T.R. Wabeke Msc[*]

Op weg naar een veiliger .nl-domein met machine learning

Twee uitdagingen waaraan SIDN de komende tijd werkt

138

Trefwoorden:
domeinnamen, machine learning, misbruik, hacken, nepwebwinkels

In dit artikel bespreek ik kort wat machine learning is en worden twee vraagstukken geïntroduceerd waarbij SIDN Labs machine learning inzet voor het terugdringen van domeinnaammisbruik. SIDN Labs is het researchteam van SIDN, de beheerder van het .nl-domein. Dit onderzoek wil bijdragen aan het verhogen van de veiligheid en weerbaarheid van het internet.[1]

Introductie

Het terugdringen van domeinnaammisbruik is een belangrijke internet-brede uitdaging, ook voor het .nl-domein. Een voorbeeld van domeinnaammisbruik zijn nepwebwinkels. Deze webwinkels zijn een probleem, omdat ze geen producten leveren, namaakproducten leveren of creditcardfraude plegen. Een ander voorbeeld zijn domeinen waaraan een gehackte website is gekoppeld. Deze websites vormen een risico voor nietsvermoedende internetters, bijvoorbeeld omdat ze de computers van gebruikers kunnen infecteren met malware die hen bijvoorbeeld hun gebruikersnaam en wachtwoord afhandig maakt.

Het proactief detecteren van domeinnaammisbruik voor .nl is meestal ingewikkeld. Dit komt onder andere door de schaal: er zijn meer dan 5,8 miljoen .nl-domeinnamen waardoor het handmatig monitoren van ieder domein onhaalbaar is. Daarnaast is misbruik vaak lastig te herkennen: typerende eigenschappen van domeinen, waaraan een gehackte website is gekoppeld zijn lastig te definiëren, bijvoorbeeld omdat de kwaadwillenden die erachter zitten regelmatig van tactiek veranderen.

SIDN doet daarom onderzoek naar het gebruik van machine-learning-methodes om domeinnaammisbruik te detecteren en terug te dringen. Deze techniek is gericht op het automatisch extraheren van regels en patronen uit grote hoeveelheden data, waarmee beslissingen genomen of voorspellingen gedaan kunnen worden. Bijvoorbeeld de inschatting dat een domeinnaam waarschijnlijk gehackt is.

Hier volgt eerst een korte uitleg over machine learning. Vervolgens bespreken we de manier waarop we in ons onderzoek nepwebwinkels en gehackte websites detecteren in meer detail.

Definities
- DNS: Domain Name System, zorgt ervoor dat een domeinnaam zoals uitgeverijparis.nl wordt omgezet in een IP-adres zoals 192.0.2.112
- TLD: top-level-domein, bijvoorbeeld .com, .nl, .amsterdam
- Machine learning: techniek die automatisch regels en patronen extraheert uit grote hoeveelheden data waarmee beslissingen of voorspellingen gemaakt worden
- Voorbeeld: een datapunt waarbij de verwachte uitkomst of ground truth bekend is
- Model: de regels en patronen die zijn geëxtraheerd uit voorbeelden via machine learning
- Ground truth: een bekende uitkomst of categorie, wordt gebruikt om de inschattingen en classificaties van een model te evalueren

Wat is machine learning?

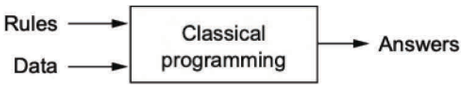
In de komende alinea's wordt machine learning uitgelegd aan de hand van een botnetdetector. Een botnet is een verzameling van computers die zijn geïnfecteerd met malware en die gezamenlijk opereren, bijvoorbeeld om een Distributed-Denial-of-Service- (DDoS-)aanval uit te voeren.[2] Hierbij proberen de computers die onderdeel uitmaken van een botnet tegelijkertijd een server te bereiken. Hierdoor wordt de server overbelast en onbereikbaar. Een botnetdetector helpt DDoS-aanvallen voorkomen door het netwerkverkeer van computers te analyseren. De detector staat normaal netwerkverkeer toe, maar verdacht verkeer wordt automatisch geblokkeerd. Een botnetdetector heeft regels nodig om normaal netwerkverkeer van DDoS-verkeer te onderscheiden.

