Measuring Illegal Online Gambling Defacement in Indonesia

Luqman Muhammad Zagi [1*], Girindro Pringgo Digdo[2], Wervyan Shalannanda[3]
[1]Radboud University, Nijmegen, the Netherlands
[2]Cyber Army Indonesia, Bandung, Indonesia
[3]Institut Teknologi Bandung, Bandung, Indonesia
*E-mail: luqman.zagi@ru.nl

ABSTRACT

Website defacement remains a persistent cybersecurity challenge, with attackers leveraging compromised platforms to promote illicit activities, disseminate malicious content, or undermine the credibility of trusted institutions. In recent years, government agencies and educational institutions in Indonesia have increasingly become targets of such attacks. One particularly concerning trend involves the systematic insertion of illegal online gambling (IOG) content into legitimate websites. Reports (DarkRadar, 2025) indicate that more than 2,500 Indonesian websites were compromised within a four-month period by actors seeking to exploit institutional trust and visibility to advertise gambling services. Such incidents not only degrade the reliability of online services but also carry reputational, financial, and legal consequences for the affected entities. These developments underscore the urgent need for practical, scalable, and reproducible approaches to measuring and detecting website defacement, particularly in environments where sophisticated defenses may not be feasible.

Globally, studies highlight both the sophistication of gambling operators and the challenges of regulation. Research in Asia has revealed coordinated infrastructures, advanced promotion strategies, and extensive abuse of third-party services, while European work has demonstrated that regulatory measures (such as Spain's Royal Decree 958/2020) can reduce new gambling accounts, marketing expenditures, and money wagered (Aonso-Diego et al., 2025; Huang et al., 2022; Yang et al., 2019). At the technical level, machine learning and deep learning models have achieved strong performance for defacement detection but are computationally intensive, require extensive training data, and degrade under domain shift (Nguyen et al., 2021; Vinayagam et al., 2024). These limitations restrict their use for continuous, real-world monitoring.

To address this gap, we present a lightweight and reproducible methodology based on keyword-driven dorking combined with systematic crawling. The method leverages DuckDuckGo's Tracker Radar Collector (TRC), an open-source crawling framework widely used in measurement studies. We implemented custom collectors for keywords, HTML, snapshots, and internal links, enabling both first-time detection and repeated monitoring. The data collection pipeline was orchestrated by two handlers: the *list handler*, which operated hourly to capture newly reported defaced URLs from DarkRadar, and the *internal list handler*, which executed daily to crawl third-party internal links extracted from defaced pages. This dual approach ensured both breadth and depth in capturing gambling-related compromises.

Data was collected between 27 May and 25 June 2025 using a virtual machine in Jakarta with eight CPU cores and 8 GB RAM. During this period, the system identified 453 webpages potentially affected by defacement. After filtering and verification, 346 pages were confirmed as defaced, corresponding to 147 unique websites. The number of defaced pages per domain ranged from 1 to 33, with a mean of 2.35 and a median of 1. Domain-level analysis revealed that defacements spanned eight of the 13 Indonesian top-level domains (TLDs). The most affected were *ac.id* (117 pages across 65 domains), *go.id* (115 pages across 39 domains), and *co.id* (46 pages across 12 domains). This distribution indicates that academic and governmental institutions were disproportionately targeted due to their visibility and credibility.

Keyword analysis showed that the most prevalent was *slot* (95.2%), followed by *bet* (90.5%), while *zeus* was least common (27.2%). Some keywords produced false positives, especially *bet*, due to substring matches with tokens like *between* and *beta*. Overall, the false positive rate was 20.3%, but crawling and keyword counting reduced misclassifications effectively.

The persistence of defacement varied widely. We categorized cases as *repeat defacement* (150), *fixed* (129), *fluctuating*, and *constant*. A total of 102 pages were defaced only once, while others cycled repeatedly between compromise and repair, with the highest reaching 112 cycles. Redirection was identified on 28 pages across 13 sites, sometimes involving multiple gambling domains. Title manipulation was observed in 84 cases, though none displayed explicit “hacked by” signatures.

