IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Evaluation of India's Most Visited Websites in Aspects of Security & Structure

Irshad Alam [1], Satwinder Singh [2*], Gurpreet Kaur[3]

[1, 2]Department of Computer Science and Technology, School of Engineering and Technology, Central University of Punjab, Bathinda, India
[3]Department of Law, Bathinda College of Law, Bathinda, India

[*]*Corresponding Author: satwinder.singh@cup.edu.in,*

DOI: https://doi.org/10.26438/ijcse/v7i5.985991 | Available online at: www.ijcseonline.org

Accepted: 15/May/2019, Published: 31/May/2019

***Abstract*—** Web applications play a significant role in today's digital age. Their uses in our lives have become indispensable. It has made web applications a favorite target for attackers and has increased web security risk. This study focuses on finding structural aspects and vulnerabilities present in India's 50 websites which were categorized into five categories of 10 most visited sites, i.e., e-commerce, news, entertainment, education, and other scanned as an ordinary user to consider safety assessment of these websites. The knowledge about these sites, such as technologies used and infrastructure they have, the vulnerabilities they possess, has been investigated using penetration tests in this study. As a result of this research, web server information and operating system information from 86% to 66% respectively of the reviewed websites are identified. Medium and low degree vulnerabilities have been present in all scanned websites. Some of them even have High vulnerabilities also. With the vulnerability screening tests, their degree of vulnerabilities graph revealed, and information about the most identified weaknesses was given.

***Keywords*—**Web Applications, Penetration Testing, Penetration Testing Tools, Weakness Analysis, Web Security.

I. INTRODUCTION

Applications on the internet have made our daily lives easier with their simple and fast access, ignoring time and place; they have become crucial. It made Web applications a common target for malicious users and increased web security risk. Websites are critical pathways to facilitate customer service, input procurement, e-commerce, and employee connectivity, and they continue to influence important penetration in several business areas. Businesse, either small or large, are hosting web services, and over more than 50% of small businesses are now offering web accessibility. As such, internet security has become critically vital for many organizations, and the prevention of security compromises enabled by web vulnerabilities is gaining the attention of company leadership and the broader security community increasingly. Nevertheless,

Web vulnerabilities are one of the many possible causes of recently happened security breaches contributing to large revelation of user's personal data, leakage of sensitive business information, and causing great losses.

Recent years have witnessed the rise of web apps and their code complexity. Web applications have evolved from sets of straightforward, static web content, to feature-rich web 3.0 applications which often integrate third-party services. Unfortunately, there is a lot of features that a present-day web application possesses, the larger attack surface it usually exposes to attackers. With the incorporation of client-side JavaScript code, attackers will inject malicious JavaScript code in user inputs and exploit numerous ways in which web browsers invoke their JavaScript engines. In cases wherever web applications behave otherwise to administrators and normal users, attackers will break authentication processes and access sensitive data and functionality. In recent years, the integration of third-party services has become common, giving attackers a new chance to take advantage of miscommunications between servers and service providers. The best practice for web security is to develop an application with security in mind end-to-end. However, several web developers are unacquainted with numerous attack vectors, particularly application-specific ones.

Application-specific vulnerabilities in web apps are especially difficult to detect. When developing a web app, developers often have a clear picture of how the ideal application should be in their minds. Unfortunately, in real, the implemented application often does more or less than what it is intended to do. So purpose of this study is to find out web vulnerabilities present in the most visited websites

of India in aspects of their structure and security. Most visited websites data has been collected from www.alexa.com and Google.

Rest of the paper is organized as follows. Section II contain the related work, Section III gives information about penetration testing and methodology used, Section IV contain results and discussions, the last section concludes the paper.

