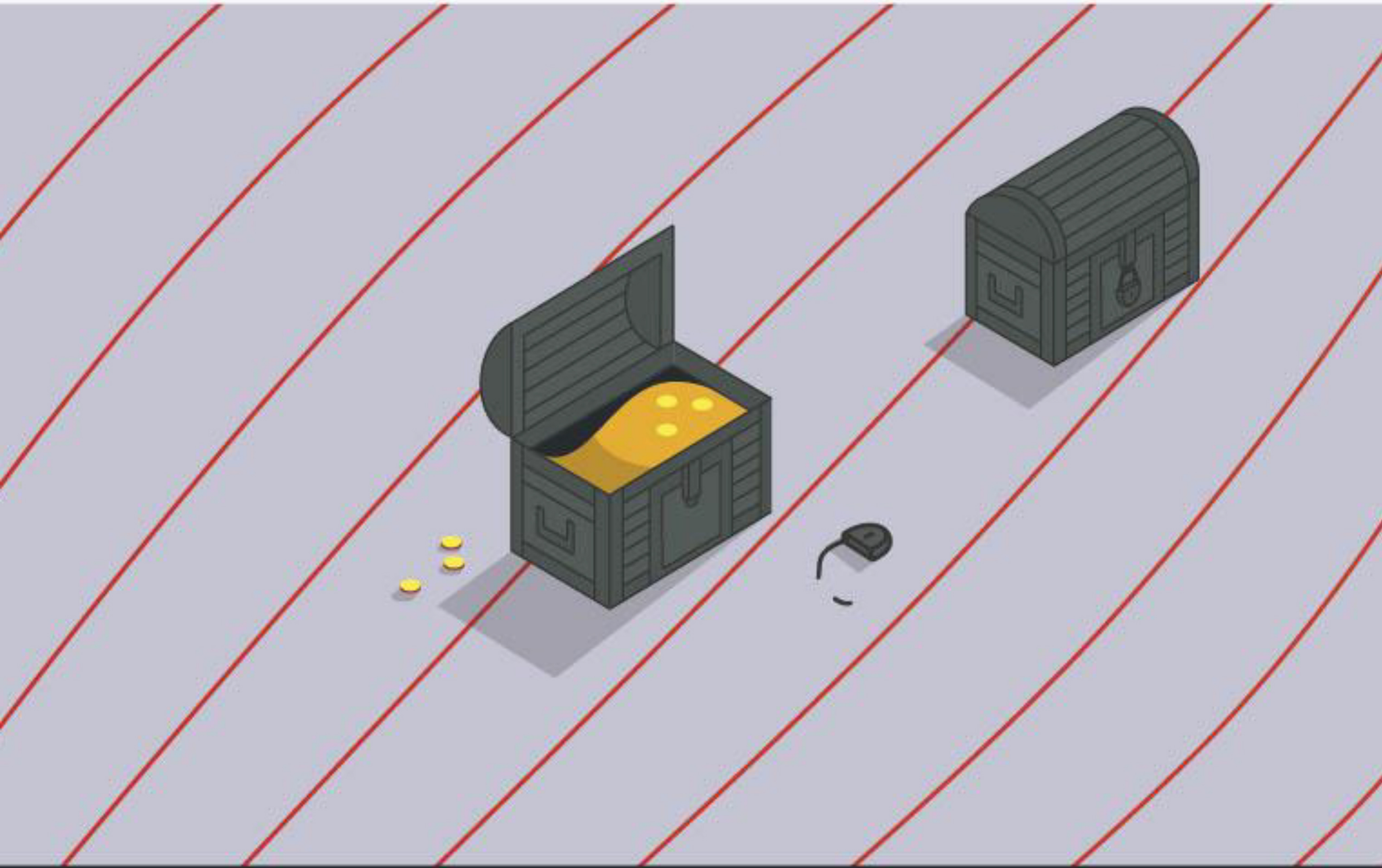
INQUEST



InQuest® Data Loss Discovery

Driven by Deep File Inspection® (DFI)

InQuest Data Loss Discovery

Data Loss Discover Driven by Deep File Inspection (DFI)

Contents

InQuest Data-Loss Discovery

Data-Loss Discovery Driven by Deep File Inspection (DFI)

Executive Summary

The purpose of this whitepaper is to describe the capabilities provided by the InQuest platform related to identifying the exposure of sensitive information. With the recent explosion of data breach reports in the news, preventing the loss of sensitive data has become an area of focus for many organizations. If an attacker gains access to a protected network and begins exfiltrating sensitive information, the longer the breach goes undetected, the higher the damage to the organization. To evade detection of data exfiltration, hackers commonly obfuscate and embed stolen data within benign files and encrypted network flows. It is essential that data leakage is detected as soon as possible to minimize financial, reputational, intellectual property damage, and exposure.[1]

