Securing Web Applications from malware attacks using hybrid feature extraction

Subramaniyaswamy V.*, Gopireddy Venkata Kalyani, Naladala Likhitha

School of computing, SAstra Deemed to be University, Thanjavur, Tamil Nadu

*Corresponding Author

Abstract:
In this technological era, many of the applications are taking the utilization of services of internet in order to cater to the needs of its users. With the rise in number of internet users, there's a substantial inflation within the internet attacks. Because of this hike, Web Services give rise to new security threats. One among the major concerns is the susceptibility of the internet services for cross site scripting (XSS). More than three fourths of the malicious attacks are contributed by XSS. This article primarily focuses on detection and exploiting XSS vulnerabilities. Generally, improper sanitization of input results in these type of susceptibilities. This article primarily focuses on fuzzing, and brute forcing parameters for XSS vulnerability. In addition, we've mentioned the planned framework for contradicting XSS vulnerability.

Keywords: Cross Site Scripting attacks, WAF detection, web application security, fuzz testing.

1. Introduction:
Cross Site Scripting (XSS) is a completely, a generally exploited vulnerability which could be very extensively unfold and easily detectable. These days it is one of the unusual software stage attacks that hackers use to sneak into web packages. This results in compromise of privateness of clients of a selected net site that can totally breach the safety where customer details are stolen or manipulated. These days, net applications have come to be an essential part of our existence and culture. Almost half of all websites have high protection vulnerabilities. Cross site scripting is one such predominant attack [11-20]. It is a manner of injecting malicious JavaScript code to the trusted and legitimate websites at client side. This snippet of malicious JavaScript is then achieved by way of the sufferer who is journeying the goal site and consequently the net application is attacked even without the knowledge of users [21-26]. While a user go to the infected or a mainly-crafted hyperlink, it will execute the malicious JavaScript. An XSS vulnerability will allow attackers to do phishing assaults, session information hijacking, theft of cookies, and web application will function abnormally [27-32]. The web browser takes the facts which are not trustworthy without any proper validation and sanitization [4] and thus the XSS attacks arises. So in XSS assaults three events are worried- the attacker, the consumer and the website. After this assault arises, the web server can no longer guarantees that produced pages are well encoded to prevent the unintentional execution of scripts.
Figure 1. Possibility of a website having a vulnerability by class

Figure 1 explains the percentage of possibility of type of attacks. The statistics show that almost 65% of total attacks are contributed by XSS where as 47% percentage is contributed by Information leakage and 30% by content spoofing. Authorization, SQL injection, Resource Location combined together contributes to only 50 %

1.1 Steps in Exploiting the Vulnerability:
- A payload is built suitably by fuzzing a parameter.
- The parameters are brute forced with the payloads.
- The commands of a WAF/Filter are reverse engineered.
- The framework detects the presence of WAF depending on the error code.
- Using filter Checker, Reflected XSS vulnerability can be determined.
- Using the payloads crafted, Blind XSS vulnerability can be determined.
- Opens the Proof of Concept (POC) in a browser window.

1.2 Various kinds of XSS vulnerabilities
XSS vulnerability is classified as:
Attacker finds a website with vulnerability to inject script

Attacker injects script to steal

User’s cookie is sent to attacker

Website visitor

Malicious script is activated, each time the website is visited

Attacker finds a website with vulnerability to inject script

---

**Figure 2. Stored XSS attacks**

Figure 2 illustrates the Stored XSS attacks. If the targeted servers permanently stored the injected scripts in the form of database or message forum, visitor log, comment field, it is classified as Stored XSS attacks. As shown in Figure 2, the stored information is requested and then the malicious script is retrieved by the victim. This kind of attack is often classified as Persistent attack and also known as Type-I XSS.

**Reflected XSS Attacks:**
Figure 3 describes Reflected XSS attacks. If the script that is injected is reflected off the web server, for example in an error message, search result, or any other response which includes some or all of the input sent to the server as part of the request, it is classified as Reflected XSS. Figure 3 depicts these type of attacks, where attacks are delivered to victims via another form, likely in an email message, or on some other web site. When a user is tricked into clicking on a malicious link, submitting a specially crafted form, or even just browsing to a malicious site, the injected code travels to the vulnerable web site, which reflects the attack back to the user’s browser. The browser then executes the code assuming that it originated from a “trusted” server. This is often classified as Non-Persistent and is also known Type-II XSS.

