Phishing Website Detection using Machine Learning Rules with Cryptography Technique

S. Thamizhazhaki[1], J. Pragathi[2]

[1]PG Student, Computer Science and Engineering, Sir Issac Newton College of Engineering and Technology, Nagapattinam, Tamilnadu, India

[2]Assistant Professor, Computer Science and Engineering, Sir Issac Newton College of Engineering and Technology, Nagapattinam, Tamilnadu, India

Abstract: ***Malicious URLs and websites pose a frequent and serious risk to online safety. Search engines naturally become the foundation of information management. However, the proliferation of fraudulent websites on search engines has put our users at serious risk. Most current techniques for detecting rogue websites concentrate on a particular assault. At the same time, numerous websites are unaffected by the readily available browser extensions based on blacklists. Since the server cannot deduce any useful information from the masked data, it is crucial that any data leaving the client side be effectively masked. The initial PPSB service is here suggested. Strong security guarantees are provided, something that is lacking in the current SB services. The proposed approach detects the malicious URL access with the help of blacklist storage. The input URL (given by user) was classified with the help of SVM classification. SVM is a type of machine learning algorithm that accurately detect, whether the URL is safe or unsafe. In particular, it carries over the capacity to recognize harmful URLs while safeguarding the browsing history and proprietary information of the blacklist provider (the list of unsafe URLs) as well as the user's privacy. In order to protect user privacy from outside analysts and service providers, a model that encrypts sensitive data was presented in this work. Additionally fully supports selective aggregate functions for analysing online user behaviour and ensuring differential privacy. Data about users' online behaviour is encrypted using the Homomorphic RSA technique.***

Keywords: Malicious URL Detection, Blacklist Creation, History encryption using Homomorphic RSA, URL Recommendation, Key verification, History Access

1. Introduction

1.1 Network Security

Managing security is the process of understanding the risks and identifying how plenty danger is appropriate. Different stages of security are suitable for specific organizations. No network is fully secure, so don’t intention for that level of protection. If you attempt to stay up-to-date on each new danger and every virus, you’ll soon be anxiety and strain. Here present more benefits of computer networks and the Internet. Connecting your network to the Internet presents access to a full-size amount of facts and lets in you to share information on a first rate scale. However, the communal nature of the Internet, which creates so many advantages, also offers malicious users access to several targets. The Internet is only as secure because the networks it connects, so we all have a responsibility to ensure the safety of our networks.

Information safety is the process of securing statistics records from unauthorized access, use, change, tempering, or disclosure. With the extended use of electronic media in our personal lives as well as corporations, the opportunity of security breach and its main effect has elevated. The robbery of personal identity, credit card data, and other critical facts the uses of hacked user names and passwords havegrown to be common place these days. In addition, the theft of exclusive business statistics might also lead to loss of enterprise for industrial groups.

1.2 Web Security

The Internet is a risky location, with extremely good regularity, customers listen about websites becoming unavailable due to denial of provider attacks, or showing modified (and regularly unfavorable) information on their homepages. In different excessive-profile cases, hundreds of thousands of passwords, electronic mail addresses, and credit score card details were leaked into the public area, exposing website customers to both private embarrassment and financial danger.

The purpose of internet site safety is to prevent those (or any) kinds of attacks. The more formal definition of website safety is the act/practice of protective websites from unauthorized access, use, modification, destruction, or disruption.Effective internet site security requires layout attempt throughout the entire of the website: to your net utility, the configuration of the web server, your rules for growing and renewing passwords, and the patron-facet code. While all that sounds very ominous, the coolest information is that in case you're the usage of a server-aspect net framework, it will almost actually enable "with the aid of default" robust and properly-thought-out protection mechanisms in opposition to some of the extra not unusual assaults. Other attacks can be mitigated thru your net server configuration, for instance with the aid of permitting HTTPS. Finally, there is publically available vulnerability scanners gear that let you find out in case you've made any obvious mistakes.

1.3 Safe Browsing

Malicious SB service issuer wants to recognize whether a person is journeying a specific web page, e.g., some political information. One way to gain that is that the web browser sends all the visited URLs to a far off server, either inside the plaintext, hash cost or encrypted layout. However, this behavior may be detected by tracking and analyzing the browser, e.g., the usage of the taint evaluation technique. Specifically, as a way to track a particular URL the SB carrier issuer can insert the 32-bit hash prefixes of all its decompositions, e.g., c01e362f, after which push this newly up to date prefix filter out to the customers. Later, once a user visits the internet page (or comparable URLs that percentage a few decompositions), the matched hash prefixes might be sent to the far flung SB server. Based on the prior information of the prefix filter (i.e., the mappings among the hash prefixes and their corresponding URLs), the server can infer the URL (or area) navigated by means of the user. It gives strong protection ensures that are lacking in present SB services. In precise, it inherits the capability of detecting dangerous URLs, at the same time as on the identical time protects both the person’s privacy (surfing records) and blacklist provider’s proprietary belongings (the list of risky URLs). This approach has a few disadvantages along with; developing metadata of URLs fails while the server gets multiple prefixes for a URL and there may be a threat that other URLs may additionally have the equal hash prefixes this makes collision among URLs.

