PDF association

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

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.

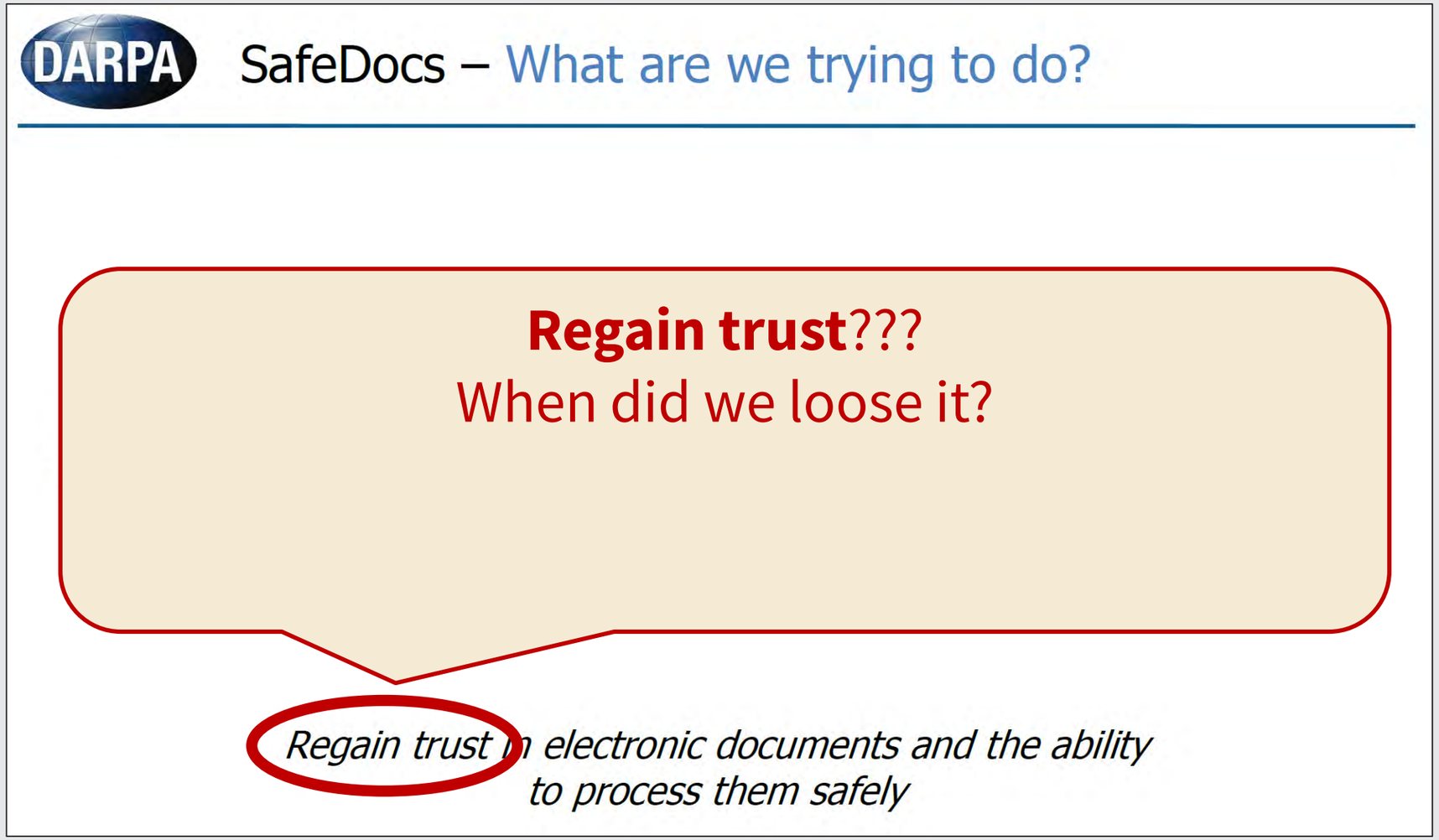
DARPA Safe Documents (SafeDocs)

SafeDocs – What are we trying to do?

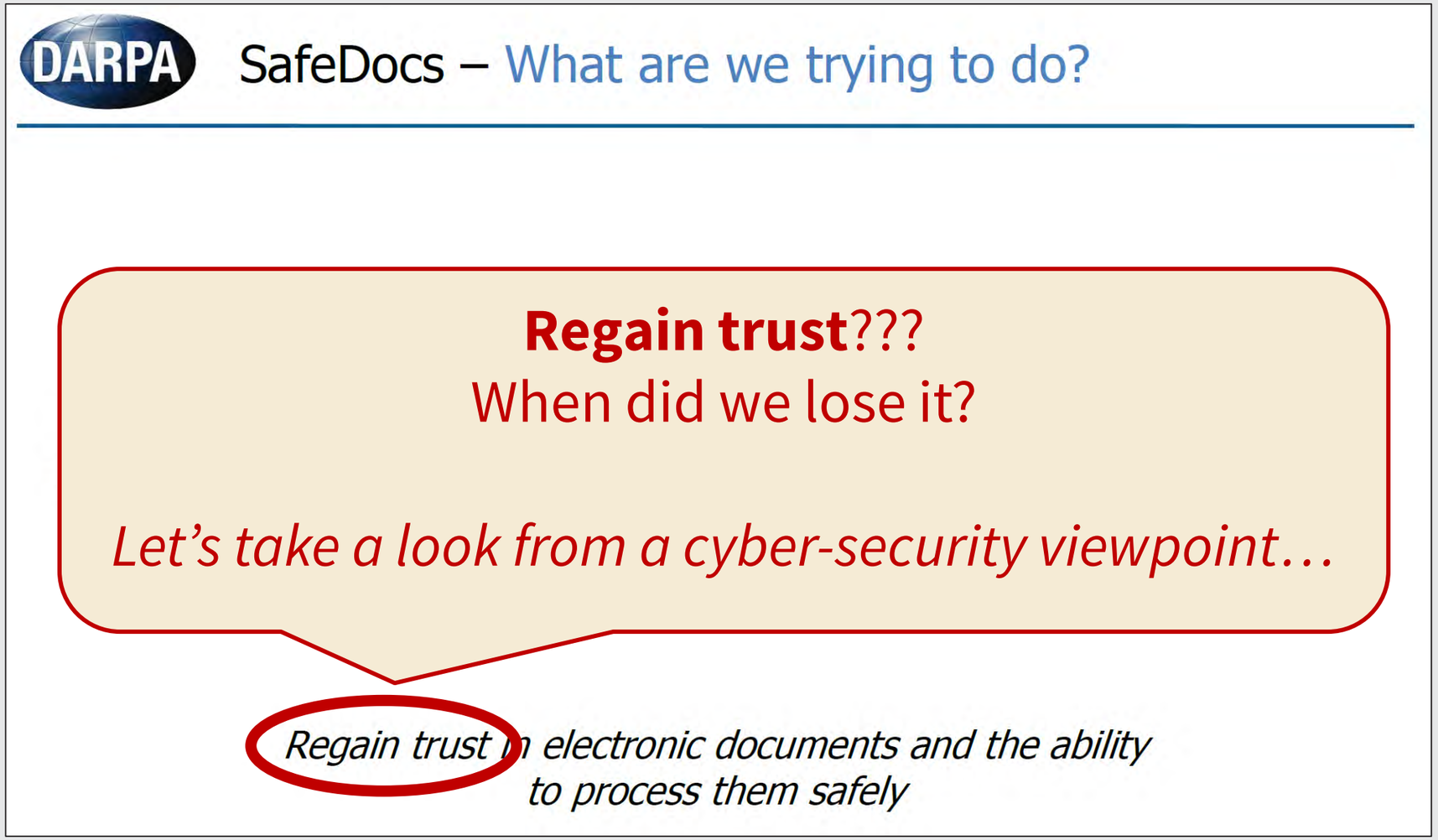
Reduce electronic document complexity and build verified parsers to radically improve software's ability to reject invalid and malicious data

