Prediction and Anomaly Detection for Enterprise Access Control

by

Andre D'Souza
University of Toronto

April 2019

Supervised by Dr. Yuri Lawryshyn

Engineering Science Thesis Executive Summary

Introduction

The scale and complexity of large organizations causes the administration of network security to be time consuming, error prone and expensive [1]. Role Based Access Control has emerged as the best practice for access control in organizations over 500 members, and operates in a two tier system mapping network permissions to roles and then users to various roles based on their access needs. Each permission grants a resource (network system such as a database, file access...) to an employee. When organizations possess dynamic user bases with changing employee responsibilities, Role Based Access Control requires excess role creation to minimally grant sufficient resource access, becoming a costly and inefficient access control method [1].

The objective of this research is to detect anomalies in permissions granted to users for existing access control configurations, and apply recommendation systems to predict future access control mapping and maximize enterprise network security. To accomplish this goal, nonessential permissions in user resource access control mapping are identified, and a resource approval predictor for granting new permissions to users is applied. An improved access control mapping technique will better administer minimal permission granting to users, improving cybersecurity within large organizations. In the cases of employee error or system compromise, enterprise systems will be less vulnerable to data loss compared to the implementation of traditional Role Based Access Control. Predicting user resource permissions will also reduce access control administration costs and employee downtime.

Background

Role Based Access Control is easier maintained when fewer generalized roles are defined, although network security is enhanced by minimizing the permissions granted to each user. Overly customizing roles to specific users requires extensive efforts from network administrators, making Role Based Access Control costly and inefficient to maintain. Excess role creation is mitigated by linking multiple roles to single users, leading to permission set covering issues [1]. The omission of necessary permissions creates barriers to user functionality, while granting excess permissions during user-role mapping creates the security risks outlined above. Attribute Based Access Control has emerged an improved access control method by considering employee attributes and request context as well. However, the attribute based version has not been widely adopted due to upgrade cost outweighing the potential benefits received.

Recommendation systems are actively used to connect users to products (ie. Netflix, Amazon) based on patterns in user activity, attributes and item interaction [2]. The recommendation system approach may be used to predict employee access control mapping as an improved method over Role Based Access Control. The implementation of a hybrid recommender system through an ensemble algorithm is proven to have the most accurate output [3], and has the potential to become an enterprise access

control solution. The PyOD toolkit for Python also contains over 20 scalable outlier detection algorithms, and can be applied to detect anomalies in access control mapping [4].

Methodology

The processes used in this thesis can be grouped into three stages: data preparation, prediction of access control mapping, and outlier identification (see Figure 1). For both data preparation and prediction, inspiration was drawn from publically available solutions to Kaggle's Amazon Access Challenge [5]. However, all methodology and results obtained pertain to analysis conducted on an anonymized dataset provided by the project sponsor, a Canadian financial institution.

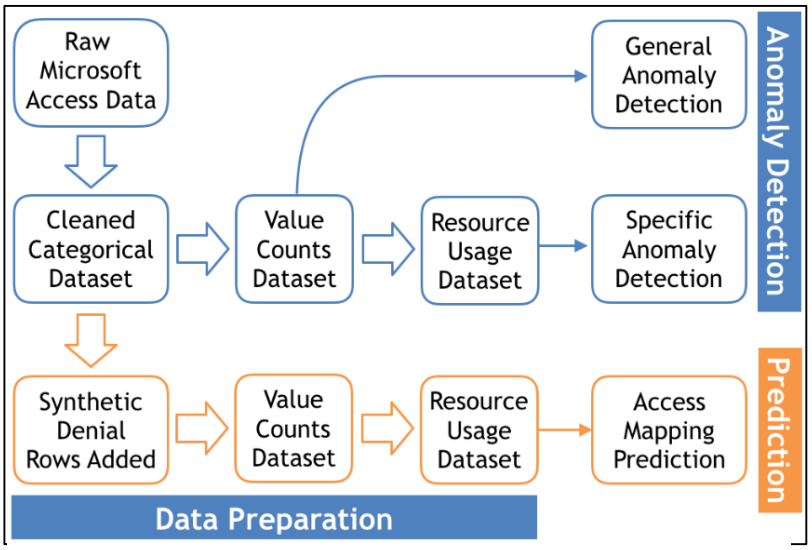


Figure 1. Methodology Overview

The project sponsor dataset contains existent access control mapping, including various employee attributes (such as Title, Department, Manager…). The raw data was parsed into a tabular format, with each row representing a unique employee resource pairing, along with that employee's respective attributes. The cleaned dataset includes 956 employees, 8k resources, and 66.3k unique employee resource pairing records. Each record indicates a resource provided to an individual employee, so while both resources and employees may appear in multiple records, a unique pairing will only appear once. Each entry in the dataset was then converted into value counts, or number of unique appearances in the dataset. Similarly, resource usage percentages were calculated for each entry, indicating the percentage of that particular attribute's occurrences in the dataset related to the specific resource. To create a training set for access control prediction, 4k synthetic resource denial rows containing unique employee resource pairings absent from the original dataset were created and appended. The addition of synthetic rows labelled the outcome for each access control record as approved or denied, with a 94% approval rate.

