ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

«Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ» КЕАҚ

Студенттер мен жас ғалымдардың «ǴYLYM JÁNE BILIM - 2024» XIX Халықаралық ғылыми конференциясының БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ XIX Международной научной конференции студентов и молодых ученых «ǴYLYM JÁNE BILIM - 2024»

PROCEEDINGS of the XIX International Scientific Conference for students and young scholars «ǴYLYM JÁNE BILIM - 2024»

2024
Астана

УДК 001
ББК 72
G99

«ǴYLYM JÁNE BILIM – 2024» студенттер мен жас ғалымдардың XIX Халықаралық ғылыми конференциясы = XIX Международная научная конференция студентов и молодых ученых «ǴYLYM JÁNE BILIM – 2024» = The XIX International Scientific Conference for students and young scholars «ǴYLYM JÁNE BILIM – 2024». – Астана: – 7478 б. - қазақша, орысша, ағылшынша.

ISBN 978-601-7697-07-5

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов имолодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 001
ББК 72
G99

ISBN 978-601-7697-07-5

18. Digital Forensics Magazine. – https://www.digitalforensicsmagazine.com/ (Accessed August 25, 2023).

19. Cybercrime Magazine. – https://cybersecurityventures.com/cybercrime-magazine/ (Accessed August 30, 2023).

20. The Sleuth Kit and Autopsy. Open Source Digital Forensics Tools. – https://www.sleuthkit.org/ (Accessed September 5, 2023).

INVESTIGATING ADVANCED DATA SECURITY TECHNIQUES AND ALGORITHMS TO COMBAT EMAIL SPAM

Kurmanbek Sagynysh Nurhatkyzy
sagynyshkurmanbek@gmail.com
Master's student of «Eurasian national university named after L.N. Gumilyov»
«Information security» department, Astana, Kazakhstan
Supervisor - Tokseit D.K.

1. Introduction

The rise of email spam constitutes a continuing concern in the digital age. With the advent of networked networks and the rising reliance on email for communication, the threat posed by spam has expanded enormously. Beyond merely cluttering inboxes, spam emails can hold harmful payloads, such as malware and phishing links, risking the security and privacy of receivers. Moreover, the economic toll of spam is tremendous, with firms investing large resources on filtering and reducing its impacts. Consequently, there is a critical need for resilient and flexible solutions to resist this scourge. By focusing on sophisticated data security approaches and algorithms, this study tries to answer this requirement and contribute to the ongoing efforts to increase email security.

In the ever-evolving landscape of the digital era, email communication acts as a cornerstone, enabling connectivity and facilitating information transmission. However, this crucial medium is continually threatened by the subtle and adaptive nature of spam. The rising volume and sophistication of spam need a continuous research and refining of advanced data security strategies and algorithms. This imperative is underscored by a compendium of scholarly works authored by researchers such as Altulaihan et al. (2023), Mohammed et al. (2019), Rafat et al. (2022), Yasin and Azmi (2023), Karim et al. (2019), Ferreira et al. (2021), and Xia (2020), collectively shedding light on the dynamic challenges associated with email spam and emphasizing the urgent need for innovative security measures.

Altulaihan et al. (2023) establish the groundwork by presenting a complete study of tools and approaches important for examining email security vulnerabilities. Their work acts as a cornerstone for our study, offering a platform for the practical application of forensic approaches. Our investigation intends to extend beyond the discovery of security vulnerabilities, diving into the real-world deployment of investigative techniques to secure the integrity of email communication. Aim is to make a positive impact on the advancement of security solutions that are proactive and flexible.

Building on this foundation, our analysis draws inspiration from the work of Mohammed et al. (2019), who acknowledge the intricacies introduced by multi-natural language emails. The proliferation of varied language patterns in spam emails needs an inventive approach to anti-spam methods. Our research strives to bring unique viewpoints that harness linguistic diversity, exploring innovative approaches to boost detection skills and keep ahead of the developing strategies adopted by spammers. Our goal is to strengthen email communication's ability to withstand the continuous barrage of sophisticated spam by doing this.

In this setting, it is vital to define the key terminology pertinent to the conversation. Email spam refers to unsolicited messages distributed in masse, generally for commercial or harmful motives. Advanced data security approaches comprise a range of technologies targeted at detecting, assessing, and mitigating security risks, including spam. The objective of this article is twofold: firstly, to explore the nuances of email spam and its impact on individuals and companies, and secondly, to analyze the usefulness of advanced security methods in countering this issue. By clarifying the underlying workings of these strategies, we want to provide stakeholders with the knowledge necessary to defend their email infrastructure against spam attacks.

2. Literature review

In the ever-evolving digital ecosystem, email communication stands as a keystone, promoting global connectivity and information exchange. However, the ongoing threat of spam poses a continuing challenge, needing tireless exploration and improvement of advanced data security techniques and algorithms. A thorough examination of recent literature offers insight on the complexity of email security and the novel techniques researchers are taking to address the growing nature of spam.