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

DARPA Safe Documents (SafeDocs)



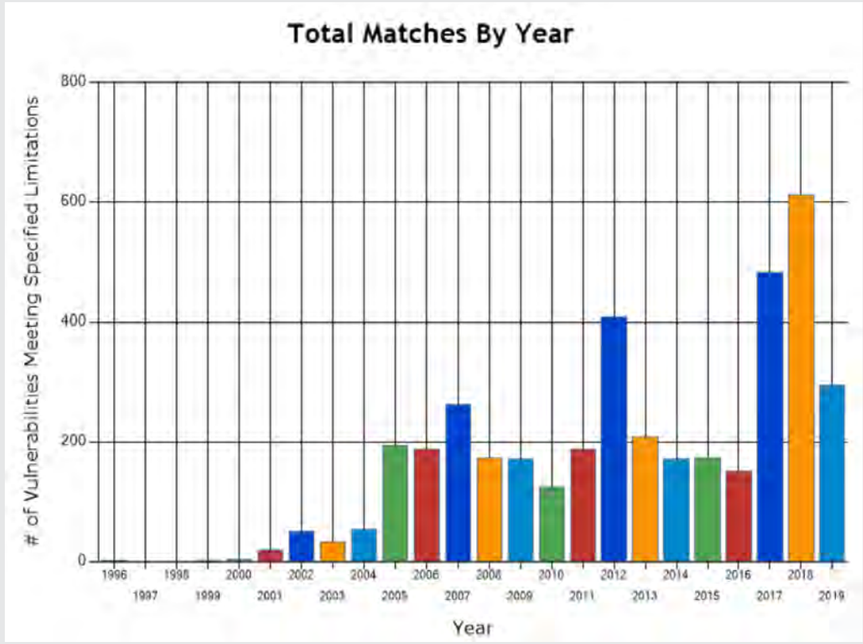
DARPA Safe Documents (SafeDocs)

