Stealthy and Seductive: A Survey on Online Illicit Promotion

Lu Zhang*, Yeonjoon Lee**

*Major in Bio Artificial Intelligence, Hanyang University (Graduate student)

** Dept. of Computer Science and Engineering, Hanyang University (Professor)

Abstract

With the growth of netizens and the popularity of the Internet and smart devices, illicit promotional practitioners have been trending to promote online instead of offline. However, social media platform companies, operators, and government have their detection systems and censorship to solve such promotion. While they use newer techniques to deal with them, the illicit promotional practitioners also use newer techniques to evade the detection system and censorship; the process is like an adversarial process. In this paper, we look at this phenomenon and analyze how practitioners use techniques to do illicit promotion online and how to mitigate this phenomenon.

I. Introduction

Since ancient times, promotion has become a meaningful way to make potential users know about the goods. The time going, this behavior became more online instead of offline[1].

Online Illicit Promotion (OIP)'s history can be traced back to the beginning of the Internet era. In the early days, miscreants promoted by spamming[2] or posting on online forums[3]. Nowadays, OIP has become more stealthy with new carriers and techniques.

With the widespread use of artificial intelligence (AI) in recent years, platform companies, operators, etc., tend to use AI to detect to save manpower. Escape AI scrutiny became a target for OIP. OIP's promotion form is usually also combined with social engineering and human-computer interaction design to achieve the best effect. OIP mainly influences regular netizens through Explicit Advertisements (for those who are interested in illicit content) and Seduction (for those who are not interested in illicit content initially). This paper focuses on these two aspects and concludes with current typical OIP methods and mitigations towards them.

1.1 Explicit Advertisement

Drugs[1], porn[5], and other illicit content have a large underground market, but they cannot be promoted explicitly through popular legal online channels (e.g., social media) due to strong AI detection systems. Miscreants are increasingly employing new techniques to escape AI detection for displaying explicit content through popular legal online channels.

Search Engine Optimization (SEO) raises a website's visibility among the relevant results returned by major search engines, making it more visible to its intended audience.

1.2 Seduction

Search Engine Optimization (SEO) raises a website's visibility among the relevant results returned by major search engines, making it more visible to its intended audience. Information on authoritative or benign websites and apps is more convincing. Such information can seduce people to make wrong decisions more easily. Based on this, an illegal shortcut is to hack into popular websites, plant promotional content and referring links for the illicit pages. Due to its cost-effectiveness, promotional website defacements have been widely used for black hat SEO[4]. The other shortcut is to build authoritative- or benign-looking websites and apps.

II. OIP methods

2.1 Explicit Advertisement

Image is the most direct way to promote. However, an image with explicit improper content can be detected by AI easily. Nevertheless, AI does not always work: miscreants use adversarial images[6] to evade AI detection.

Yuan et al.[5] proposed Adversarial Promotional Porn Image (APPI), a kind of image that is essentially an adversarial image and makes use of misclassification of models to evade AI detections (e.g., Yahoo Open NSFW model[16], Google Cloud vision API[17]) by rotation, noising, and other image processing methods, images with improper and promotional contents can evade existing state-of-the-art models.

Jargon is also a method to evade detection. Potential users can recognize them easily, but non-potential users and AI have no idea.

Yang et al.[4] found that miscreants use two main methods - homophonic jargon and homomorphic jargon - to change the original context to evade AI. For humans, it is still readable and understandable. Wang et al.[1] Systematically discovered a new type of underground illicit drug promotion on local search services. Miscreants pollute local business listing providers' purchased data to contaminate knowledge bases to promote illicit drug.

2.2 Seduction

Internationalized Domain Names (IDNs) [7]were introduced and standardized in 2003 to support Unicode characters from various languages. Hu et al.[7] found that IDN can be used to impersonate other domain names for phishing purposes. For example, the Latin character "a" resembles the Cyrillic character "a." It makes use of homomorphic characters or letters. Lin et al. [8] found that phishing websites use similar logos and UIs to cheat users.

Wang et al.[9] discovered an underground industrial e-commerce fraud activities chain mainly of crowdsourcing. E-commerce miscreants advertise their attack toolkits and services using group chat, seek further collaborations and share purchase links through IM-based social networks such as Telegram, QQ, etc.

Lee et al.[10] found a new promotional way through potentially-harmful illicit UI(PHI-UI) which are hidden behind benign UIs so that it can bypass the review of the APP store.

III. Defense

3.1 Against Explicit Advertisement

Yuan et al.[5] built a measurement tool, Malena, to measure APPIs. They use R-CNN[21] to locate porn areas and ResNet-50[22] to determine if the target

contains porn image content. As for inserted promotional content, the text part can be extracted by PixelLink[20]. Moreover, they built a tool for QR codes.

Yuan et al.[11] built Semantic Comparison Model based on Word2Vec neural network[12] to detect jargon for cybercrime purposes. Ke et al.[13] built a Chinese jargon unsupervised detection model based on BERT[14].

Yang et al.[4] used several techniques towards homophonic jargon and homomorphic jargon. For homophonic jargon, they normalized them by understanding the sentiment by context and using Pinyin similarity distance. For homomorphic jargon, they used Four-Corner Method[18], which has proven very efficient for such problems related to Chinese character shape. Wang et al.[1] built a measurement tool and used a graph mining algorithm to exploit illicit drug promotional content.

