Blacklists and whitelists in the framework of a domain reputation system

Cristian-Alexandru GHEORGHIȚĂ[1,2], Dragoş SMADA[1], Adrian-Victor VEVERA[1], Mihail DUMITRACHE[1,3,4], Ionuț-Eugen SANDU[1], Carmen-Ionela ROTUNĂ[1,2]

[1] National Institute for Research & Development in Informatics – ICI Bucharest, Romania

[2] University Politehnica of Bucharest, Romania

[3] University of Bucharest, Faculty of Letters, Romania

[4] Academy of Romanian Scientists, Bucharest, Romania

alexandru.gheorghita@ici.ro, dragos.smada@ici.ro, victor.vevera@ici.ro, mihail.dumitrache@ici.ro, ionut.sandu@ici.ro, carmen.rotuna@ici.ro

Abstract: The current research approaches the implementation of a real-time domains monitoring platform to determine their reputation, which involves data collection, machine learning analysis, and creating a real-time monitoring interface. Collaboration between various parties and continuous platform improvement are essential to meet evolving user needs and cybersecurity demands. This integrated approach strives to ensure a safer online environment by monitoring and enhancing the reputation of internet domains. The purpose of a domain name reputation system is to assess the trustworthiness of internet domain names. It relies on key components like whitelists, blacklists, registry data and domain rating tools for evaluation and security. Requirements for developing a domain name reputation system highlights the essential prerequisites for an effective domain name reputation system. A significant emphasis is placed on the identification and use of blacklists and whitelists to distinguish trusted from untrusted domains.

Keywords: Domain, Reputation System, Blacklists, Whitelists, Domain Security.

Listele negre și listele albe în cadrul unui sistem de reputație a domeniilor

Rezumat: Cercetarea actuală abordează implementarea unei platforme de monitorizare a domeniilor în timp real pentru a determina reputația, care implică colectarea de date, analiza de învățare automată și crearea unei interfețe de monitorizare în timp real. Colaborarea între diverse părți și îmbunătățirea continuă a platformei sunt esențiale pentru a răspunde nevoilor în evoluție ale utilizatorilor și cerințelor de securitate cibernetică. Această abordare integrată urmărește să asigure un mediu online mai sigur prin monitorizarea și îmbunătățirea reputației domeniilor de internet. Scopul unui sistem de reputație a numelor de domenii este de a evalua gradul de încredere al numelor de domenii de internet. Acesta se bazează pe componente cheie, cum ar fi listele albe, listele negre, datele de registru și instrumentele de evaluare a domeniilor pentru evaluare și securitate. Cerințele pentru dezvoltarea unui sistem de reputație a numelor de domenii evidențiază premisele esențiale pentru un sistem eficient de reputație a numelor de domenii. Se pune un accent semnificativ pe identificarea și utilizarea listelor negre și a listelor albe pentru a distinge domeniile de încredere de cele care nu sunt de încredere.

Cuvinte cheie: Domeniu, Sistem de reputație, Liste negre, Liste albe, Securitatea domeniului.

1. Introduction

In the last decade, the number of cyber-attacks has increased exponentially, especially the number of distributed denial-of-service (DDoS), phishing and ransomware attacks. To prevent such attacks or to minimize their impact, it is necessary for the Registry to implement a series of security measures at the physical, organizational, network and application levels. Proactive detection of security incidents is necessary as it allows the discovery of malicious activities through the use of monitoring tools. Monitoring can be internal, through your own systems, or external, by contracting specialized services that signal intrusions before the affected entities become aware of the problem.

Proactive security involves using certain measures to prevent an attack or identify an attack as it occurs. For a TLD (Top Level Domain) operator, the integrity of the Registry is critical to its operations. Depending on its operating model, a TLD may provide services to its partners (Registrars) and/or domain name users (Registrants). The integrity of the Registry depends on the ability of both the Registrar and the Registrant to protect their system access credentials. Like other

https://doi.org/10.33436/v33i4y202303

service providers that allow users access to online systems, TLD operators apply enhanced security mechanisms, such as two-factor authentication and access based on IP addresses declared by partners (whitelisting).

In the event of a cyber-attack affecting systems or information, the Registry must notify clients of the detected intrusion and its consequences. A potential threat to Registries is the compromise of Registrars' or Registrants' accounts, resulting in attackers infiltrating the domain management system. Attackers tend to target high-traffic domain names and redirect users to clone sites used to collect personal information or deploy and spread malware.

Other attack vectors to the security of Domain Registries are phishing and malware distribution via Doppelgänger domains. Phishing is a popular social engineering attack method, consisting of stealing user data such as login credentials, personal data and banking information. Doppelgänger sites are generally used to distribute malware and run phishing schemes, usually by impersonating financial institutions or government agencies, to collect valuable personal information that can be used to steal identities and rob bank accounts. These types of attack involve new domain registrations that mimic existing domains, usually with heavy traffic, with similar spellings or easy-to-mistake permutations.

The current research paper proposes a system for determining the reputation of domain names which uses Registry data, whitelists, blacklists and other domain related data sources.

Chapter 2 delves into the concept of internet domain reputation, defined as the evaluation and assignment of trust to a web domain based on quality, integrity, and performance. Factors involved in establishing a domain's reputation include data collection, analysis, monitoring of domain activity, security, authenticity, content quality, user feedback, performance, and availability. The reputation value is determined through rating algorithms, aiding users in decision-making regarding domain interaction.

