PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Bacharelado em Ciência da Computação

Lara Brígida Rezende Souza
Matheus Moreira Sorrentino
Raul da Cruz Fonseca
Victor Cabral de Souza Oliveira

DETECÇÕES DE ANOMALIAS EM ARQUIVOS CRIPTOGRAFADOS

Belo Horizonte

2023

Lara Brígida Rezende Souza
Matheus Moreira Sorrentino
Raul da Cruz Fonseca
Victor Cabral de Souza Oliveira

DETECÇÕES DE ANOMALIAS EM ARQUIVOS CRIPTOGRAFADOS

Projeto de Pesquisa apresentado na disciplina Trabalho Interdisciplinar III - Pesquisa Aplicada do curso de Ciência da Computação da Pontifícia Universidade Católica de Minas Gerais.

Belo Horizonte

2023

RESUMO

Podemos resumir que nesta pesquisa os autores irão abordar o tema de detecção de anomalias em arquivos criptografados, trazendo os problemas principais, os objetivos e a revisão bibliográfica como grandes enfoques.

A revisão bibliográfica contará com dez artigos relacionados ao tema e os autores trarão a correlação daquilo que está sendo dito nos artigos com o tema abordado por eles nesta pesquisa.

Palavras-chave: detecção; anomalias; arquivos criptografados; pesquisa; revisão bibliográfica; objetivos; problemas.

SUMÁRIO

1 INTRODUÇÃO

Escolhemos o tema de detecção de anomalias em arquivos criptografados, pois é um tema relacionado com a disciplina de Algoritmos e Estrutura de Dados III do terceiro período do curso de Ciência da Computação da Pontifícia Universidade Católica de Minas Gerais. Sendo essa uma matéria que todos os autores da pesquisa estão cursando.

Além do mais, é um tema deveras interessante, afinal, aborda como identificar erros em arquivos criptografados, o que leva a busca de algoritmos para executarem tal função.

Alguns problemas identificados para o tema da pesquisa são:

1. **Dificuldade na identificação de anomalias criptografadas -** Como identificar anomalias dentro de arquivos criptografados sem acessar seu conteúdo?

2. **Altas taxas de falso positivo ou negativo -** Como reduzir as taxas de falso positivo/negativo na detecção de anomalias em arquivos criptografados?

3. **Adaptação a mudanças nas estratégias de ataque -** Como garantir que os sistemas de detecção de anomalias possam se adaptar a novas técnicas de ataque que tentam contornar a criptografia?

4. **Dimensionamento e desempenho -** Como garantir que os algoritmos de detecção de anomalias possam lidar com grandes volumes de arquivos criptografados de forma eficiente e com baixa latência?

5. **Criptografia forte x Detecção de anomalias -** Como equilibrar a criptografia robusta para proteger os dados e ao mesmo tempo permitir a detecção eficaz de anomalias?

Sugerimos os seguintes problemas pelo fato de serem os mais relevantes quando se trata do tema de detecção de anomalias em arquivos criptografados. Afinal, todos os tópicos são demasiados importantes pelo fato de só ser possível executar os objetivos contornando os problemas citados.

Este trabalho está organizado da seguinte forma. A seção 1.1 representa os objetivos procurados na pesquisa. O capítulo 2 apresenta o referencial teórico usado neste trabalho.

1.1 Objetivos

O objetivo geral deste projeto é estudar a detecção de anomalias em arquivos criptografados através de alguns artigos de referência para ter um referencial de discussão sobre os problemas que podem surgir acerca do tema e gerar possíveis soluções.

1.1.1 Objetivos específicos

Os objetivos específicos deste projeto são:

1. **Desenvolvimento de algoritmos eficientes -** Criar algoritmos que possam detectar anomalias em arquivos criptografados sem exigir descriptografia completa.
2. **Aprimoramento da precisão -** Melhorar a precisão na detecção de anomalias para reduzir falsos positivos/negativos, aumentando a confiança nas detecções.
3. **Adoção de técnicas de aprendizado de máquina -** Utilizar técnicas de aprendizado de máquina para aprender padrões de comportamento normal e identificar desvios significativos.
4. **Melhoria da escabilidade -** Otimizar os algoritmos e a infraestrutura para garantir a detecção eficiente de anomalias em grandes volumes de arquivos criptografados.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta os artigos que os autores utilizaram como referência para realizar a pesquisa acerca do tema de detecção de anomalias em arquivos criptografados. Ao todo foram utilizados dez artigos que a seguir serão correlacionados com o tema abordado na pesquisa.

