University of BRISTOL

CyBOK

Cyber Security Body of Knowledge

Forensic Knowledge Area

Issue 1

Author - **Vassil Roussev**

Presentation - **Russell May**

bristol.ac.uk

University of BRISTOL

CyBOK

The CyBOK project would like to understand how the CyBOK is being used and its uptake. The project would like organisations using, or intending to use, CyBOK for the purposes of education, training, course development, professional development etc. to contact it at contact@cybok.org to let the project know how they are using CyBOK.

bristol.ac.uk

Introduction

- Digital Forensics
- Legal Aspects
  (Limited to general principles)

Definitions

Digital forensics is considered the application of science to the identification, collection, examination, and analysis of data while preserving the integrity of the information and maintaining a strict chain of custody for the data. Data refers to distinct pieces of digital information that have been formatted in a specific way.

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

"Guide to integrating forensic techniques into incident response." 2006

Definitions and Conceptual Models

- **Locard's exchange principle**
  Physical contact between objects inevitably results in the exchange of matter, leaving *traces* that can be analysed to (partially) reconstruct the event.
- **Cyber Domain**
  A persistent *digital (forensic) trace* is not inevitable.
- **Digital Traces**
  Although they frequently exist, these traces of cyber interactions are the result of conscious engineering decisions that are *not* usually taken to specifically facilitate forensics.

UK Computer Law and Digital Evidence

- Types of computer misuse
- ACPO Good Practice Guide when gathering digital evidence.
- Forensic Science Regulator

bristol.ac.uk

Types of Computer Misuse

- Hacking.
- Data misuse and unauthorised transfer or copying of data.
- Copying and distributing copyrighted software, music and film.
- Email and social media abuses.
- Pornography, child abuse.
- Identity and financial abuses.
- Viruses and malware.

Forensic examination is also equally relevant to non-cyber crimes.

UK Computer Law

The **Computer Misuse Act 1990** is intended to secure computer material against unauthorised access or modification. Offences:

- **S1** - Unauthorised access to computer material.
- **S2** - Unauthorised access with intent to commit or facilitate commission of further offences.
- **S3** - Unauthorised acts with intent to impair or with recklessness as to impairing operation of a computer.
- **S3ZA** - Unauthorised acts causing, or creating a risk of, serious damage.
- **S3A** - Making, supplying or obtaining articles for use in offence under Section 1, 3, or 3ZA

UK Computer Law

Police & Criminal Evidence Act 1984
and
Criminal Justice & Police Act 2001

- Address computer-specific concerns with respect to warrants, search and seizure.
- Presentation of computer records in court.

Regulation of Investigatory Powers Act 2000

- Specifies the circumstances under which individuals are legally required to disclose encryption keys.

Note: There is a separate Law and Regulation CyBOK knowledge area.

Good Practice Guide for Digital Evidence

Four Principles

1. No action taken by law enforcement agencies, persons employed within those agencies or their **agents** should change data which may subsequently be relied upon in court.
2. In circumstances where a person finds it necessary to access original data, that person must be **competent** to do so and be able to give evidence explaining the relevance and the implications of their actions.
3. An **audit trail** or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
4. The person in charge of the investigation has overall responsibility for ensuring that the law and these principles are adhered to.

bristol.ac.uk

ASSOCIATION OF CHIEF POLICE OFFICERS

Forensic Science Regulator

ISO Standards

In the UK, the **Forensic Science Regulator** mandates that any **provider of digital forensic science** must be accredited to;

- **ISO 17020** (2012) for any crime scene activity
- **ISO 17025** (2005) for any laboratory function

The ISO certification attests to the quality and rigour of the processes followed in performing the forensic examination.

Conceptual Models

- Cognitive Task Model
- Bottom-up Processing
- Top-down Processing

Conceptual Models

- **STATE Centric**

  The starting point for state-centric approaches is a snapshot of the state of the system of interest; for example, the current content of a hard drive or another storage medium.