II. RELATED WORK

Internet and web security are increasing every day due to millions of users and exist in all areas of life from shopping to traveling, from finance to health, from communication to entertainment and from office to outdoor works. The Internet has become an integral part of our daily lives, providing unprecedented convenience through web and mobile applications say Phung et al. [1]. Kochare and Chalurkar [2] said web applications are open to all, including hackers, because of their definition, the security of these applications is troublesome. Since nowadays the information security is important, there are many studies in the literature. Polat et al. [3] mentioned the importance of penetration testing, which should be done intermittently for information security, especially for the information security, by talking about the types of infiltration tests, the study methodology, and the application forms.

Web penetration test methods detect vulnerabilities using different attack types. Among these types, SQL injection attacks and XSS attacks are very common. Ruse et al. [4] have handled these SQL injections and XSS attack techniques, which are among the most common attack types. They also discussed about the protection methods in details and mentioned their detection methods. Huang et al. [5] analyzed the current situation of Chinese websites using 57,122 web spoofing attacks from the year 2012 to 2015 presented by researchers. According to them, the data were collected in four different groups, which include companies listed in the stock exchange market, educational institutions, government institution, and new companies. They developed an automatic classifier model for differentiating web security vulnerabilities and found the most common 15 security vulnerabilities and their distribution. They concluded that the start-up companies had serious security vulnerabilities present in them while government and educational institutions showed more interest in this area. Stiawan et al. [6] analyzed cyberattack techniques and the penetration test anatomy for assisting security officers in performing an appropriate self-security assessment on their network systems. Sandhya et al. [7] focused on solving the problem of the threat of expose of data issue by surveying various tools for penetration testing. Besides, they provided a sample for basic penetration testing using Wireshark. Nixon and Haile [8] used some penetration tests on WLAN security protocols and MAC Filtering. They used a computer with Kali Linux operating system for this aim. As a result of various experiments, they observed that there are many loopholes in WLAN and proposed a solution to secure the WLAN using Pseudo Random MAC Address Generation Algorithm called PRMACGA. Bullee et al. [9] investigated the extent of persuasion principles are used in successful social engineering attacks. They extracted 74 scenarios from social engineering literature and analyzed. Each scenario was split into attack steps, containing single interactions between offender and target. For each attack step, persuasion principles were identified. As a result of the scenario analysis, they determined how to exploit the human element in security. Wu et al. [10] analyzed the measures that a social planner, such as the government or industry association, controls firms' security decisions. The obtained results show that taken precautions measures are not always effective. They recommend to social planners to enhance or attenuate the controlling level of the two security decisions based on realistic security and business environments. Cisar et al. [11] discussed the assessment of information system security. The authors focused on three major features of the system for the security of an information system: availability, integrity, and confidentiality. The paper presents a wide-ranging overview of possible uses, benefits, and drawbacks of the Kali Linux Operating System.

This study focuses on showing what kind of information and vulnerabilities can be found by an ordinary internet user as Web penetration tests are indispensable for evaluation of web sites in aspects of security. The predetermined web sites are scanned using web penetration test methods via analysis sites, and open source programs in kali to collect some information about the technologies and infrastructure they use.

III. PENETRATION TESTING AND METHODOLOGY

Penetration tests are so critical for evaluation of web sites in aspects of structure and security purposes. Penetration tests, also called pen testing is the practice of testing a computer, network or web applications to find security vulnerability to protect technologies infrastructure and institution's data flow by an attacker (Hacker, former employee, Script Kiddie, etc.) or malware (worm, virus, Trojan horse, spyware, etc.). The purpose of these penetration tests are to find the weaknesses and eliminates them to prevent malicious people's unauthorized access. Penetration tests consist of 5 phases as data collection, weakness scan and analyze exploitation, continuous access and reporting.