Figure 1: Exposure Score

In the first section, we provide a birds-eye view of the InQuest mission, team, and platform. We examine the functionality that empowers analysts with the ability to quickly and efficiently identify data exfiltration across their organization's boundaries. The InQuest solution to Data Leakage consists of four main steps: Observe, Dissect, Identify, and Alert. Then, we walk through some specific use cases relating to data-loss discovery, and capabilities used to provide additional visibility into sessions and files that traverse your continuous security monitoring points. Finally, we discuss the threat intelligence supporting breach detection and Command and Control (C2) events that are indicative of a compromise within your enterprise.

[1] https://inquest.net/blog/2020/04/21/Deep-File-Inspection-for-Data-Leakage

InQuest Company and Solution Overview

The InQuest platform provides high-throughput Deep File Inspection (DFI) for threat prevention, data-leakage discovery, and threat hunting.[2] We ingest, dissect, catalog, and retain files for real-time and retrospective analysis, leveraging the power of hindsight to apply today's threat intelligence to yesterday's data. Built by SOC analysts for SOC analysts, we empower defenders to save their organizations' most precious and limited commodity, human cognition, by democratizing advanced malware analysis skills to reduce analyst fatigue and frustration resulting in an increased return on investment with regards to personnel.

We aim to automate and scale the expert knowledge of a typical SOC analyst. Available on-premise, cloud-based, or as a service, the InQuest platform leverages a variety of sources in our automated decision-making engine. Sources include bi-directional orchestration with multi-scanning and sandbox platforms, unique threat intelligence sources, and a seasoned signature development team augmented by machine learning.[3]

The InQuest leadership and engineering teams are comprised of passionate security researchers hailing from both the public and private sectors.[4] Our mission is to deliver our decades of lessons-learned to protect users and organizations everywhere. We strive to maintain our hands-on familiarity with a wide range of prevention, analysis, and monitoring solutions, continuously exploring, examining, and validating security vendors. We continue our community involvement through contributions by way of talks, publications, open-source software, and threat research collaboration. Two unmatched advantages fuel our team's differentiators:

1. We've worked with thousands of real-world exploits from vulnerability discovery and exploitation specialists from all around the world.
2. We've vetted nearly every major security vendor under the Sun, having hands-on experience with the best of breed Commercial Off The Shelf (COTS) and Open Source Software (OSS) solutions across the spectrum.

Regardless of how the InQuest solution is deployed, our goals are to:

1. Reduce analyst frustration and fatigue by acting as a force multiplier to support the needs and scale of businesses ranging from small offices to global enterprises.
2. Expose deeply embedded malicious logic through novel methods, automating human-intensive tasks to democratize expert-level skill sets to a broader audience.
3. Utilize all analyst and threat intelligence resources available in customer environments to automate the identification and validation of threats and data theft.

For further details, or to get in touch, please visit us at www.inquest.net.

[2] https://inquest.net/empower-your-operations/deep-file-inspection
[3] https://inquest.net/blog/2018/11/14/Ex-Machina-Man-Plus-Machine
[4] https://inquest.net/company/leadership

Data-Loss Discovery

A variety of general patterns for sensitive and personally identifiable information are incorporated within the platform--for example, Social Security numbers, classified document watermarks, financial information, and more[5]. Forethought doesn't cover all data leakage. In such cases where sensitive information was leaked, and defenders want to tie that data back to the related network stream, RetroHunting can help. A supplementary user-defined signature with the relevant leaked keywords can be applied to the system. If these keywords are found anywhere in semantic, meta, or Optical Character Recognition (OCR) output layers, an alert is produced. The InQuest platform also offers an option to catalog all captured email bodies for historical analysis, providing yet another layer of analytics for the identification of data leakage.

Figure 2: Severe Data Loss

The InQuest platform provides functionality that empowers analysts with the ability to quickly and efficiently identify data exfiltration across their network boundaries. The InQuest solution to Data Leakage consists of four main steps: Observe, Dissect, Identify, and Alert/Block[6].

Observe

The InQuest Collector can be deployed on-premise, in the cloud, or through a SAAS offering to collect all traffic passing through the network boundary of a protected network. As traffic passes through the network boundary, the Collector captures it and reassembles network sessions from the captured packets. Once reconstructed, these sessions are passed on to InQuest's post-processing modules for dissection and analysis.

