AI for

Cybersecurity

By: Tang Kai Jie

2200741C

Cover Page

a) Model Training

1. introduction to my model

2. Model and accuracy

b) Develop an attack plan based on phishing

c) Mapping to MITRE ATLAS Framework

d) Develop a defense strategy plan

Data source: https://www.kaggle.com/datasets/eswarchandt/phishing-website-detector

Model Training

1. Introduction to my model

Topic I have chosen:

The topic I have chosen is phishing. Phishing attacks are a simple yet effective method for obtaining sensitive information from unsuspecting users, targeting critical data such as usernames, passwords, and bank account details. Cybersecurity experts are now seeking reliable detection techniques for identifying phishing websites. This paper explores the use of machine learning to detect phishing URLs by analysing features of both legitimate and phishing URLs, employing algorithms such as Decision Tree, Random Forest, and Support Vector Machine. It aims to identify phishing URLs and determine the most effective machine learning algorithm by comparing their accuracy rates, false positive rates, and false negative rates. The pervasive nature of the Internet has not only enhanced connectivity but also provided opportunities for malicious activities like phishing. Phishers use social engineering or create fake websites to illegally obtain sensitive information. Despite various detection methods, phishers adapt to evade these mechanisms. Machine learning has proven to be a powerful tool in identifying common characteristics of phishing attacks.

**Dataset Description:**

1. The dataset for a “.txt” file is with no headers and has only the column values.
2. The actual column-wise header is described above and, if needed, you can add the header manually if you are using '.txt' file.If you are using '.csv' file then the column names were added and given.
3. The header list (column names) is as follows :  
   [ 'UsingIP', 'LongURL', 'ShortURL', 'Symbol@', 'Redirecting//',  
   'PrefixSuffix-', 'SubDomains', 'HTTPS', 'DomainRegLen', 'Favicon',  
   'NonStdPort', 'HTTPSDomainURL', 'RequestURL', 'AnchorURL',  
   'LinksInScriptTags', 'ServerFormHandler', 'InfoEmail', 'AbnormalURL',  
   'WebsiteForwarding', 'StatusBarCust', 'DisableRightClick',  
   'UsingPopupWindow', 'IframeRedirection', 'AgeofDomain',  
   'DNSRecording', 'WebsiteTraffic', 'PageRank', 'GoogleIndex',  
   'LinksPointingToPage', 'StatsReport', 'class' ]

**Brief Description of the features in data set**

● UsingIP (categorical - signed numeric) : { -1,1 }  
● LongURL (categorical - signed numeric) : { 1,0,-1 }  
● ShortURL (categorical - signed numeric) : { 1,-1 }  
● Symbol@ (categorical - signed numeric) : { 1,-1 }  
● Redirecting// (categorical - signed numeric) : { -1,1 }  
● PrefixSuffix- (categorical - signed numeric) : { -1,1 }  
● SubDomains (categorical - signed numeric) : { -1,0,1 }  
● HTTPS (categorical - signed numeric) : { -1,1,0 }  
● DomainRegLen (categorical - signed numeric) : { -1,1 }  
● Favicon (categorical - signed numeric) : { 1,-1 }  
● NonStdPort (categorical - signed numeric) : { 1,-1 }  
● HTTPSDomainURL (categorical - signed numeric) : { -1,1 }  
● RequestURL (categorical - signed numeric) : { 1,-1 }  
● AnchorURL (categorical - signed numeric) : { -1,0,1 }

1. Model and accuracy

After I analysis the data (EDA) I split the data and tested them using different ways. The ways that I have used:

1. Logistic Regression
2. KNN algorithm
3. Decision Tree
4. Support Vector Machine Algorithm
5. Gradient Boosting
6. AdaBoosting Classifier
7. Random Tree

The results for the models are:

Logistic Regression Accuracy: 0.9243292131444076

K-Nearest Neighbour Accuracy: 0.9424178474525173

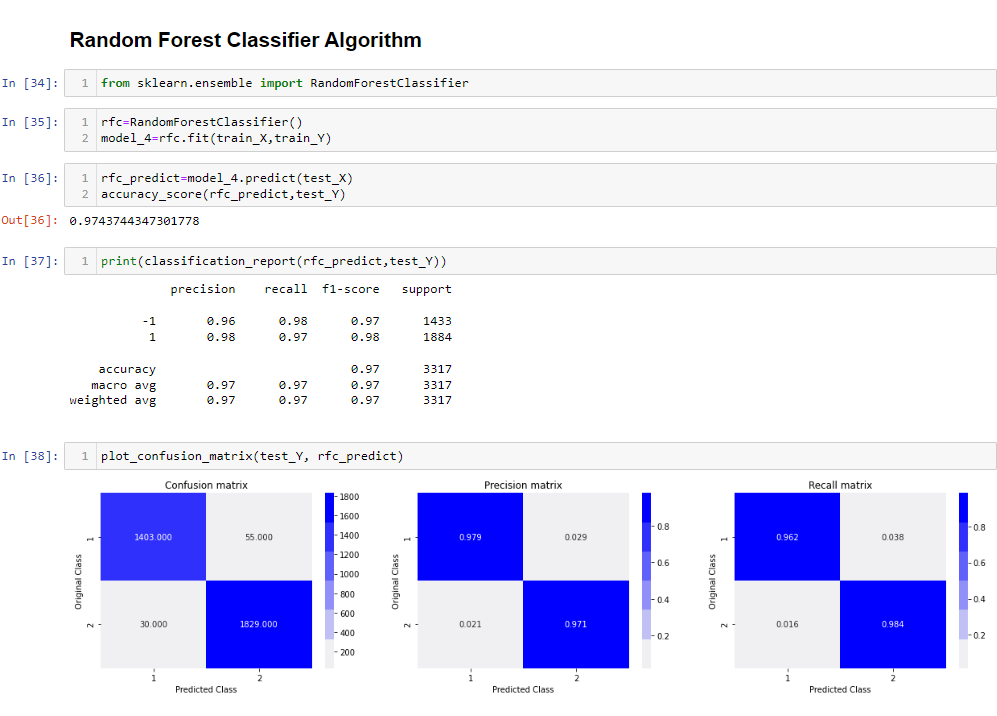
Decision Trees Classifier Accuracy: 0.9614589596329812

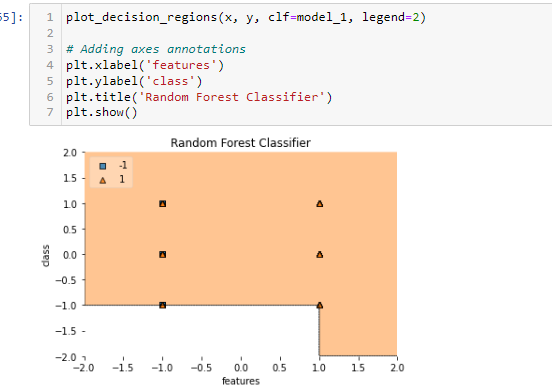
Random Forest Classifier Accuracy: 0.9743744347301778

support Vector Machine Accuracy: 0.9472414832680133

Adaboost Classifier Accuracy: 0.9104612601748568

So inconclusion the best to the worst model is: Random Forest Classifier Accuracy, Decision Trees Classifier Accuracy, support Vector Machine Accuracy, K-Nearest Neighbour Accuracy, Logistic Regression Accuracy, Adaboost Classifier Accuracy.





Develop an attack plan based on my model.

Before I start developing an attack plan on my model using phishing here are some assumptions, I can take notice to further help me in planning my attack.

First Attack:

Assumptions:

The ML model is hosted on a publicly accessible server.

The architecture of the ML model is exposed, providing details on how the model processes input and generates output.

The data source used by the ML model is exposed, making it possible to understand the type of data the model interacts with.

The target users include system administrators, data scientists, and other stakeholders with access to the ML model and its data.

Attack Plan:

1. Reconnaissance

It involves gathering as much information as possible about the target to craft a convincing phishing campaign.

