Security intelligence door aggregatie open-source intelligence

AFSTUDEERSCRIPTIE

Sjors Haanen

JANUARI 2017

Titelblad afstudeerscriptie

Gegevens afstudeerder	
Naam	S.C.W. Haanen
Studentnummer	2166252
Afstudeerrichting	ICT & Cyber Security (voltijd)
Afstudeerperiode	september 2016 tot februari 2017 (85 dagen)

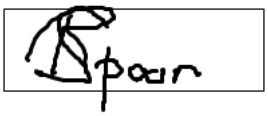
Gegevens organisatie	
Naam	SURFnet B.V.
Afdeling	Security Innovatie & Exploitatie
Bezoekadres	Kantoren Hoog Overborch (Hoog Catharijne) Moreelsepark 48 3511 EP Utrecht
Bedrijfsbegeleider	Rogier Spoor (productmanager)

Gegevens docentbegeleider	
Naam	Stefan Roijers
Organisatie	Fontys Hogescholen ICT

Gegevens afstudeerscriptie	
Titel	Security intelligence door aggregatie open-source intelligence
Datum uitgifte	10 januari 2017

Getekend voor gezien door bedrijfsbegeleider:

Datum: 10 januari 2017



Rogier Spoor

Voorwoord

Het afgelopen schoolsemester heb ik een afstudeerproject mogen uitvoeren bij SURFnet in Utrecht. In dit project heb ik openbare security intelligence bronnen op het internet onderzocht en gecombineerd tot een tool die informatie kan tonen over actuele ICT-kwetsbaarheden binnen organisaties. Ik heb mijn onderzoek bij SURFnet uitgevoerd omdat ik van tevoren had gehoord over de prettige open werksfeer die er heerst en de grote mate van deskundigheid waarover de medewerkers beschikken. Ook heeft SURFnet meegewerkt aan een aantal zeer succesvolle innovatieve projecten zoals het ontstaan van het internet in Nederland. Hierdoor wist ik dat SURFnet een uitstekende plek is om veel te leren en mijn opdracht deels naar eigen wensen te vormen.
Ik wil alle SURFnet medewerkers bedanken voor de hulp die ik gekregen heb tijdens mijn afstudeerproject. In het bijzonder wil ik Rogier Spoor bedanken voor de kansen die hij me heeft gegeven om mijn project bekend te maken bij een groot aantal cyber security deskundigen en mijn project te betrekken bij de samenwerking tussen SURFnet en GDI Foundation.

Utrecht, januari 2017
Sjors Haanen

Samenvatting

Er zijn diverse openbare bronnen op het internet die informatie kunnen geven over kwetsbaarheden in computersystemen of gelekte gevoelige informatie op het internet. Dit zijn veelal websites die de informatie verzamelen en dit aanbieden aan eindgebruikers. Bij SURFnet is er behoefte aan een manier om met behulp van deze bronnen nieuwe kwetsbaarheden of gelekte informatie uit de systemen van SURFnet en aangesloten instellingen overzichtelijk en laagdrempelig te tonen.

In de situatie voor de start van dit project waren de bronnen die voor dit project gebruikt zijn niet geaggregeerd en moesten deze afzonderlijk geraadpleegd worden. Er kon per instelling nog geen compleet beeld gegeven worden van de kwetsbaarheden die zichtbaar waren voor de buitenwereld. Hierdoor werd onnodig risico gelopen waarbij kwetsbaarheden eenvoudig konden worden geëxploiteerd door criminele organisaties. Er miste een manier om gevonden kwetsbaarheden uit verschillende openbare bronnen overzichtelijk weer te geven.

Deze probleemstelling heeft geleid tot de volgende onderzoeksvraag: “Hoe kan informatie uit openbare bronnen worden gekoppeld en welke eisen worden daaraan gesteld door de belanghebbenden, zodat de kwaliteit van de security intelligence van SURFnet en de aangesloten instellingen wordt verbeterd?” Om deze probleemstelling en onderzoeksvraag te beantwoorden is er onderzoek gedaan en aan de hand van de onderzoeksresultaten is een Proof of Concept opgeleverd.

Het onderzoek heeft laten blijken dat de bronnen Shodan, Censys, Zoomeye, IpInfo en @dumpmon geschikt zijn om in een tool te integreren. Uiteindelijk zijn Shodan, Censys en IpInfo gebruikt in de Proof of Concept. De informatie uit deze bronnen kan op aanvraag of automatisch opgevraagd worden en wordt gestructureerd opgeslagen in een Elastic Stack omgeving bestaande uit Logstash, Elasticsearch en Kibana. Veldonderzoek in de vorm van interviews is gedaan bij de belanghebbenden om vervolgens rekening te houden met hun wensen bij het ontwerpen van de Proof of Concept. Eindgebruikers van de tool kunnen met Kibana, die de verzamelde data visualiseert, dashboards configureren die een algemeen beeld op hoog niveau geven over gevonden kwetsbaarheden binnen SURFnet en aangesloten instellingen. Ook kan informatie over getroffen IP-adressen op detail bekeken worden zodat men over de nodige informatie beschikt om vervolgacties te kunnen ondernemen.
Dit heeft een praktische tool opgeleverd waarin per instelling te zien is welke informatie zichtbaar is voor de buitenwereld. In de toekomst kan SURFnet de Proof of Concept inzetten voor gebruik door de instellingen en kan GDI Foundation deze in hun eigen project gebruiken.

Abstract

There are several public sources on the internet which can provide information about vulnerabilities in computer systems or leaked sensitive information on the Internet. These are usually websites which gather the information and present it to end users. SURFnet is in need for a way to show the vulnerabilities or leaked information from the systems of SURFnet and affiliated institutions in a clear and approachable way.

In the situation before the start of this project, the sources used in this project were not aggregated and had to be consulted separately. There could not yet be made a complete picture per institution of the vulnerabilities visible to the outside world. This was an unnecessary risk in which vulnerabilities could be easily exploited by criminal organizations. There lacked a way to give a clear overview of vulnerabilities found from various public sources.

This problem led to the following research question: "How can information from public sources be combined and which demands do stakeholders have, so that the quality of the security intelligence SURFnet and its affiliated institutions is improved?" To answer this problem and research question there has been conducted research and the results have been used to provide a Proof of Concept.

Research has shown that the sources Shodan, Censys, Zoomeye, IpInfo and @dumpmon are suitable to integrate into a tool. Ultimately, Shodan, Censys and IpInfo have been used in the Proof of Concept. The information from these sources can be retrieved on demand or automatically and are stored in a structured form in an Elastic Stack environment consisting of Logstash, Elasticsearch and Kibana. Field research in the form of interviews has been performed on the stakeholders for finding their wishes, and these wishes were taken into account when the Proof of Concept was being designed. End users of the tool can use Kibana, which visualizes the collected data, to configure dashboards which give a general picture at high level about found vulnerabilities within SURFnet and affiliated institutions. Also, information on affected IP addresses can be viewed in detail so the needed information can be gathered to be able to take further action.
This led to a practical tool in which information visible to the outside world can be viewed per institution. In the future SURFnet can use the Proof of Concept for use by the institutions, and GDI Foundation can use it in their own project.

Inhoudsopgave

Verklarende woordenlijst

Begrip	Verklaring/definitie
Application Programming Interface	Een manier om met een programma te communiceren oftewel data uit te wisselen. Zo heeft Shodan er een om via programmeercode zoekopdrachten te doen op de data van Shodan.
Databasedump	Een of meerdere bestanden die waardevolle en vaak gevoelige informatie van een database bevat. Een database dump is oorspronkelijk bedoeld als een back-up middel, maar is tegenwoordig meer in het nieuws doordat kwaadwillenden de gevoelige informatie van een bedrijf via een databasedump onrechtmatig hebben verkregen.
Hacker/Hacken	De meest gangbare en de in dit document gehanteerde betekenis van een hacker is iemand die met kwaadaardige bedoelingen probeert in te breken in computersystemen, oftewel hacken.
Kwetsbaarheid	Eigenschap van een samenleving, organisatie of informatiesysteem (of een onderdeel daarvan) die een kwaadwillende partij de kans geeft om de legitieme toegang tot informatie of functionaliteit te verhinderen en te beïnvloeden, of om die ongeautoriseerd te benaderen.
Persisteren	In database termen betekent persisteren het opslaan van informatie voor later gebruik.
Virtuele Machine	Een computerprogramma dat een computer nabootst, waar andere programma's op kunnen worden uitgevoerd. In de praktijk kan het beschouwd worden als een volledige computer. In dit project wordt met virtuele machines de computers bedoeld die onderdeel uitmaken van de Proof of Concept.

1 Inleiding

1.1 Leeswijzer

Hoofdstuk 1 is de inleiding van het verslag. In hoofdstuk 2 wordt SURF als organisatie beschreven. Hoofdstuk 3 beschrijft de opdracht met het onderzoeksaspect en de belanghebbenden van het project. Vervolgens gaat hoofdstuk 4 in op de inrichting en planning van het proces. Hoofdstuk 5 begint met een beschrijving over de samenwerking met externe belanghebbenden, met daarna de uitvoering van de opdracht met de kernpunten van de onderzoeksresultaten. In hoofdstuk 6 is opbrengst van de Proof of Concept beschreven. Hoofdstuk 7 is de conclusie waarin de hoofdvraag en deelvragen en daarmee de probleemstelling kernachtig beantwoord wordt. Ten slotte bevat hoofdstuk 8 de evaluatie waarin gereflecteerd wordt op de afstudeerstage met de kritische zelfreflectie.

1.2 Aanleiding

Er zijn diverse openbare security intelligence bronnen op het internet die informatie kunnen geven over kwetsbaarheden in computersystemen of gelekte gevoelige informatie op het internet. Dit zijn websites die de informatie verzamelen en dit aanbieden aan bijvoorbeeld onderzoekers of security officers. Echter is niet uit te sluiten dat kwaadwillenden hier ook gebruik van maken. De bronnen komen aan deze informatie door bijvoorbeeld scanners te gebruiken die aan het internet verbonden apparaten scannen voor aangeboden ICT-services. Een ander voorbeeld van hoe de bronnen aan hun informatie komen is door actief gevoelige informatie te zoeken op het internet dat ooit gestolen of per ongeluk uitgelekt is. Hierbij moet men denken aan klantgegevens, persoonsgegevens of intellectuele eigendommen. De websites overtreden hierbij niet de wet omdat ze de informatie zelf niet hebben gestolen en niet direct misbruiken om hiermee iets illegaals te doen. Veelal publiceren ze de gevoelige informatie alleen aan de desbetreffende getroffen organisaties, via authenticatie op domeinnaam niveau.

SURFnet wil deze beschikbare bronnen niet links laten liggen omdat SURFnet en aangesloten instellingen (de klanten) hiermee kunnen achterhalen welke informatie er van hun publiekelijk beschikbaar is. De bronnen zijn veelal van kwalitatief hoog niveau en kunnen specifieke kwetsbaarheid aantonen. Daarnaast kunnen ze een algemeen beeld geven van de staat van ICT-beveiliging van organisaties, waarbij gevonden kwetsbaarheden vervolgens tot op detailniveau bekeken kunnen worden. Dat maken deze bronnen tot een belangrijke aanvulling op de traditioneel gebruikte bronnen. Het gebruik van de bronnen door SURFnet moet wel ethisch verantwoord zijn: de bronnen moeten er zijn om het internet veiliger te maken en moeten geen gevoelige informatie zoals persoonsdata delen met derden.

Bij SURFnet is er behoefte aan een manier om met behulp van deze bronnen nieuwe kwetsbaarheden of gelekte informatie uit de systemen van SURFnet en aangesloten instellingen overzichtelijk en laagdrempelig te tonen. SURFnet heeft al langer een goede samenwerking met Fontys Hogescholen en de opleiding ICT & Cyber Security sluit goed aan bij de opdracht. Dit heeft de aanleiding gevormd om Sjors Haanen van deze opleiding als afstudeerder aan te nemen.

2 De organisatie

SURF is een ICT-samenwerkingsorganisatie van Nederlandse onderwijs- en onderzoeksinstellingen. Succesvolle innovatieprojecten op ICT-gebied worden gerealiseerd in landelijke ICT-voorzieningen zodat de aangesloten instellingen hier gebruik van kunnen maken. Daarnaast heeft SURF grote invloed gehad op het succes van de Amsterdam Internet Exchange (AMS-IX) en de software achter het DigiD systeem. SURF bestaat uit drie gespecialiseerde onderdelen die in dit hoofdstuk verder beschreven worden.

2.1 SURFnet

SURFnet, gevestigd in Utrecht, richt zich op het stimuleren, ontwikkelen en exploiteren van een hybride netwerk, een vertrouwde identiteit en een samenwerkingsomgeving, en zorgt er hiermee voor dat onderzoekers, docenten en studenten kunnen samenwerken met behulp van ICT. Dit doet men door een dienstverlening op twee gebieden. Enerzijds zorgt SURFnet voor een netwerkinfrastructuur dat efficiënt datatransport realiseert. Anderzijds biedt SURFnet een samenwerkingsinfrastructuur aan die systemen, instrumenten en mensen verbindt.

2.2 SURFmarket

SURFmarket, gevestigd in Utrecht, is een licentieorganisatie die namens de aangesloten instellingen onderhandelt met ICT-leveranciers en uitgevers om campuslicenties af te sluiten voor software, content en hardware. Deze biedt SURFmarket vervolgens aan de aangesloten instellingen zodat zij kunnen profiteren van de best mogelijke voorwaarden voor de aanschaf van onder meer software, clouddiensten en digitale content.

2.3 SURFsara

SURFsara, gevestigd in Amsterdam en Almere, faciliteert wetenschappelijk onderzoek door diensten te leveren op het gebied van supercomputers, netwerken, dataopslag en hoogwaardige visualisatie. Oorspronkelijk alleen voor de Vrije Universiteit Amsterdam, de Universiteit Amsterdam en het Mathematisch Centrum. Al vele jaren mogen ook andere universiteiten en onderzoeksinstituten gebruik maken van de diensten van SURFsara.

3 Opdrachtomschrijving

3.1 Gewenste situatie

De gewenste situatie is dat instellingen die aangesloten zijn bij SURFnet beschikking hebben over een snel in te zetten tool die openbare bronnen gebruikt om een compleet overzicht te geven over gevonden ICT-kwetsbaarheden en gelekte informatie per instelling. Hierdoor worden instellingen bewust van gelekte gevoelige informatie en kunnen beveiligingsrisico's tijdig gedicht worden om (verder) misbruik te voorkomen.

