Analyzing the Impact of GDPR on Storage Systems

Aashaka Shah, Vinay Banakar, Supreeth Shastri
Melissa Wasserman and Vijay Chidambaram
General Data Protection Regulation (GDPR)

May 25, 2018
Adopted after 2 years of public debate.
All but 2 EU countries have legislated.

Fundamental right
Grants all European people a right to protection and privacy of personal data

Personal data
Any information relating to a natural person;
Broad in scope unlike FERPA, HIPAA

Covers entire lifecycle
Collection, processing, protection, transfer and deletion; Regulated via 99 articles

Hefty penalty
Max penalty of 4% of global revenue or €20 million, whichever is greater
GDPR Entities

Data Subject (e.g., Spotify user)

Controller (e.g., Spotify)

Processor (e.g., Google cloud)

Supervisory Authority

Other Controllers (e.g., SoundCloud)

- Personal data
- GDPR queries

allow data sharing

provide personal data

exercise GDPR rights

send personal data for external processing

store and process personal data internally

report GDPR violations

audit and investigate

notify data breaches

Supervisory Authority

Data Subject (e.g., Spotify user)

Controller (e.g., Spotify)

Processor (e.g., Google cloud)

Other Controllers (e.g., SoundCloud)

- Personal data
- GDPR queries

allow data sharing

provide personal data

exercise GDPR rights

send personal data for external processing

store and process personal data internally

report GDPR violations

audit and investigate

notify data breaches
By the end of 2018 [Gartner 2018]

<50% estimated compliance

94,622 complaints from people

In the first 9 months of GDPR rollout
Analyzing GDPR: Two Key Observations

31 of the 99 GDPR articles directly pertain to storage systems.

GDPR’s goal of data protection by design and by default conflicts with the traditional system design goals of performance, cost, and reliability.
Investigate how GDPR-compliance impacts Storage Systems

- What effort is needed to make a modern storage system, GDPR-compliant?
- What is the resulting performance impact?
- Is it possible to achieve strict compliance in an efficient manner?
Key GDPR Articles concerning Storage Systems

**Rights of data subjects**

- [15] Right of **Access**
- [16] Right to **Rectification**
- [17] Right to Be **Forgotten**
- [20] Right to **Portability**
- [21] Right to **Object**

**Responsibilities of Data Controllers**

- [5] **Purpose / Storage** limitations
- [24] Responsibility of the controller
- [25] Protection by **Design** & by **Default**
- [30] **Records** of Processing activity
- [33] Notification of **Data Breaches**
## Translating GDPR Articles into Storage Features

<table>
<thead>
<tr>
<th>GDPR article</th>
<th>Key requirement</th>
<th>Storage feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Conditions for data collection</td>
<td>Store metadata associated with personal data</td>
</tr>
<tr>
<td>17</td>
<td>Right to be forgotten</td>
<td>Find and delete groups of data</td>
</tr>
<tr>
<td>25</td>
<td>Protection by design and by default</td>
<td>Safeguard and restrict access to data</td>
</tr>
<tr>
<td>30</td>
<td>Records of processing activity</td>
<td>Store audit logs of all operations on data</td>
</tr>
</tbody>
</table>

... complete table in the paper
# Features of GDPR-Compliant Storage

<table>
<thead>
<tr>
<th><strong>Timely deletion</strong></th>
<th><strong>Metadata indexing</strong></th>
<th><strong>Encryption</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Associate TTL to all personal data; it can be static value or a policy criterion</td>
<td>Provide quick and efficient access to groups of data</td>
<td>Encrypt data at rest, and while in transit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Manage data Location</strong></th>
<th><strong>Access control</strong></th>
<th><strong>Monitoring &amp; Logging</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to find and control the location of personal data at all times</td>
<td>Limit access to permitted entities, for established purposes, and for predefined duration of time</td>
<td>Save the audit trail of all internal actions and external interactions</td>
</tr>
</tbody>
</table>
GDPR-Compliance is a Spectrum

**Response Time**
- **Real-time**
  - Complete GDPR tasks synchronously in real-time
- **Eventual**
  - Complete GDPR tasks asynchronously

**Capability**
- **Full**
  - Support all GDPR features natively
- **Partial**
  - Support for some GDPR features is lacking or coarse-grained
Despite needing to implement a small set of new features for GDPR-compliance, storage systems would experience significant performance impact.
## Redis’ support for GDPR features

<table>
<thead>
<tr>
<th>Feature</th>
<th>FULL</th>
<th>PARTIAL</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monitoring &amp; Logging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage data Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timely deletion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata indexing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Encryption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GDPR-Compliant Redis: Monitoring & Logging

Three built-in options

- **MONITOR** debug command
- Configure **slowlog** option
- Piggyback on **AOF**

*modified AOF code to include read/scan operations*

Even fully supported features can cause significant **performance overheads**
GDPR-Compliant Redis: **Timely Deletion**

Three options to delete

- **DEL** and **UNLINK**
- **FLUSH{DB | ALL}**
- **EXPIRE** and **EXPIREAT**

Redis erases expired keys using a lazy randomized algorithm. We changed it to a static scheme (== sub-second latency for up to 1M keys).

System internals should be carefully analyzed to determine the **degree of compliance**.
GDPR-Compliant Redis: Encryption

No native support

- Encryption at rest w/ LUKS
- Encryption in transit w/ STunnel

Investigated key-level encryption using Themis (similar performance overhead)

Retrofitting new features not aligned with the core design principles of the system will result in excessive performance overheads.
Concluding Remarks

**GDPR-compliant Redis**
Performance impact of GDPR on a modern storage system

**Research challenges**
Efficient Logging; Efficient Deletion; Efficient Metadata indexing

**Beyond GDPR**
California's CCPA is going into effect 1/1/2020

We want to hear from you!
https://utsaslab.github.io/research/gdpr/