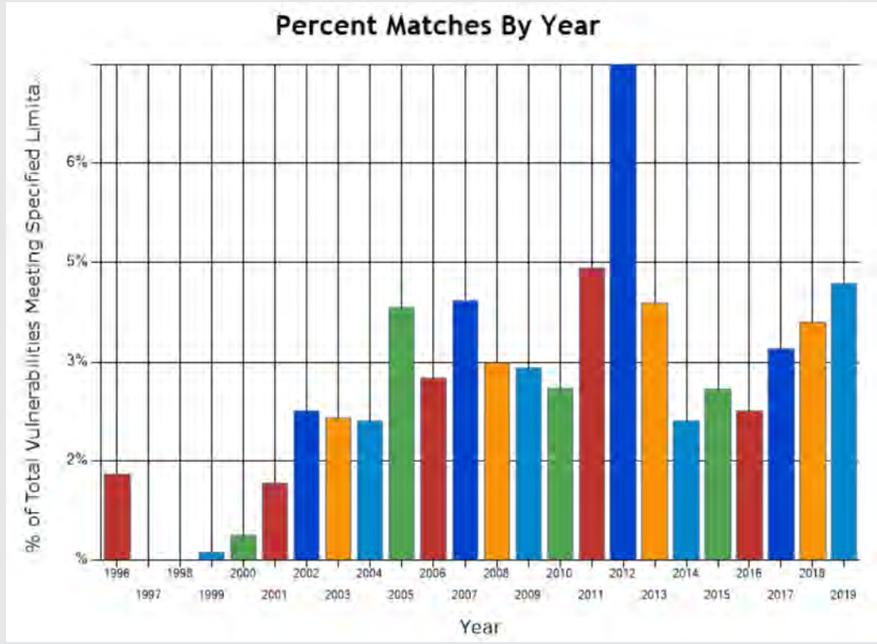


NIST Vulnerability data…

- “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

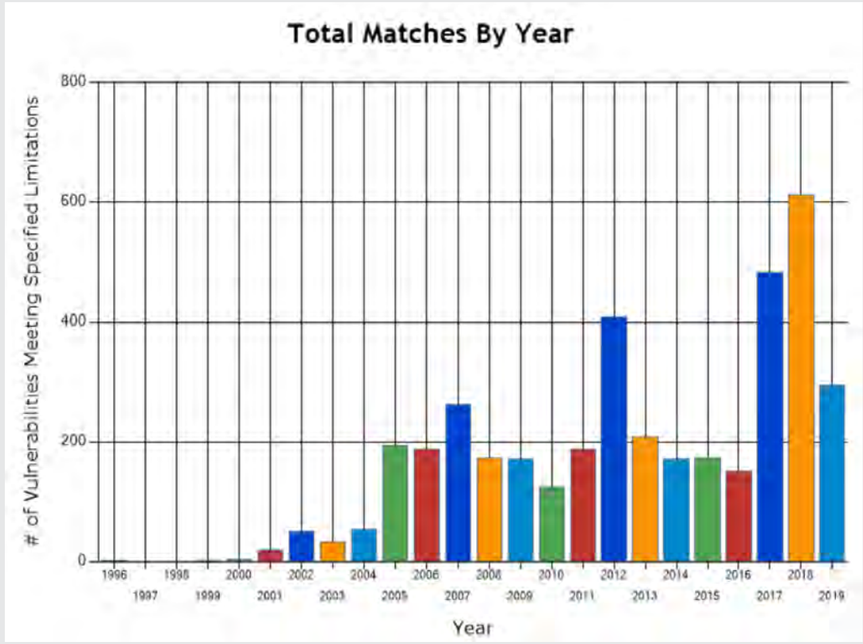


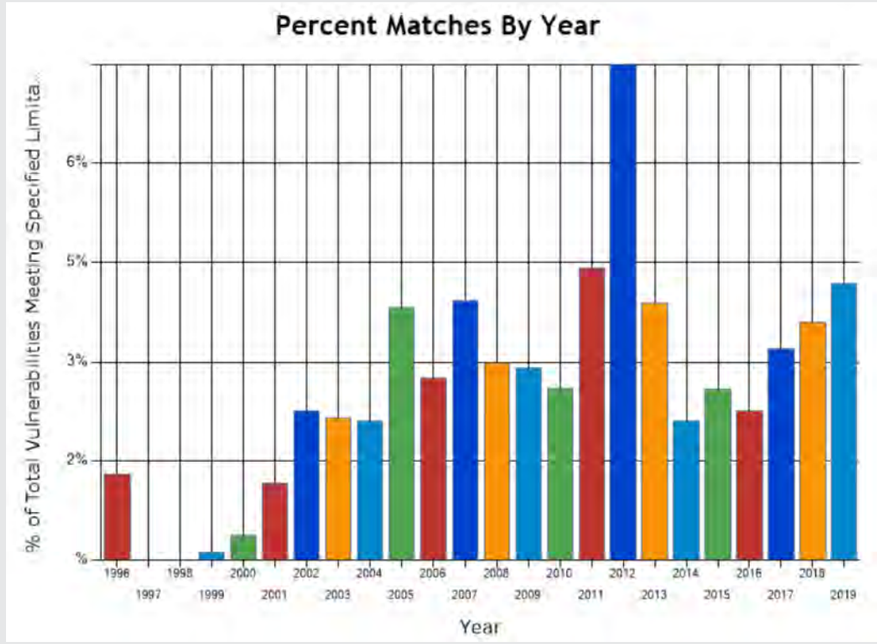


NIST Vulnerability data…

- "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?

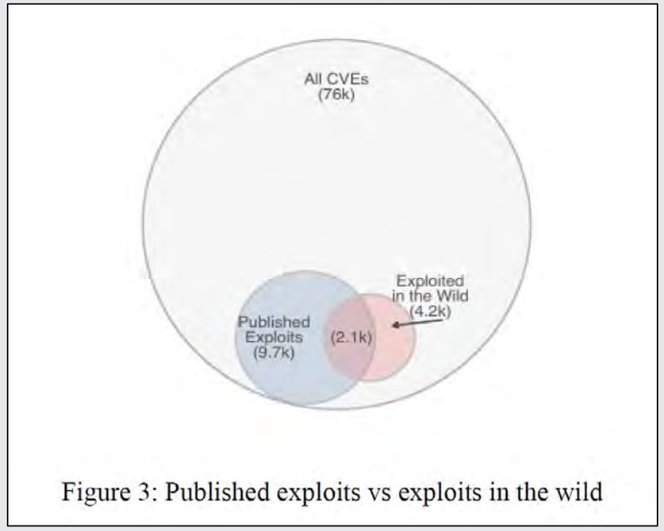


Figure 3: Published exploits vs exploits in the wild

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

What does that really mean?

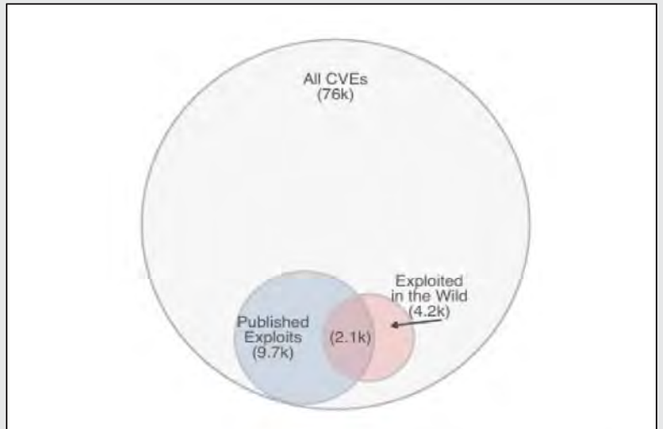


Figure 3: Published exploits vs exploits in the wild

"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?