3.2 Against Seduction

Considering that direct comparison of website screenshots is not only less efficient but also less accurate, Lin et al.[8] proposed a deep learning-based detection tool that extracts brand icon information from the page. To address the problem that traditional image similarity detection methods cannot detect unknown graphics, Abdelnabi et al.[15] innovatively used a ternary convolutional neural network model to implement VisualPhishNet, a similarity-based method for detecting phishing attacks based on screenshots of websites. Wang et al.[9] mined the industry chain and built a chatbot on mining the industrial chain more deeply meanwhile, Phishing the miscreants.

Lee et al.[10] built a measurement tool, Chameleon-Hunter, to find hidden UIs related to semantic inconsistency.

For IDN, browsers typically implement rules to detect homographs that may impersonate homomorphic IDNs that may impersonate legitimate domain names. Once identified, browsers will no longer display Unicode but its Punycode to alert the user. Hu et al.[7] systematically evaluated browser-level defenses against homograph IDNs. They verified that all major browser rules have blind spots that can be circumvented through automated tests. Another way to defend against Seduction is blacklisting. The Chinese government encourages citizens to install an app called ''National Anti-Fraud Center''[19], which is advertised to effectively prevent illicit seductions. However , this is not considered a good prevention method because of permission abuse.

IV. Conclusion

It is a constant adversarial process to employ new strategies and techniques to

Paper	IOP method	Defense
Wang et al. [1]	Contaminate knowledge bases	Graph mining
Yang et al. [4]	Homophonic and homomorphic jargon	Pinyin similarity + Four-corner method
Yuan et al. [5]	APPI	Computer vision-based measurement
Yuan et al. [11]	Underground dark jargon	Word2Vec-based Semantic Comparison Model
Ke et al. [13]	Chinese underground jargon	BERT-based unsupervised detection
Hu et al. [7]	Homograph IDN	Measure major browsers' defense against IDN
Lin et al. [8]	Similar logos and UIs	Deep learning-based detection
Wang et al. [9]	Underground E-commerce	Data mining + Chatbot
Lee et al. [10]	PHI-UI	Semantic inconsistency
Abdelnabi et al.[15]	Phishing websites	Similarity-based detection

Table 1. Analysis of IOP methods and Defenses

make OIP avoid being detected and to develop new strategies and techniques to detect OIP. To fight against current OIP, many works discovered it and proposed defenses. Although some work's defenses are weak, they are still meaningful because they measured specific OIP targets and mitigate them more or less.

This paper introduces the main techniques used in OIP and the existing defense methods from Explicit Advertisement and Seduction (Table 1).

[References]

1. Wang, Peng, Zilong Lin, Xiaojing Liao, and XiaoFeng Wang. "Demystifying Local Business Search Poisoning for Illicit Drug Promotion."
2. Cranor, L.F. and B.A. LaMacchia, Spam! Communications of the ACM, 1998. 41(8): p. 74-83.
3. Shin, Y., et al., A link graph-based approach to identify forum spam. Security and Communication Networks, 2015. 8(2): p. 176-188.
4. Yang, R., et al. Scalable Detection of Promotional Website Defacements in Black Hat {SEO} Campaigns. in 30th USENIX Security Symposium (USENIX Security 21). 2021.
5. Yuan, K., et al. Stealthy porn: Understanding real-world adversarial im- ages for illicit online promotion. in 2019 IEEE Symposium on Security and Privacy (SP). 2019. IEEE.
6. Hendrycks, D. and K. Gimpel, Early methods for detecting adversarial images. arXiv preprint arXiv:1608.00530, 2016.
7. Hu, H., et al. Assessing Browser-level Defense against {IDN-based} Phishing. in 30th USENIX Security Symposium (USENIX Security 21). 2021.
8. Lin, Y., et al. Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages. in 30th USENIX Security Symposium (USENIX Security 21). 2021.
9. Wang, P.W., et al. Into the Deep Web: Understanding E-commerceFraud from Autonomous Chat with Cybercriminals. in Proceedings of the ISOC Network and Distributed System Security Symposium (NDSS), 2020. 2020.
10. Lee, Y., et al., Understanding Illicit UI in iOS apps Through Hidden UI Analysis. IEEE Transactions on Dependable and Secure Computing, 2019. 18(5): p. 2390-2402.
11. Yuan, K., et al. Reading Thieves' cant: automatically identifying and understanding dark jargons from cybercrime marketplaces. in 27th USENIX Security Symposium (USENIX Security 18). 2018.
12. Church, K.W., Word2Vec. Natural Language Engineering, 2017. 23(1): p. 155-162.
13. Ke, L., X. Chen, and H. Wang. An Unsupervised Detection Framework for Chinese Jargons in the Darknet. in Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022.
14. Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
15. Abdelnabi, S., K. Krombholz, and M. Fritz. VisualPhishNet: Zero-day phishing website detection by visual similarity. in Proceedings of the 2020 ACM SIGSAC conference on computer and communications security. 2020.
16. https://github.com/yahoo/open_nsfw18
17. https://cloud.google.com/vision
18. https://en.wikipedia.org/wiki/Four-Corner_Method
19. https://www.ft.com/content/84b6b889-ae03-47f7-9cd0-bd604b21d5de
20. Deng, Dan, Haifeng Liu, Xuelong Li, and Deng Cai. "Pixellink: Detecting scene text via instance segmentation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. 2018.
21. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.
22. Zhong, Zilong, Jonathan Li, Lingfei Ma, Han Jiang, and He Zhao. "Deep residual networks for hyperspectral image classification." In 2017 IEEE international geoscience and remote sensing symposium (IGARSS), pp. 1824-1827. IEEE, 2017.