A malicious user would possibly leverage PPSB to degrade the consumer-facet consumer experience, like putting a number of faux or secure URLs or increasing the server-aspect delay. To cope with this capability difficulty, PPSB presents a flexible mechanism for customers to add or eliminate blacklist providers. Admin ought to add the fake URL and keyword to this blacklist storage. User can also allowed suggesting the malicious internet site info concerning black list. In this gadget malware detection machine makes use of a supervised machine gaining knowledge of technique for discovering malwares. The SVM classification with malware detection system extends the idea of signature primarily based detection system with a aggregate of conduct tracking approach. It utilizes static and dynamic evaluation of malwares with the aid of taking the run time traces of the executables. Image based malicious detection also provide to compare the image functions based totally on original internet site and malicious website. This version also affords seek records security which encrypts the users’ sensitive statistics to save you privateness from both outside analysts and the aggregation provider. Also, completely helps selective combination functions for on-line consumer conduct analysis and ensuring differential privateness.

2. Related Work

Cuiet al.propose the fingerprint techniques and locality-sensitive hashing to transform the hassle of NDD into the keyword seek. Initially, the CPs encrypts their information objects with a general encryption scheme, e.g., AES. And the ISP will connect these metadata together with corresponding encrypted data storage and supply them to ISs that are close to the users. The user might be capable of access to the encrypted data with the aid of the CP and generate encrypted queries for secure NDD along with her very own key. Stage Two - Secure Detection: In order to find near replica records items from the encrypted in-network storage, the user wishes to generate an encrypted query tq from the fascinated records with her own key and ship it to the closest IS. Stage Three - In order to in addition improve the exceptional of query consequences, the IS needs to filter out the ability fake positives from the placed candidates. Hence, the IS and the ESP will behavior a comfortable evaluation system through Yao’s garbled circuits protocol. In specific, the ESP (because the garbled-circuit generator) prepares a garbled circuit for the IS (because the garbled-circuit evaluator), wherein the circuits characteristic assessments if the query item and each candidate are certainly close to-duplicates primarily based on a particular distance metric. For those certified applicants, the IS would return them at once to the consumer.

Cuiet al. propose a privacy based malware detection carrier for Android, in which the privateness (or property) of telephone companies, customers, and protection carrier companies are covered. It detects malicious application present in mobile dealer’s app stores and on users’ phones, without immediately sharing apps, apps’ runtime behaviors, and malware signatures to different parties. Proposed design purpose is to permit the SPs carry out malware detection without holding the apps. Thus, the above-mentioned capabilities, together with the permissions, behavioral footprints, and record hashes are always minimum leakage approximately the apps. The SPs can't (without difficulty) recover the original value to the one-manner belongings of the chosen cryptographic hash functions. Even a SP may additionally moreover guess the unique values with the assist of brute-pressure attack, the price of the manner and the price of the recovered talents do no longer deserve this strive. Remember the code of the app is not shared and cannot be recovered at all.

Shengshanet al. implement the primary practical tool for privateness primarily based pass-media retrieval thru using trusted processors. Proposed scheme permits secure aggregation of the statistics from distinct parties, and secures canonical correlation evaluation (CCA) over collaborated records to reap semantic models. Verification mechanisms are designed to shield toward active assaults from a malicious adversary. SGX offers records sealing and unsealing features to protect personal facts outdoor the boundary of an enclave. The sealing facilities offer each enclave with keys that are precise to the processor and the enclave digest. When the enclave is closed, the private facts can be sealed and stored on the platform in the encrypted shape. It aims at addressing the hassle of the way to securely behavior cross-media retrieval over encrypted records strategically the use of a far off SGX-enabled server.

Wanget al. proposed system investigates the problem of excessive fine phrase vectors over big-scale encrypted data with the privacy-maintaining collaborative neural network algorithms. It carries multiple users (i.e., records owners), a primary faraway server S and a crypto provider provider

CSP. More specifically, there are n customers in overall, denoted as ui (i = 1 . . . N), every of which owns a non-public data document fi and desires to perform collaborative neural network getting to know with all different participating customers. In different words, they will make a contribution their personal information files in the encrypted shape to the imperative remote server. After receiving the encrypted records, the far off server S plays training over the contributed records and produces amazing word vectors in conjunction with a model, which may be used later for various natural language processing (NLP) duties.

