Preliminary Report

Challenge 1

Phishing Detection in Cybersecurity

Arsénio Ferraz - 1010137
Bruno Camarneiro - 1250422
César Vieira - 1241523
Gustavo Lima - 1221349
Rui Soares - 1221283

Contents

Work Objectives

The objective of the present work consists in developing an expert system capable of identifying malicious URLs and phishing pages. The system has been designed to provide practical efficiency and decision explainability, all whilst remaining adaptable to emerging attack techniques. This has been accomplished by leveraging a combination of technical indicators, integration with external information sources and the application of symbolic reasoning and heuristic methods. Furthermore, the system received contributions from an expert in the field and recent literature, complementing the quality of its knowledge base.

Specific objectives include:

- Model and formalize the core characteristics of phishing URLs and pages using both practical experience and academic insights
- Establish a robust set of rules and techniques for automated and transparent risk assessment
- Investigate the integration of symbolic reasoning platforms (such as Drools) for hybrid analytical approaches
- Validate the proposed methods with real-world examples supplied by the expert, and simulate complex scenarios such as redirections, DOM manipulation, homograph attacks, and shortened URLs

Knowledge Sources and Expert Characterization

Academic and Technical Sources

- IEEE, arXiv, ACM, Scopus and ResearchGate database research papers
- ENISA and Word Economic Forum landscape reports
- OpenPhish, PhishTank, VirusTotal APIs documentation
- Meetings and Message trading with the expert

Expert Profile

David Marques – Head of Cybersecurity at Grupo Nabeiro / Delta Cafés

Professional Background:

With over 15 years of experience in advanced cybersecurity strategies, incident response, and awareness programs, the professional background involves leading multidisciplinary teams in complex enterprise environments. This experience includes designing adaptive policies and providing guidance on responding to threats such as phishing, supply chain attacks, identity management challenges, and application vulnerabilities.

Key Expertise Areas:

Key areas of expertise include real-world incident analysis and response coordination, implementing enterprise-scale phishing defense solutions, and developing risk communication strategies that prioritize thorough assessment over preventive blocking. Additionally, there is a strong focus on creating internal awareness programs, including the production of video content and interactive workshops, as well as developing cybersecurity portals and driving employee training initiatives.

Contribution to Project:

The contribution to the project included providing real attack examples and validating detection thresholds. Operational challenges and strategies for managing false positives were shared. Recommendations were made for implementing adaptive cycles for rule updates based on lessons learned from incidents. The work also emphasized the importance of identity management, supply chain security, and integrating innovation processes.

Knowledge Acquisition Sessions

Session 1: Preparation and Framework Definition

- Structured interview guide development focusing on real attack patterns
- Technical heuristics identification and operational challenge mapping
- Expert validation of proposed detection methodologies

Session 2: Expert Interview and Validation

- In-depth discussion of real-world phishing incidents and attack vectors
- Analysis of Domain Generation Algorithms (DGAs), CDN abuse, shortener manipulation
- Validation of detection rules and threshold recommendations
- Review of organizational response strategies and false positive handling

Session 3: Rule Consolidation and Technical Implementation

- Translation of expert insights into implementable Prolog and Drools rules
- Threshold calibration based on operational experience
- Integration strategies for external threat intelligence sources

Note: This session is going to be schedule with the Expert during this week.

Session 4: Iterative Refinement

- Continuous feedback integration from expert recommendations
- Rule optimization based on simulated attack scenarios
- Alignment of technical implementation with enterprise operational requirements

Note: Session that we intend to schedule in the upcoming weeks.

Acquired Knowledge Representation

The system architecture follows a two-phase approach validated by expert experience:

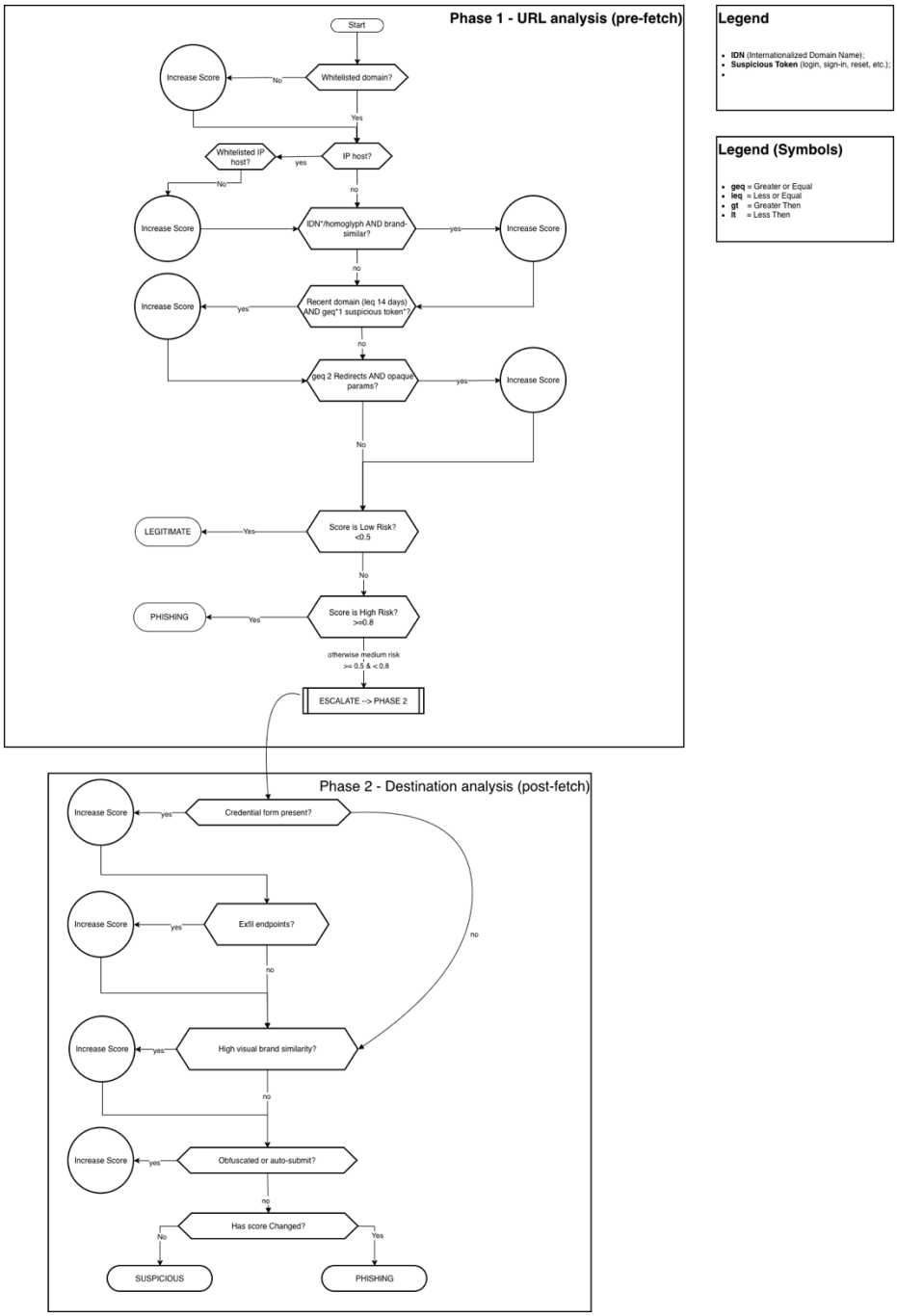

Figure 1 - Diagram

Background Story

Let me take you through the journey, as if you're inside our anti-phishing engine, discovering threats and making decisions in real time.

The Story of Two Links: The Battle Between Legit and Phishing

It's Monday morning, and our security system receives two new web addresses to check. On the surface, both seem harmless. But as the engine begins to analyze, the true story unfolds...

First: The Suspicious Shortcut Adventure

A user clicks:
http://bit.ly/3xZpF8a

Our engine springs into action. First stop,**R1.C - URL Shortener Detection**—this is a classic shortcut: *bit.ly*! It's flagged instantly.

The engine follows the redirect path:

- **R1.A - Redirect Depth Analysis**: One, then two quick jumps: redirect1.com, then finally landing at hack-paypal.com/login. Multiple redirects—something attackers love.
- **R1.B - Domain Diversity**: The domains are all different. That's the attacker's trick—hop across unrelated servers to stay hidden.

As soon as our engine sees hack-paypal.com, it does a quick check:

- **R1.E - Blacklist Check**: Red alert. hack-paypal.com is on our blacklist of known phishers.

All these alarms ring together. Our system scores each rule and combines them, concluding with a bold red warning:
"Phishing Detected - High Confidence"

Second: The Trusted Internal Chain

Another link comes through:
https://secure.intranet.corp.com/app

It quickly hops, just once, to internal-redirect.corp.com/home.

Our engine runs the same rules:

- **R1.A - Redirect Depth Analysis**: Only one redirect, the most common for internal navigation.
- **R1.B - Domain Diversity**: Two domains, but both have the same corporate root—safe so far.
- **R1.C - URL Shortener Detection**: Not detected.
- **R1.E - Whitelist Check**: Both domains well known, trusted and validated by our system.

