TI 3 - Revisão Bibliográfica

Lara Souza, Matheus Sorrentino, Raul Fonseca, Victor Oliveira

Setembro 29, 2023

1 - Tema

Detecções de Anomalias em Arquivos Criptografados.

2 - Integrantes

2.1 - Lara Brígida Rezende Souza

[1] Recentemente, muitas empresas migraram seus dados para a nuvem usando serviços de sincronização e compartilhamento de arquivos (FSS), que foram implantados para usuários móveis. No entanto, as soluções Bring-Your-Own-Device (BYOD) para dispositivos móveis cada vez mais implantados também levantaram um novo desafio sobre como evitar que os usuários abusem do serviço FSS. Neste artigo, abordamos esse problema usando um novo modelo de sistema envolvendo abordagens de detecção, rastreamento e revogação de anomalias. A solução apresentada aplica um novo sistema criptográfico baseado em chave pública de limite, denominado criptografia hierárquica parcialmente ordenada (PHE), que implementa uma hierarquia de chaves de ordem parcial e é semelhante à hierarquia de papéis amplamente utilizada no RBAC. O PHE fornece dois mecanismos principais de segurança, ou seja, rastreamento de traidores e revogação de chaves, o que pode melhorar muito a eficiência em comparação com abordagens anteriores. A análise de segurança e desempenho mostra que o PHE é uma criptografia de limite comprovadamente segura e oferece os seguintes benefícios importantes de gerenciamento e desempenho: pode prometer rastrear com eficiência todas as possíveis coalizões de traidores e apoiar a revogação pública não apenas para os usuários, mas também para os grupos especificados.

[2] Havia recursos estáticos e dinâmicos distintos de arquivos de programas empacotados e criptografados; uma amostra de treinamento é criada com base em sua entrega conjunta. Métodos de aprendizado de máquina foram usados para construir um classificador para detecção de arquivos empacotados ou criptografados.

[3] Com o desenvolvimento das empresas e o seu crescimento gradual, os seus terminais de dispositivos continuam a expandir-se em termos de tipos, números e gamas de aplicações. A forma de proteção da segurança do terminal está se tornando cada vez mais severa e vulnerabilidades e vírus dos terminais surgem indefinidamente. Uma rede corporativa e um ambiente de terminal de alta qualidade, eficiente e seguro são uma garantia importante para o bom desenvolvimento das empresas. No entanto, os métodos de monitorização comummente utilizados dos terminais de equipamentos existentes, especialmente os métodos de detecção de tráfego encriptado, têm sido incapazes de satisfazer as necessidades de algumas empresas de monitorização em tempo real,

identificação rápida e bloqueio atempado de comportamentos de alto risco dos terminais. Neste artigo, é proposto um método de monitoramento de tráfego criptografado para usuários finais para realizar a detecção anormal de tráfego de usuários. O modelo de rede neural profunda é usado para extrair recursos de dados de comunicação e recursos de tráfego anormal para comparação de similaridade, a fim de julgar se é tráfego anormal.

2.2 - Matheus Moreira Sorrentino

[4] Consideramos o problema de detecção de anomalias com um pequeno conjunto de exemplos de anomalias parcialmente rotuladas e um conjunto de dados não rotulados em grande escala. Este é um cenário comum em muitas aplicações importantes. Os métodos relacionados existentes ajustam-se exclusivamente aos exemplos limitados de anomalias que normalmente não abrangem todo o conjunto de anomalias ou prosseguem com a aprendizagem não supervisionada a partir dos dados não rotulados. Propomos aqui, em vez disso, uma abordagem baseada em aprendizagem por reforço profundo que permite uma otimização ponta a ponta da detecção de anomalias marcadas e não marcadas. Esta abordagem aprende a anormalidade conhecida interagindo automaticamente com um ambiente de simulação tendencioso para anomalias, enquanto estende continuamente a anormalidade aprendida para novas classes de anomalias (ou seja, anomalias desconhecidas), explorando ativamente possíveis anomalias nos dados não rotulados. Isto é conseguido otimizando conjuntamente a exploração dos dados de pequenas anomalias rotuladas e a exploração das raras anomalias não rotuladas. Extensos experimentos em 48 conjuntos de dados do mundo real mostram que nosso modelo supera significativamente cinco métodos concorrentes de última geração.

[5] A detecção de anomalias baseada em log foi extensivamente estudada para ajudar a detectar anomalias complexas de tempo de execução em sistemas de produção. No entanto, as técnicas existentes apresentam vários problemas comuns. Primeiro, eles dependem fortemente de registros rotulados por especialistas para discernir padrões de comportamento anômalos. Mas rotular manualmente dados de log suficientes para treinar redes neurais profundas com eficácia pode demorar muito. Em segundo lugar, eles dependem da previsão do modelo numérico com base na entrada do vetor numérico, o que faz com que as decisões do modelo sejam em grande parte não interpretáveis pelos humanos, o que exclui ainda mais a correção de erros direcionada.
Nos últimos anos, testemunhamos avanços inovadores em grandes modelos de linguagem (LLMs), como o ChatGPT. Esses modelos provaram sua capacidade de reter o contexto e formular respostas perspicazes em conversas inteiras. Eles também apresentam a capacidade de conduzir uma aprendizagem rápida e contextualizada com capacidade de raciocínio. À luz dessas habilidades, é natural explorar sua aplicabilidade na compreensão do conteúdo dos logs e na condução da classificação de anomalias entre os logs do sistema de arquivos paralelos.

