JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.5 No.2 May 2023: 93-96

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v5i2.9998

# Phishing Detection Applications for Website and Domain at Browser using Virustototal API

**Nadia[1*], Wellson Leewando[2], Javier Paulus[3], Valentino Nooril[4]**

[1-4] Cyber Security Program, Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
nadia@binus.edu; wellson.leewando@binus.ac.id; javier.paulus@binus.ac.id;
valentino.nooril@binus.ac.id

*Correspondence: nadia@binus.edu

***Abstract –*** *The purpose of this research is to create a browser extension-based application that can detect malicious sites to minimize phishing attacks. The research method used is to conduct a literature study and collect data from the questionnaire results. Research testing methods are blackbox testing, performance testing using 100 URL with precision and recall method, and comparison between two other simillar applications. The results of this study indicate that this application has good functionality and can reduce phishing attacks on users. The conclusion that can be drawn from this research is that the malicious site detection feature in browser extensions can enhance user protection from phishing attacks.*

***Keywords:*** *Browser Extension; Phishing Detection; Phising*

## I. INTRODUCTION

Currently, the use of information technology in Indonesia has shown a significant increase from year to year. According to an analysis from Purwiantono (2017), the existence of this convenience can be a security hole for internet users who are still unfamiliar with the security of transactions in cyberspace which can be used by internet criminals to obtain confidential information such as personal data, e-mail, passwords, even financial information such as credit card data and online banking unnoticed by internet users. One method that can be used by criminals is phishing. Clearly, a phishing site is a site designed by internet criminals in such a way as to resemble an authentic site (appearance, content, domain URL or other) to trick the victim (internet user) by making it appear as if the victim is accessing a site page from a legitimate source (Purwiantono, 2017).

From the results of research by Dongsong Zhang and friends, buying and selling sites and online banking are the sites most frequently targeted by internet criminals for phishing, because the profit potential that internet criminals can reap is quite large when compared to other sites (Zhang et al., 2014). According to Anti-Phishing Working Group or APWG (2022), public awareness of phishing sites is increasing from year to year, but the number of phishing sites and the losses they cause are also growing faster. This can cause fear and decrease the trust of internet users, even though in Indonesia digital technology continues to develop and can be accessed by everyone. It can be concluded that what is needed now is a system capable of accurately detecting phishing sites to prevent and avoid losses caused by phishing sites to internet users, especially in Indonesia.

To overcome this problem, here we will introduce our application “PhishOff” which is in the form of a browser extension for chromium-based browsers. The purpose of this extension is to help ordinary people avoid phishing site attacks. By installing this browser extension,  user can be avoided from dangerous sites. The trick is to block access to dangerous site pages and give users the choice to continue or not.

## II. METHODS

The following is an overview of the thought process that the author went through when developing the “PhishOff” application, from start to finish (Show in Figure 1).

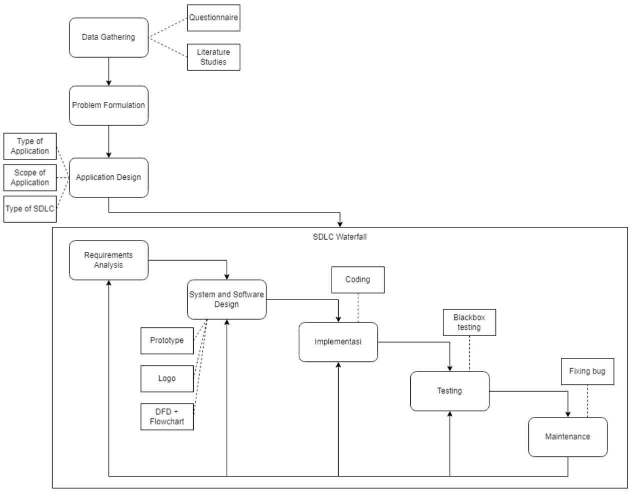

**Figure 1.** Flow Diagram of Application Development

Data gathering is the first step to be carried out in this study. The purpose of this stage is to prepare the needs of the author in conducting research and it is stated at this stage only. The process in this stage is conducting literature studies from various sources such as articles, journals, and so on as well as conducting community surveys to collect data that can support facts.

The problem formulation is the second step carried out in this study, namely analyzing the needs that were previously prepared. The current problem is the lack of public literacy regarding phishing and the high incidence of phishing attacks in society, especially between family or friends. The habit of using the internet in everyday life, which is difficult for people to let go of, causes a high risk of cyber-attacks, especially phishing.

Application design is the third step carried out in this study. At this step, the author determines the type of application to be made, the things that can be done by the application, and the type of SDLC that will be used to develop the application. The type of application to be made is in the form of a browser extension-based application and the application can check the URL being accessed based on the VirusTotal rules and API and then can grant or block access to the URL being accessed. The type of SDLC that will be used to develop applications is a waterfall which includes several stages such as Requirements Analysis, System & Software Design, Implementation, Testing, and Maintenance (Sommerville, 2015). The following is description of waterfall stages used at developing application.

