THE QUEUE-POLL DATASHIELD EXTENSION

Standardised, Monitored, Indirect, and Secure Access to Sensitive Data

Julian Gruendner - Chair of Medical Informatics, Erlangen

12.09.2019
DataSHIELD Workshop 2019
Background

• Biomedical research requires large amounts of data
• Multi institutional networks necessary to provide such data pools
• One of the major hurdles are the restrictions placed on medical data (data security, patient privacy, governance ... )
Background – MIRACUM

- MIRACUM consortium funded by MII (medical informatics initiative) seeks to analyse data across hospitals
- MIRACUM Use Case 2: Development of clinical predictive models across hospitals
- Data protection policy: Analyze data without the data leaving the hospitals
DataSHIELD does **not** require data to leave hospitals, but it **requires requests to enter and responses to exit** the hospital network.

**Challenges:**

- Application provider is **not** system admin.
- Firewalls are usually restrictive in a hospital setting.
- Joining and exiting an existing data analysis network should be easy.
Background

Problem

STRICT FIREWALLS of hospitals allow no direct access to Opal server from outside network

Solution

QUEUE-POLL implementation to allow indirect access from outside
DS – Queue-Poll Extension
DS – Queue-Poll Extension

Poll Service and Opal Server (VM)

- **DS POLL GUI Server Container**
- **DS POLL Container**: GUI Web Service and Running Poll Service
- **Opal Server Container**
  - Data & Metadata Module
  - R Module
  - DataSHIELD
- **Web Services**
- **Mongo DB**
- **R Server Container (RServe)**

Networks:
- **datashield_docker_opal_net – docker network**
- **HTTPS** connections

**FIREWALL**
- HOSPITAL DMZ
- HOSPITAL NETWORK

**CLIENT**

**QUEUE**

**HTTPS** connections
DS – Configurable Circle of Trusted IPs
DS – Queue-Poll Extension

Poll Service and Opal Server (VM)

- DS POLL GUI Server Container
- DS POLL Container: GUI Web Service and Running Poll Service
- Opal Server Container
- Data & Metadata Module
- R Module
- DataSHIELD
- Web Services
- Mongo DB
- R Server Container (RServe)

CLIENT

HTTPS

QUEQUE

HTTPS

HOSPITAL NETWORK

HTTPS

HOSPITAL DMZ

FIREWALL

FIREWALL

Research Database

prepared dataset

project proposal

UAC (Use and Access Committee)
How is extra security provided?

- Queue installed in **DMZ behind firewall**, which only allows incoming requests on port 443
- Queue **only accepts Connections from authorized peers**, configured via IP address
- Opal has built-in **roles and rights management**, which further restricts data access
- **Data Provisioning only on official request to** the hospital UAC (use and access committee)
Performance of Q-P

- Impact of Q-P decreases with larger number of data records in dataset:
  - 5000 data records = 2.8 times direct request time
  - 50,000 data records = 2.2 times direct request time
  - => percentage cost of Q-P decreases with larger data sets
The Installation Process

• **Requirements**: Ubuntu 18 or > (other versions not tested)

• **Fully automated installation** and provisioning process includes:
  
  • The Q-P extension
  
  • Opal and Datashield
  
  • Test data, which is automatically loaded into opal
  
  • An R server
Summary

• General Q-P Implementation, which reverses request handling from push to pull

• Q-P allows only specific requests to specific applications

• Created a privacy preserving data analysis network, which can be easily joined and exited

• Q-P and DataSHIELD allows hospitals to stay in charge of their patient data and their participation in research

• Paper available here:

Demonstration/ Live Installation
Demonstration/ Live Installation

• Where to find it:
  - https://github.com/juliangruendner/ds_develop

• Installation packages:
  - https://github.com/juliangruendner/ds_develop/tree/master/build_from_images
DS – Configurable Circle of Trusted IPs
Demonstration/ Live Installation

FIREWALL

HOSPITAL DMZ

Analysis Server (VM)

- ANALYSIS
- NGINX

Queue Server (VM)

- QUEUE
- NGINX

HTTPS

443

FIREWALL

HOSPITAL NETWORK

Poll Service and Opal Server (VM)

- Opal Server Container
  - Data & Metadata Module
  - R Module
  - DataSHIELD
  - Web Services
  - Mongo DB
  - R Server Container (RServe)

- DS POLL Container: GUI Web Service and Running Poll Service

HTTPS
Questions?