



Thema: cyberweerbaarheid versus cybersecurity

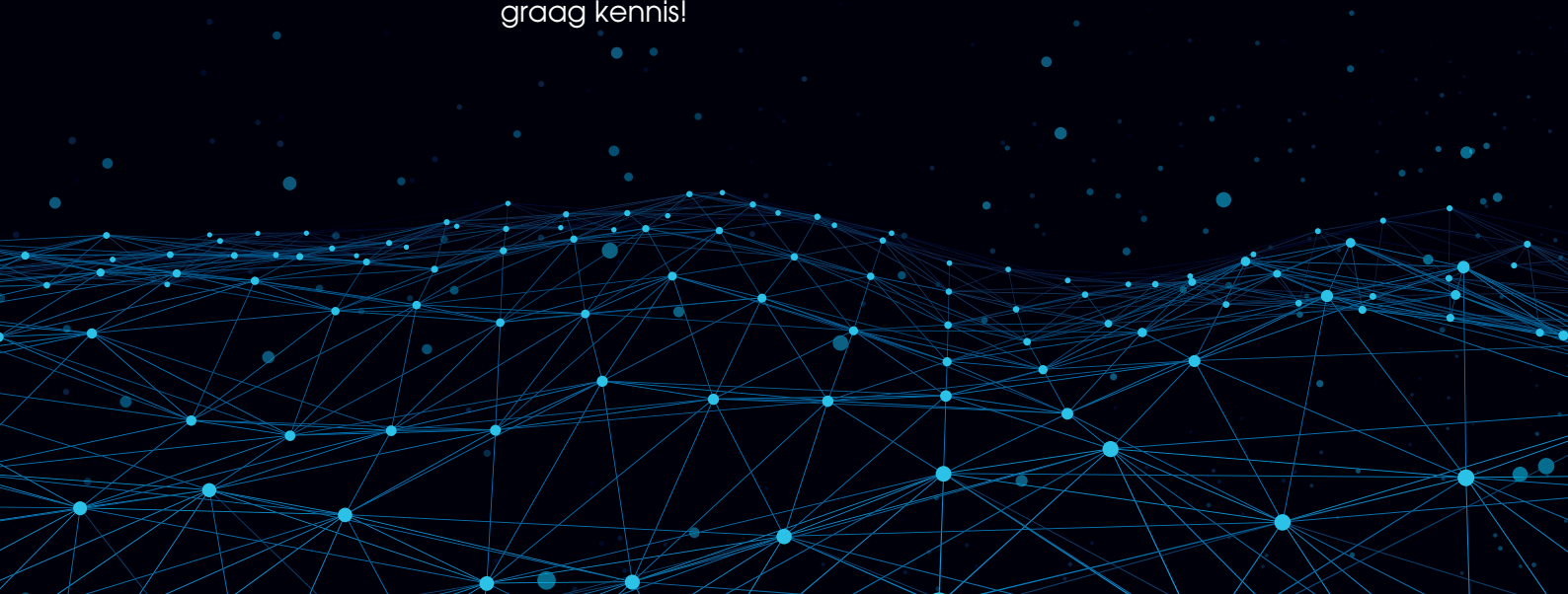
- ◆ Een groot digitaal dilemma
- ◆ Een veerkrachtige enterprise-architectuur vraagt om DEI
- ◆ AI ontwikkelen of toepassen: red teaming vereist!

Op zoek naar verbinding

Het IB Magazine van PvlB zoekt nieuwe redactieleden. Heb je passie voor informatiebeveiliging en wil je meebouwen aan hét magazine dat vakgenoten verbindt? Schrijf je graag of denk je mee over onderwerpen? Dan pas je perfect in ons team!

Ervaring is mooi meegenomen, maar vooral jouw motivatie telt. Samen zorgen we voor een sterk en relevant magazine, dóór en vóór professionals.

Nieuwsgierig? Mail ons via ibmagazine@pvib.nl – we maken graag kennis!



Op een scheiding van wegen



Chris de Vries

In deze uitgave laten wij Rick van der Kleij, lector Cyberveiligheid bij Avans, aan het woord. Rick, die sinds 8 mei jl. lector is, heeft als standpunt - in zijn lectorale rede en in zijn artikel hier - dat het tijd is voor een fundamenteel andere aanpak. De essentie: neem afscheid van de term 'cybersecurity'. De schrik slaat de lezer(es) wellicht om het hart, maar de term cyberweerbaarheid zal u geruststellen. Het gaat hierbij om acceptatie van de realiteit, te weten: absolute veiligheid bestaat niet meer en stimuleer herstelvermogen, reageer adequaat op incidenten en leer van wat er gebeurt. Voor eenieder verplichte lectuur.

Henk van Ee, voorzitter van de Stichting Cyberbrein, gaat in zijn artikel erop in dat er juist een gebrek is aan goed opgeleide cybersecurityspecialisten. Dit terwijl er kinderen zijn die op steeds jongere leeftijd digitale vaardigheden ontwikkelen en de ouderen inhalen. Dit gekoppeld aan Rick's paradigma maakt ons duidelijk dat onze problemen in de digitale maatschappij groot zijn.

Swarna Kumarswamy-Das en Ruggero Montalto werken beide bij TNO aan het Open-Source Software Libraries Security (OSSLS) binnen het kader van het PCSI-project. Zij gaan in op het gevaar dat marktcompetitie eist dat Open-Source Software Libraries (let op: OSSLS) ingezet wordt in casu dat de afhankelijkheid daarvan toeneemt en dus het gevaar op cyberattacks. Ook in ons vakgebied geldt dat de marktwerking onze data-beveiliging kan verstoren.

Nicole van der Meulen, lijkt een uitdaging aan ons allen voor te leggen met de titel: "Omarm het ongemak". Dat spreekt mij persoonlijk zeer aan, gelet mijn zware griep, die ik graag zou uitruilen, maar dat kan niet. Dus dan maar een paar weken rustig aan doen. Dat kan ook fijn zijn. Op Nicole terugkomende, ook zij gaat op cyberweerbaarheid in en koppelt dat aan een persoonlijk inkijkje uit haar leven. Compliment aan haar voor deze analogie.

Samaa Mohammad-Ulenberg en Linda Li duiken vervolgens in het thema van 'Red teaming'. Ik zou de lezer(es) willen adviseren ook het artikel van Dr. Jonathan Kwik ernaast te leggen. Daar waar 'Red teaming' binnen de eigen organisatie lijkt te vallen, omvat het thema van Jonathan: "Het toekomstige oorlogsfront!?" met als subtitel: "Uw netwerk, uw (sociaal) leven thuis en op straat" Dan komt het wel erg dicht aan (t)huis. Dit maakt het realiseren van een magazine als deze zo interessant, al deze samenhang, creativiteit en inzichten.

Ik wens onze trouwe lezers veel genoeg met dit nummer.

Chris

IN DIT NUMMER

- 03 Voorwoord - Op een scheiding van wegen
- 04 Een groot digitaal dilemma
- 12 PCSI Open-Source Software Libraries Security (OSSLS)
- 15 Omarm het ongemak
- 16 Het toekomstige oorlogsfront!?
- 20 Een complete gids: bedreigingsmodellering voor beveiligingsdeskundigen, architecten en ontwikkelaars

- 25 Een veerkrachtige enterprise-architectuur vraagt om DEI
- 26 AI ontwikkelen of toepassen: red teaming vereist!
- 31 Water naar de zee dragen, maar dan anders
- 32 Vroegsignalering van jonge cyberbreinen: Investeren in digitaal talent voor een weerbare toekomst

Auteur: Rick van der Kleij is lector Cyberweerbare Organisaties bij Avans Hogeschool. Zijn leeropdracht is het cyberweerbaarder maken van de Nederlandse maatschappij tegen een alsmaar toenemende cyberdreiging.



Een groot digitaal dilemma

En waarom we nu toch echt afscheid moeten nemen van cybersecurity

Cybersecurity legt de focus op veiligheid. Deze aanpak richt zich op het versterken van kwetsbare onderdelen van het bedrijf om voorspelbare dreigingen te kunnen weerstaan. Een aanpak die vaak de metafoor oproept van een kasteel met hoge muren en diepe slotgrachten: een ondoordringbare barrière tegen externe aanvallen. Maar bovenal een aanpak van cyberrisico's die niet langer voldoet, want het aantal incidenten en de impact ervan neemt alleen maar toe. Volgens Rick van der Kleij is een doorbraak naar een nieuw paradigma nodig. Dit nieuwe denkkader kunnen we vinden in het idee dat bedrijven weerbaar moeten zijn als reactie op cyberrisico's. We noemen deze kijk op de aanpak van cyberrisico's ook wel cyberweerbaarheid.

Dit artikel is een bewerking van de rede die Rick van der Kleij uitsprak bij de publieke aanvaarding van de functie van lector Cyberweerbare Organisaties aan het Centre of Expertise Veiligheid & Veerkracht, Avans Hogeschool op 8 mei 2025.

Zowel statelijke actoren als criminelen hebben het op onze digitale infrastructuur en onze gegevens gemunt, zo valt te lezen in de Veiligheidsstrategie voor het Koninkrijk der Nederlanden van de NCTV. De kans op cyberincidenten neemt toe. Wat precies door mag gaan als een cyberincident verschilt voor elk bedrijf. Maar over het algemeen is iets een cyberincident als het bedrijfsprocessen, financiën of reputatie ernstige schade kan toebrengen (1). Dit kan variëren van een hack (digitale inbraak) en gijzelsoftware tot een datalek en een softwarefout.

Niet alleen de kans op cyberincidenten neemt toe. Bedrijven worden daadwerkelijk voortdurend aangevallen. Soms zelfs zonder dat ze het

zelf in de gaten hebben. Vaak gaat het om cybercriminelen die snel en makkelijk geld willen verdienen. Of om scholieren die cyberaanvallen doen voor de lol. Maar we hebben ook te maken met voorbereidingshandelingen voor sabotage, spionage en ongewenste beïnvloeding door andere landen. Denk aan de bemoeienis van Russische schepen met onderzeese infrastructuur in de Noordzee, zoals onlangs nog in het nieuws was. We zijn echter niet klaar voor wat er de komende vier of vijf jaar op ons afkomt, aldus Rutte in zijn eerste grote toespraak als secretaris-generaal van de NAVO. Onze aanpak schiet tekort.

De opbouw van cyberrisico's

In het licht van bovenstaande is het dan ook niet verwonderlijk dat de Cyber Security Raad - een nationaal en onafhankelijk adviesorgaan van het kabinet dat is samengesteld uit vertegenwoordigers van publieke en private organisaties en de wetenschap - het demissionaire kabinet adviseert om meer te investeren in digitale veiligheid. De raad waarschuwt dat de huidige inspanningen en investeringen in cyber-

security niet genoeg zijn om de groeiende digitale dreigingen vanuit statelijke actoren en cybercriminelen het hoofd te bieden. Los van het feit dat meer investeringen nodig zijn, rijst de vraag waar dan in te investeren? Om die vraag te beantwoorden is het nodig om eerst naar het begrip cyberrisico te kijken en hoe daarmee om te gaan.

Algemeen gesteld is risico de kans op een incident maal de impact. Binnen de mogelijkheid dat een cyberincident optreedt (kans) is het voor cybersecurity nuttig om een onderscheid te maken tussen dreiging en kwetsbaarheid. Daar waar de politie en het Openbaar Ministerie zich vooral richten op het bestrijden van dreigingen via opsporing en vervolging van criminelen en het verstoren van criminele activiteiten, nemen bedrijven traditioneel vooral maatregelen om kwetsbaarheden te verhelpen en zich te beschermen tegen cyberdreiging.

Deze focus op preventie zien we dan ook terug in definities van cybersecurity: het beschermen van computers, servers, mobiele apparaten, elektronische systemen, netwerken en gegevens op die systemen tegen specifieke dreigingen die kwetsbaarheden proberen te misbruiken om gevolgen te bewerkstelligen die economische of politieke impact hebben. Inspanningen op het gebied van cybersecurity richten zich dan ook voornamelijk op het verminderen van kwetsbaarheden en, in mindere mate, op de gevolgen van een succesvolle aanval, bijvoorbeeld door het maken van back-ups. Het uitvoeren van risicobeoordelingen draagt hier in belangrijke mate aan bij.

Het uitvoeren van een risicobeoordeling is een steeds grotere uitdaging geworden. Dit komt deels door de moeilijkheid om adequaat rekening te houden met mogelijke cascade-effecten die voortvloeien uit storingen (2). Een complicerende factor is dat klassieke risicobeoordelingsmethoden niet kunnen worden toegepast als data niet beschikbaar is, incompleet is, of anderszins niet kan worden bepaald. Een risicobeoordeling vereist dat risico's kunnen worden geïdentificeerd ex ante. Door de snelle ontwikkelingen in het dreigingslandschap ontstaan echter telkens nieuwe dreigingen die op voorhand niet kunnen worden voorzien of voorkomen. Het gebrek aan data over deze nieuwe dreigingen, bijvoorbeeld over de frequentie van voorkomen en mogelijke impact op de bedrijfsvoering, maakt het uitvoeren van risicobeoordelingen lastig. Hierdoor is een cybersecuritystrategie slechts gedeeltelijk effectief.

Naar een nieuwe aanpak van cyberrisico

Een nieuwe risicomanagementbenadering is nodig. Een aanpak van cyberrisico's die rekening houdt met de toename aan complexiteit in het dreigingslandschap en die ook bescherming biedt tegen toekomstige cyberdreigingen. Denk hierbij aan plotselinge vormen van cyber-

criminaliteit die in een keer heel groot worden, zoals whatsappfraude. Het onverwachte karakter hiervan maakt het lastig om ons ertegen te beschermen. Het idee dat bedrijven cyberweerbaar (in het Engels: cyber resiliënt) moeten zijn als reactie op cyberincidenten is de afgelopen jaren dan ook opgekomen en populair geworden.

Het algemene idee van weerbaarheid is dat in plaats van alle inspanningen te richten op het buitenhouden van criminelen uit bedrijfsnetwerken, het beter is aan te nemen dat zij uiteindelijk door de verdediging van de systemen zullen breken. Dat er dus ook aandacht moet zijn voor het verminderen van de impact van deze gebeurtenis. De overheid stapt al over van de notie dat Nederland beschermd moet worden tegen cyberaanvallen, naar de invalshoek assume breach: ga ervan uit dat er al een kwaadwillende in je netwerk zit. In het verlengde hiervan zei justitieminister David van Weel in augustus 2024, nadat meerdere overheidsinstanties getroffen waren door een storing in het digitale krijgsmacht netwerk Nafin: 'Get used to it.' Van Weel: 'Kwetsbaarheden gaan we altijd houden. De kwestie is hoe snel kunnen we zorgen dat we kunnen doorwerken als er een storing is. En hoe snel kunnen we de storing vinden en zorgen dat het netwerk weer online komt.' Van Weel benadrukte ook dat overheid en bedrijfsleven weerbaarder moeten worden.