Chapter 3 outlines the essential requirements for the development of an effective Domain Name Reputation System tailored for a Domain Registry. Key prerequisites include accurate data collection on domain names, encompassing registration dates, associated IP addresses, and relevant historical usage data. Real-time monitoring capabilities are crucial for swiftly detecting malicious or suspicious activity. The integration of machine learning algorithms for data analysis to identify patterns indicative of fraudulent or abusive use is emphasized.

Chapter 4 focuses on the identification and description of blacklists and whitelists which can support the development of an Internet domain name reputation system. Blacklists consist of items like domain names, email addresses, or IP addresses considered harmful, primarily used to block or filter undesirable content while whitelists encompass trusted and safe domain names.

2. State of the art

Internet domain reputation refers to evaluating and assigning a value or trust to a web domain based on its quality, integrity and performance. It is a process of collecting and analysing relevant information about the domain to determine how it is perceived and rated by users and the online community (Banciu et al., 2019).

Establishing a domain's reputation involves several factors, including:

- Collect relevant data - the first step in establishing a domain's reputation is to collect relevant data about it. This data may include domain owner information, domain history, registration data, associated IPs, page content and more;
- Data analysis - after the data has been collected, it is analysed to identify the key characteristics of the domain and to assess its performance in various relevant areas such as security, authenticity and content. This analysis can be carried out through the use of automated analysis technologies, machine learning algorithms or through the intervention of experts;
- Domain activity monitoring - is important to detect any significant changes in its behaviour or activity. This may be done by monitoring website traffic, detecting cyber-attacks or spam, analysing user feedback or other relevant methods;
- Security - a domain with a good reputation is associated with a high level of security. It

should be protected against cyber threats such as hacking or phishing and provide a safe experience for users (Vevera et al., 2021);

- Authenticity - the reputation of a domain is influenced by its authenticity. Users trust domains that are legitimate and follow relevant rules and regulations. Correctly registering the domain and providing the appropriate information about its owner helps increase trust in its reputation;
- Trusted content - domains with quality and trusted content are more likely to have a good reputation. This may include providing accurate, relevant and useful information, avoiding spam or questionable content, and complying with relevant standards and ethics;
- User feedback - user ratings can play an important role in establishing a domain's reputation. Users' opinions and experiences of the domain, expressed through reviews, comments or ratings, can influence its perception and reputation;
- Performance and availability - the reputation of a domain can be affected by the performance and availability of the website. Users expect a fast-loading time, a well-functioning website and a smooth domain interaction experience;
- Rating algorithms and scores - based on the data collected and the analysis performed, rating algorithms can be used to assign a score or ranking to the domain based on its level of reputation. These algorithms can be trained to identify relevant patterns and trends and provide an objective reputation assessment;
- Assign a reputation value - based on data analysis and monitoring of domain activity, a reputation value can be assigned to the domain. This can be a numerical value or a qualitative value such as "good", "mediocre" or "risky". This value can be used to guide users in making decisions about interacting with the domain.

For organizations, cybercriminals taking control of an owned domain is a major concern, as they can exploit it for malicious purposes, such as hosting malware and obtaining user credentials. In these cases, the domains' reputation is compromised (OWASP, 2020).

ICANN (Internet Corporation for Assigned Names and Numbers) is responsible for the oversight and management of top-level domains (TLDs) in the global domain name system. It doesn't register individual domain names but plays a crucial role in establishing and regulating TLDs. Some of the well-known TLDs under ICANN's purview include .com, .org, .net, .gov, .edu, and many more.

ICANN's role involves accrediting domain registrars that can register domain names under these TLDs. It also establishes policies and guidelines to maintain the security, stability and integrity of the domain name system.

Individuals, businesses and organizations can register domain names through accredited registrars for the TLDs overseen by ICANN. These domain names can then be used for websites, email addresses, and various online applications. ICANN's role is essential in ensuring that domain names are allocated and managed in a fair and consistent manner across the internet (ICANN, 2005) (ICANN, 2016).

If not properly secured, the Domain Name System (DNS) can be vulnerable to exploitation by malicious individuals. Malware developers are aware of the importance of DNS accessibility and are actively looking for ways to disrupt the operation of DNS and adjacent servers (Scalzo, 2017).

A crucial element in domain security is the implementation of a blacklist system capable of identifying and filtering malicious websites, even those that are not yet known. Fukushima et al., propose the development of a blacklist system with the ability to analyse the characteristics of malicious websites using their domain information, such as autonomous system (AS), IP address block, IP address, domain and registrar, and propose a blacklist which combines IP address blocks and low-reputation registrars, which are frequently used by attackers (Fukushima et al., 2011).

However, creating and maintaining these blacklists is difficult, leading to errors and omissions. To solve this, Lison & Mavroeidis (2017) proposed a machine learning model based on deep neural architecture that automatically detects whether domain names and IP addresses are malicious. Because it was trained on an extensive passive DNS database, the model achieves a high detection rate of 95% (Lison & Mavroeidis, 2017).

Using blacklists only to determine the reputation of a domain name proves to be extremely ineffective in identifying both known and newly generated malicious URLs. Moreover, it relies on human input and proves to be a time-consuming process, especially in real-time environments (Vinayakumar et al., 2018).

The complexity of identifying the classification of malicious domains can be addressed by using contemporary learning and adaptive methodologies. Unsupervised intelligent domain classification can be achieved by using metaheuristics-based search algorithms such as Cuckoo Search. Object-oriented engineering can be used to implement the proposed model, with future extensions a possibility (Sarkar et al., 2013).