2.1 Primeiro Artigo

A detecção automática de anomalias na mineração de dados é altamente relevante para a detecção de comportamentos anômalos em arquivos criptografados. Assim como em séries temporais, identificar atividades suspeitas ou não autorizadas em arquivos criptografados é essencial para a segurança e a integridade dos dados. O uso de métodos avançados, como autoencoders, permite aprender padrões eficientes nos arquivos, distinguindo entre o comportamento normal e possíveis comportamentos anômalos. Essa abordagem é fundamental para fortalecer a proteção de arquivos criptografados e prevenir possíveis ataques ou violações de segurança.

[1] Unsupervised Anomaly Detection in Time Series Using LSTM-Based Autoencoders (PROVOTAR; LINDER; VERES, 2019)

2.2 Segundo Artigo

A análise de logs é crucial na detecção de anomalias em arquivos criptografados, revelando possíveis violações de segurança. Nossa abordagem inovadora, baseada em mineração de dados com Apache Hadoop, permite detectar novos tipos de violações automaticamente, mantendo baixas taxas de erro. Isso fortalece a segurança e integridade dos sistemas.

[2] Anomaly Detection from Log Files Using Data Mining Techniques (BREIER; BRANIŠOVÁ, 2015)

2.3 Terceiro Artigo

A complexidade crescente dos sistemas de computador torna a detecção manual de falhas inviável. A análise automatizada, usando modelos como autoencoders com unidades LSTM, é essencial. Propomos um modelo que gera pontuações de anomalia, indicando a raridade de eventos de log, sem depender de dados rotulados. Isso é relevante para a detecção de anomalias em arquivos criptografados, onde a análise automatizada é crucial dada a complexidade e o volume de atividades.

[3] Anomaly Detection from Log Files Using Unsupervised Deep Learning (BURSIC; CUCULO; D'AMELIO, 2020)

2.4 Quarto Artigo

Na segurança cibernética, a detecção de intrusões é vital para enfrentar ameaças crescentes. Propomos uma metodologia dinâmica para detecção de anomalias em arquivos de log, considerando a natureza dinâmica dos sistemas e suas interdependências. Essa abordagem gradativamente agrupa e analisa linhas de log em janelas de tempo, detectando anomalias relacionadas a frequências e alterações periódicas. Ao considerar cenários reais, nosso protótipo demonstrou eficiência na detecção de alterações dinâmicas nas linhas de log, com baixa taxa de falsos alarmes. Essa adaptabilidade sugere aplicabilidade relevante na detecção de anomalias em arquivos criptografados.

[4] Dynamic log file analysis: An unsupervised cluster evolution approach for anomaly detection (LANDAUER et al., 2018)

2.5 Quinto Artigo

Na busca por eficácia na detecção de anomalias em sistemas computacionais, os registros de log se destacam como fontes cruciais de informações. Duas abordagens predominam: detecção em lote, mais computacionalmente exigente, e detecção em streaming, proporcionando alertas imediatos. Este trabalho apresenta um framework não supervisionado que viabiliza a detecção de anomalias em tempo real, demonstrando sucesso ao identificar anomalias em arquivos de log hdfs, com uma pontuação F-1 de 83%. Esta proposta oferece potencial aplicabilidade na detecção de anomalias em arquivos criptografados.

[5] An unsupervised anomaly detection framework for detecting anomalies in real time through network system's log files analysis (LANDAUER et al., 2018)

2.6 Sexto Artigo

A detecção não supervisionada de anomalias em fluxos de dados é vital para diversas aplicações práticas, especialmente quando não é viável obter dados de treinamento rotulados. Este estudo propõe a adaptação do classificador Online Evolution Spiking Neural Network (OeSNN) para a detecção não supervisionada de anomalias (OeSNN-UAD). O OeSNN-UAD emprega um inovador método de detecção de anomalias em duas etapas, demonstrando superioridade em relação a outras abordagens não supervisionadas e semissupervisionadas na detecção de anomalias. Essa proposta possui relevância direta para a detecção de anomalias em arquivos criptografados, contribuindo para a segurança e integridade dos dados.

[6] Unsupervised Anomaly Detection in Stream Data with Online Evolving Spiking Neural Networks (MACIąG et al., 2021)

2.7 Sétimo Artigo

A detecção de anomalias é essencial para a segurança em sistemas computacionais, e os registros de log são uma fonte valiosa de dados para esse propósito. A abordagem proposta, Logsy, emprega métodos de classificação para aprender representações de log que distinguem dados normais de log de anomalias. Ao usar conjuntos de dados de log auxiliares, acessíveis via internet, Logsy melhora a representação dos dados normais e, consequentemente, a detecção de anomalias. Essa proposta inovadora tem implicações significativas na detecção de anomalias em arquivos criptografados, fortalecendo a segurança e confiabilidade dos sistemas.

