THREAT FEED AGGREGATOR (TFA)

Name: Isaac Isalwa

Date of Submission: April 5th 2024

Acknowledgment

I would like to express my heartfelt gratitude to Madam Sekento for her invaluable guidance and mentorship throughout the development of this proposal. Her expertise and support have been pivotal in shaping the clarity and depth of the project. Additionally, I extend my thanks to my fellow IT friends from the Information Communication Technology (ICT) department for their collaborative spirit and contributions, which have enriched the proposal with diverse perspectives and insights.

Abstract

The Ethical Threat Feed Aggregator (TFA) project aims to develop an integrated platform for aggregating, filtering, and prioritizing threat intelligence data. This proposal outlines the systematic approach, utilizing the Waterfall methodology, to design, develop, and deploy the TFA. The project encompasses phases such as system analysis, design, development, verification, validation, integration, deployment, operation, and maintenance. By leveraging industry-standard tools, qualitative and quantitative research methods, and iterative development practices, the TFA project strives to deliver a robust and user-friendly solution to enhance threat detection and response capabilities.

Declaration

I, Isaac Isalwa, declare that the content of this proposal titled "Ethical Threat Feed Aggregator (TFA): Design and Development Using Waterfall Methodology" is original and has been prepared based on my research, analysis, and understanding of the subject matter. Any references or sources used in this proposal have been appropriately cited and credited.

I acknowledge that this proposal is submitted as part of academic evaluation, and is intended solely for the intended audience's review and consideration. I take full responsibility for the accuracy, completeness, and integrity of the information presented herein.

Name: Isaac Isalwa

Organization: Umma University

Date : 4th April 2024

Table of Contents

Chapter One Introduction

1.1 Background of the study

The cybersecurity landscape is awash in data. Threat feeds, advisories, vulnerability reports, and news articles stream forth from countless sources, each offering valuable insights but overwhelming even the most seasoned security teams. Sifting through this fragmented landscape to identify critical threats and prioritize risks proves a colossal task.

The rise of Security Information and Event Management (SIEM) systems offered a glimmer of hope, aiming to centralize and analyze security data. However, these systems often struggle to integrate disparate threat feeds efficiently, leaving gaps in coverage and requiring time-consuming manual configuration.

Recognizing this challenge, this project proposes the development of an ethical threat feed integrator.

1.2 Problem Statement

Security professionals today face an overwhelming deluge of threat intelligence information from disparate sources. Hundreds of threat feeds churn out data 24/7, covering diverse threat landscapes and attack vectors. Manually sifting through and analyzing this vast amount of data is impractical, inefficient, and often ineffective. As a result, security teams struggle to:

1. Filter through noise and identify high-priority threats: The sheer volume of alerts leads to alert fatigue and missed critical threats.
2. Gain actionable insights: Raw data from various feeds requires tedious analysis and correlation to extract meaningful intelligence.
3. Prioritize effectively: Limited resources necessitate efficient allocation of attention to the most relevant threats facing their organization.
4. Maintain situational awareness: Keeping up with the constantly evolving threat landscape is a significant challenge.

This information overload poses a significant risk to organizational security, increasing the likelihood of successful attacks, data breaches, and financial losses. Traditional security tools often lack the capabilities to effectively aggregate, analyze, and prioritize threat data, leaving security teams scrambling to keep pace

1.3 General Objectives

To connect to a wide range of ethical and authorized threat feeds, eliminating the need for manual integration for individual platforms.

1.4 Primary objectives

To aggregate and prioritize threat data from diverse sources, reducing alert fatigue and improving security team response times.

To employ advanced data analysis and machine learning to provide contextualized and actionable threat insights for targeted decision-making.

To enhance the overall situational awareness and proactive security posture of organizations compared to traditional SIEM solutions.

1.5 Research Questions

1.Can an ethical threat feed integrator effectively aggregate and prioritize threat data from diverse sources, reducing alert fatigue and improving security team response times?

