National College of Ireland

Supervised Learning on Active directory with overcoming cybersecurity challenges

Practicum 2
MSc in Cybersecurity

Mabika Mabika
Student ID: X211321135

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Mabika Mabika
Student ID:	X211321135
Programme:	MSc Cyber Security **Year:** 2024
Module:	Practicum 2
Supervisor:	Vikas Sahni
Submission Due Date:	03/01/2025
Project Title:	Supervised Learning on Active directory with overcoming cybersecurity challenges
Word Count:	5614 **Page Count** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Mabika Mabika
Date:	28/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Supervised Learning on Active directory with overcoming cybersecurity challenges

Mabika Mabika

X211321135

Abstract

This research presents an analysis on the application of supervised learning techniques for threat detection in cybersecurity. With the increasing number of cases of cyber threats, many organizations face challenges in detecting these threats in time before they inflict harm. This research evaluates the integration of supervised learning models to address key cybersecurity challenges in Active Directory environments, such as unauthorized access, privilege escalation and anomalous behavior detection. Key findings reveal that various supervised learning algorithms such as support vector machine (SVM) networks in the detection of anomalous patterns, user behavior, and unauthorized access attempts. Additionally, the research discusses strategies to overcome challenges like feature selection and false positive rates, ultimately providing possible scenarios to experiment on supervised learning effectiveness using possible threats that can be inflicted on Active Directory, such as password spraying to check if supervised learning models are effective in minimizing the risks of cyber threats.

Keywords: Active Directory, Supervised Learning, cyberthreats and fraudulent detection.

1. Introduction

Active Directory has become a gateway of various security incidents in recent years through methods such as unauthorized access, Zero Days Attacks, social engineering and phishing. The ability to defend Active Directory can play a significant role in saving the company from financial loss, customer confidence and employee loss. In most organizations Active Directory is the center of operations in terms of IT employee management and the ability to be able to recover Active Directory in the event of the unforeseen disaster is a very big strategy which organizations help themselves from ransomware attacks. Active Directory (AD) is a widely used directory service that provides centralized authentication, authorization and management of resources on a network domain. AD runs on Windows Server machines or on cloud-based platforms such as Azure AD. AD stores sensitive information about users, computers, groups, policies, and other objects on the network making it a critical service for

many organizations. AD works on a network domain structure and its main purpose is to provide authorization, authentication and providing measures to protect Information Technology (IT) and computing services for organizations (McDonald, Papadopoulos, Pitropakis, Ahmad & Buchanan, 2022). Furthermore McDonald et al. (2022), highlight AD is also vulnerable to various attacks that aim to compromise its integrity, availability, or confidentiality. These attacks include ransomware, malware, credential theft, privilege escalation, domain takeover, and more. Such attacks can cause significant damage to the organization, such as data loss, service disruption, unauthorized access, or identity theft. Over time, attacks using AD have changed, attackers target several capabilities and functionalities offered by AD (Mokhtar, Jurcut, ElSayed, & Azer, 2022). Therefore, it is essential to understand the security challenges of AD and how to prevent or mitigate them.

Information Technology (IT) and computing play an integral part of day-to-day operations within enterprises and organizations, increasing productivity in the modern workplace. This has made IT services critical infrastructure to businesses and organizations, however when the IT systems become compromised businesses can lose large amounts of money and in some cases vital information. Cybercriminals have caught onto this and are continuously finding ways to cause harm, through data destruction, causing downtime through malware and ransomware (McDonald et al., 2022). The integration of machine learning encompasses various types of learning algorithms, supervised learning is a subset of machine learning which is trained on labelled data. This paper aims to look at how labelled data handle, stores information and protects AD from unauthorized access from hackers. The motivation for this study comes from various recent literature on AD and how various organizations have used and are using supervised learning in protecting their assets/infrastructure and are able to detect zero-day attacks and most important overcoming the challenges encountered while using AD.

2.2 Motivation

In many organizations, Active Directory plays a major role in protecting the network resources, allocating resources & privileges, authenticating and authorizing. Due to the increasing and complexity of cybersecurity threats worldwide, this has exposed many organizations particularly in their network infrastructure mainly the AD, which play the pivotal role in network security. Supervised learning which is a part of machine learning, has been pivotal in emerging technologies to determine potential threats by detecting anomalies and enhancing network security. This study aims to propose supervised learning techniques to overcome cybersecurity challenges within the AD.

1.2 Research Question

How can supervised learning techniques be used to detect and prevent password spraying attacks within Active Directory?

This research question attempts to focus on using machine learning to enhance cybersecurity by addressing a prevalent attack vector furthermore it allows an exploration of the effectiveness of supervised learning models in identifying and mitigating password spraying attacks, which is a common threat to Active Directory security.