Altulaihan et al. (2023) present a comprehensive analysis of email security challenges, tools, and strategies used in investigations. The report underlines the essential need for comprehensive approaches to tackle the increased sophistication of phishing attempts, social engineering, and other security concerns within the email domain. By diving into forensic tools and investigative approaches, the researchers give vital insights into the evolving world of email security concerns.

Moving beyond the overarching investigation of email security, Mohammed et al. (2019) focus on a single component — the identification of spam in emails containing multi-natural language content. Traditional spam filters can struggle with various linguistic patterns, resulting to increased false positives and negatives. Mohammed et al. solve this difficulty by providing an anti-spam detection approach that harnesses the capabilities of machine learning and natural language processing. Their work emphasizes the necessity of adaptive technologies in effectively identifying and reducing the different linguistic strategies deployed by spammers.

In a mathematics and bioscience-driven exploration, Rafat et al. (2022) give ideas about dodging obscure communication within spam emails. By introducing ideas from these disciplines, the researchers attempt to untangle the nuances of spam patterns and behaviors. This interdisciplinary approach indicates a break from standard techniques and highlights the relevance of multiple views in understanding and combating the intricate strategies of spammers.

Shifting the focus to difficulties and solutions in email spam filtering, Yasin and Azmi (2023) give a detailed assessment of filtering strategies. The paper identifies the continuing issues faced in this domain and suggests strategies to enhance the efficacy of spam filters. By addressing concerns relating to false positives, false negatives, and the adaptive nature of spammers, Yasin and Azmi contribute to the ongoing discourse surrounding the refining of filtering systems.

3. Advanced data security Techniques against spam

Advanced data security techniques and algorithms are complex strategies used to safeguard sensitive information and systems against a range of cyber threats, specifically email spam. Due to the widespread use of digital communication, email has become a prime target for spammers who aim to distribute harmful content, engage in phishing efforts, and carry out fraudulent schemes. Conventional spam filters frequently find it challenging to keep up with the ever-changing strategies used by spammers, resulting in a growing influx of spam emails into users' inboxes.

To tackle this difficulty, it is imperative to employ sophisticated data security approaches and algorithms that can offer stronger and more efficient spam detection and mitigation systems. These methods utilize state-of-the-art technology like machine learning, deep learning, and natural language processing to examine the content of emails, the behavior of the sender, and other pertinent metadata. Data security specialists can accurately differentiate between legitimate messages and potential threats by utilizing powerful algorithms to detect trends, anomalies, and suspicious signs associated with spam emails.

Advanced data security techniques are crucial for protecting individuals, companies, and key infrastructure from the harmful consequences of email spam. By employing these methods, we may improve the security of our emails, safeguard confidential data, and reduce the threats posed by malevolent individuals attempting to take advantage of weaknesses in our digital communication platforms.

4. Methodology

To investigate sophisticated data security approaches and algorithms for combating email spam, the initial step is to gather a varied and extensive dataset. This dataset should comprise a sufficient number of both spam and legitimate email samples to ensure the effectiveness and robustness of the study. To do this, several sources can be employed, including public spam repositories, email honeypots, and real-world email traffic. Spam emails can be collected from repositories that specialize in collecting and curating spam communications, while email honeypots can be set to attract and capture spam emails in real-time. Additionally, authentic emails can be sourced from respectable email providers or organizations, ensuring a balanced representation of diverse email types and domains.

Once the dataset is acquired, it undergoes preprocessing to assure its quality and consistency. This preprocessing step covers numerous tasks, including data purification and standardization. Duplicate emails are eliminated to avoid redundancy, while unnecessary metadata and formatting issues are handled to strengthen the uniformity of the collection. Furthermore, email content is preprocessed by tokenizing, stemming, and deleting stop words to standardize text representations and assist further analysis. Additionally, essential elements are collected from email metadata, such as sender addresses, subject lines, timestamps, and attachment details, to provide valuable insights for the spam detection process.

Following preprocessing, feature engineering is undertaken to strengthen the discriminative ability of the dataset and improve the performance of the spam detection algorithms. This involves undertaking exploratory data analysis to uncover significant features and trends in the dataset. Feature engineering techniques may involve the extraction of linguistic features from email content, such as word frequencies, n-grams, and grammatical structures, as well as the incorporation of sender reputation scores and email routing information. By building informative features, the dataset becomes more representative and conducive to the construction of successful spam detection models.

Once the dataset is generated, the following step comprises the creation and evaluation of improved data security approaches and algorithms for combating email spam. This comprises the selection and implementation of appropriate machine learning, deep learning, and natural language processing algorithms, according to the features of the dataset and the aims of the research. Supervised learning methods, such as Support Vector Machines (SVM), Naive Bayes, and Random Forest, may be applied for binary classification of emails into spam and valid categories. Deep learning algorithms, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can be leveraged for feature extraction and classification tasks, using the hierarchical representations and sequential connections inherent in email data. Natural language processing tools, such as sentiment analysis, topic modeling, and Named Entity Recognition (NER), may potentially be integrated to assess the semantic and syntactic features of email content and boost the identification of spamming attempts.