- **LOG Centric**

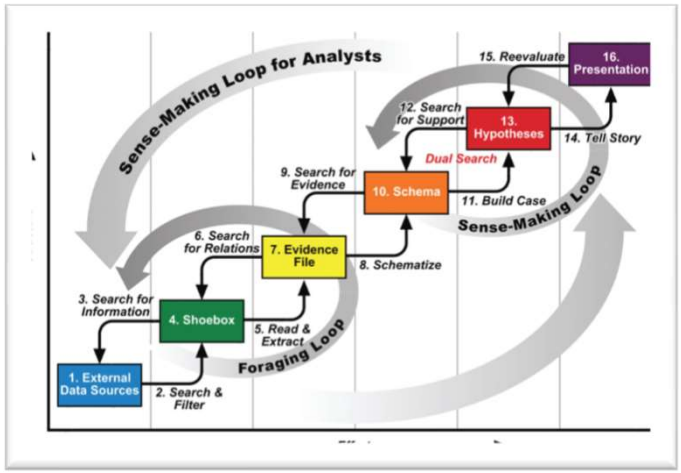
  An OS (Operating System) maintains a variety of monitoring logs that detail various aspects of the operation of the OS kernel and different applications.

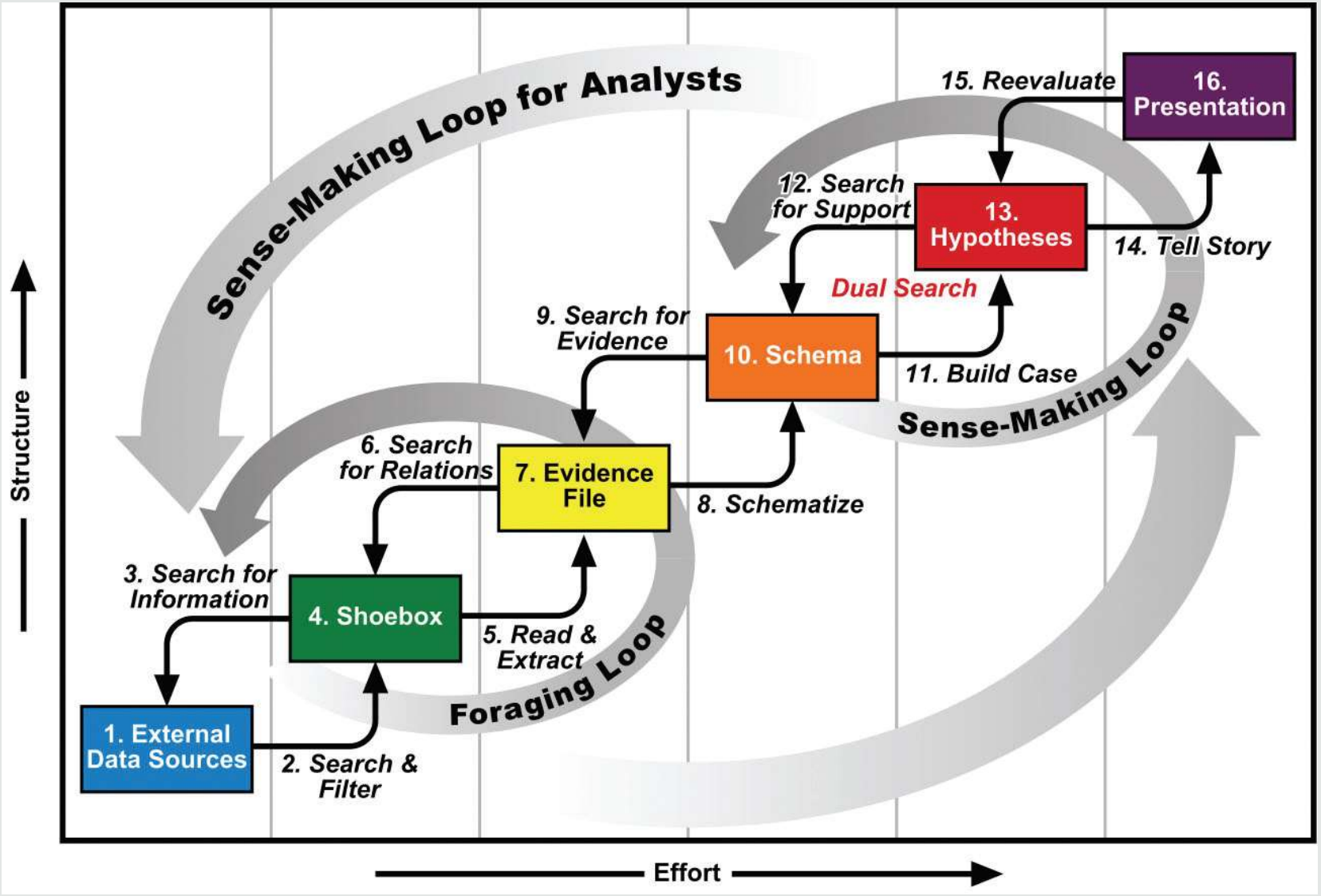
Cognitive Task Model

The benefit of using this model is that:

- It provides a reasonably accurate description of the investigative processes and allows us to map the various tools to the different phases of the investigation
- It provides a suitable framework for explaining the relationships of the various models developed within the area of digital forensics
- It can seamlessly incorporate information from other lines of the investigation.