[7] Self-Attentive Classification-Based Anomaly Detection in Unstructured Logs (NEDELKOSKI et al., 2020)

2.8 Oitavo Artigo

No contexto de detecção de anomalias em arquivos criptografados, a migração de dados para a nuvem e o uso de serviços de compartilhamento de arquivos apresentam desafios de segurança. Este artigo propõe um modelo criptográfico chamado criptografia hierárquica parcialmente ordenada (PHE), que permite rastreamento de traidores e revogação eficiente de chaves. Essa abordagem fortalece a segurança dos dados em ambientes de nuvem.

[8] PHE: An Efficient Traitor Tracing and Revocation for Encrypted File Syncing-and-Sharing in Cloud (ZHU et al., 2018)

2.9 Nono Artigo

No âmbito da detecção de anomalias em arquivos criptografados, abordagens de aprendizado de máquina são empregadas para desenvolver classificadores capazes de identificar arquivos empacotados ou criptografados. Esses métodos utilizam tanto recursos estáticos quanto dinâmicos dos arquivos para criar amostras de treinamento e, assim, aprimorar a capacidade de detecção de possíveis anomalias ou comportamentos indesejados nesses tipos de arquivos.

[9] Detection of packaged and encrypted PE files with the use of machine-learning algorithm (GEVORGYAN; ABRAMOV, 2018)

2.10 Décimo Artigo

No contexto da detecção de anomalias em arquivos criptografados, é essencial garantir a segurança dos terminais de dispositivos, especialmente diante da expansão contínua em tipos e aplicações. A detecção anormal de tráfego de usuários, proposta neste artigo, emprega um modelo de rede neural profunda para extrair características dos dados de comunicação e analisar o tráfego de forma a identificar comportamentos de alto risco nos terminais. Isso se relaciona ao tema, pois aborda a detecção de padrões anômalos em comunicações criptografadas, contribuindo para a segurança e eficiência no contexto empresarial.

[10] An anomaly detection method of encrypted traffic based on user behavior (XIN et al., 2021)

REFERÊNCIAS

BREIER, J.; BRANIŠOVÁ, J. Anomaly detection from log files using data mining techniques. In: KIM, K. J. (Ed.). Information Science and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015. p. 449–457. ISBN 978-3-662-46578-3.

BURSIC, S.; CUCULO, V.; D'AMELIO, A. Anomaly detection from log files using unsupervised deep learning. In: SEKERINSKI, E. et al. (Ed.). Formal Methods. FM 2019 International Workshops. Cham: Springer International Publishing, 2020. p. 200–207. ISBN 978-3-030-54994-7.

GEVORGYAN, R. A.; ABRAMOV, E. S. Detection of packaged and encrypted pe files with the use of machine-learning algorithm. In: Proceedings of the 11th International Conference on Security of Information and Networks. New York, NY, USA: Association for Computing Machinery, 2018. (SIN '18). ISBN 9781450366083. Disponível em: <https://doi.org/10.1145/3264437.3264481>.

LANDAUER, M. et al. Dynamic log file analysis: An unsupervised cluster evolution approach for anomaly detection. Computers Security, v. 79, p. 94–116, 2018. ISSN 0167-4048. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404818306333>.

MACIąG, P. S. et al. Unsupervised anomaly detection in stream data with online evolving spiking neural networks. Neural Networks, v. 139, p. 118–139, 2021. ISSN 0893-6080. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0893608021000599>.

NEDELKOSKI, S. et al. Self-attentive classification-based anomaly detection in unstructured logs. In: 2020 IEEE International Conference on Data Mining (ICDM). [S.l.: s.n.], 2020. p. 1196–1201.

PROVOTAR, O. I.; LINDER, Y. M.; VERES, M. M. Unsupervised anomaly detection in time series using lstm-based autoencoders. In: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). [S.l.: s.n.], 2019. p. 513–517.

XIN, G. et al. An anomaly detection method of encrypted traffic based on user behavior. In: Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics. New York, NY, USA: Association for Computing Machinery, 2021. (ICCIR '21), p. 51–56. ISBN 9781450390231. Disponível em: <https://doi.org/10.1145/3473714.3473724>.

ZHU, Y. et al. Phe: An efficient traitor tracing and revocation for encrypted file syncing-and-sharing in cloud. IEEE Transactions on Cloud Computing, v. 6, n. 4, p. 1110–1124, 2018.