DOM Based XSS:
DOM Based XSS attack involves no HTTP request. Modifying the Document Object Model of the target site in the user side code in the victim’s browser results in injection of script and is the malicious code is then executed.

2. Related Work
guard against XSS assaults. As an independent component or with generally utilizes plans like filtering, their approach can give a powerful resistance against XSS assaults. Abdalla Wasef Marashdih [5] concentrated on the methodologies used to wipe out XSS vulnerability from the source program. There are two methods that leads to the disposal phase of XSS on web applications based on Java. Along these lines, it closed saying that more examination is required in the field of weakness points from the source program of the applications. Since PHP is the most broadly utilized web innovation, the scientists needed to focus on including an elimination phase of cross site scripting in web applications that are built using PHP. Malviya [9] proposed an examination to solidify the comprehension of XSS, their cause and appearance, sorts of risks and alleviation endeavours of XSS. Bates [6] proposed an enhanced outline for a client side XSS filter. This configuration accomplishes elite and high loyalty by mediating on the interface between the program's HTML parser and JavaScript engine. This execution is implemented as default in Google Chrome. Mishra [3] has discovered that security in web applications is frequently broken from users’ information. The sort of assaults that web application is vulnerable incorporates SQL Injection, Cross Site Scripting (XSS) and Denial-of-Service (DoS). With a specific end goal to keep these attacks, both ASP.NET and PHP advancements have rich capacities and libraries that are equipped for sifting users inputs against different parameters. Shar, L.K. [8] proposed properties that are related to hybrid and dynamic code examination, which describe input validation and cleansing code patterns for anticipating SQL infusion and XSS vulnerabilities. Martin Johns [4] depicted XSSDS a server-side Cross-website Scripting identification system, which utilizes two novel recognition approaches that depend on bland perceptions of XSS assaults and web applications. A prototypical usage showed that this current approach's abilities to dependably distinguish XSS attacks while keeping up a mediocre false positive rate. Gupta M.K [2] proposed an order of software security approaches used to create secure programming in different period of software development life cycle and furthermore compressed different static examination approaches that identify vulnerabilities in coding due of SDL.

2.1 Challenges faced in web services:
The definition for Security is the system’s quality which guarantees the absence of manipulation or unauthorized access. The protection threats turn up because of exploitation of vulnerabilities, throughout development of the system. There are many reasons for such vulnerabilities, in which one can allude to the complexity of systems. The key challenges are [1]:

- Highly secured Input validation mechanisms are employed in the web applications.
- The web applications belonged to HTML5 are lacked in XSS defensive frameworks
- Absence of context-sensitive cleaning within the existing XSS sanitization-based outputs
- High rate of false positives are encountered.

2.2 Working of XSS attack:
There is no limit for XSS attacks. In XSS attack, malicious script will be sent to a user by an assailter. The browser of end user is unaware that the script is not a trusted one and hopes that the script is from the
source which is a trusted one. Then, it will execute the script which is harmful. When the malicious script is executed by the browser, the attacker can access cookies, session tokens, the victim will be redirected to some other web pages which will be controlled by the attacker or other delicate information that is held by the browser. The content of the HTML page can even be rewritten by these scripts. Cross-Site Scripting (XSS) attacks arise when:

- Data is intruded into an online application from the sources which are not safe.
- The data is fringed into the dynamic content and is sent to the web user even before it is checked for the presence of any content which is malignant.

### List of escape codes [1]

<table>
<thead>
<tr>
<th>Display</th>
<th>Hexadecimal code</th>
<th>Numerical code</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;</td>
<td>&amp;#x22;</td>
<td>&amp;#34;</td>
</tr>
<tr>
<td>#</td>
<td>&amp;#x23;</td>
<td>&amp;#35;</td>
</tr>
<tr>
<td>&amp;</td>
<td>&amp;#x26;</td>
<td>&amp;#38;</td>
</tr>
<tr>
<td>'</td>
<td>&amp;#x27;</td>
<td>&amp;#39;</td>
</tr>
<tr>
<td>(</td>
<td>&amp;#x28;</td>
<td>&amp;#40;</td>
</tr>
<tr>
<td>)</td>
<td>&amp;#x29;</td>
<td>&amp;#41;</td>
</tr>
<tr>
<td>/</td>
<td>&amp;#x2F;</td>
<td>&amp;#47;</td>
</tr>
<tr>
<td>;</td>
<td>&amp;#x3B;</td>
<td>&amp;#59;</td>
</tr>
<tr>
<td>&lt;</td>
<td>&amp;#x3C;</td>
<td>&amp;#60;</td>
</tr>
<tr>
<td>&gt;</td>
<td>&amp;#x3E;</td>
<td>&amp;#62;</td>
</tr>
</tbody>
</table>