Steps:

Server Identification:

Use tools like Shodan or Censys to identify the server hosting the ML model. Look for specific open ports, services, and software versions that might reveal additional details about the server.

Model Architecture Analysis:

Examine the exposed architecture to understand how the model processes inputs and generates outputs. This might include studying any public documentation, API endpoints, and open-source components.

Data Source Investigation:

Investigate the exposed data source to understand the types of data being processed by the ML model. This includes reviewing any accessible databases, file storage systems, or data pipelines.

Stakeholder Identification:

Collect information about potential targets, such as system administrators and data scientists, through social media (LinkedIn, Twitter) and professional networking sites. This information will help in crafting personalized phishing emails.

Tools: Shodan, Censys, Nmap.

1. Crafting a phishing email

Create a convincing phishing email that appears legitimate and relevant to the stakeholders.

Steps:

Scenario Development:

Create a realistic scenario that would prompt immediate action, such as an urgent software update, a detected anomaly in the ML model, or a critical security alert.

Email Template Design:

Design an email template that mimics the organization's communication style, including logos, signatures, and formatting.

Malicious Link/Attachment:

Include a link or attachment that appears to be a legitimate request for information or action. Ensure that the link directs to the phishing website, or the attachment contains malware designed to capture credentials.

1. Setting up a phishing website

Develop a fake website that mimics the legitimate portal used by the stakeholders.

Steps:

Website Cloning:

Clone the legitimate login page or create a new page that closely resembles the official portal. Pay attention to details such as logos, colour schemes, and text formatting.

Domain Selection:

Register a domain name like the official one (e.g., using a slight misspelling or additional characters) to avoid suspicion.

Functionality Implementation:

Ensure the phishing website captures login credentials and any other input data, storing it securely for later use.

1. Email Delivery   
   Send the phishing email to the target list.

Steps:

Email Service Selection:

Use a reputable email delivery service to avoid spam filters and increase the chances of the email being delivered to the inbox.

Timing:

Send the email during business hours when the targets are likely to check their emails. Timing can increase the likelihood of the email being read and acted upon.

Personalization:

Personalize the email content by addressing the recipient by name and referencing specific details relevant to their role or the ML model.

1. Harvesting Credentials

Capture the login credentials and other sensitive information entered by the targets.

Steps:

Monitor the Phishing Website:

Set up monitoring on the phishing website to log incoming credentials and other data entered by the victims.

Data Storage:

Store the captured data securely for analysis and further exploitation.

1. Post-Exploitation

Use the harvested credentials to gain unauthorized access to the ML model and its data.

Steps:

Login Attempt:

Attempt to log into the legitimate server using the stolen credentials.

System Exploration:

Once access is gained, explore the system for additional vulnerabilities or opportunities to extract more sensitive data.

Maintain Access:

Create backdoors or other persistent mechanisms to maintain access to the system even if the credentials are later changed.

1. Expected Outcome
2. Credentials Compromised: Login details (usernames and passwords) of stakeholders will be captured.
3. Unauthorized Access: Unauthorized access to the ML model and its data sources.
4. Data Exfiltration: Extraction of valuable data for further use or sale on the dark web.
5. Increased Awareness: If detected, the organization may increase security measures, providing valuable insight into their incident response protocols.

By meticulously gathering and analysing this information, attackers can craft highly targeted and convincing phishing campaigns that are more likely to succeed in tricking stakeholders into divulging sensitive information.

Second Attack:

Malvertising (Malicious Advertising):

Malvertising involves injecting malicious advertisements into legitimate online advertising networks and websites. These ads can deliver malware or redirect users to phishing websites. In this attack, I will use malvertising to target users within the organization hosting the ML model, exploiting their exposure to online ads to deliver a payload that facilitates unauthorized access.

Assumptions:

Employees of the target organization frequently browse the web and interact with online ads.

The target organization does not have robust ad-blocking or web filtering solutions in place.

The target organization allows external ad content to be displayed within their network.

Attack Plan:

1. Ad Network Infiltration

Gain access to a legitimate online ad network to serve malicious ads.

