UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DIRETORIA DE PESQUISA E PÓSGRADUAÇÃO
CURSO DE ESPECIALIZAÇÃO EM INDÚSTRIA 4.0

JEFFERSON MEDEIROS DA SILVA

APLICAÇÃO DE MACHINE LEARNING PARA DETECÇÃO DE ANOMALIAS EM REDE SCADA

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

PONTA GROSSA
2020

JEFFERSON MEDEIROS DA SILVA

APLICAÇÃO DE *MACHINE LEARNING* PARA DETECÇÃO DE ANOMALIAS EM REDE SCADA

Trabalho de Conclusão de Curso de Especialização apresentada como requisito parcial à obtenção do título de Especialista em Industria 4.0, da Universidade Tecnológica Federal do Paraná, Câmpus Ponta Grossa.

Orientador: Prof. Dr. Rui Tadashi

PONTA GROSSA

2020

Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ CÂMPUS PONTA GROSSA
Diretoria de Pesquisa e Pós-Graduação

TERMO DE APROVAÇÃO DE TCCE

Aplicação de machine learning para detecção de anomalias em rede SCADA

Jefferson Medeiros da Silva

Este Trabalho de Conclusão de Curso de Especialização (TCCE) foi apresentado em oito de fevereiro de 2020, como requisito parcial para a obtenção do título de Especialista em Indústria 4.0. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

_____

Prof. Rui Tadashi Yoshino

Membro titular

_____

Prof. Marcelo Carvalho

Membro titular

_____

Profa. Fernanda Tavares Treinta

Membro titular

- A Folha de aprovação encontra-se assinada na secretaria do curso -

Dedico este trabalho à Deus, que nos criou e foi criativo nesta tarefa. Seu fôlego de vida em mim me foi sustento e me deu coragem para questionar realidades e propor sempre um novo mundo de possibilidades, à minha família, pelos momentos de ausência.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço ao meu orientador Prof. Dr. Rui Tadashi, pela sabedoria com que me guiou nesta trajetória.

Aos meus colegas de sala.

Gostaria de deixar registrado também, o meu reconhecimento à minha família, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

.

RESUMO

Silva, Jefferson Medeiros. **Aplicação de machine learning para detecção de anomalias em rede SCADA :**. 2020. 26 f Monografia (Especialização em Engenharia de Produção) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2018.

Os sistemas SCADA são importantes em processos industriais. No princípio, esses sistemas eram isolados e sem conectividade externa. Atualmente, os modelos de sistemas SCADA baseiam-se em conectividade e em sistemas abertos e estão sendo conectados às intranets corporativas e à Internet visando o aumento da eficiência e da produtividade. Essa integração com a internet acarretou múltiplos problemas relacionados com segurança. Entretanto, sistemas para detecções destas anomalias podem ser capazes de detectar possíveis ataques enviados a esses sistemas. O monitoramento dessas redes de forma automática se faz cada vez mais necessária utilizando ferramentas de machine learning para que o modelo possa ir se adaptando as novas configurações da rede.

Palavras-chave: Machine Learning. Cybersecurity. SCADA Systems

ABSTRACT

SILVA, Jeffrson Medeiros. **Machine learning application for anomaly detection in SCADA network:** 2020. 26 p. Monograph (Especialization in Production Engineering) - Federal Technology University - Paraná. Ponta Grossa, 2018.

SCADA plays an important role into industrial process. In the beginning, these systems were standalone models, with closed architectures and no external connectivity. Nowadays, SCADA needs connectivity and open systems and are connecting to corporate intranets and to the Internet for improve efficiency and productivity. This integration with the internet has brought several security issues. However, anomaly detection systems would be able to detect possible attacks on those systems, The monitoring of these networks automatically becomes increasingly necessary using machine learning tools so that the model can adapt to the new network configurations