2.How can advanced data analysis and machine learning be leveraged within an ethical framework to provide contextualized and actionable threat insights for targeted decision-making?

3.To what extent can integrating ethical and authorized threat feeds enhance the overall situational awareness and proactive security posture of organizations compared to traditional SIEM solutions?

1.6 Scope of the Study:

The scope of a study explains the extent to which the research area will be explored in the work and specifies the parameters within the study will be operating.[Editage insights]

This project focuses on the development and evaluation of an ethical threat feed integrator designed to address the challenges of fragmented threat intelligence landscapes. Its scope encompasses:

1.6.1 Data Sources:

Integration with a curated selection of ethical and authorized threat feeds. This includes official government advisories, vulnerability reports, security forums, and other approved sources, adhering to responsible data collection practices.

1.6.2 Functionality:

1. Seamless aggregation: Connects to multiple threat feeds, eliminating the need for individual integrations and manual configuration.
2. Advanced data processing: Employs techniques like entity recognition, correlation analysis, and machine learning to extract valuable insights, identify threat actors and campaigns, and prioritize the most relevant threats.
3. Actionable intelligence: Presents data in a user-friendly format with contextual information and actionable recommendations, empowering security teams to make informed decisions quickly.
4. Customization: Allows user-defined filtering and prioritization based on specific security needs and industry verticals.

1.6.3 Evaluation:

Performance testing: Measures the system's efficiency in data aggregation, processing, and threat prioritization.

User experience evaluation: Assesses the usability and effectiveness of the user interface and information presentation.

Security posture improvement: Evaluates the impact of the system on reducing alert fatigue, improving response times, and enhancing overall security posture through real-world testing with pilot organizations.

1.6.4 Expected Outcomes:

Reduced alert fatigue: Focus on relevant and prioritized threats, enabling faster and more effective response.
Improved situational awareness: Comprehensive view of the threat landscape for proactive security posture enhancement.
Enhanced decision-making: Actionable insights empower security teams to prioritize critical threats and mitigate risks efficacement.

1.7 Justification of the Study:

Justfication of the study is the rationale for one's research is the justification for undertaking a given study.[CW Authors]. The escalating volume and complexity of cyber threats overwhelm security teams with a constant barrage of alerts. Traditional SIEM solutions struggle to keep pace, lacking efficient integration with diverse threat feeds and requiring time-consuming manual configuration. This project addresses this critical challenge by developing an ethical threat feed integrator offering:

Unified Platform: Streamlined access to a wide range of ethically sourced threat feeds, eliminating the need for individual integrations.
Enhanced Data Analysis: Utilizes advanced techniques to correlate data, identify critical connections, and present a holistic view of the threat landscape.
Intelligent Prioritization: Employs machine learning and expert-curated rules to prioritize threats based on severity, exploitability, and relevance to specific environments.

Actionable Insights: Delivers clear and concise information, empowering security teams with contextual understanding for rapid decision-making.

This project prioritizes ethical data collection practices, ensuring compliance with legal and privacy regulations. By leveraging publicly available information, collaborating with trusted security organizations, and respecting intellectual property rights, with an aim to contribute to a responsible and secure cybersecurity ecosystem.

1.8 Limitations of the Study:

The limitations of a study are its flaws or shortcomings which could be the result of unavailability of resources, small sample size, flawed methodology, etc.[English Editing Services]

This project focuses on integrating ethically sourced threat feeds, limiting the scope of information compared to unrestricted or potentially compromising sources.

The effectiveness of the prioritization engine depends on the quality and comprehensiveness of integrated threat feeds.

Extensive user testing and real-world evaluation are crucial to refine the system's performance and optimize its impact.

Despite these limitations, this project offers a significant contribution by addressing the critical need for efficient and ethical threat feed integration. By empowering security teams with actionable insights and improved situational awareness, the proposed threat feed integrator holds the potential to enhance organizational security posture and contribute to a safer digital environment.

