Vulnerabilities in PDF

Kyu Doun Sim

kdsim@connect.ust.hk

**Motivation**

Phil Ydens, Adobe’s VP Engineering for Document could, estimates that there are 2.5 trillion PDF documents in the world.[1] A lot of contents including academic journals and paper works are in pdf format. It is so widely used that people will send their resumes or cover letters in a pdf format for job applications. It would be one of the most widely used document file format in the world. It was Adobe’s goal to make PDF universal and has achieved its goal. Just from being a text format, PDF allows digital signatures, file attachments, and metadata to work in government organizations. The success of PDF is undoubtful. However, it generally led me to wonder if there are any vulnerabilities the PDF file format. A lot of security vulnerabilities often come from maliciously crafted user inputs that could take advantage of the implementation flaws of a software. What users can at most do with PDF files are to either read or write texts on them. This generally led me to investigate what security vulnerabilities that it could have and what we could learn from it.

**Introduction**

The following written report explores the cybersecurity aspect in PDF file formats. It will start off my introducing what PDFs are and any relevant terminologies that need to be explained in order to explain without hinderance. I will further introduce patches done by the company that created the file format, Adobe, and third party users’ or organization’s contribution to help users to use the file format without security vulnerabilities.

**Relevant Terminologies**

Portable Document Format (PDF)

The Portable Document Format, in short PDF, is a file format created by Adobe, a firm famously known for its range of software that it provides. Any file that has a .pdf extension will be available to be opened by various software. There are numerous software that reads PDF depending on the operating system. It was developed by Adobe in the 1990s and the company has been pushing the public to use PDF as a standard. One of the main problems of document files back in the 1990s was if users were opening documents in a different operating system, document viewing software, and even a different version of the document reader, the fonts and graphics will not be consistent in the these readers. Adobe has been pushing to solve this main concern that users had back then. Newspaper companies were very excited with such technology because printing these documents were also a problem back then. As the file format were able to successful to give consistent text and image quality regardless of the software reading the file and the operating system, more people found the appeal to it and started to widely adopt the format.

Portal Document Format, pdf, is a type of file created from the software company Adobe. The company has created various software including Photoshop, Adobe Premiere and has been creating a lot of software that are used for creators. PDF is no exception, and it is probably one of the most widely used text format. The following is the internal structure of the format.

|  |
| --- |
| Header |
| Body |
| Cross-Reference |
| Trailer |

Figure 1. Structure of the PDF. Analogous to HTML, there are different

The header shows what version of the PDF it is. Body component will often be the actual content of the file. Notice how there could be JavaScript incorporated in the body section of the PDF. The capabilities of JavaScript will further be discussed as many of the exploits

JavaScript

Java script is a scripting language that mostly used on frontend of the Internet. It is gaining its popularity over in the backend too. JavaScript has been added to the PDF in 2000 [reference] allowing the users to create a robust document that could incorporate coding capabilities and interact with any web page more easily.

Figure 2. A screenshot of a PDF document that incorporates JavaScript code.

**Vulnerabilities Discovered / Impactful Events**

The following discusses prominent vulnerabilities or hacks that are involved with pdf. It is not a comprehensive list of all vulnerabilities and patches, and has been selected widely from different categories and different portions of the file exploited.

There are various ways how a malicious PDF can reach a person. It is usually via the Internet. There are numerous PDF documents that could be the textbook or the paper that you needed for.

There are 2 components to read a PDF. You would need the PDF file itself, and a reading software that could be Adobe Acrobat, Smallpdf, or even your web browser. Adversaries could put in malicious code in the PDF file itself, or it could exploit the reader software. The following screenshot is from Adobe’s website showing the vulnerabilities that the company has discovered and patched so far.

Vulnerability 1

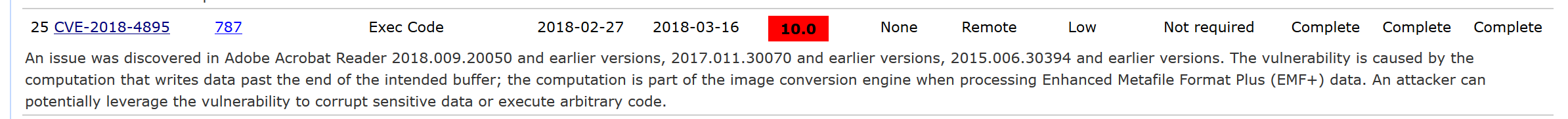
1. What
2. When
3. Where
4. Payload
5. Exploited vulnerability
6. Patches

Vulnerability 2

1. What
2. When
3. Where
4. Payload
5. Exploited vulnerability
6. Patches

Vulnerability 3

1. What
2. When
3. Where
4. Payload



An issue was discovered in Adobe Acrobat Reader 2018.009.20050 and earlier versions, 2017.011.30070 and earlier versions, 2015.006.30394 and earlier versions. The vulnerability is caused by the computation that writes data past the end of the intended buffer; the computation is part of the image conversion engine when processing Enhanced Metafile Format Plus (EMF+) data. An attacker can potentially leverage the vulnerability to corrupt sensitive data or execute arbitrary code.