* **Thymen Wabeke werkt als research engineer bij SIDN Labs en is redactielid van *P&I*. Thymen is bereikbaar via thymen.wabeke@sidn.nl.**

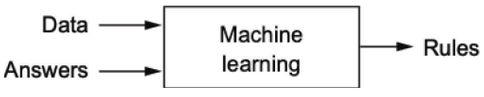
1 Een eerdere versie van dit artikel verscheen online op www.sidnlabs.nl/a/weblog/op-weg-naar-een-veiliger-internet-met-machine-learning.
2 www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/.

Er zijn twee manieren om deze regels te bepalen (zie figuur 1). De traditionele manier is om regels kennis-gedreven te programmeren. Een cybersecurity-expert gebruikt dan zijn kennis over het netwerkverkeer van botnets om de regels van een detector te bepalen. De expert weet bijvoorbeeld uit ervaring dat bepaalde eigenschappen van netwerkverkeer vaak voorkomen bij DDoS-aanvallen. Hij kan deze eigenschappen omzetten in regels waaraan nieuw netwerkverkeer wordt getoetst.

Daarnaast is het mogelijk om regels op een data-gedreven manier te definiëren. Dit is het paradigma achter machine learning. Hiervoor zijn grote hoeveelheden data nodig waarbij de verwachte uitkomst of 'ground truth' bekend is. Datapunten waarbij een ground truth bekend is worden 'voorbeelden' genoemd. In het geval van de botnetdetector is dit historisch netwerkverkeer waarbij een cybersecurity-expert ieder netwerkpakket als normaal of verdacht heeft gemarkeerd. Machine learning zorgt er vervolgens voor dat de computer de relatie tussen data en uitkomst herkent, ofwel de relatie tussen een netwerkpakket en de uitkomst als normaal of verdacht herkent. De geëxtraheerde relaties vormen het model en nieuw netwerkverkeer kan hieraan worden getoetst.



Figuur 1a. Kennis-gedreven programmeren



Figuur 1b. Data-gedreven programmeren

Bron figuren: Deep Learning with Python (François Chollet)

De afgelopen jaren zijn er veel doorbraken geweest op het gebied van machine learning.[3] Denk bijvoorbeeld aan AlphaGo van DeepMind die het spel Go beter speelt dan de menselijke kampioen of aan slimme assistenten zoals Google Assistent en Siri. Deze doorbraken zijn het resultaat van een toename in het onderzoek naar betere algoritmes. Ook is steeds meer data beschikbaar waarop deze algoritmes toegepast kunnen worden. Een uitgebreide uitleg over machine learning is te vinden in het tweede hoofdstuk van het whitepaper *Tijd voor implementatie van verantwoorde datadiensten* van TNO en het boek *Learning From Data*.[4]

Machine learning voor een nog veiliger .nl

SIDN past machine learning toe om een bijdrage te leveren aan de veiligheid en weerbaarheid van het internet en in het bijzonder voor het beheer van het .nl-domein. Machine learning is daarbij vooral geschikt voor vraagstukken waarbij regels niet handmatig bepaald kunnen worden terwijl er wel voorbeelden voorhanden zijn. In dit artikel worden twee van deze vraagstukken geïntroduceerd die gericht zijn op het terugdringen van domeinnaammisbruik. De overige vraagstukken zijn te vinden in een eerder verschenen blog (zie voetnoot 1).

Vraagstuk 1: Detectie van nepwebwinkels

De detectie van nepwebwinkels staat al een paar jaar op de onderzoeksagenda van SIDN. Het detecteren van nepwebwinkels gaat steeds beter, maar er is ook verbetering mogelijk.

Oplichters zitten niet stil en passen de webwinkels zo nu en dan aan. Het plan is dit op te vangen door een adaptief detectiemodel te ontwikkelen. Dit betekent dat er een systeem wordt ingericht dat detectiemodellen automatisch evalueert en her-traint. De verwachting is daarnaast dat veranderingen sneller herkend worden als de boordelingen van analisten gebruikt worden om het model te verbeteren. Het is hierbij een uitdaging om te bepalen welke beoordelingen het beste resultaat opleveren (active learning).

