CERIAS

the center for education and research in information assurance and security

Information About the Author

# Role Mining and Policy Misconfigurations

Download our SACMAT 2010 Paper

Ian Molloy, Ninghui Li, Jorge Lobo, Yuan (Alan) Qi, and Luke Dickens

## Motivation

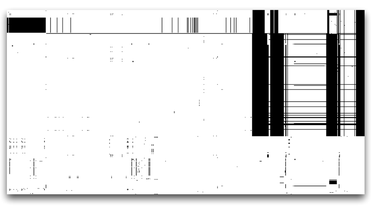
- RBAC is a popular, attractive model
- Constructing an RBAC state is difficult
- Role mining leverages existing data, reducing the role engineering costs
- How do we know the data is correct?
- Can we predict an unknown decision?
- What if we have incomplete data?

## Role Mining

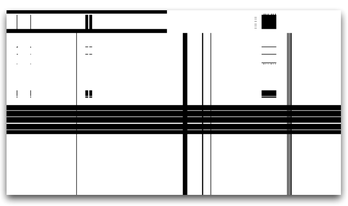
- Convert an existing access control state to RBAC
- Minimize Structural Complexity [SACMAT 2008]
- *HierarchicalMiner*, Based on Formal Concepts
- Compact results but assumes clean data

$$\text{wsc}(\gamma, W) = w_r * |R| + w_u * |UA| + w_p * |PA| + w_h * |t\_reduce(RH)| + w_d * |DUPA|$$

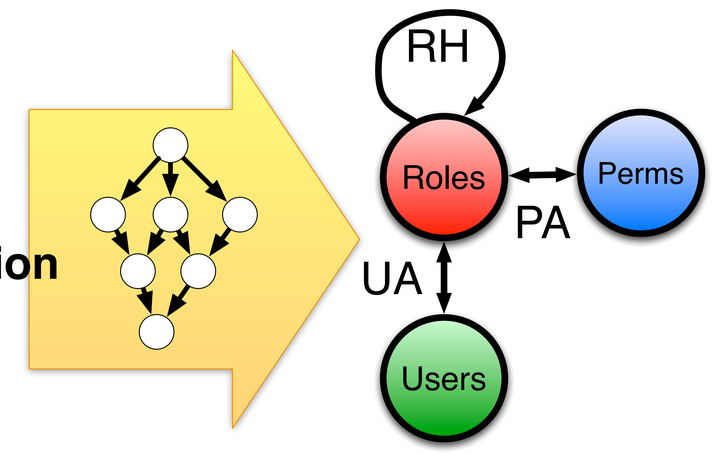
## User-Permission Relation



## User-Attribute Relation



## RBAC

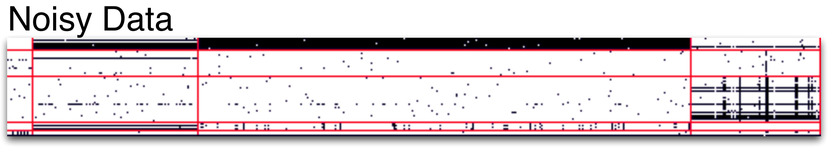



## Noisy Data

- Real data is noisy, and may contain errors
- Permissions aren't revoked when job functions change
- Separate Tasks: First clean data, then convert to RBAC

Clean Data

Noisy Data



## Approach

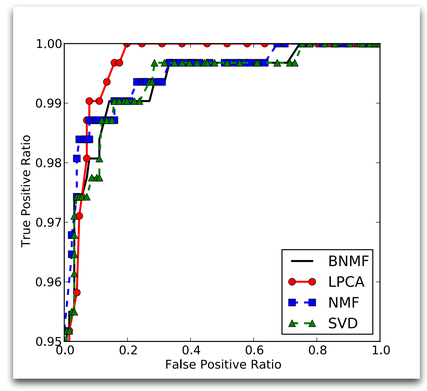
- Use known access control decisions
- Leverage user-permission and user-attributes
- Matrix Decomposition and Relation Learning
- "Netflix Prize of Access Control"
- e.g., SVD, NMF, *LPCA*

$$UP = f(UA \times PA) \quad f(\theta) = (1 + e^{-\theta})^{-1}$$

$$D(X \parallel \hat{X} = f(AB^T)) = \sum_{ij} \left( X_{ij} \log \frac{X_{ij}}{\hat{X}_{ij}} + (1 - X_{ij}) \log \frac{1 - X_{ij}}{1 - \hat{X}_{ij}} \right)$$

## Prediction and Reconstruction

- Decomposition is inexact, correcting errors
- Missing values are approximated
- Useful for AC exception models
- Mining from incomplete data (logs)
- Probabilistic models estimate *"risk"*

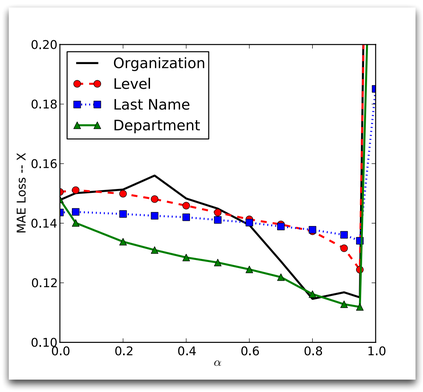



## Collective Matrix Factorization

- Leverages attributes to improve accuracy
- Cluster users based on permissions *and* attributes
- Select attributes by entropy and cardinality

$$h(p_j \mid A) = -\sum_{a \in A} P[A = a] \sum_{s \in \{0,1\}} P[p_j = s \mid A = a] * \log_2 P[p_j = s \mid A = s]$$

| Attribute | Order | Total Entropy | Pred. Improv. |
|---|---|---|---|
| ***Last Name*** | ***2224*** | ***17.86*** | ***6.6%*** |
| Manager | 298 | 34.25 | 17.5% |
| Department | 192 | 41.83 | 24.4% |
| Title | 527 | 42.74 | 15.2% |
| Location | 53 | 62.20 | 17.6% |
| ***Organization*** | ***12*** | ***82.11*** | ***22.5%*** |
| ***Level*** | ***17*** | ***102.29*** | ***17.3%*** |
| Contractor | 2 | 104.15 | 12.0% |
| | * | 107.34 | |




## References

- "Mining Roles with Noisy Data" SACMAT 2010
- "Approximate Inference for Nonparametric Bayesian Matrix Factorization" AISTATS 2010
- "Mining Roles with Semantic Meaning" SACMAT 2008

PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center