Scoring and Judging Artifacts in Autopsy

Brian Carrier

Basis Technology

SANS DFIR Summit 2021

AUTOPSY DIGITAL FORENSICS

CYBER TRIAGE

Problem: Data Overload

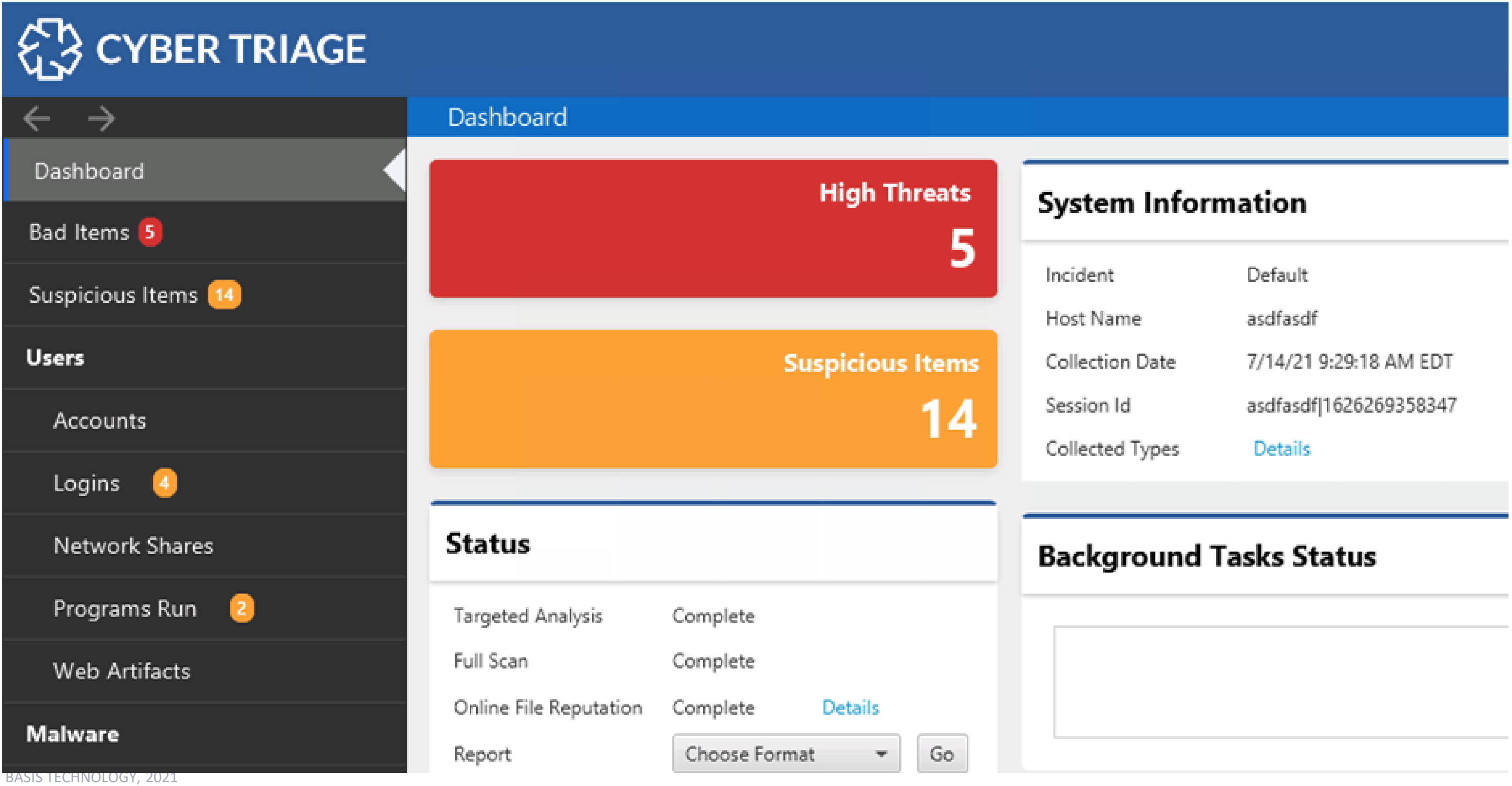
- Cases involve an increasing number of devices
- Devices are getting bigger and bigger
- We’ve been talking about this for years…
- We need better solutions to deal with it

We’re Not Alone…

- This is not unique to DFIR.
- Information Retrieval has various ranking algorithms to put most relevant results on top.
- Google PageRank revolutionized this for web searches.

- We need to bring these concepts into DFIR to quickly highlight the evidence.

