Scriptie onderzoek Forecasting

Naam afstudeerder: Bart Dekker
Studentnummer:
Eerste examinator: Ernst Mak
Tweede examinator: Ander de Keijzer
Naam bedrijf: RoutIT
Status: Definitief
Datum: 25-05-2018
Versie: 1.0

Versiebeheer

Versie	Datum	Wijzigingen	Status
0.1	26-03-2018	Eerste opzet scriptie.	Gereed
0.2	03-04-2018	Redigeren teksten, uitbreiden theoretisch kader.	Gereed
0.3	04-04-2018	Toevoegen informatie netwerksegmenten.	Gereed
0.4	09-04-2018	Eerste concept naar opdrachtgever	Gereed
0.5	16-04-2018	Opmerkingen opdrachtgever verwerkt. Eerste concept naar HU.	Gereed
0.6	18-04-2018	Toevoegen conclusie, aanbevelingen, etc..	Gereed
0.7	20-04-2018	Toevoegen managementsamenvatting.	Gereed
0.8	15-05-2018	Invoegen bijlagen. Tweede feedback ronde opdrachtgever verwerkt. Tweede concept.	Gereed
1.0	25-05-2018	Scriptie definitief	Gereed

Distributielijst

Functie	Naam	Organisatie	Datum
Docentbegeleider	Ander de Keijzer	Hogeschool Utrecht	
Bedrijfsbegeleider	Winfred Hofman	RoutIT	
Bedrijfsbegeleider	Erica Schonewille	RoutIT	
Team Monitoring	Kristiaan Vos	RoutIT	

Goedkeuringen

Naam	Functie	Handtekening	Datum
Ander de Keijzer	Docentbegeleider		
Winfred Hofman	Bedrijfsbegeleider		
Erica Schonewille	Bedrijfsbegeleider		
Kristiaan Vos	Team Monitoring		

Auteur

Naam	Initialen	Student nummer	Telefoon	E-mail
Bart Dekker	B.J.M.			

Voorwoord

In dit voorwoord wil ik allereerst mijn Opa, Ger Vrijenhoek, bedanken, omdat hij mij essentiele gaven heeft bijgebracht voor het volgen van een HBO-opleiding. Het betreft hierbij de betekenis van het woord en discussievaardigheden. Wat betreft de begeleiding die ik nodig heb gehad tijdens mijn studie wil ik graag de decaan, Truus Bläcker, en mijn persoonlijk begeleider Wilfred Woutersen (Krul) bedanken. Beide hebben mij geholpen als ik niet wist om te gaan met studenten en docenten. Van het docententeam wil ik in het bijzonder Marcel Philip en Tim Jansma bedanken. Beide docenten hebben mij veel geleerd en deelden mijn passie voor techniek. En niet onbelangrijk, ik kon ook goed met ze lachen. Wat betreft mijn afstuderen wil ik graag RoutIT en in het bijzonder mijn begeleiders, Erica Schonewille, Winfred Hofman en Kristiaan Vos bedanken. Daarnaast wil ik graag mijn directe collega's van het NOC bedanken, ik heb er enorm van genoten om met jullie over Linux, BSD en honden te praten. Als laatste, alhoewel niet de laatste in mijn gedachten, wil ik Ander de Keijzer bedanken voor zijn kritische blik en positieve begeleiding tijdens het afstuderen.

Ede, 22 mei 2018

Managementsamenvatting

RoutIT is een *Internet Service Provider* (ISP) gericht op de zakelijke markt. Ten einde opstoppingen in het netwerk voor te zijn en een goede inschatting van de toekomstige eisen bij het aanschaffen van nieuwe apparatuur te kunnen maken, heeft RoutIT de afstudeerder Bart Dekker een onderzoeksopdracht toegewezen. Deze onderzoeksopdracht betreft het startpunt van het opzetten van proactieve in plaats van reactieve monitoring zoals de situatie nu is.

Om opstoppingen in het netwerk, ook wel netwerkcongestie genoemd, te kunnen voorspellen is data en kennis van het RoutIT netwerk nodig. Zonder enige informatie voorspellingen doen is niet mogelijk.

Het RoutIT netwerk is opgedeeld in netwerksegmenten. Ieder netwerksegment bestaat uit een cluster van apparatuur met dezelfde taak. Tussen deze netwerksegmenten kan door verschillende factoren congestie ontstaan. De volgende factoren zijn geïdentificeerd:

- Belangrijke nieuwsfeiten
- DDoS aanvallen;
- Economische groei;
- Extreem weer;
- Meerdaagse sportevenementen;
- Nieuwe klanten en migraties.
- *Software updates*;
- *Streaming* media en computerspellen;
- Terroristische aanslagen.

Door het gebruik van verschillende binnen RoutIT aanwezige relevante datasets is netwerkcongestie te voorspellen. De volgende relevante datasets zijn geïdentificeerd:

- IRMA
- RTK
- Zabbix

Door foutieve data in de eerste twee datasets zijn deze bij dit onderzoek niet gebruikt. Het advies van de afstudeerder is om het opschonen van deze datasets wel ter hand te nemen, aangezien deze data veel kan zeggen over de toekomstige netwerkcongestie. Met de data uit Zabbix zijn voorspellingen gedaan over enkele dagen in de toekomst, echter Zabbix heeft op dit moment te weinig data tot zijn beschikking om al over langere tijd voorspellingen te doen. Dit probleem zal zich automatisch en zonder ingrijpen oplossen.

Het analyseren van data is het terrein van *Business Intelligence* (BI) oplossingen. Aan de hand van het enquêteren van collega's en het Gartner *Magic Quadrant* is een lijst van 27 BI-oplossingen samengesteld. Uit deze lijst is na een gefaseerd selectieproces Splunk geadviseerd als mogelijke BI-oplossing voor het voorspellen van netwerkcongestie. Vervolgens is Splunk langdurig getest. De testen met Splunk betreffen een simulatie, waarbij voorspellingen werden vergeleken met de daadwerkelijke meetwaarden.

Inhoudsopgave

Voorwoord 3
Managementsamenvatting 4
1 Inleiding 7
2 Geheimhouding en anonimisering van de data 7
3 Organisatorische context en de positie van de afstudeerder 8
3.1 RoutIT als organisatie 8
3.2 Producten en RoutIT partners 8
3.3 Het RoutIT netwerk 8
3.4 De RoutIT organisatiestructuur 9
3.5 De Network Operations Center (NOC) afdeling 9
4 Formulering van de opdracht 10
4.1 Aanleiding voor de opdracht 10
4.2 Aanleiding en uitleg van de deelvragen 10
4.3 De doelstelling voor dit onderzoek 11
5 Onderzoeksvraag en deelvragen 12
6 Het theoretisch kader 12
6.1 Netwerkcongestie 12
6.2 *Business intelligence* en *data warehouse* 13
6.3 Machine Learning (ML) 13
6.4 Seasonal data en *forecasting* 13
6.5 Het ARIMA of Box-Jenkins filter 14
6.6 Het Kalman filter 15
6.7 Zabbix en de Zabbix functie "*forecasting*" 15
7 Globale aanpak 15
7.1 Enquêtering van collega's en selecteren mogelijke oplossing 15
7.2 Geplande aanpak aangaande te lezen literatuur ..... **Error! Bookmark not defined.**
8 Netwerksegmenten en de netwerkcongestie ertussen 16
8.1 Netwerksegmenten per team 16
8.1.1 Team Infra 17
8.1.2 Team *Voice* 18
8.2 Beïnvloeding netwerksegmenten op netwerkcongestie onderling 19
8.2.1 Team Infra 19
8.2.2 Team Voice 19
9 Netwerkcongestie factoren 20
9.1 Voorspelbare versus onvoorspelbare netwerkcongestie factoren 20
9.2 Onvoorspelbare netwerkcongestiefactoren 20
9.2.1 DDoS aanvallen 20
9.2.2 Terroristische aanvallen, rampen en extreem weer 21

9.3 Voorspelbare netwerkcongestiefactoren 21
9.3.1 Nieuwe klanten en migraties 21
9.3.2 Meerdaagse sportevenementen 21
9.3.3 *Streaming* media en computerspellen 22
9.3.4 Nieuwe versies van *Operating Systems* (OS) en applicaties 22
9.3.5 Economische groei 23
10 Datasets en de op te vragen informatie 24
10.1 Informatievragen aan de dataset 24
10.2 Beoordeling van de datasets 25
10.2.1 Centraal Bureau voor de Statistiek (CBS) 25
10.2.2 Beoordeling van de IRMA en RTK-datasets 25
10.2.3 Beoordeling van de Zabbix dataset en de implementatie binnen RoutIT 26
11 Beoordeling van Zabbix voor het doen van voorspellingen 27
11.1 Zabbix *forecasting* 27
11.1.1 Het verschil tussen historische en trend data in Zabbix 27
11.1.2 De Zabbix *forecasting* functionaliteit beschreven 27
12 Onderzoek naar een *Business Intelligence* (BI) oplossing 29
12.1 Voorselectie pakketten 29
12.2 Selectie op basis van de MoSCoW analyse 30
12.3 Integratie met de bestaande productiefaciliteiten 30
12.4 Beoordeling voorspellingsmogelijkheden 31
13 Beschrijving van Splunk en de gemaakte keuze binnen Splunk 32
13.1 Contact met Splunk vertegenwoordigers 32
13.2 Opvragen monitoringdata 32
13.3 De Splunk *Machine Learning Toolkit* (MLT) 32
13.4 Kalman vs. ARIMA 34
13.5 Splunk *forecast* model in MLT 34
13.6 Uitvoeren simulaties 35
14 Conclusie 38
15 Aanbevelingen aan RoutIT als vervolg op dit onderzoek 41
16 Evaluatie procesgang en conformering aan de kwaliteitscriteria 41
16.1 Evaluatie procesgang 41
16.2 Conformering aan de kwaliteitscriteria 42
17 Ethische afwegingen 42
Bibliografie 43

1 Inleiding

Om te kunnen blijven voortbestaan en te kunnen groeien is het voor een dienstverlenend bedrijf belangrijk om de kwaliteit van de diensten goed te borgen en indien nodig bij te sturen. Voor RoutIT als *Internet Service Provider* (ISP) is het noodzakelijk dat de kwaliteit van het netwerk van hoog niveau is. Door RoutIT wordt aangenomen dat de hoeveelheid door klanten ingenomen netwerkcapaciteit zal groeien en bij ongewijzigd beleid tot netwerkcongestie zal leiden. De vraag is nu hoe de netwerkcongestie te voorkomen. RoutIT maakt voor de monitoring van het netwerk gebruik van het pakket Zabbix. Vanwege technische beperkingen wordt de monitoringdata slechts 14 dagen gedetaileerd bewaard door Zabbix waarna de data gecomprimeerd wordt en details verloren gaan. Door deze compressie blijven de grote lijnen van de verbruikte bandbreedte beschikbaar. Verder zijn er, naast die van Zabbix andere datasets binnen RoutIT die gebruikt kunnen worden om de ingenomen netwerkcapaciteit te voorspellen. Tijdens dit onderzoek zijn deze datasets geanalyseerd en beoordeeld, waarna de resultaten beschreven zijn in deze scriptie.

Het RoutIT netwerk bestaat uit netwerksegmenten waarvan de componenten, individuele routers en switches, samen een taak uitvoeren. Het is van belang om te weten welke netwerksegmenten er binnen het RoutIT netwerk zijn en hoe deze elkaar beïnvloeden in het ontstaan van netwerkcongestie.

Voor de analyse van al deze data en het doen van voorspellingen zal deze data bijeen gebracht moeten worden in een pakket. Verder zal dit pakket mogelijkheden moeten hebben om de data te analyseren, anders zou een standaard database server ook volstaan. Voor meer informatie over het type oplossing, zie het theoretisch kader.

Aangezien het door RoutIT gebruikte monitoringpakket, Zabbix, zelf ook al beschikt over mogelijkheden tot voorspellen van toekomstig netwerkverkeer, zal ook onderzocht worden in hoeverre deze functionaliteit bruikbaar is. Deze vraag is met name interessant omdat indien Zabbix al kan bieden wat RoutIT wil, er geen kennis van een totaal nieuw pakket hoeft te worden opgegedaan.

2 Geheimhouding en anonimisering van de data

De afstudeerder heeft richting RoutIT de plicht om gevoelige informatie geheim te houden en de *privacy* van diens medewerkers te waarborgen. Daarom zullen slechts algemene omschrijvingen van medewerkers opgenomen worden in dit document. Voor RoutIT is informatie over zijn netwerk, waaronder verbruikte bandbreedte, concurrentiegevoelige informatie. Daarom zal in deze scriptie geen bandbreedte informatie opgenomen worden die herleidbaar is naar een specifiek moment in de tijd. Ook zal geen informatie verstrekt worden over de totale capaciteit van (geaggregeerde) lijnen. Voornoemde informatie is wat betreft de afstudeerder niet noodzakelijk ten behoeve van de onderbouwing van dit onderzoek en hoeft dan ook niet in de scriptie beschreven te worden.

3 Organisatorische context en de positie van de afstudeerder

In dit hoofdstuk worden RoutIT als organisatie, verkochte producten, de RoutIT partners, de afdeling en de positie van de afstudeerder op deze afdeling beschreven.

3.1 RoutIT als organisatie

RoutIT is een ISP gericht op de zakelijke markt. De kern van de dienstverlening bestaat uit het bieden van een internetverbinding en telefonie. De klanten van RoutIT stellen hoge eisen aan de dienstverlening, zowel wat betreft de snelheid en de beschikbaarheid als de vertrouwelijkheid en integriteit van de dienstverlening, daarom is RoutIT ISO27001 (ISO, 2018) gecertificeerd. Onderdeel van de veiligheidsprocedures is onder andere toegangscontrole op apparatuur en in het gebouw. Kenmerk van RoutIT is dat het een dynamisch bedrijf is. De visie en missie van RoutIT luidt als volg: “Samen met onze partners elke ondernemer succesvoller maken met vernieuwende en hoogwaardige ICT-oplossingen: dat is onze missie. Wij geloven in de kracht van de partner, in eenvoud en in de kracht van ondernemen!” (RoutIT, 2018).

3.2 Producten en RoutIT partners

Alle RoutIT producten zijn *white label* producten. Deze producten verkoopt RoutIT aan IT-bedrijven, zogenaamde partners, die daar hun eigen naam opplakken. Zo is een internetverbinding door een eindgebruiker alleen te bestellen via een RoutIT partner. Deze partner ondersteunt de eindgebruiker, maar de eindgebruiker maakt onderwater gebruik van de RoutIT infrastructuur. RoutIT doet hierbij niet aan koppelverkoop van diensten. Het is prima mogelijk om telefonie van RoutIT te betrekken en internet via een andere provider. Dit is echter niet aan te raden omdat de kwaliteit van de telefonie dan niet gegarandeerd kan worden. Indien klanten de benodigde apparatuur nog niet in hun bezit hebben dient deze via andere leveranciers dan RoutIT aangeschaft te worden. RoutIT levert zelf geen apparatuur aan klanten.

3.3 Het RoutIT netwerk

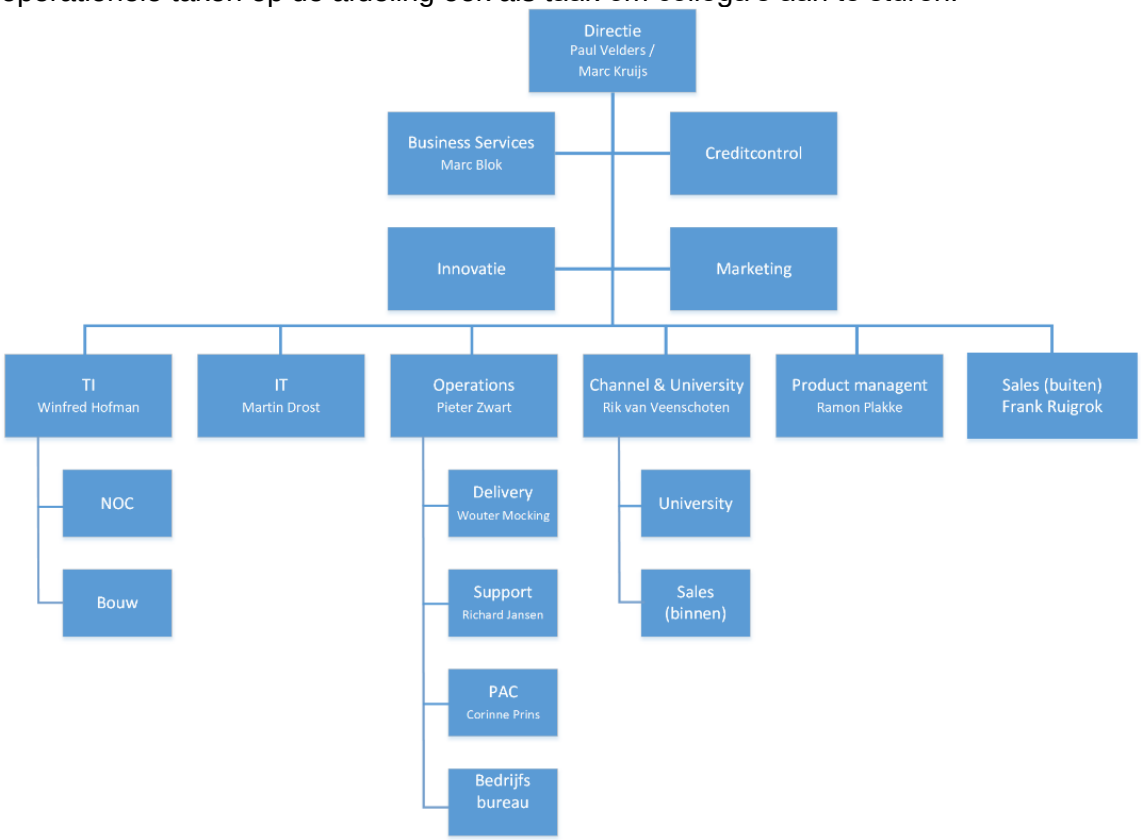
De apparatuur van RoutIT is geplaatst bij datacentra verspreid over het land. RoutIT huurt in deze datacentra al naar behoefte één of meer serverracks. Apparatuur geplaatst in datacentra maakt normaliter gebruik van de glasvezelkabels die het datacentrum verbinden met de buitenwereld. RoutIT daarentegen huurt van een derde leverancier eigen glasvezelkabels en sluit daar zelf apparatuur op aan.

De door RoutIT gebruikte netwerkapparatuur komt uitsluitend van Cisco. Bij het nieuw aanschaffen van netwerkapparatuur betreft het de Cisco ASR1000 en ASR9000 series *router*. Vanuit de historie zijn er ook nog andere typen *routers* en *switches* van Cisco in gebruik. De Cisco ASR-producten betreffen een combinatie van *switching* en *routing* voor de ISP-markt, maar worden door RoutIT *router* genoemd. Een ASR is een modulair apparaat dat uitbreidbaar is. RoutIT heeft op ASR1000 *routers* 6 poorten van 10Gbps en op de ASR9000 meer dan 50 poorten van ofwel 10Gbps of 100Gbps in gebruik. Voor de netwerkverbindingen in het netwerk wordt uitsluitend van glasvezel gebruik gemaakt. De glasvezelkabels worden met meerdere kleuren belicht om zodoende de capaciteit van de glasvezelkabel te verhogen. Een glasvezelkabel kan tussen de 50m en 40km overbruggen. De ASR9000 variant met de minste functionaliteit beschikt al over een capaciteit van maximaal 44Gbps. De ASR9000 variant met de meeste functionaliteit beschikt over een capaciteit van maximaal 160Tbps. In het geval van de duurdere ASR9000 modellen koopt een klant een *router* zonder poorten. De poorten dienen als *line cards* los aangeschaft te worden. Een *line card* is te vergelijken met een enorme *Network Interface Card* (NIC) (Cisco, 2018b). Een *line card* is namelijk geschikt om net als een NIC hardwarematig geschikt om gebruikt te worden als *routed interface* zonder tussenkomst van software.

Ten behoeve van de verbinding met andere ISP's voor het verbinden met apparatuur welke zich niet in het eigen netwerk bevindt, maakt RoutIT gebruik van het *Border Gateway Protocol* (BGP). Op deze manier kan RoutIT informatie uitwisselen met andere ISP's (Rekhter, Li, & Hares, 2006). Het uitwisselen van informatie wordt *peering* genoemd. Op die manier is het voor klanten van RoutIT ook mogelijk om bijvoorbeeld websites te bezoeken die hun internetconnectiviteit niet bij RoutIT betrekken. RoutIT maakt via BGP-verbinding met ongeveer 100 *peers*, dit gebeurt op de AMS-IX en NL-IX *Internet eXchanges* (IX). Een alternatief voor *peering* is *transit*. In het geval van *transit* wordt bij internetverkeer voor een bestemming buiten het RoutIT netwerk of de ISP's waar RoutIT mee *peert* het desbetreffende verkeer naar de *transit provider* gestuurd die vervolgens zorgdraagt voor een goede afhandeling. Uit redundantie oogpunt betrekt RoutIT bij drie *transit providers* een *transit* dienst.

3.4 De RoutIT organisatiestructuur

In afbeelding 1 wordt de RoutIT organisatiestructuur weergegeven. RoutIT heeft daarmee een vrij platte organisatie. De afdelingsmanagers vallen direct onder de directie. De afdeling "*Operations*" is de enige afdeling met meerdere managers. Deze managers hebben naast de operationele taken op de afdeling ook als taak om collega's aan te sturen.



Afbeelding 1: Organogram met de RoutIT organisatiestructuur

3.5 De Network Operations Center (NOC) afdeling

Het NOC, waar de afstudeerder werkt, heeft als doel om op technisch vlak nieuwe diensten te ontwikkelen, maar ook om bestaande diensten en het netwerk te onderhouden, te beheren en incidenten op te lossen. Door RoutIT ontwikkelde diensten worden op commercieel niveau continue geëvalueerd en indien nodig worden de diensten verbeterd of verwijderd uit het dienstenportofolio. Ideeën voor nieuwe diensten worden aangedragen door een andere afdeling, die klantcontact heeft. Deze afdeling doet ook de commerciële evaluatie. Het NOC komt daarnaast ook in actie als de derdelijns helpdesk van RoutIT een probleem van een klant niet kan oplossen.

Binnen het NOC is het de bedoeling dat mensen van hetzelfde team elkaar kunnen vervangen. In de praktijk heeft iedereen zijn specialisme. Door de aard van de opdracht is de afstudeerder gekoppeld aan Kristiaan. Hij is verantwoordelijk voor het monitoringgedeelte. De door de afstudeerder uitgevoerde opdracht moet het werk van Kristiaan vergemakkelijken.

4 Formulering van de opdracht

In dit hoofdstuk wordt de aanleiding voor RoutIT om dit onderzoek te starten beschreven, en de redenen voor de formulering van de verschillende (deel)vragen.

4.1 Aanleiding voor de opdracht

Voor RoutIT is het een gegeven dat de huidige beschikbare netwerkcapaciteit in de toekomst ontoereikend is. Op dit moment heeft RoutIT een reactieve manier van monitoren. Het doel van dit onderzoek is om de eerste stap te zetten naar proactieve monitoring waarbij voorspeld kan worden wanneer er problemen gaan ontstaan, opdat klanten geen last ondervinden van netwerkcongestie. Voor het zetten van deze eerste stap naar pro-actieve monitoring heeft RoutIT een opdracht uitgeschreven met als doel een methode te vinden, welke op basis van de binnen RoutIT beschikbare datasets, voorspellingen kan doen over wanneer en waar de netwerkcapaciteit ontoereikend is.

4.2 Aanleiding en uitleg van de deelvragen

Op dit moment kunnen knelpunten en groeiverwachtingen al handmatig in kaart gebracht worden, maar het formaliseren en automatiseren van dit proces heeft voor RoutIT grote voordelen. Op die manier kan sneller geacteerd worden op organisatorische veranderingen, de klantenaantallen en de opkomst van concurrenten en concepten in de ICT-branche.

RoutIT heeft de beschikking over diverse datasets. Eén van die datasets is de monitoringdata uit Zabbix. Een andere dataset betreft uitstaande orders en opleverdata van door partners aangevraagde diensten. Maar er is nog meer. Tijdens deze opdracht moet antwoord gegeven op de volgende en aanverwante vragen:

- Welke datasets kunnen behulpzaam zijn bij het voorspellen van de benodigde capaciteit?
- Welke bewerkingen moeten worden uitgevoerd op de datasets?
- Hoe kunnen voornoemde data gecombineerd en geanalyseerd worden?
- In welke netwerksegmenten treedt wanneer de congestie op?

Het RoutIT netwerk bestaat uit verschillende segmenten met elk hun eigen taken. Deze segmenten interacteren met elkaar en hebben zodoende invloed op netwerkcongestie bij elkaar. Bij het selecteren van de datasets is het belangrijk om te onderzoeken welke informatie noodzakelijk is voor de desbetreffende collega van het NOC die verantwoordelijk is voor het bijbehorende netwerksegment. De netwerksegmenten zijn op dit moment deels geïdentificeerd, maar hier ligt voor de afstudeerder ook nog een taak. Daarmee is het verzamelen van gegevens over het netwerk en de netwerkcongestiefactoren waarschijnlijk het belangrijkste deel van deze opdracht. Misschien komt wel uit dit onderzoek dat ten behoeve van de proactieve monitoring meer informatie moet worden vastgelegd of dat er transformaties moeten plaatsvinden. Oftewel welke voorwaarden moeten er gesteld worden aan de huidige omgeving om het door de afstudeerder uitgevoerde onderzoek te kunnen volbrengen?

Op dit moment heeft RoutIT een *data warehouse* oplossing van Microsoft. Op het vlak van *data warehousing* en *business intelligence* is Microsoft leidend wat betreft innovativiteit en visie (Sallam et al., 2017). Het in gebruik zijnde pakket bij RoutIT zal meegenomen worden in de zoektocht naar een voorspellingsmethodiek, echter het staat de afstudeerder vrij om voor het NOC een ander pakket te selecteren. In dat geval zal het NOC een ander pakket gebruiken dan de afdeling *Business Intelligence* (BI) van RoutIT. Deze uitkomst is verdedigbaar op grond van de verschillende doelen van het NOC en de afdeling BI.

Van Zabbix is bekend dat het aangaande het netwerkverkeer beschikt over een voorspellende functie. Het is mogelijk om delen van de analyse uit te besteden aan Zabbix. Of dit ook raadzaam is dient tijdens dit onderzoek door de afstudeerder onderzocht te worden. Het onderzoek wordt op voorhand niet door de opdrachtgever beperkt aangaande kosten of leveranciers. Wel is het wenselijk om de oplossing dusdanig te selecteren dat de oplossing geïntegreerd kan worden of samenwerken met de bestaande systemen binnen RoutIT.

Wegens de complexiteit van deze opdracht is het niet de bedoeling dat de afstudeerder een implementatie doet. Een *Proof of Concept* (PoC) met als onderdeel een simulatie is daarentegen wel onderdeel van deze opdracht. RoutIT heeft voor de simulatie en de PoC onder andere hardware en een virtueel platform ter beschikking. Wegens de overdrachtelijkheid en voortzetting van het onderzoek en de PoC zal relevante kennis en ervaring worden opgenomen in de scriptie en in bijlage E en bijlage F bijgevoegd PoC ontwerp en testplan.

