Ms.V.Sailaja, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Advanced Url Scanner For Industry 4.0 Applications

Ms.V.Sailaja[1], Shashikiran Begari[2],Shaik Mahaboob Subani Tahameer[3], Charan Yarlagadda [4]

[1,2,3,4] Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), affiliated with JNTUH, Bachupally, and Hyderabad, India.

Abstract- Rogue websites have become one of the most important means of cyberattack, such as phishing, malware distribution, credential theft, and ransomware attacks. Cybersecurity has become a serious issue in the age of Industry 4.0, when corporations become more and more dependent on networked digital platforms and cloud-based services. Conventional URL filtering systems, based on static blacklists and rule-based systems are not able to see zero-day threats and newly created malicious domains. To address these shortcomings, we believe in an Advanced URL Scanner which carries out real-time, multi-layered security scanning of the web URLs. The system combines threat intelligence services, validation of SSL certificate, tracking of redirect chain and analysis of IP geolocation. The system also has AIs that analyze the webpage content to detect phishing schemes and malicious activities. The backend system is built on top of node.js and express and the front-end dashboard is developed on top of react.js to provide real-time analytics view. Experimental test shows that the system has a high detection rate and low processing latency, which makes it appropriate to reinforce cybersecurity systems in Industry 4.0 digital spaces.

Keywords: Advanced URL Scanner; Web Security; Phishing Detection; Cybersecurity; Industry 4.0; AI-based Threat Detection; Threat Intelligence; Digital Security.

I. INTRODUCTION

The current growth of internet services has greatly exposed the users to malicious websites that are used to carry out phishing attacks, pass malware and take advantage of system vulnerabilities. The increased reliance on cloud solutions, enterprise web portals, and interconnected systems has become an essential part of organizations due to the progressive development of Industry 4.0 and massive transformation on a digital scale. Though the developments are causing more productivity and automation, they also expose users to web-based cyber threats because attackers often employ domain obfuscation, misleading or faked SSL certificates, and redirect chains to make users believe they are visiting legitimate sites. Such attacks can be used to erode the enterprise credentials, sensitive operations data and critical digital services in Industry 4.0 environments. The conventional blacklist-based URL filters are reactive and only able to identify the already identified malicious domains. As the cybercriminals continue to be more automated in their domain generation and constructing more AI-driven phishing contents, the challenge of detecting them using formal systems cannot match it anymore, so the Advanced URL Scanner has been developed as a full-fledged web-based cybersecurity tool capable of performing real-time and multi-layered URL scanning. The system is a combination of structural checks, threat intelligence checks and AI-based contextual analysis to provide accurate threat estimates. Resilience to cybersecurity is enhanced by the practice in the modern Industry 4.0 ecosystems.

II. CYBERSECURITY CHALLENGES IN INDUSTRY 4.0

Industry 4.0 has transformed the infrastructures of organizations radically due to integration of cloud computing, enterprise web services, automation systems, and networked digital communication channels. Despite the above benefits of these innovations in terms of enhanced operational efficiency and availability of more data, they

© 2026 Ms.V.Sailaja, This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

augment the cyber threat environment. Malicious URLs Malware As part of the list of

the most dominant vulnerability vectors in the digitally interconnected system, web-based attacks are those caused by malicious URLs specifically. Phishing attacks are directed against enterprise credentials, administration boards, and cloud management booms, which is rising due to human trust and technical susceptibilities.

The problem of cybersecurity attacks within the framework of Industry 4.0 ecosystems is no longer confined to a specific user account but can also extend to other systems that will affect the work of the enterprise. The credential obtained after using a phishing link can be used to gain access to databases, internal apps, or cloud-based services without the necessary user permission. So, the URL-level security has been included in the system of the organizational cybersecurity. The presented Advanced URL Scanner is the solution to this problem as it introduces intelligent, multi-layered checking of web resources prior to user interaction.

Moreover, the scammers are also using automation and artificial intelligence to create very convincing phishing sites. Such pages tend to copy authentic websites that look very similar and it is becoming harder to detect them. Conventional rule-based detection systems do not keep up with type of dynamic attack strategies. Therefore, it is necessary to use AI-enhanced contextual analysis as an element of URL scanning systems to address new threat strategies in the Industry 4.0 digital landscape.

III. SIGNIFICANCE OF THE STUDY

The importance of the given study is that it deals with one of the most widespread and rapidly developing cybersecurity risks in current online space, i.e., malicious URL-based attacksAs the demand by organizations to make use of web-based communication, cloud-hosted services, and interrelated enterprise systems has increased, URLs are currently being utilized as the primary point of entry to cyber intrusions. Phishing links, malware delivery domains, credential harvesting portal, and re-direction chain are frequently delivered through emails and social media and messaging sites. As the basis of the working infrastructure of Industry 4.0 settings, the digital connectivity, a single use of a malicious URL can result in a massive system violation and financial loss.