[6] Com um número crescente de ataques à rede utilizando comunicação criptografada, a detecção de anomalias no tráfego criptografado é de grande importância para garantir uma operação confiável da rede. No entanto, os métodos existentes de extração de recursos para detecção de anomalias de

tráfego criptografado apresentam dificuldades na extração de recursos, resultando em sua baixa eficiência. Neste artigo, propomos uma estrutura de detecção de anomalias de tráfego criptografado baseada na extração automática paralela de recursos, chamada detecção de tráfego criptografado profundo (DETD). O DETD proposto usa um autoencoder de pilha multicamada paralelo de pequena escala para extrair recursos de tráfego local do tráfego criptografado e, em seguida, adota um algoritmo de seleção de recursos baseado em regularização L1 para selecionar o conjunto de recursos mais representativo para a tarefa final de detecção de anomalias de tráfego criptografado. Os resultados experimentais mostram que o DETD tem robustez promissora na extração de recursos, ou seja, a eficiência de extração de recursos do DETD é 66% maior do que a do autoencoder empilhado convencional, e o desempenho de detecção de anomalias é tão alto quanto 99,998% e, portanto, o DETD supera o estrutura profunda de gama completa e outros algoritmos de detecção de anomalias de rede neural.

2.3 - Raul da Cruz Fonseca

[7] A proteção eficaz contra ataques cibernéticos requer monitorização e análise constantes dos dados do sistema numa infraestrutura de TI, tais como ficheiros de registo e pacotes de rede, que podem conter informações privadas e sensíveis. Os centros de operações de segurança (SOC), que são estabelecidos para detectar, analisar e responder a incidentes de segurança cibernética, muitas vezes utilizam modelos de detecção para tipos conhecidos de ataques ou para anomalias e os aplicam aos dados do sistema para detecção. O SOC também está motivado a manter seus modelos privados para capitalizar os modelos que são sua experiência de propriedade e para proteger suas estratégias de detecção contra aprendizado de máquina adversário. Neste artigo, desenvolvemos um protocolo para avaliar de forma privada modelos de detecção nos dados do sistema, no qual a privacidade dos dados do sistema e dos modelos de detecção é protegida e o vazamento de informações é totalmente evitado ou diminuído de forma quantificável. Nossa abordagem principal é fornecer criptografia ponta a ponta para os dados do sistema e modelos de detecção utilizando criptografia baseada em rede que permite operações homomórficas sobre texto cifrado. Empregamos conjuntos de dados recentes em nossos experimentos que demonstram que o sistema de detecção de intrusão que preserva a privacidade proposto é viável em termos de tempos de execução e requisitos de largura de banda e confiável em termos de precisão.

[8] Os ataques cibernéticos são omnipresentes e a sua rápida detecção é crucial para a segurança do sistema. A detecção de intrusão baseada em assinatura monitora os sistemas em busca de indicadores de ataque e desempenha um papel importante no reconhecimento e prevenção de tais ataques. Infelizmente, ele é incapaz de detectar novos vetores de ataque e pode ser evitado por variantes de ataque. Como solução, a detecção de anomalias emprega técnicas de aprendizado de máquina para detectar eventos de log suspeitos sem depender de assinaturas predefinidas. Embora a visibilidade dos ataques no tráfego de rede seja limitada devido à criptografia dos pacotes de rede, os dados de log do sistema estão disponíveis em formato bruto e, portanto, permitem análises granulares. No entanto, o processamento de logs do sistema é difícil, pois envolve diferentes

formatos e eventos heterogêneos. Para facilitar a detecção de anomalias com base em logs, apresentamos o AMiner, uma ferramenta de código aberto na caixa de ferramentas AECID que permite análise, análise e alertas rápidos de logs. Neste artigo, descrevemos a arquitetura modular do AMiner e demonstramos sua aplicabilidade em três casos de uso.

2.4 - Victor Cabral de Souza Oliveira

[9] A detecção de anomalias, também conhecida como detecção de valores discrepantes ou detecção de novidades, tem sido uma área de pesquisa duradoura, porém ativa, em várias comunidades de pesquisa há várias décadas. Ainda existem algumas complexidades e desafios únicos que exigem abordagens avançadas. Nos últimos anos, o aprendizado profundo permitiu a detecção de anomalias, ou seja, a detecção profunda de anomalias, emergiu como uma direção crítica. Este artigo examina a pesquisa de detecção de anomalias profundas com uma taxonomia abrangente, cobrindo avanços em 3 categorias de alto nível e 11 categorias refinadas dos métodos. Revisamos suas principais intuições, funções objetivo, suposições subjacentes, vantagens e desvantagens e discutimos como elas abordam os desafios mencionados acima. Discutimos ainda um conjunto de possíveis oportunidades futuras e novas perspectivas para enfrentar os desafios.