Most existing domain name reputation systems do not provide a dynamic, real-time, ML-based monitoring solution that can be used by domain ccTLD (Country Code Top Level Domain) or their registrars to scan domain names every day, thereby enabling the detection of compromised domain names.

The system proposed in this research paper will use domain history information combined with information from the Registry database along with a series of external tools such as blacklists to determine a reputation score for a .ro domain.

The solution will be proactive and allow, through machine learning techniques, to detect a malicious domain pattern. The action will allow any suspicious domain to be selected and quarantined for its rehabilitation or removal.

Currently there are a number of systems designed to determine the reputation of domain names. Exposure is a system designed to detect malicious domains by analysing DNS data. It employs a large-scale passive DNS analysis method to identify domains associated with malicious activity. (Bilge et al., 2011). Notos is a DNS reputation system that identifies malicious domains based on their unique characteristics, using passive DNS query data and analysing network and zone characteristics of domains. The data is used to develop models of trusted domains and malicious domains (Antonakakis et al., 2010). Kopis uses a passive approach to observe DNS traffic at higher levels of the DNS hierarchy. By analysing worldwide DNS query resolution patterns, it can pinpoint malicious domains. Unlike previous DNS reputation systems such as Notos and Exposure, which depend on monitoring traffic from local recursive DNS servers, Kopis provides a fresh perspective and incorporates new traffic features that leverage the global visibility gained by observing network traffic (Antonakakis et al., 2011).

Compared to the existing domain name reputation systems such as Notos, Kopis, Exposure or MaldomDetector, the proposed system architecture enables the dynamic evaluation of a domain's reputation in real time.

3. Requirements for developing a domain name reputation system

A domain name reputation system for a Domain Registry needs to meet certain requirements to be effective. Therefore, developing and implementing an architecture for a real-time domain monitoring platform to determine their reputation involves collecting and analysing data, using machine learning algorithms, and creating an interface for real-time domain monitoring. This process requires an integrated approach and close collaboration between the various parties involved, as well as continuous improvement of the platform to meet the ever-changing needs of users and the cyber environment.

The key requirements are:

- to accurately collect data about domain names, including the date they were registered, the IP addresses associated with them, and any historical data related to their use;
- to monitor domain names in real time to detect any malicious or suspicious activity as soon as possible;
- to use machine learning algorithms to analyse data and identify patterns of behaviour that may indicate fraudulent or abusive use;
- to use collaborative filtering techniques to identify relationships between domains, such as those that belong to the same owner or are used for similar purposes;

- to integrate with threat intelligence feeds to provide additional context for domain name behaviour;
- to have a user-friendly interface that will allow domain registrars and law enforcement agencies to easily access and interpret the data;
- to protect the privacy of applicants registering domains while providing sufficient data for analysis;
- to provide timely responses to domain registrars and law enforcement agencies when suspicious activity is detected so that appropriate action can be taken to prevent abuse;
- to handle large volumes of data and will be scalable to accommodate a growing number of domain names and users;
- to incorporate strong security measures to protect against unauthorized access and data breaches (Holland, 2019).

Currently, there are several techniques that can be used to improve the accuracy and effectiveness of a domain name reputation system. These include:

- Real-time monitoring - involves continuously monitoring the behaviour of domains in real-time and updating their reputation scores accordingly. This can help to quickly identify and mitigate new threats as they emerge (Rotună et al., 2022);
- Whitelists and blacklists - are lists of domains known to be either malicious or safe, respectively. Using these lists as a starting point, a domain name reputation system can quickly classify domains as safe or risky;
- IP Reputation - can help identify malicious activities that may be associated with a specific IP address.

In general, an effective domain name reputation system must incorporate a combination of techniques that, along with advanced machine learning algorithms and real-time monitoring capabilities, work together to identify potentially malicious domains. By doing so, such a system can provide accurate and reliable information on the reputation of a given domain name, helping to protect users from potential threats and ensuring a resilient and secure online environment.

4. Blacklists and whitelists

In the context of the Internet domain name reputation system, the terms "blacklist" and "whitelist" refer to lists containing IP addresses or domain names considered unsafe or safe, respectively. These lists are used to filter and control e-mail traffic and access to online resources in order to prevent spam and other unwanted activities.

An example of such a list is the Domain Name System Blacklist (DNSBL). This is a service that allows email servers to check whether an email server's IP address is listed as being associated with sending spam or other unwanted activities.

On the other hand, a "whitelist" is a list containing IP addresses or domain names considered safe and reliable. These lists are used to allow access to online resources or to receive emails from certain sources. Whitelists are often used in combination with blacklists to provide better control over email traffic and access to online resources.

To develop and implement an architecture for a real-time domain monitoring platform to determine their reputation, malicious domain detection techniques used in Notos, Exposure and Kopis solutions can be used.

Given the specific aspects to be monitored, such as traffic and DNS infrastructure, as well as blacklists and whitelists, technologies and methods such as data analysis and machine learning (ML) algorithms can be used for classifications and predictions (Dumitrache et al., 2023).

5. Blacklists

Blacklists are lists of items, typically domain names, email addresses, IP addresses, or keywords, that are considered undesirable or harmful in a specific context. These lists are used for various purposes, primarily to block, filter, or otherwise take action against items included in the list.

