International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Detecting Mobile Malicious Webpages in Real time

RAVI T
Assistant Professor
Department Of Information Technology
Panimalar Engineering College Chennai,
Tamil Nadu, India ravi_it@panimalar.ac.in

MOHIT R
Department Of Information Technology
Panimalar Engineering College Chennai,
Tamil Nadu, India
mohitrpani@gmail.com

NITHISH KANNA M
Department Of Information Technology
Panimalar Engineering College Chennai,
Tamil Nadu, India
mnithipecit01@gmail.com

MOWLESWARAN L
Department Of Information Technology
Panimalar Engineering College Chennai,
Tamil Nadu, India
lmowleswaran8747@gmail.com

Abstract - **The mobile technology expansion and the growth of the mobile technology are two sides of a coin. A huge rush has been formed by the availability of high speed internet connections in the mobile web access of people. People now use they can do all the necessary things using their smartphones add web banking and electronic payment solutions and business communication and healthcare access. The count has been expanded by the increasing digital world of targets that are now assaulted by cybercriminals evil web pages, which they utilize in assaulting mobile device users. These pages are designed in such ways that they are able to do phishing and steal user credentials whilst they also redirect the users to non-approved destinations and distribute evil programs. The existing web security systems that employ blacklist and signature based methods are ineffective against new threats and zero-day attacks. The research study introduces a detection system that is immediate applies machine learning in mobile web page protection. Threat discovery using website behavior analysis. Its working mechanism is by ensuring 24-hour surveillance of Webpage layouts and webpage running URL properties system data in order to have better identification of websites results. The technique provides rapid detection outcomes along with low computing needs which protect mobile phones with an efficient. security system.**

The keywords include Mobile Security, Malicious Webpages, etc. Phishing Attack, Live Attack Detection, Machine Web Threat Analysis, Learning.

I.INTRODUCTION

The widespread adoption of smartphones has transformed dramatically how the users access the internet. Mobile webpages also vary with desktop webpages in terms of layout, content and functionality. These variations pose some difficulties to the conventional malicious webpage detection systems, which are mostly set-up in a desktop setup.

The mobile web pages can also incorporate features like redirections, embedded scripts, and dynamic content which are also integrated in legitimate pages. This makes it hard to differentiate between bad and good behavior.

Also, mobile devices offer extra features including direct calling and location access, and this can be used by the attackers. Thus, a specialized system should be developed, which will be able to identify malicious mobile webpages in real time.

II. RELATEDWORK

Research in detection of malicious webpages has found extensive coverage in the area of cybersecurity, especially in relation to phishing and malware. Initial methods were more desktop oriented based and used blacklist methods, heuristic analysis and visual similarity methods. Blacklist systems like Google Safe browsing are based on the previously known malicious web addresses, but are unable to identify the 0-day attacks and new malicious web pages.

A number of studies have examined machine learning-based solutions to phishing. These approaches are used to draw out attributes on the Uniform resource Locators, UI contents and webpages format and categorize the websites as benign or malicious. A systematic literature review shows that machine learning tools, in particular, Random Forest and Deep Learning models, are very popular and can have high efficiency in identifying phishing sites.

The conventional desktop-based detection methods lack effectiveness in the mobile context because of the variation in the size of the screen, the structure of the content and the way a user interacts with the screen. It has been found that in most instances of mobile web pages it is evident that a mobile webpage is redirected and scripted several times even in the instances where the redirecting process is done with good intentions. It is hard to differentiate between a malicious process with the use of traditional approaches.

To solve this problem, there are a number of mobile specific detection systems suggested. APuML (Anti-Phishing using Machine Learning) is one of these methods, where both the feature of the site and the popularity of the site are used to identify such dangerous websites.

According to this system, classifiers like Support Vector Machines, Decision Trees and Random Forest are used and the accuracy of the classification is about 93.85%.
It can also identify high-end attacks including clickjacking and drive-by downloads. Other studies have also investigated Android-based phishing detection systems, in which machine learning models are embedded into mobile apps to forecast and block the access to harmful web sites. These systems extract URL properties and webpage contents to improve the detection features in a mobile setting.

In addition to the research in the domain of mobile safety point out that the level of mobile cyberattacks has been growing over the past years and that strong detection systems should be developed to serve the mobile platforms.

Regardless of these improvements, the current solutions have certain drawbacks like high false positives, failure to respond to emerging threats, and poor real-time performance. The kAYO proposed system solves these issues through the creation of a mobile-specific technique of static analysis which takes into account peculiarities in the form of phone call links, mobile APIs, and, thus, enhances the detection rates and efficiency.

