International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1353-1357

Knowledge-Based Approach to Detect Potentially Risky Websites

Meiyalakan K 1st, Naveenkumar C 2nd, Sridharan P 3rd , Srikanth R 4th, Tamilarasan E 5th

1st Assistant Professor, 2nd, 3rd, 4th, 5th UG Scholar (B.E), Department of Computer Science and Engineering, Mahendra Institute of Technology, Mahendhirapuri.

ABSTRACT

Nowadays, fraudulent and malicious websites are emerging as a harmful and very common problem on the Internet. It causes huge money losses and irreparable damage for both companies and particulars. To face this situation, governments have approved multiple law projects. This way, the legality on the Internet is being enforced and sanctions to those offenders who develop illegal or malicious activities are being imposed. However, governments still need a way to simplify the classification of websites into risky or non-risky, since most of this work is manual. This paper presents the DOmains Classifier based on RIsky Websites (DOCRIW) framework to detect domains that contain possible fraud or malicious content. It is based on two main components. The first component is a previously built knowledge base containing information from risky websites. The second one complements the system with a binary classifier able to label a website (as risky or not) considering just its domain. The system makes use of web information sources and includes host-based variables. It also applies similarity measures, supervised learning algorithms and optimization methods to enhance its performance. The presented work is experimental, rendering promising outcomes.

Date of Submission: 06-06-2022 Date of acceptance: 21-06-2022

I. INTRODUCTION

Public bodies that prosecute fraudulent and malicious web sites dedicate a significant amount of time and resources to detect scam and malware on the Internet [2]. Most of this work is usually manual, which translates into hard and inefficient efforts. For this reason, it has become essential to develop systems able to automate the classification of websites into potentially risky or non-risky according to the features of these sites. In this context, a risky website is one with malicious, unsafe or fraudulent content with dangerous intentions against their visitors [1].

The study by Spanish Information Security Observa tory (OSI) captures the magnitude of the risky websites prob lems in Spain [3]. Among the main results and conclusions of the study, it should be noted that a 53.1% of Spanish Internet users claimed to have been victims of an attempt (not neces sarily consummated) of fraud in the last three months. The invitations to visit some suspicious website (34.4%). In the analyzed period, 95.2% of Spanish Internet users share that they have not suffered economic damage in the last three months as a result of a fraud via Internet, while 4.8% have suffered losses. Besides, the empirical analysis of the equipment shows that 39.8% of the computers host some type of Trojan, 6.8% host banking Trojans (malicious code snippets intended to intercept electronic banking credentials of specific entities) and a 5.8% suffer a rogue-ware infection (or fake antivirus). Furthermore, 81.8% of Internet users who have suffered an incident of this type have not changed their habits surfing websites, compared to 5% who have abandoned this activity and 13.2% who have reduced the use of Internet. The Span ish Observatory of Computer Crimes (OEDI) have reported 110, 613 cyber-crimes in Spain in 2018, 74% of them have been fraud [4].

In this paper, two main contributions have been made. The first one consists of a novel Knowledge-Based System (KBS)

II. LITERATURE SURVEY

In service-oriented computing, services are built as an assembly of pre-existing, independently developed services. Hence, predicting their dependability is important to appropriately drive the selection and assembly of services, to get some required dependability level. We present an approach to the dependability prediction of such services, exploiting ideas from the Software Architecture- and component-based approaches to software design. In the Service-Oriented Computing (SOC) paradigm, an application is built as composition of components and services (including both basic services, e.g. computing, storage, communication, and

"advanced" services that incorporate some complex business logic) provided by several independent providers. A basic requirement for SOC is that support should be given to automatically discover and select the services to be assembled. The "Malicious Web Service s" and "Grid computing" frameworks represent standardization efforts in this area. An important issue for applications built in this way is how to assess, as much as possible automatically to remain compliant with the SOC requirements, their quality, for instance their performance or dependability characteristics. In this paper, we focus on dependability aspects, and provide an approach that lends itself to automatization to predict the service reliability, defined as a measure of its ability to successfully carry out its own task. The main goal of this approach is to define a compositional way for predicting the service reliability that reflects the underlying structure of a service realized within the SOC framework. To this purpose, we exploit ideas taken from Software Architecture and Component-based approaches to software design.

