ESTUDO SOBRE A APLICAÇÃO DE REGRAS DE ASSOCIAÇÃO PARA SEGURANÇA E AUDITORIA DE BANCOS DE DADOS

LUCAS BAYER DE ARAUJO[1]; ROGÉRIO DA COSTA ALBANDES[1]; ANA MARILZA PERNAS[1]

[1]Universidade Federal de Pelotas – {lbdaraujo; rcalbandes; marilza} @inf.ufpel.edu.br

1. INTRODUÇÃO

Bancos de dados (BDs) possibilitam gerenciar informações de forma eficiente, sendo empregados em áreas como saúde, finanças e governo. Frequentemente guardam conteúdos sensíveis, o que torna indispensável adotar medidas de segurança para protegê-los contra violações de privacidade, que podem gerar prejuízos financeiros, jurídicos e de reputação (OMOTUNDE; AHMED, 2023).

A relevância da segurança em BDs foi reforçada pela Lei Geral de Proteção de Dados (LGPD), em vigor desde setembro de 2020, que estabelece princípios de proteção às informações pessoais no Brasil. A lei busca garantir direitos de liberdade e privacidade, regulando o tratamento de dados por pessoas físicas ou jurídicas, públicas ou privadas, sob fiscalização da Autoridade Nacional de Proteção de Dados (ANPD), que pode aplicar sanções em caso de descumprimento (BRASIL, 2018).

A defesa de sistemas de armazenamento baseia-se nos princípios de confidencialidade, integridade e disponibilidade, protegendo-os de ameaças como injeção SQL, softwares maliciosos e ataques de negação de serviço (DoS) . Para isso, emprega mecanismos como autenticação, controle de acesso, criptografia, auditoria e sistemas de detecção de intrusão (OMOTUNDE; AHMED, 2023). Nesse contexto, a auditoria em BDs busca garantir conformidade com regras definidas, identificando anomalias através de análise de logs de transação (ALBANDES, 2024).

Contudo, a auditoria enfrenta desafios, especialmente em cenários de Big Data, em que é necessário lidar com grandes volumes de registros e identificar padrões implícitos. A mineração de regras de associação surge como alternativa, ao possibilitar a detecção eficiente de padrões nos dados, que podem ser empregados na descoberta de abusos de privilégio, tentativas de invasão e fragilidades em políticas de segurança (CHENG; XU; GONG, 2016).

Assim, o objetivo geral deste trabalho é comparar algoritmos de regras de associação, como Apriori (AGRAWAL; SRIKANT, 1994), FP-Growth (HAN; PEI; YIN, 2000) e Eclat (ZAKI, 2002), no contexto da auditoria de BDs voltada à segurança. Busca-se avaliar aspectos como complexidade, tempo de execução e escalabilidade, a fim de identificar quais algoritmos são mais eficazes para detectar padrões suspeitos. Para tanto, será conduzido um estudo conceitual e experimental, contemplando implementação, testes e comparação dos métodos, de modo a avaliar sua eficiência em cenários práticos.

2. METODOLOGIA

Inicialmente, realizou-se uma revisão bibliográfica sobre auditoria e segurança em BDs, visando identificar requisitos para um ambiente seguro e compreender o papel dos mecanismos de monitoramento na proteção das

informações. A partir disso, foram estudados métodos de verificação de integridade, como sistemas de auditoria e detecção de intrusão.

Na sequência, a pesquisa foi direcionada para o estudo de regras de associação e dos algoritmos empregados em sua obtenção. Essa análise teve como finalidade explorar o potencial do uso de mineração de padrões como ferramenta para identificação de comportamentos anômalos em sistemas de armazenamento, contribuindo para a detecção de falhas de confiança.

Por fim, direcionou-se a análise para trabalhos que exploram a aplicação da mineração de regras de associação no contexto de auditoria de bancos de dados, permitindo observar como essa técnica tem sido utilizada em cenários práticos para apoiar a detecção de anomalias e fortalecer a segurança das informações.

Os passos seguintes buscam implementar e testar os algoritmos estudados em um cenário prático, visando avaliar sua eficiência na auditoria de BDs.

3. RESULTADOS E DISCUSSÃO

Até o momento foi concluída a etapa de revisão bibliográfica, na qual foram estudados os seguintes temas: regras de associação; algoritmos de regras de associação, como Apriori, FP-Growth e ECLAT; e aplicação de regras de associação na auditoria e segurança em BDs.