### 2.3 Determining if the web application is vulnerable:

To dispose the XSS blemishes can be troublesome. The most ideal approach to discover blemishes is to play out a security survey of the code and look for all spots where contribution from a HTTP ask for could advance into the HTML yield. Note that a wide range of HTML labels can be utilized to transmit a noxious JavaScript. Nessus, Nikto, and some other accessible apparatuses can help examine a site for these imperfections, yet can just begin to expose what's underneath. In the event that one a player in a site is helpless, there is a high probability that there are different issues too.

It's pivotal that you kill HTTP TRACE bolster on all web servers. An aggressor can take treat information through JavaScript notwithstanding when document. Cookie is handicapped or not upheld on the customer. This assault is mounted when a client presents a noxious content on a gathering so when another client taps the connection, a non-concurrent HTTP Trace call is activated which gathers the client's treat data from the server, and after that sends it over to another malevolent server that gathers the treat data so the assailant can mount a session seize assault. This is effortlessly alleviated by evacuating support for HTTP TRACE on all web servers.
4. Proposed Work

Figure 4 depicts the overall flowchart of the proposed work as per algorithm 4.1. The framework starts with detection of presence of web Application Firewall as depicted in algorithm 4.2. To check for reflected XSS, filter checker is proposed as shown in algorithm 4.4. In order to check Blind XSS, payload generation is used as depicted in algorithm 4.7. After suitable payloads are injected as in algorithm 4.5, Browser with the attacked site is displayed.
### 4.1 Algorithm: Detection and Exploitation

**Input**: URL

**Output**: Browser displaying the site with the given URL with a particular vulnerability exploited

1. **Start**
2. Initialize an array with a list of sanitized XSS attack payloads.
3. Target←URL
4. param_parser(Target, parameter_data, GET, POST)
5. if “=” in target,
   1. GET←True, Post←False
   2. parameter_data←NULL;
   3. parameter_parser(target,parameter,GET,POST)
   4. Initiator(target,GET,POST)
6. Else
   1. GET←False, POST←True
   2. parameter_data←post data
   3. parameter_parser(target,parameter,GET,POST)
   4. Initiator(target,GET,POST)

**End if**
7. Check for the status of web Application Firewall
8. Check for Reflected XSS
9. Detection of Blind XSS
10. Insert payloads and crafts the URL
11. Display the vulnerable site with the particular vulnerability exploited
12. **End**
4.2 Algorithm: Web Application Firewall Detector

Start
Initialize fuzz with XSS checker with noise containing alert message
noise ← quote_plus("<script>alert()</script>"

If GET in the URL ∈ web application
response ← br.open(url+fuzz)
Else
response ← br.open(url+fuzz)
Print 'web application' firewall is offline
End if

For each http response code in error
If “406” or “501” in error msg,
WAFName ← “Mod-Security”
else if “999” in error msg,
WAFName ← “WebKnight”
else if “419” in error msg,
WAFName ← “F5 BIG IP”
else if “403” in error msg,
WAFName ← “UnKnown”
else
print “web application firewall is offline”
End if
End for each

Web application firewall detector: As depicted in algorithm 4.2, the web servers are verified for the presence of firewall for its security. Here, the fuzz is initialized with an alert message which is enclosed in a script tag and sent as an input parameter. The URL of the given web application is then checked for any presence of GET method. If it is present, the browser will simply display the given alert message as a response. If there is no such GET method, the status of the firewall will be ‘offline’. Later, the errors which will be displayed as http response is examined. If the error message is as follows, the corresponding firewall names are given:

<table>
<thead>
<tr>
<th>Error message</th>
<th>WAFName</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. “406” or “501”</td>
<td>Mod-Security</td>
</tr>
<tr>
<td>2. “999”</td>
<td>WebKnight</td>
</tr>
<tr>
<td>3. “419”</td>
<td>F5 BIG IP</td>
</tr>
<tr>
<td>4. “403”</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