The engine weighs every signal. This time, the outcome is calm and reassuring:
“Legitimate - No risk detected”

Conclusion:

Our ruleset isn’t just a checklist, it’s a story engine, reading each link’s background, motives, and behaviors. It catches the crafty villain (“hack-paypal.com” hiding behind shortcuts and misdirection) and frees the hero (an everyday internal workflow) without false alarms.

That’s the power of combinatorial logic applied to real-world cybersecurity. Every click, every redirect, our system tells the story and gets the ending of the happy ever after.

Phase 1: Static URL Analysis

Technical Implementation: Prolog-based rule engine

- Domain composition analysis (age, WHOIS, IP presence, subdomains)

- Encoding pattern detection and suspicious token identification

- Homograph and IDN abuse detection

- URL shortener chain analysis

- TLD reputation scoring

Key Rules Formalized:

```Drools
rule "Malicious URL Detection"
when
   $domain : DomainInfo(url == $url, age < 30)
   $pattern : SuspiciousPattern(url == $url)
   $token : CredentialToken(url == $url)
then
   // Ação: marcar ou registrar URL como malicioso
   System.out.println("Malicious URL detected: " + $url);
   // Ou executar lógica de negócio relevante
end```

## Phase 2: Dynamic DOM Analysis

Technical Implementation* Drools-based decision engine

- Rendered HTML inspection for sensitive fields

- External link ratio calculation

- JavaScript exfiltration pattern detection
```

- Page entropy analysis

- Visual brand impersonation detection

Scoring System:

- Risk scores: 0-49 (legitimate), 50-79 (suspicious), 80-100 (malicious)

- Weighted aggregation of multiple indicators

- Adaptive thresholds based on source context

Integration Framework

- External API integration (VirusTotal, PhishTank, Cisco Talos)

- Dynamic whitelist/blacklist management

- Exception handling for legitimate edge cases (CDNs, corporate shorteners)

Challenges, Limitations and Future Perspectives

Technical Challenges

- Sophisticated evasion techniques using legitimate infrastructure (CDNs, cloud platforms)
- High false positive rates in complex enterprise environments
- Real-time processing requirements vs. comprehensive analysis depth
- Dynamic threshold adjustment based on threat landscape evolution

Expert-Identified Limitations

- Traditional blacklist approaches insufficient for zero-day attacks
- Need for continuous rule adaptation based on incident learning
- Balance between security enforcement and user experience
- Integration challenges with existing enterprise security stacks

Technological Adequacy Assessment

The chosen symbolic reasoning approach proves particularly suitable for:

- **Explainable decisions**: Critical for enterprise security operations requiring audit trails
- **Rule adaptability**: Enables rapid response to new attack patterns identified by expert
- **Performance scalability:** Meets real-time processing requirements for URL analysis

Future Development Directions

Connect the system to threat intelligence services such as VirusTotal and PhishTank to keep updated information about dangerous sites and automate the handling of legitimate exceptions (like trusted internal domains). Finally, the system should be available on an easy-to-use web platform so that rules for detecting phishing can be checked, tested, and updated in a centralized way. These steps help make the solution more practical, adaptable, and effective for daily protection against phishing threats.

Bibliography

Academic Sources:

1. Patil, D. R., Dhage, S. N.: Framework for anti-phishing: URL feature extraction and classification. In: Proceedings of International Conference on Information Security and Privacy, pp. 105–112 (2019)
2. Hranický, R., Horák, A., Polišenský, J., Ondryáš, O., Jeřábek, K., Ryšavý, O.: Spotting the Hook: Leveraging Domain Data for Advanced Phishing Detection. In: 2024 20th International Conference on Network and Service Management (CNSM), Oct. 2024, pp. 1–7. doi:10.23919/CNSM62983.2024.10814617.
3. Aravindhan, K., Sethumadhavan, M., Krishnan, P.: Enhanced phishing detection through comprehensive URL analysis and machine learning techniques. Journal of Information Security and Applications 45, 123–135 (2019)
4. Bayer, J., Maroofi, S., Hureau, O., Duda, A., Korczynski, M.: Building a Resilient Domain Whitelist to Enhance Phishing Blacklist Accuracy. In: APWG eCrime, pp. 1–14 (2023). doi:10.1109/eCrime61234.2023.10485549.
5. Ishida, Y., Hanada, M., Waseda, A., Kim, M. W.: Analysis of DNS Graph of Phishing Websites Using Digital Certificates. In: ICACT, pp. 174–179 (2023). doi:10.23919/ICACT56868.2023.10079566.
6. Nishitha, U., Kumar, S., Reddy, P.: Phishing Detection Using Machine Learning Techniques. In: 2023 IEEE International Conference on Computing and Communications Technologies, pp. 234–241. IEEE, New York (2023)
7. Hranický, R., Horák, A., Polišenský, J., Jeřábek, K., Ryšavý, O.: Unmasking the Phishermen: Phishing Domain Detection with Machine Learning and Multi-Source Intelligence. In: IEEE NOMS, pp. 1–5 (2024). doi:10.1109/NOMS59830.2024.10575573.
8. A. O'Mara, I. Alsmadi, A. Aleroud, D. Alharthi: Phishing Detection Based on Webpage Content: Static and Dynamic Analysis. In: 3rd Intelligent Cybersecurity Conference (ICSC), pp. 39–45 (2023).
9. Rose, S., Johnson, M., Chen, L.: Real-time phishing detection through browser extension monitoring of DOM and network activity. Computers & Security 118, 102–115 (2022)
10. K. Bajaj, K. Pattabiraman, A. Mesbah: LED: Tool for Synthesizing Web Element Locators. In: 30th IEEE/ACM Int. Conf. on Automated Software Engineering (ASE), pp. 848–851 (2015).
11. Omolara, A.E., Jantan, A., Abiodun, O.I., et al.: DaE2: Unmasking malicious URLs by leveraging diverse and efficient ensemble machine learning. Computers & Security 147, 103–118 (2025)
12. Abad, S., Rahman, M., Thompson, J.: Classification of Malicious URLs Using Machine Learning. PMC Journal of Cybersecurity Research 8(2), 45–62 (2023)
13. I. Alsmadi, A. O'Mara, A. AlEroud: Generative Adversarial Analysis of Phishing Attacks on Static and Dynamic Content of Webpages. In: IEEE Intl Conf on Parallel & Distributed Processing with Applications (ISPA/BDCloud/SocialCom/SustainCom), pp. 1657–1662 (2021).
14. Çelik, L., Yilmaz, E., Hassan, A.: Enhancing Phishing Detection in Financial Systems through Natural Language Processing. arXiv preprint arXiv:2507.04426 (2025)
15. Hranický, R., Bujlow, T., Čejka, T.: Multi-source threat intelligence integration for robust phishing detection. Journal of Network Security 28(3), 189–205 (2024)
16. Alsabah, M.; Nabeel, M.; Boshmaf, Y.; Choo, E. Content-Agnostic Detection of Phishing Domains using Certificate Transparency and Passive DNS. In: Proceedings of RAID 2022, Association for Computing Machinery (ACM), 2022. doi: 10.1145/3545948.3545958.
17. Abyaa, Y.; Hureau, O.; Duda, A.; Korczynski, M. LogoTrust: Leveraging BIMI to Build a Validated Dataset of Brands, Domain Names, and Logos. In: IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 132–137 (2025). doi: 10.1109/EuroSPW67616.2025.00021.

Industry Reports:

- ENISA Threat Landscape 2024/2025
- World Economic Forum Global Cybersecurity Outlook 2025

Technical Documentation:

- OpenPhish, PhishTank, VirusTotal, Cisco Talos API documentation
- MXToolbox analysis tools documentation
- Puppeter: A JavaScript web scrapping tool
- Dns Twister: Anti-phishing domain name search engine and DNS monitoring service

Specific Terminology

Core Concepts: Phishing, Domain Generation Algorithm (DGA), Redirect Chains, Homographs, Document Object Model (DOM), Lexical Analysis, Data Exfiltration, Domain name Server (DNS)

Technical Implementation: Prolog, Drools, Fuzzy Logic, Scrapping, Reputation API, Blacklist/Whitelist, Adaptive Scoring

Operational Terms: Supply Chain Security, Identity Management, Incident Response, Risk Communication, Threat Intelligence, False Positive Management

Summary

This preliminary report demonstrates the team's systematic approach to knowledge acquisition, combining rigorous academic research with practical expertise from enterprise cybersecurity operations. The expert collaboration with David Marques has been instrumental in validating theoretical approaches against real-world attack scenarios and operational constraints.

The chosen technical architecture effectively addresses the specific characteristics of phishing detection through its two-phase analysis model, while the symbolic reasoning implementation ensures both performance and explainability requirements. The continuous feedback loop with industry expertise positions the project to deliver practical value beyond academic achievement, contributing to the evolution of enterprise cybersecurity defense capabilities.

The knowledge representation framework developed captures not only technical detection mechanisms but also the operational context necessary for effective deployment in enterprise environments, reflecting the team's synthesis capability and deep understanding of the cybersecurity challenge domain.