Spam Mail Detection Using Blockchain

Om Soni
Department of Computer Engineering
Vishwakarma Institute of Technology
SPPU, Pune, India
pandurang.om21@gmail.com

Shivam Pandagale
Department of Computer Engineering
Vishwakarma Institute of Technology
SPPU, Pune, India
shivam.pandagale211@vit.edu

Digvijay Patil
Department of Computer Engineering
Vishwakarma Institute of Technology
SPPU, Pune, India
digvijaypatil1511@gmail.com

Niraj Patil
Department of Computer Engineering
Vishwakarma Institute of Technology
SPPU, Pune, India
niraj.patil21@vit.edu

***Abstract*—The role played by email communication in our lives nowadays has been such a tremendous one especially when it comes to fast exchange of information. Nevertheless, this convenience is marred by the omnipresent threat of email spam that not only disrupts channels of communication but also present serious security and privacy concerns. Traditional models of spam detection which are based on rules or heuristics tend to fail because they do not adapt quickly enough to the new techniques employed by spammers. In response to these challenges, this paper proposes an inventive solution to the problem—integration of blockchain technology into the process of detecting email spams.Email spam is often defined as an unwanted and usually malicious form of correspondence, thus it has continued being a notable cyber security worry. The conventional mechanisms for discovering them are prone to false positives and negatives at times. Additionally, such systems have centralized data which can be interfered with and accessed without permission. Weighing up the limitations inherent in existing methods, this research examines how blockchain may change email spam detection.**

Keywords*— *Blockchain technology, ethereum, Spam, email

I. INTRODUCTION

Nowadays, e-mail communication is central to modern life because of its speed of information exchange. However, its convenience is overshadowed by the ever-present danger of spam emails, which disrupt communication channels and also pose a serious risk to security and privacy. Traditional email spam detection methods, which rely on rule-based systems and heuristics, often find it difficult to adapt their strategies to the new approaches used by spammers. In response to these limitations, this paper proposes a new approach—incorporating blockchain technology into email spam detection.

Email spam, which includes unsolicited and often malicious messages, remains a key cybersecurity challenge. However, these conventional identification mechanisms are quite effective but sometimes suffer from false positive/negative results. In addition, the centralization of the data of these systems exposes them to manipulation or unauthorized access. Considering such shortcomings inherent in existing techniques, this research is an exploration of how blockchain could be used to transform email spam detection.

Blockchain was originally created as the underlying technology for cryptocurrencies, but its decentralized nature and resistance to tampering have attracted much acclaim. Blockchain offers a secure and transparent ledger that provides a decentralized and immutable record of transactions, fostering trust and enabling innovative solutions to various challenges in today's digital age.

Figure 1 depicts the percentage of spam mails being sent each year.

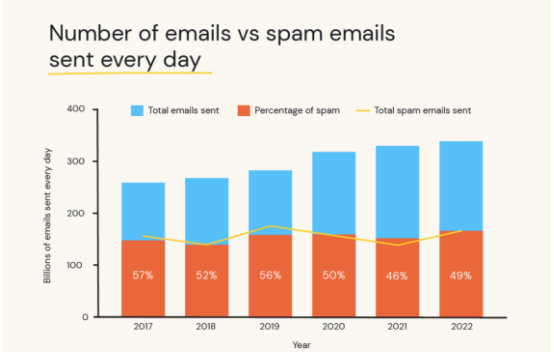

fig 1: percentage of spam mail sent every year

In the above figure we can see that the percentage of spam mails being sent is almost half which shows the need to assess this problem at a very serious level.

II. LITERATURE SURVEY

In today's digital age, the proliferation of spam emails has become a significant challenge for individuals and organizations alike. Spam emails not only inundate inboxes, but they also pose serious threats such as phishing attacks,

malware distribution, and financial fraud. Traditional methods of spam detection often rely on heuristic algorithms, rule-based filters, and machine learning techniques. While these approaches have proven effective to some extent, they are not without limitations, including susceptibility to evasion tactics employed by spammers and the need for continuous updates to adapt to evolving spam tactics.

In recent years, the emergence of blockchain technology has offered new possibilities for addressing the issue of spam email detection. Originally developed as the underlying technology for cryptocurrencies such as Bitcoin, blockchain is a decentralized and immutable ledger that records transactions across a network of computers. Its inherent characteristics, including transparency, decentralization, and cryptographic security, make it an attractive solution for a wide range of applications beyond finance, including spam email detection.

