Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 9
September 2024

A Novel Approach to Website Security: Enhanced Packet Sniffing with Artificial Neural Network

Deepa Singh Sisodiya[1*]· Dr. Priyank Jain[2]· *Dr.* Ritu Tiwari[3]*Dr. ShriyaSahu*[4]

[1]Department of Computer Science and Engineering, IIIT-Pune, India
[2]Department of Computer Science and Engineering, IIIT-Pune, India
[3]Department of Computer Science and Engineering, IIIT-Pune, India
Corresponding author's E-mail: Deepa Singh Sisodiya[1]

Abstract Artificial intelligence (AI) technology is a broad, cutting-edge discipline that is evolving, as well as an essential study path in the domains of science, computer engineering, and technology. It has the potential to transform the traditional working style of artificial hosting by promoting the computer server to do continuous intelligent analysis, allowing technicians to work less and achieve greater data processing efficiency. This is the key reason for proposing an AI based solution for website security in this paper.

This research focuses on identifying the existing challenges on website security based on the user feedback, literature review and analysis of the security posture of a few websites. The user feedback was collected through the survey of more than 150 users. Based on the identified challenges in website security an Artificial Intelligence (AI) based framework proposed to mitigate the possible risks of those identified challenges. AI is capable of handling massive amounts of data, clarifying data correlations, and cope with extremely complicated situations. Artificial intelligence technology can imitate the human brain's cognitive mode and perform several activities such as data deep analysis and learning, effective data processing, real-time human computer interaction, and so on.

Keywords Packet Sniffing, Website Security, Artificial Neural Networks, Artificial Intelligence, Online Security

1 Introduction

The fifth-most significant risk in the year 2020, cyberattacks are anticipated to increase in both severity and frequency in public & private sectors [5]. This means many new users are possibly subjected to existing security risks if it is not addressed properly. In the last few years, there has been drastic growth in newly discovered vulnerabilities and website hacking [2] which can also lead to security breaches. Therefore, a framework is required to provide a systematic approach to protect confidential data from being hacked and misused. Benefit of the mass, the victim of database breach would be the users of that website, thus they have the right to be notified, in a noticeable way.

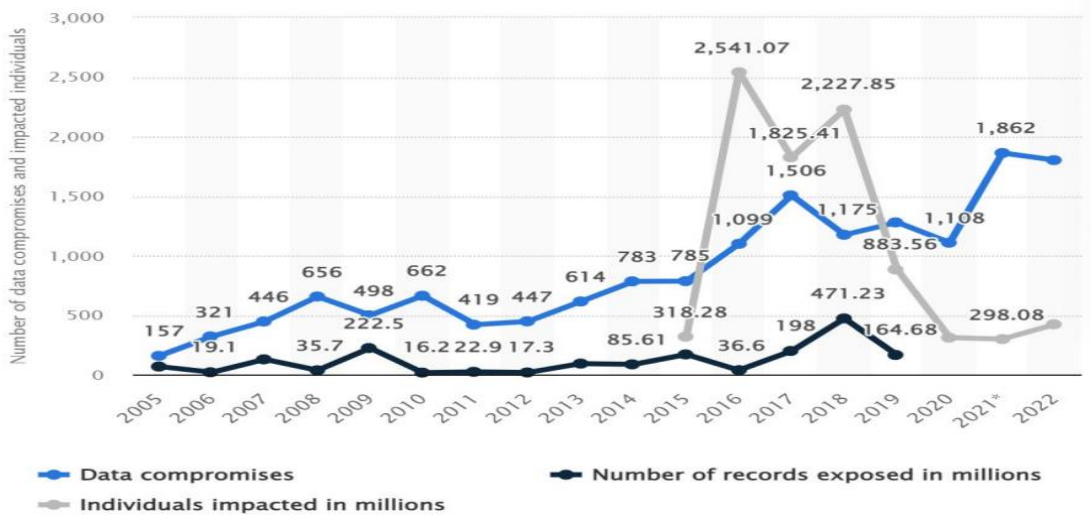


Fig.1 The price of data loss and the number of Americans affected between 2005 and 2022 [10]

The organizational attack surface is enormous and, as a result of digitalization, it is continuing to expand and evolve rapidly as digitalization allows remote access to organizational data. Due to huge web-based threat attack vector, Analyzing and enhancing website security is not a human-scale challenge. In reply to this unprecedented challenge, Artificial Intelligence (AI)-based technologies for website security / cyber defense have grown to assist information security teams in efficiently and efficaciously decreasing threat of data breach and upgrading their security system.

Eight out of the ten users believe that the existing communication approach between the web service provider and end-user is not efficient. AI and machine learning (ML) are becoming key technologies in IS (Information Security) because millions of events can be rapidly analyzed by them and a wide range of threats can be identified, from malware exploiting zero-day vulnerabilities to risky behavior identification that could proceed to a phishing attack or malicious code download. Relying on the past, these systems learn over time, to determine new types of risk and present threats. Behavior histories provide profiles for the users, assets, and networks, allowing AI to recall and acknowledge deviations from standard practices.

Problem statement: In today's digital landscape, cybersecurity threats targeting websites have become increasingly sophisticated, posing significant risks to sensitive data and user privacy. Traditional methods of packet sniffing for website security often lack the agility and accuracy needed to detect and mitigate these evolving threats effectively. Therefore, there is a pressing need to develop advanced techniques that leverage artificial neural networks (ANNs) to enhance packet sniffing capabilities for website security.

This **research aims** to address this gap by investigating the feasibility and effectiveness of employing ANNs as a novel approach to packet sniffing, with the goal of improving the detection and prevention of malicious activities, vulnerabilities, and unauthorized access attempts targeting websites.

The **objective** of this research can besummaries as follows:

1. Investigate the feasibility and efficacy of employing artificial neural networks (ANNs) for packet sniffing in the context of website security.
2. Evaluate the performance and accuracy of ANN-based packet sniffing compared to traditional methods in detecting and mitigating website security threats.
3. Explore the potential of ANN algorithms to enhance real-time monitoring and anomaly detection capabilities for website security.
4. Assess the practical implications and deployment considerations of integrating ANN-based packet sniffing into existing website security frameworks for improved threat intelligence and incident response.

This research starts with the collection of relevant data through various ways, starting with the literature review, subsequently followed with the data collection through analysis of the security posture of different websites and survey. In the literature review more than 28 research papers were cited to demonstrate the requirement of research in this topic.

The survey is the major data collecting tool in this study since it enables data gathering from a large number of individuals without actual interaction during the epidemic. Participants can complete the survey at their leisure. Therefore, a survey was conducted with the help of 150 respondents to understand the detailed data for this topic.

Finally, the data was collected through the analysis of the online portal of different institutes to get the in-depth understanding of the existing challenges and the possible area of improvement. The security posture of the online portal was analyzed through penetration testing to discover the hidden security flaws. This approach of data collection also helps to understand the organization’s approach and their understanding for the security of online portals.

At last, the collected data is further analyzed to conceptualize the deliverable in the form of framework. Because the framework is a great tool for minimizing loss and enhancing quality by removing numerous inherent and unintentional

complexities[14]. The suggested framework aids in process maturity, accountability, risk management, and performance monitoring with improved security, among other things.

In this paper, literature review was done with the help of numerous research papers and journals to understand the existing research done in this topic. After this to understand the existing challenges in website security, a total of five public website security was analyzed through the penetration testing (after getting written consent from the owner). Followed with the online survey of various website stakeholders to get in-depth understanding of current challenges in website security and possible area of improvement. Based on the collected data, a framework is proposed later to address the existing challenges and also to provide guidance to the website developers and stakeholders for the key areas to focus to secure a website. The reliability of the proposed framework was ensured with the help of 1-2-1 interview with the survey responders, who were agreed to help further. The research methodology is graphically summarized below.

