

BACHELORSCHRIJF
INFORMATICA



RADBOUD UNIVERSITEIT

Vertrouwelijkheid in het iRN

Auteur:
Lars Bade
s4051513

Inhoudelijk begeleider:
ing. R. J. M. Wichers Schreur
ronny@cs.ru.nl

Tweede lezer:
G. Alpár
gergely@cs.ru.nl

15 juli 2014

Samenvatting

Alle gegevens die binnen het Internet Research and Investigation Network (iRN) worden verzameld, worden op dit moment onversleuteld opgeslagen. Hierdoor is de vertrouwelijkheid van de opgeslagen gegevens niet in alle omstandigheden gegarandeerd. Doordat dit network door overheidsdiensten zoals de Nederlandse Politie en de Belangstingdienst voor opsporing en onderzoek wordt gebruikt, is de vertrouwelijkheid van deze gegevens echter erg belangrijk en door de actuele manier van data opslag dus de privacy van de verdachten bedreigt.

In deze scriptie zal ik een methode voorstellen om de gegevens die in het iRN worden opgeslagen te beschermen om de privacy van de verdachten wel te kunnen garanderen. Hiervoor zal ik ook de voorwaarden noemen waaraan een dergelijk methode moet voldoen en laten zien, dat de van mij voorgestelde manier van data opslag inderdaad aan deze voorwaarden voldoet.

Inhoudsopgave

1	Inleiding	2
2	Voorkennis	4
2.1	Internet research and investigation network (iRN)	4
2.1.1	Geschiedenis, doel, werkwijze	4
2.1.2	Architectuur	5
2.2	Cryptografie	6
3	Onderzoek	8
3.1	Onderzoeksvraag	8
3.2	Aanvalscenarios	9
3.3	Voorwaarden	10
3.4	Notaties	10
3.4.1	Versleuteling	10
4	Gerelateerd onderzoek	12
5	Mogelijke Oplossing	14
5.1	Algemene beschrijving	14
5.1.1	Backup van sleutels	15
5.2	Database	15
5.3	Operaties op het dataset	16
5.3.1	Lezen van dossiers	17
5.3.2	Bewerken van dossiers	17
5.3.3	Toevoegen van gebruikers aan een dossier	17
5.3.4	Toegang tot een dossier weigeren	18
5.4	Correctheid	18
5.5	Problemen	18
6	Conclusies	20

Hoofdstuk 1

Inleiding

Het Internet Research and Investigation Network (iRN) is een netwerk voor overheidsdiensten dat voor opsporing, onderzoek, toezicht en/of handhaving kan worden gebruikt. Dit netwerk bestaat inmiddels uit meer dan 2.200 computers, die op verschillende locaties bij overheidsdiensten binnen Nederland geplaatst zijn. Tot deze computers hebben op dit moment ongeveer 9.500 ambtenaren toegang. Dit houdt ambtenaren van verschillende instanties in, zoals politieagenten of medewerkers van de Belastingdienst. Het netwerk werd ingericht, omdat het internet een steeds vaker gebruikt medium is en internetverkeer inmiddels ook als bewijsmateriaal bij strafzaken gebruikt kan worden. Omdat de IP-adressen van de computers binnen dit netwerk anoniem zijn, kunnen de ambtenaren het netwerk gebruiken om het internet doelgericht en zonder herkend te worden te doorzoeken. Hierbij wordt het hele internetverkeer opgeslagen en kan later voor de rechter als bewijsmateriaal worden ingezet. [1]

