POSTER: Do You Sell This? Utilizing Product Searches to Find SEO-driven Fake Shopping Sites

Mizuho Hasegawa[*], Akihide Saino[*], Akira Fujita[‡,*], Kazuki Takada[†,*], Rui Tanabe[*,§], Carlos H. Gañán[¶,*], Michel van Eeten[¶,*], Katsunari Yoshioka[*]
[*]Yokohama National University, [‡]National Institute of Information and Communications Technology, [†]Hitachi Systems, Ltd., [§]Juntendo University, [5]Delft University of Technology

***Abstract*—We propose a method to identify SEO-driven fake shopping sites by utilizing product search queries. Our approach leverages two sources of search queries: those generated by actual users and those for popular products on legitimate shopping platforms. Utilizing search queries from 2,093 users and popular products on two major shopping platforms in Japan, we detected 38,966 domains of potentially fake shopping sites over a 11-month period from February 2024 to January 2025. A manual investigation of 100 random samples revealed that 97% of the sites exhibited distinctive characteristics of fake shopping sites, including inconsistencies in displayed and implemented payment methods or mismatched shop operator information, such as postal addresses, contact information, and representative names.**

I. Introduction

Fake shopping sites exploit Search Engine Optimization (SEO) techniques to rank prominently in product search results, attracting potential victims looking to purchase items online [1]. These sites often utilize a "*springboard*"—webpages optimized specifically for search engine indexing to achieve high rankings across various product searches. When users search for products, the springboard appears prominently in search results. Upon visiting the springboard, users are redirected to fake shopping sites, where product information is replicated from legitimate shopping platforms.

Effective detection is crucial for addressing countermeasures against fake shopping sites. However, efficiently identifying these sites within the dynamic and vast expanse of the web remains a significant challenge. A recent study proposed crawling newly registered domains as a method to collect fake shopping sites [3]. In contrast, this study takes a different approach, focusing on the inherent nature of SEO-driven fake shopping sites—that is, they are "*designed to be found*" through product searches.

In this study, we propose a method to identify SEO-driven fake shopping sites by utilizing product search queries from two sources: queries generated by actual users and those for popular products on legitimate shopping platforms. Using search queries from 2,093 users and popular products on two major shopping platforms in Japan, we detected 38,966 potentially fake shopping site domains over a 11-month period. To evaluate our method, we randomly selected 100 samples and manually examined them for common characteristics of fake shopping sites, including mismatches between displayed and implemented payment methods, and suspicious company information such as inconsistencies in company names, postal addresses, contact information, and representative names.

The evaluation revealed that 97 sites displayed clear characteristics of fake shopping sites. Two sites, initially detected as shopping-related, had switched to online casino content by the time of manual review. Only one site was identified as a legitimate shopping site, representing a false positive. These results highlight the high accuracy of our proposed method in detecting SEO-driven fake shopping sites.

To assess its utility, we matched the detected sites with web access logs from 2,096 users recorded between December 1, 2024, and January 7, 2025 (37 days). The results revealed 177 access attempts to 14 different identified sites by 9 users, which could have been preemptively flagged, demonstrating the blocklist's effectiveness in user protection.

II. Methodology

Our proposed method employs two types of search queries and their associated Search Engine Result Pages (SERPs):

User Queries: These are search queries and SERPs from popular search engines, including Google, Yahoo!, and Bing. They are collected from real users who installed a browser extension developed for a web security research project [4].

Popular Product Queries: These queries are based on searches for popular products. Information on these products is sourced from two major shopping platforms in Japan: Shopping Site A and Shopping Site B. Shopping Site A was chosen due to its significant association with product misuse in fake shopping sites, while Shopping Site B was selected for its publicly accessible data, formatted for direct use in queries.

The utilized search queries were in Japanese language, enabling the detection of Japanese fake shopping sites. However, we anticipate that the proposed method can be adapted to detect fake shopping sites in other languages, provided they replicate product information from legitimate shopping sites.

By combining these two sources, the method enables a more comprehensive collection of SEO-driven fake shopping sites. Detection based on User Queries reflects the behavior of real users participating in the Web security research project, allowing for the identification of fake sites relevant to their search habits. Conversely, detection based on Popular Product Queries leverages broader insights into popular products on

Network and Distributed System Security (NDSS) Symposium 2025
23–28 February 2025, San Diego, CA, USA
ISBN 979-8-9894372-8-3
https://dx.doi.org/10.14722/ndss.2025.[23|24]xxx
www.ndss-symposium.org

major shopping platforms, which are likely to attract the attention of fake shopping site operators.