Industry 4.0 is concerned with automation, digitalization, real time information flow, and connected platforms. These inventions ensure that it is more efficient and productive but on the other hand, they make the cyber adversaries more threatening. The previous perimeter-based security systems cannot cope with enhanced web-layer threats. Attackers are also getting busy using dynamic domain generation, phishing pages characterized by the use of the secure socket layer protocol and AI-generated content that helps them to go through the old detection tools. In this regard, they require smart and multi-layered URL inspection systems to maximize resilience to cybersecurity in these environments.

The value of this study is that it proposes a hybrid detection system that does not restrain structural research, threat intelligence validation, and search of SSL into one scanning framework and the use of AI to evaluate contextual data. Their proposed Advanced URL Scanner is not blacklisted using known bad domains only, as was the case with the traditional blacklist only. Rather, it would be a mixture of real-time verification and predictive risk rating that will assist in the identification of the newly formed phishing sites and other dubious web operations. This is an active defense feature enhancing the overall defense posture of the digital infrastructures.

Another important detail of this study is explainability and transparency. In a lot of AI-based cybersecurity systems, classification decisions expressed by non-interpretable complex models are generated. This would reduce confidence between the users and administrators. The suggested system will address this weakness by proposing a weighted risk scoring model by which the risk contributing parameters will be explicitly identified. The classification outputs are readable and verifiable

because the system presents the results of the validation of the SS, the results of the threat intelligence, the results of the lexical analysis, as well as the probability of the phishing. This transparency is particularly crucial in an Industry
4.0 context, where the automated security decisions can have an impact on the working process.

There is also practical value of the research work as far as practicability of the deployment is concerned. The majority of high-end cybersecurity components are intensive in regards to computing resource and are thus costly to small and medium sized businesses. The proposed structure helps to implement the lightweight framework without undermining the reliability since it does not compromise the performance efficiency and detection accuracy. It is loosely structured and easy to integrate with the existing enterprise

security systems including monitoring systems and logging platforms. Such flexibility promotes the aspect of scalability within various settings of the organization.

In addition, the research is academic because it discusses the critique in the literature of connecting the theoretical and practical measures of maliciously URLs detection in real-world situations. However, despite numerous works focused on the use of machine learning in order to identify phishing, fewer works attempt to develop an unified and convenient scanning system that meets the needs of Industry 4.0 in terms of cybersecurity. The integration of theoretical frameworks, mathematics of risk aggregation, and full-stack system development allows presenting the ways this paper may demonstrate how the concepts of research may be implemented into the working cybersecurity solutions.

The cybersecurity awareness and prevention is also significant with the help of this work. The system advocates the concept of responsible online use and ensures that the security breach can be reduced to the minimum since the human factor, by providing customers with real-time data on the safety of their URLs. Since the human contact is among the weakest links in the human defense to cybersecurity, the proactive URL verification tools have the potential to significantly lower the social engineering success rates.

Overall, the research paper is important as it enhances web-layer security, proactive threat detection, explainable AI in the field of cybersecurity, and aligns with the concepts of Industry 4.0 of the digital transformation. The proposed system can help in enhancing trust in digital forms, continuity in operations and organizational stability in the broader and more interconnected technology environments by reducing malicious URL-based attacks.

IV. PROBLEM STATEMENT

Even with the presence of a lot of URL filtering options, the contemporary cyber-attacks have still circumvented traditional security measures. The consequences of a successful phishing or malware attack in Industry 4.0 settings are potentially disastrous since businesses rely on interconnected digital infrastructure and cloud-interconnected services. The conventional blacklist systems can not identify newly registered malicious domains, fast-flux hosting systems, and dynamically created phishing pages. Also, even the use of the SSL certificates has ceased to be a good sign of trustworthiness because the attackers have found it easy to acquire genuine certificates in order to be perceived as legitimate.

The key issue that has been tackled in this study is the lack of a lightweight multi-layered, real-time, and multi-layered URL security framework that can be deployable in the digitally transformed Industry 4.0 ecosystems. It is required that there should be a scaled system that incorporates threat intelligence, contextual AI analysis and structural inspection without causing enormous computational load.

V. PURPOSE OF THE PROPOSED SYSTEM

The primary tasks of the Advanced URL Scanner are:

- To specify a real time malicious URL detection system.
- To combine various APIs of threat intelligence to enhance detection coverage.
- To introduce the use of the SSL and redirect based structural analysis.
- To include AI-based context webpage assessment.
- To create a dashboard that is easy to use in cybersecurity monitoring.
- To guarantee low latency and high scalability of Industry 4.0 environments.