III. RESEARCH METHODOLOGY

The proposed system is aimed at identifying malicious mobile webpages in a systematic approach which is a combination of data gathering, feature identification, and machine learning-based classification. The methodology will be such that the accuracy and real-time performance is high.

A. Data Collection

A dataset of mobile web pages on a large scale is gathered by different sources, both benign and malicious web pages. This data set comprises of over 350,000 webpages thus being diverse in terms of webpage structure, content, and behavior. The bad examples are phishing websites, scam websites and malware-carrying web pages whereas the good ones come through reputable domains.

B. Data Preprocessing

The obtained data undergoes preprocess in which noise and unrelated information are eliminated. This step includes:

- Eliminating repetitive web pages.
- Cleaning HTML content
- Normalizing URLs
- Blocking the partial or blocked pages.
- Preprocessing makes the dataset consistent, and capable of extracting features and training models.

C. Feature Extraction

It is a system that extracts the static features of the mobile webpages without running them. The reason why the static analysis is preferable is that it is efficient and can be used to detect in real-time.

The features obtained are:

- URL-related features: URL length, the presence of special characters, suspicious domain.
- HTML characteristics Counter of iframes, hidden elements, form tags.
- JavaScript has: Redirections, obfuscated scripts.
- Mobile-specific features Presence of click-to-call links, mobile API use.
- Such qualities would be used to differentiate between innocent and malignant web pages.

D. Feature Selection

Not every features that have been extracted are of equal usefulness to classification. Thus, the feature selection methods are used to reveal the most important features. This helps in:

- Improving the complexity of computation.
- Enhancing the model performance.
- Eliminating redundant data

E. Model Training

Training a classification model using the selected features is done using machine learning algorithms. Algorithms that are commonly used are:

- Decision Trees
- Random Forest
- Support Vector Machines (SVM)

To be able to test the quality of the model, this dataset is divided into trainings and testing the sets.

F. Classification

The trained model takes the webpages and puts them under two categories:

- Benign webpages
- Malicious webpages

This categorization is done in real time using the features that have been extracted out of the web page that is under access.

G. System Implementation

The suggested system is used in the form of a mobile device browser extension. The extension:

- Checks webpages as the browsing takes place
- Applies the trained model
- Notifies users about malicious webpage

H. Performance Evaluation

Standard performance measures are used to evaluate the system as:

- Accuracy
- Precision
- Recall
- F1-Score

The proposed system is accurate in the vicinity of 90 per cent with a reflection of its efficiency as an indicator of malicious mobile webpages.

Factor	**Cantina**[56][59]	**kAYO**
Designed to detect	*Phishing*	Mobile web threats
Detects pages written in	*English-Only*	*Any Language*
Avg.feature extraction time	*2.82 seconds*	*0.016 seconds*
Evaluation set size	*200*	*34914*
True positive rate	*97%*	*89%*
False positive rate	*6%*	*8%*
Detects pages missed by google safe browsing?	*No*	*yes*

IV. Performance Evaluation Metrics

To test how effective the proposed system is when it comes to monitoring the malicious mobile webpages, a count of marginal ability metrics are employed. Such metrics are based on the confusion matrix which comprises of the following elements:

- True Positive (TP): Great webpages that were rightly classified as malicious.
- True Negative (TN): The Webpages that were correctly characterised as benign.
- False Positive (FP): Webpages that are wrongly identified as malicious.
- False Negative (FN): Evil Web pages that have been wrongly labeled as innocent.

1. Accuracy

The accuracy is a measure that is used to gauge the general correctness of the model in classifying webpages.
Accuracy = \ frac {TP + TN} {TP + TN + FP + FN}
It is the ratio of the number of correctly recognized cases to the total number of cases.

2. Precision

Precision measures the number of the webpages that are predicted to be malicious but are actually malicious. Low precision implies reduced false alarms.

3. Recall (Sensitivity)

Recall always uses to measure how the model is capable to detect malicious webpages.
Recall = TP/TP+FN

Recalling is high so that a majority of the malicious webpages are identified.

4. F1-Score

The mean of recall and precision is called F1-Score, and it gives a balance between them.
F1-Score = 2 X Precision X Recall / Precision + Recall
It comes in handy where the false positives are significant as well as false negatives.

5. False Positive Rate (FPR)

This measure is used to show the rate of false alarms of benign webpages.
'FPR' = FP / FP +TN
A smaller FPR would be preferable in order to prevent unwarranted warnings.