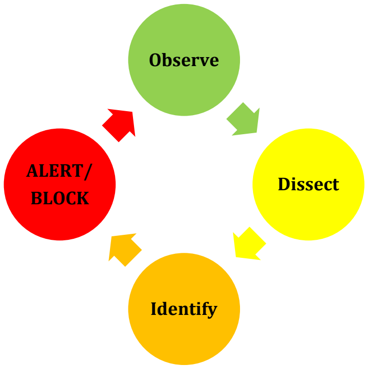


Figure 3: Methodology

Dissect

InQuest has developed proprietary dissection technology capable of processing the most common file types. This technology automatically identifies where data can be hidden within these file structures. The file dissection utility natively supports a variety of compression, encoding, and obfuscation techniques and automatically extracts embedded and obfuscated data hidden in files for further analysis. File dissection and post-processing are run recursively so that each extracted piece of hidden content is analyzed, providing protection against attackers using multiple levels of obfuscation or recursion to conceal data and guarantees that all hidden content is exposed for analysis.

[5] https://inquest.net/defend-the-enterprise/data-loss-discovery
[6] https://inquest.net/resources#use_case_14

Identify

Once dissection is complete, each piece of revealed data is tested against the full signature library of the InQuest system. In addition to the Data Leakage signatures provided by InQuest Labs, customers also can define and deploy custom signatures based on their specific needs. These user-defined signatures can be used for detecting sensitive data-in-transit or data-at-rest while enabling analysts to quickly identify and pinpoint the location of an attempted data exfiltration crossing their organization's boundaries.

User-defined signatures can be defined based on proprietary, sensitive, etc. information known only to the internal organization. Simple signatures may alert on the detection of typical markings for documents containing sensitive data ("FOUO", "SECRET", "PROPRIETARY", etc.). Other potential signatures may include account credentials, Social Security Numbers, or different types of Personally Identifiable Information (PII). The possibilities are endless and can be tailored to meet the needs of a particular organization.

Alert/Block

InQuest provides an intuitive and powerful user interface to enable analysts to quickly access data passing through their enterprise. The automated alerting functionality will notify an analyst if any of the currently defined Data Leakage signatures have triggered if the configuration to block the session is not enabled. Additionally, their associated data exposure levels are provided with immediate access to the associated network sessions, files, and post-processing tool results.

The Inquest User Interface also provides powerful search and query functionality against all of the data observed passing through the organizational boundary as well as the results of analysis engines. The UI can be used in the development and testing of new signatures to explore relationships among data and alerts and to determine the possible impact of a detected breach.

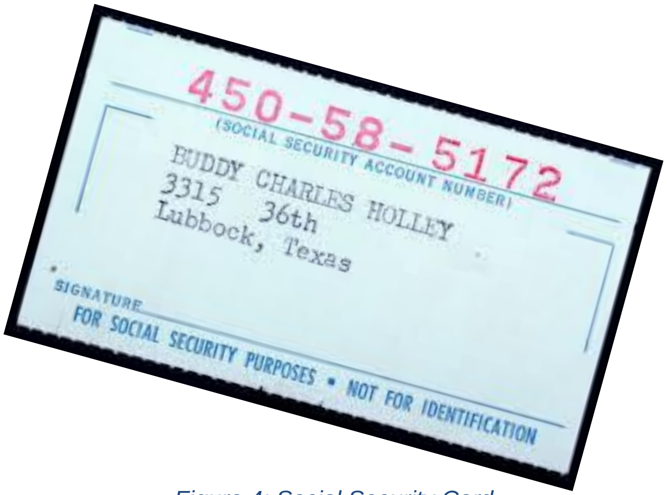
Managed Services

InQuest offers Threat Hunting Operations Augmentation for your organization. We supplement your staff with our team of subject matter experts that have been on the frontlines and in the trenches identifying and mitigating threats for nearly the last two decades. Services include 24x7 Continuous Threat Hunting and Monitoring while leveraging the InQuest Platform, and strategic integrations within your environment and more.[7]