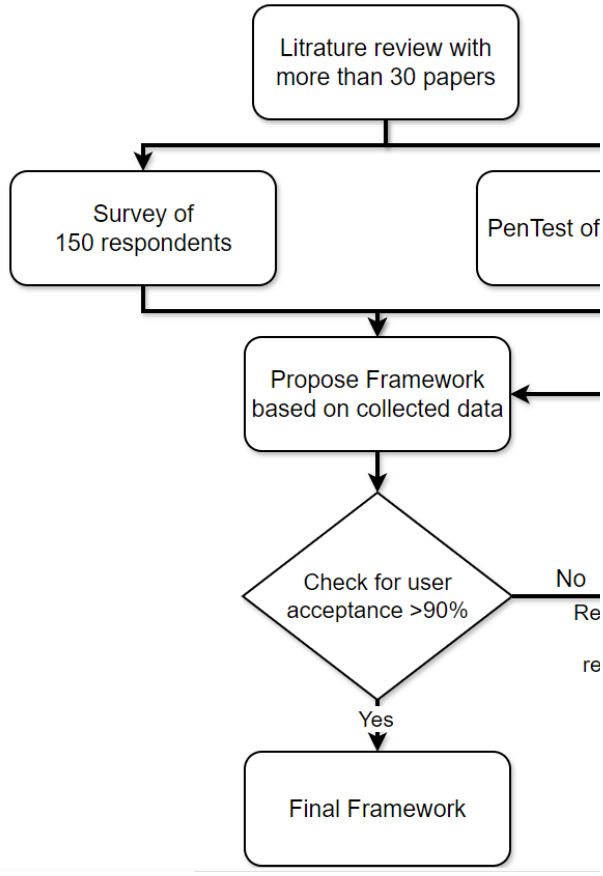


Fig.02 Research methodology

2 Literature review

A literature review's aim is to get an understanding of the current studies and discussions related to a specific topic or field of study and to transfer that knowledge in the form of a report. Performing a literature review will help in expanding awareness in a particular subject.

Phishing is a type of social engineering attack in the digital age where hackers obtain sensitive data or information by using the internet with carelessness or insensitivity. It has been demonstrated that defenses already in place, like anti-phishing applications and analytic methods for spotting phishing activity, work. AI-based schemes are the mainstay of contemporary defenses against phishing efforts. As a result, they advised using meta- learners while developing phishing threat detection systems[1]. For many years, Captcha, which prevents computer programs from hacking websites, has served as the most essential security tool. From the viewpoint of attackers, a wider character set offers more solution space and theoretical security. However, the security of Captchas with big character sets in the actual world has not yet been thoroughly explored [2]. Attackers employ exploit kits (EKs) to transmit malware constantly and discreetly. Their technique also finds malicious webpages involved in EKs, which can boost analysis efficiency even more[3]. Of the most significant responsibilities of privacy-enhancing technology is to provide confidentiality and privacy to Web users' web activity. A local passive eavesdropper can forecast the target user's browsing behavior although if she employs anonymous techniques such as VPNs, IPsec, and Tor. Deep learning advancements have recently enabled attackers to undertake more precise website fingerprinting attacks. Adversarial Website Adaptation (AWA) is a novel security mechanism presented in this study against website fingerprinting attacks utilizing adversarial deep learning algorithms. AWA generates a transformer set for every run, ensuring that each website has its own transformer [5]. Network traffic analysis is rapidly being utilized in a variety of applications to safeguard or expose systems, data, and individuals. One type of passive traffic assessment attack is called webpage fingerprinting those compromises online browsing privacy. It is a collection of algorithms for detecting patterns in a sequence of data packets created when a person navigates across several webpages. Despite the fact that a number of safeguards have been put up to

stop such passive attacks, new assaults have taken advantage of them demonstrating their incompetency and/or impracticality. In this paper, they presented a unique protection mechanism to combat website fingerprinting attacks [6]. Although a number of strategies for detecting fraudulent websites have been suggested, it is becoming increasingly difficult for such methods to produce satisfactory results nowadays. With different Web spam strategies, many rogue websites may still avoid detection. In this article, they described three categories of Website spam tactics used by malevolent web pages. including hidden Frame spam, redirection spam, and content concealing spam. They also demonstrated a novel detection system that collects snapshots of fraudulent webpages and assumes the perspective of users to reject Web spam. As a classification method, the suggested detection approach employs a Convolutional Neural Network (CNN), a type of deep neural network. Two distinct trials were carried out to validate the method's usefulness. These experimental findings show that the suggested strategy performs better and is relevant to a real-world Web context[7]. When seeking a web-based service, consumers frequently fail to configure the website's security settings in accordance with their personal privacy preferences. Users often struggle because they are overwhelmed by the variety of options available, lack expertise in associated technology, or are ignorant of their own privacy preferences. Privacy configuration prediction techniques are especially well-suited to addressing all of these issues[9].

One kind of investigation into traffic exploit is called website fingerprinting (WF) that allows a local passive eavesdropper to infer the victim's activities even when the traffic is secured by a VPN or anonymity scheme such as Tor. A WF attacker may obtain over 98% accuracy on Tor traffic by using a deep-learning classifier. In this study, they presented Mockingbird, a method for producing traces that may withstand adversarial training by wandering randomly in the space of possible traces rather than following more predictable gradients [13]. The website fingerprinting (WF) attack allows a local eavesdropper to determine which website a client is browsing while connected to an encrypted network. snWF determines if a user is visiting a monitored website in a large open-world context with 400,000 websites, with a true positive rate of 98.1% and a false positive rate of 5.7%. They also tested snWF in a more genuine attack scenario known as wide world to see if it can accurately identify websites that even an enemy has never seen before, and they discovered that snWF outperforms state-of-the-art attacks in this new context [14]. Many protection measures against website fingerprinting attacks have been presented in recent years. One of the most adequate protection methods is a walkie-talkie (WT) constructed on top of the Tor network. This study investigates the source of the higher page loading time and proposes a defense strategy known as Tail Period (TT), which solves the issue by restricting the maximum time that a current request may block future requests. Their testing findings showed that the suggested TT defense might greatly cut page loading time while maintaining defensive performance[16].

To compromise user privacy, website fingerprinting attacks involve statistical analysis of network data. This code performs a cache side-channel attack in order to detect additional websites being visited. They explored one such approach and discovered that when employed in the Tor Browser, it lessens the efficacy of the assault and totally removes it[17]. For safeguarding communication channels, the projects employ security. technologies such as digital certificates, Secure Socket Layer (SSL), and HTTPS (Secure HyperText Transport Protocol) [18]. Phishing attacks are among the most common methods of compromising user data and carrying out unwanted activity. It was discovered in the mid-1990s and is still one of the most serious cybercrime methods for stealing user data today. To counter this assault, several algorithms have been developed. To deploy and test the identification of phishing websites, many Machine Learning methods are applied. The issue is that there is no extensive analysis of URLs, domain types, origin, and other essential factors available. Although there isn't a precise method for determining if a website is legitimate or connected to phishing, understanding its structure and pattern will aid in detection [26]. Malicious actors have utilized cyberspace to conduct cyberattacks as information technology (IT) has improved over the past decade. Artificial intelligence is a branch of science that may be used to tackle huge and complicated issues by learning from previous events. This AI capacity can be utilized to create software or a framework that can self-improve in order to safeguard cyberspace [28].Below is the summarized list of research papers followed to identify the existing progress in this topic