After extracting search results from SERPs, we use the web browser automation tool Puppeteer [2] to access each search result page. If access is redirected via the location header or JavaScript, the redirected pages are also visited. The HTML source of each website, along with associated resources (e.g., JavaScript and image files), is retrieved. Finally, each search result page is evaluated based on the following conditions:

- The search result page automatically redirects visitors to another page.
- The final destination page directly references resources, such as product image files, hosted on predefined legitimate shopping sites to replicate their product information.
- The domain of the final destination page differs from the legitimate shopping site whose resources are referenced.

If all these conditions are met, the final destination page is classified as a fake shopping site, and the search result page is designated as a springboard. To monitor trends effectively, the entire process—from data collection to fake shopping site identification—is executed daily.

III. Evaluation

We collected the User Queries and associated SERPs from 2,093 users who performed at least one web search between February 10, 2024, and January 7, 2025.[1] Moreover, approximately 300 Popular Product Queries were generated daily using product information extracted from two major shopping platforms in Japan. Then, Google SERPs with top 100 results were collected using these Popular Product Queries. We predefined 22 legitimate shopping sites in Japan.

A. Number of Potential Fake Shopping Sites Collected

We conducted fake shopping site detection from February, 2024, to January, 2025, collecting a total of 38,966 fake shopping sites, aggregated by their domain names.

- Using User Queries, the system crawled an average of 11,474 URLs daily, detecting approximately 29 new fake shopping sites per day.
- Using Popular Product Queries, the system crawled an average of 21,861 URLs daily, identifying around 105 new fake shopping sites per day.

A total of 1,899 sites (5% of the total) were detected from both sources. This highlights that combining the two sources improves detection coverage by identifying largely distinct groups of fake shopping sites.

B. Accuracy of Collected Fake Shopping Sites

The accuracy of the proposed method was evaluated through manual inspection of the detected sites. Of the 38,966 domains identified between February 2024 and January 2025, 1,164 remained accessible during the inspection period in January 2025. From these, 100 accessible sites were randomly selected for manual examination to identify common characteristics of fake shopping sites.

Among the selected sites, two had shifted their original shopping-related content, detected in April 2024, to online casino platforms by January 2025. As only the top pages of these two sites were retained at the time of detection, further analysis of their content was not possible. The remaining 98 sites continued to operate as shopping sites. For these, we conducted a detailed analysis, comparing payment methods described on their web pages with their actual implementations, as well as examining the company information hosting the sites, including company names, postal addresses, contact phone numbers and emails, representative person's names, and legal registration details, where available.

Common indicators of fake shopping sites included mismatches between displayed and implemented payment methods, fraudulent or copied company information, and shared details across multiple detected sites. Based on these indicators, we determined that 97 of the 98 remaining sites were clearly fake shopping sites. However, one site had consistent and verifiable information across all examined criteria and was classified as a legitimate shopping site, thus representing a false positive. This manual investigation confirms the high accuracy of our method.

C. Effectiveness as a Blocklist

We evaluated the effectiveness of the detected sites when used as a blocklist for preventing user access to fake sites. We matched the 38,966 detected domains with web access logs from 2,096 users recorded over a 37-day period (December 1, 2024, to January 7, 2025). We consider that access to the identified fake shopping sites could have been prevented if our method had detected them at least one day prior to the access, assuming the blocklist is updated and distributed daily. Our findings indicate that 177 access attempts to 14 identified sites by 9 users could have been mitigated, demonstrating the effectiveness of our approach in enhancing user protection.

IV. Conclusion

We revealed the effectiveness of utilizing product search queries to detect SEO-driven fake shopping sites. Additionally, the detected sites can be used as a blocklist to protect users from accessing fake shopping sites. One of our future works is to provide this blocklist to users who participate in the project.

Acknowledgements. A part of this research was conducted in the WarpDrive project, supported by the NICT, Japan. This work was supported by JSPS KAKENHI Grant Numbers 21H03444 and 21KK0178.

[1] Studies utilizing web access logs from the WarpDrive project [4] were reviewed by the project's Ethical Review Board to ensure compliance with ethical standards and the protection of user privacy. Participants are required to agree to the browser extension's terms, which outline the data collection process for security research, prior to installation.

