Anonymisation and privacy User Manual

SPHINX

A Universal Cyber Security Toolkit for Health-Care Industry

Table of contents

Table of figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

1 Introduction

Chimera is a component for collecting and anonymizing data during the collection. For example, the anonymization framework can parse data from databases or text files and discard or pseudonymize data and convert specific values to hash values in order to enhance privacy. This tool is frequently used for conducting analysis and it can also collect network traffic directly from a network interface and proceed at the anonymization processes as well. There are two different options for enabling the chimera. The first option is to enable it to parse specific sources and by using the Web UI to create models for processing the data according to specific rules. The second option is to enable an agent which collects data from files, network interfaces or databases and sends them to another endpoint (e.g., the SIEM). These data are also parsed using the predefined rules that will discard or anonymize specific data.

Chimera is a dataflow application, integrated in a Web User Interface that can communicate with the Orchestration-Frameworks APIs allowing a user to manipulate knowledge and data generated by other tools. This tool provides que current functionalities:

1. Standalone as GUI with CHIMERA_STUDIO for data exploration & data workflow design
2. WebService Integrated with OF for CHIMERA queries & data workflows
3. WebService Integrated with OF for exporting Microsoft Outlook PST Files into a folder (attachments) & json file (metadata & messages)

2 Installation/Deployment

A service for the web UI can be executed using a docker image ready to build using a docker at port 3000 using the web browser. If the service is already deployed the chimera component is possible to be accessed from the web browser to the appointed HTTP port. If no model is required by using the Web UI, the agents must be installed. The procedure for the agent's installation is the same as the SIEM, since the same agent is being used.

3 Use Case 01: Filter data

You have a CSV file and only want to select the second column. In the next 3 tables the query, input and the result are presented.

3.1 Query

```
split(,) | puts $1
```

3.2 Input

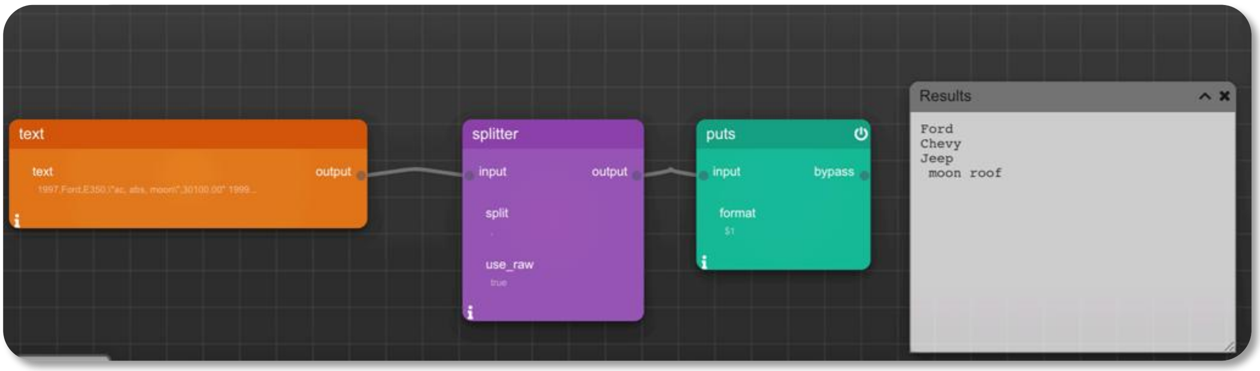
```
1997,Ford,E350,\"ac,abs, moon\",30100.00"
1999,Chevy,"Venture ""Extended Edition""",,49000.00
1996,Jeep,Grand Cherokee,"MUST SELL!
air22, moon roof, loaded",479699.00
```

3.3 Output

The chimera component looks like the image below. A text is provided as input the splitter separates the values and the puts extracts the 1st column after the commas (e.g., Ford) etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

```
Ford
Chevy
Jeep
moon roof
```



```
Ford
Chevy
Jeep
 moon roof
```

Figure 1 Chimera model for extracting the second column only from a csv

4 Use Case 02: Healthcare Data encryption

In the following model in the orange box, we put the path osf the csv file we want to encrypt/anonymize. Then we apply the splitter according to the commas (since this is a csv file).