According to the Software Architecture approach, an application is seen as consisting of a set of components that offer and require services, connected through suitable connectors. In particular, special emphasis is given to the connector concept, that embodies all the issues concerning the connection between offered and required services. Hence, a connector can also represent a complex architectural element carrying out tasks that are not limited to the mere transmission of some information, but could also include middleware services such as security and fault-tolerance. Using different types of connectors to assemble the same set of services we can easily experiment the impact on the overall system QoS of different ways of architecting the service assembly.

The problem of composing services to deliver integrated business solutions has been widely studied in the last years. Besides addressing functional requirements, services compositions should also provide agreed service levels. Our goal is to support model-based analysis of service compositions, with a focus on the assessment of non-functional quality attributes, namely performance and reliability. We propose a model driven approach, which automatically transforms a design model of service composition into an analysis model, which then feeds a probabilistic model checker for quality prediction. To bring this approach to fruition, we developed a prototype tool called ATOP, and we demonstrate its use on a simple case study. Service-Oriented Architectures (SOAs) provide a new paradigm for the creation of business applications. This paradigm enforces decentralized developments and distributed systems compositions: new added-value services may be created by composing independently developed services. Malicious Web Service s are an increasingly important and practical instance of SOAs, supported by standards and by specific technology. Typically, service compositions can be orchestrated by using a workflow language, like the Business Process Execution Language (BPEL).

An SOAs can benefit from the Model Driven Development (MDD) paradigm. In essence, this means that models are built to support software engineers in reasoning on the software architecture. As a satisfactory solution is built at the model level, transformation steps (possibly automated) derive the final, platform-specific implementation. In this specific case, model-level analysis should support the early QoS assessment of a service composition. The composition may be assessed at design time, before a concrete binding from the workflow to the externally invoked services is established. It is requested, however, that a specification of the external services in terms of their functional and non-functional attributes is available. The actual binding to concrete services may then be established dynamically at run time, provided that the selected concrete services fulfil their specification. This may be enforced by a suitable QoS-driven binding mechanism. The model may be used also to support evolution of the software architecture. It can also be useful to devise suitable reconfiguration strategies for the dynamic contexts where the application will be deployed. Once the application is running, model-based .

III. EXISTING METHOD

In existing, they have focused on predicting reliability of various factors involved in building enterprise application, nonetheless, considered reliability of remote Malicious Web Service as constants. For remote Malicious Web Service s the vender will provide probabilistic details about the flow of executing user requests. Here we use CSPN (Cryptanalysis of Substitution-Permutation Networks) model technique This model deals more with design time problems and does not reflect the impact of problems that occur at runtime. They have studied in detail transactional dependency among different type of Malicious Web Service s. CSPN is Substitution-permutation network (SPN). In cryptography, an SP-network, or SPN, is a series of linked mathematical operations used in block cipher algorithms such as AES.

IV. PROPOSED SYSTEM

A novel method for Quos metrification based on Hidden Markov Models (HMM), which further suggests an optimal path for the execution of user requests. The users can weigh their options directly and individually, for themselves. We use Hidden Markov Models for Building a directed graph among hidden states of component Malicious Web Service s used in composition. Analyzing the current status of each vertex of

directed graph i.e., underlying hidden states. Predicting hidden states behaviour in terms of response time during nth time interval t. Finally, selecting optimal Malicious Web Service s used in composition based on hidden states behaviour. A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (*hidden*) states. An HMM can be presented as the simplest dynamic Bayesian network. The mathematics behind the HMM were developed by L. E. Baum and co-workers. It is closely related to an earlier work on the optimal nonlinear filtering problem by Ruslan L. Stratonovich, who was the first to describe the forward-backward procedure.

In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a HMM, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly.

Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

4.3 The agency role of voice assistants

In their recommender agent function, VAs attempt to predict which items a goal consumer will like primarily based totally on expressed possibilities or implicit behaviors (Shen, 2014). This shape of recommender machine might also additionally update conventional decisionmaking while clients experience time constraints or apprehend the referrer as a specially informed source (Olshavsky & Granbois, 1979). End-customers usually examine a digital agent on its cappotential to customize hints that fulfill their needs. Consumers undertake algorithmic recommender structures if they may be believed to in shape their interests (Abdollahpouri et al., 2019). Higher accuracy of hints from a platform interprets into now no longer handiest an growth in patron pleasure however additionally their ordinary agree with withinside the technology (Li & Karahanna, 2015). In this context, advice effects might also additionally correspond to patron possibilities greater carefully than in the event that they had selected independently (André et al., 2018). Due to their primary function in a complicated commercial enterprise network (Snehota & Hakansson, 1995), VAs do now no longer don't forget customers because the handiest stakeholders profiting from the advice outcome. The strategic desires of the retailer, merchant, advertiser, and voice assistant itself, might also additionally fluctuate from the ones of end-customers. Thus, the consumer isn't the only awareness of a advice in nearly each transaction at the VA. For instance, a VA would possibly suggest a personal label over a patron emblem following the retailer's goal to unexpectedly develop its stocks in a selected product category. Thus, the goals of numerous events want to coexist (Abdollahpouri et al., 2019). The last purpose of advice personalization is the automation of the shopping for experience. Throughout the gathering of considerable volumes of private and behavioral information, VAs can push customers to automate repurchase, for instance, via "subscribe & save" promotional activities, an increasing number of famous at the e-trade websites. According to André et al. (2018), this strength of legal professional toward VAs is going on the price of higher-order mental tactics along with feelings and ethical judgments. In the context of buy automation, clients would possibly have aspirational possibilities that fluctuate from the possibilities counseled via way of means of their beyond behavior. These meta possibilities, additionally known as possibilities over possibilities (Jeffrey, 1974), are obvious withinside the case of an environmentally conscious man or woman who desires to use much less bottled water however is often reminded to shop for plastic bottles. The inherent anxiety among the actual-self and the ideal-self represents a boundary for the ones clients who comply with VAs' hints to automate repurchases.