There are several domain blacklists checking tools available to check if a domain has been blacklisted. Some of the most commonly used such tools are:

Spamhaus

It is one of the leading providers of blacklists for identifying and blocking spam, malware and other malicious activities. The Spamhaus Block List (SBL) includes a comprehensive database of known spam sources, while the Exploits Block List (XBL) focuses on identifying the IP addresses of infected machines participating in botnet activities.

Spamhaus also offers other lists such as the Policy Block List (PBL) for identifying dynamic IP addresses and the Domain Block List (DBL) for blocking domains associated with spam or malware.

Spamhaus DBL is a list of disreputable domain names. This is published in DNSBL (Domain Name System Block List) format of domains. These domain reputations are calculated based on several factors and are kept in a database which in turn feeds the DBL area.

The DBL zone can be queried publicly, with real-time responses, just like other Spamhaus DNSBL zones. Those using e-mail server software capable of scanning e-mail headers and message bodies for URIs can use DBL to identify, classify, or reject e-mail containing DBL-listed domains. Security professionals and researchers also use domain reputation in their work.

The DNSBL (also known as the "Blocklist") is a collection of records queried in real time by the email servers in order to determine the source of the emails that are received. A DNSBL's role, like Spamhaus' SBL/XBL/PBL advisory system, is to offer a perspective to anyone who requests it, as to whether a given IP address complies with Spamhaus's own inbound mail acceptance policy (Figure 1).

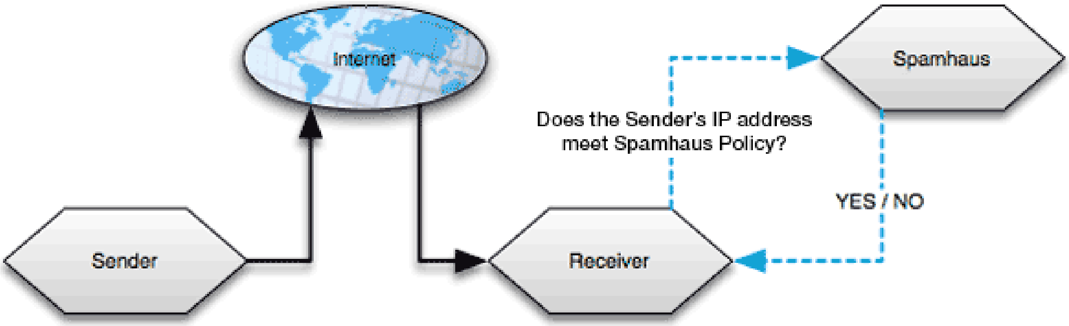

Figure 1. Spamhaus architecture (Spamhaus, 2023)

A team of dedicated specialists maintains the reputation database by using diverse data from multiple sources to create and maintain a large set of rules that control an automated system that constantly analyses a large portion of the world's e-mail flow.

Most DBL listings are done automatically, although, if necessary, Spamhaus researchers add or remove listings manually. DBL data is subject to exchange with other Spamhaus systems, which may result in new DBL listings or the inclusion of IP addresses in other Spamhaus zones.

DBL works as both a domain URI block list and a right-hand side block list (RHSBL). It is effective for filtering mail during the SMTP session for all header checks - rDNS, HELO, MAIL FROM, From, Reply-To and Message-ID domains - as well as URLs in messages. DBL is widely used by many parties, both large and small, for checks involving domain reputation in email and other applications.

DBL is managed as a near-zero false positive list that can be safely used by production email systems to reject emails that are flagged by it. In addition, it can be used for labelling, scoring or folding.

DBL includes domains that are used in mass spam, including phishing, fraud, or sending or hosting malware or viruses, as well as other domains with a bad reputation due to many heuristics. DBL offers a disposal system that is monitored, automated and self-service. DBLs also expire automatically when domains are no longer in use or when reputation variables change (The Spamhaus Project, 2023).

Spam URI Real-Time Blocklist (SURBL)

The Spam URI Real-Time Blocklist (SURBL) has the purpose to prevent domains and URLs detected in spam messages from being used. It operates by collecting URLs from email content and comparing them to a database of spam-related sites. SURBL detects dangerous websites in real time and can assist prevent users from visiting hazardous websites linked in spam emails.

SURBL focuses in delivering reputation data in near-real-time streams. This information can be used to protect mail flow, SMS messages, Internet access (safe browsing, DNS firewalls) and a variety of other activities. The most well-known data set is Multi, which is freely available. These are not restricted to e-mail operations. SURBL now provides cryptographic addresses, phone numbers, an email source list, and streams of freshly active domains.

Many of the streams can be accessed via RPZs (Response Policy Zones). RPZ statistics can be used to trigger actions in DNS firewalls, such as blocking access to potentially harmful websites (Surbl, 2023).

Multi

Domains of malicious or abusive websites are listed in this dataset. Regardless of the sender's IP addresses, it can be used to filter or classify unsolicited messages based on links in the message content. Filtering based on known malicious sender IPs is best complemented by this.

Fresh

Fresh is a list of domains that have recently been added to TLD zone file delegations. This includes a timestamp of when we last detected changes and therefore gives an indication of recently delegated domains. Since newer domains are more likely to be abusive, this can be used as one of several factors to indicate domain reputation.

HashBL

HashBL is a list of cryptographic hashes for various Internet abuse related items. This data set is constantly growing, with new categories being added.

URI Query (UriQ)

URI Query (UriQ) is an API for checking full URIs, especially for legitimate but hacked or abusive sites that cannot be enumerated at the host (domain or IP) level in the main dataset (Surbl, 2023).