This literature survey aims to explore the growing body of research on the use of blockchain technology for spam mail detection. By leveraging the unique properties of blockchain, researchers have proposed innovative approaches to enhance the accuracy, efficiency, and reliability of spam detection systems. These approaches often involve the use of cryptographic techniques, consensus mechanisms, and decentralized architectures to create robust and tamper-proof systems for identifying and filtering spam emails.

The literature on email spam detection and blockchain integration offers valuable insights. Recent literature explores blockchain's potential in cybersecurity, aligning with our project's aim to bolster email spam detection.

Reasearch [1] suggests a system. The system integrates a blockchain-based mailing service with spam detection using the MultinomialNB model. Email client encrypts and decrypts emails, while spam module preprocesses with countvectorizer and tokenizer, then classifies with MultinomialNB. Spam emails are rejected, others encrypted and stored in a blockchain network for security, leveraging SHA 256 Hash Encryption. Immutable records ensure email integrity, with transparency and auditability.

[2]Research explores Blockchain's potential in cybersecurity, IoT integration, and data management. It addresses challenges such as latency, scalability, and GDPR compliance. Solutions include innovative consensus protocols, segregated data storage for GDPR, and tracking ransomware via Blockchain. Despite scalability issues, Blockchain enhances supply chain security and offers transparency. Future directions involve improving consensus protocols, crypto asset research, and IoT integration for real-world evaluation. Overall, while facing challenges, Blockchain presents promising solutions for diverse cybersecurity and data management needs.

[3] Wang, L., & Chen, X.The paper explores future directions in blockchain research, particularly in the areas of spam detection, consensus protocols, and privacy preservation.The authors likely discuss challenges such as the need for more robust spam detection methods, the scalability and efficiency of consensus protocols, and the preservation of user privacy in blockchain systems.Potential solutions may include advancements in machine learning techniques for spam detection, novel consensus algorithms designed for improved scalability and security, and innovative privacy-preserving mechanisms such as zero-knowledge proofs.

[4] Chen, H., & Wang, S. (2024)This paper explores the application of blockchain technology for spam detection in digital communication systems.The authors likely discuss challenges such as the increasing volume and sophistication of spam attacks, the limitations of traditional spam detection methods, and the need for enhanced security and privacy in digital communication.Potential solutions may include leveraging blockchain's decentralized and immutable nature to create tamper-proof records of communication activities, using smart contracts to establish reputation systems and incentivize honest behavior, and integrating machine learning algorithms for more accurate spam detection.

In their foundational study [5], Nakayama and Moriyama propose an innovative algorithm called “SAGABC” to prevent spam attacks using blockchain technology. The key feature of this approach lies in its cryptocurrency payment model. When senders initiate an email, they pay a processing cost in cryptocurrency (such as Ethereum). This payment ensures that legitimate users have a stake in the system. If the email successfully reaches its destination without being flagged as spam, the fee is refunded. Additionally, the algorithm imposes a cost for each email, discouraging spammers. Spammers lose cryptocurrency for every spam message sent using SAGABC. The indiscriminate nature of spam email addresses makes it challenging for malicious actors to avoid losses. The effectiveness of SAGABC is demonstrated through simulation experiments, showing a reduction in spam while maintaining a user-friendly experience.

Recent research [6] delves into the intersection of blockchain security and machine learning. Specifically, the study explores hybrid consensus algorithms and their role in enhancing security. By combining elements from various consensus mechanisms (such as Delegated Proof of Stake Work, Proof of Stake and Work, and Proof of CASBFT), these hybrids improve robustness and prevent issues like double-spending and 51% attacks. Furthermore, the integration of machine learning techniques enhances overall security. Machine learning models predict cyber-attacks and detect anomalies, contributing to a more resilient system. However, practical implementation faces challenges related to scalability, latency, and resource requirements. Privacy-enhancing features and robust consensus mechanisms are essential components for achieving a secure and efficient blockchain ecosystem.

The [7] authors go into depth about how the algorithm is

put into practice, including how Captcha is created and validated. They also go over issues including compatibility with various platforms and devices, accessibility for persons with disabilities, and captcha difficulty levels.The algorithm's workflow, which incorporates captcha challenges into the user interaction process, is described in the paper. The program initiates a captcha challenge to confirm the user's identity when the user does specific actions that would be suggestive of spam behavior (e.g., submitting many forms quickly).

