Paper: Onderzoeksvoorstel

Verbetering van bedrijfsnetwerken met machine learning-gebaseerde Intrusion Detection Systemen (IDS) voor nauwkeurige detectie van cyberaanvallen.

Research Methods, 2022-2023

Jasper Van de Kerkhove

E-mail: jasper.vandekerkhove@student.hogent.be
Project repo: https://github.com/hogenttin/rm-2223-paper-rmvandekerkhovej

Samenvatting

Intrusion Detection Systemen (IDS) worden door veel bedrijven gebruikt als beveiliging voor hun netwerken. Enkele voorbeelden hiervan zijn Snort en OSSEC. Beide deze systemen zijn open source wat wil zeggen dat er geen licentie- of aankoopprijs aangekoppeld is. Een nadeel dat verboden zijn met dergelijke IDS is dat deze vaker een false positive kan geven. Dit betekent dat het IDS systeem een melding gaat geven dat er een virus gedetecteerd is in het netwerk. Maar er dat het eigenlijk niet het geval is. Omdat men dan dergelijke meldingen moet gaan controleren heeft men dan steeds een menselijke controle nodig. In dit onderzoek zal er Machine Learning worden gebruikt om false positives te vermijden en te optimaliseren. Hierdoor kan men de werklast van een systeembeheerder drastisch gaan verminderen. Niet alleen gaat Machine Learning voor minder false positives zorgen maar zullen de algoritmen ook meer geavanceerde attacks gaan herkennen dan standaard IDS die deze attacks simpelweg niet zou herkennen. In mijn opleiding Toegepaste Informatica heb ik gekozen om mij te specialiseren in cybersecurity. Dit onderzoek sluit dus perfect aan bij mijn opleiding. De uitvoering van het onderzoek is gedaan met het systeem voorzien van Machine Learning. Het systeem is voor een bepaalde tijd in een veilige test omgeving geplaatst. Op deze manier leert het systeem malware kennen en kan het zijn algoritme gaan beginnen opbouwen en samenstellen. Tenslotte gaan we dan een standaard IDS naast een IDS voorzien van machine learning gaan plaatsen gedurende een bepaalde tijd. Op het einde worden beide netwerken dan vergeleken op basis van hoeveel geslaagde aanvallen er nog waren en hoeveel false positives er nog opgetreden zijn. In de conclusie is het duidelijk dat het aantal false positives verminderd is bij het systeem met Machine Learning. Het implementeren van Machine Learning in huidige IDS is dus zeker een positieve vooruitgang in de beveiliging van netwerken.

Keuzerichting: System & Network Administrator
Sleutelwoorden: Security, AI, Machine Learning

Inhoudsopgave

1. Inleiding

Vandaag bevinden er zich veel verschillende soorten gevaren op het internet . Enkele voorbeelden hiervan zijn DDOS attacks en malware. Een netwerk kan beveiligd worden tegen dit soort aanvallen door gebruik te maken van Intrusion Detection System. Een IDS systeem kan aanvallen detecteren en waarschuwing genereren. In dit onderzoek wordt er onderzocht om dit efficiënter te maken met behulp van het gebruik van Machine Learning. Er zal onderzocht worden of deze intrusion detection systemen met behulp van machine learning dezelfde performance kunnen geven met minder manueel werk nodig. In deze paper zal er ook onderzocht welke algoritmen wel werken en welk niet of minder goed. De doelgroep die we willen bereiken zijn grote bedrijven die veel belangrijke info bijhouden waardoor zij het beste IDS systeem kunnen gebruiken. Het resultaat dat ik wil creëren is een proof-of-concept die kan aantonen dat bepaalde onderdelen van het IDS met machine learning al op bepaalde aanvallen kan reageren. Om dit doel te bereiken zullen we daarmee een reeks testuitvoeren die enkele weken zal duren. We zullen de resultaten meten op onder andere detectiesnelheid en correctheid.

2. Literatuurstudie

1 Inleiding

1.1 Onderzoeksdoel

Het onderzoeksdoel is om IDS systemen performanter en efficienter te maken. In dit onderzoek gaan we gebruik maken van Machine learning om IDS systemen dus neuwkeuriger cyberaanvallen te laten detecteren en voorkomen.