Regras de associação são representadas como x → y, em que x e y são itens ou conjuntos de itens, e indicam a tendência de *y aparecer junto à x* (AGRAWAL; SRIKANT, 1994). A análise dessas regras é baseada em métricas como suporte, que mede a frequência de ocorrência de itens ou regras em relação ao total de transações, e confiança, que avalia a relevância da associação. O uso conjunto dessas medidas, associado a limiares mínimos de suporte (*minsupp*) e confiança (*minconf*), possibilita aprendizado de regras e a identificação de padrões nos dados (ZHANG; ZHANG, 2002).

A mineração das regras de associação é dividida em duas etapas: a identificação dos *itemsets* frequentes (grandes), cujo o suporte é maior que *minsupp*; e a geração das regras relevantes, formadas pelos *itemsets* grandes, que atendem ao *minconf* (KOTSIANTIS; KANELLOPOULOS, 2006). Para isso, definem-se os valores de *minsupp* e *minconf*. Em seguida, calcula-se a frequência (suporte) de todos os *itemsets*, começando pelos unitários (1-*itemsets*) e avançando para os 2-*itemsets*, que são combinações dos anteriores. Esse processo ocorre de forma iterativa, progredindo para os k-*itemsets* até que não haja mais conjuntos frequentes. Por fim, a partir dos *itemsets* obtidos, extraem-se as regras de associação válidas, cuja confiança é maior ou igual a *minconf*.

Para realizar esse processo, foram propostos diversos algoritmos como o Apriori, que tenta otimizar o problema de geração de itens candidatos, o FP-Growth, que utiliza uma estrutura de árvore para evitar esse problema completamente e o Eclat, que adota uma estratégia de intersecção de conjuntos e torna mais eficiente a busca por *itemsets* grandes (SRINADH, 2022).

O Apriori apresenta uma modificação na etapa de localização de *itemsets* frequentes, com o objetivo de reduzir o número de candidatos testados quanto ao suporte mínimo. Para isso, utiliza a propriedade anti-monotônica, segundo a qual todo subconjunto de um *itemset* frequente também será frequente, e, inversamente, k-*itemsets* devem ser formados apenas por itens frequentes. Assim, após identificar os 1-*itemsets* grandes, o algoritmo combina apenas conjuntos aprovados nas etapas anteriores, diminuindo o total de candidatos. Esse procedimento é denominado *Prune* (Poda)(AGRAWAL; SRIKANT, 1994).

Já o Frequent Pattern Growth (FP-Growth) busca evitar os altos custos da explosão combinatória na geração de candidatos, e utiliza a Frequent Pattern Tree (FP-Tree), uma estrutura que guarda transações e frequências de forma compacta. Ela é construída a partir dos 1-*itemsets* frequentes e organizada em ramos que são unidos quando compartilham prefixos, reduzindo nós e garantindo crescimento linear, em relação ao exponencial dos algoritmos baseados em candidatos, e economia de tempo e memória (HAN; PEI; YIN, 2000).

Além disso, o algoritmo possui um mecanismo próprio de aprendizado de regras, o Pattern Fragment Growth, utiliza uma tabela de cabeçalho que referencia as aparições de cada item na árvore. Essa tabela é percorrida dos itens menos aos mais frequentes, identificando relações relevantes com base no *minsupp* (OZAWA et al., 2020).

Por fim, o Equivalence Class Transformation (ECLAT) busca evitar o alto consumo de memória na geração de conjuntos candidatos por meio de sua principal característica, o uso de um banco de dados vertical, a *tid-list*. Essa tabela associa cada k-*itemset* às transações em que aparece, permitindo obter *itemsets* frequentes por meio da intersecção dessas listas e reduzindo a necessidade de múltiplas varreduras no banco (ZAKI, 2002). O processo inicia com a criação da *tid-list* para todos os 1-*itemsets*, em que cada item é vinculado aos identificadores das transações em que ocorre. Assim, o suporte é obtido pela contagem desses identificadores. Em seguida, a *tid-list* dos 2-*itemsets* é formada pela intersecção das listas de seus itens. Esse procedimento é repetido até que não se encontrem mais k-*itemsets* frequentes (SRINADH, 2022).