### 2.1. Requirements Analysis

In this stage, the authors determine the features needed in the application based on the data that has been collected at the data collection stage.

### 2.2. System & Software Design

In this stage, the author makes the application flow by making UML. In addition, the author also makes application designs and application logos.

### 2.3. Implementation

In this stage, the author creates applications using HTML, CSS, JavaScript, Django (Python Framework), Firebase, and VirusTotal API.

### 2.4. Testing

In this stage, the author performs black box testing on the application by trying to check the 100 URLs used.

### 2.5. Maintenance

In this stage, the author will fix the bugs found in the testing stage.

## III. RESULTS AND DISCUSSION

The results of testing the solution made in the form of a browser extension application are carried out using three testing methods, namely black-box testing, testing of 100 URLs, and comparisons between two similar applications. In black-box testing, the application made is tested using 7 test classes which aim to find out whether the function of the application is in accordance with what is expected. The results of black-box testing are show in Table 1.

**Table 1.** Black-box Testing

| No | Scenario | Expected Result | Output |
|---|---|---|---|
| 1 | Users can use the toggle on/off feature in the application | The app can check URLs when the user is toggled on and vice versa | Success |
| 2 | The user accesses a URL | The app can check the URL in the address bar that is currently active | Success |
| 3 | The user presses the allowed URL button on the warning page | The app redirects to a URL that is detected as dangerous | Success |
| 4 | The user presses the allowed domain button on the warning page | The app redirects to a URL that is detected as dangerous | Success |
| 5 | The user accesses a link that has been granted allowed URL access | The app does not block URLs | Success |
| 6 | Users access links that have been given allowed domain access | The app does not block URLs with domains that have been granted access | Success |
| 7 | The user closes the warning page tab without pressing allowed URL or allowed domain and opens the same URL | The app still blocks URL access and redirects to a warning page | Success |

Next is testing of 100 URLs which aims to measure the accuracy of the detection system for the URL's hazard level. The measurement results use precision and recall and f-measure which functions to display the accuracy value by comparing the predicted results with the actual results. The following is the result of testing 100 URLs:

- Precision: 94.3%
- Recall: 100%
- F-measure: 97.1%

From the results of testing 100 URLs, the results of which were analyzed using precision and recall and f-measure, the accuracy value achieved was 97.1%, which means that the application created can detect and block URLs that are considered dangerous.

Then the next test is a comparison between two similar applications which can be seen in the following Table 2.

**Table 2.** Comparison of Similar Applications

| No | Feature | PhishOff | Ublock Origin | McAfee WebAdvisor |
|---|---|---|---|---|
| 1 | Instalation | Import Files into Chrome Extension | Install from Chrome Store | Install the Windows app and the browser app |
| 2 | Source Check | Local Cache, Database, & VirusTotal analysis dan Rules | Various URL lists from security researchers | McAfee Labs |
| 3 | Update References | Real-time for latest data and every 7 days since analyzed for existing data | Every 4 hours for the latest list | Proprietary |
| 4 | Data Volume | Scalable due to Firebase and VirusTotal analysis | Limited depending on the URLs in the List | Proprietary |

The user evaluation show in Figure 2 was carried out on 35 respondents which aimed to get a review of the research carried out which had the results: 86% Positive and 14% Nagative.

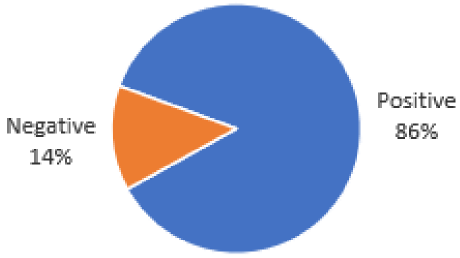

**Figure 2.** Graph of User Evaluation Results

## IV. CONCLUSION

Based on the results of research on PhishOff, a browser extension-based phishing site and malicious domain detection application, which uses the VirusTotal API and artificial rules. Following are the conclusions that can be drawn:

- From the results of black-box testing, our application has completed all scenarios with successful results.
- From the results of performance testing using the local cache, database, and VirusTotal API analysis and rules, it can be seen from the precision value of 94.3%, the recall value of 100%, and the f-measure value of 97.1%. With the results of these calculations, it can be seen that the application is capable of detecting phishing sites and malicious domains with an accuracy rate of up to 97.1%.
- From the analysis of similar applications, the application that we made has advantages in the Installation and Checking Source and Data Volume features, while the weakness of our application is in the Reference Update.
- From the results of user evaluations of the PhishOff browser extension application, our application received a positive response from all respondents. Regarding operations, respondents were satisfied with our application. Regarding user experience, respondents got a good experience with our application. Regarding the perceived impact, respondents felt the security benefits of daily internet activities. Regarding interest, respondents conveyed a good impression.

