Data and Network Security

(Master Degree in Computer Science and Cybersecurity)

Lecture 10

SAPIENZA
UNIVERSITÀ DI ROMA

Outline for today

- Literature Analysis – update
- Recap last lecture
- GDPR
- GDPR – right to access data
- How to prepare the Presentation

Examination

The final grade is calculated as follows:

- **45% Literature Analysis and active participation**

- **45% Written research report**

- **10% Active participation to Question & Answer section.**

***If for any reason students do not turn in most of the required above tests, then those students will be required to take an oral exam on the entire course programme (TDB)**

The weight of this final oral exam = 100%.

Presentation - Update

- Prepare a 25-30 minute presentation (divide your work equally and efficiently)

- First 5-10 minutes dedicate to background then follow with the research papers

- In the last minutes hint at the possible future directions.

- 10 minutes of questions (from your colleagues, this counts on their 10% of active participation also)

Outline for today

- Literature Analysis
- **Recap last lecture**
- GDPR
- GDPR – right to access data
- How to prepare the Presentation

Digital Forensics

Definition: branch of forensic science that deals with the identification, preservation, examination, analysis, and presentation of digital evidence derived from electronic devices and digital media.

Goal: To explain current state of a digital artifact

Digital Forensics

Definition: branch of forensic science that deals with the identification, preservation, examination, analysis, and presentation of digital evidence derived from electronic devices and digital media.

Goal (here): To find out the data type of an artifact

Entropy estimation is not enough!

Most data types (e.g., text, images, audio)
are information-rich and highly structured

=> low entropy!

Entropy estimation is not enough!

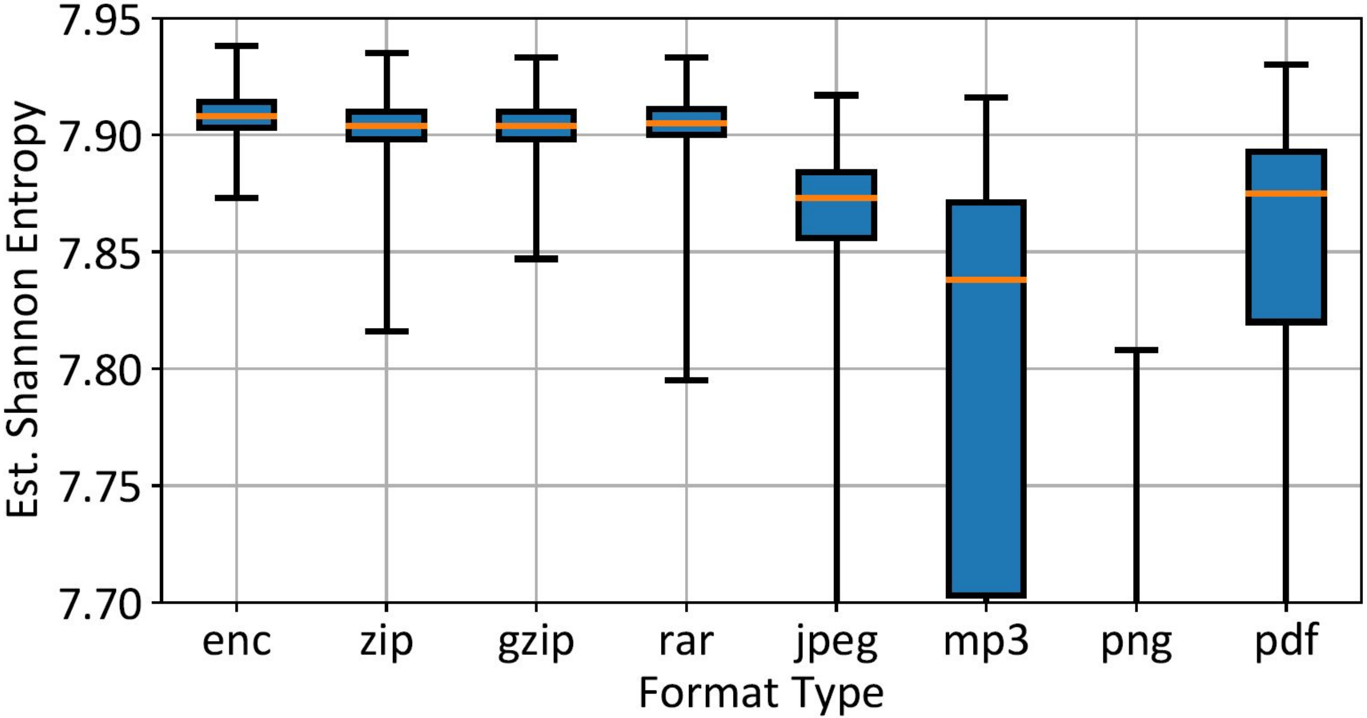
~~Most data types (e.g., text, images, audio) are information-rich and highly structured~~

~~=> low entropy!~~

Modern CPUs can efficiently decompress data for processing, and compress it back for storage or transmission!

=> Most data formats use compression!

Entropy estimation is not enough!



NIST SP800-22 & χ^2 Test

The NIST SP800-22 describes a suite of tests to evaluate the quality of random number generators:

- ➢ 15 distinct tests, which analyze various structural aspects of a byte sequence.
- ➢ Commonly employed as a benchmark for distinguishing compressed and encrypted content.

The X^2 Test is a simple statistical test to measure goodness of fit.