The information transformation processes in the two loops can be classified into bottom-up (*organising data to build a theory*) or top-down (*finding data based on a theory*) ones.





Cognitive Task Model

Cognitive Tasks - Bottom-up Processes

- **Search and filter**
  Data sources are searched to eliminate irrelevant data.
- **Read and extract**
  Collections in the shoebox are analysed to extract individual facts and relationships that can support or disprove a theory.
- **Schematise**
  For example, Timeline Analysis
- **Build case**
  Analysis yields theories or hypotheses that can explain the evidence
- **Tell story**
  Final report of forensic examination.

Cognitive Tasks - Top-Down Processes

- **Re-evaluate**
  Feedback may necessitate re-evaluating evidence or pursuing alternative theories.
- **Search for support**
  A hypothesis may need more facts to be of interest and, ideally, would be tested against possible alternative hypotheses.
- **Search for evidence**
  Analysis of theories may require the re-evaluation of evidence to ascertain its significance/provenance, or it may trigger the search for more/better evidence.
- **Search for relations**
  Pieces of evidence in the file can suggest new searches for facts and relations on the data.
- **Search for information**
  The feedback loop from any of the higher levels can ultimately cascade into a search for additional information; this may include new sources, or the re-examination of information that was filtered out during previous passes.

Forensic Process

Results of a forensic investigation must be admissible in a court of law.

- Data Provenance and Integrity
- Scientific Methodology
- Forensic Procedure
- Tool Validation
- Triage

Data Provenance and Integrity

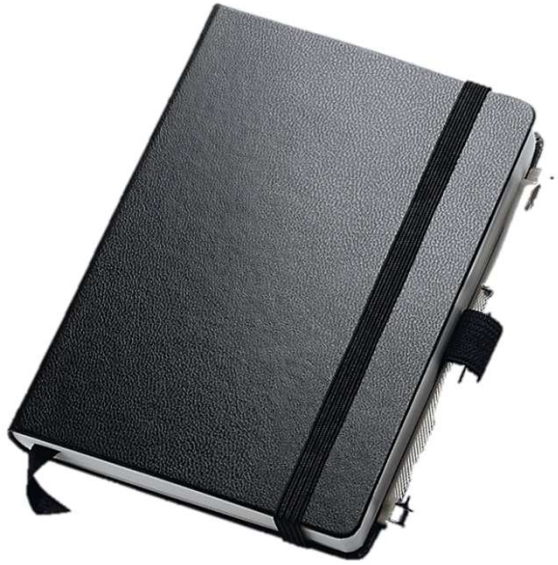
- The traditional gold standard is a bit-level copy of the forensic target media, which can then be analysed using knowledge of the structure and semantics of the data content.
- Copy(s) to be made with validated forensic tools.
- As storage devices increase in complexity and encryption becomes the default data encoding, it is increasingly infeasible to obtain a true physical copy of the media and a logical or partial acquisition may be the only possibility.

ACPO Principle 1 applies.

Scientific Methodology

- The notion of ***reproducibility*** is central to the scientific validity of forensic analysis.

- A third party starting with the same data and following the same process described in the case notes should obtain the same result.



ACPO Principle 3 applies

Forensic Procedure

- The organisational aspect of the forensic process, which dictates how evidence is acquired, stored, and processed is critical to the issue of admissibility in a court of law.

Triage - Devices

Digital forensic triage of individual live computer devices, using forensic tools, can assist in determining whether;

- A device is likely to contain evidence pertinent to the investigation.
- A suspect, witness and or victim has logged into the device.
- A specific file or file type exists on the storage media.
- Any of the storage media is encrypted and will require decryption keys. If the media is found to be encrypted a decision can be made as to whether;
  - To shutdown the device.
  - Do a live acquisition.
  - Export the encryption keys (if present).
  - Ascertain password(s) or backup encryption keys.

Triage - Data

- The volume of data contained by a forensic target typically far exceeds the amount of data relevant to an inquiry.
- Therefore, in the early stages of an investigation, the focus of the analysis is to quickly identify the relevant data and filter out the irrelevant.
- Legally, there can be several constraints placed on the triage process;
  - The nature of the investigation.
  - Inherent privacy rights.
  - Jurisdiction.