Weerbaarheid is meer dan een modewoord

Zoals gezegd is het idee dat organisaties cyber resiliënt moeten zijn als reactie op cyberincidenten de afgelopen jaren opgekomen en populair geworden. Het woord resilience is afgeleid van het Latijnse woord resillire wat 'to leap back' betekent volgens Oxford English Dictionary, ofwel terugveren. Het eerste geregistreerde gebruik van resilience is volgens de Oxford Living Dictionary uit 1529. In de context van cyberdreigingen is echter de term weerbaarheid gangbaarder dan terugveren. Dat heeft deels te maken met het feit dat weerbaarheid het actieve vermogen benadrukt van een systeem om weerstand te bieden aan dreigingen. Het legt meer de nadruk op de capaciteit van systemen om niet alleen terug te veren, maar ook om zichzelf te verdedigen en aan te passen aan voortdurende en toekomstige dreigingen.

Weerbaarheid is een complex concept, aangezien dit geen lineair proces of enkele handeling is, maar eerder vereist dat er competenties aanwezig zijn in verschillende fasen, zowel vóór als na een potentiële schokgebeurtenis, dat makkelijker te observeren is als het ontbreekt dan als het er is (3). Van Harmelen beschrijft weerbaarheid zelfs als een emergent fenomeen (4). Weerbaarheid is volgens haar geen ding of vaste eigenschap, maar refereert aan een complex dynamisch proces van positieve adaptatie aan een stressor dat zich ontvouwt over tijd.

In de praktijk is weerbaarheid sterk geïnspireerd door organische systemen, zoals het menselijk immuunsysteem

Van Harmelen stelt dat we weerbaarheid niet kunnen zien of testen, maar wel kunnen afleiden uit de reactie op een schokgebeurtenis, dat een bedrijf weerbaarheid laat zien. Dientengevolge zijn er meerdere definities van weerbaarheid in omloop, die allemaal enigszins van elkaar verschillen, afhankelijk van de vakdiscipline waarbinnen ze worden gebruikt of het niveau van aggregatie. In de kern beschrijft weerbaarheid het vermogen van een systeem, individu of gemeenschap om te herstellen van en zich aan te passen aan verstoringen of veranderingen, zoals natuurrampen of psychologische stress. De basisprincipes blijven vergelijkbaar: de capaciteit om te absorberen, aan te passen, en terug te keren naar een normale of verbeterde toestand na een verstoring.

Weerbaarheid in het veiligheidsdomein: cyberweerbaarheid

In de veiligheidskunde werd vanaf de jaren 2010 het begrip pas populair. Dit werd mede veroorzaakt doordat een groeiend aantal wetenschappers, maar ook praktijkmensen, inzag dat de traditionele definitie van veiligheid in de zin van de afwezigheid van ongevallen niet langer functioneel was en wel een revisie kon gebruiken (5).

In de veiligheidspraktijk is weerbaarheid sterk geïnspireerd door organische systemen, zoals het menselijk immuunsysteem, die zich aan veranderende omstandigheden kunnen aanpassen en zichzelf kunnen herstellen. Net zoals biologische systemen immuniteit ontwikkelen om te reageren op infecties en andere aanvallen, moeten systemen zich aanpassen, zo is de gedachte, aan de steeds veranderende bedreigingen. Net als het menselijk lichaam zijn systemen niet hermetisch afgesloten. Het immuunsysteem is ook niet in staat alle aanvallen buiten de deur te houden (6). Systemen moeten in staat zijn om vitale systeemfuncties te blijven beschermen en veerkrachtig te herstellen van de gevolgen van incidenten nu en in de toekomst. Dit idee sluit aan bij het idee om systemen in het veiligheidsdomein te bezien vanuit de socio-techniek, waarbij alle onderdelen van een systeem – zoals mensen, technologie en organisatorische structuren – in samenhang worden gezien. Samenwerking tussen deze componenten is vereist ten behoeve van gezamenlijke optimalisatie. Het streven is naar een veerkrachtig systeem dat niet alleen problemen voorkomt, maar ook steeds beter wordt in het omgaan met en herstellen van verstoringen.

Het figuur in deze bijdrage geeft het vermogen om te herstellen van cyberincidenten weer aan de hand van de zogenaamde weerbaardheidscurve (zie ook (7)). De y-as geeft de systeemprestatie weer en de x-as laat zien dat cyberweerbaarheid over de tijd verschillende uitkomsten kan hebben. De vier weerbaardheidsdoelen - voorbereiden, reageren, herstellen en aanpassen - staan weergegeven boven de x-as. Deze doelen spelen een belangrijke rol vóór, tijdens en na een verstoringende gebeurtenis. Onder de x-as staan een aantal gebeurtenissen vermeld. Op t_0 vindt een cyberincident plaats. Op t_1 leidt dit tot een aantasting van de systeemprestatie. De tijd tussen t_0 en t_1 wordt bepaald door de robuustheid van het systeem. Roubuustheid refereert aan de mate waarin een systeem in staat is om onverwachte nadelige effecten het hoofd te kunnen bieden zonder degradatie van systeemprestatie tot gevolg. Hoe groter de robuustheid, hoe groter de tijd tussen t_0 en t_1 . Op t_2 begint het herstel van de systeemprestatie. Op t_3 start het aanpassen van het systeem. De cyberweerbaarheid van een systeem is de oppervlakte die ligt besloten in het 'dal' tussen t_1 en t_3 . Hoe kleiner de oppervlakte, hoe weerbaarder het systeem.

Recent onderzoek suggereert dat een verstoringende gebeurtenis adaptatie of transformatie kan katalyseren, wat kan leiden tot een sterk verbeterde systeemprestatie na een crisis, bekend als 'bouncing forward', 'bouncing back better' (8). Idealiter leidt het herstel en de adaptatie dan ook tot een hoger algemeen weerbaardheidsniveau. In het figuur is dit weergegeven als scenario #4: het optimale cyberweerbaardheidsniveau. Dit scenario kan ook worden omschreven als bouncing back better. Het hogere weerbaardheidsniveau kan een positief effect hebben op de systeempresentatie. Door te leren van incidenten kan in de toekomst beter worden omgegaan met vergelijkbare verstoringen. Aanpassingsmaatregelen zouden dan ook tenminste gericht moeten zijn op een incrementele verhoging van de cyberweerbaarheid tot een optimaal niveau. Maar ook andere scenario's zijn denkbaar. Een verstoringende gebeurtenis kan leiden tot blijvende aantasting van de systeemprestatie en een nieuw lager normaal. Dit is weergegeven als scenario #2. Ook is denkbaar dat het systeem de verstoring niet te boven komt en in elkaar stort. Voor een bedrijf betekent dit dat zij failliet gaat. Dit is weergegeven als scenario #1. Op t_4 start de transitie terug naar de voorbereidende fase (zie ook (3)).



Rick van der Kleij tijdens het uitspreken van zijn lectorale rede aan Avans Hogeschool op 8 mei 2025.

Een systemische kijk op cyberweerbaarheid

Cyberweerbaarheid op het systeemniveau omvat meer dan alleen een technische infrastructuur; het is een samenhangend geheel van hardware, software, netwerken, gegevens, processen en mensen die nodig zijn om zich voor te bereiden op, te kunnen reageren op, herstellen van, en zich aan te passen aan cyberincidenten. Deze bredere benadering omvat een holistische of systemische kijk. Een aanpak waarbij weerbaarheid niet kan worden gereduceerd tot enkel de som van alle technische maatregelen die een bedrijf neemt, maar het resultaat is van constante interacties tussen alle elementen die noodzakelijk zijn om essentiële functies en diensten te leveren. Daarnaast klinkt steeds vaker door dat cyberdreigingen ook in samenhang moeten worden bekeken met andere dreigingen, zoals militaire of terroristische dreigingen. Niinistö (9) noemt dit een all hazard aanpak waarvan het belang wordt onderstreept door het demissionaire kabinet.

Cyberweerbaarheid gedefinieerd

Vanuit het hierboven geschetste brede en holistische perspectief volgt de in deze bijdrage door mij bepaalde definitie van cyberweerbaarheid:

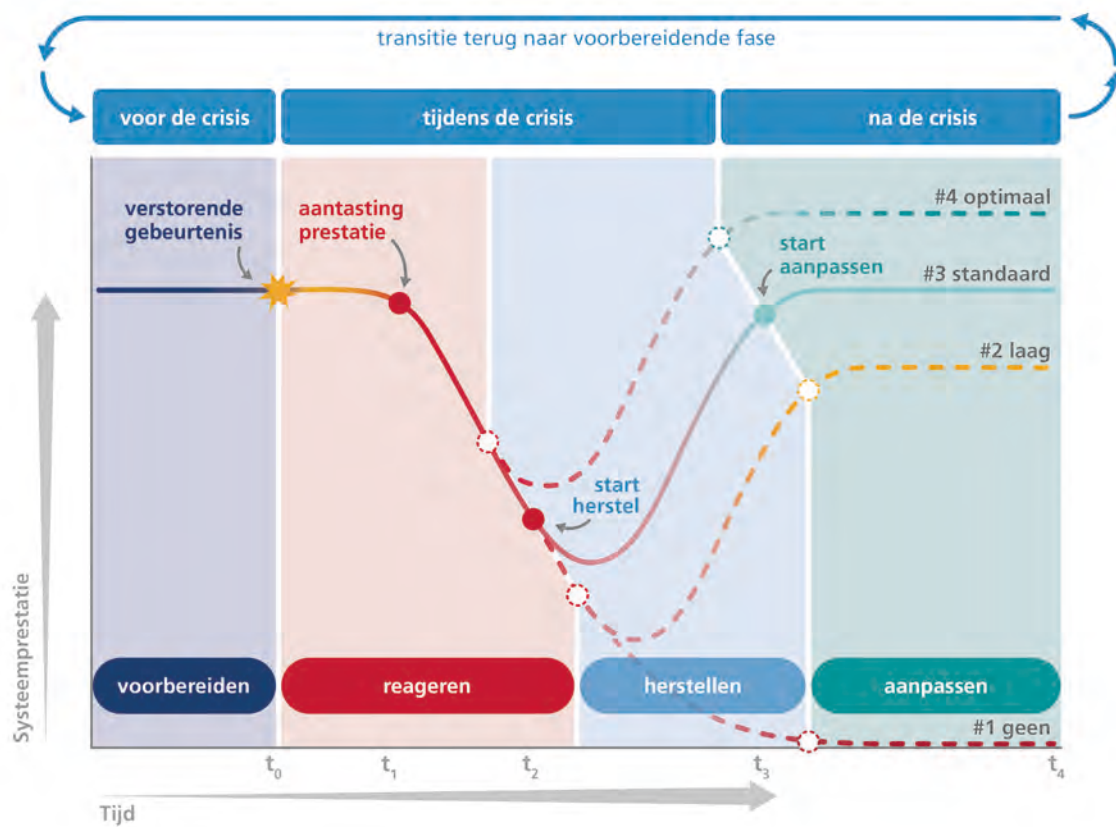
'Het vermogen van een systeem, dat bestaat uit een samenhangend geheel van hardware, software, netwerken, gegevens, processen en mensen; om onverwachte, positieve mogelijkheden te benutten om digitale processen te verbeteren en zich voor te bereiden op, te kunnen reageren op, te herstellen van en zich aan te passen aan een verstoring van de digitale omgeving die kan leiden tot een aantasting van het functioneren.'

Samenvattend kunnen we stellen dat het toenemende gebruik van de term cyberweerbaarheid een groeiend bewustzijn weerspiegelt dat een hermetisch afgesloten 'kasteel' onmogelijk te bouwen is. Cyberincidenten zijn niet te voorkomen. Het is belangrijk om te leren met tegenslag om te gaan. Dat is zeker niet eenvoudig. Het wordt daarom belangrijker om te accepteren dat incidenten (en kansen) zich voordoen, adequaat te reageren, ervan te herstellen waar nodig, te evalueren en te leren voor een volgende keer. Dat laat onverlet dat de basis – het 'kasteel' – zo stevig mogelijk moet zijn (6).

De cyberweerbaarheidsopgave

Ik kom dan nu tot wederom een belangrijk deel van mijn betoog. Hoe gaan we het groot digitaal dilemma waar bedrijven voor staan aanpakken? Hoe kunnen we zorgen dat bedrijven cyberweerder worden? Want het ontstaat niet vanzelf. Als een bedrijf meer cyberweerbaarheid nuttig en nodig acht, vraagt dit om gerichte inspanningen. Je komt er niet met een strategie van *laissez-faire*. In het huidige tijdsgewricht, met een verslechterende veiligheidssituatie, ontkomen we er niet aan om een flinke stap extra te zetten. We moeten onze cyberweerbaarheid verhogen door inzichtelijk te maken wat aanvullend nodig is, ofwel de cyberweerbaarheidsopgave.

Deze opgave kan worden begrepen als een regelproces van voorbereiden, organiseren en controleren van de middelen van een systeem om ervoor te zorgen dat het naar behoren functioneert (5). De doelen zijn ofwel het benaderen van een nieuwe en meer wenselijke toestand of het vermijden of ontwijken van een bestaande maar ongewenste toestand van cyberweerbaarheid. Naast het benaderen of vermijden van iets, kan de reden voor een verandering ook zijn om de huidige



Cyberweerbaarheidscurve

positie te handhaven. Dit lijkt misschien een paradox, maar dat is het niet. Als een bedrijf zich op de beoogde of gewenste positie bevindt, wil ze daar uiteraard blijven. Omdat de omstandigheden - zowel intern als extern - in de praktijk nooit perfect stabiel zijn, zullen er veranderingen nodig zijn om te compenseren voor alles wat de status quo kan beïnvloeden (5).

Hollnagel gebruikt als voorbeeld het plannen van een reis om het managen van cyberweerbaarheid te beschrijven. Deze (reis)metafoor is handig, omdat het helpt om te kunnen controleren hoe iets beweegt en van positie verandert of de reis nu fysiek of abstract is en of het onderwerp tastbaar of ontastbaar is. De metafoor is ook nuttig, omdat het wijst op de noodzaak van drie verschillende soorten kennis. Het is noodzakelijk om te weten:

1. wat de huidige positie is,
 2. wat het doel is, en
 3. welke middelen nodig zijn, d.w.z. welke veranderingen nodig zijn en hoe deze te maken om in de richting van het doel te bewegen (5).
- Ik licht dit in het vervolg van dit artikel toe.

Het bepalen van de huidige positie

Een belangrijke eerste stap in de cyberweerbaarheidsopgave waar bedrijven voor staan is het meten van hun huidige cyberweerbaarheid. Zonder goede metingen is het lastig te bepalen of maatregelen de cyberweerbaarheid verbeteren of juist verminderen. De CSR pleit in een beleidsreactie van 25 november 2024 voor een jaarlijkse meting. Huidige metingen zoals de Cybersecuritymonitor van het CBS meten vooral cybercrime-gerelateerde incidenten en de maatregelen die getroffen worden om deze incidenten te voorkomen. Dit zijn uitdrukkelijk geen metingen van cyberweerbaarheid. Maar eerder van cybersecurity. Hoe moet het dan wel?

Eerder refereerde ik echter al aan de opvatting dat weerbaarheid een emergent fenomeen is. Van Harmelen (4) stelt dat weerbaarheid niet refereert aan een vaste eigenschap die je kan meten zoals temperatuur met een thermometer, of gewicht met een weegschaal. Weerbaarheid refereert aan het proces van dynamische adaptatie aan een schokgebeurtenis. Dat betekent dat we weerbaarheid alleen kunnen vaststellen uit de reactie op de schokgebeurtenis. Een goede

meetmethode dient derhalve in staat te zijn om de dynamische interacties over het verloop van tijd in kaart te brengen.