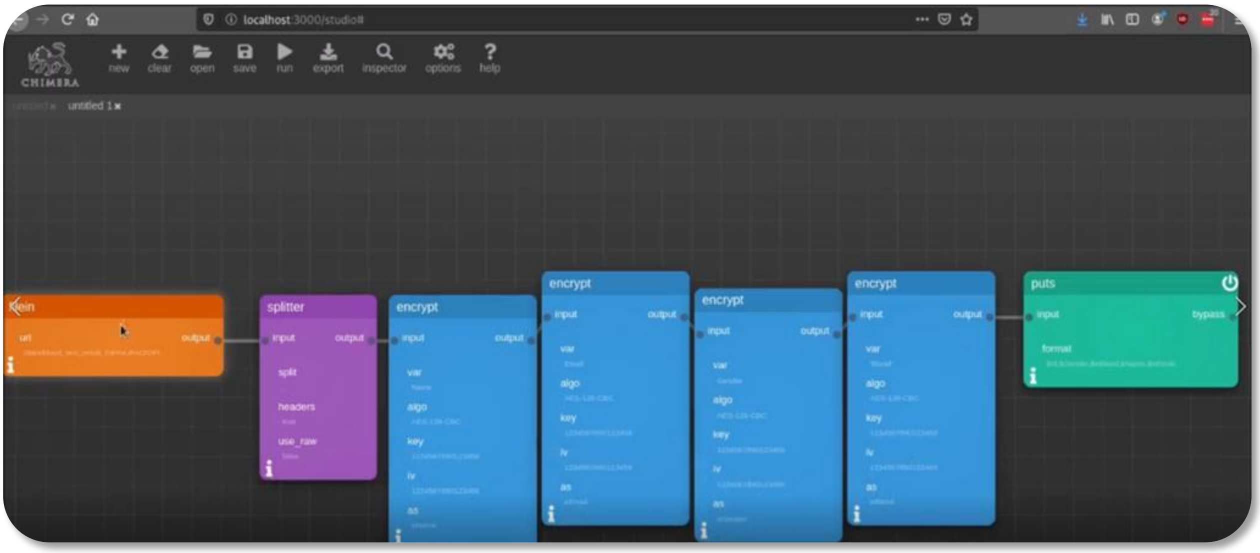
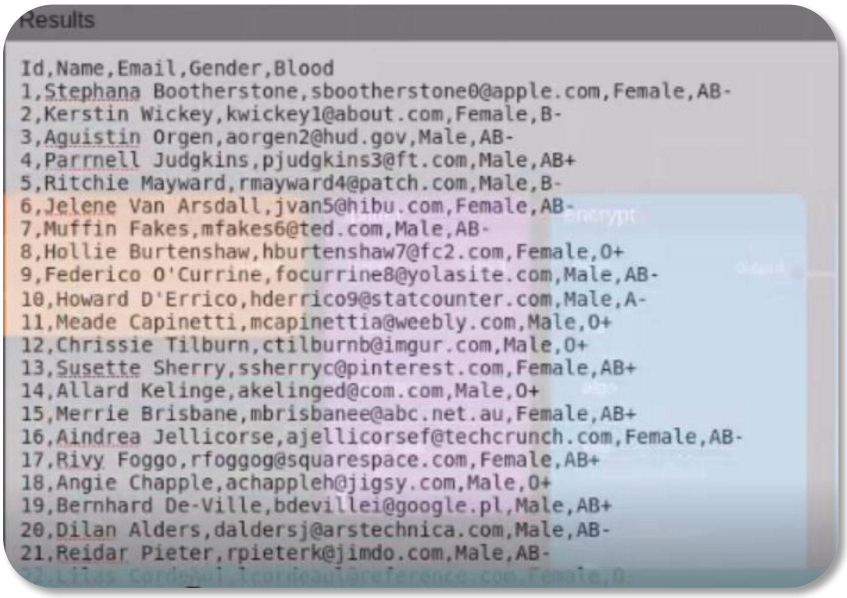


Figure 2 Chimera model for encrypting healthcare data from a csv

Afterwards we select which data tags we want to encrypt (blue boxes – Name, Email, Gender, Blood-Type). Then we extract the data using the puts or we can redirect the output to another file. The csv before the anonymization is presented in **Figure 3**.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