[7] https://inquest.net/services

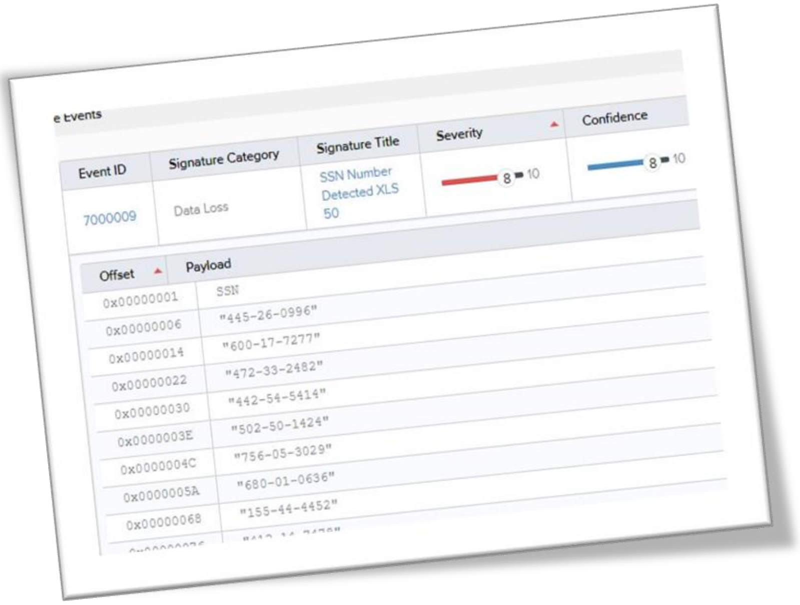
Personally Identifiable Information (PII)

Detecting Personally Identifiable Information (PII) is often regarded as the crux of use-cases relating to data-loss detection. Benchmark data breaches like Equifax's compromised personal info containing 145.5 million social security numbers have driven the bearing of organizational security posture.[8]



450-58-5172
(SOCIAL SECURITY ACCOUNT NUMBER)
BUDDY CHARLES HOLLEY
3315 36th
Lubbock, Texas
SIGNATURE
FOR SOCIAL SECURITY PURPOSES • NOT FOR IDENTIFICATION

Figure 4: Social Security Card



e Events

Event ID	Signature Category	Signature Title	Severity	Confidence
7000009	Data Loss	SSN Number Detected XLS 50	8 ▬ 10	8 ▬ 10

Offset	Payload
0x00000001	SSN
0x00000006	"445-26-0996"
0x00000014	"600-17-7277"
0x00000022	"472-33-2482"
0x00000030	"442-54-5414"
0x0000003E	"502-50-1424"
0x0000004C	"756-05-3029"
0x0000005A	"680-01-0636"
0x00000068	"155-44-4452"

Figure 5: SSNs Detected

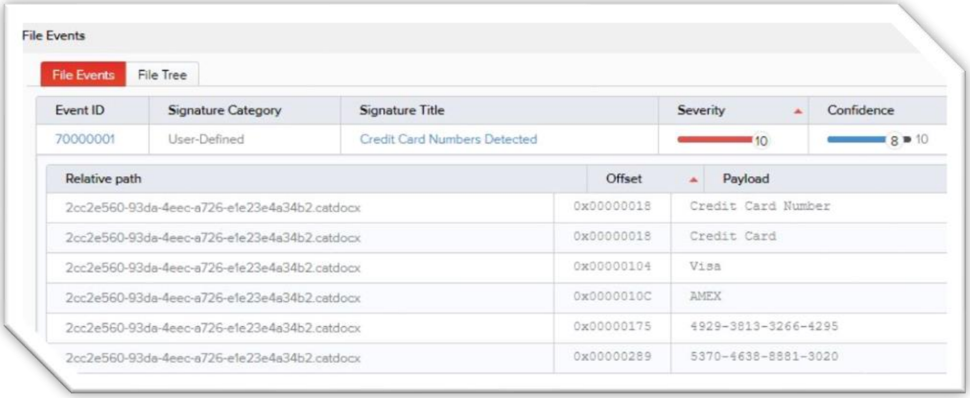
InQuest Data-Loss signatures have incorporated the ability to detect social security numbers and other instances of the disclosure of personal information. Relative to the threshold of occurrences within a session, the data loss score increases. The relative severity is beneficial to have for instances of single detection like a user emailing his W2 to himself during tax season. However, the severity increases when a significant culmination of sensitive information (like military deployment roster) is emailed unencrypted or to an external domain

Payment Card Information (PCI)

Exfiltration of Credit Card numbers and associated account information is often recognized as a prime objective of a threat actor when a breach has occurred. Regarding the 2007 benchmark case where TJX was breached, and at least 94 million Visa and MasterCard accounts were exposed and resulted in an estimated $256+ million.[9] Other notable credit card compromises include The Home Depot (56 Million), Heartland Systems (160 Million), and Capital one (106 Million).[10]

[8] https://www.popsci.com/social-security-number-equifax-leak/
[9] http://www.nbcnews.com/id/21454847/ns/technology_and_science-security/t/tjx-breach-could-top-million-accounts/.
[10] https://www.investopedia.com/news/5-biggest-credit-card-data-hacks-history/

While the detection of credit card numbers doesn't always signify the occurrence of a data breach, their appearance can identify poor practices, accidental exposure, storage, or transmission. The InQuest data-loss signatures have been tuned to determine the format of the respective vendors to reduce false-positives as well as incorporate a variety of different formatting techniques that are often used.