Een groot probleem bij het opslaan van dit internetverkeer is, dat de vertrouwelijkheid van de opgeslagen gegevens gegarandeerd moet worden om de privacy van de verdachten niet te bedreigen. Omdat het om bewijsmateriaal van strafzaken gaat en de privacy van de verdachten beschermd moet worden, mag niet zo maar iedereen toegang hebben tot deze gegevens. Hierbij is belangrijk dat het opgeslagen internetverkeer op zich geen gevoelige gegevens bevat, want alle gegevens die met behulp van het iRN verzameld kunnen worden, staan in principe openbaar en dus voor iedereen toegankelijk in het internet. De reden ervoor dat deze gegevens echter wel gevoelig zijn is, dat deze vaak bepaalde (groepen van) personen betrekken en de combinatie van de opgeslagen gegevens zowel als het feit dat de overheid naar gegevens over deze personen rechercheert, wel gevoelige informatie bevatten. Omdat het bij de opgeslagen data dus om gevoelige informatie gaat, mag niet iedereen die toegang tot het netwerk heeft tegelijkertijd toegang tot alle opgeslagen gegevens hebben. Er mag steeds maar diegene toegang tot de gegevens van een zaak hebben, die ook zelf met het verkennen van die zaak

bezig is.

Op dit moment worden alle gegevens binnen het iRN onversleuteld opgeslagen en kan dus iedereen met toegang tot de database alle opgeslagen gegevens lezen. Hierdoor is de privacy van de verdachten bedreigd. Er moet dus onderzocht worden, welke mogelijkheden er zijn om een nieuwe methode van data opslag te construeren, waarmee de gegevens opgeslagen kunnen worden, zonder dat gevoelige informatie kan lekken. Een voor de hand liggende mogelijkheid om dit voor elkaar te krijgen is gebruik te maken van cryptographie en de opgeslagen gegevens dusdanig te versleutelen, dat alleen maar bevoegde ambtenaren de bij een zaak horende gegevens kunnen inzien.

In deze scriptie zal ik onderzoeken aan welke voorwaarden een nieuw systeem voor de data opslag moet voldoen om de vertrouwelijkheid van de opgeslagen gegevens te garanderen, zonder de werkwijze van het iRN te beperken. Vervolgens zal een methode worden beschreven die gebruik maakt van cryptographie en sleutelbeheer om ervoor te zorgen dat niemand in staat is om gegevens te achterhalen die niet voor hem bedoeld zijn. Tot slot zal ik laten zien, dat het door mij voorgestelde systeem inderdaad aan alle opgestelde voorwaarden voldoet.

Hoofdstuk 2

Voorkennis

2.1 Internet research and investigation network (iRN)

2.1.1 Geschiedenis, doel, werkwijze

In 2003 begon de ontwikkeling van het iRN als project van de politie regio Gelderland-Zuid om opsporing en onderzoek voor verschillende overheidsinstanties in het internet mogelijk te maken. Dit netwerk kan worden gebruikt om internetverkeer te verzamelen en dusdanig op te slaan, dat deze gegevens achteraf op een forensische manier gebruikt kunnen worden. Het is een afgesloten netwerk waarin zich inmiddels ongeveer 2.200 computers bevinden die bij verschillende overheidsinstanties op meer dan 460 locaties zijn geplaatst. Op deze computers hebben ondertussen ongeveer 9.500 mensen toegang. Tegenwoordig wordt het iRN niet alleen voor opsporing en onderzoek gebruikt, maar ook voor toezicht en handhaving. [1]

Het hoofddoel van het iRN is om het internetverkeer van de computers binnen het netwerk op te slaan zodat het achteraf voor de rechter als bewijsmateriaal kan worden gebruikt. Hierbij is heel belangrijk, dat zowel de integriteit als de authenticiteit van de data gegarandeerd worden, omdat de rechter anders niet ervan uit kan gaan, dat de voorgedragen gegevens ook kloppen. Iedere ambtenaar met toegang tot het netwerk kan zelf dossiers aanmaken en vervolgens data hierin verzamelen. Het verzamelen van data kan daarbij op twee manieren worden gedaan:

1. **Handmatig:** De gebruiker kan zelf het internet gaan doorzoeken naar gegevens over de zaak die gekoppeld is aan het dossier.
2. **Automatisch:** De gebruiker kan zoekwoorden specificeren, waardoor het internet automatisch naar bij de zoektermen passende gegevens wordt doorzocht. Hiervoor worden niet alleen websites gecheckt, maar ook diverse feeds, van bijvoorbeeld twitter of facebook, in real time opgeslagen. Dit automatische systeem is als het project iColumbo bekend.