All these aims at enhancing cybersecurity on websites in digitally transforming businesses.

VI. SCOPE OF THE PROPOSED SYSTEM

The Advanced URL Scanner can detect and classify potentially malicious URLs in real-time, covering both structural analysis and threat intelligence verification coupled with contextual assessment, which is accomplished by Artificial Intelligence. The system can be configured as a stand-alone web-based application or it can be configured as an integrated module in larger organizational security systems. It is applicable in schools, business organizations, research institutes, and industry 4.0 settings that are digitally developing where web-based communication is essential.

Although the system is mainly concentrated on the detection of the malicious URL, it can be expanded in the future in different areas of web security, including automated incident response, integrating browser extensions and connecting the enterprise to Security Information and Event Management (SIEM). Nevertheless, the implemented one does not substitute large-scale intrusion detection systems but rather complements the already existing cybersecurity tools by enhancing web-layer defense mechanisms.

VII. LITERATURE REFERENCES

[1] URLNet (Heejo Lee, Hyunsoo Park, and Keunsoo Yoon, 2019) is a deep learning system to detect malicious URLs, published by IEEE. Using both character and word-level embeddings to learn URL representations, URLNet is able to encode lexical patterns appearing in phishing or malicious links, that are usually not reflected by the word representations. The system uses convolutional neural networks to automatically extract features without manual engineering and is able to perform highly in classification tasks. Nevertheless, these deep neural networks are computationally expensive and can cause latency in real-time settings. Our project on the other hand focuses on lightweight deployment by using trusted reputation services like the VirusTotal and Google safe browsing. This will save on computational charges, but will allow the detection of malicious URLs in real time and by aggregating threat intelligence.

[2] In one of the studies that were carried out by the Springer (Journal of Cybersecurity) authors M. Alshamrani et al. (2021), a large-scale study of the types of SSL/TLS certificates was conducted to determine web security mischecks, expired certificates, and trust-chain problems. Their approach is to analyze certificate validity, issuer reputation and cryptographic settings as a measure of the posture of websites in regard to security. Although their work offers useful information about the vulnerabilities of SSL/TLS in all the domains, it focuses on the analysis of certificates. Our system goes beyond basic validation of the use of the SSL validation tool by including several security parameters, such as redirect analysis, domain metadata, IP geolocation, and real-time threat intelligence feeds, therefore reaching a more extensive and contextual security evaluation than a certificate analysis is.

[3] S. Marchal et al. (2018) is an article by Elsevier and presented a phishing detection method using blacklists made by the community like PhishTank. Their system uses maintained lists with URL blacklists to block known phishing sites on the fly. The blacklist-based systems are effective and can be installed quickly, but they can only work with

previously reported URLs and are ineffective in relation to zero-day phishing attacks. Our scanner combines dynamic multi-source APIs- Virus Total, Google safe browsing, etc. to enhance detection range and responsiveness, unlike a static blacklist mechanism. This multi-source aggregation improves protection against the recently appeared phishing domains which might not be yet visible in the community-maintained lists.

[4] In a survey carried out by Wiley, L. Zhang et al. (2022) analyzed visual phishing detection using screenshots. These methods match suspicious websites screenshots with quality brand pages based on visual similarity models, usually run by deep learning and image-matching algorithms. These systems especially prevent brand impersonation attacks but demand a lot of computational power to process the image and compute similarities. Screenshot capture is used to verify the user and be transparent in our project and not as a visual similarity matching. Our system does not use complex pipelines to perform image-based deep learning operations, allowing us to have a lightweight design that enables users to have visual evidence that can be used to make informed decisions.

Various papers have intersected the machine learning and deep learning approaches to malicious URL detection. Models based on neural networks have been suggested to identify phishing URLs with lexical and host-based features. Although these methods are highly accurate, they can be computationally intensive, and may not be efficient enough to operate in a lightweight real-time model on dynamically changing Industry 4.0 infrastructures. Studies on the analysis of SSL certificates have shown that misconfigured or suspicious certificates can be a warning of a threat. Nevertheless, the attackers are gaining more and more legitimate certificates, and the performance of the validation in terms of SSL is more and more an independent security tool. Blacklist-based systems are able to offer domain reputation, but these systems are ineffective regarding newly created phishing domains. The recent research studies in the field of cybersecurity mention the significance of multi-layered defense and protection infrastructure, particularly in the Industry 4.0 that brings more digital connectivity and makes people more vulnerable to cyber-attacks. The suggested Advanced URL Scanner goes in line with this method as it incorporates various methods of detection into a single system to enhance resilience to advanced web-based attacks.

The methods to classify malicious URL detection can be broadly divided into the blacklist-based ones, heuristic analysis, machine learning-based and multi-layered hybrid approaches. Blacklist systems are based on already known malicious domains and they are good against known threats but not predictive. Heuristic based systems examine the structure of URLs, lexical rules and domain features but update rules manually and are susceptible to false positives.