URI Blacklist (URIBL)

URI Blacklist (URIBL) is a comprehensive blacklist service that targets domains associated with spam, malware and other malicious activity. URIBL maintains several lists, including the ABUSE, MULTI and RHS lists, each of which has different areas of interest. The ABUSE list identifies domains associated with abuse, while the MULTI list detects spam, malware and other malicious activity. The RHS list focuses on identifying domains that do not meet the appropriate domain registration standards. URIBL public lists are:

- black.uribl.com - contains domain names owned and used by spammers, but not limited to those appearing in URIs found in unsolicited bulk and/or commercial emails (UBE/ UCE). This list aims for zero false-positive responses. This area is rebuilt frequently as new data is added;
- grey.uribl.com - contains domains found in UBE/UCE. It may include ESPs that allow customers to import their recipient lists and may not have any control over subscription methods. This list can and probably will cause false positives depending on how you define UBE/UCE. This area is rebuilt several times a day as needed;
- red.uribl.com - contains domains that are actively appearing in the mail stream, are not listed on URIBL black and are either: monitored or very young (domain age via whois) or use whois privacy features to- and protect identity. This list is automated in nature;
- white.uribl.com - contains legal domain names that we do not want to appear on other

URIBL lists. This list is fairly static with only a few changes per day. This area is rebuilt as needed;

- multi.uribl.com - checks if a domain is on one of the lists. This area is rebuilt if any of the above areas are rebuilt except the white area.

Barracuda Reputation Block List (BRBL)

BRBL is a blacklist widely used by Barracuda Networks security products. This focuses on identifying IP addresses associated with spam activities. BRBL ensures real-time spam blocking by checking the reputation of the sender IP address in its database

BRBL employs a standard DNSBL (Domain Name System Block List) implementation that may be applied to decrease spam on a variety of popular email services, including Microsoft Exchange, IBM Lotus Domino, sendmail, Postfix and qmail, along with many antispam solutions, including Barracuda Spam & Virus Firewall.

DNSBL systems list IP addresses, which are frequently those observed by the list operator as sending spam or hosting spammers. By including DNSBL entries in the settings of a mail server, that mail server can choose to reject connection attempts coming from the listed IP addresses or use the list to enforce a suitable filtering policy.

DNSBLs are also known as RBLs (Realtime Blackhole Lists) or simply BLs (Block/Black lists). DNSBL is not preventing anyone from sending email; it only prevents it from being delivered to the recipient according to the sender's instructions. DNSBLs are defensive instruments which cannot be used to cause offensive damage such as denial-of-service (DDoS) attacks.

BRBL provides a list of IP addresses that send spam. The Barracuda Reputation system uses automated collection methods to add and delete IP addresses from the BRBL.

The automated spam capture, collection and classification system automatically adds IP addresses to the list when spam is detected. When an email is received, the connection is automatically analysed to determine whether the connecting machine is either an open proxy or a node of a spamming botnet. In either case, the IP address is immediately added to BRBL's block list.

Most IP addresses are listed as a result of sending spam or viruses directly to the Barracuda Reputation System detectors. The Barracuda Reputation System detects spam by using honeypots, special addresses created to receive only spam that do not belong to any real user, and by analyzing captive spyware protocol activity. In addition, BRBL leverages some data derived from the Barracuda Blocklist (BBL) that is delivered to the Barracuda Spam & Virus firewalls. A single infected computer sending spam through a network using NAT can cause e-mail to be blocked from the entire local network (BarracudaCentral.org, 2023).

PhishTank

PhishTank is a community-driven blacklist specifically designed to detect and block phishing sites. It relies on user suggestions and verification to identify and blacklist phishing domains. PhishTank provides an API that allows developers and security systems to integrate their services for proactive protection against phishing attacks.

It works as a collaborative platform where users can submit and verify suspected phishing URLs or websites. PhishTank then consolidates this information into a comprehensive database that is made available to the public.

Users can register suspicious phishing URLs or websites with PhishTank for analysis. Submissions can be made through the PhishTank website or through its API. Each record includes details about the suspected phishing site, such as the URL and any comments or additional evidence. PhishTank maintains a comprehensive database of verified phishing URLs and websites. This database is regularly updated and made available to the public through the website and API. Users can query the database to check if a particular URL is flagged as a phishing site.

PhishTank's API allows integration with various security products, web browser, and email clients. These integrations enable proactive protection against phishing attacks by comparing URLs visited or emails received against the PhishTank database. If a match is found, appropriate warnings or blocks can be implemented to prevent users from accessing phishing content.

Malwaredomains.com is a website that provides information about malicious domains and also provides a list of domains that are known to host malware (Malware domains, 2009).

Fukushima Blacklist

Fukushima et al., (2011) propose the development of a blacklist system with the ability to analyse the characteristics of malicious websites using their domain information, such as Autonomous System (AS), IP address block, IP address, domain and Registrar, and propose a blacklist that combines IP address blocks and low-reputation Registrars, which are frequently used by attackers (Fukushima et al., 2011).

6. Whitelists

It is important to note that the effectiveness and appropriateness of whitelists may vary depending on the context and specific requirements of an organization or system. Maintaining and managing them often requires continuous updates and careful analysis to ensure that trusted domains are accurately identified and added.