We’ve Been Doing This in Cyber Triage



Incident	Default
Host Name	asdfasdf
Collection Date	7/14/21 9:29:18 AM EDT
Session Id	asdfasdf|1626269358347
Collected Types	Details

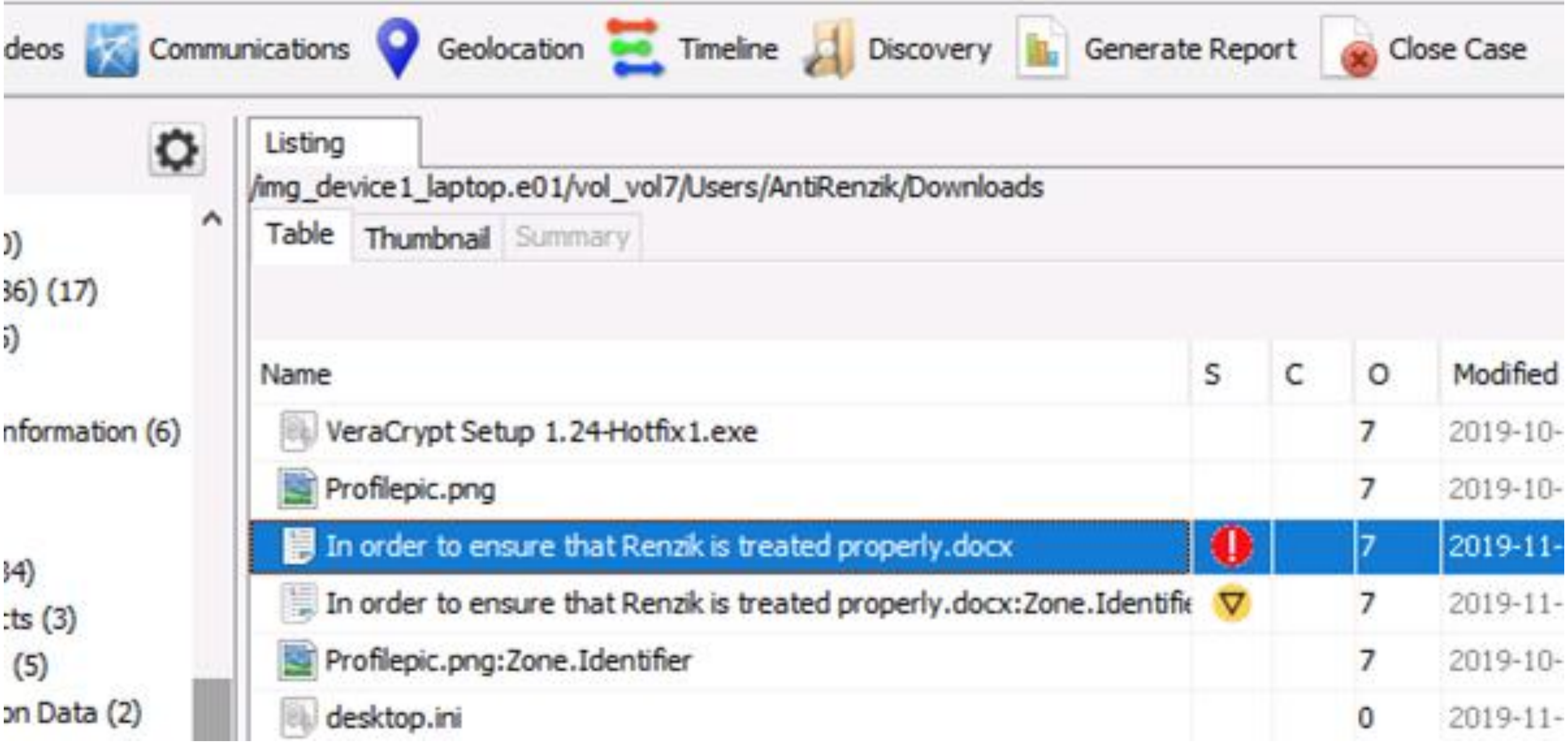
Targeted Analysis	Complete	
Full Scan	Complete	
Online File Reputation	Complete	Details
Report	Choose Format ▾	Go

How Does Cyber Triage Do It?

- Various modules analyze artifacts and assign scores.
- The user is shown the items with the highest scores.

Module Results	**Final**
Malware Module: 25 engines flag it as malicious (Bad)	Bad
Path Module: Executable in AppData (Suspicious) Malware Module: 0 engines flag it and first seen 1 year ago (OK)	OK

And Now We've Moved It to Autopsy



/img_device1_laptop.e01/vol_vol7/Users/AntiRenzik/Downloads

Name	S	C	O	Modified
VeraCrypt Setup 1.24-Hotfix1.exe			7	2019-10-
Profilepic.png			7	2019-10-
In order to ensure that Renzik is treated properly.docx	!		7	2019-11-
In order to ensure that Renzik is treated properly.docx:Zone.Identifie	▽		7	2019-11-
Profilepic.png:Zone.Identifier			7	2019-10-
desktop.ini			0	2019-11-

