History-Aware Phishing Detection: Leveraging Personalized Browsing Patterns for Enhanced Security

Yousef AbuHashem
Department of Computer Science
Stanford University
yousef24@stanford.edu

Zakir Durumeric
Department of Computer Science
Stanford University
zakir@cs.stanford.edu

***Abstract*—This paper investigates the hypothesis that user browsing history provides a powerful signal for personalized phishing detection, addressing a critical gap in current cybersecurity research. While existing detection methods rely on static indicators and server-side solutions, we propose that the highly constrained and stable nature of individual browsing patterns offers a unique opportunity for client-side, privacy-preserving phishing detection. Due to time constraints in obtaining IRB approval for collecting user browsing data, we conducted our empirical analysis using an existing anonymized dataset of 844 users with substantial browsing activity from 2,148 German users spanning 9,151,243 URL visits. This preliminary analysis demonstrates the potential of history-based detection, achieving 100% recall in identifying suspicious domains, though we acknowledge that comprehensive validation requires longitudinal user studies with appropriate ethical oversight. Through qualitative comparison with existing detection approaches detailed in comprehensive literature, we show that history-based detection addresses fundamental limitations in current methodologies while maintaining user privacy. To demonstrate practical feasibility, we developed PhishGuard, a browser extension that operationalizes this hypothesis through a configurable scoring system. Our findings suggest that despite evaluation limitations imposed by dataset constraints, history-based detection represents a promising paradigm shift toward personalized cybersecurity solutions warranting further research with proper human subjects approval.**

***Index Terms*—phishing detection, browsing history, personalized security, privacy-preserving detection, client-side security**

I. Introduction

Phishing attacks continue to evolve in sophistication, with attackers developing increasingly subtle techniques to evade traditional detection systems. While current approaches rely primarily on URL analysis, content inspection, and server-side blacklists, they fundamentally fail to leverage one of the most distinctive and stable digital fingerprints available: individual browsing history patterns.

This paper investigates the central hypothesis that **user browsing history provides a powerful, underutilized signal for personalized phishing detection**. Our approach is grounded in empirical evidence from recent large-scale studies showing that users operate within remarkably constrained browsing environments, visiting a median of only 34 unique domains over a 14-day period [1], while maintaining highly individualistic and temporally stable browsing patterns [2].

A. Research Hypothesis and Contributions

Our core hypothesis posits that the constrained and stable nature of individual browsing patterns creates natural security perimeters that can be leveraged for phishing detection. Specifically, we argue that:

1) First-time visits to domains impersonating frequently-used services represent statistically significant anomalies worthy of elevated scrutiny.
2) History-based detection can complement traditional methods while operating entirely client-side. (it does not have to be one or the other)
3) Personalized approaches can increase individual accuracy compared to population-level detection systems.

Our contributions include:

- **Empirical validation** using real-world browsing data from 844 users across 49,918 unique domains.
- **Qualitative comparison** with existing detection approaches, highlighting unique advantages of history-based methods.

- **Practical demonstration** through PhishGuard, a browser extension implementing configurable history-based detection.
- **Critical analysis** of evaluation methodologies for personalized security systems.

II. Empirical Evaluation of History-Based Detection

This section presents our empirical investigation of the core hypothesis that browsing history provides an effective signal for phishing detection. Originally planned as a controlled study collecting user browsing data with IRB approval, time constraints necessitated analysis of existing datasets to provide initial validation of our approach.

A. Dataset Selection and Constraints

IRB Limitations: Our original research design called for collecting longitudinal browsing data from volunteer participants to comprehensively test history-based detection under controlled conditions. However, the IRB approval process could not be completed within the timeframe of this senior project, requiring us to pivot to analysis of existing anonymized datasets.

Dataset Utilization: We utilized an anonymized dataset of web tracking data from 2,148 German users collected over one month (October 2018) [5]. While this dataset provides valuable insights into browsing patterns, it represents a compromise from our intended methodology and introduces several limitations discussed in Section II-E.