Het betreft hier dus een ontwerp-/adviesopdracht.

4.3 De doelstelling voor dit onderzoek

Het doel van dit onderzoek is het doen van voorspellingen aangaande toekomstige netwerkcongestie. Hiervoor is het van belang om te kijken welke factoren van invloed zijn op netwerkcongestie in (delen) van het netwerk van RoutIT en hoe de samenhang is.

5 Onderzoeksvraag en deelvragen

De voor dit onderzoek gedefinieerde onderzoeksvraag luidt als volgt:

In hoeverre en op welke manier is het mogelijk om met over de jaren afnemende betrouwbaarheid te voorspellen wanneer netwerkcongestie optreedt in het netwerk van RoutIT teneinde proactief verbeteringen aan te kunnen brengen?

Ter beantwoording van de hierboven genoemde onderzoeksvraag zijn de volgende deelvragen gedefinieerd:

1. Welke netwerksegmenten zijn betrokken bij netwerkcongestie binnen de RoutIT core-dienstverlening en hoe beïnvloeden die elkaar?
2. Welke voor dit onderzoek relevante datasets heeft RoutIT tot zijn beschikking en in hoeverre zijn deze bruikbaar voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie in de core-dienstverlening te voorkomen?
3. Welke mogelijkheden aangaande *business intelligence* zijn beschikbaar en kunnen geïntegreerd worden met de bestaande RoutIT productiefaciliteiten aangaande het doen van voorspellingen betreffende het netwerkverkeer ten einde netwerkcongestie in het netwerk te voorkomen?
4. In hoeverre is de Zabbix functie "*forecasting*" bruikbaar voor toepassing door RoutIT voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie te voorkomen?
5. In hoeverre zijn de onder deelvraag 3 geselecteerde producten geschikt om meerjarige voorspellingen, met afnemende betrouwbaarheid, aangaande de verwachte groei van het netwerkverkeer te doen ten einde netwerkcongestie te voorkomen?

6 Het theoretisch kader

In dit hoofdstuk worden termen en technieken besproken die essentieel zijn voor het begrijpen van de onderzoeksvraag en de deelvragen alsmede het begrijpen van dit onderzoek.

6.1 Netwerkcongestie

Zodra RoutIT eindklanten meer bandbreedte in het RoutIT netwerk genereren dan waarvoor het netwerk ontworpen is, kan er netwerkcongestie ontstaan. Netwerkcongestie betreft een ophoping van netwerkpakketten bij de overgang naar een ander netwerk(segment) waarbij bandbreedte verschillen optreden (Nagle, 1984). Dit kan als gevolg hebben dat bijvoorbeeld het laden van webpagina's niet lukt of dat de kwaliteit van media *streams* laag is. Aangezien netwerkcongestie een negatieve invloed heeft op de RoutIT dienstverlening, is het van belang om netwerkcongestie te kunnen voorspellen en daarmee voor te zijn.

6.2 *Business intelligence* en *data warehouse*

Binnen RoutIT is al veel data beschikbaar. Uit het vooronderzoek, waarover later meer, is gebleken dat er oplossingen bestaan om die data te bijeen te brengen en te analyseren, dit wordt *Business Intelligence* genoemd. *Business Intelligence* (BI) is het analyseren en representeren van data op een dusdanige manier dat de data ingezet kan worden om inzichten te verwerven of voorspellingen aangaande de toekomst mee te doen (Stodder, 2015). Vroeger was BI sterk ICT-gerelateerd en was de kennis van een programmeer- of scriptingtaal nodig om met BI te kunnen werken. Op dit moment is de ontwikkeling gaande om BI ook voor minder technisch aangelegde gebruikers toegankelijk te maken doormiddel van gebruiksvriendelijke visuele software. *Business intelligence* betreft alleen het analyse stuk. Het opslaan van de data gebeurt in een *data warehouse*. Een *data warehouse* is een vorm van opslag om data vanuit verschillende bronnen samen te brengen. Een belangrijke functionaliteit van een *data warehouse* is om de aangeleverde data op te schonen, te catalogiseren en om te zetten naar een uniform formaat (Patil, Rao, & Patil, 2011). Het *data warehouse* biedt vervolgens de data in een uniform formaat aan ten behoeve van toepassingen als, maar niet uitsluitend, *business intelligence*.

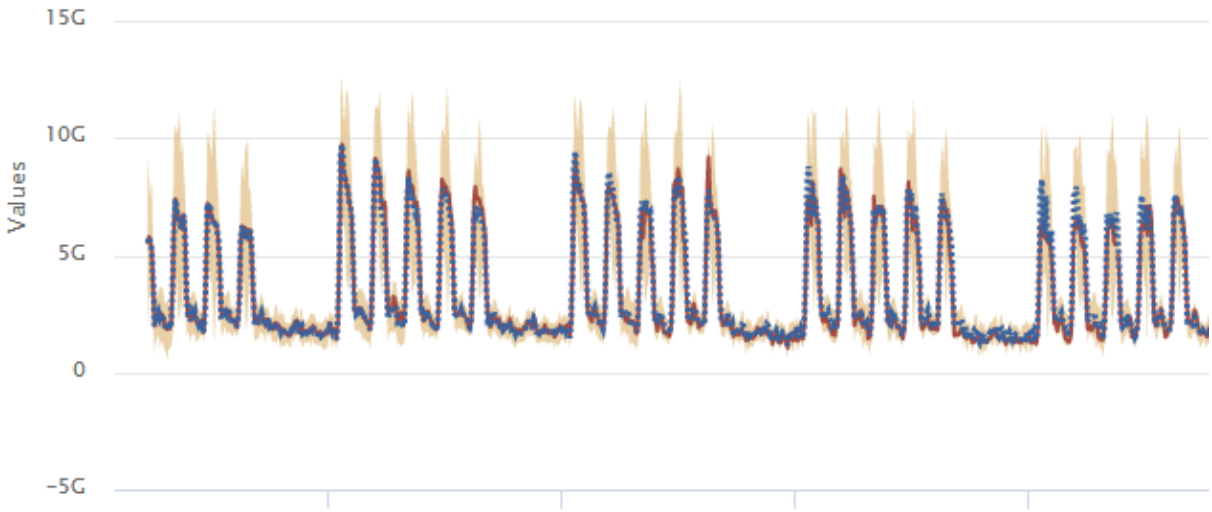
6.3 *Machine Learning* (ML)

Computers zijn ontworpen om instructies op te volgen. Echter het domweg opvolgen van instructies is weinig flexibel en het kost relatief veel tijd om een computer complexere zaken te laten doen. *Machine Learning* (ML) is een computertechniek waarbij een computer kan leren van voorbeelden. Indien een computer wil leren vertalen, dan geef je de computer bijvoorbeeld documenten waarbij de inhoud hetzelfde, maar de taal anders is. Dit kunnen bijvoorbeeld overheidsdocumenten zijn van de Europese Unie (EU) (Smola & Vishwanathan, 2008). Daarnaast kan *Machine Learning* leren van door zichzelf gemaakte fouten en de leermomenten vervolgens weer meenemen bij het aanpakken van opvolgende problemen (Michalski, Carbonell, & Mitchell, 1984).

De hierboven gegeven definities van ML zijn slechts gedeeltelijk van toepassing op dit onderzoek, aangezien de voorgaande definities gericht zijn op wetenschappelijk onderzoek naar onder andere taal. Tijdens de zoektocht naar een BI-oplossing, welke bruikbaar is voor dit onderzoek, is de term *Machine Learning* veelvuldig voorgekomen. In ons geval houdt ML in dat we een ML-programma een hoeveelheid data over een bepaalde netwerklink geven, waarna deze data door een algoritme gehaald wordt. Op basis van het vinden van een statistisch verband, waarover hieronder meer, kan vervolgens een patroon gevonden worden. Het ML algoritme kan dit patroon met zijn specifieke kenmerken vervolgens doortrekken naar toekomstige momenten (McCrea, 2018).

6.4 *Seasonal data* en *forecasting*

Bij het analyseren van bandbreedte statistieken, zijn er vanuit de statistiek meerdere typeringen van dergelijke statistieken. Deze paragraaf gaat over de op de netwerkstatistieken van RoutIT van toepassing zijnde beschrijving, namelijk *seasonal data*, en geeft uitleg over *forecasting* op basis van die data. *Seasonal data* is data die over tijd hetzelfde patroon laat zien (The Pennsylvania State University, 2018). Als ISP gericht op de zakelijke markt laat het op grond van de hoeveelheid verbruikte bandbreedte altijd hetzelfde patroon zien. Dit patroon bestaat eruit dat de vijf werkdagen aan hun hoge pieken te herkennen zijn en dat het weekend te herkennen is door een vrijwel vlakke lijn, zie afbeelding 2.



Afbeelding 2: *Seasonality* van het RoutIT netwerkverkeer.

Het hierboven beschreven patroon zorgt ervoor er bij het kiezen van een *forecasting* algoritme gekeken moet worden of het desbetreffende algoritme om kan gaan met *seasonal data*. *Forecasting* is een techniek waarbij op basis van een patroon in de beschikbare data een voorspelling gedaan kan worden (Tableau, 2018). Daarmee is *forecasting* een toegepaste vorm van ML zoals besproken in de voorgaande paragraaf.

6.5 Het ARIMA of Box-Jenkins filter

Een methode die gebruikt kan worden voor het doen van *forecasts* is de ARIMA-methode. De ARIMA-methode is een methode voor het voorspellen oftewel *forecasten* van tijdgebonden reeksen. ARIMA wordt vaak de Box-Jenkins methode genoemd omdat George Box en Gwilym Jenkins de praktische toepassing van ARIMA in 1976 beschreven hebben (NCSS Statistical Software, 2018). Modernere methoden betreffen de Sigma en *Particle* filters. Zowel het Sigma als het *Particle* filter leren door het doen van heel veel voorspellingen over een enkele toekomstige waarde en deze voorspellingen vervolgens te vergelijken met de echte waarde. Vervolgens leren beide filters door een wegingsproces steeds beter te voorspellen (McClure, 2018). Deze manier van werken vergt echter veel rekenkracht. Daardoor schalen de Sigma en *Particle* filters slechter dan ARIMA of Kalman, zie volgende paragraaf. Alhoewel ARIMA en de implementatie van George Box en Gwilym Jenkins licht verschillen worden de termen ARIMA en Box-Jenkins door elkaar gebruikt om naar hetzelfde te verwijzen.

De ARIMA methode is alleen met veel handwerk en controle in staat tot het omgaan met *seasonal data* (NCSS Statistical Software, 2018). Dit komt doordat ARIMA bij missende waarden juist voor de bij dit onderzoek behorende *seasonal data* slecht in staat is om een goed alternatief aan te dragen. In het geval van RoutIT kan er data missen door een herstart of het kapot gaan van een apparaat. De missende waarde wordt door ARIMA ingevuld door de eerste voorgaande en opvolgende waarden te middelen, wat juist bij seasonal data onbetrouwbaar is doordat de data een vreemde piek of dal gaat vertonen. Om die reden is bij het gebruik van ARIMA grondige kennis van het algoritme vereist ten einde de data en uitkomsten handmatig te kunnen controleren en repareren.

6.6 Het Kalman filter

Een andere methode die gebruikt kan worden voor het doen van *forecasts* is het Kalman filter. Het Kalman filter is genoemd naar zijn bedenker Rudolf Kálmán en is ruim 50 jaar oud (Faragher, 2018). Het Kalman filter daarentegen vraagt juist weinig rekenkracht, waardoor het met de toenmalige techniek gebruikt kon worden om de maanlanding en terugkeer van de astronauten mogelijk te maken (Faragher, 2018).

Tijdens het Proof of Concept (PoC) is gebleken dat het Kalman filter over verschillende methoden beschikt. Deze methoden zijn onder te verdelen in methoden voor seasonal en lineaire data. Zonder deze methoden is het Kalman filter niet in staat om *seasonal data* te verwerken. Het Kalman filter wordt verder gebruikt in bijvoorbeeld *touchepads* van laptops om zo de muisaanwijzer een soepeler baan te laten beschrijven. Tijdens de PoC zullen deze methoden onderzocht en beoordeeld moeten worden samen met de andere *forecast* methoden in de desbetreffende BI-oplossingen.

6.7 Zabbix en de Zabbix functie *"forecasting"*

Zabbix is een Open Source Software (OSS) monitoringpakket en tevens niet gebonden aan één type *Operating System* (OS) of netwerkapparaat (Zabbix, 2018a). Voor het monitoren van servers en netwerkapparatuur maakt Zabbix gebruik van onder andere het *Simple Network Management Protocol* (SNMP) of de Zabbix *Agent* (Zabbix, 2018b). Binnen RoutIT wordt Zabbix gebruikt om het netwerk en andere productiefaciliteiten te monitoren. Zabbix bevat onder andere gegevens over het actuele en historische bandbreedte verbruik door eindgebruikers. Hieronder vallen, wellicht ten overvloede, in het geval van RoutIT ook telefoniegesprekken. Voor het doen van voorspellingen bevat Zabbix de functie *forecasting*, deze functie maakt gebruik van statistische formules (Zabbix, 2015). *Forecasting* binnen Zabbix heeft als nadeel dat voor deze functionaliteit geen gebruik gemaakt kan worden van Zabbix *Trends* data, maar dat de volledige historische data benodigd is.

7 Globale onderzoeksaanpak

In dit hoofdstuk zal uitgelegd worden welke onderzoeksmethoden gebruikt worden tijdens dit onderzoek. De theorie die gebruikt zal worden is MoSCoW (McCabe, 2007). De voor RoutIT toegepaste MoSCoW-analyse wordt in bijlage C weergegeven. In de onderstaande paragrafen wordt het informatievergaringsproces beschreven.

7.1 Enquêtering van collega's en selecteren mogelijke oplossing

Voor het verzamelen van informatie over het RoutIT netwerk en de optredende netwerkcongestie zal de afstudeerder enquêtes houden bij de collega's van het NOC. Met deze enquêtes kan de afstudeerder zich oriënteren binnen RoutIT en de wensen van de verschillende teams op het NOC bepalen. Het betreft daarbij korte vragenlijsten om, op advies van de opdrachtgever, te voorkomen dat de afstudeerder te veel onderwerpen bij dit onderzoek gaat betrekken indien collega)s ongelimiteerd hun wensen kunnen geven. De leden van het NOC zijn opgedeeld in verschillende teams. Deze teams zijn ieder voor zich verantwoordelijk voor een deel van de dienstverlening. De teams zijn als volgt:

- Team Data
- Team Infra
- Team Monitoring
- Team Voice

Per team heeft de afstudeerder een teamspecifieke vragenlijst opgesteld en deze in een gesprek aan de teams voorgelegd. Door de aard van de opdracht is de afstudeerder defacto ingedeeld bij team Monitoring. Het contact met team Monitoring is daardoor op het 'spar-niveau', daarom is team Monitoring niet officieel geënquêteerd.

Voortvloeiend uit de gehouden enquêtes zal gekeken worden welke voorwaarden aan de oplossing van het netwerkcongestie probleem verbonden zijn en hoe dit te vertalen is naar een concrete oplossing. De manier van selecteren hangt af van de op afstand beschikbare informatie over de verschillende oplossingen. Voor het prioriteren van de eisen zal een MoSCoW analyse gebruikt worden. De geselecteerde oplossing is te allen tijde een voordracht die nog door RoutIT geaccordeerd moet worden.

7.2 Aanpak lezen en beoordelen literatuur op bruikbaarheid

Tijdens dit onderzoek zal zoveel mogelijk gebruik gemaakt worden van gezaghebbende literatuur. De literatuur die voornamelijk gebruikt zal worden zijn artikelen die gevonden zijn door middel van Google Scholar, databanken welke beschikbaar gesteld zijn door de Hogeschool Utrecht, en documentatie van leveranciers. Bij het gebruik van een generieke zoekmachine wordt de bron beoordeeld op zijn gezaghebbendheid voor opname in de scriptie. Ten einde het onderzoek niet door een commerciële partij te laten beïnvloeden en het onderzoek reproduceerbaar te maken, wordt gebruik gemaakt van DuckDuckGo om zodoende niet in de *filter bubble* terecht te komen (Pariser, 2011).

8 Netwerksegmenten en de netwerkcongestie ertussen

Voor het doen van voorspellingen is het noodzakelijk om te achterhalen hoe het RoutIT netwerk in elkaar steekt en waar de netwerkcongestie optreedt. Voorspellingen kunnen zonder "kristallen bol" alleen gedaan worden met behulp van informatie. Deze informatie moet gaan over die plaatsen in het netwerk waar de netwerkcongestie verwacht wordt. De benodigde informatie is beschikbaar, maar het is de kunst om de informatie te filteren op relevantie. Aangezien het RoutIT netwerk uit een veelheid aan netwerksegmenten bestaat, dienen deze netwerksegmenten eerst in kaart gebracht te worden. Een netwerksegment is een verzameling apparaten in het RoutIT netwerk met een en dezelfde taak. Het kan hierbij gaan om een groep *routers* of *firewalls* of om bijvoorbeeld een groep servers zoals telefooncentrales. Inventarisatie van netwerksegmenten die last hebben van netwerkcongestie is nodig om te weten welke vragen aan de datasets, waarover later meer, gesteld moeten worden. Informatie uit deze datasets kan vervolgens weer gebruikt worden voor het doen van voorspellingen aangaande het optreden van netwerkcongestie.

8.1 Netwerksegmenten per team

Binnen het NOC zijn verschillende teams die elk hun eigen specialiteit hebben. Ieder team beheert een eigen subset aan netwerksegmenten. In deze paragraaf worden de netwerksegmenten in alfabetische volgorde per team weergegeven. Deze lijst is verkregen met behulp van de voornoemde gehouden enquêtes en de in de NOC Wiki gevonden netwerksegmenten.

Gekozen is om alleen netwerksegmenten die een verbindende taak hebben op te nemen. Netwerksegmenten die een dienst leveren bovenop verbindende netwerkblokken zoals Pinnen over IP (PoIP) zijn daarom niet meegenomen. Het *Out Of Band* (OOB) netwerk, een netwerk om in noodgevallen apparatuur te kunnen bereiken als het reguliere netwerk buiten werking is, heeft wel een verbindende werking, maar hier mag geen productieverkeer overheen. Daarom is het OOB-netwerk niet opgenomen in dit overzicht.

De delen van het netwerk die slechts een eindpunt voor het verkeer behelzen zijn wegens voornoemde beperking niet meegenomen. Daarom zijn de door team Data beheerde netwerksegmenten niet meegenomen in dit overzicht. Afgaande op de team Data enquête treedt hier tevens ook geen netwerkcongestie op. De monitoring is alleen voor team *Voice* in het overzicht meegenomen, omdat het monitoring netwerksegment bij team *Voice* ook een verbindende werking kan hebben. Dit in tegenstelling tot de monitoring bij team Infra.

8.1.1 Team Infra

Team Infra is verantwoordelijk voor alle *routers*, *switches*, en andere netwerkapparatuur van RoutIT. Het netwerk van RoutIT is zoals gebruikelijk bij grotere netwerken opgedeeld in meerdere lagen. Dit gelaagde netwerkarchitectuurmodel is door meerdere netwerkapparatuurleveranciers ieder voor zich ontwikkeld. Om de leesbaarheid van deze lijst van netwerksegmenten te verhogen is bij team Infra deze gelaagde aanpak weergegeven. Per laag zijn de netwerksegmenten weergegeven met bijbehorende uitleg.

- *Access* laag
    - Op de *Access* laag komen de verbindingen van eindklanten binnen.
    - ASW-*switches* koppelen de ATM-*routers* met de core-laag. Deze tussenstap is nodig vanwege de beperkte technische mogelijkheden van de ATM-*routers*.
    - BNG-*routers* termineren de klantverbindingen. Het betreft hier *Open Systems Interconnect* (OSI) laag 2 verbindingen zoals xDSL, FTTH, etc. De verschillende BNG's koppelen vervolgens via de BBNA-*routers* met de rest van het RoutIT netwerk.
    - DAS/COLO-*switches* koppelen klanten in datacentra met het internet en met het RoutIT netwerk.
    - LIS/LAS-*switches* verzorgen de koppeling tussen de kleinere leveranciers van verbindingen en RoutIT op laag 2. Deze speciale koppeling is nodig vanwege de gebruikte MPLS-techniek waardoor een directe koppeling met de BNG niet mogelijk is.
    - SAS-*switches* verbinden de netwerkapparatuur ten behoeve van de *firewall* dienst met elkaar en het RoutIT netwerk. Deze dienst schermt netwerken van eindklanten af van het internet, zodat een eindklant niet zelf een *firewall* hoeft te implementeren en beheren.
- Aggregatie / Core laag
    - De Aggregatie laag verbindt de afzonderlijke netwerksegmenten en netwerklagen met elkaar.
    - BBNA-*routers* aggregeren de binnenkomende klantverbindingen vanuit de BNG op de *Access* laag en leiden deze naar de core *routers* toe of buiten de core *routers* om naar specifieke bestemmingen. Dit gebeurt op laag 3 van het OSI-model.
    - Core-*routers* vormen het hart van het RoutIT netwerk. Deze *routers* verbinden de verschillende diensten en netwerklagen met elkaar.
- DDoS protectie
    - DDoS protectie is noodzakelijk om eindklanten geen hinder te laten ondervinden van een DDoS aanval.
    - Bij een DDoS aanval is er keuze uit verschillende opties om deze DDoS aanval te weerstaan. De eerste optie is om het verkeer op een goede manier door het netwerk te geleiden. De tweede optie is om het legitieme netwerkverkeer te scheiden van het aanvalsverkeer met speciale hardware. De laatste optie is hetzelfde als de vorige optie, maar dan wordt het verkeer door een externe partij gescheiden.

- Internet laag
    - De internet laag verzorgt de communicatie met netwerken die niet onder het beheer van RoutIT vallen. De IAR-*routers* zijn de koppeling tussen RoutIT, *transit* providers en de overige *peers* en aggregeren deze verbindingen. Deze *routers* hebben fysieke koppelingen met meerdere Internet *eXchanges* (IX).

8.1.2 Team *Voice*

Team *Voice* is verantwoordelijk voor het telefonie deel van de RoutIT dienstverlening. Telefoongesprekken gevoerd tussen RoutIT eindklanten worden in het RoutIT netwerk afgehandeld. Alle gesprekken zowel in- als uitgaand die gevoerd worden door RoutIT eindklanten met telefoonnummers die geen gebruik maken van RoutIT telefonie worden afgehandeld door KPN. Hierbij is wel de voorwaarde dat de RoutIT eindklant zelf beschikt over een Nederlands nummer. Als een eindklant gebruik maakt van een niet-Nederlands nummer worden inkomende gesprekken afgehandeld via een andere partij.

- *Access* SBC
    - Is het centrale aanspreekpunt voor telefoons van eindklanten die zich in het RoutIT netwerk bevinden. De *Access* SBC kijkt onder andere of een telefoon die zich in het RoutIT netwerk bevindt gebruik mag maken van de telefoniedienst.
- Monitoring
    - OCOM is een platform t.b.v. het *troubleshooten*. Met OCOM kunnen SIP-connecties en telefonie specifieke netwerkkarakteristieken geanalyseerd worden.
    - TAP monitort de *security* van het *Voice*-platform en de prestaties van *port mirrors* (SPAN) die nodig zijn voor verschillende functionaliteiten in het *Voice*-platform.
    - Zabbix *Proxies* monitoren de host machines van het *Voice*-platform op zaken als *uptime*, CPU-*load*, etc. gelijk aan alle andere door RoutIT beheerde machines.
- MSC
    - MSC geeft een belsignaal af aan de telefoon voor gespreksinitiatie en handelt vervolgens het audiosignaal af. Daarmee is de MSC de daadwerkelijke telefooncentrale.
- PC
    - Aanspreekpunt voor telefoonapplicaties die zich buiten het RoutIT netwerk bevinden, maar waarbij de gebruiker wel telefonie afneemt bij RoutIT. Het gaat hier bijvoorbeeld om mobiele telefoons met een RoutIT simkaart. Tevens faciliteert de PC de integratie met chat- en CRM-applicaties.
- PE
    - De PC oftewel *peering* SBC handelt gesprekken af met telefoons die zich niet in het RoutIT netwerk bevinden. Hieronder vallen alle telefoons die gebruik maken van een andere telefonieprovider. Via de PE geeft RoutIT deze door aan KPN die de verdere verbinding met de desbetreffende telefonieprovider regelt.
- *Voice Firewall*
    - Beschermt het *Voice*-platform netwerk tegen malafide verkeer vanaf het internet. Op de *Voice-firewall* is de Cisco IOS-functionaliteit TCP-*intercept* geconfigureerd. TCP-*intercept* op de *Voice-firewall* voorkomt

een (D)DoS SYN-aanval wat inhoudt dat een TCP-connectie wordt geïnitieerd maar niet wordt voltooid.

8.2 Beïnvloeding netwerksegmenten op netwerkcongestie onderling

In deze paragraaf wordt per team een beschrijving van de netwerkcongestie en de beïnvloeding van de netwerksegmenten onderling beschreven.

8.2.1 Team Infra

Uit de bij team Infra gehouden enquête is bovenstaande lijst met netwerklagen en segmenten voortgekomen. Daarnaast heeft de afstudeerder frequent de netwerkstatistieken bekeken. Hieruit is gebleken dat het pad van het publieke internet, vertegenwoordigd door de Internet laag, via de Aggregatie laag naar de *Access* laag het meest vatbaar is voor netwerkcongestie. Binnen de Internet laag zijn vooral de netwerksegmenten die verbinden met de *Access* laag, dus niet met servers in een datacentrum, hoofdzakelijk verantwoordelijk voor netwerkcongestie.

Wat betreft de *peering* op de Internet laag maakt RoutIT voor het overgrote deel van het internetverkeer gebruik van meerdere Internet *eXchanges* (IX). Deze IX'en koppelen het RoutIT netwerk met de rest van het internet. Als bij een meerdaags evenement RoutIT eindklanten massaal via internet TV gaan kijken, dan ontstaat er netwerkcongestie in het voornoemde netwerksegmentenpad. Tijdens de Olympische Winterspelen was er in de monitoring te zien dat er veel TV gekeken werd vanaf de werkplek door eindklanten. Meerdere verbindingen met IX'en en verbindingen op de *Access* laag richting eindklanten liepen vol. Door redundantie in het netwerk heeft dit echter niet voor problemen gezorgd omdat verkeer werd omgeleid via minder drukke netwerkpaden. Wel tonen dergelijke statistieken de noodzaak voor dit onderzoek aan.

