As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 7(09): 182-185, 2021

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0709029
Available online at: http://www.ijmtst.com/vol7issue09.html

The Conflict-Free Replicated Datatype Algorithm for Detecting E-Commerce Phishing Website

Shonal Rath

BA Economics (Hons), University of Delhi

To Cite this Article
Shonal Rath. The Conflict-Free Replicated Datatype Algorithm for Detecting E-Commerce Phishing Website. *International Journal for Modern Trends in Science and Technology* 2021, *7* pp. 182-185. https://doi.org/10.46501/IJMTST0709029

Article Info
Received: 22 August 2021; Accepted: 16 September 2021; Published: 19 September 2021

ABSTRACT

By posing as a trustworthy company through electronic communication, a fraud effort is made to get sensitive and personal information such as passwords, usernames, and financial data such as credit/debit card numbers. The phishing website will seem just like the real website and will send the visitor to a page on the malicious website where they will be asked to submit personal information. Machine learning techniques may be used to increase the accuracy of phishing website prediction. The suggested technique predicts URL-based phishing websites using W3C-defined characteristics and provides the highest level of accuracy. This technique makes advantage of URL characteristics. Phishing site urls have certain characteristics, which we discovered.

We are all experiencing a worldwide epidemic, which has resulted in a significant shift in people's lifestyles. People were completely unprepared for such an event. As a result, everyone has had a tough time dealing with the circumstance. However, as we all know, life moves on, and individuals have changed their lifestyles and adapted to the present circumstances. As a result, the internet and online apps have become one of the most essential aspects of people's working environments. People's jobs are becoming increasingly reliant on internet applications. As a result, the main concept is to alleviate the challenges that programmers face in the IT industry.

KEYWORDS: Phishing, Algorithm, Legitimate, Prediction, peer-to-peer, code-editor

INTRODUCTION

Phishing is a type of cybercrime, and the reason phishers do it is because it is simple, inexpensive, and successful. Phishers may easily obtain anyone's email address; it is quite easy to locate an email address these days, and you can send an email to anyone for free anywhere in the world. These phishers spend relatively little money and effort in order to obtain important information fast and efficiently. Phishing scams result in virus infections, data loss, identity theft, and money theft, among other things. The important information of a user, such as the account password, OTP, credit/ debit card numbers CVV, and sensitive data connected to a user, is the data that these cyber criminals are interested in.

When working on development projects, numerous programmers collaborate in groups. Within the same project file, every programmer with access to the project can create, edit, and add code. To minimise code duplication, programmers must synchronise, and to overcome this synchronisation problem, integrated real-time cooperation in a single environment is necessary. The Integrated Development Environment (IDE) is designed to provide a collaborative

environment for programming teams with real-time text editing capabilities. Multiple users can collaborate while editing a document thanks to the ability to update text in real time.

Google Docs, for example, is one of several free programmes that enable real-time text editing. This functionality not only facilitates user cooperation, but it may also be highly useful in programming. There are also a multitude of web-based collaboration tools available. Ether Pad, for example, offers real-time text editing.

Ace, Code Mirror, and Monaco – Editor are all web-based text editors that may be integrated into an IDE or application. Project or software development necessitates coordination and cooperation among programmers, therefore collaboration solutions are extremely beneficial in increasing project efficiency. Collaboration in programming may increase the efficiency and quality of a project or piece of software. Real-time collaborative programming allows programmers to work on the same programming file. Without a direct instruction from the programmer, a real-time system will automatically merge the code written by the programmer (such as update, commit). Multiple programmers can access and modify the same source code directory, even if they are working on it at the same time. During collaborative programming, a programmer can cooperate with other programmers by entering and exiting a real-time session. Using the join protocol with two-way communication, there are procedures to join and exit the session.

A collaborative real-time editor is a type of collaborative software or web application that allows multiple users on different computers or devices to edit the same digital document, computer file, or cloud-stored data – such as an online spreadsheet, word processing document, database, or presentation – in real time. As a result, a Real-Time Collaboration Code-Editor based on the CRDT Algorithm may assist users in efficiently collaborating while working on a project even while they are offline.