Yuanet al. propose the frequency hiding query scheme which lets in the server to look the flattened question distribution most effective. To enhance the scalability, further design the result sharing query scheme, which approaches a small portion of query points and stocks the effects with different close by factors. Besides, we installation a strict constraint to carefully pick question factors to reap "as-strong-as-possible" ensures. It consists of three parties, i.e., the consumer of the legal user, the statistics owner who has the source dataset and the server inside the public cloud. Before the use of our device for secure similarity be a part of queries, a setup manner is needed. The facts proprietor will construct an encrypted LSH-based index I, encrypt the dataset S, and add them to the cloud server. After that, the consumer will pre-system the question set Q and generate relaxed tokens t from LSH hash values of question points. When the server gets t, it's going to system them over I to get a fixed of ids of collided records factors. When the variety of information factor's hash collisions to a question factor is more than a pre-described collision threshold α, those information and question factors can be considered as a candidate pair.

3. Existing Methodologies

Phishing is the fraudulent activity to get sensitive records inclusive of usernames, passwords and credit score card info, frequently for malicious motives, with the aid of disguising as a sincere entity in an digital conversation. Phishing attack can be carried out in various form like Email phishing, Website phishing, spear phishing, Whaling, Tab napping, Evil dual phishing and many others. To avoid this phishing attack various anti-phishing applications have to be use. There are diverse anti phishing solutions inclusive of Blacklist, heuristic, visible similarity, machine learning techniques and many others.

This is maximum usually used method in which list of phishing URL is stored in database after which if URL is found in database, it's miles referred to as phishing U and offers warning otherwise it's far referred to as legitimate. This technique is simple and quicker to put in force as it see URL is in db or not. But limitations are small trade in URL is sufficient to skip the list based totally technique and frequent replace of listing is important to counter new attack.

Phishing imitates the traits and functions of emails and makes it appearance similar to the original one. It appears much like that of the legitimate supply. The user thinks that this electronic mail has come from a genuine business enterprise or an business enterprise. This makes the consumer to forcefully go to the phishing website via the links given inside the phishing email. These phishing websites are made to mock the arrival of an unique organisation internet site. The phishers force consumer to top off the personal facts through giving alarming messages or validate account messages and so forth so they replenish the desired information which can be utilized by them to misuse it. They make the situation such that the person isn't left with another option but to go to their spoofed website.

In the training phase, we ought to use the categorized facts wherein there are samples consisting of phish region and valid area. If we try this then category will now not be a hassle for detecting the phishing area. To do a operating detection version it's miles very critical to use records set inside the schooling section. We need to use samples whose lessons are regarded to us, which means that the samples that we label as phishing ought to be detected best as phishing. Similarly the samples which are classified as legitimate might be detected as legitimate URL. The dataset to be used for machine learning without a doubt consist those features. There so many machine learning algorithms and every set of rules has its very own operating mechanism which we have already seen in the previous work. The existing device uses any one of the best machine learning algorithms for the detection of phishing URL and predicts its accuracy. The present device has good accuracy but its miles still not the satisfactory as phishing assault is a totally crucial; we ought to find a fine approach to put off this. In the currently existing device, most effective one device mastering algorithm is used to be expecting the accuracy; the usage of most effective one algorithm isn't a terrific method to improve the prediction accuracy.

4. Malicious URL Detection with History Encryption Approach

A malicious user might leverage PPSB to degrade the client-side consumer experience, like putting a number of fake or safe URLs or growing the server-side postpone. To deal with this ability problem, PPSB gives a reliable mechanism for users to feature or put off blacklist providers. Admin should add the unsafe URL and keyword to this blacklist storage system. User also can allowed suggesting the malicious internet site info regarding black listing. In this proposed application malware detection system makes use of a supervised learning method for discovering malwares. The SVM based malware detection device extends the concept of signature based detection with a mixture of behavior tracking approach. It makes use of static and dynamic evaluation of malwares by way of taking the run time traces of the executable. Keyword based malicious detection also provide to compare the keyword capabilities based on unique website and malicious website. This version also gives seek data protection which encrypts the customers' access history data to prevent privateness from both outside analysts and the aggregation provider issuer. Also, completely helps selective combination functions for on line user conduct evaluation and making sure differential privacy.