8.2.2 Team Voice

Uit de enquête met team *Voice* is naar voren gekomen dat de telefonie infrastructuur geen last ondervindt van netwerkcongestie zoals beschreven in het theoretisch kader. Wel bestaat de mogelijkheid dat de betrokken netwerkapparatuur de grote hoeveelheid kleine pakketten niet aankan. Veel kleine netwerkpakketten zorgen voor een hoge *frame rate*. Deze *frame rate* geeft aan hoeveel frames er in aantallen, dus niet in grootte, verstuurd worden. Op dit moment wordt het aantal *frames* per seconde niet bijgehouden in de monitoring en is in overleg met de opdrachtgever besloten om tijdens dit onderzoek netwerkcongestie als gevolg van een overvloedig aantal *frames* per seconde niet mee te nemen in dit onderzoek. Het probleem van een hoge *frame rate* is ondervangen door het plaatsen van apparatuur met een overcapaciteit wat betreft de *frame rate* afgezet tegen de benodigde *frame rate*. De benodigde *frame rate* is een beredeneerde schatting op basis van de op specifieke netwerkapparatuur binnen de betrokken netwerksegmenten beschikbare statistieken.

De koppeling van een eindklant met het RoutIT netwerk wordt verzorgd door een modem. Tussen het modem van de eindklant en de desbetreffende BNG is bandbreedte gereserveerd voor telefonie. Het is dus niet mogelijk voor een klant om telefonieverkeer door internetgebruik te verstoren. De ervaring leert dat in het core-netwerk altijd voldoende capaciteit voorhanden is om telefoniegesprekken af te handelen. Dit komt omdat een telefoniegesprek slechts 100Kbps in beslag neemt. Indien telefonie wordt weggedrukt in het core-netwerk zijn er grotere problemen op dat moment dan alleen een verstoring van telefonie.

Voor team Voice spelen in het verlengde van dit onderzoek wel andere zaken. Op dit moment kost het veel tijd om genoeg telefonielicenties voorradig te hebben. Team *Voice* zou dan ook geholpen zijn met een *forecasting tool* voor dergelijke licenties. Net als de *frame rate* vallen licenties op dit moment buiten het bereik van dit onderzoek.

9 Netwerkcongestie factoren

Onderdeel van dit onderzoek is om datasets te beoordelen op bruikbaarheid bij het voorspellen van netwerkcongestie. Ten behoeve van dat deel van het onderzoek is het van belang om te weten welke vragen met behulp van de datasets moeten worden beantwoord. Voor het stellen van die vragen is het van belang om te weten welke netwerkcongestiefactoren er zijn. Het is namelijk de ervaring van de leden van het NOC dat netwerkcongestie niet vanuit het niets komt, maar dat netwerkcongestie altijd oorzaken heeft. Om de oorzaken te achterhalen en wat de verschillende teams onder netwerkcongestie verstaan, is aan ieder team in een gesprek een enquête voorgelegd.

Naar aanleiding van de enquêtes heeft de afstudeerder verder nagedacht over mogelijke extra factoren. Deze factoren zijn geverifieerd en aangevuld door de opdrachtgever.

9.1 Voorspelbare versus onvoorspelbare netwerkcongestie factoren

Netwerkcongestiefactoren vallen uiteen in voorspelbare en onvoorspelbare factoren. Bij voorspelbare factoren kan vooraf extra bandbreedte gereserveerd worden. Bij onvoorspelbare factoren is het zaak om voldoende bandbreedte reserves te hebben. Het RoutIT beleid stelt dat op een normale dag 60% (in de toekomst 50%) van de netwerkcapaciteit in gebruik mag zijn. Wordt meer dan 60% netwerkcapaciteit verbruikt op een normale dag, dan wordt de beschikbare netwerkcapaciteit uitgebreid. Het is daarmee zaak om ervoor te zorgen dat voorspelbare factoren niet aan de reservecapaciteit tornen, omdat er anders geen reservecapaciteit meer is om de onvoorspelbare factoren op te vangen indien die samen met een voorspelbare factor optreden.

9.2 Onvoorspelbare netwerkcongestiefactoren

Voor de volledigheid van dit onderzoek zijn de onvoorspelbare netwerkcongestiefactoren ook opgenomen en beschreven. De onvoorspelbaarheid maakt het lastig om deze factoren mee te nemen in het onderzoek, omdat zowel de omvang als het moment van de netwerkcongestie zich niet laten voorspellen. Daarom zijn onvoorspelbare netwerkcongestiefactoren wel opgenomen en beschreven,maar niet meegenomen in de uiteindelijke doelstellingen van dit onderzoek.

9.2.1 DDoS aanvallen

Een *Distributed Denial of Service* (DDoS) aanval is een aanval waarbij meerdere computers samen een doelwit aanvallen. Voor een ISP zoals RoutIT of een datacentrum zijn externe DDoS aanvallen dagelijkse kost. Om eindklanten geen hinder te laten ondervinden van een DDoS aanval heeft RoutIT zoals aangegeven in het vorige hoofdstuk DDoS protectiemaatregelen genomen. Een DDoS met een externe bron is daarmee voor RoutIT vervelend, maar goed te doorstaan.

Een interne DDoS aanval heeft als bron computers van eindklanten, dergelijke DDoS aanvallen komen voort uit slecht geconfigureerde klantapparatuur waardoor bijvoorbeeld een *Simple Network Management Protocol* (SNMP) of *Network Time Protocol* (NTP) server verkeerd staat ingesteld. Het goed configureren van dergelijke apparatuur oftewel *Customer Premises Equipment* (CPE), is de verantwoordelijkheid van de eindklant en de bijbehorende RoutIT partner. Een interne DDoS aanval heeft grote gevolgen voor het RoutIT netwerk, maar is net als een externe DDoS aanval niet voorspelbaar.

9.2.2 Terroristische aanvallen, rampen en extreem weer

Terroristische aanvallen zijn door een bedrijf als RoutIT niet te voorspellen, maar hebben wel impact op het netwerkverkeer. Bij het uit de lucht schieten van vlucht MH17 (BBC, 2018) zag RoutIT het aantal telefoongesprekken abrupt dalen en het netwerkverkeer abrupt stijgen. Mensen wilden weten wat er gebeurd was, hoeveel slachtoffers er waren en of het Nederlanders betrof en gingen daarom op zoek naar audiovisueel materiaal om zich een beeld te vormen.

Dit alles geldt ook in meer of mindere mate voor rampen en extreem weer. Bij rampen zoals een door technische mankementen neergestort vliegtuig of een overstroming als gevolg van extreem weer, zullen veel mensen net als bij een aanslag opzoek gaan naar audiovisueel materiaal.

9.3 Voorspelbare netwerkcongestiefactoren

Bij de voorspelbare netwerkcongestiefactoren is het bij de factoren van tijdelijke aard, zoals de Olympische Spelen, vooraf duidelijk wanneer de desbetreffende factor plaats heeft. Andere voorspelbare factoren houden een trend in de samenleving in, zoals de verschuiving van geschreven documentatie naar audiovisuele handleidingen. Deze paragraaf gaat dieper in op de op dit moment onderkende voorspelbare netwerkcongestiefactoren.

9.3.1 Nieuwe klanten en migraties

De groei van het RoutIT klantenbestand komt voor rekening van bedrijven die overstappen van andere ISP's, zoals bijvoorbeeld Ziggo, naar een RoutIT partner. De nieuwe eindklant wordt door een partner aangemeld in IRMA. IRMA is het ordersysteem van RoutIT, in dit ordersysteem worden onder andere opleverdata en lijnsnelheden opgenomen.

Een RoutIT partner is niet gebonden aan het exclusief aanbieden van RoutIT diensten. Het staat een partner vrij om eveneens met RoutIT concurrerende diensten te verkopen. Vanuit de historie kan de partner ook nog klanten hebben die via een andere ISP zijn aangesloten op het internet. Een RoutIT partner kan om allerlei, voor RoutIT onbekende redenen, besluiten al zijn klanten via RoutIT van internet of telefonie te gaan voorzien. Dit soort nieuwe eindklanten vallen onder het kopje migraties. Beide soorten eindklanten houden voor RoutIT een groei in het klantenbestand in, waarbij het verschil is dat bij eindklantmigraties het vaak grote aantallen nieuwe eindklanten in één keer zijn en dat het bij totaal nieuwe klanten om een meer gelijkmatige stijging kan gaan.

Zowel wat betreft migraties als andere nieuwe eindklanten is op een zeker moment bekend wanneer er nieuwe verbindingen bijkomen, die tevens ook voor een verkeerstoename gaan zorgen. Deze informatie is interessant omdat het betrekken van bijvoorbeeld opleverdata bij de voorspellingen iets kan zeggen over de verkeerstoename in het netwerk.

9.3.2 Meerdaagse sportevenementen

Meerdaagse sportevenementen, zoals de Olympische Spelen en de *Tour de France* zorgen, indien de wedstrijden tijdens Nederlandse kantooruren worden uitgezonden, voor een toename in het netwerkverkeer. Het verschil tussen meerdaagse en ééndaagse sportevenementen is dat bij meerdaagse sportevenementen mensen de wedstrijden bij de koffieautomaat bespreken en zodoende meer kijkers trekken. Indien meerdaagse sportevenementen in Nederland tijdens kantooruren worden uitgezonden, ziet RoutIT een toename in het netwerkverkeer omdat eindklanten schijnbaar onder werktijd naar de wedstrijden kijken. Dit komt doordat veel mensen emotioneel betrokken zijn bij sporters en bij sportwedstrijden wegens nationale trots. Door de eis vanuit de consument voor een steeds hogere beeldkwaliteit, neemt ook de bandbreedte behoefte op een dergelijk moment toe.

9.3.3 *Streaming* media en computerspellen

Vanuit de historie, als zijnde een ISP gericht op de zakelijke markt, hoeft RoutIT niet in te spelen op het toenemende netwerkverkeer als gevolg van populaire series op bijvoorbeeld Netflix of HBO. Op dit moment ziet RoutIT een verschuiving in het hiervoor beschreven beeld. Zo is er een effect op het netwerkverkeer zichtbaar bij de uitkomst van een nieuwe *Game of Thrones* aflevering, of de uitkomst van een populaire *videogame*. Indien dit samenvalt met een update van Windows (zie hieronder), dan is dit helemaal goed te zien in het netwerkverkeer.

Een voorbeeld van deze verschuiving op de werkvloer is dat tijdens de Olympische Winterspelen op het NOC één van de monitoringschermen op basis van een Ziggo Go abonnement (Ziggo, 2018) geschikt gemaakt was om de schaatswedstrijden te volgen. Ziggo Go is een mogelijkheid van Ziggo om TV te kijken via het internet en is inbegrepen bij een Ziggo TV-abonnement. Daarnaast zijn er TV-diensten als Knippr (Knippr, 2018) die onafhankelijk van ISP's als Ziggo en KPN een TV-dienst via internet aanbieden. Dergelijke diensten zijn eveneens goed bruikbaar om op het werk TV te kijken. Het opkomen van diensten als Knippr past in een trend waarbij mensen hun TV-abonnement opzeggen en overgaan op diensten als Netflix en Knippr voor het kijken van audiovisueel vermaak, in de Verenigde Staten worden deze mensen ook wel *cord cutters* (Griffith, 2018) genoemd. Op 1 juni 2018 stopt Knippr als dienst (van Miltenburg, 2018). Het is niet aannemelijk dat het stoppen van Knippr effect zal hebben op de toename van het netwerkverkeer binnen het RoutIT netwerk aangezien Knippr stopt wegens het niet halen van het gewenste aantal klanten. Daarmee zijn door de omvang van het Knippr klantenbestand de effecten op het onststaan van netwerkcongestie binnen RoutIT verwaarloosbaar.

Uit eigen ervaring en uit die van collega's puttend kan gesteld worden dat veel documentatie tegenwoordig in de vorm van video wordt aangeboden in plaats van een PDF-bestand of webpagina. Het kijken van dergelijke video's vraagt voor een langere tijd bandbreedte dan bijvoorbeeld het downloaden van een PDF-bestand of het bekijken van een webpagina. Daarnaast kan het zijn dat meerdere collega's tegelijkertijd naar één en dezelfde documentatievideo kijken en dus allebei voor langere tijd bandbreedte nodig hebben. *Streaming* wordt dan ook een steeds belangrijker element in het netwerkverkeer voor RoutIT.

9.3.4 Nieuwe versies van *Operating Systems* (OS) en applicaties

Om veilig te blijven in de digitale wereld zijn *updates* van *Operating Systems* (OS) en applicaties noodzakelijk. Via de zoekmachine DuckDuckGo en Google Scholar zijn geen analyses te vinden van de download omvang van populaire OS'en en applicaties. Besloten is om een steekproef te doen naar de download omvang van *updates* voor het populairste OS en populaire *social media apps*. Aangezien RoutIT op dit moment alleen klanten in Nederland heeft gelden de statistieken in deze paragraaf, tenzij anders aangegeven, alleen voor Nederland. Het meest gebruikte OS is Microsoft Windows 10 (Statcounter, 2018a), en het meest gebruikte social medium is Facebook (Statcounter, 2018c). Gezien de inburgering van het woord "*appen*" als synoniem voor het sturen van een bericht via WhatsApp is een steekproef naar download omvang van *apps* ongeloofwaardig als de WhatsApp *app* daarin niet is meegenomen. Wereldwijd heeft WhatsApp namelijk 1,5 miljard actieve gebruikers (Statista, 2018).

Voor de download omvang van de Facebook en WhatsApp *app* is alleen gekeken in de Google *Play Store* en de Apple *App Store*, omdat deze twee *app stores* horen bij de twee meest gebruikte mobiele OS'en die tevens de smartphonemarkt bijna volledig in handen hebben (Statcounter, 2018b).

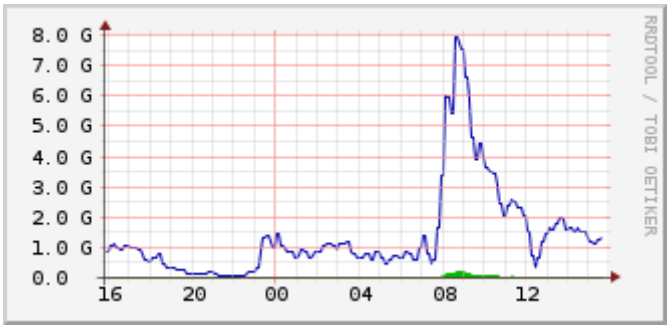
Apple *App Store*:

- Facebook: 294,4MB
- WhatsApp: 106,5MB

Google *Play Store*:

- Facebook: 74,0MB
- WhatsApp: 61,0MB

Wat betreft de Microsoft Windows 10 *updates* ziet RoutIT dat deze een steeds grotere belasting voor het netwerk vormen. De Windows 10 Anniversary *update* was ongeveer 3GB groot (J., 2016). RoutIT merkt de effecten, zoals in afbeelding 3 te zien is, van Windows en van de *updates* voor mobiele apparaten in het netwerkverkeer. Toen het verschijnen van Windows *updates* samenviel met de Olympische Winterspelen werden de interne bandbreedterecords gebroken.



Afbeelding 3: Windows update piek in het RoutIT netwerk.

9.3.5 Economische groei

Bij economische voorspoed is het aannemelijk dat bedrijven meer mensen aannemen. Deze nieuwe medewerkers hebben afhankelijk van hun taakomschrijving ook een werkplek met computer, telefoon en internettoegang nodig. Het is voor RoutIT niet mogelijk om te zien hoeveel van het internet gebruik makende apparaten er bij een klant op locatie staan. Het aantal apparaten is daarentegen wel een factor bij de groei van het netwerkverkeer. Als een eindklant meer medewerkers krijgt, zijn er meer medewerkers die deel kunnen nemen aan de hierboven beschreven factoren. Tevens is het zo dat hoe welvarender eindgebruikers worden hoe meer data zij gaan verstoken, omdat er dan duurdere abonnementen voor *streaming* diensten worden gekocht. Op die manier gaat het bandbreedte verbruik omhoog.

10 Datasets en de op te vragen informatie

Dit hoofdstuk behandelt de informatievragen en beschrijft en beoordeelt de gevonden interne en externe datasets. De onderkende datasets zijn als volgt:

- Extern:
  - Centraal Bureau voor de Statistiek (CBS)
- Intern:
  - IRMA
  - LogNet
  - RTK
  - Zabbix

10.1 Informatievragen aan de dataset

Hieronder volgen de informatievragen aan de datasets. Aangezien er veel combinatievragen bij zitten is ervoor gekozen om de vragen niet te groeperen op dataset. Tevens is het uitvragen en combineren van de datasets geen taak voor de afstudeerder of het NOC, maar een taak van de *Software Development* afdeling. Een uitzondering op de regel betreft de Zabbix data. Hiervoor heeft de afstudeerder, zoals hieronder beschreven, een Python script geschreven. Het “type order” is wat in de IRMA *Graphical User Interface* (GUI) het product wordt genoemd.

- De verdeling van Connectivity orders over handoverpoints gegroepeerd op type order (bandbreedte);
- De verdeling van Connectivity orders over de BSR / BNG’s gegroepeerd op type order (bandbreedte);
- Verdeling Connectivity orders over leveranciers gegroepeerd op type order (bandbreedte);
- Geplande opleverdatum van een Connectivity order samen met de bijbehorende type order (bandbreedte);
- Snelheden per handoverpoint gegroepeerd op best-effort en premium SLA;
- Snelheden per BSR / BNG gegroepeerd op best-effort en premium SLA;
- Betreft de nieuwe Connectivity order een opwaardering van een bestaande order;
- Betreft de nieuwe Connectivity order een afwaardering van een bestaande order;
- De bandbreedte van een nieuwe Connectivity order afgezet tegenover de vervallende order;
- Het aantal opgezegde Connectivity orders waarbij de eindklant geen vervangende RoutIT dienst neemt voor de desbetreffende order afgezet tegen nieuwe en bestaande orders op het moment van uitvoeren van de zoekopdracht gegroepeerd per leverancier en type order (bandbreedte);
- De geplande datum waarop een Connectivity order gedeactiveerd wordt en er dus geen verkeer meer mogelijk is vanuit de desbetreffende eindklant;
- Meetwaarden bandbreedte verbruik van en naar het internet;
- Meetwaarden bandbreedte verbruik per BSR / BNG;
- Meetwaarden bandbreedte verbruik op de interfaces aan de randen van de core;
- Meetwaarden bandbreedte verbruik op de BSR / BNG’s per lijn gegroepeerd tegen de verkochte order typen (bandbreedte);

10.2 Beoordeling van de datasets

In deze paragraaf worden in sub paragrafen de afzonderlijke datasets beoordeeld. Wegens de gevoeligheid van de data zijn er geen illustrerende afbeeldingen, waardoor informatie gelekt kan worden, opgenomen in de volgende sub paragrafen.

10.2.1 Centraal Bureau voor de Statistiek (CBS)

Het Centraal Bureau voor de Statistiek (CBS) voert onder andere analyses uit van de staat van de Nederlandse economie. De data van het CBS is via een *Application Programming Interface* (API) beschikbaar voor de buitenwereld. Bij het bekijken van de data in de API is gebleken dat de data vaak over voorgaande jaren gaat en dat de laatste prognoses niet zijn meegenomen. Voor dit onderzoek zijn de aantallen medewerkers en de groei daarvan van belang, omdat dit vervolgens iets kan zeggen over de groeiverwachting wat betreft afgenomen bandbreedte per lijn en het gemiddelde gebruik van een dergelijke lijn eventueel in combinatie met andere datasets. Na het bekijken van de door het CBS aangeboden dataset is gebleken dat de informatie over de trend aangaande groei in aantallen werknemers niet beschikbaar is.

10.2.2 Beoordeling van de IRMA en RTK-datasets

Het ordersysteem (IRMA) en het systeem met technische netwerkinformatie (RTK) kunnen gecombineerd veel van de voorgaande vragen beantwoorden. De bij deze datasets behorende vragen zijn voorgelegd aan het hoofd van de *Software Development* afdeling en deze heeft de vragen weer in handen gegeven van anderen. Met de personen die de informatie hebben aangeleverd zijn meerdere gesprekken gevoerd waarbij zij voornoemde aanvullingen hebben gegeven wat betreft de formulering opdat deze bruikbaar wordt voor de *Software Development* afdeling.

Op dit moment zijn niet alle vraagstukken te beantwoorden door deze datasets. Dat komt doordat de daarvoor benodigde informatie zoals informatie over de BSR's en BNG's niet wordt vastgelegd in IRMA en RTK.

Na analyse van de door de *Software Development* afdeling aangedragen antwoorden, moet geconcludeerd worden dat deze data te ernstig vervuild is om al tijdens dit onderzoek gebruikt te worden. De vervuiling bestaat bijvoorbeeld maar niet uitsluitend uit datums die elkaar tegenspreken en logischerwijs niet kunnen kloppen en het opnemen van de omschrijving van een invoerveld ("Kies hier uw *handoverpoint*") als *handoverpoint* waar een eindklantverbinding op moet worden aangesloten. In overleg met de opdrachtgever is besloten om te stoppen met het analyseren van de aangeleverde data en het opschonen van de datasets over te laten aan de *Software Development* afdeling. Wel is het van belang om in een later stadium de aangeleverde data uit de datasets IRMA en RTK toe te voegen aan de dataverzameling op basis waarvan de voorspellingen gedaan worden.

10.2.3 Beoordeling van de Zabbix dataset en de implementatie binnen RoutIT

Het monitoringpakket Zabbix wordt door RoutIT gebruikt om het gehele RoutIT netwerk in de gaten gehouden. Zabbix houdt onder andere statistieken bij over het netwerkverkeer en de configuraties, zoals geconfigureerde netwerkinterfaces en bestandssystemen. Deze statistieken kunnen we verdelen in verschillende meetwaarden. Iedere eindklant krijgt een virtuele interface op een BSR / BNG. Wegens de enorme hoeveelheid van virtuele *interfaces* worden hier geen statistieken van bijgehouden. Allereerst omdat het invoeren in Zabbix van de desbetreffende virtuele *interfaces* een enorm project is, ten tweede omdat het in de gaten houden van statistieken over de virtuele interfaces de monitoring te zwaar belast en ten derde omdat het gemakkelijk gebeurt dat een virtuele *interface* in de monitoring achterblijft na het opzeggen of verhuizen van een eindklantverbinding. Daarnaast is de kans groot dat een virtuele *interface* over het hoofd te wordt gezien tijdens de invoer waarmee de dataset onbetrouwbaar wordt. Wat RoutIT betreft weegt het invoeren van alle virtuele interfaces om voornoemde redenen op dit moment niet op tegen de nadelen. Daarmee zal Zabbix geen gemiddelden kunnen aandragen van het bandbreedte verbruik per internetverbinding door een eindklant. Deze vraag kan wel beantwoord worden indien in IRMA-informatie wordt opgenomen over de BSR's en BNG's, aangezien er in Zabbix wel het aantal actieve internetverbindingen van eindklanten wordt bijgehouden per BSR en BNG.

Zabbix kan op dit moment nog niet alle informatievragen beantwoorden. Dit komt zowel door de configuratie van Zabbix als door die van IRMA en RTK. Onder de beperkingen van de Zabbix dataset valt ook de periode waarover Zabbix op dit moment data bevat, voor meer informatie daarover zie volgend hoofdstuk. Wat Zabbix wel kan, is het aanleveren van onbewerkte meetwaarden betreffende gewone netwerkinterfaces op netwerkapparatuur. Als een eerste stap, waarbij de volgende stap gezet moet worden na dit onderzoek, is dergelijke data zeer bruikbaar om voorspellingen mee te doen. Dergelijke statistieken kunnen door het in het theoretisch kader besproken Kalman en ARIMA-filter gebruikt worden om voorspellingen te doen. Deze data zal met behulp van een Python script worden opgevraagd bij de Zabbix API en als invoer dienen voor verdere testen.

Historisch gezien is ten behoeve van de weergave bij iedere *interface* de uitgaande bandbreedte omgezet naar een negatieve waarde. Op die manier was het mogelijk om in Zabbix de in- en uitgaande bandbreedte leesbaar weer te geven. Aangezien dit bij andere projecten dan dit onderzoek tot problemen leidde, werd al tijdens de afstudeerperiode deze conversie niet meer toegepast. Door deze conversieslag moet de monitoringdata bewerkt worden. Voor meer informatie over deze bewerking zie hieronder.

De lijst met na transformatie geschikte datasets is als volgt:

- IRMA
- RTK
- Zabbix

11 Beoordeling van Zabbix voor het doen van voorspellingen

Uit navraag bij team Monitoring is gebleken dat het voornoemde Zabbix al beschikt over *forecasting* mogelijkheden. Aangezien bij een project 20% van de kosten gaat zitten in de implementatie van een product of methode en 80% in het beheer daarvan (McCabe, 2007), is het vanuit het oogpunt van efficiëntie en kostenbesparing raadzaam om eerst de Zabbix *forecasting* functionaliteit te bekijken.

11.1 Zabbix *forecasting*

In deze paragraaf wordt allereerst het één en ander uitgelegd over de omgang van Zabbix met de verzamelde data en de keuzen van RoutIT daarin. Vervolgens wordt dieper ingegaan de *forecasting* functionaliteit zelf.

11.1.1 Het verschil tussen historische en trend data in Zabbix

De via Zabbix door RoutIT verzamelde data valt uiteen in historische data, welke twee weken bewaard wordt, en trend data welke 365 dagen bewaard wordt. Het verschil tussen historische data en trends is dat historische data iedere gemeten waarde bevat en dat trends alleen de gemiddelde, minimale en maximale waarden over een tijdsbestek van een uur bevat (Zabbix, 2018c). Het voordeel van trends ten opzichte van historische data is dat er veel minder opslagruimte nodig is, en dat voor het weergeven van de data in een grafiek veel minder systeembronnen en tijd nodig zijn. De Zabbix API heeft, zo leert de ervaring, een limiet ingesteld op de hoeveelheid data die weergegeven kan worden door andere tools zoals Grafana. Grafana wordt door RoutIT gebruikt omdat Grafana voor RoutIT meer mogelijkheden biedt om de data goed weer te geven in grafieken en zo de analyse te vergemakkelijken.

Wegens de recente migratie van het oude naar het nieuwe Zabbix platform, is er weinig data beschikbaar. De oudst beschikbare data gaat terug tot begin dit jaar. Het oude Zabbix platform, van voor de herimplementatie van de monitoring, is nog wel beschikbaar. Wegens een actie om te voorkomen dat het oude Zabbix platform onbruikbaar werd, bevat het oude Zabbix platform slechts de data van november 2017 tot en met begin januari 2018. Oudere versies van Zabbix hadden, volgens collega's, moeite de hoeveelheden data die RoutIT Zabbix liet verzamelen, te beheren. Dit heeft ervoor gezorgd dat bij deze oude Zabbix versies de *database* sneller gevuld werd dan dat de *database* opgeruimd kon worden. Hierdoor werd Zabbix dermate traag dat Zabbix niet meer gebruikt kon worden voor het *troubleshooten* van problemen. Verder heeft de Zabbix API op het oude platform geen ondersteuning om de gewenste data uit te lezen. Deze beperkingen zorgen ervoor dat op dit moment de termijn die voorspeld kan worden beperkt is.