Table 1 Literature Review Table

Paper	Paper Title	Contribution	Technique used	Limitations	Remark
[1]	AI meta-learners and extra-trees algorithm for the detection of phishing websites	This study shows four (4) meta-learner models are proposed (AdaBoost-Extra Tree (ABET), Bagging - Extra tree (BET), Rotation Forest - Extra Tree (RoFBET) and LogitBoost-Extra Tree (LBET)) developed using the extra-tree base classifier.	AI meta-learners and extra-trees	Mainly focus limited to performance standard for phishing website detection methods.	
[2]	A security analysis of captchas with large character sets	To tackle these rapidly evolving Captchas, a simple, fast, and successful deep learning solution is shown.	Deep learning technique	Proposed a 3D image-based scheme as the replacement of Captchas.	
[3]	An exploit kits detection approach based on http message graph	They propose an approach that works effectively in both the recent actual scenarios and ground truth datasets. Their technique also finds harmful websites involved in EKs, which can boost examination accuracy even more.	exploit kits (EK)	Use HTTP Message Graph detect exploit kits (EKs) which are used by attackers to distribute malware silently.	
[4]	An unsupervised deep learning model for early network traffic anomaly detection	The D-PACK mechanism, which comprises of a Convolutional Neural Network and an unsupervised deep learning model (e.g., Autoencoder) for auto-profiling traffic patterns and filtering aberrant traffic, was introduced in this paper.	Convolutional Neural Network (CNN) and an unsupervised deep learning model	Lack of extensive evaluation on diverse datasets representing various network environments	
[5]	Awa: Adversarial website adaptation	In this paper the extra-tree basis classifier was taken to create four (4) meta-learner models: Rotation Forest - Extra Tree (RoFBET), AdaBoost-Extra Tree (ABET), LogitBoost-Extra Tree (LBET)) and Bagging - Extra Tree (BET).	AI meta-learners and extra-trees	Flaws in the research design, such as lack of randomization, control groups, or blinding, can undermine the validity of the results.	
[6]	Bimorphing: A bi-directional bursting defense against website	To fight these rapidly evolving Captchas, a simple, efficient, and successful machine learning solution is shown.	mathematical optimization	Lack of validation of proposed defense	

	fingerprinting attacks			algorithm to counteract the website fingerprinting attacks.	
[7]	CNN based malicious website identification by invalidating several web spams	In this paper an approach that works well in both recent practical scenarios and ground truth databases was suggested by them. Their technique also finds harmful websites involved in EKs, which can boost analytical efficiency even more.	Convolutional Neural Network (CNN)	Use a CNN to analyze captured webpage images for malicious websites detection, which will not be effective in many cases.	
[8]	Using Ensemble-Based Feature Selection and Classification Models to Detect DNS Typo-Squatting	The D-PACK mechanism, which comprises of a Convolutional Neural Network (CNN) and an unsupervised deep learning model (e.g., Autoencoder) for auto-profiling traffic patterns and filtering aberrant traffic, was introduced in this paper.	Convolutional Neural Network (CNN) and an unsupervised deep learning model	It proposes a framework to address typo-squatting vulnerability. Paper focus on DNS-based attack	
[9]	Explainable machine learning for predicting default privacy settings	The extra-tree basis classifier was used to create four (4) meta-learner models (LogitBoost-Extra Tree (LBET), Rotation Forest - Extra Tree (RoFBET), Bagging - Extra Tree (BET), and AdaBoost-Extra Tree (ABET)).	AI meta-learners and extra-trees	Focus on user's privacy setting prediction tools	
[10]	Exploiting Internet of Things Protocols for Malicious Data Exfiltration Activities	To fight these rapidly evolving Captchas, a simple, fast, and successful deep learning solution is shown.	Machine learning algorithms	Small sample that may not be representative of the population being studied. This can affect the generalizability of the findings.	
[11]	Lucene-P $^ 2 $2: A Distributed Platform for Privacy-Preserving Text-Based Search	They suggest an approach that runs well in both ground truth datasets and recent actual scenarios. Their technique also finds harmful websites involved in EKs,	exploit kits (EK)	Proposed framework which performs real privacy-preserving tasks.	

		which can boost analysis efficiency even more.			
[12]	Flamingo search algorithm: A new swarm intelligence optimization algorithm	The D-PACK mechanism, comprises of a Convolutional Neural Network (CNN) and an unsupervised deep learning model (e.g., Autoencoder) for auto-profiling traffic patterns and filtering aberrant traffic, was introduced in this paper.	Convolutional Neural Network (CNN) and an unsupervised deep learning model	It involves practical challenges related to integration with existing systems, user acceptance, and maintenance.	
[13]	Mockingbird: Defending against deep-learning-based website fingerprinting attacks with adversarial traces	In this an extra-tree basis classifier was used to create four (4) meta-learner models (AdaBoost-Extra Tree (ABET), Bagging - Extra Tree (BET), Rotation Forest - Extra Tree (RoFBET), and LogitBoost-Extra Tree (LBET)).	AI meta-learners and extra-trees	Proposed solution has implementation challenges that must be addressed before it could be practically deployed.	
[14]	snWF: Website Fingerprinting Attack by Ensembling the Snapshot of Deep Learning	To handle these rapidly evolving Captchas, a simple, fast, and successful deep learning solution is shown.	Deep learning	This research may raise ethical concerns related to privacy, consent, fairness, and transparency. Failure to address these ethical considerations can lead to negative societal impacts or backlash.	
[15]	PPSB: An open and flexible platform for privacy-preserving safe browsing	They suggest an approach that works well in both recent actual scenarios and ground truth datasets. Their technique also finds harmful websites involved in EKs, which can boost analysis efficiency even more.	Searchable encryption	Dynamic Threat Landscape: with evolving threats, vulnerabilities, and attack techniques research findings may become	

				outdated quickly.	
[16]	Tail Time Defense Against Website Fingerprinting Attacks	Tail Timeout (TT), a novel delay-conscious security mechanism against website fingerprinting assaults based on WT, is among the most successful countermeasures.	Tail Timeout Mechanism	Attackers may attempt to evade detection or manipulate experiments, leading to biased results or misleading conclusions.	
[17]	Website fingerprinting through the cache occupancy channel and its real world practicality	An alternative approach of the attack is used. In this the attacker delivers JavaScript code to the PC of the target user This code performs a cache side-channel attack so that additional websites being visited can be identified.	Machine learning	Evaluation of the effectiveness of proposed solutions can be challenging due to the lack of standardized evaluation criteria.	
[18]	Developing certificate-based projects for web security classes	A web security project that secures online applications using certificate-based techniques. The initiatives entail simulating attacks and defending against them.	Secure Socket Layer (SSL), Digital certificates, and HTTPS	Proposed solution is good for eavesdropping, but not much effective for spoofing.	
[19]	Improving Security of Internet of Vehicles Based on Post-Quantum Signatures with Systolic Divisions	The main operations are to Enhance the security of IoVs-based post-quantum signatures that are resistant to quantum computer assaults Divisions in a finite field.	Systolic architecture	Sampling technique: used sample is limited which may lead to inappropriate and biased result.	
[20]	Scanner++: Enhanced Vulnerability Detection of Web Applications with Attack Intent Synchronization	Scanner++ as a proxy-based architecture with intent synchronization through packages. Scanner++ initially employs a purification method to collect and refine attack intentions, which are comprised of attack surfaces and thread vectors. These can be derived from the request packets of the basic scanners.	Intent synchronization mechanism	The proposed solution integrates multiple scanner which can negatively impact website performance.	

