***Abstract* – This document provides a comprehensive analysis of the patent US11483343B2, which pertains to a phishing detection system and method of use. The analysis will delve into various aspects of the patent, including its technological underpinnings, the novelty of the invention, its potential applications. A high-quality summary of the document is presented, highlighting the key elements that contribute to its significance in the field of cybersecurity.**

The analysis is beneficial for security professionals, IT experts, and stakeholders in various industries, providing them with a distilled essence of the patent and its utility in enhancing cybersecurity measures. It serves as a valuable resource for understanding the patented technology's contribution to the ongoing efforts to combat phishing and other cyber threats.

I. INTRODUCTION

The patent US11483343B2, titled "Phishing Detection System and Method of Use," focuses on an advanced system and methodology for identifying and mitigating phishing attacks. This patent proposes a specific architecture for a phishing detection system that scans messages for suspicious URLs and analyzes the corresponding webpages to identify phishing attempts

II. INDUSTRIES

The phishing detection system and method are applicable across a wide range of industries and sectors that rely on digital communications and are vulnerable to phishing attacks:

A. Technology Sector:

- Technology companies, especially those providing software, cloud services, social media platforms, and e-commerce, are prime targets for phishing attacks seeking user data and credentials.
- The technology sector would benefit from improved phishing detection to protect their platforms, customers, and reputation.

B. Financial Services:

- Financial institutions like banks, investment firms, insurance companies, and fintech startups handle sensitive financial data and transactions.
- Phishing attacks often impersonate financial services to steal account credentials, payment details, and commit fraud.
- The financial sector has a strong need for effective phishing detection to secure customer accounts and comply with regulations.

C. Healthcare:

- Healthcare organizations like hospitals, clinics, insurance providers, and pharmaceutical companies maintain personal health information and insurance/payment data.
- Phishing attacks may seek to steal patient data, commit insurance fraud, or disrupt operations.
- Protecting against phishing is critical for HIPAA compliance and patient trust in the healthcare sector.

D. Education:

- Educational institutions from schools to universities have moved many services online and hold student personal and financial data.
- Phishing attacks may target students, faculty, and staff to steal identities, academic records, or research data.
- Schools and universities need anti-phishing measures to safeguard educational data and intellectual property.

E. Government:

- Government agencies at the federal, state, and local levels are also targeted by phishers seeking sensitive data or to disrupt services.
- Improved phishing detection can help secure public sector systems and data.

III. THE PROPOSED SOLUTION

This patent proposes a multi-stage phishing detection system that scans messages, resolves embedded URLs, extracts webpage features, and applies machine learning to identify phishing attempts. While it offers more proactive and comprehensive coverage than traditional methods, it may face performance and accuracy challenges in the evolving landscape of phishing attacks. Nonetheless, it represents a significant step towards automated, real-time phishing detection and prevention.

The proposed phishing detection system and method identify phishing attempts in electronic messages and aims to proactively detect and block such malicious messages.

A. Key Components of the Proposed Solution:

Phishing Detector: The core component is a phishing detector module that analyzes messages for suspicious content. It consists of two main subcomponents:

- **Scan Engine**: Scans the message body and attachments to identify any URLs (web addresses) present. Extracts these URLs for further analysis.

- **Fetcher Component**: Takes the URLs found by the scan engine and resolves them to the actual webpages they point to. Retrieves the HTML source code of these webpages.

Feature Extraction: The phishing detector then extracts two types of features from the retrieved webpages:

- **URL-based Features**: Analyzes the structure and components of the URL itself, such as length, special characters, IP address usage, etc. Suspicious patterns may indicate a phishing attempt.
- **Hyperlink-based Features**: Examines hyperlinks present in the webpage source code. Looks at target URLs, anchor text, and other link attributes for signs of deception.

Machine Learning Models:

- **Hybrid Feature Set**: The URL and hyperlink features are combined into a hybrid feature set representing each webpage. This provides a comprehensive characterization of the page's suspiciousness.
- **Machine Learning Models**: The hybrid feature sets are used to train machine learning classifiers to distinguish between phishing and legitimate webpages. Models are trained on large datasets of known phishing and benign examples.