11.1.2 De Zabbix *forecasting* functionaliteit beschreven

De Zabbix *forecasting* functionaliteit is een op statistische functies gestoelde functionaliteit (Zabbix, 2015). Deze functionaliteit is onderdeel van de *trigger* oftewel de alarmeringsfunctionaliteit. Bij een alarm is het in Zabbix zo dat bij iedere opgehaalde meetwaarde wordt gekeken of de drempelwaarde van het bijbehorende alarm wordt overschreden. Op dat moment wordt de status van het alarm aangepast (Zabbix, 2018d). Aan een alarm kunnen meerdere drempelwaarden worden gehangen met ieder hun eigen status. Doordat de *forecasting* functie gebaseerd is op de alarmeringsfunctionaliteit, is het niet mogelijk om een grafiek weer te geven zonder de hulp van een *calculated item.* Een *calculated item* berekent op basis van een interval de gewenste voorspelling, waarna deze kan worden weergegeven in een grafiek, zie voor een voorbeeld afbeelding 4.

Afbeelding 4: *Forecast* aangaande vrije schijfruimte samen met de daadwerkelijke waarde (Zabbix, 2015).

Door de opzet van de Zabbix *forecasting* functie, namelijk op basis van alarmering, is het niet mogelijk om een grafiek te maken voor de toekomst. In de documentatie (Zabbix, 2015), staan wel grafieken maar deze grafieken gaan ook altijd tot het nu. *Forecasting* in Zabbix houdt daarmee in dat bij een nieuw binnengekomen waarde er gekeken wordt of er een alarm moet afgaan omdat verwacht wordt dat er een drempelwaarde overschreven gaat worden. De verwachte waarde kan op basis van een *calculated item* worden weergegeven in een grafiek samen met de werkelijke waarde. Zabbix heeft hiervoor altijd een concrete waarde nodig. Een voorspelling op basis van historische data over langere tijd is daarmee niet mogelijk.

Voor het goed instellen van Zabbix *forecasting* is grondige kennis van statistiek nodig en verificatie aan de hand van bijvoorbeeld MatLab. Simpelweg zeggen “de voorspelde en gemeten waarde komen overeen” geeft geen garanties aangaande de juistheid van de configuratie (Zabbix, 2015).

Door het kennisintensieve traject bij het configureren van Zabbix *forecasting* is de beheersbaarheid laag. Op vele plekken kunnen tijdens het configureren fouten gemaakt worden die gevolgen hebben voor de betrouwbaarheid van de voorspelling. Door de ondoorzichtigheid van dit proces is het goed mogelijk dat een fout pas ontdekt wordt als er al netwerkcongestie optreedt of juist grote (onnodige) investeringen gedaan zijn. Daarbij komt dat de Zabbix *forecasting* functionaliteit in zijn meer geavanceerde vorm veel systeembronnen verbruikt. RoutIT is voor het herkennen van storingen net als andere bedrijven afhankelijk van het gebruikte monitoringsysteem. Daarom is het voor RoutIT onaanvaardbaar als de stabiliteit van de monitoring in gevaar komt door het voorspellen van de toekomst. Het repliceren van de monitoringdata naar een tweede Zabbix instantie zou tot de mogelijkheden behoren, ware het niet dat het dan nog steeds niet mogelijk is om andere datasets dan alleen de monitoringdata mee te nemen in de voorspelling. Aangezien 80% van de kosten van een oplossing naar het beheer van de oplossing gaat (McCabe, 2007), is het niet verstandig om voor het doen van voorspellingen te kiezen voor een functie met dusdanig veel onzekerheden en beperkingen. In overleg met de opdrachtgever is om voornoemde redenen besloten om op zoek te gaan naar een oplossing die deze beperkingen niet heeft.

12 Onderzoek naar een *Business Intelligence* (BI) oplossing

Het combineren en analyseren van datasets is zoals beschreven in het theoretisch kader het expertise gebied van *Business Intelligence* (BI) oplossingen. Een BI-oplossing betrekt de te analyseren data, eveneens beschreven in het theoretisch kader, van een *data warehouse*. Tijdens het onderzoek is gebleken dat de verschillende BI-oplossingen vaak gecombineerd komen met een *data warehouse*. Voor het verkrijgen van de namen van BI-oplossingen zijn twee verschillende methoden gebruikt. Allereerst is aan de collega's van het NOC gevraagd of zij geschikte BI-oplossingen kennen. Vervolgens is het Gartner *Magic Quadrant* (Sallam et al., 2017) aangaande BI-oplossingen geraadpleegd. Uit dit proces is een lijst van 27 namen voortgekomen. Ten behoeve van de doorlooptijd is besloten om niet nog via een generieke zoekmachine als DuckDuckGo te zoeken naar andere BI-oplossingen.

In de volgende paragrafen wordt het proces van de pakketselectie en de MoSCoW analyse beschreven. Het uiteindelijk gekozen pakket zal in een apart hoofdstuk beschreven worden.

12.1 Voorselectie pakketten

Op basis van de hierboven beschreven onderzoeksmethoden is de volgende lijst met bedrijven samengesteld. Indien een bedrijf meerdere BI-oplossingen heeft staat achter het bedrijf een productnaam vermeld.

- Alteryx
- Birst
- Board International
- ClearStory Data
- Datameer
- Domo
- Elastic
- IBM
- Information Builders
- Logi Analytics
- Microsoft Azure ML
- Microsoft PowerBI
- MicroStrategy
- Oracle
- Pentaho
- Pyramid Analytics
- Qlik
- Salesforce
- SAP
- SAS
- Sisense
- Splunk
- Tableau
- ThoughtSpot
- TIBCO Software
- Yellowfin
- Zoomdata

Op basis van de bovenstaande lijst is een voorselectie gemaakt. De websites van de verschillende leveranciers zijn doorzocht op de termen "*forecasting"* en "*predictive*" en hun vervoegde varianten. Op deze manier zijn de pakketten weggefilterd die geen affiniteit hebben met het doen van voorspellingen.

12.2 Selectie op basis van de MoSCoW analyse

In deze paragraaf wordt beschreven hoe de MoSCoW analyse uit bijlage C gebruikt is om dit onderzoek uit te voeren. De MoSCoW analyse is opgesteld door de afstudeerder en geaccordeerd door de opdrachtgever. Zoals hierboven beschreven is de in de voorgaande paragraaf weergegeven lijst gefilterd op basis van het voorkomen van de woorden "*forecasting"* en "*predictive*" samen met hun vervoegingen. Aangezien de *Must haves* van een MoSCoW analyse de enige eisen bevatten zonder welke een pakket niet ingezet kan worden, zijn alle overgebleven pakketten op basis van een leveranciersdocumentatiestudie gevalideerd op de aanwezigheid van de *Must haves*. Tijdens dit onderzoek zijn punten toegekend op de verschillende onderdelen van de *Must have*. Indien een pakket alleen de minimale mogelijkheden heeft op het desbetreffende gebied, zijn minder punten toegekend dan wanneer het pakket uitgebreide mogelijkheden heeft. Hiervoor is gekozen om zo een onderscheid te maken tussen flexibele en minder flexibele inzetbare pakketten. De top vijf van pakketten die het best uit de *Must have* documentatiestudie komt is als volgt:

- Datameer
- Elastic
- Logi Analytics
- Splunk
- TIBCO Software

In overleg met de opdrachtgever is besloten om van de pakketten die op alle *Must haves* punten behaald hebben, de vijf hoogst scorende aan een markt verkennende PoC te onderwerpen. Op deze manier kan beter beoordeeld worden of de *Must haves* ook daadwerkelijk aanwezig zijn en is het ook beter dan middels een documentatiestudie mogelijk om de overige delen van de MoSCoW analyse te beoordelen. Ten behoeve van deze fase van het onderzoek is een Testplan opgesteld, zie bijlage F. Voor het ontwerp van de markt verkennende PoC zie bijlage E. Tijdens de markt verkennende PoC zijn ten behoeve van de grondigheid de *Must haves* opnieuw beoordeeld. De gehele pakketselectie is weergegeven in bijlage G.

Van voornoemde lijst is, op Elastic na, contact geweest met alle leveranciers. Bij Elastic was dit contact niet nodig omdat het verkrijgen van licenties via een geautomatiseerd proces verliep. Logi Analytics gaf aan dat zij de door ons gewenste functionaliteit nog in ontwikkeling hebben en wilde daarom niet meedoen aan deze PoC.

12.3 Integratie met de bestaande productiefaciliteiten

Integratie van de BI-oplossing met de bestaande productiefaciliteiten van RoutIT is zowel vanuit het beveilingsoogpunt als bij het importeren van de data in de BI-oplossing van belang. De veiligheid van de koppeling tussen de BI-oplossing en de productiefaciteiten is van belang in verband met de concurrentiegevoeligheid van de te importeren data. Deze informatie dient goed beveiligd te zijn tegen ongeautoriseerde toegang. Voor veilig inloggen op systemen maakt RoutIT gebruik van het *Lightweight Directory Access Protocol* (LDAP). De eis voor het ondersteunen van LDAP is daarom opgenomen als *Must have* in de MoSCoW-analyse.

Op dit moment worden al data exports van IRMA en RTK gedaan, het exportformaat betreft CSV. Met behulp van Python is het tevens mogelijk om de monitoringdata van Zabbix op te vragen en weg te schrijven in het CSV-formaat. Daarom is het kunnen importeren van het CSV-formaat als *Must have* opgenomen in de MoSCoW-analyse.

12.4 Beoordeling voorspellingsmogelijkheden

Zoals in het testplan beschreven, is er een Python script geschreven voor het genereren van fictieve data. Deze fictieve data is vervolgens gebruikt om de voorspellingsmogelijkheden te onderzoeken. Op basis van deze fictieve data is in het pakket gekeken of er *forecasting* mogelijkheden in het pakket zitten. Zowel op basis van de beschikbare documentatie als door het pakket te onderzoeken, bleek *forecasting* alleen bij Elastic en Splunk een mogelijkheid te zijn. De andere pakketten bleken geen *forecasting* functionaliteit ingebouwd te hebben. Datameer had wel een aantal functies die kunnen assisteren bij het doen van voorspellingen, maar dat zorgt alsnog voor heel veel extra handwerk.

Wat betreft het doen van voorspellingen bleek uiteindelijk dat het goed inregelen van *forecasting* bij Elastic geen intuïtief proces is. Voor het juist configureren van Elastic is het nodig om rechtstreeks tegen de API aan te praten, een taak die eigenlijk door een GUI of script uitgevoerd moet worden. De noodzaak om via de API Elastic juist te configureren maakt Elastic op basis van de eerder vermelde 80% / 20% regel (McCabe, 2007) een arbeidsintensieve en onbeheersbare optie. Samen met de af te dragen licentiegelden is het inzetten van de productcombinatie van Elastic voor RoutIT niet aantrekkelijk. Om basis handelingen als het importeren van data en het koppelen met LDAP te implementeren is al gespecialiseerde kennis nodig. De eis voor gespecialiseerde kennis maakt het lastig voor andere leden van team Monitoring om bij uitval van de Elastic specialist door bijvoorbeeld ziekte over te nemen als gewenst door RoutIT. Licentiekosten en het volgen van trainingen zijn voor RoutIT niet onoverkomelijk, maar dan moet het product wel beheersbaar en bij uitval van de specialist het beheer over te nemen zijn. Om de producten van Elastic efficiënt en betrouwbaar te gebruiken en te beheren is dagelijks gebruik van de producten noodzakelijk. Tevens zijn de instelmogelijkheden van de *forecasting* zeer beperkt, en daarmee niet inzichtelijk of beïnvloedbaar. Om de voornoemde redenen vallen de producten van Elastic af als mogelijke BI-oplossing.

Het pakket Splunk van de gelijknamige leverancier, is in staat gebleken om binnen een redelijk tijdsbestek en zonder al teveel specialistische kennis of inwerktijd in te zetten voor het doen van voorspellingen. Het proces om de *forecasting* mogelijkheden van Splunk te benutten bleek intuïtief te werken met als uitkomst een eenvoudig model. Voor de complexere modellen welke door RoutIT gewenst zijn, bestaan trainingen om de bijbehorende taal, genaamd SPL, onder de knie te krijgen. Het volgen van dergelijke trainingen is voor RoutIT geen belemmering. Indien eenmaal een model geschreven is, kan het desbetreffende model ook door andere leden van team Monitoring gelezen en begrepen worden zonder dat dagelijks gebruik van Splunk noodzakelijk is.

De voorgaande analyse heeft geleid tot een voordracht van Splunk als te kiezen BI-oplossing door RoutIT. Dit advies is door de opdrachtgever bevestigd en overgenomen. In het volgende hoofdstuk worden de tests en gesprekken met Splunk consultants beschreven.

13 Beschrijving van Splunk en de gemaakte keuze binnen Splunk

In dit hoofdstuk worden het proces en de beoordeling van Splunk op de bruikbaarheid voor RoutIT beschreven. Splunk is tijdens het beoordelen grondig getest wat betreft voorspellingsmogelijkheden.

13.1 Contact met Splunk vertegenwoordigers

Tijdens het testen met Splunk zijn er verschillende vragen gerezen aangaande de in de Splunk GUI gebruikte termen en het Splunk licentiemodel. Daarom is contact opgenomen met de *Sales* afdeling van Splunk in de Verenigde Staten (VS) en het Verenigd Koninkrijk (VK). De veelheid aan gecontacteerde vertegenwoordigers en kantoren komt voort uit het constant doorschuiven van vragen van de afstudeerder naar een nieuw persoon door Splunk. Ondertussen is het onderzoek naar de voorspellingsmogelijkheden op basis van de beschikbare documentatie vanuit Splunk vervolgd. Op basis daarvan heeft de afstudeerder contact gezocht met een bedrijf dat Splunk in Nederland samen met ondersteuning en trainingen verkoopt. Wat betreft de aanschaf is het niet mogelijk dit direct bij Splunk te doen, dit moet via een dergelijke wederverkoper. Het desbetreffende bedrijf is geselecteerd door de opdrachtgever. Dit bedrijf is SMT.

Een consultant van SMT is naar RoutIT gekomen, waar de afstudeerder samen met de opdrachtgever een gesprek heeft gehad met voornoemde consultant. Deze consultant heeft de vragen van de afstudeerder vervolgens ter beantwoording doorgestuurd binnen SMT. De opdrachtgever is nu met SMT in overleg over de aanschaf van Splunk.

13.2 Opvragen monitoringdata

Voor dit onderdeel van het onderzoek is gekozen om met echte data te werken. Door met echte data te werken is het mogelijk om de voorspelling te vergelijken met de daadwerkelijke meting. In tegenstelling tot Zabbix, waar het vergelijken van voorspellingen met echte data wordt afgeraden is dit bij Splunk niet het geval. Tijdens het doorlezen van de Splunk documentatie werd het vergelijken van voorspellingen met echte metingen zelfs geadviseerd. Het is de afstudeerder echter niet gelukt om de bewuste webpagina terug te vinden, waardoor een bronvermelding hier ontbreekt. De Zabbix software beschikt over een API om monitoringdata uit te lezen. Iemand van team Data heeft een Python module geschreven waarmee de monitoringdata is uit te lezen. De desbetreffende *engineer* weet veel van de Zabbix API en Python, daarom is omwille van de doorlooptijd ervoor gekozen om dit niet door de afstudeerder te laten doen.

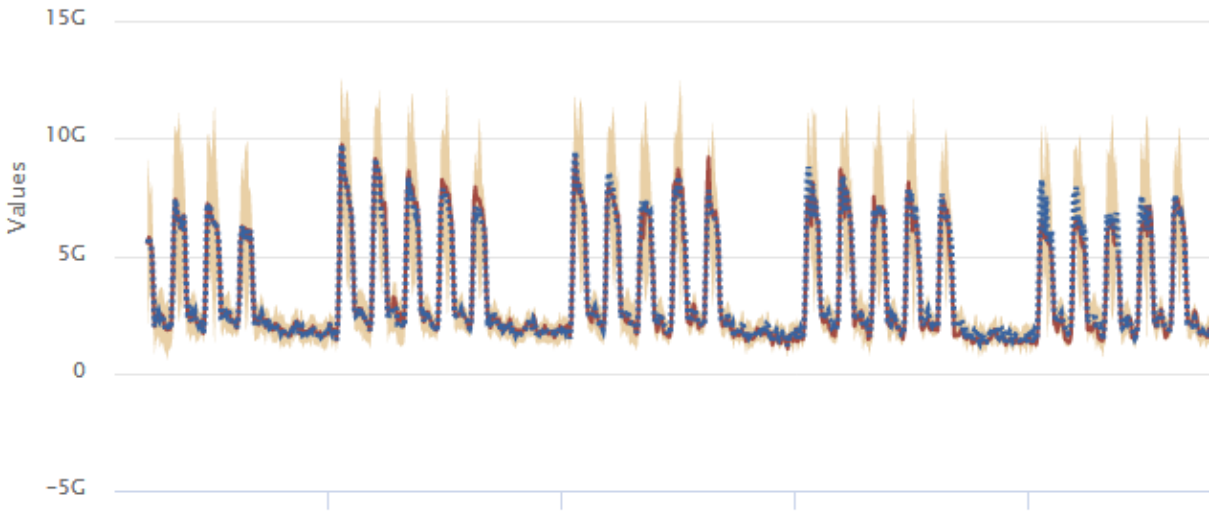
De afstudeerder heeft vervolgens verschillende andere Python scripts geschreven. Allereerst twee scripts om zowel trend als historische monitoringdata op te vragen en weg te schrijven in CSV-formaat, daarnaast is een conversie script geschreven om negatieve monitoringwaarden om te zetten naar positieve waarden, zie bijlage H. Dit is gedaan door alle voorloop mintekens uit de bandbreedtekolom te verwijderen. Op deze manier was de monitoringdata toch bruikbaar om mee te testen.

13.3 De Splunk *Machine Learning Toolkit* (MLT)

Middels de Splunk Machine Learning Toolkit (MLT) app is het mogelijk om voorspellingen te doen op basis van historische monitoringdata. MLT is een *app* die alleen beschikbaar is voor Splunk Enterprise. Daarnaast is het mogelijk om data uit andere datasets mee te nemen in het *forecasting* proces. MLT bevat hiervoor een uitbreiding van SPL en een aantal *assistants* om het proces te vergemakkelijken. Tijdens de test is gebruik gemaakt van *Forecast Time Series assistant*. Deze *assistant* wordt binnen de Splunk GUI aanbevolen voor het voorspellen van netwerkverkeer, zie afbeelding 5.

Afbeelding 5: Voorbeelden *Forecast Time Series assistant*.

Tevens gaat het bij de door Zabbix gegenereerde data inderdaad om tijdgebonden data, oftewel *time series* waardoor de voornoemde *assistant* de beste mogelijkheden biedt. Tijdgebonden data is data die niet los gezien kan worden van het moment waarop de observatie plaats had (Anderson, 1994). Dit is bij uitstek het geval bij de monitoringdata aangaande het netwerkverkeer, omdat de data, zie afbeelding 6, iedere week ongeveer hetzelfde patroon laat zien en daarmee tijdgebonden is.



Afbeelding 6: Tijdgebonden monitoringdata.

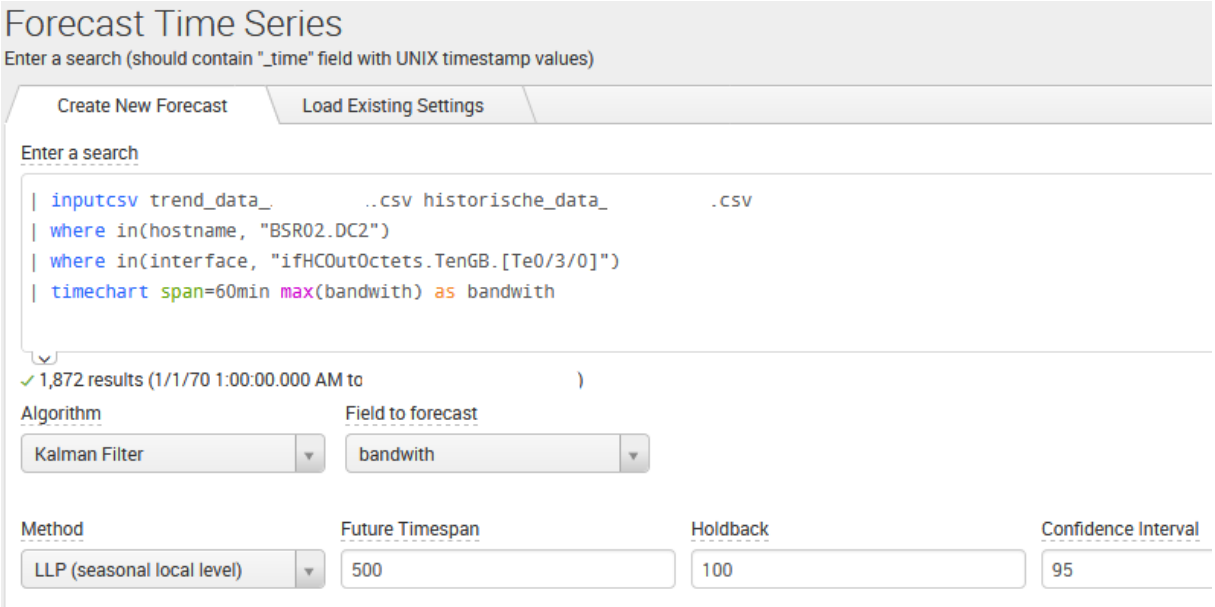
Indien een deel van deze grafiek los bestudeerd zou worden, hebben we alleen een getal zonder enige relatie met andere gebeurtenissen. Bestuderen we dat getal samen met de tijd dan is bijvoorbeeld de overgang van dag naar nacht of van werkweek naar weekeinde te herkennen.

13.4 Kalman vs. ARIMA

Binnen Splunk MLT zijn er twee algoritmes beschikbaar ten behoeve van het werken met tijdgebonden data, het Kalman en het ARIMA-filter. Voor meer informatie over beide filters zie het theoretisch kader. Het ARIMA-filter heeft, zoals beschreven in het theoretisch kader, moeite met de door RoutIT aangeboden *seasonal data* en vraagt daarnaast veel statistische kennis waardoor het gebruik van het ARIMA-filter niet intuïtief is maar specialistische kennis vraagt. Indien de keuze valt op het ARIMA-filter wordt ook het argument om de producten van Elastic niet te adviseren tenietgedaan. Het Kalman filter is binnen Splunk een stuk toegankelijker en op een intuïtieve manier te gebruiken. Naast alle technische beperkingen als beschreven in het theoretisch kader komt hier ook weer de al eerder beschreven 80% / 20% regel (McCabe, 2007) om de hoek kijken. Om deze redenen is er in overleg met de opdrachtgever voor gekozen om het Kalman filter te gebruiken voor het doen van voorspellingen.

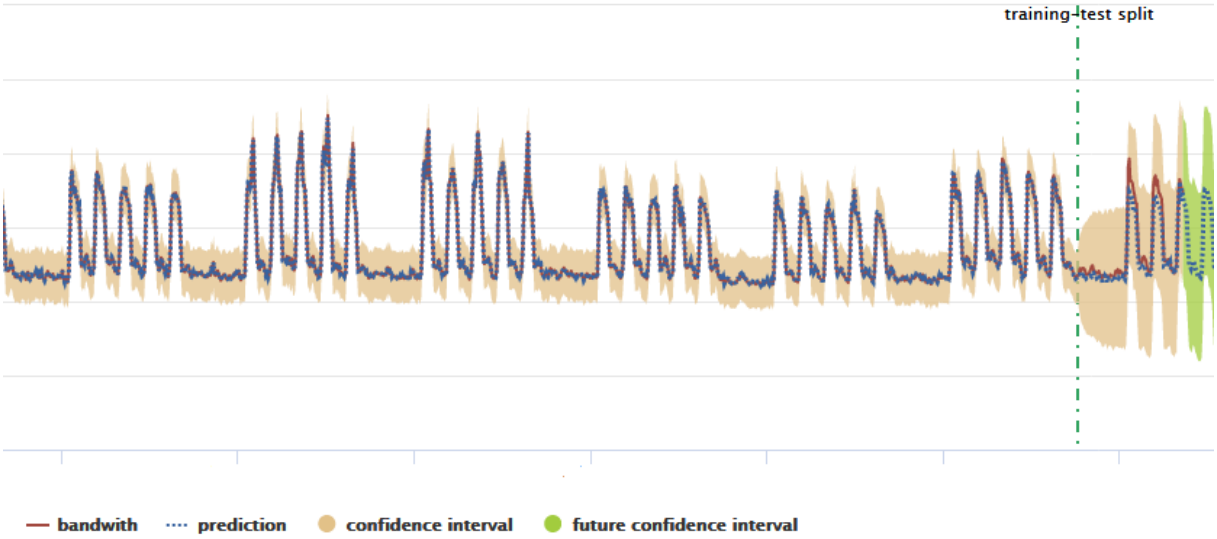
13.5 Splunk *forecast* model in MLT

Als uitgangspunt is het voorbeeld voor het voorspellen van internetverkeer uit de Splunk MLT genomen. Dit voorbeeld is vervolgens aangepast aan de RoutIT wensen. Zie voor het model afbeelding 7. Dit model importeert allereerst de dataset waarna het een specifieke interface van een specifieke router selecteert. Vervolgens wordt de data per 60 minuten samengevat middels de “span” optie. Voor 60 minuten is gekozen omdat de Zabbix *trend* data al samengevat is per 60 minuten. De *span* optie is verplicht voor het werken met het *timechart* commando en voorkomt dat de browser door een veelheid aan weer te geven data vastloopt. Dat de browser vastloopt bij een korte *span* tijd is proefondervindelijk vastgesteld. De manier waarop dit samenvatten gebeurt wordt aangegeven met de optie “max”, door het opgeven van deze optie wordt per uur de maximale waarde gepakt. Andere mogelijkheden zijn het nemen van de minimale en gemiddelde waarde. Gekozen is voor de max optie omdat RoutIT er zeker van wil zijn dat er altijd voldoende capaciteit is, dit wordt bereikt door met de hoogste getallen te werken.



Afbeelding 7: Splunk *forecasting* model in de MLT.

Zoals in afbeelding 7 te zien is, zijn er meerdere methoden beschikbaar binnen het Kalman filter. De enige methoden die geschikt zijn voor de *seasonal data* van RoutIT zijn LLP en LLP5, deze geschiktheid is te zien in het selectiemenu. Door het vergelijken van de voorspellingen van meerdere dagen, zowel voor LLP als LLP5, is gebleken dat LLP de werkelijkheid het beste benadert. Het verschil met de werkelijkheid is voor LLP kleiner dan LLP5. Deze verschillen zijn Splunk niet aan te rekenen aangezien er gewoon te weinig monitoringdata beschikbaar is op dit moment. De *Future Timespan* optie bepaalt hoeveel voorspellingen Splunk mag doen, waar de *Holdback* optie aangeeft hoeveel data er achter gehouden moet worden voor het gekozen algoritme (Splunk, 2018). Het achterhouden van data geeft de onderzoeker de mogelijkheid om de daadwerkelijke data met de echte data in Splunk te vergelijken, zie afbeelding 8.



Afbeelding 8: Splunk *forecast* met achtergehouden data. Met aan de linkerkant tot aan de verticale lijn de data welke is gebruikt door het voorspellingsalgoritme.