Goals of This Talk

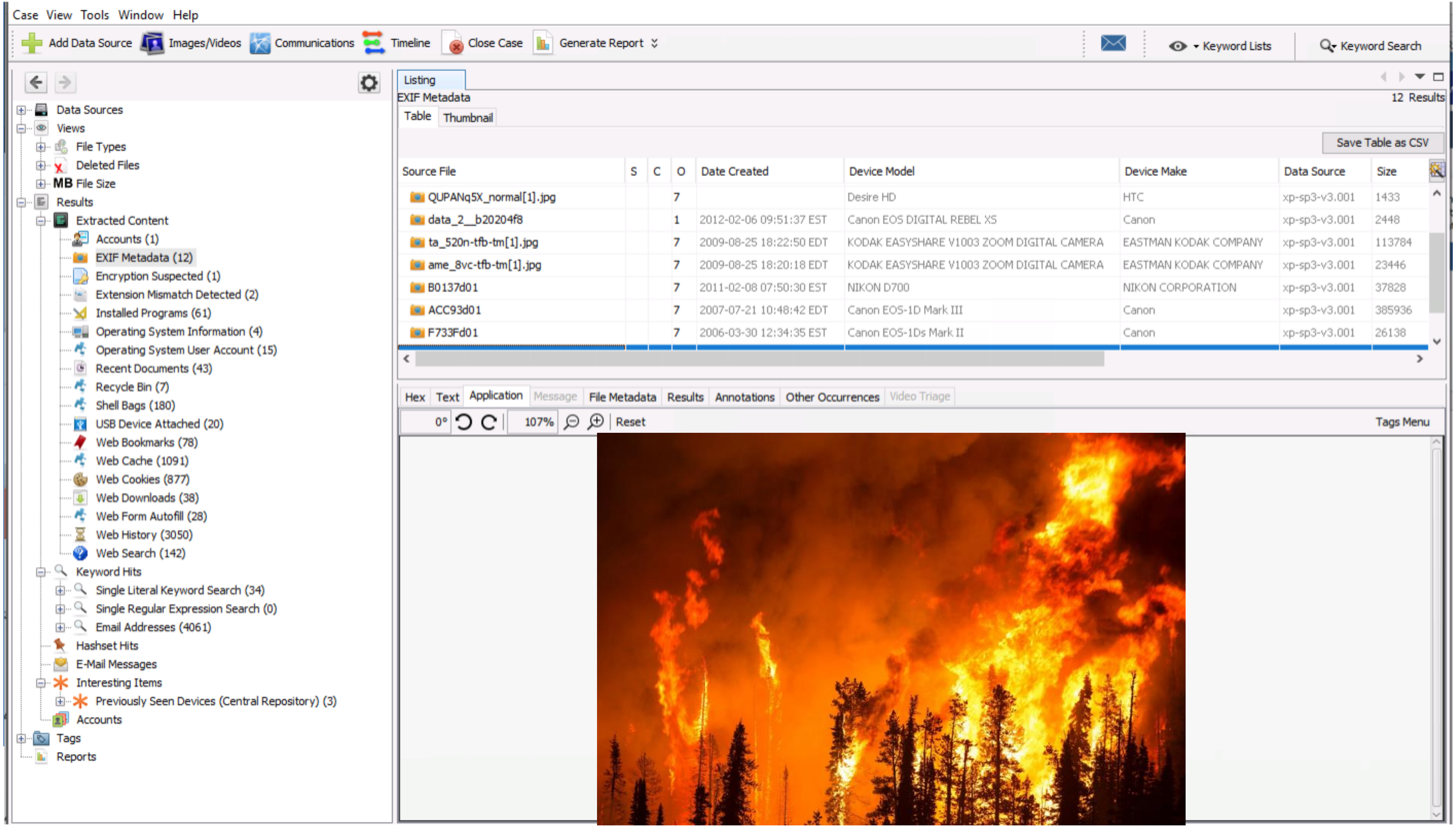
1) Introduce the general scoring concepts

1) Outline changes to Autopsy
   a) Storage
   b) User Interface
   c) Module Writing

What is Autopsy?

- Open source digital forensics platform.
- General purpose for all kinds of investigations.
- Has been designed for:
  - Ease of use
  - Showing results ASAP
  - Extensibility (many plug-in frameworks)
- Has the features you need.
- Free to download:

  http://www.autopsy.com



EXIF Metadata

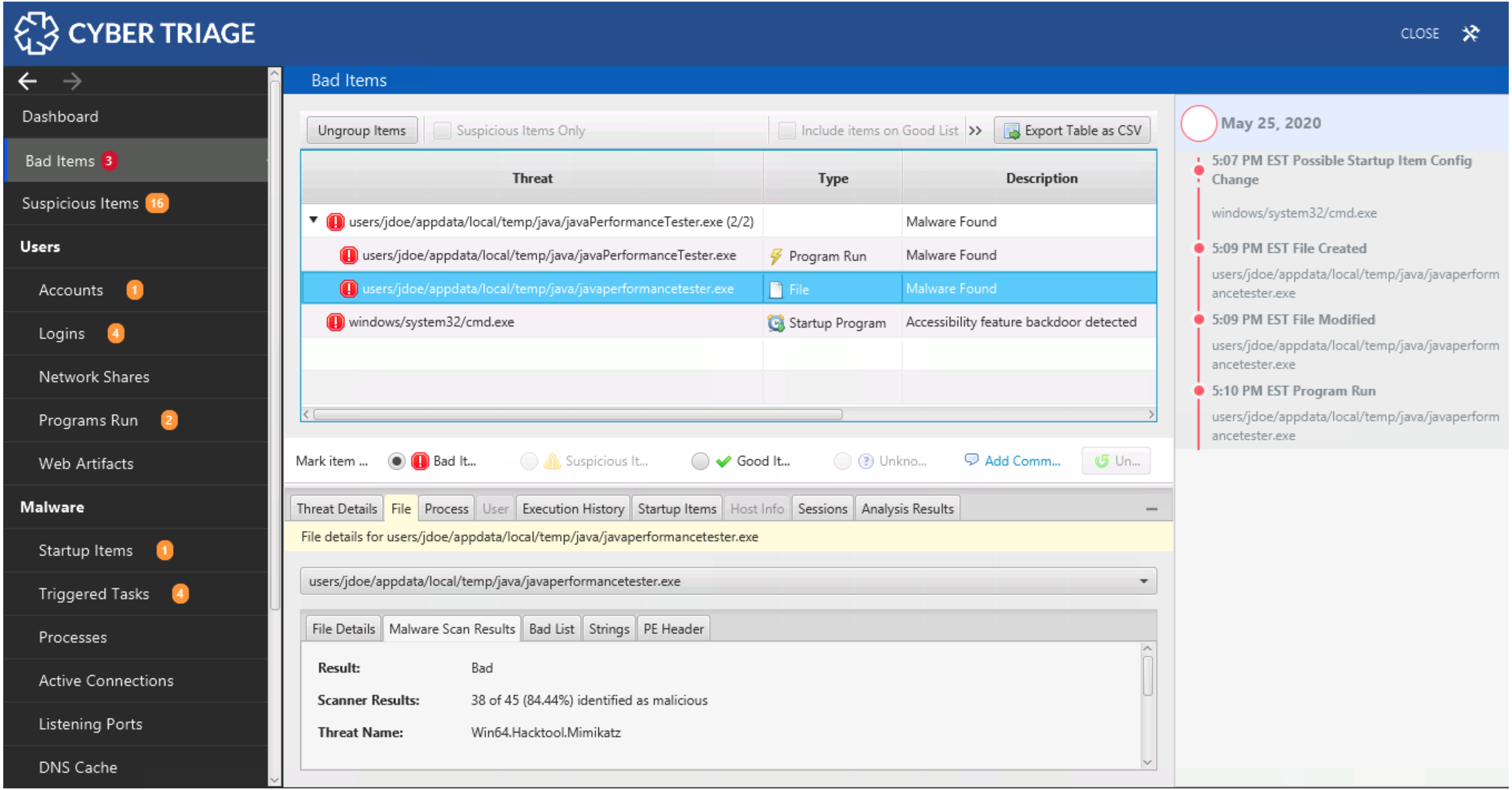
12 Results

Source File	S	C	O	Date Created	Device Model	Device Make	Data Source	Size
QUPANq5X_normal[1].jpg			7		Desire HD	HTC	xp-sp3-v3.001	1433
data_2__b20204f8			1	2012-02-06 09:51:37 EST	Canon EOS DIGITAL REBEL XS	Canon	xp-sp3-v3.001	2448
ta_520n-tfb-tm[1].jpg			7	2009-08-25 18:22:50 EDT	KODAK EASYSHARE V1003 ZOOM DIGITAL CAMERA	EASTMAN KODAK COMPANY	xp-sp3-v3.001	113784
ame_8vc-tfb-tm[1].jpg			7	2009-08-25 18:20:18 EDT	KODAK EASYSHARE V1003 ZOOM DIGITAL CAMERA	EASTMAN KODAK COMPANY	xp-sp3-v3.001	23446
B0137d01			7	2011-02-08 07:50:30 EST	NIKON D700	NIKON CORPORATION	xp-sp3-v3.001	37828
ACC93d01			7	2007-07-21 10:48:42 EDT	Canon EOS-1D Mark III	Canon	xp-sp3-v3.001	385936
F733Fd01			7	2006-03-30 12:34:35 EST	Canon EOS-1Ds Mark II	Canon	xp-sp3-v3.001	26138

What is Cyber Triage?

- Hyper-focused Intrusion Forensics Tool
- Started as an Autopsy module, became entirely independent, and now shares common database.
- Collects artifacts, scores them (malware, etc.), and recommends new ones.
- Learn more at:

  http://www.cybertriage.com



Threat	Type	Description
▼ users/jdoe/appdata/local/temp/java/javaPerformanceTester.exe (2/2)		Malware Found
users/jdoe/appdata/local/temp/java/javaPerformanceTester.exe	Program Run	Malware Found
users/jdoe/appdata/local/temp/java/javaperformancetester.exe	File	Malware Found
windows/system32/cmd.exe	Startup Program	Accessibility feature backdoor detected

File details for users/jdoe/appdata/local/temp/java/javaperformancetester.exe

users/jdoe/appdata/local/temp/java/javaperformancetester.exe

Result: Bad

Scanner Results: 38 of 45 (84.44%) identified as malicious

Threat Name: Win64.Hacktool.Mimikatz

May 25, 2020

5:07 PM EST Possible Startup Item Config Change

windows/system32/cmd.exe

5:09 PM EST File Created

users/jdoe/appdata/local/temp/java/javaperformancetester.exe

5:09 PM EST File Modified

users/jdoe/appdata/local/temp/java/javaperformancetester.exe

5:10 PM EST Program Run

users/jdoe/appdata/local/temp/java/javaperformancetester.exe

General Concepts (Tool Agnostic)

Investigative Question

What you’ve been tasked to answer using digital data.