Forensic Analysis

- Operating Systems
- Storage Media
- Data Abstraction Layers
- Acquisition Methods
- Encryption Concerns
- File System Data Recovery
- Memory Acquisition
- Application Forensics

Operating System Analysis

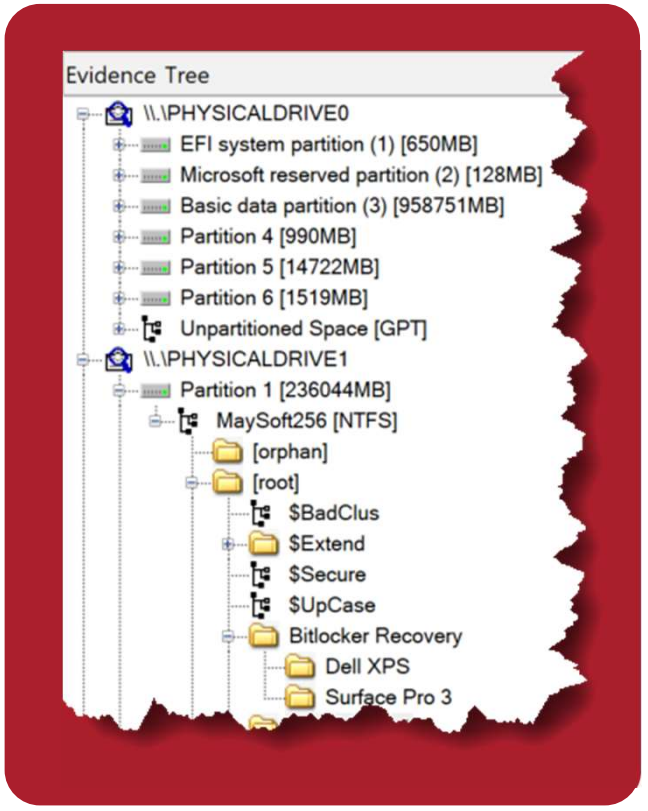
Operating System Analysis employs knowledge of how operating systems function in order to reach conclusions about events and actions of interest to the case.

Storage Forensics

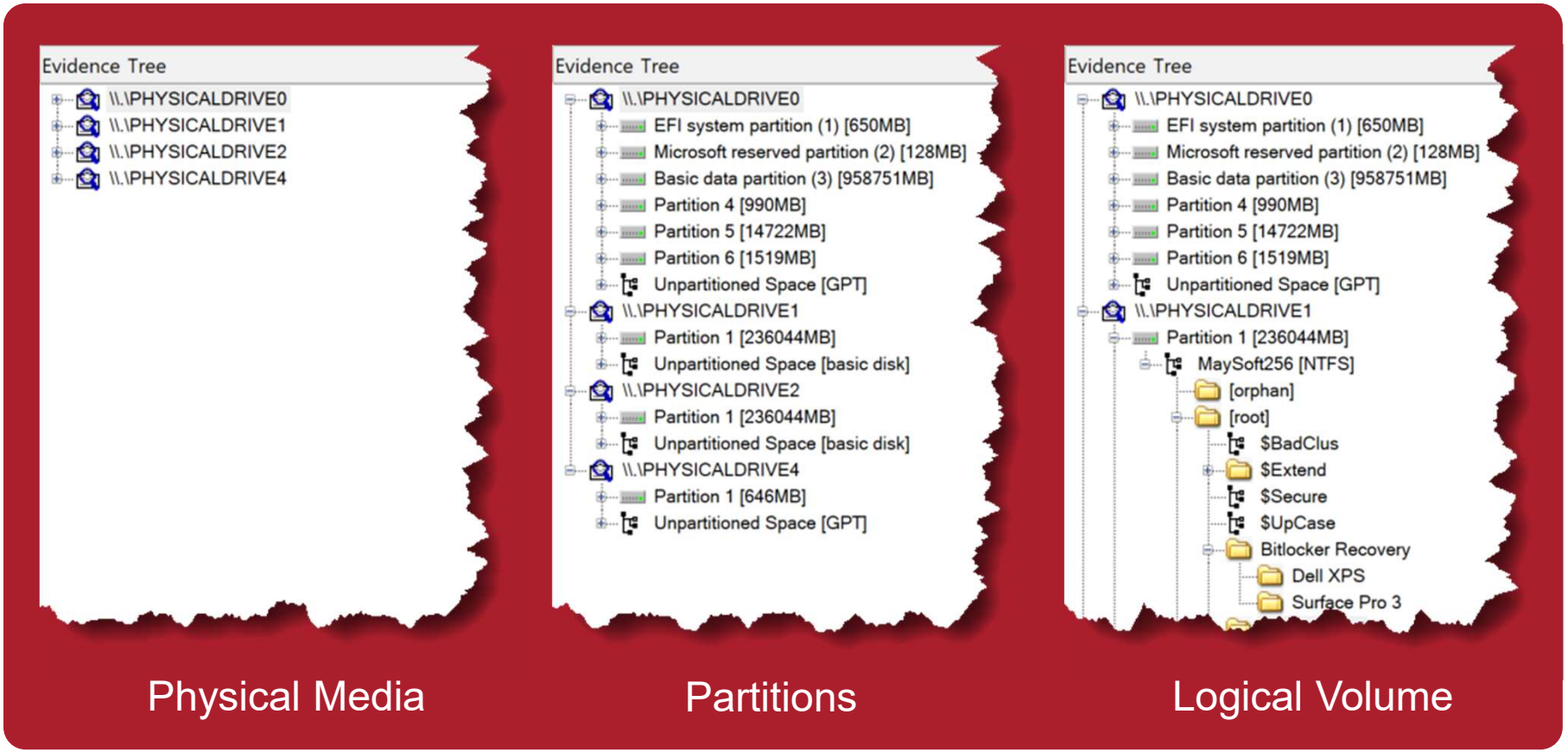
Persistent storage in the form of Hard Disk Drives (HDDs), Solid State Drives (SSDs), optical disks, external storage media etc. is the primary source of evidence for most digital forensic investigations.

