Operating Honeypots and Ensuring GDPR Alignment

A Practical Approach

CIRCL
Computer Incident
Response Center
Luxembourg

Team CIRCL *TLP:WHITE*

7th May 2018

Honeypots

Definition (Honeypot[1])

- Honeypots are resources **meant to be probed or attacked**
- **Network traffic is often recorded** to/from these (otherwise) unused resources

Collected data

- Misconfigured systems
  - Badly configured printers
  - Errors in DNS resolver configuration leaking DNS requests
- Backscatter traffic as service response to spoofed IPs of the honeypot
- Attack traffic: (automated) exploit code, brute force attacks, amplification attacks

[1]

Motivation for honeypot operation

- Detect **new attack trends**
- Collect information about **attack techniques**
- Collect **malicious software** for further analysis
- Measure (distributed) **denial of service attacks**
- Estimate size of **botnets**
- Discover misconfigured machines → **inform security point of contacts**

Objectives of this talk

Identify favourable argumentation within GDPR to ensure these operations

Identification of personal data and data subject

- At time of collection, it is not known what **kind of data** is collected
- Depends on the **service attacked** and the **capabilities** of the honeypot
- There are doubts how data should be **interpreted**, i.e. endianness, parsing of uninitialized memory
- **Spoofed** network packets
- **NAT** (Network address translation) → many machines behind an IP address
- Automated attacks → reverse Turing test
    - Scanning for new targets
    - Brute force attacks
    - Automated exploits, i.e. DDoS botnets

Operating honeypots in your own networks for your own interests

Article 13: Regulation 2016/679

```
    Information to be provided where personal data are collected from the data
     subject

    (d) where the processing is based on point (f) of Article 6(1), the legitimate
     interests pursued by the controller or by a third party;
```

- Inform data subjects that honeypots are included in your networks
- Explain the reasons

Operating honeypots in distributed open networks

Information to be provided where personal data have not been obtained from the data subject

Article 14 : Regulation 2016/679

```
5. Paragraphs 1 to 4 shall not apply where and insofar as:
(b) the provision of such information proves impossible or would involve
a disproportionate effort, in particular for processing for archiving
purposes in the public interest, scientific or historical research
purposes or statistical purposes, subject to the conditions and safeguards
referred to in Article 89(1) [...]
```

- It is not obvious to identify data subjects
- Although there is a strong need to identify responsible point of contact to remediate situations such as fixing misconfigured systems, information leaks, or others

Data subject rights

- Rights
  - Right of access by the data subject (article 15)
  - Right of rectification (article 16)
  - Right of erasure (article 17)
  - Right of restriction of processing (article 18)
  - Right of data portability (article 20)
  - Right to object and automated individual decision making (article 21 & 22)
- Implementation
  - Data subjects using honeypot services → standard contractual implementations
  - Data about data subjects within collected datasets → challenge

Data subject rights

Challenges

- How to prove the identity of a data subject?
- How to prove that the requested data belongs to the data subject?
- How to ensure no interference with ongoing investigations by law enforcement (i.e. evidences collected by the honeypot)

Article 11: Regulation 2016/679

Processing which does not require identification

```
    2.  Where, in cases referred to in paragraph 1 of this Article, the
  controller is able to demonstrate that it is not in a position to identify the
   data subject, the controller shall inform the data subject accordingly, if
  possible. In such cases, Articles 15 to 20 shall not apply except where the
  data subject, for the purpose of exercising his or her rights under those
  articles, provides additional information enabling his or her identification.
```

In doubt, demonstrate to be not in position to identify data subjects (i.e. spoofed IP packets, compromised machines)

In case of doubts on the data subject or data ownership

Article 12: Regulation 2016/679

```
Without prejudice to Article 11, where the controller has reasonable doubts
    concerning the identity of the natural person making the request referred to
    in Articles 15 to 21, the controller may request the provision of additional
    information necessary to confirm the identity of the data subject.
```

Article 12: Regulation 2016/679

```
5. ... Where requests from a data subject are manifestly unfounded or excessive, in
     particular because of their repetitive character, the controller may either:
(a) charge a reasonable fee taking into account the administrative costs of
    providing the information or communication or taking the action requested; or
(b) refuse to act on the request.
```

Operating honeypots in distributed open networks

Article 89: Regulation 2016/679

```
Safeguards and derogations relating to processing for archiving purposes in the
      public interest, scientific or historical research purposes or statistical
      purposes

1.    Processing for archiving purposes in the public interest, scientific or
      historical research purposes or statistical purposes, shall be subject to
      appropriate safeguards, in accordance with this Regulation, for the rights and
       freedoms of the data subject. Those safeguards shall ensure that technical
      and organisational measures are in place in particular in order to ensure
      respect for the principle of data minimisation. Those measures may include
      pseudonymisation provided that those purposes can be fulfilled in that manner.
       Where those purposes can be fulfilled by further processing which does not
      permit or no longer permits the identification of data subjects, those
      purposes shall be fulfilled in that manner.
```

Operating honeypots in distributed open networks

Notes on pseudonymisation

Article 4: Regulation 2016/679

```
(5) pseudonymisation means the processing of personal data in such a manner that
    the personal data can no longer be attributed to a specific data subject
    without the use of additional information, provided that such additional
    information is kept separately and is subject to technical and organisational
    measures to ensure that the personal data are not attributed to an identified
    or identifiable natural person;
```

Honeypots as legitimate interests for CSIRTs

Recital 49: Regulation 2016/679

```
The processing of personal data to the extent strictly necessary and proportionate
     for the purposes of ensuring network and information security, i.e. the
     ability of a network or an information system to resist, at a given level of
     confidence, accidental events or unlawful or malicious actions that compromise
      the availability, authenticity, integrity and confidentiality of stored or
     transmitted personal data, and the security of the related services offered by
     , or accessible via, those networks and systems, by public authorities, by
     computer emergency response teams (CERTs), computer security incident response
      teams (CSIRTs), by providers of electronic communications networks and
     services and by providers of security technologies and services, constitutes a
      legitimate interest of the data controller concerned. This could, for example
     , include preventing unauthorised access to electronic communications networks
      and malicious code distribution and stopping denial of service attacks and
     damage to computer and electronic communication systems.
```

Conclusions

- GDPR supports internal honeypot operations
- GDPR supports distributed honeynet operations
- GDPR enables information sharing of honeypot data

- Contact: info@circl.lu
- `https://www.circl.lu/`
- `https://github.com/CIRCL/compliance`