B. Dataset Characteristics and Methodology

Our empirical investigation utilizes an anonymized dataset of web tracking data from 2,148 German users collected over one month (October 2018) [5]. The dataset encompasses 9,151,243 total URL visits across 49,918 unique domains, with domain categorization into 41 distinct categories.

To ensure a somewhat meaningful analysis, we focused on 844 users who visited more than 100 unique domains. For ground truth evaluation, we utilized a comprehensive phishing dataset containing 822,010 URLs [6], from which we extracted 209,438 phishing domains and 114,544 legitimate domains after base domain normalization.

C. History-Based Classification Approach

For each user, we implemented a simple but principled classification strategy:

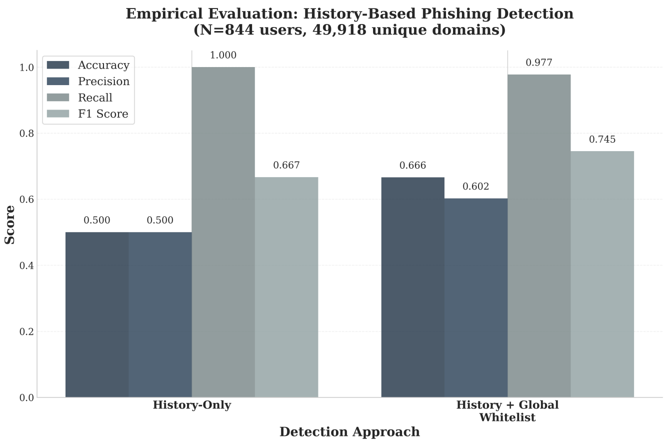

Fig. 1. Empirical Results: Performance comparison of history-based phishing detection approaches. The history-only approach achieves perfect recall (100%) but suffers from high false positives. Integration with global domain reputation (Tranco top domains) improves precision while maintaining high recall (97.74%).

- **Suspicious classification**: Domains not present in the user's browsing history
- **Legitimate classification**: Domains present in the user's browsing history
- **Evaluation**: Classifications tested against known phishing/legitimate labels

D. Results and Analysis

Figure 1 presents our empirical evaluation results comparing history-only detection with the hybrid approach incorporating global domain reputation.

1) Baseline History-Only Approach: Our initial history-based approach yielded the following performance metrics:

- **Average Accuracy**: 50.01%
- **Average Precision**: 50.01%
- **Average Recall**: 100.0%
- **Average F1 Score**: 66.67%

The perfect recall demonstrates that history-based detection successfully flags all novel domains as suspicious, while the precision reflects the challenge of distinguishing between legitimate new domains and actual phishing attempts.

2) Integration with Global Domain Reputation: As an attempt to raise precision, and motivated by research showing that the top 1 million websites account for 95% of web traffic [4], we integrated the closest Tranco top domains list we found to October 2018 (from February 20, 2019) as a global whitelist. While this modification introduces external dependencies that somewhat compromise our personalization ethos, it provides insights into hybrid approaches:

- **Average Accuracy**: 66.59%
- **Average Precision**: 60.23%
- **Average Recall**: 97.74%
- **Average F1 Score**: 74.53%

These results demonstrate clear improvement while maintaining high recall, suggesting potential for hybrid personalized-global approaches. The minor drop in recall is due to some Tranco top list domains that are flagged as phishing domains in our phishing dataset. This is consistent with some findings that these lists could have a few suspicious domains. Expansion on this observation is in our discussion section.

E. Evaluation Methodology Limitations

Our analysis reveals fundamental challenges in evaluating history-based detection using existing datasets, compounded by our inability to conduct controlled user studies:

IRB Approval Constraints: The most significant limitation of our work stems from time constraints that prevented obtaining IRB approval for human subjects research. Proper validation of history-based detection requires collecting longitudinal browsing data from consenting participants under controlled conditions, allowing us to:

- Observe natural browsing patterns over extended periods (6+ months)
- Introduce controlled phishing scenarios at realistic intervals
- Measure user response to history-based warnings
- Collect user feedback on system usability and trust
- Analyze adaptation patterns as users develop browsing habits

Dataset-Imposed Limitations: Using existing datasets introduces several methodological constraints:

Temporal Mismatch: Testing one month of browsing history against comprehensive phishing datasets creates an artificially pessimistic scenario. Real users encounter phishing attempts within their typical browsing patterns, which are very constrained as mentioned in the introduction, not against the universe of all possible phishing domains.