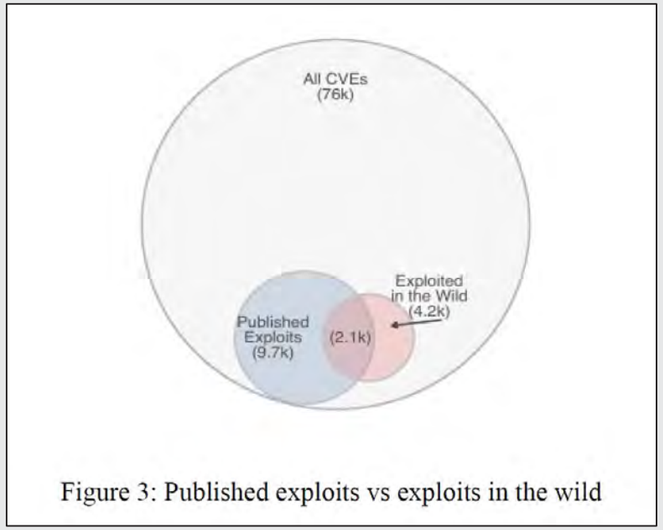
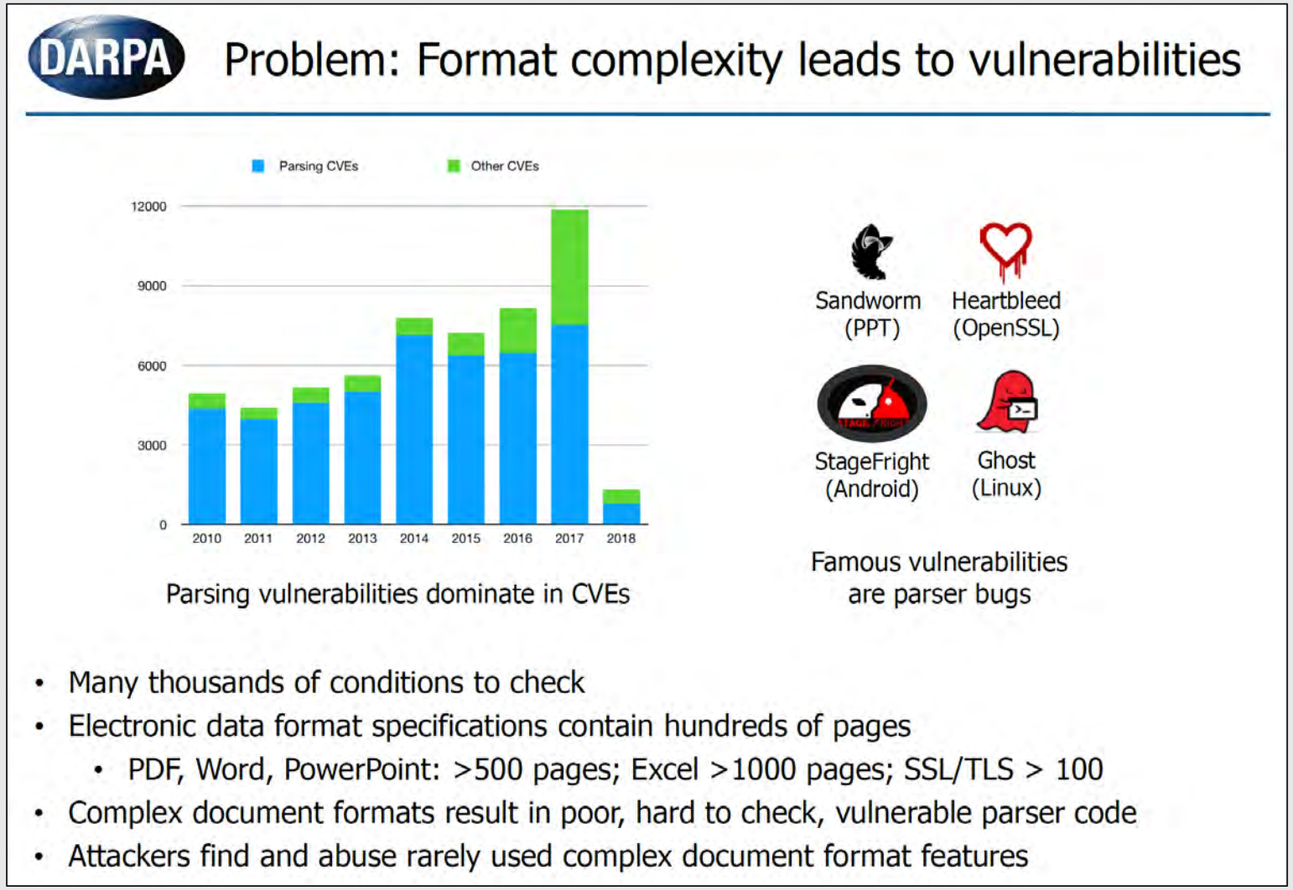


Figure 3: Published exploits vs exploits in the wild

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

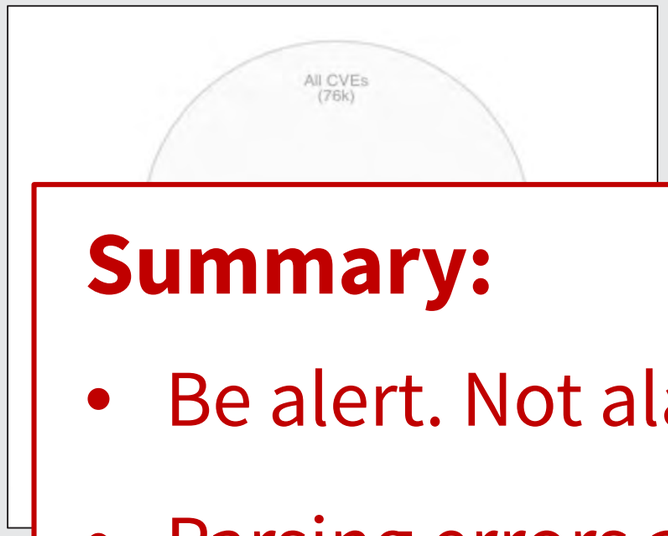
"*A Broad View of the Ecosystem of Socially Engineered Exploit Documents*", Blond et al, 2017.



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?



Summary:

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

“VirusTo
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 CVEs Other CVEs

- 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

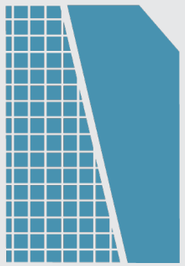
Payload

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

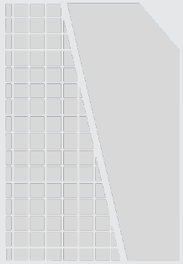
Payload

DoS

Zero-Day

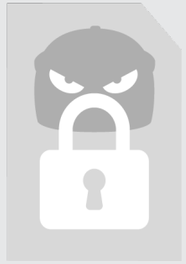
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking

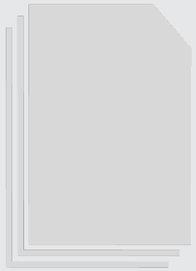


UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...



Spam

Phishing

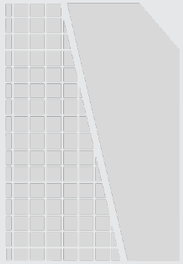
Payload

DoS

Zero-Day

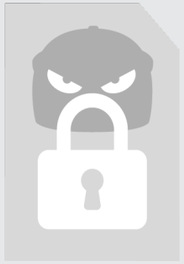
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking

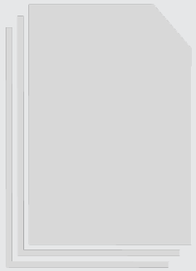


UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...