1.3 Research Objectives

The following research objectives are targeted to be achieved:

- To determine the critical features and indicators of password spraying attacks from Active Directory logs and related data sources.
- To assess the effectiveness of Support Vector Machines as a supervised learning algorithms in identifying password spraying attempts with high accuracy.
- To identify automated mitigation strategies based on the detection of password spraying attacks, and evaluate their impact on cybersecurity.
- To conduct experiments and case studies to validate the proposed framework in a real-world Active Directory scenario, demonstrating its practical applicability and scalability.

2. Related Work

In this section, existing research work in AD structure/architecture, AD security through supervised learning and overcoming challenges encountered in AD security is reviewed and classified.

2.1 Architecture, functionality of AD and how it relates to security

The authors Kotlaba, Buchovecká & Lórencz (2021) highlighted that AD is a hierarchical and distributed database that organizes and manages the network objects, such as users, computers, groups, policies and services. AD is based on the Lightweight Directory Access Protocol (LDAP), which is a standard protocol for accessing and modifying directory information over a network. Obimbo & Ferriman (2011) highlights that Lightweight Directory Access Protocol (LDAP) are servers widely used to authenticate users in an organization. AD typically has and is used for a broad range of identity-related services in a domain which include authentication, authorization, and accounting services for users and computers. Microsoft is the proprietary service which developed AD, it has further evolved to integrate both Windows and Non-Windows gadgets. The use of AD has elevated to a new

level that it is now hybrid and on public cloud environments with products such as Azure AD and AWS Directory Service in McDonald et al. (2022) the authors ascertain that the security of AD depends on the security of its components and the network infrastructure that supports them. AD relies on several security mechanisms and protocols, such as Kerberos, Secure Sockets Layer (SSL), Transport Layer Security (TLS), encryption, hashing, digital signatures, certificates, and access control lists (ACLs). These mechanisms and protocols are designed to protect the communication, data, and identity of the AD objects and users. However, they are also subject to various vulnerabilities and limitations, such as configuration errors, mismanagement, outdated software, weak passwords, spoofing, tampering, replay, and man-in-the-middle attacks.

In a study conducted by Binduf, Alamoudi, Balahmar, Alshamrani, Al-Omar & Nagy (2018) which related to security issues in AD environment, the authors highlighted that AD plays an important aspect in the business environment allowing greater control of the business. The study showcased vulnerabilities within AD and stressed the critical elements that must be hardened to protect AD resources from attacks. Furthermore, in a study by Pektas & Basaranoglu (2017) the study presented a penetration testing on AD. This study came up with a structure that had 10 stages that could be performed by attackers on AD which was called MSDEPTM (Microsoft Domain Environment Penetration Testing Methodology) to provide essential/important metrics for Microsoft domain penetration testing. MSDEPTM is a penetration methodology for Windows domain environment which provides key metrics for Microsoft domain penetration testing. This research compared their structure viz a vis other standard penetration testing and outlined the advantages of their structure. The research provided methods that can be used to secure AD from attackers.

2.2 Active Directory security through Machine Learning

Machine Learning was developed to provide systems to automatically learn how to detect, decide and execute tasks. Machine Learning (ML) can either be supervised, unsupervised, semi-supervised and reinforcement. This research focuses on supervised learning, according to Trivedi & Khadem (2022) Supervised learning is defined as a process in which labelled datasets are trained to predict the continuous value target feature. Apruzzese, Laskov, Montes de Oca, Mallouli, Brdalo Rapa, Grammatopoulos & Di Franco (2023) suggest that approaches based on supervised learning are expensive to deploy as they are reliable and provide excellent results since the reliance is from good quality labels.

In a study done by Apruzzese et al. (2023) the authors highlighted a holistic approach of ML in the entire cybersecurity domain, showcasing advantages of ML with respect to human driven detection methods. The authors proposed various intrinsic factors affecting real ML deployments in cybersecurity. The study conducted two real case studies describing industrial applications of ML as a defense mechanism against cyber-threats.

The authors Matsuda, Fujimoto & Mitsunaga (2018) studied and proposed a technique to detect attacks on AD involving the domain administrator accounts using Windows event logs. These events are supervised and they prove that they are efficient in detecting attacks. However, in the study, the drawback highlighted is that the detections are based on an explicit list of commands which attackers can run giving limited flexibility for adaption when the attackers use a different method or different angle of attack to carry out malicious activities. In a separate study done by Pektas & Basaranoglu (2017) to detect Advanced Persistent Threat (APT) attacks using ML techniques, the results yielded highlighted that ML is useful and effective. Although the study was unsupervised machine learning they also used event logs as a data source and emphasized that detecting attacks using process creation in detecting APT attacks in AD environments is crucial. From the study the limitation of the approach is that false positives can be detected if the administrator uses the same commands as the attackers.

