IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Question Answering AI System using SQUAD

Mushraf Shaikh, Vedant Sawant, Akshata Bidwe, Arafat Shaikh
K. K. Wagh Polytechnic Nashik, Maharashtra, India
mashaikh@kkwagh.edu.in, vedantsawant208@gmail.com, akshatabidwe06@gmail.com,
arafatshaikh244@gmail.com

Abstract: *In recent years, one needs answer to the question from a huge data on finger tips. Artificial Intelligence Question Answering is about making a computer program that could answer questions in natural language. It can be achieved using SQUAD (Stanford Question Answering Dataset) which will include questions asked by humans from the given comprehension. This paper aims at creation of a system specifically using BERT (Bidirectional Encoder Representations from Transformers) algorithm where user can input a question from the passage of text containing the answer, then span of text corresponding to the text will get highlighted and user will get the most relevant answer. Question answering is at the heart of natural language processing and is composed of two sections: Reading Comprehension and Answer Selection. Question Answering were based on statistical methods and researchers generated set of features based on text input. Answer Selection is a fundamental task in Question Answering, also a tough one because of the complicated semantic relations between questions and answers. Attention is a mechanism that has revolutionized deep learning community. These techniques are widely used among search engines, personal assistant applications on smart phones, voice control systems and a lot more other applications. The BERT Model is superior in all aspects of answering various types of questions.*

Keywords: BERT (Bidirectional Encoder Representations from Transformers), SQUAD (Stanford Question Answering Dataset)

I. INTRODUCTION

As the Internet is growing, most of the end users are not aware of the data security and privacy. Internet is not a secured medium to exchange the information because most of the applications have the vulnerability. Due to vulnerability various kinds of attacks could happen. This attack is due to poor coding and design. To provide security website developer must be aware of the issues related with security and privacy. The three pillars of data security are Confidentiality, Integrity and availability. Confidentiality allows the be safe access to authorized users only ,Integrity allows the data of the end user should be in complete form as well as ensures the authorized users can modify the information .Availability ensures that timely access of information by authorized person i.e. data is available when authorized user wants to access it.
To check the presence of vulnerabilities developer can use two main approaches of software testing [1]:
Web Scanner will consist of following vulnerability detection:

- White-box testing: The source code of the application is examined to keep the track of vulnerable code. It uses control structure to design the test cases and removes the general errors.
- Black-box testing: The source code is not examined directly. Instead, special input test cases are generated and sent to the application. Then, the results returned by the application are examined for unexpected behaviors which indicate the vulnerabilities. Web Scanner will consist of following vulnerability detection:

Working of Web Application:[1]
Figure 1 shows how the web application work for SQL injection which includes client side and server side components.

1. The client side components include static HTML pages with scripting languages
2. In Server side request are processed by web server using dynamic HTML pages through execution part i.e. Java Servlet and the interpreter gives response to the client request. Most web application store the information in databases.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Web Scanner will consist of following vulnerability detection:

1. **Website Defacement:** Website defacement refers to any unauthorized changes made to the appearance of either a single page, or an entire site. In some cases, a website is completely taken down and replaced by something new. In other instances, a hacker may inject code in order to add images, popups, or text to a page that were not previously present.
2. **Malware Detection:** Malware refers to software programs designed to damage or do other unwanted actions on a computer system. Attackers inject javscript codes or server side codes to gather sensitive information or try gain access to private computer systems.
3. **Port Scanning:** It systematically scan a server for open ports. Administrators frequently used this type of scanning to verify security policies of their networks to identify running services on a server.
4. **SQL Injection:** This attack focuses on the database of website or application. SQL injection attacks are caused when attacker inserts a malicious SQL query into database for manipulating it or gaining access of it.

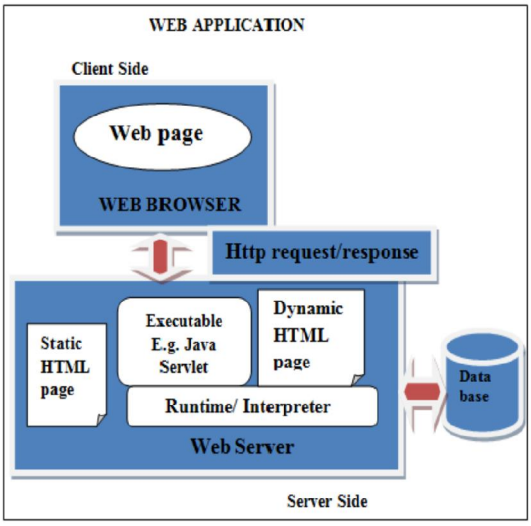

Fig 1: Diagram for Working Of Web Application.