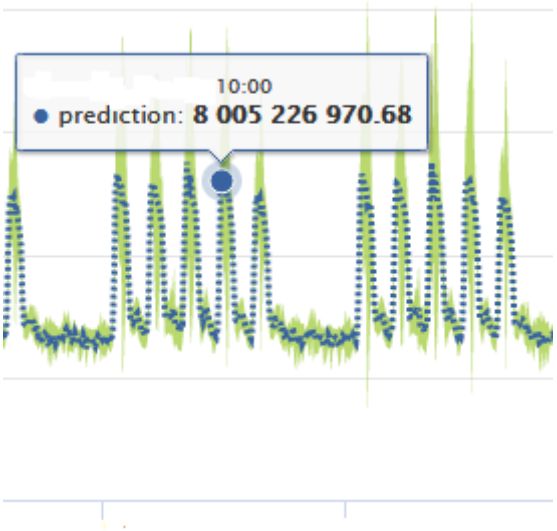
De *Confidence Interval* uit afbeelding 7, moet, blijkens de omschrijving in de Splunk GUI een waarde zijn tussen de 0 en 99 en geeft aan hoe tolerant Splunk is ten opzichte van onzekerheden in de voorspelling. Hoe hoger de *Confidence Interval*, hoe toleranter Splunk is aangaande onzekerheden. Tijdens de testen is gebleken dat het ophogen van voornoemde waarde naar 99 slechts een marginaal betere *forecast* geeft. Daarom is deze waarde op de standaardinstelling gelaten.

13.6 Uitvoeren simulaties

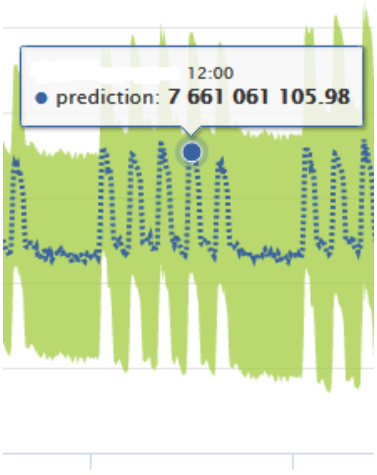
Het woord simuleren duidt erop dat er met fictieve data een forecast gedaan wordt. In overleg met de opdrachtgever is er echter voor gekozen om echte data in de PoC te stoppen. Hiervoor is gekozen om te onderzoeken hoe Splunk reageert op een dataset waarbij het patroon iedere dag hetzelfde is, maar de waarden iedere dag anders zijn. Te zien is dat het netwerkverkeer een groei laat zien. Een dergelijk patroon zou te vangen zijn geweest in een Python script voor het genereren van fictieve data ten behoeve van de gewenste simulatie. Echter zou dit aan de ene kant veel tijd vergen wegens de bestudering van de echte data en aan de andere kant wegens het schrijven en valideren van het Python script. Wegens de doorlooptijd is daarom gekozen om te werken met echte data. Tevens geeft het gebruik van echte data ook de mogelijkheid om te onderzoeken hoe Splunk omgaat met eenmalige uitschieters in het netwerkverkeer.

Zoals bij de beoordeling van de datasets beschreven, is door de overgang naar een nieuwe Zabbix omgeving en beperkingen in de Zabbix software van de oude omgeving niet genoeg data beschikbaar ten behoeve van het doen van meerjarige voorspellingen. Gekozen is om de monitoringdata van een *router* te gebruiken die het langst in de monitoring van de nieuwe Zabbix omgeving staat.

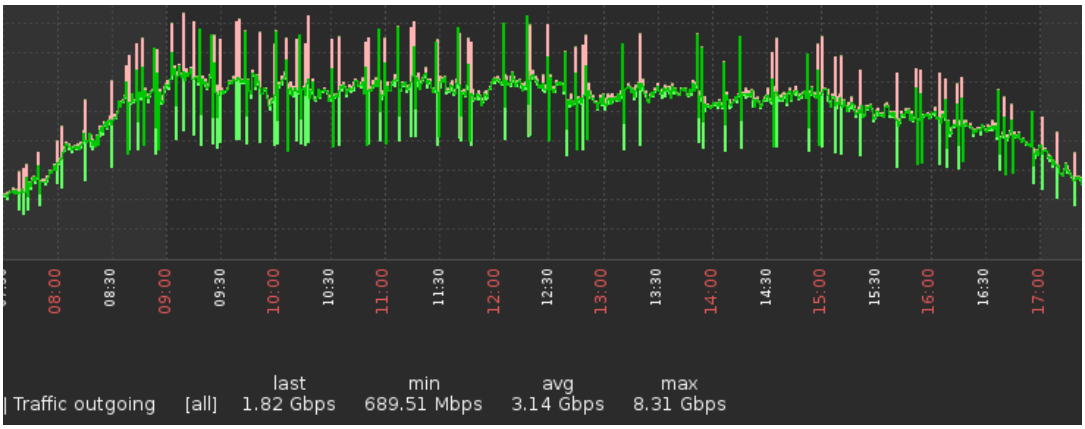
Gekozen is om de data uit de nieuwe Zabbix te halen en vervolgens een model te schrijven zoals dat hierboven besproken is. Vervolgens is gekeken of de voorspellingen van Splunk overeenkomen met de monitoringdata uit Zabbix. Tijdens de simulatie is één enkele keer de data uit Zabbix gedownload ten einde te voldoen aan de voor onderzoeken geldende eis om telkens maar één variabele te veranderen. Deze variabele betreft de gebruikte methode uit het Kalman filter. De gedane voorspellingen gaan over dagen waarvan geen data in de Splunk software beschikbaar was. Onderstaand zijn van dezelfde dag zowel de Splunk *forecast* als de grafiek van Zabbix over die dag weergegeven. Wegens de geheimhouding is tijdsinformatie verwijderd. In afbeelding 9 is de grafiek te zien voor LLP, in afbeelding 10 is de grafiek te zien voor LLP5 en in afbeelding 11 is de grafiek voor Zabbix te zien. Alle grafieken geven dezelfde dag weer.



Afbeelding 9: Splunk voorspelling met LLP. Het getal geeft de voorspelde piek in bits per seconde weer.



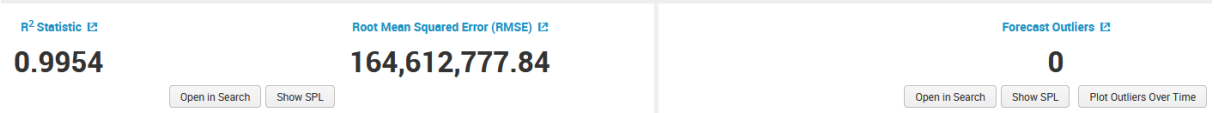
Afbeelding 10: Splunk voorspelling met LLP5. Het getal geeft de voorspelde piek in bits per seconde weer.



Afbeelding 11: Werkelijke data Zabbix.

Bij het vergelijken van afbeelding 9, 10 en 11 is te zien dat de voorspellingen de piek van de dag slecht kunnen voorspellen. Aan SMT, de Splunk wederverkoper, is de vraag gesteld hoeveel data noodzakelijk is voor het doen van voorspellingen. Hierop is nog geen reactie gekomen. Wel is uit eigen tests met de hoeveelheid data gebleken dat de voorspelling verbetert bij het gebruik van meer data. Tijdens deze simulatie is namelijk begonnen met het ingeven van louter Zabbix *History data* wat door de configuratie van Zabbix slechts 14 dagen bedraagt. Gebleken is dat met een dusdanige hoeveelheid data het weekend niet goed werd voorspeld. Bij het ingeven van meer data verbeterden de voorspellingen en werd overgang van werkweek naar weekend wel voorspeld.

In afbeelding 12 en 13 worden indicatoren weergegeven die de juistheid van de voorspelling volgens het Kalman filter beschrijven. Het gaat om de R2 *Statistic*, de *Statistic Root Mean Squared Errors* (RMSE) en de *Forecast Outliers*. Afgaande op de informatie in Splunk beschrijft de R2 *Statistics* de zekerheid van het Kalman filter over de gedane voorspelling op een schaal van 0 tot 1, waarbij hoger beter is, de RMSE waarde geeft aan hoeveel de voorspelde waarde verschilt ten opzichte van de data op basis waarvan het algoritme getraind wordt (Lacey, 2018). Het veld *Forecast Outliers* geeft aan hoeveel waarden te ver verwijderd zijn van het statistische patroon om bruikbaar te zijn voor het doen van voorspellingen.Bij het vergelijken van bovenstaande afbeeldingen is te zien dat de LLP-methode de betere voorspellingen geeft, terwijl afgaande op de RMSE en de R^2 *Statistics* scores uit afbeeldingen 12 en 13 juist LLP5 betere voorspellingen geeft. Alle beoordelingsscores zijn voor LLP5 beter dan voor LLP, maar tijdens de hierboven beschreven datavergelijkingen is juist gebleken dat LLP de betere voorspellingen doet.



Afbeelding 12: Beoordelingsscores voor LLP5.

Afbeelding 13: Beoordelingsscores voor LLP.

14 Conclusie

In dit hoofdstuk worden de conclusies van het onderzoek weergegeven. Allereerst zullen de deelvragen beantwoord worden waarna de onderzoeksvraag beantwoord wordt.

1. **Welke netwerksegmenten zijn betrokken bij netwerkcongestie binnen de RoutIT core-dienstverlening en hoe beïnvloeden die elkaar?**

Het pad van netwerksegmenten met de grootste gevoeligheid voor netwerkcongestie loopt vanaf de Internet laag via de aggregatie laag naar de *Access* laag. Deze netwerksegmenten vallen onder de verantwoordelijkheid van team Infra. De netwerksegmenten van de teams Data, Monitoring en *Voice* zijn niet betrokken bij netwerkcongestie of vallen buiten dit onderzoek.

Voornoemde netwerksegmenten beïnvloeden elkaar doordat eindklanten tegelijk grote hoeveelheden data gaan opvragen op de *Access* laag. De data komt binnen op de Internet laag en gaat vervolgens via de aggregatie laag terug naar de *Access* laag. Door de grote hoeveelheden data ontstaat netwerkcongestie in dit pad. Factoren voor netwerkcongestie betreffen onder andere meerdaagse sportevenementen, *software updates,* terroristische aanslagen en extreem weer. Indien de datastroom in dit pad afneemt verdwijnt ook de netwerkcongestie.

2. **Welke voor dit onderzoek relevante datasets heeft RoutIT tot zijn beschikking en in hoeverre zijn deze bruikbaar voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie in de core-dienstverlening te voorkomen?**

De voor RoutIT relevante datasets in het kader van dit onderzoek zijn als volgt:

- IRMA
- RTK
- Zabbix

De datasets van IRMA en RTK bevatten te veel vervuiling om bruikbaar te zijn tijdens dit onderzoek. Na opschoning door de afdeling *Software Development* is de verwachting dat deze data zeer bruikbaar is, omdat daarmee informatie over geplande groei gebruikt kan worden bij het doen van voorspellingen. Door de huidige vervuiling zijn de datasets IRMA en RTK in overleg met de opdrachtgever niet gebruikt bij dit onderzoek. Zabbix is na het uniformiseren van de data doormiddel van het transformeren van eventuele negatieve waarden naar positieve waarden, zeker bruikbaar voor het doen van korte termijn voorspellingen. Deze beperking komt doordat Zabbix op dit moment te weinig data bevat om voorspellingen te doen over de middellange termijn. Dit probleem wordt met het verstrijken van de tijd automatisch opgelost doordat vanzelf meer monitoringdata naar Zabbix stroomt.

3. **Welke mogelijkheden aangaande *business intelligence* zijn beschikbaar en kunnen geïntegreerd worden met de bestaande RoutIT productiefaciliteiten aangaande het doen van voorspellingen betreffende het netwerkverkeer ten einde netwerkcongestie in het netwerk te voorkomen?**

De volledige lijst van gevonden *Business Intelligence* (BI) oplossing leveranciers is in het desbetreffende hoofdstuk te vinden en het herhalen van deze lijst voegt niets toe aan de conclusie. De volgende lijst bevat alle leveranciers die tijdens het uitvoeren van de documentatiestudie bij de MoSCoW-analyse punten hebben gescoord voor de *Must haves* aangezien de bijbehorende pakketten aan de essentiële eisen voldaan hebben. Onder de essentiële eisen vallen onder andere integratie eisen. Deze pakketten zijn tevens door de voorselectie heen gekomen waarbij een eis was om gewag te maken van voorspellende mogelijkheden.

- Datameer
- Elastic
- Logi Analytics
- MicroStrategy
- SAS
- Splunk
- TIBCO Software

4. **In hoeverre is de Zabbix functie "*forecasting*" bruikbaar voor toepassing door RoutIT voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie te voorkomen?**

De Zabbix *forecasting* functionaliteit is om meerdere redenen ongeschikt voor het voorspellen van netwerkcongestie. De redenen zijn als volgt:

- Geen mogelijkheid om voorspellende grafieken aangaande de toekomst weer te geven;
- Maakt alleen gebruik van *History values*, wat een te groot beslag legt op systeembronnen;
- Brengt de stabiliteit van de monitoring in gevaar door de behoefte aan systeembronnen bij het berekenen van voorspellingen;
- Vergt een gespecialiseerde statisticus om in te richten, te controleren en te beheren.

5. **In hoeverre zijn de onder deelvraag 3 geselecteerde producten geschikt om meerjarige voorspellingen, met afnemende betrouwbaarheid, aangaande de verwachte groei van het netwerkverkeer te doen ten einde netwerkcongestie te voorkomen?**

Van de onder deelvraag 3 geselecteerde producten zijn in overleg met de opdrachtgever, in verband met de doorlooptijd, de hoogst scorende vijf getest tijdens een *Proof of Concept* (PoC). Het betreft de volgende leveranciers:

- Datameer
- Elastic
- Logi Analytics
- Splunk
- TIBCO Software

Van voornoemde leveranciers heeft Logi Analytics aangegeven niet mee te willen doen met deze fase van het onderzoek. De leveranciers Elastic en Splunk hadden als enige functionaliteit voor het doen van voorspellingen. Bij Elastic zijn de opties van voornoemde functionaliteit beperkt en het implementeren van Elastic vraagt veel tijd en specialistische kennis, wat vervanging van de specialist door andere leden van team Monitoring lastig maakt. Uiteindelijk is wat betreft mogelijkheden en doorlooptijd aangaande implementatie en beheer Splunk als enige geschikt bevonden. Deze geschiktheid is geverifieerd aan de hand van de in deze scriptie beschreven simulatie.

Tot slot, het antwoord van de afstudeerder op de onderzoeksvraag:

> *In hoeverre en op welke manier is het mogelijk om met over de jaren afnemende betrouwbaarheid te voorspellen wanneer netwerkcongestie optreedt in het netwerk van RoutIT teneinde proactief verbeteringen aan te kunnen brengen?*

Gebleken is dat de combinatie van *Business Intelligence* (BI) met de binnen RoutIT beschikbare datasets potentie heeft om over meerdere jaren voorspellingen te doen aangaande het optreden van netwerkcongestie. Het is met inachtneming van de beperkingen als beschreven in deze scriptie op dit moment al mogelijk om voorspellingen te doen met behulp van BI-oplossingen. De verwachting is dat na het beschikbaar komen van meer data de voorspellingen nauwkeuriger worden en over langere perioden gedaan kunnen worden conform de hoofdvraag. Op dit moment kan slechts enkele dagen vooruit gekeken worden.

De aan RoutIT geadviseerde BI-oplossing betreft Splunk *Enterprise*. Splunk scoort het hoogst bij de tijdens dit onderzoek uitgevoerde MoSCoW-analyse. Tevens is het tijdens de uitgevoerde simulatie zoals in het desbetreffende hoofdstuk beschreven, in staat gebleken om de voor dit onderzoek geschikte en beschikbare datasets om te zetten in een voorspelling aangaande netwerkcongestie.

15 Aanbevelingen aan RoutIT als vervolg op dit onderzoek

Om de implementatie van Splunk soepel te laten verlopen is het aanbevolen om degenen die de implementatie en het beheer van Splunk moeten gaan uitvoeren, de bij de taken behorende trainingen op het gebied van Splunk te laten volgen. Op die manier kan sneller een kwalitatief beter model geschreven worden.

Voor het doen van lange termijn voorspellingen is het noodzakelijk dat de afdeling *Software Development* de data uit IRMA en RTK controleert en de aanwezige fouten in deze datasets herstelt. Tevens is het van belang dat nieuwe foutieve invoer wordt voorkomen door het aanpassen van de IRMA- en RTK-software. Door het gebruik van de data uit IRMA en RTK is het mogelijk om verder en preciezer vooruit te kijken. Op dit moment beveelt de afstudeerder aan om bij het doen van investeringen een marge van +10% van de maximale lijncapaciteit behorende bij de desbetreffende voorspelling te nemen.

Tijdens dit onderzoek is alleen gekeken naar netwerkcongestie. Splunk kan daarentegen voor veel meer worden ingezet. Een logisch vervolg is om op basis van Splunk inzicht te krijgen in de huidige staat van de digitale veiligheid betreffende RoutIT apparatuur ten einde het management en afdelingen te kunnen adviseren. Daarnaast is het met Splunk mogelijk om voorspellingen te doen over in gebruik zijnde licenties door team *Voice*.

16 Evaluatie procesgang en conformering aan de kwaliteitscriteria

In dit hoofdstuk wordt de procesgang geëvalueerd in combinatie met de conformering aan de kwaliteitscriteria.

16.1 Evaluatie procesgang

Als maatstaf voor de procesgang wil de afstudeerder graag het opvolgen van de planning gebruiken. Het Plan van Aanpak (PvA) was ruim voor de *deadline* opgeleverd. Dit PvA is besproken met en op inhoud goed bevonden door Ander de Keijzer van de HU en Winfred Hofman van RoutIT tijdens het bezoek van Ander aan RoutIT op 06-02-2018. Het PvA had op 16-02-2018 definitief moeten zijn. De volgende fase van het onderzoek was het verzamelen van informatie. Deze fase verliep dermate snel dat de afstudeerder ruim twee weken voor liep op planning. In overleg met de opdrachtgever is vervolgens afgesproken om deze extra tijd te gebruiken voor een markt verkennende PoC. Deze had meer tijd nodig dan gepland waardoor er minder tijd was voor andere *deadlines*. In overleg met de opdrachtgever is besloten om die tijd te nemen. Toen de uiteindelijke PoC met de geadviseerde oplossing ging starten, bleek dat de installatie en het verkennen van de geadviseerde oplossing overgeslagen kon worden waardoor de volgende *deadlines* allemaal gehaald zijn.

De communicatie met de opdrachtgever is goed verlopen. De afstudeerder heeft regelmatig documenten overlegd en de opdrachtgever uitgenodigd voor voortgangsgesprekken. Daarnaast is de opdrachtgever ook tijdens de werkzaamheden voorzien van informatie. Tijdens dit onderzoek heeft de afstudeerder veelvuldig contact gehad met collega’s om aan informatie en benodigdheden als VM’s te komen of om problemen op te lossen.

Het halen van de planning was soms een uitdaging. Dit is goed gelukt door tijdige afstemming tussen opdrachtgever en afstudeerder op aangeven van de afstudeerder. De planning en de bijbehorende werkzaamheden zijn tijdig aangepast. Waardoor het onderzoeksproces ruim binnen de tijd is afgerond.

16.2 Conformering aan de kwaliteitscriteria

Het kwaliteitscriterium als beschreven in het PvA betreft de tevredenheid van de opdrachtgever en Ander de Keijzer over de op te leveren producten. Zowel de opdrachtgever als Ander de Keijzer hebben vragen gesteld en opmerkingen geplaatst bij de verschillende producten. Deze vragen en opmerkingen zijn naar tevredenheid door de afstudeerder verwerkt waarmee aan de kwaliteitscriteria is voldaan. De opdrachtgever is dermate tevreden dat de opdrachtgever de geadviseerde BI-oplossing nu ook daadwerkelijk aan aanschaft en de afstudeerder een contract heeft aangeboden. Tevens is de afstudeerder gevraagd om na afronding van zijn studie de geadviseerde oplossing te implementeren en uit te bouwen voor RoutIT.

17 Ethische afwegingen

In dit hoofdstuk worden de gedachten van de afstudeerder en de bijbehorende afwegingen op ethisch vlak beschreven. Dit onderzoek is de aanzet voor proactieve monitoring van het netwerk in al zijn facetten. De gebruikte analyse software kan, zoals in de onderstaande alinea's beschreven, in principe gebruikt worden voor inbreuken op de persoonlijke levenssfeer. Echter zoals ook beschreven is dit door de benodigde investeringen en koers van RoutIT niet waarschijnlijk.

Een ethische afweging tijdens dit onderzoek is of het gedrag van een eindklant middels dit onderzoek in kaart gebracht en vermarkt kan worden. Strikt genomen is dat mogelijk door het koppelen van *Domain Name System* (DNS) server logs met de klantgegevens vanuit IRMA waar tevens de IP-adressen in worden bijgehouden. Door RoutIT worden geen DNS-logboeken bijgehouden voor andere doeleinden dan *troubleshooting*. Het op deze manier gaan inzetten van klantdata vraagt een investering in tijd en geld in combinatie met een fundamentele koerswijziging van RoutIT als bedrijf waardoor de afstudeerder deze mogelijkheid uitsluit. Tevens zullen klanten gezien de prijsstelling van de RoutIT diensten een dergelijke inbreuk op de persoonlijke levenssfeer niet accepteren waardoor RoutIT dan zijn bestaansrecht verliest.

Bovenstaand voorbeeld betreft het verzamelen van veel meer data dan nu door RoutIT gedaan wordt, maar is met de uitkomsten van dit onderzoek zeker mogelijk. Echter, daarvoor is het noodzakelijk een volledig aparte afdeling op te zetten en een koerswijziging door te voeren. Aangezien RoutIT de nadruk legt op zoveel mogelijk automatiseren met zo min mogelijk mensen is het onwaarschijnlijk dat RoutIT voornoemde investeringen zal willen doen ook in combinatie met het risico op vertrekkende klanten.

Bibliografie

Anderson, T. W. (1994). *The Statistical Analysis of Time Series*. Retrieved from https://books.google.nl/books?hl=en&lr=&id=rCOzXIC8ZLkC&oi=fnd&pg=PR11&dq=about+time+series&ots=It04mlWzLu&sig=xmEX7iaIWyQQVd9R2vR3l2cYD0c&redir_esc=y#v=onepage&q=about%20time%20series&f=false

BBC. (2018, September 28). MH17 Ukraine plane crash: What we know. Retrieved from http://www.bbc.com/news/world-europe-28357880

Cisco. (2018a, January 19). How Voice over IP (VoIP) Works. Retrieved January 19, 2018, from https://www.cisco.com/c/en/us/solutions/small-business/resource-center/serve-customers-better/voip-how.html

Cisco. (2018b, January 25). Cisco ASR 9000 Series Aggregation Services Routers Simple, Scalable, Secure. Retrieved January 25, 2018, from https://www.cisco.com/c/en/us/products/routers/asr-9000-series-aggregation-services-routers/index.html#~stickynav=4

Faragher, R. (2018). Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation. *IEEE SIGNAL PROCESSING MAGAZINE*, 128–132.

Griffith, E. (2018, February 5). The Ultimate Cord Cutter's Guide. Retrieved March 7, 2018, from https://www.pcmag.com/article2/0,2817,2478213,00.asp

ISO. (2018, January 16). ISO27001. Retrieved January 16, 2018, from https://www.iso.org/isoiec-27001-information-security.html

J., D. (2016, August 16). What is the size of the Windows 10 Anniversary update? Retrieved February 6, 2018, from https://answers.microsoft.com/en-us/windows/forum/windows_10-other_settings/what-is-the-size-of-the-windows-10-anniversary/1ae1dfd0-0b40-476d-a6b6-2434494c7e50

Knippr. (2018, March 7). Knippr. Retrieved March 7, 2018, from https://www.knippr.nl/

McCabe, J. D. (2007). *Network Analysisi, Architecture, and Design* (3rd ed.). Burlington: Morgan Kaufmann Publishers.

McClure, T. (2018, April 4). How Kalman Filters Work, Part 1. Retrieved April 4, 2018, from http://www.anuncommonlab.com/articles/how-kalman-filters-work/

McCrea, N. (2018, April 3). An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples. Retrieved April 3, 2018, from https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1984). *Machine Learning An Artificial Intelligence Approach*. Springer-Verlag.

Nagle, J. (1984, June 1). Congestion Control in IP/TCP Internetworks (RFC896). IETF. Retrieved from https://tools.ietf.org/html/rfc896

NCSS Statistical Software. (2018). ARIMA (Box-Jenkins).

Pariser, E. (2011). *The Filter Bubble: What The Internet Is hiding From You*. Viking. Retrieved from https://books.google.nl/books?hl=en&lr=&id=-FWO0puw3nYC&oi=fnd&pg=PT3&dq=filter+bubble&ots=g4HtIvvXN0&sig=F7uX563PCacylQhIYfJ-zyeIcqI&redir_esc=y#v=onepage&q=filter%20bubble&f=false

Patil, P. S., Rao, S., & Patil, S. B. (2011). Optimization of Data Warehousing System: Simplification in Reporting and Analysis. *IJCA Journal*, (9), 33–37.

Rekhter, Y., Li, T., & Hares, S. (2006, January 1). A Border Gateway Protocol 4 (BGP-4) (RFC4271). Retrieved from https://tools.ietf.org/html/rfc4271

RoutIT. (2018, January 16). RoutIT. Retrieved January 16, 2018, from https://routit.nl/

RoutIT. (2018, May 25). Over RoutIT. Retrieved May 25, 2018, from https://routit.nl/over-routit/

Sallam, R. L., Howson, C., Idoine, C. J., Oestreich, T. W., Richardson, J. L., & Tapadinhas, J. (2017). *Magic Quadrant for Business Intelligence and Analytics Platforms* (Magic Quadrant No. G00301340) (p. 82). Gartner.

Smola, A., & Vishwanathan, S. V. N. (2008). *Introduction to Machine Learning*. Retrieved from http://alex.smola.org/drafts/thebook.pdf

Splunk. (2018, April 9). Predict. Retrieved April 9, 2018, from https://docs.splunk.com/Documentation/Splunk/7.0.3/SearchReference/Predict

Statcounter. (2018a, February 5). Desktop Windows Version Market Share Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from http://gs.statcounter.com/windows-version-market-share/desktop/netherlands/#monthly-201701-201801-bar

Statcounter. (2018b, February 5). Mobile & Tablet Operating System Market Share Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from http://gs.statcounter.com/os-market-share/mobile-tablet/netherlands/#monthly-201701-201801-bar

Statcounter. (2018c, February 5). Social Media Stats Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from http://gs.statcounter.com/social-media-stats/all/netherlands/#monthly-201701-201801-bar

Statista. (2018). Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions). Retrieved February 5, 2018, from https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/

Stodder, D. (2015). *Visual Analytics for Making Smarter Decisions Faster*.