In many circumstances, a hybrid solution that combines whitelisting and blacklisting can provide a more robust and comprehensive security architecture, allowing enterprises to strike a balance between trusted access and protection against malicious or hacked domains. These lists are of three types: application-specific, reputation-based and domain-based.

Application-specific whitelists are tailored to specific applications or platforms and allow only pre-approved domains or entities to interact with the system. Examples include:

- Email whitelists: Email servers or clients may have whitelists that allow email to be sent only from trusted domains or senders;
- Website whitelists: Browsers or security software may have whitelists to allow access to certain websites considered safe.

Reputation-based whitelists are based on the reputation and trustworthiness of domains, often determined by various evaluation methods. Examples include:

- Domain Reputation Services (DRS). Services such as Cisco's SenderBase or Proofpoint's Domain Reputation offer reputation-based whitelists that identify trusted domains based on their behaviour, sending practices and other factors;
- Trusted vendors or partners: Organizations can maintain their own internal whitelists, consisting of domains associated with trusted vendors, partners, or known entities with which they have established relationships.

Application-specific whitelists are tailored to specific applications or platforms and allow only pre-approved domains or entities to interact with the system. Examples include:

- Email whitelists. Email servers or clients may have whitelists that allow email to be sent only from trusted domains or senders;
- Website whitelists. Browsers or security software may have whitelists to allow access to certain websites considered safe.

7. Conclusions

The purpose of developing a Domain Name Reputation System is to establish a system that assesses the trustworthiness and credibility of domain names on the internet. This system will rely on key components such as whitelists, blacklists, registry data and domain rating tools. These components will serve as the foundation for evaluating domain reputation and security.

The requirements for developing a domain name reputation system identified in this research paper are the prerequisites for an effective domain reputation system. This step will ensure that the system is built to meet specific criteria and objectives for assessing and enhancing the reputation of domain names. In this preparatory phase, a significant focus is placed on the utilization of blacklists and whitelists as essential tools to categorize and differentiate trusted and untrusted domains.

Subsequent research efforts will concentrate on the architecture of the system, outlining how various components will interact and function together. Additionally, the research will delve into the selection of technologies for system implementation. This crucial decision will determine the tools and frameworks used to build the system, ensuring its efficiency and effectiveness.

The final stage of development will focus on creating the technical solution, where the system will be built and configured according to the established architecture and technological choices. This solution will be designed to serve as a reliable means of assessing domain name reputation, contributing to a safer and more secure online environment.

Developing and implementing an architecture for a real-time domain monitoring platform to determine their reputation involves collecting and analysing data, using machine learning algorithms, and creating an interface for real-time domain monitoring. This process requires an integrated approach and close collaboration between the various parties involved, as well as continuous improvement of the platform to meet the ever-changing needs of users and the cyber environment (Rotună et al., 2023).

Acknowledgements

This research work was supported by a grant of the Romanian Ministry of Innovation and Research, Nucleu Programme, project code: PN 2338 02 01, project name: "Architecture – platform of an intelligent system for monitoring Internet domains by developing a dynamic reputation establishment system (TLDRep)" and by a grant of Romanian Academy of Sciences through the competition "AOSR-TEAMS-II" EDITION 2023-2024 - "Digital transformation in the sciences", project name: "Digital transformation tools for eGovernment through the use of .ro domains".

REFERENCES

Antonakakis, M., Perdisci, R., Dagon, D., Lee, W. & Feamster, N. (2010) Building a dynamic reputation system for DNS. *The 19th USENIX Security Symposium, August 11-13, 2010, Washington, DC, USA*. pp.273–290.

Antonakakis, M., Perdisci, R., Lee, W., Vasiloglou, N. & Dagon, D. (2011) Detecting malware domains at the upperDNS hierarchy. *Proceedings of the 20th USENIX SecuritySymposium, August 10-12, 2011, San Francisco, CA*. pp.1-16.

Banciu, D., Petre, I. & Dumitrache, M. (2019) Electronic system for assessing and analysing digital competences in the context of Knowledge Society. In: *Proceedings of the 11th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2019, June 27-29, 2019, Pitesti, Romania*. IEEE. doi: 10.1109/ECAI46879.2019.9042151.

BarracudaCentral.org, (2023) *Technical Insight for Security Pros.* https://www.barracudacentral.org/ [Accessed 17th September 2023]

Bilge, L., Kirda, E., Kruegel, C. & Balduzzi, M. (2011) *EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis*. https://sites.cs.ucsb.edu/~chris/research/doc/ndss11_exposure.pdf. [Accessed 17th September 2023]

Dumitrache, M., Sandu, I.E., Udroiu, A.M. & Gheorghiță, C.A., (2023) Considerații teoretice privind stabilirea reputației unui domeniu Internet (Theoretical considerations about establishing the Internet domain reputation). *Revista Română de Informatică și Automatică (Romanian Journal of Information Technology and Automatic Control).* 33(1), pp. 81-92.doi:10.33436/v33i1y202307.

Fukushima, Y., Hori, Y. & Sakurai, K. (2011). *Proactive Blacklisting for Malicious Web Sites by Reputation Evaluation Based on Domain and IP Address Registration*. doi:10.1109/TrustCom.2011.46.