Steps:

Ad Network Research:

Identify ad networks commonly used by websites frequented by the target organization’s employees.

Account Creation:

Create accounts on these ad networks, posing as a legitimate advertiser.

Campaign Setup:

Design an ad campaign with attractive and relevant content to increase the likelihood of clicks.

1. Malicious Ad Creation

Develop ads that appear benign but still stay hidden in the malicious code.

Steps:

Design Attractive Ads:

Create visually appealing ads related to popular topics, products, or services.

Embed Malicious Code:

Use techniques such as steganography to hide malicious scripts within the ad images or use JavaScript to redirect users to a malicious website.

1. Payload Development

Develop a payload that executes once the ad is clicked or viewed.

Steps:

Malware Development:

Develop or customize malware designed to exploit browser vulnerabilities or perform drive-by downloads.

Phishing Website Setup:

Alternatively, create a fake login page or data capture form that the ad redirects to.

Testing:

Ensure the payload bypasses common security measures and is difficult to detect.

1. Targeting and Distribution

Ensure the malicious ads are displayed to the intended targets.

Steps:

Geotargeting:

Use the ad network’s targeting options to focus on regions where the target organization operates.

Behavioural Targeting:

Leverage behavioural data to display ads to users with specific browsing patterns typical of employees in the target organization.

Budget Allocation:

Allocate the ad budget to maximize visibility and engagement.

1. Execution and Monitoring

Deploy the ads and monitor for interactions.

Steps:

Ad Deployment:

Launch the ad campaign on the chosen ad networks.

Monitoring Tools:

Use analytics tools provided by the ad network to monitor ad performance and interactions.

Payload Activation:

Track when and where the payload is activated, gathering data on compromised systems.

1. Post-Exploitation

Use the compromised systems to gain further access to the target organization’s network and data.

Steps:

Initial Access: Utilize the initial foothold gained through malware or captured credentials.

Privilege Escalation:

Escalate privileges to gain broader access within the network. Lateral Movement: Move laterally across the network to identify and compromise additional systems.

Data Exfiltration:

Extract sensitive data, focusing on credentials, internal documents, and access to the ML model.

1. Expected Outcome

Initial Compromise:

Compromise of user systems within the target organization through malicious ads.

Credential Harvesting:

Collection of login credentials if the payload involves a phishing website.

Malware Deployment:

Installation of malware on user systems, providing a backdoor for further access.

Network Access:

Gaining a foothold within the target organization’s network, allowing for further exploitation and data exfiltration.

By planning and executing this malvertising attack, attackers can potentially bypass traditional security measures and gain access to sensitive information within the target organization.

Third Attack:

Attack Plan: Credential Stuffing Attack

A credential stuffing attack leverages previously breached username and password pairs to gain unauthorized access to user accounts. This attack exploits the common user practice of reusing passwords across multiple sites. By automating login attempts using these credentials, attackers can gain access to multiple accounts, including those within the target organization.

Assumptions:

The target organization's users reuse passwords across different services. Previously breached credential databases are accessible to the attacker. The target organization's login system does not have robust defences against credential stuffing, such as multi-factor authentication (MFA) or advanced rate-limiting.

Attack Plan:

1. Reconnaissance

Identify user accounts and gather relevant breached credentials.

Steps:

Breach Database Access:

Obtain access to databases of breached credentials from forums, dark web markets, or other sources.

Employee Email Collection:

Collect a list of email addresses associated with the target organization. This can be done through LinkedIn, corporate websites, and email harvesting tools.

Credential Matching:

Cross-reference the collected email addresses with the breached credential databases to identify matching username-password pairs.

1. Tool Preparation

Set up tools and infrastructure needed to execute the credential stuffing attack.

Steps:

Credential Stuffing Tools:

Use tools like Sentry MBA, Snipr, or custom scripts to automate login attempts.

Proxy Setup:

Configure a pool of proxies to distribute login attempts, avoiding IP-based rate limiting and detection.

Login Page Analysis:

Analyse the target organization's login page to understand the required parameters and response behaviours.

1. Attack Execution

Perform automated login attempts using the collected credentials.

Steps:

Script Configuration:

Configure the credential stuffing tool with the target login URL, necessary parameters, and the list of credentials.

Proxy Integration:

Integrate the proxy list with the tool to ensure login attempts come from different IP addresses.

Rate Limiting Awareness:

Implement timing controls to avoid triggering account lockouts or rate limiting. Execution: Run the tool to perform automated login attempts

1. Monitoring and Harvesting

Monitor the success of login attempts and harvest valid credentials.

Steps:

Success Logging:

Log successful login attempts and store the valid credentials securely.

Session Hijacking:

Capture session tokens or cookies, if possible, to maintain access without triggering alerts.

Credential Verification:

Manually verify a subset of the successful logins to ensure the credentials are still valid.

1. Post-Exploitation

Use the harvested credentials to further penetrate the target organization's systems.

Steps:

Privileged Account Access:

Identify and prioritize access to privileged accounts, such as administrators or executives.

Lateral Movement:

Use the compromised accounts to explore the internal network and gain additional access.

Data Exfiltration:

Extract sensitive data, focusing on critical business information, customer data, and intellectual property.

Persistent Access:

Establish backdoors or persistent access mechanisms to maintain long-term access to the compromised accounts.

1. Expected Output

Account Compromise:

Unauthorized access to multiple user accounts within the target organization.

Privileged Access:

Access to high-value accounts with administrative or sensitive data privileges.

Data Exfiltration:

Extraction of valuable data, including internal documents, customer information, and proprietary data.

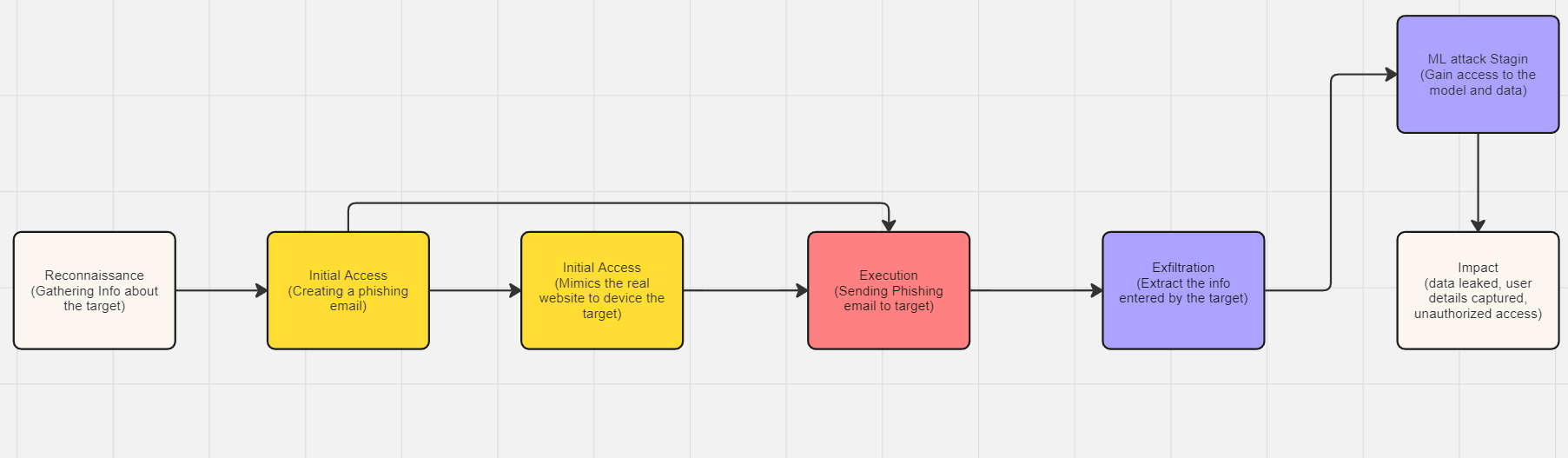
Network Penetration:

Broader access to the organization's internal network, potentially compromising additional systems.