References

[1] Kramar, B. Avast researchers detect a surge in fake e-shops. https://blog.avast.com/avast-researchers-detect-surge-in-fake-e-shops, 2024.
[2] Puppeteer. Puppeteer. https://pptr.dev/, Last Access: 2025/01/25.
[3] Sakai, K., Takeshige, K., Kato, K., Kurihara, N., Ono, K., and Hashimoto, M. An automatic detection system for fake japanese shopping sites using fasttext and lightgbm. In *IEEE Access* (2023), IEEE.
[4] WarpDrive. Warpdrive. https://warpdrive-project.jp, Last Access: 2025/01/25.

POSTER: Do You Sell This? Utilizing Product Searches to Find SEO-driven Fake Shopping Sites

Mizuho Hasegawa[1], Akihide Saino[1], Akira Fujita[2,1], Kazuki Takada[3,1],
Rui Tanabe[1,4], Carlos H. Gañán[5,1], Michel van Eeten[5,1], Katsunari Yoshioka[1]
[1]Yokohama National University, [2]NICT, [3]Hitachi Systems, Ltd., [4]Juntendo University, [5]TU Delft

Introduction

- Fake shopping sites exploit Search Engine Optimization (SEO) techniques.
- However, exploring fake shopping sites remains a significant challenge.
- We propose a method to **identify SEO-driven Fake Shopping Sites** using search queries generated by real users (**User Queries**) and those for popular products on legitimate shopping platforms (**Popular Product Queries**).

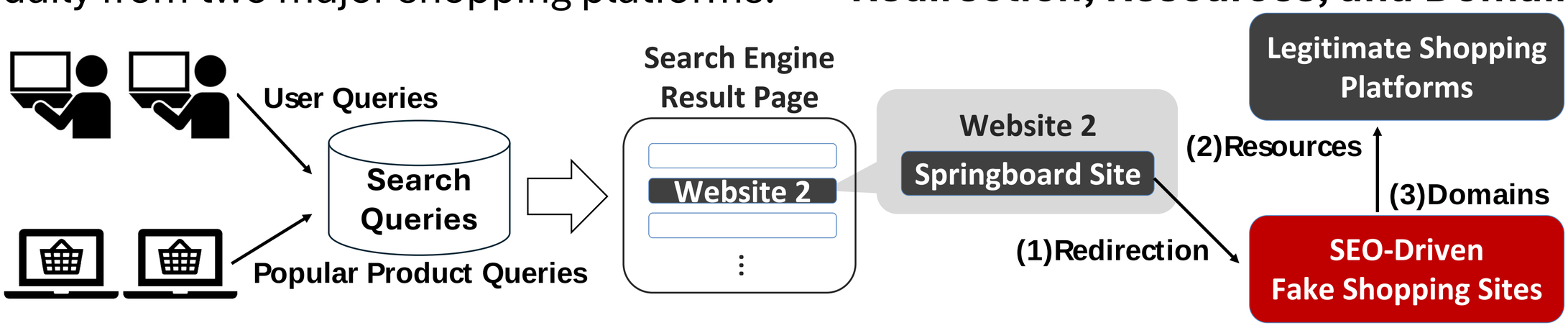
Methodology

Data Sources

- **User Queries:** Collected via browser extensions from 2,093 actual users.
- **Popular Product Queries**: Extracted daily from two major shopping platforms.

Detection Process:

- Crawling Search Engine Result Pages using a web browser automation tool.
- Detect fake shopping sites based on **Redirection, Resources, and Domains**.