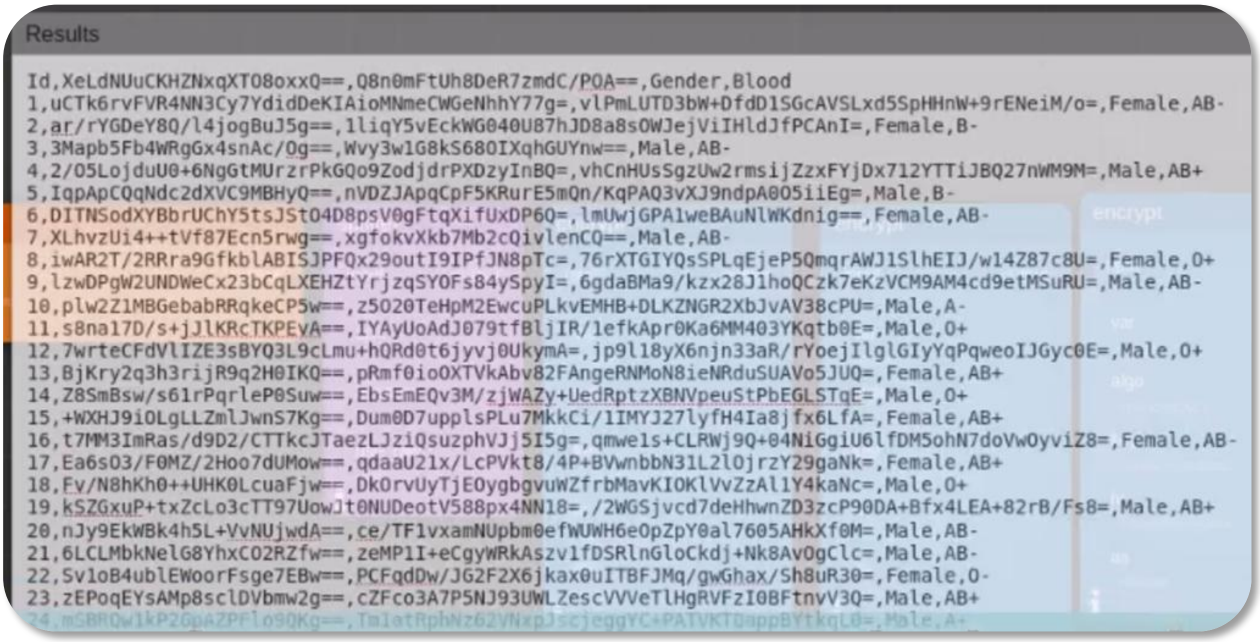


Results

```
Id,Name,Email,Gender,Blood
1,Stephana Bootherstone,sbootherstone0@apple.com,Female,AB-
2,Kerstin Wickey,kwickey1@about.com,Female,B-
3,Aguistin Orgen,aorgen2@hud.gov,Male,AB-
4,Parrnell Judgkins,pjudgkins3@ft.com,Male,AB+
5,Ritchie Mayward,rmayward4@patch.com,Male,B-
6,Jelene Van Arsdall,jvan5@hibu.com,Female,AB-
7,Muffin Fakes,mfakes6@ted.com,Male,AB-
8,Hollie Burtenshaw,hburtenshaw7@fc2.com,Female,O+
9,Federico O'Currine,focurrine8@yolasite.com,Male,AB-
10,Howard D'Errico,hderrico9@statcounter.com,Male,A-
11,Meade Capinetti,mcapinettia@weebly.com,Male,O+
12,Chrissie Tilburn,ctilburnb@imgur.com,Male,O+
13,Susette Sherry,ssherryc@pinterest.com,Female,AB+
14,Allard Kelinge,akelinged@com.com,Male,O+
15,Merrie Brisbane,mbrisbanee@abc.net.au,Female,AB+
16,Aindrea Jellicorse,ajellicorsef@techcrunch.com,Female,AB-
17,Rivy Foggo,rfoggog@squarespace.com,Female,AB+
18,Angie Chapple,achappleh@jigsy.com,Male,O+
19,Bernhard De-Ville,bdevillei@google.pl,Male,AB+
20,Dilan Alders,daldersj@arstechnica.com,Male,AB-
21,Reidar Pieter,rpieterk@jimdo.com,Male,AB-
22,Lilas Cordeaul,lcordeaul@reference.com,Female,O-
```

Figure 3 Data from CSV before anonymizing

After the model processes the csv data the following data are presented (**Figure 4**). The second column is anonymized while the Gender and the Blood type is still there.



