabbix agent installed and configured.

Then, the Zabbix (ver. 6.0 LTS) monitoring tool, which can be installed not only in the cloud but also in a local environment, was installed on a virtual machine with the same parameters as Wazuh.

During the implementation of Zabbix, a LAMP stack (Linux, Apache, MySQL, PHP) was utilized to install the Zabbix server, front-end, agent, and MySQL server. After adding several instances, including a virtual machine with Wazuh, a comprehensive configuration of Zabbix was performed. Templates with custom items were created to monitor CPU, RAM, and disk usages, and overall host availability. Triggers were also configured to receive alerts when resources fell below specified thresholds.

The second important instance, which can be hosted in both public and private clouds, was a virtual machine with Wazuh. Wazuh is a security information and event management (SIEM) system that integrates extended detection and response (XDR) capabilities. The purpose of XDR is to collect and correlate, in real time, data from the installed agents (Fig. 3). SIEM is tasked with monitoring, detecting, and alerting security-related events. It helps identify potential security gaps before attackers can harm the organization. SIEM detects anomalies using artificial intelligence. The basic architecture of SIEM is depicted in Fig. 4. Figure 5 illustrates the SIEM system, where additional rules have been created, including those defending against brute-force attacks. Since the term IaaS refers to cloud environments, it is also important to implement the VPC flow logs service. This tool allows to monitor network packet traffic within a VPC in GCP. VPC Flow Logs is a solution similar to Azure Network Watcher. It records such information as source and destination IP addresses, ports, protocols, session duration, and more. This allows to analyze network traffic, conduct audits, and detect network anomalies. After configuring VPC Flow Logs, network traffic data can be forwarded to tools like Cloud Logging, Cloud Monitoring, BigQuery, or other data analytics software. Additionally, cloud provider tools were used in the project, but the goal was to create an infrastructure that could also work in hybrid environments.

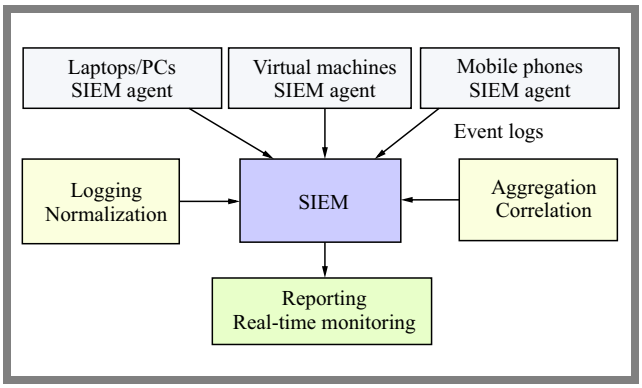

Fig. 4. SIEM architecture [12].

```
<command>
  <name>firewall-drop</name>
  <executable>firewall-drop</executable>
</command>

<active-response>
  <command>firewall-drop</command>
  <location>all</location>
  <rules_group>authentication_failed|authentication_failures</rules_group>
  <timeout>700</timeout>
  <repeated_offenders>30,60,120</repeated_offenders>
</active-response>
```

Fig. 5. Example of Wazuh rules.

At the heart of the project was Cloud IDS (CIDS) from Google. It is a native solution that detects network threats, such as command and control attacks, malware and spyware. CIDS is based on an engine of the renowned Palo Alto Networks cybersecurity company. Managing CIDS is very simple, and the system allows to control it via a graphical user interface (GUI) or a command line interface (CLI), as well as through application programming interfaces (APIs). Importantly, CIDS automatically scales up and down, so there is no need for an advanced design to increase performance. The tool not only monitors traffic to and from the Internet, but also verifies internal communication within and between the VPCs.

Due to the fact that cloud architecture is of the distributed variety, it is a target of attacks undertaken by multiple intruders. Clouds are therefore primarily vulnerable to attacks at the network layer. The most common of these include address resolution protocol (ARP) spoofing, DNS poisoning, port scanning, IP spoofing, man-in-the-middle attacks, as well

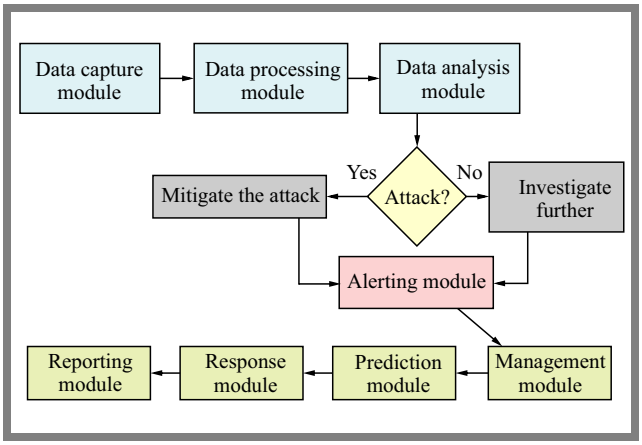

Fig. 6. Cloud IDS concept.

as denial of service (DoS) and distributed denial of service (DDoS) attacks [13].