[21]	Deep Learning-based Network Traffic Prediction for Secure Backbone Networks in Internet of Vehicles	In this paper the subject of end-to-end network traffic forecasting in IoV backbone networks is investigated, and a deep learning-based technique is suggested. The developed approach takes into account the spatiotemporal characteristics of network traffic and can capture network traffic's long-term reliance.	Deep Learning	Evaluating the effectiveness of cybersecurity solutions or techniques can be challenging due to the difficulty of replicating real-world attack scenarios in controlled environments.	
[22]	Web of Digital Twins	Explores the Web of Digital Twins (WoDT), a larger perspective in which the digital twin paradigm is used for the ubiquitous softwarisation of potentially large-scale interconnected physical realities.	Web of Digital Twins (WoDT)	Data limitation: only two case studies are considered which might not good sample size.	
[23]	Hiding Message Using a Cycle Generative Adversarial Network	In this study five networks at the same time are trained: a steganographic network, an inverse steganographic network, a hidden message reconstruction network, and two discriminative networks, which form a secret message cycle generative adversarial network (HCGAN).	Hidden message cycle generative adversarial network (HCGAN)	Proposed solution may impact performance negatively due to higher processing at network with slowdown browsing speed.	
[24]	Botnet attack detection in Internet of Things devices over cloud environment via machine learning	The purpose of this study paper is to look at cyber security in the age of B-IDS, DDOS, and virus assaults.	Machine learning	Use of nine sensor devices over N-BaIoT datasets will significantly reduce the performance.	
[25]	Chatbots: Security, privacy, data protection, and social aspects	The paper goes on the scenarios in which such type of data may be used and how chatbots controls it. Many bots function on a social/messaging network, this is having its own data rules and guidelines.	End-to-End Encryption (E2EE)	Data Quality and Bias: the chatbots data for this research gathered from few companies but the transparency of these data is	

				question.	
[26]	An Intelligent System for Securing Network From Intrusion Detection and Prevention of Phishing Attack Using Machine Learning Approaches	This study recommends estimating phishing websites and doing a comprehensive investigation referring pre-existing dangerous URLs, this aids in efficiently filtering the websites and creating a comparison of all the qualities in them.	Machine Learning Approaches	Susceptible to zero-day attack as research proposal is based on detailed analysis comparing pre-existing malicious URLs.	
[27]	Review of Machine Learning Techniques Used for Intrusion and Malware Detection in WSNs and IoT Devices	This research explains the various forms of attacks as well as the IoT authentication methods. Botnet attacks in IoT applications pose a significant risk for the security.	Machine Learning Approach for Intrusion and Malware Detection	Evaluating the effectiveness of solutions can be challenging due to the difficulty of replicating real-world attack scenarios in controlled environments.	
[28]	*Applications of Artificial Intelligence in Cyber Security*	This research focus on power of AI can be used to make a software or framework capable of advancing itself to make cyberspace secure.	Intrusion detection system	Lack of supporting evidence.	

3 Evaluation of collected data.

For this research the analysis of fivewebsites was done to verify how much a website is secure to protect confidential data and from malicious attacks. Only five websites were analyzed, because for this research security analysis was done on the live websites and it was hard to get approval for conducting penetration testing as it can cause partial service interruption.Also, trust was the second major issue in getting approval for this.

This analysis was conducted through penetration testing using Kali based penetration tools. These websites are the online portal of different educational institutes and small IT service provider companies, which contains data of their students and staffs. The key confidential data on these websites includes Personally Identifiable Information (PII) like address and contact information, salary and bank account details, performance reports etc. To analyze the security posture of these web portals, penetration testing was conducted. Starting with the identification of open ports on the target system using Nmap as shown below as sample.

```
Host is up (0.00037s latency).
Not shown: 65504 closed ports
PORT      STATE SERVICE     VERSION
21/tcp    open  ftp         vsftpd 2.3.4
|_ftp-anon: Anonymous FTP login allowed (FTP code 230)
22/tcp    open  ssh         OpenSSH 4.7p1 Debian 8ubuntu1 (protocol 2.0)
| ssh-hostkey:
|   1024 60:0f:cf:e1:c0:5f:6a:74:d6:90:24:fa:c4:d5:6c:cd (DSA)
|_  2048 56:56:24:0f:21:1d:de:a7:2b:ae:61:b1:24:3d:e8:f3 (RSA)
23/tcp    open  telnet      Linux telnetd
25/tcp    open  smtp        Postfix smtpd
|_smtp-commands: metasploitable.localdomain, PIPELINING, SIZE 10240000, VRFY, ETRN, STARTTL
ES, 8BITMIME, DSN,
| ssl-cert: Subject: commonName=ubuntu804-base.localdomain/organizationName=OCOSA/stateOrPr
no such thing outside US/countryName=XX
| Not valid before: 2010-03-17T14:07:45
|_Not valid after:  2010-04-16T14:07:45
|_ssl-date: 2019-08-22T13:31:52+00:00; +1s from scanner time.
53/tcp    open  domain      ISC BIND 9.4.2
| dns-nsid:
|_  bind.version: 9.4.2
80/tcp    open  http        Apache httpd 2.2.8 ((Ubuntu) DAV/2)
|_http-server-header: Apache/2.2.8 (Ubuntu) DAV/2
|_http-title: Metasploitable2 - Linux
111/tcp   open  rpcbind     2 (RPC #100000)
| rpcinfo:
|   program version   port/proto  service
|   100000  2            111/tcp  rpcbind
|   100000  2            111/udp  rpcbind
|   100003  2,3,4       2049/tcp  nfs
|   100003  2,3,4       2049/udp  nfs
|   100005  1,2,3      42648/udp  mountd
|   100005  1,2,3      59475/tcp  mountd
|   100021  1,3,4      46979/tcp  nlockmgr
|   100021  1,3,4      47652/udp  nlockmgr
|   100024  1          35795/udp  status
|_  100024  1          47809/tcp  status
139/tcp   open  netbios-ssn Samba smbd 3.X (workgroup: WORKGROUP)
445/tcp   open  netbios-ssn Samba smbd 3.X (workgroup: WORKGROUP)
512/tcp   open  exec        netkit-rsh rexecd
513/tcp   open  login
514/tcp   open  tcpwrapped
1099/tcp  open  java-rmi    Java RMI Registry
1524/tcp  open  shell       Metasploitable root shell
2049/tcp  open  nfs         2-4 (RPC #100003)
```

Fig.03 Nmap scan result on the targeted web portal

The Nmap command can assist in determining the open ports required to carry out the different attacks. Consider a DoS approach; the TCP port 80, which is a HTTP port and open in above screenshot, is sufficient for the attack. Any unsecure ports or services like TCP/21 & TCP/80, which is open in above screenshot will allow man-in-middle attack through packet inspection as shown below in packet capture screenshot.

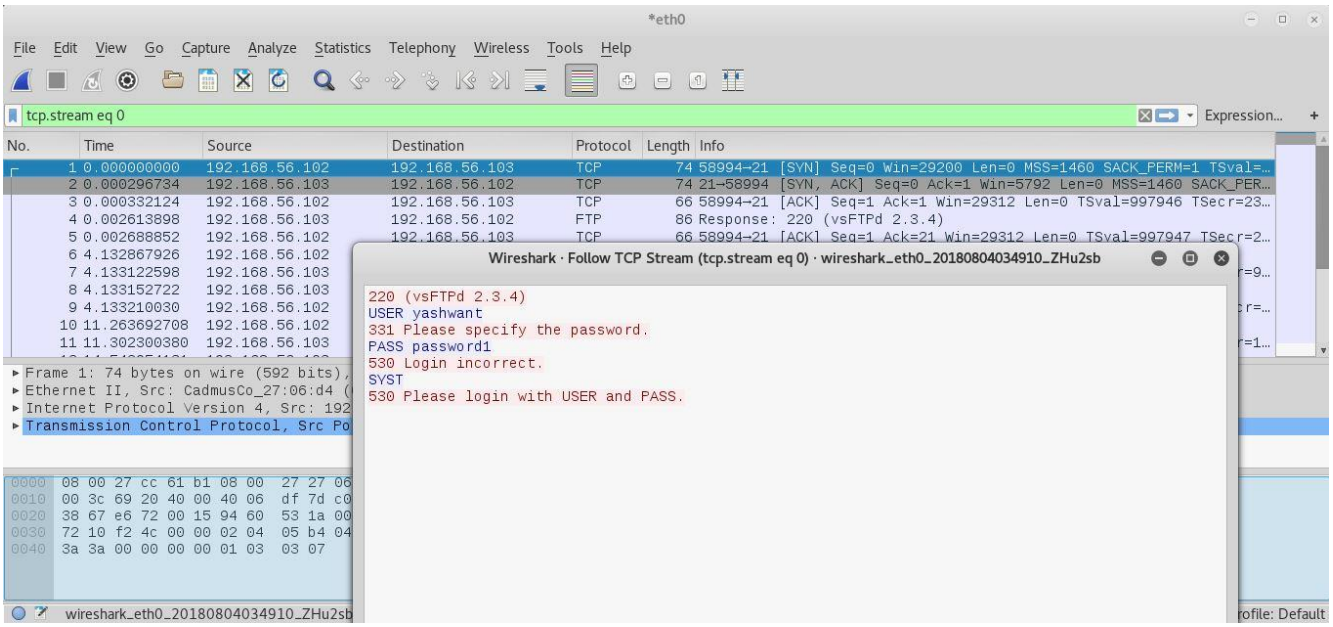


Fig.04 Credential disclosure through unsecure protocols

After Nmap, Nessus and OpenVAS are used to get more details of the vulnerabilities on the targeted

system. And finally, penetration testing was conducted using various tools on the targeted system to identify the existing security posture and the severity level if the identified weakness will be misused by any attacker as listed in below table.

