VERSLAG

SECURITY CAFÉ – BIG DATA & SECURITY

Gerco Kanbier is directeur van Trust in People – the information protection company.
Hij is te bereiken via gerco.kanbier@trustinpeople.com

Big data wordt door Gartner gedefinieerd als "high volume, high velocity and high variety". Is big data een nieuwe hype of een andere naam voor Security Information Event Management? Welke organisaties zijn bezig met big data? Wat zijn de kansen en risico's van big data? Kun je op basis van het Safe-Harbour convenant tussen VS en EU opvragen wat er over mij persoonlijk is vastgelegd bij de NSA? Als privacy-data wordt gestolen door een interne medewerker en dit op eigen houtje publiek maakt, is de organisatie daarvoor dan eindverantwoordelijk en kan hiervoor een boete worden opgelegd door de toezichthouder? Bij websites testen we vaak of via SQL-injection data te manipuleren is. Hoe bewaak je de integriteit van big data-bronnen als dat gebaseerd is op NoSQL-technologie zoals Hadoop? Kortom, een hoop vragen waar het panel in 45 minuten op los kon gaan.

Vier maal per jaar organiseert Trust in People het Security Café. De editie van 25 juni 2013, werd wederom bij Koetjes & Kalfjes te Zoetermeer georganiseerd. Op LinkedIn hadden de inmiddels 700 leden van de groep voor het onderwerp "Big Data & Security " gestemd. In het expertpanel Michiel Toes, commercieel directeur van SMT en specialist op het gebied van Big Data en Security management oplossingen. Hans Teffer, CISA en Register IT auditor, gespecialiseerd in het ontwerpen en creëren van holistische security monitoring omgevingen en Sander Klous, hoofd Data & Analytics bij KPMG en binnenkort hoogleraar Big Data Ecosystems aan de Universiteit van Amsterdam. Hier volgt een impressie van de discussie.

Sander Klous stelt als eerste dat er een verschil is tussen "Big Data for Security" en "Security of Big Data". "Big Data for Security" biedt mogelijkheden op het gebied van fraudedetectie zoals bij zorgverzekeraars en preventie zoals bij terreurbestrijding bij de NSA. Bij "Security of Big Data" gaat het om de betrouwbaarheid van die informatie en wie toegang heeft tot die informatie.

De meeste big data initiatieven zijn al enige jaren geleden opgestart, aldus Hans Teffer. Google is in 2004 met hun BigTable database gestart. Facebook, Amazon en de NSA begonnen met hun varianten daarop in 2007. Hans ziet big data niet als een hype, maar als een 'mainstream' implementatie voor overheden en grote bedrijven.

Logfiles kunnen een schat aan informatie te bevatten, mits je weet waar je naar op zoek bent. ArcSight, LogLogic, Splunk en Hadoop zijn voorbeelden van software applicaties om grote logverzamelingen te analyseren. ArcSight (HP) en LogLogic zijn in de markt vooral gepositioneerd als SIEM-oplossing en richten zich op realtime security en compliance waarbij vooraf bekend is wat je zoekt. Splunk en Hadoop (Apache) zijn gebaseerd op NoSQL-technologie en worden vaak gebruikt als ongestructureerd data-archief waarop analyses realtime, maar ook achteraf alsnog uitgevoerd worden. Dit wordt zowel voor beveiligingsdoeleinden, fraudedetectie als analyses gebruikt. Het mooie van Splunk, alsdus Michiel Toes, is dat analyses achteraf nog uitgevoerd kunnen worden, omdat je alle logdata opslaat. Soms weet je naar aanleiding van (APT) incidenten pas veel later waarop je moet letten in de logfiles. Bewezen analyses zoals aankoopgedrag of surfgedrag kunnen vervolgens weer gestructureerde input zijn voor een datawarehouse.

> Soms weet je pas later waarop je moet letten in de logfiles

Het analyseren en vastleggen van klantengedrag is gekoppeld aan privacy-wetgeving. Met big data zijn er nogal wat uitdagingen op gebied van privacy. Indien klanten bijvoorbeeld toestemming hebben gegeven voor een 'cookie' ten behoeve van commerciële en/of analysedoeleinden, dan is er nog geen toestemming om andere bronnen - waar ook toestemming voor is - aan elkaar te koppelen. Als er namelijk nieuwe informatie over deze klant ontstaat, door verbanden te leggen en bronnen aan elkaar te koppelen, dan moet dit vooraf en duidelijk worden verteld aan die klant. Deze klant moet een overzicht van zijn gegevens kunnen opvragen en eventueel eisen dat zijn gegevens verwijderd worden. De vraag is hoe die verantwoordelijkheid in de organisatie belegd is en of hier toezicht op is. Het is

namelijk gemakkelijk om dit technisch te realiseren zonder dat een klant dit door heeft. Volgens Sander Klous is het voor big data-systemen belangrijk om toegangsmaatregelen te ontwerpen op het niveau van rapportages, zoekfunctionalteit en business logica. Hiermee moet voorkomen worden dat een analist kan inzoomen op de details van een individuele gebruiker, als daar geen toestemming voor is verleend. Dit wordt ook wel 'Granular Access Control' genoemd.