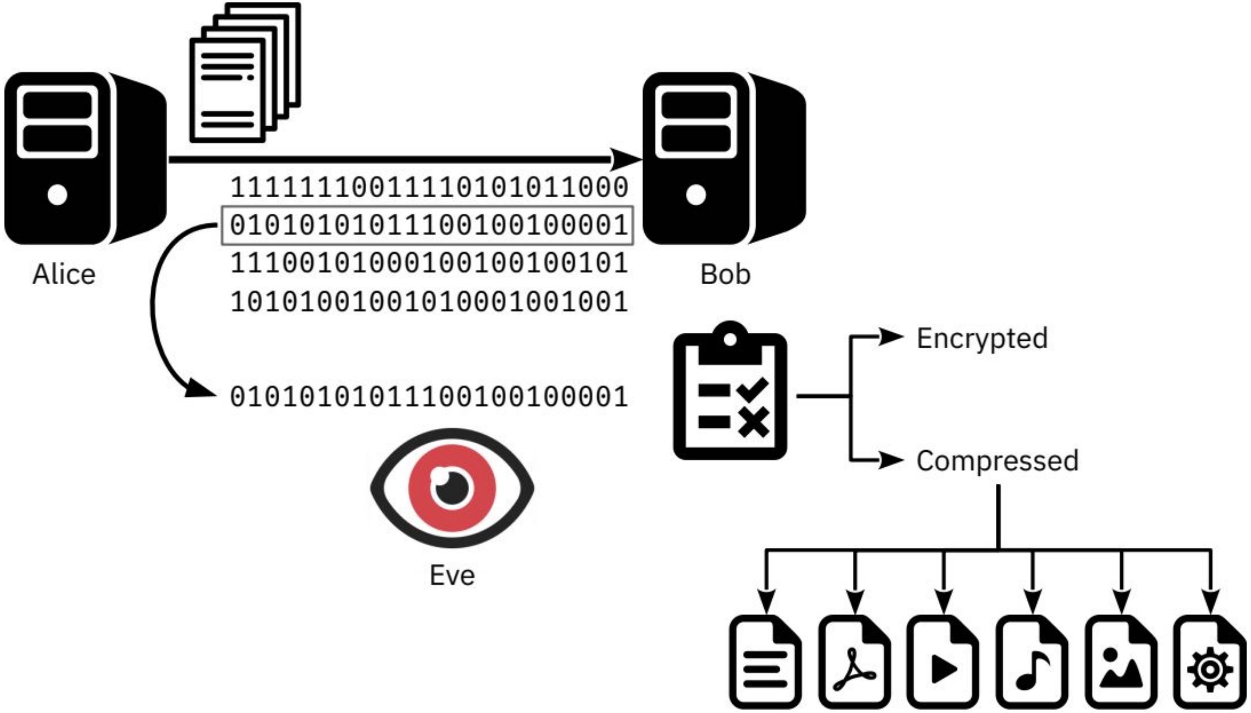
- ➢ It has been widely applied to distinguish compressed and encrypted content.

Combine what we know?

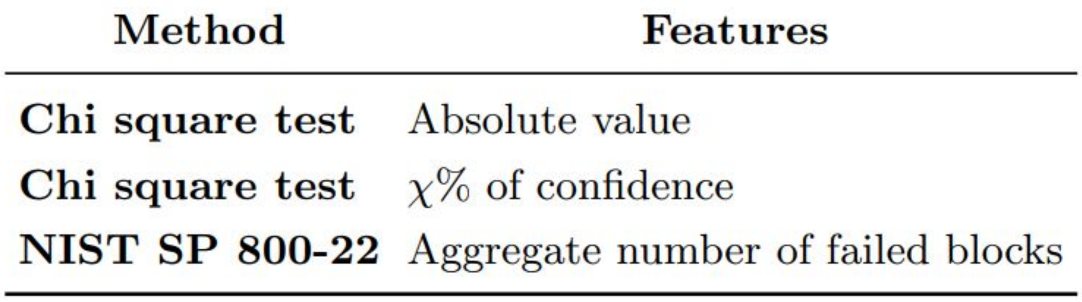
Simultaneously incorporate *three methods* to distinguish between compressed and encrypted fragments:

- X^2 test with a threshold.
- X^2 confidence interval.
- Subset of NIST 800-22:
  - ➢ frequency within block test
  - ➢ cumulative sums test
  - ➢ approximate entropy test

Feature selection



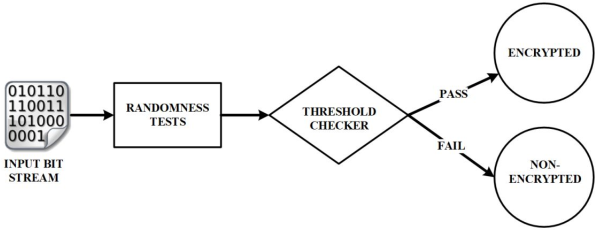
Feature selection

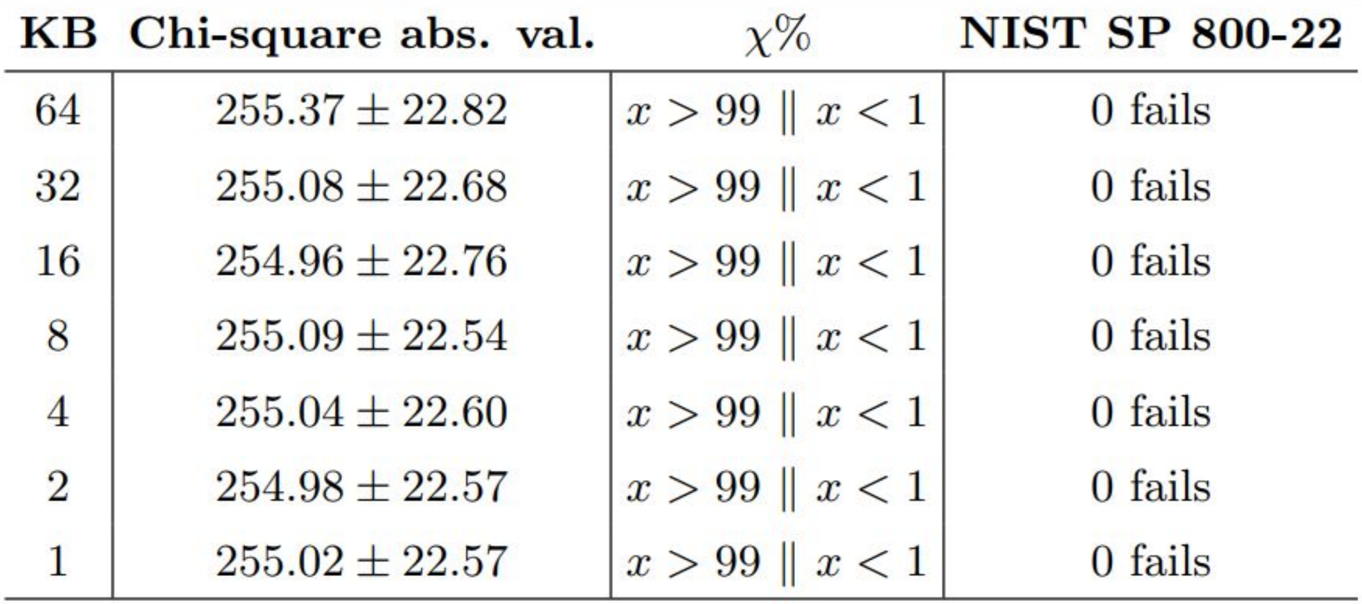


Method	Features
Chi square test	Absolute value
Chi square test	χ% of confidence
NIST SP 800-22	Aggregate number of failed blocks

- A test of goodness of fit establishes whether an observed frequency distribution differs from a theoretical distribution.
- A test of homogeneity compares the distribution of counts for two or more groups using the same categorical variable
- A test of independence assesses whether observations consisting of measures on two variables are independent of each other.