Spam

Phishing

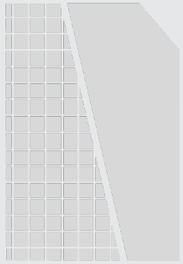
Payload

DoS

Zero-Day

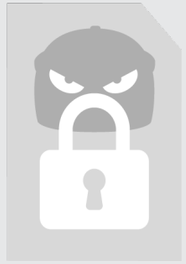
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking



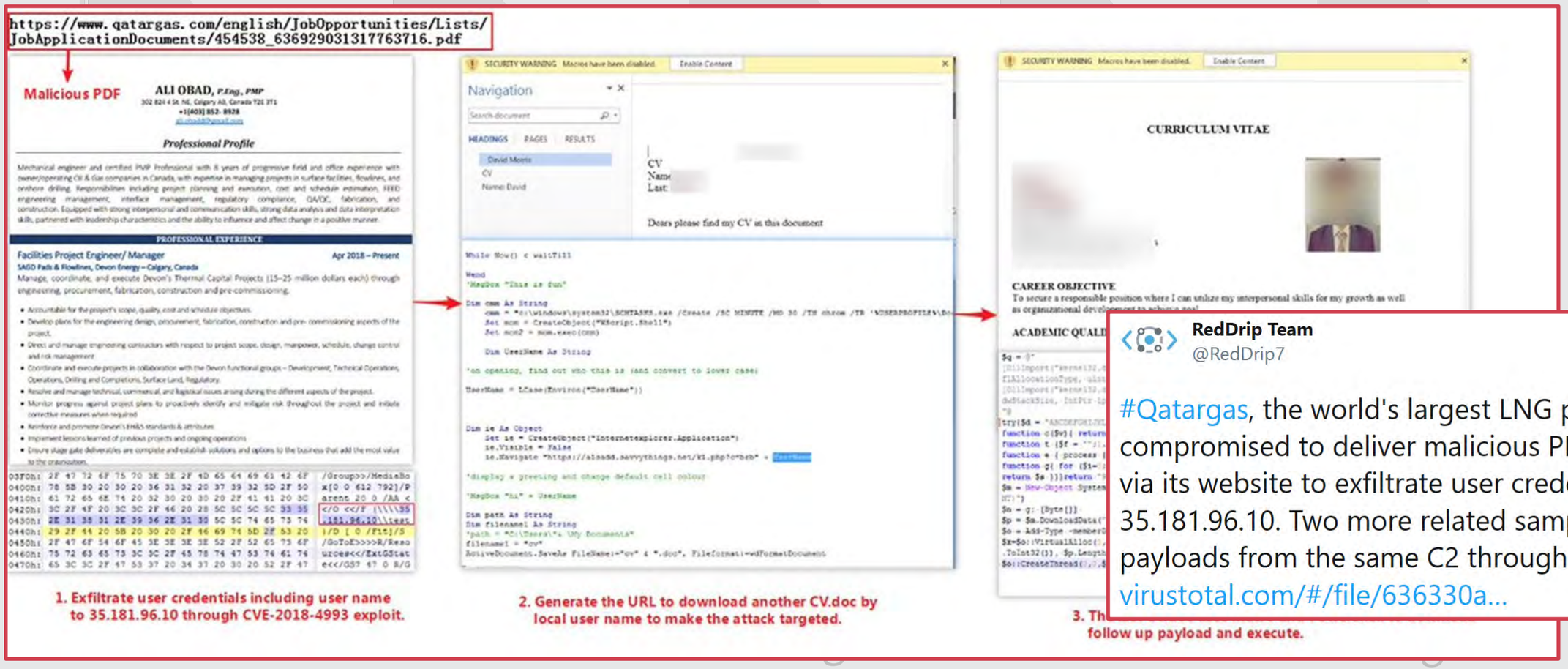
UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...