If there are no any such http error responses, the firewall will be considered as ‘offline’ and that particular application is more susceptible to XSS attacks.
4.3 Algorithm: Fuzz Testing

Initialize the fuzzer array with fuzz vectors for each type
Fuzzer <-- fuzz vectors
For each vector in the fuzzer array
   Quote each vector in the fuzzer vector
   Data_to_be_injected <-- parameter. Replace(XSS Checker, fuzz)
   For each vector in the fuzzer array,
      append the vector with the URL
   End For each
End For each

Fuzz Testing: Algorithm 4.3 explains the payloads in the form of 'fuzz vectors' are initialized to a 'fuzzer array'. quote_plus() quotes the values, which means spaces are quoted as a '+' character and '/' characters are encoded as %2F, which follows the standard for GET requests. Data to be injected is the XSS Checker being replaced with fuzz already crafted from the array. This parameter value is appended with the URL and fuzz testing is performed.

4.4 Algorithm: Filter Checker

Start
1. strength <-- NULL
2. lstring <-- parameter. Replace(XSS Checker, quote_plus('"<svg/onload=(confirm())">'))
   If '"<svg/onload=(confirm())>"' in lstring
      filter_strength <-- low
   End if
3. mstring <-- parameter. Replace(XSS Checker, quote_plus('"<zz/on xx=yy>"'))
   If '"<zz on xx=yy>"' in mstring
      filter_strength <-- medium
      Else
         filter_strength <-- high
   End if
Return filter_strength
End
Filter Checker: This is implemented as described in algorithm 4.4 in order to check if any filter is present. Initially, strength is initialized to NULL. Replace XSSChecker with ‘<svg/onload=(confirm())>’ using quote_plus and initialized to lstring and if it is present in lstring, the filter_strength is ‘low’. Also replace XSS Checker with ‘<zz/on xx=yy>’ using quote_plus and initialized to mstring. If if it is present in mstring, the filter_strength is ‘medium’ else it is ‘high’.

### 4.5 Algorithm: Injection for Reflected XSS

Start
For each occurrence in izip(occurrence number, occurrence location)
  if payload_to_check and payload_to_compare matches with (" ")
    Print “Double quotes (") are allowed”
  Else
    Print “Double quotes (") are not allowed”
  End if
  if payload_to_check and payload_to_compare matches with (" ")
    Print “single quotes (”) are allowed”
  Else
    Print “single quotes (”) are not allowed”
  End if
  if payload_to_check and payload_to_compare matches with (" <> ")
    Print “angular brackets (<>) are allowed”
  Else
    Print “angular brackets (<>) are not allowed”
  End if
End For each
End

Algorithm 4.5 describes the Injection for Reflected XSS. Occurrence number is no. of times, a particular payload is occurred
If payload_to_check and payload_to_compare matches with (" ")
  Then double quotes (") are allowed else not allowed.
  ➢ If payload_to_check and payload_to_compare matches with (" ")
    Then single quotes (’) are allowed else not allowed.
  ➢ If payload_to_check and payload_to_compare matches with (" <> ")
    Then angular brackets (<>) are allowed else not allowed.
4.6 Algorithm: DOM Tree Generation

**Input:** set of HTTP response

**Output:** Generated DOM tree of each requested web page

Start
Stack <-- -1
Node <-- Null
Tag <-- Null
For each http response R
  W <-- webpage
  T <-- extract_tag(W)
  Stack.push(T)
  Node.add_node(T,NULL)
  H <-- T ∪ Tag
  while( Stack! = " ")
    loc <-- extract_tag(W)
    If(opening_tag(loc))then
      A <-- stack.size()
      B <-- DFS(Node,A)
      Node.add_node(loc,B)
      Stack.push(loc)
      T <-- loc ∪ T
    Else If(closing_tag(loc)) then
      Stack.pop()
    End If
  End While
End For each
Return node
End

**DOM Tree Generation:** Each removed module of the web application is sent to this part. It is in charge of the development of DOM tree for the got module by actualizing the calculation.