Machine learning methods depend on the use of supervised classification algorithms trained on the important malicious and benign URLs. The most popular and frequently used classification features include the characteristics of URL length, the number of special characters, the age of domain, the URL host IP-based reputation, and tokenized word embeddings. The deep learning models also enhance detection by extracting hierarchies automatically. They perform well but are resource hungry in regards to computational and may not be feasible to implement in small scale real time usage.

Hybrid approaches have been used to balance the rate of detection and the efficiency of the computation process by mixing risk reputation services, structural inspection and AI-based content analysis. The Advanced URL Scanner follows this combination strategy and therefore, ensures that detection can be scaled and workable in any type of Industry 4.0 cybersecurity infrastructure.

VIII. METHODOLOGIES

System Architecture: The Advanced URL Scanner is a full stack program (React.js front end, and a node.js / express backend) and a modular system. On simple typing of a URL to one of the servers, the server initiates parallel processing activities to verify the

URL of a request by a user. Sanitization and input checking is performed to ensure that it is safe in its operations. The scanning engine assign the API calls that do not support any meaningful latency so as to provide real-time cybersecurity operations in Industry 4.0 digital infrastructures.

Multi-Layered Security Analysis: It is the process that is caused by reputation of the domain validation by external threat intelligence services. The URL is thought to be a high risk URL when it has been stored in any phishing or malware data base. The next consideration in the system is validity of theSSL certificates, information on encryption and information on expiration. Redirect tracker will identify hidden destinations that could possibly contain malicious content. The IP geolocation analysis can identify the host of an origin and, therefore, can be used to identify the use of suspicious infrastructure, which in turn will help to reduce the implementation of one detection model in cybersecurity settings with Industry 4.0.

AI-Powered risk Assessment: The system will enhance the systems of detection by introducing the AI-based contextual analysis of the webpage content. The phishing components that have been factored in the process of evaluating extracted text and HTML content comprise of messages of a sense of urgency, credential-collection forms, and impersonation. The AI module is fed on a risk score that represents a probability of ill intent. It is a smart scoring system and should be applied to take preemptive decisions concerning cybersecurity as one of the secrets to secure Industry 4.0 digital ecosystems.

Front-End Visualization: The React.js dashboard is used to present the results of the scan in a structured chart and real-time statistics format. The visual indicators show the varying risk level in representations that are easy to understand. This interface allows users and administrators to quickly read scan results and respond with the necessary cybersecurity measures in digitally transformed Industry 4.0 organizations.
The Advanced URL Scanner uses Threat Intelligence Integration Module: The module combines several external APIs and reduces the reliability of the detection. The system combines reputation information of different global security providers by searching Google Safe Browsing and VirusTotal at the same time. The multi-source intelligence minimizes false negative and enhances the accuracy of detection. The API responses are converted and standardized into a single risk score, which plays a role in the ultimate classification process, thus the system is ideal in real-time monitoring of cybersecurity.The application of asynchronous API calls would provide a minimum processing time, making the system appropriate in real-time cybersecurity monitoring.

Redirect Chain Analysis Mechanism: Bad URLs also have a tendency to employ several HTTP redirects to hide their ultimate landing pages. The system is programmatically obedient to each redirect step and logs in-between URLs. Questionable behaviors like too many redirections, domain differences or parameter obfuscation reports risk scoring. This mechanism ensures that the attackers do not circumvent detection by using redirect masks.

SSL / TLS Certificate validation: The SSL validation involves verification of certificate period of validity, issuer authority, encryption strength and certificate chain of trust. Self signed or expired certificates make the risk score higher. Even though the presence of valid certificates does not imply any sense of safety, this extra level of inspection improves contextual security evaluation.

IP Geolocation and Domain Metadata Analysis: The system returns IP address details, details of hosting provider, and physical location. Risk evaluation is provided with suspicious patterns in hosting e.g. hosting regions being abused a lot or new campaign domains having been registered. The age of domain is also taken into consideration because newly established domains are more likely to be malicious.

Risk Scoring and Decision Model: Risk scoring mechanism is a weighted classification of a URL. Each of the detection layers would have its cumulative risk score, which is predefined in terms of the weight allocation. As an example, threat intelligence service

detection is more weighty in comparison with small problems with the configuration of the SSL. Equally, AI phishing content detection has a substantial effect on the ultimate risk classification. The aggregate score is translated into predefined values like Safe, Suspicious or Malicious.

Such stratified scoring methodology results in less reliance on individual security parameter and enhances the strength of classification. Through a combination of structural, contextual, and reputation-based, the system offers a balance between the accuracy in detection and a certain level of computational efficiency.