For anomaly detection purposes, 21 feature pairings were created using each possible combination of the seven employee attribute resource usages. The 21 feature pairings were then analyzed with three classifiers: Isolation Forest, Histogram Based Outlier Detection and Cluster Based Local Outlier Factor (see Figure 2). The classifiers yielded 63 outlier classifications for each entry, which were aggregated into an Outlier Score. Furthermore, for the prediction of access control mapping an 80/20 training testing split was applied to the dataset with synthetic rows added. The preliminary recommendation system composed of Random Forest, Extra Trees and Gradient Boosting, and yields a probability of resource approval for each entry (grant resource to employee or deny).

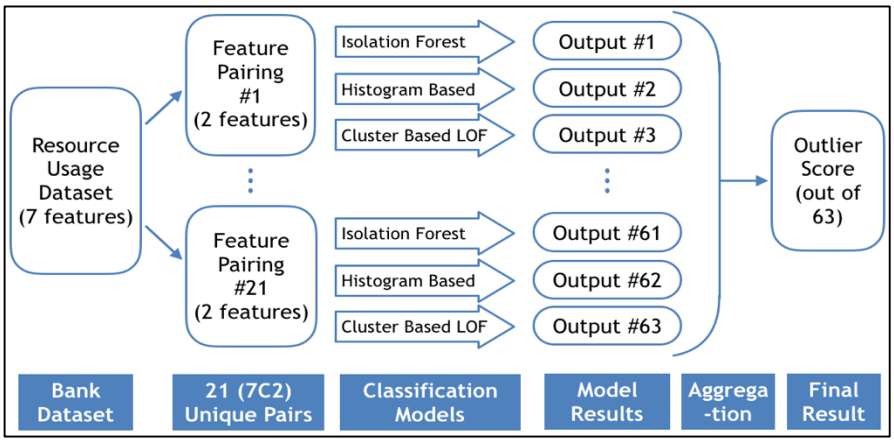
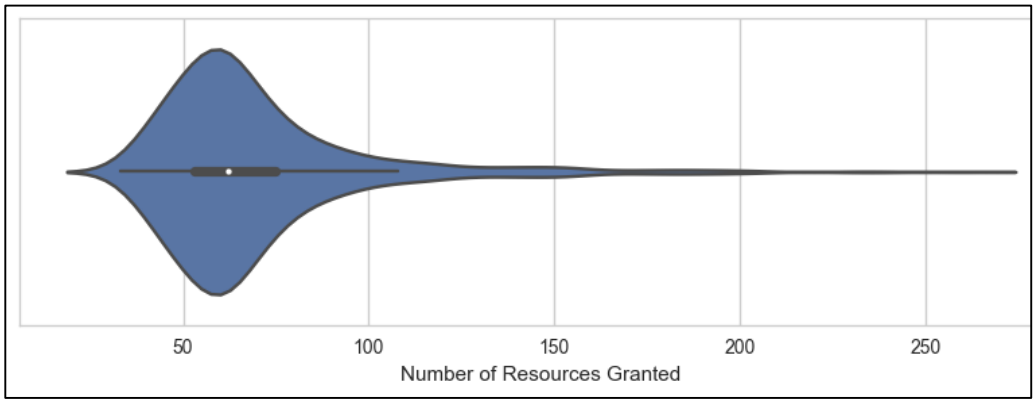


Figure 2. Outlier Score Detailed Methodology

Results

To identify employees with excess access control, the number of resources granted to each employee was analyzed as shown by Figure 3. While all 956 employees can access at least 25 resources, the dataset included employees granted up to 260 resource permissions with a median of only 68. Similarly, manager occurrences in the dataset were also analyzed to possible indicate whether certain managers are overly lenient when granting access to underlying employees. While possible to flag outlying employees using value counts, this does not indicate which individual employee resource pairings are the outlying occurrences.



Resources Granted Cutoff	Frequency of Employees
25	956
50	770
100	95
150	27
200	5

Figure 3. Distribution of Number of Resources Granted to Each Employee

To obtain more granular results, feature pairings were classified using the PyOD toolkit and aggregated into an Outlier Score out of 63 (see Figure 2). As shown by Figure 4, over 60% of the 66.3 records were not flagged by a single classifier (deemed ordinary entries). Further analysis dictated that an Outlier Score of 16 or higher is likely to reduce the presence of falsely labelled outliers. Using 16 as the threshold encompasses around 5% of the dataset, with records linked to 144 employees. The maximum Outlier Score assigned to a record was 57 out of 63.

