10º Encontro de Ensino Pesquisa e Extensão

Patrocínio, MG, outubro de 2023

UM ESTUDO DAS TECNOLOGIAS DE DETECÇÃO E MITIGAÇÃO DE AMEAÇAS DE ENGENHARIA SOCIAL EM REDES SOCIAIS PROFISSIONAIS

Mateus Paláuro Queiroz da Silva, Júnio Moreira
<mateusffxv@gmail.com>,<juniomoreira@iftm.edu.br>
Instituto Federal do Triângulo Mineiro
Modalidade: Pesquisa
Formato: Artigo Completo

Resumo

À medida que as redes sociais profissionais se tornam cada vez mais importantes para a interação profissional, também aumenta a preocupação com os ataques de engenharia social nesses ambientes. Este artigo destaca inovações recentes em cibersegurança, como o uso de algoritmos de aprendizado de máquina para identificar padrões de comportamento suspeitos, análise de linguagem natural para detectar mensagens enganosas e monitoramento contínuo das atividades de conta para identificar atividades não autênticas. A análise e discussão dos resultados permite demonstrar que a integração de avaliações de risco em tempo real e a colaboração entre profissionais de segurança e plataformas de redes sociais desempenham um papel fundamental na evolução das estratégias de prevenção, essenciais para proteger a integridade das redes sociais profissionais e garantir a segurança das informações compartilhadas pelos usuários.

Palavras-chaves: Cibersegurança. Engenharia Social. Redes Sociais Profissionais. Aprendizado de Máquina. Linguagem Natural.

Introdução

O avanço contínuo da tecnologia, a proliferação de dispositivos conectados à internet e a disseminação das redes sociais tornou-se um elemento fundamental da

vida moderna. Hoje, é comum que indivíduos em diferentes partes do mundo se conectem e comuniquem através dessas plataformas, seja para fins pessoais ou profissionais. A ascensão da internet também transformou a maneira como nos comunicamos, deslocando-se do tradicional e-mail para um ecossistema de comunicação mais dinâmico e interativo proporcionado pelas redes sociais.

No seu início, as redes sociais eram uma ocorrência rara, restrita a um grupo limitado de pessoas com acesso à internet e dispositivos capazes de suportar essas plataformas. No entanto, à medida que a internet se tornou mais acessível e os dispositivos mais difundidos, testemunhamos a proliferação de redes sociais, tanto para uso pessoal quanto profissional.

A comunicação via redes sociais atraiu uma audiência crescente, incluindo empresas, funcionários e consumidores, devido à sua simplicidade e eficiência. Com o tempo, grupos diversos desenvolveram software de automação para otimizar processos e interações nas redes sociais. No entanto, essa automação nem sempre é usada para fins legítimos, frequentemente recorrendo a táticas de engenharia social para persuadir os destinatários a compartilharem informações pessoais.

Esses esforços maliciosos de automação frequentemente se disfarçam como ofertas de emprego, cobranças fraudulentas, ou promessas de vantagens inacreditáveis. A engenharia social é usada para ganhar a confiança das vítimas e obter informações confidenciais. Tais interações podem ser completamente automatizadas ou envolver algum grau de intervenção humana.

À medida que a tecnologia avança, os métodos de engenharia social também evoluem, tornando-se cada vez mais sofisticados e difíceis de serem detectados ou mitigados. Portanto, torna-se imperativo encontrar soluções para proteger as pessoas contra golpes que, muitas vezes, se disfarçam tão bem que conseguem enganar indivíduos de diversas origens e classes sociais, resultando em prejuízos significativos com apenas um simples clique.

Este artigo se propõe a analisar esses métodos de engenharia social, examinar as medidas adotadas por algumas redes sociais para detectar e mitigar essas ameaças, e avaliar o estado atual das estratégias de prevenção e combate à engenharia social, seja na estabilidade, ou em momentos de instabilidade social, quando essas táticas

maliciosas costumam se intensificar.