File Events

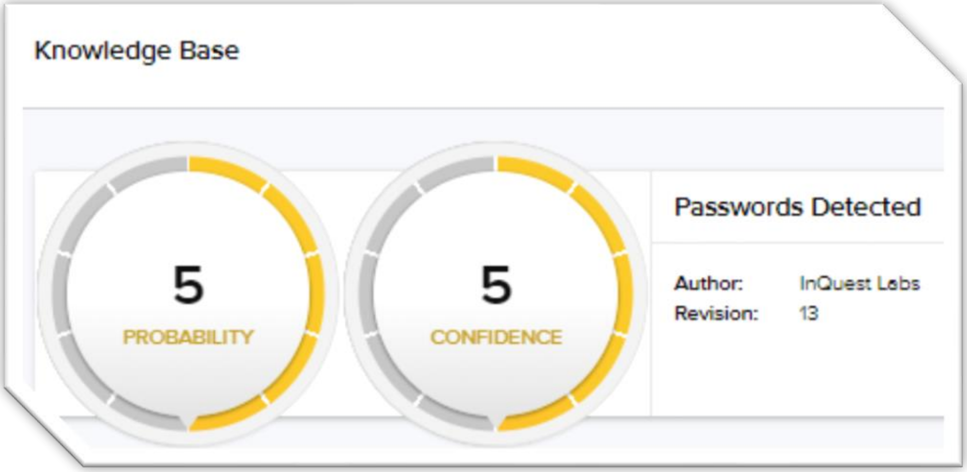
Event ID	Signature Category	Signature Title	Severity	Confidence
70000001	User-Defined	Credit Card Numbers Detected	10	8 10

Relative path	Offset	Payload
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x00000018	Credit Card Number
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x00000018	Credit Card
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x00000104	Visa
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x0000010C	AMEX
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x00000175	4929-3813-3266-4295
2cc2e560-93da-4eec-a726-e1e23e4a34b2.catdocx	0x00000289	5370-4638-8881-3020

Figure 6: Credit Card Numbers Detected

Account Credentials

The prevalence of using an unencrypted document as a password safe has been observed within a few customer environments. While this is a poor security practice for that individual, it can be more concerning when credentials for internal accounts are saved locally to a file or transported insecurely within a file. InQuest has pre-defined signatures that can successfully detect the presence of username and password combinations at-rest or in-motion. While these types of signatures have traditionally been prone to false positives as well as false negatives, InQuest Labs consistently tunes/tightens these signatures to ensure their accuracy and fidelity.