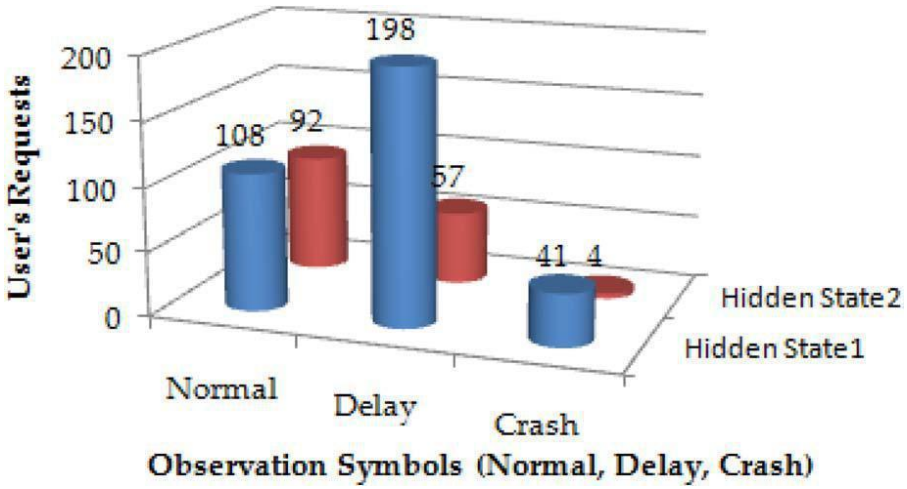

Fig. 5.1 Hidden states observation patterns in terms of number of requests

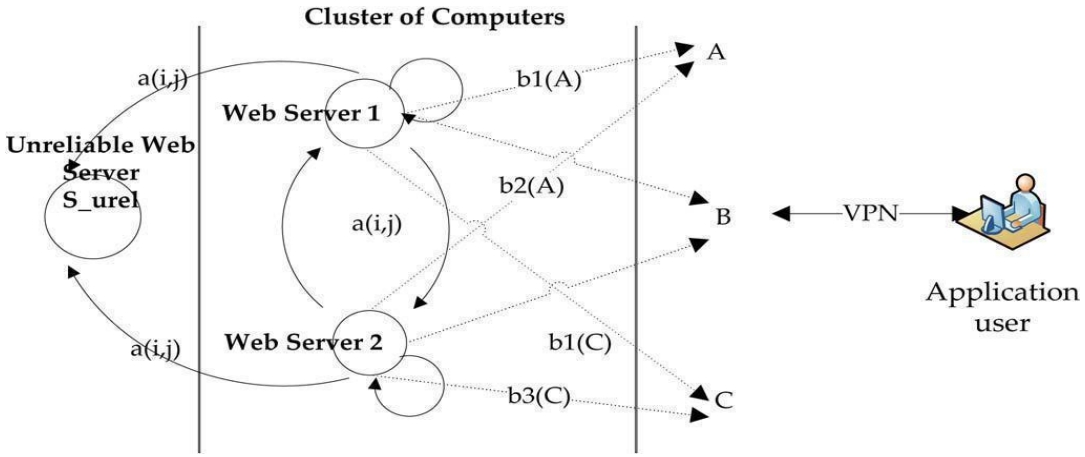

Fig. 5.2. Hidden states and corresponding observation symbol

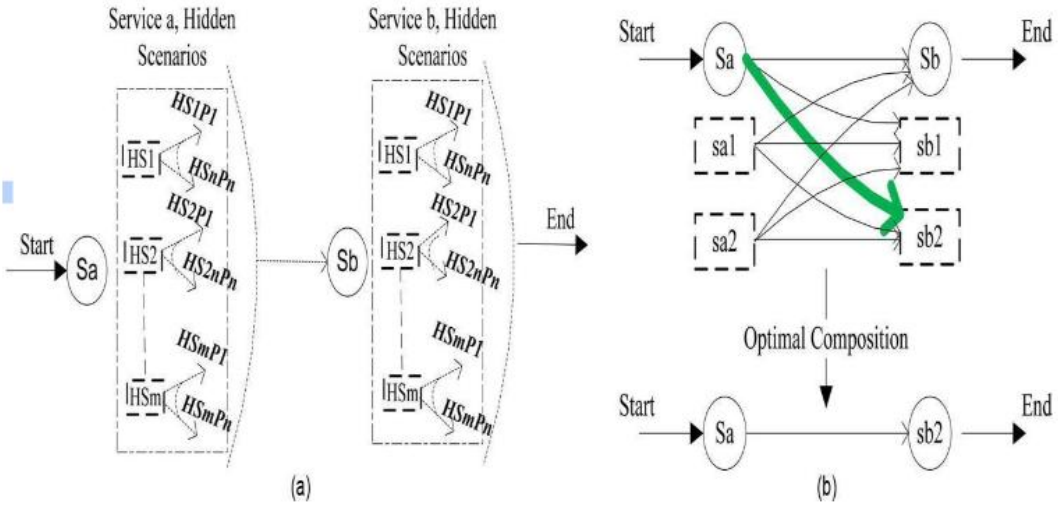

Fig. 5.3. Example of a simple case in composite Malicious Web Service (CWS) with various hidden scenarios. (a) Basic composition (simple case) (b) Optimal composition of Malicious Web Service s