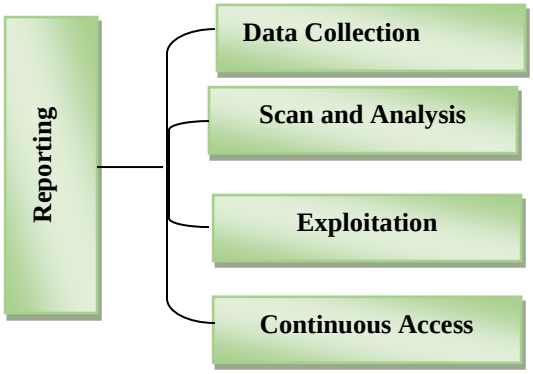
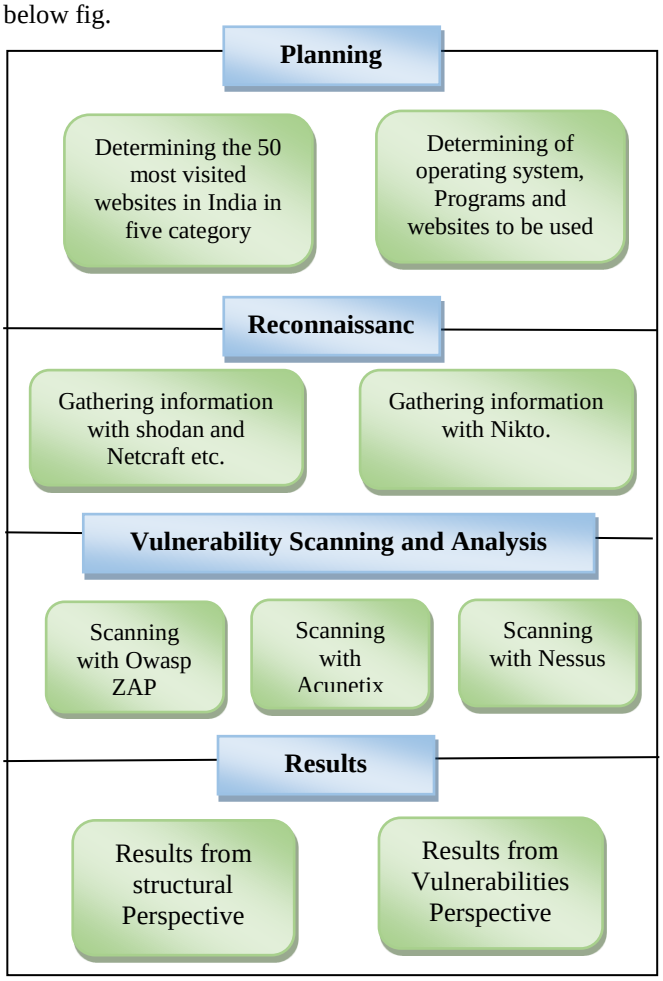


Fig 1. Steps involved in Penetration Testing

The method followed for this research work is shown in below fig.



For working purposes websites are determined based on https://www.alexa.com website statistics and are grouped into five different groups shown in table1. For carrying out the studies, Kali Linux in the virtual box and Microsoft Windows 10 operating systems are used. In order to gather information about websites, Shodan and Netcraft analysis websites are used. Also, open source coded Nikto application under Kali Linux is used. For weakness scan popular Owasp ZAP v2.7.0 and Acunetix v12 (https://www.acunetix. com trial version) program which works on Microsoft Windows and Nessus scanner's trial version8.3.1 (https://www.tenable.com) are used.

Table 1. Grouping of determined websites

Group-1	E-Commerce
Group-2	News
Group-3	Entertainment
Group-4	Education
Group-5	Other

This study includes information collection from penetration tests methods, weakness scanning and analysing operation. The method followed is shown in Fig 2. In the light of method followed; first selected websites' information about infrastructure and their implied technology is gathered. For this Shodan and Netcraft which are analysis websites are used and then Nikto in kali is used. Comparisons were made via weakness scan. Information which is to be collected from the perspective of structure and technology they imply are shown in table 2.

