IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FILE TRACKING SYSTEM USING XSS AND XXE INJECTION

Prince Patel[1]
UG Student, St. Peter's Engineering College,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

G.Sai Raghav[2]
UG Student, St. Peter's Engineering College,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

Bala Pavan Manikantha.K[3]
UG Student, St. Peter's Engineering College,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

Dr.K.Satish Kumar[4]
Professor, St. Peter's Engineering College,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

***Abstract*— Cyberattack is increasing very rapidly, as most of the information in recent times are transferred through online mode using different forms of files like pdf, document files, excel sheets. It is almost impossible to protect it completely when we are sharing with someone over the internet like e-mail, social media, cloud storage, etc. There is no guarantee that it will be received to the receiver with no breaches within the middle and you may not be able to know if such breaches had leaked your records and inside your files, these leaked records may cause disaster in the future, the file tracking system is a tool that it will help in early detection of such breaches are easily detected when accessed by an attacker thus the owner of the file can send their files to anyone without any fear and if there are any data breaches in the middle he/she may be able to get the information of the attacker who had breached it, this may assist the owner of the file to take early precautions before something wrong occurs.**

Keywords*— cyberattack, XSS, XEE, token injection, tracking, data breach

I.INTRODUCTION

File Tracking System is a Tool makes files such as word document, pdf files trackable such that if there was any attempt made by the attacker to steal this data, the user will be notified through email with the details of the attacker which includes location and IP address of the place where this incident occurred, so that the user preventive steps before anything wrong happens from that data leak.

This type of attack is a chief reason for statistics breaches. A records breach is a security violation in which touchy, included or exclusive records is copied, transmitted, considered, stolen or used by a character unauthorized to accomplish that."[1] records breaches may additionally involve financial facts together with credit score card & debit card details, bank details, private fitness statistics (phi), in my view identifiable records (PII), trade secrets and techniques of organizations, or intellectual belongings. Maximum information breaches involve overexposed and prone unstructured information – files, documents, and sensitive facts.[2]an example showing the common cost of facts breaches is proven in determine 1.

According to the non-profit patron corporation privacy rights clearinghouse, a total of 227,052,199 person data containing touchy non-public statistics had been involved in safety breaches within the u. S. A. Among January 2005 and can 2008, excluding incidents in which touchy records became now not actually uncovered.[3]

File tracking System currently supports Word Document Files(.docx) and Portable Document Format files(.pdf).

This tracking is done by embedding tokens(honeytokens) into the file by using approaches such as XSS payload injection and external XML Entity injection in those files, and these tokens will generate an alert when the file is accessed.

Honey Tokens are unique tokens that are injected into fake documents to trap the attacker and when unknown access has occurred these tokens Detect and alert the Tracking.[4]

II. IMPLEMENTATION

This project is implemented as a web application made with different popular technology stacks. Brief information about all the technologies used is Python, Flask, Front End Technologies(HTML, CSS, JavaScript), MYSQL(Database)..

A. Python:

Python is a high-level Programming language and is interpreted. Python's layout philosophy emphasizes code readability with its wonderful use of large indentation. Its language constructs in addition to its object-oriented technique intended to help programmers to put in writing smooth, logical code for small and big scale tasks.[5].python strives for a less complicated, much less cluttered syntax and grammar at the same time as giving developers a preference of their coding technique. Inside the assessment of Perl's "there may be a couple of ways to do it" motto, python embraces a "there want to be one— and preferably best one —obvious way to do it" layout philosophy.[6]

Python3.6 is the version of python which is currently used to run our application. You can check the version of your python by running “python –version” in the command prompt. For more detail about python3.6 Reference check the official Documentation released by docs.python.org.[7]

The minimum system requirements to run this version of python is an OS with windows 7/10 or MAC OS or Linux installed with x86 64-Bit CPU architecture(Intel/AMD) having 4GB RAM and 5GB Free Disk Space.[8]

Our software is currently built on Linux Ubuntu 64bit based System python 3.6 version, so during running on Windows / MAC OS based system there will be some change in dependencies to be installed depending on the OS it will be running on.