Keywords: Machine Learning. Cybersecurity. SCADA Systems

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

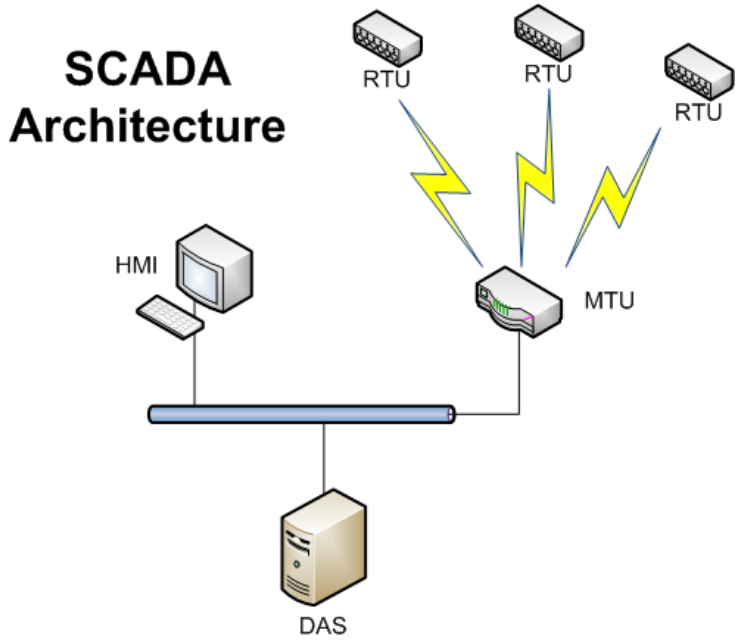
Os sistemas SCADA (*Supervisory Control and Data Aquisition* - Controle Supervisório e de Aquisição de Dados) realizam funções que incluem a supervisão e o controle de processos locais ou remotos em tempo real. Esses sistemas, em geral bastante complexos, são constituídos por computadores, aplicativos e dispositivos utilizados para a aquisição de dados e para atuar sobre os processos. Todos os equipamentos são interligados através de uma rede de comunicação. (SERGIO, 2004).

A maioria dos protocolos industriais utilizados na rede SCADA foram projetados inicialmente há algumas décadas, quando o sistema era fisicamente isolado de qualquer outra rede, o que contribui para a falta de demanda por serviços de segurança básicos, como autenticação e verificação de integridade nesses protocolos.

Os protocolos de comunicação utilizados em sistemas SCADA são projetados para garantir confiabilidade, operações em tempo real preciso e eficiência na comunicação entre componentes. Funções que prejudicam sua eficiência, incluindo funções de segurança, são desconsideradas na sua implementação, tornando-os menos seguros contra ameaças. Entre os diversos protocolos disponíveis para comunicação em sistemas SCADA, os mais amplamente utilizados são o Modbus e o DNP3 (EITELVEIN, 2015)

Em geral, os sistemas SCADA são construídos com uma configuração de unidade central de processamento, chamada de Estação Central ou de MTU (*Master Terminal Unit)* e de estações remotas, chamadas de RTU (*Remote Terminal Unit*). Esses componentes trocam informações através de um meio de comunicação. Os meios de comunicação podem ser links através de satélite, links de rádio, cabos metálicos ou fibras ópticas. A Figura 1 mostra a arquitetura simplificada de um sistema SCADA.

Figura 1 - Arquitetura simplificada de um sistema SCADA. O sistema é composto por uma Estação Central e várias Estações Remotas, interligadas através de um Meio de Comunicação.



Fonte: http://itech.fgcu.edu/faculty/zalewski/CEN3213/CEN3213security-6.html

A implementação de um IDS baseado em Detecção de Anomalias requer a realização da classificação, durante a execução, do tráfego de dados do sistema. Algoritmos de Aprendizado de Máquina - *Machine Learning* (ML) têm sido utilizados para a classificação de tráfego de dados por possuírem diversas características vantajosas para a tarefa. Enquanto técnicas de classificação de tráfego tradicionais dependem da inspeção do conteúdo de pacotes, mecanismos baseados em ML classificam os dados através de atributos que podem ser observados externamente, como tamanho dos pacotes e tempo entre a chegada de pacotes, criando padrões estatísticos. Há benefícios consideráveis nessa abordagem: o campo de dados do pacote deixa de ser obrigatoriamente visível (os dados podem estar criptografados, por exemplo), e o classificador não precisa conhecer a sintaxe dos dados nos pacotes de cada aplicação (ARMITAGE, 2008). Os algoritmos de ML podem ser classificados em três categorias:

- Aprendizado supervisionado
- Aprendizado não supervisionado
- Reforço

Conforme citado por (SIMÕES, 2006) para aprendizado supervisionado, os parâmetros internos da rede são determinados através da diminuição de uma

medida de erro definida entre valores de saída dados em exemplos mapeados entrada-saída e saídas atingidas pela rede para as mesmas entradas. Já para o aprendizado não supervisionado, a rede determina seus parâmetros sem o conhecimento da resposta desejada, tipicamente com base nas informações de números de classes desejadas e topologia da rede. Para o aprendizado por reforço alguma função heurística é utilizada para descrever a qualidade da resposta a uma dada entrada.

2 SEGURANÇA DE REDES CORPORATIVAS

As redes SCADA, como mencionadas anteriormente, são situadas no tratamento de processos. Desta forma, os procedimentos de segurança visam inibir que uma falha de segurança resulte em consequências mais severas para e garantir, também, que o sistema seja confiável e controlável Confiabilidade é a garantia de que o sistema e seus componentes realizarão suas funções sob determinadas condições pelo tempo em que essas condições forem mantidas e controlabilidade é a garantia de que, se bem manipulados seus condicionantes, o sistema e seus componentes realizarão as funções para as quais foram projetados. (SÉRGIO, 2004)

Segundo (TURCATO, 2015) ataques em redes são considerados anomalias definidas como ações diferentes observadas no comportamento normal do tráfego esperado, que podem ser indicativos de ataques, abuso (mau uso) na rede, eventos de falha, problemas de infraestrutura na coleta de dados, entre outros. Assim, conclui que, nem toda anomalia pode ser considerada um ataque, mas sempre representa uma informação suspeita que deve ser analisada.

Os mecanismos de detecção utilizados em IDSs são classificados em Detecção Baseada em Assinaturas - *Signature Based Detection* (SBD) e Detecção Baseada em Anomalias - *Anomaly Based Detectio*n (ABD). Técnicas de SBD consistem em registrar perfis de ataques maliciosos conhecidos, sem conhecimento do comportamento normal do sistema. Ao realizar a análise do tráfego, as informações são comparadas com os perfis de ataques existentes, a fim de identificar ameaças conhecidas. Já mecanismos de ABD se fundamentam em distinguir o comportamento normal esperado do sistema. Ao monitorar o tráfego da rede, os dados analisados são qualificados em tráfego normal, se são condizentes com o comportamento aguardado pelo sistema, ou anormal. (AXELSSON, 2000)

3 MACHINE LEARNING

Segundo (PIASSA, 2019) o aprendizado de máquina (*Machine learning*) é composto por modelos computacionais agrupados em base de dados que podem determinar e descrever o evento. Estes algoritmos empregam a experiência e treinamento para aperfeiçoar sua performance nas realização de previsões com um grau de definição mais precisas baseadas nas informações disponíveis para resolução de problemas práticos

Dentro do aprendizado de máquina basicamente se trabalha com dois modelos:

Aprendizado supervisionado (*Supervised Learning*), quando um algoritmo recebe uma base de dados para treinamento contendo amostras classificadas podendo, assim, efetuar previsões novos pontos.

Aprendizado não supervisionado (*Unsupervised Learning*), quando um algoritmo é alimentado apenas com dados não classificados e efetua o aprendizado e previsões baseados nesses dados para novos pontos.