Table 2. Information to be collected about websites

1. Web Servers/ Proxy Servers of determined websites
2. Operating Systems of hosts
3. Working Frameworks
4. Security Equipment they used
5. Web Trackers

For weakness scan popular Owasp ZAP v2.7.0 and Acunetix v12 (https://www.acunetix. com trial version) program which works on Microsoft Windows and Nessus scanner's trial version8.3.1 (https://www.tenable.com) are used. Information to be gathered for vulnerability analysis is:

1. Degree of vulnerabilities present in them
2. Weakness level of websites
3. Weakness evaluation

IV. RESULTS AND DISCUSSIONS

Information gathered after analysing the websites in terms of structure and the technology they use are listed below. The websites are also tested in the perspective of weakness analysis. Results listed below are summarized at the end of the studies.

A. Structural Results

The information of web server or proxy server used by websites, their operating system, the platform they work on, Security equipment they used, location and web trackers in the 50 sites which are predetermined, is summarized below in Tables 3,4,5,6 and 7. It is the known fact that finding out the operating system used in server and web server software can be helpful to information gathering which is the first step of the attack. Thus, Web Application Firewall (WAF) software hinders the information gathering procedures called footprint. Hence, no information was gathered about some websites' operating system and web hosts.

Table 3. Web Servers or Proxy Servers of determined websites

Group	AkamaiGhost	Nginx	Cloud front	Cloud flare	Varnish	Microsoft10.0	Apache
1	5	2	-	1	-	-	1
2	3	4	-	-	-	-	2
3	-	2	3	-	-	-	4
4	-	2	-	3	2	-	-
5	1	2	-	4	-	-	1
Total	9	12	3	8	2	1	8

As seen in Table 3, the percentage of Nginx (current version 1.16.0) choosers as web host software are 24%, and all of them are using old version except one website, one of the sites from Group 2 is still using 1.1.19 version. The outdated versions of Nginx webserver could be a reason for some weaknesses like remote exploit. 9 out of the 50 websites scanned uses AkamaiGHost. The older version of AkamaiGHost allows remote attackers to execute arbitrary code via unspecified vectors. Apache is used by 16% of the websites, and the Cloudflare is used by 16% as proxy server. Another web host server used is Openresty by one website of Group-1, Varnish by two websites of Group-4, ECS Icy by two websites of Group-4, and one website of Group-4 uses GHS and one website of Group- 5 uses Microsoft 10.0.

Table 4. Operating Systems of hosts

Group	Win 2003	Win2012	Linux	Undetected
1	-	-	7	3
2	-	-	9	1
3	1	-	8	1
4	-	-	5	5
5	-	-	3	7
Total	1	-	32	17

As seen in Table 4., the web server uses Linux as Operating system with the percentage of 66, and it has been preferred by news websites the most as 28.12%. A website has been detected which is still using Windows 2003 server for which Microsoft has stopped providing support since July 14, 2015, and it will not provide any new security patch anymore.

Table 5. Working Frameworks

Group	PHP	.NET	J2EE	Adobe enterprise cloud	Undetected
1	6	-	3	2	2
2	6	1	-	-	4
3	7	-	-	1	2
4	5	1	1	2	2
5	3	2	2	-	6
Total	27	4	6	5	16

According to Table 4.3, PHP is the most used framework with 54%. Especially it has been preferred by websites of Group-3 then followed by websites of news and e-commerce. The J2EE framework is used by 6 of the websites. A very few sites have also used ruby on rails token and Akamai bot manager.

Table 6. Security Equipment

Group	Citric Net-Scalar	Cloud flare	AWS WAF	Imperva Secure Sphere	Detected	Un-detected
1	1	1	1	1	1	5
2	-	-	-	-	1	9
3	-	1	-	1	2	6
4	-	2	-	-	3	5
5	-	2	-	2	2	4
Total	1	6	1	4	9	29

As shown in Table 6, 21 (42%) websites use security equipment. It has been seen that one website uses Citrix Netascaler and six websites use Cloudflare's enterprise-class product devices as Web Application Firewall. WAF protects web applications by filtering and monitoring HTTP traffic between a web app and the internet. It protects web apps especially against injection and XSS attacks. One website from group-1 uses AWS WAF, and four websites use Imperva SecureSphere as their WAF. Out of total selected websites, nine websites use WAF which are not disclosed.