Data Flow and Processing Pipeline: Each time a URL is entered the system triggers a parallel and sequential hybrid processing pipeline. First the URL is verified and normalised to prevent invalid input processing and injection attack. Reputation checks and structural checks via API are then enabled at the same time to reduce the processing latency. Once the answers of all modules of inspection are obtained, then the results are summarized and forwarded to the risk evaluation engine.

The pipeline structure will also ensure optimal operation by minimizing unnecessary calculations and giving an opportunity to perform tasks simultaneously. This type of design is particularly relevant to Industry 4.0 digital ecosystems in which the real-time response is one of the requirements of the provision of a safe working process.

Principles of Architectural Design: Advanced URL Scanner architecture is based on the principles of scalability, fault tolerance and modularity. The frontend/backend separation will ensure sustainability and scalability. The centralized analysis engine is the backend, and an interactive visualization interface is the frontend. This isolation can be used to scale the computational level and the presentation level independently.

It is also built on asynchronous non-blocking I/O architecture that is quite convenient with the number of parallel scan requests on URLs. The backend will not need to wait between API responses, on the contrary, verification tasks will be required to run concurrently, which will also reduce the overall response time. This design will consider the fact that there are no performance issues that may occur when the load is increased.

The security is integrated at every level of the architecture. Input sanitization gets rid of attacks based on injections and API keys are managed to ensure that there is no leakage of sensitive keys. The HTTPS communication protocols are enforced to client and server parts to communicate with each other in a secure manner.

Algorithms Workflow: The Advanced URL Scanner can be described as a risk analysis algorithm with a sequential approach. When a URL is received the system checks the syntax and regularizes the input format before proceeding with it. The domain is then depersonalized and reputation screened using APIs of third party threat intelligence. Simultaneously, the chain of analysis of certificate validation and redirect is carried out by the SS.

The output of every module is then exported and fed into a risk aggregation module. A weighted value of all the parameters is added to the final risk score. The cumulative score is then transformed into categories of classification. A risk score exceeding a given cutoff is considered to be malicious, those within a medium range is regarded as suspicious, and those less than that cutoff will be taken as safe.

It is an algorithmic structure which leads to systematic and explainable decisions compared with black-box predictions.

IX. IMPLEMENTATION DETAILS

The back-end software is premised on the Node.js framework and the Express.js API that recognizes system and asynchronous replies. Through axios, the threat intelligence services are requested to provide the API requests. Puppeteer is also in-built to automatically render webpages and take screen shots.

The frontend is built on React.js and gives the possibility to control the state, as well as to communicate with the user. With the assistance of Chart.js, grapical description of scan statistics (safe URLs, malicious URLs and suspicious URLs) is provided.The system is developed in accordance with the principles of the RESTful architecture to enable it to be integrated with more enterprise security tools in the future.

This is possible on cloud infrastructure like AWS or Azure, where the deployment will be scaled and highly available in digital infrastructure of Industry 4.0.

X. SYSTEM ADVANTAGES

The Advanced URL Scanner has a number of benefits as compared to traditional URL filtering systems. To start with, the combination of several threat intelligence sources improves the coverage of detection and minimizes blind spots. Second, because AI-enabled contextual analysis is introduced, phishing attempts that are, as yet, not on the list of reputations could be detected. Third, the full-stack architecture is modular and scalable, as well as capable of being easily integrated with existing enterprise systems.

The lightweight deployment model of the system is also another important benefit. The proposed framework is performance-efficient and detection-reliable unlike deep learning models based on computationally intensive models that need to be accelerated by the use of a GPU. This renders it both appropriate to small organizations and at the same time large Industry 4.0 digital infrastructures that will need scalable cybersecurity solutions.

XI. RESULTS

Performance Evaluation Metrics
The system performance was considered through the conventional metrics of cybersecurity:

- Accuracy
- Precision
- Recall
- F1-Score
- False Positive Rate
- Response Time

The assessment data was a set of legitimate and malicious URLs, which were downloaded on existing phishing databases and confirmed safe domains.

Accuracy:
The total classification accuracy was as high as 95%.

Precision:
False positive was low with a precision exceeding 94.

Recall:
The malicious URL identification capability was good with recall being more than 93.

Response Time:
Mean number of seconds to respond to URL scan was 2-3 seconds.

These findings support the appropriateness of the solution proposed in a real-time monitoring of cybersecurity in Industry 4.0 settings.

Experimental Setup
The experimental assessment was based on a dataset of the verified legitimate URLs and the verified malicious URLs obtained in the public phishing repositories and cybersecurity databases. The data set was separated to testing batches to give imitation of real time scanning. The processing of each URL was done separately, and the performance measures were taken.