Bepalen van het doel

Het Cybersecuritybeeld (CSBN) uit 2021 van de NCTV kopt 'cyberaanvallen tasten zenuwstelsel maatschappij aan' en stelt dat cyberweerbaarheid in onze maatschappij nog onvoldoende is, waarbij een groot verschil tussen bedrijven wordt geconstateerd in de mate hiervan. Het CSBN spreekt in dit verband over een 'cyberweerbaarheidskloof'. Ook recente versies van het CSBN bevestigen dit beeld. Het CSBN uit 2023 stelt bijvoorbeeld: 'Er gaapt een groeiende kloof tussen organisaties die de cyberveiligheid op orde hebben en organisaties die dat niet hebben.' Het CSBN uit 2023 constateert daarbij dat ondanks beste bedoelingen deze 'achterblijvers' niet altijd actie ondernemen in het verhogen van hun huidige cyberweerbaarheidsniveau naar een optimaal niveau, passend bij het risicoprofiel van het bedrijf. De CSR stelt in de hierboven genoemde beleidsreactie dat vooral in het mkb relatief veel achterblijvers zijn.

Deloitte (10) beschrijft de cyberweerbaarheidskloof aan de hand van het huidige cyberweerbaarheidsniveau versus het optimale cyberweerbaarheidsniveau van een bedrijf. De huidige cyberweerbaarheid van een bedrijf is het vertrekpunt. De optimale cyberweerbaarheid is het beoogde doel dat nodig is om vastgestelde risico's afdoende te mitigeren en digitaal weerbaar te zijn. Het totaalpakket aan risico's (het risicoprofiel) wordt onder andere bepaald door de aard van een bedrijf, de sector en de keten waarin het actief is; de digitale infrastructuur, de gevoeligheid van data en het dreigingslandschap. Het optimale cyberweerbaarheidsniveau is dus afhankelijk van het risicoprofiel en verschilt daarom ook per bedrijf. Bedrijven waarbij het huidige niveau van cyberweerbaarheid ook het optimale niveau van cyberweerbaarheid betreft (of daar heel dichtbij zit) hebben hun cyberrisico's onder controle, terwijl er ook bedrijven zijn waarbij het huidige cyberweerbaarheidsniveau ver achterblijft bij het optimale niveau. Voor deze laatste groep betekent dit dat bedrijven:

- de huidige en/of optimale cyberweerbaarheidsniveaus niet kennen of
- deze niveaus wel kennen, maar er bewust niet voor kiezen of onvoldoende in slagen de cyberweerbaarheid te versterken.

Deloitte stelt daarmee dat het optimale cyberweerbaarheidsniveau afhankelijk is van het type bedrijf. Een bakker hoeft in absolute termen niet even cyberweerbaar te zijn als een internationaal technologiebedrijf in de medische sector. Als bedrijven onderling worden vergeleken, is dus het verschil tussen het optimale en huidige niveau van belang. Dit verschil definieert dus of een bedrijf voorloopt of achterblijft in vergelijking met andere bedrijven. Een bedrijf dat vooroploopt hoeft dus niet per se de meeste cyberbeveiligingsmaatregelen te hebben geïmplementeerd.

Welke middelen nodig zijn

Het mag duidelijk zijn dat de cyberweerbaarheid van een systeem wordt beïnvloed door vele verschillende factoren, vaak op een complexe en tegenstrijdige manier. Het verkleinen van het verschil tussen het optimale en het huidige niveau van weerbaarheid vereist dat deze factoren zorgvuldig worden beheerd of benut. Tegelijkertijd verschillen systemen onderling, waardoor het niet mogelijk is om een uniforme set maatregelen aan te bieden om de cyberweerbaarheid te verbeteren (3). De bakker uit het voorbeeld hierboven heeft bijvoorbeeld vooral behoefte aan een beveiligd kassasysteem, terwijl een bedrijf in de gezondheidszorg vooral behoefte heeft aan begin-tot-eindversleuteling van patiëntgegevens en gesegmenteerde netwerken. Meer over welke middelen nodig zijn, kunt u vinden in de uitgeschreven tekst van mijn rede (11).

Tot besluit

Een paradigmaverschuiving is nodig in ons denken over cyberveiligheid. Het huidige denkkader dat we cybersecurity noemen is steeds minder afdoende in onze aanpak van cyberrisico's. Cybersecurity richt zich op het versterken van kwetsbare onderdelen van het bedrijf om voorspelbare dreigingen te kunnen weerstaan. Maar dat is, zoals ik heb laten zien in deze bijdrage, allang niet meer passend. Geopolitieke spanningen, verschuivende internationale machtsverhoudingen en technologische innovaties zorgen voor een toename van complexe en onvoorspelbare dreigingen. Bedrijven worden daardoor steeds kwetsbaarder. Een doorbraak naar een nieuw paradigma is nodig. Het denkkader dat bedrijven weerbaar moeten zijn als reactie op cyberrisico's is de afgelopen jaren opgekomen en populair geworden. We noemen deze nieuwe kijk op de aanpak van cyberrisico's ook wel

We noemen deze nieuwe kijk op de aanpak van cyberrisico's ook wel cyberweerbaarheid.

cyberweerbaarheid. Het managen van cyberweerbaarheid omvat niet alleen een aanpak gericht op het voorkomen van schade. Het omvat ook het vermogen om onverwachte, positieve mogelijkheden te benutten om de veiligheid te verbeteren en te kunnen reageren op, te herstellen van en zich aan te passen aan ongunstige gebeurtenissen of activiteiten. De nadruk bij weerbaarheid ligt op het versterken van de adaptieve capaciteiten van het systeem, zodat zij zich onder wisselende omstandigheden kan aanpassen, in plaats van enkel negatieve uitkomsten te verminderen (12).

Praktijkgericht onderzoek vervult een belangrijke rol in de doorbraak naar dit nieuwe paradigma, omdat deze doorgaans wordt afgedwongen doordat bepaalde nieuwe wetenschappelijke inzichten steeds opnieuw proefondervindelijk worden gestaafd. We mogen niet uit het oog verliezen dat de belangrijke vragen zijn: wat werkt daadwerkelijk in de praktijk, onder welke omstandigheden – en hoe kunnen we dit empirisch vaststellen? Praktijkgericht onderzoek, gebaseerd op data en onderzoek in plaats van op anekdotes, zorgt voor oplossingen die direct toepasbaar zijn en aansluiten bij de behoeften van de praktijk. Voor oplossingen die wel werken (13). Het is naar mijn mening dan ook doorlopend nodig. Zodoende creëren we een doorbraak naar het nieuwe paradigma van cyberweerbaarheid en een situatie waar het werkveld voldoende geïnformeerd is om de juiste besluiten te nemen over het toepassen van de laatste stand van kennis in hun beveiligingspraktijk.

Referenties

- (1) World Economic Forum. (2024). Unpacking Cyber Resilience (White paper).
- (2) Roelofsma, P. (2024). Cascade Cyber Risk Management: Between Rule and Reality (Lectorale rede). Haagse Hogeschool.
- (3) Eken, M., Lucas, R., Hoorens, S., & Nederveen, F. (2024). Strengthening societal resilience. A perspective on strategic priorities for the Netherlands amidst geopolitical tensions. RAND Europe.
- (4) Van Harmelen, A-L. (2022). Veerkracht bestaat niet (Inaugurale rede). Universiteit Leiden.
- (5) Hollnagel, E. (2022). Systemic potentials for resilient performance. In F. Matos, P. M. Selig & E. Henriqson (Eds.), Resilience in a digital age: Global challenges in organisations and society (pp. 7-17). Cham: Springer International Publishing.
- (6) Van Boheemen, P., Munnichs, G., Kool, L., Diercks, G., Hamer, J., & Vos, A. (2020). Cyberweerbaar met nieuwe technologie: Kans en noodzaak van digitale innovatie. Den Haag: Rathenau Instituut.
- (7) Sarker, P., & Lester, H. D. (2019). Post-disaster recovery associations of power systems dependent critical infrastructures. *Infrastructures*, 4(2), 30.
- (8) Grøtan, T. O., Antonsen, S., & Haavik, T. K. (2022). Cyber resilience: a pre-understanding for an abductive research agenda. In F. Matos, P. M. Selig & E. Henriqson (Eds.), Resilience in a Digital Age: Global Challenges in Organisations and Society (pp. 205-229). Cham: Springer International Publishing.
- (9) Niinistö, S. (2024). Safer Together: Strengthening Europe's Civilian and Military Preparedness and Readiness. European Commission.
- (10) Deloitte. (2024). Cyberweerbaarheidskloof. Aanbevelingen voor een cyberweerbaar mkb en het verkleinen van de cyberweerbaarheidskloof in Nederland. Amsterdam: Deloitte.
- (11) Van der Kleij, R. (2025). Nu veiligheid er niet meer is: Naar een aanpak van cyberrisico's die wel werkt (Lectorale rede). Avans Hogeschool.
- (12) Ebert, N., Schaltegger, T., Ambuehl, B., Schöni, L., Zimmermann, V., & Knieps, M. (2023). Learning from safety science: A way forward for studying cybersecurity incidents in organizations. *Computers & Security*, 103435.
- (13) Leukfeldt, E. R. (2024). Meer dan alleen een goed idee: Naar een empirisch onderbouwde aanpak van cybercrime (Inaugurale rede). Universiteit Leiden.

Auteurs: Swarna Kumarswamy-Das is a cybersecurity researcher at TNO, with an expertise in software security and (autonomous) cyber resilience. She is also involved in Horizon Europe projects such as eFORT, with a focus on OT security. Ruggero Montalto is Projectmanager Cyber Security Technologies ook werkzaam bij TNO en hij is bereikbaar via: ruggero.montalto@tno.nl



PCSI Open-Source Software Libraries Security (OSSLS)

A collaborative effort to deploy cost-effective fuzzing in software delivery pipelines

The Partnership for Cybersecurity Innovation (PCSI) is a public-private partnership among ABN AMRO, Achmea, Belastingdienst, ING, and TNO. PCSI partners work together on innovative cybersecurity solutions that can be used to protect Dutch companies and organizations against tomorrow's cyberattacks.

PCSI-OSSLS (1) was an innovation project which ran from 2022 to 2024. The project team was composed by experts from ASML, ING, Achmea and TNO. The project addressed one of the PCSI security radar themes identified as: "increasing dependency on Open-Source Software Libraries (OSSs)". Early-stage discovery (shift-left) of exploitable vulnerabilities reduces security risks in systems and applications. PCSI-OSSLS combines security testing with regular functional software testing to improve code quality with increased coverage, and thus also increasing the security of software applications against cyberattacks that leverage vulnerabilities in OSSs.

The business drivers behind our solution

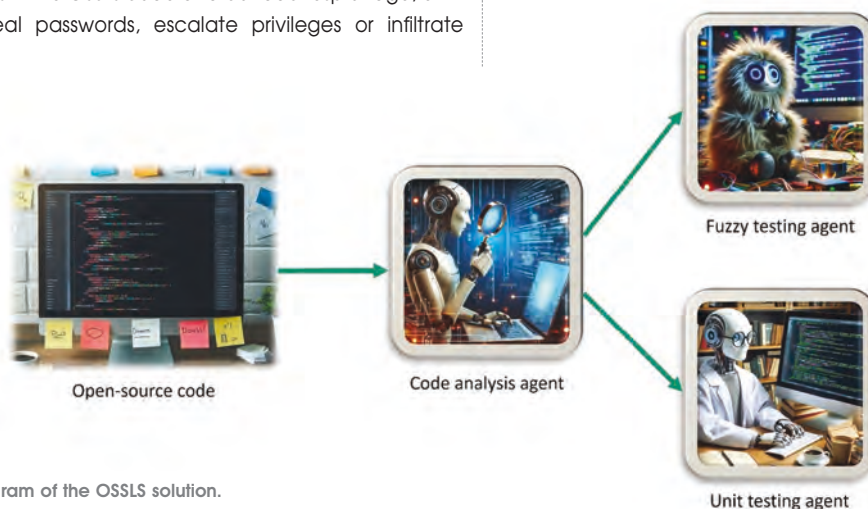
In the current state of practice, use of OSSs is a precondition for market competitiveness. This has created an increased dependency on OSSs and hence an increased risk of cyberattacks. Threat actors can leverage (or actively introduce) vulnerabilities in the source code of OSSs that may remain undetected and therefore unpatched. Threat actors can then break into systems using these flaws in the OSSs code and conduct espionage, exfiltrate data, steal passwords, escalate privileges or infiltrate networks.

Cyberattacks can hinder business and operations considerably, by heavily impacting an organisation financially (costs), legally, and reputationally. As a way of mitigating risks introduced by the adoption of OSSs, organizations can improve code quality and code security by deploying effective software testing techniques. PCSI-OSSLS delivers an innovative solution aimed at DevOps who regularly use OSSs in their applications.

How we developed our solution

In PCSI-OSSLS we integrated automated software security testing techniques into existing development pipelines to help improve code quality and security, increase robustness and reliability of software applications and systems. PCSI-OSSLS uses advanced testing techniques such as fuzzing, that help in discovering unknown vulnerabilities in software code, thereby increasing reliability on OSSs.

Along with popular software testing techniques such as SAST (static application security testing), SCA (software composition analysis) and IAST (interactive application security testing), fuzzing



Process flow diagram of the OSSLS solution.

While developing our solution, we have gained a good understanding about an array of software testing techniques including fuzzing

is used for security testing of open-source libraries by companies within their applications and systems. Fuzzing provides a good overall picture of the quality of the target system and software. By using fuzzing, you can improve robustness and security posture of the system and software under test. One of the barriers to a larger-scale implementation of fuzzing is the costly human work needed in preparing specific code used for fuzzing tests called "harnesses". In PCSI-OSSLS, we leverage the power of generative AI, to automatically generate these harnesses required for fuzzing, considerably reducing manual efforts and hence making the deployment of fuzzing in pipelines both more accessible and cost-effective.

In a current experiment a concept of AI agents is used. The AI agent tries to encapsulate and replace a human specific activity, which is a highly specialized component that uses a mixture of generative AI and traditional programming techniques.

Method:

1. Code analysis agent analyses input from data structures and prepares specifications
2. Fuzzy testing agent uses the results generated in the previous step as input to generate a fuzzing harness
3. Another agent (not depicted in picture) is responsible for compilation, receiving feedback and fixing potential compilation errors. This is also the agent used for code coverage and quality

The outcome is a network of AI agents that automate tasks using generative AI, while traditional programming handles, their internal logic and external communication.

Envisioned roadmap and future works

Experiments conducted so far have involved small and simple open-source libraries. Our main goal here was to keep the process automated which helped to incorporate into software delivery pipelines. In the next phase, we plan to use bigger

libraries and to fuzz entire business applications. The final step will be to automatically remove all unused functionalities from open-source libraries to reduce and limit the potential attack surface. All data which is produced from the agents (as described above), will have significant impact on helping in securing applications thereby improving the security posture of any organization where (continuous) software delivery is relevant to (core) business continuity.