3.2 Doelstelling

Het doel van dit project is om de security intelligence van SURFnet en de aangesloten instellingen te vergroten met behulp van een tool die informatie uit verschillende openbare bronnen verzamelt en aggregeert. Onder security intelligence verstaan we inzicht op security risico's vergaard door het analyseren van verzamelde informatie. De tool geeft een 'health check' per instelling die de beschikbare informatie over kwetsbaarheden en gelekte informatie laagdrempelig en uniform inzichtelijk maken. De focus ligt daarbij op het tonen van de meeste relevante en belangrijkste kwetsbaarheden waarbij false positives (loze alarmen) vermeden worden.

Voorafgaand aan de tool wordt onderzoek gedaan naar de benodigde informatie om de tool te ontwerpen. Het gehele onderzoek begint met een functionele analyse zodat duidelijk is waaraan de tool moet voldoen, maar omvat ook het maken van de tool zelf. De onderzoeksresultaten worden gerapporteerd en dient als advies voor SURFnet. In de context van deze afstudeeropdracht zal de tool gelden als Proof of Concept (verder: PoC). Dit is een realisatie van de conclusies uit de onderzoeksresultaten dat gedemonstreerd wordt bij de eindpresentaties. SURFnet kan na dit project ervoor kiezen om de PoC in gebruik te laten nemen door de aangesloten instellingen. Daarbij kan de PoC aangepast worden in een iteratief proces naar aanleiding van nieuwe feedback van belanghebbenden. Daarom ontvangt SURFnet ook documentatie van de PoC voor de overdraging, dat in dit document toegevoegd is als bijlage 5.

Doelstelling voor het onderzoek is onder andere om inzicht te geven in welke bronnen een bijdrage kunnen leveren aan het vergroten van de security intelligence van SURFnet en de aangesloten instellingen naar aanleiding van de eisen van belanghebbenden. Het onderzoek zal tijdens het ontwerpen van de PoC doorlopen zodat te zien is of de resultaten daadwerkelijk een oplossing zijn voor de probleemstelling.

Als internet serviceprovider is het voor SURFnet van belang om het SURFnet netwerk zo schoon mogelijk te houden. Het doel voor de aangesloten instellingen is dat ze snel en accuraat geïnformeerd worden over bedreigingen.

3.3 Probleemstelling

Op dit moment zijn er een aantal openbare bronnen die nog niet standaard gebruikt worden bij SURFnet. Ook worden de bronnen niet geaggregeerd en moeten ze afzonderlijk geraadpleegd worden. Tevens zorgen afzonderlijke bronnen vaak voor false positives. Er kan nog geen integraal beeld gegeven worden van de kwetsbaarheden van een organisatie. Hierdoor wordt onnodig risico gelopen waarbij kwetsbaarheden eenvoudig kunnen worden geëxploiteerd door criminele organisaties. Zij gebruiken namelijk dezelfde en/of vergelijkbare bronnen en beschikken over genoeg tijd en expertise om de hieruit vergaarde informatie te misbruiken. Concluderend mist er dus een manier om gevonden kwetsbaarheden uit verschillende openbare bronnen overzichtelijk weer te geven.

3.4 Onderzoek

3.4.1 Onderzoeksvraag

Om een oplossing te vinden voor de geschetste probleemstelling is de volgende onderzoeksvraag opgesteld.

“Hoe kan informatie uit openbare bronnen worden gekoppeld en welke eisen worden daaraan gesteld door de belanghebbenden, zodat de kwaliteit van de security intelligence van SURFnet en de aangesloten instellingen wordt verbeterd?”

Om een volledig antwoord te kunnen geven op de bovenstaande hoofdvraag moeten de volgende deelvragen beantwoord worden:

- Wat zijn de wensen van de belanghebbenden met betrekking tot het verhogen van de kwaliteit van security intelligence?
- Welke bronnen kunnen van toegevoegde waarde zijn?
- Hoe wordt informatie uit de bronnen gepersisteerd zodat deze effectief geraadpleegd kan worden?
- Welke producten kunnen zorgen voor een gebruikersomgeving die voldoet aan de wensen van de belanghebbenden en hoe staan die producten in relatie tot elkaar?
- Welke koppelingen van de informatie uit de bronnen kunnen de kwaliteit van de security intelligence verhogen?

De deelvragen zijn uitgewerkt in de paragrafen 5.2 t/m 5.6. De conclusie van de onderzoeksvraag is opgenomen in hoofdstuk 7.

3.4.2 Onderzoeksopzet

Als onderzoeksmethode zijn bij elke deelvraag geschikte onderzoeksstrategieën gezocht. Dit zijn de strategieën van de ‘Methodenkaart Praktijkonderzoek’. Deze kaart legt een verbinding tussen de verschillende onderzoekstradities die ten grondslag liggen aan het vakgebied en het ontwikkel- of maakproces in de ICT. De kaart kent 5 strategieën: veld, bieb, werkplaats, lab en showroom. Zie de bron voor verdere toelichting van deze strategieën [1].

Tabel 1: Gebruikte methoden en strategieën per deelvraag.

Deelvraag	Methode	Strategie
1. Wensen belanghebbenden	Interviews met belanghebbenden	Veldonderzoek
2. Bronnen	Internetonderzoek, overleggen met Victor Gevers, experimenteren met gevonden bronnen in testomgeving	Biebonderzoek Werkplaatsonderzoek
3. Informatie persisteren	Internetonderzoek, experimenteren met software in testomgeving	Biebonderzoek Werkplaatsonderzoek
4. Producten voor gebruikersomgeving	Internetonderzoek, experimenteren met (software)producten in testomgeving	Biebonderzoek Werkplaatsonderzoek
5. Koppelingen van informatie	Internetonderzoek, experimenteren in testomgeving	Biebonderzoek Werkplaatsonderzoek

Ook vinden er gesprekken met medewerkers van SURFnet om van hun ervaringen gebruik te maken, bijvoorbeeld voor de keuze van bepaalde software of ‘best practices’ bij het onderzoek.

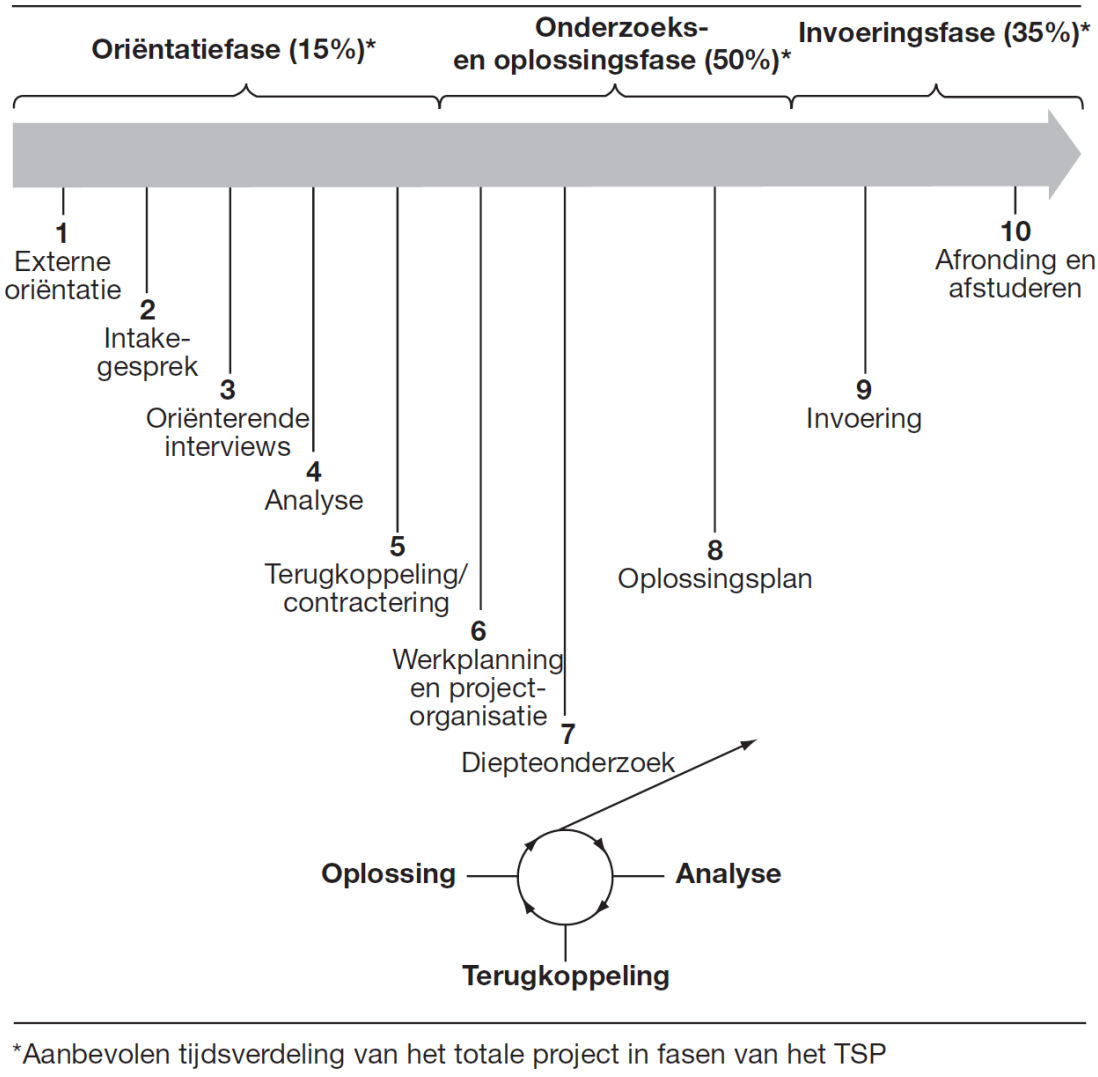
3.5 Belanghebbenden

De belanghebbenden zijn de mogelijke eindgebruikers van de uiteindelijke PoC: mensen verantwoordelijk voor de ICT-beveiliging bij SURFnet en aangesloten instellingen zoals security officers. Hierbij horen naast de technische werknemers die kwetsbaarheden willen verhelpen ook het management die op een hoger niveau willen zien hoe de organisatie ervoor staat. Verder is het aan de instellingen zelf wie toegang krijgt tot de PoC en wie verantwoordelijk is voor het nemen van eventuele vervolgacties naar aanleiding het vinden van nieuwe kwetsbaarheden.

Ook is GDI Foundation een belanghebbende van de resultaten van dit onderzoek. GDI Foundation is een non-profit organisatie dat zich inzet om het internet veiliger te maken voor iedereen door met name gevonden kwetsbaarheden in computersystemen van burgers en organisaties verantwoord te onthullen en deze te adviseren in vervolgacties. Ze proberen op deze manier kwaadwillenden voor te zijn.
Ten Slotte is Remco Verhoef een belanghebbende van dit project. Remco is een ervaren securityspecialist en programmeur die open-source security tools beschikbaar stelt. Meer over GDI Foundation, Remco Verhoef en de samenwerking met dit project is de lezen in paragraaf 5.1.

4 Proces (inrichting en planning)

De inrichting van dit project is afgeleid van het ‘Tien Stappen Plan’ [2]. Dit is een leidraad voor het methodisch uitvoeren van afstudeeropdrachten dat de afstudeerkwaliteit beheersbaar kan maken. Zie figuur 1 voor een overzicht van de stappen.



Figuur 1: Tien Stappen Plan van het afstudeerproject.

Stap 1 t/m 5 zijn al deels doorlopen voorafgaand aan de start van de afstudeeropdracht. Zo heeft er voor de start van de stageperiode al een gesprek plaatsgevonden tussen Sjors Haanen, de opdrachtgever en de 1e assessor met als input een volledig ingevuld gespreksformulier om de opdracht al meer vorm te kunnen geven. Ook de gesprekken met belanghebbenden tijdens de eerste weken van de stage zijn onderdeel van deze stappen. Stap 6 ‘werkplanning en projectorganisatie’ betreft het schrijven van het projectplan en eindigt met de vereiste goedkeuringen. Daarna begint stap 7 ‘diepteonderzoek’ met een onderzoek naar de hoofd- en deelvragen dat nodig is om SURFnet een

onderbouwd advies te kunnen leveren en om de PoC te kunnen ontwerpen. Het maken van de PoC start ook in deze fase. Gevonden bronnen als input voor de PoC en software waar de PoC uit zal bestaan worden uitgeprobeerd in de testomgeving om bevindingen in het onderzoek te onderbouwen. Stap 8 ‘oplossingsplan’ betekent in dit project het afronden van de PoC, waarbij regelmatig met de opdrachtgever wordt besproken of dat de PoC de goede kant op gaat. Rond deze tijd begint ook het schrijven van het afstudeerverslag. Stap 9 uit het Tien Stappen Plan is niet van toepassing aangezien de daadwerkelijke uitrol van de PoC niet bij de scope van dit project hoort. Ten slotte wordt er in stap 10 ‘afronding en afstuderen’ een presentatie met demo van de PoC bij SURFnet en een presentatie bij de afstudeerzitting gegeven. De faseringen die zijn ontstaan uit de bovenstaande inrichting met de bijbehorende tijdsaanduidingen zijn te vinden in bijlage 1.

5 Uitvoering

5.1 Samenwerking

Tijdens het gehele onderzoek heeft wekelijks een kort overleg plaatsgevonden met Victor Gevers en Vincent Toms van GDI Foundation en Rogier Spoor van SURFnet, om de voortgang en nieuwe ideeën te bespreken. GDI Foundation is zelf van plan om een project op te zetten genaamd ‘Internet & Cyber Security Health Check’. Met dat project willen ze hun huidige tool en informatieverstrekking verder optimaliseren zodat meerdere partijen in staat worden gesteld om tijdig adequate tegenmaatregelen te treffen tegen cyber crime mogelijkheden. De scope van hun project is echter op nationaal niveau terwijl de scope van dit project alleen SURFnet en aangesloten instellingen betreft. GDI Foundation kan op basis van de PoC van dit project verder gaan door deze PoC op landelijk niveau op te schalen. Dit project heeft GDI geholpen bij het aanvragen van financiële steun bij het SIDN fonds. Dit is gebeurd door in de pitch voor het SIDN fonds een demo te geven van de PoC. Het SIDN fonds heeft mede naar aanleiding van de presentatie met demo toegestemd op financiering van het bedrag van €73.000,-. In combinatie met het ontwerp van de PoC zal dit een goede basis zijn voor het project van GDI Foundation.

Ook is er samengewerkt met Remco Verhoef. Remco wil graag een bijdrage leveren aan dit project en dit heeft hij gedaan door zijn tool ‘IpInfo’ beschikbaar te stellen als bron voor de PoC. Deze tool kan meer informatie geven over een gegeven IP-adres. Tijdens de integratie van IpInfo in de PoC heeft hij in overleg verschillende verbeteringen aangemaakt aan zijn tool zodat deze sneller data kan leveren. Deze bron komt verder aan bod in subparagraaf 5.3.3.