A aplicação das regras de associação na auditoria de bancos de dados é dividida em três partes. A primeira é o agente de auditoria, que coleta os registros (*logs*) de operações realizadas, como inserções, exclusões, alterações e leituras, os usuários envolvidos e os aspectos temporais. Esses dados são obtidos por meio de *triggers* que fazem inserções automáticas na tabela de auditoria. Em seguida, o módulo de análise de dados aplica técnicas de mineração sobre os *logs* coletados, gerando regras de associação que permitem identificar padrões nos dados armazenados, além de desvios, como acessos em horários incomuns ou falhas repetidas de *login*. Por fim, o banco de padrões guarda os resultados obtidos e monitora em tempo real, emitindo alertas sempre que atividades divergirem das tendências estabelecidas (CHENG; XU; GONG, 2016).

Na prática, ALBANDES (2024) aplicou as regras de associação em um banco de dados de saúde, extraindo padrões de acesso de médicos e pacientes. Foram identificadas regras como {tbPerson, manhã} → {dia útil}, com 80% de confiança, revelando que acessos à tabela de pessoas no período da manhã ocorriam majoritariamente em dias de semana. Durante os testes, três alertas de anomalias foram registrados, evidenciando a aplicabilidade do método.

De forma semelhante, WU; HUANG (2009) exploraram regras de associação para detectar invasões e comportamentos suspeitos, como ataques de personificação ou tentativas de abuso por usuários legítimos. Os resultados mostraram taxas de detecção de 70% para *logins* com credenciais comprometidas e de até 90% em casos de tentativas de contornar controles de acesso. Esses achados reforçam o potencial das regras de associação para fortalecer a segurança de bancos de dados.

4. CONCLUSÕES

O presente trabalho, ainda em desenvolvimento, iniciou-se com um estudo bibliográfico sobre regras de associação, os algoritmos Apriori, FP-Growth e Eclat, e sua aplicação em auditoria de bancos de dados. Até o momento, os algoritmos foram descritos e analisados sob a perspectiva teórica, destacando diferenças conceituais e práticas, além da descrição do uso de regras de associação no processo de auditoria. A próxima etapa consistirá na implementação e avaliação desses algoritmos, comparando seu desempenho e aplicabilidade.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules in Large Databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, San Francisco, CA, USA. **Proceedings. . .** Morgan Kaufmann Publishers Inc., 1994. p.487–499. (VLDB '94).

ALBANDES, R. d. C. **Abordagem IoT PD-RPM**: Promovendo Aderência à Diálise Peritoneal Considerando o Cenário da Internet das Coisas. 2024. Tese de Doutorado — Universidade Federal de Pelotas, Pelotas.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. 2018. Acesso em: 12 ago. 2025, Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <https://www.planalto.gov.br/ccivil03/ato2015−2018/2018/lei/L13709compilado.htm > .

CHENG, M.; XU, K.; GONG, X. Research on audit log association rule mining based on improved Apriori algorithm. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA ANALYSIS (ICBDA), 2016., 2016. China. **Anais...**, 2016. p.1–7.

HAN, J.; PEI, J.; YIN, Y. Mining frequent patterns without candidate generation. **SIGMOD Rec.**, New York, NY, USA, v.29, n.2, p.1–12, May 2000.

KOTSIANTIS, S.; KANELLOPOULOS, D. Association rules mining: A recent overview. **GESTS International Transactions on Computer Science and Engineering**, v.32, n.1, p.71–82, 2006.

OMOTUNDE, H.; AHMED, M. A comprehensive review of security measures in database systems: Assessing authentication, access control, and beyond. **Mesopotamian Journal of CyberSecurity**, v.2023, p.115–133, 2023.

OZAWA, S.; BAN, T.; HASHIMOTO, N.; NAKAZATO, J.; SHIMAMURA, J. A study of IoT malware activities using association rule learning for darknet sensor data. **International Journal of Information Security**, v.19, p.83–92, 2020.

SRINADH, V. Evaluation of Apriori, FP growth and Eclat association rule mining algorithms. **International journal of health sciences**, n.II, p.7475–7485, 2022.

WU, G.; HUANG, Y. Design of a new Intrusion detection system based on database. In: INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING SYSTEMS, 2009., 2009, Singapore. **Anais...** IEEE, 2009. p.814–817.

ZAKI, M. J. Scalable algorithms for association mining. **IEEE transactions on knowledge and data engineering**, v.12, n.3, p.372–390, 2002.

ZHANG, C.; ZHANG, S. Association rule mining: models and algorithms. Springer, 2002.