B. Method of Use:

- **Message Scanning**: When a new message arrives, the phishing detector's scan engine identifies any URLs present in the content.
- **URL Resolution**: The fetcher component resolves the found URLs to their target webpages and retrieves the page source code.
- **Feature Extraction**: URL and hyperlink-based features are extracted from each webpage.
- **Classification**: The pre-trained machine learning models are applied to the extracted feature set. The models classify the webpage as phishing or legitimate.
- **Action**: If a webpage is deemed a phishing attempt, the original message can be quarantined or blocked. Alerts may be generated for administrators or the intended recipient.

IV. Process Flow

The key process flow involves the scan engine extracting URLs from messages, the fetcher resolving those URLs to webpages, analyzing the URL and hyperlink features of those pages, and applying ML models to detect phishing attempts, leading to automatic deletion of phishing messages. This multi-stage analysis allows proactive, real-time filtering of phishing content based on the destination webpage characteristics, going beyond traditional URL or content-based filtering methods.

This process flow covers the end-to-end lifecycle of the proposed solution and focuses on the requested aspects:

A. Scan Engine and Fetcher:

- The scan engine scans incoming messages to identify and extract any URLs present in the message body or attachments.
- The fetcher component then resolves the extracted URLs to the actual webpages they point to and retrieves the HTML source code of those webpages.

B. URL Detection and Resolution:

- The scan engine is responsible for detecting URLs embedded in messages. It scans the message content and attachments to identify URL strings.
- Once URLs are detected, the fetcher component resolves them to their target webpages. This involves following redirects and retrieving the final webpage that the URL ultimately points to.
- The fetcher retrieves the full HTML source code of the resolved webpage for further analysis.

C. Webpage Analysis:

- The retrieved webpage HTML is analyzed to extract two types of features:
    - **URL-based features**: Analyzing the URL string itself for suspicious patterns like length, special characters, IP address usage, etc.
    - **Hyperlink-based features**: Examining the hyperlinks in the webpage source, looking at target URLs, anchor text, and link attributes.
- These URL and hyperlink features are combined into a hybrid feature set representing the webpage's suspiciousness.
- Pre-trained machine learning models are applied to this feature set to classify the webpage as phishing or legitimate.

D. Phishing Detection Criteria:

- The key phishing detection criteria are the URL and hyperlink features extracted from the resolved webpage.
- Suspicious URL patterns can include excessive length, random character strings, IP addresses, URL shorteners, etc.
- Hyperlink features like mismatched target URLs, suspicious anchor text, or links to known malicious sites can indicate phishing.
- The machine learning models are trained on datasets of known phishing and legitimate webpages to learn the distinguishing patterns.
- A webpage is classified as phishing if the model determines its URL and hyperlink features match learned patterns of malicious pages.

E. *Message Deletion:*

- If a webpage linked in a message is determined to be a phishing attempt, the original message can be quarantined or deleted automatically.
- This prevents the user from engaging with the malicious content and potentially compromising their information.
- Message deletion can happen as soon as the phishing determination is made, before the message reaches the user's inbox.
- Alternatively, suspicious messages could be flagged for review before deletion, in case of potential false positives.

V. Benefits, Drawbacks and Significance of Proposed Solution

A. *Key benefits*

The key benefits of this phishing detection system are its ability to automatically delete phishing messages, avoid reliance on potentially stale external blacklists, improve detection accuracy through machine learning, prevent phishing in real-time before messages reach inboxes, and integrate with existing email infrastructure for multi-layered defense. These capabilities represent a significant advancement over traditional phishing prevention methods.

1) *Automated message deletion for phishing:*

- If a webpage linked in a message is determined to be a phishing attempt, the original message can be automatically quarantined or deleted
- This prevents the user from engaging with the malicious content and potentially compromising their information
- Message deletion can happen as soon as the phishing determination is made, before the message reaches the user's inbox

2) *Reduced reliance on external blacklists:*

- The system avoids reliance on external blacklists or databases that may become stale
- It uses only URL and hyperlink-based features extracted from the webpage source code itself, without relying on third-party services
- This allows it to detect new and evolving phishing attempts that may not yet be included in blacklists

3) *Improved phishing detection accuracy:*

- Combining URL and hyperlink analysis provides more comprehensive coverage and accuracy compared to traditional methods
- Machine learning models are trained on large datasets of known phishing and benign examples to learn distinguishing patterns
- This enables adaptable, automated classification and reduces false positives compared to rule-based approaches

