Typosquatting Potential on the Official Website (An Empirical Study of Popular Webites in Indonesia)

Dimas Sasongko[1,*] Agung Suprapto[2] Purwono Hendradi[3]

[1,3]*Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Magelang, 56172, Indonesia*
[2]*Department Islamic Broadcast and Communication Studies, Faculty of Da'wa, Institut Agama Islam Negeri Salatiga, 50716, Indonesia*
[*]*Corresponding author. Email: dimassasongko@ummgl.ac.id*

ABSTRACT
At the time of typing the website address (URL), the users often make mistakes (typos) that are unable to visit the desired website. The situation becomes worse when the wrong URL pointing to other websites that do not wish to be visited by the user and may be made available for the aim of cyber-crime. This condition is called Typosquatting. Typosquatting is often used to carry out dangerous cyber-attacks. Research conducted will focus on the potential Typosquatting popular website domain in Indonesia based on Similar web and Alexa rankings. The purpose of this research is to determine the potential for Typosquatting attacks on the popular website domain. Potential observation for Typosquatting was carried out using the Domain Security Test owned by Immuniweb. The result obtained from observation is the number of active domains contained in Typosquatting potential domain has a percentage of 90 % and 26 % potential domain of Typosquatting visited by the users. The most identified Typosquatting model is substitution and addition, meanwhile the most widely used of the website purpose is Hit Stealing and Parking Domain / For Sale. The letters on the left and right of the type target letter have the potential to be more often cause typo if it is compared to the letters which are above and below the type target letter.

Keywords: *Typosquatting, Website, Domain Name System, Cyberattacks, Typo.*

1. INTRODUCTION

At these time billions of devices to access Internet and allow the users to connect and exchange information. One important part of the internet infrastructure is Domain Name System (DNS). DNS allows the users to use easier text-based domain names to read when compared using the numeric address form provides IP address to an Internet access service as the service website. At the end of 2015 data showed nearly three hundred million website domain names were registered [1].

Use of the Internet has been developed rapidly around the world resulted in an increasing number of websites and the number of visits on the website. Some ways done by the users to visit a website is to type Uniform Resource Locators (URL) in the browser. At the moment clicking ethics URL, users often make mistakes (typos) that are unable to visit the desired website. The situation is made worse when the wrong URL leads to another website that the user does not want to visit [2], [3]. This condition is called Typosquatting.

Typosquatting is often used to carry out dangerous cyber-attacks. Hackers create fake websites that imitate the appearance and purpose nuance of the original website which is intended by the user so that he does not realize is visiting a fake website. Thus, the risk of a typo is causing a phishing attack. Sometimes the fake website created with the aim to steal users' personal information, such as credit cards, PIN, passwords, and other personal information of the users [2].

Research about Typosquatting website domain had been done by researchers. A research conducted by Janos Szurdi discusses about the most effective Typosquatting Technique. In the conducted research is obtained information that the most effective Typosquatting technique is implemented in a website URL is reorder the URL to the different sequence and do substitution order from the website URL. In the conducted research is also obtained information that, to avoid typo and dangers of

Typosquatting is by learning the knowledge to the users [4].

Another study that will be undertaken by Dastagir Husain Muhammed analyzes Typosquatting scenario against existing popular domain in Bangladesh. The domain that has potential to make a typo is then checked whether it is active or not. Furthermore, the domains are analyzed for the content displayed to see if they aim to take advantage of user mistakes in writing the domain with spam, scam, or other media that allow to commit cyber-crimes [5].

In the previous research, there were no observation about potential visitors accessing the domain with potential for typos and mapping of potential typos based on the target letters on the keyboard. The research conducted will focus on the potential of popular Typosquatting website domain in Indonesia based on Similar web and Alexa rankings.

The purpose of this research is to know the potential Typosquatting attacks on the popular website domain. There are some research questions, they are: (i) what is the potential and status of the website domain for typo? (ii) what is the typo model and purpose of the typo domain? (iii) what is the potential for user visits to mistyped domains? (iv) and what is the potential for typos based on the target letter on the keyboard? The research questions will be discussed through research that will be carried out based on the research method in Chapter 2.

2. METHOD

The research method begins with the research stages as shown in Figure 1. the study begins with literature review about Typosquatting in particular on a website URL. Furthermore, the website domain is chosen to be used as the Typosquatting test material, in this study we will use the popular website domains in Indonesia which are ranked in the top 10 versions of Similarweb and Alexa. After selecting the website domain, the next step is to map the potential Typosquatting on the website domain. The next stage is to check the status and objectives of the mapping results of potential Typosquatting website domains. The last stage is to analyze the mapping findings and test the typo website domain. The results of the analysis are expected to provide recommendations to the official website domain owners to be more aware and able to take preventive measures from Typosquatting attacks.