A traditional firewall performs well against external network attacks, but for internal security incidents or external attacks, such as DoS or DDoS, IDS would prove to be a more effective solution [14]. In the context of cloud computing, the role of IDS is unquestionable. It is designed to serve as both an additional preventive layer and to effectively handle DoS/DDoS attacks and more sophisticated incidents [15]. An IDS system captures data and notifies about registered events. It is important to keep in mind that not all alerts generated by IDS are always related to intrusions. Often, they are the so-called false positives which are erroneously indicating that an event occurred when, in reality, it never did [16]. To respond promptly to network threats detected by CIDS, custom workflows have been created to take adequate, alert-based action.

The systems have been configured and integrated, including with the Splunk platform, so that data generated by Cloud IDS can be used for threat investigation and correlation within SIEM and for responding to threats using security orchestration, automation, and response (SOAR). Splunk is another important component of this chain. This tool allows to analyze logs and monitor the infrastructure. It serves as an SIEM system, similar to Microsoft Sentinel in Microsoft Azure cloud. Splunk also enables the processing of very large datasets, contributing to the frequent utilization of this tool in Big Data scenarios.

5. Intrusion Detection System Concept

Three types of cloud intrusion detection systems (CIDSs) differ in their approach to threat detection. The first one is a host-based intrusion detection system (HIDS). Here, tests are performed on the host computer to detect incidents. HIDSs operate on the premise that intruders always leave traces behind. In the past, HIDSs served as the first line of defense, but today, there is a shift towards choosing more advanced and comprehensive security systems.

Network-based IDS (NIDS) is the second approach which identifies any intrusions by monitoring and capturing all network traffic. The last solution – distributed intrusion detection system (DIDS) – combines the functionalities of both HIDS and NIDS.

As far as detection methods are concerned, three different techniques may be distinguished: signature-based detection, anomaly-based detection, and hybrid detection [17]. Signature-based detection involves comparing network communication with a predefined list of access controls. Packets are analyzed based on defined rules, and contextual analysis is often relied upon as well. Packets are tracked over a longer period of time. Additionally, in this method, a technique is sometimes employed where a deliberately crafted response is sent to the intruder to make them believe that they have not been detected.

The second method is anomaly-based detection. In this case, there are no predefined patterns. The operation of a NIDS that utilizes anomaly detection typically involves heuristic analysis and anomaly analysis. Heuristic analysis uses algorithms that define specific behaviors as anomalies. In the case of anomaly analysis, network traffic that deviates from the norm is detected. Hybrid detection is the last method. It combines anomaly-based and signature-based detection techniques.

Cloud IDS solutions that employ hybrid detection achieve significantly better results than other techniques [18]. Using machine learning (ML) algorithms allows for a significant improvement in cloud monitoring effectiveness, especially when it comes to detecting attacks and anomalies. ML algorithms can analyze logs and all network traffic to search for irregularities that may indicate potential attacks. Unfortunately, a solely ML-based approach can be vulnerable to data poisoning. Due to the distributed and open nature that characterizes cloud computing and processing it relies on, the cloud is unfortunately highly susceptible to external attacks. This is why it is crucial for every cloud service provider, including own solutions e.g., OpenStack, to offer a cloud-based IDS.

There are many IDS solutions available on the market that anyone can implement for free, using their own skills and expertise. One of the most well-known is undoubtedly Snort which, when combined with other open-source tools and with in-house developed approaches, may serve as an excellent alternative to proprietary solutions [19].

The cloud-based IDS proposed in this work consists of several modules tasked with data capture, data processing, analysis, alerting, management, prediction, response, and reporting (Fig. 6).