While developing our solution, we have gained a good understanding about an array of software testing techniques including fuzzing. These valuable lessons further help us to develop implementation guidelines, which can be used by organisations within their software development pipelines. It is important to involve the right people and functions within each organisation at every stage, such as developers, testers, security experts, product owners, risk and change management experts. This will ensure a successful and thorough adoption of our solution.

We are looking into implementing our solution at different PCSI partners and beyond. This will give us further insights on how to overcome challenges when applying different software testing techniques within organisations for increasing security, resilience and robustness. From the lessons learnt, we continue to update our implementation guidelines. Finally, we intend to continue researching into areas such as use of generative AI for clustering of results as obtained from fuzzing. This will further help in lowering barriers to include fuzzing into testing pipelines.

References and External links

(1) <https://pcsi.nl/nl/projecten/open-source-software-libraries-security-threat-intel-sharing/>

(PCSI Partnership for Cyber Security Innovation: <https://pcsi.nl/nl/>)

(Software Security Testing Techniques and Tools (tno.nl): <https://repository.tno.nl/SingleDoc?docId=59416>)

(PCSI security radar: <https://pcsi.nl/en/cyber-security-radar-en/>)



Dr. Nicole van der Meulen is expert op het gebied van cybersecurity en emerging technologies en werkzaam bij SURF als Cybersecurity Innovation Lead, ns.vandermeulen@gmail.com

Omarm het ongemak

Binnen de cybersecuritywereld houden we ervan te denken dat wij uniek zijn. Alsof onze uitdagingen alleen te maken hebben met de complexe wereld van cyberbeveiliging. Maar de waarheid is dat ik me daar steeds meer aan begin te ergeren. Begrippen als complexiteit worden te vaak misbruikt als excuus, als een manier om te verhullen dat we simpelweg niet goed omgaan met ongemakkelijke waarheden.

In april gaf ik een keynote in Praag over deze *inconvenient truths* in de cybersecuritywereld. Toen ik de titel koos, moest ik niet alleen aan de beroemde documentaire van Al Gore denken, maar ook aan mijn eigen ervaring. Ooit woog ik boven de 110 kilo. Mijn BMI was veel te hoog. Er moest iets veranderen.

Hoewel ik pas 20 jaar oud was en geen directe gezondheidsklachten had, wist ik dondersgoed dat ik ongezond was. Ik begon gestaag met afvallen. Lange tijd bleef ik rond de 90 kilo hangen, totdat ik er serieus werk van ging maken. De sportschool werd geen optie meer, maar een vast onderdeel van mijn routine.

Toen ik terugdacht aan deze persoonlijke reis, zag ik opvallende parallellen met cybersecurity. Mijn routine om gezond en op gewicht te blijven, brengt bepaalde ongemakken met zich mee. Ik moet concessies doen op het gebied van eten en tijd. Cybersecurity is in wezen ook een ongemak. Nagenoeg alle maatregelen die we nemen kosten tijd, geld en maken dingen 'onhandiger'. Dat is nu eenmaal zo.

Maar in plaats van daartegen te strijden of in ontkenning te schieten, moeten we deze *inconvenient truths* omarmen. Ja, security is een gedoe — maar het is een gedoe dat we belangrijk vinden, omdat het ons beschermt tegen grotere problemen. Security heeft een intrinsieke waarde, juist omdat de voordelen ervan vaak onzichtbaar en niet tastbaar zijn.

Een andere *inconvenient truth* is dan ook dat deze onzichtbaarheid ons soms in de weg zit. Terwijl ik bij het afvallen visuele vooruitgang zag, blijft veiligheid zelf vaak verborgen. Onveiligheid daarentegen is altijd pijnlijk zichtbaar.

Waar ik naartoe wil: laten we stoppen met het herhalen van clichés over complexiteit en een steeds veranderende werkelijkheid. Door daarop te blijven focussen, blijven we hangen. In plaats daarvan moeten we ons richten op de noodzaak van noeste arbeid binnen security. Het draait om consistentie en continuïteit. Misschien weinig charmant of het podium waard, maar wel effectief.

Heel hard rennen na een incident om daarna een jaar lang achterover te leunen, helpt niemand. Dat is een soort crashdieet: tijdelijk lijkt het effectief, maar uiteindelijk richt het meer schade aan dan het voorkomt.

We hebben misschien geen volledige controle over onze omgeving of externe dreigingen, maar we kunnen wel besluiten om security te prioriteren — ondanks de ongemakken. Het is tijd om de dagelijkse, saaie handelingen te omarmen die ervoor zorgen dat cyberweerbaarheid geen illusie blijft, maar een duurzame realiteit wordt.

Auteur: Dr. Jonathan Kwik is een onderzoeker internationaal recht werkende bij het T.M.C. Asserinstituut en hij is bereikbaar via: j.kwik@asser.nl



Het toekomstige oorlogsfrent

Uw netwerk, uw (sociaal) leven thuis en op straat

De aanloop naar de NAVO-top van juni 2025 was een spannende tijd. Vooral overheidsinstanties en veiligheidsdiensten maakten zich zorgen om cyberaanvallen en cyberspionage door vijandelijke staten. Toch klonken ook stemmen die -terecht- de aandacht trokken naar de veiligheid van privé-actoren: 'Niet alleen overheden of vitale organisaties lopen risico; via keteneffecten kan in feite ieder bedrijf doelwit worden. (1) Zo werd gewaarschuwd voor cybercriminelen die van de afleiding gebruik zouden kunnen maken om zich op zwakkere doelwitten te richten.

Maar de dreiging komt niet slechts vanuit cybercriminelen. De civiele sector wordt ook een steeds aantrekkelijker doelwit voor spionage en manipulatie door staten. Zo voert Rusland al sinds het begin van de oorlog in 2022 aanvallen uit tegen Europese bedrijven betrokken bij steun aan Oekraïne. Echter, in de toekomst zal het niet blijven bij privébedrijven waarvan de militaire en politieke samenhang voor de hand ligt. Recente ontwikkelingen rond het gebruik van kunstmatige intelligentie (AI) – zowel op het slagveld als in de informatieruimte – suggereren dat het heel moeilijk gaat worden voor bedrijven om uit het vizier te blijven van vijandelijke statelijke actoren, ongeacht de branche waarin het bedrijf zich bevindt. In dit artikel ga ik het hebben over AI, defensie, moderne oorlogsvoering en de burgerbedrijven die ertussenin knel zitten. Daarbij is een korte introductie handig over hoe AI en opkomende technologieën het moderne slagveld transformeren. Vervolgens nemen we twee casussen, anti-AI tegenmaatregelen en gerichte cognitieve oorlogsvoering, ter illustratie van deze onontkoombare verstrengeling tussen defensie en civiel.

AI rukt op binnen defensie

AI kan praktisch niet meer ontkoppeld worden van het moderne slagveld. Een kleine vier jaar geleden werd nog soms bezorgd gespeculeerd over een mogelijke 'AI-wapenwedloop' en hoe dat voorkomen zou kunnen worden. Dat station is inmiddels gepasseerd. AI wordt niet alleen grootschalig ingezet, maar ook voor een ongelofelijk brede variatie in taken. Zowel Oekraïense als Russische drones maken gebruik van AI-aangedreven beeldclassificatie om vijandige soldaten en tanks te herkennen, en

tijdens de India-Pakistancrisis in april werd de Harop-drone gebruikt om autonoom radarinstallaties te herkennen en te vernietigen. AI wordt ook steeds vaker ingezet bij planning en besluitvorming, zoals bij de organisatie van logistiek, het prioriteren van doelwitten en het vormen van operatieplannen. Behalve op het fysieke toneel, maakt AI ook veel mogelijk op het cognitieve slagveld. Zo gebruikt Oekraïne software die automatisch posts op sociale media bijhoudt en trends analyseert om te bepalen hoe kinetische aanvallen impact hebben op het Russische moreel (zogenoemd sentiment analysis software) (2). AI beperkt zich niet tot het verzamelen en analyseren van data: AI kan ook creëren. De opmars van Generatieve AI, waarvan ChatGPT waarschijnlijk het bekendste voorbeeld is, is niet onopgemerkt gebleven bij strijdmachten. Nederland is bezig een eigen alternatief te maken, genaamd DefGPT, voor exclusief gebruik door Defensie. Generatieve AI wordt ook steeds meer gebruikt voor het maken van deepfakes en desinformatie, een cruciaal element in de moderne informatie-oorlogsvoering.

Uitgehongerd

Binnen militaire en strategische kringen is AI dus een Zwitsers zakmes, een allemansvriend. Als er voor een bepaalde taak een effectief AI-programma ontwikkeld kan worden, is het in hun voordeel dit te doen. Maar wat heeft dit te maken met bedrijven, die wellicht niets met defensiezaken te maken (willen) hebben? De nadruk hierboven lag op 'effectief' en 'kunnen' zelfs voor de best gefinancierde defensie-organisaties, met alle capaciteiten tot hun beschikking, is het namelijk niet altijd gemakkelijk om goed-functionerende AI-systemen te ontwikkelen. Het grootste obstakel waar defensie-organisaties mee worstelen

is data. Moderne AI-systemen zijn datagedreven: dat wil zeggen dat hun 'intelligentie' afhangt van de kwaliteit en kwantiteit van de data, die wordt gebruikt tijdens het trainen van de modellen. En laat kwaliteit en kwantiteit nu juist dun gezaaid zijn als het gaat om militaire data. Zo zijn er op internet nu eenmaal veel minder hoge kwaliteit foto's te vinden van Russische tanks, dan van katten.

Het 'grijpen van wat maar te grijpen valt' via een zogenaamde *dragnet* wordt dan een aantrekkelijke oplossing. Hierbij moet ook de kanttekening gemaakt worden dat 'militair-relevant' niet per se hoeft te betekenen 'militair-gerelateerd'. Gegevens van bedrijven en instituties die op eerste blik niets te maken hebben met defensiezaken kunnen evengoed interessant zijn. Bijvoorbeeld het Israëlisch leger is verwoed datasets aan het verzamelen van tekst en alledaagse gesprekken in het Palestijns-Arabisch dialect, om een verbeterd AI-surveillancesysteem te ontwikkelen voor de bezette Palestijnse gebieden (3).

In sommige gevallen wordt meer 'geleend' van de privésector dan slechts datasets. De modernste AI-tools zoals fundatiemodellen (waaronder ChatGPT) zijn praktisch niet meer in-house te ontwikkelen. GPT-5 van OpenAI telt bijvoorbeeld tientallen triljoenen (= 19 nullen!) parameters. Weinig organisaties ter wereld hebben toegang tot zo'n hoeveelheid data en zelfs als ze dit wél hadden, kost het trainen van zulke kolossale AI-modellen duizelingwekkend veel energie en tijd. Wat organisaties ervoor in de plaats doen, is bestaande (door privé-actoren getrainde) modellen nemen en deze ombouwen. Dit is ook wat Defensie doet met bovengenoemde DefGPT: 'Gelukkig hebben we DefGPT niet zelf vanaf nul hoeven op te bouwen. Eigenlijk is dat vrijwel ondoenlijk, omdat je gigantische hoeveelheden data nodig hebt om generatieve AI überhaupt enigszins te laten werken. We hebben daarom gebruikgemaakt van een opensourcemodel' (4).

Defensieorganisaties kijken dus met steeds meer interesse naar data buiten traditioneel-militaire kaders. Data die gegenereerd wordt in civiele contexten, of die opgeslagen wordt in servers van civiele bedrijven, is onmisbaar geworden bij het trainen van nieuwe AI-modellen. En omdat het vanaf nul trainen van fundatiemodellen onhaalbaar is, worden civiele fundatiemodellen (waar uw en mijn data waarschijnlijk ook al in verwerkt zit) hertraind. Ten slotte is civiel-gegenereerde data ook belangrijk in het gebruik van de AI-systemen. Oekraïne's sentiment analysis-algoritme en Israël's surveillancesysteem analyseren tijdens hun inzet telefoon- en sociale mediagesprekken in real-time.

Het idee dat civiele data gebruikt wordt voor militaire doeleinden, is voor sommigen wellicht al onaangenaam genoeg. Nog verontrustender is dat vijandelijke actoren veel extra incentives hebben om civiele bedrijven als doelwit te nemen. In plaats van defensie-installaties en -servers direct aan te vallen (die vaak veel beter beveiligd zijn), bieden civiele bedrijven en netwerken nieuwe en toegankelijke aanvalsvectoren voor vijandelijke staten. Laten we hier twee representatieve casussen van nemen.

De hallucinerende drone

Wat kun je doen als jouw soldaten op het slagveld, zelfs in het donker en met *camo-gear*, met verbijsterende accuraatheid worden gespot door vijandelijke AI-drones en meedogenloos opgejaagd en neergeschoten worden? Anti-drone-afweer (vaak ook AI-aangedreven) lijkt in het huidige Russisch-Oekraïense conflict één van de populairdere antwoorden te zijn, maar een wat meer technologisch vernuftige tegenmaatregel is *Counter-AI* (CAI). Dit zijn technieken die misbruik maken van kwetsbaarheden van AI (zoals het feit dat ze geen gezond verstand bezitten en slechts statistische associaties/voorspellingen maken) om de AI-systemen onschadelijk te maken, of zelfs tegen hun eigenaren op te zetten. Een populair voorbeeld hiervan is de 'adversarial example'. Zo kan je een T-shirt met een QR-code dragen om jezelf onzichtbaar te maken voor gezichtsherkenningssystemen (5) of banden plaatsen op de vleugels van je gevechtsvliegtuig om overvliegende drones in de war te brengen.

De vergiftigingsaanval is een CAI-techniek die wat meer voorbereiding vereist, maar ook veel beter uitbetaalt. Een vergiftigingsaanval komt neer op het compromitteren van een AI-systeem terwijl het nog in zijn trainingsfase zit. Door de 'jonge' AI foute dingen te leren, kan de aanvaller daar later profijt uit halen. Via een vergiftigingsaanval kan een vijandelijke staat bijvoorbeeld een aanvalsdrones 'leren' om burgers als vijandige soldaten te herkennen of een AI-adviseur constant adviezen te laten genereren die zijn gebruiker wil horen (zogenaamde 'kontlikkende AI'), in plaats van wat zijn gebruiker moet horen (6).

Het gevaar voor de civiele sector ontstaat door het feit dat het manipuleren van datasets dé manier is voor de vijandelijke actor om een vergiftigingsaanval uit te voeren. Foto's van alledaagse burgers kunnen bijvoorbeeld met de label 'Russische soldaat' voorzien worden. Hiermee bereikt de aanvaller dan inderdaad zijn doel om de AI-drones van de kaart te brengen, maar datasetmanipulaties maken geen onderscheid: dezelfde (foutief-

gelabelde of gemanipuleerde) data kan evengoed door een civiel bedrijf gebruikt worden om hun eigen AI-modellen te trainen, met onverwachte (en mogelijk sociaal en economisch schadelijke) gevolgen.

Breindata voor het grijpen

'Cognitieve oorlogsvoering' is een relatief recent concept in NAVO-cirkels, maar bestaat al eeuwen in de vorm van propaganda en psychologische oorlogsvoering. Het verschil tussen oudere informatie-operaties en cognitieve oorlogsvoering ligt in de context: moderne technologie verbreedt aanzienlijk de mogelijkheden voor informatieverzameling en -verspreiding. Het digitale advertentiesysteem is hier een illustratieve microkosmos van. Commerciële bedrijven verzamelen onze clicks, zoekopdrachten en reacties op sociale media om zo goed mogelijk producten te kunnen adverteren. Deze advertenties zijn gericht, op maat gemaakt en worden direct 'geleverd' aan het doelwit via smartphones en browsers, die vrijwel iedereen nu tot zijn beschikking heeft.