In this system malware detection approach uses a supervised machine learning approach for discovering malwares. The

SVM based malware detection system extends the concept of signature based detection system with a mixture of behavior tracking approach. Also, completely supports selective combination capabilities for on line user behavior evaluation and making certain differential privacy. A malicious user would possibly leverage PPSB (Privacy Preserving Safe Browsing) to degrade the consumer-aspect user revel in, like inserting a number of fake or secure URLs or increasing the server-side delay. To deal with this potential problem, PPSB affords a flexible mechanism for users to add or eliminate blacklist providers. Admin could add the faux URL and key-word to this blacklist storage. User also can allowed suggesting the malicious internet site information regarding black list.

The cause of internet site security is to prevent these (or any) sorts of attacks. Effective website protection calls for layout attempt throughout the entire of the website: in net application, the configuration of the internet server, user policies for creating and renewing passwords, and the patron-aspect code. The proposed assignment detects Malicious or Fake URLs to prevent the customers accessing from Unsafe URLs. Also offer comfortable encryption method to encrypt the consumer search records earlier than saved at the server.

5. Algorithm

Homomorphic RSA ALGORITHM

Homomorphic RSA is a public key cryptosystem. RSA is an asymmetric cryptographic algorithm named after its founders Rivest, Shamir &Adelman. It generates two keys: public key for encryption and private key to decrypt message.RSA algorithm consists of three steps. Step one is key generation which is to be used to generate keys to encrypt and decrypt data. Step two is encryption, where actual process of conversion of plaintext to cipher text is being carried out. Step three is decryption, where encrypted text is converted in to plain text at other side.RSA is based on factoring problem of finding product of two large prime numbers. Key size is 1024 to 4096 bits.

Key Generation Algorithm

This is the original algorithm.

1) Generate two large random primes, p and q, of approximately equal size such that their product n=pq is of the required bit length, e.g. 1024 bits.
2) Compute n=pq and $\phi=(p-1)(q-1)$.
3) Choose an integer e, $1<e<\phi$, such that gcd (e, ϕ) =1.
4) Compute the secret exponent d, $1<d<\phi$, such that $ed\equiv 1 \bmod \phi$.
5) The public key is (n, e) and the private key (d, p, q). Keep all the values d, p, q and ϕ secret.

n is known as the modulus.
e is known as the public exponent or encryption exponent or just the exponent.
d is known as the secret exponent or decryption exponent.

Encryption

Sender does the following:-

1) Obtains the receiver's public key (n, e).
2) Represents the plaintext message as a positive integer m with $1<m<n$.
3) Computes the ciphertext $c = m^e \bmod n$.
4) Sends the ciphertext c to receiver.

Decryption

Receiver does the following:-

1) Uses the private key (n, d) to compute $m = c^d \bmod n$.
2) Extracts the plaintext from the message representative m.

Support Vector Machine

Support Vector Machine (SVM) is a supervised algorithm based on machine learning which can be used for both classification and regression problems. However, it's far ordinarily used in classification work. In this work, plot each data item as a point in n-dimensional space with the value of every feature being the count of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is best for segregates the two classes (hyper-plane/ line). The hyperplane is the line with the biggest margin to both groups.

Support Vector Machines has higher effectiveness in higher dimensional spaces. It is even very effective on data sets where number of dimensions is greater than the number of samples. This is mainly because of the kernel trick, which we talk about it later. Further advantages of Support Vector Machines are the memory efficiency, speed and general accuracy in comparison to other classification methods like k-nearest neighbor or deep neural networks.
Step1: Malicious URLs and keywords have been collected and stored on blacklist storage.
Step 2: For the collection of Malicious URLs number of features could be used like URL length, the number of dots, ip Address, SSL connection, at symbol(@) and dash symbol(-).
Step 3: The selected features identified from URL then stored on blacklist.
Step 4: User could enter the URL or Keyword for searching details.
Step 5: Input data classified with trained dataset with the help of SVM classifier.
Step 6: SVM classifier returns either an URL is phishing or non-phishing.