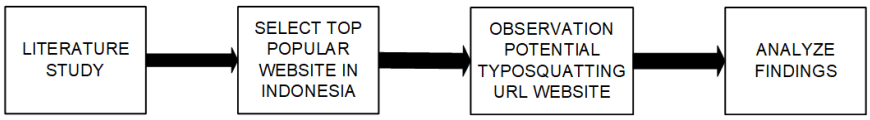

Figure 1 Research workflow

2.1. Literature Study

The initial stage of the research conducted was literature study. The study in literature consists of similar research and theoretical studies used in conducting research. Similar research related to Typosquatting that has been carried out by previous researchers was conducted to obtain the authenticity of the research and determine the contribution of the research, a discussion of the research carried out by previous studies can be found in Chapter 1.

Typosquatting domain is the act of registering a domain name that is very similar to an existing and legitimate domain, in an attempt to capture some of the traffic destined for domain originals. The annoyance of typing domains exploits the tendency of users to make typos when typing in domain names and is often used for financial gain. For example, someone who registers for go oge.com will immediately receive traffic in a large volume intended for google.com. This traffic can in turn be monetized, by displaying advertisements as seen in Figure 2. Typosquatting domains have proven to be profitable, while requiring no technical expertise [6].

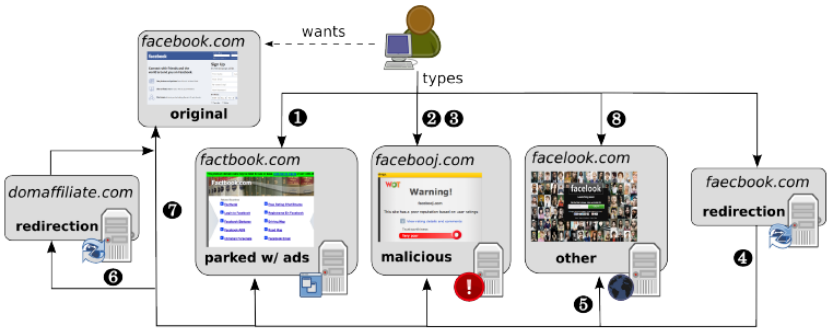

Figure 2 The typosquatting ecosystem with various monetization techniques [7]

In 2006, Wang categorized Typosquatting mistakes into five different categories [8], additionally from cisecurity.org divides Typosquatting mistakes into six different categories [9]. For ease of understanding, assume the example.com domain is a legitimate domain:

1. Typo by eliminating dot on domain writing after write "www", for example as follows wwwexample.com.
2. Typo in one letter character missing (omission), for example exmple.com which should be example.com.
3. Typo of interchangeable two-character (permutation), samples hnya exmaple.c om which should be example.com.
4. Typo of interchangeable two-character with other letter (substitution), usually because the letters are located close together on the keyboard layout, for example ez ample.com which should be example.com.
5. Typo of writing duplicated letter character (duplication), for example exammple.com which should be example.com.

In the white paper document issued by cisecurity.org in 2018, there is an addition of a typo model made by the users, namely:

1. Typo by adding the letter character (addition), for example examples.com which should be example.com.
2. Typo by adding a hypenation character, for example ex-ample. com which should be example.com.
3. Typo by using characters that are considered same (homoglyph), for example the capital letter "i" will be same/similar to the small letter "l". For example, exampIe.com by writing the capital letter "i" which should be example.com by writing the small letter "l".

2.2. Select Top Indonesia Popular Website

Selection of website domain as research subject is a website domain in Indonesia which was included in ten-rank popular websites by Similarweb and Alexa version in January 2020. List of popular website domains in Indonesia as shown in Table 1. There are eight domains of Indonesian company (Okezone, Tribunnews, Kompas, Grid, Detik, Tokopedia, Liputan6, and Sindonews) and seven domains from international companies (Google, Youtube, Facebook, Instagram, WhatsApp, Ucweb, and Twitter).