Gezien het succes van dit model in de advertentieruimte, moet het niet als verrassing komen dat ook statelijke actoren hier kansen ruiken voor militaire, strategische of geopolitieke doeleinden. Fysieke oorlogen kunnen immers op het mentale slagveld gewonnen worden – bijvoorbeeld door de bevolking van een land te overtuigen dat doorvechten zinloos is. Rusland is hier enorm sterk in: al jaren gebruiken ze sociale media, traditionele media en online platforms om de rechtstaat in Europa te verzwakken, bevolkingen te polariseren, vertrouwen in overheden te beschadigen en democratische principes te ondermijnen (7).

Nieuwe technologieën zullen de mogelijkheden voor cognitieve aanvallen steeds vergroten. We hebben het al gehad over het gebruik van AI tijdens de 'analysefase' (zoals AI-aangedreven software voor sentimentanalyse) en de 'leveringsfase' (zoals AI-gedreven nepaccounts, die volledige en steeds overtuigendere gesprekken kunnen voeren). Maar cruciaal voor cognitieve oorlogsvoering is ook de 'extractiefase', d.w.z. het verzamelen van geïndividualiseerde data van toekomstige doelwitten. De smartphone biedt al een schat van data over levenspatronen en voorkeuren, maar toenames in het gebruik van internet-of-things (IoT) en andere endpoints zal enorm veel data genereren, via apparaten die vaak slecht beveiligd zijn.

Als we naar de toekomst van endpoints kijken, kunnen we speculeren dat brein-computer interfaces (BCI) een van de aantrekke-

lijkste doelwitten zullen worden voor spionage en extractie. BCI's kunnen zowel invasief als niet-invasief geplaatst worden en lezen onze elektrische breinsignalen. Daarmee kunnen gebruikers de thermostaat, prothetische armen en zelfs drones met hun brein bedienen. Daartegenover staat de rechtstreekse informatiewissel tussen de BCI-sensoren en de elektrische breinsignalen van gebruikers – de heilige graal voor cognitieve oorlogsvoering. Dit directe inzicht in de gedachtenpatronen van doelwitten zal de vijandelijke actor in staat stellen cognitieve aanvallen op maat te fabriceren van ongekende precisie en effectiviteit, die precies inspelen op de zwakheden, cognitieve vertekeningen en vooroordelen van het doelwit (8).

Een gewaarschuwd bedrijf...

Dit artikel heeft inzicht gegeven in de verstrengeling tussen civiele data en militaire doeleinden. In het AI-tijdperk is vrijwel alle data eigenlijk militair-relevant geworden. Dat wil uiteraard niet zeggen dat we allemaal direct al een schietschijf op de rug hebben, maar het is goed te beseffen dat bedrijven en organisaties die zich volledig in het civiele domein wanen, steeds vaker slachtoffer kunnen worden van cyberoperaties door statelijke actoren met veel financiële en technische backing. Cyberveiligheid wordt dus alleen maar crucialer, zowel voor opgeslagen data als voor de endpoints in de handen van de eindgebruiker, zoals IoT en BCI's.

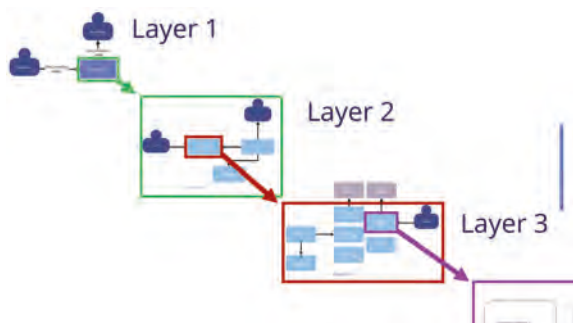
Referenties

- (1) <https://samendigitaalveilig.nl/nieuws/cyberaanvallen-rond-de-navo-top-waarom-de-5-basisprincipes-voor-digitale-weerbaarheid-juist-nu-cruciaal-zijn/#:~:text=Op%2024%20en%2025%20juni,het%20maken%20van%20politieke%20statements.>
- (2) <https://www.economist.com/science-and-technology/2024/04/08/how-ukraine-is-using-ai-to-fight-russia>
- (3) <https://www.theguardian.com/world/2025/mar/06/israel-military-ai-surveillance>
- (4) <https://www.digitaleoverheid.nl/cio-rijk/sneak-peek-cio-cafe-generatieve-artificial-intelligence/>
- (5) <https://arxiv.org/abs/1910.11099>
- (6) <https://jonathankwik.com/books/digital-yes-men-how-to-deal-with-sycophantic-military-ai-2/>
- (7) <https://www.kew.org.pl/en/2024/11/27/destroy-from-within-russias-cognitive-warfare-on-eu-democracy/>
- (8) <https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-HFM-377/MP-HFM-377-09.pdf>

Een complete gids

Bedreigingsmodellering voor beveiligingsdeskundigen, architecten en ontwikkelaars

Bedreigingsmodellering identificeert beveiligingsrisico's om verdedigingsprioriteiten vast te stellen, terwijl het C4-framework deze risico's organiseert in verschillende systeemlagen, van architectuur tot code. Samen vormen ze een allesomvattende, aanpasbare aanpak voor het beveiligen van applicaties.



Deel 1: Inleiding: bedreigingsmodellering en C4

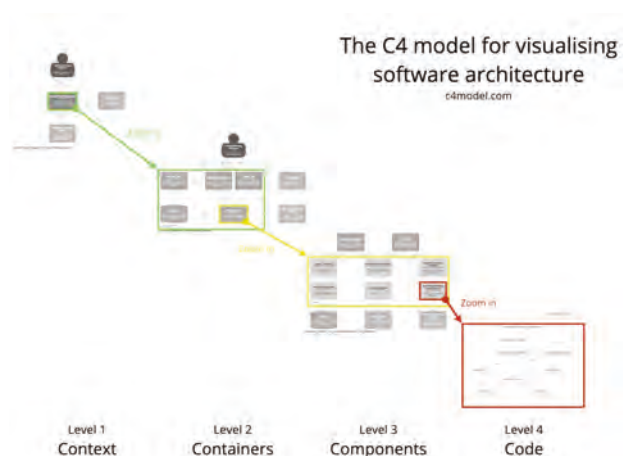
Wat bedreigingsmodellering is en waarom het meest belangrijk is

Bedreigingsmodellering is een manier om beveiligingsrisico's in een systeem, applicatie of omgeving te overwegen. Het gaat om het identificeren van bedreigingen (1), het evalueren hoe ernstig deze problemen kunnen zijn en vervolgens het plannen van verdedigingen om deze bedreigingen op afstand te houden. Door potentiële risico's in een vroeg stadium uit te zoeken, kunnen teams beschermingen aanbrengen voordat die risico's echte problemen worden.

In een notendop helpt bedreigingsmodellering u een duidelijk beeld te krijgen van waar uw applicatie of systeem kwetsbaar kan zijn, zodat u intelligente beveiligingsbeslissingen kunt nemen en een veerkrachtiger systeem kunt bouwen. Het houdt de focus op wat het belangrijkste is, waardoor het eenvoudiger wordt om

prioriteit te geven aan beveiligingsinspanningen en -middelen waar deze de grootste impact zullen hebben.

Inleiding tot het C4-kader



Figuur 2: C4-model van C4model.com

Het C4-raamwerk - een afkorting van **Context**, **Container**, **Component** en **Code** - is een krachtig hulpmiddel om de verschillende lagen van een systeem te organiseren en te visualiseren. Elke laag helpt ons om het systeem vanuit een andere hoek te bekijken, waardoor we potentiële beveiligingsrisico's kunnen ontdekken die anders door de mazen van het net zouden glippen.

De gelaagde aanpak van C4 lijkt een beetje op inzoomen met een camera: we beginnen breed om het grote geheel te zien en gaan dan dichterbij om specifieke details te bekijken. Deze gelaagde aanpak zorgt ervoor dat we zowel de grote risico's op hoog niveau als de gedetailleerde technische kwetsbaarheden binnen het systeem vastleggen.

Lees meer over C4 (2).

Waarom C4 gebruiken?

Wat het C4-raamwerk vooral waardevol maakt, is de flexibiliteit. En omdat we veel ontwikkelen binnen de aviation sector, is het een perfecte methode die zowel door architecten als ontwikkelaars kan worden begrepen. Het werkt zowel vanuit een **top-down als bottom-up** perspectief, wat betekent dat het nuttig is voor **architecten en strategische denkers** die beginnen met de zakelijke impact en de systeemstructuur (de "top-down" benadering), maar ook **voor ontwikkelaars en technici** die direct in de technische details en code duiken (de "bottom-up" benadering).

C4 brengt deze verschillende gezichtspunten samen en geeft alle betrokkenen een gemeenschappelijke structuur om van daaruit te werken. Dus of je nu een architect bent die de bredere doelen van het systeem definieert of een ontwikkelaar die zich verdiept in de technische risico's, met het C4-framework kunt u zien hoe elke laag voortbouwt op de volgende. Uiteindelijk maakt deze afstemming tussen architecturale overzichten op hoog niveau en technische details op grondniveau het eenvoudiger voor teams om ervoor te zorgen dat ze alle beveiligingsaspecten afdekken, zodat het systeem van boven tot onder wordt beschermd.

Moet ik C4 gebruiken?

Nee, je hoeft geen C4 te gebruiken. Je kunt andere standaarden gebruiken om je architectuur of code te modelleren. Bij het modelleren van bedreigingen is het echter goed om de gelaagde aanpak van C4 aan te houden. De verschillende lagen zijn nodig om te weten waar de bedreiging zich bevindt en hoe deze zich verhoudt tot onderdelen op hoger of lager niveau van het hele plaatje. Je wilt code-gerelateerde bedreigingen niet oplossen met grote architectuurbeslissingen noch andersom.

Deel 2: Onze benadering van het proces van dreigingsmodellering

In dit gedeelte zoomen we in op bedreigingsmodellering vanuit een zakelijk perspectief op hoog niveau.

Voor ingenieurs

Deze top-down benadering is perfect voor architecten die beveiligingskwesties willen verbinden met bedrijfsdoelstellingen, beginnend vanuit de grote context en geleidelijk afwerkend naar specifieke kwetsbaarheden in de code (3).

In dit geval begin je met stap 1 en werk je naar beneden naar stap 5. Het is zeer waarschijnlijk dat je de hulp van ontwikkelaars en engineers nodig hebt zodra je voorbij stap 3 of 4 bent, afhankelijk van de case en de context.

Voor ontwikkelaars

Als je een ontwikkelaar of een ingenieur bent, is het misschien beter om bottom-up te werken in plaats van top-down. Begin dan bij stap 5 en werk terug naar stap 1. Rond stap 1 of 2 is het waarschijnlijk dat je wat input van een architect nodig hebt.

Stap 1: Bedrijfsrisico's en doelstellingen definiëren



Figuur 3: voorbeeld voor een fictief bedrijf genaamd "PromotionForce".

Om te beginnen met bedreigingsmodellering, breng in kaart de **belangrijkste bedrijfsrisico's** die van invloed kunnen zijn op het SaaS-platform.

Denk aan zorgen op hoog niveau zoals:

- **Downtime:** Als het platform uitvalt, heeft dit invloed op zowel de inkomsten als de reputatie.
- **Datalekken:** Verlies van gevoelige klantgegevens kan leiden tot financiële schade en reputatieschade.
- **Reputatierisico's:** Alles waardoor klanten het vertrouwen in het platform kunnen verliezen, zoals verkeerde gegevensverwerking of storingen.

Hier definieer je ook wie de **primaire belanghebbenden** zijn (bijv. klanten, interne teams) en verduidelijk je welke rol de applicatie speelt bij het creëren van bedrijfswaarde. Als je deze basiselementen begrijpt, kun je de prioriteiten van je bedreigingsmodellering bepalen.

Stap 2: Contextlaag - Creëer de scène op hoog niveau met gegevensstroomdiagrammen



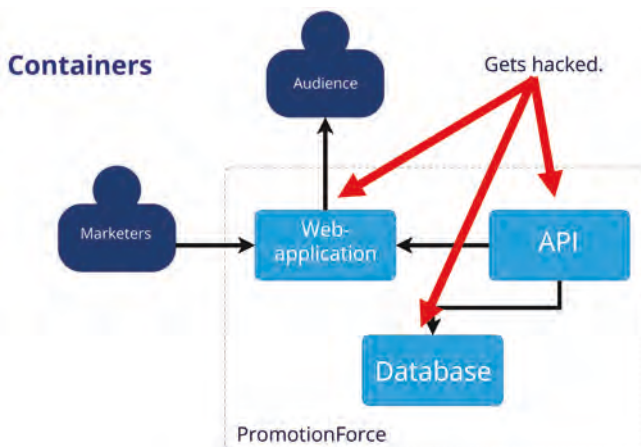
Figuur 4: voorbeeld contextbedreigingen voor "PromotionForce".

Nu zoomen we een beetje in op de **contextlaag**. Hier gaat het allemaal om de manier waarop de applicatie functioneert binnen de grotere bedrijfsomgeving:

- Schets het **doel van de applicatie** en definieer de verschillende **gebruikersrollen en interacties**.
- Breng de **belanghebbenden en workflows** in kaart - wie gebruikt het platform en waarvoor?
- Houd rekening met potentiële bedreigingen op hoog niveau, zoals **ongeautoriseerde toegang (4)** of **misbruik door insiders**.

Deze laag is cruciaal voor het identificeren van de externe factoren en brede risico's die de gehele applicatie beïnvloeden.

Stap 3: Containerlaag: breng de belangrijkste functionele componenten in kaart



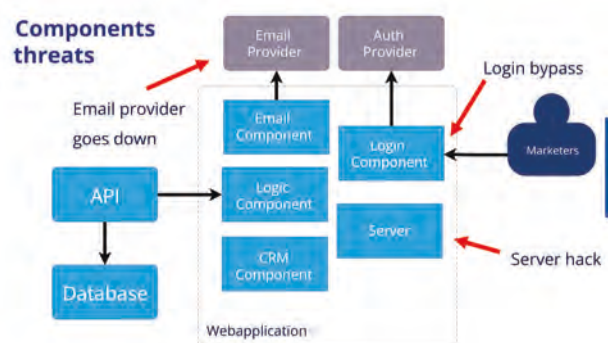
Figuur 5: voorbeeld containerbedreigingen voor "PromotionForce".

Laten we vervolgens naar de **containerlaag** gaan. Hier kijken we naar de **belangrijkste functionele containers** waaruit de applicatie bestaat:

- Identificeer kernelementen zoals **webinterfaces**, **API's**, **databases** en andere hoofdcomponenten.
- Schets hoe **gegevens tussen deze containers stromen** om te begrijpen waar gevoelige informatie kwetsbaar kan zijn.
- Het maken van een gegevensstroomdiagram (Data Flow Diagrammen (DFD)) kan helpen bij het visualiseren van de gegevensstroom door het systeem en het identificeren van mogelijke kwetsbaarheden en vertrouwensgrenzen.
- Begin met het beoordelen van **bedreigingen op container-niveau** die van invloed kunnen zijn op het systeem als geheel. Bedenk bijvoorbeeld of de API kan worden blootgesteld aan onbevoegden of dat er een kwetsbaarheid zit in de databaseconfiguratie.