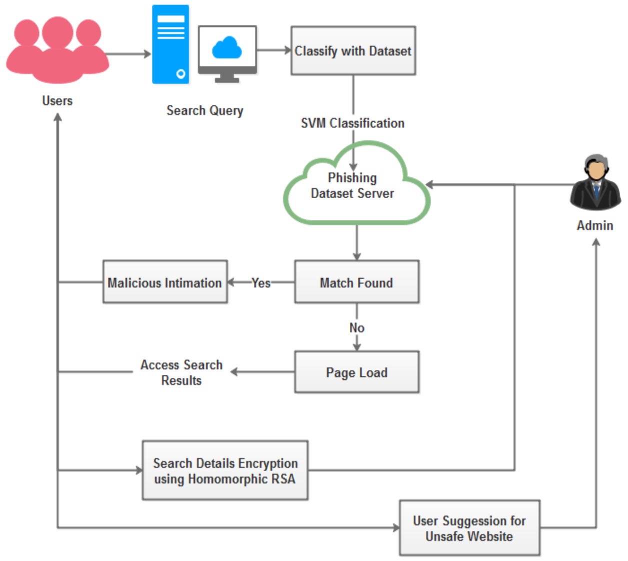


Figure 4.1: Proposed work Architecture

The blacklist providers have the incentive to collect and publish unsafe URLs and keywords for helping users to avoid websites that contain malware or phishing and deceptive content, e.g., for the better marketing purpose. Assume that the current blacklist providers and the corresponding PPSB servers are semi-trusted. They faithfully perform the designed procedures, i.e., the database preparation/ update. But they should not be aware of the queried URLs from users. In proposed work a service provider that owns a high-quality blacklist, which may be more frequently updated or simply contains more items. User also allowed to directly sharing blacklists with servers in an uncontrollable way could make these dataset be obtained by every user, including the competitors. The client needs to search into the list of unsafe URLs and keywords. The searched URL could be matched with blacklist providers. Once match could be found, that will show the malicious alert to the user. Otherwise the page will be loaded to show the details to searched user. In further the searched URLs or Keywords are stored in encrypted format, which will not reveal the users sensitive information to the server or unknown person.

The proposed work consists of the following modules,

- Framework Construction
- URL Encryption
- User Registration and Login
- Unsafe URL Detection
- History Access
- Malicious URL Suggestion

Module Descriptions

Framework Construction

The detection of malicious URLs limits web-based attacks by preventing web users from visiting malicious URLs and warning web users prior to accessing content located at a malicious URL. Thus, malicious URL detection protects computing system hardware/software from computer viruses, prevents execution of malicious or unwanted software, and helps avoid accessing malicious URLs web users do not want to visit. This proposed framework uses SVM classification models to detect a malicious URL and categorize the malicious URL as one of a phishing URL.

URL Encryption

The blacklist storage models by using a set of training data (unsafe URLs and keywords) and machine learning algorithms. The training data includes a known set of unsafe URLs and a known set of malicious keywords. This framework also supports URL encryption process, to avoid the unauthorized prediction of URL details. In blacklist storage, keyword and URL will be encrypted and stored in the intermediate. AES encryption algorithm used for encrypting data before stored on blacklist. This encryption also provides the security for user search history.

User Registration and Login

Users have to register with their name, password and Email id. These details will be saved in the database. The user have to login with the name and password. The entered data will be compared with the available data. If match found, the user can proceed. If match not found during search, the user have to re-enter the details again. This process will helps to protect from unauthorized access in search engine and also helps to predict the user who add the blacklist.

Once the login procedure is succeeded, the user can search details using URLs and keywords. The user will enter a URL or keyword in the search box and click the submit button. When the user clicks the search button, the request was processed and related details are shown to the user.

Unsafe URL Detection

The verification of URLs and keywords is very essential in order to ensure that user should be prevented from visiting malicious websites. SVM mechanisms have been proposed to detect the malicious URLs. One of the basic features that a mechanism should possess is to allow the fake URLs that are requested by the client and prevent the malicious URLs before reaching the user. This is achieved by notifying the user that it was a malicious website. The SVM technique extract URL (or) keyword features associated with the known URLs, and use the machine learning algorithms to train the classification models to detect and categorize an unknown malicious URL. A database updation is performed every time the systems come across a new URL. Here, the new URL will be matched and tested with every previously known malicious URL in the black list. The update has to be made in black list whenever system comes across a new malicious URL. This also allows users to provide suggestions to add malicious URLs.

History Access

All the users of the same application will not be allowed to access the user search data. Privileges are given by the administrator to the users. A malicious user in this case can be any user in the network who is not authorized to access the users search data but can intrude by using others credentials. This problem of intrusion can be overcome by providing Secret Key Sharing process generated by the system which will be forwarded only to authorized users. The intruder cannot be able to access the database without the shared secret key. If user wants to access search history, they will share request to the admin for access permission.

After getting permission from admin, they will allow accessing their search history in plain text format. This approach will enhance the security of history protection process.

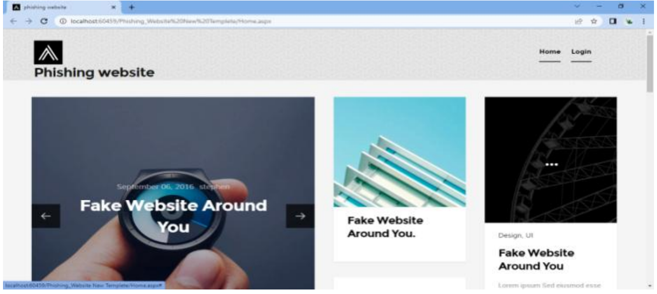
Malicious URL Suggestion

In proposed work, the URL suggestion process could be implementing to enhance the performance of blacklist storage. When user finds any malicious URL during searching process, they will allow to suggestion process. Here user should send URL details to admin, to add blacklist storage. This frequent update in blacklist improves the performance of unsafe (or) malicious URL detection.