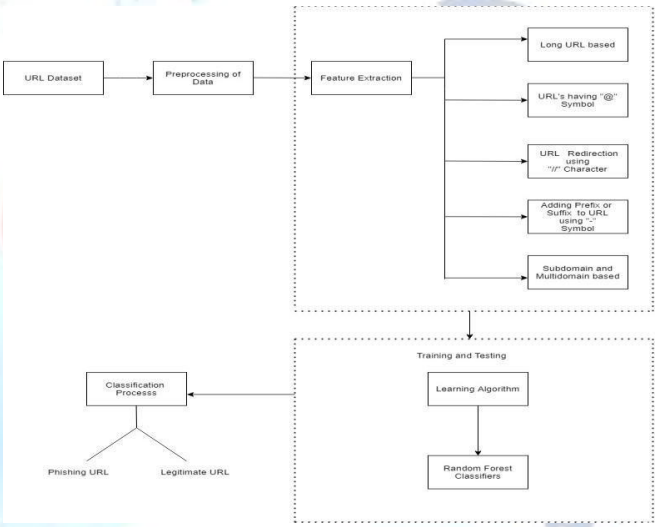
METHODOLOGY

DATA SETS

The URL data was taken from the Phish tank website, which is an anti-phishing service.

It has a total of 2905 urls in an unstructured format. Our major goal is to determine whether the URL is phishing or not. There are only a few webpage connections in this collection (Some of them are legitimate websites and some are fake websites). Pre-processing the data before creating a model, as well as extracting characteristics from the data depending on certain criteria. We need to segment the data based on the URL's characteristics.

MODEL USED:



We gathered URL data in an unstructured format from the Phishtank website. During preprocessing, five characteristics are created from unstructured data. These characteristics include URL length, prefix/suffix, number of dots, number of slashes, and subdomain length. After that, a structured dataset is produced, with binary values (0,1) for each feature, which is then sent to the various classifiers. After that, we train the Random Forest classifier and compare their accuracy. The classifier then recognises the supplied URL based on the training data, indicating whether the site is phishing or not.

1. Every character in CRDT is given a unique ID. After then, these characters are kept in a doubly linked list.
2. Lamport Timestamps are the unique identifiers. They are made up of a one-of-a-kind user identity and a logical clock that grows with each character inserted.

3. When a user types "ABC" from left to right, the following actions are performed. Insert(0, "A"), insert(1, "B"), insert(2, "C"), insert(3, "D"), insert(4, "E"), insert(5 (2, "C")

4. When a user pastes a large amount of material into a document, it is represented by a single Item.

5. Real-Time functionality may be implemented using editors such as Monaco Code-Editor and Code-Mirror.

6. The CRDT Algorithm may be integrated into the open source code editor after learning the ideas of the CRDT Algorithm and how the open source editor works.

7. Offline Editing is a helpful feature that can be implemented with the aid of the CRDT Algorithm, allowing the user to continue working on the code editor even if the internet connection is lost.

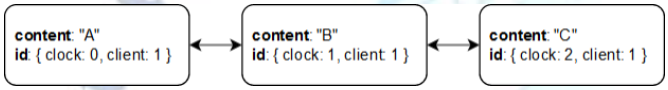


Fig 1: CRDT Node Representation

Source: P. Molli, G. Oster, and M. Rusinowitch.Proving Correctness of Transformation Functions inreal-Time Groupware

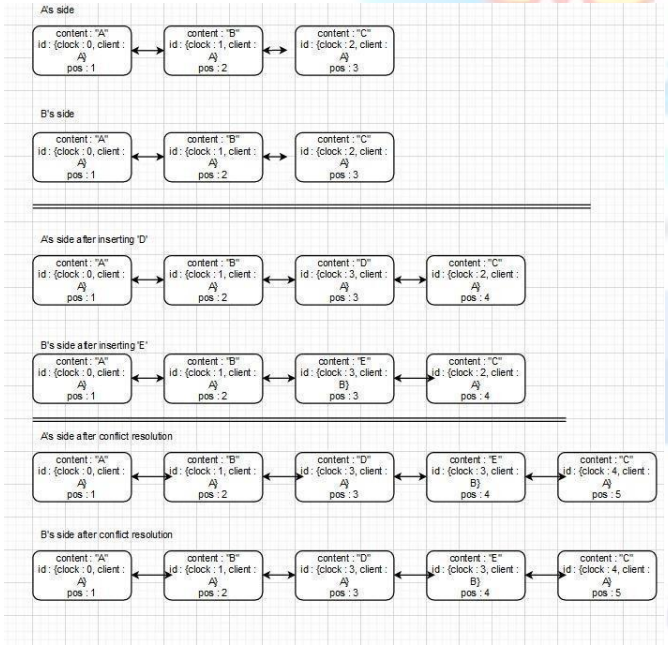


Fig 2: Multiple Users at same placeinsertion

Source: Pujara, P., & Chaudhari, M. B. (2018). Phishing Website Detection using Machine Learning: A Review

STRUCTURE OF PAPER

As discussed in the earlier sections, we have used one classifier to predict and detect if the website is phishing or legitimate.