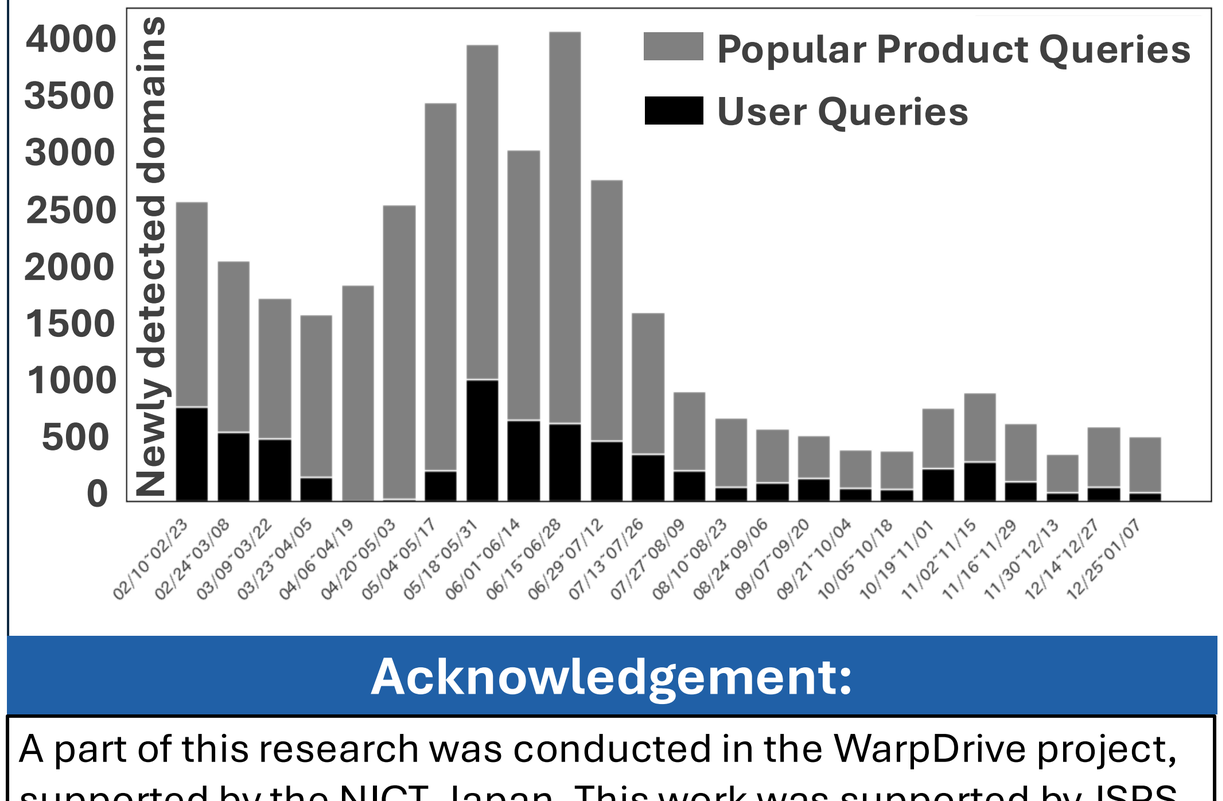


*(1) automatically redirects visitors to another, (2) directly references resources like product image files hosted on predefined legitimate shopping sites to replicate their product information, and (3) the domain differs from the legitimate shopping site whose resources are referenced.

Evaluation Result

Number of Collected Sites

- We detected **38,966** potential domains from February 2024 to January 2025.
- We identified **29** new sites with User Queries and **105** with Popular Product Queries on a daily average.
- Note that 5% were detected from both.



Accuracy of Detected Sites

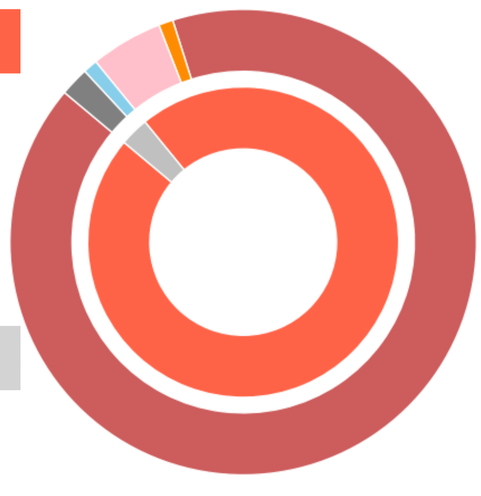
- We randomly selected 100 detected sites and confirmed that **97** displayed the characteristics of fake shopping sites below.
- Mismatched payment methods (displayed vs implemented).
- Fraudulent/copied company information (company name, addresses, etc.).

Fake shopping sites: 97/100

- Mismatch Payment Methods: 91
- Address with Other Companies: 5
- Address Not Associated with Any Company: 1

Others: 3/100

- Content Changed: 2
- False Positives: 1



Effectiveness as a Blocklist

- Among 2,096 users, 177 attempted to access **14** fake shopping sites in 37 days.
- We plan to provide our blocklist.

Acknowledgement:

A part of this research was conducted in the WarpDrive project, supported by the NICT, Japan. This work was supported by JSPS KAKENHI Grant Numbers 21H03444 and 21KK0178.