Table 2 List of Attacks Performed

Attack	**Tools Used**	**DurationofAttack**	**Attacker**	**Target**
DoS attack	Hping and SYN Flood	21 hours	KaliLinux	Metasploitable
Web Application attack	Slowloris	28 hours	KaliLinux	Windows Web Server
Brute Force attack	FTP-Patator	36 hours	KaliLinux	Metasploitable
Infiltration	Nmap and Hping	20 hours	KaliLinux	Metasploitable

Based on the penetration testing on the targeted system, system owner must require focusing on analysis of the security posture of webservers through regular analysis and review of security posture. Keep track of recommended secure ciphers and threat signatures and based on the recommendation harden the system accordingly. After the analysis of the security posture of web-portals, further data is collected from the website developer, white-hat hackers, and the end-users. The sample size of the survey is 150 respondents. To understand the security challenges, in the survey questions were asked to identify their experience with the frequency of security flaws, key drivers for security weaknesses and how much company is focusing on website security?

Out of these 150 respondents, the majority of the respondents are end-users with vast IT experience followed by website developers and white-hat hackers. Somewhat less than 30% of participants manage 1-5 webpages, while 18% manage 5-15 sites. Below is the list of the summarized findings through the survey data collection.

Table 3 Survey responders' background

Background of responders	**Percentage in survey**
IT professionals	36%
Students	8%
White hat attackers	14%
Website developers	20%
Non-IT users	18%
Business owner	4%

Table 4 Summarized survey findings

Key Areas	**Low Impact**	**High Impact**	**Total response**
Absence of security policies	42	108	150
Lack of security awareness	36	114	150
Secure website plugin	31	119	150
Human error	86	64	150
Absence of backup policy	57	93	150
Reliable solution for security report	63	87	150

for decision making			
Well defined terms and conditions	68	82	150

More than seventy-six percent of the study respondents are concerned about website security. Nearly 46% of respondents have witnessed an increase in cyberattacks on the websites for which they are accountable. Surprisingly, 38 respondents had no idea whether or not the attacks had grown. We also observed that 15% of respondents had seen a compromised website in the previous month before doing the study. This provides us a fair idea of the scope of the challenge. Infected websites are used to launch further attacks against many other sites and businesses. Hacked websites are frequently used to drive traffic to dangerous sites, steal credit card data and, in some circumstances, infect the systems of visitors.

AI – based website security mechanism is required to be developed through which many of these challenges are able to be handled. There are technologies available to collect data continuously and autonomously from across the corporate organization's information systems. This information is then assessed and used to conduct pattern correlation across abundance of data pertinent to the organizational attack surface.

As more of our lives become digital, a company's contemporary portfolio of software assets grows across all sectors like APIs for web apps, microservices, and serverless computing. Because of the increased complexity of code and architecture, establishing and sustaining the security of all of these has increasingly tough.

For the website security there are multiple aspects which need to be considered. Cyber security is one of the key considerations of it as it defines the security aspects for online access and to protect the online system from remote users or hackers. Below are listed some of the major points for the cybersecurity of a website.

- **Defects in the level code,** which means malfunctioned code for the application which may allow backdoor entry for hackers.
- **Trustworthy computing and opportunistic user and system behavior** highlights the importance and loyalty and trustworthiness of the stakeholders for the safety of online resources.
- **Imperfections in the security infrastructure** focus on security flaws in organization's online assets. There can be multiple reasons for such kind of flaw like inappropriate security procedures, lack of monitoring and security compliance, absence of regular security review etc.
- **Inadequacies in confidentiality**, focus on CIA triad is crucial for the user trust and the safety of online resources.
- **Problems with usefulness** highlights that many time deliverables are not appropriate for the actual requirements mainly because the analysis of the requirement was not done appropriately.
- **Inadequate security measures**, it has been observed that organizations sometime focus on attracting customers while the security get neglected. This will not be beneficial in the long run as it adversely affect the user trust and the organization's reputation.

4 Identified security gap.

Absence of standard and guidelines for the web designing services is one of the key reasons for bad user experience like no proper communication, terms andconditions not defined in user friendly manner. As there is no assurance given by most of the web service providers, it is necessary to provide guidance in the form of framework for end-users to validate that the service provider has good security measure in place. Transparency and communication are important for building trust and to improve user experience. Therefore, in the proposed framework this is one of the keyconsiderations. It is anticipated that security measures and features will be added to the website system in the future.

For the safety of data-in-transit or safety of database, Web service providers don't give any assurance. Database breach of a big website, phishing attacks, cyber vulnerabilities, securing from malicious cyber-attacks is going fast with the advent of digitalization. Cyber assaults have been named the fifth biggest critical danger for 2020, and they are projected to become the new normal in both the private and public sectors. (Source-https://www.embroker.com/blog/cyber-attack-

statistics/).

To examine the top important cyber security challenges which if addressed with the right solutions may be able to assist small and large businesses in overcoming broad-level data breaches and phishing attacks. Some of the key cyber security challenges are as follows:

- Third parties may illegally use the 5G network's potential.
- Malware on mobile devices is becoming more prevalent.
- Artificial Intelligence (AI) is controlling Cyber Security Systems somewhere.
- IoT devices are increasing in popularity.
- Attacks on critical business aspects are being targeted by ransomware.
- Phishing and spear-phishing attacks are out of control.
- Hacktivism is on the rise.
- Drone jacking is a new trend that concerns Cyber experts.
- Social engineering preventative measures.
- Attacks on cloud systems.
- Cyber-attacks are increasing on IoT (Internet of Things) devices.
- Increase in phishing attacks.
- Spreading malware via Ads.

5 Use of AI to address the identified website security challenges.

AI, in conjunction with machine learning, is completely equipped to resolve some of our most difficult challenges, and website security is surely any of them. Machine Learning and AI may be used to "keep up with the bad guys," automating threat detection and acknowledging more efficaciously than traditional technology driven techniques in today's ever-changing cyber-attacks and proliferation of endpoints.

Artificial intelligence (AI) with machine learning (ML) may be used for the security in a variety of methods, including malware identification via fast and spontaneous packet's payload analysis based on the saved signature, attack's prediction, and gore clustering security events. It has the potential to discover previously unknown attacks that lack a recognized signature. AI updates the payload signature through self-learning through ongoing traffic and help to protected from the zero-day vulnerabilities. The AI and ML technologies are used to deal with a large number of input variables and to draw a valid result.