Na de goedkeuring van de eerste definitieve versie van het projectplan is gestart met de fase ‘onderzoek en PoC’. De kernpunten van de onderzoeksresultaten komen in de volgende hoofdstukken aan bod.

5.2 Functionele analyse PoC

Om een functionele analyse te maken is de volgende deelvraag beantwoord: “Wat zijn de wensen van de belanghebbenden met betrekking tot het verhogen van de kwaliteit van security intelligence?” Voor dit veldonderzoek zijn de volgende belanghebbenden geïnterviewd:

- Rogier Spoor, de opdrachtgever van dit project;
- Victor Gevers en Vincent Toms van GDI Foundation;
- Ewald Beekman, voorzitter van SCIRT (de CERT-gemeenschap van SURFnet);
- Alf Moens, de Chief Information Security Officer (CISO) van SURF.

De kernpunten van dit veldonderzoek zijn samenvattend verwerkt als functionele en niet-functionele eisen. Zie bijlage 2 voor uitwerkingen van de interviews.

5.2.1 Functionele eisen

- De gebruiker kan de resultaten per organisatie raadplegen;
- De gebruiker ziet een algemeen overzicht kan van de ICT-kwetsbaarheden, waarbij de informatie laagdrempelig en overzichtelijk getoond wordt;
- De gebruiker kan een ICT-kwetsbaarheid op detailniveau bestuderen;
- De gebruiker kan informatie over openstaande services naar buiten bekijken;
- De informatie uit de PoC kan op aanvraag en periodiek ververst worden;
- De PoC kan veranderingen ten opzichte van eerdere raadplegingen van de PoC laten zien;
- De PoC kan alle beschikbare data van een bepaalde ASN (groep IP-reeksen), IP-reeks of IP-adres uit de bronnen opvragen.

5.2.2 Niet-functionele eisen

- De data moet juist, relevant en tijdig zijn;
- False positives moeten zoveel mogelijk vermijd worden;
- De PoC moet effectief te raadplegen zijn waarbij de informatie snel getoond wordt;
- De PoC moet stabiel zijn zodat de data zo veel mogelijk beschikbaar is;
- De PoC moet schaalbaar zijn omdat de data en het aantal eindgebruikers zullen groeien.

5.3 Potentiële bronnen

Voor het onderzoek zijn geschikte openbare bronnen nodig. In deze paragraaf wordt antwoord gegeven op de vraag: “Welke bronnen kunnen van toegevoegde waarde zijn?”

Binnen deze deelvraag worden alle onderzochte openbare bronnen behandeld waarbij ook uitgelegd wordt waarom deze in het onderzoek zijn meegenomen. De daadwerkelijke opbrengsten van geïntegreerde bronnen in de PoC zijn beschreven in hoofdstuk 6. Sommige specifieke bronnen die onder een gezamenlijk begrip vallen zijn samengepakt in een paragraaf.

5.3.1 IOT-scanners

Om kwetsbaarheden in ICT-services te vinden zoekt men met scanners welke services er aan het internet hangen. Er zijn een aantal open-source (met openbare broncode) scanners die gratis te gebruiken zijn voor iedereen. Het gebruiken van deze tools vergt echter expertise en de nodige inspanning en zijn ontwikkeld om op kleine schaal IP-adressen te scannen. Om het internet in kaart te brengen en om informatie over publiekelijke ICT-services laagdrempelig aan te bieden zijn er sinds 2009 ‘internet of things scanners’ (verder: IOT-scanners) ontstaan.

IOT-scanners zijn zoekmachines die het internet scannen voor verbonden apparaten. Apparaten kunnen servers zijn maar bijvoorbeeld ook netwerkapparatuur of camera’s. Via zoekfilters kan er gericht gezocht worden naar bijvoorbeeld een specifieke service of reeks IP-adressen. Deze scanners worden steeds populairder omdat ze verschillen met traditionele kwetsbaarheidsscanners doordat ze op globale schaal scannen en snel doorzoekbaar zijn. Onder de motorkap gebruiken ze open-source scanners en structureren ze de resultaten van de scans om deze data vervolgens aan te bieden aan eindgebruikers. Een kwaadwillende zou de informatie kunnen misbruiken om zwakheden te vinden. Daarom is het voor organisaties van belang om als preventieve maatregel deze scanners te gebruiken om hun eigen systemen te controleren. De IOT-scanners die hier aan bod komen zijn Shodan, Censys en Zoomeye. Shodan en Censys werden al handmatig gebruikt door verschillende SURFnet medewerkers en Zoomeye is aangeleverd door Victor Gevers van GDI Foundation.

Shodan.io
Shodan is de eerste IOT-scanner op het internet die in 2009 is opgezet door John Matherly. Deze verzamelt informatie door de banners van services op te vragen. “*Een banner is tekstuele informatie dat een service op een apparaat beschrijft*” [3]. Met zoekfilters kunnen er vervolgens specifieke banners opgevraagd worden aan Shodan.

Shodan biedt een Application Programming Interface (verder: API) aan die ervoor zorgt dat er data opgevraagd kan worden via programmacode. Ook is er een Python library (bruikbare computercode) beschikbaar om met de API te communiceren. Python is een programmeertaal die gebruikt wordt binnen dit project.

Om optimaal gebruik te kunnen maken van Shodan als bron is de afname van een abonnement noodzakelijk. Hiermee heeft men toegang tot alle filters, ontvangt men maandelijks krediet om meer zoekopdrachten uit te kunnen voeren en kan men gebruik maken van extra functionaliteiten. Voor dit onderzoek is er een gratis ‘Educational API plan’ aangevraagd dat beschikbaar is voor geverifieerde studenten. Deze is met name nodig informatie van grotere IP-reeksen te kunnen opvragen. E-mail ondersteuning is aanwezig en de API-documentatie is compleet en wordt netjes bijgehouden. Om deze redenen is Shodan geïntegreerd in de PoC. Na dit onderzoek zal SURFnet een abonnement bij Shodan moeten afnemen. De opdrachtgever heeft al aangegeven dat dit geen probleem is.

Censys
Censys is een IOT-scanner die in tegenstelling tot Shodan geheel gratis is. Ook Censys biedt een API en Python library aan. Om restricties op te heffen is dit project geverifieerd bij Censys. Ook zijn er voor dit onderzoek per e-mail enkele vragen gesteld. In beide gevallen is er snel vanuit Censys gereageerd, en het feit dat Censys als gratis dienst snel kosteloos support levert draagt bij aan de bruikbaarheid van deze IOT-scanner voor dit onderzoek. Een bijkomend voordeel is dat het opvragen van de data via SQL gaat, waardoor het mogelijk is om flexibele zoekopdrachten uit te voeren. Waar Shodan op dit moment zo'n 50 zoekfilters ondersteunt kan er met Censys gezocht worden op 2390 velden. Door deze redenen is Censys ook geïntegreerd in de PoC.

Zoomeye
Een andere grote IOT-scanner is Zoomeye, een Chinese zoekmachine die sinds 2013 actief is. Zoomeye is net als Censys in zijn geheel gratis en heeft ook een API beschikbaar. De makers onderhouden echter geen Python library.

De voordelen van Zoomeye zijn naast het gratis gebruik dat het ook webapplicaties identificeert en dat het volgens Victor Gevers meer resultaten kan hebben omdat Zoomeye scanners minder vaak actief geblokkeerd worden dan Shodan scanners. Dit komt omdat Zoomeye nog niet zo bekend is.

Tijdens het uitproberen van deze bron is opgevallen dat Zoomeye (nog) niet alles op orde heeft. Zo zijn er een aantal problemen geweest bij de registratieprocedure door slordige fouten in de website van Zoomeye. Op e-mails die er gestuurd zijn naar aanleiding van deze problemen wordt niet gereageerd. Ook liep het Zoomeye account dat nodig is voor dataverzoeken al gauw tegen limieten aan. Hun website geeft aan dat er dan een aanvraag per e-mail gestuurd kan worden, maar ook hier blijft een reactie uit.

5.3.2 Databasedumps

Organisaties maken gebruik van databases om hun gegevens digitaal op te slaan voor latere raadpleging. De meeste databases bevatten deels gevoelige gegevens die afgeschermd moeten worden, zoals intellectuele eigendommen en persoonsgegevens. Van databases worden databasedumps gemaakt voor back-up doeleinden. Het komt voor dat deze databasedumps publiekelijk beschikbaar worden. Soms gebeurt dit omdat een medewerker de databasedump per ongeluk vrij te downloaden aanbiedt aan het internet. Vaker gebeurt dit met intentie, bijvoorbeeld door een wraakactie van een (ex-)medewerker of door criminele hackers die onrechtmatig toegang hebben verkregen en vervolgens de databasedump stelen. De hackers verkopen deze op het internet in het criminele circuit. Uiteindelijk lekken deze gegevens vaak zover uit dat ze het nieuws bereiken. Dan verschijnen de databasedumps op het gewone internet.

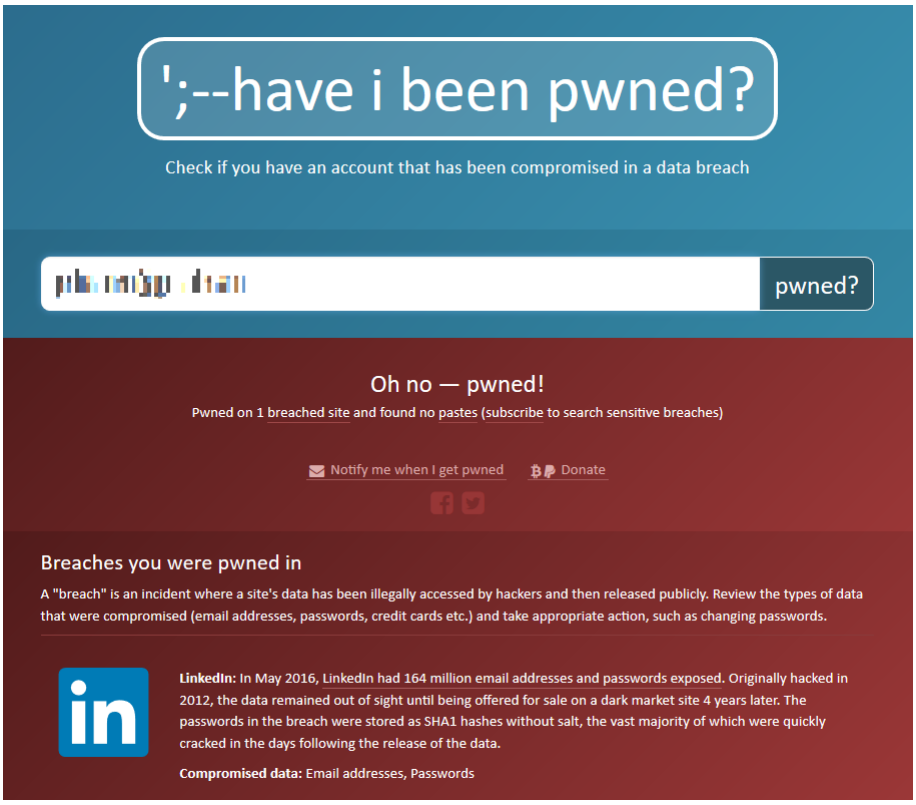
Het bovenstaande probleem vormt een bedreiging voor organisaties. Datalekken kunnen leiden tot imagoschade en/of economische schade. Bovendien kunnen uitgelekte gegevens van medewerkers misbruikt worden om daarmee in te breken op andere systemen. Veelal worden hiervoor de opgeslagen wachtwoorden in de databasedumps gebruikt. Sinds 2016 kan het College Bescherming Persoonsgegevens boetes opleggen aan organisaties wanneer daar data gelekt is en de organisatie blijkt geen passende maatregelen te hebben genomen om de persoonsgegevens te beveiligen. In de praktijk zien we dat veel organisaties hier niet aan voldoen en dit wordt pas zichtbaar wanneer de schade al geleden is. Inloggegevens komen dan op straat te liggen, en uit een recent internationaal onderzoek is gebleken dat 61% van de ondervraagden hun wachtwoorden hergebruiken terwijl ze weten dat dit risico's met zich meebrengt [4]. Dit betekent dat organisaties ingelicht moeten worden als

inloggegevens van hun medewerkers voorkomen in zo'n lek, aangezien er dan een aannemelijke kans is dat medewerkers op het werk dezelfde inloggegevens gebruiken.

Er bestaan verschillende websites op het internet die deze uitgelekte databasedumps verzamelen om hiermee een service aan te bieden. Eindgebruikers kunnen er controleren of hun e-mailadressen erin voorkomen. Het nadeel hiervan is dat ook kwaadwillenden de service kunnen gebruiken, maar dan om gericht wachtwoorden op te zoeken van andere e-mailadressen. Het enige wat organisaties hieraan kunnen doen is eerder van een datalek op de hoogte zijn. Daarom volgen nu een aantal van deze door de opdrachtgever aangeleverde databasedump services en wordt er gekeken of het de moeite waard is om deze te integreren in de PoC.

HaveIBeenPwned

De bekendste van deze services is 'HaveIBeenPwned?' (verder: HIBP). Deze website is in 2013 gemaakt door de Australische web security expert Troy Hunt en bevat over een verzameling van 1.8 miljard accounts (oktober 2016) afkomstig van databasedumps. Gebruikers kunnen via de website door deze verzameling zoeken door een e-mailadres of gebruikersnaam in te voeren. Alle functionaliteiten die HIBP aanbiedt zijn gratis. In figuur 2 is een voorbeeld te zien van een zoekresultaat op de website.



Figuur 2: Een zoekresultaat op de website van HaveIBeenPwned.com.

Nieuwe datalekken worden steeds toegevoegd en gebruikers kunnen zich met hun e-mailadres abonneren waardoor ze een e-mail krijgen zodra er een nieuwe lek met hun e-mailadres erin bekend is. Ook kan er met een compleet domein geabonneerd worden. Bij dit laatste moet er wel geverifieerd worden dat de abonnee eigenaar is van het betreffende domein. Deze functionaliteit is het meest waardevol voor de doelgroep van dit onderzoek, aangezien SURFnet en aangesloten instellingen na

een simpel verificatieproces in de toekomst altijd op de hoogte zullen zijn wanneer de website een nieuw relevant lek vindt. Dit maakt de integratie van HIBP in de PoC overbodig.

Leakedsource
LeakedSource is een soortgelijke website als HIBP. De eigenaars beweren over meer accountgegevens te beschikken dan HIBP, dat het interessant maakt voor verder onderzoek. Gebruikers kunnen hier ook hun e-mailadres controleren, maar om te zien over welke data het precies gaat wordt er om een betaling gevraagd. Ook is er tegen betaling een API beschikbaar voor organisaties. Dit zou voor SURFnet interessant kunnen zijn. Er wordt onderscheid in prijs gemaakt aan de hand van het aantal verzoeken en het gebruik van wildcard functionaliteiten. Met het laatste kan men voor een heel domein tegelijkertijd zoeken.

