Role Mining Data for RBAC Programming Challenge

Yanhong A. Liu* Scott D. Stoller*

RBAC and role mining. Role-Based Access Control (RBAC) is a security policy framework for controlling user access to resources based on roles [4, 14]. It is extremely important for reducing the cost of policy administration, especially in large organizations. An ANSI standard was created for RBAC [5, 1].

Role mining is the problem of analyzing given user-permission relation, to determine a set of roles with user-role relation and permission-role relation, such that the RBAC policy grants the same permissions as given [3].

RBAC challenge with role mining. An RBAC programming challenge was created for LPOP 2018 [18, Appendix B, page 14-17]. The problems include updates—for actions and transactions—and queries—for checking, analysis, optimization, and planning—in the presence of constraints, naturally organized into a set of components for ease of use by the applications.

Role mining is described as two of the four functionalities in the last component—administrative RBAC. The first one is to minimize the combined size of user-role relation and permission-role relation. The second one allows the use of a role hierarchy to inherit permissions and is to minimize the combined size of user-role relation, permission-role relation, and role hierarchy. Note that both functionalities take some given user-role and permission-role relations instead of a single user-permission relation, and ask that the join of the two resulting relations (plus hierarchy) equals the join of the two given relations. However, a user-role relation and a permission-role relation can be built from a user-permission relation trivially by either making a role for each user, or making a role for each permission.

Data for role mining. Realistic test data for RBAC is challenging to find. We describe the best data we know that is available for role mining. Note that other components of RBAC are much less challenging than administrative RBAC, and different relations can be easily generated to show scalable solutions, e.g., in [8, 7].

Ene et al. [3] collected several real-world access control lists to use for evaluation of their role mining algorithm. They made this dataset available to other researchers, and it has become the most widely used dataset for this purpose. Papers that use it include [12, 11, 13, 16, 9, 17, 19, 6, 10, 15, 2]. Figure 1 shows the sizes of the policies. Ene et al. describe the policies included in the dataset as follows [3]:

*Authors' contact: Computer Science Department, Stony Brook University, Stony Brook, New York. Email: {liu,stoller}@cs.stonybrook.edu

Dataset	$\|U\|$	$\|P\|$	$\|UP\|$
firewall-1	365	709	31951
firewall-2	325	590	36428
americas-small	3477	1587	105205
americas-large	3485	10127	185294
apj	2044	1146	6841
emea	35	3046	7220
healthcare	46	46	1486
domino	79	231	730
customer	10021	277	45427

Figure 1: Policy sizes. $|U|$, $|P|$ and $|UP|$ are the numbers of users, permissions, and user-permission assignments, respectively.

We applied our algorithms to some network access control rules used in Hewlett Packard (HP) to manage external business partner connectivity. We obtained two user profiles (americas small and americas large) from Cisco firewalls that authenticate external users and provide them with limited HP network access based on their user profiles. We also got similar, smaller datasets apj and emea. The healthcare dataset was obtained from the US Veteran's Administration, which has developed a comprehensive list of the healthcare permissions that may be assigned to licensed or certified providers. The domino graph is from a set of user and access profiles for a Lotus Domino server. customer is an access control graph obtained from the IT department of an HP customer.

This data is available at `http://lpop.cs.stonybrook.edu/rbac-challenge`.

References

[1] ANSI INCITS. Role-Based Access Control. ANSI INCITS 359-2004, American National Standards Institute, International Committee for Information Technology Standards, Feb. 2004.

[2] C. Cotrini, L. Corinzia, T. Weghorn, and D. Basin. The next 700 policy miners: A universal method for building policy miners. In *Proceedings of the 2019 ACM Conference on Computer and Communications Security (CCS)*, pages 95–112, 2019.

[3] A. Ene, W. G. Horne, N. Milosavljevic, P. Rao, R. Schreiber, and R. E. Tarjan. Fast exact and heuristic methods for role minimization problems. In *Proceedings of the 13th ACM Symposium on Access Control Models and Technologies*, pages 1–10, 2008.

[4] D. Ferraiolo and R. Kuhn. Role-based access control. In *Proceedings of the 15th NIST-NSA National Computer Security Conference*, pages 554–563, 1992.

[5] D. F. Ferraiolo, R. Sandhu, S. Gavrila, D. R. Kuhn, and R. Chandramouli. Proposed NIST standard for role-based access control. *ACM Transactions on Information and Systems Security*, 4(3):224–274, 2001.

[6] M. Frank, J. M. Buhmann, and D. A. Basin. Role mining with probabilistic models. *ACM Transactions on Information and System Security*, 15(4):1–28, 2013.

[7] M. Gorbovitski, Y. A. Liu, S. D. Stoller, and T. Rothamel. Composing transformations for instrumentation and optimization. In *Proceedings of the ACM SIGPLAN 2012 Workshop on Partial Evaluation and Program Manipulation*, pages 53–62, 2012.

[8] Y. A. Liu, C. Wang, M. Gorbovitski, T. Rothamel, Y. Cheng, Y. Zhao, and J. Zhang. Core role-based access control: Efficient implementations by transformations. In *Proceedings of the ACM SIGPLAN 2006 Workshop on Partial Evaluation and Program Manipulation*, pages 112–120, 2006.

[9] H. Lu, J. Vaidya, V. Atluri, and Y. Hong. Constraint-aware role mining via extended Boolean matrix decomposition. *IEEE Trans. Dependable Sec. Comput.*, 9(5):655–669, 2012.

[10] B. Mitra, S. Sural, V. Atluri, and J. Vaidya. The generalized temporal role mining problem. *Journal of Computer Security*, 23(1):31–58, 2015.

[11] I. Molloy, H. Chen, T. Li, Q. Wang, N. Li, E. Bertino, S. B. Calo, and J. Lobo. Mining roles with multiple objectives. *ACM Transactions on Information and Systems Security*, 13(4):36:1–36:35, 2010.

[12] I. Molloy, N. Li, T. Li, Z. Mao, Q. Wang, and J. Lobo. Evaluating role mining algorithms. In *Proceedings of the 14th ACM Symposium on Access Control Models and Technologies (SACMAT)*, pages 95–104, 2009.

[13] I. Molloy, N. Li, Y. A. Qi, J. Lobo, and L. Dickens. Mining roles with noisy data. In *Proceedings of the 15th ACM Symposium on Access Control Models and Technologies (SACMAT)*, pages 45–54, 2010.

[14] R. Sandhu, E. Coyne, H. Feinstein, and C. Youman. Role-based access control models. *IEEE Computer*, 29(2):38–47, 1996.

[15] S. D. Stoller and T. Bui. Mining hierarchical temporal roles with multiple metrics. *Journal of Computer Security*, 26(1):121–142, 2018.

[16] J. Vaidya, V. Atluri, Q. Guo, and H. Lu. Role mining in the presence of noise. In *Proceedings of the 24th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy (DBSec)*, volume 6166 of *Lecture Notes in Computer Science*, pages 97–112. Springer, 2010.

[17] N. V. Verde, J. Vaidya, V. Atluri, and A. Colantonio. Role engineering: From theory to practice. In *Proceedings of the Second ACM Conference on Data and Application Security and Privacy (CODASPY)*, pages 181–192, 2012.

[18] D. S. Warren and Y. A. Liu. LPOP: Challenges and Advances in Logic and Practice of Programming. *Computing Research Repository*, arXiv:2008.07901 [cs.PL], Aug. 2020.

[19] Z. Xu and S. D. Stoller. Algorithms for mining meaningful roles. In *Proceedings of the 17th ACM Symposium on Access Control Models and Technologies (SACMAT)*, pages 57–66, 2012.