Data protection by means of fragmentation

Summer school on real-world crypto and privacy

Katarzyna KAPUSTA

Self introduction

- PhD Student at **Telecom ParisTech Universite Paris-Saclay**
- Supervisor: Gerard MEMMI
- Subject: data fragmentation and dispersal as a way of data protection

- Education and previous experience:
  - M.Eng. Telecom ParisTech Universite Paris-Saclay, Paris, France
  - M.Sc. AGH University of Science and Technology, Cracow, Poland
  - Previous work experience :
    - Security consultant, E&Y, Paris
    - Software developer at Thales Communications & Security, Paris
    - Software developer intern at CERN, Geneva

Why do we need fragmentation?

- The security of encrypted data depends on the chosen algorithm, as well as on the strength and the secure storage of its key

- Fragmenting data into multiple fragments and dispersing these fragments over various locations aims at frustrating an attacker

- Nowadays, fragmentation is enabled by the cloud environment (large number of servers, multiple data centers) and already used for data resilience purposes (RAID, Hadoop)

Our division of data fragmentation

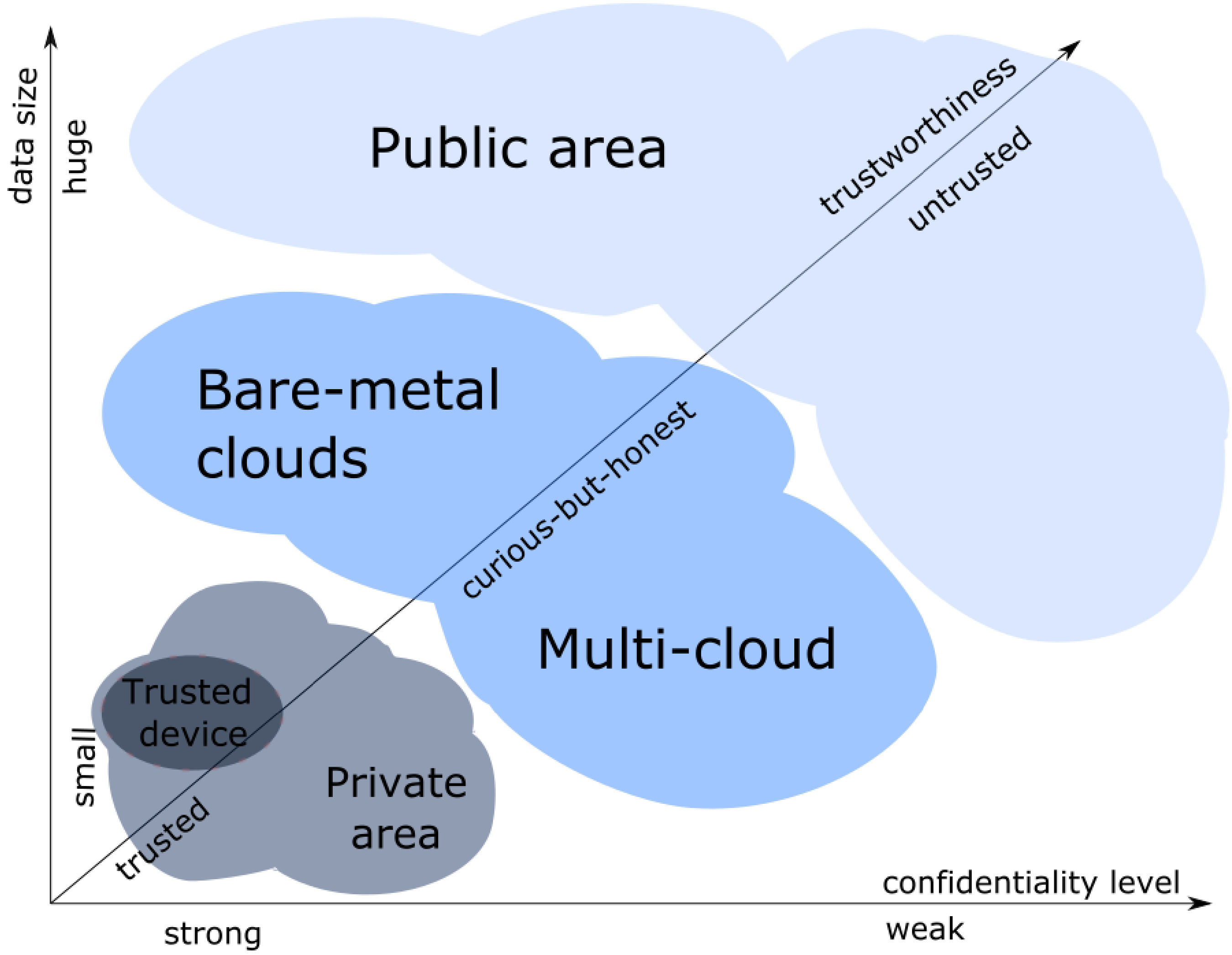
- **Bitwise**: fragmenting data without any consideration for their structure, their semantics, or their uneven level of confidentiality
  - *Techniques: perfect or computational secret sharing, information dispersal algorithms*