Leakbase
Leakbase is een website die dezelfde functionaliteiten aanbiedt als HIBP en LeakedSource. Begin 2017 beweert Leakbase een verzameling van 1,9 miljard accounts te hebben. Ook zij hebben een e-mail notificatie service beschikbaar, maar om resultaten te zien is een abonnement nodig. De abonnementen zijn niet duur; een maand kost $18 (de prijs in januari 2017). Overigens geeft de website nergens aan dat het zoeken naar complete domeinen mogelijk is.

@dumpmon
Cybercriminelen die gehackte databasedumps willen verkopen zetten vaak een klein stukje van deze dump op pastebin websites. Dit zijn type websites waar anoniem tekst online gezet kan worden. Zo kan een hacker aan zijn potentiële klanten bewijzen dat hij daadwerkelijk een databasedump heeft.

In mei 2013 heeft beveiligingsonderzoeker Jordan Wright een Twitter bot (computerprogramma dat geautomatiseerd taken uitvoert) genaamd ‘@dumpmon’ online gezet. Deze bot houdt de pastebin websites ‘Pastebin’, ‘Slexy’ en ‘Paste’ in de gaten voor stukjes databasedumps van hackers door op trefwoorden te zoeken als e-mailadressen en wachtwoorden. Een voordeel van deze bot is dat het geautomatiseerd kleinere stukjes databasedumps vindt, terwijl de mensen achter de andere websites zich focussen op alleen de grote vangsten. Zo vindt @dumpmon data die de andere websites vaak niet hebben.
Jordan Wright heeft de Python code van @dumpmon online gepubliceerd onder de MIT-licentie. Deze licentie staat toe dat de code door iedereen te gebruiken en aan te passen is. SURFnet zou deze bot dus zelf kunnen uitvoeren en de bevindingen ervan in de PoC kunnen laten weergeven.

Concluderend zijn er verschillende websites die kunnen uitzoeken of accountinformatie voorkomt in een uitgelekte databasedump. HIBP doet dit gratis terwijl er bij Leakedsource en Leakbase een abonnement nodig is. Het voordeel van Leakedsource is dat ze over de meeste accountdata lijken te beschikken. Aangezien HIBP en Leakedsource email notificaties kunnen sturen wanneer een e-mailadres uit een domein gevonden is, is integratie in de PoC van deze websites overbodig. De Twitter bot @dumpmon zou wel nog een toevoeging kunnen zijn aangezien veel van de data die @dumpmon vindt niet in HIBP, Leakedsource en Leakbase te vinden is.

5.3.3 IpInfo

IpInfo is een tool die van een gegeven IP-adres informatie verzamelt uit verschillende publieke bronnen. Op het moment van schrijven zijn de volgende bronnen in IpInfo geïntegreerd:

- Real-time Blackhole Lists: geeft aan of het IP-adres actief misbruikt wordt;
- Whois: geeft informatie over de beheerder van het IP-adres;
- Geolite: geeft geografische coördinaten;
- Check.torproject.org: checkt of het IP-adres onderdeel is van het Tor netwerk, een open netwerk voor anonieme communicatie.

Het toevoegen van IpInfo heeft een aantal voordelen. Het zorgt ervoor dat de maker Remco Verhoef betrokken wordt in het project. Zo kan Remco vanuit zijn expertise advies geven en IpInfo verbeteren ten behoeve van de PoC. Een bijkomend voordeel is dat de PoC direct profiteert van alle toekomstige uitbreidingen aan IpInfo.

Time	type	ip	reverse	exit_node	location.geo	blacklist
▸ December 13th 2016, 11:40:41.930	ipinfo	194.53.92.60	gasten-en-bezoekers.etz.nl.	false	{ "lon": 5.1063, "lat": 51.5769 }	{ "text": "Client host blocked using /?r=1&ip=194.53.92.60", "list": "b.barracudacentral.org" }, { "text": "Spam Received See: http://

Figuur 3: Een voorbeeld van IpInfo informatie over een IP-adres.

5.4 Het persisteren van de data

In deze paragraaf wordt antwoord gegeven op de volgende deelvraag: “Hoe wordt informatie uit de bronnen gepersisteerd zodat deze effectief geraadpleegd kan worden?”

Bij het persisteren van de informatie moet met de volgende dingen rekening gehouden worden:

- De PoC moet effectief te raadplegen zijn, dus de informatie moet snel getoond kunnen worden;
- De data moet zo veel mogelijk beschikbaar zijn, dus de PoC moet stabiel zijn;
- Bij het groeien van de data en eindgebruikers moet de PoC schaalbaar zijn.

Tijdens het onderzoek is duidelijk geworden dat Elasticsearch voldoet aan bovenstaande eisen. Elasticsearch is een open source op tekst gebaseerde zoekmachine van Elastic. Op het moment van schrijven staat het bovenaan in de ranglijst van zoekmachines door DB-Engines [5]. Bij SURFnet wordt het al in verschillende projecten gebruikt. De kennis is dus aanwezig voor ondersteuning bij het bouwen van de PoC en het onderhoud wanneer deze in gebruik genomen is. Ook Sjors Haanen heeft al enige ervaring met Elasticsearch opgedaan bij zijn opleiding. Daarbij heeft Elasticsearch onderscheidende kenmerken die aansluiten op de eisen die bij dit vraagstuk horen. Het kan heel snel zoekresultaten weergeven. Daarbij is de software schaalbaar: wanneer de hardware tegen zijn grenzen aanloopt kunnen er gemakkelijk machines aan toegevoegd worden of kan een bestaande machine een hardware upgrade krijgen. Zo ontstaat er een cluster van Elasticsearch nodes (actieve Elasticsearch programma’s) die de software draaiende houden. Het clusteren heeft een positief effect op de snelheid omdat de nodes het werk onder elkaar verdelen. Bovendien draagt het bij aan de stabiliteit en daarmee de beschikbaarheid van de data. Nodes kunnen namelijk ook replica’s van elkaars data bevatten dat zorgt voor redundantie en minder kans op uitval bij het wegvallen van een node.

Elasticsearch wordt samen met de programma’s Logstash, Kibana en Beats gratis aangeboden onder de naam ‘Elastic Stack’. Logstash wordt gebruikt om data uit verschillende bronnen te structureren en te verrijken alvorens het naar Elasticsearch gestuurd wordt. Kibana is een grafische gebruikersomgeving die de data in Elasticsearch kan visualiseren en komt aan bod in paragraaf 5.5. Beats is een lichtgewicht programmaatje dat data naar Logstash kan verzenden. Deze is echter niet nodig voor dit project omdat we geen data opvragen uit onze eigen computers.

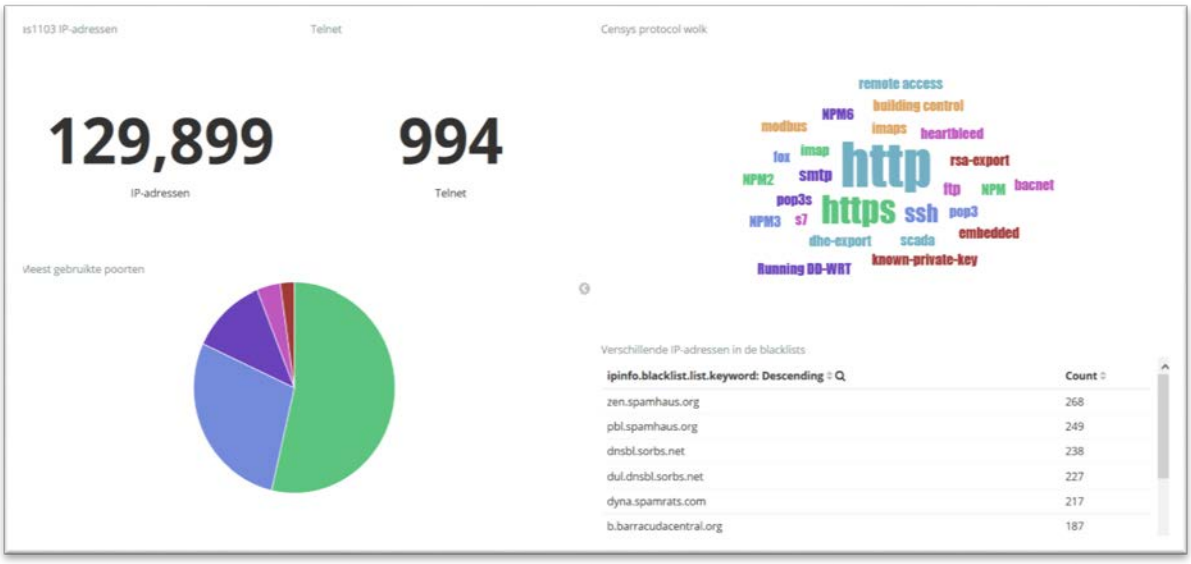
Wanneer SURFnet de PoC van dit project in de toekomst wil uitrollen is de aanschaf van de betaalde X-Pack uitbreiding van Elastic Stack zeer gewenst al dan niet noodzakelijk. De voordelen van X-Pack voor deze PoC zijn in bijlage 3 beschreven.

5.5 Software voor de gebruiksomgeving

In deze paragraaf wordt de volgende deelvraag behandeld: “Welke producten kunnen zorgen voor een gebruikersomgeving die voldoet aan de wensen van de belanghebbenden en hoe staan die producten in relatie tot elkaar?”

Er is software nodig die de data uit Elasticsearch zo kan visualiseren dat de eindgebruikers een gebruikersvriendelijk dashboard te zien krijgen. Een van de wensen van de opdrachtgever is dat de informatie laagdrempelig en overzichtelijk getoond wordt. In de eerste instantie moet er dus een algemeen overzicht te zien zijn waar vervolgens op specifieke elementen verder ingezoomd kan worden.

Omdat voor persistentie al gebruik gemaakt wordt van Elasticsearch ligt Kibana voor de hand. Kibana is een data visualisatie aanvulling voor Elasticsearch om de data in Elasticsearch te doorzoeken, bekijken en te visualiseren. Het doorzoeken van de Elasticsearch data kan met verschillende zoekfilters. Daarnaast kan er ook per veld gesorteerd worden. De belanghebbenden die geïnteresseerd zijn in het grotere plaatje en de details niet hoeven te zien kunnen gebruik maken van de dashboards in Kibana. Deze dashboards zijn schermen waarin visualisaties geplaatst kunnen worden als cirkeldiagrammen, histogrammen, grafieken, tabellen en gemiddelden. Deze zijn volledig configureerbaar en eenvoudig in gebruik. De belanghebbenden die geïnteresseerd zijn in het grotere plaatje hoeft de details niet te zien. Kibana kan deze wens vervullen met dashboards.



Figuur 4: Een Kibana dashboard met verschillende visualisaties.

De technische belanghebbenden willen de data uit de PoC nader kunnen analyseren. Bijvoorbeeld naar aanleiding van opvallende bevindingen afkomstig van de dashboards. Zij kunnen dan naar de ‘discovery’ pagina navigeren waar per IP-adres alle beschikbare informatie nader onderzocht kan worden. Zoekfilters kunnen blijven gelden bij het navigeren tussen beide pagina’s. Heb je bijvoorbeeld in het dashboard gefilterd op een bepaalde service, dan zie je in de discovery pagina alleen de IP-adressen die deze service aanbieden.
Kibana kan via X-Pack uitgebreid worden met verschillende functionaliteiten die een aantal wensen van de belanghebbenden kunnen vervullen. Een van die wensen is dat er verschillen op termijn

gezien kunnen worden om te kijken of er verbetering is qua beveiliging. Dit is mogelijk met de ‘reporting’ functionaliteit.

Ten slotte kan Kibana met de opslagmogelijkheid van zoekfilters focussen op het tonen van de meest relevante en belangrijkste kwetsbaarheden. Eindgebruikers kunnen in de toekomst nieuwe filters voor een kwetsbaarheid opslaan zodat anderen deze ook kunnen gebruiken. Zo kunnen ze zelf altijd nieuwe kwetsbaarheden detecteren.

5.6 Koppelen van informatie

In deze paragraaf wordt beschreven hoe de informatie uit de verschillende bronnen door koppeling in Elastic Stack verrijkt kan worden. Daarmee wordt antwoord gegeven op de deelvraag: “Welke koppelingen van de informatie uit de bronnen kunnen de kwaliteit van de security intelligence verhogen?”

