SOUTHERN ILLINOIS UNIVERSITY Carbondale

A Comparison of Database Insider Attack Monitoring Algorithms using Page Ranks with In/Out Links

Alan Y. Park. Advisor: Sam Chung
School of Information Systems & Applied Technologies, College of Applied Sciences and Arts, Southern Illinois University Carbondale

Summary

The purpose of this research is to compare two database insider attack monitoring approaches that use page rank algorithms: PageRank and Weighted PageRank. By calculating the weight of queries, we predict the users' pattern in the database system.

Keywords

- Insider Attack Monitoring,
- Web Mining,
- PageRank Algorithm,
- Weighted PageRank Algorithm,
- Markov Chain

Motivation

Nowadays, an average attacker takes less than ten seconds to hack information systems. Specially, the insider attacks that any malicious attack on the database systems performed by an entrusted group of people having authorized access are more dangerous than the outsider attacks since it has more entry points of malicious attacks than ones from outsider attacks.

Problem Statement

Which approach is better to monitor insider attacks from users who can access database systems with legitimate access controls between using PageRank and Weighted PageRank algorithms?

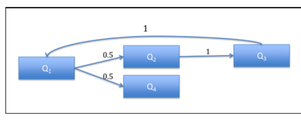
Approach

By using both PR and WPR algorithm, we calculate the weight of queries and distributes rank scores based on the popularity of the queries. Our case study shows that two algorithms brought different outcomes to our query rank based database insider attack monitoring. From the value of the weight of queries, we are able to select suspicious queries that will not occur regularly on the database system. Based on both the PR and the WPR algorithm, we are able to extract monitoring information from users' past behavior on the database system.

PageRank

- Defining the Query Transition Probability Matrix K as

$$K = \{k_{ij} \mid i,j \in I, n = |I|,$$
k_{ij} = 0 if there is no link between Q_i and Q_j
k_{ij} = 1/the number of out links of Q_i
$\sum_{j=1}^{n} k_{ij} = 0 | 1\ for\ the\ fixed\ i\}.$



$$K = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- Building the Query Rank Probability Matrix M as

$$M = \{m_{ij} \mid i,j \in I, n = |I|,$$
$$\sum_{j=1}^{n} m_{ij} = 1\ for\ the\ fixed\ i\}$$
$$r_i = \sum_{j=1}^{n} k_{ij},\ \delta = (1-p)/n$$
$$m_{ij} = \begin{cases} pk_{ij} + \delta & : r_i = 1 \\ \frac{1}{n} & : r_i = 0 \end{cases}$$

$$M = \begin{bmatrix} 0.0375 & 0.4625 & 0.0375 & 0.4625 \\ 0.0375 & 0.0375 & 0.8875 & 0.0375 \\ 0.8875 & 0.0375 & 0.0375 & 0.0375 \\ 0.2500 & 0.2500 & 0.2500 & 0.2500 \end{bmatrix}$$

- We assume that all queries are ranked equally at the beginning.

$$\lambda_0 = [0.2500 \quad 0.2500 \quad 0.2500 \quad 0.2500]$$
$$\lambda_1 = \lambda_0 M^1 = [0.3031 \quad 0.1969 \quad 0.3031 \quad 0.1969]$$
$$\lambda_2 = \lambda_1 M^1 = \lambda_0 M^1 M^1 = \lambda_0 M^2$$
$$\lambda = \lambda_0 M^{17} = [0.3078 \quad 0.2138 \quad 0.2646 \quad 0.2138]$$

Weighted PageRank

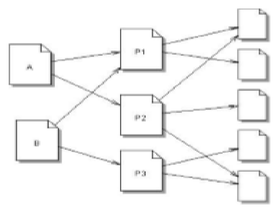
- The weight of link(v,u) calculated based on the number of in-links.

$$W^{in}_{(v,u)} = \frac{I_u}{\sum_{p \in R(v)} I_p}$$

- The weight of link(v,u) calculated based on the number of out-links.