Table 1. Top Indonesia Popular Website

No	Popular Website Based on Similarweb Rangking	Popular Website Based on Alexa Rangking
1	Google.com	Okezone.com
2	Youtube.com	Google.com
3	Facebook.com	Tribunnews.com
4	Tribunnews.com	Youtube.com
5	Detik.com	Detik.com
6	Instagram.com	Liputan6.com
7	Whatsapp.com	Kompas.com
8	Ucweb.com	Grid.id
9	Kompas.com	Tokopedia.com
10	Twitter.com	Sindonews.com

2.3. Observation Potential Typosquatting Domain Website

Observation of potential Typosquatting was carried out on website domain as shown in Table 1 was carried out using the Dark Web Exposure Test owned by https://www.immuniweb.com as shown in Figure 3. Users only need to enter the website domain address then the system will automatically search for potential typosquatting domains.

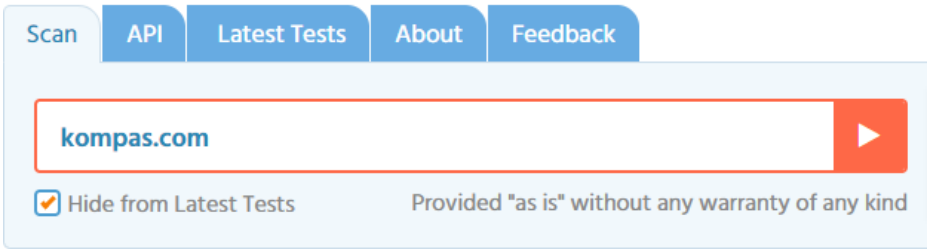

Figure 3 Immuniweb Dark Web Exposure Test

At this stage, the website domain is scanned to find potential typos in the URL as shown in Figure 4. After finding a list of potential typos in the URL, the next step is to check the status and purpose of each URL. The status check is conducted to find out the status of the active domain or not, and the purpose of the URL check is to find out the purpose of the URL if it is active, for example it is used for advertising, adult content, the domain is for sale, or the URL is forwarded to an official domain. The final step is to detect user visits to the potential Typosquatting domain, and to detect user visits using the features of https://www.statshow.com.

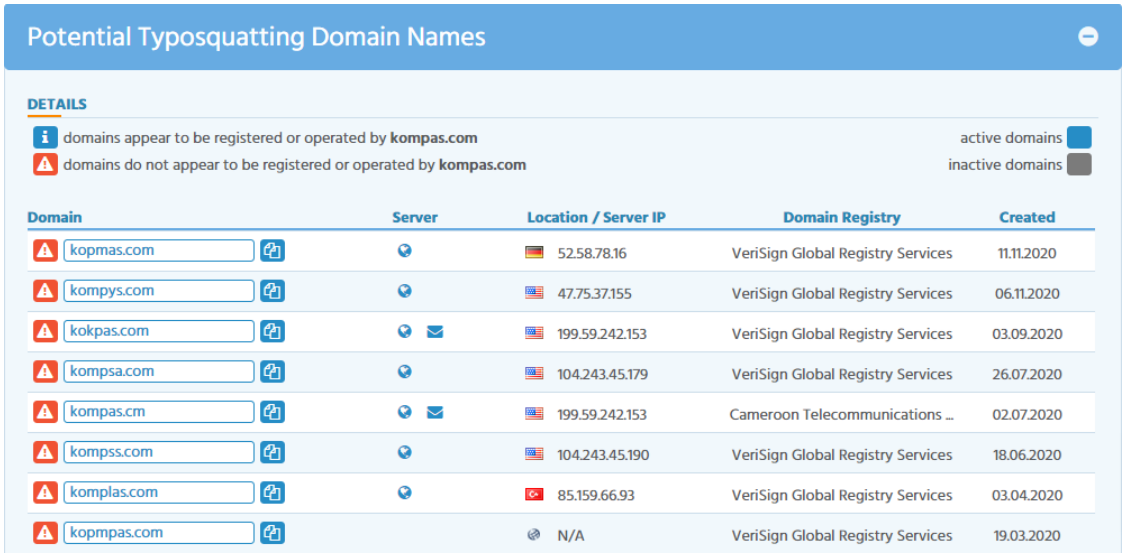

Figure 4 Results Potential Typosquatting Domain Names

2.4. Analyze Findings

The final stage of this research is to analyze the findings from the observation results of the potential Typosquatting website domain and check the status of the Typosquatting website domain against the dangers of cyber-attacks. In addition, analysis was conducted on the potential for user visits to typos and potential typos based on the target letters on the keyboard. Top Indonesian popular website in Table 2.