Tableau. (2018, March 7). How Forecasting Works in Tableau. Retrieved March 4, 2018, from http://onlinehelp.tableau.com/current/pro/desktop/en-us/forecast_how_it_works.html

The Pennsylvania State University. (2018, April 3). Seasonal ARIMA models. Retrieved April 3, 2018, from https://onlinecourses.science.psu.edu/stat510/node/67

van Miltenburg, O. (2018, March 21). T-Mobile stopt met tv-dienst Knippr wegens te weinig klanten. Retrieved from https://tweakers.net/nieuws/136509/t-mobile-stopt-met-tv-dienst-knippr-wegens-te-weinig-klanten.html

Zabbix. (2015). Forecasting trigger functions (p. 10). Zabbix. Retrieved from http://zabbix.org/mw/images/1/18/Prediction_docs.pdf

Zabbix. (2018a, January 16). Zabbix. Retrieved January 16, 2018, from https://www.zabbix.com/

Zabbix. (2018b, January 19). Configuring a host. Retrieved January 19, 2018, from https://www.zabbix.com/documentation/3.4/manual/config/hosts/host

Zabbix. (2018c, January 30). 3 History and trends. Retrieved January 30, 2018, from https://www.zabbix.com/documentation/3.4/manual/config/items/history_and_trends

Zabbix. (2018d, January 31). 2. Definitions. Retrieved January 31, 2018, from https://www.zabbix.com/documentation/3.4/manual/definitions

Ziggo. (2018, March 7). Ziggo Go. Retrieved March 7, 2018, from https://www.ziggo.nl/televisie/online-tv-kijken/

Bijlage A – Plan van Aanpak (PvA)

Automatiseren voorspelling netwerkcongestie in het core-netwerk van RoutIT

Naam afstudeerder: Bart Dekker
Studentnummer:
Naam bedrijf: RoutIT
Datum: 12-02-2018
Status: Definitief
Versie: 1.0

Versiebeheer

Versie	Datum	Wijzigingen	Status
0.1	15-01-2018	Eerste opzet PvA	Gereed
0.2	19-01-2018	Toevoegen opdracht en theoretisch kader	Gereed
0.3	20-01-2018	Toevoegen globale aanpak	Gereed
0.4	23-01-2018	Verwerken opmerkingen RoutIT	Gereed
0.5	29-01-2018	Eerste concept	Gereed
1.0	12-02-2018	Feedback Ander verwerkt	Gereed

Distributielijst

Functie	Naam	Organisatie	Datum
Docentbegeleider	Ander de Keijzer	Hogeschool Utrecht	
Bedrijfsbegeleider	Winfred Hofman	RoutIT	
Bedrijfsbegeleider	Erica Schonewille	RoutIT	
Team Monitoring	Kristiaan Vos	RoutIT	

Goedkeuringen

Naam	Functie	Handtekening	Datum
Ander de Keijzer	Docentbegeleider		
Winfred Hofman	Bedrijfsbegeleider		
Erica Schonewille	Bedrijfsbegeleider		
Kristiaan Vos	Team Monitoring		

Auteur

Naam	Initialen	Student nummer	Telefoon	E-mail
Bart Dekker	B.J.M.			

Voorwoord

In dit document treft u mijn Plan van Aanpak (PvA) ten behoeve van mijn afstuderen bij RoutIT. Het afstuderen is het laatste vak dat een student aan de Hogeschool Utrecht moet behalen om in het bezit te komen van zijn diploma. RoutIT is een *Internet Service Provider* (ISP) gericht op de zakelijke markt.

Op dit moment is de analyse van knelpunten in het netwerk een maandelijks uitgevoerd handmatig karwei. Ten einde dit proces te automatiseren en het voorspellen wanneer knelpunten gaan optreden over de jaren heeft RoutIT een afstudeeropdracht uitgeschreven.

Dit PvA beschrijft een plan, dat wil zeggen dat er in het verloop van het project veranderingen of wijzigingen zich kunnen voordoen aangaande de beschreven plannen. Voor het verloop van het project zal het PvA wel aangehouden worden.

Dit PvA heeft als doel om een beeld te schetsen van de activiteiten waaruit het afstudeertraject zal bestaan en de afspraken die gemaakt worden met de begeleider vanuit de Hogeschool Utrecht, Ander de Keijzer en met RoutIT in de persoon van Winfred Hofman en Erica Schonewille.

Ede, 15 januari 2018.

Inhoudsopgave

Voorwoord 49
1 Inleiding 7
2 Organisatorische context en de positie van de afstudeerder 8
2.1 RoutIT als organisatie 8
2.2 Producten en RoutIT partners 8
2.3 Het RoutIT netwerk 8
2.4 De Network Operations Center (NOC) afdeling 9
3 Formulering van de opdracht 10
3.1 Aanleiding voor de opdracht 10
3.2 Aanleiding en uitleg van de deelvragen 10
3.3 De doelstelling voor dit onderzoek 11
3.4 Afbakening en randvoorwaarden 55
4 Onderzoeksvraag en deelvragen 55
5 Het theoretisch kader 12
5.1 Netwerkcongestie 12
5.2 *Voice over Internet Protocol* (VoIP) 56
5.3 Zabbix en de Zabbix functie "*forecasting*" 15
5.4 *Business intelligence en data warehouse* 13
6 Globale aanpak 15
6.1 Uitleg MoSCoW analyse 57
6.2 Enquêtering van collega's en selecteren mogelijke oplossing 15
6.3 Geplande aanpak aangaande te lezen literatuur ..... **Error! Bookmark not defined.**
7 Producten en kwaliteitscriteria 57
8 Projectorganisatie 58
9 Algemene planning 59
10 Risicoanalyse 60
Bibliografie 60

1 Inleiding

Om voort te kunnen blijven bestaan en te kunnen groeien is het voor een dienstverlenend bedrijf belangrijk om de kwaliteit van de diensten goed te borgen en indien nodig bij te sturen. Voor RoutIT (RoutIT, 2018) als *Internet Service Provider* (ISP) is het noodzakelijk dat de kwaliteit van het netwerk van hoog niveau is. Het is een gegeven dat de hoeveelheid door klanten ingenomen netwerkcapaciteit zal groeien en bij ongewijzigd beleid tot netwerkcongestie zal leiden. Op basis van historische data uit het gebruikte monitoringpakket wil RoutIT kunnen voorspellen wanneer en waar de netwerkcongestie op gaat treden. Het doen van voorspellingen is op dit moment een handmatig proces dat maandelijks uitgevoerd wordt en waar veel tijd in gaat zitten.

RoutIT maakt voor zijn netwerkapparatuur uitsluitend gebruik van Cisco, daarmee is de keuze voor het monitoringpakket van Cisco zelf een logische. Echter, aangezien RoutIT naast netwerkapparatuur van Cisco ook servers op basis van Windows, Linux en vSphere, waarbij voornoemde *Operating Systems* (OS) niet goed integreren met het monitoringpakket van Cisco. Daarom is gekozen voor Zabbix als monitoringpakket.

RoutIT heeft aan de afstudeerder Bart Dekker van de opleiding *System and Network Engineering* (SNE) aan de Hogeschool Utrecht (HU) gevraagd om voornoemde problemen te onderzoeken en met een oplossing te komen. De oplossing moet gebruik maken van de monitoringdata uit Zabbix.

2 Organisatorische context en de positie van de afstudeerder

In dit hoofdstuk wordt RoutIT als organisatie, de afdeling en de positie van de afstudeerder op deze afdeling beschreven.

2.1 RoutIT als organisatie

RoutIT is een ISP gericht op de zakelijke markt. De kern van de dienstverlening bestaat uit het bieden van een internetverbinding en telefonie. De klanten van RoutIT stellen hoge eisen aan de dienstverlening zowel wat betreft de snelheid en de beschikbaarheid als de vertrouwelijkheid van de dienstverlening, daarom is RoutIT ISO27001 (ISO, 2018) gecertificeerd. Onderdeel van de veiligheidsprocedures is onder andere toegangscontrole op apparatuur en in het gebouw en het bijhouden van welke medewerker toegang heeft gehad tot welk wachtwoord.

Kenmerk van RoutIT is dat het een dynamisch bedrijf is. Het *Network Operations Center* (NOC) waar de afstudeerder werkt heeft als doel om op technisch vlak nieuwe diensten te ontwikkelen, maar ook om bestaande diensten en het netwerk te onderhouden en te beheren. Door RoutIT ontwikkelde diensten worden op commercieel niveau continue geëvalueerd en indien nodig worden de diensten verbeterd of verwijderd uit het dienstenportofolio. Ideeën voor nieuwe diensten worden aangedragen door een andere afdeling, die klantcontact heeft, deze afdeling doet ook de commerciële evaluatie. Het NOC komt daarnaast ook in actie als de derdelijns helpdesk van RoutIT een probleem van een klant niet kan oplossen.

2.2 Producten en RoutIT partners

Alle RoutIT producten zijn *white label* producten. Deze producten verkoopt RoutIT aan IT-bedrijven, zogenaamde partners, die daar hun eigen naam opplakken. Zo is een internetverbinding door een eindgebruiker alleen te bestellen via een RoutIT partner. Deze partner ondersteund de eindgebruiker, maar de eindgebruiker maakt onderwater gebruik van de RoutIT infrastructuur. RoutIT doet hierbij niet aan koppelverkoop van diensten. Het is prima mogelijk om telefonie van RoutIT te betrekken en internet via een andere provider. Dit

is echter niet aan te raden omdat de kwaliteit van de telefonie dan niet gegarandeerd kan worden.

2.3 Het RoutIT netwerk

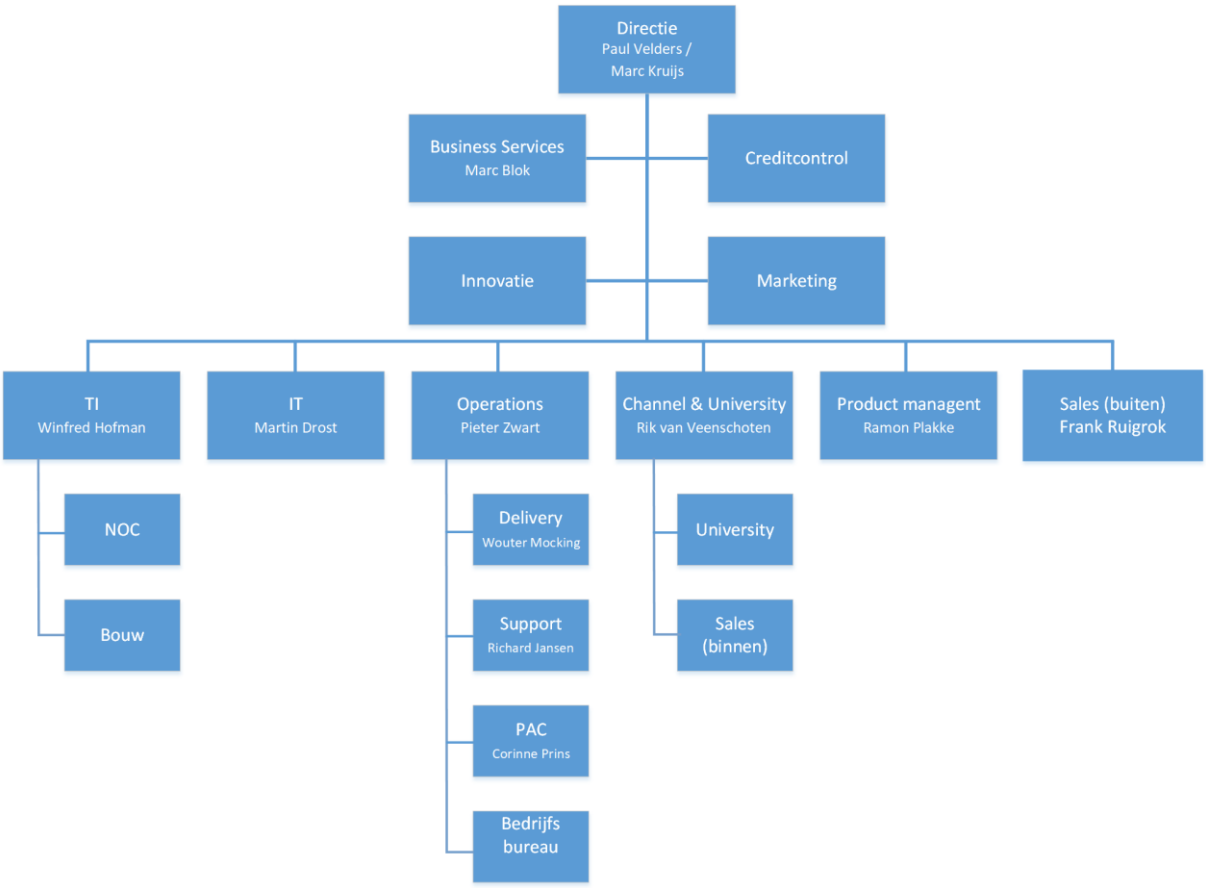
De apparatuur van RoutIT is geplaatst bij datacentra verspreid over het land. RoutIT huurt in deze datacentra al naar behoefte een of meer serverracks. Apparatuur geplaatst in datacentra maakt normaliter gebruik van de glasvezelkabels die het datacentrum verbinden met de buitenwereld. RoutIT huurt echter van een derde leverancier eigen glasvezelkabels tussen de datacentra.

RoutIT maakt uitsluitend gebruik van Ciscoapparatuur. Bij het nieuw aanschaffen van netwerkapparatuur betreft het de Cisco ASR1000 en ASR9000 series *router*. Vanuit de historie zijn er ook nog andere typen routers en switches van Cisco in gebruik. De Cisco ASR-producten betreffen een combinatie tussen *switching* en *routing* voor de ISP-markt, maar worden door RoutIT *router* genoemd. Een ASR is een modulair apparaat dat uitbreidbaar is. RoutIT heeft op ASR1000 *routers* 6 poorten van 10Gbps en op de ASR9000 meer dan 50 poorten van ofwel 10Gbps of 100Gbps in gebruik. Voor de netwerkverbindingen in het netwerk wordt uitsluitend van glasvezel gebruik gemaakt. De glasvezelkabels worden met meerdere kleuren belicht om zodoende de capaciteit van de glasvezelkabel te verhogen. Een glasvezelkabel kan tussen de 50m en 40km overbruggen. De ASR9000 met de minste functionaliteit beschikt al over een capaciteit van maximaal 44Gbps. De variant met de meeste functionaliteit beschikt over een capaciteit van maximaal 160Tbps. In het geval van de ASR9000 modellen met meer functionaliteit koopt een klant een *router* zonder poorten. De poorten dienen als *line cards* los aangeschaft te worden. Een *line card* is te vergelijken met een enorme *Network Interface Card* (NIC). (Cisco, 2018b)

Een ISP kan op een directe manier via het *Border Gateway Protocol* (BGP) informatie uitwisselen met andere ISP's (Rekhter et al., 2006). Dit wordt *peering* genoemd. Op die manier is het voor klanten van RoutIT ook mogelijk om bijvoorbeeld websites te bezoeken die hun internetconnectiviteit niet bij RoutIT betrekken. RoutIT maakt via BGP-verbinding met ongeveer 100 *peers*. Een alternatief voor *peering* is *transit*. In het geval van *transit* wordt bij internetverkeer voor een bestemming buiten het RoutIT netwerk of de ISP's waar RoutIT mee *peered* het desbetreffende verkeer naar de *transit provider* gestuurd die vervolgens zorgdraagt voor een goede afhandeling. Uit redundantie oogpunt betrekt RoutIT bij drie *transit providers* een *transit* dienst.

2.4 De RoutIT organisatiestructuur

In afbeelding 1 wordt de RoutIT organisatiestructuur weergegeven. RoutIT heeft daarmee een vrij platte organisatie. De afdelingsmanagers vallen direct onder de directie. De afdeling “Operations” is de enige afdeling met meerdere managers. Deze managers hebben naast de operationele taken op die afdeling ook de taak om collega’s aan te sturen. Deze managers zijn dus meewerkend voorwerp op de afdeling. Dat geld ook voor de andere managers op de afdelingen.



Afbeelding 1: Organogram met de RoutIT organisatiestructuur

2.5 De Network Operations Center (NOC) afdeling

De afstudeerder werkt op het NOC aan zijn opdracht. Op het NOC heeft iedereen altijd storingsdienst. Tijdens deze storingsdienst is de *engineer* verantwoordelijk voor zijn eigen werk. Indien de engineer robuust werk aflevert, dan heeft de *engineer* buiten kantoortijden in principe niets te doen. Indien de *engineer* fouten gemaakt heeft, dan is het de taak van de *engineer* deze fouten op te lossen.

Binnen het NOC is het de bedoeling dat mensen elkaar kunnen vervangen. In de praktijk heeft iedereen zijn specialisme. Door de aard van de opdracht is de afstudeerder gekoppeld aan Kristiaan. Hij is verantwoordelijk voor de monitoring. De door de afstudeerder uitgevoerde opdracht moet het werk van Kristiaan vergemakkelijken.

3 Formulering van de opdracht

In dit hoofdstuk wordt de opdracht beschreven samen met de aanleiding voor RoutIT om deze opdracht uit te schrijven.

3.1 Aanleiding voor de opdracht

Voor RoutIT is het een gegeven dat de huidige netwerkcapaciteit in de toekomst ontoereikend is. Opdat klanten geen last ondervinden van netwerkcongestie, wil RoutIT op tijd de capaciteit van het netwerk kunnen uitbreiden of knelpunten verhelpen. Daarom heeft RoutIT een opdracht uitgeschreven om naar een methode te zoeken om op basis van de binnen RoutIT beschikbare datasets, waaronder monitoringdata en geplande opleveringen van verbindingen, voorspellingen kan doen over wanneer de netwerkcapaciteit ontoereikend is.

3.2 Aanleiding en uitleg van de deelvragen

Op dit moment kunnen knelpunten en groeiverwachtingen al handmatig in kaart gebracht worden, maar het formaliseren en automatiseren van dit proces heeft voor RoutIT grote voordelen. Op die manier kan sneller geacteerd worden op organisatorische veranderingen, de klantenaantallen en de opkomst van concurrenten en concepten in de ICT-branche.

RoutIT heeft de beschikking over een verscheidenheid aan datasets, een van die datasets is de monitoringdata uit Zabbix een andere dataset betreft uitstaande orders en opleverdata van door partners aangevraagde diensten, maar er is nog veel meer. Sommige datasets kunnen gebruikt worden voor het voorspellen van de benodigde capaciteit over tijd in het netwerk ten einde netwerkcongestie te voorkomen. Maar welke datasets zijn relevant? En wat voor bewerkingen moeten er op de datasets worden losgelaten voordat de desbetreffende data bruikbaar is om netwerkcongestie te voorspellen? Om deze en gerelateerde vragen te beantwoorden heeft RoutIT een opdracht uitgeschreven die in deze paragraaf besproken wordt.

Het RoutIT netwerk bestaat uit verschillende segmenten met elk hun eigen taken. Deze segmenten interacteren met elkaar, en hebben zodoende invloed op netwerkcongestie bij elkaar. Belangrijk bij het selecteren van de datasets is te kijken welke informatie belangrijk is voor de desbetreffende collega van het NOC die verantwoordelijk is voor het bijbehorende netwerksegment. De netwerksegmenten zijn op dit moment deels geïdentificeerd, maar hier ligt voor de afstudeerder ook nog een taak. Daarmee is het verzamelen van gegevens over het netwerk en de netwerkcongestiefactoren waarschijnlijk het belangrijkste deel van deze opdracht. Misschien komt wel uit dit stuk dat voor dit onderzoek meer informatie moet worden vastgelegd of dat er transformaties moeten plaatsvinden, oftewel welke voorwaarden moeten er gesteld worden aan de huidige omgeving om het door de afstudeerder uitgevoerde onderzoek te volbrengen?

Op dit moment heeft RoutIT een *data warehouse* oplossing van Microsoft, op het vlak van *data warehousing* en *business intelligence* is Microsoft leidend wat betreft innovativiteit en visie (Sallam et al., 2017). Het is dan ook niet aan de orde voor de afstudeerder om een migratie van het huidige *data warehouse* naar een *data warehouse* van een andere fabrikant voor te stellen. Wel is het mogelijk om ten behoeve van de analyse specifieke datasets op een continue basis te exporteren naar een voor de analyse software bruikbare en toegankelijke opslag.

Van Zabbix is bekend dat het aangaande het netwerkverkeer beschikt over een voorspellende functie. Het is mogelijk om delen van de analyse uit te besteden aan Zabbix, of dit ook raadzaam is dient tijdens dit onderzoek door de afstudeerder bekeken te worden. Het onderzoek wordt op voorhand niet beperkt aangaande kosten of leveranciers door de opdrachtgever, wel is het wenselijk om de oplossing dusdanig te selecteren dat de oplossing geïntegreerd kan worden of samenwerken met de bestaande systemen binnen RoutIT.

Wegens de complexiteit van deze opdracht is het niet de bedoeling dat de afstudeerder een implementatie doet. Een *Proof of Concept* (PoC) met als onderdeel een simulatie is daarentegen wel onderdeel van deze opdracht. RoutIT heeft voor de simulatie en de PoC onder andere hardware en een virtueel platform ter beschikking. Wegens de overdrachtelijk- en uitbreidbaarheid van het onderzoek en de PoC zal relevantie kennis en ervaring worden opgenomen in de scriptie en een bijgevoegd PoC ontwerp.

Het betreft hier dus een ontwerp-/adviesopdracht.

3.3 De doelstelling voor dit onderzoek

Het doel van dit onderzoek is om te kijken welke factoren van invloed zijn op netwerkcongestie in (delen) van het RoutIT netwerk en hoe de samenhang daartussen is. Tevens is het van belang om de netwerkcongestie inzichtelijk te maken en voorspellingen over de netwerkcongestie te doen.

3.4 Afbakening en randvoorwaarden

Deze opdracht wordt uitgevoerd door een student van de Hogeschool Utrecht (HU). Indien dit onderzoek goed wordt uitgevoerd kent de HU de student zijn laatste 30 *European Credits* (EC) toe. Na het behalen van deze 30 EC kan de student zijn diploma aanvragen en daarmee de HBO-opleiding SNE afronden. De 30 EC staan voor 840 studielasturen die worden doorgebracht bij RoutIT.

Het is niet de taak van de afstudeerder om de huidige productiefaciliteiten te migreren of een migratie aan te bevelen. Wel is het de bedoeling om te onderzoeken welke producten de huidige productiefaciliteiten in het kader van de hierboven besproken opdracht kunnen aanvullen. Belangrijke eis aan de voor te stellen oplossing is dat de kwaliteit van de monitoring niet onder de voorgestelde oplossing mag leiden.

De volgende randvoorwaarden zijn van toepassing op dit project:

- Beschikbaarheid van een geschikte werkomgeving bij RoutIT;
- Middelen om een PoC en simulatie op te kunnen zetten;
- Kennis van Zabbix;
- Kennis van *data warehousing, business intelligence* en *machine learning* of anders de mogelijkheid deze kennis te vergaren;
- Hulp van collega's bij het analyseren van de datasets samen met hun inhoud;
- Toegang tot de benodigde productie- en testfaciliteiten;
- Goede communicatie binnen de projectorganisatie.

Dit project bevat geen deelopdrachten, tevens is dit project geen onderdeel van een groter project. Wel is het mogelijk dat tijdens dit project door RoutIT een separaat en aanverwant project gestart wordt dat gelijktijdig met dit project wordt uitgevoerd. In dat geval heeft de afstudeerder niets van doen met dat andere project.

4 Onderzoeksvraag en deelvragen

De voor dit onderzoek gedefinieerde onderzoeksvraag luidt als volgt:

In hoeverre en op welke manier is het mogelijk om met over de jaren afnemende betrouwbaarheid te voorspellen wanneer netwerkcongestie optreedt in het netwerk van RoutIT teneinde proactief verbeteringen aan te kunnen brengen?

Ter beantwoording van de hierboven genoemde onderzoeksvraag zijn de volgende deelvragen gedefinieerd:

1. Welke netwerksegmenten zijn betrokken bij netwerkcongestie binnen de RoutIT core-dienstverlening en hoe beïnvloeden die elkaar?
2. Welke voor dit onderzoek relevante datasets heeft RoutIT tot zijn beschikking en in hoeverre zijn deze bruikbaar voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie in de core-dienstverlening te voorkomen?
3. Welke mogelijkheden aangaande *business intelligence* zijn beschikbaar en kunnen geïntegreerd worden met de bestaande RoutIT productiefaciliteiten aangaande het

doen van voorspellingen betreffende het netwerkverkeer ten einde netwerkcongestie in het netwerk te voorkomen?

4. In hoeverre is de Zabbix functie "*forecasting*" bruikbaar voor toepassing door RoutIT voor het doen van voorspellingen aangaande de groei van het netwerkverkeer ten einde netwerkcongestie te voorkomen?
5. In hoeverre zijn de onder deelvraag 3 geselecteerde producten geschikt om meerjaren voorspellingen, met afnemende betrouwbaarheid, aangaande de verwachte groei van het netwerkverkeer te doen ten einde netwerkcongestie te voorkomen?

5 Het theoretisch kader

In dit hoofdstuk worden termen en technieken besproken die essentieel zijn voor het begrijpen van de onderzoeksvraag en de deelvragen alsmede het begrijpen van dit onderzoek.

5.1 Netwerkcongestie

Netwerkcongestie betreft een ophoping van netwerkpakketten bij de overgang naar een ander netwerk(segment) waarbij bandbreedte verschillen optreden (Nagle, 1984). Dit kan als gevolg hebben dat bijvoorbeeld het laden van webpagina's niet lukt of dat de kwaliteit van telefonie over internet (VoIP, zie hieronder) ondermaats is. Aangezien de overdracht van webpagina's en telefonie tot de *core business* van RoutIT behoren is netwerkcongestie voor RoutIT niet acceptabel.

5.2 Voice over Internet Protocol (VoIP)

Voice over Internet Protocol (VoIP) is een manier om te telefoneren via een bestaand datanetwerk (Cisco, 2018a). Het voordeel aan VoIP tegenover *Plain Old Telephony System* (POTS) oftewel traditionele telefonie is dat er geen aparte bekabeling nodig voor het aansluiten van een telefoon. De VoIP technologie maakt gebruik van hetzelfde datanetwerk als een computer. Daarmee kunnen bedrijven hun operationele kosten verminderen. Het is bij VoIP mogelijk om te beeldbellen en om met meerdere mensen een gesprek te voeren, zogenaamde *video conferencing* (Cisco, 2018a).

5.3 Zabbix en de Zabbix functie *"forecasting"*

Zabbix is een *Open Source Software* (OSS) monitoringpakket en tevens niet gebonden aan één type *Operating System* (OS) of netwerkapparaat (Zabbix, 2018a). Voor het monitoren van servers en netwerkapparatuur maakt Zabbix gebruik van onder andere *Simple Network Management* Protocol (SNMP) of de Zabbix *Agent* (Zabbix, 2018b). Binnen RoutIT wordt Zabbix gebruikt om het netwerk en andere productiefaciliteiten te monitoren. Zabbix bevat onder andere gegevens over het actuele en historische bandbreedte verbruik door eindgebruikers. Hieronder vallen, misschien ten overvloede, ook telefoniegesprekken. Voor het doen van voorspellingen bevat Zabbix de functie *forecasting*, deze functie maakt gebruik van statistische formules (Zabbix, 2015). *Forecasting* binnen Zabbix heeft als nadeel dat voor deze functionaliteit geen gebruik gemaakt kan worden van Zabbix Trends data, maar dat de volledige historische data benodigd is. Tevens is Zabbix Trends *resource* intensief bij het doen van voorspellingen op de lange termijn zoals gewenst door RoutIT.