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



Social Engineering

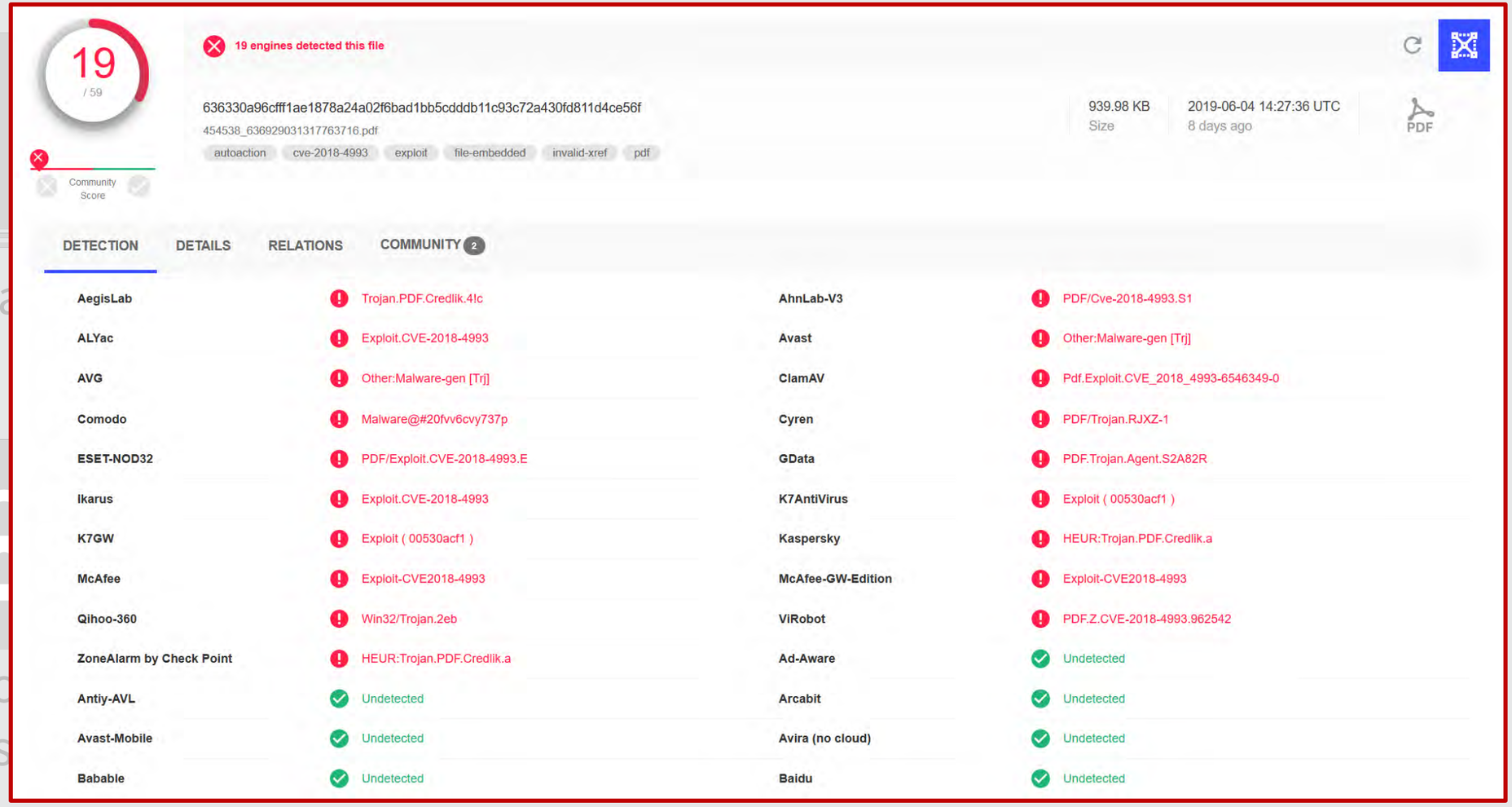
RedDrip Team
@RedDrip7

#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...

Hiding

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

How cyber-security researchers think about PDF...



19 / 59

19 engines detected this file

636330a96cfff1ae1878a24a02f6bad1bb5cdddb11c93c72a430fd811d4ce56f

454538_636929031317763716.pdf

autoaction | cve-2018-4993 | exploit | file-embedded | invalid-xref | pdf

939.98 KB Size

2019-06-04 14:27:36 UTC 8 days ago

PDF

Community Score

DETECTION | DETAILS | RELATIONS | COMMUNITY 2

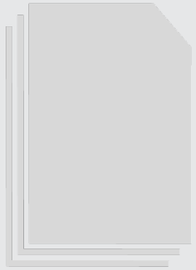
AegisLab	Trojan.PDF.Credlik.4!c	AhnLab-V3	PDF/Cve-2018-4993.S1
ALYac	Exploit.CVE-2018-4993	Avast	Other:Malware-gen [Trj]
AVG	Other:Malware-gen [Trj]	ClamAV	Pdf.Exploit.CVE_2018_4993-6546349-0
Comodo	Malware@#20fvv6cvy737p	Cyren	PDF/Trojan.RJXZ-1
ESET-NOD32	PDF/Exploit.CVE-2018-4993.E	GData	PDF.Trojan.Agent.S2A82R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

Social Engineering

Information Hiding

https://www.virustotal.com/gui/file/636330a96cfff1ae1878a24a02f6bad1bb5cdddb11c93c72a430fd811d4ce56f/detection

How cyber-security researchers think about PDF...



Spam

Phishing

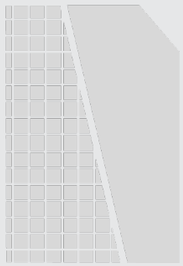
Payload

DoS

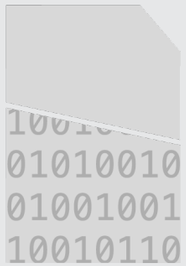
Zero-Day

Social Engineering

Misinformation/ Mis-trust



Polyglot



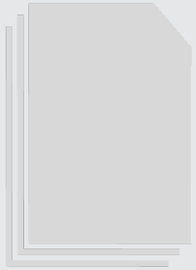
Content Masking

UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...



Spam

Misinformation/
Mis-trust

"***1 Trillion Dollar Refund – How To Spoof PDF Signatures***" by Mladenov et al.

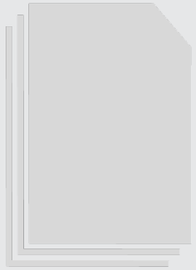
Attacks take an already digitally signed PDF then modify the PDF (*introduce errors*) such that software does not detect the modifications and reports the digital signature as still valid.

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

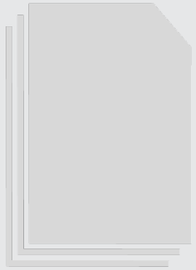
Misinformation/
Mis-trust

Root Causes:

- complexity of PDF specification

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

How cyber-security researchers think about PDF...



Spam

Misinformation/
Mis-trust

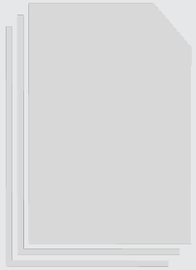
Root Causes:

- complexity of PDF specification

^and it's nested formats

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

How cyber-security researchers think about PDF...



Spam

Misinformation/
Mis-trust

Root Causes:

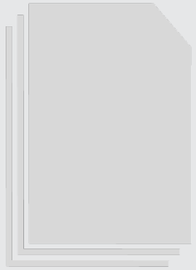
- complexity of PDF specification

^and it's nested formats

- tolerance to errors (permissiveness)

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

How cyber-security researchers think about PDF...



Spam

Phishing

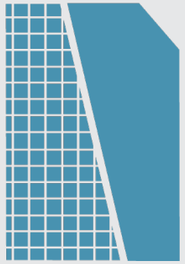
Payload

DoS

Zero-Day

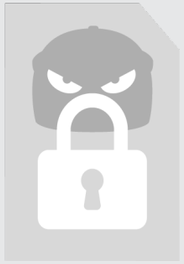
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking

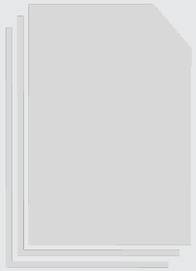


UI Forgery

Privacy Leakage

Information Hiding

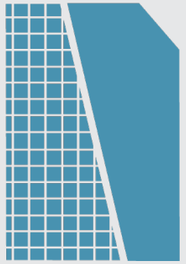
How cyber-security researchers think about PDF...



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"

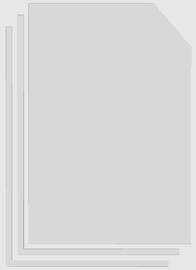
Masking

Forgery

Leakage

Hiding

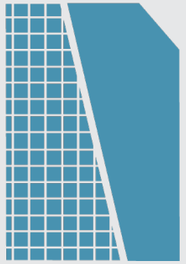
How cyber-security researchers think about PDF...



