

Team 4: Bi-Weekly Report 9

Project Title: PEACH Data Mining

Date: 9 March 2018

Gavin Shek, Saqib Jahangir, David Stepanovs

Overview

During these 2 weeks, we have made improvements to the anonymisation tool. We have communicated with a potential user of the software application and have taken feedback on board and added features and made fixes to certain parts of the user interface. Next, we have also now deployed the analytics system along with the visualisation tool on an Azure virtual machine.

Meetings

Numerous Meetings During and Outside of Lab Sessions

Unfortunately, we have faced problems with deploying the system on the distributed operating system DC/OS. This is the operating system that is recommended with the analytics infrastructure but since we have been unable to deploy successfully, and after much debugging, we decided to deploy the system on an Ubuntu Server virtual machine instead.

Completed Tasks

- Made improvements to the anonymisation application, based on user feedback
- Deployed the analytics system on Azure
- Ensured that Kibana is able to display graphs and visualisations with the data output by the core analytics

Problems to be Solved

- Write user manuals, and technical documentation on the deployment of the analytics system
- Provide a simple API for users to be able to upload data into the analytics system

Plan

- Create the documentation on using the software and deploying it
- Create a REST API along with a web application which allows users to upload files directly to the analytics system

- Implement a feature into the anonymisation solution which allows generated and anonymised files to automatically be uploaded to the core analytics system

Individual Section

Gavin Shek

During these 2 weeks, I have written scripts to allow the core analytics system to communicate with the Elastic stack. The data is uploaded and stored in Elasticsearch having been output by Apache NiFi.

Saqib Jahangir

I have attempted to write NiFi templates.

David Stepanovs

I have made improvements to the data anonymiser tool, by fixing problems and introducing features based on user feedback.