1.8 CONCLUSION

Join us in creating a cybersecurity environment where information overload is no longer a barrier, and threats are identified and mitigated before they can cause harm. This ethical threat feed integrator is a crucial step towards achieving that goal.

This project contributes to a responsible and secure cybersecurity ecosystem by providing an ethical threat feed integrator that empowers security teams to navigate the ever-evolving threat landscape effectively.

Chapter 2: Literature Review

2.1. Introduction

This literature review chapter delves into the critical role of threat intelligence (TI) in safeguarding organizations within the dynamic cybersecurity landscape. The ever-evolving nature of cyber threats necessitates proactive measures, and TI serves as a cornerstone for effective anticipation, detection, and response strategies.

However, traditional manual approaches to gathering and analyzing TI data are hampered by the sheer volume of information. This review addresses this challenge by exploring how automated solutions, specifically Threat Feed Aggregators (TFAs), are revolutionizing threat intelligence.

2.2 Review of Existing Literature on TFAs

Several studies have explored the functionalities and benefits of TFAs:

- Yu et al. (2023) discuss the potential challenges associated with TFAs, including information overload and the need for skilled personnel to interpret and utilize the aggregated data effectively.

- Gustavo González-Granadillo * , Susana González-Zarzosa and Rodrigo Diaz on SIEMs amalysis paper empahisize that one step forward for cyber threat detection, mitigation, and prevention is to consider AI/ML in SOAR solutions which would be ideally integrated in SIEM platforms. AI/ML- powered defense systems are able to analyze large amount of data and identify suspicious patterns in real-time (or near real-time). The main targets for AI/ML applications include intrusion detection (network-based attacks), phishing and spam (emails), threat detection and characterization, and user behavioral analytics [4].
- Adaptive security intergration paper on Adaptive Security Intergration, highlighted that the next generation of SIEMs must integrate evolved and adaptive SOAR solutions with advanced capabilities that enable dynamic interactions at all phases of the incident workflow to quickly

deal with existing and emerging threats [127,128]. Examples of enriched adaptive SOAR include the NextGuard Adaptive security Operations suite from Nokia NextGuard[5].

- In the paper ‘From Cyber Security Information Sharing to Threat Management’, the authors emphasize that threat intelligence refers to more complex cyber threat information that has been acquired or inferred through the analysis of existing information. Information such as the different malware families used over time with an attack or the network of threat actors involved in an attack, is valuable information and can be vital to understanding and predicting attacks, threat developments, as well as informing law enforcement investigations. This information is also actionable, but on a longer time scale. Moreover, it requires action and decision-making at the human level. They also highlighted the need for effective intelligence management platforms to facilitate the generation, refinement, and vetting of data, post sharing. [6]They expounded on the key challenges that exist include in design of such a system which stated that the system: working with multiple intelligence sources, combining and enriching data for greater intelligence, determining intelligence relevance based on technical constructs, and organizational input, delivery into organizational workflows and into technological products, would be a challenge.(Brown, Gommer, Serrano 2012)

2.3 Existing Landscape of TFAs

The TFA landscape is diverse, offering a range of solutions with varying functionalities. Some prominent examples include:

2.3.1 Open-source:

AlienVault Open Threat Exchange (OTX) .The AlienVault Open Threat Exchange (OTX) is a valuable resource for cybersecurity professionals. This free, community-driven platform fosters collaboration by allowing researchers and professionals to share the latest threat indicators. OTX provides a rich collection of Indicators of Compromise (IOCs) such as malicious URLs, IP addresses, and file hashes. This centralized repository keeps security teams informed about

emerging threats. However, OTX data often lacks context about the specific threats or attacks the IOCs are associated with. Additionally, the accuracy of community-submitted information can vary, and OTX's functionality is primarily focused on data sharing with limited analytical capabilities.(as discussed in 2024)[2].