6. False Negative Rate (FNR)

This is used to measure the frequency of missed malicious webpages by the system.
'FNR' = FN / FN + TP
FNR of lower importance is essential to security applications.

7. Model Efficiency

Besides measures of accuracy, the system is tested by:

- Detection Time: This is the time taken to categorize a webpage.
- Scalability: Capability of processing big data.

V. Research Analysis

The study is aimed at overcoming the weaknesses of the conventional malicious webpage detection systems, which are mostly targeted to desktop. The layout, functionality, and patterns of interaction of mobile webpages vary greatly, thus the current techniques are ineffective.

The proposed system has a mobile-adapted system that involves applying the feature analysis at rest with the use of the static features. The study shows that the characteristics like redirections and iframes which are good pointers of ill intent in desktop contexts act unlike those in mobile webpages.

Key Research Findings
Mobile web pages need different detection mechanical analysis because of the dissimilarities in their format.
Static features such as:

- Number of iframes
- URL patterns
- JavaScript behavior
- Mobile-specific features (e.g. click-to-call links) are strongly correlated with evil action.
- The system was tested by a dataset of more than 350,000 webpages.

The system is effective in detecting threats that are overlooked by the available tools like Google safe browsing and Virus total.

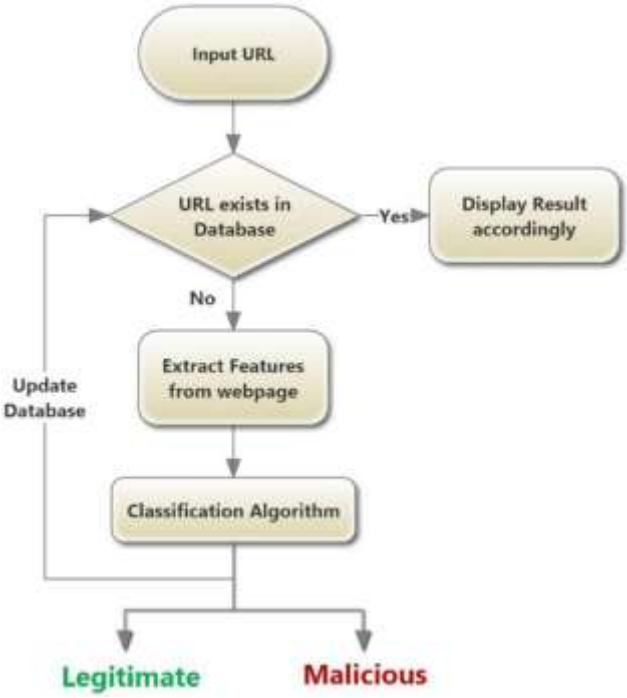

Fig.1 Workflow of the website

VI. Performance Analysis

The proposed system is tested with respect to standard classification metrics and experimental validation of the performance of the suggested system.

1. Accuracy Performance

The system has about 90 percent accuracy of detecting malicious mobile web pages.

- Shows good classifying of both benign and malicious pages.
- Similar or more superior to conventional detection systems.

2. True Positive rate (Detection rate)

This is the proportion of the number of infected individuals correctly identified as infected by the test to the total number of infected persons detected by the test.<|human|>True Positive rate (Detection rate):

This is the ratio of the number of infected persons who were accurately identified as infected by the test to the total number of infected persons who were detected by the test.
Reaches approximately 89% true positive, i.e. majority of the malicious pages are identified with accuracy.

3. False Positive Analysis

- Good false positive means that the correct webpages are not wrongly identified.
- Critical to keep the user trust and usability.

4. Comparative Performance

The system will be contrasted with the current solutions:

- System
- Detection Capability
- Limitations
- Traditional Blacklists
- Detect known threats
- Cannot detect new attacks
- Machine Learning Models
- High accuracy

It is more efficient in identifying mobile threats and unknown attacks.

5. Efficiency Analysis

- The method of the static analysis makes sure that it is detected more quickly than dynamic ones.
- Compared to conventional methods, feature extraction is much faster.
- Appropriate in real-time application in mobile browsers.

6. Scalability

Supports huge amounts of data (350,000 and more webpages) well.
Can be extended to:

- Larger datasets
- Different mobile platforms
- Real-world deployment

7. Robustness

- Capable of detecting:
- Phishing attacks
- Scam webpages
- Call-based attacks of a fraudulent nature.
- Detects even malicious pages that were not detected before.