3. Methodology

The methodology used in this study was experimental research, Em (2024) highlights that experimental research is a scientific method that is used to investigate cause and effect relationships between variables. This research methodology was selected as it manipulated one or more independent variables to observe the effect on a dependent variable. Furthermore, the experimental research sought to investigate the cause-and-effect relationships between variables, independent variables being password complexity, authentication attempts, user behavior and security policies whilst dependent variables are security breaches, incident response time and user compliance.

a. **Manipulation of variables:** This study the researcher manipulated the security of the Active Directory by using password spraying as an attack on Active Directory. Van der Stede (2014) ascertains that the researcher intentionally changes or manipulates one or more independent variables to observe an effect on the dependent variable. The manipulation creates an environment which allows the researcher to assess the result.

b. **Controlled Conditions:** This study was done in a controlled environment which is a home lab using the VMware software. According to a study done by (Ross & Morrison, 2013) experimental research is conducted under controlled conditions to ensure that redundant variables do not misperceive the results. Controlling variables makes researchers be able to isolate the effects of the independent variable on the dependent variable.

c. **Measurement of Variables**: In the study Sysmon is used as an event monitor to obtain important data which can be used to measure the independent and dependent variable. An experimental research evaluation of the relationship is obtained by measuring both the independent and dependent variables. Accurate data collection is obtained by using valid and trustworthy measuring instruments (Bhandari, 2022).

d. **Causal Inferences**: In experimental research it is important to determine the casual links between variables as this is the main objectives of experimental research. Conclusions can be drawn from the casual relationship between variables by manipulating an independent variable and monitoring the impact that the variable will have on the dependant variable (Kuang, Li, Geng, Xu, Zhang, Liao & Jiang, 2020).

e. **Procedures in Conducting Experimental Research**:
   According to Ross & Morrison (2013) methodical procedures and techniques used by researchers to carry out experimental research. They include study design, participant selection, intervention implementation, data collection and result analysis.

In the study the methodical procedure was as follows:

1. Creation of a home lab - Active Directory using Virtual Machine (Oracle VirtualBox 7.0).
2. Selected Kali Linux 6.11.2-amd64 as my attacker and Windows Server 2019 as the victim.
3. Performed attacks on the Domain Controller of the Active Directory from Kali Linux machine.
4. Chose password spraying as an intervention method to be able to manipulate the dependent.
5. Used Sysmon v15.15 for data collection on the Event Viewer.
6. Analysed data using extracted information from Event viewer and supplemented using public data sets and existing research.
7. Finally draw conclusions based on the findings.

3.1 Design Specification

The design specification and process flow of my proposed methodology are discussed and presented in detail in this section. The process flow diagram below describes the actions taken to achieve the primary objective of detecting and mitigating attacks on AD, including password spraying. The methodology is built on the approach below:

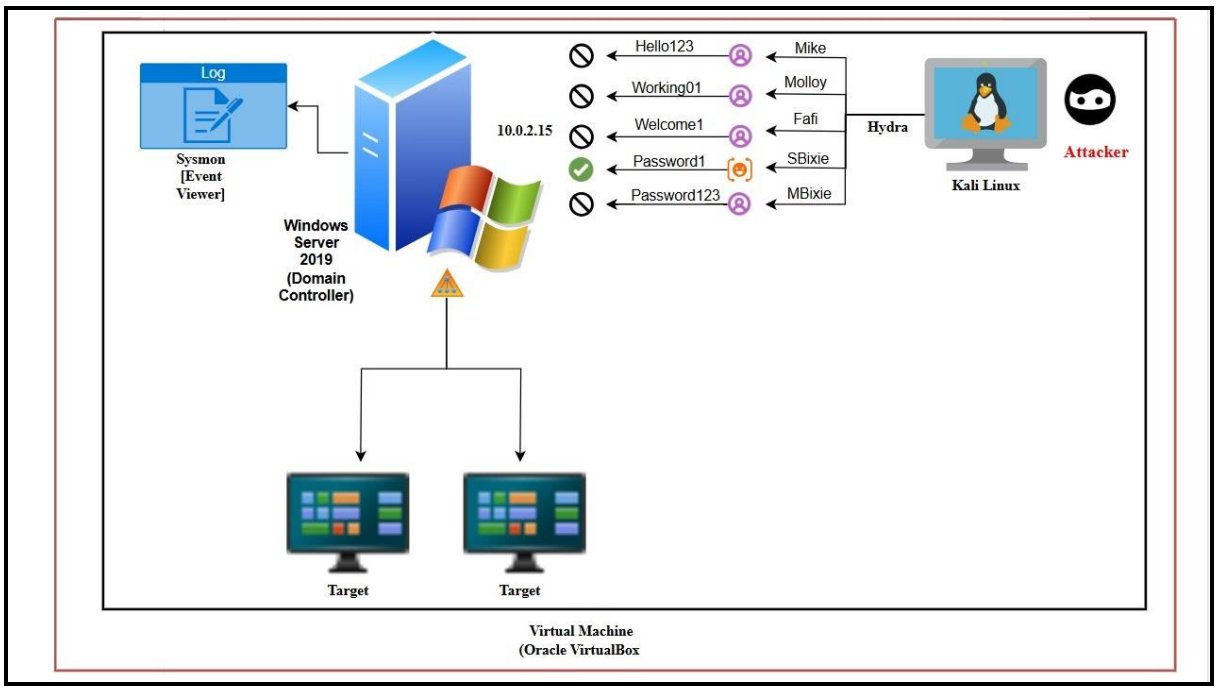

Figure 1: Process flow diagram

The methodology also incorporated a simulated attack which was password spraying, to validate the model's detection capability. The combination of supervised learning, detailed log analysis, and practical attack scenarios is aimed to improve the overall security posture of AD environments.

3.2 Supervised Learning Model Framework

The study leveraged on Support Vector Machines (SVM) as part of Supervised learning model which is trained on labeled datasets. The datasets are generated from controlled environments where both normal and malicious activities are simulated.

- Leveraging Sysmon for advanced logging and event tracking
- Implementing supervised machine learning models trained to identify anomalous behaviors indicative of potential attacks.

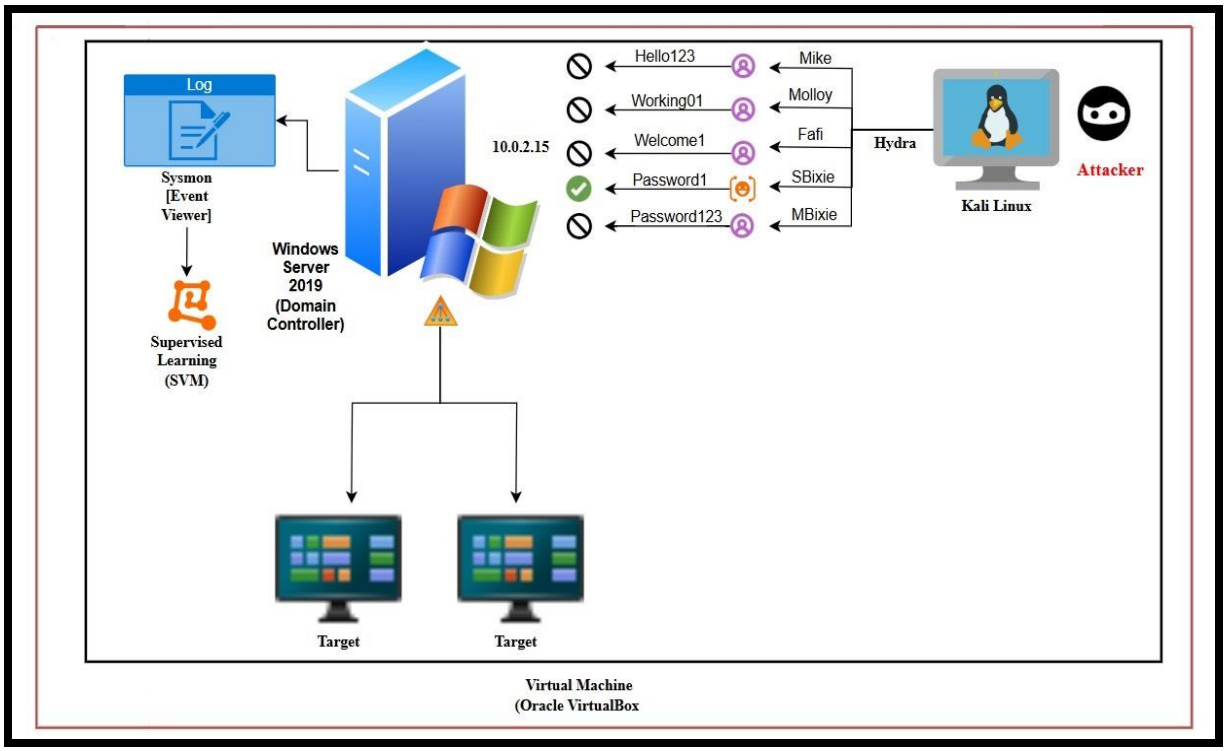

Figure 2: Process flow for Supervised Learning (SVM) detecting attacks on Active Directory

3.3 Password Spraying Attack

In the study password spraying was implemented on the AD using Hydra, from the attacking machine. Password spraying was used in the study as it is one of the prevalent attacks amongst the rest used by attackers for targeting AD environments. In the study it is very important to observe changes in Sysmon (event logs) and system behavior during the attack.