Het doel is hier om potentiële storingspunten te schetsen voor de belangrijkste applicatiecomponenten.

Stap 4: Componentlaag - Subcomponenten analyseren



Figuur 6: voorbeeldcomponenten bedreigingen voor "PromotionForce"

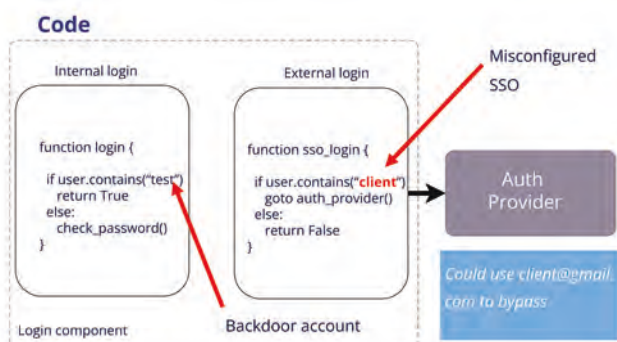
Nu we een idee hebben van de hoofdcontainers, is het tijd om **verder in te zoomen** op de **componentenlaag**:

- Kijk naar specifieke elementen binnen elke container, zoals **authenticatie**, **e-mailverwerking** of **opslagsubsystemen**.
- Het analyseren van deze subcomponenten helpt bij het identificeren van potentiële beveiligingsrisico's binnen elk onderdeel van de applicatie.
- Identificeer component-specifieke bedreigingen die je high-level doelstellingen kunnen beïnvloeden, zoals een **authenticatie omzeiling** of **onveilige e-mail afhandeling**.

In deze laag gaat het om het vinden van meer gerichte kwetsbaarheden binnen elk onderdeel van de applicatiestructuur,

waarbij ze worden teruggekoppeld naar bredere systeem- en bedrijfsdoelen.

Stap 5: De coderingslaag - samenwerken aan risico's op codeniveau



Figuur 7: voorbeeld van logische aanmeldingsbedreigingen voor "PromotionForce".

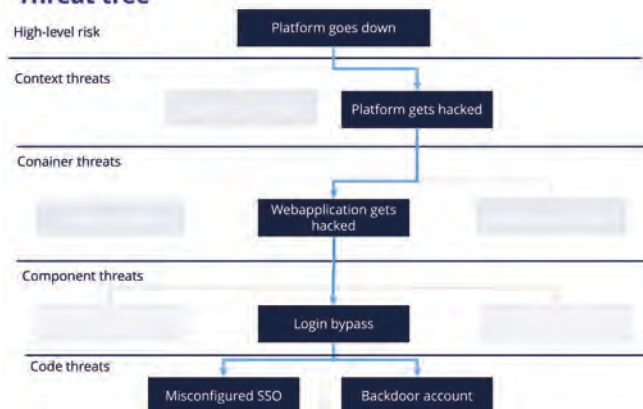
Tot slot komen we bij de **coderingslaag**, waar het tijd is om ons te richten op **specifieke kwetsbaarheden in de code**. Hier is samenwerking met ontwikkelaars en beveiligingsteams van groot belang:

- Samenwerken met het ontwikkelingsteam en de beveiligingsteams om mogelijke problemen in de codebase te identificeren, zoals **hardcoded credentials, onbeveiligde API-sleutels of logische fouten**.
- Zorg ervoor dat de risico's op codeniveau overeenkomen met de bredere risico's die je hebt geïdentificeerd, zodat ontwikkelaars begrijpen welke kwetsbaarheden het belangrijkst zijn om aan te pakken.

Hier is het doel om technische risico's te identificeren die zowel de functionaliteit van de applicatie als de bedrijfsdoelstellingen kunnen beïnvloeden. Het helpt ook bij het overbruggen van de kloof tussen beveiligingsdoelen en praktische implementatie.

Alles samenbinden: maak een bedreigingsboom

Threat tree



Figuur 8: voorbeeld Threat Tree voor "PromotionForce".

Zodra je elke laag hebt doorlopen, is het tijd om alles samen te brengen door een **bedreigingsboom** te creëren. Deze boom is in wezen een visuele kaart die zakelijke risico's op hoog niveau koppelt aan specifieke technische kwetsbaarheden en mogelijke risicobeperkende stappen:

- Begin bovenaan met een risico op hoog niveau (zoals het risico op een datalek).
- Breng vervolgens in kaart hoe dit risico wordt gefilterd door de verschillende lagen - context, container, component en uiteindelijk code.
- Deze boom biedt een gestructureerd overzicht van **waar de kwetsbaarheden liggen** en welke **mitigaties** aansluiten bij uw belangrijkste doelstellingen.

De bedreigingsboom is niet alleen een geweldige visuele samenvatting van het proces van bedreigingsmodellering, maar ook een handig hulpmiddel voor het prioriteren van beveiligingsinvesteringen en het bijhouden van de voortgang in de loop van de tijd. Door de boom af te stemmen op bedrijfsdoelen kunnen architecten een duidelijk en gestructureerd pad creëren om het SaaS-platform van bovenaf te beveiligen.

Deel 3: Perspectieven integreren voor een alomvattend bedreigingsmodel

In dit laatste deel brengen we zowel de high-level als de technische kijk op bedreigingsmodellering samen. Door de top-down- en bottom-up benaderingen te integreren, creëren we een uniform bedreigingsmodel dat zowel bedrijfsrisico's als specifieke technische kwetsbaarheden aanpakt. Deze gecombineerde visie helpt teams om effectiever prioriteiten te stellen voor

Het doel is om een stappenplan op te stellen voor waar beveiligingsinspanningen de grootste impact zullen hebben.

beveiligingsinspanningen en zorgt ervoor dat er geen kritieke gebieden over het hoofd worden gezien.

Top-down en bottom-up bedreigingstress op elkaar afstemmen

Nu we twee perspectieven hebben - één van de architecten gericht op bedrijfsrisico's en één van de ontwikkelaars gericht op technische bedreigingen - is het tijd om deze op één lijn te brengen in één bedreigingsboom door gebruik te maken van bedreigingsmodellen. Bedreigingsmodellen spelen een cruciale rol bij het combineren van bedrijfsrisico's op hoog niveau met technische kwetsbaarheden en zorgen zo voor een allesomvattende aanpak voor het identificeren van bedreigingen en het stellen van prioriteiten.

Dit is hoe je het moet aanpakken:

- **Combineer bedrijfsrisico's op hoog niveau met technische kwetsbaarheden:** begin met de bedrijfsrisico's die door de architecten zijn geïdentificeerd en verbind deze met specifieke technische risico's die door de ontwikkelaars aan het licht zijn gebracht.
- **Maak een kaart die bedrijfsdoelstellingen koppelt aan bedreigingen:** deze boom moet laten zien hoe elk risico op hoog niveau doorsijpelt naar het codeniveau, zodat er een duidelijk pad ontstaat van overkoepelende doelen naar gedetailleerde technische problemen.

Het doel is om een stappenplan op te stellen voor waar beveiligingsinspanningen de grootste impact zullen hebben. Deze gebundelde bedreigingsboom laat niet alleen zien waar de risico's zich bevinden, maar ook hoe elk risico zich verhoudt tot het grotere geheel, waardoor het eenvoudig is om te zien waar beveiligingsinspanningen de meeste waarde zullen hebben.

Call to action

Beste mensen ik vertrouw erop dat u met dit artikel de eerste stap durft te zetten om een bedreigingsmodulering vorm te geven. Mijn suggestie: begin met een 'brainstorm', nodig jouw betrokken medewerkers erbij uit en neem bovenstaande aanbevelingen als een leidraad ter hand. Ik ben benieuwd wie dat doet en wat de ervaringen zijn. Diegene onder u zou ik willen uitdagen/uitnodigen om in een volgend artikel uw ervaring met ons allen te delen en daarmee een voorbeeld voor anderen te zijn. Ik zal met alle plezier mijn positief kritisch gezichtspunten met u delen.

Referenties en Bijlagen

- (1) <https://www.securityscientist.net/blog/14-answers-on-threats/>
- (2) <https://c4model.com/>
- (3) <https://www.securityscientist.net/blog/test-executable-code-to-identify-vulnerabilities-and-verify-compliance-11-practices/>
- (4) <https://www.securityscientist.net/blog/protect-all-forms-of-code-from-unauthorized-access-and-tampering-6-practices/>

Waar informatie over Open Worldwide Application Security Project (OWASP):

- **OWASP Top 10:** Uitgebreide inspiratie voor de beveiliging van web toepassingen. <https://owasp.org/www-project-top-ten/>
- **OWASP API top 10:** Inspiratie voor API-beveiliging <https://owasp.org/www-project-api-security/>
- **MITRE ATT&CK:** Gedetailleerde kennisbank van tactieken en technieken van tegenstanders op basis van waarnemingen uit de echte wereld. <https://attack.mitre.org/>
- **OWASP gidsen voor bedreigingsmodellering:** Extra inspiratie en richtlijnen voor bedreigingsmodellering. https://owasp.org/www-community/Threat_Modeling

Lex Borger is security consultant bij Tesorion en oud-hoofdredacteur van IB Magazine. Lex is bereikbaar via lex.borger@tesorion.nl



Een veerkrachtige enterprise-architectuur vraagt om DEI

Wat organisaties vandaag het meest nodig hebben in hun enterprise architectuur, is veerkracht. De razendsnelle ontwikkeling van AI, veranderende geopolitieke verhoudingen, dominantie van cloudplatformen en strengere wet- en regelgeving (zoals NIS2 en DORA) vormen risico's die traditionele architecturen moeilijk kunnen opvangen. Architectuur is vaak homogeen, statisch en gericht op beheersbaarheid. Eenmaal gekozen voor één leverancier of platform, zit je er meestal lang aan vast. Juist die rigide beheersbaarheid is nu een bedreiging.

Dit zien we aan de dominantie in de cloud. Drie leveranciers uit de VS bepalen hier de markt. In veel sectoren in Nederland is sterk geconvergeerd naar één van deze leveranciers. Ik heb veel respect voor de stappen die deze leveranciers zetten op het gebied van beveiliging en hun pogingen om EU-gegevens binnen de EU te houden. Uiteindelijk is het resultaat dat er weinig Europese spelers zijn, en degenen die er zijn, krijgen moeilijk een voet aan de grond.

We moeten weer wendbaar en weerbaar worden. Dit vraagt om een ander denkkader, gebaseerd op DEI: Diversiteit, Evenwicht en Integratie.

Diversiteit – Een diverse architectuur is in staat om verschillende technologieën, leveranciers en platformen te combineren. Dit beperkt risico's, voorkomt afhankelijkheid en maakt ruimte voor innovatie. Denk aan een organisatie die verschillende SaaS-applicaties gebruikt voor haar kernprocessen, op locatie ook kritieke systemen heeft draaien die zonder internetverbinding functioneel blijven en kan experimenteren met lokale AI-modellen of edge computing voor productieprocessen.

De diversiteit moet wel doordacht zijn. Dit vraagt om het herkennen van **functionele clusters**: kleine, samenhangende eenheden binnen de organisatie (zoals klantenservice, inkoop, of risicobeheer) die hoge interne cohesie en hun eigen veilige (zero-trust) schil hebben. Door elk cluster onafhankelijk aan te pakken en in te vullen, kun je sneller innoveren zonder dat het hele landschap tegelijk moet worden aangepast.

Evenwicht – Door architectuur op te bouwen uit kleinere clusters, wordt elke ontwerpkeuze beheersbaarder. Je kunt continu afwegen: hoe verlagen we kosten, beperken we risico's of verbeteren we prestaties? Dit verhoogt de **operationele balans**, zelfs bij verstoringen. Bijvoorbeeld: als een cloudleverancier faalt, blijft een lokaal cluster met kritieke data operationeel. Als een platform of besturingssysteem een kritieke kwetsbaarheid heeft en direct offline moet, blijft de rest gewoon draaien. Zo voorkom je dat één verstoring de hele keten raakt.

Integratie – Een architectuur gebaseerd op diversiteit en evenwicht vraagt om **naadloze integratie**. Niet standaardisatie van applicaties, maar standaardisatie van interfaces (API's, event streams) vormt de kern. Beheer, toegangsbeveiliging (identiteiten en autorisaties), logging en monitoring moeten geregeld zijn in het cluster en geïntegreerd kunnen worden met de beheer- en beheersomgeving.

AI kan hierin een versneller zijn: het helpt bij het interpreteren van integratiestromen, detecteren van knelpunten en optimaliseren van verbindingen tussen systemen.

Hoe beter je in staat bent om diverse technieken, platformen en processen te integreren met je bedrijfsvoering én je bewijsvoering (compliance), hoe weerbaarder en wendbaarder je architectuur wordt. Veerkracht ontstaat niet vanzelf – het is het resultaat van bewuste ontwerpkeuzes in DEI.

Auteurs: *Hoofdauteur:* Samaa Mohammad-Ulenberg is zelfstandig onafhankelijk expert op digitale ethiek en project-leider. Tot april was zij bestuurslid bij stichting Algorithm Audit en initiator van de Dutch AI Ethics Community. Ze is te bereiken op [Linkedin.com/in/samaam](https://www.linkedin.com/in/samaam). *Co-auteur:* Linda Li is voorzitter van stichting PerspethiX. Zij zet zich in voor verantwoord AI waar de mens centraal staat en is te bereiken op <https://www.linkedin.com/in/linda-li>

AI ontwikkelen of toepassen: red teaming vereist!

Responsible AI gaat ook het security-vak aan

'Wie heeft er gehoord van AI red teaming?' vraagt Samaa aan het publiek tijdens haar presentatie over de raakvlakken tussen security, AI en ethiek op de PVIb-themabijeenkomst over responsible AI in februari 2024, waar de implementatie van de AI Act en de rol van security werd besproken. Er gaan enkele handen aarzelend de lucht in. Dan stelt ze de vervolgvraag: 'Wie heeft er een 'AI red team' binnen de organisatie?'. Geen reactie, het blijft stil.



Dit maakte duidelijk dat er nog te weinig aandacht is voor AI red teaming. En dat is zorgelijk, want zonder AI red teams kunnen kritieke risico's onopgemerkt blijven. Bestaande principes uit Responsible AI en het security vak combineren een interdisciplinair AI red team als een krachtig middel om AI-risico's – zoals datalekken, bias of onbedoelde schadelijke effecten – te identificeren en te mitigeren. De boodschap is helder: elke organisatie die een AI-systeem ontwikkelt of toepast, zou een AI red team moeten hebben.