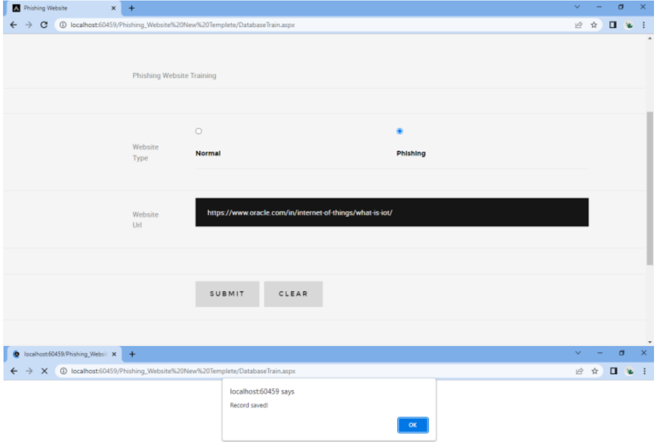
6. Experimental Results

An experimental result shows the process of secure web search with access history encryption. This proposed approach was implemented using ASP.NET as front end and SQL Server as back end software.

Home Page



Add Phishing URL on Blacklist

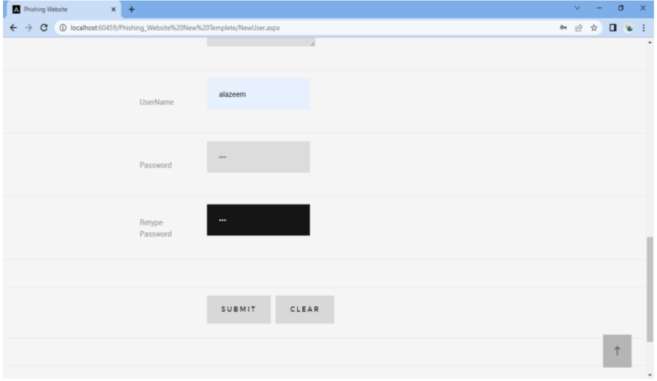

localhost:60459/Phishing_Website%20New%20Templete/DatabaseTrain.aspx
Phishing Website Training
Website Type
Normal
Phishing
Website Url
SUBMIT
CLEAR
OK

Website Url: https://www.oracle.com/in/internet-of-things/what-is-iot/

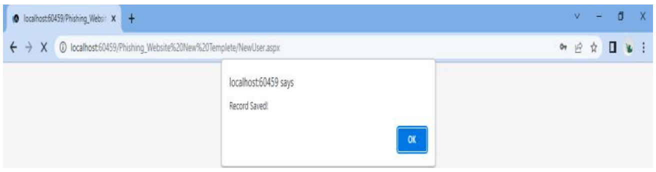
localhost:60459 says

Record saved!

User Register


localhost:60459/Phishing_Website%20New%20Templete/NewUser.aspx
UserName
Password
Retype-Password
SUBMIT
CLEAR