4 OBJETIVO

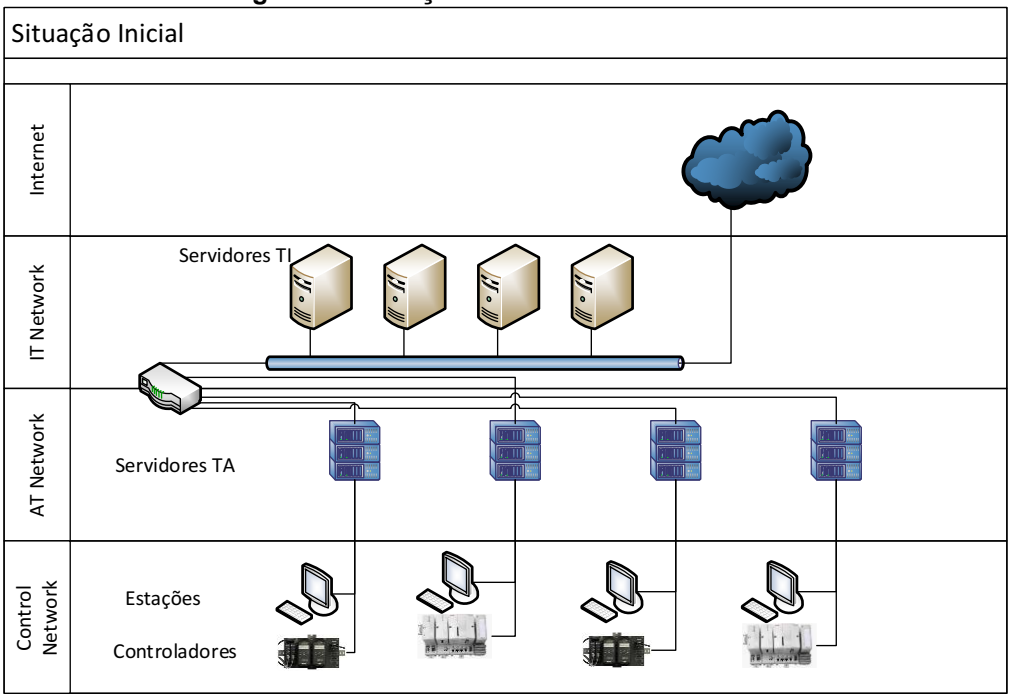
O trabalho tem como objetivo principal o estudo de aplicação de software para detecção de anomalias, em redes de computadores industriais, e avaliar sua eficiência quanto para monitoramento em redes SCADA.

5 METODOLOGIA

O trabalho tem como objetivo principal o estudo de aplicação de software para detecção de anomalias e monitoramento de rede SCADA para um ambiente escolhido em uma fábrica na região dos campos gerais – PR

Para realização do trabalho foi realizado um levantamento da situação inicial da rede analisada (Figura 2) onde foi evidenciado cinco servidores que não possuíam um sistema de controle específico.

Figura 2 - Situação atual da rede analisada.

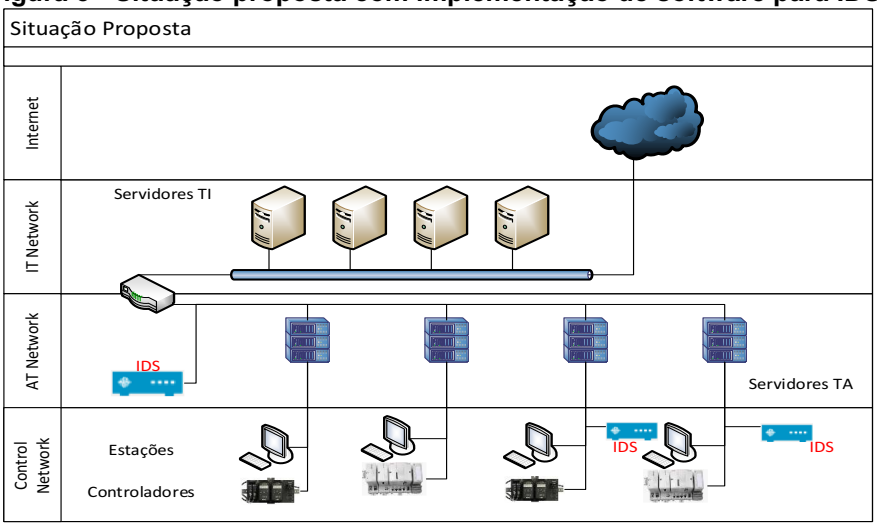


Fonte: Autoria própria

Após os estudos foi proposto (Figura 3) um novo arranjo contendo, uma nova DMZ para servidores de parceiros, implementação de um novo roteador e três módulos de IDS nos quais o software de monitoramentos irá coletar os dados

durante quinze dias como fase de aprendizado do ML (*Machine Learning*) e posteriormente iniciariam os comparativos entre o cenário apreendido e possíveis anomalias