4) *Real-time phishing prevention:*

- The system proactively detects phishing by analyzing destination webpages, not just message content
- URLs are resolved and webpages analyzed in real-time as messages arrive
- This allows phishing attempts to be blocked before they reach the user's inbox, preventing engagement with malicious content

5) *Integration with mail transfer agents or client software:*

- The phishing detection system can be integrated into mail transfer agents (MTAs) or email client software
- Integration with MTAs allows scanning and blocking of phishing messages during the email delivery process
- Integration with email clients provides last-mile protection at the user's device level
- This enables a multi-layered defense, protecting at both the server and endpoint

B. *Limitations*

While the proposed system offers improvements over traditional methods, it still faces challenges in terms of computational efficiency, adaptability to new threats, accuracy trade-offs, dependency on external factors, language coverage, and user behavior. Addressing these limitations will be key to providing robust, real-time phishing protection in the face of ever-evolving attacks.

1) *Computational cost and scalability:*

- Resolving URLs and analyzing webpages at scale can be computationally expensive
- The system needs to handle a high volume of messages and URLs, which may impact performance and scalability
- Possible delays in message delivery due to the scanning process could affect user experience

2) *Constant arms race with attackers:*

- Phishers constantly adapt their techniques to evade detection, leading to an ongoing arms race
- The system may struggle to keep up with new phishing patterns and zero-day attacks
- Attackers may find ways to obscure phishing content or mimic benign pages to bypass detection

3) *Handling false positives and negatives:*

- The system may generate false positives, incorrectly flagging legitimate messages as phishing
- False negatives, where phishing attempts go undetected, are also a risk
- Balancing accuracy and minimizing false positives/negatives is challenging and impacts user trust

4) *Dependency on external data sources:*

- The system relies on third-party data like WHOIS records, PageRank, etc. for webpage analysis
- Changes or disruptions to these external data sources could affect the system's accuracy and reliability

5) *Language and internationalization:*

- Phishing attempts in different languages or targeting specific regions may be harder to detect
- The system may need adaptation and training for multilingual and international coverage

6) *User behavior and social engineering:*

- No technical solution can completely prevent users from falling for well-crafted social engineering attempts
- User curiosity, distraction, or lack of caution can lead to clicks on phishing links despite warnings
- Continuous user education and awareness are still necessary to complement any technical detection system

7) *Potential privacy concerns:*

- Analyzing user messages and browsing activity for phishing detection may raise privacy questions
- Balancing user privacy with effective threat detection needs to be considered

8) *Staying ahead of evolving threats:*

- As phishing tactics evolve, the detection system needs continuous updating and retraining
- Adapting to new phishing patterns and attack vectors requires ongoing effort and resources

C. *Significance*

The key significance of this phishing detection system is its ability to improve detection accuracy through machine learning, prevent phishing in real-time before messages reach inboxes, automatically delete phishing messages, avoid reliance on stale blacklists, and integrate with existing email infrastructure for comprehensive, multi-layered protection. These capabilities represent a significant advancement over traditional phishing prevention methods in the ongoing battle against increasingly sophisticated phishing attacks.

1) *Improved Phishing Detection Accuracy:*

- Combining URL and hyperlink analysis provides more comprehensive coverage and accuracy compared to traditional methods
- Machine learning models are trained on large datasets of known phishing and benign examples to learn distinguishing patterns, enabling adaptable, automated classification and reducing false positives

2) *Real-Time Phishing Prevention:*

- The system proactively detects phishing by analyzing destination webpages, not just message content, in real-time as messages arrive
- This allows phishing attempts to be blocked before they reach user inboxes, preventing engagement with malicious content

3) *Automated Message Deletion:*

- If a webpage linked in a message is determined to be a phishing attempt, the original message can be automatically quarantined or deleted before reaching the user's inbox
- This prevents users from engaging with the malicious content and potentially compromising their information

4) *Reduced Reliance on External Blacklists:*

- The system avoids reliance on potentially stale external blacklists by using only URL and hyperlink features extracted from the webpage source code itself
- This allows it to detect new and evolving phishing attempts that may not yet be included in blacklists

5) *Integration with Email Infrastructure:*

- The phishing detection system can be integrated into mail transfer agents or email client software for server-side scanning or last-mile endpoint protection
- This enables a multi-layered defense, protecting at both the email delivery and user device levels