The interval values on different size





KB	Chi-square abs. val.	$\chi\%$	NIST SP 800-22
64	255.37 ± 22.82	$x > 99 \parallel x < 1$	0 fails
32	255.08 ± 22.68	$x > 99 \parallel x < 1$	0 fails
16	254.96 ± 22.76	$x > 99 \parallel x < 1$	0 fails
8	255.09 ± 22.54	$x > 99 \parallel x < 1$	0 fails
4	255.04 ± 22.60	$x > 99 \parallel x < 1$	0 fails
2	254.98 ± 22.57	$x > 99 \parallel x < 1$	0 fails
1	255.02 ± 22.57	$x > 99 \parallel x < 1$	0 fails

Encryption/Compression Distinguisher

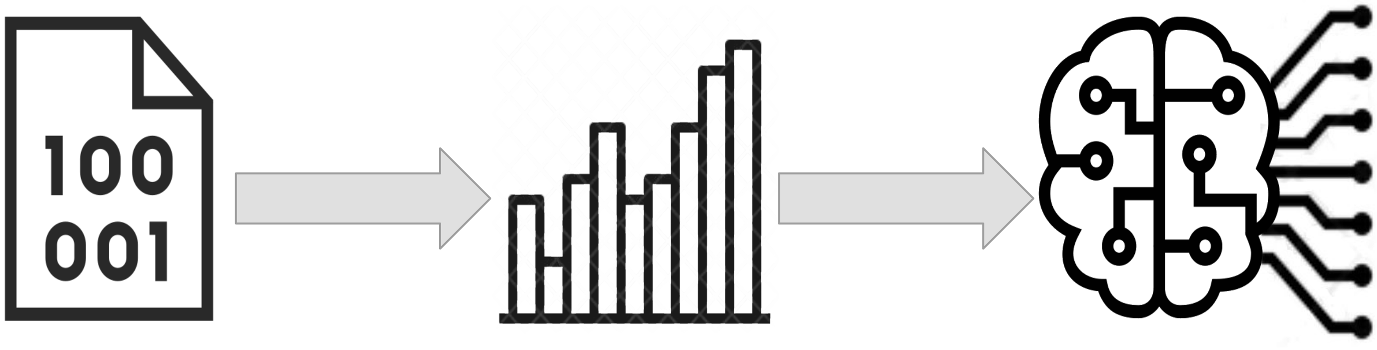
Tests based on byte-value distribution can distinguish some encrypted and compressed content, ***but have accuracy issues.***

WHY?

These tests lose information about the shape of the data *distribution*!

Deep Neural Networks can consider the entire discrete distribution and can learn to recognize complex distributions!

How it works?