Spam

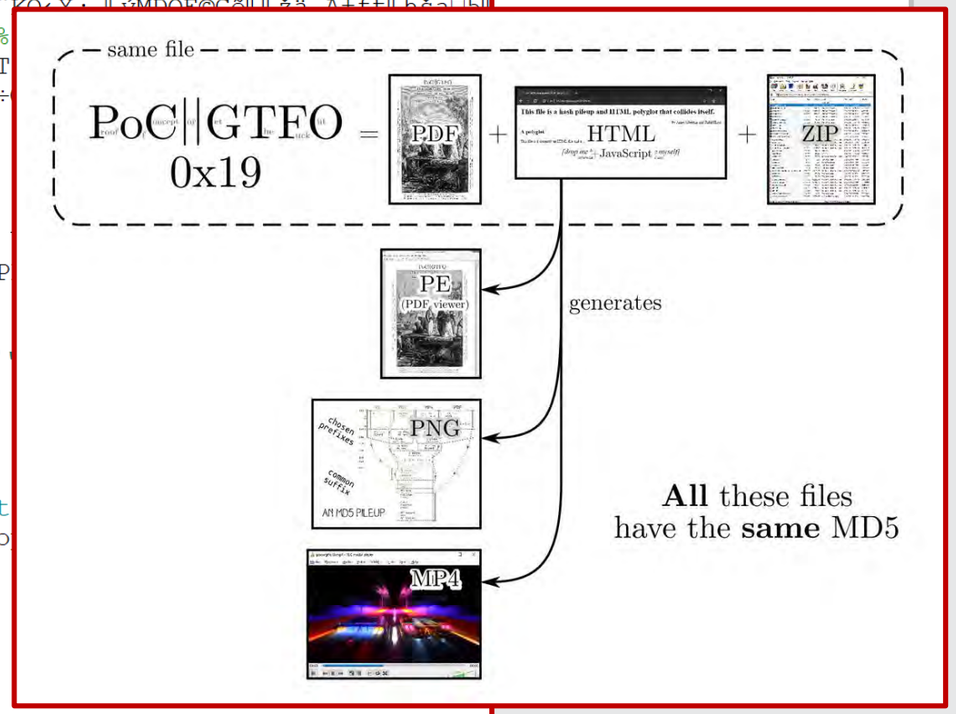
Phishing

Misinformation/
Mis-trust

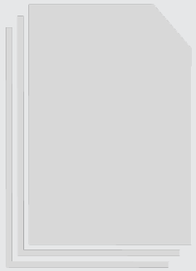


Polyglot

```
%PDF-1.3
%ÂµÂ¶

1 0 obj
<</Length 2 0 R>>
stream
-----=-------=\|  MD5 pileup  \|  PDF prefix  ||  Ange & Marc  /\-=-    jµçd-Ô
▯ÒòLù%)4"ªkÑx ▯šOÏœV †óv)S{♠©ü«Ì«Œèl‹\|?{Ï ™▯[-Ö}†+z▯ÖŽší-¨«ö.FÍ{ï▯▯L-@½•▯ò\"^Žñ
C▯ Î¤□è')▯ ž€æàTäà'▯@¸lÓV▯ åFþÇ ÒD„▯▯Ú»▯ŽGc.á ▯e¨KÕ‹Ý. ▯‹MÐÕÈ©Cã▯▯šä Ä+f+▯hšø□5▯
h3•¯¿ê▯ŒÕ¾Iî£Hé▯°N▯Šïn▯ *V ŒbDµâ▯pñ¤ýÖk¥O℗▯.▯%
Z‹œíQXe℗NÚŒžë▯Q"õ×LÒë  €ÉôÂ▯œ§?-ïáD´ÿ?ß×ç  ÞT
¤▯Û—©I†F«▯;^Ì▯℗X"  ||³▯$ÕsóŒ▯ÈUã …7>G0▯™¡ ³FIù÷
0„'▯²▯Gè▯0|×Æ$¢¢ P—>ªpÇ▯¥?zí-Ã¤▯          Î
±X ¦Ùß° j¨Ž³¦rÝÛÌÓ
h%n¨
ŸK▯1Ã-S▯§¹V▯▯>ôÞHÂ▯f-å▯W▯9      -▯
'âV¢Ÿþ8û▯z¨ù÷Ú*çqá FÑ ½A'™?"W°¶˜ZÒs…) Ù¸ö▯»Ÿp
dÖP Ø▯▯©,û´õw $L▯free        ▯ýÂ^  !N¤aNGE    P
  p   °   ð+ PiL  ,  pL   @ ▯     ▯ ▯     ▯

$      'ÑIÔÖ°'‡Ö°'‡Ö°'‡Eþ¿‡Ð°'‡Í-¹‡î°'‡ßÈ¤‡Ð°'
  p   °   ð+ PiL  ,  pL   @ ▯     ▯ ▯     ▯
       ▯▯œ£GÑ9ýFl¦NL Fi   J   &! E-->
<div id='mypage'>
<span id='dropme'
  ondrop="event.preventDefault();dropped(event
  ondragover="event.preventDefault();event.sto

<style>
  body {
    /* top left&right bottom*/
    margin: 10px 20px 0px;
  }
```



How cyber-security researchers think about PDF...



Spam

Phishing

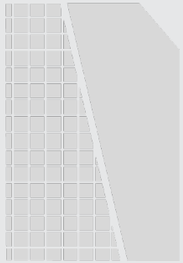
Payload

DoS

Zero-Day

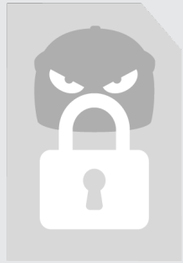
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking



UI Forgery

Privacy Leakage

Information Hiding

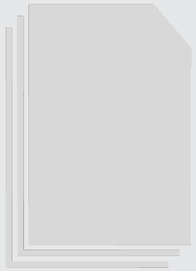
How cyber-security researchers think about PDF...

"*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 <u>not</u> visible in the content

How cyber-security researchers think about PDF...