3.4 Sysmon for Data Collection

The study deployed Sysmon to capture detailed event logs on critical AD components. These logs include information about process creation, network connections, and other activities that can signal malicious behavior. The logs are then ingested into a machine learning model for analysis. Sysmon's key features in enhanced monitoring include detailed event capture, which includes granular details for identifying anomalous patterns. Sysmon's configuration files also allow for customization and fine-tuning of captured events to reduce noise and focus on critical data.

The table below illustrates Sysmon configurations for data collection.

Configuration	Description
HashAlgorithms	SHA256 for integrity verification.
LogonSessionIDs	Correlation of session data with processes
FileCreateTime	Monitoring sensitive file modifications.

Table 1: Sysmon Configurations

3.5 Supervised Learning for Threat Detection

The design of our model employs supervised learning. Supervised learning is employed to detect malicious activities and anomality activity in AD environments. As suggested by Pang, Shen, Cao & Hengel (2022) anomaly detection is intended to detect any abnormal or unusual activity observed in the data.

Supervised learning is a model that is trained on labeled datasets containing both benign and malicious events. The features of the model include Sysmon logs such as process names, network activity and event timestamps. Supervised learning algorithms could include Random Forest and Support Vector Machine (SVM). In our approach we have chosen three representative algorithms of the kind, we used Support Vector Machine (SVM) for novelty detection. The process flow involved preprocessing logs to reduce noise and standardize the formats. The relevant features are then identified and transformed to create a structured dataset. The algorithms being the principle point of supervised learning are trained on historical data with known attack labels. Furthermore, real-time analysis is implemented using the trained algorithms to identify potential attacks. The performance metrics of the model is evaluated on Accuracy, Precision and Recall and the False Positive Rate. Accuracy helps classify events as either malicious or benign, while Precision and Recall measures the models' effectiveness in identifying true positives. The False Positive Rate metric helps us in minimize false alarms in the system as it identifies threats.

3.6 Implementation

This section explains the implementation of supervised learning models and Sysmon in a lab environment. The implementation was carried out in two phases.

3.6.1 Tools Used

i. Virtual Machine

Oracle VirtualBox V7.0 was used as a virtualization software to create and manage the virtual machines for the experiment environment.

ii. Kali Linux – Attacker

The attacking machine was installed with Kali Linux where password spraying attacks were simulated using Hydra.

iii. Windows Server 2019 - Domain Controller

Domain Controller was the main primary target from the attacker, it contained the Sysmon logs.

iv. Sysmon Monitoring - Sysmon

Sysmon is a Windows system service used to collect detailed logs of system activities from Domain Controller which was the target machine.

3.6.2 Phase 1: Environment Setup

Lab Setup: An AD domain was configured with a simulated organizational structure.

Sysmon Installation: Sysmon was installed on Domain Controller (Windows 2019 Server) to monitor log activity.

Dataset Generation:

Normal activity: Routine user and system behaviors.

Attack simulation: Password spraying.

3.6.3 Phase 2: Supervised Learning Models

The logged data was preprocessed to generate feature vectors for training the models. Features include:

- Login frequency and duration.
- Service ticket patterns.
- Network activity anomalies.

3.6.4 Phase 3: Technology

For practical implementation, AD was implemented in a controlled home lab in a Virtual Machine on the Oracle VirtualBox V7.0. Virtual Machine allows us to setup multiple

machines which include the attacking machine (Kali Linux 6.11.2-amd64) and the victim which is the Domain Controller running Windows Server 2019.

On the attacker (Kali Linux), Hydra was used to perform password spraying. Hydra is an open source, password brute-forcing tool designed around flexibility and high performance in online brute-force attacks. Hydra is a parallelized login cracker which supports numerous protocols to attack (kali, 2024), furthermore it is very fast and flexible, and new modules are easy to add. Hydra is a tool that makes it possible for researchers and security consultants to show how easy it would be to gain unauthorized access to a system remotely. Hydra provides brute-forcing capabilities for various protocols such as LDAP, http and SMP. It was designed to be parallelized, meaning multiple threads can operate in parallel to optimize efficiency and speed up the brute-forcing process.

On the client (Windows Server 2019) which has our Active Directory, Sysmon was installed to monitor all logs and any suspicious activity. As illustrated below Hydra was used to attack the LDAP on Active Directory. Assuming that we have one of the Users from AD, we can use the Password.txt file to try get the user's password. It can either be successful or unsuccessful. In our experiment the firewall blocked connection of the attacker this was due to the group policy rules.

