BLOCKCHAIN AGAINST FAKE NEWS AND DEEP FAKES

[1]Sathiyapriya P, [2] Kohila R, [3] G Durga Prasad, [4] J Sumanth, [5] V Pavan Kumar Reddy

[1]Assistant Professor, Department of CSE(Cyber Security), Muthayammal Engineering College.

[2]Assistant Professor, Department of CSE(Cyber Security), Muthayammal Engineering College.

[3,4,5]Student, Department of CSE(Cyber Security), Muthayammal Engineering College.

[1]sathiyapriya18mec@gmail.com, [2] kohilamca@gmail.com, [3]durgaprasadgadiparthi4@gmail.com, [4]pkreddy824@gmail.com, [5]muthyalasumanth8555@gmail.com,

ABSTRACT - The rise of ubiquitous deepfakes, misinformation, disinformation, and post-truth, often referred to as fake news, raises concerns over the role of the Internet and social media in modern democratic societies. Due to its rapid and widespread diffusion, digital deception has not only an individual or societal cost, but it can lead to significant economic losses or to risks to national security. Blockchain and other distributed ledger technologies (DLTs) guarantee the provenance and traceability of data by providing a transparent, immutable, and verifiable record of transactions while creating a peer-to-peer secure platform for storing and exchanging information. This overview aims to explore the potential of DLTs to combat digital deception, describing the most relevant applications and identifying their main open challenges.

KEYWORDS - Fake News, Disinformation, Deepfakes, Blockchain Technology, Distributed Ledger Technologies (DLT), Cybersecurity, Content Authentication, Digital Trust, Decentralized Verification, Threat Detection.

1. INTRODUCTION

The Gartner Predicts That the majority of individuals in developed economies will consume more false than true information by 2022.1 Digital deception is commonly recognized as deceptive or misleading content created and disseminated to cause public or personal harm (e.g., post- truth, populism, and satire) or to obtain a profit (e.g., click baits, cloaking, ad farms, and identity theft). In the context of mass media, digital deception originates either from governments or non-state actors that publish content without economic or educational entrance barriers system as a consequence, these horizontal and decentralized communications cannot be controlled or stopped with traditional centralized tools.

In addition, this lack of supervision allows for security attacks (e.g., social engineering). Moreover, the veracity of information seems to be sometimes negotiable for the sake of profit, as the competition is increasingly tough.

1.1 SCOPE OF THIS PROJECT

This project aims to:

- Detect fake news and deep fakes using blockchain technology
- Store news data securely using Distributed Ledger Technology (DLT).
- Verify the authenticity and source of online content.
- Provide a simple web interface for users to check news credibility.
- Maintain trust and reputation scores for news publishers.
- Identify groups of fake news using keyword analysis.
- Design the system so it can be expanded in the future.
- Detect and verify fake news and deepfakes using blockchain technology.

The framework is designed for users to verify news authenticity, track content sources, and reduce the spread of fake news and deepfakes using blockchain technology.

2. PROPOSED WORK

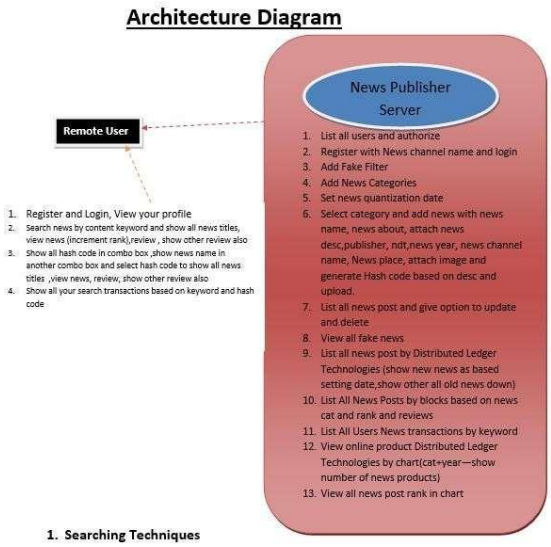
Traditional content moderation methods, such as flagging and notice-and-take-down, depend on centralized authorities that have full control over content removal. In Distributed Ledger Technologies (DLTs), especially permissionless blockchains, there is no central authority. Any participant can act as a validator, which requires the implementation of additional consensus mechanisms to manage and moderate content in a decentralized manner.