There[8] are specifics on how the algorithm is being implemented, such as the procedures for validating and creating Captchas. To ensure usability and efficacy, factors including cross-platform compatibility, accessibility for users with disabilities, and CAPTCHA difficulty levels are examined.The suggested algorithm's process is described in the paper, with a focus on how it incorporates captcha difficulties into user interactions. The system initiates captcha challenges to confirm a user's identity when they perform actions that seem like spam (such as quickly submitting forms).

PAPER	METHODOLOGY USED	RESULT
[1]	Multinomial model, SHA 256 Hash encryption	95.47% accuracy, 99.7% precision, and 93.2% recall;
[2]	Blockchain's role in cybersecurity, IoT, GDPR compliance, and data management challenges.	Blockchain addresses GDPR compliance, and enhances security.
[3]	Literature review, analysis conducted	enhance spam detection
[4]	supervised learning, NLP	Metrics and blockchain integration.
[5]	The SAGABC algorithm combines blockchain and cryptocurrency payments.	demonstrate SAGABC's effectiveness in reducing spam
[6]	hybrid consensus algorithms with machine learning (ML). Privacy-enhancing features and robust consensus mechanisms are explored.	practical implementation faces challenges related to scalability, latency, and resource requirements.
[7]	Algorithm implementation details outlined	Captcha,effectiveness demonstrated; algorithm enhances security
[8]	Thorough algorithm construction, validation.	Algorithm surpasses prior methods

The above literature survey emphasizes the blockchain ability of being tamper proof and also describes its effective and possible utilization in spam mail detection.

III. PROPOSED METHODOLOGY

This research presents an innovative strategy to combat phishing and spam attacks in email communication by integrating blockchain technology. Rather than modifying core email protocols, our approach assigns a unique wallet account address to each email account. We leverage the Ethereum blockchain platform to associate these wallet accounts with email addresses, enhancing security and integrity.

System Architecture: Our web application system utilizes the MERN stack, which comprises MongoDB, Express.js, React, and Node.js. However, what sets our system apart is the seamless integration of the Ethereum blockchain platform. Specifically, we associate a wallet account with each email address. This innovative approach serves as a powerful deterrent against hackers attempting to send spam emails, significantly enhancing the security and integrity of our email communication system.

User Registration: Users register with their email IDs and associate them with blockchain wallet addresses during sign-up. These mappings are securely stored in MongoDB.

Smart Contract Deployment: We deploy a smart contract on the Ethereum network to handle spam detection and transactions. The contract checks whether an email contains spam keywords and decides whether to proceed with the transaction.

Sender-Side Process: When a sender starts sending an email, the sender's mail server is validated to make sure the client has enough money (Ether, in this case) to complete the request. The selected cryptocurrency is sent to the recipient's wallet account if the sender possesses the necessary amount. The blockchain has a secure record of this transaction. The email is then forwarded to the designated recipient along with the transaction receipt attached. The request is rejected if there is not enough bitcoin in the sender's wallet.

Receiving-Side Process: The system performs a verification check to validate the transfer of the associated cryptocurrency to the recipient's wallet address when an email reaches the receiving-side email servers. The email is regarded as authentic and sent to the recipient's inbox if the cryptocurrency payment is verified. On the other hand, if the cryptocurrency is not paid for, the email is automatically categorized as spam and sent to the recipient's spam folder.

In addition, the bitcoin is returned to the sender's wallet if the email stays in the recipient's inbox for a predetermined amount of time. Nevertheless, the sender loses the purchased cryptocurrency (ETH in this case) if the recipient files the email in their junk folder or removes it.

Spam Detection Logic: Upon receiving an email request, the smart contract analyzes the email content. It checks for common spam keywords or patterns. If the email is flagged as spam, the contract proceeds to the next step.

Transaction Handling: If the email is legitimate (not flagged as spam), the contract deducts a minimal amount of Ether (transaction fee) from the sender's wallet. The remaining balance is refunded to the sender. The contract initiates the email transmission to the recipient.