Results

```
Id,XeLdNUuCKHZNxqXTO8oxxQ==,Q8n0mFtUh8DeR7zmdC/POA==,Gender,Blood
1,uCTk6rvFVR4NN3Cy7YdidDeKIAioMNmeCWGeNhhY77g=,vlPmLUTD3bW+DfdD1SGcAVSLxd5SpHHnW+9rENeiM/o=,Female,AB-
2,ar/rYGDeY8Q/l4jogBuJ5g==,1liqY5vEckWG040U87hJD8a8sOWJejViIHldJfPCAnI=,Female,B-
3,3Mapb5Fb4WRgGx4snAc/Og==,Wvy3w1G8kS68OIXqhGUYnw==,Male,AB-
4,2/O5LojduU0+6NgGtMUrzrPkGQo9ZodjdrPXDzyInBQ=,vhCnHUsSgzUw2rmsijZzxFYjDx712YTTiJBQ27nWM9M=,Male,AB+
5,IqpApCQqNdc2dXVC9MBHyQ==,nVDZJApqCpF5KRurE5mQn/KqPAQ3vXJ9ndpA0O5iiEg=,Male,B-
6,DITNSodXYBbrUChY5tsJStO4D8psV0gFtqXifUxDP6Q=,lmUwjGPA1weBAuNlWKdnig==,Female,AB-
7,XLhvzUi4++tVf87Ecn5rwg==,xgfokvXkb7Mb2cQivlenCQ==,Male,AB-
8,iwAR2T/2RRra9GfkblABISJPFQx29outI9IPfJN8pTc=,76rXTGIYQsSPLqEjeP5QmqrAWJ1SlhEIJ/w14Z87c8U=,Female,O+
9,lzwDPgW2UNDWeCx23bCqLXEHZtYrjzqSYOFs84ySpyI=,6gdaBMa9/kzx28J1hoQCzk7eKzVCM9AM4cd9etMSuRU=,Male,AB-
10,plw2Z1MBGebabRRqkeCP5w==,z5O20TeHpM2EwcuPLkvEMHB+DLKZNGR2XbJvAV38cPU=,Male,A-
11,s8na17D/s+jJlKRcTKPEvA==,IYAyUoAdJ079tfBljIR/1efkApr0Ka6MM403YKqtb0E=,Male,O+
12,7wrteCFdVlIZE3sBYQ3L9cLmu+hQRd0t6jyvj0UkymA=,jp9l18yX6njn33aR/rYoejIlglGIyYqPqweoIJGyc0E=,Male,O+
13,BjKry2q3h3rijR9q2H0IKQ==,pRmf0ioOXTVkAbv82FAngeRNMoN8ieNRduSUAVo5JUQ=,Female,AB+
14,Z8SmBsw/s61rPqrleP0Suw==,EbsEmEQv3M/zjWAZy+UedRptzXBNVpeuStPbEGLSTqE=,Male,O+
15,+WXHJ9iOLgLLZmlJwnS7Kg==,Dum0D7upplsPLu7MkkCi/1IMYJ27lyfH4Ia8jfx6LfA=,Female,AB+
16,t7MM3ImRas/d9D2/CTTkcJTaezLJziQsuzphVJj5I5g=,qmwe1s+CLRWj9Q+04NiGgiU6lfDM5ohN7doVwOyviZ8=,Female,AB-
17,Ea6sO3/F0MZ/2Hoo7dUMow==,qdaaU21x/LcPVkt8/4P+BVwnbbN31L2lOjrzY29gaNk=,Female,AB+
18,Fv/N8hKh0++UHK0LcuaFjw==,DkOrvUyTjEOygbgvuWZfrbMavKIOKlVvZzAl1Y4kaNc=,Male,O+
19,kSZGxuP+txZcLo3cTT97UowJt0NUDeotV588px4NN18=,/2WGSjvcd7deHhwnZD3zcP90DA+Bfx4LEA+82rB/Fs8=,Male,AB+
20,nJy9EkWBk4h5L+VvNUjwdA==,ce/TF1vxamNUpbm0efWUWH6eOpZpY0al7605AHkXf0M=,Male,AB-
21,6LCLMbkNelG8YhxCO2RZfw==,zeMP1I+eCgyWRkAszv1fDSRlnGloCkdj+Nk8AvOgClc=,Male,AB-
22,Sv1oB4ublEWoorFsge7EBw==,PCFqdDw/JG2F2X6jkax0uITBFJMq/gwGhax/Sh8uR30=,Female,O-
23,zEPoqEYsAMp8sclDVbmw2g==,cZFco3A7P5NJ93UWLZescVVVeTlHgRVFzI0BFtnvV3Q=,Male,AB+
24,mSBRQw1kP2GpAZPFlo9QKg==,TmlatRphNz62VNxpJscjeggYC+PATVKTOappBYtkqL0=,Male,A+
```

Figure 4 Data from CSV file after anonymization

The above procedures are the executed periodically and can be used to databases, network interfaces or other data sources (text files, csv, XML etc.)