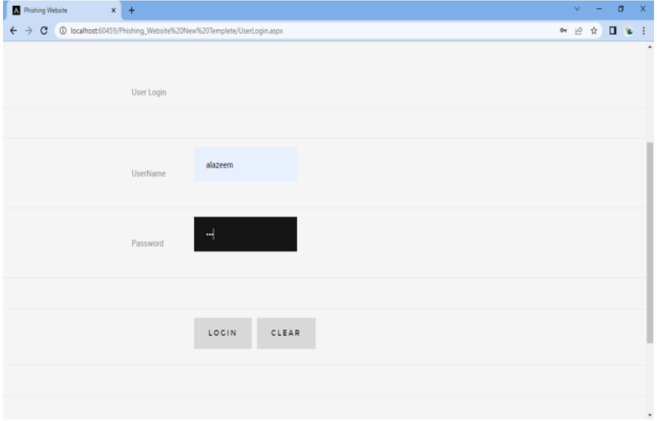
UserName: alazeem


localhost:60459/Phishing_Website%20New%20Templete/NewUser.aspx
OK

localhost:60459 says

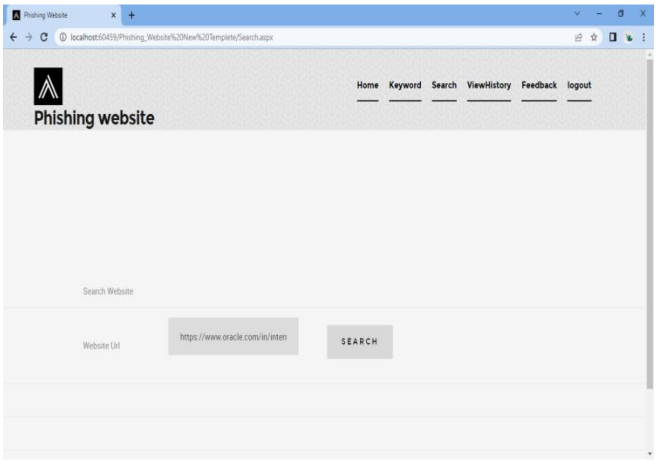
Record Saved!

User Login


localhost:60459/Phishing_Website%20New%20Templete/UserLogin.aspx
User Login
UserName
Password
LOGIN
CLEAR

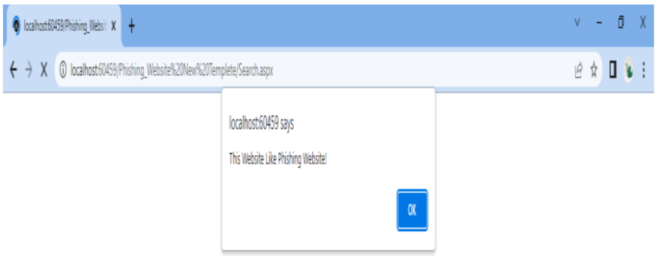
UserName: alazeem

Search URL



Website Url: https://www.oracle.com/in/intern

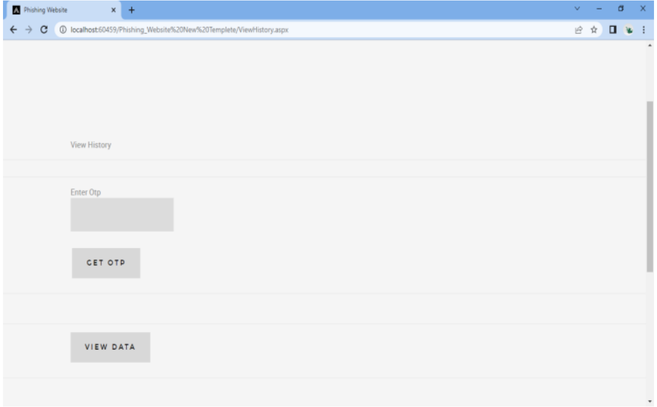
Malicious URL Detection

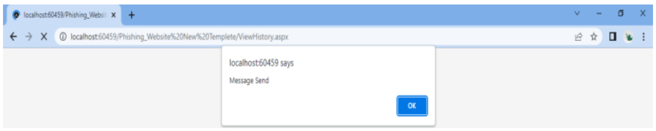

localhost:60459/Phishing_Website%20New%20Templete/Search.aspx
OK

localhost:60459 says

This Website Like Phishing Website!

Key Verification for History Access

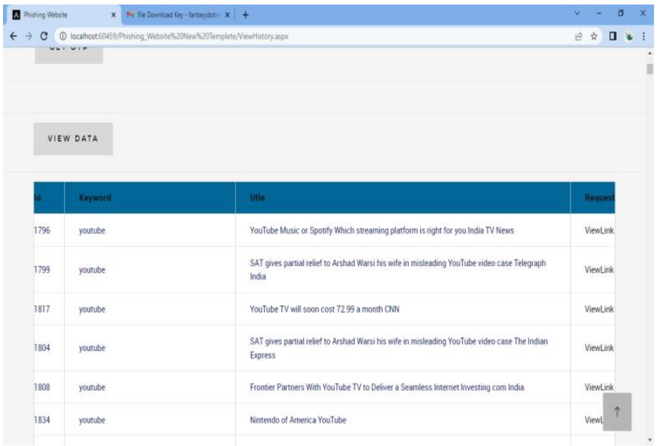

localhost:60459/Phishing_Website%20New%20Templete/ViewHistory.aspx
View History
Enter Otp
GET OTP
VIEW DATA


localhost:60459/Phishing_Websit
localhost:60459/Phishing_Website%20New%20Templete/ViewHistory.aspx