Security en de AI Act

Sinds 1 augustus 2024 is de Europese AI Verordening (hierna: AI Act) in werking getreden. De AI Act onderscheidt AI-systemen in drie risicocategorieën: verboden praktijken (onder andere voor emotieherkenning op de werkplek en in het onderwijs, social scoring en schadelijke manipulatie), hoog-risico (voor rechts-handhaving, essentiële particuliere en openbare diensten en bepaalde HR-toepassingen) en risico op misleiding (denk aan generatieve AI, chatbots en deepfakes). AI-systemen die een hoog risico vormen voor de gezondheid, veiligheid en fundamentele rechten zullen aan strenge eisen moeten voldoen. De AI Act vereist continu risicomanagement (artikel 9), datakwaliteit en representativiteit (artikel 10), transparantie (artikel 13), menselijk toezicht (artikel 14), technische robuustheid en cybersecurity (artikel 15) en behoorlijk wat meer.

Hier komt de link met security om de hoek kijken. Artikel 15 richt zich op de technische betrouwbaarheid van een AI-systeem. Het beschrijft het belang van technische robuustheid en weerstand tegen aanvallen. AI-systemen moeten niet alleen bestand zijn tegen duidelijke cyberaanvallen om het systeem te hacken of te manipuleren, maar ook tegen, wat minder bekende, AI-specifieke aanvallen. Artikel 15 noemt het risico op manipulatie van de trainingsdataset ("datavervuiling") of van vooraf getrainde componenten die in de training zijn gebruikt ("modelvervuiling"), van input die is gecreëerd om fouten van het model te veroor-

zaken ("vijandige voorbeelden" of "modelontwijking"), van aanvallen om vertrouwelijke gegevens te bemachtigen en van tekortkomingen van het model (zoals "bias"). De AI Act verplicht om technische oplossingen in te voeren die een hoog risico AI-systeem bestand maken tegen deze AI-specifieke kwetsbaarheden.

AI-systemen die niet onder een van de risicocategorieën vallen, worden niet gereguleerd onder de AI Act. Dit betekent echter niet dat deze AI-systemen geen risico kunnen vormen voor de gezondheid, veiligheid of fundamentele rechten van mensen en het milieu; of beïnvloed kunnen worden door externe dreigingen. Het is raadzaam om bij de ontwikkeling of toepassing van elk AI-systeem (en eigenlijk ook bij elk algoritmische, technologische en digitale toepassing) de risico's na te gaan en af te vragen of de ontwikkeling of toepassing van het AI-systeem mag, kan en gewenst is. Dit gaat een stap verder dan slechts de technische haalbaarheid en het volgen van wetgeving: dit gaat om het aanmoedigen van ethische, (maatschappelijk) verantwoorde, deugdelijke en integere AI (offewel: Responsible AI).

Responsible AI en security

AI-systemen zijn niet objectief; mensen hebben veel invloed op de resultaten ervan. Voorbeelden van menselijke keuzes liggen onder andere in het verzamelen en selecteren van data, het kiezen van modellen en features en het identificeren en afwegen van risico's. Onbedoeld kunnen hierdoor vooroordelen en fouten insluipen. Daarnaast zijn de meeste keuzes te herleiden naar waarden. Door deze waarden expliciet te maken, is het mogelijk om de keuzes en de daaruit volgende positieve of negatieve effecten (zoals aantastingen van fundamentele rechten als privacy, non-discriminatie en vrijheid) tegen elkaar af te wegen. Met Responsible AI is het uiteindelijke doel dan ook om AI-systemen te creëren die deugdelijk zijn, een positieve impact hebben en eventuele schadelijke gevolgen kunnen minimaliseren. Responsible AI vereist een holistische en interdisciplinaire benadering waarin technologie, cultuur en governance samen-



Twitter-bot "Microsoft Tay" degradatie van onschuldige naar haatzaaiende bot in 16 uur.)

Onderwerpen die binnen Responsible AI veel aandacht vragen zijn onder meer inclusiviteit, gelijke behandeling, privacy, uitlegbaarheid, veiligheid, technische robuustheid, duurzaamheid en transparantie.

Deze onderwerpen overstijgen vaak de wettelijke verplichtingen en raken aan fundamentele ethische vragen: wat is in deze specifieke context écht het juiste om te doen, ook als de wet niets expliciet voorschrijft?

Traditionele red teaming versus AI red teaming

In het security-vak is 'red teaming' een bekend middel om de verdediging van een organisatie tegen externe dreigingen te testen, risico's te identificeren, aannames te valideren, tekortkomingen bloot te leggen en de algehele beveiliging en infrastructuur te verbeteren. Traditionele red teams proberen systemen aan te vallen zoals een kwaadwillende dat zou doen. Hierbij maken ze gebruik van social engineering, waarbij ze psychologische trucs gebruiken om de mens te misleiden opdat ze het systeem kunnen penetreren.

AI red teams gaan een paar stappen verder. Ten eerste pakken ze ook AI-specifieke risico's aan, zoals de hierboven beschreven voorbeelden. Ten tweede houden ze rekening met de veranderlijkheid van AI-systemen en willekeur van outputs. AI-systemen ontwikkelen zich constant, zowel door de ontwikkelaars als door automatische feedback loops. Bovendien is het in het geval van generatieve AI (genAI) mogelijk dat twee dezelfde inputs verschillende outputs opleveren. Ten derde overlapt AI red teaming met de praktijk van Responsible AI, omdat het nodig is schadelijke uitkomsten met betrekking tot content (zoals desinfor-

matie en verheerlijking van geweld) en eerlijkheid (zoals stereotyperen en discriminatie) te onderzoeken. Door bovengenoemde principes van Responsible AI over te nemen, kunnen AI red teams problemen eerder identificeren om ze adequaat aan te pakken. Tot slot gaan AI red teams ervan uit dat dreigingen zowel kwaadaardig als onopzettelijk kunnen zijn. AI-systemen kunnen ook door onschuldige interacties met reguliere gebruikers op negatieve wijze veranderen. AI red teams vallen aan vanuit het perspectief van niet voor de hand liggende personas om deze kwetsbaarheden te vinden. Ze kijken ook naar social engineering scenario's: hoe kunnen kwaadwillenden gebruik maken van de vooroordelen of zwakke plekken in de menselijke controlelagen?

Een voorbeeld van een onopzettelijke dreiging met een negatief effect, is de onverwachte transformatie die Microsoft Tay in 2016 had gemaakt. Microsoft Tay was een Twitter bot met aanvankelijk vriendelijke Tweets. Als AI-chatbot leerde Tay van de Tweets van gebruikers die ermee interacteerden. Binnen enkele uren interacteerden duizenden Twittergebruikers op ongepaste wijze met Tay, wat ertoe leidde dat Tay zulke reacties ging spiegelen (zie figuur 1). Dit was niet de bedoeling: Tay is zestien uur na de lancering uit de lucht gehaald.

Uit dit voorbeeld kunnen drie lessen voor red teaming worden gehaald.

- ten eerste kan dit soort datavervuiling dus veroorzaakt worden door gebruikers, zonder dat de AI-bot de intentie had dat te doen. Als AI red teams aan de slag gaan, moeten ze ook vanuit allerlei 'personas' op zoek gaan naar manieren om het AI-systeem te kraken.

AI-ontwikkelaars: dit zijn mensen met diepgaande kennis van verschillende AI-technieken.

- ten tweede benadrukt dit weer dat Responsible AI een belangrijk onderdeel is van AI red teaming, want dit helpt om alle mogelijke ethische risico's en negatieve maatschappelijke gevolgen te zoeken en er gericht naar te testen.

- als laatste is het, in het verlengde van de andere twee lessen, belangrijk om vanuit een divers en interdisciplinair team te red team'en, zodat er vanuit zo veel mogelijk perspectieven kan worden getest. Dit vermindert het risico op groepsdenken. Als een groep zodanig wordt beïnvloed door groepsprocessen, wat de kwaliteit van groepsbesluiten vermindert, dan missen ze kritische overwegingen aangaande de feiten en hun blinde vlekken.

Waardevolle deelnemers voor een interdisciplinair team

Een ideaal AI red team komt voort uit een onafhankelijke organisatie. Als dit niet haalbaar is, dan is het nuttig om collega's buiten het team, van het beoogde AI-systeem, te vragen om deze rol te vervullen. Om een AI red team ideaal te vertegenwoordigen, zijn de volgende waardevolle rollen denkbaar:

1) AI-ontwikkelaars: dit zijn mensen met diepgaande kennis van verschillende AI-technieken. Zij begrijpen hoe deze modellen getraind worden, hoe ze reageren op bepaalde input en hoe hun architectuur eventueel misbruikt kan worden. Ze zijn in staat om dreigingen als "vijandige voorbeelden" te ontwerpen en zwakke plekken in de modellen bloot te leggen. Daarnaast zouden zij AI-modellen goed moeten kennen om te begrijpen op welke manieren vooringenomenheid in de AI-levenscyclus kan voorkomen.

2) Ethische hackers, cyber security en informatiebeveiliging specialisten: zij zijn vaardig in het zoeken naar beveiligingslekken, het uitvoeren van penetratietests en het bedenken van scenario's waarin kwaadwillenden het systeem proberen te ondermijnen. Ze kunnen bijvoorbeeld proberen vertrouwelijke gegevens te extraheren, input-naar-output manipulaties te doen of backdoors in de AI-infrastructuur te ontdekken. Informatiebeveiliging specialisten kunnen ook bijvoorbeeld een generatieve AI chatbot uitdagen om datalekken uit te lokken.

3) Data- en privacy-experts: dit zijn deskundigen die analyseren of het AI-systeem onbedoeld gevoelige gegevens kan lekken, of er vooringenomenheid in de data zit en of de kwaliteit van de data goed genoeg is. Bovendien bewaken zij de naleving van privacy-wetgeving en compliance (zoals de Algemene Verordening Gegevensbescherming).

4) Ethiek experts, gedragswetenschappers, antropologen en sociologen: deze professionals hebben kennis van ethische kaders, gedrag en maatschappij, mensenrechten en maatschappelijk verantwoord innoveren. Zij onderzoeken de maatschappelijke impact van het AI-systeem, identificeren mogelijke vormen van discriminatie, stigmatisering of ongewenste beïnvloeding van gebruikers en dragen alternatieve ontwerpkeuzes of richtlijnen aan. Ook kunnen zij ondersteunen in het vinden van mogelijkheden voor social engineering en manieren waarop menselijke interacties tot onbedoelde dreigingen leiden.

5) Domeinexperts: dit zijn personen met diepgaande kennis van de specifieke sector waarin het AI-systeem wordt ingezet (bijvoorbeeld de financiële sector, de gezondheidszorg of de energie-sector). Zij begrijpen de context, de toepasselijke regelgeving, de

AI ontwikkelen of toepassen: red teaming vereist!

typische use cases en risico's binnen hun vakgebied, wat kan helpen om domein specifieke kwetsbaarheden te ontdekken.

6) Juristen en beleidsmakers: zij zijn experts op het gebied van wet- en regelgeving, contracten, aansprakelijkheidskwesties en intellectueel eigendom. Zij waarborgen dat eventuele bevindingen worden getoetst aan juridische kaders, en kunnen aanbevelingen doen om AI-systemen zodanig aan te passen dat deze compliant zijn met de relevante wet- en regelgeving. Dit voorkomt dat de organisatie voor juridische verrassingen komt te staan.

Deze rollenlijst is niet uitputtend. De kern blijft om een interdisciplinair en divers team samen te stellen.

AI red teaming; waar te beginnen?

Een AI red team oprichten is makkelijker gezegd dan gedaan. De inzet ervan hoeft niet van de één op de andere dag te gebeuren. Desalniettemin kan elke organisatie klein beginnen en inspiratie halen uit elementen en principes van AI red teaming. Enkele aanbevelingen zijn:

1) Zet teams op die divers (zowel qua identiteit, als in expertise) en interdisciplinair zijn. Stel vast welke rollen nodig zijn, zorg ervoor dat de verschillende expertises ingeschakeld kunnen worden en maak helder waarvoor deze perspectieven nodig zijn.

2) Richt privacy- data- en AI-governance in. Identificeer ook welke AI-systemen onder de AI Act vallen en of ze aan structurele vereisten moeten voldoen, zoals voor continu risicomanagement (artikel 9), transparantie (artikel 11) en cybersecurity (artikel 15). AI-systemen die niet onder de AI-act vallen zijn niet vanzelfsprekend veilig, dus wees kritisch en zorg ervoor dat ze onder beheersing blijven.

3) Hanteer aanvullend een Responsible AI aanpak en definieer wat dit betekent voor de keuzes die de organisatie moet maken. Neem niet alleen wet- en regelgeving in acht, maar onderzoek zo breed mogelijk de ethische en maatschappelijke gevolgen. Stel bij elk AI-systeem de vraag of de ontwikkeling en toepassing ervan mag, kan en gewenst is.

4) Krijg de zaken helder rondom de autorisatie en authenticatie van functionarissen die toegang hebben tot de data en het model.

5) Zorg voor bewustzijn en expertise op alle mogelijke vormen van vooringenomenheid in de volledige AI-levenscyclus en in het gebruik.

6) Onderzoek regelmatig vooringenomenheid van medewerkers. Bereid voor op de mogelijkheden van social engineering en groepsdenken, stel personas op en maak medewerkers bewust van hun blinde vlekken en kwetsbaarheden.

7) Zorg voor actuele training en expertise op alle privacy- beveiliging- en AI-specifieke dreigingen en onbedoelde uitkomsten. Maak een plan om de risico's te mitigeren en zorg voor duidelijkheid over eindverantwoordelijkheid.

8) Borg AI red teaming organisatorisch in de evaluatiecyclus.

9) Klaar om AI red teaming uit te proberen? Draai een pilot casus met één AI-systeem dat representatief is voor de organisatie. Definieer duidelijke doelen: welke risico's worden opgespoord en welk gedrag wordt getest? Ontwikkel het team op iteratieve wijze.

Er zijn nog weinig organisaties die een volwassen AI red team hebben ontwikkeld. Een voorbeeld is te vinden bij Microsoft. Op learn.microsoft.com deelt dit team documentatie en inspiratie uit de praktijk. Ook kleinere organisaties kunnen baat hebben bij deze hulpmiddelen.

Conclusie

Responsible AI en security liggen in elkaars verlengde: waar security zich voorbereidt op misbruik van buitenaf, richt Responsible AI zich op het gevaar dat het AI-systeem zelf is of kan worden. Beide domeinen focussen op veiligheidsrisico's. Security focust op de betrouwbaarheid, beveiliging en veiligheid van het systeem en Responsible AI focust op deugdelijkheid en de veiligheid van de mens tegenover het systeem. Deze twee benaderingen vullen elkaar sterk aan en hebben met elkaar gemeen dat ze op systematische wijze risico's opsporen. In een AI red team komen deze twee zoekstrategieën samen. Door AI red teaming worden tekortkomingen en blinde vlekken op de techniek, mens, en governance duidelijk. Wij vinden het belangrijk dat AI red teaming meer aandacht krijgt in Nederland. We raden elke organisatie die met AI werkt aan om de oprichting van een AI red team te realiseren. Wees een trendsetter en begin direct!

Call for action

Bent u geïnspireerd door ons artikel en wilt u erop reageren? Wij zijn bereikbaar op LinkedIn en zetten de discussie graag voort. Zodra red teaming binnen AI verder uitgediept wordt, komen wijzelf hierop terug.