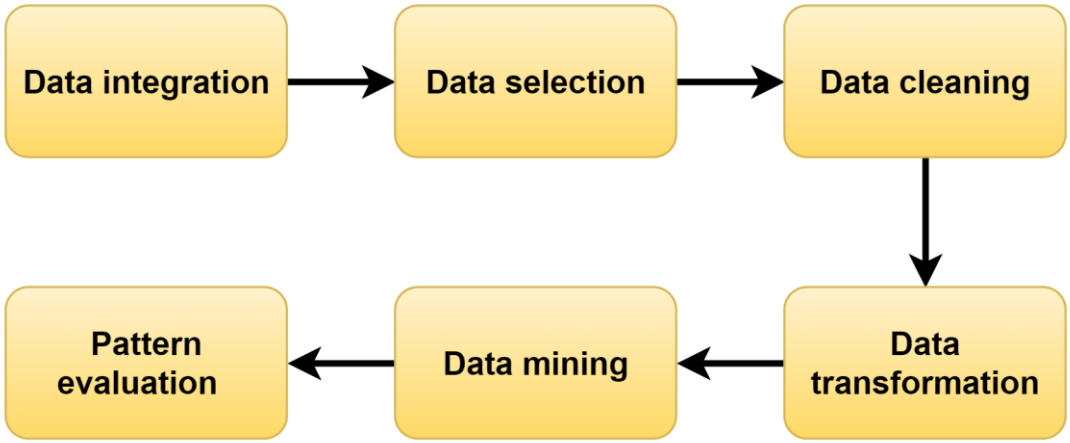


Fig.05 – Steps involved in data mining used in machine learning

It can also be helpful to detect abnormality on the traffic to particular website and help to protect from the DoS/DDoS attacks. AI will be helpful for regular audit of the security posture of the system. It will analyze the existing security posture on a regular basis based on the security template and alert the owner if any difference from the ideal configuration.

6 Proposed framework for the identified challenges

Below is the pictorial representation of the key components for website security and how an attacker or a user accesses the website online. For better security and availability, these days

organizations use load-balancer to distribute traffic and to ensure availability along with the web application firewall. These are some of the key considerationswhen planning for the website security.

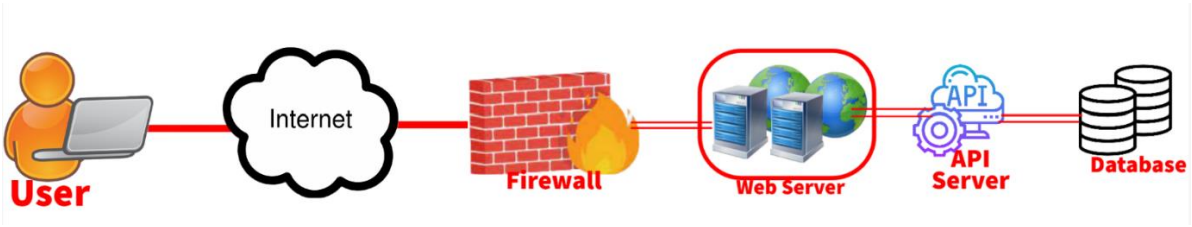
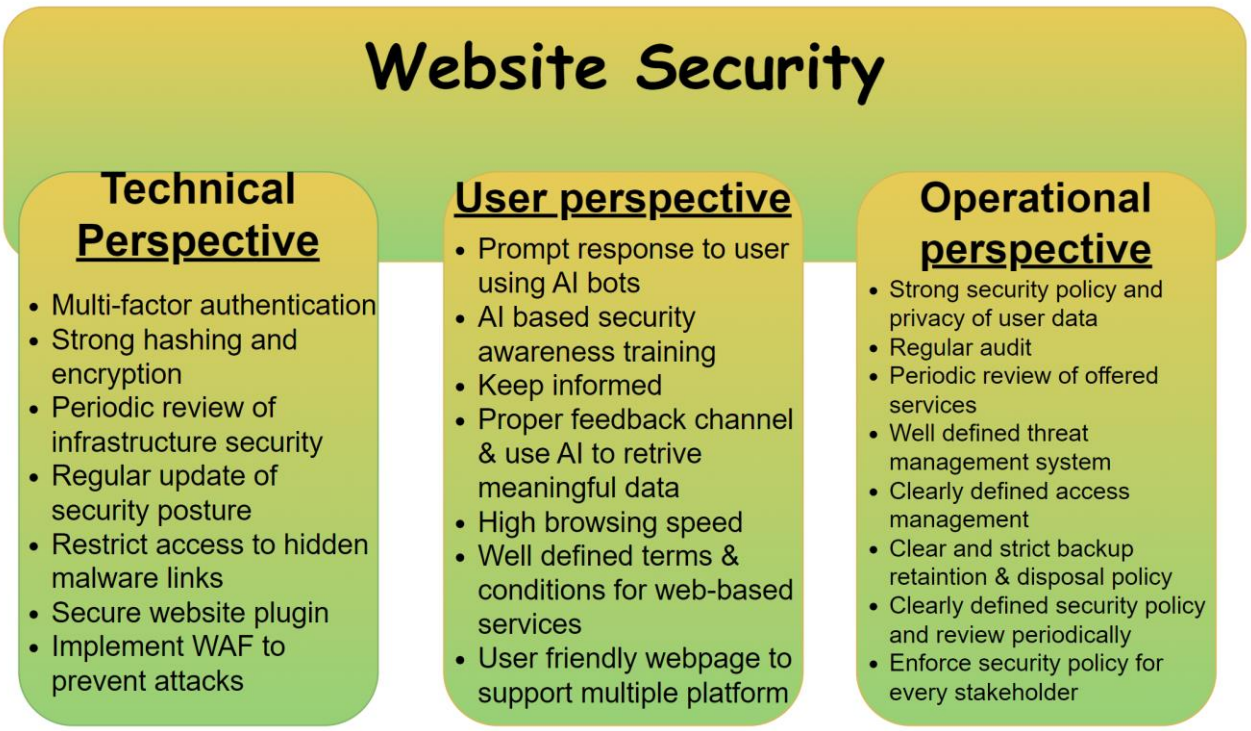


Fig.06 – Simplified view of an attack

Among the various approaches to deliver the proposed solution, the proposal is delivered in the form of framework.As the framework allows content management and helps to deliver solution in simple and structured way, which a non-technical user can also easily follow. It also guides users on the key considerations while deciding among the various web-based services.



7Fig.07 – Framework for Website Security

This framework will be useful for the website developers as well as for system owner to guide them from the security perspective to secure the data and to maintain confidentiality and trustworthiness. Below is the summarized explanation of the proposed framework.

The technical perspective of this framework focus on the technical aspects of the website security which includes the multi-factor authentication of users to validate their identity, strong hashing and encryption is to save the transit data from man-in-middle attacks, periodic review of infrastructure security to ensure the security measures are up to standard and not out dated, regular update of security posture to ensure that system is up-to-date and avoid any unknown vulnerability in the system, analyze and restrict access to the malware link because if the

website is compromised then it can be used to perform further attacks, secure website plugin as it is identified that plugins are the common source of backdoor entry for an attacker, implement web application firewall (WAF) to analyze the traffic at application layer and prevent attacks.

Similarly, the user perspective of this framework focus on the prompt response to the users general queries using AI bots, AI based security awareness training like use of IVR so that user can virtualize to understand and remember it for longer time, keep informed for the any relevant changes or cyberattacks to build trust with users, multiple feedback channel so that users can express easily and use AI to fetch the relevant data from it to improve the deliverables, high browsing speed for better user experience, terms and conditions for the online services must be clear and offered in a convenient manner then it will be simple for users to know and accept it and to avoid any future conflicts, also the webpage should support multiple platform like laptop, mobile and tablet application etc.

Finally, the operational perspective of this framework focus on the robust security policy and user data privacy as it is basic requirement for safety of a system, the regular audit by external party will help to validate the existing security measures and reliability of security policies, periodic review of existing services using AI (based on available metadata) will help to analyze the relevance of existing services and the possible improvement, appropriate threat and access management is important to ensure the CIA triad for a website, backup must be maintained regularly and the safe keeping of the backup file is also important to avoid any possible misuse of it, at last a well-defined security policy for every stakeholder is must and should enforce properly.