Holland B. (2019) *TLD Operator Perspective on the Changing Cyber Security Landscape*. https://www.cigionline.org/articles/tld-operator-perspective-changing-cyber-security-landscape/ [Accessed 17th September 2023]

ICANN (2005) *GAC Principles and Guidelines for the Delegation and Administration of Country Code Top Level Domains - Role of Government or Public Authority*. https://gac.icann.org/contentMigrated/gac-principles-and-guidelines-for-the-delegation-and-administration-of-country-code-top-level-domains-role-of-government-or-public-authority [Accessed 5th October 2023]

ICANN (2016) *Uniform Domain-Name Dispute-Resolution Policy - ICANN*. https://www.icann.org/resources/pages/help/dndr/udrp-en [Accessed 5th October 2023]

Lison, P. & Mavroeidis, V. (2017). Neural reputation models learned from passive DNS data. In: *2017 IEEE International Conference on Big Data (Big Data), 11-14 December, 2017, Boston, MA, USA*. pp. 3662-3671. doi:10.1109/ BigData. 2017.8258361.

Malware Domains. Malware Domain Block List. https://riskanalytics.com/.

OWASP (2020) *Open Source Foundation for Application Security*. https://www.owasp.org [Accessed 7th October 2023].

Rotună, C.I., Gheorghiță, C.A., Sandu, I.E., Dumitrache, M., Udroiu, A.M. & Smada, D. (2023) A Generic Architecture for Building a Domain Name Reputation System. *Studies in Informatics and Control*. 32(2), 39-49. doi: 10.24846/v32i2y202304.

Rotună, C.I., Dumitrache, M. & Sandu, I.E. (2022) Evaluarea algoritmilor de învățare automată pentru monitorizarea automata (Assessment of Machine Learning algorithms for automated monitoring). *Revista Română de Informatică şi Automatică (Romanian Journal of Information Technology and Automatic Control)*. 32(3), 73-84. doi:10.33436/v32i3y202206.

Sarkar, M., Banerjee, S. & Hassanien, A. (2013) Searching DNS for malicious domain registration: identification through hybrid cuckoo search metaphor and object-oriented implementation. *International Journal of Reasoning-based Intelligent Systems.* 5(4), 280 - 289. doi:10.1504/IJRIS.2013.058773.

Scalzo F. (2017*) DNS-based threats: DNS reflection and amplification attacks, Verisign*. https://blog.verisign.com/security/dns-based-threats-dns-reflection-amplification-attacks/ [Accessed 7th October 2023]

Surbl.org. (2023) SURBL. https://www.surbl.org/ [Accessed 9th October 2023]

The Spamhaus Project (2023) Understanding DNSBL Filtering. https://www.spamhaus.org/whitepapers/dnsbl_function/ [Accessed 15th October 2023]

Vevera, A.V., Georgescu, A. & Cîrnu, C.E. (2021) Opportunities for Cybersecurity Research in the New European Context. *Romanian Cyber Security Journal.* 3(1), 79-88.

Vinayakumar, R., Soman, K. & Prabaharan, P. (2018) Evaluating deep learning approaches to characterize and classify malicious URL's. *Journal of Intelligent & Fuzzy Systems*. 34(3), 1333-1343. doi:10.3233/JIFS-169429.

Cristian-Alexandru GHEORGHIȚĂ graduated the Faculty of Informatics, University Bucharest in 2013. He is a Ph.D. student at Faculty of Computer Science and Automatics, Politehnica University of Bucharest. Currently he works as Researcher at ICI Bucharest. His main areas of interest are Cyber Security, Big Data, Cloud Computing, Cloud-Native, DevOps. He is involved in research projects specific to the Information Society. His research was published in journal articles and proceedings of conferences.

Cristian-Alexandru GHEORGHIȚĂ a absolvit Facultatea de Informatică din cadrul Universității din București, în anul 2013. Este doctorand la Facultatea de Automatică și Calculatoare, Universitatea Politehnica din București. În prezent lucrează ca cercetător în cadrul Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București. Principalele sale domenii de interes sunt: Cyber Security, Big Data, Cloud Computing, Cloud-Native, DevOps. Este implicat în proiecte de cercetare specifice societății informaționale. Cercetările sale au fost publicate în articolele revistelor de specialitate și în lucrările conferințelor științifice.

Dragoş SMADA graduated the Faculty of Electronics, Telecommunications and Information Technology at the Polytechnic University of Bucharest. He has a Master's Degree in Information and Documents Management at the University of Bucharest. He is currently a scientific researcher at ICI Bucharest. His main areas of interest are Big Data, Internet of Things, software engineering, information security, software architecture. He participated in both national and international research projects in the IT&C. He published as author and co-author of journal articles and scientific presentations at conferences.

Dragoş SMADA a absolvit Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Universitatea Politehnica din București. Deține o diplomă de „Master în Managementul Informațiilor și Documentelor” obținută în cadrul Universității București. În prezent este cercetător științific la ICI București. Principalele sale domenii de interes sunt: Big Data, Internetul obiectelor, inginerie software, securitate informatică, arhitectură software. A fost implicat în proiecte de cercetare naționale și internaționale din domeniul TIC și a publicat rezultatele cercetărilor în articole de specialitate din reviste și conferințe IT.

Adrian-Victor VEVERA is the General Director, Senior Researcher II and member of the Scientific Council of the National Institute for Research and Development in Informatics. Mr. Vevera holds a Ph.D. in military and information sciences, being both a lawyer and a nuclear physics engineer. He has extensive experience in the field of national security, fulfilling various positions, over time, in numerous managerial and counselling positions in different state-run organisations. He has published numerous articles and papers on national and international security issues, energy security, cybercrime, critical infrastructure protection, and has been the coordinator of numerous projects of national interest.