B. Flask:

Flask is a lightweight web Framework built on python[13]. It is categorized as a microframework as it does not require unique tools or libraries. It has no database abstraction layer, shape validation, or different components where pre-existing 1/3-celebration libraries provide commonplace skills. However, flask enables extensions that could add software talents as though they were implemented in flask itself. Extensions exist for object-relational mappers, shape validation, add handling, various open authentication eras, and numerous common framework-associated equipment.[9]

Some major components of flask are Werkzeug, jinja , MarkupSafe, itsDangerous..

Werkzeug is a flask library to achieve Web Server Gateway Interface and it is licensed under BSD License. It supports all python 3.5 and later versions.

Jinja is a template engine for python that handles templating inside HTML files using jinja syntax which is similar to python, it is also licensed under the BSD License.

Markupsafe is string copying with the library for python programming language which is also licensed beneath BSD. The eponymous markup-safe extends the python string type and marks the contents as “secure” combining markup-safe with everyday strings routinely.

ItsDangerous is a secure information serialization library for the python programming language, licensed beneath a BSD license. It's far used to keep the consultation of a flask software in a cookie without allowing users to tamper with the consultation contents

C. MySQL:

MySQL is an open-source Relational Database Model System(RDBMS) that follows Structured query language(SQL) as its syntax and is licensed under GNU, it is publicly available to everyone. MySQL is used by around 80% of existing business applications including big companies like Facebook, NetFlix, Google, Amazon as its Database System for storing most of their data.[10]

D. Front-End:

Front-End technology is very important as it is the user interface which actually a user will see on the client-side, technologies used as front-end are given below.[14]

- HTML is the backbone of websites, it acts as the skeleton of any web application the latest version of HTML is HTML5 published in October 2014 by W3 recommendations.
- CSS controls the style of the page which makes the components of the website look clear and easily accessible to the user without any problem on any device.
- Javascript enables event-driven tasks and makes the page dynamic, to improve the user experience when used various events like a mouse click, scroll, etc.

The goal of frontend technology is to provide easy access to the tool without any problem.[11]

III. WORKING MODEL

This software is currently deployed as a web application so it becomes easier for everyone to access with any device which has browser support.

These web application working steps are given below:

1. User will open the website and generate a token. (Tokens consist of a unique identifier that can be embedded in a file. Whenever that file is accessed, we send a notification email to the address tied to the token.)
2. The token embedded file download link will be generated for the User.
3. When someone tries to access this token embedded file, he will be tracked and details will be sent to the user who has generated the token in the form of email and web notification.
4. Fig 1 represents the complete Working Model.

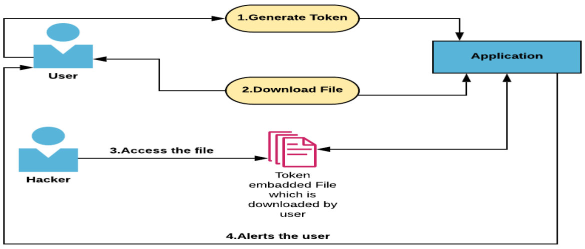


Fig 1. Flow diagram of file Tracking System

The token embedding process differs from one file to another, we have currently embedded tokens in .docx and .pdf files. We have given the embedding token procedure on the docx file in Fig 2.

When any docx file is unpacked, it displays all the XML files by which this docx was made. We have written python script to achieve XML External Entity injection into one of the XML files which were unpacked and Token is injected in this process. Once the token is injected, these XML files are again zipped back and the docx file is recovered which includes the injected token in the original one. Similarly, we have written python script to inject tokens into pdf files also by using cross-side scripting payload injection.