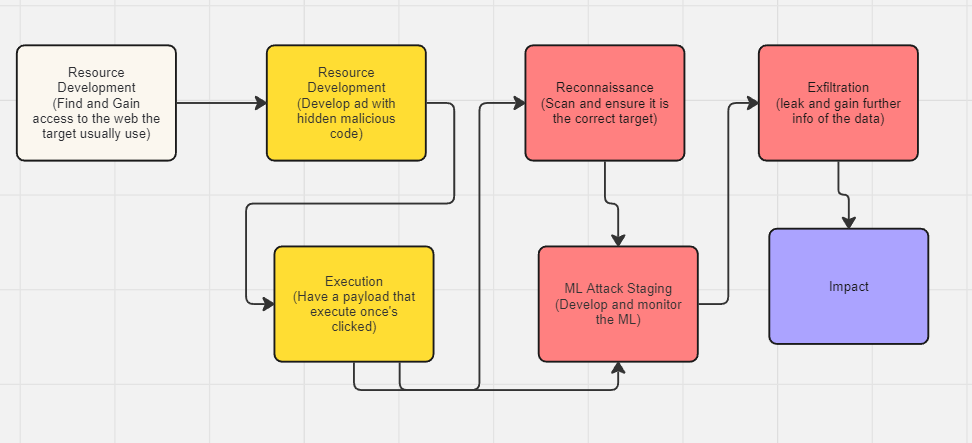
By planning and executing a credential stuffing attack, attackers can exploit weak password practices and insufficient security measures to gain unauthorized access to user accounts and sensitive data within the target organization.

Mapping to MITRE ATLAS framework

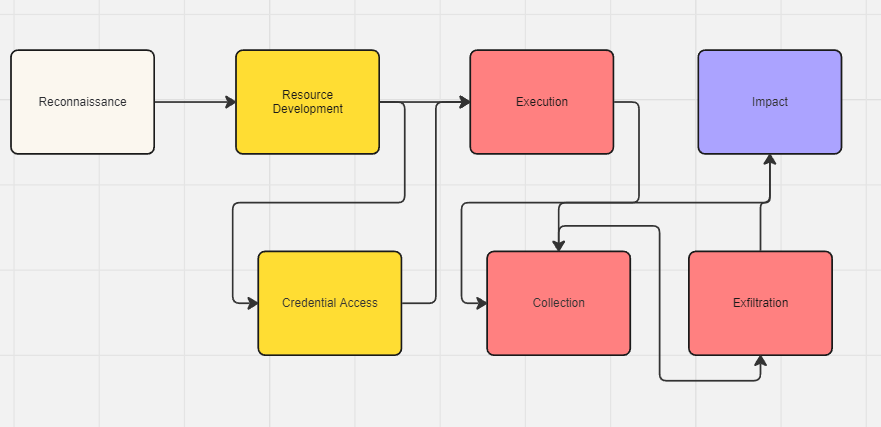
First Attack:



Second Attack:



Third Attack:



Develop a defence strategy plan.

This defence strategy plan aims to protect against the phishing, malvertising, and credential stuffing attacks. It includes technical measures, administrative policies, and user awareness initiatives to ensure a comprehensive security posture.

1. Technical Measures

* 1. Network Security

Firewalls and Intrusion Detection Systems (IDS)

* + 1. Implementation: Deploy advanced firewalls and IDS to monitor network traffic and block suspicious activities. Use Intrusion Prevention Systems (IPS) to automatically respond to detected threats.
    2. Maintenance: Regularly update firewall rules and IDS/IPS signatures to recognize the latest threats.

Web Filtering and Ad-Blocking

* + 1. Web Filtering: Implement web filtering solutions to block access to known malicious websites and ad networks.
    2. Ad-Blocking: Deploy ad-blocking software across all user systems to prevent exposure to malicious advertisements (malvertising).
  1. Email Security

Email Filtering

* + 1. Spam Filters: Use email filtering solutions to detect and block phishing emails. Implement Advanced Threat Protection (ATP) to scan email attachments and links for malware.
    2. Authentication Protocols: Use DMARC, SPF, and DKIM protocols to authenticate incoming emails and prevent email spoofing.