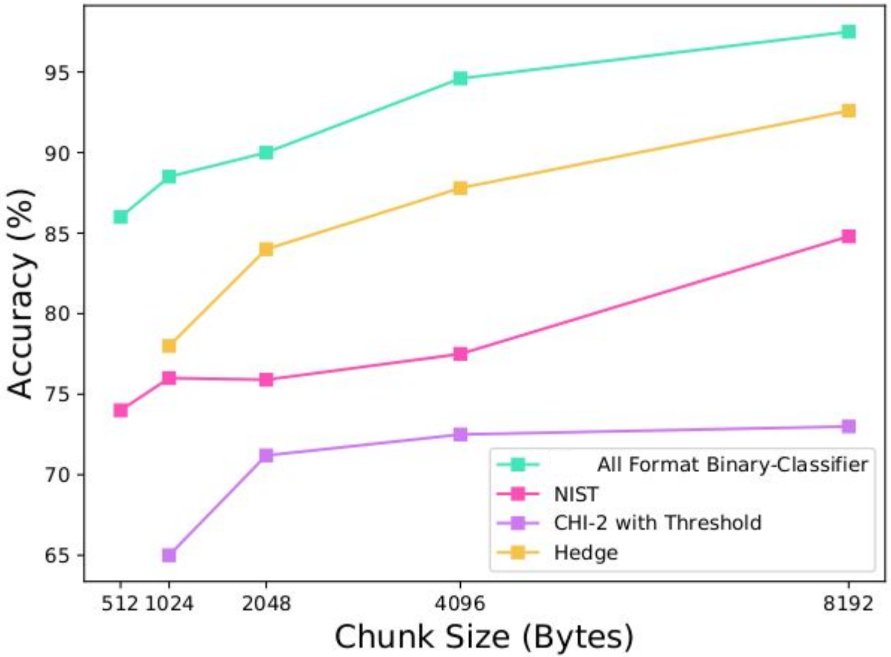
Data fragment

Histogram

Classification

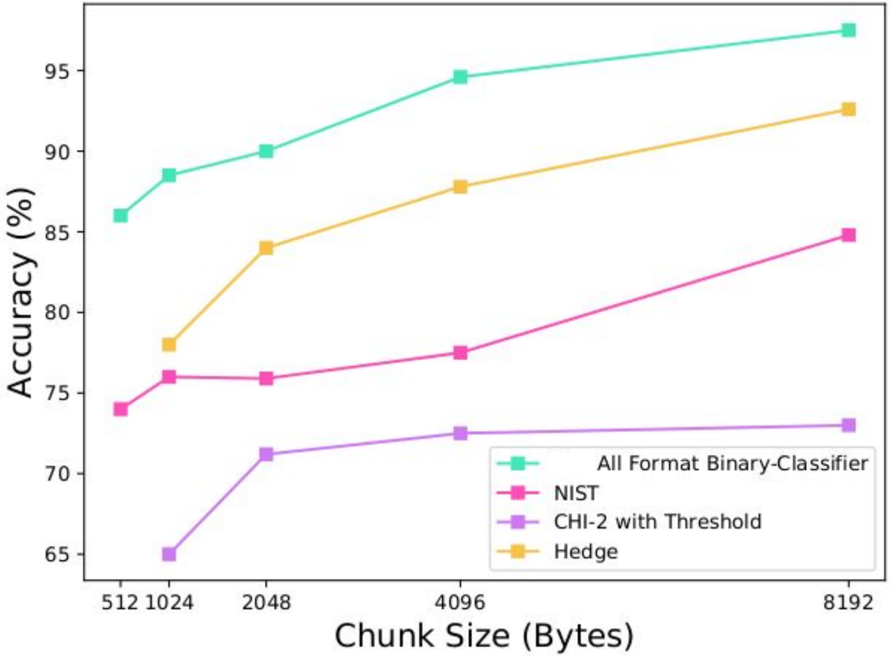
Evaluation

Q: *Is a given data fragment compressed or encrypted?*

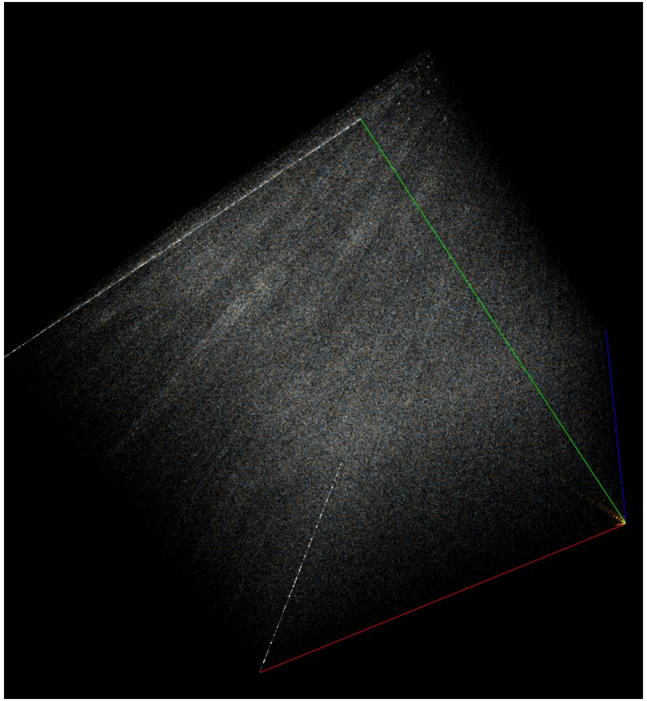


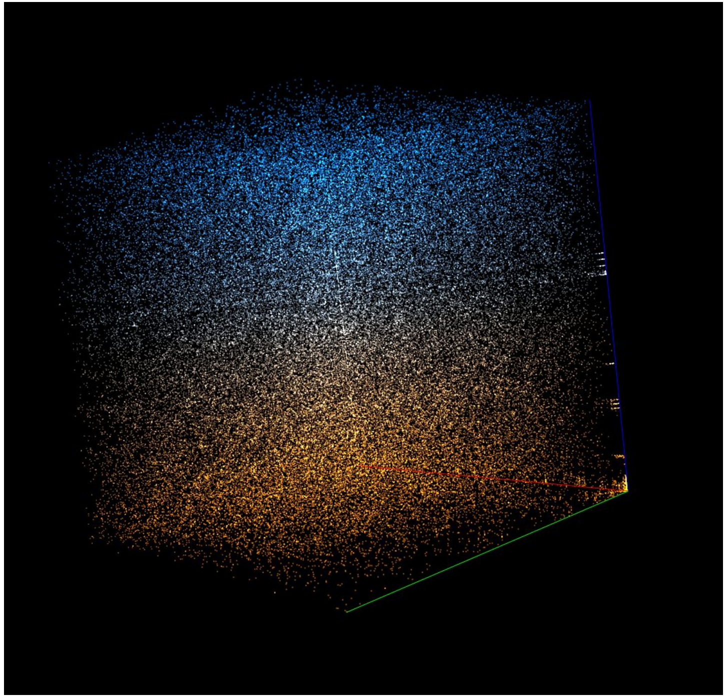
Other possible approaches

Q: *Can we get more information looking at something else than just the histogram?*



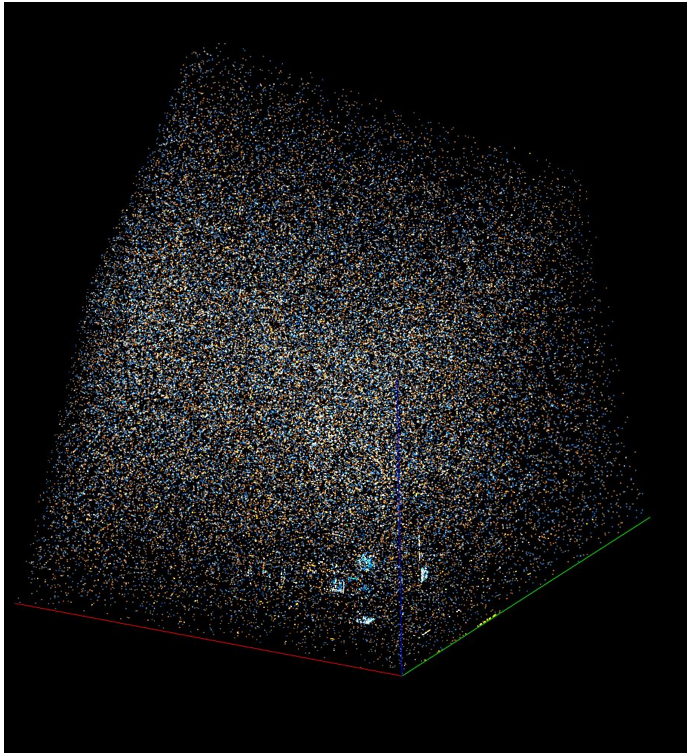
Q: *Can we get more information looking at something else than just the histogram?*