Examples:

- Is this computer compromised?
- Was this computer used to commit fraud?
- Does this computer contain child exploitation images?

Relevant Artifacts

The artifacts (or files) that will help you answer your Investigative Question.

Examples:

- Intrusion: Malware files, login events, ...
- Fraud: Spreadsheets, emails, ...
- Child Exploitation: Pictures, web history, ...

Analysis Techniques

Methods applied to artifacts to determine their relevance.

Examples:

- Lookup a file’s MD5 hash in known hashset (good or bad).
- Keyword search artifacts for relevant words
- Static analysis on executables to look for malware signatures, etc.

Analysis Results

The outcome of an analysis technique on a given artifact.

Examples:

- File hash has found in a known good hash set
- E-mail message has a phrase associated with fraud
- Executable has a known ransomware signature

Analysis Result Relevance Score

Likelihood that the artifact is relevant to your investigative question based on the analysis result conclusion.

Examples:

- Good Hash Set Hit -> Not Relevant
- Keyword Hit -> Maybe Relevant
- Ransomware Signature Hit -> Relevant

Artifact Aggregate Score

Combination of all of the analysis result scores for a given artifact.

Example:

- An executable file with two analysis results:
  - 1) In a suspicious folder -> Likely Relevant
  - 2) With a malware signature -> Relevant
  - Aggregate Score: Relevant

We’ll talk about the Autopsy algorithm later on.

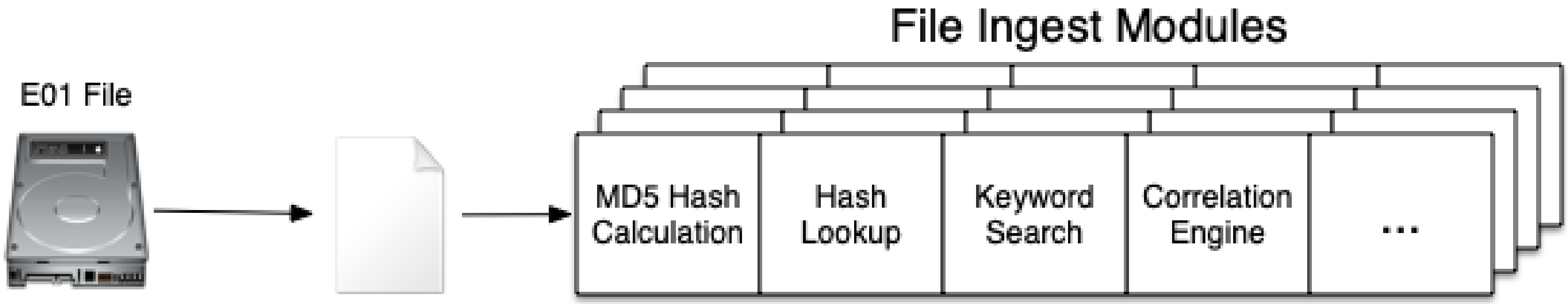
Summary

- Every artifact and file can have:
  - One or more analysis results
  - An aggregate score based on the individual results
- You can focus on the most relevant artifacts first.
- Then, the likely relevant, etc.

Storing Scores in Autopsy

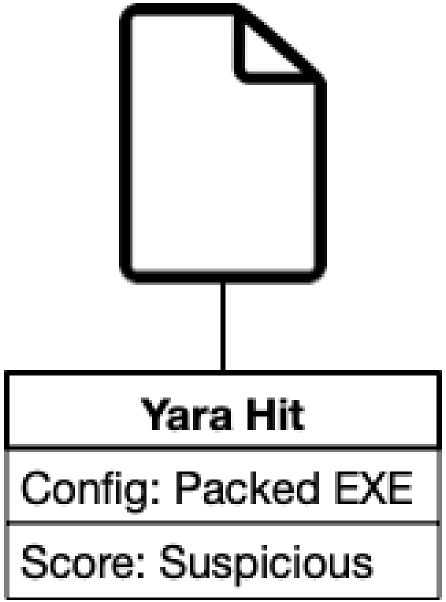
New Analysis Result Type

- Autopsy now has a concept of "Analysis Result" with a score.
- Ingest modules continue to perform the "Analysis Techniques" and create the results (and extract data)



Analysis Result Details

- Are associated with a file or artifact.
  - **Type**: Such as “Hash Set Hit”
  - **Score**: How relevant the item is to the investigation
  - **Configuration** (optional): Such a hash set name
  - **Conclusion** (optional): Such as “Signed, but untrusted”