The working procedure of the calculation is depicted as takes after:
At first, it separates HTTP reaction website page as W. Right off the bat, it finds first HTML tag and stores it in T and furthermore pushes it into the stack. At that point, T is added to the unfilled DOM tree as its root hub. All the distinguished labels are put away into the archive tag and in addition stretched out to the Stack.
4.7 Payloads generated for comparison:

Payloads generated for Blind XSS:

1. `<img src=l onerror=(prompt)() x>`
2. `<! -- <img src = -- <img src=x onerror=(prompt)`
3. `<details open on toggl e=confi/ua0072m()>
4. `<A id=x href=javascript;co/lu006efiru006d1>&quot;1&quot;)&gt;Click</A>`
5. `<img src=x onerror=(confirm)() x>`
6. `<script x=confirm` `onerror=eu0076al(src)>
7. `<script x=confirm` `
8. `<svg onload=(confirm)()>
9. `<script src=//14.rs>`
10. `<img src=x:confirm`` onerror=eval(src)>
11. `<script x=confirm` 
12. `<script x=confirm`` onload(=confirm)()>
13. `<y on mou sedown=((alert))()&gt;ClickHere!`
14. `<a href=javascript&colon;co/lu006efiru006d1>&quot;1&quot;)&gt;Clickme</a>`
15. `<img src=x onerror=co/lu006efiru006d1` `>`
16. `<svg onload=co/lu006efiru006d1` `>`

4.8 Event Handlers and their corresponding tags used:

<table>
<thead>
<tr>
<th>Event Handlers</th>
<th>JavaScript functions used for popup</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>onerror</code>: ['object', 'img', 'video']</td>
<td>confirm()</td>
</tr>
<tr>
<td><code>onload</code>: ['svg', 'body']</td>
<td>prompt()</td>
</tr>
<tr>
<td><code>onstart</code>: ['marquee']</td>
<td>find(confirm)</td>
</tr>
<tr>
<td><code>onmouseover</code>: ['d3v', 'a', 'iframe', 'body']</td>
<td></td>
</tr>
<tr>
<td><code>onfocus</code>: ['d3v', 'body']</td>
<td></td>
</tr>
<tr>
<td><code>onclick</code>: ['d3v', 'body']</td>
<td></td>
</tr>
<tr>
<td><code>onToggLe</code>: ['deTails']</td>
<td></td>
</tr>
</tbody>
</table>

5. Usage statistics for Web Application:

A web application utilizes program and web advances to execute assignments through a system with the assistance of a web program [5]. Dissimilar to work area programming that is started by a working framework, sites ought to be opened through a web program. A web program utilizes the web server to interface with the instruments associated with the system. The fundamental program of the web is put
away on a web server, where all their code and data are put away. In this manner, end clients don't require investing extra energy in introducing programming on their hard drives. A portion of the outstanding advancements that encourage programming engineers to make progressively created site pages are ASP.NET, PHP, JavaserverPages (JSP), PERL, and Python [3].

### Usage statistics

![Usage statistics chart](image)

6. **Experimental Results:**

<table>
<thead>
<tr>
<th>Web Applications</th>
<th>Vulnerability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. dramaonline.pk/search.php?q=ok</td>
<td>No of reflections found: 8</td>
</tr>
<tr>
<td>2. <a href="http://www.occidentalleather.com/search.php?Q=d3v$E=1&amp;X=0">www.occidentalleather.com/search.php?Q=d3v$E=1&amp;X=0</a></td>
<td>WAF Detected: Mod_Security</td>
</tr>
<tr>
<td>6. alphaonenow.org/info.php?id=131</td>
<td>No of reflections: 4</td>
</tr>
<tr>
<td>8. <a href="http://www.f10products.co.za/index.php?id=5">www.f10products.co.za/index.php?id=5</a></td>
<td>100% efficient payload is found</td>
</tr>
</tbody>
</table>
The above results show that web applications having reflected XSS has been found using filter checker are found and number of reflections are also shown. First any web application checks for Application firewall. All the above results has their application firewall offline except for the 2nd example. Once reflected XSS is found to be not present, the framework checks for Blind XSS using payload generation. Payloads that are possible injections are given in the table above.

7. Conclusion:
Since technology is increasing, there is an increase in need for securing web Applications. Due to the hike in Internet Users the web services results in giving new challenges. One of major challenge is to secure against malicious attacks. Among which XSS is the major contributions. So this framework focuses on detecting XSS vulnerability of all flavors. First WAF status is detected based on the error code. Then filter checker is applied to determine the presence of Reflected XSS vulnerability and Payloads are generated to check for Blind XSS. The experimental results show that there lot of sites of vulnerable to XSS and the type of vulnerability is also mentioned.

References:


