Digital Forensics

Marian Svetlik
svetlik@df-pro.cz
svetlik@fi.muni.cz
www.digital-forensic.pro

Digital Forensics Course Concept

Marian Svetlik

- Expert Witness in Digital Forensics
- Information Security Expert
- Vice-president a CEO of The Academy of Forensic Sciences
- Digital Forensic Review - Journal Editor
- ISMS Lector at University of Economics Prague
- Comuter Crime Lector at University of Finance and Administration Prague
- Cybercrime Lector at CEVRO Institute
- Digital Forensic Special Expert C4e at MUNI
- Programme Committee member of the DFRWS EU
- IDFA Management Board Member

Course Content

- DF definition, relation to the cybersecurity and to the cybercrime
- Digital Traces & Digital Evidence, properties, documentation
- Sources, Handling, Gathering and Protection
- DF Examination Principles
- DF Lab creation and management, Assessment, Certification, Accreditation
- DF in Law, Electronic Evidence

Recap

- Digital Forensic vs. Digital Forensics
- Digital Forensic Science Definition
  - **Digital Forensic Science** is an exact forensic science that examines the processes and patterns of the origin, existence and extinction of digital information and interprets this knowledge to explain the processes associated with it for the purposes of forensic examination.
- Digital Forensic Analysis Definition
  - **Digital Forensic Analysis** is the process of applying scientifically justified and proven methods to examine digital traces for decision-making by government agencies (e.g. police investigators, prosecutors and judges but other state bodies too) and other legal entities (e.g. organizations and private persons) for purposes of legal acts.
- Digital Forensic Analysis Basic Properties
  - Independency, Professionalism, Repeatability, Reviewability, Integrity, Legality, Documentation
- Expert vs. Forensic Expert

Today outline

- Digital Traces
- Digital Evidence, properties, documentation

Digital Trace

A digital trace constitutes a cultural artefact in a virtual settlement. **It is some form of evidence of the activity of a user.** Digital traces may be records stored in databases or log files which can be accessed by researchers...

https://www.igi-global.com/dictionary/excavating-business-intelligence-from-social-media/39719

1st Break

... Digital trace is some form of evidence...

Information

Information is a term that characterizes reducing the uncertainty of a particular condition/situation

[translated from: https://cs.wikipedia.org/wiki/Information]

Digital information is digitized information - information encoded in **digital coding**

Information

The **information is immaterial**, but some kind of material carrier is required to convey information from the source to the target (eg from the suspect to the investigator or from the investigator to his supervisor).

Does not matter, what kind of material carrier is used, information is the same...

Information Encoding

Encoding information is another important term that requires explanation. Indeed, in order to transmit (or otherwise process) the information, the above-mentioned material presentation (interpretation) of the information (its recording on a material carrier) is not sufficient. **There must be an agreement between the source and the target of the information in what "language" they will pass on the information.** In other words, how and by what means they will encode the information transmitted.

Another example of encoding information is **written text**. In this case too, **people had to first agree on the language used, but also on the alphabet (character set) used, the system of writing letters and words**. This is how discrete coding originated.

All of this seems self-evident, automatic, and nobody actually thinks about it. We have experienced it for hundreds and thousands of years. All the more so because the above-mentioned **ways of communicating information (or ways of communicating) have been conceived by humanity to be directly perceived by the human senses**.

Information - Basic Attributes

The information must be to the recipient:
- recievable/processable
- relevant
- understandable

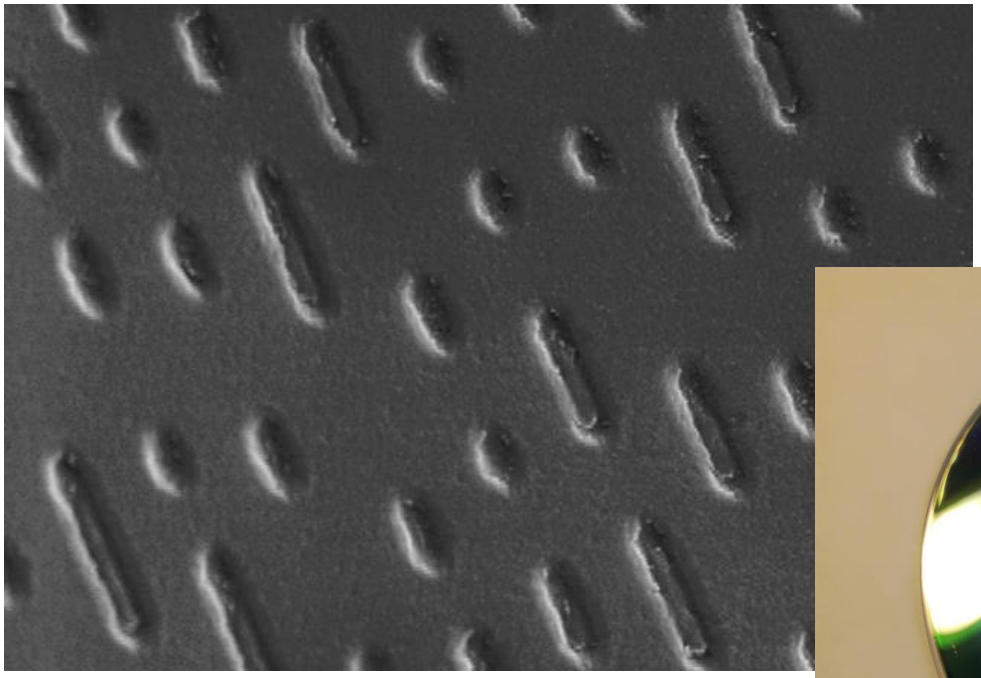
[Other attributes, such as truthfulness, timeliness and others, as appropriate, could also be mentioned]

Information and Carrier

- Text on paper, as well as holes in the punched tape, magnetic record on the hard disk, sound waves or electric current in the network connection cable (and others) are only different material records (material interpretations) of immaterial information
- Paper, punched tape, computer hard disk or connection cable (and many others) are only (potential) carriers of information. The type of carrier itself does not even need to be indicative of the manner of recording and coding of the information.

2nd Break

2nd Break



- The same information coud be recorded on a different carriers
  - hand-written letter could be stored on a paper and in the same time on harddisk as a scan in PDF format -> as the digital record
- The same information could be stored on the same type carrier in different ways of coding
  - information on the paper could be stored as a plain text, picture or eritten row of “0” and “1”

The same info in different coding

Kakost (Geranium) je nevětší rod jednoletých a vytrvalých rostlin z čeledi kakostovitých, který je tvořen asi 430 druhy (v české přírodě se jich objevuje okolo 20). Vyskytují se hlavně v mírném pásmu a to ve všech světadílech, jisté druhy rostou i v horách tropických oblastí. Tak široký rod obsahuje druhy s různými ekologickými nároky, teplomilné druhy vyžadují propustné půdy a slunná stanoviště, stínomilné naopak hlubší a těžší půdy které jsou dostatečně vlhké. Popis: Kakosty jsou většinou byliny a jen řídce keře vysoké 10 až 70 cm, velmi často porostlé různými druhy chlupů. Listy vyrůstají nejčastěji vstřícně, mají palisty, bývají dlanitě dělené až dlanitě laločnaté a ojediněle lichozpeřené.

01001011	01100001	01101011	01101111	01110011	01110100	00100000	00101000
01000111	01100101	01110010	01100001	01101110	01101001	01110101	01101101
00101001	00100000	01101010	01100101	00100000	01101110	01100101	01110110
11101100	01110100	10011010	11101101	00100000	01110010	01101111	01100100
00100000	01101010	01100101	01100100	01101110	01101111	01101100	01100101
01110100	11111101	01100011	01101000	00100000	01100001	00100000	01110110
01111001	01110100	01110010	01110110	01100001	01101100	11111101	01100011
01101000	00100000	01110010	01101111	01110011	01110100	01101100	01101001
01101110	00100000	01111010	00100000	11101000	01100101	01101100	01100101
01100100	01101001	00100000	01101011	01100001	01101011	01101111	01110011
01110100	01101111	01110110	01101001	01110100	11111101	01100011	01101000
00101100	00100000	01101011	01110100	01100101	01110010	11111101	00100000
01101010	01100101	00100000	01110100	01110110	01101111	11111000	01100101
01101110	00100000	01100001	01110011	01101001	00100000	00110100	00110011

K	a	k	o	s	t		(
G	e	r	a	n	i	u	m
)		j	e		n	e	v
•	t	•	•		r	o	d
	j	e	d	n	o	l	e
t	•	c	h		a		v
y	t	r	v	a	l	•	c
h		r	o	s	t	l	i
n		z		•	e	l	e
d	i		k	a	k	o	s
t	o	v	i	t	•	c	h
,		k	t	e	r	•	
j	e		t	v	o	•	e
n		a	s	i		4	3

In most cases, information and its material interpretations have no direct connection.

Digital information

- Digital information is the record of (immaterial) information in digital form on a material medium that is capable of carrying or transmitting such kind of record.

3rd Break

Information must be recievable/processable (seizable/accesible), relevant and understandable.
Is this true for digital information too?

0010011100010000

Latent and Coded

Digital information is information like any other, but it is not directly detectable by common user means and also not directly understandable. **Digital information is therefore latent and coded.** In order to bring the digital information into a usable state, it is necessary to make the digital record accessible and decode. Only after the digital record is made available and decoded its relevance could be assessed.

Relevancy

Relevancy means significance or importance, usually in relation to a goal or purpose.

[https://en.wikipedia.org/wiki/Relevance]

Relevancy - Case Study

1.
In order to investigate crime, it is necessary to examine the computer of the suspect and to determine whether there are photographs in the computer that would show certain facts relevant to the investigation. If such photographs are found, it is necessary to find out additional information on how the photographs were taken.

Relevancy - Case Study

2.
Data from the computer was seizued and decoded. The expert assessed the relevance of the resulting data in relation to the requirement to identify formats of image data (digital photographs). A contiguous series of photo files named IMG01001.JPG through IMG01020 was found, totaling 20 files. EXIF information was obtained from the photos, confirming that the series of photos was taken in a continuous time series at one location and one camera. These results were transmitted to the investigator.

Relevancy - Case Study

3.
The investigator subsequently assessed the relevance of the photographs to the crime under investigation and selected three photographs that were relevant to the case with the requirement to examine in greater detail the credibility of all the information relating to the photographs.

Relevancy - Case Study

4.

The expert therefore analyzed the image data using a special software that calculates the specific characteristics of the image sensor chip for group identification of the photographic device that took the digital photographs. The expert then determined the relevance of the data obtained in relation to the data obtained from all EXIF information of all images and concluded that the investigator's three photographs were very likely to be falsified because the calculated characteristics of the image sensor chips for the three photographs were significantly different . It was found that the three photos in question were taken by another camera under different lighting conditions and that the EXIF information of the three photos was altered.

4th Break

- How is it possible that the expert did not immediately examine the possibility of falsifying photographs?

Locard's Exchange Principle

- Locard's Principle holds that the perpetrator of a crime will bring something into the crime scene and leave with something from it, and that both can be used as forensic evidence
- Dr. Edmond Locard (13 December 1877 – 4 May 1966) was a pioneer in forensic science who became known as the Sherlock Holmes of France. He formulated the basic principle of forensic science as: "**Every contact leaves a trace**"
- Physical evidence cannot be wrong, it cannot perjure itself, it cannot be wholly absent. **Only human failure to find it, study and understand it, can diminish its value.**

Locard's Principle in Digital World

- Every activity in digital systems leaves digital traces
  - Only sometimes we are not able to identify, seize or interpret it
- Digital trace is a set of digital data with understandable information
  - Digital data without information is not a digital trace
- Criminalistics Digital Trace is Digital Trace with information, which could be connected (directly or indirectly) with criminal case
  - Independence evidence assessment

Summary

- Information:
  - is immaterial but need a material carrier
  - is latent and coded
  - does not depend on its carrier
  - must be seizeable, understandable and relevant

- Digital information is information recorded in digital form
- Locard's Principle in Digital World

5th Break

Digital Trace and its Properties

- Material substance of digital trace
- Latency
- Time identification
- Information value
- Lifetime
- Quality of Archives
- Big volumes of data
- Big density of information
- Dynamics of development of ICT
- Speed
- Complexity
- Territorial dimension
- Data protection
- Automatisation
- Hiding of identity
- Restoration of data
- Low confidence in digital evidence weight
- …

Material Carrier

Time identification

- Digital record in standard information system is identified with time stamp (directly or indirectly).
- Theoreticly we have to be able to identify the time of each event in information technology.
  - often we do not know how
  - problem of time zones
  - problem of time formats

Information value

- An enormous amount of information of a diverse nature is stored in information systems. Digital data has a very high information value and their potential is far from being used efficiently and comprehensively.

Lifetime

- Very long (years)
- Very short (miliseconds)

Quality of Archives

- Backup policy
  - relevant technology
  - audited procedures
  - storage media

- 3rd parties responsible for storing - cloud
  - not physical access to the data

Data volumes

- Common automatic forensic analysis systems have limits of data volumes (1-5 TB)
- For “Big Data” analysis special computer clusters are necessary - high price
- Special data clusters are developed (HADOOP)

Big density of information

- In Big Data we search for a needle in a haystack. Sometimes we search 1 bit in lot of TerraBytes of data

Dynamic Development

- Lot of new systems, vesrions, applications and new technologies need specialisation
- Coplexity grows
- Variability grows
- Capacity grows
- Speed grows
- …

Speed

- You can “delete” all disk in 1/10 of a second
- Computer power gives us an opportunity to provide complex data analysis in a reliable time

Coplexity and Heterogenity

- IT environment becomes very complex and heterogenous
- Simplest user interface cost a huge complexity in the backgroud of the programme code

Territorial dimension

- Data is moving in a second over all world
  - we do not know what data is moving out from our device, ve only belive to providers
  - we do not know where our data is stored
    - cloud
    - voice data from our personal assistants or data from our wearables
  - remember Locard's Principle ;)

Data Protection

- Encryption - the biggest problem of the forensic examination
- Data hiding - steganography
- Access Control
- Password Protection

Automatisation

- High probability of effective usage of automatic data analysis
- Real exeptions
  - images (machine learning?)
  - automatic programme code analysis?
- Problem of the reliability and trustworthiness of the results - and how use AI in court?

Hiding Identity

- Identification:
  - individual (personal) identification
  - group identification
- Using biometric systems - trustworthiness
- Perhaps only the use of DNA identification will help to use individual identification in computer technology.

Restoration of data

- Once stored data are stored (nearly) forever and only the way of deleting is to overwrite them.
- There is big chance to restore even deleted data

Problematic(?) weight of the digital evidence

- What is the weight of the digital evidence?
- Real assessmet of the weight of digital evidence could be done only by digital (forensic) expert, but he is not alowed to do it, because only judge is alowed to do it.
- Problem of the material true and expert opinion