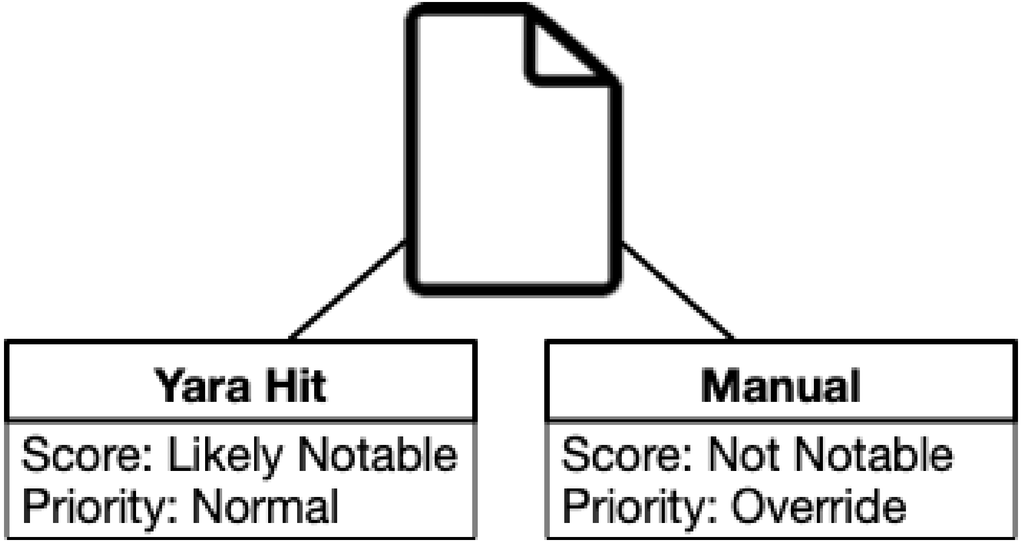


Scoring Details: Significance

- A score has two fields: Significance and Priority
- Significance is about relevance:
    - NOTABLE: The artifact is relevant (i.e. check it out now)
    - LIKELY_NOTABLE: The artifact is likely relevant (i.e. check it out soon)
    - LIKELY_NONE: The artifact is likely not relevant
    - NONE: The artifact is not relevant (i.e. ignore it)
    - UNKNOWN: It's unclear if the artifact is relevant or not

Scoring Details: Priority

- Sometimes you need to overrule an analysis result.
  - Normal Priority: Default for automated analysis techniques
  - Override: Reserved for manual adjustments.
- Example:
  - A Yara module marked a file as LIKELY_NOTABLE.
  - The user can override that score.



Yara Hit
Score: Likely Notable Priority: Normal

Manual
Score: Not Notable Priority: Override

Aggregate Score

- The aggregate score for each item is stored in a separate database table.
- By default, every file and artifact has an aggregate score of “Unknown”.

Aggregate Score Algorithm

- If any “override” scores exist, use only them.
- Prioritize NOTABLE (Bad) over NONE (Good) and high confidence over low confidence.
    - NOTABLE > NOT_NOTABLE > LIKELY_NOTABLE > LIKELY_NOT_NOTABLE > UNKNOWN
- Example:
    - Malware Scan: NOT_NOTABLE
    - Executable Path: LIKELY_NOTABLE
    - Aggregate Score: NOT_NOTABLE

Example Scores

Analysis Result Type	Significance
Hash Set Hit	Notable or Not Notable
Keyword Hit	Likely Notable
Interesting File or Artifact	Likely Notable
Encryption Detected	Notable
Encryption Suspected	Likely Notable
File Extension Mismatch	Likely Notable
Yara Hit	Notable
Tags	Notable or Likely Notable

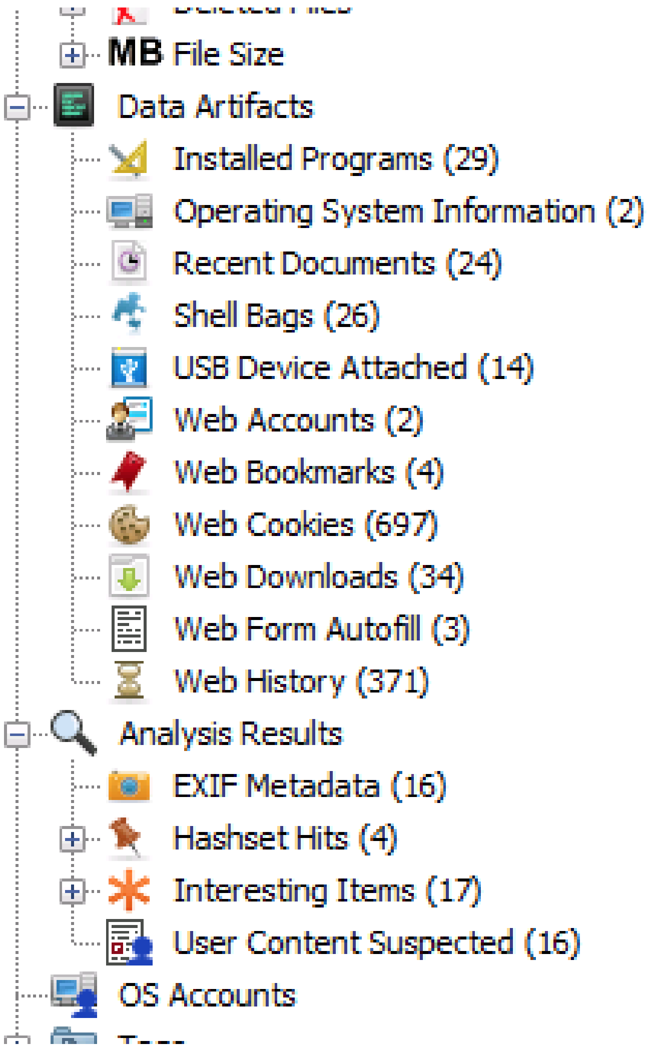
Data Artifacts

- New term to refer to “extracted data”.
    - I.e. a web artifact extracted from a SQLite database
- Autopsy used to store analysis results and data artifacts the same way.
- They are now different:
    - Analysis Results have scores, etc.
    - Data Artifacts have OS Accounts, etc.