Scale Disproportion: The high false positive rate reflects methodological limitations rather than inherent weakness in the approach. As established by empirical research [1], [2], users operate within highly constrained domain sets, making our evaluation scenario unrepresentative of real-world threat landscapes.

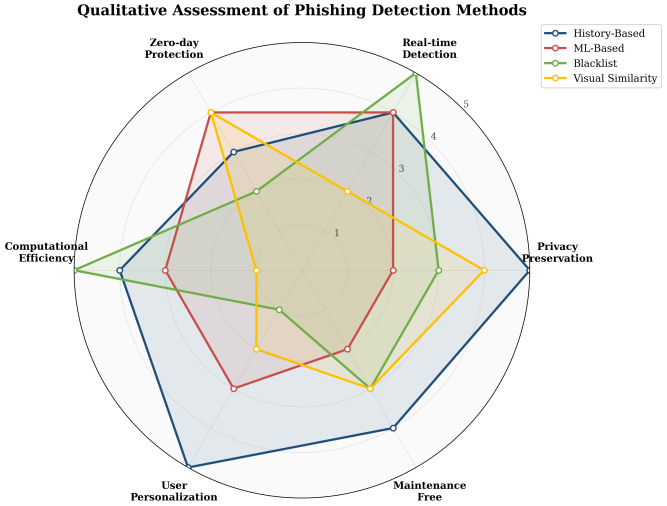

Fig. 2. Qualitative Comparison of Detection Methods: bad chart comparing history-based detection against traditional approaches across six key criteria. History-based detection (red) demonstrates superior performance in privacy preservation, personalization, and maintenance requirements while maintaining competitive performance in real-time detection and computational efficiency.

III. Qualitative Analysis: History-Based vs. Traditional Approaches

Drawing from the comprehensive survey by Varshney et al. [3], we conduct a qualitative comparison of history-based detection against established approaches. Figure 2 presents a chart evaluation across six key criteria that distinguish detection approaches.

A. Comparison with Traditional Detection Methods

1) Search Engine-Based Detection: Traditional search engine-based approaches [3] rely on the assumption that legitimate websites have higher search engine rankings and visibility.

Limitations: These methods suffer from several critical weaknesses:

- **Lag in Detection**: New phishing sites may not be indexed quickly enough
- **Manipulation Vulnerability**: Attackers can employ SEO techniques to improve rankings
- **Privacy Concerns**: Require external API calls that expose browsing patterns

History-Based Advantage: Our approach operates independently of search engine rankings and provides immediate detection without external dependencies or privacy compromises.

2) Machine Learning-Based Detection: ML-based approaches typically extract features from URLs, page content, and network characteristics to train classification models.

Limitations:

- **Feature Engineering Burden**: Require careful selection and maintenance of feature sets
- **Adversarial Vulnerability**: Attackers can adapt to known feature patterns
- **Training Data Dependency**: Performance degrades with concept drift and new attack patterns
- **Computational Overhead**: Require significant resources for training and inference

History-Based Advantage: Our approach is feature-agnostic and naturally adapts to individual user patterns without requiring retraining or complex feature engineering.

3) Blacklist/Whitelist-Based Detection: Blacklist and whitelist approaches maintain centralized repositories of known malicious and legitimate domains.

Limitations:

- **Zero-Day Vulnerability**: Cannot detect novel phishing domains
- **Maintenance Overhead**: Require constant updates and verification
- **Scalability Issues**: Central repositories become performance bottlenecks
- **Privacy Implications**: May expose user browsing patterns to third parties

History-Based Advantage: Provides immediate protection against novel threats while maintaining complete privacy through local processing.