Knowledge Base

5 PROBABILITY

5 CONFIDENCE

Passwords Detected

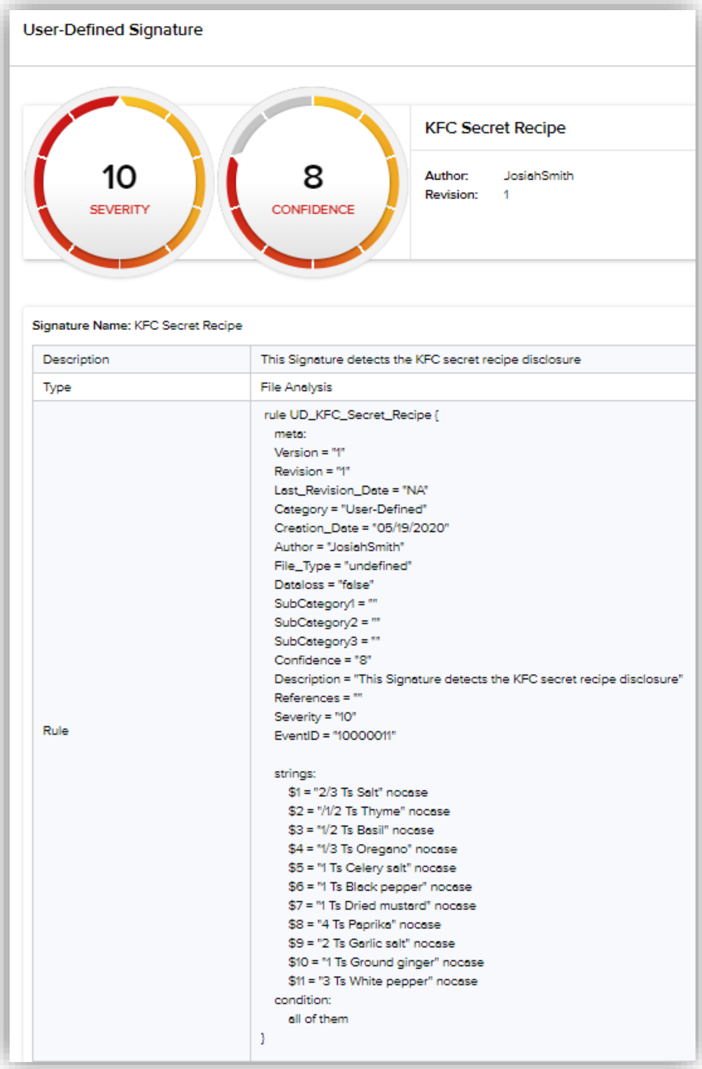
Author: InQuest Labs
Revision: 13

Figure 7: Passwords Detected

Proprietary Data / User-Defined

Not all information deemed sensitive to an organization relates to credit card numbers or social security numbers. Consider proprietary data or the type of “Crown Jewels” your company is safeguarding.

Exfiltration/negligent disclosure can be detected in these situations by developing user-defined signatures to alert on the specifics. Take this user-defined signature that identifies the secret recipe for KFC chicken. The data loss strings were defined, and the occurrence of all of them will produce medium-high confidence, high severity data-loss alert.



User-Defined Signature

10 SEVERITY

8 CONFIDENCE

KFC Secret Recipe

Author: JosiahSmith
Revision: 1

Signature Name: KFC Secret Recipe

Description	This Signature detects the KFC secret recipe disclosure
Type	File Analysis

Rule

```
rule UD_KFC_Secret_Recipe {
  meta:
  Version = "1"
  Revision = "1"
  Last_Revision_Date = "NA"
  Category = "User-Defined"
  Creation_Date = "05/19/2020"
  Author = "JosiahSmith"
  File_Type = "undefined"
  Dataloss = "false"
  SubCategory1 = ""
  SubCategory2 = ""
  SubCategory3 = ""
  Confidence = "8"
  Description = "This Signature detects the KFC secret recipe disclosure"
  References = ""
  Severity = "10"
  EventID = "10000011"

  strings:
     $1 = "2/3 Ts Salt" nocase
     $2 = "/1/2 Ts Thyme" nocase
     $3 = "1/2 Ts Basil" nocase
     $4 = "1/3 Ts Oregano" nocase
     $5 = "1 Ts Celery salt" nocase
     $6 = "1 Ts Black pepper" nocase
     $7 = "1 Ts Dried mustard" nocase
     $8 = "4 Ts Paprika" nocase
     $9 = "2 Ts Garlic salt" nocase
     $10 = "1 Ts Ground ginger" nocase
     $11 = "3 Ts White pepper" nocase
  condition:
     all of them
}
```

Figure 8: User-Defined Signature

RetroHunting

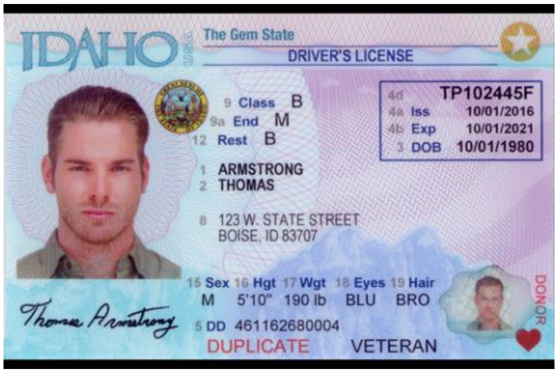
A useful feature InQuest provides for detecting Data-Loss is RetroHunting. RetroHunting provides visibility into the 4th dimension of time by scanning previously captured files with user-defined or newly released signatures from InQuest Labs. RetroHunt® can detect previously data-loss and/or threat events, in either a manual or automated fashion.[11]

Another use case for RetroHunting is utilizing the platform to test the performance of a data-loss signature on production-grade data without causing network degradation or overwhelming your security operations team with false positives.[12] Potential signatures can be added to the platform and tested against real-world data for tuning and acceptance. Results from the test can be reviewed, signatures tweaked, then re-ran for further improvements as needed. An integrated solution for organizations who wish to replicate a workflow commonly utilized by IDS/IPS signature developers, but lack the resources required to implement a custom solution.