Autopsy UI Changes

New Tree Layout

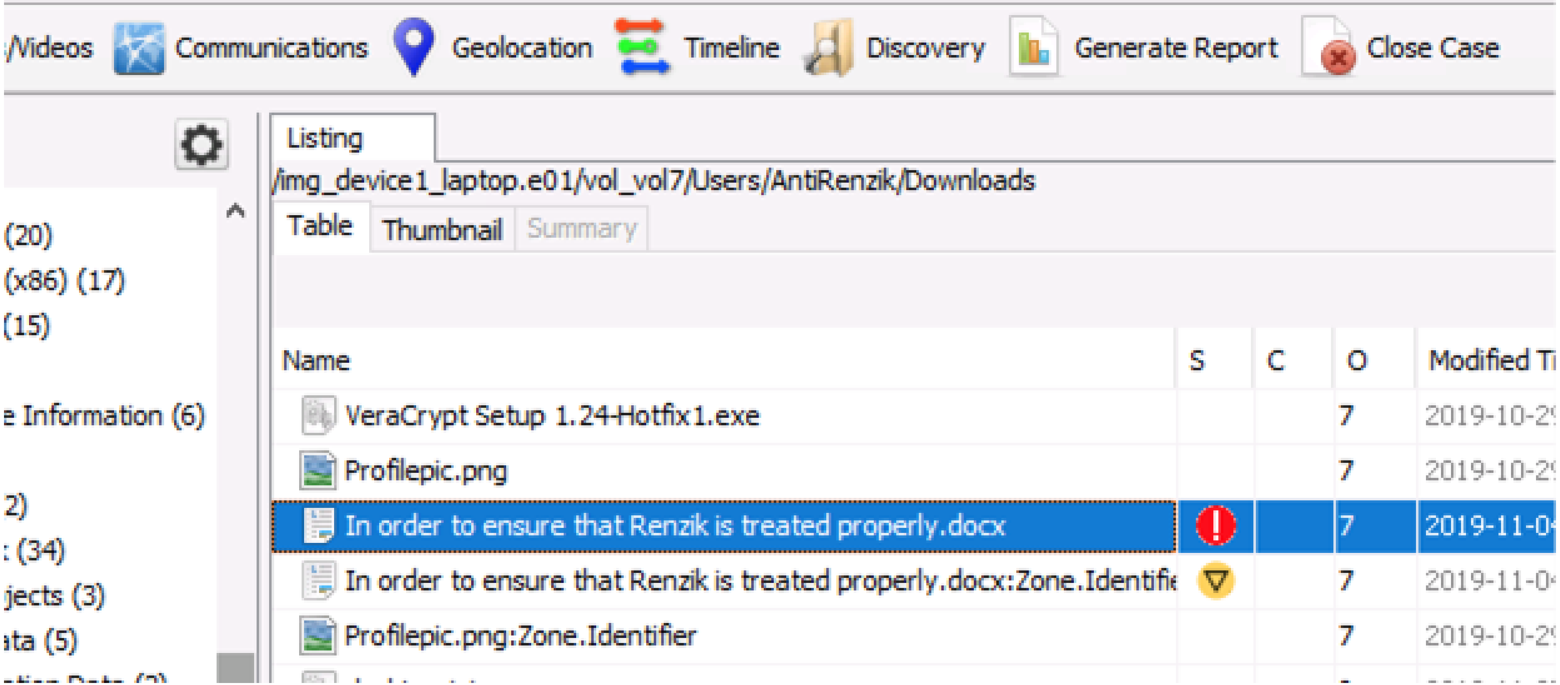
Separate sections for Data Artifacts from Analysis Results



- MB File Size
- Data Artifacts
  - Installed Programs (29)
  - Operating System Information (2)
  - Recent Documents (24)
  - Shell Bags (26)
  - USB Device Attached (14)
  - Web Accounts (2)
  - Web Bookmarks (4)
  - Web Cookies (697)
  - Web Downloads (34)
  - Web Form Autofill (3)
  - Web History (371)
- Analysis Results
  - EXIF Metadata (16)
  - Hashset Hits (4)
  - Interesting Items (17)
  - User Content Suspected (16)
- OS Accounts