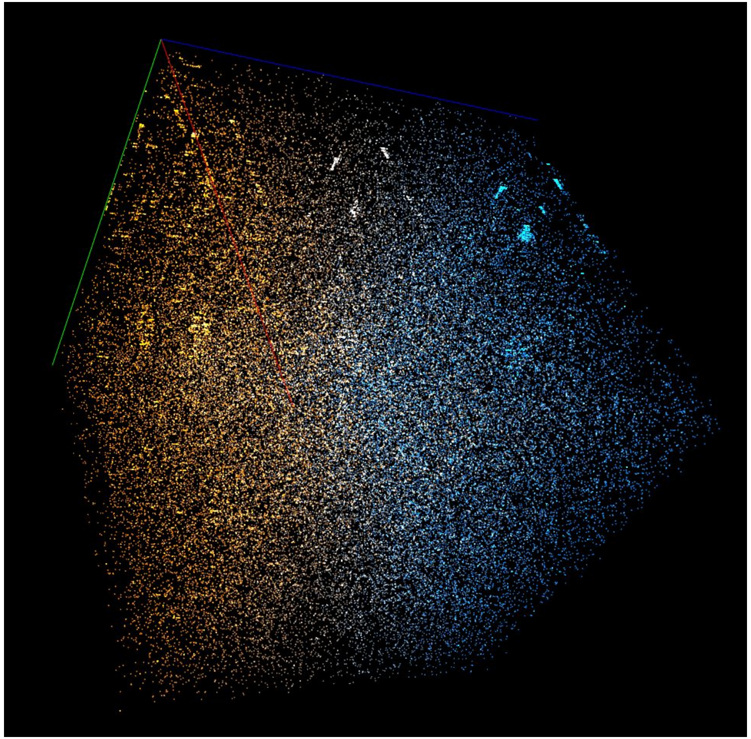




JPG file

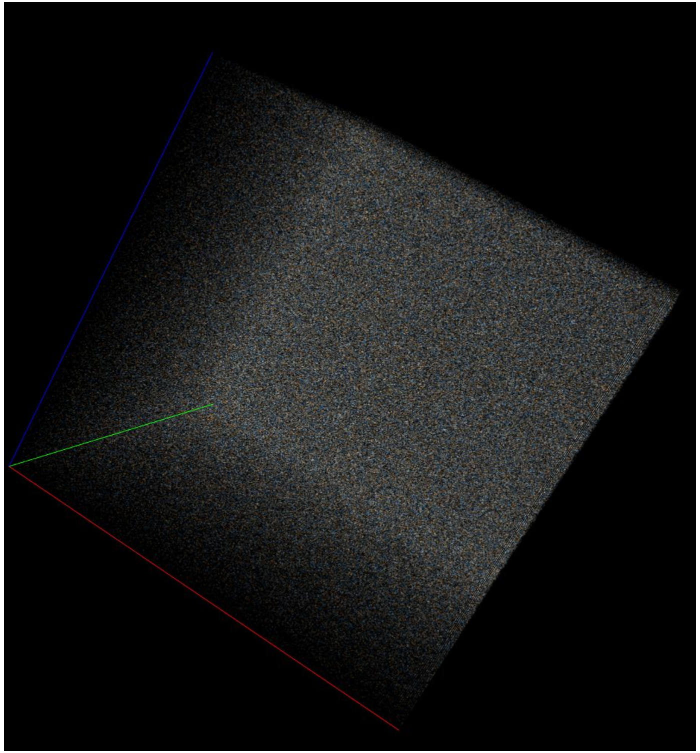
Q: *Can we get more information looking at something else than just the histogram?*

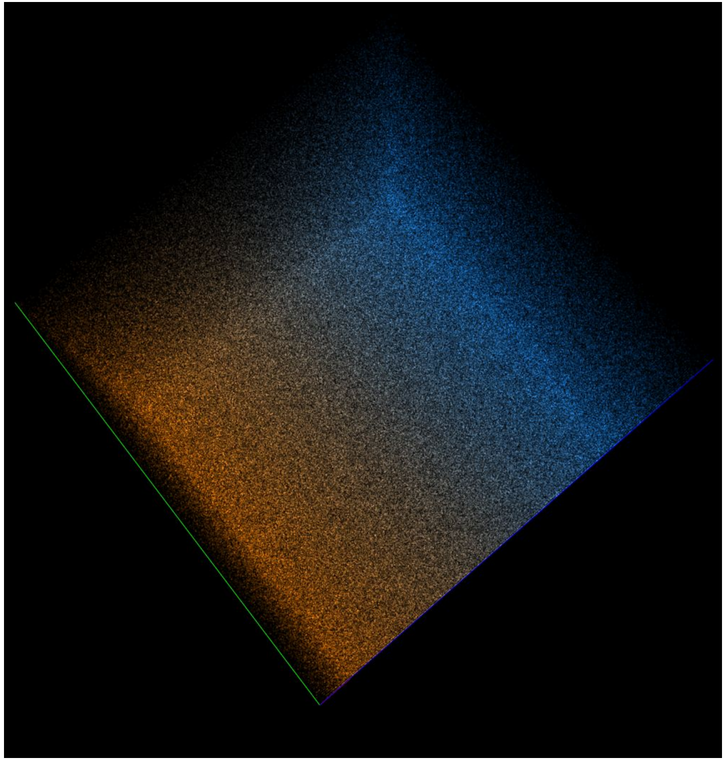




Pdf file

Q: *Can we get more information looking at something else than just the histogram?*





Encrypted file

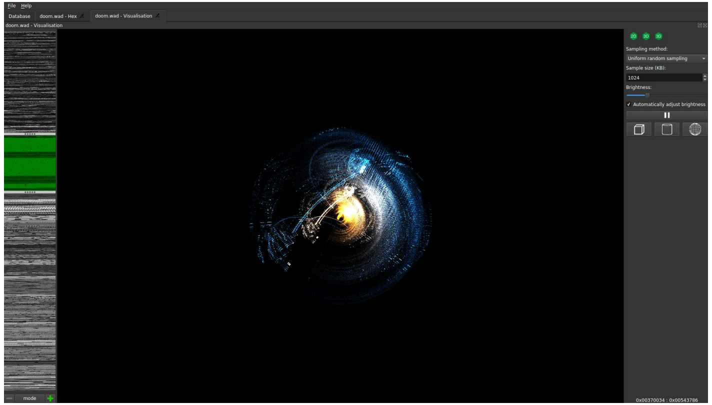
VELES

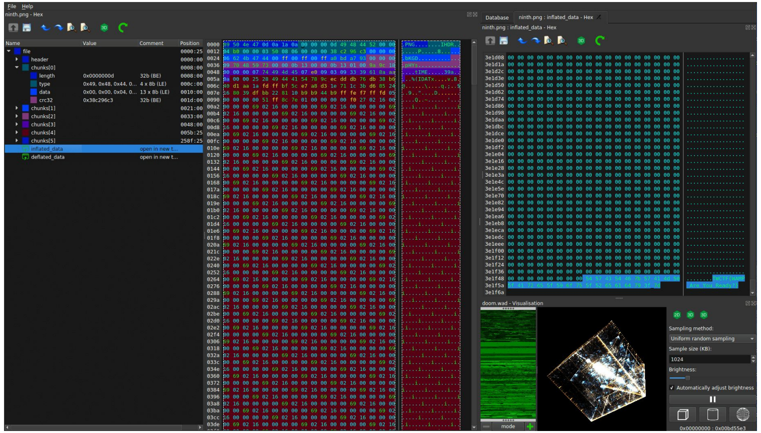
Humans sometimes have it difficult to notice subtle patterns in large amounts of binary data. Visualisation can help.
Veles combines advanced hex explorer and data visualization features with an extensible framework for binary data analysis.

- Reverse engineering binaries?
- Exploring file system images?
- Steganography?