[11] https://inquest.net/blog/2018/05/09/retrohunting-with-inquest
[12] http://go.inquest.net/whitepapers/RetroHunting.pdf

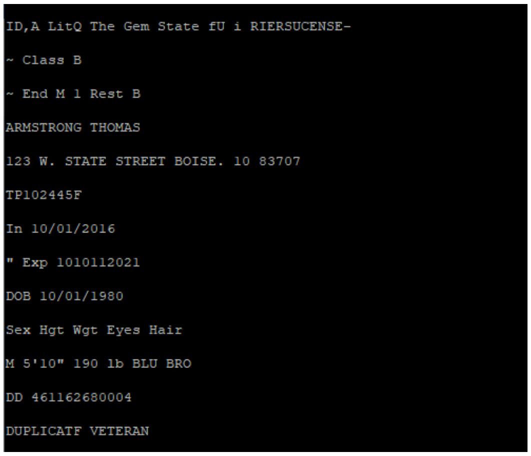
Optical Character Recognition

Beyond the capability of identifying, extracting, and alerting on sensitive content from hundreds of file types, InQuest Deep File Inspection (DFI) utilizes machine vision and optical character recognition (OCR) to identify relevant information within scanned documents, and other images used innocuously or in attack lures.[13] Some practical examples of this technology would be to alert on watermarked documents or pictures of identification cards such as Social Security Cards or Drivers Licenses. Another exemplar falls into scanned documents, from financial applications, security clearance questionnaires, or a myriad of other materials that were filled out by hand, scanned, and delivered. Data-loss signatures can be anchored on a variance of contextual data. Consider the following image; its derived OCR text highlights how you could alert on "DOB," license number, or the document discriminator. Pairing the data points together, or the format they are in, the confidence of the alert can increase within the monitoring environment.



IDAHO USA The Gem State
DRIVER'S LICENSE
9 Class B
9a End M
12 Rest B
4d TP102445F
4a Iss 10/01/2016
4b Exp 10/01/2021
3 DOB 10/01/1980
1 ARMSTRONG
2 THOMAS
8 123 W. STATE STREET
BOISE, ID 83707
15 Sex 16 Hgt 17 Wgt 18 Eyes 19 Hair
M 5'10" 190 lb BLU BRO
5 DD 461162680004
DUPLICATE VETERAN
DONOR

Figure 9: Driver's License Image



```
ID,A LitQ The Gem State fU i RIERSUCENSE-

~ Class B

~ End M 1 Rest B

ARMSTRONG THOMAS

123 W. STATE STREET BOISE. 10 83707

TP102445F

In 10/01/2016

" Exp 1010112021

DOB 10/01/1980

Sex Hgt Wgt Eyes Hair

M 5'10" 190 lb BLU BRO

DD 461162680004

DUPLICATF VETERAN
```

Figure 10: OCR Output

Data-at-Rest and Compliance

Many regulations require you to scan your data stores or file-shares for unencrypted or unprotected data-at-rest, so you might wonder why a data-loss methodology and execution strategy wouldn't start there. But the truth is, auditors are more concerned with the fact that you are complying with the surveys, than what your findings are. So, scanning for data-at-rest is essential for compliance, but not the primary objective or value of your data-loss control.

Scanning data-at-rest also presents the opportunity to identify staging areas used by attackers before they attempt to exfiltrate your sensitive data. They are often placed on arbitrary file-shares en masse at a centralized location. Recognizing the sensitive information at this phase can break the chain and stop the attackers from achieving their

[13] https://inquest.net/blog/2020/05/12/Detecting-Coercive-Lures-with-OCR

objectives. Of course, there is also accidental placement and forgotten data repositories that were acquired and stored long ago