Anti-Phishing Tools

1. Detection Tools: Deploy anti-phishing tools that detect and warn users about potential phishing attempts.
   * 1. Sandboxing: Use sandbox environments to open and analyze email attachments and links safely. C.
   1. Authentication and Access Control

Multi-Factor Authentication (MFA)

* + 1. Requirement: Enforce MFA for all user accounts, especially for privileged accounts. MFA reduces the risk of account compromise through credential stuffing and phishing.

Password Policies

* + 1. Complexity and Rotation: Implement strong password policies that require complex passwords and regular changes. Encourage the use of password managers to store and manage passwords securely.

Single Sign-On (SSO)

* + 1. Integration: Use SSO solutions with MFA to simplify and secure user authentication processes, reducing the number of credentials that need to be managed.
  1. System and Application Security

Patch Management

* + 1. Routine Updates: Regularly update and patch all systems, applications, and software to mitigate vulnerabilities that could be exploited.

Application Whitelisting

* + 1. Controlled Execution: Implement application whitelisting to ensure that only approved applications can run on the network.
  1. Network Segmentation

Critical Systems Segmentation

* + 1. Isolate Critical Systems: Segment critical systems and sensitive data into isolated network segments to limit access and reduce the impact of potential breaches.

Least Privilege Principal

* + 1. Access Control: Apply the principle of least privilege to restrict user access to only what is necessary for their role, minimizing the risk of privilege escalation.

1. Administrative Measures
   1. Incident Response Plan

Development and Maintenance

* + 1. Comprehensive Plan: Develop a comprehensive incident response plan (IRP) to quickly detect, respond to, and recover from security incidents.
    2. Regular Drills: Conduct regular drills and updates to the IRP to ensure its effectiveness and the readiness of the response team.

Incident Response Team

1. Establishment: Form an incident response team with clearly defined roles and responsibilities to handle security incidents promptly.
   1. Security Policies and Procedures

Policy Development

* + 1. Security Policies: Develop and enforce security policies covering password management, email usage, internet browsing, and data handling to establish clear guidelines for employees.
    2. Audits and Assessments Regular Reviews: Conduct regular security audits and risk assessments to identify and address vulnerabilities in the organization’s infrastructure and practices.

1. User Awareness and Training
   1. Security Awareness Training

Regular Training Sessions

* + 1. Education: Conduct regular security awareness training for all employees to educate them on the latest phishing techniques, safe email practices, and how to recognize suspicious activities.

Phishing Simulations

1. Testing and Improvement: Regularly perform phishing simulations to test and improve employee awareness and response to phishing attacks.
   1. Reporting Mechanisms

Establish Channels

* + 1. Clear Reporting: Create clear channels for employees to report suspected phishing emails or security incidents promptly.

Encourage Reporting

1. Culture of Reporting: Encourage a culture of proactive reporting without fear of reprimand to ensure that potential threats are addressed swiftly.

Summary of Defence Measures

Technical Measures:

* Advanced firewalls and IDS/IPS
* Web filtering and ad-blocking
* Email filtering and anti-phishing tools
* MFA and strong password policies
* Regular patch management
* Network segmentation and least privilege principle

Administrative Measures:

* Comprehensive incident response plan and team
* Detailed security policies and regular audits

User Awareness and Training:

* Regular security awareness training and phishing simulations
* Clear reporting mechanisms and encouragement for reporting

By implementing this multi-layered defence strategy, the organization can significantly reduce the risk and impact of phishing, malvertising, credential stuffing, and other cyber-attacks described in Section B.

Conclusion

The rapidly evolving landscape of cyber threats necessitates a comprehensive and multi-faceted defence strategy. In this project, we explored various attack vectors, including phishing, malvertising, and credential stuffing, each posing significant risks to organizational security. Through a detailed examination of these attack strategies, we identified the critical steps and measures attackers might take to compromise systems and data.

