

Predicting and visualising ATM attacks

R. Fasel and B. Regnerus

Faculty of Electrical Engineering, Mathematics and Computer Science

University of Twente

7500 AE Enschede, The Netherlands

I. INTRODUCTION

ATM attacks and fraud continue to make headlines. The compromise of ATMs is a very lucrative criminal business. ATMs typically exist as physical contact points of the banking infrastructure that are exposed to public sight. This makes ATMs a very accessible target for exploitation, vulnerable to a large variety of attack scenarios. The European ATM Security Team report that the amount of ATM attacks both physical- and logical attacks only increase [1], therefore securing ATMs but also predicting ATM attacks are more important then ever.

This paper describes the process of both visualising ATM attacks and predicting cases where ATM attacks could occur. The main focus of this of this research is the exploration of using tools provided in the DM- and DPV topics provided in the Data Science course. Furthermore other tools such as the Python machine learning library Scikit-learn and JavaScript data visualisation library D3.js will be used.

A. The Dataset

Provided for this project is a dataset from the European TREsPASS project [2] that was developed in collaboration with Spanish banks. The dataset contains information of 723 ATMs, attacks on those ATMs and the information about those attacks. In more detail, the dataset provides: geographical details about ATMs (such as latitude and longitude coordinates, distance to a highway), demographic details (such as per capita income, population density, median population age), logging details of the attacks per ATM, the profile of the involved attacker, the information on how much amount was recovered and when did recovery event took place.

For this research the most important features of the dataset are both the geographical- and demographical features. Those features can be used for both visualising attacks and predicting attacks using a classifier with those features as an input.

B. Research Questions

With the DataScience DM and DPV topics in mind this lead to the following research questions:

How can a dataset containing geographical- and demographical features of ATM attacks be predicted and visualised?

This question is answered with the support of the following questions:

- How can factors that contribute to attack frequency and attacker success be plotted geographically?

- Which factors contribute most to attack frequency and attacker success?
- How can ATM attacks be predicted using a Naive Bayes classifier?

II. METHOD

To give an accurate answer the above mentioned research questions multiple different tasks have to be conducted. The tasks have been further separated in the application of topic DPV (II-A) and the application of topic DM (II-B).

A. Application of Topic DPV

In this section we try to identify a good method for obtaining visualisation from the raw ATM attack dataset. Raw data was available in an ARFF format and a good description of each data column was also available. The data which was available was already in quite a clear format and did not need many transformations to make it available for visualisation. Therefore Pentaho Kettle was not used, instead basic transformations were performed to transform the ARFF file to a GeoJSON format using the Python language, since the data which is encoded consists of geographic data points this format seemed appropriate. For the visualisation we wanted to also show the contours of the freguesias (municipalities of Lisbon). This data was not available in the ATM attack dataset, however it was available under an Creative Commons licence from the Lisbon Open Data website [4].

Using the two GeoJSON formatted datasets a visualisation was created in Javascript using D3.js. [5] The use of the D3.js library was chosen over other tools such as Tableau since it allows for more flexibility when developing visualisations. D3.js is able to read from the GeoJSON formatted files, using them as the datasource therefore no external database warehouse, such as MySQL, was needed.

B. Application of Topic DM

(TODO)

III. RESULTS

Following from II (the method) a visualisation was developed in D3.js (III-A) together with a machine learning model created using a Naive Bayes classification method (III-B).

A. Results of Topic DPV

The visualisation which was developed contains the outlines of the freguesias as a grey shape, on top of those shapes the ATMs are plotted geographically as black dots. ATMs which have been attacked have a red circle around them. The visualisation can be further explored by turning on- and off other visualisation layers. There can be chosen from a list of nine different features that will be visualised as circles on the map. Most prominently, the predicted attack probability can be visualised, this probability is the outcome of II-B (the Application of Topic DM) and will be further discussed in III-B (Results of Topic DM). Other features that can be displayed are, inter alia, motorway distance, or police distance. These features were all present in the ATM attacks dataset. The scale of the circles is explained by a circle legend giving the values of the corresponding circle size. The source code and full visualisation can be found on Github [4][6].