Threat feed aggregators can address these limitations of OTX by enriching data with additional context such as threat actor information or associated vulnerabilities. Aggregators can also employ validation techniques to ensure the accuracy of IOCs and prioritize threats based on an organization's specific needs. By offering advanced threat analysis tools, aggregators can transform OTX data into more actionable threat intelligence, improving an organization's ability to proactively defend against cyberattacks.

2.3.2Commercial:

Commercial threat intelligence platforms like Anomali ThreatStream, Palo Alto Networks AutoFocus, and McAfee Threat Intelligence Exchange (TIE) offer valuable features but come with limitations. While they provide extensive threat data feeds, often with rich context, these platforms typically require paid subscriptions, limiting accessibility for some organizations. Additionally, vendor lock-in and limited customization can restrict flexibility. Security teams might find themselves locked into a single platform and unable to fully tailor threat intelligence to their specific needs.(Basumalik 2022)

2.4 Gaps in Existing Threat Intelligence Approaches

Traditional methods of threat intelligence gathering and analysis face limitations:

- **Manual Processes:** Manually collecting and analyzing vast amounts of threat data is time-consuming and inefficient. Human error can also lead to inaccurate or incomplete information.
- **Limited Scope:** Organizations may rely on a single source or a limited range of sources for threat intelligence, potentially missing critical threat indicators.
- **Lack of Context:** Raw threat data often lacks context about the specific threats or attacks, making it difficult to assess the true risk.
- **Inaccurate Information:** Community-driven platforms like OTX might have inconsistencies in data accuracy due to reliance on user-submitted information.
- **Limited Analysis Capabilities:** Basic data sharing without advanced threat analysis tools hinders proactive defense strategies.

2.5 How TFAs Bridge the Gap:

Threat Feed Aggregators (TFAs) can effectively address these limitations of OTX. By aggregating data from various sources, including OTX, TFAs provide a more comprehensive view of the threat landscape. They can enrich OTX data with additional context, such as threat actor information or associated vulnerabilities, giving security teams a clearer picture of the potential risks.

Furthermore, TFAs can employ validation techniques to ensure the accuracy of IOCs. This helps to mitigate the risk of acting on false positives and wasting valuable resources.

Most importantly, TFAs go beyond simple data sharing by offering advanced threat analysis tools. These tools can automate tasks like threat identification, prioritization, and investigation. This frees up security teams to focus on more strategic initiatives and allows them to make informed decisions based on a deeper understanding of the threats they face.

2.6 Functionalities and Benefits of TFAs

TFAs offer several key functionalities:

- Centralized Collection: TFAs aggregate data from various sources, including commercial feeds, open-source intelligence (OSINT), threat research reports, and government advisories. This eliminates the need to manually access and analyze data from multiple sources, saving time and effort.
- Data Filtering and Enrichment: TFAs filter raw data, removing irrelevant information and enriching it with additional context, such as threat actor attribution, associated vulnerabilities, and mitigation strategies.
- Normalization and Standardization: TFAs often normalize data into standardized formats, such as Structured Threat Information Expression (STIX) or TAXII, enabling seamless integration with security information and event management (SIEM) systems and other security tools.
- Alert Generation and Prioritization: TFAs can generate alerts based on user-defined criteria, allowing security teams to prioritize and focus on the most relevant threats.

2.7 Future Directions of TFAs

TFAs are expected to evolve alongside the cyber threat landscape. Key future directions include:

- Integration with Artificial Intelligence (AI) and Machine Learning (ML): AI/ML can enhance data analysis, automate threat detection, and personalize threat intelligence based on an organization's specific needs.
- By using machine learning algorithms to analyze data, organizations can detect potential threats before they occur and stay ahead of the game[9]. SIEM solutions that utilize predictive analytics offer several benefits over traditional SIEM, including early detection of threats, better accuracy, increased efficiency and scalability.(Joy Wang 2023)
- Enhanced Data Sharing and Collaboration: Greater collaboration between organizations and threat intelligence communities can enrich TFA data and improve overall threat visibility.
- Standardization and Interoperability: Standardization across TFA platforms will facilitate seamless information exchange and improve the overall effectiveness of the technology.