Table 7. Web Trackers

Group	CDN	Analysis	Widget	Advertisement
1	3	8	8	13
2	13	24	10	12
3	11	6	10	13
4	6	8	3	-
5	10	11	4	8
Total	43	57	35	46

Web tracker are the means by which websites identify and collect information about the users. It tracks and share location, shopping habits, area of interests, and personal browsing habits of the web users with third parties, implementation is lowest at education websites. All the websites of every group are using them, and they are used for statistic/analysis, CDN (Content Distribution Network), widget, and advertising. Most websites prefer to use analysis web trackers with 57 out of 181 trackers found during scanning.

B. *Results from weakness perspective*

Owasp ZAP, Acunetix, and Nessus programs have been used to find the Vulnerabilities present in these websites. Owasp ZAP and Acunetix, programs categorizes vulnerabilities in four-level. These are at high, medium, low, and informational level. Nessus categorizes vulnerabilities into five levels, which are critical, high, medium, low, and informational. Information level can be ignored. While the high level is critical and must be taken care immediately to prevent the loses, which may cause great damage. In this study, selected Websites were scanned in the computer lab by Owasp ZAP, Acunetix, and Nessus programs from 15th March to 18th April 2019.

a) Evaluation of scan results with Owasp ZAP:

A total of 50 websites were scanned with the Owasp ZAP program. The number of vulnerability information found in the results of scanning with the Owasp ZAP program is shown in fig 3.

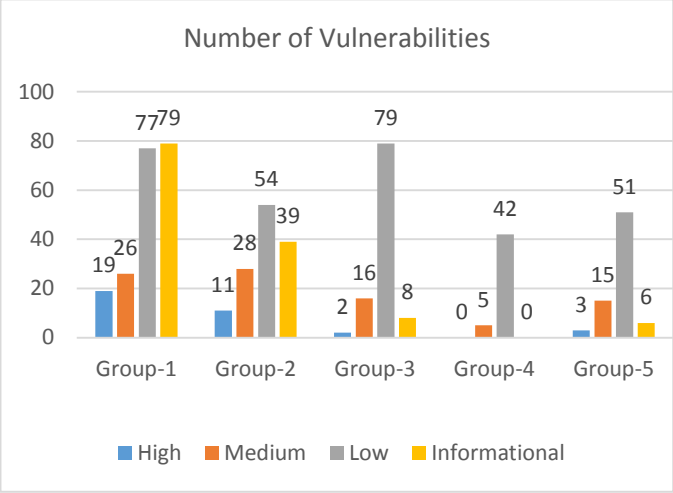


Fig 3. Number of vulnerabilities found by Owasp ZAP

As shown above, the most high-risk vulnerability has been found in Group-1 websites, then followed by Group-2 websites. Medium vulnerability is found most in the Group-2. Low-grade vulnerabilities are found in close proximity to each other in groups. It is found that all the websites have weaknesses when we examine fig 3.

b) Evaluation of scan results with Acunetix:

Determined Websites are scanned using Acunetix V12. The number of vulnerability information found in the results of scanning with the Acunetix program are shown in fig 4.