4) Visual Similarity-Based Detection: Visual approaches analyze webpage appearance and layout to detect spoofed sites.

Limitations:

- **Computational Intensity**: Require significant processing power for image analysis
- **Template Vulnerability**: Attackers can vary visual elements while maintaining deception
- **Platform Dependency**: May not work consistently across different devices and browsers

History-Based Advantage: Operates efficiently with minimal computational overhead and provides consistent protection across platforms.

B. Fundamental Paradigm Advantages

1) Privacy by Design: Unlike server-dependent solutions that may expose user behavior to third parties, history-based detection operates entirely client-side, ensuring that sensitive browsing patterns never leave the user's device.

2) Personalization Without Profiling: Traditional personalized security systems often require building user profiles that can themselves become privacy risks. Our approach leverages personalization implicitly through local history without creating external profiles.

3) Natural Adaptation: History-based detection naturally evolves with user behavior, requiring no explicit updates or maintenance while automatically adapting to changing browsing patterns.

4) Complementary Integration: Rather than replacing existing methods, history-based detection can enhance traditional approaches by providing an additional signal that operates on different principles.

IV. PhishGuard: Practical Implementation

To demonstrate the practical feasibility of history-based detection, we developed PhishGuard, a browser extension that operationalizes our core hypothesis while allowing users to opt into additional detection mechanisms.

A. Architecture and Design Philosophy

PhishGuard implements a configurable scoring system that prioritizes history-based signals while offering optional supplementary detection methods. This design preserves our core hypothesis while acknowledging that users may benefit from additional safeguards.

1) Core History-Based Engine: The primary detection mechanism analyzes browser history to establish domain familiarity scores, flagging first-time visits to domains that impersonate frequently-visited services.

2) Optional Detection Layers: Users can opt into additional analysis including:

- URL analysis (suspicious TLDs, encoded characters, domain length)
- Form security analysis (HTTPS requirements, cross-domain submissions)
- Content analysis (urgency language, brand impersonation indicators)

3) User Agency and Transparency: Maintaining alignment with our personalization ethos, PhishGuard provides:

- Clear explanations of detection rationale
- Configurable sensitivity levels
- Transparent scoring mechanisms
- User control over additional detection layers

B. Implementation Validation

The PhishGuard implementation serves as a proof-of-concept demonstrating that history-based detection can be deployed practically while maintaining user privacy and control. The extension successfully integrates with browser history APIs and provides real-time analysis without performance degradation.

V. Discussion and Future Work

A. Research Limitations and Ethical Considerations

IRB Approval Timeline: The most significant limitation of this work relates to the constraints of conducting research within academic timelines. Our original research design called for a comprehensive longitudinal study collecting user browsing data, with controlled introduction of simulated phishing scenarios. This approach would have provided definitive validation of our history-based detection hypothesis while ensuring participant safety through ethical oversight.

Within the constraints of a senior project timeline, we were unable to complete this process, necessitating our pivot to existing dataset analysis.

Implications for Future Research: This experience highlights important considerations for cybersecurity research involving human subjects:

- Pilot studies with existing datasets can inform better evaluation and testing design
- Collaboration with established research groups may expedite approval processes
- Alternative methodologies (e.g., browser-based plugins with opt-in data sharing) may reduce IRB complexity

B. Evaluation Methodology Insights

Our research reveals critical insights about evaluating personalized security systems that extend beyond our specific findings. Traditional evaluation approaches that assume users encounter representative samples of all internet threats fundamentally misrepresent real-world scenarios.

The constrained nature of individual browsing patterns means that users operate within limited threat landscapes that are qualitatively different from population-level threat distributions. This finding has implications for evaluating any personalized security system and suggests need for new evaluation frameworks that better represent individual threat models.