Score Column Shows Aggregate Score

Red = Notable. Yellow = Likely Notable

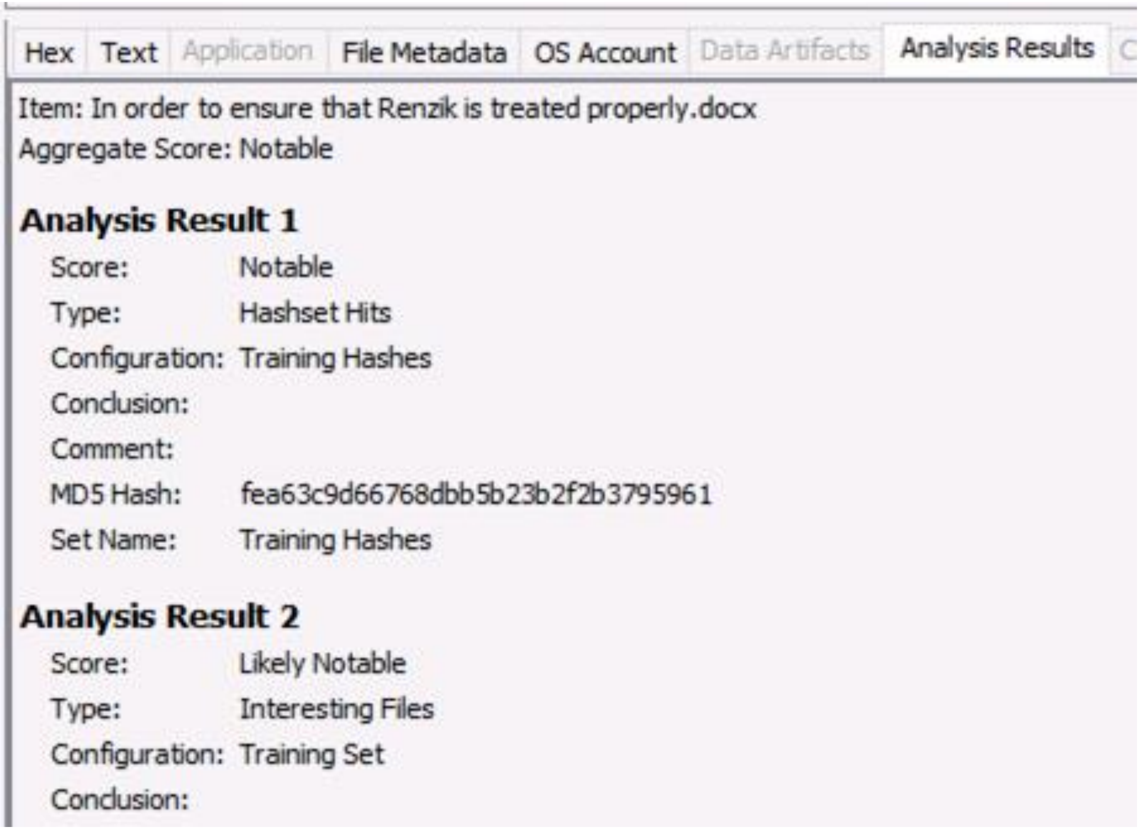


/img_device1_laptop.e01/vol_vol7/Users/AntiRenzik/Downloads

Name	S	C	O	Modified Ti
VeraCrypt Setup 1.24-Hotfix1.exe			7	2019-10-2
Profilepic.png			7	2019-10-2
In order to ensure that Renzik is treated properly.docx	!		7	2019-11-0
In order to ensure that Renzik is treated properly.docx:Zone.Identifie	▽		7	2019-11-0
Profilepic.png:Zone.Identifier			7	2019-10-2

New Analysis Result Viewer

Shows aggregate score and details of each analysis result



Item: In order to ensure that Renzik is treated properly.docx
Aggregate Score: Notable

Analysis Result 1

Score: Notable
Type: Hashset Hits
Configuration: Training Hashes
Conclusion:
Comment:
MD5 Hash: fea63c9d66768dbb5b23b2f2b3795961
Set Name: Training Hashes

Analysis Result 2

Score: Likely Notable
Type: Interesting Files
Configuration: Training Set
Conclusion:

Changes For Autopsy Module Writers

Why Make An Autopsy Module...

- We make it really easy to write modules
  - Copy and paste one of our templates
  - Autopsy will pass in files to you (you can ignore images vs ZIP file vs carved, etc.)
  - You do your fancy work and save results to the database.
  - Your results show up in the UI automatically
  - Your results show up in reports automatically
- You can focus on the fancy stuff

Creating Analysis Results

- You used to call newBlackboardArtifact()
- Now you call newAnalysisResult() and pass in a score
- That’s it. The aggregate score, UI, etc. all get automatically updated.

```
file.newAnalysisResult(TSK_EXT_MISMATCH_DETECTED,
    LIKELY_NOTABLE, “txt not allowed for JPEG”)
```

Summary

The Future….

- This is all in Autopsy 4.19.0 (out next week)
- The basic infrastructure was the first step.
- Next Steps:
  - More modules will start making scores
  - More UIs will use the score for display and search
  - Allow user to override the scores

www.autopsy.com

Additional Learning Opportunities

- OSDFCon is Oct 20 and will be hybrid.

  http://osdfcon.org

- Divide and Conquer DFIR Process. Free 3-hour course.

  http://cybertriage.com/training

- Video-based Autopsy training is available.

  http://autopsy.com/training