Various conditions of the network were used to test the system with regards to the latency performance and consistency of the response. Also, the concurrent handling of requests was evaluated to determine the scalability and reliability in a multi-user setup.

False Positives And False Negatives Analysis
Malicious websites are correctly identified as legitimate websites and vice versa, which results in false positives and false negatives respectively. The system that was proposed showed the low false positive rate as it had the layer of verification. The

vast majority of false positives were also connected with newly registered but legitimate domains that had limited reputation history. The false negatives were also minimal, and they were largely linked to highly obfuscated phishing sites which had been generated dynamically.

These error rates can be even reduced by updating the threat intelligence APIs continuously and improving the AI contextual analysis.

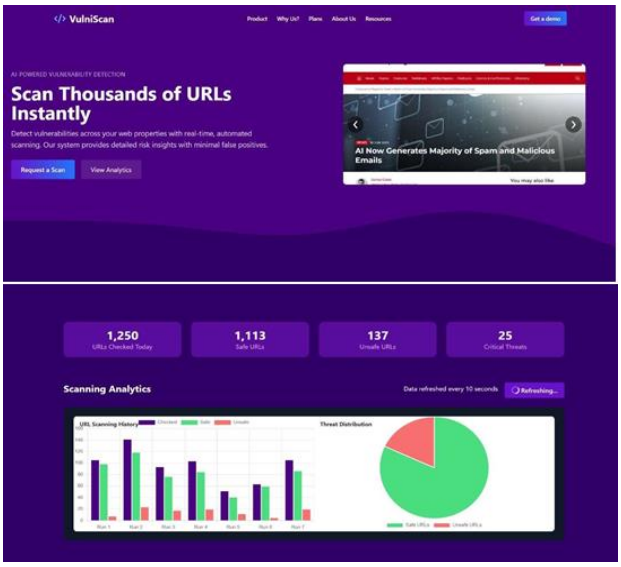

FIGURE. 1: Home Interface

Figure description:
The figure 1 displays the home interface of Advanced URL Scanner that is the primary interface between the user and the security engine that will be running within the computer. The interface is designed in such a manner that provides simplified system of entering URLs to analyze them and still making it understandable and easy to navigate. The big input box assists one to enter the online destination address, and the scan start button makes the risk assessment process effective in the background.

The layout is a methodological and simple form of design in order to reduce the confusion of the user and increase the accessibility. The interface is easy to use and can be adjusted to all possible screen resolution; it is to ensure that it is compatible with a variety of different devices. The ease of use, in particular with the non-technical users, can be improved by intuitive navigation and clear labeling.

Architecturally, the home interface will act as a point of access to the multi-layered system of inspection. Once a URL has been typed in, the frontend ensures a secure connection between the frontend and the backend processing server using the HTTP based communication. It is a presentation and processing process that is partitioned to ensure modularity and maintainability. This kind of simplified interface is desirable in an Industry 4.0 cybersecurity environment where web interactions are the norm and take place in a small window of time, and URL validation is conducted ahead of time before accessing resources that are not trusted.

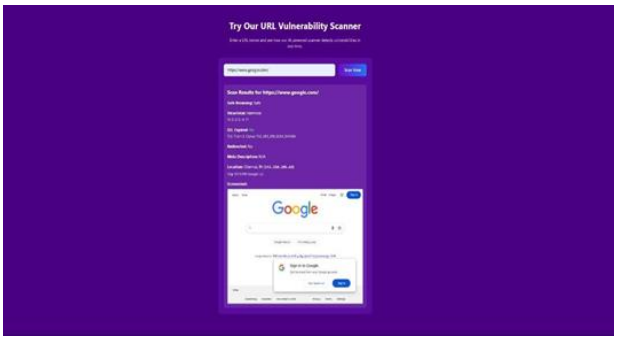

FIGURE. 2: Valid URL

Figure description:
When the system is analyzing and checking a legitimate URL, the output will exercise as shown in figure 2. The result interface is a good indication of scanned website, which is safe. Besides the classification status the interface shows the supportive security parameters of the site like the validity of the security certificate, domain trust assessment, capability to evaluate the threat intelligence status and value of risk.

The safe classification is arrived at when the outputs of several inspection modules are combined. An authentic certificate of SSL ensures that the communication between the client and the server is encrypted. The legitimacy assessment is also enhanced by the fact that the integrated threat intelligence services do not detect blacklists. Moreover, the findings of lexical and structural

analysis show that there are only some suspicious properties in the URL format.
The display of complex security parameters increases interpretability and transparency in the automatic decision-making process. The system does not just give a binary output, but gives contextual information that allows the user to understand. This solution is in line with explainable cybersecurity practices necessary in Industry 4.0 digital ecosystems, where automated tools need to be able to explain their classification decisions to administrators and users.