[10] Ao contrário das técnicas de detecção de intrusão baseadas em assinatura ou uso indevido, a detecção de anomalias é capaz de detectar novos ataques. No entanto, o uso da detecção de anomalias na prática é dificultado por uma alta taxa de alarmes falsos. Foi demonstrado que as técnicas baseadas em especificações produzem uma baixa taxa de alarmes falsos, mas não são tão eficazes quanto a detecção de anomalias na detecção de novos ataques, especialmente quando se trata de sondagens de rede e ataques de negação de serviço. Este artigo apresenta uma nova abordagem que combina detecção de intrusão baseada em especificações e detecção de intrusão baseada em anomalias, mitigando os pontos fracos das duas abordagens e ampliando seus pontos fortes. Nossa abordagem começa com especificações de máquinas de estado de protocolos de rede e aumenta essas máquinas de estado com informações sobre estatísticas que precisam ser mantidas para detectar anomalias. Apresentamos uma linguagem de especificação na qual todas essas informações podem ser capturadas de forma sucinta. Demonstramos a eficácia da abordagem nos dados de avaliação de detecção de intrusão do Lincoln Labs de 1999, onde somos capazes de detectar todos os ataques de investigação e negação de serviço com uma baixa taxa de alarmes falsos (menos de 10 por dia). Considerando que a seleção de recursos foi uma etapa crucial que exigiu muito conhecimento e conhecimento no caso de abordagens anteriores de detecção de anomalias, mostramos que o uso de especificações de protocolo em nossa abordagem simplifica esse problema. Além disso, o componente de aprendizado de máquina da nossa abordagem é robusto o suficiente para operar sem supervisão humana e rápido o suficiente para que nenhuma técnica de amostragem precise ser empregada. Como mais uma prova de eficácia, apresentamos resultados da aplicação da nossa abordagem para detectar vírus de e-mail furtivos em um ambiente de intranet.

Referências

[1] Y. Zhu, G. Gan, R. Guo and D. Huang, "PHE: An Efficient Traitor Tracing and Revocation for Encrypted File Syncing-and-Sharing in Cloud," in IEEE Transactions on Cloud Computing, vol. 6, no. 4, pp. 1110-1124, 1 Oct.-Dec. 2018, doi: 10.1109/TCC.2016.2573299.

[2] R. A. Gevorgyan and E. S. Abramov. 2018. 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 (SIN '18). Association for Computing Machinery, New York, NY, USA, Article 28, 1–2. https://doi.org/10.1145/3264437.3264481

[3] Gong Xin, Zhao Xixi, Xin Haoguang, Gu Liang, Meng Yaning, Ma Xin, Dong Chenni, Duan Xiaorong, Sun Haichuan, and Wang Liguo. 2021. An anomaly detection method of encrypted traffic based on user behavior. In Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics (ICCIR '21). Association for Computing Machinery, New York, NY, USA, 51–56. https://doi.org/10.1145/3473714.3473724

[4] Guansong Pang, Anton van den Hengel, Chunhua Shen, and Longbing Cao. 2021. Toward Deep Supervised Anomaly Detection: Reinforcement Learning from Partially Labeled Anomaly Data. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD '21). Association for Computing Machinery, New York, NY, USA, 1298–1308. https://doi.org/10.1145/3447548.3467417

[5] Chris Egersdoerfer, Di Zhang, and Dong Dai. 2023. Early Exploration of Using ChatGPT for Log-based Anomaly Detection on Parallel File Systems Logs. In Proceedings of the 32nd International Symposium on High-Performance Parallel and Distributed Computing (HPDC '23). Association for Computing Machinery, New York, NY, USA, 315–316. https://doi.org/10.1145/3588195.3595943

[6] Long, Gang, and Zhaoxin Zhang. "Deep Encrypted Traffic Detection: An Anomaly Detection Framework for Encryption Traffic Based on Parallel Automatic Feature Extraction." Computational Intelligence and Neuroscience 2023 (2023): 3316642-12. Web.

[7] Karaçay, Leyli, Erkay Savaş, and Halit Alptekin. "Intrusion Detection Over Encrypted Network Data." Computer Journal 63.4 (2020): 604-19. Web.

[8] Max Landauer, Markus Wurzenberger, Florian Skopik, Wolfgang Hotwagner, and Georg Höld. 2023. AMiner: A Modular Log Data Analysis Pipeline for Anomaly-based Intrusion Detection. Digital Threats 4, 1, Article 12 (March 2023), 16 pages. https://doi.org/10.1145/3567675

[9] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. 2021. Deep Learning for Anomaly Detection: A Review. ACM Comput. Surv. 54, 2, Article 38 (March 2022), 38 pages. https://doi.org/10.1145/3439950

[10] R. Sekar, A. Gupta, J. Frullo, T. Shanbhag, A. Tiwari, H. Yang, and S. Zhou. 2002. Specification-based anomaly detection: a new approach for detecting network intrusions. In Proceedings of the 9th ACM conference on Computer and communications security (CCS '02). Association for Computing Machinery, New York, NY, USA, 265–274. https://doi.org/10.1145/586110.586146.