Snort allows for the integration of a preprocessor engine, enabling packet modification and preliminary packet analysis. This functionality may be utilized in our IDS concept, for example in data analysis, data capture or data processing modules. The detection engine used in Snort can also be harnesses as a data analysis module. The Wazuh tool, implemented in our infrastructure, may also serve – in conjunction with Snort – as a very effective foundation. Thanks to its log management feature, Wazuh is capable of analyzing and storing logs from multiple sources, similarly to the dedicated Microsoft Azure platform – the Sentinel. In the concept of our own cloud IDS, it can be utilized as a data capture module. Since Wazuh, an open-source XDR, combines SIEM and

IDS/IPS functionalities, it can seamlessly be adopted as a prediction module, a response module, or a reporting module. With its integrated IDS/IPS feature, Wazuh can effortlessly analyze, monitor, and respond to any incidents [12]. Utilized during the infrastructure creation phase, Splunk, owing to its excellent properties as a platform for visualization, analysis, and processing of data from various sources, can serve exceptionally well as a foundation for a management and analysis module. By incorporating other open-source tools and implementing custom solutions, we are essentially able to transform the cloud IDS provided by the cloud provider into our own solution. This adaptation is applicable not only in public cloud environments, but also in private, hybrid, or multi-cloud scenarios.

Data capture module. This module captures outgoing and incoming packets from various TCP/IP network protocols, such as TCP, IP, FTP, HTTP, SMTP, POP3, and IMAP. Other data, such as system logs and application data, are collected as well. The received data is placed in a queue, where it is subsequently processed by the data processing module.

Data processing module. This module processes the data collected at an earlier stage. It initially attempts to perform a selection stage to determine whether a given behavior is normal or constitutes a potential security incident, utilizing a signature-based detection model for this purpose. After normalizing the processed data, they are then sent to the next module of cloud IDS.

Data analysis module. This block is expected to thoroughly analyze the received data and identify potential threats. ML is employed at this stage to train a model based on historical data, where both normal network traffic and attack cases are well-known. Once our model is trained, it is used to analyze current data. Thanks to the model, any anomalies and suspicious network activities that may indicate attacks can be detected.

Alerting module. If our analytical model recognizes or deems a certain behavior to be an attack, it will trigger an alert by generating a suitable notification.

Management module. It allows to manage the IDS system (configuration, updates, data access). This module allows to control functionalities, configure settings, and access previously processed data. The management module facilitates log management, signature updates, and rule administration.

Prediction module. The prediction module is one of the most critical components of any modern IDS system. It leverages advanced predictive algorithms learned from the data analysis module to forecast potential incidents using such techniques as: long short-term memory, decision trees, linear regression, support vector machines. The appropriately trained models are then used in this module to predict potential events and incidents.

Response module. It oversees all previous operations in the cloud-based IDS. If a threat is classified and detected, this module takes actions to isolate and neutralize the threat. In the case of a threat, the response module blocks the attacker's actions.

Reporting module. It is the final component of the cloud-based IDS. This module collects all data regarding detected incidents, system performance, and overall operations. It enables the creation of advanced reports and analyses for the IT security team.

6. Results of the Work

The objective of the research was to create a flexible and scalable infrastructure in the public cloud and implement monitoring systems that are well-known in both "on-prem" environments and those created by cloud service providers and available exclusively in their computational cloud. An especially crucial aspect here is the approach to infrastructure creation itself. In this project, it was not developed using the traditional method, i.e., through the GUI. The infrastructure was built entirely using the IaC paradigm, with the help of the Terraform tool. As part of the work, a comprehensive plan was developed, utilizing the most popular and best-in-class solutions available in the market. The first step consisted in a thorough analysis of the requirements, followed by the practical implementation of the infrastructure in GCP using the IaC approach with Terraform. Zabbix and Wazuh monitoring systems were deployed and configured on two virtual machines, while Zabbix and Wazuh agents were installed on the remaining machines for monitoring purposes.

The implementation of monitoring mechanisms that would enable to track and analyze the performance of applications and resources was another important aspect of the project. A special template with suitable items responsible for collecting data on resource utilization, network load, service availability, and other relevant metrics, was created. The information is available in real time and can assist in optimizing the deployed applications and infrastructure. Additionally, notification and alerting functionalities were configured to inform about potential issues or threshold value breaches. Such an approach allows to quickly respond to events and to implement corrective measures. Security aspects were also taken care of in the project. Security was ensured at various levels, including during authorization, authentication, data encryption, and access control stages. Additionally, the implemented SIEM tools provide information if specific events occur, enhancing security monitoring and incident response capabilities.