De onthullingen van Edward Snowden over het publieke geheim dat NSA 'ons' bespioneert is echt een mediahype. Voor welke doeleinden de NSA namelijk binnen- en buitenland mag bespioneren, is gewoon bij Amerikaanse wet geregeld. Je kan je afvragen waarom sommige doeleinden in dit soort wetgeving is aangenomen, maar dat is anno 2013 een feit. Het verschil met een totalitaire staat, is dat dit valt en staat met toezicht op misbruik of oneigenlijk gebruik door derden. Dit soort wetgeving geldt overigens bijna voor alle inlichtingendiensten wereldwijd. Gegeven deze wetgeving is de wens voor een onafhankelijke internationale toezichthouder geboren. Inlichtingendiensten mogen net zomin als banken 'valsspelen', ook al is dat beschreven in de kleinste lettertjes van een wetsartikel. Het is op sommige punten immoreel en onethisch voor lokale belangen. Gezien de recente leerpunten uit de bancaire sector, is aangescherpt en internationaal toezicht bij inlichtingendiensten ook nodig als het gaat om bescherming tegen misbruik van gegevens. De overheid en inlichtendiensten hebben een serieuze verantwoordelijkheid dat informatie nooit in handen valt van iemand met andere/kwade bedoelingen. Ik verwacht namelijk niet dat de hacker straks toestemming gaat vragen bij het College Bescherming Persoonsgegevens voor een big data-store van de NSA, omdat de doeleinden van de nieuwe eigenaar zijn veranderd.

Inlichtingendiensten mogen net zomin als banken 'valsspelen'

Als data wordt gestolen door een interne medewerker en dit tegen de contractuele afspraken op eigen houtje publiek maakt, is de organisatie daarvoor dan eindverantwoordelijk? We kunnen technisch, juridisch en organisatorisch veel dingen beveiligen, maar incidenten blijven gebeuren. Een CD met alle zwitserse bankrekeningen, een USB-stick met alle medische gegevens van patiënten of alle diplomatieke documenten over een dubieuze oorlog, het gebeurt gewoon. Dit probleem is misschien alleen te verzekeren. Echter, ik heb nog geen juridische claims gezien van slachtoffers naar de organisatie of overheid die eindverantwoordelijk is. Noch heb ik toezichthouders boetes zien uitdelen.

Bij big data wordt vaak gebruik gemaakt van gedistribueerde systemen, met in veel gevallen duplicatie van data, wat vaak ten goede komt aan de beschikbaarheid van informatie. Deze data wordt ongestructureerd opgeslagen, echter de exacte lokatie van data elementen is moeilijk te bepalen. Juridisch kan dat een probleem zijn, als je zelf niet weet waar de data opgeslagen is. Meestal wordt dan wetgeving gehanteerd die geldt waar het hoofdkantoor gevestigd is. Daarnaast is het nog maar de vraag of logfiles op die verschillende locaties beschermd zijn.

Ongestructueerde data zoals log-files zijn over het algemeen niet beveiligd zoals gestructueerde databases. Je kan dus eenvoudig log-bestanden aanpassen met onjuiste analyses en beslissingen tot gevolg. De integriteit en vertrouwelijkheid van deze big data-stores moet dus anders beveiligd worden. Daarom dient er een vorm van security monitoring opgezet te worden rond Big Data-stores waarin gelet moet worden op de integriteit van log-files, wie toegang heeft tot welke data, het soort analyses dat wordt uitgevoerd en de volledigheid van informatiebronnen monitoren, alsdus Hans Teffer.

Tot slot een voorbeeld van een big data-ecosystem: Talend-software kan je gebruiken om verschillende databronnen aan een big data-systeem te koppelen op basis van NO-SQL Splunk-technologie. Rapportages en analyses uit Splunk vormen dan weer input voor een datawarehouse.

CC by Rena Schild / Shutterstock.com

Links

Security Cafe: informatie beveiliging community in Nederland www.trustinpeople.com/security_cafe.php