3.6.5 Phase 4: Dataset Preparation

The dataset used for the experiment is non-synthetic, originating from our AD environment of a Virtual Machine home lab with few daily users. It consists of Windows Events 18 857 collected over the time span of 4 days. As the 18 857 Events belongs to the overall numerous events logged in a Windows domain, the dataset was further filtered and 9309 Events were highlighted with critical, verbose and warning. The Events were exported to CSV file so that they could be processed and presented in graph form for analysis. Data contains only regular traffic from the environment. A few crafted malicious events, which are to represent potential Password spraying activity, were injected into validation and testing datasets.

4. Evaluation

The use of Kali Linux to attack Domain Controller stimulated the experimental environment. Kali Linux was installed on the VirtualBox as an attacking virtual machine. Windows Server

2019 was also installed on VirtualBox as the machine to be attacked. Hydra, which was preinstalled on Kali Linux was used to password spray the Domain Controller and Sysmon installed on the Windows server captured the Event logs. The main goal was to password spray the Domain Controller. This section evaluated the system's ability to detect attacks effectively.

4.1 Experiment 1

Testing Environment Setup

For the password spraying to be done, 2 Virtual Machines were setup for the Password spraying attack. Kali Linux being the attacker and Windows Server 2019 (Domain Controller) being the client. Domain Controller was setup/configured with over 1000 users. The below were considered in the environment setup:

i. The attacking machine (Kali Linux) in our study we used Hydra as its attacking agent which comes as preinstalled.
ii. Domain Controller is designed to keep away unauthorized entry of attackers. The target machine was setup with Active Directory.
iii. On the Domain Controller, Active Directory had 1000 users were configured.
iv. The password spraying was performed and Sysmon used to collect data logs.

Detecting Password spraying

Domain Controller system was subjected to password spraying attacks. The password spraying attacks were to detect failed login. The attacker used a list of passwords to use against a different combination of usernames on the target system. To perform password spraying, logging into the Kali Linux machine is the first step so we can make use of Hydra to access the password spraying protocols. Install Hydra, if Hydra is not already installed. The following commands can be used to install Hydra.

sudo apt-get install hydra

We create text files containing the usernames (users.txt) and common passwords (password.txt) you want to test. In the Kali Linux terminal use the following command to configure Hydra for password spraying against the LPAD protocol on the target machine:

hydra -L users.txt -P password.txt -u -t 1 <target_ip> ldap

To execute password spraying, run the Hydra command. Hydra will attempt to log in with each username from users.txt using each password from passwords.txt. The -t 1 option specifies that Hydra should use only one task (or connection) at a time, which helps avoid overwhelming the LPAD service. Hydra terminal will display its progress, showing any successful logins or failures. Hydra rejected connection between the attacker and the Domain Controller after a certain threshold of failed attempts. The connection was likely rejected due to account lockout policies or rate-limiting mechanisms. To detect the password spraying attempts, two mechanisms were utilized, Monitoring Logs and Behavioral Analysis. For the monitoring logs, we checked the Domain Controller's event logs for multiple failed logins attempts with different usernames and a common password.

Behavioral analysis, we used a security tool (for example, SIEM systems) to detect patterns indicative of password spraying.

Real-Time Monitoring

Sysmon logs were installed and used to analyze logs in real-time. Sysmon logs were exported to csv file so that they are analyzed further using Excel. Initial data was captured over a total of 3 days based on LDAP protocol. Data was captured on period starting Wednesday 18 December 2024 and ended Friday 20 December 2024. Wednesday no attack was performed, and the logs were normal, Thursday 19 December 0055am password spraying was attempted, the logs showed that there was a process create and finally on Friday 20 December was a normal day.

4.2 Experiment 2

Using the CICIDS2017 dataset which contains benign and the most up-to-date common attacks and which resembles the true real-world data (PCAPs). The results of the network traffic analysis using CICFlowMeter with labelled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files) generated a realistic background traffic. The dataset was built on the abstract behavior of 25 users based on the HTTP, HTTPS, FTP, SSH, and email protocols. And the data was captured on period starting at 9 a.m., Monday, July 3, 2017 and ended at 5 p.m. on Friday July 7, 2017, for a total of 5 days. Monday is the normal day and only includes the benign traffic. The implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. They have been executed both morning and afternoon on Tuesday, Wednesday, Thursday and Friday. In the published paper by Sharafaldin, Lashkari & Ghorbani (2018) three common information retrieval evaluation metrics:

Precision (Pr) or Positive Predictive value: It is the ratio of correctly classified attacks flows (TP), in front of all the classified flows (TP+FP).

Recall (Rc) or Sensitivity: It is the ratio of correctly classified attack flows (TP), in front of all generated flows (TP+FN).

F-Measure (F1): It is a harmonic combination of the precision and recall into a single measure. Pr = T P T P+FP, Rc = T P T P+FN, F1 = 2 1 Pr + 1 Rc.