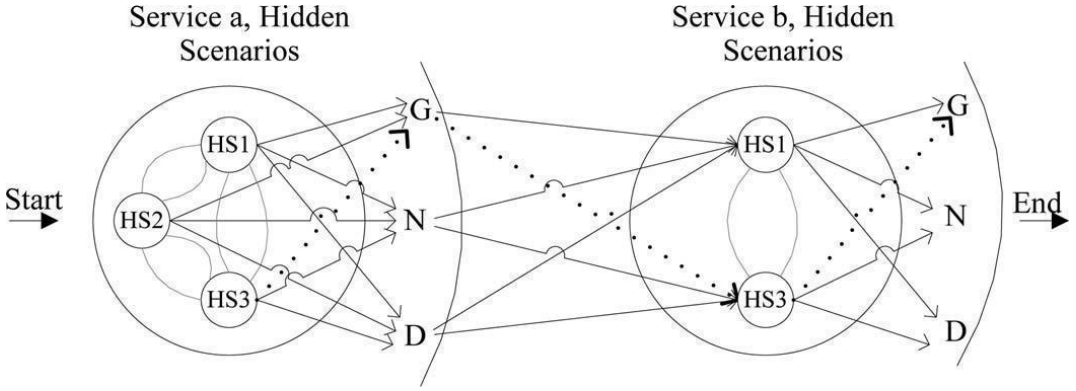

Fig. 5.4. Directed graph of composition in terms of hidden states

VII. CONCLUSION

The System probabilistic model for predicting response time of Malicious Web Service and then selected an optimal Malicious Web Service at runtime from the list of functionally equivalent Malicious Web Service s. To know the probabilistic insight of WSs we have used HMM. In our model we have assumed that WS is deployed on a cluster of web servers and sometime the delay or crash during WS invocation is because the bad node in sever clustering responds to users‟ requests. With the help of HMM we have predicted the probabilistic behavior of these web servers and then selected the WS based on their probabilistic value. Usage of the machine learning algorithm provides the better result and accuracy. Any machine learning algorithm such as **ANN , NB, SVM** can be used to predict the accurate result of malicious risky websites from the given input testing dataset

REFERENCES

[1]. Ahmed W, „„Architecture-Based Reliability Prediction for Service- Oriented Computing,‟‟ in Architecting Dependable Systems III. Berlin, Germany: Springer- 2005, pp. 279-299.
[2]. Boumhamhi k, Ghezzi C, Mirandola R, and. Tamburrelli G, „„Quality Prediction of Service Compositions through Probabilistic Model Checking,‟‟ IEEE Quality Software- Architecture, Models Architecture, 2008, pp. 119-134.
[3]. Cristescu M, „„Composing Malicious Web Service s: A QoS View,‟‟ IEEE Internet Computer volume 8, no. 6, pp. 80-90, November 2004.
[4]. Leilei C, Zhao W, and Bouguettaya A, „„QoS Analysis for Malicious Web Service Compositions Based on Probabilistic QoS,‟‟ in Service-Oriented Computing. Berlin, Germany: Springer-Verlag, 2011, pp. 47-6Bouguettand R.L. Michael, „„Collaborative Reliability Prediction of Service-Oriented Systems,‟‟ in Proceeding 32nd ACM/IEEE volume 1, pp. 35-44.
[5]. Maheswari S, Macedo M, G, and Lima L, „„An Approach for Estimating Execution Time Probability Distributions of Component-Based Real-Time Systems,‟‟ Journal Universal Computer Science, volume 15, no. 11, pp. 2142-2165, 2009.
[6]. Menasce D.A and Ciovica L, „„Estimation of the Reliability of Distributed Applications,‟‟ Inf. Economic, volume 14, no. 4, pp. 19-29, 2010.
[7]. Perrone R,X. Xu, „„Reliability Prediction and Sensitivity Analysis of WS Composition,‟‟ in Petri Net: Theory and Applications, V. Kordic, Ed. Rijeka, Croatia: Intech, 2008, pp. 459-470.
[8]. Salfner F and Jarir Z, „„A Flexible Approach to Compose Malicious Web Service s in Dynamic Environment,‟‟ International Journal Digital Social, volume 1, no. 2, pp. 157-163, 2010.
[9]. Stefano G, Yue Z, and Kwei-Jay L, „„Efficient Algorithms for Malicious Web Service s Selection with End-to-End QoS Constraints,‟‟ ACM Transaction Web, volume 1, no. 1, p. 6, May 2007.
[10]. Tao Y, Z, and M.R. Lyu, „„WSPred: A Time-Aware Personalized QoS Prediction Framework for Malicious Web Service s,‟‟ in Processing IEEE 22nd ISSRE, 2011, pp. 210-219.
[11]. Zaki H, „„QoS Based Efficient Malicious Web Service Selection,‟‟ Eur. Journal Science volume 66, pp. 428-440, 2011.
[12]. Zheng H, Wang Q, W. Xu, and Zhang L, „„Evaluating the Survivability of SOA Systems Based on HMM,‟‟ in Process IEEE Malicious Web Service , 2010, pp. 673-675.
[13]. Zhong D, „„Predicting Failures with Hidden Markov Models,‟‟ in Process 2005, pp. 41-46.
[14]. Zibin Z, A, and B, „„Malicious Web Service s Reputation Assessment Using a
[15]. Hidden Markov Model,‟‟ in IEEE Service-Oriented Computer, 2009, pp. 576-591.