Fundamentação Teórica

A engenharia social em ambientes online, especialmente em redes sociais e plataformas de Inteligência Artificial (IA), tem se tornado uma preocupação crescente devido à sua capacidade de manipular e explorar a confiança das pessoas. A qualidade e a integridade dos dados usados no treinamento de IA desempenham um papel crítico em determinar se a IA terá um comportamento benéfico ou prejudicial. Conforme observado em (BLAUTH; GSTREIN; ZWITTER, 2022), a IA pode adquirir comportamentos questionáveis e maliciosos se for alimentada com dados duvidosos ou mal-intencionados. Isso pode incluir tanto os criadores da IA que fornecem dados problemáticos quanto outros usuários que injetam conteúdo malicioso ou questionável durante o treinamento.

Um aspecto importante da engenharia social em IA envolve a criação de gatilhos que alteram o comportamento da IA em determinadas condições. Esse conceito é ilustrado pelo caso do chatbot de namoro conhecido como *Cyber-Lover*, que foi lançado em 2007. Esse chatbot atraiu usuários para salas de bate-papo e os persuadiu a clicar em links fraudulentos, extraindo informações valiosas (BLAUTH; GSTREIN; ZWITTER, 2022).

As redes sociais, devido à sua natureza interativa e de comunicação direta, tornaram-se alvos frequentes de ataques de engenharia social. De acordo com (YOO; CHO, 2022), as defesas contra ataques de phishing em redes sociais frequentemente mostram ineficiência. Os ataques em redes sociais são muitas vezes demorados e envolvem a criação de laços emocionais com as vítimas antes que informações sejam reveladas. Tais ataques são frequentemente relatados após o incidente, tornando a detecção precoce desafiadora. O estudo também destaca que os golpes de engenharia social resultaram em perdas significativas, como os 54,7 milhões de dólares em golpes de phishing em redes sociais na Coreia do Sul em dois anos (YOO; CHO, 2022).

A teoria do engano é frequentemente empregada em ataques de engenharia social, onde informações verdadeiras são fornecidas inicialmente para construir confiança,

seguidas por informações enganosas. No trabalho em (ABRI et al., 2022), os atacantes usam métricas de custo de cooperação para alcançar seus objetivos. A eficácia desses ataques depende em grande parte do grau de confiança estabelecido. Para mitigar essas ameaças, uma abordagem proposta é aumentar o custo para os atacantes, tornando as operações de engenharia social menos viáveis.

A pandemia COVID-19 trouxe um aumento significativo no número de incidentes relacionados a ataques de engenharia social. De acordo com (ARBERTAVICIUS; CARMESINI, 2022), as empresas do Reino Unido enfrentaram tentativas de roubo de dados a cada 47 segundos durante a pandemia. O Brasil também foi identificado como um dos líderes em crimes digitais baseados em engenharia social. Esses eventos destacam a importância de considerar fatores humanos e estar preparado para enfrentar ameaças mesmo em um ambiente tecnologicamente avançado (ARBERTAVICIUS; CARMESINI, 2022).

Pesquisas em (MATA; DIAS; SALLES, 2021) demonstraram que modelos de IA de longa duração, mesmo com treinamento constante, sofrem uma queda de acurácia ao longo do tempo. Uma abordagem sugerida para mitigar esse problema envolve modelos adaptativos e abordagens híbridas, mantendo uma maior acurácia ao longo do tempo. Além disso, a avaliação a partir de contas humanas é fundamental para manter a eficácia dos detectores de ameaças de IA.

Em resumo, a detecção e mitigação de ameaças de engenharia social em redes sociais profissionais e ambientes alimentados por IA são questões complexas e em constante evolução. A compreensão dos mecanismos de ataque, a qualidade dos dados de treinamento da IA, a identificação de gatilhos comportamentais, a proteção de redes sociais contra phishing e a adaptação de modelos de IA de longa duração são áreas críticas de pesquisa e ação para enfrentar essas ameaças.