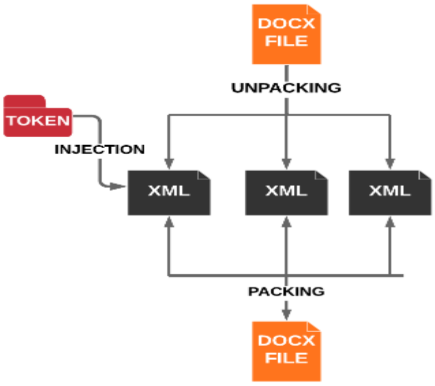
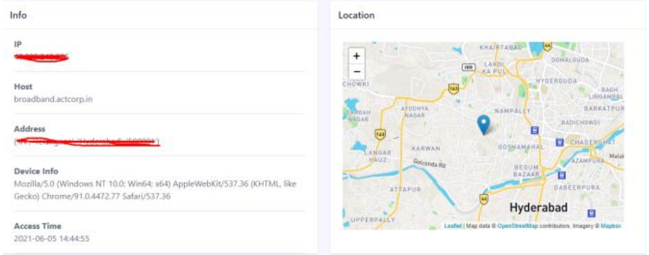


Fig 2. Injecting token into Docx file

IV.OUTPUT

When the user uploads supported type of files, he/she will get a downloadable link for the trackable generated file.

This trackable file is nothing but an original file that is injected with a token. When these trackable files are accessed by anyone at that point the owner of the file will immediately get an alert mail about the details regarding the access on the file. The details such as the IP Address, Location of the attacker who had accessed the file will be shown on the tracking page of the user dashboard as given in Figure 3.



Info

IP
[redacted]

Host
broadband.actcorp.in

Address
[redacted]

Device Info
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.77 Safari/537.36

Access Time
2021-06-05 14:44:55

Location

Fig 3. The output of tracking details

V.SUMMARY

File tracking system tool takes files like a word document and pdf files as its input and as a result, generates the trackable file. These trackable files are embedded with a unique token that will be used in the future for alerting the owner if the file when unknown access has occurred on his/her file.

When an attacker tries to open this trackable file, the token will immediately get activated and triggers an alert to the user in form of an email, web notification. This trigger will include information such as location, time, and IP address of the computer from which file was accessed to reveal the identity of the attacker so the user can take action and preventive steps before something goes wrong with that breached data.

This tool currently supports word document(.docx) and Portable Document Format(.pdf) for generating trackable file. To run this software there are a few dependencies and modules which has to be installed as mentioned in Implementation.

To Deploy this tool as a web application, the server must meet supported requirements as mentioned in Implementation, the source code and working link of this project will be released on https://github.com/20prince12/tracki-file.

V.CONCLUSION

File tracking system is a tool that can help in detecting the data breaches which can happen on these files. Our main focus was to take early measures when your data is breached by an attacker. Even though this tool will not help the owner of the file to stop hacker from a data breach but it will help owner for early detection of the attack and take preventive measures on the data which was breached so in future even if the hacker tries to expose the data, this may not cause a disaster for the owner.

There are few limitations like it supports only specific types of file formats currently which can be solved in the future by improving the tool to support other types of file formats like images, excel sheets.

References

[1] United States Department of Health and Human Services. Retrieved 2015-09-01.2015-09-01.

[2] Panama Papers Leak "The New Normal?" Retrieved 2016-08-20.

[3] “Data Breaches Chronologies" by Privacy Rights Clearinghouse.

[4] White Paper: "Honeypot, Honeynet, Honeytoken: Terminological issues

[5] A Python Book: Beginning Python, Advanced Python by Dave Kuhlman, revised Version October 05, 2014.

[6] "How to Think Like a Computer Scientist" by Jeffrey Elkner, first Edition April 2002

[7] “How to Become a Core Developer” by Python Software Foundation

[8] "Enthought Python Minimum Hardware Requirements" by Jonathan March.Retrieved (February 13, 2020).

[9] "Flask Foreword". Sourced from the original on 2017-11-17.

[10] "Keeping Up" by Jason Sobel(Friday, December 21, 2007) was archived from the Facebook blog on 1 June 2019.

[11] "What is front-end development?" by Ivan Codesido(28-10-2009).

[12] "2010 Annual Study: German Cost of a Data Breach" (PDF) - Symantec(March 2013)

[13] .An introduction to the Flask Python web app framework by Nicholas Hunt-Walker Retrieved 02 Apr 2018.

[14] The web standards model - HTML CSS and JavaScript by W3C.