5. **Local File Inclusion (LFI):** LFI is the process of including files on a server through the web browser. This vulnerability occurs when a page include is not properly cleaned, and allows directory traversal characters to be injected.

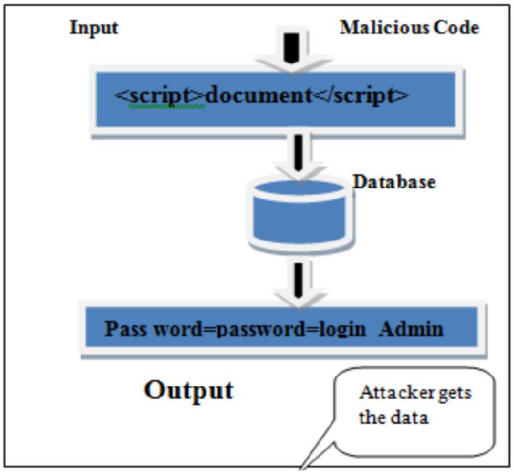

Fig: 2 XSS In Vulnerability

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

6. **Remote File Inclusion (RFI):** RFI type of vulnerability most often found on websites. It allows an attacker to include a remote script file on the web server. The vulnerability occurs due to the use of user-supplied input without proper validation.
7. **Cross Site Scripting (XSS):** XSS is very similar to SQL-Injection. In SQL-Injection by injecting SQL Queries as user inputs the vulnerability can be exploited. In XSS, we inject code (client side scripting) to the remote server.

Aim of this paper is to review results of current website scanning tools, their limitations, and future of the research on web application scanning methods.

II. LITERATURE SURVEY

Lots of research has been done on Web Security Scanner. Manual vulnerability checking of all web applications is very difficult which includes processing a large volume of data. Common Vulnerabilities and Exposures database [2]. Automated web application scanner can scan your website, identify all the files accessible from the internet [3]. It requires the depth skill and ability to keep track of large volumes of code used in a web application. Hackers are continuously searching new methods to exploit your web-vulnerability statistics [4] Internet, but still exist in the web application, and can thus still be exploited [5]. Acunetix's engineers have developed tools for Web site analysis and vulnerability detection, application minimizes risk by identifying vulnerabilities in the network so the user can protect the network before an attack occurs.,AppDetective discovers database applications within an infrastructure and assesses their security strength, IA² automatically reports on the Defense Information Systems Agency's (DISA) Security Readiness Review, third party vulnerability scanner results, and DISA's Security Checklists, BigFix provides organizations with the ability to assess their managed systems against Open Vulnerability Assessment Language (OVAL)-based vulnerability definitions, Computer Oracle and Password (COPS) is a vulnerability toolkit that examines a system for a number of known weaknesses, CORE IMPACT Pro is a comprehensive software solution for assessing the security of network systems, endpoint systems, email users, and Web applications[6].

SecuBat vulnerability scanner consists of three main components[8] :

First, the crawling component collects a set of target web pages. Then, the attack component presents the configured attacks against these targets. Finally, the analysis component determine whether an attack was successful by examining the results returned by the web applications

- **Crawling Component:** By using a root web address as a starting point, the crawler collect all the pages and steps down the link tree and included web forms. Similar to a typical web crawler, SecuBat has configurable options for the maximum number of pages per domain to crawl, maximum link depth, maximum crawling time, and the option of dropping external links. Conceptual ideas is taken from existing ststem for the implementation of this component, especially from Ken Moody's and Marco Palomino's SharpSpider [9], and David Cruwys' spider [10].
- **Attack Component:** In attack component each page of web forms is scanned because the fields of web forms constitute the entry points to web applications. For each web form, they extract the method (i.e., GET or POST) and the action address which are used to submit the form content. Also, the form fields as well as its corresponding CGI parameters are gathered . Then, depending on the actual, appropriate values for the form fields are selected. Finally, the form content is uploaded to the server specified by the action address (using either a GET or POST request). As defined in the HTTP protocol [11], In response to web request the attacked server responds by sending back a response page .
- **Analysis Modules:** The analysis module parse and interpret the server response after launching the attack. To check the attack was successful an analysis module calculates a confidence. As number of websites are scanned a care needs to be taken in calculation so that false positives are reduced.