Proposta

Neste estudo, apresentamos uma proposta de estratégias de prevenção de ameaças de engenharia social em redes sociais profissionais, com base nos resultados e discussões previamente apresentados. Estas estratégias visam proteger a integridade

das redes sociais profissionais e garantir a segurança das informações compartilhadas pelos usuários.

Para conduzir nossa pesquisa, utilizamos uma abordagem baseada em análise de materiais científicos e tecnológicos. A metodologia adotada envolveu a revisão e análise de artigos científicos, documentos teóricos e práticos, bem como informações disponíveis em fontes confiáveis na internet.

Uma das principais estratégias para a prevenção de ameaças de engenharia social em redes sociais profissionais é a integração de avaliações de risco em tempo real. Isso envolve o uso de algoritmos de aprendizado de máquina para identificar padrões de comportamento suspeitos e atividades não autênticas. As plataformas de redes sociais devem monitorar continuamente as interações dos usuários e identificar desvios significativos de padrões normais de comportamento.

A análise de linguagem natural desempenha um papel crucial na detecção de mensagens enganosas e atividades maliciosas. As plataformas de redes sociais profissionais devem empregar algoritmos avançados de processamento de linguagem natural para identificar padrões de discurso que possam indicar tentativas de engenharia social. Essa análise pode ser usada para detectar mensagens que tentam manipular ou coagir outros usuários.

As plataformas de redes sociais profissionais devem conduzir um monitoramento proativo das tendências de ataque e ameaças emergentes. Isso envolve a análise contínua de dados e o acompanhamento de novos métodos de engenharia social. A capacidade de adaptação rápida às novas ameaças é essencial para manter a segurança da plataforma.

É crucial estabelecer uma colaboração eficaz entre profissionais de segurança cibernética e as próprias plataformas de redes sociais profissionais. As empresas de redes sociais devem estar dispostas a compartilhar dados e informações relevantes sobre ameaças e atividades suspeitas com as equipes de segurança cibernética. Da mesma forma, os profissionais de segurança devem contribuir com insights e expertise para aprimorar as medidas de detecção e mitigação de ameaças.

Por fim, este estudo é classificado como qualitativo, pois nossa pesquisa baseia-

se na análise crítica e comparativa de materiais teóricos e práticos. Ao analisar diversos conceitos, conclusões e abordagens, buscamos criar uma síntese que contribua para a compreensão do problema em questão. A pesquisa é exploratória, pois nosso objetivo principal é descobrir como diversos autores pesquisaram, por que o fizeram, como conduziram seus experimentos e como relataram seus resultados. Ao final deste estudo, esperamos contribuir para o entendimento das tecnologias de detecção e mitigação de ameaças de engenharia social em redes sociais profissionais.

Resultados

Este artigo discutirá os resultados de nossa pesquisa nas tecnologias de detecção e mitigação de ameaças de engenharia social em redes sociais profissionais. Também discutiremos as implicações desses resultados e como eles afetam a segurança dessas plataformas, além de apresentar alguns fatores preventivos.

A discussão dos resultados destaca a importância da adaptação contínua das estratégias de segurança à medida que as ameaças de engenharia social evoluem. A colaboração entre profissionais de segurança e plataformas é fundamental para identificar e responder às ameaças de maneira eficaz. Além disso, a conscientização dos usuários sobre os riscos de engenharia social desempenha um papel crítico na prevenção de ataques.

Portanto, apresentamos uma visão geral dos resultados de nossa pesquisa e enfatizamos a importância contínua de investir em tecnologias de detecção e mitigação de ameaças de engenharia social em redes sociais profissionais. A segurança dessas plataformas é fundamental para preservar a confiança dos usuários e a integridade das interações profissionais.

Uso de Algoritmos de Aprendizado de Máquina

Um dos principais resultados de nossa pesquisa é a crescente adoção de algoritmos de aprendizado de máquina para identificar padrões de comportamento suspeitos em redes sociais profissionais. Esses algoritmos são capazes de analisar grandes volumes de dados de usuários e identificar atividades que se desviam do comportamento autên-

tico, a partir do treinamento e processamento de informações dos algoritmos. Isso inclui a detecção de contas falsas, atividades automatizadas e tentativas de manipulação.