$$W^{out}_{(v,u)} = \frac{O_u}{\sum_{p \in R(v)} O_p}$$

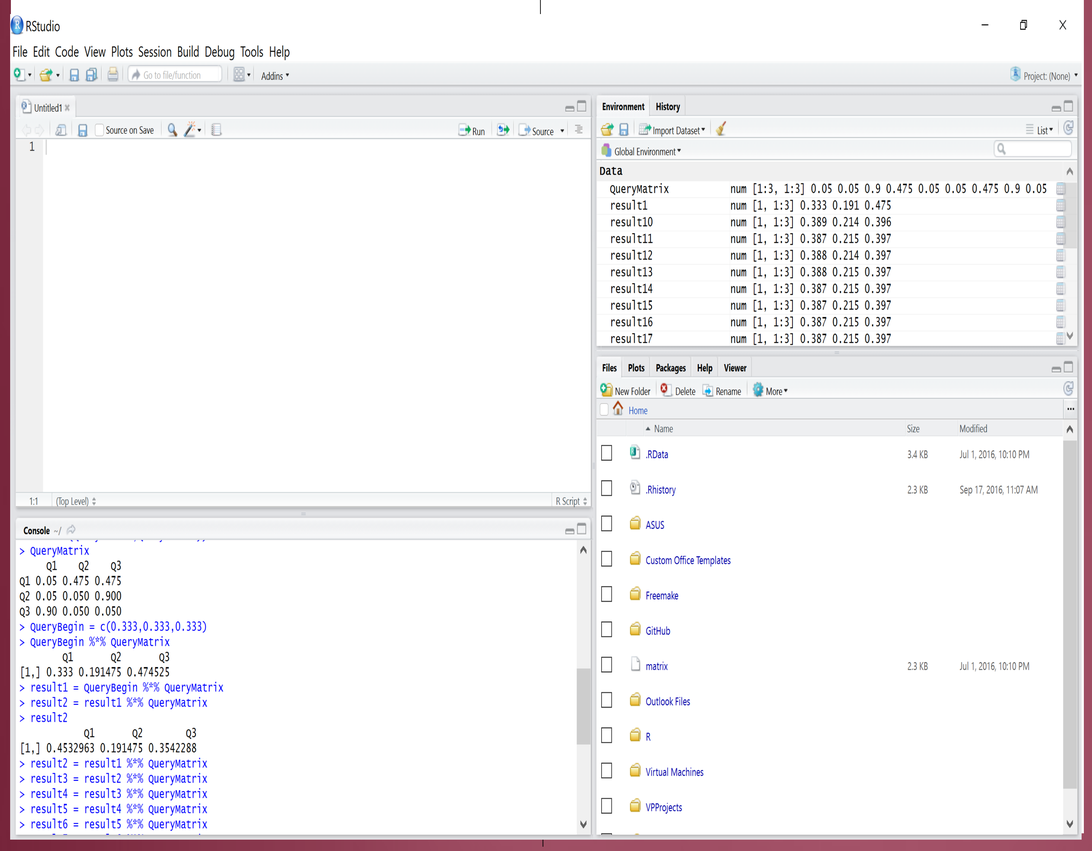
- Page A has two reference pages p1 and p2. The inlinks of p1 is 2 and p2 is 1, outlinks of p1 is 2 and p2 is 3



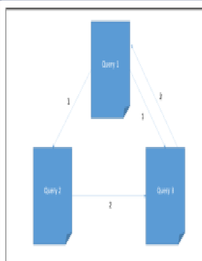
- The original PageRank formula is modified as

$$PR(u) = (1-d) + d \sum_{v \in B(u)} PR(v) W^{in}_{(v,u)} W^{out}_{(v,u)}$$

Calculation using RStudio

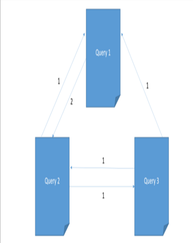


Comparison Criteria



	PageRank	Weighted PageRank
Query 1	0.387	0.222
Query 2	0.215	0.337
Query 3	0.397	0.441

Case 1



	PageRank	Weighted PageRank
Query 1	0.333	0.433
Query 2	0.433	0.232
Query 3	0.234	0.183

Case 2

Evaluation and Conclusion

Consequently, Since the PageRank algorithm only takes the number of out-links as a factor while the WPR algorithm takes number of both in-links and out-links, we have the different value of the weight of queries.

Future Works

- Applying both algorithms to the database system.
- Visualization of the query weight by using Rstuido.

Reference

- **Cheolmin Sky Moon, Sam Chung, Barbara Endicott-Popovsky, Multiple-Criteria Query Statement Probabilities Based Database Insider Attack Monitoring System, University of Washington Tacoma MS Capstone, 2014.**
- **Larry Page, and Sergey Brin, Rajeev Motwanit, Terry Winograd, "The PageRank Citation Ranking: Bring Order to the Web", Technical report in Standford University, 1998.**
- **Wenpu Xing and Ghorbani Ali, "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR '04), IEEE, 2004.**

SIU CARBONDALE