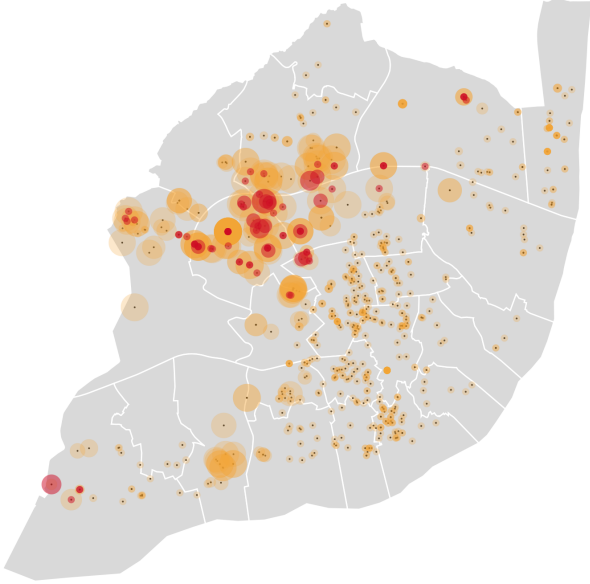


Fig. 1: Visualisation of ATM attacks (red circles), Predicted ATM attacks (orange circles) and ATM locations (black dots) on the map of Lisbon. [3]

B. Results of Topic DM

(TODO)

TABLE I: Performance results

	Precision	Recall	F1-score	Support
XX	0%	0%	0%	0
XX	0%	0%	0%	0
avg / total	0%	0%	0%	0

IV. CONCLUSION AND DISCUSSION

According to the results from both the data visualisation and machine learning prediction, it is possible to accurately predict locations for ATM attacks and visualise the attacks

with the geographical- and demographical features provided by the ATM attacks dataset.

Since the data was labeled with latitude and longitude coordinates the visualisation created in D3.js proves that it is possible to plot attack frequency and attacker success geographically. There are however a few drawbacks from the visualisation. Since some locations, for example a bank, have multiple ATMs at the same location, the circles overlap and therefore can distort the visualisation. In addition all features are labeled to a specific ATM location, therefore some locations have a lot of demographical data available while other places have little data available.

(EXPLAIN FACTORS WHICH CONTRIBUTE TO ATTACKS)

(EXPLAIN RESULTS CLASSIFIER)

Further research should be done to optimise the machine learning model even further. More datasources could be sourced as input features for the model. Furthermore there could be experimented with other types of classifiers to further improve the classification.

REFERENCES

- [1] European ATM Crime Report Archives — European ATM Security Team, European ATM Security Team. [Online]. Available: <https://www.european-atm-security.eu/tag/european-atm-crime-report/>. [Accessed: 13-Apr-2017]
- [2] The TRESPASS Project. [Online]. Available: <https://www.trespass-project.eu>. [Accessed: 13-Apr-2017]
- [3] ATM Attacks. [Online]. Available: <https://regnerus.github.io/atm-attack-data-visualisation/>. [Accessed: 13-Apr-2017]
- [4] Cmara Municipal de Lisboa - Organisationer - Portal Dados Abertos. [Online]. Available: <http://dados.cm-lisboa.pt/sv/organization/camara-municipal-de-lisboa>. [Accessed: 14-Apr-2017]
- [5] M. Bostock, D3.js - Data-Driven Documents. [Online]. Available: <https://d3js.org/>. [Accessed: 14-Apr-2017]
- [6] regnerus, regnerus/atm-attack-data-visualisation, GitHub. [Online]. Available: <https://github.com/regnerus/atm-attack-data-visualisation>. [Accessed: 14-Apr-2017]