Spam

Phishing

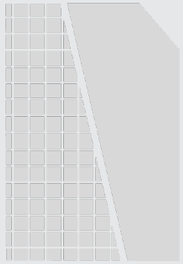
Payload

DoS

Zero-Day

Social Engineering

Misinformation/ Mis-trust



Polyglot

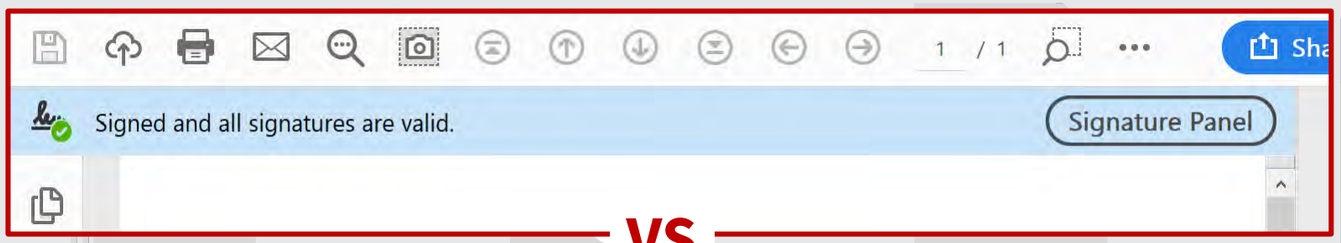
Content Masking

UI Forgery

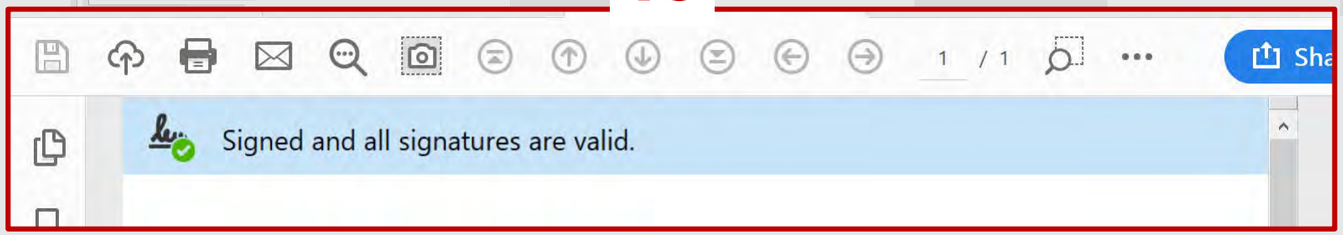
Privacy Leakage

Information Hiding

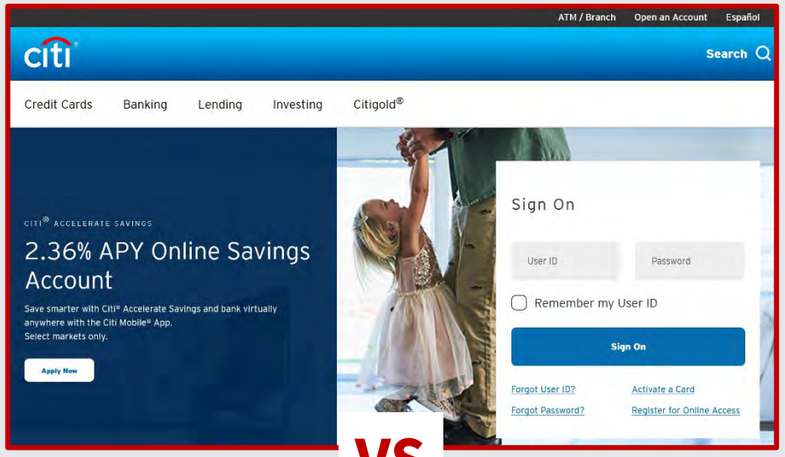
How cyber-security researchers think about PDF...



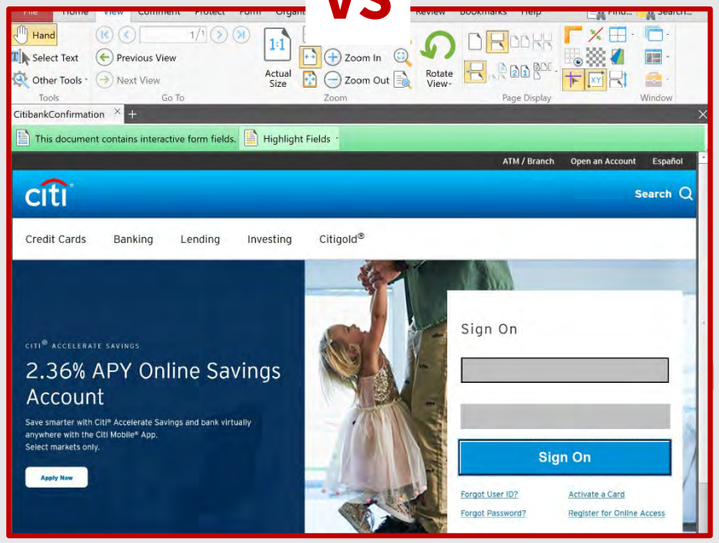
VS



DoS



VS



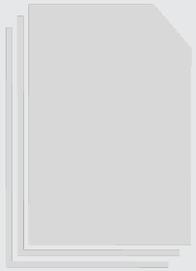
Misinformation/
Mis-trust

Polyglot

Content
Masking

UI
Forgery

How cyber-security researchers think about PDF...



Spam

Phishing

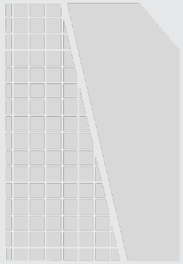
Payload

DoS

Zero-Day

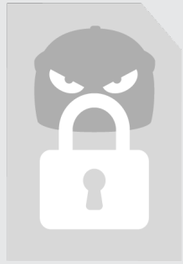
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking

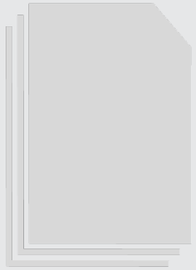


UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...



Spam

Phishing

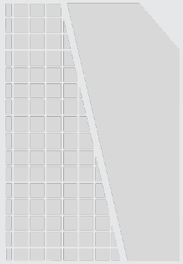
Payload

DoS

Zero-Day

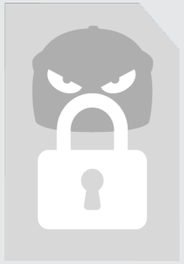
Social Engineering

Misinformation/ Mis-trust



Polyglot

Content Masking



UI Forgery

Privacy Leakage

Information Hiding

How cyber-security researchers think about PDF...

Spam

Phishing

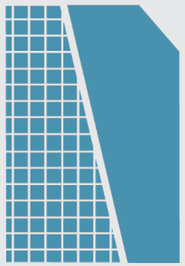
Payload

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?***
- **Goal:** “Regain trust in electronic documents and the ability to process them safely”

^and it's nested formats

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.