We also analyzed third-party and internal URLs. In total, 8,837 unique URLs from 5,930 domains were injected into defaced pages. Most appeared only once, though nine URLs were reused across 21 different pages. The most frequent domains were *heylink.me* (16), *linklst.bio* (9), and *lazada.co.id* (9), the first two being link-shortening services often used to hide redirection paths. From 8,837 unique URLs, 5,830 internal URLs were successfully queried. Of these, 567 were hosted behind Cloudflare—sometimes associated with IOG-related infrastructure—1,266 returned errors, 2,340 led directly to IOG websites, and the rest were unrelated. This demonstrates that not all internal links from defaced pages resolve to gambling sites, suggesting selective link manipulation by attackers.

Response times were inconsistent. Of 279 URLs with measurable recovery, 43.4% were fixed within 24 hours of first defacement, compared to 36.2% when considering average cycles. Conversely, 13.3% required more than a week initially, while only 10.4% took longer than a week on average. The mean initial reaction time was 74.7 hours, with an overall average of 75.3 hours, reflecting uneven and often delayed remediation. Some domains were suspended, others partially repaired, and many fell into repeated cycles of compromise.

Our findings offer three main insights. First, attackers are diversifying techniques, relying on hidden URLs, redirection, and link injection in addition to direct content manipulation. Second, responses from institutions are inconsistent, with many acting quickly once but struggling to sustain defenses during repeated attacks. Third, lightweight methods such as dorking and crawling, while less sophisticated than machine learning, provide a scalable, reproducible, and resource-efficient means of capturing defacement dynamics.

Future work should extend the observation period beyond one month to capture seasonal patterns, analyze internal links in greater depth, and incorporate technology stack profiling to explore correlations between specific frameworks and susceptibility to defacement.

In conclusion, this study shows that Indonesian websites, particularly academic and governmental domains, face persistent threats from gambling-related defacement. By combining dorking with crawling, we provide a simple yet powerful methodology for measuring defacement at scale. The results underscore the need for continuous monitoring and provide actionable evidence to strengthen institutional responses against the growing threat of illicit online gambling.

Keywords: *Measurement, Illegal Online Gambling, Defacement, Indonesia*

References

Aonso-Diego, G., García-Pérez, Á., & Krotter, A. (2025). Impact of Spanish gambling regulations on online gambling behavior and marketing strategies. *Harm Reduction Journal*, *22*(1), 107. https://doi.org/10.1186/s12954-025-01219-7

DarkRadar. (2025). *Laporan Insiden Defacement Domain go.id*. CyberArmy.id. https://www.linkedin.com/posts/girindigdo_darkradar-laporan-insiden-defacement-domain-activity-7329053959891968000yJgT?utm_source=social_share_send&utm_medium=member_desktop_web&rcm=ACoAACJWCQ0Bl-DrNBA9x9XWVwT8nSEUc2ztAPk

Huang, R.-T., Shih, C.-H., Lin, T.-C., & Lin, W.-Y. (2022). The Study on Illegal Online Gambling Investigation in Taiwan. *Procedia Computer Science*, *207*, 2901–2910. https://doi.org/10.1016/j.procs.2022.09.348

Nguyen, T. H., Hoang, X. D., & Nguyen, D. D. (2021). Detecting Website Defacement Attacks using Web-page Text and Image Features. *International Journal of Advanced Computer Science and Applications*, *12*(7). https://doi.org/10.14569/IJACSA.2021.0120725

Vinayagam, V. M., Sathish, A., Sreenivasulu, K., Sreenu, D., Snehit, R., & Reddy, Y. A. (2024). Web Defacement Identification and Detection System. *2024 International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)*, 1013–1019. https://doi.org/10.1109/I3CEET61722.2024.10994031

Yang, H., Du, K., Zhang, Y., Hao, S., Li, Z., Liu, M., Wang, H., Duan, H., Shi, Y., Su, X., Liu, G., Geng, Z., & Wu, J. (2019). Casino royale: A deep exploration of illegal online gambling. *Proceedings of the 35th Annual Computer Security Applications Conference*, 500–513. https://doi.org/10.1145/3359789.3359817