The proposed concept of a cloud-based IDS can be utilized, for instance, in the case of a private cloud or a public cloud environment. Public cloud providers often do not pass much information to their clients (e.g. attempted or detected attacks) due to reputation-related concerns. The proposed IDS would enable the cloud to be monitored in an independent manner, and would provide information that is often unavailable. Importantly, only readily available open-source products were used to create our IDS. Thanks to such an approach, an in-house cloud IDS with no functionality-related limitations may be created. However, the location of the IDS is a crucial aspect that needs to be taken into consideration in such

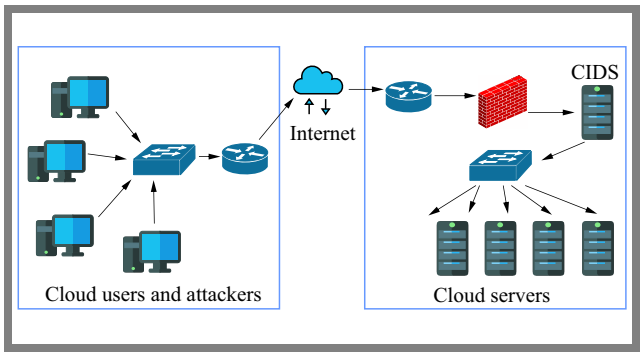

Fig. 7. Architecture of the proposed cloud IDS.

a situation. In cloud computing, the concept of virtualizing all resources is commonly used. This means that many cloud users are hosted on a single physical machine located in the cloud provider's data center. Hence, the implementation of a cloud IDS as HIDS may not be the best idea due to performance issues, as discussed in [17]. High traffic volumes will result in packet loss and will overload the hosting machine. Therefore, the best solution for a cloud IDS would be to deploy it as a network-based intrusion detection system (NIDS) and to place it in the back-end part of the cloud infrastructure. This would enable it to detect intrusions both from external and internal sources (Fig. 7).

7. Conclusions

Creating infrastructure, especially in the cloud, is not an easy process. Many factors need to be taken into account, such as security, scalability, costs, compliance and legal regulations, management, and monitoring. Cloud infrastructure differs from on-premises infrastructure, primarily in terms of costs. Everything should be thoroughly examined, analyzed, and continuously monitored to avoid unnecessary expenses or risk. Security is often overlooked, yet it proves to be the most crucial aspect of the process. Therefore, this work presents a security approach and highlights the most important factors to consider when planning cloud migration. Understanding the cloud service model, having the ability to identify potential threats, defining security, access control, data encryption, monitoring and recovery policies relied upon in the event of failures or attacks – all these are the very basics of cloud security.

The proposed concept of building your own IDS may be a successful in understanding how many mechanisms, which are typically hidden in a standard cloud IDS, need to be taken care of. Your own CIDS may be created, for instance, for the needs of a private cloud, based on an open-source platform, such as OpenStack [20]. However, it is important to keep in mind that when using major IT cloud providers such as AWS, Microsoft Azure or Google Cloud, we are not confined solely to their security solutions. Thanks to the open-source concept, there are numerous tools available on the market that can effectively replace commercial solutions. This provides us with greater flexibility, allowing us to expand and tailor solutions to our own needs without incurring additional costs, rather than being solely dependent on the cloud provider's offering. This is another advantage of cloud computing and one of the reasons why many organizations benefit from this particular approach. Despite ready-made commercial solutions being available, each cloud user has the choice of choosing the right software, and should not be prevented from implementing their own solutions.

References

[1] G. Rodrigues *et al.*, "Monitoring of Cloud Computing Environments: Concepts, Solutions, Trends, and Future Directions", *SAC '16: Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pp. 378–383, 2016 (https://doi.org/10.1145/2851613.2851619).

[2] M. Birje, P. Challagidad, R.H. Goudar, and M. Tapale, "Cloud Computing Review: Concepts, Technology, Challenges and Security", *International Journal of Cloud Computing*, vol. 6, no. 1, pp. 32–57, 2017 (https://doi.org/10.1504/IJCC.2017.083905).

[3] M.G. Azam, "Application of Cloud Computing in Library Management: Innovation, Opportunities and Challenges", *International Journal of Multidisciplinary*, vol. 4, no. 1, pp. 2–11, 2019 (https://doi.org/10.5281/zenodo.2536637).

[4] L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud Computing: An Overview", *IEEE International Conference on Cloud Computing*, pp. 626–631, 2009 (https://doi.org/10.1007/978-3-642-10665-1_63).

[5] L. Devadass, S.S. Sekaran, and R. Thinakaran, "Cloud Computing in Healthcare", *International Journal of Students Research in Technology and Management*, vol. 5, no. 1, pp. 25–31, 2017 (https://doi.org/10.18510/ijsrtm.2017.516).

[6] D. Sullivan, "Overview of Google Cloud Platform", in: *Official Google Cloud Certified Associate Cloud Engineer Study Guide*, John Wiley & Sons, pp. 1–14, 2019 (https://doi.org/10.1002/9781119564409.ch1).