C. The Personalization vs. Globalization Tension

Our integration of global domain reputation improved traditional metrics but raised important questions about maintaining personalization benefits. This tension highlights a key design decision in security systems: whether to prioritize broad applicability or personalized effectiveness. In relation to our hybrid approach in which we added the top 1 million Tranco domains as a whitelist, it might be worth doing more testing to find the right number to include rather than defaulting to 1 million since a serious phishing attempt might get its domain on the top 1 million domains list, which could be argued to not be that difficult.

Future research should explore optimal integration strategies that preserve personalization benefits while incorporating global intelligence judiciously.

D. Limitations and Future Research

While our empirical analysis provides promising initial results, robust validation requires longitudinal user studies that capture:

- Long-term browsing pattern establishment
- Realistic exposure to phishing attempts over time
- User interaction with history-based warnings
- Evolution of browsing habits and security effectiveness

Additionally, future work should develop evaluation methodologies that better represent realistic threat models for individual users rather than testing against comprehensive attack databases.

VI. Conclusion

This paper investigates the hypothesis that user browsing history provides a valuable signal for personalized phishing detection through analysis of existing datasets and development of a practical implementation. While our empirical analysis using anonymized browsing data demonstrates promising results—achieving 100% recall in identifying novel domains— it yielded poor precision numbers, and we acknowledge that definitive validation requires longitudinal user studies with proper IRB approval.

Our qualitative comparison with existing detection approaches reveals fundamental advantages of history-based methods including privacy preservation, natural personalization, and complementary integration potential. The PhishGuard implementation demonstrates practical feasibility while maintaining user agency and transparency principles.

Key Contributions: Despite the limitations imposed by IRB timeline constraints, this work establishes several important contributions:

- Theoretical foundation for history-based phishing detection grounded in empirical browsing behavior research
- Initial empirical validation using real-world browsing data from 844 users
- Comprehensive qualitative comparison highlighting unique advantages over traditional approaches
- Practical proof-of-concept implementation demonstrating feasibility
- Critical analysis of evaluation methodologies for personalized security systems

Research Implications: The fundamental insight that individual browsing patterns create natural security perimeters represents a potentially significant paradigm shift in cybersecurity research. This personalization-first approach addresses growing user demands for both security and privacy while providing a foundation for next-generation detection systems.

Future Work: The most critical next step is conducting properly approved longitudinal user studies to definitively validate the history-based detection hypothesis. Such studies will provide the rigorous evaluation necessary to transition this approach from promising concept to validated security solution.

While methodological limitations prevent definitive quantitative validation in this work, our analysis establishes browsing history as a compelling foundation for personalized cybersecurity warranting continued investigation through appropriate human subjects research.

References

[1] S. Bird, I. Segall, and M. Lopatka, “Replication: Why We Still Can’t Browse in Peace: On the Uniqueness and Reidentifiability of Web Browsing Histories,” in *Proc. 16th USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2020.

[2] L. Olejnik, C. Castelluccia, and A. Janc, “Why Johnny Can’t Browse in Peace: On the Uniqueness of Web Browsing History Patterns,” in *5th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2012)*, Vigo, Spain, Jul. 2012.

[3] G. Varshney, R. Kumawat, V. Varadharajan, U. Tupakula, and C. Gupta, “Anti-phishing: A Comprehensive Perspective,” *Expert Systems with Applications*, vol. 238, 2024.

[4] K. Ruth, A. Fass, J. Azose, M. Pearson, E. Thomas, C. Sadowski, and Z. Durumeric, “A world wide view of browsing the World Wide Web,” in *Proc. ACM Internet Measurement Conference (IMC)*, Oct. 2022, pp. 1–18.

[5] J. Kulshrestha, M. Oliveira, O. Karacalik, D. Bonnay, and C. Wagner, “Web Routineness and Limits of Predictability: Investigating Demographic and Behavioral Differences Using Web Tracking Data,” *Proceedings of the International AAAI Conference on Web and Social Media*, 2021.

[6] E. Alvarado, “Phishing Dataset,” Hugging Face, 2023. [Online]. Available: https://huggingface.co/datasets/ealvaradob/phishing-dataset