2.8 Conclusion

TFAs play a crucial role in enhancing cyber threat detection and mitigation by streamlining the collection, analysis, and utilization of TI. As the cyber threat landscape continues to evolve, TFAs are expected to become increasingly sophisticated, incorporating AI/ML and fostering collaboration to empower security teams in the fight against cyberattacks.

Chapter 3: Agile Development for Ethical Threat Feed Aggregator (TFA)

3.1. Introduction

This document proposes the adoption of an Agile development methodology for the Ethical Threat Feed Aggregator (TFA) project. This approach prioritizes flexibility, collaboration, and iterative delivery, ensuring a responsive and user-centric solution that effectively addresses cybersecurity threats.

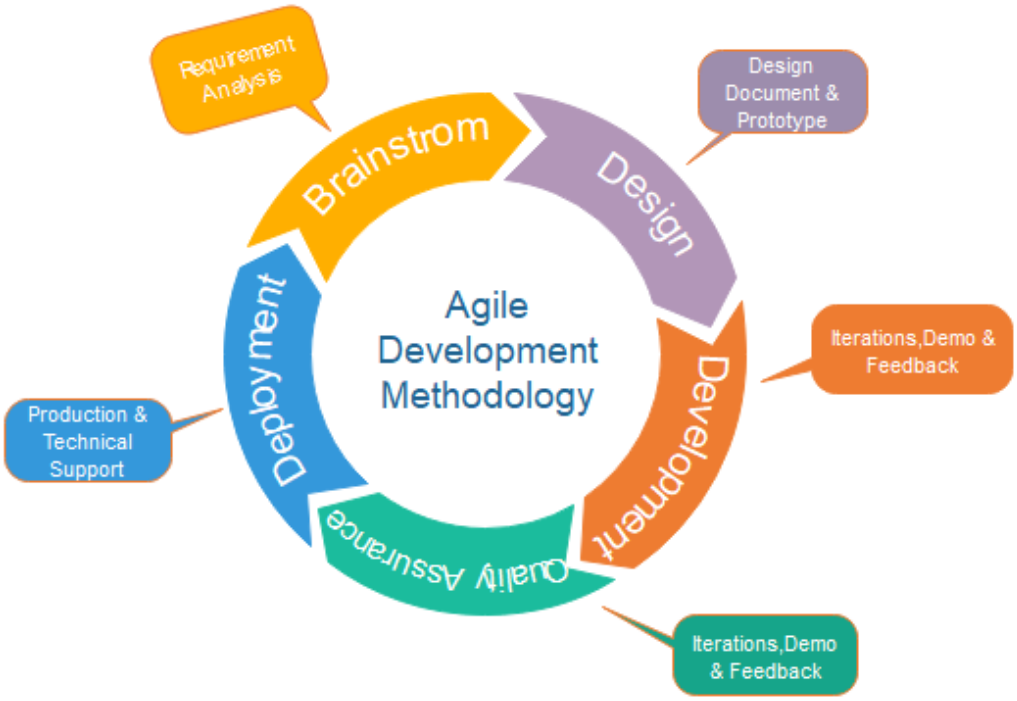

Fig. Agile Model

3.2. Benefits of Agile Development

- Adaptability:Agile allows adaptation to evolving threats, data sources, and changing user requirements throughout the project lifecycle.
- Continuous Improvement: Incremental feature delivery facilitates continuous feedback and refinement of the TFA system.
- Stakeholder Satisfaction: Close collaboration with stakeholders fosters transparency and ensures the final product aligns with their expectations.
- Faster Time to Market: Delivering features in short sprints allows for early deployment and user feedback, accelerating project completion.
- Reduced Risk: Agile promotes iterative testing and early detection of issues, minimizing project risks.