No estudo de (SARAIVA, 2022), foi validado o potencial do machine learning para classificar emails como legítimos ou de phishing. Inicialmente, foram coletadas amostras de emails tanto legítimos quanto ilegítimos e realizou-se um tratamento dos dados para padronizá-los em um único idioma. Em seguida, aplicou-se um algoritmo de Processamento de Linguagem Natural (PLN) que, entre suas funcionalidades, extraiu um conjunto de características relacionadas a aspectos linguísticos. O resultado desse processo foi um conjunto de dados final composto por 72 características distintas.

Para a seleção de recursos, foram testados dois algoritmos: o método "*feature importance*"com o algoritmo Random Forest (RF) e o método SelectBest SKB. Notavelmente, o algoritmo RF se destacou, produzindo resultados superiores com uma precisão de 92,13%, o que pode ser considerado excelente. Por fim, conclui que "Pessoas, Processos e Tecnologias são os três pilares de uma visão holística da cibersegurança.

Análise de Linguagem Natural na Detecção de Mensagens Enganosas

Outro resultado significativo é o uso efetivo da análise de linguagem natural para detectar mensagens enganosas. Essa técnica permite que as plataformas de redes sociais profissionais identifiquem mensagens que tentam manipular ou coagir outros usuários. A capacidade de compreender o contexto e o significado por trás das mensagens é fundamental para a detecção precoce de ameaças.

O estudo realizado em (MIRHOSEINI; VAHEDI; NASIRI, 2020) envolveu pré-processamento de conteúdo de e-mails, seguido pela aplicação dos resultados a um classificador de spam. Além disso, utilizou-se um conjunto de dados da Apache chamado SpamAssassin. Esse conjunto de dados continha uma extensa coleção de e-mails classificados como spam (e-mails indesejados) e ham (e-mails legítimos), que foram divididos em duas partes: uma para treinamento, utilizando 90% do conjunto de dados, e outra para teste, abrangendo os restantes 10%. Aplicando a técnica de Máquina de Vetores de Suporte (SVM), o método implementado alcançou uma acurácia de 0.90, demonstrando ser altamente promissor.

Monitoramento Contínuo das Atividades de Conta

Nossos resultados também destacam a importância do monitoramento contínuo das atividades de conta para identificar atividades não autênticas. Isso inclui o rastreamento de interações incomuns, como o compartilhamento excessivo de conteúdo ou o envolvimento em comportamento automatizado. O monitoramento contínuo permite que as plataformas identifiquem ameaças em tempo real e tomem medidas preventivas imediatas.

No que se destaca, o estudo mencionado em (SALAHDINE; KAABOUCH, 2019) apresenta uma variedade de técnicas de detecção utilizando softwares de detecção e monitoramento contínuo. Antes da instalação do aplicativo em si, algumas ferramentas analisam o fluxo de informações, enquanto outras interrompem imediatamente o fluxo de dados ao detectar qualquer anomalia, com base no que já foi aprendido e comparado com seu banco de dados. O estudo conclui que os mecanismos de defesa baseados em inteligência artificial são os mais eficazes na proteção contra ataques de engenharia social.

Colaboração entre Profissionais de Segurança e Plataformas

Um dos resultados mais destacados é a necessidade de colaboração estreita entre profissionais de segurança cibernética e as próprias plataformas de redes sociais profissionais. Essa colaboração é essencial para compartilhar informações sobre ameaças emergentes e desenvolver estratégias eficazes de detecção e mitigação. A parceria entre especialistas em segurança e desenvolvedores de plataforma é fundamental para a evolução das estratégias de prevenção.

Os resultados de nossa pesquisa demonstram que as redes sociais profissionais estão cada vez mais expostas a ameaças de engenharia social. A evolução das estratégias de prevenção é essencial para proteger a integridade dessas plataformas e garantir a segurança das informações compartilhadas pelos usuários. A integração de tecnologias avançadas, como algoritmos de aprendizado de máquina e análise de linguagem natural, desempenha um papel fundamental nesse processo.