Qayyum et al. introduced the concept of proof-of-truthfulness, where any node in the network can verify whether a piece of content exists on the blockchain. The content is stored using a Merkle tree structure, where each node contains hash pointers to its child nodes.

In this approach, reliable fact-checkers are identified and rewarded for validating content using incentives such as tokens. As fact-checkers provide accurate validations, their reputation score increases, leading to higher rewards. Similarly, content creators are encouraged to submit their content for verification to build credibility and trust within the system.

Reputation systems assign credibility scores to content publishers to help readers identify potential bias or misinformation.

2.1 System Architecture



3.1 Data Processing & Preprocessing

To ensure accurate results, the framework incorporates:

- **Data Collection:** Gathers news content, metadata, publisher details, timestamps, and keywords from users and news sources.
- **Content Hashing:** Generates cryptographic hash values to uniquely identify content and detect tampering.
- **Metadata Validation:** Verifies source information, publication time, and category to improve traceability.
- **Blockchain Storage:** Stores hashes and metadata securely on the Distributed Ledger (DLT).

3.2 Content Verification Techniques

3.2.1 Fake News Detection:

Utilizes Compares content hash values with blockchain records to detect unauthorized changes. Analyzes keyword patterns and repetition to identify misinformation campaigns. Uses reputation scores of publishers to assess content credibility.

3.2.2 Deepfake Verification:

Validates multimedia metadata against blockchain-stored-records. Detects inconsistencies in content origin and timestamp information. Applies AI-assisted checks to identify manipulated or altered media.

3.3 TOOLS AND LIBRARIES

The proposed system uses a combination of modern tools and technologies to ensure secure, reliable, and efficient detection of fake news and deepfakes. Blockchain and Distributed Ledger Technology (DLT) are used to store content hashes and metadata in an immutable and transparent manner, ensuring data integrity and traceability.

The framework employs the following technologies:

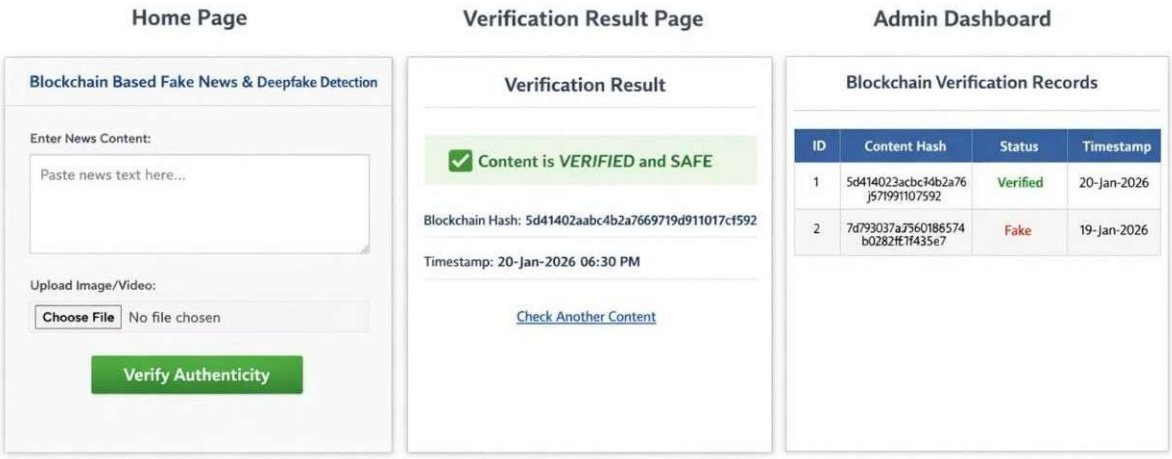
- **Java:** Core programming language for application development.

- **J2EE (JSP & Servlets):** Web framework for backend processing and content management.
- **Blockchain / Distributed Ledger Technology (DLT):** Used for secure, immutable storage of content hashes and metadata.
- **MySQL:** Database for storing user data, news content, and transaction records.
- **HTML, CSS, JavaScript:** Front-end technologies for the user interface and user interaction.
- **SHA Cryptographic Hashing:** Ensures data integrity and detects content tampering.
- **Machine Learning Algorithms:** Analyze content patterns and detect fake news trends.