8. Limitations in Performance

- Dynamic or runtime attacks can not be detected by use of a static analysis.
- Quality of training dataset is related to performance.
- May needs periodical model revision.

VII. CONCLUSION

The study examined in this paper offers a mobile- centric method, which is a real-time detector of malicious webpages based on both a static feature analysis and machine learning methodology. Systems that are used in traditional detection systems do not work well on mobile webpages since they have differences in structure, user interaction and functionality.

The system proposed is a solution to these issues, as it identifies the relevant statistic features, including URL properties, HTML composition, JavaScript functionality, and mobile features such as click to call links. The system is trained using a large set of more than 350,000 webpages where it is trained to correctly identify webpages as benign or malicious.

The records of the analysis denotes that the system is about 90 percent accurate, which proves its efficiency to detect mobile-based threats. Moreover, user security can be increased through the use of a browser extension to detect in real-time, thereby increasing the level of security when browsing is done on a mobile device.

In general, the suggested system offers a scalable, lightweight, and efficient system of malicious mobile webpage detection, as well as makes a significant contribution to enhancing mobile web security.

VIII. Future Work

In spite of the promising results of the proposed system, a number of areas should be improved and enhanced in the future:

1. Dynamic Analysis Integration.

Dynamic analysis methods can be used in future work to identify the pattern of script execution and user interaction to further improve the detection of sophisticated attacks.

2. Deep Learning-Based Models

Developed models like Deep Neural Networks (DNN), LSTM, and Transformer-based models can be considered to improve the accuracy of detection and the detection of more sophisticated attack patterns.

3. Real-Time Cloud Integration

The system can also be combined with the use of cloud-based services in order to result in the constant updates, the extensive threat intelligence and the enhanced performance of the detection.

4. Cross-Platform Compatibility

Future applications can be implemented on to various platforms including:

- Android and iOS applications.
- Different mobile browsers
- Progressive and hybrid web applications.

5. Minimization of False Positives.

The false positives can be minimized with the help of additional optimization methods, so that the users will experience a greater level of confidence and satisfaction with the system.

6. The Adaptive Learning Mechanism.

The system will be made to update itself automatically through incorporation of online learning or adaptive models by adhering to new threats and emerging attack patterns.

REFERENCES

[1] C. Amrutkar and Y. S. Kim, "Detecting Mobile Malicious Webpages in Real Time," IEEE Trans. Mobile Comput., 2017.
dl.acm.org

[2] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," ACM Computing Surveys, 2019.
researchgate.net

[3] Mobile Malicious WebPages Detection in Real-Time, Int. J. Res., 2021.
ijracse.com

[4] Detecting Mobile Malicious Pages in Real Time, IJRaset, 2020.
IJRASET

[5] S. Jain and A. Jain, "Real-Time Detection of Malicious URLs by Utilizing Machine Learning Techniques," Proc. IEEE Conf., 2024. Scribd

[6] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," arXiv, 2017 — Comprehensive survey on ML techniques for malicious URL detection, covering feature extraction and algorithm design.
researchgate.net

[7] N. Reyes-Dorta, P. Caballero-Gil, C. Rosa-Remedios, "Detection of Malicious URLs Using Machine Learning," Wireless Networks, 2024 — Examines ML techniques and even quantum ML approaches for malicious URL detection.
researchgate.net

[8] Yichen Wang, "Malicious URL Detection: An Evaluation of Feature Extraction and Machine Learning Algorithm," Highlights in Science, Engineering and Technology, 2022 — Discusses feature extraction strategies and ML models.
researchgate.net

[9] A. Indian, "Identifying Malicious URLs Using Deep Learning Based Approaches," ScienceDirect, 2025 — Deep learning based model for malicious URL detection.
ScienceDirect

[10] Apoorva Joshi, Levi Lloyd, Paul Westin, Srini Seethapathy, "Using Lexical Features for Malicious URL Detection — A Machine Learning Approach," arXiv, 2019 — Shows lexical feature based ML for fast detection.
arXiv

[11] Hung Le, Quang Pham, D. Sahoo, S.C.H. Hoi, "URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection," arXiv, 2018 — Uses deep learning to learn URL patterns automatically. arXiv

[12] Ehsan Nowroozi et al., "An Adversarial Attack Analysis on Malicious Advertisement URL Detection Framework," arXiv, 2022 — Explores adversarial ML and detection robustness.
arXiv

[13] M. Maftoun et al., "Malicious URL Detection using optimized Hist Gradient Boosting Classifier," arXiv, 2024 — Recent work applying optimized ML models for URL detection.
arXiv