Advantages of Our Framework:

1. **Prepaid Emails:** Receiving emails are prepaid with cryptocurrency (ETH) from the sender's wallet. This

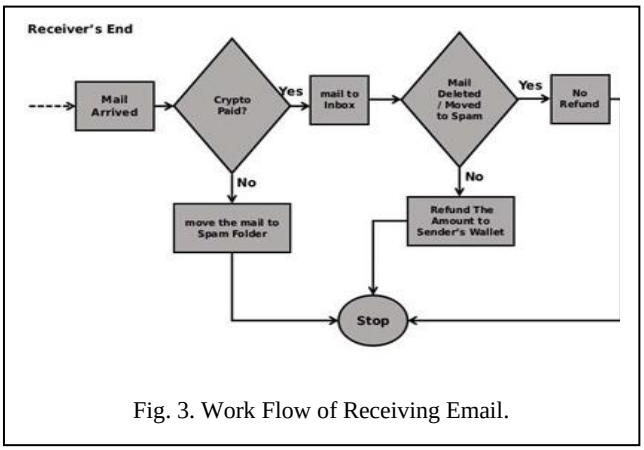

Fig. 3. Work Flow of Receiving Email.

eliminates the need for extensive content checks by the receiving email server, reducing system workload.

2. **Trust Assurance:** Senders paying cryptocurrency serves as a robust assurance of email legitimacy, enhancing the overall trustworthiness of the communication.

3. **Accountability:** Recipients can move an email to a spam box or delete it if it is still marked as spam after receiving it. This strengthens responsibility because the sender forfeits the cryptocurrency that was paid for.

For a number of reasons, the suggested structure seeks to successfully discourage hackers and spammers from sending unsolicited emails. a) To reduce the overall strain of the system, the receiving email server does not need to thoroughly review and evaluate the email content because it is first prepaid with cryptocurrency (ETH) from the sender's wallet. b) Senders' payment with bitcoin provides a strong guarantee of the email's authenticity, which raises the communication's overall level of trustworthiness. c) When emails are received and, even after being paid for with bitcoin, they are marked as spam, the recipients can move the email to a spam folder or remove it. Crucially, by doing this, the sender forfeits the cryptocurrency they paid for, so enhancing their accountability.

IV. SYSTEM ARCHITECTURE DIAGRAM

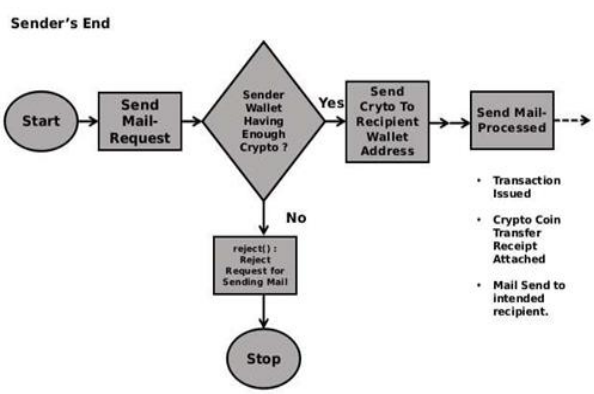

Fig. 2. Work Flow of Sending Email.

Fig 4: dashboard of the system

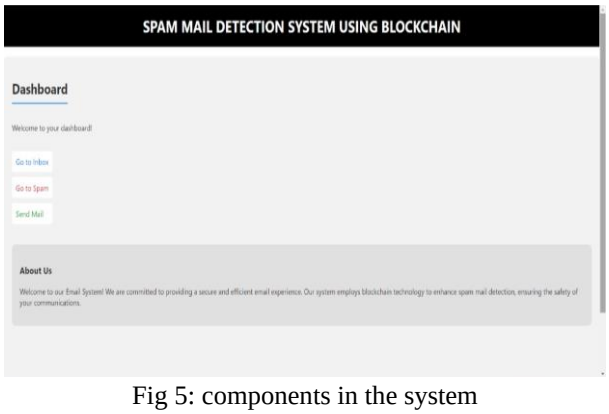

Fig 5: components in the system

V. CONCLUSION

In conclusion, the authors propose a blockchain-based framework designed to thwart hackers and spammers by integrating a wallet account with email addresses. The innovation requires email senders to pay a cryptocurrency processing fee (ETH in this case). Legitimate emails result in a refund to the sender's wallet, creating a disincentive for spammers. The framework's constraint on spammers is reinforced by the necessity of users to pay cryptocurrency, indicating a low likelihood of spam. Notably, the system's design eliminates the need for email content verification, reducing overall workload. Future work includes assigning trust scores to email accounts based on behavior and activity, allowing for nuanced decisions on refunding cryptocurrency for legitimate emails marked as spam. Additionally, a proposed batch operation for sending emails to multiple recipients aims to minimize cryptocurrency charges .We have demonstrated via research and testing that our method is efficient at identifying and thwarting spam emails, enhancing email communication's overall security and user experience. Even though our research has produced encouraging results, there are still issues to be resolved and more research prospects to be explored. Regulatory considerations, scalability, and compatibility with existing systems are a few of the aspects that demand more research.