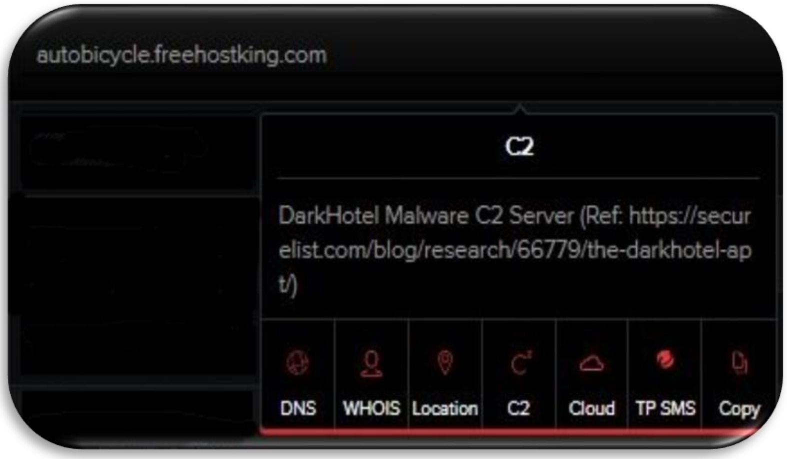
Threat Intelligence

Coupled with DFI, InQuest Labs publishes threat intelligence updates on a regular weekly or as-needed basis. These updates include IP, domain, and SSL certificate Indicators of Compromise (IoC), as well as additions/modifications/removals to the InQuest data-loss and threat detection knowledge base. We regularly update the InQuest scoring algorithm, which is the component responsible for producing an all-encompassing data-loss or threat score (ranging from 0 through 10).

The InQuest Labs team sources threat intelligence through a variety of methods and origins, including proprietary harvesting methods, commercial feeds and partnerships, unique partnerships, and open-source intelligence (OSINT).

Command and Control Communications

Malicious software often seeks to gain control of your systems and establish command-and-control communications to initiate processes such as exfiltrating targeted data. If a zero-day exploit has been used, there is typically no signature that can be utilized to identify the exploit and stop it before it compromises your systems. Detecting anomalous command-and-control communications is key to dealing with attacks of this type to provide your SOC staff with the information they need to identify the breach quickly and perform incident response promptly.



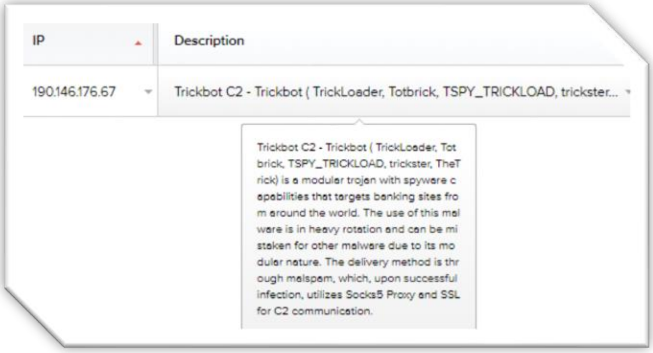
autobicycle.freehostking.com

C2

DarkHotel Malware C2 Server (Ref: https://securelist.com/blog/research/66779/the-darkhotel-apt/)

Figure 9: C2 DNS Reference

InQuest's platform continuously monitors command and control (C2) communications (DNS and IP) for signs of unusual activity. Keeping abreast of the latest adversary C2 infrastructure through threat intelligence is critical for detecting this activity. Our C2 detection engine alerts you if any of those nodes are seen touching your network. So, we not only focus on what is being said but also who is saying it. The InQuest Labs Team publishes daily updates of known C2 IP addresses and domains globally, which are then flagged in our UI for further investigation.



IP	Description
190.146.176.67	Trickbot C2 - Trickbot (TrickLoader, Totbrick, TSPY_TRICKLOAD, trickster...

Trickbot C2 - Trickbot (TrickLoader, Tot
brick, TSPY_TRICKLOAD, trickster, TheT
rick) is a modular trojan with spyware c
apabilities that targets banking sites fro
m around the world. The use of this mal
ware is in heavy rotation and can be mi
staken for other malware due to its mo
dular nature. The delivery method is thr
ough malspam, which, upon successful
infection, utilizes Socks5 Proxy and SSL
for C2 communication.

Figure 10: C2 IP Reference

The ability to identify anomalies in C2 communications quickly enables your SOC staff the ability to rapidly respond to prevent exfiltration of sensitive information like company proprietary information, account credentials, PII, etc.

Conclusion

In this paper, we provided readers with high-level background information regarding InQuest as a company and platform. We introduced a methodology for data-loss discovery and then dove into some specific use-cases showcasing the platform's ability to leverage data leakage signatures that are developed and maintained by InQuest Labs as well as user-defined signatures developed by our customers

The competencies that are augmented with RetroHunting and Optical Character Recognition (OCR) compliment the detection of historical events, tuning signatures, and contribute unparalleled insight into the text embedded within graphical assets.

Implementation of actionable Threat Intelligence coupled with indicators of compromise in the form of Command and Control alerting is pivotal for timely detection and response to a data breach within your organization.

In conclusion, the InQuest platform is capable of identifying data loss and C2 activity associated with a variety of threat actor groups by performing Deep File Inspection (DFI), behavioral analytics, and leveraging unparalleled threat intelligence.