Intrusion Detection Systems IDS:

IDS is één van de manieren dat organisaties momenteel gebruiken om te vechten tegen cyberaanvallen. Met intrusion detection ID kunnen we verdachte netwerk verkeer identificeren vooraleer dat het informatie van het bedrijf kan bereiken. Het is een proces waarbij dat het security brezaches gaat identificeren door events die gebrueren in het netwerk te onderzoeken. Omdat vandaag de dag iedereen constant en overal met het internet wil verbonden zijn is een gebrek aan netwerk breik niet meer verstaanbaar. Zo een uitgebreid netwerk geeft aanvallers meer en meer mogelijkheden om aanvallen ui te voeren. Daarom is het belangrijk om netwerk apparaten veilig te houden van een aanval. Dit gaat dan ook over hoe je succesvol gaat beschermen tegen gekende of niet gekende aanvallen. Dit wordt alleen maar moeilijker door het stijgende aantal cyber aanvallen.

Welke algoritmes zijn beter?

Er zijn verschillende mogelijheden van machine learning algoritmes die kunnen gebruikt worden. Enkele voorbeelden zijn decision tree, K-nearest neighbor, K-mean clustering, AE AutoEncoder, ANN Artificial Neural Network, DNN Deep neural Network en CNN Convolutional Neural Network. Uit een studie is gebleken dat AE AutoEncoder en DNN Deep neural network de meest gebruikte en beste algoritmes zijn om te gebruiken voor een IDS systeem.(Zeeshan Ahmad, 2020)
Een andere studie toont aan dat K-nearest neighbor het beste algoritme was in een test omgeving maar ook in een echte omgeving.
(Syam Akhil Repalle, 2017)

Belang dat er genoeg en correcte datasets worden gegeven om een algoritme te trainen.

Dit onderzoek toonde aan dat het van belang is dat er bij real-life testen moderne datasets worden gebruikt. In test omgevingen werden veelal oude datasets zoals KDD Cup'99 en NSL-KDD datasets gebruikt. Maar in een echte omgeving is het belangrijk dat er moderne datasets worden gebruikt om dat aanvallers ook constant nieuwe technieken ontwikkelen. (Zeeshan Ahmad, 2020)

Deze studie heeft een simpel decision tree algoritme gebruikt een openbaar beschikbaar KDD dataset. Ze hebben deze vergeleken met andere state-of-the-art algoritmes en hebben ontdekt dat de decision tree algoritme beter scoorde op het vlak van precisie. (Mahbooba, 2021)

Waarom kan Machine Learning niet altijd vertrouwd worden?

Een voorbeeld uit deze paper van Vanin en Newe toont aan dat je kan gebruik maken van unsupervised Machine Learning. Wanneer je gebruik maakt van unsupervised ML is er in principe geen interventie meer nodig van een persoon. Dat wil zeggen dat het volledig op zichzelf leert en verbeterd. Maar omdat dit dan ook een hoeger training level nodig heeft verbruikt het meer energie. Het brengt ook een hoger risico van onjuiste resultataen en valse detecties. (Patrick Vanin, 2022).

3. Methodologie

article

Fase 1: Selecteren van datasets

Doelstelling: Bepalen welke datasets gebruikt zullen worden voor het trainen van de machine learning modellen.

Aanpak:

- De datasets die gebruikt zullen worden moeten aan bepaalde voorwaarden voldoen om representatief te zijn. Ze bevatten verschillende netwerkverkeergegevens zoals de netwerkprotocollen, gebruikers, locaties, verkeer van applicaties en apparaten.

- Om de datasets te kunnen trainen voor machine learning moeten ze ook verschillende typen aanvallen bevatten, zoals DoS, MITM, brute force, enzovoort.

Resultaat, deliverable(s): Een lijst van datasets die gebruikt zullen worden voor het trainen van de machine learning modellen.

Fase 2: Implementeren van IDS met Machine Learning

Doelstelling: Het implementeren van een IDS met machine learning.

Aanpak:

- Er zal een IDS geïmplementeerd worden met machine learning. Deze zal getraind worden met de datasets die geselecteerd zijn in fase 1.

- De geschikte algoritmen zullen gekozen worden aan de hand van de literatuurstudie.

- Ik ga gebruik maken van een open source IDS zoals Suricata of Snort.

- Ik ga maar bepaalde onderdelen van het IDS implementeren en testen zoals Dos en mitm aanvallen. En hiervoor een proof of concept maken.

Resultaat, deliverable(s): Een IDS met machine learning die getraind is met de datasets die geselecteerd zijn in fase 1.

Fase 3: Implementatie van traditioneel IDS zonder machine learning

Doelstelling: Het implementeren van een traditioneel IDS zonder machine learning.