- **Structurewise**: exploiting data structures, multi-level confidentiality, and machine trustworthiness
  - *Techniques: database fragmentation, selective encryption*

Fragmentation in the cloud: issues

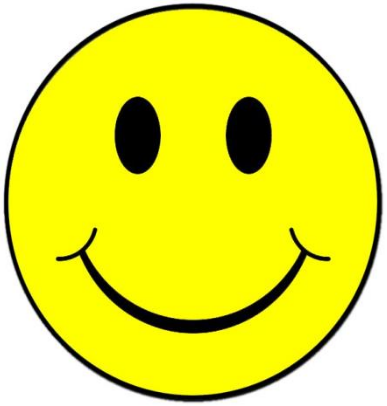
- Physical location control vs. virtualization
  - How to ensure secure data separation? Bare-metal cloud? Special agreement? Hybrid cloud? Coarse-grained solution: multi-cloud
- Latency problems: combining fragmentation with parallelization
- Defining security levels without user interaction for fragmentation of structured data

Fragmentation in the cloud: desired architectural traits

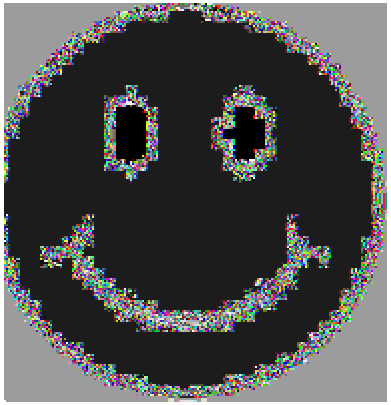


Improving information dispersal

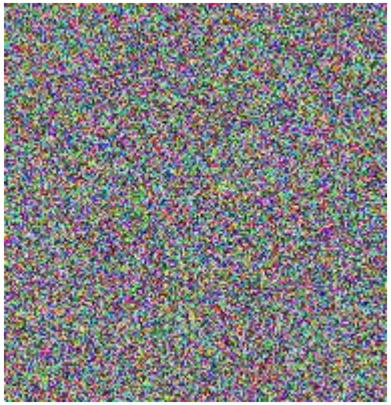
- Information Dispersal Algorithms (IDAs): a space-efficient keyless way to fragment and add resilience to data at the same time
- Used mostly in transmission scenarios
- **Problem**: lack of data protection, patterns are preserved inside the fragments
- **Solution**: a dispersal scheme that keeps the main properties of the IDAs while improving data protection (and also performance)