localhost:60459 says

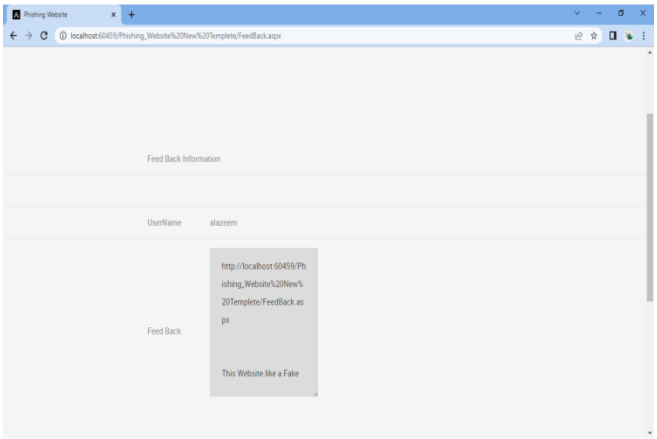
Message Send

View Access History


Phishing Website
file Download Key - fantasydotn
localhost:60459/Phishing_Website%20New%20Templete/ViewHistory.aspx
VIEW DATA

Id	Keyword	title	Request
1796	youtube	YouTube Music or Spotify Which streaming platform is right for you India TV News	ViewLink
1799	youtube	SAT gives partial relief to Arshad Warsi his wife in misleading YouTube video case Telegraph India	ViewLink
1817	youtube	YouTube TV will soon cost 72 99 a month CNN	ViewLink
1804	youtube	SAT gives partial relief to Arshad Warsi his wife in misleading YouTube video case The Indian Express	ViewLink
1808	youtube	Frontier Partners With YouTube TV to Deliver a Seamless Internet Investing com India	ViewLink
1834	youtube	Nintendo of America YouTube	ViewL

Feedback about Malicious Websites


Phishing Website
localhost:60459/Phishing_Website%20New%20Templete/FeedBack.aspx

Feed Back Information

UserName alazeem

Feed Back http://localhost:60459/Phishing_Website%20New%20Templete/FeedBack.aspx

This Website like a Fake

7. Conclusion

In this proposed work, implement a Malicious URL Detection process using machine learning techniques. This focuses on detecting unsafe website URLs and keywords with the help of encrypted blacklist storage. According to few selected features can be used to differentiate between legitimate and malicious web pages. These selected features are many such as URLs and Keywords. In this proposed safe browsing website a service provider that has a high-quality blacklist, which may be more frequently updated or simply contains more items. User also allowed to directly sharing blacklists with servers in an uncontrollable way could make these dataset be obtained by every user. With the help of efficient classification approach will detect the fake websites accurately and prevent the users from accessing that websites. This also provides the secure encryption approach avoid the unknown access of search history. The security is provided to the search data which has been stored in the database.

References

[1] Cui, Helei, Xingliang Yuan, YifengZheng, and Cong Wang. "Towards Encrypted In-Network Storage Services with Secure Near-Duplicate Detection." IEEE Transactions on Services Computing (2018).

[2] Cui, Helei, Yajin Zhou, Cong Wang, Qi Li, and KuiRen. "Towards Privacy-Preserving Malware Detection Systems for Android."In 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS), pp. 545-552.IEEE, 2018.

[3] Hu, Shengshan, Leo Yu Zhang, Qian Wang, Zhan Qin, and Cong Wang. "Towards private and scalable cross-media retrieval." IEEE Transactions on Dependable and Secure Computing (2019).

[4] Wang, Qian, Minxin Du, Xiuying Chen, Yanjiao Chen, Pan Zhou, Xiaofeng Chen, and Xinyi Huang. "Privacy-preserving collaborative model learning: The case of word vector training." IEEE Transactions on Knowledge and Data Engineering 30, no. 12 (2018): 2381-2393.

[5] Yuan, Xingliang, Xinyu Wang, Cong Wang, Chenyun Yu, and SaranaNutanong. "Privacy-preserving similarity joins over encrypted data." IEEE Transactions on Information Forensics and Security 12, no. 11 (2017): 2763-2775.

[6] Ramezanian, Sara, TommiMeskanen, MasoudNaderpour, Ville Junnila, and ValtteriNiemi. "Private membership test protocol with low communication complexity." Digital Communications and Networks (2019).

[7] Keelveedhi, Sriram, MihirBellare, and Thomas Ristenpart. "DupLESS: server-aided encryption for deduplicated storage." In Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13), pp. 179-194. 2013.

[8] Armknecht, Frederik, Jens-Matthias Bohli, Ghassan O. Karame, and Franck Youssef. "Transparent data deduplication in the cloud." In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 886-900.ACM, 2015.

[9] Demir, Levent, Amrit Kumar, Mathieu Cunche, and CédricLauradoux. "The pitfalls of hashing for privacy." IEEE Communications Surveys & Tutorials 20, no. 1 (2017): 551-565.

[10] Gerbet, Thomas, Amrit Kumar, and CédricLauradoux. "A privacy analysis of Google and Yandex safe browsing." In 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 347-358. IEEE, 2016.