Classifie rs	Precisio n	Recall	F1	AUC	Accuracy (%)
Random Forest Classifie r	0.9	0.8	0.85	0.87	85.6

Using random classifiers, we were able to obtain the necessary results in determining whether or not the site is phishing. The precise findings may be seen in the graph below. The AUC, precision, recall, and F1 score achieved by employing a classifier are given in the graph. The Conflict-free Replicated Datatype Algorithm was used to successfully incorporate Real-Time functionality into the Monaco Code - Editor, as well as the offline editing capability. As a Real Time Code Editor, this code editor can be quite beneficial in today's epidemic scenario.

Collaboration saves users time since several users may work on a same project at the same time, exchange ideas, and share thought processes. This will reduce the amount of time the user spends getting to the location for an offline meeting and making it a far more efficient way of working. Because everything is done from home or at a distant location, it pushes people to be more productive. Implementing offline capability in the Code-Editor is a very useful and helpful feature since it allows the user to continue working on the project even if their internet connection is unreliable. Once they rejoin to the internet, they may resume working on the project.

In the histogram graphical form, the graph shows the accuracy attained by applying different classifiers.

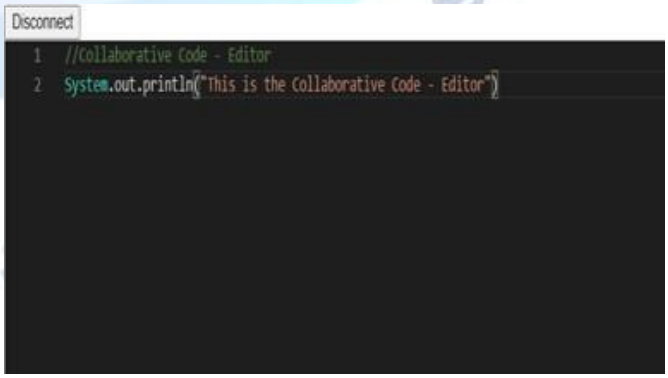


Fig 3: Result of Code

SOURCE: C. A. Ellis and S. J. Gibbs. Concurrency Control ingroupware Systems. SIGMOD Record:

Proceedings ofthe ACM SIGMOD Conference on the Management ofdata - SIGMOD '89

CONCLUSION

It has been discovered that phishing assaults are quite significant, and we must develop a system to identify them. Because such phishing websites might expose the user's sensitive and personal information, it's even more vital to address this problem. Any machine learning method with a classifier can be used to tackle this problem. We currently have classifiers that provide a decent prediction rate for phishing websites, but our poll revealed that it would be better to utilise a hybrid method for prediction in order to increase the accuracy of phishing website prediction.

REFERENCES

[1] Conflict-free Replicated Datatype Algorithm (https://en.wikipedia.org/wiki/Conflict-free replicated datatype algorithm)

[2] C.-L. Ignat, G. Oster, H.-G. Roh, and P. Urso. M. Ahmed-Nacer, C.-L. Ignat, G. Oster, H.-G. Roh, and P. Urso. Crdts for Real-timedocument Editing are being evaluated. Pages 103–112 in Proceedings of the 11th ACM Symposium on Document Engineering. ACM 2011 in New York

[3] SNS2009, pages 46–52, Nürnberg, Germany, March 2009. 3. Eurosys Workshop on Social Network Systems - SNS2009, pages 46–52, Nürnberg, Germany, March 2009.

[4] Collins-Sussman, Fitzpatrick, and Pilato, B. Collins-Sussman, B. W. Fitzpatrick, and Pilato, C. M. Subversion is a version control system. 2004. O'Reilly & Associates, Inc.

[5] S. J. Gibbs and C. A. Ellis. Controlling Concurrency in Groupware Systems SIGMOD The ACM SIGMOD Proceedings is a record of the proceedings of the ACM SIGMOD.

[6] Y. Ding, N. Luktarhan, K. Li, and W. Slamu (2019). For identifying phishing webpages, a keyword-based combination technique is used. 256-275 in Computers & Security.

[7] Marchal, S., Saari, K., Singh, N., and Asokan, N. Marchal, S., Saari, K., Singh, N., and Asokan, N. (2016, June). Know your phish: Cutting-edge methods for detecting phishing sites and their intended victims. IEEE's 36th International Conference on Distributed Computing Systems (ICDCS) took place in 2016. (pp. 323-333). IEEE.

[8] A. Hodi, J. Kevri, and A. Karadag (2016). Machine learning approaches for phishing website categorization are compared. Icesos'16: International Conference on Economic and Social Studies.

[9] Shekokar, N. M., Shah, C., Mahajan, M., and Rachh, S. Shekokar, N. M., Shah, C., Mahajan, M., and Rachh, S. (2015). An excellent method for detecting and preventing phishing scams. 82-91 in Procedia Computer Science.

APPENDIX