[7] A. Vázquez, C. Dafonte, and Á. Gómez, "Open Source Monitoring System for IT Infrastructures Incorporating IoT-Based Sensors", *Proceedings*, vol. 54, no. 1, art. no. 56, 2020 (https://doi.org/10.3390/proceedings2020054056).

[8] M. Copeland, "Other Azure Security Services", in:*Cloud Defense Strategies with Azure Sentinel*. Apress, pp. 39–75, 2021 (https://doi.org/10.1007/978-1-4842-7132-2_2).

[9] T. Svoboda, J. Horalek, and V. Sobeslav, "Behavioral Analysis of SIEM Solutions for Energy Technology Systems", *Context-Aware Systems and Applications, and Nature of Computation and Communication, 9th EAI International Conference, ICCASA 2020, and 6th EAI International Conference, ICTCC 2020*, pp. 265–276, 2021 (https://doi.org/10.1007/978-3-030-67101-3_21).

[10] A. Mycek, D. Grzonka, and J. Tchórzewski, "Agent-Based Simulation and Analysis of Infrastructure-as-Code Process to Build and Manage Cloud Environment", *ECMS 2023: Proceedings of the 37th ECMS International Conference on Modelling and Simulation*, pp. 513–520, 2023 (https://doi.org/10.7148/2023-0513).

[11] D. Gupta, S. Bhatt, M. Gupta, O. Kayode, and A.S. Tosun, "Access Control Model for Google Cloud IoT", *2020 IEEE 6th Intl Conference on Big Data Security on Cloud, BigDataSecurity 2020, 2020 IEEE Intl Conference on High Performance and Smart Computing, HPSC 2020 and 2020 IEEE Intl Conference on Intelligent Data and Security, IDS 2020*, Baltimore, USA, pp. 198–208, 2020 (https://doi.org/10.1109/BigDataSecurity-HPSC-IDS49724.2020.00044).

[12] H. Zahid, S. Hina, M.F. Hayat, and G. A. Shah, "Agentless Approach for Security Information and Event Management in Industrial IoT", *Electronics*, vol. 12, no. 8, art. no. 1831, 2023 (https://doi.org/10.3390/electronics12081831).

[13] C.N. Modi, D.R. Patel, A. Patel, and R. Muttukrishnan, "Bayesian Classifier and Snort based Network Intrusion Detection System in Cloud Computing", *Third International Conference on Computing,*

Communication and Networking Technologies, Coimbatore, India, pp. 1–7, 2012 (https://doi.org/10.1109/ICCCNT.2012.6396086).

[14] C.N. Modi *et al.*, "A Survey of Intrusion Detection Techniques in Cloud", *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013 (https://doi.org/10.1016/j.jnca.2012.05.003).

[15] Y. Mehmood, A. Shibli, U. Habiba, and R. Masood, "Intrusion Detection System in Cloud Computing: Challenges and Opportunities", *Conference Proceedings - 2013 2nd National Conference on Information Assurance, NCIA 2013*, Rawalpindi, Pakistan, pp. 59–66, 2013 (https://doi.org/10.1109/NCIA.2013.6725325).

[16] U. Oktay and O.K. Sahingoz, "Proxy Network Intrusion Detection System for Cloud Computing", *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)*, Konya, Turkey, pp. 98–104, 2013 (https://doi.org/10.1109/TAEECE.2013.6557203).

[17] W. Elmasry, A. Akbulut, and A.H. Zaim, "A Design of an Integrated Cloud-based Intrusion Detection System with Third Party Cloud Service", *Open Computer Science*, vol. 11, no. 1, pp. 365–379, 2021 (https://doi.org/10.1515/comp-2020-0214).

[18] A. Patel, M. Taghavi, K. Bakhtiyari, and J. Celestino Junior, "An Intrusion Detection and Prevention System in Cloud Computing: A Systematic Review", *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 25–41, 2013 (https://doi.org/10.1016/j.jnca.2012.08.007).

[19] A. Tasneem, A. Kumar, and S. Sharma, "Intrusion Detection Prevention System Using SNORT", *International Journal of Computer Applications*, vol. 181, no. 32, pp. 21–24, 2018 (https://doi.org/10.5120/ijca2018918280).

[20] A. Sagala and R.M. Hutabarat, "Private Cloud Storage Using OpenStack with Simple Network Architecture", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 4, no. 1, pp. 155–164, 2016 (https://ijeecs.iaescore.com/index.php/IJEECS/article/download/5803/4578).

Andrzej Mycek, M.Sc.
Department of Computer Science
https://orcid.org/0000-0002-2720-6071
E-mail: andrzej.mycek@pk.edu.pl
Cracow University of Technology, Cracow, Poland
http://www.pk.edu.pl