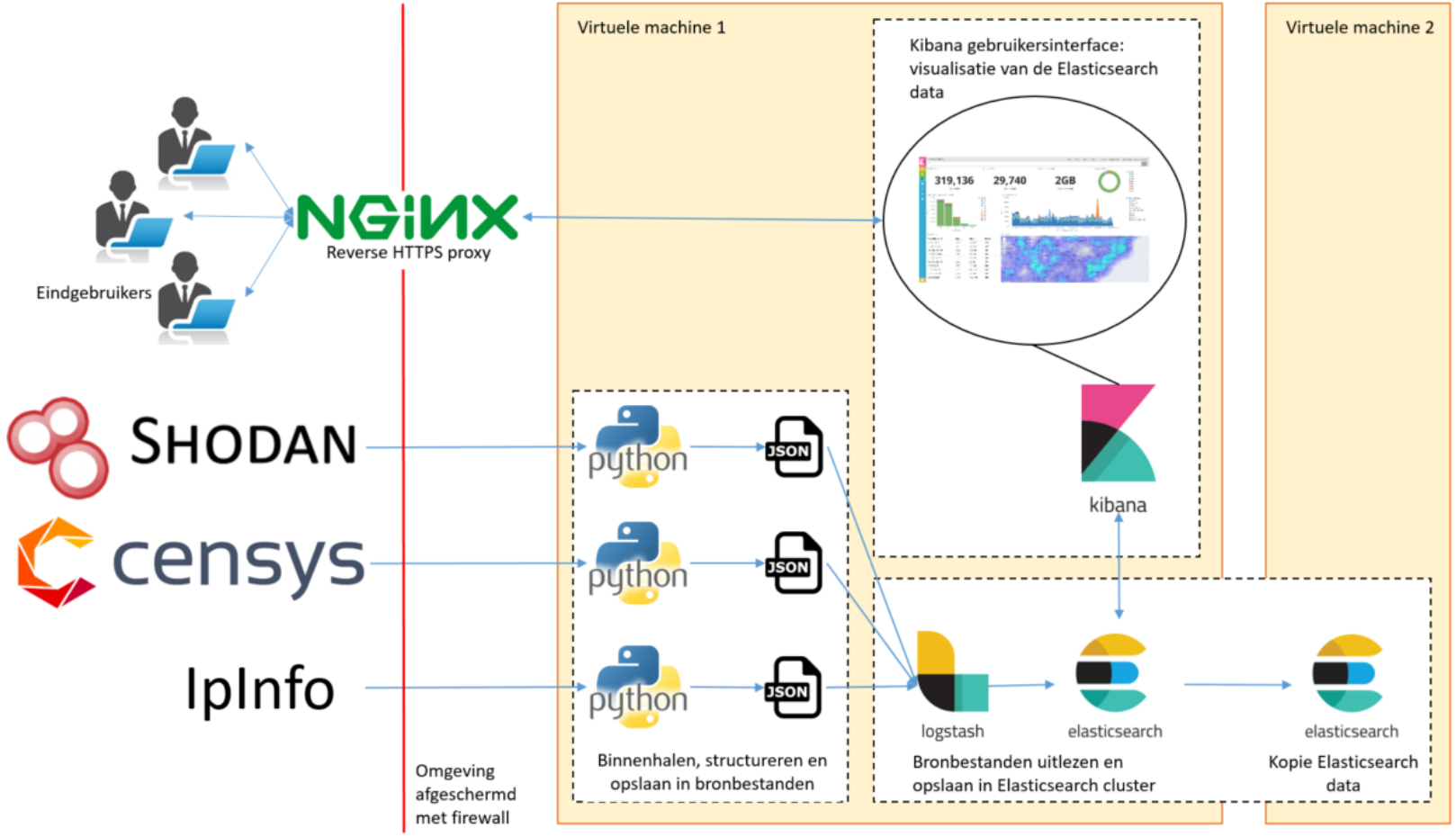
Initieel werd de data in de PoC niet gecombineerd tot één element. Zoeken op een specifiek IP-adres resulteerde in verschillende resultaten voor elke bron. Tijdens het bouwen van de PoC werden de nadelen hiervan duidelijk. De data uit verschillende bronnen werden wellicht onder elkaar getoond in één tool, maar de scheiding zorgde ervoor dat raadpleging in Kibana alsnog afzonderlijk gebeurde. Ook was het zoeken ingewikkelder en minder effectief omdat zoekfilters specifiek per bron waren doordat elke bron een andere structuur had. Deze bevindingen hebben geleid tot de beslissing om alle elementen per uniek IP-adres te combineren tot één element. Zie figuur 5 voor een voorbeeld. Dit betekent dat alle bronnen geconverteerd moeten worden naar één structuur. Daarmee is ook besloten dat het structureren met Python scripts moet gebeuren. Argumentatie over de keuze van Python staat in bijlage 4. Eerst gebeurde de structurering in Logstash, maar het is duidelijk geworden dat Python flexibeler is. Zo wordt de data na het ophalen dus al direct door hetzelfde script gestructureerd alvorens het in een bestandje in JSON-formaat weggeschreven wordt. JSON is een gestandaardiseerd gegevensformaat waarin Shodan, Censys en IpInfo de data ook al aanbieden. Ook Elasticsearch kan goed met dit formaat overweg. Een bijkomend voordeel van het structureren in Python is dat er in dit proces al direct data verwijderd kan worden die niet relevant is. Dit resulteert in minder gebruikte opslagruimte van de bronbestanden voor Elasticsearch en een snellere verwerkingstijd door Logstash.

ip	censys.tags	ipinfo.reverse	shodan.https.product	shodan.isp
131.174.70.38	http, https, imaps, pop3s	mail.ru.nl., autodiscover.ru.nl.	Microsoft IIS	Radboud Universiteit Nijmegen

Figuur 5: Een IP-adres met informatie uit elke bron.

De koppeling van informatie uit verschillende bronnen is een belangrijk onderdeel geweest voor dit onderzoek. Koppelingen leiden namelijk tot verrijking van informatie omdat de bronnen elkaar aanvullen. Ook kunnen hiermee false positives vermeden worden omdat verschillende bonnen elkaar kunnen bevestigen. Daarbij kunnen soortgelijke velden van verschillende bronnen gesynchroniseerd worden tot één veld dat leidt tot gemakkelijker zoeken en minder redundantie (het overbodig vaker opslaan van dezelfde data).

6 Opbrengst Proof of Concept



Figuur 6: De demo omgeving van de PoC.

De PoC is gemaakt aan de hand van de functionele analyse. Het ontwerp van de PoC is gevisualiseerd in figuur 6. Het proces begint bij de Python scripts die data uit Shodan, Censys en Ipinfo halen. Deze scripts kunnen handmatig of geautomatiseerd in termijnen uitgevoerd worden en staan onder versiebeheer op Github [6]. Meer informatie over het project op Github is te vinden in bijlage 4. De scripts converteren de binnengehaalde data naar een (JSON-) formaat die direct geaccepteerd wordt door Elasticsearch. De resulterende gestructureerde data wordt opgeslagen in tekstbestanden. Logstash houdt deze bestanden in de gaten en stuurt nieuwe data direct door naar de Elasticsearch cluster. Ten behoeve van de stabiliteit bestaat deze cluster uit 2 identieke Elasticsearch nodes waarvan er eentje op een aparte VM staat. De data uit Elasticsearch wordt voor de eindgebruikers gevisualiseerd in Kibana. Tussen Kibana en de buitenwereld met eindgebruikers draait een NGINX reverse proxy. Deze zorgt ervoor dat eindgebruikers Kibana kunnen bereiken via een veilige HTTPS verbinding. Dit wil zeggen dat de verbinding tussen Kibana en de eindgebruikers met versleuteling beveiligd wordt tegen afluisteren en manipulatie. Eindgebruikers kunnen Kibana raadplegen voor een algemeen overzicht van kwetsbaarheden middels interactieve visualisaties. Tevens kan alle beschikbare informatie over een IP-adres op detailniveau bekeken worden.

De bronnen Zoomeye en @dumpmon zijn in het onderzoek uitgekomen als geschikte bronnen voor de PoC. Bij Zoomeye is er vanuit de website beheerders nooit gereageerd op de vraag of restricties opgeheven worden, waardoor Zoomeye niet meegenomen kon worden in dit afstudeerproject. De daadwerkelijke integratie van @dumpmon paste uiteindelijk niet meer in het tijdsbestek van dit project, maar de onderzoeksresultaten wijzen erop dat het in de toekomst zeker waard is om deze bron alsnog te integreren.

7 Conclusie

In dit hoofdstuk worden de hoofdvraag en deelvragen en daarmee de probleemstelling kernachtig beantwoord.

“Hoe kan informatie uit openbare bronnen worden gekoppeld en welke eisen worden daaraan gesteld door de belanghebbenden, zodat de kwaliteit van de security intelligence van SURFnet en de aangesloten instellingen wordt verbeterd?”

Onder de bruikbare openbare bronnen die onderzocht zijn vallen in ieder geval Shodan, Censys, IpInfo, Zoomeye en @dumpmon.

Deze bronnen kunnen worden gekoppeld door het verzamelen van alle informatie in een Elastic Stack omgeving waarin de informatie uit alle bronnen wordt gestructureerd per IP-adres. De informatie uit @dumpmon is echter niet altijd gebonden aan een IP-adres, en manieren om deze informatie te koppelen kan dus input zijn voor vervolgonderzoek. De kwaliteit van de security intelligence van SURFnet en de aangesloten instellingen kan vervolgens verbeterd worden door de visualisatie dashboards in Kibana te raadplegen en steeds door security experts filters te maken voor nieuwe ICT-kwetsbaarheden.

De kern van eisen van de belanghebbenden is als volgt:

1. Er kan op twee niveaus naar de data gekeken worden:
    a. Het kan een integraal beeld geven;
    b. Er kan vervolgens op detail gekeken worden naar specifieke elementen.
2. False positives moeten zoveel mogelijk vermeden worden.

De eis over de false positives is deels opgelost door het combineren van verschillende bronnen in één IP-adres element. Zo is op detailniveau in één opslag te zien of verschillende bronnen elkaar kunnen bevestigen doordat de informatie verrijkt is en bij elkaar staat. In de toekomst zullen door de toevoeging van nieuwe bronnen specifieke kwetsbaarheden nader geanalyseerd worden om deze te valideren. Hiermee wordt de kwaliteit van de security intelligence nog verder verbeterd.

Verdere filtering van false positives ligt bij de deskundigheid van de eindgebruiker. Er is namelijk alleen informatie gebruikt die vanuit de buitenwereld te zien is, en sommige false positives kunnen alleen gefilterd worden wanneer er ook informatie over de afgeschermde delen van een netwerk bekend is. Het is dus belangrijk dat security experts de juiste zoekopdrachten maken en delen omdat zij kennis hebben over de herkenning van false positives.
Vervolgens kunnen bij gevonden en bevestigde kwetsbaarheden de beheerders van de bijbehorende IP-adressen of eigenaars van de gelekte data op de hoogte gebracht worden. Met het periodiek aanroepen van Python scripts kan er op termijn gekeken worden of meldingen van kwetsbaarheden verdwijnen en of de situatie daadwerkelijk verbeterd.

Het maken van filters voor nieuwe kwetsbaarheden en het detecteren en dichten hiervan zouden besproken kunnen worden in de SCIRT-bijeenkomsten, dat ook ten goede komt aan de security intelligence van de gehele SURFnet community.

8 Evaluatie

Uit de kritische zelfreflectie voor de afstudeerstage bleek dat ik me nog kon verbeteren in de prestatie-indicator ‘onderbouwing en verantwoording’. Ik had aan de start van een project soms moeite met het helder formuleren van het hoofdprobleem, de hoofdvraag en de huidige en gewenste situatie. Dit komt omdat ik me op dat moment nog niet goed een beeld kon schetsen van die situaties en zo het gevoel kreeg in het diepe gegooid te worden.
In dit project heb ik de gewenste situatie uitvoerig besproken met de opdrachtgever en goed geanalyseerd wat de huidige situatie was alvorens aan het project te beginnen. Zo heb ik tijdens het maken van mijn projectplan al verschillende belanghebbenden van mijn project geïnterviewd zodat ik de probleemstelling beter kon formuleren. Ook heb ik de opdrachtomschrijving van mijn projectplan laten reviewen door verschillende SURFnet collega’s dat ook heeft bijgedragen aan een betere formulering. Ten slotte mag het projectplan nog aangepast worden tijdens de uitvoering, en dit heb ik gedaan (resulterend in projectplan versie 1.1) om de scope en focus van de opdracht te veranderen van ‘bronnen’ naar ‘openbare bronnen’. Zo kwam ik uiteindelijk beter uit met het tijdsbestek dat ik had, en heb ik met mijn onderzoek en PoC kunnen laten zien welke informatie over een organisatie direct te zien is vanuit de buitenwereld.
Een verbeterpunt dat ik vanuit mijn ervaring bij SURFnet mee ga nemen is dat ik in het vervolg de scope van de opdracht van tevoren beter ga inschatten zodat ik aan het begin van een project niet te veel beloof in mijn projectplan. Zo komt het resultaat uiteindelijk beter overeen met wat afgesproken is zonder dat ik deze scope tussendoor hoef te veranderen.

Bibliografie

[1] Proeven van onderzoek. Geraadpleegd op 12 september 2016, van http://www.ralphniels.nl/pubs/jacobs-proevenvanonderzoekboek.pdf.

[2] Competent afstuderen met het Tien Stappen Plan. (2007, februari). Geraadpleegd op 14 september 2016, van http://hoadd.noordhoff.nl/sites/7745/_assets/7046d91.pdf

[3] Complete Guide to Shodan. (2016, oktober). Geraadpleegd op 23 november 2016, van https://leanpub.com/shodan

[4] The password paradox and why our personalities will get us hacked. (2016, oktober). Geraadpleegd op 30 november 2016. http://prod.cdata.app.sprinklr.com/DAM/434/LastPass_ExecutiveSummary_fina-88e8a5a2-00cb-4a09-b363-e01a45f829d6-1389898992.pdf

[5] DB-Engines Ranking of Search Engines. (2016, november). Geraadpleegd op 7 november 2016, van http://db-engines.com/en/ranking/search+engine

[6] Vulnerabilityfinder. Geraadpleegd op 2 januari 2017, van https://github.com/sjorsng/vulnerabilityfinder

Bijlagen

Bijlage 1: Projectplan

Projectplan

Security intelligence door aggregatie open-source intelligence

SURFnet

Utrecht

Datum	**:**	**Datum**
Versie	**:**	**1.1**
Status	**:**	**Definitief**
Bestandsnaam	**:**	**Projectplan_Sjors_Haanen**
Auteur	**:**	**Sjors Haanen**

Revisies

Dit document kent de volgende versies:

Versie	Datum	Auteur(s)	Wijzigingen	Status
0.1	14-9-2016	Sjors Haanen	Eerste invulling	Concept
0.2	21-9-2016	Sjors Haanen	Feedback verwerkt van Rogier Spoor	Concept
0.3	27-9-2016	Sjors Haanen	Feedback verwerkt van Bart Bosma	Concept
1.0	7-10-2016	Sjors Haanen	Feedback verwerkt van Stefan Roijers en Bart Bosma	Compleet
1.1	23-11-2016	Sjors Haanen	De scope veranderd van ‘diverse’ naar ‘openbare’ bronnen	Compleet

Goedkeuring

Dit document kent de volgende goedkeuringen:

Versie	Datum	Aan
1.0	7-10-2016	Rogier Spoor Stefan Roijers
1.0	19-10-2016	Casper Schellekens
1.1	28-11-2016	Rogier Spoor Stefan Roijers

Review

Dit document is ter review voorgelegd aan de volgende personen:

Versie	Datum	Naam	Rol
0.1	14-9-2016	Rogier Spoor	Opdrachtgever
0.2	21-9-2016	Bart Bosma Xander Jansen Stefan Roijers	Informatieverstrekker Informatieverstrekker 1e assessor
0.3	27-9-2016	Bart Bosma Xander Jansen Stefan Roijers	Informatieverstrekker Informatieverstrekker 1e assessor
1.0	7-10-2016	Stefan Roijers Casper Schellekens	1e assessor 2e assessor
1.1	23-11-2016	Rogier Spoor Stefan Roijers Casper Schellekens	Opdrachtgever 1e assessor 2e assessor

Inhoudsopgave

Achtergrond

SURF

SURF is een ICT-samenwerkingsorganisatie van Nederlandse onderwijs- en onderzoeksinstellingen. Succesvolle innovatieprojecten op ICT-gebied worden gerealiseerd in landelijke ICT-voorzieningen zodat de aangesloten instellingen hier gebruik van kunnen maken. Daarnaast heeft SURF grote invloed gehad op het succes van de Amsterdam Internet Exchange (AMS-IX) en de software achter het DigiD systeem. SURF bestaat uit drie gespecialiseerde onderdelen die in dit hoofdstuk verder besproken worden.