Original



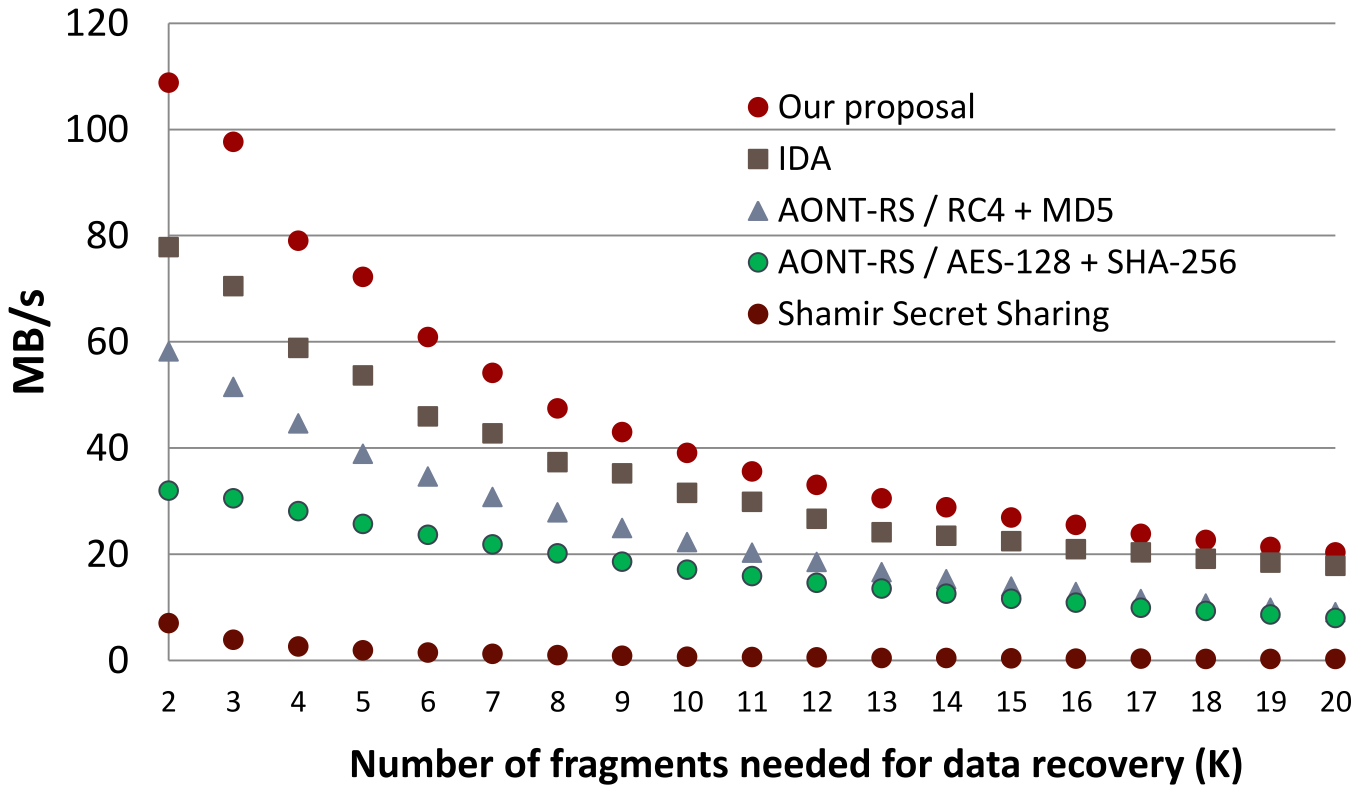
IDA



Our algorithm

Improving information dispersal

- Performance comparison with relevant works in an IoT scenario



References:

- K. Kapusta, G. Memmi, and H.Noura, ”POSTER: A Keyless Efficient Algorithm for Data Protection by Means of Fragmentation”, in ACM CCS 2016, Vienna, 2016.
- K. Kapusta and G. Memmi, ”Data protection by means of fragmentation in several distributed storage systems”, in CFIP-Notere, Paris, 2015.
- G. Memmi, K.Kapusta, and H.Qiu, ”Data protection by means of fragmentation in several distributed storage systems”, in Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC), 2015
- G. Memmi, K.Kapusta, and H.Qiu, ”Data Protection: Combining Fragmentation, Encryption, and Dispersion, an intermediary report”, ITEA2-CAP WP3 Intermediary Report, June 2015.

Thank you! ☺