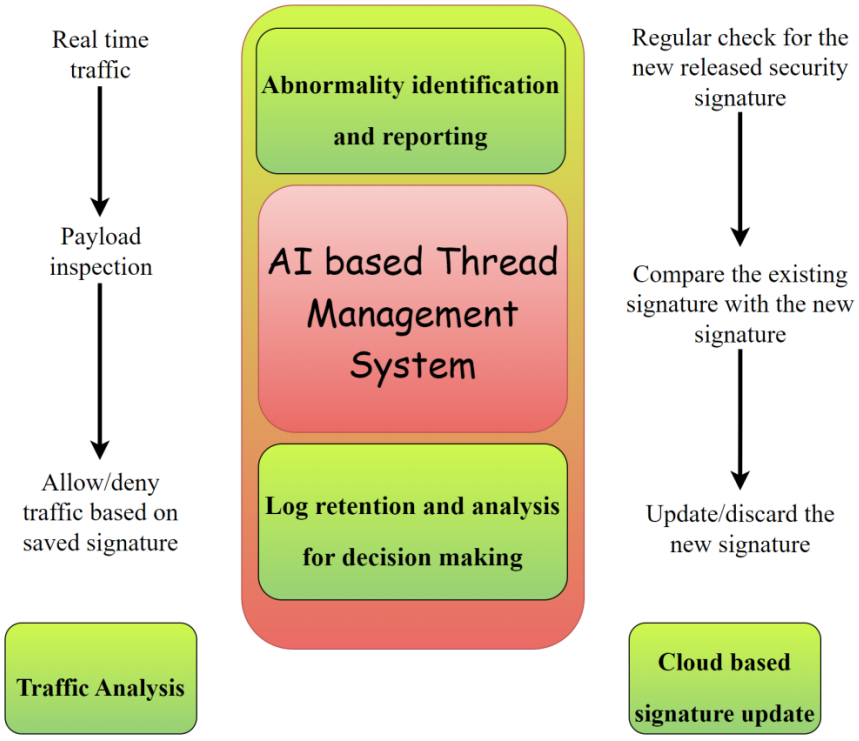


Fig.08 – AI based threat management system

AI will be highly useful for each and every aspect of the threat management system. AI will help to analyze real-time traffic and make quick decisions whether to allow or deny traffic based on the existing security signature. It will also be helpful to automate the security signature update and the system posture update based on the recommendation. Such recommendations, AI can easily fetch from the cloud automatically. Based on the real-time traffic logs, AI can generate reports which will be useful for the decision-making process. The summarized benefits of AI in

threat management system are depicted in above figure.

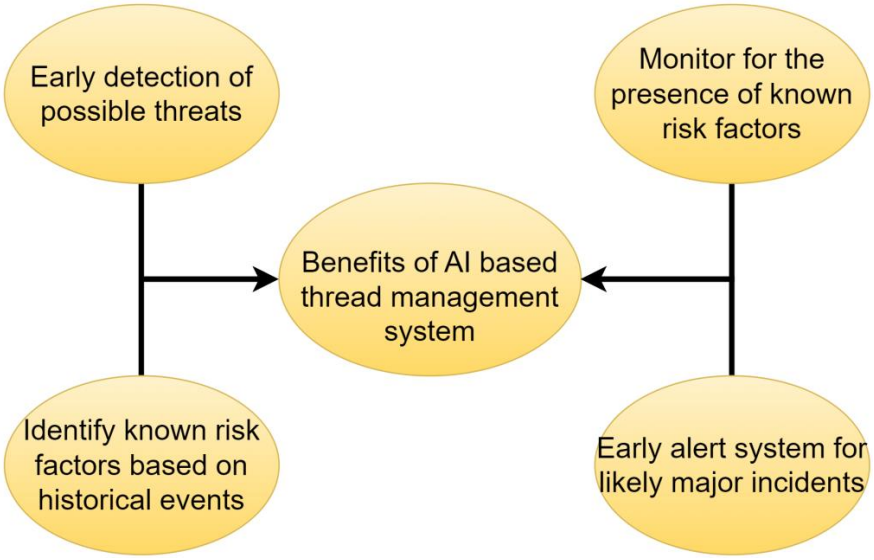


Fig.09 – Benefits AI based threat management system

Along with the above highlighted use of AI in threat management system, token-based user session is also important to ensure that the existing session will not hijack and misuse. It is an important role played by AI for maintaining the uniqueness of each session and maintaining logs, which can be further analyzed to understand the user interest and to provide customized services based on it. These unique session logs will also be useful for the audit purpose.Organizations can implement their own threat management system which can perform packet inspection and trigger the alert to the system owner if any abnormalities identified based on the metadata. This thread management system can check the newly released security signature online and update its database accordingly. The proposed solution will be helpful to ensure high availability of website, protect user data, control website traffic,

protect trust and reputation of the website owner, and security automation through AI. Below is a snapshot of the thread management system, which was designed through deep learning over the 5+ years of inhouse metadata. In this system can also manually block the malicious domains based on the security advisory as shown below.

		RECEIVE TIME	TYPE	THREAT ID/NAME	FROM ZONE	TO ZONE	SOURCE ADDRESS	DESTINATION ADDRESS	TO PORT	APPLICATION	ACTION	SEVERITY
		04/20 12:04:07	vulnerability	Microsoft Windows win.ini Access Attempt Detected	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	80	web-browsing	reset-server	high
		04/20 12:03:22	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	49155	incomplete	reset-server	critical
		04/20 12:03:12	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	49231	incomplete	reset-server	critical
		04/20 12:03:02	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	2179	incomplete	reset-server	critical
		04/20 12:02:52	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	5985	incomplete	reset-server	critical
		04/20 12:02:42	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	139	incomplete	reset-server	critical
		04/20 12:02:32	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	111	incomplete	reset-server	critical
		04/20 12:02:27	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	49202	incomplete	reset-server	critical
		04/20 12:02:17	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	49235	incomplete	reset-server	critical
		04/20 12:02:07	vulnerability	Apache Log4j Remote Code Execution Vulnerability	inside	BM_Inter...	1[redacted]?0	1[redacted]?5	9850	incomplete	reset-server	critical
		04/15 01:13:39	flood	TCP Flood	outside	outside	0.0.0.0	0.0.0.0	0	not-applicable	random-drop	critical
		04/15 01:13:29	flood	TCP Flood	outside	outside	0.0.0.0	0.0.0.0	0	not-applicable	random-drop	critical

Fig.10 – AI based threat management system.

This threat management system classifies packets based on in-house metadata and also on the security signature from online repository. It performs the deep packet inspection by checking on the payload of the data-in-transit and if the payload seems different from the service, then packet is categorized as malicious. With the benefit of a high level of security for the internet traffic, one of the major drawbacks of this system is possible slowness if not implemented correctly. This is because payload inspection requires lots of processing and computational efficiency. Therefore, organizations need to implement it wisely.

7 Conclusion

This proposed framework for website security is useful to meet the existing challenges, but it needs to be evaluated and evolve further based on the changing requirement. It needs to be aware of the website and web services user for the possible security risk and guide them for the basic solution

for those risks. Web-based service providers must boost user confidence through effective communication and visibility. We need to continually reinvent online security technology in order to deliver better intelligent services to consumers, and artificial intelligence is the essential technology that can tackle this requirement.

The approaches are to introduce the application of artificial intelligence techniques in website security. Main aim towards the research area chosen will be helpful to overcome several security loopholes and also it helps to spread awareness about emerging security threats. Consider safe, alert,
and resilient cyber models capable of managing risks and driving innovation in the cyber environment.

The key limitation for this research was getting approval for the available websites to analyze the security of it through penetration testing. This is mainly because of the lack of confidence of the owner as they suspect that if any loophole is identified then it may be misused in future. Also, this activity may cause slowness for end users and possibly denial-of-services. The second limitation for this research was the sample size, as for this search data was initially collected from a survey of 150 respondents followed with the literature review. If there are a higher number of respondents, then it will be possible to get more reliable info.