SURFnet

SURFnet, gevestigd in Utrecht, richt zich op het stimuleren, ontwikkelen en exploiteren van een hybride netwerk, een vertrouwde identiteit en een samenwerkingsomgeving, en zorgt er hiermee voor dat onderzoekers, docenten en studenten kunnen samenwerken met behulp van ICT. Dit doet men door een dienstverlening op twee gebieden. Enerzijds zorgt SURFnet voor een netwerkinfrastructuur dat efficiënt datatransport realiseert. Anderzijds biedt SURFnet een samenwerkingsinfrastructuur aan die systemen, instrumenten en mensen verbindt.

SURFmarket

SURFmarket, gevestigd in Utrecht, is een licentieorganisatie die namens de aangesloten instellingen onderhandelt met ICT-leveranciers en uitgevers om campuslicenties af te sluiten voor software, content en hardware. Deze biedt SURFmarket vervolgens aan de aangesloten instellingen zodat zij kunnen profiteren van de best mogelijke voorwaarden voor de aanschaf van onder meer software, clouddiensten en digitale content.

SURFsara

SURFsara, gevestigd in Amsterdam en Almere, faciliteert wetenschappelijk onderzoek door diensten te leveren op het gebied van supercomputers, netwerken, dataopslag en hoogwaardige visualisatie. Oorspronkelijk alleen voor de Vrije Universiteit Amsterdam, de Universiteit Amsterdam en het Mathematisch Centrum. Tegenwoordig mogen ook andere universiteiten en onderzoeksinstituten gebruik maken van de diensten van SURFsara.

Doel van dit document

Dit projectplan heeft als doel om de afstudeeropdracht van Sjors Haanen bij SURFnet te specificeren. De eerste complete versie vereist goedkeuring door bedrijfsbegeleider Rogier Spoor en docentbegeleider Stefan Roijers en is een vereiste om te kunnen beginnen aan de onderzoeks- en oplossingsfase van het Tien Stappen Plan (TSP) van het afstudeertraject. Alle betrokken partijen kunnen dit projectplan raadplegen om gemaakte afspraken terug te lezen.

Projectopdracht

Aanleiding

Er zijn diverse openbare bronnen die informatie kunnen geven over kwetsbaarheden in computersystemen of gelekte informatie op het internet. Bij SURFnet is er behoefte aan een manier om met behulp van deze bronnen nieuwe kwetsbaarheden of gelekte informatie uit de systemen van SURFnet en de aangesloten instellingen overzichtelijk en laagdrempelig te tonen.

Gewenste situatie

De gewenste situatie is dat instellingen die aangesloten zijn bij SURFnet beschikking hebben over een snel in te zetten tool die openbare bronnen gebruikt om een analyse over kwetsbaarheden en gelekte informatie te maken en daarmee een compleet overzicht te geven over de bevindingen. Hierdoor worden instellingen bewust van bepaalde risico's en kunnen kwetsbaarheden tijdig gedicht worden om misbruik te voorkomen.

Doel van het project

Het doel van dit project is om de security intelligence van SURFnet en de aangesloten instellingen te vergroten met behulp van een tool die informatie uit verschillende openbare bronnen verzamelt en aggregeert. Onder security intelligence verstaan we inzicht op security risico's vergaard door het analyseren van verzamelde informatie.

De tool geeft een dagelijkse 'health check' van elke individuele instelling. De health check moet de beschikbare informatie over openstaande services, kwetsbaarheden en gelekte informatie laagdrempelig, uniform en per organisatie inzichtelijk maken. De focus ligt daarbij op het tonen van de meeste relevante en belangrijkste kwetsbaarheden waarbij false positives niet worden getoond.

Voorafgaand aan de tool wordt onderzoek gedaan naar de benodigde informatie om de tool te ontwerpen. Dat wil niet zeggen dat het onderzoek eindigt bij de start van het bouwen van de tool: Het onderzoek omvat namelijk ook het maken van de tool, maar het kent een zwaartepunt als vooronderzoek om de tool te kunnen realiseren.
De onderzoeksresultaten worden gerapporteerd in een onderzoeksrapport, dat onderdeel is van het eindresultaat. Het rapport geldt dan als een advies voor SURFnet. In de context van deze afstudeeropdracht zal de tool gelden als Proof of Concept (verder: PoC).

Doelstelling voor het onderzoeksrapport is onder andere om inzicht te geven in welke bronnen een bijdrage kunnen leveren aan het vergroten van de security intelligence van SURFnet en de aangesloten instellingen naar aanleiding van de eisen van belanghebbenden.
Als internet serviceprovider is het voor SURFnet van belang om het netwerk dat ze beheren zo schoon mogelijk te houden. Het doel voor de aangesloten instellingen is dat ze snel en accuraat geïnformeerd worden over bedreigingen.

Probleemstelling

Op dit moment zijn er een aantal openbare bronnen die nog niet standaard gebruikt worden bij SURFnet. Ook worden de bronnen niet geaggregeerd en moeten ze afzonderlijk geraadpleegd worden. Tevens zorgen afzonderlijke bronnen vaak voor false positives. Er kan nog geen integraal beeld gegeven worden van de kwetsbaarheden van een organisatie. Hierdoor wordt onnodig risico gelopen waarbij kwetsbaarheden makkelijk kunnen worden geëxploiteerd door criminele organisaties. Zij gebruiken namelijk dezelfde bronnen en beschikken over genoeg tijd en expertise om de hieruit vergaarde informatie te misbruiken. Tevens kunnen belanghebbenden nog wensen hebben over functionaliteiten die in de huidige situatie nog ontbreken. Concluderend mist er dus een manier om gevonden kwetsbaarheden uit verschillende openbare bronnen overzichtelijk weer te geven.

Onderzoeksplan

Het onderzoek onder andere gaat duidelijk maken welke soort informatie in ieder geval getoond moeten worden en welke bronnen de PoC gaat gebruiken. Het onderzoek zal tijdens het ontwerpen van de tool doorlopen en de resultaten hiervan worden gerapporteerd in het eindproduct ‘onderzoeksrapport’.

Hoofd- en deelvragen

Het onderzoek moet antwoord geven op de volgende hoofdvraag:

‘Hoe kan informatie uit openbare bronnen worden gekoppeld en welke eisen worden daaraan gesteld door de belanghebbenden, zodat de kwaliteit van de security intelligence van SURFnet en de aangesloten instellingen wordt verbeterd?’

Om een volledig antwoord te kunnen geven op de bovenstaande hoofdvraag zijn de volgende deelvragen opgesteld:

1. Wat zijn de wensen van de belanghebbenden met betrekking tot het verhogen van de kwaliteit van security intelligence?
2. Welke bronnen kunnen van toegevoegde waarde zijn?
3. Hoe wordt informatie uit de bronnen gepersisteerd zodat deze effectief geraadpleegd kan worden?
4. Welke producten kunnen zorgen voor een gebruikersomgeving die voldoet aan de wensen van de belanghebbenden en hoe staan die producten in relatie tot elkaar?
5. Welke koppelingen van de informatie uit de bronnen kunnen de kwaliteit van de security intelligence verhogen?

Deelvraag	Methode	Strategie*
3. Wensen belanghebbenden	Interviews met belanghebbenden	Veldonderzoek
4. Bronnen	Internetonderzoek, overleggen met Victor Gevers, experimenteren met gevonden bronnen in testomgeving	Biebonderzoek Werkplaatsonderzoek
5. Informatie persisteren	Internetonderzoek, experimenteren met software in testomgeving	Biebonderzoek Werkplaatsonderzoek

Deelvraag	Methode	Strategie*
6. Producten voor gebruikersomgeving	Internetonderzoek, experimenteren met (software)producten in testomgeving	Biebonderzoek Werkplaatsonderzoek
7. Koppelingen van informatie	Internetonderzoek, experimenteren in testomgeving	Biebonderzoek Werkplaatsonderzoek

***** Toelichting strategieën: http://www.ralphniels.nl/pubs/jacobs-proevenvanonderzoekboek.pdf

Ongetwijfeld zullen er als methode ook gesprekken plaatsvinden met medewerkers van SURFnet om van hun ervaringen gebruik te maken, bijvoorbeeld voor de keuze van bepaalde software.

Onderzoeksstrategie (methodiek)

De werkwijze van dit project is afgeleid van het 'Tien Stappen Plan' (Kempen & Keizer, 2006). Daarmee zijn stap 1 t/m 5 al deels doorlopen tijdens de sollicitatieprocedure. Zo heeft er voor de start van de stageperiode al een gesprek plaatsgevonden tussen de projectmanager, de opdrachtgever en de 1e assessor met als input een volledig ingevuld gespreksformulier om de opdracht al meer vorm te kunnen geven. Ook de gesprekken met de informatieverstrekkers en belanghebbenden tijdens de eerste weken van de stage vallen onder deze stappen. Stap 6 'werkplanning en projectorganisatie' betreft het schrijven van dit projectplan en zal eindigen met de vereiste goedkeuringen beschreven in paragraaf 'Fase Schrijven projectplan'. Daarna begint stap 7 'diepteonderzoek' met een onderzoek naar de hoofd- en deelvragen dat nodig is om SURFnet een onderbouwd advies te kunnen leveren en om de PoC te kunnen opzetten. Het maken van de PoC zal ook al starten in deze fase. Gevonden bronnen als input voor de PoC en software waar de PoC uit zal bestaan zullen namelijk uitgeprobeerd worden in de testomgeving om bevindingen in het onderzoek te onderbouwen. Stap 8 'oplossingsplan' zal in dit project het afronden van de PoC betekenen, waarbij nog steeds met de opdrachtgever gekeken wordt of dat de PoC de goede kant op gaat. Rond deze tijd wordt ook het afstudeerverslag geschreven. Stap 9 uit het Tien Stappen Plan is niet meegenomen aangezien de daadwerkelijke uitrol niet bij dit project hoort. Ten slotte wordt er in stap 10 'afronding en afstuderen' een presentatie met demo van de PoC voor SURFnet en een presentatie bij de afstudeerzitting gegeven.

Belanghebbenden

De belanghebbenden zijn de uiteindelijke eindgebruikers van de tool. Dit zijn de mensen die de over security gaan van de instelling waarvoor ze werken. Dit zullen voor het merendeel security officers zijn, maar dat staat echter niet vast. De instellingen zullen zelf kiezen wie van de tool gebruik gaat maken binnen hun organisatie. Ook kunnen mensen met een managersfunctie geïnteresseerd zijn in de 'health check' omdat dit een globaal overzicht kan geven op hoger niveau. Daarom vallen alle potentiële eindgebruikers onder de verzamelnaam 'belanghebbenden'.

Begrenzing

Tot het project behoort:	Tot het project behoort niet:
1 Projectplan	1 Het beschikbaar stellen van de tool aan de bij SURFnet aangesloten organisaties (dat wil zeggen: de daadwerkelijke uitrol van de PoC)
2 Onderzoeksrapport	2
3 PoC	3
4 Afstudeerverslag (scriptie)	4

Randvoorwaarden

Aan sommige bronnen die van toegevoegde waarde kunnen zijn voor de tool zijn kosten verbonden. Er is in overleg financiering beschikbaar voor de dekking hiervan. Mochten de kosten van een bron te hoog zijn dan kan deze bron niet gebruikt worden.

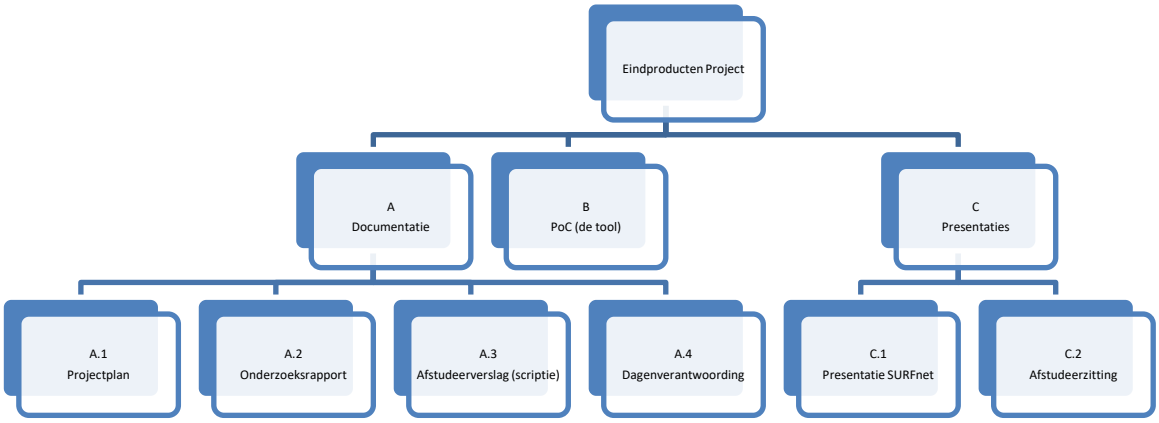
Uiteindelijk moet bij de uitrol de tool voldoen aan de volgende eisen:

- Medewerkers van de aangesloten instellingen bij SURFnet kunnen federatief inloggen op de PoC omgeving (via SAML);
- De tool voldoet aan de Handreiking Security bij het Juridisch Normenkader van SURF;
- De tool voldoet aan de relevante standaarden uit de lijst open standaarden van het forum standaardisatie;

In de scope van dit project is het wenselijk om vanaf het begin al rekening te houden met deze eisen bij het ontwikkelen van de PoC. Het is echter onwaarschijnlijk dat de PoC uit dit project in deze versie zonder aanpassingen (of overname in andere projecten als het GDI project) uitgerold zal worden. De bovenstaande randvoorwaarden zijn dus niet per se een vereiste om dit project succesvol af te sluiten.