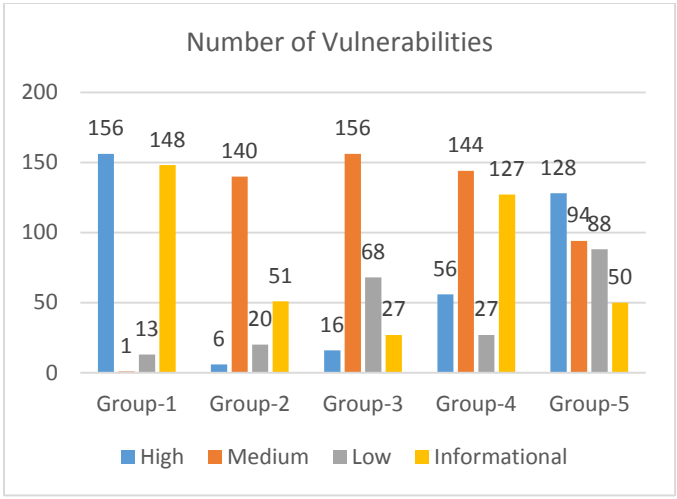


Fig 4. Number of vulnerabilities found by Acunetix

As shown above, the most high-risk vulnerability has been found in Group-1 websites, then followed by Group-5 websites. Medium vulnerability is highest in the Group-3. Low-grade vulnerabilities are lowest in Group-1 and highest in Group-5. It is found that all the websites have weaknesses.

c) Evaluation of Scan results with Nessus:

Determined websites were scanned with Nessus program. The Number of vulnerabilities is shown in fig 5.

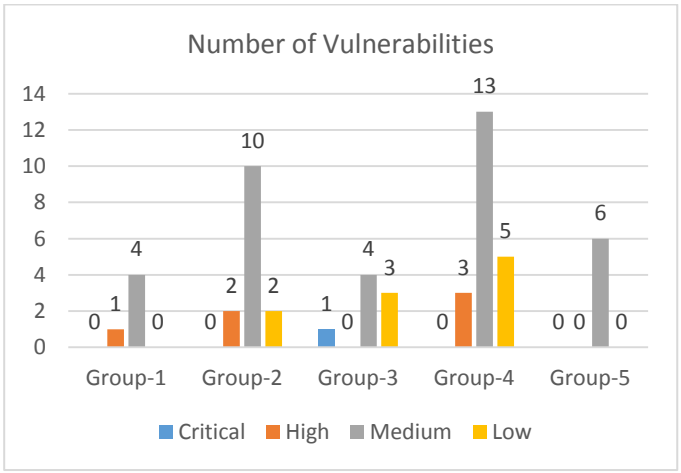


Fig 5. Number of vulnerabilities found by Nessus

As shown in fig 5, only one entertainment site from Group-3 has Critical risk vulnerability. High-risk vulnerability is found in all the groups except Group-3 and Group-5. Medium level risk vulnerabilities are most in Group-4, then followed by Group-2 and Group-5. The low-level threat is at highest in Group-4 and none in Group-1 and Group-5.

The result obtained from the study shows that:

- Linux is the most preferred operating system with 38%.
- For the web server, Nginx software is used by 28% of the websites and AkamaiGHost by 18%.
- PHP is the most preferred with 54% when it comes to the platform used.
- Determined websites are using security equipment with 42%.
- According to Owasp ZAP results obtained, Anti CSRF Tokens Scanner is the high-level threat which is present in most of the websites.
- Reverse tabnabbing, according to Owasp ZAP is the medium level threat which is present in many websites of almost all groups.
- According to Acunetix software scan results "HTML Form without CSRF Protection" is the most common weakness at medium level risk, highest at 96% in Group-3 and 88.46% in the Group-4.
- "Clickjacking: X-Frame-Option Header Missing" is the most common weakness found in low-level risk in all group websites as shown by Acunetix scan results.
- Nessus scan results showed that "Web Application Potentially Vulnerable to Clickjacking" is the medium level risk present in group-2 and group-5.
- For removing considerable number of vulnerabilities from the web applications, they should be tested for penetration testing at regular period of time to determine possible attacks or threats beforehand, to see deficiencies present in them and take precautions to safeguard the web apps.
- Websites which are most visited uses a firewall which is managed by specialists contain small amount of vulnerability in them, and the information that can be available to hackers is less than the sites without having a firewall.
- It has been found that collecting information from websites which imply WAF is problematic. Use of WAF is recommended to avoid gathering of the information required for attackers.
- The differences in the security of the group is a result of their different needs, different business policies followed by them.
- Using ready codes increases weaknesses.
- Precautions against information gathering should be taken by the websites as it is the first step of attacks.