1. Exploited vulnerability
2. Patches

Vulnerability 4

Time bomb

1. What
2. When
3. Where
4. Payload
5. Exploited vulnerability
6. Patches

**Third Party Contributions**

One prominent project created by jesparza is the “peepdf” tool created in Python and is used to analyze whether a pdf could be harmful or not. There are numerous libraries and work done by people to mitigate the potential damage a PDF file could cause.

Web browser’s have also constantly been pushing to support a more secure software not to get compromised any malicious PDFs.

**Defenses against Attacks**

As discussed in the vulnerabilities portion, often the vulnerabilities comes form the software that allows you to read the pdf. If users are using Adobe Reader to read their PDFs, they should make sure they are updating their software whenever there are relevant security patches. If users are using their web browsers to read PDF files, they should regularly update their browsers to ensure that the code will not exploit the browser as well. Another way users could be careful is to be aware of the sources of the PDF that they are downloading from. Websites that pirate textbook pdfs or information could often contain such malicious codes. On the other hand, this could be coupled with a spear phishing attack that could come from e-mail attachments. Therefore, users should always check whom send the e-mail, and also to double check the format of the attached files.

**Conclusion**

So far the prominent vulnerabilities and patches introduced by Adobe, and efforts from third party users to detect harmful pdf were discussed. PDF will continuously use be successful as it has been pushing its boundaries in terms of the graphical capabilities of it and the

Figure x. A rough screenshot what Google News return when “pdf hack” was searched. The materials are still very recent and relevant. Notice malicious e-mails are coupled with PDF.

The long history and exposure of PDFs often let the user’s guard down. The static nature of a PDF is rather deceiving, as the vulnerabilities exposed shows how the users would not be too suspicious of the file they received. These PDF vulnerabilities coupled with social engineering or phishing attacks even makes it more dangerous as the users do not suspect that the sender would send PDF files that has a malicious intent. This means that people will continuously try to exploit PDFs.

**References**

1. <https://itextpdf.com/en/blog/technical-notes/do-you-know-how-many-pdf-documents-exist-world>
2. <https://www.decalage.info/file_formats_security/pdf>
3. <http://www.malwaretracker.com/pdfthreat.php>
4. <https://www.investintech.com/resources/blog/archives/84-13-ways-the-pdf-is-vulnerable.html>
5. <https://www.welivesecurity.com/2018/05/15/tale-two-zero-days/>
6. <https://securityxploded.com/pdf_vuln_exploits.php>
7. <https://security.stackexchange.com/questions/64052/can-a-pdf-file-contain-a-virus>
8. <https://www.foxitsoftware.com/blog/understanding-security-vulnerabilities-in-pdfs/>
9. <http://pralab.diee.unica.it/sites/default/files/aisec09f-corona.pdf>
10. <https://www.zdnet.com/article/windows-10-or-mac-user-patch-adobe-reader-and-acrobat-now-to-fix-9-critical-security-flaws/>
11. <https://www.brightfort.com/pdffix.html>
12. <https://techcrunch.com/2010/07/23/an-analysis-of-a-pdf-exploit/>
13. <https://techxplore.com/news/2019-10-major-vulnerabilities-pdf.html>
14. <https://www.menlosecurity.com/blog/critical-vulnerability-in-firefox-built-in-pdf-viewer>
15. <https://en.wikipedia.org/wiki/History_of_the_Portable_Document_Format_(PDF)>
16. <https://www.websecurity.digicert.com/en/hk/security-topics/what-is-ssl-tls-https>
17. <https://github.com/jesparza/peepdf>
18. <http://www.h-online.com/security/features/CSI-Internet-PDF-timebomb-1038864.html>
19. <https://www.sentinelone.com/blog/malicious-pdfs-revealing-techniques-behind-attacks/>
20. <https://proprivacy.com/privacy-news/pdf-security-guide-vulnerable-exploit>
21. <https://www.cvedetails.com/vulnerability-list/vendor_id-53/product_id-497/Adobe-Acrobat-Reader.html>
22. <https://www.cvedetails.com/product/497/Adobe-Acrobat-Reader.html?vendor_id=53>