Eindproducten

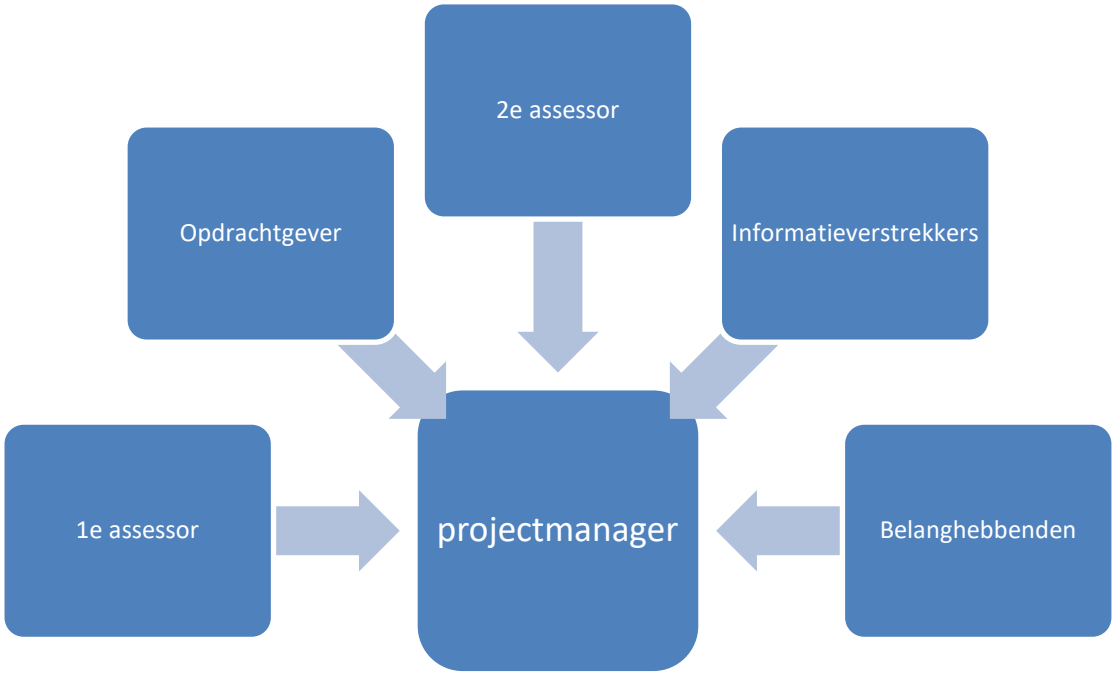
Product Breakdown Structure project



Product		Omschrijving
A.1	Projectplan	Het projectplan is dit document. Hierin staan alle afspraken omtrent de afstudeeropdracht.
A.2	Onderzoeksrapport	De resultaten van het onderzoek en tevens het onderzoeksaspect van de afstudeeropdracht. Dit en de PoC tellen mee voor 40% in de beoordeling van het afstuderen.
A.3	Afstudeerverslag (scriptie)	Een verslag over de opdracht, de werkzaamheden en de opbrengst voor SURFnet. Dit document telt voor 40% mee in de beoordeling van het afstuderen.
A.4	Dagenverantwoording	De dagen die besteed zijn aan de afstudeerstage, met een korte beschrijving van de werkzaamheden
B	PoC	De tool die ontworpen is aan de hand van de resultaten uit het onderzoeksrapport.
C.1	Presentatie SURFnet	Een presentatie voor SURFnet met een advies en een demonstratie van het PoC.
C.2	Afstudeerzitting	Presentatie en verdediging. Telt voor 20% mee in de beoordeling van het afstuderen.

Projectorganisatie

Teamleden & belanghebbenden



Persoon	Rol/taken	Beschikbaarheid
Sjors Haanen *+316 11293804* *s.haanen@student.fontys.nl*	*Projectmanager*	*5 dagen per week, gedurende het hele project*
Rogier Spoor *+31-887873000* *rogier.spoor@surfnet.nl*	*Opdrachtgever*	*5 dagen per week, gedurende het hele project*
Stefan Roijers *08850-77348* *s.roijers@fontys.nl*	*1e assessor*	*Op aanvraag en tijdens bedrijfsbezoeken*
Casper Schellekens *08850-73223* *c.schellekens@fontys.nl*	*2e assessor*	*Op aanvraag en tijdens afstudeerzitting*
Victor Gevers *victor@gdi.foundation*	*Informatieverstrekker en belanghebbende, met name voor bruikbare bronnen*	*Via Skype en op aanvraag*
Bart Bosma *+316 57877993* *Bart.bosma@surfnet.nl*	*Informatieverstrekker: vertegenwoordiger van SOC belangen bij aangesloten instellingen.*	*Op aanvraag*
Gijs Rijnders *+316 34095803* *Gijs.rijnders@surfnet.nl*	*Informatieverstrekker, met name voor de opslag van verzamelde data*	*Op aanvraag*
Xander Jansen *+31 887873000* *xander.jansen@surfnet.nl*	*Informatieverstrekker en belanghebbende*	*Op aanvraag*

Communicatie

Soort overleg	Frequentie	Aanwezig
Voortgangsbespreking met opdrachtgever	Op aanvraag (minimaal 1 keer per week)	Projectmanager Opdrachtgever
Gesprek met Bart Bosma	Op aanvraag	Projectmanager Bart Bosma
eerste bedrijfsbezoek	Eenmaal: In week 39 (lesweek 5)	Projectmanager Opdrachtgever 1e assessor
tweede bedrijfsbezoek	Eenmaal: In week 3 in 2017 (lesweek 18)	Projectmanager Opdrachtgever 1e assessor
Presentatie bij SURFnet	Eenmaal	Projectmanager Opdrachtgever Belangstellenden
Afstudeerzitting	Eenmaal: In week 4 of 5 in 2017 (lesweek 19 of 20)	Projectmanager Opdrachtgever 1e assessor 2e assessor Externe deskundige

Besluitvorming

De resultaten van het onderzoek vormen de basis voor besluitvorming binnen het project. Deze resultaten en besluiten zullen tussentijds in een voortgangsbespreking besproken worden zodat de opdrachtgever invloed kan hebben op de besluitvorming.

Wijzigingen in dit projectplan zullen eerst besproken worden in een voortgangsbespreking en vereisen goedkeuring van de opdrachtgever alvorens deze doorgevoerd mogen worden. Een wijziging resulteert in een nieuwe versie van dit document en zal opnieuw verspreid worden onder de belanghebbenden. Indien relevant zal de wijziging ook goedgekeurd moeten worden door de assessoren.

Activiteiten en tijdplan

Opdeling en aanpak van het project

Eerst wordt dit projectplan geschreven waarbij feedback verwerkt wordt van de assessoren en de opdrachtgever. Tijdens het wachten op feedback wordt er al een begin gemaakt aan het onderzoek. Tijdens het onderzoek zal ook al aan de PoC gewerkt worden zodra er genoeg informatie vergaard is om hieraan te beginnen. Ten slotte wordt het afstudeerverslag geschreven. De projectmanager krijgt feedback van de 1e assessor op een conceptversie hiervan. Ook kan er feedback gevraagd worden aan de opdrachtgever. Na het opleveren van de definitieve versie van het afstudeerverslag wordt de presentatie voor de afstudeerzitting voorbereid.

Overall tijdplan

Fasering	Start	Gereed
1 Schrijven projectplan	1-9-2016	In week 39 (lesweek 5, vóór eerste bedrijfsbezoek)
2 Onderzoek en PoC	Week 39	Week 3 in 2017 (lesweek 17)
3 Schrijven afstudeerverslag	Week 45 (lesweek 10)	Week 3 in 2017 (lesweek 17)
4 Voorbereiding presentaties	Week 4 in 2017 (lesweek 18)	Op datums van presentaties

Fase Schrijven projectplan

Omschrijving en aanpak

Het projectplan wordt geschreven door de projectmanager. Overleg met de opdrachtgever is noodzakelijk omdat hierin afspraken gemaakt worden voor het hele project. In deze fase wordt ook al overlegd met mensen waarmee wellicht samengewerkt kan worden. Ook is GDI foundation bezig met de opstart van een soortgelijk project.

Eindproducten

Het eindproduct van deze fase is het projectplan.

Startvoorwaarden

Deze fase kan van start gaan als de projectmanager aan het begin van de stage ingewerkt is en beschikking heeft over zijn eigen werkplek. Ook moet er een officiële goedkeuring zijn van de examenkamer om aan de afstudeerstage te beginnen.

Activiteitenlijst

Activiteit	Wie	Start	Gereed
1 Schrijven projectplan	Projectmanager	1-9-2016	In week 39 (lesweek 5, vóór eerste bedrijfsbezoek)
2 Feedback op projectplan	Opdrachtgever 1e assessor	/	/
3 Goedkeuring projectplan	Opdrachtgever 1e assessor 2e assessor	In week 39 (lesweek 5)	In week 39 (lesweek 5)

Fase Onderzoek en PoC

Omschrijving en aanpak

In deze fase wordt onderzoek gedaan naar de hoofd- en deelvragen en wordt tegelijkertijd de PoC ontworpen.

Eindproducten

De eindproducten van deze fase zijn het onderzoeksrapport en de PoC.

Startvoorwaarden

Na goedkeuring van het projectplan kan aan deze fase gestart worden.

Activiteitenlijst

Activiteit	Wie	Start	Gereed
4 Onderzoek doen	Projectmanager	Week 39	Week 3 in 2017 (lesweek 17)
5 PoC maken	Projectmanager	Week 39	Week 3 in 2017 (lesweek 17)

Fase Schrijven afstudeerverslag

Omschrijving en aanpak

In deze fase wordt het afstudeerverslag gemaakt.

Eindproducten

Het eindproduct van deze fase is het afstudeerverslag.

Startvoorwaarden

Het afstudeerverslag kan pas volledig afgerond worden wanneer het onderzoek klaar is.

Activiteitenlijst

Activiteit	Wie	Start	Gereed
6 Schrijven afstudeerverslag	Projectmanager	Week 45 (lesweek 10)	Week 3 in 2017 (lesweek 17)
7 Feedback op afstudeerverslag	Opdrachtgever 1e assessor	Week 46	Week 51 (lesweken 11 t/m 16)

Risico's en afhankelijkheden

Afhankelijkheden

Tot zover bekend is dit project niet afhankelijk van andere projecten. De medewerking van informatieverstrekkers en eventuele sleutelfiguren van bruikbare bronnen voor de PoC kunnen wel invloed hebben op de kwaliteit en de voortgang van het project.

Projecten die van dit project afhankelijk zijn

De resultaten van dit project kunnen wellicht bijdragen of overgenomen worden door het geplande project 'Internet & Cyber Security Health Check' van GDI Foundation. In dat project wil GDI Foundation hun huidige tool en informatieverstrekking verder optimaliseren zodat meerdere partijen instaat worden gesteld om tijdig adequate tegenmaatregelen te treffen tegen cyber crime mogelijkheden. De projectleden van GDI Foundation zijn benieuwd naar de resultaten van het onderzoek van dit project en kunnen de bijbehorende PoC wellicht gebruiken in hun eigen project. De projectmanager zal zorgen dat de PoC overdraagbaar is.

Risico's en uitwijkactiviteiten

Risico	Activiteiten ter voorkoming opgenomen in plan	Uitwijkactiviteiten
1 De doelstellingen van het project blijken niet specifiek genoeg. De verwachtingen van betrokkenen kunnen daardoor verschillen.	In de Voortgangsbesprekingen samen met de opdrachtgever waken of het project de goede kant uitgaat.	De doelstellingen in dit projectplan aanpassen in overleg met de opdrachtgever.
2 Er wordt langs elkaar gewerkt met soortgelijke projecten doordat er niet genoeg wordt samengewerkt. Het risico is dat dit project aan het einde weinig meerwaarde heeft omdat de andere projecten hetzelfde de probleemstelling al oplossen.	Contact blijven houden met soortgelijke projecten en samenwerken waar kan.	Opnieuw afstemmen met de opdrachtgever om ervoor te zorgen dat dit project uiteindelijk alsnog meerwaarde heeft.
3 De PoC blijkt niet volledig door de verkeerde inschatting in haalbaarheid.	Op tijd beginnen aan de PoC en van tevoren goed bespreken met de opdrachtgever welke functionaliteiten haalbaar zijn binnen de tijdsduur van dit project.	Zorgen dat in ieder geval het onderzoeksrapport genoeg inhoud heeft zodat SURFnet dit zelf kan gebruiken in toekomstige projecten.
4 De projectmanager mist de expertise om een onderdeel te voltooien.	De projectmanager vergaart zelf de expertise tijdens de stage benodigd om het onderdeel te kunnen voltooien.	Expertise wordt extern opgezocht, binnen SURFnet of bij een externe organisatie.

Bijlage 2: Uitwerkingen interviews

“Wat zijn de wensen van de belanghebbenden met betrekking tot het verhogen van de kwaliteit van security intelligence?” Deze deelvraag vereist volgens het projectplan interviews met belanghebbenden van de PoC. Deze interviews zijn hieronder uitgewerkt.

Ten eerste is Rogier Spoor, de opdrachtgever, een belanghebbende van de tool en zijn wensen zijn al in het hoofdstuk ‘projectopdracht’ van het projectplan verwerkt. Hieronder volgen de kernpunten uit Rogiers wensen voor de tool:

- Rogier wil graag niet meer alle bronnen afzonderlijk te hoeven raadplegen, maar ziet dit graag geautomatiseerd gebeuren. Het liefst ziet hij een tool die op aanvraag of automatisch periodiek scant en de resultaten toont;
- De tool moet de informatie laagdrempelig en overzichtelijk tonen. Dit zorgt ervoor dat de informatie daardoor ook geschikt is om aan het management van SURFnet of de instellingen voor te leggen. Op die manier kan men aangeven dat iets aandacht nodig heeft. Ook kan dan een algemene indruk geven van de kwetsbaarheden of een of ‘health check’ doen van de beveiliging van het netwerk;
- De belangrijkste kwetsbaarheden moeten in ieder geval getoond worden met zo weinig mogelijk false positives. Dit kan gerealiseerd worden door slim gebruik te maken van beschikbare functionaliteiten voor specifieke kwetsbaarheid detectie op bekende aanvallen. Een manier om de false positives te vermijden is om data uit verschillende bronnen met elkaar te vergelijken.

Victor Gevers is medeoprichter van GDI Foundation. Omdat GDI Foundation geïnteresseerd is in de PoC en omdat Victor veel technische details weet over het vinden van ICT-kwetsbaarheden is Victor ook gevraagd naar zijn wensen en ideeën voor de tool:

- Victor hoopt dat andere eindgebruikers de tool later gaan gebruiken om hun eigen kwetsbaarheden te vinden en zelf te dichten, zodat GDI Foundation zich op andere dingen kan richten. Op het moment doet hij namelijk aan responsible disclosure (het verantwoord melden) van kwetsbaarheden die men eigenlijk al eerder zelf had moeten vinden. Deze kwetsbaarheden zijn eenvoudig te vinden via de openbare bronnen waar de tool gebruik van gaat maken, en kunnen door de tool dus snel aan het licht gebracht worden;
- Ook Victor geeft aan dat het filteren van false positives een goed aandachtspunt is voor de tool.

Ewald Beekman is voorzitter van SCIRT: de CERT-gemeenschap van SURFnet. Hierin zitten afgevaardigden van de incident response teams van de aangesloten instellingen die regelmatig bijeenkomsten hebben om kennis te delen over voorgekomen incidenten om zo van elkaar te leren. Naast SCIRT is Ewald lid van een aantal andere security clubjes waardoor hij goed op de hoogte is van wat er op het moment speelt in de wereld van security. Daarom is Ewald ook gevraagd om zijn inbreng:

- Ewald geeft ten eerste aan dat hij geïnteresseerd is in exposure naar buiten, dus services die publiekelijk open staan. Instellingen moeten er goed van op de hoogte zijn welke services in hun netwerk daaronder vallen, want iedere service kan een deur naar binnen zijn voor kwaadwillenden;
- Ten tweede zou Ewald het handig vinden om veranderingen ten opzichte van vorige raadplegingen te kunnen visualiseren, om zo de voortgang te kunnen bewaken;

- Ten slotte geeft Ewald ook aan dat het wenselijk is om foutieve informatie en false positives te filteren.

Ten slotte is Alf Moens geïnterviewd. Alf is de Chief Information Security Officer van SURF en is onder andere ook lid van de SCIRT-gemeenschap. Hij is vooral bezig met het beleid van informatiebeveiliging in plaats van de techniek.

- Alf zou meldingen van afwijkingen in het interne netwerkverkeer willen zien omdat dat een indicatie is dat er iets aan de hand is. Zo kunnen aanvallen in een vroeg stadium gedetecteerd worden. Hierop effectief monitoren is echter moeilijk voor elkaar te krijgen omdat het verkeer bij aangesloten instellingen als universiteiten heel divers is. Dit maakt het lastig om te bepalen wat normaal verkeer is.
  Een tijd na dit interview is de scope van dit project veranderd naar openbare bronnen, en daarom kan met deze wens geen rekening gehouden worden. Afwijkingen in het interne netwerkverkeer kunnen namelijk alleen binnen het interne netwerk van een organisatie gedetecteerd worden, en dus niet met open-source intelligence.
- Ten slotte geeft Alf aan dat bij het gebruik van bronnen voor de tool gelet moet worden op de juistheid, tijdigheid en relevantie van de hieruit vergaarde informatie.

Bijlage 3: X-Pack voor de PoC

Elastic biedt als Elastic Stack uitbreiding ook een betaald softwarepakket genaamd X-Pack. Dit pakket bevat een aantal extra programma's die met name zorgen voor beveiliging en monitoring van Elastic Stack:

- Security: Autorisatie en authenticatie in Elastic Stack met rollen en groepen;
- Alerting: Notificaties instellen op veranderingen in de data;
- Monitoring: Monitort de gezondheid van Elastic Stack door factoren te tonen als geheugengebruik en opslagruimte voor de data;
- Reporting: Genereren en delen van rapportages;
- Graph: Relaties tussen de data visualiseren.

De data die de PoC verzamelt kan gevoelig zijn voor misbruik. Het feit dat de tool bronnen combineert en verbanden aan het licht kan brengen versterkt deze gevoeligheid. Daarom moet de tool zelf beveiligd worden. Voor dit afstudeerproject zijn de volgende beveiligingsmaatregelen van kracht:

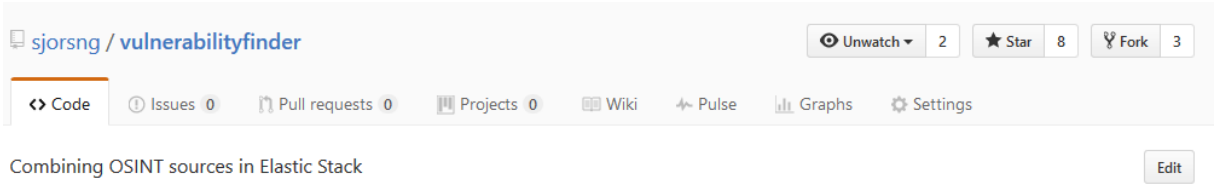
- De gehele PoC opstelling staat achter een firewall die alleen maar webverkeer voor toegang tot de tool en versleuteld verkeer voor onderhoud toestaat naar het internet;
- Het webverkeer gaat altijd via een HTTPS verbinding die uitwisseling van data versleutelt zodat alleen de gebruiker en de server de verzonden data kan uitlezen;
- Toegang tot de tool vereist een geldige gebruikersnaam in combinatie met een wachtwoord.

De tool is dus al beschermd van de buitenwereld. Wanneer deze tool uiteindelijk ook gebruikt gaat worden door aangesloten instellingen is verdere afscherming van specifieke data noodzakelijk. Instellingen hebben namelijk niets te maken met elkaars data. Hiervoor is authenticatie en autorisatie op basis van rollen en groepen nodig, en dit is een van de voornaamste redenen om gebruik te maken van X-Pack. Wanneer de tool tegen die tijd als een service aangeboden wordt moet er tevens gestreefd worden naar een zo hoog mogelijke beschikbaarheid, en daarbij komt de monitoringsmogelijkheid van pas. Tevens zijn de rapportagefunctionaliteiten van X-Pack handig voor de instellingen om automatisch rapporten te ontvangen.
Bovenstaande redenen maken de aanschaf van X-Pack aantrekkelijk en wellicht noodzakelijk om het uiteindelijk als een service aan te kunnen bieden. De uitrol valt echter buiten de scope van deze afstudeeropdracht en voor dit onderzoek is dus alleen van Elastic Stack gebruik gemaakt.

Bijlage 4: Argumentatie Python en versiebeheer op Github

Hoe de data opgehaald wordt uit de bronnen is geen deelvraag van het onderzoek geweest maar is wel noodzakelijk om te weten om PoC te kunnen maken. Tijdens het onderzoek naar de bronnen is duidelijk geworden dat de meeste bronnen een API ter beschikking stellen om data verzoeken op te doen. Ook is opgevallen dat over het algemeen Python het populairst is om bij de bronnen data op te vragen.
Python is een programmeertaal die ontwikkeld is met het oog op leesbare code. Het heeft ingebouwde functies die al veel onder de motorkap doen waardoor de programmeur hier geen aandacht aan hoeft te besteden en met weinig code veel kan bereiken. Dit bevordert de leesbaarheid van de code. Ook bestaan er voor de veelgebruikte functionaliteiten libraries die met één commando geïnstalleerd en gebruikt kunnen worden. Dit maakt Python een populaire programmeertaal waardoor er veel over op te zoeken is.
De meeste onderzochte bronnen hebben zelf een Python library die API verzoeken versimpelt. Zo'n library kan geïnstalleerd worden in Python waarna functies uit de library aangeroepen kunnen worden die op hun beurt de API aanroepen. Dit scheelt werk en tijd aangezien deze libraries voor het API aanroepwerk zorgen en dat deze geüpdatet worden zodra de software of de tool waarvoor het geschreven is veranderd. Ook zijn er op deze manier meerdere mensen die van de library gebruik maken waardoor er een soort community ontstaat waar mensen elkaar helpen bij problemen.

Uiteindelijk wordt er op de volgende manier binnen het project gebruik van Python gemaakt: Op de VM waar de PoC op staat kunnen de Python scripts (bestanden met Python code) handmatig of geautomatiseerd uitgevoerd worden. Voor iedere bron is er een script die informatie uit de bronnen haalt, structureert en wegschrijft naar bestanden op de VM. Deze bestanden kunnen vervolgens uitgelezen worden door de andere software van de PoC.
Van de Python scripts vindt versiebeheer plaats op Github. Hiermee wordt de geschiedenis van de scripts bijgehouden. In overleg met de opdrachtgever mogen de scripts publiekelijk beschikbaar zijn. De scripts zelf zijn namelijk generiek en niet gebonden aan SURFnet specifieke dingen. De gevoelige data en visualisaties van de kwetsbaarheden staan hier dus niet in, maar zijn veilig afgeschermd van de buitenwereld in de rest van de PoC. Het voordeel van publiekelijk versiebeheer is dat andere geïnteresseerden buiten SURFnet ook van de scripts gebruik kunnen maken en zelfs op eigen initiatief verbeteringen kunnen aanbrengen.



Bijlage 4, figuur 1: De Python scripts staan onder publiekelijk versiebeheer op Github.

In figuur 1 is te zien dat er al een aantal gebruikers zijn die dit project een ster hebben gegeven of 'geforkt' hebben. Dat wil achtereenvolgens zeggen dat het project als favoriet is op geslagen of dat de gebruiker een kopie van het project heeft overgenomen onder zijn eigen Github account om zelf aan te passen. Verbeteringen door anderen kunnen vervolgens na overleg doorgevoerd worden in de PoC waar SURFnet op zijn beurt ook weer op vooruit gaat.

Bijlage 5: Documentatie Proof of Concept

(De IP-adressen zijn in dit document vervangen door xxx.xxx.xxx.xxx)

Bereikbaarheid services

Kibana: *xxx.xxx.xxx.xxx*

Marija: *xxx.xxx.xxx.xxx* (experimentele Elasticsearch visualisatietool van Remco Verhoef)

Basic authentication credentials: <User>:<Pass> (vraag Rogier Spoor voor credentials)

Shell toegang tot de VM's

De tool draait op 2 Ubuntu VM's: 'ES1' en 'ES2'. Beide staan op de Openstack omgeving van utrecht.surfcloud.nl. SSH toegang tot deze machines gaat met SSH keys en kan geregeld worden met Rogier Spoor. Op ES1 draait praktisch de hele tool, op ES2 draait een extra Elasticsearch node die bijdraagt aan de stabiliteit van de Elasticsearch cluster.

ES1: *xxx.xxx.xxx.xxx*
ES2: *xxx.xxx.xxx.xxx*
Uername: ubuntu
Password: SSH key

Gebruikte software installeren & configureren

NGINX proxy

Nginx webserver wordt hier gebruikt als proxy die je verbinding beveiligd met HTTPS via Let's Encrypt certificaten die zelf steeds vernieuwd worden via een CRON-job. Daarvoor is deze tutorial gevolgd. Hier is vervolgens password authentication aan toegevoegd met deze tutorial.

'Python 3' scripts

De Python**3** scripts staan onder versiebeheer op Github: https://github.com/sjorsng/vulnerabilityfinder
Neem contact op met Sjors Haanen ('sjorsng' op Github) om collaborator te worden van dit project.

De scripts staan ook op ES1, onder /home/ubuntu/pyscripts/vulnerabilityfinder
Deze zijn uitvoerbaar in Python versie **3**.x. Hiervoor zijn volgende python**3** modules zijn nodig (te installeren met easy_install of pip voor Python**3**):

- Elasticsearch
- Censys
- Shodan
- Netaddr

Mochten er in de tussentijd nog gebruik gemaakt worden van andere Python modules, dan geeft Python de benodigde modules aan wanneer een van de scripts uitgevoerd wordt.

Meer uitleg over de scripts zelf is te vinden in de README.md en/of de wiki van het Github project.

Logstash 5.1.1

Voor het installeren van Logstash is deze tutorial gevolgd.

De configuratiebestanden van Logstash ook onder versiebeheer, in hetzelfde Github project als aangegeven bij 'Python 3' scripts.

Logstash kan uitgevoerd worden met een van de volgende commando's:

- sudo -E /usr/share/logstash/bin/logstash -f /etc/logstash/conf.d/10-logstash-shodan.conf
  Voert Logstash uit met het gespecificeerde configuratiebestand.
- sudo -E /usr/share/logstash/bin/logstash --path.settings=/etc/logstash/conf.d
  Voert Logstash uit met alle configuratiebestanden in /etc/logstash/conf.d
  Handig om hierin symlinks te plaatsen naar de Github configuratiebestanden met:
  sudo ln -s <locatie_configbestand> /etc/logstash/conf.d/<link_naam>

Handige extra opties voor bij bovenstaande commando's:

-t Checkt alleen de aangegeven configuratiebestanden op syntax errors;

--verbose Extra meldingen voor debugging.

Als Logstash een lange opstarttijd heeft kan dit waarschijnlijk verholpen worden met deze instructies.

Meldingen van Logstash zijn terug te lezen in de console of in /var/log/logstash/ logstash-plain.log.

Elasticsearch 5.1.1

Voor het installeren van Elasticsearch via de Apt repository is deze tutorial gevolgd.

Het configuratiebestand van Elasticsearch staat op ES1: /etc/elasticsearch/elasticsearch.yml

Herstarten kan met: sudo systemctl restart elasticsearch.service

Elasticsearch luistert op *xxx.xxx.xxx.xxx* voor requests, requests naar *localhost* werken niet.

Kibana 5.1.1

Voor het installeren van Kibana via de Apt repository is deze tutorial gevolgd.

Kibana is gewoon via zijn webinterface in te stellen, via Management -> Advanced Settings. Daar is het handig om 'timepicker:timeDefaults' te veranderen naar een langer tijdsbestek, zoals:

{"from": "now-90d","to": "now","mode":"quick"}

Dit is de basisfilter voor de 'discovery' pagina. Alle data zal steeds binnen 3 maanden wel opnieuw opgevraagd en geüpdatet moeten worden zodat de informatie recent genoeg is.

Tegenwoordig heeft Kibana een 'Dev tools' pagina, waar je gemakkelijk requests kan uitvoeren op Elasticsearch, zoals nieuwe indexen aanmaken. Dit werkt fijner dan bijvoorbeeld via een Shell.

Verdere algemene uitleg over de functionaliteiten in Kibana zijn hier te lezen.

Nieuwe organisatie (Elasticsearch index) toevoegen

Om data van verschillende organisaties gescheiden te houden wordt er per organisatie een aparte index gemaakt. Elasticsearch heeft de juiste instellingen nodig om de data te verwerken. Dat gaat door een template aan te maken en dat gaat het gemakkelijkste via Kibana -> Dev tools -> Console:

```
PUT _template/template_osint
{
  "template" : "*",
  "settings" : {
    "number_of_shards" : 1,
    "index.mapping.total_fields.limit": 7000
  },
  "mappings" : {
      "_default_": {
    "properties": {
      "ip": {
        "type": "ip"
      },
      "ip_int": {
        "type": "long"
      },
      "ipinfo.location": {
        "properties": {
          "geo": {
            "type": "geo_point"
          }
        }
      },
      "ipinfo.whois.person.last_modified": {
        "type": "text"
      },
      "ipinfo.whois.last_modified": {
        "type": "text"
      }
    }
   }
  }
}
```

Bovenstaand request maakt een template aan die voor iedere toekomstige nieuwe index geldt. Het IP-adres wordt gebruikt als uniek ID (aangegeven in Logstash). Een element bevat dus informatie over alle bronnen over een specifiek IP-adres. Met "index.mapping.total_fields.limit" wordt het maximaal aantal mogelijke veldjes verhoogd zodat geen elementen verloren gaan. Dit getal kan in de toekomst nog groeien, dus verhoog het getal als Logstash meldingen dit aangeven. Van “ipinfo.location.geo” wordt een geo_point gemaakt zodat het mogelijk is om er in Kibana een wereldkaart van te maken. De overige mappings in de template zorgen ervoor dat Elasticsearch specifieke veldjes opslaat als de juiste datatype.