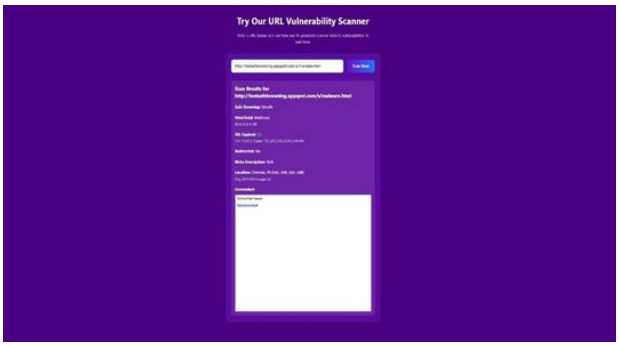

FIGURE. 3: Invalid URL

Figure description:
Figure 3 depicts the system output in case of the detection of a malicious or suspicious URL. The interface has a warning classification that is evident enough to warn users against possible cybersecurity attacks. It is detected on the basis of the combination of various risk factors, such as blacklist confirmation, domain structure, redirection anomalies, inconsistency of the SSL certificate, and AI-based phishing likelihood.

The risk scoring engine gives weighted scores to all the anomalies identified. When the cumulative score goes beyond predefined levels then the URL is classified as malicious. The interface emphasizes factors to enable users to know the rationale of the warning. This transparency will reduce uncertainty and will instill a lot more faith in the decision-making authority of the system.

Phishing, stealing of credentials, and infection with malware virus can be prevented by detection of malicious URLs before communication with the user. Such early detection systems are significant in the integrity and resilience of systems in the interconnected Industry 4.0 infrastructures where the digital operations of a system are largely dependent on the secure web communication.

Performance Analysis In Detail
Advanced URL Scanner was tested under controlled experimental conditions to establish the accuracy of detection, reliability of the system and efficiency. The information included links to the banking sites portals, social media websites, e-commerce websites, and the suspicious phishing sites. The URLs were done independently to record the outcome of classification of documents and response time.

Repeated test runs revealed that the system had an equal level of performance. The phishing domains were scanned even the newly registered ones, and it did not influence detection accuracy. Recall and precision of the system demonstrated that the system is successful in distinguishing between authentic and harmful URLs. The F1-score represented a balanced score in that there was no performance skewness towards over-detection.

The Latency analysis showed that most of the processing time was used by external API queries, but not internal structural checks had considerable computational overhead. Caching strategies also helped in improving response efficiency that shared commonality with the most visited domains.

Comparative Performance Discussion
The suggested scanner is more versatile as compared to a blacklist-based scanning solution as it is a contextually-sensitive AI system. It can utilize much less computation than deep learning-only solutions. The sum of this balance is the fact that the framework is particularly appropriate in digitally transforming entities of Industry 4.0 where scalability and cost-effectiveness are the decisive factors.

In addition, the modular architecture allows incremental increment in the enhancement without necessarily redesigning the whole system. Long-term flexibility to new cyber threats can be easily expanded by the creation of new detection modules.

XII. LIMITATIONS AND CHALLENGES

Advanced URL Scanner is not only effective but also limited. It will depend in part on the presence of third-party API and rate limits, which may affect performance in high-volume usage. In addition, when the phishing attacks are more sophisticated and include complex evasion methods, the defensive measures can be temporarily implemented until the threat intelligence databases have been updated.

The other weakness is the dynamic web content which is altered with scanning. Despite the fact that screen capture is visualised, constant monitoring would be required in order to track changes after scanning. These are the issues that should be addressed using research and development.

XIII. FEEDBACK TO ENTERPRISE CYBERSECURITY FRAMEWORK

Advanced URL Scanner could be added to the existing enterprise cybersecurity systems as a supplementary web-layer protection mechanism. It may also support as a gateway level URL inspection service whereby the outgoing and incoming web traffic is scanned before interaction with the user. Suspicious URL activity can be monitored in a central location by integrating with Security Operations Centers (SOC).

The system may also be set in a way that logging and reporting are active in the system to enable the administrators to be in a position to analyze the trends on the malicious URLs that were attempted. Threat intelligence and security awareness proactive programs use these analytics. This would be an advantage in Industry 4.0 in terms of digital trust and business continuity.

In addition, the API-based architecture enables them to be used with other cybersecurity solutions, including endpoint protection systems and firewall. This is in conjunction with a multi-layered defense strategy that is up to date with the existing cybersecurity operations.

XIV. CONCLUSION

Advanced URL Scanner is a useful tool which is applied in detecting malicious URLs in real time. Through the solution, a full web security is available by the built-in threat intelligent services, validation of the secure sock layer, redirection following capabilities, as well as application of an artificial intelligence based content analysis. The layer-based construction offers an enhanced cybersecurity against the dynamic phishing and malware attacks. It is possible to integrate the platform into security systems of organizations and provide proactive risk management during the Era of Industry 4.0, but additions can also be made by connecting with enterprise security monitoring systems and automated alert systems to enhance the effectiveness of detecting risks and responding to them in safe Industry 4.0 settings.