Adrian-Victor VEVERA este Director General, cercetător științific gradul II și membru în Consiliul Științific al Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București. Doctor în științe militare și informații, fiind la bază atât jurist cât și inginer specializat în fizică nucleară, deține o vastă experiență în ceea ce înseamnă securitatea națională, ocupând, de-a lungul timpului, numeroase poziții manageriale și de consiliere în diverse organisme ale statului. A publicat numeroase articole și lucrări pe teme de securitate națională și internațională, securitate energetică, criminalitate informatică, protecția infrastructurilor critice și a fost coordonatorul a numeroase proiecte de interes național.

Mihail DUMITRACHE graduated from Politehnica University of Bucharest, the Faculty of Electrical Engineering with the specialization „Computer Assisted Engineering”, he is an engineer and holds a Ph.D. degree in Electrical Engineering. In between, he obtained two Master's Degrees, one in Electrical Engineering at Politehnica University of Bucharest and one in Electronic Public Administration, at the University of Bucharest. His professional career started at the National Institute for Research and Development in Informatics – ICI Bucharest in 2002 as a computer programmer. Currently, he is a Scientific Researcher II and Head of the .ro Domain Administration Department (RoTLD) – ICI Bucharest and also a Lecturer at the University of Bucharest. He is the author and co-author of several scientific studies and articles.

Mihail DUMITRACHE este absolvent al Facultății de Electrotehnică, Universitatea Politehnica din București, specializarea „Inginerie Asistată de Calculator”, inginer și doctor în Inginerie Electrică. Deține două diplome de master în specializarea „Inginerie Electrică”, Universitatea Politehnica din București și în specializarea „Administrație Publică Electronică”, Universitatea din București. Și-a început activitatea profesională în cadrul Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București în anul 2002, ca programator. În prezent este cercetător științific gradul II, Șef la Departamentul „Administrare domenii RoTLD” – ICI București și Lector Universitar la Universitatea din București. Este autor și coautor al unor studii și articole de specialitate.

Ionuț-Eugen SANDU graduated university with a BS in Computer and Systems' Science (2006) and obtained a Master's Degree in Electronic Public Administration in 2007. In 2010, he became Scientific Researcher within the .ro Domain Administration Department (RoTLD) of the National Institute for Research and Development in Informatics – ICI Bucharest, and since 2015 is Scientific Researcher grade III and Head of the Technical Division of RoTLD, with responsibilities in systems' administration, development of new services, development of communication infrastructures. He is also in charge with maintaining a close relationship with partners. Currently, he is Technical Director of National Institute for Research & Development in Informatics – ICI Bucharest.

Ionuț-Eugen SANDU este licențiat în Știința Sistemelor și a Calculatoarelor (2006), obține master în Administrație Publică Electronică în anul 2007. Din anul 2010 devine cercetător științific la Departamentul de Administrare Domenii .ro din cadrul Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București, iar începând cu anul 2015 devine Șef Serviciu Tehnic RoTLD și cercetător științific gradul III în cadrul aceluiași Institut. Domeniile sale principale de interes sunt: administrare sisteme, dezvoltare de noi servicii, dezvoltare a infrastructurii de comunicații, precum și relația cu partenerii. În prezent este Director Tehnic al Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București.

Carmen-Ionela ROTUNĂ is a Ph.D. student at the Politehnica University of Bucharest, in the field of Systems Engineering and graduated with a Master's Degree from the Faculty of Mathematics and Computer Science of the University of Bucharest. Currently, she is a Scientific Researcher at the National Institute for Research and Development in Informatics – ICI Bucharest, where she conducts research activities in eGovernment, eServices, Cloud, Big Data and AI, also being the author and co-author of various articles published in specialized journals and conference proceedings recognized nationally and internationally, of project deliverables and books. She was also a team member in national and European projects in the IT&C area: SPOCS - Simple Procedures Online for Cross-border Services (CIP-ICT PSP), eSENS - Electronic Simple European Networked Services (CIP ICT), Cloud for Europe C4E (FP7), TOOP - The "Once-Only" Principle Project (H2020), where she was the national coordinator for the WP2: Architecture work package and national coordinator for WP3: Project Piloting. She is currently WP3 leader in EUROCC – National Competence Centres in the framework of EuroHPC project.

Carmen-Ionela ROTUNĂ este Doctorand la Universitatea Politehnica din București, domeniul „Ingineria Sistemelor" și a absolvit programul de master la Facultatea de Matematică și Informatică din cadrul Universității din București. În prezent este cercetător științific în cadrul Institutului Național de Cercetare-Dezvoltare în Informatică – ICI București, unde desfășoară activități de cercetare în domeniile: eGovernment, eServices, Cloud, Big Data și AI, fiind autor și coautor al unor articole publicate în reviste de specialitate și volume de conferință recunoscute la nivel național și internațional, precum livrabile de proiect și cărți. Totodată a participat la proiecte naționale și europene din aria IT&C: SPOCS – Simple Procedures Online for Cross-border Services (CIP-ICTPSP), eSENS - Electronic Simple European Networked Services (CIP ICT), Cloud for Europe C4E (FP7), TOOP - The "Once-Only" Principle Project (H2020), unde a avut rolul de coordonator la nivel național pentru pachetul de lucru WP2: Arhitectură și WP3: Pilotare. În prezent este implicată în proiectul EUROCC – National Competence Centres in the framework of EuroHPC cu rol de coordonator în cadrul pachetului de lucru WP3.