5.4 Business intelligence en data warehouse

Business Intelligence (BI) is het analyseren en representeren van data op een dusdanige manier dat er voorspellingen kunnen worden gedaan over de toekomst (Stodder, 2015). Vroeger was BI sterk ICT-gerelateerd en was de kennis van een programmeer- of scriptingtaal nodig om met BI te kunnen werken. Op dit moment is de ontwikkeling gaande om BI ook voor minder technisch aangelegde gebruikers toegankelijk te maken doormiddel

van gebruiksvriendelijke visuele software. *business intelligence* betreft alleen het analyse stuk, de data wordt opgeslagen in een *data warehouse*. Een *data warehouse* is een vorm van opslag om data vanuit verschillende bronnen. Een belangrijke functionaliteit van een *data warehouse* is om de aangeleverde data op te schonen, te catalogiseren en om te zetten naar een uniform formaat (Patil et al., 2011). Het *data warehouse* biedt vervolgens de data in een uniform formaat aan ten behoeve van toepassingen als, maar niet uitsluitend, *business intelligence*.

6 Globale aanpak

In dit stuk zal uitgelegd worden welke methode gebruikt wordt in dit project. De theorie die gebruikt zal worden is MoSCoW en zal ook in een volgend hoofdstuk nader uitgelegd worden. De literatuur die gebruikt zal worden in dit project zal ook aanbod komen.

6.1 Uitleg MoSCoW analyse

De MoSCoW-methode is een techniek om prioriteiten te stellen binnen een onderzoek. De eisen aan het eindresultaat worden hiermee ingedeeld op basis van prioriteiten. Deze prioriteiten zijn: *Must haves*, *Should haves*, *Could haves* en *Won't haves* (ook wel *Would haves* genoemd) (McCabe, 2007).

M – *Must haves*: Deze eisen dienen in het eindresultaat te zitten anders is het eindresultaat onbruikbaar.
S – *Should haves*: Deze eisen zijn zeer gewenst maar zonder deze eisen is het eindresultaat ook bruikbaar.
C – *Could haves*: Deze eisen komen aanbod als er genoeg tijd over is.
W – *Won't haves*: Deze eisen zullen niet tijdens dit onderzoek aan bod komen, maar kunnen mogelijk in de toekomst interessant zijn voor een vervolgonderzoek.

6.2 Enquêtering van collega's en selecteren mogelijke oplossing

Voor het verzamelen van informatie over het RoutIT netwerk en de optredende netwerkcongestie zal de afstudeerder enquêtes afnemen van de collega's van het NOC. Het betreft daarbij korte vragenlijsten om, op advies van de opdrachtgever, te voorkomen dat de afstudeerder te veel bij dit onderzoek gaat betrekken.

Voortvloeiend uit de gehouden enquêtes zal gekeken worden welke voorwaarden aan de oplossing van het netwerkcongestie probleem verbonden zijn en hoe dit te vertalen naar een concrete oplossing. De manier van selecteren hangt af van de op afstand beschikbare informatie over de verschillende oplossingen. Voor het prioriteren van de eisen zal een MoSCoW analyse gebruikt worden. De geselecteerde oplossing is te allen tijde een voordracht die nog door RoutIT geaccordeerd moet worden.

6.3 Geplande aanpak aangaande te lezen literatuur

Tijdens dit onderzoek zal zoveel mogelijk gebruik gemaakt worden van gezaghebbende literatuur. De literatuur die voornamelijk gebruikt zal worden zijn artikelen die gevonden zijn door middel van Google Scholar, databanken welke beschikbaar gesteld zijn door de Hogeschool Utrecht, en documentatie van leveranciers. Bij het gebruik van een generieke zoekmachine wordt de bron beoordeeld op zijn gezaghebbendheid voor opname in de scriptie.

7 Producten en kwaliteitscriteria

Aan het einde van het project zal er door de afstudeerder de volgende producten opgeleverd worden:

- Dit Plan van Aanpak;
- Een inventarisatie van het netwerk voor zover noodzakelijk voor dit onderzoek;
- Een eisenpakket samen met een prioritering aan de hand van een MoSCoW analyse;
- Scriptie met een beschrijving van het onderzoek en een onderbouwing van de oplossing;
- Een ontwerp van de PoC;
- Een PoC getest aan de hand van een simulatie;
- Een testplan ten behoeve van de PoC.

Het opleveren van de producten geschied in nauw overleg met de opdrachtgever. Indien de opdrachtgever het eens is met de kwaliteit van de producten, worden de producten voorgelegd aan Ander de Keijzer van de Hogeschool Utrecht. Indien zowel Ander de Keijzer als RoutIT akkoord zijn met de kwaliteit van het product, is het product af. In de planning zijn verschillende beoordelingsmomenten in de vorm van *milestones* opgenomen.

Ten behoeve van de PoC zal naast het ontwerp ook een testplan geschreven worden om zodoende de kwaliteit van de PoC te waarborgen en reproduceerbaar te maken.

8 Projectorganisatie

In tabel 1 zijn de bereikbaarheidsgegevens van de projectorganisatie weergegeven in combinatie met de communicatieafspraken. Verder staat ook de naam van de docentbegeleider namens de HU vermeld.

Bedrijfsbegeleider #1	Erica Schonewille		
Bedrijfsbegeleider #2	Winfred Hofman		
Team Monitoring	Kristiaan Vos		
Docentbegeleider	Ander de Keijzer		
Afstudeerder	Bart Dekker		
Bereikbaarheid van de betrokkenen:			
Naam	**Studentnr.**	**Tel. Nr.**	**E-mailadres**
Erica Schonewille	N.v.t.		
Winfred Hofman	N.v.t.		
Kristiaan Vos	N.v.t.		
Ander de Keijzer	N.v.t.		
Bart Dekker			
Bedrijfsgegevens RoutIT			
Adres	**Postcode**	**Plaats**	**Website**
Maxwellstraat 51	6716BX	Ede	https://routit.nl/
Communicatieafspraken • Alle documenten worden in de met de projectorganisatie gedeelde map op Box geplaatst; • Communicatie met de opdrachtgever vindt plaats tijdens vergaderingen en via E-mail of door even langs te lopen; • Communicatie met Ander de Keijzer (HU) vindt plaats via E-mail, WhatsApp, telefonisch en tijdens vergaderingen.			

- De afstudeerder controleert dagelijks zijn RoutIT- en HU-mail en WhatsApp;
- Voortgangsbewaking: Middels planning.

Tabel 1: De projectorganisatie.

9 Algemene planning

Voor het verloop van het project zal de planning uit tabel 2 gehanteerd worden.

Datum	Type	Omschrijving	Uren (geschat)	Inleveren bij
15-01-2018	Info	Aanvang afstuderen	N.v.t.	N.v.t.
26-01-2018	Doc	Concept PvA ter goedkeuring aanbieden	64	RoutIT
29-01-2018	Doc	Concept PvA inleveren	8	Ander de Keijzer
02-02-2018	*Milestone*	Netwerksegmenten inventarisatie	32	RoutIT
16-02-2018	Doc	PvA definitief	8	Ander de Keijzer
16-02-2018	*Milestone*	Datasets verzamelt en beoordeelt	64	RoutIT
02-03-2018	*Milestone*	Eisenpakket en prioritering	64	RoutIT
07-03-2018	*Milestone*	Voordracht BI-pakket	24	RoutIT
13-03-2018	Actie	PoC opzetten	24	N.v.t.
21-03-2018	Actie	Simulatie afronden	40	N.v.t.
04-04-2018	Actie	PoC afronden	64	N.v.t.
17-04-2018	Doc	Eerste concept scriptie	64	Ander de Keijzer
08-05-2018	Doc	Tweede concept scriptie	96	Ander de Keijzer
29-05-2018	Doc	Inleveren definitieve scriptie	96	Ander de Keijzer
06-06-2018	Doc	Mailen bedrijfsbeoordeling	N.v.t.	1e examinator & Ander de Keijzer
11-06-2018 t/m 22-06-2018	Info	Periode waarin de afstudeerzitting plaats heeft.	N.v.t.	N.v.t.

Tabel 2: Algemene planning

In de planning worde de volgende typen onderscheiden:

Info: Informatieve vermelding.
Doc: Stap aangaande de documentatie t.b.v. de HU.
Actie: Uit te voeren actie t.b.v. een *milestone* of tussenstap.
***Milestone*:** Beslispunt in het project. Aan een *milestone* hangt altijd een *go / no go* moment.

10 Risicoanalyse

Tijdens het project zijn er een aantal risico factoren waar de afstudeerder rekening mee moet houden. Met deze factoren moet rekening gehouden worden om te voorkomen dat het project op een dood spoor terechtkomt.

Reeds geïdentificeerde risico's zijn afgebeeld in tabel 3 hieronder.

Risico	**Kans**	**Impact**	**Prioriteit**	**Risico beperkende maatregel**
Ziekte afstudeerder	10%	Project loopt uit, niveau 2	Gemiddeld	N.v.t.
Ziekte of afwezigheid projectorganisatie	10%	Project loopt uit, niveau 2	Gemiddeld	Vooruitplannen van afspraken
Trage communicatie met projectorganisatie	20%	Project loopt uit, niveau 3	Hoog	Vooruitplannen van afspraken, telefonisch contact
Onvoldoende middelen t.b.v. PoC of simulatie	5%	Project kan niet onderbouwd worden, niveau 4	Hoog	Goed afstemmen opdrachtgever
Informatie is niet beschikbaar of tegenstrijdig	10%	Hoofddeel onderzoek kan niet afgerond worden, niveau 4	Hoog	Op tijd overleg met opdrachtgever ten behoeve van prioritering

Tabel 3: Geïdentificeerde risico's.

Risico: Omschrijving van het risico.
Kans: Waarschijnlijkheid dat dit risico optreedt.
Impact: Impact van het risico op het project, 1= gering, 4 = hoog).
Prioriteit: Kans en impact.
Maatregel: Beschrijving per risico van passende tegenmaatregelen.

11 Bibliografie

Anderson, T. W. (1994). *The Statistical Analysis of Time Series*. Retrieved from https://books.google.nl/books?hl=en&lr=&id=rCOzXIC8ZLkC&oi=fnd&pg=PR11&dq=about+time+series&ots=It04mlWzLu&sig=xmEX7iaIWyQQVd9R2vR3l2cYD0c&redir_esc=y#v=onepage&q=about%20time%20series&f=false

BBC. (2018, September 28). MH17 Ukraine plane crash: What we know. Retrieved from http://www.bbc.com/news/world-europe-28357880

Cisco. (2018a, January 19). How Voice over IP (VoIP) Works. Retrieved January 19, 2018, from https://www.cisco.com/c/en/us/solutions/small-business/resource-center/serve-customers-better/voip-how.html

Cisco. (2018b, January 25). Cisco ASR 9000 Series Aggregation Services Routers Simple, Scalable, Secure. Retrieved January 25, 2018, from https://www.cisco.com/c/en/us/products/routers/asr-9000-series-aggregation-services-routers/index.html#~stickynav=4

Faragher, R. (2018). Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation. *IEEE SIGNAL PROCESSING MAGAZINE*, 128–132.

Griffith, E. (2018, February 5). The Ultimate Cord Cutter's Guide. Retrieved March 7, 2018, from https://www.pcmag.com/article2/0,2817,2478213,00.asp

ISO. (2018, January 16). ISO27001. Retrieved January 16, 2018, from https://www.iso.org/isoiec-27001-information-security.html

J., D. (2016, August 16). What is the size of the Windows 10 Anniversary update? Retrieved February 6, 2018, from https://answers.microsoft.com/en-us/windows/forum/windows_10-other_settings/what-is-the-size-of-the-windows-10-anniversary/1ae1dfd0-0b40-476d-a6b6-2434494c7e50

Knippr. (2018, March 7). Knippr. Retrieved March 7, 2018, from https://www.knippr.nl/

McCabe, J. D. (2007). *Network Analysisi, Architecture, and Design* (3rd ed.). Burlington: Morgan Kaufmann Publishers.

McClure, T. (2018, April 4). How Kalman Filters Work, Part 1. Retrieved April 4, 2018, from http://www.anuncommonlab.com/articles/how-kalman-filters-work/

McCrea, N. (2018, April 3). An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples. Retrieved April 3, 2018, from https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1984). *Machine Learning An Artificial Intelligence Approach*. Springer-Verlag.

Nagle, J. (1984, June 1). Congestion Control in IP/TCP Internetworks (RFC896). IETF. Retrieved from https://tools.ietf.org/html/rfc896

NCSS Statistical Software. (2018). ARIMA (Box-Jenkins).

Pariser, E. (2011). *The Filter Bubble: What The Internet Is hiding From You*. Viking. Retrieved from https://books.google.nl/books?hl=en&lr=&id=-FWO0puw3nYC&oi=fnd&pg=PT3&dq=filter+bubble&ots=g4HtIvvXN0&sig=F7uX563PCacylQhIYfJ-zyeIcqI&redir_esc=y#v=onepage&q=filter%20bubble&f=false

Patil, P. S., Rao, S., & Patil, S. B. (2011). Optimization of Data Warehousing System: Simplification in Reporting and Analysis. *IJCA Journal*, (9), 33–37.

Rekhter, Y., Li, T., & Hares, S. (2006, January 1). A Border Gateway Protocol 4 (BGP-4) (RFC4271). Retrieved from https://tools.ietf.org/html/rfc4271

RoutIT. (2018, January 16). RoutIT. Retrieved January 16, 2018, from https://routit.nl/

RoutIT. (2018, May 25). Over RoutIT. Retrieved May 25, 2018, from https://routit.nl/over-routit/

Sallam, R. L., Howson, C., Idoine, C. J., Oestreich, T. W., Richardson, J. L., & Tapadinhas, J. (2017). *Magic Quadrant for Business Intelligence and Analytics Platforms* (Magic Quadrant No. G00301340) (p. 82). Gartner.

Smola, A., & Vishwanathan, S. V. N. (2008). *Introduction to Machine Learning*. Retrieved from http://alex.smola.org/drafts/thebook.pdf

Splunk. (2018, April 9). Predict. Retrieved April 9, 2018, from https://docs.splunk.com/Documentation/Splunk/7.0.3/SearchReference/Predict

Statcounter. (2018a, February 5). Desktop Windows Version Market Share Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from http://gs.statcounter.com/windows-version-market-share/desktop/netherlands/#monthly-201701-201801-bar

Statcounter. (2018b, February 5). Mobile & Tablet Operating System Market Share Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from

http://gs.statcounter.com/os-market-share/mobile-tablet/netherlands/#monthly-201701-201801-bar

Statcounter. (2018c, February 5). Social Media Stats Netherlands Jan 2017 - Jan 2018. Retrieved February 5, 2018, from http://gs.statcounter.com/social-media-stats/all/netherlands/#monthly-201701-201801-bar

Statista. (2018). Number of monthly active WhatsApp users worldwide from April 2013 to December 2017 (in millions). Retrieved February 5, 2018, from https://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/

Stodder, D. (2015). *Visual Analytics for Making Smarter Decisions Faster*.

Tableau. (2018, March 7). How Forecasting Works in Tableau. Retrieved March 4, 2018, from http://onlinehelp.tableau.com/current/pro/desktop/en-us/forecast_how_it_works.html

The Pennsylvania State University. (2018, April 3). Seasonal ARIMA models. Retrieved April 3, 2018, from https://onlinecourses.science.psu.edu/stat510/node/67

van Miltenburg, O. (2018, March 21). T-Mobile stopt met tv-dienst Knippr wegens te weinig klanten. Retrieved from https://tweakers.net/nieuws/136509/t-mobile-stopt-met-tv-dienst-knippr-wegens-te-weinig-klanten.html

Zabbix. (2015). Forecasting trigger functions (p. 10). Zabbix. Retrieved from http://zabbix.org/mw/images/1/18/Prediction_docs.pdf

Zabbix. (2018a, January 16). Zabbix. Retrieved January 16, 2018, from https://www.zabbix.com/

Zabbix. (2018b, January 19). Configuring a host. Retrieved January 19, 2018, from https://www.zabbix.com/documentation/3.4/manual/config/hosts/host

Zabbix. (2018c, January 30). 3 History and trends. Retrieved January 30, 2018, from https://www.zabbix.com/documentation/3.4/manual/config/items/history_and_trends

Zabbix. (2018d, January 31). 2. Definitions. Retrieved January 31, 2018, from https://www.zabbix.com/documentation/3.4/manual/definitions

Ziggo. (2018, March 7). Ziggo Go. Retrieved March 7, 2018, from https://www.ziggo.nl/televisie/online-tv-kijken/

Bijlage B: Afstudeercontract

Bijlage C – MoSCoW analyse

Hieronder treft u de door de afstudeerder opgestelde en door de opdrachtgeven geaccordeerde MoSCoW analyse aan. Deze MoSCoW analyse is opgesteld in overleg met de opdrachtgever.

Must have

- Een rechtenmodel om gebruikers toegang te geven en te ontzeggen tot geselecteerde functionaliteit;
- De mogelijkheid om gebruikers op basis van externe bronnen te authentiseren;
- De BI-oplossing dient gebruik te maken van een cliënt / server model, het mag dus geen Desktopapplicatie zijn;
- Eigen opslag van te analyseren data zodat de BI-oplossing niet constant de data uit andere systemen moet betrekken in de vorm van een stream;
- Mogelijkheid tot het doen van geautomatiseerde bewerkingen op de data voorafgaand aan het analyse proces;
- Uitgebreide importmogelijkheden voor data met tenminste ondersteuning voor een import via het CSV-formaat;
- Voldoet aan securitybeleid RoutIT;
- Exportmogelijkheden van grafieken en data ten behoeve van opname in documenten.

Should have

- Bediening via een in HTML5 geschreven web interface waarbij de volledige functionaliteit beschikbaar is;
- Commerciële supportmogelijkheden voor ondersteuning bij problemen met de BI-oplossing waar RoutIT engineers niet meer uitkomen;
- Het genereren van op het management gerichte rapportage;
- Alarmering bij overschrijden van drempelwaarden of de mogelijkheid dit via Zabbix te doen;
- Goede technische documentatie beschikbaar;
- Systeem kan grafieken genereren waarbij de weergave van de grafieken aanpasbaar is.

Could have

- Oplossing draait op een door respectievelijk Red Hat en Microsoft ondersteunde CentOS en / of RHEL of Windows Server versie;
- *Regular Expressions* (regex) ondersteuning daar waar dit het werk van de Power User vergemakkelijkt;
- Ondersteuning Python en R bij het schrijven van datamodellen of andere functionaliteiten;
- On the fly aanpassen van datamodel ten bate van troubleshooting waarbij de effecten gelijk zichtbaar zijn.

Would have

- Systeem is zelf in staat om data uit externe, niet door RoutIT beheerde, databronnen op te halen.

Deze MoSCoW is ondersteunend aan het doel van de opdracht. De hoofdeis is het kunnen doen van voorspellingen, deze MoSCoW analyse bevat dan ook alleen de aanvullende eisen voor een pakket. Het kunnen doen van voorspellingen zal in deze scriptie apart geanalyseerd worden.

Bijlage D – Zelfreflectie

Bijlage E –PoC-ontwerp

Ontwerp Proof of Concept (PoC) project Forecasting

Auteur: Bart Dekker
Naam bedrijf: RoutIT
Datum: 28-03-2018
Status: Definitief
Versie: 1,0

Inhoudsopgave

1 Inleiding 71
2 Opzet testomgeving 71
3 Test data 71

1 Inleiding

In dit document treft u het ontwerp van het markt verkennende Proof of Concept (PoC) aan. De tijdens de markt verkennende PoC geselecteerde *Business Intelligence* (BI) oplossing zal vervolgens onder dezelfde condities als tijdens de markt verkennende PoC getest worden aangaande de voorspellende functie. De volgende pakketten zijn getest tijdens de markt verkennende PoC, de weergave is alfabetisch:

1. Datameer;
2. Elasticsearch, Logstash & Kibana (ELK) *stack* met de X-*pack add-on*;
3. Logi Analytics;
4. Splunk Enterprise;
5. TIBCO Spotfire.

Tijdens de markt verkennende PoC is aan de hand van het testplan de pakketselectie voor voornoemde pakketten ingevuld.

2 Opzet testomgeving

Door RoutIT is capaciteit op het ISG VMWare gebaseerde cloud platform beschikbaar gesteld voor dit project. Per pakket is onderzocht op welke *Operating Systems* (OS) het desbetreffende pakket ondersteund wordt. Indien CentOS één van de opties was, is wegens de voorkeur van RoutIT voor CentOS gekozen als te gebruiken OS. In andere gevallen is gekeken naar welk OS door de leverancier ondersteund wordt. Aangezien de leveranciers die hun BI-oplossing alleen op Windows ondersteunen gekozen hebben voor het beschikbaar stellen van een cloud omgeving, is Windows alleen gebruikt voor het opzetten van een *Active Directory* (AD) server ten behoeve van het testen van de *Lightweight Directory Access Protocol* (LDAP) capaciteiten van het pakket. De keuze voor AD als LDAP server is gemaakt aangezien de afstudeerder vanuit de opleiding ervaring heeft met AD, waardoor de doorlooptijd van deze specifieke test verkort kan worden.

De pakketten Datameer en TIBCO zijn getest in de cloud omgeving van de desbetreffende leveranciers, omdat dit de enige beschikbaar gestelde mogelijkheid was. De pakketten ELK-*stack* en Splunk *Enterprise* zijn wel getest op CentOS. Met de leverancier van Logi Analytics is contact geweest over een probeerversie van het pakket, maar zij gaven in het gevoerde telefoongesprek aan dat zij de door RoutIT gewenste functionaliteit niet kunnen leveren op dit moment. In afbeelding 1 is de testomgeving schematisch samen met de configuratie parameters weergegeven.

Ten behoeve van de door de afstudeerder tijdens de markt verkennende PoC te schrijven Python v3 scripts is een Ubuntu *Virtual Machine* beschikbaar gesteld. Voor Python v3 in plaats van Python v2 is gekozen omdat de afstudeerder deze Python versie geleerd heeft tijdens de opleiding waardoor de doorlooptijd verkort kan worden. Gekozen is voor Ubuntu omdat op CentOS Python v3 nog niet wordt meegeleverd.

3 Test data

Ten behoeve van de testen is een Python v3 script geschreven voor het genereren van fictieve data ten einde een data lek van concurrentie gevoelige informatie te voorkomen. Het Python script is integraal opgenomen in het testplan. Aangezien uit het vooronderzoek is gebleken dat de makkelijkste manier om data uit de datasets IRMA en RTK te exporteren in het CSV-formaat, genereert het voornoemde Python script fictieve data in CSV-formaat.

<Verwijderd>
Afbeelding 1: Ontwerp Proof of Concept (PoC).

Bijlage F –PoC-ontwerp

Testplan Proof of Concept (PoC) project Forecasting

Auteur: Bart Dekker
Naam bedrijf: RoutIT
Datum: 27-03-2018
Status: Definitief
Versie: 1,0

Inhoudsopgave

1 Inleiding ..... **Error! Bookmark not defined.**

2 Het genereren van fictieve test data ..... **Error! Bookmark not defined.**

3 Must have gerelateerde testen ..... **Error! Bookmark not defined.**

3.1 Aanwezigheid rechtenmodel ..... **Error! Bookmark not defined.**

3.2 Externe gebruikersauthenticatie ..... **Error! Bookmark not defined.**

3.3 Betreft de BI-oplossing een cliënt / server model? ..... **Error! Bookmark not defined.**

3.4 Geautomatiseerde bewerkingen op de data ... **Error! Bookmark not defined.**

3.5 Importeren van data ..... **Error! Bookmark not defined.**

3.6 Securitybeleid RoutIT ..... **Error! Bookmark not defined.**

3.7 Exporteren van data en grafieken ..... **Error! Bookmark not defined.**

4 Should have gerelateerde testen ..... **Error! Bookmark not defined.**

4.1 Web interface op basis van HTML5 ..... **Error! Bookmark not defined.**

4.2 Commerciële supportmogelijkheden ..... **Error! Bookmark not defined.**

4.3 Rapportage functionaliteit ..... **Error! Bookmark not defined.**

4.4 Alarmering functionaliteit ..... **Error! Bookmark not defined.**

4.5 Technische documentatie ..... **Error! Bookmark not defined.**

4.6 Genereren grafieken ..... **Error! Bookmark not defined.**

5 Could have gerelateerde testen ..... **Error! Bookmark not defined.**

5.1 Ondersteunde besturingssystemen ..... **Error! Bookmark not defined.**

5.2 Ondersteuning voor Regular Expressions ..... **Error! Bookmark not defined.**

5.3 Ondersteuning voor Python en R ..... **Error! Bookmark not defined.**

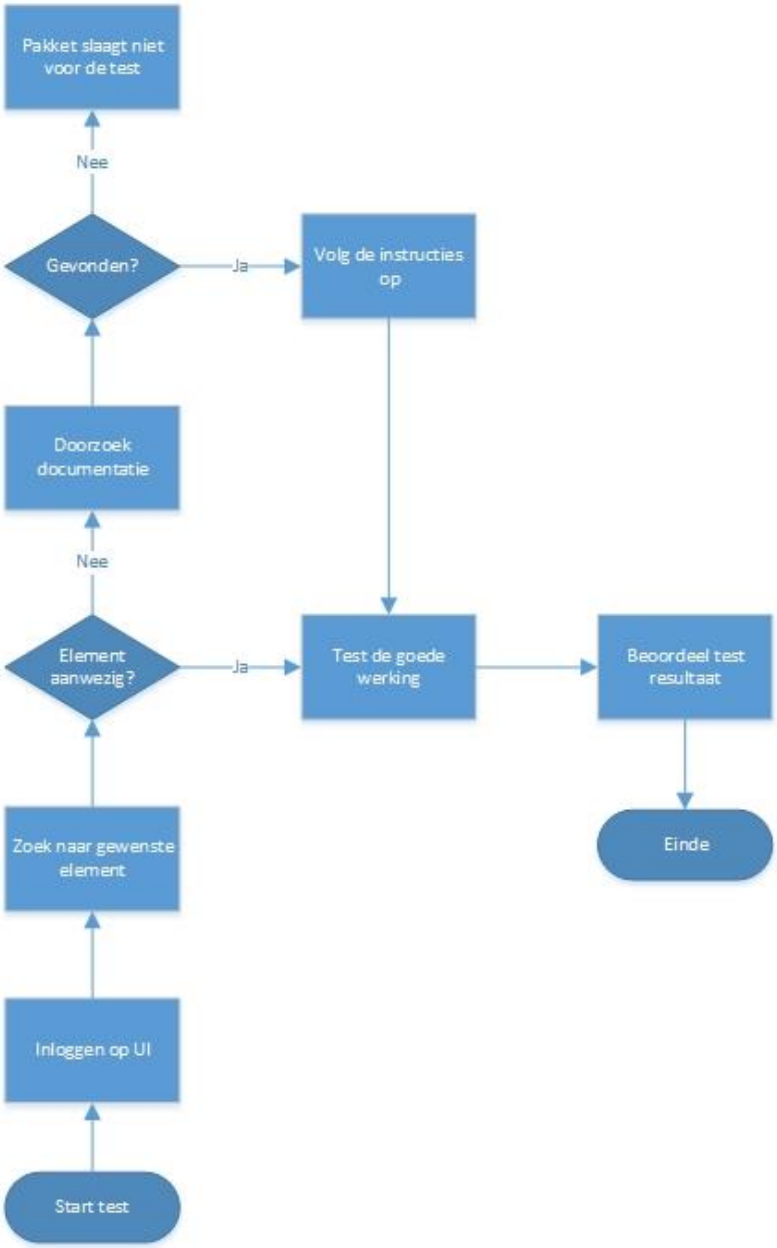
5.4 On the fly aanpassingen ..... **Error! Bookmark not defined.**

6 Would have gerelateerde testen ..... **Error! Bookmark not defined.**

1 Inleiding

In dit document treft u het testplan voor de verschillende *Business Intelligence* (BI) oplossingen ten behoeve van het onderzoek *Forecasting*. Dit testplan is gebruikt om tests uit te voeren op de top 5 van pakketten voortkomende uit de MoSCoW analyse en wordt een markt verkennend Proof of Concept (PoC) genoemd. Op basis van deze testen zal uiteindelijk één enkel pakket geselecteerd worden dat beoordeeld gaat worden op zijn voorspellende capaciteiten aangaande het netwerkverkeer. Ter voorkoming van het lekken van concurrentie gevoelige informatie treft u in het eerste hoofdstuk een in Python v3 geschreven script aan ten behoeve van het genereren van test data.