V. CONCLUSION

This research work focuses on generating a template for India's most visited websites. Both in the perspective of technology, they imply and in the perspective of their weaknesses, and sets an example to see their structure and deficiencies. The research work included the use of penetration testing tools to gather the required information. The tools have been used on kali and windows 10. With this study, it is shown what kind of information can be collected on a public domain website and which kind of vulnerability scanning can be done by an ordinary user.

Websites create a great part of today's digital era. Most of the works are now being carried out using websites. With this much use, they also constitute to a great number of security flaws as they are both open to the public and they are time and place independently. With the results obtained it can be concluded that the most visited web sites in India have a considerable number of vulnerabilities. Especially average level weaknesses cannot be ignored.

VI. REFERENCES

[1] P. Fung, *Mitigations of web applications security risks*, hong kong: Ph.D dissertation, 2014.

[2] N. Kochare, S. Chalurkar, B.B. Meshram,, "Web Application Vulnerabilities Detection Techniques Survey," *IJCSNS International Journal of Computer Science and Network Security*, vol. 13, no. 6, p. 7177, 2013.

[3] C. Polat, *Penetration Tests and Security Solutions for Corporate Networks*, Dokuz Eylul University Izmir, 2016, pp. 1-182.

[4] Ruse, M.E, *Model Checking Techniques for Vulnerability Analysis of Web Applications*, Iowa: Iowa State University, 2013.

[5] C. Huang, J. Liu, Y. Fang, Z. Zuo, "A study on Web Security incidents in China by Analyzing Vulnerability disclosure Platforms," *Computer and Security*, vol. 58, pp. 47-62, 2016.

[6] D. Stiawan, M. Idris, A. Abdullah, F. Aljaber and R. Budiarto, "Cyber-Attack Penetration Test and Vulnerability Analysis," *International Journal of Online Engineering*, vol. 13, no. 1, pp. 125-132, 2017.

[7] S. Sandhya, S. Purkayastha, E. Joshua, A. Deep, "Assessment of website security by penetration testing using Wireshark," in *4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2017.

[8] S. Nixon, Y. Haile, "Analyzing vulnerabilities on WLAN security protocols and enhance its security by using pseudo random MAC address," *International Journal of Emerging Trends \& Technology in Computer Science (IJETTCS'2017),* 2017.

[9] J.H. Bullee, L. Montoya, W. Pieters, M. Junger, P. Hartel, "On the anatomy of social engineering attacks—A literature-based dissection of successful attacks," *Journal of investigative psychology and offender profiling*, vol. 15, no. 1, pp. 20-45, 2017.

[10] Y. Wu, G. Feng, R.Y.K Fung, "Comparison of information security decisions under different security and business environments," *Journal of the Operational Research Society*, vol. 69, no. 5, pp. 747-761, 2018.

[11] P. Cisar, S.M. Maravi, I. Furstner, "Security Assessment with Kali Linux," *Banki Kozlemenyekl*, vol. 1, no. 1, pp. 49-52, 2018.

Authors Profile

Irshad Alam completed his B.E in Computer Science & Engineering from Birla Institute of Technology Mesra, Ranchi.Currently he is pursuing M.Tech in Cyber Security from Central University of Punjab. His research work focuses on Penetration Testing, Cyber Securtiy and Network Security.

Dr. Satwinder Singh had completed his Ph.D in 2014 from Guru Nanak Dev University, Amritsar. He is currently working as an Assistant Professor at Department of Computer Science and Technology, Central University of Punjab, Bathinda, Punjab, India, He has 15 years of teaching experience. He has published research papers in reputed journals and conferences. His research interests include Re-engineering of Software System, Maintenance and Fault prediction of Object Oriented Systems, Big data analytics and Text Data Analytics.

Mrs. Gurpreet Kaur had completed his Ph.D in 2016 from Punjabi University, Patiala. She is currently working in Department of Law, Bathinda College of Law, Bathinda, Punjab, India. She has 11 years of teaching experience. She has expertise in Criminal Law, International Law and Fake News.