## REFERENCES

Academy, B. (2021, October 29). Apa itu Javascript? Manfaat, Fungsi dan Contohnya. https://www.binaracademy.com/blog/apa-itu-javascript-manfaat-fungsi-dan-contohnya

Adani, M. R. (2021, June 22). Data Flow Diagram (DFD): Pengertian, Jenis, Fungsi & Contoh. https://www.sekawanmedia.co.id/blog/dfd-adalah/

Alnavar, K., Kumar, R. U., & Babu, C. N. (2021). Document Parsing Tool for Language Translation and Web Crawling using Django REST Framework. Journal of Physics: Conference Series, 1962(1), 012018. https://doi.org/10.1088/1742-6596/1962/1/012018

APWG. (2022) Phishing Activity Trends Reports Q2 2022. https://apwg.org/trendsreports/

Azizah, K. (2021, February 18). Pengertian HTML Lengkap dengan Fungsi dan Sejarah Kemunculannya. merdeka.com. https://www.merdeka.com/trending/pengertian-html-lengkap-dengan-fungsi-dan-sejarah-kemunculannya-kln.html

C., Ariata. (2022, December 14). Apa Itu CSS? Pengertian, Fungsi, dan Cara Kerjanya. Hostinger Tutorial. https://www.hostinger.co.id/tutorial/apa-itu-css

Choudhury, N. (2014). World Wide Web and Its Journey from Web 1.0 to Web 4.0. International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (6), 8096-8100.

D. Zhang, Z. Yan, H. Jiang, dan T. Kim, “A domain-feature enhanced classification model for the detection of Chinese phishing e-Business websites,” Inf. Manage., vol. 51, no. 7, hal. 845–853, Nov 2014.

Finaka, A. W., Fauzi, A., & Munir, T. S. (2018). https://indonesiabaik.id/infografis/penggunaan-komputer-saat-terhubung-internet-di-indonesia-15

Khawas, C. & Shah, P. (2018). Application of Firebase in Android App Development-A Study. https://

*Phishing Detection Applications for Website and Domain… (Nadia, et al)* 95 www.researchgate.net/profile/Chunnu-Khawas/publication/325791990_Application_of_Firebase_in_Android_App_Development-A_Study/links/5bab55ed45851574f7e6801e/Application-of-Firebase-in-Android-App-Development-A-Study.pdf

Kurniawati, P. (2018, November 2). Pengujian Sistem - SkyshiDigital. Medium. https://medium.com/skyshidigital/pengujian-sistem-52940ee98c77

Meador, D. (2020, June 19). What is Data Dictionary. https://www.tutorialspoint.com/What-is-Data-Dictionary

Narkhede, S. (2021, June 15). Understanding Confusion Matrix - Towards Data Science. Medium. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Pradiptamukherjee. (2022, November 29). What is a Flowchart and its Types? https://www.geeksforgeeks.org/what-is-a-flowchart-and-its-types/

Purwiantono, F. E. (2017). MODEL KLASIFIKASI UNTUK DETEKSI SITUS PHISHING DI INDONESIA (Master's thesis, Surabaya, 2017) (pp. 1-135). Surabaya: Institut Teknologi Sepuluh Nopember. https://repository.its.ac.id/43198/2/5215201006%20-%20Master_Thesis.pdf

Pyram, J. (2021, December 16). 9 Parts of a URL You Should Know - Joseph Pyram. Medium. https://medium.com/@joseph.pyram/9-parts-of-a-url-that-you-should-know-89fea8e11713

Simplylearn. (2023, January 11). What is a Database? Everything You Need to Know. https://www.simplilearn.com/tutorials/dbms-tutorial/what-is-a-database

Sohan, S. M., Anslow, C., & Maurer, F. (2015). A Case Study of Web API Evolution. World Congress on Services. https://doi.org/10.1109/services.2015.43

Sommerville, I. (2015). Software Engineering (10th ed.). Pearson.

Srinath, K. R. (2017). Python – The Fastest Growing Programming Language. International Research Journal of Engineering and Technology (IRJET), Vol. 4 (12), 354-357.

Statcounter. (2023). Statcounter Global Stats. https://gs.statcounter.com/browser-market-share/all/indonesia#monthly-202201-202212

Szurdi, J. (2021, November 11). A Peek into Top-Level Domains and Cybercrime. https://unit42.paloaltonetworks.com/top-level-domains-cybercrime/

Visual Paradigm. (n.d.). What is Entity Relationship Diagram (ERD)? https://www.visual-paradigm.com/guide/data-modeling/what-is-entity-relationship-diagram/

Wibisono, G., & Susanto, W. E. (2015). Perancangan Website Sebagai Media Informasi Dan Promosi Batik Khas Kabupaten Kulonprogo. EVOLUSI: Jurnal Sains Dan Manajemen, 3(2). https://doi.org/10.31294/evolusi.v3i2.630