COLUMN PRIVACY

Mr. Rachel Marbus

Water naar de zee dragen, maar dan anders

In tijden van toenemende druk op minderheden, is het belangrijk meer aandacht te besteden aan privacy. De wereld om ons heen laat veel (extreem) rechtse machthebbers zien die veiligheid op ramkoers brengen met privacy, kleine groepen binnen de samenleving worden als zondebok op het schavot gezet en weer andere groepen nemen het heft in eigen hand en voelen zich een grensbeambte die het recht heeft anderen de identiteit te controleren.

Onwaarheden over minderheidsgroeperingen worden de ether in en het web opgeslingerd, en hoewel deze onwaarheden weliswaar geen persoonsgegevens zijn, vormen zij wel een onderdeel van hoe de identiteit van een persoon gepercipieerd wordt door 'de ander'. Daarmee gaan die onwaarheden ook deel uitmaken van de identiteit. Identiteit en uitingsvrijheid daaromtrent vormen een beschermde persoonlijke ruimte en maken onderdeel uit van het Grondrecht op privacy. Je bent vrij te zijn wie je bent en relaties aan te gaan met wie je wilt. Daar staat tegenover, dat hoe iemand zelf diens identiteit waarneemt, helaas niet altijd als onderdeel van diens identiteit wordt erkend door 'de ander'.

En hoewel deze zaken misschien groter lijken dan een persoon aankan, kunnen we wel samen beter nadenken hoe we het in ons eigen onderdeel van de wereld beter kunnen maken. Vele kleine druppels maken immers ook een zee. Universiteiten kijken op dit moment samen naar soevereine technologische middelen om zo minder afhankelijk van grote machtige partijen te worden, die zeer vaak afkomstig zijn uit een land dat het niet zo nauw neemt met de privacy. En hoewel die oplossingen niet morgen gereed zijn, kun je wel alvast een nog zorgzamere pet opzetten in de praktijk. Een inschrijving voor deelname aan een boot in de queer Canal parade hoeft niet te gebeuren met behulp van middelen uit de Big Tech hoek. In Delft maken we daarvoor gebruik van het Nederlandse TopDesk, daarin kun je prima zelf een formulier maken en dat ook nog eens privacy vriendelijk inrichten.

En op een andere route proberen we transpersonen te helpen via de zogenaamde weg van de 'rechten van betrokkenen' uit de AVG. Iedereen heeft het recht om diens persoonsgegevens te laten wijzigen als die niet kloppen. Genderidentiteit gaat door middel van zelfidentificatie en de AVG vraagt nergens om bewijs dat jij daarvoor zou moeten aanleveren, de AVG vooronderstelt namelijk dat jij zelf het beste kunt aangeven wat jouw correcte persoonsgegevens zijn. Aan de verwerkingsverantwoordelijke alleen de taak om vast te stellen dat jij jezelf bent – zo voorkomen we dat iemand anders jouw persoonsgegevens kan wijzigen. Dan rest de verantwoordelijke daarna niets anders dan die wijziging door te voeren. Het zijn kleine behulpzame oplossingen die een grote impact maken precies daar waar het ertoe doet: bij jou.

Rachel



Vroegsignalering van jonge cyberbreinen

Investeren in digitaal talent voor een weerbare toekomst

De digitale weerbaarheid van Nederland staat onder druk. Terwijl de dreiging van cybercriminaliteit groeit, blijft het tekort aan goed opgeleide cybersecurityspecialisten nijpend. Tegelijkertijd zien we dat kinderen op steeds jongere leeftijd digitale vaardigheden ontwikkelen – sommigen ver boven het gemiddelde. Dit biedt kansen én risico's: jong talent kan bijdragen aan de veiligheid van morgen, maar is ook kwetsbaar voor rekrutering door criminelen.

In het kader van de Executive Masteropleiding Riskmanagement (Universiteit Twente) onderzocht ik in opdracht van Stichting Cyberbrein.nl hoe deze jonge 'cyberbreinen' al op de basisschool herkend en begeleid kunnen worden. Het onderzoek resulteerde in een set herkenbare gedragskenmerken, inzichten over de omgeving van deze kinderen en aanbevelingen voor onderwijs, beleid en praktijk.

Stichting Cyberbrein

Tezamen met Will Moonen heb ik deze stichting opgezet om jonge cyberbreinen zo vroeg mogelijk te leren (h)erkennen en begeleidingstrajecten aan te bieden. Bijvoorbeeld via *learning paths* in TryHackMe waarin jongeren in kleine groepjes aan de slag gaan onder begeleiding.

Hierbij werkt de Stichting graag samen met het onderwijs want de ervaring leert dat dat goed kan werken: de jongere kan studiepunten halen voor afgeronde *learning paths* bijvoorbeeld.

Hiernaast levert de Stichting ook vaak kennis en expertise tijdens de zogenaamde re_B00TCMP's: het eendaagse event ontwikkeld door de landelijke politie om jong hacktalent in een dag tijd kennis te laten maken met mogelijkheden om hun talent aan de goede kant van de streep in te zetten. Dit gebeurt door bijvoorbeeld jonge cyberbreinen een workshop Capture The Flag te laten faciliteren. Of door een workshop te verzorgen juist voor ouders en docenten over jong hacktalent en hoe je daar mee zou kunnen omgaan.

Ook verzorgt de Stichting zogenaamde workshopcarroussels voor onderwijsinstellingen die ingevuld worden door jonge cyberbreinen en die kunnen gaan over OSINT, CTF, programmeren met ChatGPT, Social Engineering. Deze workshops worden ook in combinatie met cybercrime specialisten vanuit de politie ingevuld en dan gaat het ook over Grenzen van de wet online.

Wat is een jong cyberbrein?

Er is geen harde definitie maar voorafgaand aan het onderzoek is gesteld: "een jong cyberbrein is vaak een kind (8-15 jaar) dat opvallend vaardig is met digitale technologie, erg nieuwsgierig is, complexe systemen intuïtief begrijpt en vaak interesse in hacken toont." Vroege signalering en positieve begeleiding zijn cruciaal om hun talent ethisch te ontwikkelen en misbruik door criminelen te voorkomen.

Onderzoek

De centrale onderzoeksvraag luidde: "Hoe zou een set aan praktische kenmerken eruit kunnen zien om het talent van jonge cyberbreinen al in groep 7/8 van de basisschool te herkennen?" Door middel van literatuuronderzoek en vragenlijsten onder ouders, leerkrachten, ICT-medewerkers en volwassen ethische hackers werd een lijst van gedragskenmerken opgesteld en gevalideerd. Een controlegroep van ouders van "gewone" kinderen diende als vergelijkingsmateriaal.

Belangrijkste bevindingen

Uit het onderzoek kwamen meerdere kenmerken naar voren die typerend zijn voor jonge cyberbreinen, waaronder:

- Nieuwsgierigheid naar alles wat met techniek te maken heeft;
- Weten snel hoe ICT werkt en zijn erg hulpvaardig;
- Analytisch denkvermogen.

Daarnaast wezen respondenten op aanvullende kenmerken zoals autodidactisch leren, out of the box denken, creatief probleemoplossend denken, gouddeerlijk en een "never give up"-mentaliteit als er focus is op iets dat de interesse wekt. Een opvallend terugkerend punt was het sterke rechtvaardigheidsgevoel dat veel cyberbreinen tonen – een aanwijzing dat zij, mits goed begeleid, goed geschikt zijn voor een ethisch pad.

Vroegsignalering van jonge cyberbreinen

De motivaties van jonge cyberbreinen zijn vaak nieuwsgierigheid en de kick om te zien of iets lukt wat ze bedacht hebben. Criminele intenties zijn niet de eerste motivatie maar als eenmaal duidelijk is wat een jong cyberbrein ontdekt heeft, ligt rekrutering door criminelen op de loer. Dat gebeurt via online platforms als Discord.

Een ander belangrijk aspect vormen rolmodellen en de ethical hacking community. Er was een opvallend hoge respons onder volwassen ethical hackers en er werd spontaan veel begeleiding aangeboden voor jonge cyberbreinen. Waarschijnlijke oorzaak voor deze hoge respons is de herkenbaarheid van de struggles van een jong cyberbrein uit de eigen jeugd van de volwassen ethical hacker. Een voorbeeld was een jongen die een spreekbeurt hield op de lagere school over firewalls en vrij snel door de juf werd afgekapd omdat het voor niemand te volgen was.

Herkenning in de praktijk

Ouders herkennen de praktische kenmerken het best omdat ze het dichtst op de leefwereld van de doelgroep zitten, gevolgd door ethische hackers. Leerkrachten en ICT-medewerkers bleken iets minder vaardig in het herkennen van praktische kenmerken. Enerzijds vanwege de eigen beperkte digitale vaardigheden (leerkrachten) anderzijds vanwege de fysieke afstand door ICT-medewerkers. De kloof tussen potentie en herkenning onderstreept in ieder geval de noodzaak van ondersteuning van professionals in de opvoedcontext van het cyberbrein.

Aanbevelingen voor praktijk en beleid

1. *Ontwikkel voorlichting over de manieren van herkenning voor de verschillende doelgroepen zoals ouders, docenten, opvoedprofessionals.*

In het algemeen zou het goed zijn om voorlichtingsmateriaal te ontwikkelen over deze specifieke groep jongeren en hoe je ze kunt leren herkennen waarbij maatwerk per doelgroep voor de hand ligt. Een medewerker ICT van een school zou jonge cyberbreinen kunnen herkennen aan feitelijk gedrag op het netwerk van de school bijvoorbeeld.

2. *Benader talentontwikkeling van jonge cyberbreinen zoals bij sport: stel dat hacken een sport zou zijn?*

Waar het bij voetbal heel normaal is dat jonge kids geïnspireerd raken door topvoetballers als Messi of Ronaldo, is dat bij hacken bijna afwezig. Hoe mooi zou het zijn dat bijvoorbeeld de deelnemers van het nationale team van Challenge The Cyber veel meer als rolmodel gaan optreden voor jonge cyberbreinen? En dat er coaches/mentoren komen vanuit de ethical hacking community die jonge cyberbreinen gaan begeleiden op

“digitale trapveldjes” zoals dat bij voetbal heel normaal is.

Trainingsmateriaal kan bestaan uit zogenaamde Capture The Flags zodat jonge cyberbreinen ook kunnen trainen.

3. *Zet jonge cyberbreinen in bij hackevents zoals Hack050, Hack010, Hack The Hague of CTRL The Port.*

Breng jonge cyberbreinen in contact met volwassen ethical hackers bij events waarbij kwetsbaarheden gezocht worden in systemen van bijvoorbeeld (leveranciers) van een gemeente of andere organisaties. Dit zou natuurlijk ook in combinatie met de eigen school kunnen.

4. *Stimuleer structurele financiering voor samenwerking tussen initiatieven zoals BringITon, DIVD Academy, Coderdojo, Hackerspaces, HackShield, Stichting Cyberbrein, Challenge The Cyber, private cybersecuritybedrijven en het COPS-team van de politie.*

Het landschap van initiatieven gericht op jongeren met hacktalent is nog niet integraal verbonden maar wat nog belangrijker is; vooral leunt op de inzet van een beperkt aantal zeer bevoegen mensen die veelal op vrijwillige basis werkzaam zijn voor het initiatief. Hierdoor blijft de capaciteit beperkt en zou er meer geïnvesteerd moeten worden om doorlopende begeleidingstrajecten in te richten. Net als in de sport waar je als jong kind ook begint bij de pupillen en kan doorgroeien.

5. *Investeer in verdiepend onderzoek en meetinstrumenten.*

Het onderzoek was kwalitatief van aard en biedt nog veel ruimte voor vervolgonderzoek: wat zijn mogelijk invloeden als je kijkt naar cultuuraspecten of demografische kenmerken als gezinssamenstelling. En in hoeverre worden de aanvullende kenmerken herkend door een grotere groep respondenten via kwantitatief onderzoek?

Conclusie

Jonge cyberbreinen vormen een unieke doelgroep: ze zijn kwetsbaar, maar ook waardevol. Hun gedrag en talenten vragen om tijdige herkenning, ethische begeleiding en passende uitdagingen. Door hun vaardigheden te erkennen en te ontwikkelen – zoals we dat ook doen met sport- of muzikalent – kunnen zij uitgroeien tot de cybersecurityprofessionals van de toekomst.

Zelf een rolmodel of coach zijn voor een jong cyberbrein? In jouw organisatie meeloopstages organiseren? Een evenement zoals een CTF organiseren? Of heb je zelf suggesties hoe je onze missie kunt helpen waarmaken? Laat het ons weten! Elke ziel telt!

Kijk op www.bringiton.tech of neem contact op met Stichting Cyberbrein.nl via Henk van Ee, henk.van.ee@cyberbrein.nl

Hoe stuur jij op security in de board room?

Kijk op [cisomasterclass.nl](https://www.cisomasterclass.nl) om uit te vinden hoe je daar grip op krijgt en schrijf je in voor de masterclass op 22, 23 en 24 september 2025.

Kennis brengt je naar de top,
skills zetten je aan het stuur!



 www.cisomasterclass.nl

 info@cisomasterclass.nl

 079-360 4268



COLOFON

IB Magazine is het huisorgaan van het Platform voor InformatieBeveiliging (PvIB) en bevat ontwikkelingen en achtergronden over onderwerpen op het gebied van informatiebeveiliging.

HOOFDREDACTEUR

Chris de Vries

REDACTIE

Alex Dingemanse
Marcel Feenstra
Maarten Hartsuijker
Fook Hwa Tan
Lilian Knippenberg
Leo van Koppen
Rachel Marbus
Chris de Vries

BLADMANAGEMENT

Congres- en Organisatiebureau Interactie BV
Zeynep Turkan
ibmagazine@pvib.nl

VORMGEVING

Neverseen Art & Design
Dimitri van den Berg

DRUK

Veldhuis Media, Meppel

UITGEVER

Platform voor InformatieBeveiliging (PvIB)
Horapark 9
6717LZ Ede
+31 85 799 0233
secretariaat@pvib.nl
www.pvib.nl

*Heb je vragen, opmerkingen of
suggesties, mail dan naar
ibmagazine@pvib.nl*

ABONNEMENTEN

De abonnementsprijs in 2025 bedraagt
€ 118,50 (exclusief btw), prijswijzigingen
voorbehouden.

ABONNEMENTENADMINISTRATIE

Platform voor InformatieBeveiliging (PvIB)
Postbus 1058
3860 BB NIJKERK
secretariaat@pvib.nl



Tenzij anders vermeld valt de inhoud van dit
tijdschrift onder een Creative Commons
Naamsvermelding-gelijkeDelen 3.0 Nederland
Licentie (CC BY-SA 3.0)
ISSN 1569-1063



Praktisch toepasbare kennis!

Heeft u al een NIS2-compliant opleidingsplan?

De nieuwe Europese cybersecurityrichtlijn stelt hoge eisen aan kennis, bewustwording en gedragsverandering binnen organisaties op het gebied van cybersecurity.

Een goed security opleidingsplan hoort daarbij. Hoe doe je dat goed?

In 7 stappen naar een gedegen NIS2-opleidingsplan

Wij helpen u graag op weg met een helder NIS2-opleidingsplan. In onze white paper leest u hoe u gestructureerd werkt aan kennisopbouw en naleving van de richtlijn: van risicobeoordeling tot trainingen voor medewerkers.

*Scan de QR code voor meer informatie.
Of ga naar:
www.securityacademy.nl/nis2-richtlijn/*