3. Proposed Agile Development Processs

3.3 System Analysis and Requirements Gathering

A comprehensive understanding of the project scope will be established through close collaboration with stakeholders to identify their needs.

Here is a breakdown of the activities:

- Stakeholder Identification: All stakeholder to be invlolved in the project will be identified, the stakeholder may include developers, security analysis, end-users and project managers.
- Data gatheringL Cnducting intervies, workshops , and online surveys to gather stakeholders needs, ecpectations, and painpoints.
- User story mapping : workshops will be conducted to visualize user workflows and define clear, measurable acceptance criteria for each feature.
- Product Backlog Creation: The gathered requirements will be prioritized and documented in a product backlog to guide development efforts.

3.4 Research Design

To ensure that the TFI effectively meets user needs, a mixed research design combining qualitative and quantitative methods will be employed:

Qualitative Methods:

- Conduct in-depth interviews with security professionals to gain insights into their current challenges and desired functionalities for threat intelligence management.
- Analyze existing user documentation and support forums to identify common pain points and user needs.

Quantitative Methods:

- Develop and distribute online surveys to a targeted sample population of security professionals.
- Utilize survey data to gain a broader understanding of user needs and preferences regarding threat intelligence integration and prioritization.

3.4.3 Research Location

The research location will depend on the chosen research methods:

- Interviews: These can be conducted in-person at participants' workplaces or virtually through video conferencing platforms.
- Online Surveys: These can be distributed electronically using survey tools and shared through email lists, industry forums, or social media groups targeting security professionals.

3.4.4 Target Population

The target population for our research will be security professionals working in IT departments, Security Operations Centers (SOCs), or threat intelligence teams. This focused population will ensure that the gathered information directly reflects the needs of those who will utilize the TFA.

3.4.5 Sample Size Selection

A statistically significant sample size will be determined based on the chosen research method:

1. Interviews: While in-depth interviews provide rich data, a smaller sample size (around 10-15 participants) can suffice due to the qualitative nature of the information gathered.

2. Surveys: For statistically significant results, a larger sample size (potentially hundreds or even thousands) is desirable. Online survey tools often provide sample size calculators to guide the selection process based on desired confidence level and population size estimations.

3.5 System Design

The system design phase focuses on translating the requirements from the previous phase into a technical blueprint. Here is a breakdown of the key activities:

- System Architecture Design: Define the overall system architecture, including hardware,software components, network topology, and data flow. This blueprint ensures scalability, security and performance of the TFA system.
- User Interface (UI) Design: Design an intuitive and user-friendly UI that caters to the needs of different user groups. Utilize design thinking principles, wire-framing tools and prototyping to create mock-ups for user feedback and validation.
- Data Flow Modeling: Development of data flow models to illustrate how data will be collected, processed and visualized within the TFA system. This ensures efficient data handling and avoids bottlenecks

3.6 System Development

The development phase follows an iterative sprint-based approach typically using scrum principles.
Here is a breakdown of the key activities within a sprint:

- Sprint Planning: The development team and stakeholders collaboratively plan the upcoming sprint, selecting features from the backlog to be developed.
- Task Breakdown: Features are broken down into smaller, more manageable tasks assigned to individual development or teams.
- Development and coding: Developer implement the planned features according to the defined requirements and technical specifications. Module development will happen here where modules such as:

    - Data Acquisition Module: Handles secure integration with various ethical threat feeds.
    - Data Processing Module: Cleans, normalizes, and enriches data, potentially utilizing entity recognition and correlation analysis.
    - Threat Prioritization Module: Ranks threats using machine learning and user-defined rules.
    - Alert Generation and Reporting Module: Generates alerts and reports with actionable insights.
    - User Interface (UI) and Reporting Dashboard: Provides a user-friendly interface for interacting with the TFA and visualizing threat intelligence data.
- Continuous Integration(CI): Frequent code commits and automated builds ensure integration and detection of potential issues.