The created models are tested using conventional performance metrics, such as accuracy, precision, recall, and F1-score, to determine their usefulness in properly recognizing spam emails while minimizing false positives and false negatives. Cross-validation approaches, such as k-fold cross-validation, may be applied to ensure the robustness and generalizability of the models over different subsets of the dataset. Furthermore, the models are submitted to rigorous testing utilizing independent validation datasets to confirm their performance under real-world settings and assess their scalability and efficiency.

Finally, the verified models are deployed and integrated into existing email security systems or frameworks to enhance their spam detection capabilities. This entails embedding the trained models within the email infrastructure, such as mail servers or spam filters, to automatically classify incoming emails in real-time. Continuous monitoring and feedback systems are built to adjust the models to evolving spamming strategies and assure their effectiveness over time. Additionally, user education and awareness initiatives may be implemented to enlighten users about email security best practices and decrease the chance of falling victim to spam and phishing assaults.

5. Conclusion

In conclusion, this thesis has examined sophisticated data security strategies and algorithms aimed at combatting email spam. By leveraging machine learning, deep learning, and natural language processing, together with robust preprocessing and feature engineering, excellent spam detection models have been constructed. These models offer promising answers to the continuous difficulty of discriminating between legal emails and spam, ultimately boosting email security for individuals and companies.

The deployment and integration of these advanced spam detection systems into operational email infrastructure constitute a key step towards minimizing the impact of spam and safeguarding users from potential dangers. Continuous monitoring, maintenance, and user education are crucial for assuring the sustained effectiveness and robustness of these systems against developing spamming strategies.

Moving forward, greater research and innovation in the realm of data security will be important to handle emerging spamming risks and boost the complexity of spam detection algorithms. By keeping at the forefront of technological breakthroughs and engaging with stakeholders across academics, industry, and government, we can collaboratively strive towards a safer and more secure email environment for all users.

Literature

1. Altulaihan, E., Alismail, A., Hafizur Rahman, M. M., & Ibrahim, A. A. (2023). Email Security Issues, Tools, and Techniques Used in Investigation. Sustainability, 15(13), 10612.

2. A.Karim, S. Azam, B. Shanmugam, K. Kannoorpatti and M. Alazab, "A Comprehensive Survey for Intelligent Spam Email Detection," in IEEE Access, vol. 7, pp. 168261-168295, 2019, doi: 10.1109/ACCESS.2019.2954791.

I. C. Ferreira, M. V. C. Aragão, E. M. Oliveira, B. T. Kuehne, E. M. Moreira and O. A. S. Carpinteiro, "The Development of the Open Machine-Learning-Based Anti-Spam (Open-MaLBAS)," in IEEE Access, vol. 9, pp. 138618-138632, 2021, doi: 10.1109/ACCESS.2021.3118901.

3. Mohammed, M. A., Mostafa, S. A., Obaid, O. I., Zeebaree, S. R., Abd Ghani, M. K., Mustapha, A., ... & AL-Dhief, F. T. (2019). An anti-spam detection model for emails of multi-natural language. Journal of Southwest Jiaotong University, 54(3).

4. Rafat, K. F., Xin, Q., Javed, A. R., Jalil, Z., & Ahmad, R. Z. (2022). Evading obscure communication from spam emails. Math. Biosci. Eng, 19(2), 1926-1943.

5. T. Xia, "A Constant Time Complexity Spam Detection Algorithm for Boosting Throughput on Rule-Based Filtering Systems," in IEEE Access, vol. 8, pp. 82653-82661, 2020, doi: 10.1109/ACCESS.2020.2991328.

Yasin, S. M., & Azmi, I. H. (2023). Email spam filtering technique: challenges and solutions. Journal of Theoretical and Applied Information Technology, 101(13), 5130-5138.

UDC: 141407

PENETRATION TESTING METHOD FOR SPECIFIC IOT DEVICE: IP CAMERA

Mektepbaeva Aruzhan, Medarov Almasbek

aruzhan.mektepbaevaa@gmail.com, amedarov59@gmail.com

students of Astana IT University, major - cybersecurity, 3-course

scientific advisor – A. Kulmuratova

Abstract. This article addressed the security vulnerabilities of IP cameras within the Internet of Things framework, emphasizing the critical need for effective penetration testing methodologies. Amidst the exponential growth of IoT devices. This research articulated the penetration testing process across various IP camera models to identify and assess the extent of potential cybersecurity threats. Employing a combination of automated tools and manual inspection techniques, such as Nikto, WhatWeb, and the Metasploit Framework, the study revealed prevalent weaknesses in password protection, firmware stability, and network configurations. Findings of rigorous analysis, it