Ook is het streven de accuratesse van de ontwikkelde detectiemodellen te verhogen. Door kennis te delen met andere partijen is een breder perspectief te verkrijgen op het probleem en zijn blinde vlekken te voorkomen. Een pilot waarbij creditcardprovider ICS Cards verdachte shops aanleverde en feedback gaf over gedetecteerde webshops, is een goed voorbeeld van zo'n samenwerking. De wens bestaat om in de toekomst ook andere partijen te betrekken zodat we van elkaar kunnen leren en nepwebwinkels effectiever offline kunnen halen. Ten slotte is het relevant om te onderzoeken hoe effectief de nepwebwinkels bestreden zijn en welke (financiële) schade hiermee te voorkomen is.

Vraagstuk 2: Detectie van gehackte .nl-websites

Websites worden soms gehackt waardoor ze in handen komen van kwaadwillenden. Deze websites vormen een risico voor nietsvermoedende internetters, omdat ze computers van gebruikers bijvoorbeeld kunnen infecteren met malware. De detectie van .nl-domeinnamen waaraan een gehackte website gekoppeld is, is op dit moment reactief. Dat wil zeggen dat de operations teams van SIDN actief worden nadat ze een melding hebben ontvangen. Deze meldingen ontvangen we meestal via anti-phishing-diensten waarop SIDN geabonneerd is.

De komende periode onderzoeken we hoe we machine learning kunnen inzetten om .nl-domeinen met gehackte websites proactief te detecteren. De eerste stap is het zoeken naar patronen bij domeinnamen die in het verleden zijn gecompromitteerd. De ENTRADA-, OpenINTEL-

3 www.theverge.com/2019/1/28/18197520/ai-artificial-intelligence-machine-learning-computational-science.
4 publications.tno.nl/publication/34627028/KALZsM/TNO-2018-datadiensten.pdf en amlbook.com/.

en DMAP-databases zijn hierbij een voor de hand liggend startpunt.[5] De eerste database bevat informatie over het DNS-verkeer dat we ontvangen voor .nl-domeinnamen. Dit is geaggregeerde data die veelal afkomstig is van internetproviders en niet van individuele gebruikers. De data zegt dus niets over unieke bezoekers, maar het geeft bijvoorbeeld wel een indicatie over de populariteit van een domeinnaam. De OpenINTEL-database bevat informatie over de infrastructuur van een domein. De DMAP-database bevat daarnaast informatie over de content. Hierdoor weten we bijvoorbeeld welke webserver gebruikt wordt en of een .nl-domein securitystandaarden zoals DNSSEC geïmplementeerd heeft.

Er is een aantal zaken dat detectie van gehackte domeinen lastig maakt. Het vermoeden bestaat bijvoorbeeld dat het tijdsaspect van belang zal zijn. Kwaadwillenden proberen zoveel mogelijk bezoekers naar een gehackte website te lokken, zodat ze meer slachtoffers maken. Een plotselinge toename in het DNS-verkeer rondom een domeinnaam zou daarom kunnen duiden op een hack. Tegelijkertijd zijn de signalen grillig. Een piek in het DNS-verkeer kan immers net zo goed veroorzaakt worden door een populaire tweet of een krantenartikel. Denk bijvoorbeeld aan een onbekend restaurant dat een goede recensie krijgt waardoor de activiteit rondom de domeinnaam plotseling sterk verandert. Het duiden van patronen vereist geavanceerde signaal- en patroonherkenning, maar ook samenwerking met andere partijen. De koppeling met aanvullende databronnen van social media, registrars of internetproviders kunnen namelijk helpen bij het duiden van signalen.

Gezamenlijk machine-learning-onderzoek

Naast de twee machine-learning-uitdagingen die hier zijn besproken zijn er uiteraard ook andere toepassingen te bedenken die bijdragen aan een veiliger en weerbaar internet. In een eerder verschenen blog is het toewijzen van economische activiteiten aan domeinen en het detecteren van verdacht Internet-of-Things- (IoT-)verkeer beschreven.[1] We zijn daarnaast erg benieuwd naar ideeën voor aanvullende toepassingen. Ook staan we open voor samenwerkingen, bijvoorbeeld met andere disciplines. Heb je suggesties voor een relevante databron, een andere invalshoek, of waardevolle machine-learning-kennis? Mail thymen.wabeke@sidn.nl.

5 Zie respectievelijk entrada.sidnlabs.nl/, openintel.nl/ en dmap.sidnlabs.nl/.