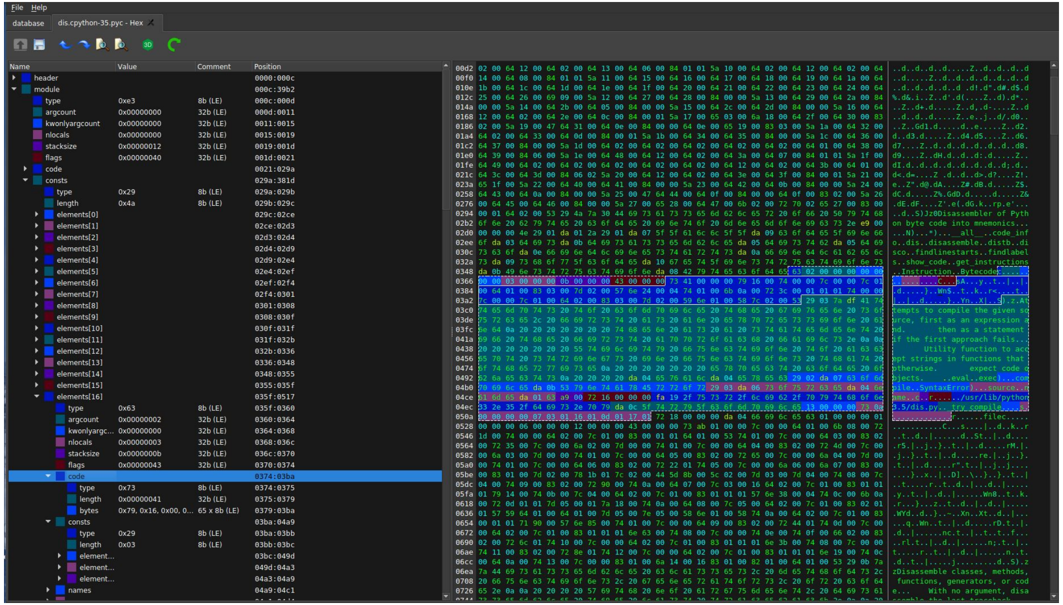
Can be a good auxiliary tool in CTF

VELES - https://codisec.com/veles/





```
Name              Value                    Comment       Position
file                                                     0000:25
  header                                                 0000:00
  chunks[0]                                              0008:00
    length        0x0000000d               32b (BE)      0008:00
    type          0x49, 0x48, 0x44, 0...   4 x 8b (LE)   000c:00
    data          0x00, 0x00, 0x04, 0...   13 x 8b (LE)  0010:00
    crc32         0x38c296c3               32b (BE)      001d:00
  chunks[1]                                              0021:00
  chunks[2]                                              0033:00
  chunks[3]                                              0048:00
  chunks[4]                                              005b:25
  chunks[5]                                              258f:25
  inflated_data                            open in new t...
  deflated_data                            open in new t...
```



```
Name                 Value                   Comment       Position
header                                                     0000:000c
module                                                     000c:39b2
  type               0xe3                    8b (LE)       000c:000d
  argcount           0x00000000              32b (LE)      000d:0011
  kwonlyargcount     0x00000000              32b (LE)      0011:0015
  nlocals            0x00000000              32b (LE)      0015:0019
  stacksize          0x00000012              32b (LE)      0019:001d
  flags              0x00000040              32b (LE)      001d:0021
  code                                                     0021:029a
  consts                                                   029a:381d
    type             0x29                    8b (LE)       029a:029b
    length           0x4a                    8b (LE)       029b:029c
    elements[0]                                            029c:02ce
    elements[1]                                            02ce:02d3
    elements[2]                                            02d3:02d4
    elements[3]                                            02d4:02d9
    elements[4]                                            02d9:02e4
    elements[5]                                            02e4:02ef
    elements[6]                                            02ef:02f4
    elements[7]                                            02f4:0301
    elements[8]                                            0301:0308
    elements[9]                                            0308:030f
    elements[10]                                           030f:031f
    elements[11]                                           031f:032b
    elements[12]                                           032b:0336
    elements[13]                                           0336:0348
    elements[14]                                           0348:0355
    elements[15]                                           0355:035f
    elements[16]                                           035f:0517
      type           0x63                    8b (LE)       035f:0360
      argcount       0x00000002              32b (LE)      0360:0364
      kwonlyargc...  0x00000000              32b (LE)      0364:0368
      nlocals        0x00000003              32b (LE)      0368:036c
      stacksize      0x0000000b              32b (LE)      036c:0370
      flags          0x00000043              32b (LE)      0370:0374
      code                                                 0374:03ba
        type         0x73                    8b (LE)       0374:0375
        length       0x00000041              32b (LE)      0375:0379
        bytes        0x79, 0x16, 0x00, 0...  65 x 8b (LE)  0379:03ba
      consts                                               03ba:04a9
        type         0x29                    8b (LE)       03ba:03bb
        length       0x03                    8b (LE)       03bb:03bc
        element...                                         03bc:049d
        element...                                         049d:04a3
        element...                                         04a3:04a9
      names                                                04a9:04c1
```

Outline for today

- Literature Analysis
- Recap last lecture
- GDPR
- GDPR - right to access data
- How to prepare the Presentation

GDPR

GDPR is a European law which went into effect on May 25, 2018

GDPR - What?

GDPR is a European law which went into effect on May 25, 2018. The GDPR lays down rules relating to the protection of fundamental rights and freedoms of persons, and in particular their right to the protection of personal data.

It aims to improve consumer protection and general levels of privacy for individuals, includes mandatory reporting of data protection breaches and has an increased emphasis on gaining explicit consent to process information.

GDPR - What?

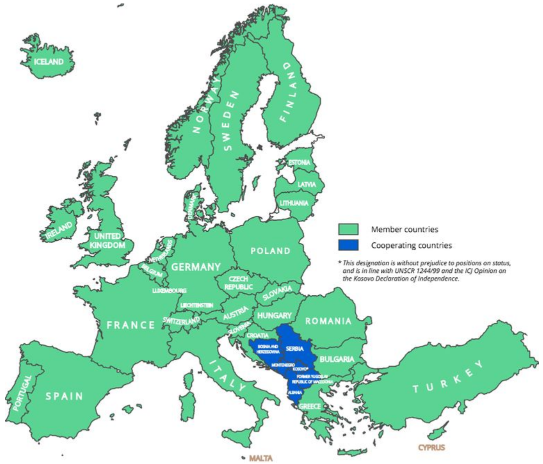
GDPR is a European law which went into effect on May 25, 2018. The GDPR lays down rules relating to the protection of fundamental rights and freedoms of persons, and in particular their right to the protection of personal data.

- Governs the type of notice that must be provided to people regarding how their identifiable data is used.
- Governs how companies are allowed to use and process identifiable data.
- Has strict requirements for using sensitive data.

GDPR - To whom it applies?

- Those who offer goods or services to persons in the EU/EEA.
- Those who control and process data about persons in the EU/EEA

  - **Personal Data** is any information that can identify a person
  - Sensitive Data:
    - race/ethnicity,
    - political opinions,
    - religious/philosophical beliefs,
    - union membership,
    - genetic data,
    - biometric data,
    - health data,
    - data regarding person's sex/orientation



* This designation is without prejudice to positions on status, and is in line with UNSCR 1244/99 and the ICJ Opinion on the Kosovo Declaration of Independence.