Algorithm	Pr	Rc	F1	Execution (Sec.)
KNN	0.96	0.96	0.96	1908.23
RF	0.98	0.97	0.97	74.39
ID3	0.98	0.98	0.98	235.02
Adaboost	0.77	0.84	0.77	1126.24
MLP	0.77	0.83	0.76	575.73
Naive-Bayes	0.88	0.04	0.04	14.77
QDA	0.97	0.88	0.92	18.79

Table 2: Performance Examination Results (Sharafaldin et al., 2018)

The table above shows the performance examination results in terms of the weighted average of evaluation metrics for the seven selected common machine learning algorithms, namely K-Nearest Neighbors (KNN), Random Forest (RF), ID3, Adaboost, Multilayer perceptron (MLP), Naive-Bayes (NB), Quadratic Discriminant Analysis (QDA) derived from the generated dataset which can be used for Supervised Learning.

- Additionally brute-force attacks came out to be in the top five initial intrusion vectors which were observed in 2023, representing 6% of intrusions. Attackers continue to leverage on emerging technologies to effectively use their tactics to gain access to target environments and conduct their operations.

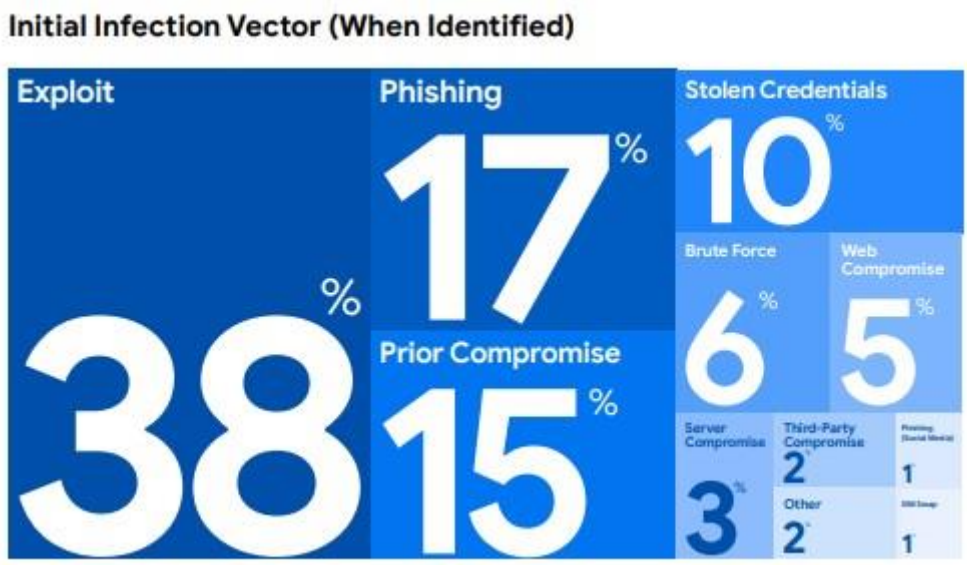

Figure 9: 2023 Intrusion Vector Statistics

4.3 Discussion

The analysis of the above study the integrated methods that can be employed to secure AD on the Domain Controller. The testing results evidently highlight that AD has to be secured. Based on the analysis above depicted in case study procedure and public dataset the testing results demonstrate that adopting machine learning techniques to improve detection of attacks. Opponents use various attacks such as password spraying to target Active Directory environments, with the goal of extracting credentials of service accounts. We utilized anomaly detection algorithms to identify unusual patterns in authentication to remote services that may indicate the possibility of password spraying attack execution this was further supported by the public dataset obtained from The Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick with their dataset CIC-IDS2017 which includes a variety of common attacks such as Brute Force FTP, SSH, DoS, DDoS and Web Attack. It also contains both benign and malicious traffic. On both, the case study and public dataset the algorithms were compared and evaluated on log data originating from a modelled network structure. The different configurations of their hyperparameters and features extracted from Windows Events have been very useful. As the results have shown, Supervised Learning approach could significantly reduce the number of false-positive detections compared to the signature-based approach although it is time consuming. At the same time, we did not observe any number of false negatives which helped improve detection capabilities. In the

study conducted by Raghavan (2019) they observed that SVM had a high training and was computational efficient making it to be a good model.