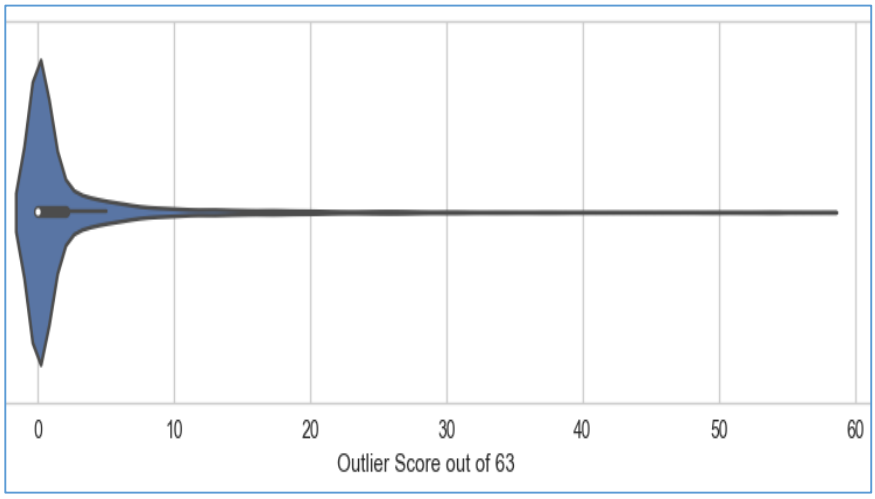


Figure 4. Distribution of Outlier Score for Each Record

Outlier Score Cutoff	Outlier Count	Inlier Count	Outlier Split (%)
0	25820	40468	38.95
5	9546	56742	14.4
10	5459	60829	8.24
15	3684	62604	5.56
20	2491	63797	3.76
25	1848	64440	2.79
30	1300	64988	1.96
40	718	65570	1.08
50	312	65976	0.47
55	21	66267	0.03

For the prediction of access control using recommendation systems, the true approval rate of the testing set was 94%. The average probability of approval on the testing set for Random Forest and Extra Trees yielded 92.3% and 91.0% respectively, while Gradient Boosting was overly sensitive at 97.5%. However, a probability threshold for each classifier can be chosen in order to match the true 94% approval rate.

Conclusions and Recommendations

The conclusion of this project is that 5% of the existent access control mappings analyzed are determined actionable outliers. It is recommended that the methodology and results produced in this project be used to aid an internal access control audit. Identified anomalies should be reviewed in order of descending Outlier Score, in order to firstly verify the most severe outliers while also minimizing the presence of false outliers. If a large portion of high Outlier Scores are deemed ordinary access control mappings, the remainder of the 5% is unlikely to be classified as anomalous and access control review would be unproductive.

The removal of unnecessary employee resource mapping would maximize enterprise cybersecurity. Minimal access granting reduces vulnerabilities and data loss caused by employee miscues, or situations in which a workstation is subject to malware or theft. Data confidentiality is also improved, ensuring only employees with sufficient clearance can access confidential information about

customers, other employees or the enterprise. Furthermore, it is possible that the dataset records could be supplemented with a resource sensitivity level or any indication of the level of importance the resource has in terms of confidentiality and operational value. If not available, the number of employees a resource has been granted to may be an alternative, as rarely granted resources likely contain more sensitive data. While a large number of excess permissions may provide resources deemed harmless, it is highly important to audit and verify access control for network resources containing highly sensitive information.

Recommendation systems can also be applied to datasets possessing existent access control mapping and user attributes. Resource request approval prediction was classified with high accuracy, proving recommendation systems as a viable method for future access control mapping. The benefits include providing a dynamic method to grant and eliminate resource access control, helping prevent employee downtime and reduce network administration costs. For future access control administration, recommendation systems could be used to assign an approval probability to all access control requests, allowing both IT administrators and managers to easier assign permissions to employees. Furthermore, all new permissions granted could be analyzed using the anomaly detection algorithm, to identify outlying access instantly and during access control audits.

References

1. C. O. a. R. J. Loomis, "Economic Analysis of Role-Based Access Control: Final Report," CSRC, 19 December 2010. [Online]. Available: https://csrc.nist.gov/publications/detail/white-paper/2010/12/19/economic-analysis-of-rbac-final-report/final.
2. Wang, H., Guo, X., Fan, Y. and Bi, J. (2014). Extended Access Control and Recommendation Methods for Enterprise Knowledge Management System. IERI Procedia, 10, pp.224-230.
3. C. C. Aggarwal, Recommender Systems The Textbook. Cham: Springer International Publishing, 2018.
4. Zhao, Y., Nasrullah, Z. and Li, Z., 2019. PyOD: A Python Toolbox for Scalable Outlier Detection. arXiv preprint arXiv:1901.01588.
5. “Amazon.com - Employee Access Challenge.” Kaggle, www.kaggle.com/c/amazon-employee-access-challenge/data.