Considerações Finais

Este artigo discutiu a importância histórica das redes sociais e da comunicação, bem como o crescimento dessas plataformas e os desafios que surgiram com o avanço tecnológico. Destacou a influência das funções e dos dados de treinamento na criação de inteligências artificiais, que podem ser usadas para diversos fins, tanto benevolentes quanto maliciosos. Além disso, enfatizou o papel significativo dos algoritmos de inteligência artificial na engenharia social, que é empregada tanto por atacantes quanto por defensores de sistemas. A tática de enganar as vítimas, seja por meio de interações simples ou pelo estabelecimento de relacionamentos falsos, foi mencionada como uma estratégia comum na engenharia social.

Pelos papéis essenciais desempenhados pelas redes sociais profissionais e as ameaças persistentes de engenharia social para o roubo de informações, é evidente a necessidade contínua de atualizar os métodos de detecção e mitigação à medida que a tecnologia avança. O desenvolvimento tecnológico também impulsiona o aprimoramento das técnicas sofisticadas de engenharia social, causando prejuízos significativos a indivíduos e organizações. Nesse contexto, o artigo apresentou dados e propostas de solução por meio de pesquisa, destacando métodos eficazes no combate à engenharia social, como o uso de aprendizado de máquina, aprendizado de linguagem natural e o monitoramento contínuo das redes sociais com ferramentas eficazes.

Embora as ferramentas baseadas em inteligência artificial sejam eficazes na luta contra a engenharia social, o elo mais frágil continua sendo o usuário. Portanto, a colaboração entre profissionais de segurança e plataformas é fundamental para evitar o roubo de informações e desenvolver métodos de defesa mais eficazes, incluindo o uso de aprendizado de máquina e processamento de linguagem natural na detecção de ameaças. Além disso, a constante evolução tecnológica exige atualizações contínuas dessas tecnologias, bem como a conscientização e treinamento dos usuários, tornando-se cada vez mais crucial diante das ameaças de engenharia social cada vez mais complexas e difíceis de identificar.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRI, F. et al. Markov Decision Process for Modeling Social Engineering Attacks and Finding Optimal Attack Strategies. **IEEE Access**, v. 10, p. 109949–109968, 2022. DOI: 10.1109/ACCESS.2022.3213711.

ARBERTAVICIUS, G.; CARMESINI, M. O. C. O aumento de casos de engenharia social durante a pandemia de COVID-19. 004, 2022.

BLAUTH, T. F.; GSTREIN, O. J.; ZWITTER, A. Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. **IEEE Access**, v. 10, p. 77110–77122, 2022. DOI: 10.1109/ACCESS.2022.3191790.

MATA, E. N. da; DIAS, G. M.; SALLES, R. M. Detecção de Bots Sociais: Uma Discussão sobre o Tempo de Vida de Abordagens Tradicionais. In: SBC. ANAIS do XXI Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais. [S.l.: s.n.], 2021. P. 337–350.

MIRHOSEINI, S. R.; VAHEDI, F.; NASIRI, J. A. **'E-mail phishing detection using natural language processing and machine learning techniques**. [S.l.], 2020.

SALAHDINE, F.; KAABOUCH, N. Social Engineering Attacks: A Survey. **Future Internet**, v. 11, n. 4, 2019. ISSN 1999-5903. DOI: 10.3390/fi11040089. Disponível em: <https://www.mdpi.com/1999-5903/11/4/89>.

SARAIVA, M. A. C. **Deteção de e-mails phishing aplicando machine learning ao conteúdo**. 2022. Diss. (Mestrado).

YOO, J.; CHO, Y. ICSA: Intelligent chatbot security assistant using Text-CNN and multi-phase real-time defense against SNS phishing attacks. **Expert Systems with Applications**, v. 207, p. 117893, 2022. ISSN 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2022.117893. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417422011435>.