3. RESULT AND DISCUSSION

The result and discussion of the research are divided into three parts, namely: observation of potential and typosquatting domain status, analysis of user visits and Typosquatting domain objectives, and analysis of typosquatting mistake models and potential typos on the keyboard. At the end of the analysis will provide recommendations to the popular domains in Indonesia in overcoming the potential of typosquatting.

Table 2. Top Indonesia Popular Website

No	Official Website	Typo Potential	Active Domains	Inactive Domains	Domains appear to be registered by Official
1	Okezone.com	21	25	2	0
2	Google.com	111	86	25	32
3	Tribunnews.com	11	10	1	0
4	Youtube.com	151	121	30	27
5	Kompas.com	37	35	2	0
6	Grid.id	1	1	0	0
7	Detik.com	21	21	0	0
8	Tokopedia.com	39	38	1	0
9	Liputan6.com	7	7	0	0
10	Sindonews.com	7	7	0	0
11	Facebook.com	135	129	6	55
12	Instagram.com	144	137	7	43
13	Whatsapp.com	86	83	3	0
14	Ucweb.com	14	13	1	0
15	Twitter.com	132	121	11	13

Typosquatting observations were carried out on 15 popular website domains in Indonesia according to Similarweb and Alexa versions. The result of observation covers the number of potential Typosquatting domain, the number of active typo domains, the number of typo domains that are not active, and the number of typo domains registered by the official website manager. The results are as shown in Table 2.

Information shown in Table 2 is Typosquatting observation result of popular website in Indonesia, data shows that 90% are active domains and 26% of active domains were detected as visited by users. One of the interesting findings from the observations is an international company like Google, Youtube, Facebook, Instagram, and Twitter registering website domain which has potential typo as a precaution against user mistakes in typing URL of the website, while the Indonesian companies such as Okezone, Detik, and Kompas do not register website domains that have potential for Typosquatting.

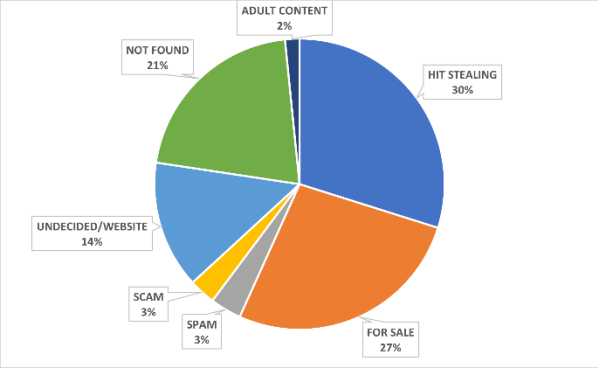

Figure 5 Aim of Potential Typosquatting Website

The information displayed in Figure 5 is the observations result of potential Typosquatting website objectives. Hit stealing and For Sale (parking domain) are the ones most often done by potential Typoquatting website owners. There are similarities in the two goals of the website, namely monetized with the main aim of making a profit.

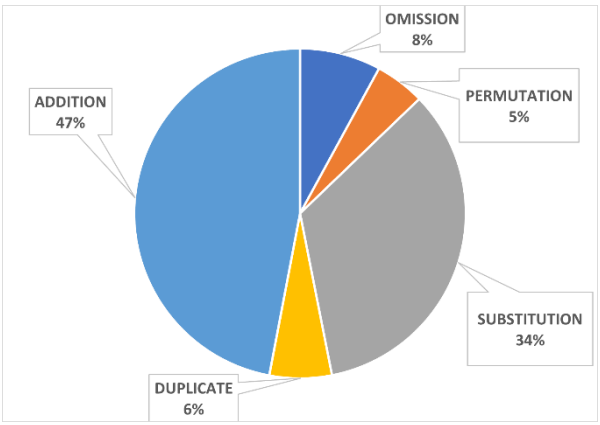

Figure 6 Findings of Typosquatting Mistake Model

The information displayed in Figure 6 is the finding of Typosquatting mistake model performed by the user. The mistake model most often made by users is the addition of letters (addition) and the letter characters that are substituted for other letters (substitution). Addition and substitution have similar mistakes when compared to other Typosquatting mistake models, so they are the most frequently used models by users. From the observation result, the layout of letters adjacent to the target letter type is the cause of two mistake models that occur the most.

