Detecting Roach Motel Dark Patterns Through Visual, Behavioral, and Semantic Analysis*

Nayeon Ryu and Seyoung Lee

Kangwon National University, ChunCheon, South Korea
01star01ek@kangwon.ac.kr, seyoung@kangwon.ac.kr

Abstract

Roach Motel is a deceptive dark pattern that restricts user choices by intentionally complicating the process of service termination and unsubscribing. It undermines transparency in digital environments by manipulating user decisions without consent, posing an increasing threat to user rights. We propose a hierarchical method that analyzes web pages' visual, behavioral, and semantic elements to systematically detect dark patterns and recommend effective measures for user protection. The proposed solution aims to enhance user autonomy and promote transparency in digital environments by discerning these harmful patterns.

Keywords: Dark Pattern Detection, Roach Motel, Hierarchical Approach

1 Introduction

Deceptive design strategies, known as dark patterns, exploit users by guiding their behavior to benefit businesses rather than the users themselves [1]. One such pattern, the Roach Motel, intentionally complicates everyday actions like unsubscribing or canceling services, effectively trapping users in unwanted situations. Unlike more overt dark patterns, the Roach Motel is subtle and difficult to detect. Research shows that the recognition rate of this pattern is only 18.6%[2]. As online subscription models continue to grow, this pattern increasingly restricts user choices and complicates subscription cancellation in a number of ways.

To address this issue, there is a pressing need for systematic methods to detect and mitigate Roach Motel patterns in web interfaces. In this paper, we propose a hierarchical approach that analyzes websites across three layers: visual, behavioral, and semantic. Our method assesses the accessibility and visibility of interface elements, evaluates user navigation to identify excessive or unnecessary steps, and detects misleading or coercive content. By integrating these layers, this approach offers a comprehensive and robust framework for identifying Roach Motel patterns and safeguarding user autonomy.

2 Our Approach

To effectively detect the deceptive Roach Motel pattern, we propose a multi-layered analysis approach that comprehensively examines the user experience across three dimensions: visual, behavioral, and semantic. Each layer provides a distinct perspective on how the website's interface may mislead users.

The **visual layer** analyzes a website's HTML and CSS code using Selenium to identify key

*Proceedings of the 8th International Conference on Mobile Internet Security (MobiSec'24), Article No. P-74, December 17-19, 2024, Sapporo, Japan.

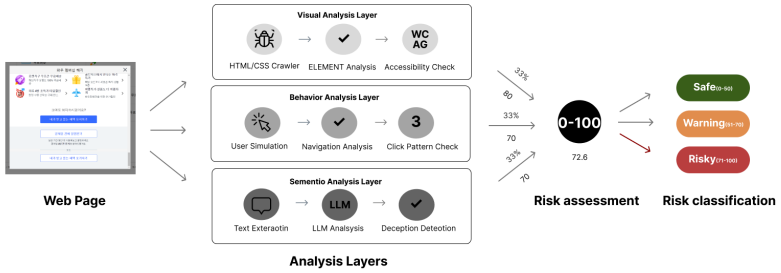

Figure 1: System Overview of the Roach Motel Dark Pattern Detection

elements like $< a >$ tags and keywords such as "delete" and "cancel." It then examines the CSS properties of these elements, including opacity, visibility, text size, and color, assessing interface accessibility according to WCAG standards, such as measuring text-to-background contrast ratios. The **behavioral layer** tracks user interactions, precisely the number of clicks and page traversals required to complete tasks. Using Selenium to simulate user scenarios, it records users' steps for actions like unsubscribing or canceling services. The system evaluates navigation paths based on the 3-Click Rule, identifying unnecessary steps that hinder users from efficiently reaching their goals. The **semantic layer** analyzes the text on a web page to detect language that may limit user choices. OCR is used to extract text near keywords like "delete" and "cancel," and Large Language Models analyze the text for manipulative or deceptive expressions, measuring the degree of ambiguity and pressure in the language. Finally, the results from each layer are combined with equal weights—33.3%, producing a final risk score. Based on the overall score, websites are then categorized into three tiers—Safe (0-50), Warning (51-70), and Risky (68-100)—providing a clear assessment of potential risks, as shown in Fig 1.

3 Conclusion

We propose a hierarchical analysis approach that efficiently detects Roach Motel dark patterns by combining visual, behavioral, and semantic analysis. This multi-layered approach overcomes the limitations of single-method detection and provides objective, quantitative risk assessments. Future work will expand the framework to detect a wider range of dark patterns, further enhancing its effectiveness in protecting user autonomy.

Acknowledgments

This work was supported by the Korea government (MOE) grant as part of the 2024 government collaboration type training project [Information Security Field] (No. 2024 Personal Information Protection-003) and the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-II221196, Regional Strategic Industry Convergence Security Core Talent Training Business).

References

[1] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021*

CHI conference on human factors in computing systems, pages 1–18, 2021.

[2] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. Towards the identification of dark patterns: An analysis based on end-user reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, pages 24–33, 2020.