In afbeelding 1 treft u de ruwe werking aan van dit testplan. Aangezien de testen van elkaar verschillen wat betreft inhoud is afbeelding 1 niet één op één te vertalen naar een specifieke test.



Afbeelding 1: Stroomdiagram met ruwe werking onderstaande testen.

2 Het genereren van fictieve test data

Het volgende Python v3 script is gebruikt om test data te genereren ten behoeve van de eerste marktverkennende PoC. Het script genereerd een CSV-bestand met de volgende kolommen:

- date
- time
- direction
- avg_speed

dit CSV-bestand bevat fictieve gegevens van een enkele *Network Interface Card* (NIC). Aangezien Zabbix dergelijke gegevens iedere halve minuut uitleest is voor een maand iedere halve minuut in beide richtingen een willekeurig gekozen waarde tussen de <verwijderd> en <verwijderd> gekozen. De waarde als uitgegeven door Zabbix is in bytes per seconde, dus de uiteindelijke waarde onder “avg_speed” reflecteert dat.

```
#!/usr/bin/python3
import csv,random,datetime

with open('zabbix_fake_data.csv', 'w', newline='') as csvfile:
    fake_data = csv.writer(csvfile, delimiter=',')
    fake_data.writerow(['date', 'time', 'direction', 'avg_speed'])
    for day in range(1,32):
        for hour in range(24):
            for minute in range(60):
                for second in range(0,60,30):
                    date = datetime.date(year=2018, month=1,day=day)
                    time = datetime.time(hour=hour, minute=minute, second=second)
                    avg_speed_in = random.randrange(<verwijderd>)
                    avg_speed_out = random.randrange(<verwijderd>)
                    fake_data.writerow([date, time, 'in', avg_speed_in])
                    fake_data.writerow([date, time, 'out', avg_speed_out])
```

3 Must have gerelateerde testen

In dit hoofdstuk worden de *Must haves* uit de MoSCoW analyse getest. Iedere *Must have* uit de MoSCoW analyse staat in een aparte paragraaf waarbij de paragraaf begint met de letterlijke omschrijving uit de MoSCoW analyse die getest gaat worden.

3.1 Aanwezigheid rechtenmodel

MoSCoW omschrijving: Een rechtenmodel om gebruikers toegang te geven en te ontzeggen tot geselecteerde functionaliteit.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar elementen die wijzen op rechtenbeheer of rollen.
3. Indien punt 2 niet slaagt wordt de documentatie geraadpleegd als verstrekt door de leverancier.
4. Gekeken wordt of gebruikers de rechten *Read Only* (RO), *Read Write* (RW) en het recht van toegang tot onderdelen van de BI-oplossing gegeven of ontzegt kan worden.

Deze test is geslaagd als punt 4 met succes is uitgevoerd.

3.2 Externe gebruikersauthenticatie

MoSCoW omschrijving: De mogelijkheid om gebruikers op basis van externe bronnen te authentiseren.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar elementen die wijzen op extern gebruikersbeheer als *Lightweight Directory Access Protocol* (LDAP).
3. Indien punt 2 niet slaagt wordt de documentatie geraadpleegd als verstrekt door de leverancier.
4. Tijdens deze test wordt gebruik gemaakt van een juist ingerichte Microsoft Windows Server *Active Directory* (AD) server ten behoeve van een test met LDAP als zijnde de defacto standaard voor gebruikersuitwisseling. Alle andere mogelijkheden als Kerberos en *Remote Authentication Dial-In User Service* (RADIUS) zijn een pré.
5. De verbinding tussen de BI-oplossing en de AD-server wordt geconfigureerd op de door de leverancier voorgeschreven manier.
6. Getracht wordt om in te loggen op de BI-oplossing met een gebruiker waarvan geverifieerd is dat deze alleen op de AD-server bestaat en niet in de BI-oplossing.

Deze test is geslaagd als punt 6 met succes is uitgevoerd.

3.3 Betreft de BI-oplossing een cliënt / server model?

MoSCoW omschrijving: De BI-oplossing dient gebruik te maken van een cliënt / server model; Het mag dus geen Desktopapplicatie zijn.

1. De BI-oplossing wordt geïnstalleerd volgens de instructies van de fabrikant in een *Virtual Machine* (VM).
2. De *User Interface* (UI) wordt op de door de fabrikant voorgeschreven manier benaderd vanaf een met de VM in verbinding staande cliëntcomputer.
3. Indien de UI van de BI-oplossing beschikbaar is vanaf de cliëntcomputer kan gesteld worden dat de test geslaagd is.
4. Indien de BI-oplossing niet beschikbaar is vanaf de cliëntcomputer maar dit volgens de documentatie wel moet kunnen, dient de tester oorzaken van technische aard uit te sluiten.
5. Indien de BI-oplossing installeert als een desktopapplicatie en niet over het netwerk te bereiken is, slaagt de BI-oplossing niet voor deze test.

Deze test is geslaagd als punt 3, 4 of 5 met succes zijn uitgevoerd.

3.4 Geautomatiseerde bewerkingen op de data

MoSCoW omschrijving: Mogelijkheid tot het doen van geautomatiseerde bewerkingen op de data voorafgaand aan het analyse proces.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar elementen die wijzen op geautomatiseerde bewerkingen.
3. Indien punt 2 slaagt wordt getracht om het veld “avg_speed” als gegenereerd door voornoemd script te vermenigvuldigen met 10 in een separaat veld.
4. Indien punt 2 niet slaagt wordt de documentatie geraadpleegd, waarna punt 3 herhaald wordt.
5. Het onder punt 3 toegevoegde veld dient uitleesbaar en hanteerbaar te zijn.

Deze test is geslaagd als punt 5 met succes is uitgevoerd.

3.5 Importeren van data

MoSCoW omschrijving: Uitgebreide importmogelijkheden voor data met tenminste ondersteuning voor een import via het CSV-formaat.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. De in de bovenstaande paragraaf gegenereerde data wordt op de door de leverancier voorgeschreven manier geïmporteerd in de BI-oplossing.
3. Gekeken wordt of de data uit te lezen en te gebruiken in andere onderdelen van de BI-oplossing is.
4. In het geval van een zelf geïnstalleerde BI-oplossing wordt de BI-oplossing herstart. Daarna wordt punt 2 herhaald.
5. In het geval van een cloud BI-oplossing wordt uitgelogd en de browser cache weggegooid. Daarna wordt punt 2 herhaald.
6. In de UI wordt gekeken of er andere mogelijkheden voor het importeren van data zijn.

Deze test is geslaagd als punt 3 en 4 of 5 met succes zijn uitgevoerd.

3.6 Securitybeleid RoutIT

MoSCoW omschrijving: Voldoet aan securitybeleid RoutIT.

Wegens de gevoeligheid van de in de BI-oplossing te importeren data moet de BI-oplossing na implementatie gekeurd worden door de *Security Officer* (SO). Aangezien de SO hier zijn eigen procedures en eisen voor heeft wordt er in dit testplan verder niets over geschreven ten einde de SO niet te belemmeren in zijn werkzaamheden.

3.7 Exporteren van data en grafieken

MoSCoW omschrijving: Exportmogelijkheden van grafieken en data ten behoeve van opname in documenten.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar elementen die wijzen op exporteermogelijkheden.
3. De kale data moet geëxporteerd kunnen worden in een open formaat als CSV of JSON, indien de data alleen in een aan de BI-oplossing eigen formaat geëxporteerd kan worden slaagt deze test niet.
4. De grafieken dienen geëxporteerd te kunnen worden in een courant afbeeldingsformaat als PNG of JPG.

Deze test is geslaagd als punt 3 en 4 met succes zijn uitgevoerd.

3.8 Should have gerelateerde testen

In dit hoofdstuk worden de *Should haves* uit de MoSCoW analyse getest. Iedere *Should have* uit de MoSCoW analyse staat in een aparte paragraaf waarbij de paragraaf begint met de letterlijke omschrijving uit de MoSCoW analyse die getest gaat worden.

3.9 Web interface op basis van HTML5

MoSCoW omschrijving: Bediening via een in HTML5 geschreven web interface waarbij de volledige functionaliteit beschikbaar is.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Aan de browser wordt gevraagd om de onderliggende HTML-code te tonen.

3. Indien de openende HTML-tag “<html>” betreft is de web interface in HTML5 geschreven.

Deze test is geslaagd indien punt 3 met succes is uitgevoerd.

3.10 Commerciële supportmogelijkheden

MoSCoW omschrijving: Commerciële supportmogelijkheden voor ondersteuning bij problemen met de BI-oplossing waar RoutIT engineers niet meer uitkomen.

1. De officiële leverancierswebsite wordt geopend in een *browser*.
2. Gezocht wordt op de leverancierswebsite naar commerciële supportmogelijkheden.

Deze test is geslaagd is geslaagd als onder punt 2 commerciële supportmogelijkheden gevonden worden.

3.11 Rapportage functionaliteit

MoSCoW omschrijving: Het genereren van op het management gerichte rapportage.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar rapportagemogelijkheden waarbij het rapport geëxporteerd moet kunnen worden in het PDF-formaat.
3. Een rapport wordt gegenereerd in de BI-oplossing en bekeken in een PDF-viewer op de cliëntcomputer.
4. Indien punt 2 niet slaagt wordt de documentatie geraadpleegd, waarna punt 2 herhaald wordt.

Deze test is geslaagd als punt 3 met succes is uitgevoerd.

3.12 Alarmering functionaliteit

MoSCoW omschrijving: Alarmering bij overschrijden drempelwaarden of de mogelijkheid dit via Zabbix te doen.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar alarmeringsmogelijkheden waarbij een alarm kan worden afgegeven via externe communicatie als email.
3. Gekeken wordt naar de opties bij het definiëren van een alarm, is het ingeven van een drempelwaarde mogelijk?
4. Indien punt 2 niet slaagt wordt de documentatie geraadpleegd, waarna punt 2 herhaald wordt.

Deze test is geslaagd als punt 3 met succes is uitgevoerd.

3.13 Technische documentatie

MoSCoW omschrijving: Goede technische documentatie beschikbaar.

1. Op de website van de leverancier wordt gezocht naar technische, op het productgerichte, documentatie. Een *case study* waarin een andere klant wordt beschreven valt hier niet onder.
2. Tijdens de installatie en het onderzoeken van de andere eisen uit de MoSCoW-analyse zal de documentatie op bruikbaarheid beoordeeld worden.

Deze test is geslaagd als punt 2 met succes is uitgevoerd.

3.14 Genereren grafieken

MoSCoW omschrijving: Systeem kan grafieken genereren waarbij de weergave van de grafieken aanpasbaar is.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt naar elementen die verwijzen naar grafieken in de UI.
3. Op basis van de gegenereerde en geïmporteerde data wordt getracht een grafiek weer te geven.
4. Bekeken wordt of er instelmogelijkheden zijn om de grafiekweergave aan te passen. Deze mogelijkheden worden ook getest op hun werking.

Deze test is geslaagd als punt 4 met succes is uitgevoerd.

4 Could have gerelateerde testen

In dit hoofdstuk worden de *Could haves* uit de MoSCoW analyse getest. Iedere *Could have* uit de MoSCoW analyse staat in een aparte paragraaf waarbij de paragraaf begint met de letterlijke omschrijving uit de MoSCoW analyse die getest gaat worden.

4.1 Ondersteunde besturingssystemen

MoSCoW omschrijving: Oplossing draait op een door respectievelijk Red Hat en Microsoft ondersteunde CentOS en / of RHEL of Windows Server versie.

1. De BI-oplossing wordt geïnstalleerd zoals aangegeven door de leverancier.
2. Indien de software daadwerkelijk geïnstalleerd is op één van de volgende besturingssystemen: CentOS, Red Hat Enterprise Linux (RHEL) of Microsoft Windows Server, wordt gekeken naar het versienummer van het besturingssysteem waarop de BI-oplossing geïnstalleerd is.
3. Gekeken wordt op de productpagina van het onder punt 2 gebruikte besturingssysteem met bijbehorend versienummer of het desbetreffende versienummer tot na afloop van dit onderzoek ondersteund is door de desbetreffende besturingssysteem leverancier.

Deze test is geslaagd als punt 3 met succes is uitgevoerd.

4.2 Ondersteuning voor Reqular Expressions

MoSCoW omschrijving: Regular Expressions (regex) ondersteuning daar waar dit het werk van de Power User vergemakkelijkt.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar invoervelden die Regular Expressions accepteren. Hierbij wordt afgegaan op al reeds met het pakket opgedane ervaring.
3. Indien punt 2 geen resultaat oplevert wordt gezocht in de documentatie van de leverancier naar Regular Expression mogelijkheden waarna punt 2 herhaald wordt.
4. De gevonden invoervelden worden getest door een Regular Expression in te voeren.
5. Indien het systeem aangeeft dat de ingevoerde Regular Expression niet correct is zal een ander Regular Expression dialect ingevoerd worden.
6. Indien het systeem aangeeft dat de invoer niet herkend wordt, is de conclusie dat de BI-oplossing geen Regular Expressions ondersteund.

Deze test is geslaagd als punt 4 met succes is uitgevoerd.

4.3 Ondersteuning voor Python en R

MoSCoW omschrijving: Ondersteuning Python en R bij het schrijven van datamodellen of andere functionaliteiten.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. Gezocht wordt in de UI naar invoervelden die Python accepteren. Hierbij wordt afgegaan op al reeds met het pakket opgedane ervaring.
3. Indien punt 2 geen resultaat oplevert wordt gezocht in de documentatie van de BI-oplossing leverancier.
4. De gevonden invoervelden worden getest door Python v2 en Python v3 in te voeren. Eén van beide Python varianten dient te werken.
5. Gezocht wordt in de UI naar invoervelden die R accepteren. Hierbij wordt afgegaan op al reeds met het pakket opgedane ervaring.
6. Indien punt 5 geen resultaat oplevert wordt gezocht in de documentatie van de BI-oplossing leverancier.
7. De gevonden invoervelden worden getest door R in te voeren.

Deze test is geslaagd als de punten 4 en 7 met succes zijn uitgevoerd. Bij het behalen van punt 4 of 7 wordt de helft van het maximale aantal punten voor dit onderdeel toegekend in de pakketselectie.

4.4 On the fly aanpassingen

MoSCoW omschrijving: On the fly aanpassen datamodel ten bate van troubleshooting waarbij de effecten gelijk zichtbaar zijn.

1. Op de *User Interface* (UI) van de BI-oplossing wordt ingelogd op de door de leverancier voorgeschreven manier.
2. De voorgaand gegenereerde grafieken worden geanalyseerd betreffende de aanpassingsmogelijkheden.
3. De desbetreffende grafiek wordt aangepast en het resultaat wordt beoordeeld. Indien het kopiëren van de grafiek noodzakelijk is, dan is deze test niet geslaagd.

Deze test is geslaagd als punt 3 met succes is uitgevoerd.

5 Would have gerelateerde testen

De *Would have* testen zullen in verband met de beschikbare tijd en de rangschikking in de MoSCoW analyse niet uitgevoerd en daarom ook niet beschreven worden. Daarnaast is het zo dat de enige *Would have* tevens in strijd is met *security best practices*.

Bijlage G –Pakketselectie

	Totaal	Rechtenmodel	Extern gebruikersbeheer	Client/srv model	Eigen opslag data	Geautomatiseerde bewerkingen	Import via CSV, etc.	Export mogelijkheden data / grafieken
Elasticsearch	20	3	3	3	3	3	3	2
Logi Analytics	20	3	3	3	3	3	3	2
Datameer	19	2	3	3	3	3	3	2
Splunk	19	3	3	3	3	3	2	2
TIBCO Software	19	3	3	3	2	3	3	2
MicroStrategy	17	1	3	3	3	2	3	2
Zoomdata	17	2	3	3	0	3	3	3
Birst	15	1	0	3	3	3	3	2
SAS	15	3	1	3	3	2	2	1
Board International	13	2	3	3	0	2	1	2
Microsoft	13	1	0	3	3	3	3	0
Sisense	13	2	1	3	3	3	1	0
Yellowfin	13	3	2	3	0	3	2	0
Salesforce	12	0	3	3	1	3	1	1
Alteryx	10	2	1	1	0	1	3	2
Pentaho	8	2	3	1	0	1	0	1
Oracle	6	0	1	1	3	0	1	0
SAP	0	0	0	0	0	0	0	0
0		Niet aanwezig of niet te vinden op afstand						
1		Wel aanwezig maar met veel hackwerk						
2		Wel aanwezig en voldoende of op afstand niet genoeg af te tasten of het een drie is.						
3		Precies zoals gewenst.						

Fictieve data generator

```
#!/usr/bin/python3
import csv,random,datetime

# Initialiseer een CSV file als doel bestand.
with open('zabbix_fake_data.csv', 'w', newline='') as csvfile:
	fake_data = csv.writer(csvfile, delimiter=',')
	fake_data.writerow(['date', 'time', 'direction', 'avg_speed'])

	# Schrijf voor de maand 01-2018 voor iedere dag per 30sec een random waarde weg. 1x voor ingaand en 1x voor uitgaand verkeer.
	for day in range(1,32):
		for hour in range(24):
			for minute in range(60):
				for second in range(0,60,30):
					date = datetime.date(year=2018, month=1,day=day)
					time = datetime.time(hour=hour, minute=minute, second=second)
					avg_speed_in = random.randrange(<verwijderd>)
					avg_speed_out = random.randrange(<verwijderd>)
					fake_data.writerow([date, time, 'in', avg_speed_in])
					fake_data.writerow([date, time, 'out', avg_speed_out])
```

Zabbix History data download script

```
#!bin/python
from get_history import *
import json,csv
import time

# Variabelen met inloggegevens Zabbix API.
zabbix_url = <verwijderd>
zabbix_usr = <verwijderd>
zabbix_pwd = <verwijderd>

# Open het CSV bestand met op te halen items.
with open ('network_int.csv', newline='') as csvfile:
	reader = csv.DictReader(csvfile)

	# Bereken de op te vragen tijden in Epoch.
	current_time = time.localtime()
	current_time_epoch =time.mktime(current_time)
	past_14days = current_time_epoch - 70 * 24 * 60**2

	# Initialiseer de doel bestanden.
	with open('historische_data.json', 'w') as json_file:
		json_file.write('[\n')
	with open('historische_data.csv', 'w') as csv_file:
		csv_file.write("interface" + "," + "_time" + "," + "itemid" + "," + "hostname" + "," + "bandwith" + "," + "ns" + "\n")

	# Loop door de lijst met op te vragen items heen en voer deze op aan de get_history library.
	for row in reader:
		api = startSession(zabbix_url, zabbix_usr, zabbix_pwd)
		host_details = getHostDetails(api, row["hostname"])
		hostid = host_details[0]['hostid']
		item_details = getItemDetails(api, hostid, row["interface_in"])
		history = getHistoryValues(api, item_details[0]['itemid'], past_14days, current_time_epoch)

		# Schrijf de data weg naar CSV en JSON. Extra code toegevoegd om human readable host- en interface name op te nemen. Deze code handelt de inkomende interfaces af.
		for value in history["history_values"]:
			value["hostname"] = row["hostname"]
```

```
        value["interface"] = row["interface_in"]
        with open('historische_data.json', 'a') as json_file:
            #json.dump(value, json_file, indent=4)
            data = json.dumps(value, indent=4)
            json_file.write(str(data) + ",\n")
        with open('historische_data.csv', 'a') as csv_file:
            csv_file.write(value["interface"] + "," + value["clock"]+ "," + value["itemid"] + "," + value["hostname"] + "," +
value["value"] + "," + value["ns"] + "\n")

    host_details = getHostDetails(api, row["hostname"])
    hostid = host_details[0]['hostid']
    item_details = getItemDetails(api, hostid, row["interface_out"])
    history = getHistoryValues(api, item_details[0]['itemid'], past_14days, current_time_epoch)

    # Schrijf de data weg naar CSV en JSON. Extra code toegevoegd om human readable host- en interface name
op te nemen. Deze code handelt de uitgaande interfaces af.
    for value in history["history_values"]:
        value["hostname"] = row["hostname"]
        value["interface"] = row["interface_out"]
        with open('historische_data.json', 'a') as json_file:
            #json.dump(value, json_file, indent=4)
            data = json.dumps(value, indent=4)
            json_file.write(str(data) + ",\n")
        with open('historische_data.csv', 'a') as csv_file:
            csv_file.write(value["interface"] + "," + value["clock"] + "," + value["itemid"] + "," + value["hostname"] + "," +
value["value"] + "," + value["ns"] + "\n")

  # Sluit de JSON file af volgens de JSON specificaties.
  with open('historische_data.json', 'a') as json_file:
```

Zabbix Trend data download script

```
#!bin/python
from get_history import *
import json,csv
import time

# Variabelen met inloggegevens Zabbix API.
zabbix_url = <verwijderd>
zabbix_usr = <verwijderd>
zabbix_pwd = <verwijderd>

# Open het CSV bestand met op te halen items.
with open ('network_int.csv', newline='') as csvfile:
  reader = csv.DictReader(csvfile)

  # Bereken de op te vragen tijden in Epoch.
  current_time = time.localtime()
  current_time_epoch =time.mktime(current_time)
  past_time = current_time_epoch - 52 * 7 * 24 * 60**2
  current_time_epoch -= 14 * 24 * 60**2

  # Initialiseer de doel bestanden.
  with open('trend_data.json', 'w') as json_file:
    json_file.write('[\n')
  with open('trend_data.csv', 'w') as csv_file:
    csv_file.write("interface" + "," + "_time" + "," + "itemid" + "," + "hostname" + "," + "bandwith" + "," + "num" + "\n")

  # Loop door de lijst met op te vragen items heen en voer deze op aan de get_history library.
  for row in reader:
    api = startSession(zabbix_url, zabbix_usr, zabbix_pwd)
```

```
        host_details = getHostDetails(api, row["hostname"])
        hostid = host_details[0]['hostid']
        item_details = getItemDetails(api, hostid, row["interface_in"])
        trend = getTrendValues(api, item_details[0]['itemid'], past_time, current_time_epoch)

        # Schrijf de data weg naar CSV en JSON. Extra code toegevoegd om human readable host- en interface name
op te nemen. Deze code handelt de inkomende interfaces af.
        for value in trend["trend_values"]:
            value["hostname"] = row["hostname"]
            value["interface"] = row["interface_in"]
            with open('trend_data.json', 'a') as json_file:
                #json.dump(value, json_file, indent=4)
                data = json.dumps(value, indent=4)
                json_file.write(str(data) + ",\n")
            with open('trend_data.csv', 'a') as csv_file:
                csv_file.write(value["interface"] + "," + value["clock"]+ "," + value["itemid"] + "," + value["hostname"] + "," +
value["value_max"] + "," + value["num"] + "\n")

        host_details = getHostDetails(api, row["hostname"])
        hostid = host_details[0]['hostid']
        item_details = getItemDetails(api, hostid, row["interface_out"])
        trend = getTrendValues(api, item_details[0]['itemid'], past_time, current_time_epoch)

        # Schrijf de data weg naar CSV en JSON. Extra code toegevoegd om human readable host- en interface name
op te nemen. Deze code handelt de uitgaande interfaces af.
        for value in trend["trend_values"]:
            value["hostname"] = row["hostname"]
            value["interface"] = row["interface_out"]
            with open('trend_data.json', 'a') as json_file:
                #json.dump(value, json_file, indent=4)
                data = json.dumps(value, indent=4)
                json_file.write(str(data) + ",\n")
            with open('trend_data.csv', 'a') as csv_file:
                csv_file.write(value["interface"] + "," + value["clock"] + "," + value["itemid"] + "," + value["hostname"] + "," +
value["value_max"] + "," + value["num"] + "\n")

    # Sluit de JSON file af volgens de JSON specificaties.
    with open('trend_data.json', 'a') as json_file:
        json_file.write(']')
```

Zabbix data transformatie script

```
#!bin/python
import csv,re

# Op de gewenste CSV file met negatieve getallen.
with open('trend_data.csv', newline='') as csvfile:
    reader = csv.DictReader(csvfile, delimiter=',')

    # Maak een nieuwe CSV file aan om de opgeschoonde data naar weg te schrijven.
    with open('trend_data_sliced.csv', 'w', newline='') as csvfile_sliced:
        csvfile_sliced.write('interface' + ',' + '_time' + ',' + 'itemid' + ',' + 'hostname' + ',' + 'bandwith' + ',' + 'num' + '\n')

        # Loop door de bron data heen.
        for row in reader:

            # Controleer het veld met negatieve getallen (bandwith) op negatieve getallen en slice de '-' er af.
            if re.search('^-', row['bandwith']):
                slice = row['bandwith']
                row['bandwith'] = slice[1:]
```

```
        csvfile_sliced.write(row['interface'] + ',' + row['_time'] + ',' + row['itemid'] + ',' + row['hostname'] + ',' +
row['bandwith'] + ',' + row['num'] + '\n')

    # Schrijf de positieve getallen zonder bewerking weg.
    else:
        csvfile_sliced.write(row['interface'] + ',' + row['_time'] + ',' + row['itemid'] + ',' + row['hostname'] + ',' +
row['bandwith'] + ',' + row['num'] + '\n')
```