GDPR - Controllers/Processors

- Controllers specify the means and purpose of the data processing. The data controller gives instructions for processing to the data processor. The controller is responsible for implementing measures to ensure that processing occurs pursuant to GDPR.

GDPR - Controllers/Processors

- Controllers specify the means and purpose of the data processing. The data controller gives instructions for processing to the data processor. The controller is responsible for implementing measures to ensure that processing occurs pursuant to GDPR.

- Processors conduct the processing under the direction of the controller. The processor cannot process personal data except upon the instructions of the controller. The processor is tasked by the text of the privacy law with helping the controller with certain tasks, including information necessary to demonstrate compliance.

GDPR - What is data processing?

- Adapting
- Altering
- Collecting
- Combining
- Consulting
- Destroying
- Disclosing
- Erasing
- Organizing
- Recording
- Retrieving
- Storing
- Structuring

GDPR - What is needed to process data?

Lawful basis is needed:

- When required for a contract
- When required for public interest
- When required to comply with a law
- When required to protect an individual’s life
- When required for the legitimate interests of a third party (no sensitive data)
- When freely given consent for a specific purpose has been provided

If sensitive data is being processed, explicit consent for those data elements is required.

GDPR - What consent is needed?

- Name and/or title of the data processor

- The purpose and basis for processing of the subject’s data

- The type of data to be processed
  - When sensitive data are going to be processed, these data elements must be explicitly listed in the consent.

- If data will be transferred to a less secure country

GDPR - Legally Effective consent?

- Clear and intelligible
- Very specific about the purpose of processing
- Unambiguous indication

GDPR - After obtaining consent?

- Processors and Controllers must ensure privacy:
  - Limit access to the data
  - Code or encrypt the data where possible
  - Limit processing to only the necessary data
  - Retain the data for the least amount of time possible
  - Incorporate data protection into the processing activities

GDPR - Subject’s rights

- Rectification of the personal data
- Get notified when their personal data is used
- Able to modify and erase
- Able to restrict how their data are processed
- Able to reject automated individual decision-making
- Able to access to their personal data collected about them
- Able to receive their data and transfer it to a third party

Outline for today

- Literature Analysis
- Recap last lecture
- GDPR
- GDPR – right to access data
- How to prepare the Presentation

GDPR - Right to access data - Article 15

- Article 15 of the law establishes one of the most fundamental rights in the Internet:
  - The users have the right to request to web sites a copy of all the personal data that they have about them. The goal of this right is to return control and awareness to the users of the personal data they share, consciously or not.

What will we see today?

- The results of a broad world-scale investigation on the actual deployment of the GDPR.
  - Step by step analysis of all the phases to a subject access request
  - Over 300 of the most popular websites according to Alexa ranking

GDPR - Selecting the providers to evaluate

- Focus on websites that store personal information to identified users:
  - Adult (7.5%), Art (5.1%), Business (7.8%),
  - Computer (9.6%), Games (8.4%), Health (1.8%),
  - Kids & Teen (2.8%), News (7.2%), Recreation (12.9%),
  - Reference (5.7%), Science (3.9%), Shopping (15.6%), Society (6%), and Sports (5.4%)

GDPR - Right to access data

- First step: create and utilize real accounts on these services
- Second step: ask for your data and evaluate

GDPR - Asking the data - focus

- While asking the data information about the relevant phases needs to be collected such as:
  - Privacy policy compliance
  - Request methodology
  - Identification
  - Response format
  - Response time
  - Information obtained

Privacy policy compliance

The GDPR states that the data collectors must inform the users about the rights of requesting a copy of, updating, or deleting their personal data owned by the data controllers.

Data collectors must provide the contact details of the data protection officer.

Users are informed about this through the privacy policy page

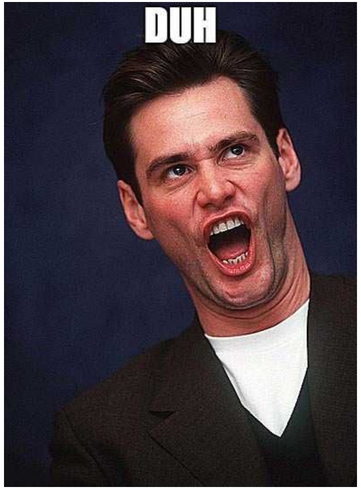
Privacy policy compliance - Results

- 6 out of 341 of them mention neither the user’s right of data access in their privacy policies nor indicate a contact point.
- 53 (15.54%) web sites do not mention the users’ rights,
- 3 (0.87%) web sites do not have any contact point for privacy information.
- Privacy policy pages of 3 web sites (0.87%) do not work.

Request methodology

Data collectors shall facilitate the exercise of subject data rights (GDPR Article 12 sec. 2)

- Asking by postal mail or via a call is not a preferable option for obvious reasons



Request methodology - results

Usually the procedure to retrieve the data is described in the privacy policy page. If not, an email to the responsible DPO was sent.

- 219 out of 332 (65.9%) data controllers
  - it is enough to send an email
- 11 (3.3%) data controllers,
  - the user also need to compile a form and send it by email.
- 96 (28.9%) data controllers
  - it is possible to request the personal data through a form on the web site
- 6 (1.8%) data controllers,
  - the only way to perform the request is by standard mail or by an phone call outside Europe

Identification

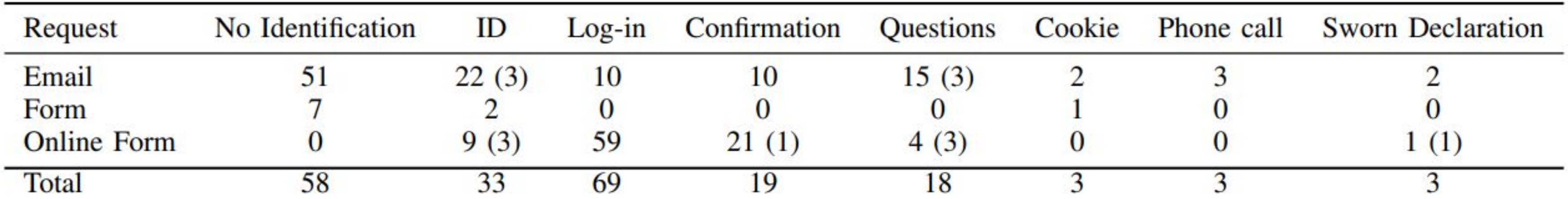
What we are asking is personal data.

Controllers should make sure that they are giving the data to the right person.

Identification - Results

What we are asking is personal data.

Controllers should make sure that they are giving the data to the right person.