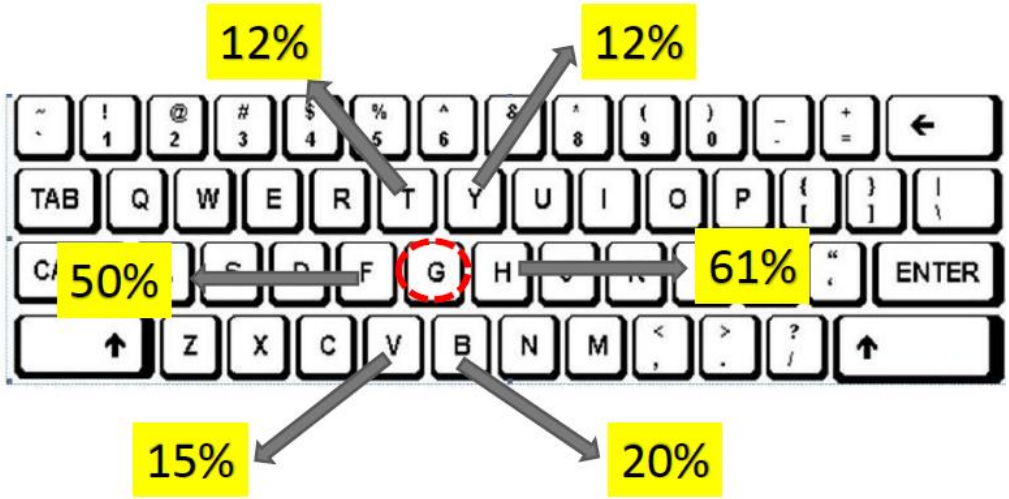

Figure 7 Example of Mapping Potential Typing Mistakes on Keyboard

The observation development of Typosquatting mistake model is the identification of potential typos based on the target letters on the keyboard. Typical mistakes that occur most often involve a one-character distance; this case is usually called fat finger distance. Typing errors occur in adjacent letters on a US English keyboard and users are more likely to mistyping letters that are closely spaced [10]. As seen in Figure 7 is an example of a mistake mapping from potential typos on the keyboard, letters that are located to the right and left of the target letter have the potential to cause typos more often than letters that are above and below the target letter type.

4. CONCLUSION

A typo is an event that can occur when a user writes a website address in the browser. One of the impacts is that users will visit unwanted websites, for example, website pages that are intentionally monetized for profit. From the observation of Domain Typosquatting, 90% are active domains and 26% of active domains are detected as being visited by users. Hit stealing and For Sale (parking domain) are the website models most often implemented by potential Typoquatting website owners. The Fat Finger Distance factor causes letters that are located next to each other on the keyboard to have the potential for typos.

ACKNOWLEDGMENTS

This paper is part of the project computer networking research group, which is conducted at the Informatics Laboratory of Universitas Muhammadiyah Magelang. The researchers are grateful to the technicians who have helped data collecting and also the reviewers. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] J. Spaulding, S. Upadhyaya, and A. Mohaisen, "You've been tricked! a user study of the effectiveness of typosquatting techniques," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 2593–2596.

[2] I. Ahmad, M. A. Parvez, and A. Iqbal, "TypoWriter: A Tool to Prevent Typosquatting," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, 2019, vol. 1, pp. 423–432.

[3] J. Spaulding, S. Upadhyaya, and A. Mohaisen, "The landscape of domain name typosquatting: Techniques and countermeasures," in *2016 11th International Conference on Availability, Reliability and Security (ARES)*, 2016, pp. 284–289.

[4] J. Spaulding, D. Nyang, and A. Mohaisen, "Understanding the effectiveness of typosquatting techniques," in *Proceedings of the fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies*, 2017, pp. 1–8.

[5] M. D. Husain and A. Iqbal, "An empirical study on typosquatting abuse in bangladesh," in *2017 International Conference on Networking, Systems and Security (NSysS)*, 2017, pp. 47–54.

[6] J. Szurdi and N. Christin, "Email typosquatting," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 419–431.

[7] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The long 'taile' of typosquatting domain names," in *23rd ${$USENIX$}$ Security Symposium (${$USENIX$}$ Security 14)*, 2014, pp. 191–206.

[8] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels, "Strider Typo-Patrol: Discovery and Analysis of Systematic Typo-Squatting.," *SRUTI*, vol. 6, no. 31–36, pp. 2–2, 2006.

[9] "MS-ISAC Security Primer - Typosquatting," *CIS*. https://www.cisecurity.org/white-papers/ms-isac-security-primer-typosquatting/ (accessed Oct. 07, 2020).

[10] J. Szurdi, "Measuring and Analyzing Typosquatting Toward Fighting Abusive Domain Registrations," PhD Thesis, Carnegie Mellon University, 2020.