Aanpak:

- Voor het implementeren van een traditioneel IDS zonder machine learning zullen we gebruik maken van een bestaand IDS. Er kan gekozen worden uit verschillende open source systemen zoals Snort, Suricata, Bro, OSSEC, enzovoort.

- Deze zullen getest worden op dezelfde manier als de IDS met machine learning.

Resultaat, deliverable(s): Een traditioneel IDS zonder machine learning.

Fase 4: Evaluatie van de prestaties

Doelstelling: Het evalueren van de prestaties van de IDS met machine learning en de traditionele IDS.

Aanpak:

- Voer een vergelijkende evaluatie uit tussen het IDS-systeem met Machine Learning en het traditionele IDS-systeem.

- Evalueer de prestaties op basis van verschillende criteria, zoals het aantal correct gedetecteerde aanvallen, het aantal false positives en de reactietijd van het systeem.

Resultaat, deliverable(s): Een evaluatie van de prestaties van de IDS met machine learning en de traditionele IDS.

Fase 5: Analyse van resultaten

Doelstelling: Het analyseren van de resultaten van de IDS met machine learning en de traditionele IDS.

Aanpak:

- De resultaten van de IDS met machine learning en de traditionele IDS zullen geanalyseerd worden.

- Er zal gekeken worden naar de nauwkeurigheid van de detectie van de aanvallen en de false positives. Er zal ook gekeken worden naar de performantie van de systemen.

- de data kan geanalyseerd worden met behulp van een tool zoals wireshark om netwerk trafic te capteren en te analyseren.

- ook gebruik maken van evaluation frameworks zoals IDMEF en CIDF om de IDS data te raporteren delen.

Resultaat, deliverable(s): Een analyse van de resultaten van de IDS met machine learning en de traditionele IDS, en een conclusie trekken uit het onderzoek.

Fase 6: Conclusie aanmaken

Doelstelling: Een conclusie maken van alle resultaten uit de vorige fasen.

Aanpak:

- De resultaten herbekijken.

- Alle resultaten samenvatten.

Resultaat, deliverable(s): Een samenvatting van alle gevonden resultaten in de vorige fases.

4. Verwachte resultaten

Er wordt verwacht dat op basis van de literatuurstudie en de methodologie de voordelen Machine Learning zichtbaar zijn in de resultaten. Er werd een vergelijking gedaan van een regulier IDS en een IDS met ML in een testomgeving. Het resultaat moet aantonen dat er een duidleijk lagere kost is omdat er minder interactie zal nodig zijn van een fysiek persoon. Door het gebruik van een algoritme in het NIDS heeft men preventiever cyberaanvallen kunnen opsporen omdat het algortime de patronen van de hackers kan herkennen. Met een juiste keuze van algoritme en trainig met morderne datasets zal het aantal incorrecte interventies ook dalen en dus een veel neuwkeuriger systeem hebben.

5. Discussie, conclusie

Ter conclusie om een goed systeem te gebruiken waarbij IDS samen met AI en machine Learning samenwerken moeten er met enkele dingen rekening gehouden worden. De correcte algoritmes. Sommige algoritmes zoals decision tree en AE of DNN zijn beter en efficiënter dan andere. Er moet een afweging gemaakt worden

welke algoritmes meer precies zijn dan andere en welk ook sneller kunnen werken of energie zuiniger zijn. De juist datasets moeten gebruikt worden om het algoritme te trainen. Dit wil zeggen moderne en recente datasets die kunnen trainen op nieuwe aanvalstechnieken. Er moet gekozen worden tussen supervised ML of semi supervised ML. De betere optie is om te kiezen voor semi-supervised ML omdat dit nog deels interactie van een echt persoon gebruikt die het algoritme op basis van de resultaten kan bijsturen.

Referenties

Mahbooba, B. (2021, januari 28). *Explainable Artificial Intelligence (XAI) to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model*. Verkregen mei 26, 2023, van https://www.hindawi.com/journals/complexity/2021/6634811/

Patrick Vanin, T. N. (2022, december 23). *Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning*. Verkregen mei 20, 2023, van https://encyclopedia.pub/entry/39074

Syam Akhil Repalle, V. R. K. (2017, december 1). *Intrusion Detection System using AI and Machine Learning Algorithm*. Verkregen mei 24, 2023, van https://IRJET-V4I12314-libre.pdf

Zeeshan Ahmad, A. S. K. (2020, oktober 16). *Network intrusion detection system: A systematic study of machine learning and deep learning approaches*. Verkregen mei 23, 2023, van https://onlinelibrary.wiley.com/doi/full/10.1002/ett.4150