Similar to the proposed solution for the website security, in-depth research needs to be done for the usefulness of AI for database security and to mitigate zero-day vulnerabilities. Future research can also focus on how to minimize the impact of cyberattacks to a website using AI.

10 Acknowledgement

First and foremost, I want to sincerely thank my supervisors, for their constant support during this journey, as well as for their compassion, inspiration, and immense knowledge. I want to thank my family and my friends, for their understanding, encouragement, and motivation. Without their help, I wouldn't have been able to get through the difficulties that I had encountered and achieve what I wanted to do. Finally, I'd want to express my gratitude to everyone who helped me with this project, even if their names aren't included here.

11 Conflict of interest

The authors have NO associations with or participation in any institution or group with any financial stake (such as charitable contributions, academic grants, speaking engagements, participation, employment, consultancies, shareholdings, or other equity ownership, and expert witnesses or patent-licensing arrangements), or non-financial participation (such as individual or business partnerships, associations, knowledge, or opinions) in the subject matter or materials presented.

References

[1] Alsariera, Y. A., Adeyemo, V. E., Balogun, A. O., &Alazzawi, A. K. (2020). Ai meta-learners and extra-trees algorithm for the detection of phishing websites. *IEEE Access*, *8*, 142532-142542.

[2] Wang, P., Gao, H., Rao, Q., Luo, S., Yuan, Z., & Shi, Z. (2020). A security analysis of captchas with large character sets. *IEEE Transactions on Dependable and Secure Computing*, *18*(6), 2953-2968.

[3] Qin, Y., Wang, W., Zhang, S., & Chen, K. (2021). An exploit kits detection approach based on http message graph. *IEEE Transactions on Information Forensics and Security*, *16*, 3387-3400.

[4] Hwang, R. H., Peng, M. C., Huang, C. W., Lin, P. C., & Nguyen, V. L. (2020). An unsupervised deep learning model for early network traffic anomaly detection. *IEEE Access*, *8*, 30387-30399.

[5] Sadeghzadeh, A. M., Tajali, B., &Jalili, R. (2021). Awa: Adversarial website adaptation. *IEEE Transactions on Information Forensics and Security*, *16*, 3109-3122.

[6] Al-Naami, K., El-Ghamry, A., Islam, M. S., Khan, L., Thuraisingham, B., Hamlen, K. W., ...& Rashad, M. Z. (2019). Bimorphing: A bi-directional bursting defense against website ingerprinting attacks. *IEEE Transactions on*

Dependable and Secure Computing, *18*(2), 505-517.

[7] Liu, D., & Lee, J. H. (2020). CNN based malicious website detection by invalidating multiple web spams. *IEEE access*, *8*, 97258-97266.

[8] Moubayed, A., Aqeeli, E., &Shami, A. (2021). Detecting DNS Typo-Squatting Using Ensemble-Based Feature Selection & Classification Models. *IEEE Canadian Journal of Electrical and Computer Engineering*, *44*(4), 456-466.

[9] Löbner, S., Tesfay, W. B., Nakamura, T., &Pape, S. (2021). Explainable machine learning for default privacy setting prediction. *IEEE Access*, *9*, 63700-63717.

[10] Vaccari, I., Narteni, S., Aiello, M., Mongelli, M., &Cambiaso, E. (2021). Exploiting Internet of Things Protocols for Malicious Data Exfiltration Activities. *IEEE Access*, *9*, 104261-104280.

[11] Uplavikar, N., Malin, B., & Jiang, W. (2020). Lucene-P $^ 2 $2: A Distributed Platform for Privacy-Preserving Text-Based Search. *IEEE Transactions on Dependable and Secure Computing*, *18*(6), 2801-2819.

[12] Zhiheng, W., &Jianhua, L. (2021). Flamingo search algorithm: A new swarm intelligence optimization algorithm. *IEEE Access*, *9*, 88564-88582.

[13] Rahman, M. S., Imani, M., Mathews, N., & Wright, M. (2020). Mockingbird: Defending against deep-learning-based website fingerprinting attacks with adversarial traces. *IEEE Transactions on Information Forensics and Security*, *16*, 1594-1609.

[14] Wang, Y., Xu, H., Guo, Z., Qin, Z., &Ren, K. (2022). snWF: Website Fingerprinting Attack by Ensembling the Snapshot of Deep Learning. *IEEE Transactions on Information Forensics and Security*, *17*, 1214-1226.

[15] Cui, H., Zhou, Y., Wang, C., Wang, X., Du, Y., & Wang, Q. (2019). PPSB: An open and flexible platform for privacy-preserving safe browsing. *IEEE Transactions on Dependable and Secure Computing*, *18*(4), 1762-1778.

[16] Liang, J., Yu, C., Suh, K., & Han, H. (2022). Tail Time Defense Against Website Fingerprinting Attacks. *IEEE Access*, *10*, 18516-18525.

[17] Shusterman, A., Avraham, Z., Croitoru, E., Haskal, Y., Kang, L., Levi, D., ...&Yarom, Y. (2020). Website fingerprinting through the cache occupancy channel and its real-world practicality. *IEEE Transactions on Dependable and Secure Computing*, *18*(5), 2042-2060.

[18] Rahman, S., Nguyen, T. A., & Yang, T. A. (2006). Developing certificate-based projects for web security classes. *Journal of Computing Sciences in Colleges*, *21*(5), 28-37.

[19] Yi, H., Chi, R., Huang, X., Cai, X., &Nie, Z. (2022). Improving Security of Internet of Vehicles Based on Post-Quantum Signatures with Systolic Divisions. *ACM Transactions on Internet Technology (TOIT)*.

[20] Yin, Z., Xu, Y., Ma, F., Gao, H., Qiao, L., & Jiang, Y. (2018). Scanner++: Enhanced Vulnerability Detection of Web Applications with Attack Intent Synchronization. *ACM Transactions on Software Engineering and Methodology*.

[21] WANG, X., NIE, L., NING, Z., GUO, L., WANG, G., GAO, X., & KUMAR, N. (2021). Deep Learning-based Network Traffic Prediction for Secure Backbone Networks in Internet of Vehicles. *ACM Trans Internet Technol*, 1-24.

[22] Ricci, A., Croatti, A., Mariani, S., Montagna, S., &Picone, M. (2022). Web of Digital Twins. *ACM Transactions on Internet Technology (TOIT)*.

[23] Shi, W., & Liu, S. (2022). Hiding Message Using a Cycle Generative Adversarial Network. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*.

[24] Waqas, M., Kumar, K., Laghari, A. A., Saeed, U., Rind, M. M., Shaikh, A. A., ... &Qazi, A. Q. (2022). Botnet attack detection in Internet of Things devices over cloud environment via machine learning. *Concurrency and Computation: Practice and Experience*, *34*(4), e6662.

[25] Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., &Ogiela, L. (2021). Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience*, *33*(19), e6426.

[26] Banik, S., Banik, S., & Mukherjee, A. (2021).

An Intelligent System for Securing Network From Intrusion Detection and Prevention of Phishing Attack Using Machine Learning Approaches. *Machine Learning Techniques and Analytics for Cloud Security*, 193.

[27] Alphonse, A. S., Priya, E. D., & Kowsigan, M. (2022). Review of Machine Learning Techniques Used for Intrusion and Malware Detection in WSNs and IoT Devices. *Design and Development of Efficient Energy Systems*, 57-65.

[28] Singh, U. Z. (2020). *Applications of Artificial Intelligence in Cyber Security* (No. 4029). EasyChair.

[29] 2019 Data Breach Investigations Report. (2019). [online] *Verizon*. Verizon. Available at: https://www.verizon.com/business/resources/reports/2019-data-breach-investigations-report.pdf

[30] Alexandra Sava, J. (2022). Virtualization. [online] Statista. Available at: https://www.statista.com/topics/6795/virtualization/#dossierKeyfigures

[31] OWASP (2021). OWASP Top 10:2021. [online] owasp.org. Available at: https://owasp.org/Top10/