Figura 3 - Situação proposta com implementação de software para IDS



Fonte: Autoria própria

Durante a fase de aprendizado foi utilizado o modelo supervisionado onde o responsável pela rede de automação validou possíveis anomalias no sistema e foi definido a situação “normal” de operação

Como resultados durante essa fase foram destacados:

Inventário de Ativos: Construir mapa de rede de automação classificando automaticamente o papel de diferentes elementos de rede *(SCADA Master, Slave, Historian, Terminal, etc*) compreendendo todos os padrões de comunicação do ambiente (protocolos e sub funções granulares) e incluindo particularidade de Sistemas operacionais/firmwares/hardware/fabricante e vulnerabilidades de todos os elementos.

Monitoramento Operacional: Monitorar processos críticos do ambiente (variáveis, valores e tempos do processo) e comportamento das diferentes funções de um processo para observar possíveis anomalias comportamentais no processo de supervisão e aquisição de dados.

Segurança Cibernética: A partir da abordagem de detecção de anomalias híbrida detector ataques, vulnerabilidades, indicadores de anormalidade de rede e proliferações de *malware* moderno como parte da prevenção e detecção de ameaças.

6 DISCUSSÃO

Com uma equipe central faz a interpretação dos alertas, ainda no modelo 8/5 (horas/dia), um número substancial de eventos está sendo evitado por estarem sendo diagnosticados antecipadamente bem como tentativas de invasão no sistema. Há potencial de maximizar estes monitoramentos e expandir o programa bem como ir melhorando seu modelo e constante atualizado de novas ameaças.

7 CONCLUSÃO

Pode-se concluir que as novas abordagens para monitoramento das redes, em especial as redes SCADA, com a utilização de *machine learning* se faz necessária visto a velocidade de surgimento de novas ameaças dentro do cenário atual e futuro, e a impossibilidade de criação de regras e monitoramento de modo manual.

REFERÊNCIAS

ARMITAGE, T. T. A Survey of Techniques for Internet Traffic Classification using Machine Learning. **IEEE COMMUNICATIONS SURVEYS & TUTORIALS**, VOL. 10, NO. 4, pp. 56-76. (FOURTH QUARTER de 2008).

AXELSSON, S. Intrusion Detection Systems: **A Survey and Taxonomy**. (14 de Março de 2000).

EITELVEIN, L. D. **Implementação e Avaliação de um Mecanismo de Detecção de Anomalias em uma Ferramenta Smart Grid.** 54 f. Dissertação (Graduação) – Instituto de informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015

FRANK, J. “Machine learning and intrusion detection**: Current and future directions**,” in Proc. National 17th Computer Security Conference, Washington,D.C., Outubro 1994

PIASSA, PEDRO, V. **Utilização de deep learning para detecção de anomalias em redes,** 31 f. Dissertação (Graduação) - Ciência da Computação da Universidade Estadual de Londrina, Londrina , 2019

SERGIO, P. &. Aspectos de segurança em sistemas SCADA **- Uma visão geral.** Universidade Federal do Rio Grande do Norte, Natal 2004

SIMÃO, A. D. **Aprendizado não supervisionado em redes neurais pulsadas de base radial**. 184 f Dissertação (Doutorado) – Escola Politécnica Universidade de São Paulo, São Paulo 2006.

TURCATO, Afonso Celso; FLAUZINO, Rogério Andrade; SESTITO, Guilherme Serpa; DIAS, André Luís; BRANDÃO, Dennis. Ataque denial of service em redes Profinet: estudo de caso. **Anais**.. Natal, RN: SBA, 2015.Disponível em: http://www.sbai2015.dca.ufrn.br/download/artigo/67

SHI, ZHONGZSHI, **Principles of Machine Learning**. International Academic Publishers, 1992.

SIMON, H., “**Why should machines learn?”** em R. S. Michalski, J. G. Carbonell, and T. M. Mitchell (editors) Machine Learning: An Artificial Intelligence Approach. Morgan Kaufmann, 1983.

WITTEN, I. e FRANK, E. Data Mining: **Practical Machine Learning Tools and Techniques with Java Implementations (Second Edition)**. Morgan Kaufmann Publishers, 2005.