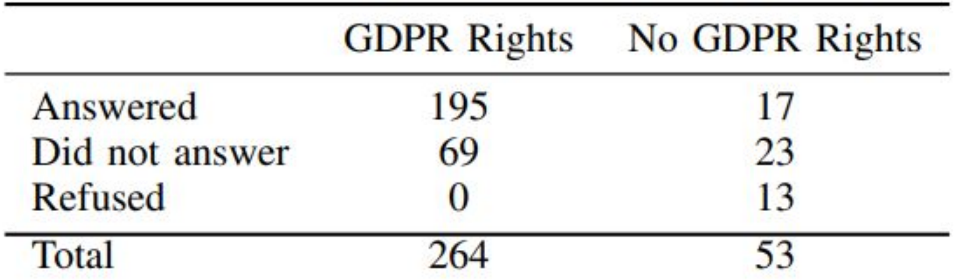
Request	No Identification	ID	Log-in	Confirmation	Questions	Cookie	Phone call	Sworn Declaration
Email	51	22 (3)	10	10	15 (3)	2	3	2
Form	7	2	0	0	0	1	0	0
Online Form	0	9 (3)	59	21 (1)	4 (3)	0	0	1 (1)
Total	58	33	69	19	18	3	3	3

Response obtained

The information retrieved by the data controllers.

Response obtained

The information retrieved by the data controllers.



	GDPR Rights	No GDPR Rights
Answered	195	17
Did not answer	69	23
Refused	0	13
Total	264	53

212 out of 326 of data controllers, were able to provide the personal data

Response time

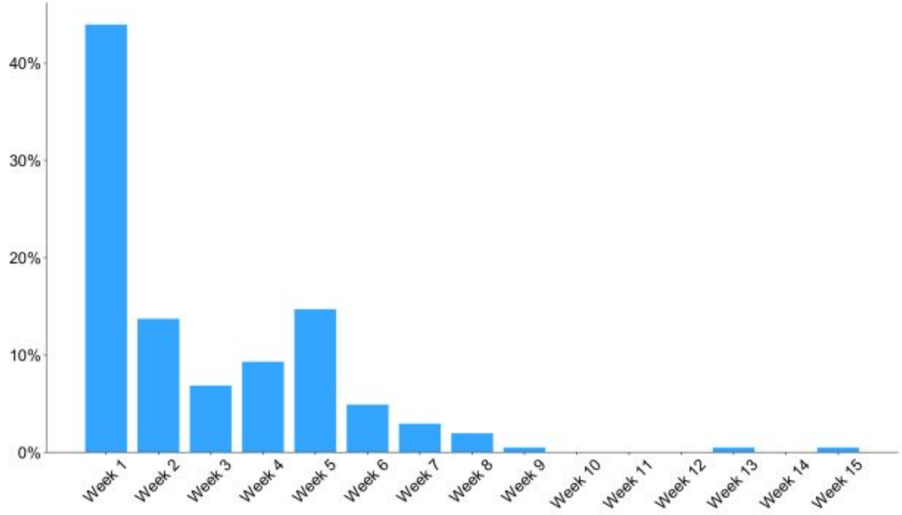
Time needed by the web sites to send back the personal data.

- The data controllers have 30 days to process the request and provide the data.
- If necessary, the controller can extend that period for additional two months but needs to notify before the end of the first month

Response time - results

Time needed by the web sites to send back the personal data.

- The data controllers have 30 days to process the request and provide the data.



Response format

- The GDPR defines two rights that allow the users to access their personal data.
  - No strict constraints about the response data format (Article 15).
  - But the data should be provided in a commonly used and machine-readable format (Article 20).

Response format - results

Time needed by the web sites to send back the personal data.

- The data controllers have 30 days to process the request and provide the data.
- If necessary, the controller can extend that period for additional two months but needs to notify before the end of the first month

- (52.7%) answered with a structured data format JSON, CSV, XLS or XML
- The rest - raw text or in a tabular form, directly typed in the email body or as a PDF attachment.

Privacy Concerns

More than 50% of data collectors that handled the request suffer from flaws that can compromise the users’ privacy

Sharing data via email and no email encryption

- Sending sensitive data as plain email or plain attachment can be risky:
    - The email can be sent to an incorrect recipient.
    - The email and the attached file are saved on the email server (an attacker gains unauthorized access to it)

Over ⅓ of the responses

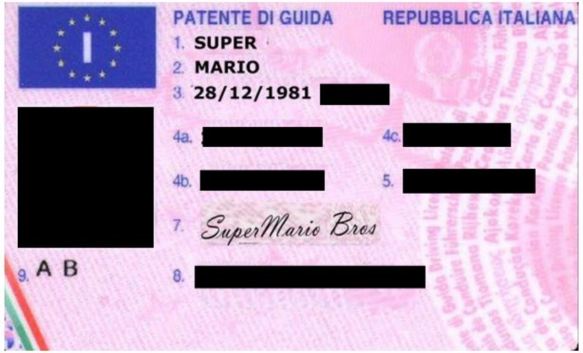
Identity card

- Prior to sending the data via email, some require the scanned ID to be sent to them.
    - No information about how should be done this scan
    - how this scan is processed or deleted after is unclear

Some collectors might not even have the data to verify that the ID scan is legitimate or not

Identity card - Tampering with the card

Some collectors might not even have the data to verify that the ID scan is legitimate or not



PATENTE DI GUIDA REPUBBLICA ITALIANA

[redacted]

1. SUPER
2. MARIO
3. 28/12/1981 [redacted]

4a. [redacted] 4c. [redacted]

4b. [redacted] 5. [redacted]

7. SuperMario Bros

8. [redacted]

9. A B

Outline for today

- Literature Analysis – update
- Recap last lecture
- GDPR
- GDPR – right to access data
- **How to prepare the Presentation**

Literature analysis - howto

- What is a literature review/analysis:
    - It is a critical summary of the works.

During the literature analysis (and hence in the presentation):

- Analyze the specific topic/issue
- Present a general background for the topic
- Present the selected works and their contribution
- Point to future directions or gaps/flaws in those works.

Literature analysis - the audience point of view

- The presentation needs to:
    - Provide a general idea of the topic (understanding, current trends etc)
    - Whether they want to read and go into more details in that topic or read those papers

Literature analysis - Papers presentation approach

- Chronological
- Thematic

Literature analysis - Papers presentation approach

- Chronological
    - Describe each work in succession

Literature analysis - Papers presentation approach

- Thematic
  - Organize the talk based on the theme or theoretical concepts that you judge to be important to understand the topic

What needs to be in the presentation?

- No slide limit - you need to finish in max 30 minutes so organize as you want.
- Provide a general background on the topic for a start and then go to the individual contributions of the papers regarding this topic
- The presentation should not be just a simple summary of the works, but more of a critical analysis, so your opinion and also comparison between works are highly suggested.

Reading Material

1. GDPR and the study related to the compliance of the major websites to GDPR, Link-1, Link-2