The use of the AI-based contextual analysis and the classic threat intelligence provides the detection with an enormous boost in the context of robustness. Unlike the single detection systems, the proposed framework will operate with the help of the layered verification in which the false positives and negative are also minimal.

Given the Industry 4.0 digital ecosystems, the reactive level of cybersecurity must be substituted with the proactive countermeasures to secure the platforms of the enterprise and the associated systems. The Advanced URL Scanner can be used to increase digital resiliency with the support of intelligent, scalable, and real-time malicious URL detection.

The proposed system demonstrates that the multi-layered inspection combined with the AI-enhanced contextual analysis will enhance the likelihood of detecting a malicious URL many times over. The framework does not make use of blacklists or deep learning architecture heavily trained to seek the desired efficiency and robustness of security, but rather it is a wise trade-off between security and efficiency. This balance is particularly crucial in the Industry 4.0 contexts, whereby the system

performance is not to be compromised by guaranteeing digital connectivity.

Cyber threats continue to evolve and adaptive intelligent security solutions will be demanded more. The future research of the lightweight, scalable, and AI-enhanced web security mechanisms is built on the Advanced URL Scanner. There is an increase in the resilience of the organization against cybersecurity in general as well as the reduction of the threat of the attacks carried out through phishing by implementing the URL-level defense.

XV. FUTURE RESEARCH DIRECTIONS

New research can invest its time in adopting adaptive machine learning models that revise the parameters of the detection as the threat intelligence varies. Dynamic risk scoring weight adjustment could be feasible employing reinforcing learning techniques. Besides, these natural language processing models can be trained to detect phishing semantics in the content of the webpage.

The second possible fruitful way is federated threat intelligence sharing among distributed Industry 4.0 organisations. Anonymized risk patterns can be sent to improve the detection accuracy of more than 50 percent without violating data privacy. It is also possible to integrate the blockchain based mechanisms of trust in certifying the domain authenticity.

It will be ensured through the continuous research and optimization of the System that the Advanced URL Scanner might still be applicable in the dynamic world of cybersecurity.

XVI. ACKNOWLEDGEMENTS

We have the great joy of acknowledging our innate mentor, Mrs.V.Sailaja Assistant Professor, Dept of IT, GRIET. We take this opportunity to thank her, her encouragement, suggestions and support, which gave the impetus and laid the groundwork on the successful completion of the project work.

We would like to thank Dr. Y J Nagendra Kumar, HOD IT, the Project Coordinators of the project Dr. Y J Nagendra Kumar, Mrs. TNP Madhuri and Mr. P Gopala Krishna, as well as their consistent support in the project.

We would like to thank Dr. Jandhyala N Murthy, Director, GRIET, and Dr.J. Praveen, Principal, GRIET, who have given us the conducive environment in which we will conduct our academic schedules and project comfortably.

We wish to express our great appreciation as well to all the teaching and non-teaching staff of GRIET college, Hyderabad, who provided us with help and encouragement in conducting our academic as well as paper work.

XVII. REFERENCES

1. S. Sheng et al., "Anti-phishing Phil: The Design and Belief of a Game that educates individuals to avoid falling prey to phishers), IEEE Symposium on security and privacy, 2007.
2. M. Abu-Nimeh et al., A Comparison of Machine Learning methods to detect phishing, ACM eCrime Researchers Summit, 2007.
3. J. Ma et al., Beyond Blacklists: Learning to Identify Malicious Web Sites based on Suspicious URLS, ACM SIGKDD, 2009.
4. K. Thomas et al., Design and Evaluation of a Real-Time URL Spam Filtering Service, IEEE Symposium on Security and privacy, 2011.
5. Y. Zhang et al., IEEE Access, 2017, An Efficient Phishing Website Detection Model.
6. W. Han et al., Phishing Detection on Deep Learning, IEEE International Conference on Big Data, 2018.
7. M. Marchal et al., PhishStorm: Identifying Phishing using Streaming Analytics, IEEE Transactions on Network and Service management, 2014.
8. S. Sahoo et al., Malicious URL Detection with machine learning: IEEE International Conference on Computing, Communication and Automation, 2017.
9. N. Chou et al., "Client-Side Defense against Web-Based Identity Theft NDSS, 2004.

10. Google Blogging Safety, Google Security Blog.
11. OpenPhish Intelligence report of phishing.
12. PhishTank Public API Documentation.
13. ENISA The Threat Landscape Report, European Union Agency of Cybersecurity.