5. Conclusion and Future Work

In this paper, the study proposes adopting machine learning techniques to improve security of Active Directory, securing it from password spraying. AD is used by most organization to store information in one place and make administrators to manage, the structured data which gives the Administrators uses of structured data which is stored as the basis for a logical, hierarchical organization of directory information. AD is prone to attacks, making it an entry point hence it must be kept secured all the time. Due to the fact that AD is home to sensitive to information it may be compromised. As a mitigating approach to fight against attacks supervised learning can be deployed to fight against attackers. According to the research, Active Directory can be attacked by using various methods however with particular reference to this study password spraying was used by the attacker. Incorporating the dataset used, the same was applied as brute force which the attacker used. The results that were taken from the Event logs conclusively highlight that supervised learning would be essential for protecting AD although attackers are constantly evolving, use of other supporting methods such as firewall, regular antivirus update, employee training and system/network monitoring can be used to the good of AD. From the evaluation observed in the previous section the results were satisfactory and SVM proved to be a better model for supervised learning. The overall study proves that the AD can be considered to be secured using supervised learning in network security. Thus, the purpose of the study was fulfilled.

5.1 Limitations

The study was restricted to network devices & its complexity and time constraints.

5.2 Future Work

If given time and better modelling of network devices a larger dataset can be obtained from various network resources giving better data that can be used in the future.

References

Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A. V., & Di Franco, F. (2023). The Role of Machine Learning in Cybersecurity. *Digital Threats: Research and Practice*, *4*(1), 1–38. https://doi.org/10.1145/3545574

Bhandari, P. (2022). A Note on Survey Research Methods Levels of Measurement: Foundational Basis for Quantitative Analysis of Survey Data. *Dhaulagiri Journal of Sociology and Anthropology*, 122–126. https://doi.org/10.3126/dsaj.v16i01.50982

Binduf, A., Alamoudi, H. O., Balahmar, H., Alshamrani, S., Al-Omar, H., & Nagy, N. (2018). Active Directory and Related Aspects of Security. *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 4474–4479. https://doi.org/10.1109/NCG.2018.8593188

Em, S. (2024). Exploring Experimental Research: Methodologies, Designs, and Applications Across Disciplines. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4801767

Kotlaba, L., Buchovecká, S., & Lórencz, R. (2021). Active Directory Kerberoasting Attack: Detection using Machine Learning Techniques. *Proceedings of the 7th International Conference on Information Systems Security and Privacy*, 376–383. https://doi.org/10.5220/0010202803760383

Kuang, K., Li, L., Geng, Z., Xu, L., Zhang, K., Liao, B., Huang, H., Ding, P., Miao, W., & Jiang, Z. (2020). Causal Inference. *Engineering*, *6*(3), 253–263. https://doi.org/10.1016/j.eng.2019.08.016

Matsuda, W., Fujimoto, M., & Mitsunaga, T. (2018). Detecting APT Attacks Against Active Directory Using Machine Leaning. *2018 IEEE Conference on Application, Information and Network Security (AINS)*, 60–65. https://doi.org/10.1109/AINS.2018.8631486

McDonald, G., Papadopoulos, P., Pitropakis, N., Ahmad, J., & Buchanan, W. J. (2022). Ransomware: Analysing the Impact on Windows Active Directory Domain Services. *Sensors*, *22*(3), 953. https://doi.org/10.3390/s22030953

Mokhtar, B., Jurcut, A., ElSayed, M., & Azer, M. (2022). Active Directory Attacks—Steps, Types, and Signatures. *Electronics*, *11*(16), 2629. https://doi.org/10.3390/electronics11162629

Obimbo, C., & Ferriman, B. (2011). Vulnerabilities of LDAP As An Authentication Service. *Journal of Information Security*, *02*(04), 151–157. https://doi.org/10.4236/jis.2011.24015

Pang, G., Shen, C., Cao, L., & Hengel, A. Van Den. (2022). Deep Learning for Anomaly Detection. *ACM Computing Surveys*, *54*(2), 1–38. https://doi.org/10.1145/3439950

Pektas, A., & Basaranoglu, E. (2017). Practical Approach for Securing Windows Environment : Attack Vectors and Countermeasures. *International Journal of Network Security & Its Applications*, *9*(6), 13–27. https://doi.org/10.5121/ijnsa.2017.9602

Raghavan, P., & Gayar, N. El. (2019). Fraud Detection using Machine Learning and Deep Learning. *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, 334–339. https://doi.org/10.1109/ICCIKE47802.2019.9004231

Ross, Steven. M., & Morrison, Gary. R. (2013). Experimental research methods. In Handbook of research on educational communications and technology. In *Experimental research methods. In Handbook of research on educational communications and technology* (pp. 1007–1029). Routledge.

Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 108–116. https://doi.org/10.5220/0006639801080116

Trivedi, R., & Khadem, S. (2022). Implementation of artificial intelligence techniques in microgrid control environment: Current progress and future scopes. *Energy and AI*, *8*, 100147. https://doi.org/10.1016/j.egyai.2022.100147

Van der Stede, W. A. (2014). A manipulationist view of causality in cross-sectional survey research. *Accounting, Organizations and Society*, *39*(7), 567–574. https://doi.org/10.1016/j.aos.2013.12.001