WebCruiser - Web Vulnerability Scanner, a compact but powerful web security scanning tool. It has a Crawler and a Vulnerability Scanner (SQL Injection, Cross Site Scripting). It can support scanning website as well as POC (Proof of concept) for web vulnerabilities: SQL Injection, Cross Site Scripting, Local File Inclusion, Remote File Inclusion, Redirect etc. More popular input validation attacks includes SQL injection and cross side scripting (XXS)[12].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

III. EXISTING SYSTEM

As day by day the popularity of the internet increases and web applications become tools of everyday use, the role of web security is also increasing .If we observe the past history there is a significant increase in the number of attacks on website. For example, security incidences like the loss of sensitive credit card information, breaching the security policies etc. Drawback of Existing web vulnerability scanners are as given below:

1. Their limitation is that their false negative and false positive rates are higher as compared to the skilled humans doing the similar type of task.
2. Authentication. Network-based vulnerability scanners are not a perfect tools . The reason is that they detect only vulnerabilities for which they have signatures even though they are properly configured,. While anonymous (unidentified) scanning can provide some benefit, scanner will reduce the its effectiveness if they fail to authenticated scanning
3. Scanner's not able to work with custom applications. CVE-based vulnerabilities are only a small subset of most organizations' overall attack surfaces. For most of the popular applications security checks may exist within your network, but what about the in-house applications or outsourced to third parties? There are no CVEs for custom apps.
4. Most web vulnerability scanning tools can identify points of vulnerabilities. While vulnerability scanners typically identify and report on issues that can be consumed as the starting point of access(entry), they are limited in identifying the complex avenues an attacker could take to compromise your network.

The advantages are:

1. Vulnerability scanner allows early detection and handling of known security problems.
2. A vulnerability scanner used to verify the assessment of the complete website. The assessment like website language platform, operating system version, plugins, open TCP-UDP ports, all links including hidden links other relevant system information. This information is useful for the security management.
3. The big advantage is that they can cover large suites of known vulnerabilities and crawl entire websites much, much faster than humans can.

Vulnerability scanners are used to scan the web applications and/or the software applications.[13].
Web Scanning can be of two types:

a) **Passive Scanning:** Passive scanning, determine whether a tool can enlist the vulnerabilities by considering the existing network.
b) **Active Scanning:** Active scanning determine whether the queries can be made to the network for the vulnerability.

There are different categories of web scanner are:

a) **Port Scanners:** Port scanners are used to scan the ports to detect the open and closed ports, operating system, services offered.
b) **Application Scanners:** Application scanners keep the track of weaknesses by assessing a specific application on the network.
c) **Vulnerability Scanners:** Vulnerability scanners find out the vulnerabilities in the system which if accessed by a unauthorized user or hacker can put the whole network system at risk

IV. ACKNOWLEDGEMENT

I would like to show my sincere gratitude towards Prof. Pramod Gosavi, HOD, Department of Information Technology, Godavari College of Engineering, Jalgaon for his valuable guidance and encouragement.

V. CONCLUSION

Web application has got tremendous changes in the past few years. Many new technologies of Web applications have been emerged in the market. The new technology provides users with a variety of Internet applications, but at the same time it introduces new security problems. The paper focuses on surveys about different web application scanners. A Many scanners detects as many number of vulnerabilities as possible.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

REFERENCES

[1]. Carlo Ghezzi, Mehdi Jazayeri, and Dino Mandrioli. Fundamentals of Software Engineering. Prentice-Hall International, 1994.
[2]. Common Vulnerabilities and Exposures. [Online]. Available: http://cve.mitre.org.
[3]. Web Application Security Scanner Evaluation Criteria. Web Application Security Consortium. [Online]. Available: http://projects.webappsec.org/Web-Application-Security-Scanner-Evaluation-Criteria.
[4]. Web Application Security Statistics. Web Application Security Consortium. [Online]. Available: http://projects. Webappsec.org/Web-Application-Security-Statistics.
[5]. Software Assurance Tools: Web Application Security Scanner Functional Specification, National Institute of Standards and Technology Std., Rev. 1.0.
[6]. Information Assurance Tools Report
[7]. Prevention from hacking attacks: Phishing Detection Using Associative Classification Data Mining
[8]. Stefan Kals, Engin Kirda, Christopher Kruegel, and Nenad Jovanovic "SecuBat: A Web Vulnerability Scanner"
[9]. Ken Moody and Marco Palomino. SharpSpider: Spidering the Web through Web Services. First Latin American Web Congress (LA-WEB 2003), 2003.
[10]. David Cruwys. C Sharp/VB - Automated WebSpider / WebRobot. http: // www.codeproject.com/csharp /DavWebSpider.asp,March 2004.
[11]. W3C World Wide Web Consortium. HTTP - Hypertext Transfer Protocol.http://www.w3.org/Protocols/, 2000.
[12]. Mr. K.Naveen Durai1, K.Priyadharsini "A Survey on Security Properties and Web Application Scanner
[13]. Sheetal Bairwa, Bhawna Mewara and Jyoti Gajrani "Vulnerability Scanners: A Proactive Approach To Assess Web Application Security".