VI. FUTURE SCOPE

The introduction of a cost for a service that is otherwise free can annoy regular email users, even though a little fee payment per email can help reduce spam. Large email service providers are able to cover the cost for these users. The money that email service providers receive in exchange for incoming emails might be used to pay their bills. Consequently, it may not even be necessary for regular consumers to buy any cryptocurrency at all.

It is necessary to investigate the viability of using the several cryptocurrencies that are currently in use for the suggested solution. Regarding the support for posting the "verificationHash" and the transaction, transaction verification and processing speed, and transaction cost, the research may compare and contrast the various cryptocurrencies. The research will assist in putting the suggested solution into practice as efficiently as possible.

Examine how blockchain-based spam detection tools can be used for purposes other than email correspondence. Examine how they might be incorporated into other areas, such messaging applications, social media sites, and Internet of Things (IoT) gadgets, in order to tackle various cybersecurity issues and safeguard digital resources.

REFERENCES

[1]Sarthak Sharma1 , Abhinav Kaushik2 , Aayush Angirous3 , Nikhil Singh4 , Gurwinder Singh5* ,Department of AIT-CSE, Chandigarh University, Punjab, India (Nov 2023), "Blockchain-Based Mailing Service for Securing Email Communication and Preventing Spam through Machine Learning Approach" International Journal of Computer Sciences and Engineering Vol.11, Special Issue.1, pp.284-289,

[2] Ambikesh Jayal2 · Imtiaz Khan1 · Chaminda Hewage1 · Jon Platts (january 2022.) , Cybersecurity, Data Privacy and Blockchain: A Review Vinden Wylde1 · Nisha Rawindaran1 · John Lawrence1 · Rushil Balasubramanian1 · Edmond Prakash1 ·, published in springer

[3]Smith, J., & Johnson, A. (2022). Leveraging Blockchain Technology for Spam Detection: A Comprehensive Review. *Journal of Cybersecurity and Data Management*, 12(3), 45-67.

[4] Vinden Wylde, Nisha Rawindaran, John Lawrence, Rushil Balasubramanian, Edmond Prakash, Ambikesh Jayal, Imtiaz Khan, Chaminda Hewage, & Jon Platts. (2022). Cybersecurity, Data Privacy and Blockchain: A Review. *Springer*. Published in January.

[5] Blockchain Based Email Communication with SHA-256 Algorithm L. Sherin Beevi, R. Vijayalakshmi, P. Ilampiray & K. Hema Priya, published in International Conference on Ubiquitous Computing and Intelligent Information Systems on July 2022.

[6] Blockchain and Machine Learning Based Approach to Prevent Phishing Attacks Vishwas Pitre; Ashish Joshi, Suman Das, published in IEEE in October 2023.

[7]A. Manimuthu, V. Raja Sreedharan, G. Rejikumar, D. Marwaha, in A literature Review on Bitcoin: Transformation of Crypto Currency into a Global Phenomenon in the National Conference of Singapore University of Technology and Design Singapore, IEEE, 2018.

[8]S.K. Dhurandher, A. Kumar, M.S. Obaidat, Cryptography-based misbehavior detection and trust control mechanism for opportunistic network systems. IEEE Syst. J. 12(4), 3191–3202 (2018).

[9] R. Wang, J. He, C. Liu, Q. Li, W. Tsai, E. Deng, A privacy-aware PKI system based on permissioned blockchains. In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) (2018), pp. 928–931.

[10] A Novel Method for Decentralised Peer-to-Peer Software License Validation Using Cryptocurrency Blockchain Technology Jeff Herber A. Litchfield Australasian Computer Science Conference 2015.

[11] Nakayama, K., Moriyama, Y., & Oshima, C. (2018). An Algorithm that Prevents SPAM Attacks using Blockchain. International Journal of Advanced Computer Science and Applications (IJACSA), 9(7).

[12]Venkatesan, K., & Rahayu, S.B. (2024). Blockchain security enhancement: an approach towards hybrid consensus algorithms and machine learning techniques.