Zelda - Data Hosting, Handling and Backup Policy

Group 09

May 2019

1 Introduction

The purpose of this report is to cover the basis of hosting, handling and backing up of data in the zelda application. Zelda is an educational information management application, hosting sensitive data relating to persons and institutional data. Thus, there is an ethical and legal need to define how to host, handle and backup this data.

2 Hosting data

Zelda's data is entirely hosted in a relational database, Postgres[1], which is hosted in a server not publicised for service providing and only allowing connections in a private network, shared by the servers available to the service infrastructure. Despite this, there is always a danger of system compromise either by malware or physical access.

To ensure the safety of data, the Postgres server must be protected with a user and password, only allowing queries with authenticated sessions. Moreover, the system user in which the server is running must also be protected by password.

3 Handling data

Data can be distributed (and handled) in two contexts, each context having its means of distribution.

This section defines and prescribes best practices for each context of distribution of data in the zelda application.

[1] `https://www.postgresql.org/`

3.1 Application Services

In the context of application services, data is requested by a client from a server, which has access to the data contained in the database. Due to this, any request could be a possible attempt to acquire access to data that the client should not have access to.

It is important that permissions for each type of users are well defined, where a type of user can only have access to a certain amount of data types. Thus, any request by an unauthenticated client will be discarded. At each request, the access permissions of the client are consulted before giving access to the data.

At this context, the means of transportation of data is the network, which is one of the possible attack vectors. All communication should be encrypted with HTTPS[2] to mitigate any attempt at intercepting the sensitive data.

3.2 Internal Handling

Internal handling of data can be done in the context of system management of the infrastructure. During handling of data, one should never transfer data without encryption. A post-handling clean-up is essential to avoid leaving behind any temporary files that may either directly compromise access of data or create vulnerabilities that may lead to the compromise of data.

4 Data backup

To avoid loss of data, the system automatically backs up the data periodically. These backups are kept in files in multiple servers, to diminish the probability of data loss at the level of backups.

4.1 Backup Frequency

The data is backed up twice a day, ensuring a backup every 12 hours. These backups are done at a time of the day where there are less activity as to not impact performance.

4.2 Backup storage

The backups are performed by each server available to the infrastructure that is running the application, stored as a file. This ensures that is replicated by many machines, preventing loss to events of data corruption, server unavailability, and other critical events that may render a machine unavailable.

[2] `https://www.w3.org/2001/tag/doc/web-https`, `https://en.wikipedia.org/wiki/HTTPS`

4.3 Backup encryption

To prevent data leaks in event that a person gains access to any of the servers either by physical access or remote access, the backups are encrypted with OpenPGP standard using GnuPG implementation[3], with a 4096-bit RSA key.

5 Backup Retention

All data must be retained, regardless of storage concerns. The backups are periodically rotated, with 7 days worth of backup state kept at anytime.

This stance of guarantee of data retention requires a careful care in PGP key maintenance. Keys must be periodically rotated, and at each rotation the available backups must be decrypted, a new key is generated, and then encryption of backups ensues.

[3] `https://www.gnupg.org/`