4. PROGRAM

```
import java.io.*;
import java.security.MessageDigest;
import java.sql.*;
import javax.servlet.http.*;

public class MainProgram extends HttpServlet {

    protected void doPost(HttpServletRequest req, HttpServletResponse res)
            throws IOException {

        String content = req.getParameter("content");

        try {
            MessageDigest md = MessageDigest.getInstance("SHA-256");
            String hash = new java.math.BigInteger(1,
                    md.digest(content.getBytes())).toString(16);

            Class.forName("com.mysql.cj.jdbc.Driver");
            Connection con = DriverManager.getConnection(
                    "jdbc:mysql://localhost:3306/fakenewsdb","root","root");

            con.createStatement().executeUpdate(
                    "INSERT INTO content(hash) VALUES('" + hash + "')");

            res.getWriter().println("Content Verified and Stored Securely");
            con.close();
        } catch (Exception e) {
            res.getWriter().println("Verification Failed");
        }
    }
}
```

5. RESULTS AND OUTPUT



- **Fake News Detection Results:** Identifies SQLi vulnerabilities along with affected queries.
- **Deepfake Verification Results:** Determines whether multimedia content has been altered based on metadata and provenance checks.
- **Content Trust Analysis Reports:** Identifies suspicious patterns and groups of fake news using keyword and reputation analysis.

6. CONCLUSION

This Provenance, consensus, and traceability can be guaranteed with DLTs when creating a P2P platform for tackling digital deception. This article analyzed some applications currently under number of additional mechanisms to control content. CSRF and Remote Code Execution. Although there are technological and practical limitations of the DLT technology when combating digital deception, the trust mechanisms provided by DLT can make it more adequate than other technologies.

REFERENCES

[1] 1. K. Panetta, Gartner Top Strategic Predictions for 2018and Beyond. Gartner, Stamford, CA, USA, 2017.

[2] 2. J. Bayer, N. Bitiukova, P. B^ard, J. Szak^acs,A. Alemanno, and E. Uszkiewicz, Disinformation and Propaganda— Impact on the Functioning of the Rule of Law in the EU and its Member State. HEC Paris Research Paper LAW-2019-1341, 2019.

[3] [3] C. Wardle and H. Derakhshan, "Information Disorder: Toward an interdisciplinary framework for research and policy making," Council of Europe Policy Report DGI(2017)09, 2017.

[4] [4] Z. Shae and J. Tsai, "AI blockchain platform for trusting news," in Proc. IEEE 39th Int. Conf. Distrib. Compute. Syst., Dallas, TX, USA, 2019, pp. 1610– 1619.5. S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380,pp. 1146–1151, 2018.6. H. Kim et al., "Deep video portraits," ACM Trans. Graph., vol. 37, no. 4, p. 163, 2018.

[5] A. Shahaab, B. Lidgey, C. Hewage, and I. Khan,"Applicability and appropriateness of distributedledgers consensus protocols in public and private

[6] sectors: A systematic review," IEEE Access, vol. 7,pp. 43622–43636, 2019.

[7] A. Qayyum, J. Qadir, M. U. Janjua, and F. Sher, "Using blockchain to rein in the new post-truth world and check the spread of fake news," IT Professional, vol.

[8] Cross-Site Scripting (XSS) – OWASP Foundation, "Understanding and Mitigating XSS Attacks," 2023.

[9] A. Seaborne and E. Prud'hommeaux, "SPARQL Query Language for RDF," W3C Recommendation, 2008. [Online]. Available: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/

[10] J. Grossman, "XSS Attacks: Exploits, Defenses, and Techniques," *IEEE Security & Privacy*, vol. 10, no. 3, pp. 56-62, 2018

[11] X. Zhang and A. A. Ghorbani, "An overview of onlinefake news: Characterization, detection, anddiscussion," Inf. Process. Manage., vol. 57, no. 2, SQLMap: Automatic SQL Injection and Database Takeover Tool, https://sqlmap.org/, Accessed: 2024.