3.7 Deployment and Quality Assurance

Automated deployment pipelines and CI/CD practices will streamline the deployment process with minimal downtime.

Rigorous testing, including unit testing, integration testing, and user acceptance testing (UAT), will be conducted to ensure each iteration meets quality standards, with specific considerations for handling and analyzing interview data.

Continuous monitoring and feedback loops from users will drive defect resolution and feature enhancements.

3.8. Project Management and Communication

Dedicated project management tools will be utilized to track progress, manage resources, and facilitate transparent communication among stakeholders.

Daily stand-up meetings, sprint planning sessions, and regular reviews will foster collaboration and ensure project alignment.

3.8. Conclusion

An Agile development methodology ensures a responsive, user-centric, and continuously improving TFA system. This approach promotes flexibility, collaboration, and rapid delivery, ultimately leading to a more effective and secure solution that empowers organizations to stay ahead of evolving cyber threats.

REFERENCES

1. Top 10 Threat Intelligence Platforms in 2022 https://www.spiceworks.com/it-security/vulnerability-management/articles/best-threat-intelligence-platforms
2. https://cybersecurity.att.com/products/ossim,
3. Akhan, M. S., Koç, C. A., & Gürses, S. M. (2020). A framework for security information and event management (SIEM) systems using threat intelligence feeds. Security and Communication Networks, 13(24), 3809-3822.

4. Chen, Y., Zhang, Z., Kang, X., & Zou, C. (2022). A Survey on Threat Intelligence: From Definitions to Applications. IEEE Communications Surveys & Tutorials, 24(3), 1711-1740. [https://ieeexplore.ieee.org/document/9453402]
5. Yu, Y., Wang, H., & Xu, L. (2023). A Survey on Threat Intelligence Feeds: Characteristics, Applications, and Challenges. IEEE Access, 11, 1136-1150. [https://ieeexplore.ieee.org/
6. Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infranstructure.
7. (https://www.nokia.com/networks/solutions/netguard/adaptive-security-operations/accessed on 7 June 2021), the Splunk adaptive Operations Framework (AOF) (https://www.splunk.com/en_us/solutions/solution-areas/security-and-fraud/adaptive-response/ initiative.html accessed on 7 June 2021), and the Integrate.
8. From Cyber Security Information Sharing to Threat Management(https://dl.acm.org/doi/abs/10.1145/2808128.2808133)
9. Project Management Institute. (2017). A Guide to the Project Management Body of Knowledge (PMBOK Guide) (Sixth Edition). Newtown Square, PA: Project Management Institute. (https://www.pmi.org/pmbok-guide-standards/foundational/pmbok)
10. From https://securityintelligence.com/ .The future of SIEM: Embracing predictive analyticshttps://securityintelligence.com/posts/the-future-of-siem-embracing-predictive-analytics/
11. A Simulation Model for the Waterfall Software Development Life Cycle by Youssef Bassil https://arxiv.org/pdf/1205.6904.pdf
12. Agile Vs. Waterfall Project Management—Which Should You Choose?

    https://www.forbes.com/sites/rachelwells/2023/12/05/agile-vs-waterfall-project-management-which-should-you-choose/
13. Waterfall model https://en.wikipedia.org/wiki/Waterfall_model
14. Everything you need to know about waterfall project management https://asana.com/resources/waterfall-project-management-methodology

15. https://en.wikipedia.org/wiki/Waterfall_model

16. Waterfall Methodology: A Complete Guide Adobe Communications Team 03-18-2022 https://business.adobe.com/blog/basics/waterfall.

17. https://www.google.com/url?sa=i&url=https%3A%2F%2Fkruschecompany.com%2Fwaterfall-software-development-methodology%2F&psig=AOvVaw3Gv8YpWUfJ2LL0DesD0B_W&ust=1712125832036000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCLD5vfTzooUDFQAAAAAdAAAAABAE

18. Waterfall model https://en.wikipedia.org/wiki/Waterfall_model