To counter these threats, we developed a robust defence strategy plan that integrates technical, administrative, and user-awareness measures. By deploying advanced firewalls, intrusion detection systems, and email filtering solutions, we can significantly mitigate the risk of phishing and malvertising attacks. Implementing multi-factor authentication (MFA) and enforcing strong password policies are essential steps in safeguarding against credential stuffing and unauthorized access.

Moreover, the importance of regular security training and awareness programs cannot be overstated. Educating employees about the latest phishing techniques and safe practices ensures that they are the first line of defence against social engineering attacks. Coupled with clear reporting mechanisms, this fosters a culture of vigilance and proactive security within the organization.

Administrative measures, including a well-defined incident response plan and regular security audits, ensure that the organization is prepared to respond swiftly and effectively to any security incident. These measures, combined with technical defences and user training, create a layered security posture that is resilient against a wide range of cyber threats.

In conclusion, the strategies outlined in this plan are designed to provide a comprehensive defence against the sophisticated and varied attacks detailed in our analysis. By implementing these measures, organizations can significantly reduce their vulnerability to cyber-attacks, protect their critical assets, and maintain the trust of their stakeholders. The proactive approach detailed here is essential for staying ahead of potential threats and ensuring long-term security and resilience in an increasingly digital world.

References

<https://atlas.mitre.org/matrices/ATLAS>

<https://www.kaggle.com/code/bhaumikreddy/phishing-website-detection>

<https://www.kaggle.com/code/kragg033/phishing-detection>

<https://www.kaggle.com/code/eswarchandt/website-phishing>

<https://www.kaggle.com/datasets/eswarchandt/phishing-website-detector/code>

<https://www.proofpoint.com/au/resources/e-books/definitive-email-security-strategy-guide?utm_source=google&utm_medium=cpc&gad_source=1&gclid=Cj0KCQjw6uWyBhD1ARIsAIMcADrW1YB6eklaw3szLa1dO33sAqYPzuZkcYJKuxAzUmnA6TgqRo0Ted0aAvkbEALw_wcB&gclsrc=aw.ds>

<https://www.agari.com/solutions/cloud-email-security?code=cmp-0000013224&ls=717710011&hsa_acc=4848471385&hsa_cam=19852257748&hsa_grp=149096357458&hsa_ad=654689643103&hsa_src=g&hsa_tgt=kwd-358669484875&hsa_kw=phishing%20email%20protection&hsa_mt=p&hsa_net=adwords&hsa_ver=3&gad_source=1&gclid=Cj0KCQjw6uWyBhD1ARIsAIMcADqh5IT-m2t-1fccKK0ncAF9fs0vniPUAg095XfBP25_wHptSPIHrF4aAuoiEALw_wcB&utm_source=google&utm_medium=ppc&utm_campaign=&utm_medium=ppc&utm_campaign=&utm_term=phishing%20email%20protection>

<https://www.yubico.com/phishing-attacks-are-getting-smarter/?gad_source=1&gclid=Cj0KCQjw6uWyBhD1ARIsAIMcADpKYW6hSyR7znsugiMNAmdrWCrThbFt7pXsyFEz2R-tcuCUNxvBMjgaAi4zEALw_wcB>

<https://umbrella.cisco.com/info/phishing-for-dummies-ebook-discover-real-risks-of-phishing?utm_medium=search-paid&utm_source=google&utm_campaign=UMB_APJC_SEA_EN_GS_Nonbrand_Threats_T1&utm_content=UMB-FY24-Q1-Phishing-for-Dummies&_bt=688693658835&_bk=phishing&_bm=p&_bn=g&_bg=161647949921&gad_source=1&gclid=Cj0KCQjw6uWyBhD1ARIsAIMcADr8OfNfFsFh15mo__nOfmj10yNGg08rfUtNASpa4qCCPUMRUib7CugaAkF0EALw_wcB>

<https://www.fortinet.com/resources/cyberglossary/types-of-phishing-attacks>

<https://www.itgovernance.eu/blog/en/the-5-most-common-types-of-phishing-attack>