Data Abstraction Layers

Computer systems organise raw storage in successive layers of abstraction – each software layer builds an incrementally more abstract data representation that is only dependent on the interface provided by the layer immediately below it. Accordingly, forensic analysis of storage devices can be performed at several levels of abstraction.



Data Abstraction Layers



Snapshots used above taken from FTK Imager.

Physical Acquisition

Physical acquisition is the process of obtaining the data directly from hardware media, at disk or partition level, using a forensic imaging tool and write blocking hardware or software.

- Obtaining data from the lowest level system interface available and independently reconstructing higher-level artifacts is considered the most reliable approach to forensic analysis.
- In some cases, it is necessary to perform additional recovery operations before a usable copy of the data is obtained. One common example is RAID storage devices.
- When possible a hardware write blocker is used between source and target media to eliminate the possibility accidental modification of the target.
- There are exceptions, in cases where it is not practical to shut down the target, a media image maybe obtained while the system is live.

File System Acquisitions

File System Acquisitions rely on a file system of the device to return files and will not include deleted, hidden or system files and no unallocated space.

- Some network storage devices are only accessible using remote access, as these may be mission critical it maybe impracticable to take them off-line for physical imaging.
- Some mobile devices provide an API either the devices file system(s) or subsets of specific file types, such as; Pictures, SMS Messages, Call Logs.
- File system acquisitions will not normally include deleted files, hidden files, system files or file fragments.
- File system acquisitions may change file metadata like file accessed time and date.

ACPO Principle 2 applies

Data Acquisition General

- Cryptographic hashes are computed for the entire image and (preferably) for every block; the latter can be used to demonstrate the integrity of the remaining evidence if the original device suffers a partial failure, which makes it impossible to read its entire contents.
- The National Institute of Standards and Technology (NIST) maintains the Computer Forensic Tool Testing (CFTT) project, which independently tests various basic tools, such as write blockers and image acquisition tools and regularly publishes reports on its findings.

Encryption Concerns

Apart from having the technical capability to safely interrogate and acquire the content of a storage device, one of the biggest concerns during data acquisition can be the presence of encrypted data.

By definition, a properly implemented and administered data security system, which inevitably employs encryption, will frustrate efforts to acquire the protected data and, by extension, to perform forensic analysis.

There are two possible paths to obtaining encrypted data

- **Technical** – Subverting the encryption system.
- **Legal** – Compelling the surrender of encryption keys.

Filesystem Analysis

What is a file system.

- Storage media organise data into sectors or pages which in turn are organised into clusters or blocks.
- A filesystem controls how data is stored and retrieved from a storage media. Data streams are stored as files in directories. As well as files directories can contain additional directories.
- A filesystem provides an interface that is used by applications to store and retrieve files by name without any concern for the physical storage method employed or the layout of the data content.
- A logical volume is a collection of one or more physical volumes presented and managed as a single unit.

File Systems – Data Recovery

Data Recovery – Spinning Disk Media

- Recovery of deleted files, yet to be overwritten by new files.
- Recovery of directories containing meta data may yield further information relating to data carved files.
- Slack Space Recovery - Slack space is the difference between the *allocated* storage for a data object, such as file, or a volume, and the storage in actual use.
- Comparing the data carved files hash with a library may assist in file type identification.
- File Signature Analysis can also be used to assist in file type identification.

File Systems – Data Recovery

Data Recovery - Solid State Media

- It has been established experimentally that file carving would only work in a narrow set of circumstances on modern Solid State Drives.
- **TRIM-aware** operating systems overwrite deleted data with a rapidity that data recovery rates in tests were almost universally zero.
- **Pre-TRIM-aware** operating systems allows for near-perfect recovery rates under the same experimental conditions.
- Solid State Drive pages need to be reset before they can be reused. This reset operation overwrites any existing data in the page.

Main Memory Forensics – Live Forensics

Live Forensics involves capturing volatile memory prior to system shutdown. There is a wealth of information about a system's run-time state that can be readily extracted, even from a snapshot in time.

- **Memory Dumps**
  - Running Processes
  - File Information
  - Network Connections
  - Artefacts and Fragments
- **Real time Analysis**
  - A trusted agent designed to allow remote access over a secure channel is pre-installed on the system.
  - The remote operator has full control over the monitored system and can take snapshots of specific processes, or the entire system.

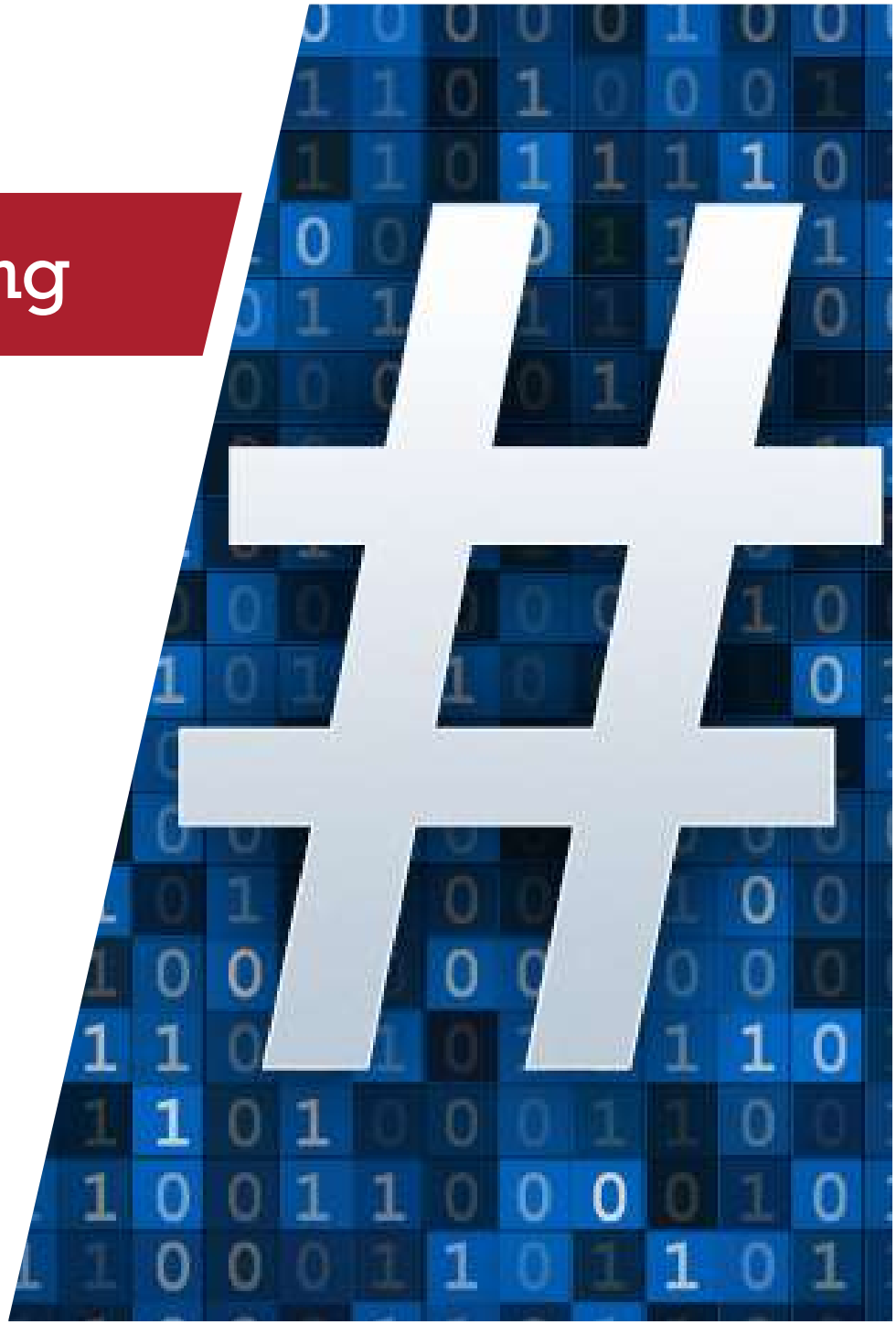
Application Forensics

Application forensics is the process of establishing a data-centric theory of operation for a specific application. For example;

- Making sense of relational database.
  - Reconstructing tables.
  - Rebuilding relational data to make it human readable.
  - Potential to recover embedded deleted record data.
- Internet Browsers
  - Visited websites
  - Local file cache
  - Files downloaded
  - Cookies
  - HTML 5 Local storage
  - Form data

Cryptographic Hashing

- Exact Match
- Approximate Matching
- Hashing for Data Carving



Cryptographic Hashing

A cryptographic hash function is a function which takes an input or arbitrary length (or 'message') and returns a fixed-size string of bytes. The string is called the 'hash value', 'message digest', 'digital fingerprint', 'digest' or 'checksum'.

The ideal hash function has three main properties:

- It is extremely easy to calculate a hash for any given data.
- It is extremely unlikely that two slightly different messages will have the same hash.
- It is extremely computationally difficult to calculate an alphanumeric text that has a given hash.

MD5	SHA1	FileNames
596efecb7e87ef5d9149190062249c33	b6fe6a26a63a7269a97d1b0d295a7dba6356e9ff	C:\\OS [NTFS]\[
11208f5385fa23fe5a4cb359adc517cd	b4e1fe914927b1960cfca7356bd1df6724252aff	C:\\OS [NTFS]\[

File Hash Analysis - Exact Matching

Finding a known file using cryptographic hashing.

If two files have identical hash values, then the files themselves are identical. This knowledge can be used to;

- **Identify known file types**
- **Identify known evidential files**

Each file has a unique hash value. Hash values are combined into sets (e.g., A digest of hash values for files comprising a single software application). Hash sets are combined into libraries (e.g., ‘Hacker Tools’).

Automated hash analysis will rapidly identify files and report the library and set they were found in.

File Hash Analysis - Proximity Matching

Sometimes called “**Fuzzy Hashing**”.

- **Resemblance queries**
  - compare two comparably-sized data objects (peers) and seek to infer how closely related they are.
- **Containment queries**
  - compare artifacts that have a large disparity in terms of size and seek to establish whether a larger one contains (pieces of) a smaller one.

The results returned for each are numeric for instance using 0 as not comparable and 100 as very comparable.

It is important for analysts to put the tool results into the correct context and to understand the performance envelope of the tools they are using in order to correctly interpret the results.

Hash Analysis - Data Carving

Block level analysis, hashing each cluster or page of unallocated storage can assist the analyst to identify known blocks before commencing any data carving.

This can improve results by reducing gaps and eliminating certain classes of false positive results.

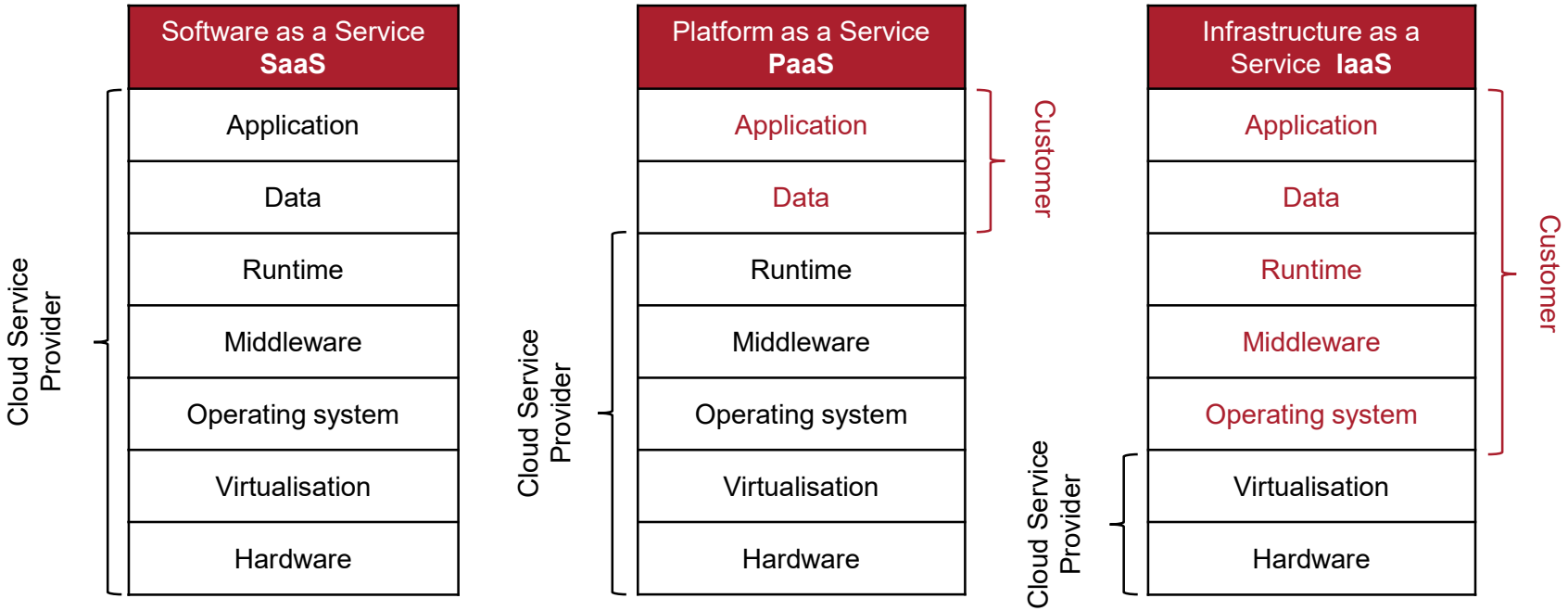
Cloud Computing

Cloud Basics

- Conceptually, cloud-based IT abstracts away the physical compute and communication infrastructure and allows customers to rent as much compute capacity as needed.
- Cloud computing services are commonly classified into one of three canonical models
  - **SaaS** - Software as a Service
  - **PaaS** - Platform as a Service
  - **IaaS** - Infrastructure as a Service

Cloud Computing

Layers of cloud computing environment owned by customer and cloud service provider on three service models



Cloud Computing Forensic Challenges

- Physical acquisition is not applicable to the cloud, logical acquisition is normal.
- The cloud is the authoritative data source.
- Logging is pervasive.
- Distributed computations are normal.

Cloud Computing - Drive Acquisition

Cloud drive services, such as ‘Dropbox’, ‘Google Drive’ and ‘Microsoft OneDrive’ are the SaaS version of the local storage device, which is central to modern digital forensics.

- **Partial replication.**

  The acquisition is of an unknown quality, subject to potentially stale and omitted data.

- **Revision acquisition.**

  Revisions reside in the cloud and clients rarely have anything but the most recent version in their cache; a client-side acquisition will clearly miss prior revisions, and does not even have the means to identify these omissions.

- **Cloud-native artifacts.**

  For example, Google Docs documents are stored locally as a link to the document which can only be edited via a web app.

Cloud Native Artefacts

Cloud artifacts often have a completely different structure from traditional snapshot-centric encoding.

For example, internally, Google Docs' documents are represented as the complete history (log) of every editing action performed on it; given valid credentials, this history is available via Google Docs' internal API.

It is also possible to obtain a snapshot of the artifact of interest in a standard format such as a PDF, via the public API. However, this is inherently forensically deficient in that it ignores potentially critical information on the evolution of a document over time.

Conclusion

Digital forensics identifies and reconstructs the relevant sequence of events that has led to a currently observable state of a target IT system or (digital) artifacts.

- The provenance and integrity of the data source and the scientific grounding of the investigative tools and methods employed are of primary importance in determining their admissibility to a court of law's proceedings.
- Digital forensic analysis is applied to both individual digital artifacts such as files and to complex IT systems comprising multiple components and networked processes.

Following the rapid cloud-based transition from Software as a Product (**SaaP**) to Software as a Service (**SaaS**), forensic methods and tools are also in a respective process of transition.

University of BRISTOL

CyBOK

Cyber Security Body of Knowledge

Forensic Knowledge Area

Issue 1

Author - **Vassil Roussev**

Presentation - **Russell May**

bristol.ac.uk