Het is ook mogelijk om gegevens weer uit een dossier te verwijderen, mocht achteraf blijken dat deze niet van belang zijn voor de bij het dossier horende zaak.

De dossiers kunnen met alle gebruikers van het netwerk worden gedeeld, waardoor er ook meerdere mensen aan dezelfde zaak kunnen werken. Hierbij blijft altijd diegene die een dossier heeft aangemaakt verantwoordelijk voor de gegevens die in dit dossier zijn opgeslagen. Daarom is deze persoon ook ervoor verantwoordelijk met wie een dossier wordt gedeeld.

Dossiers kunnen niet alleen worden gebruikt om data erin op te slaan, maar ook om de verzamelde data achteraf te bekijken. Op deze manier kunnen de gegevens als bewijsmateriaal bij strafzaken worden gebruikt.

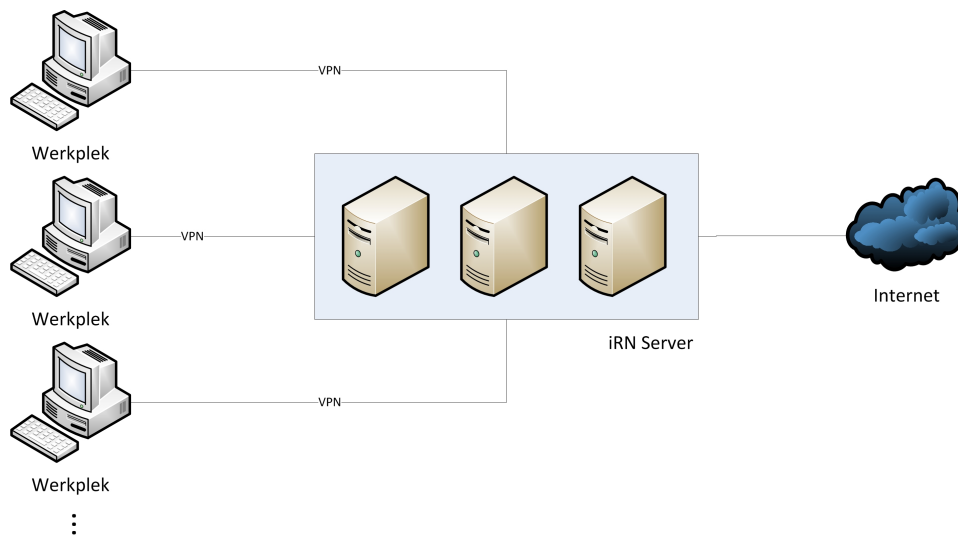
2.1.2 Architectuur

Het iRN is opgebouwd in een client/server architectuur. De 2.200 computers binnen het iRN zijn hierbij de clients die door de ambtenaren worden gebruikt om gegevens te verzamelen en de servers worden gebruikt om het internetverkeer van de clients op te slaan. De computers binnen het iRN zijn verspreid op meer dan 460 verschillende bureaus van overheidsdiensten in heel Nederland. Al deze computers zijn via een beschermde virtual private network (VPN) verbinding gekoppeld aan de servers van het iRN. Op deze manier kunnen de servers als een soort proxy worden gebruikt, waardoor het hele internetverkeer van de clients via de servers wordt omgeleid en hier kan worden opgeslagen. Een SSL versleuteling tussen de servers en de clients zorgt ervoor, dat het dataverkeer tussen deze niet uitgelezen of onderschept kan worden.

Het opslaan van internetverkeer is het enige doel van de werkplekken van het iRN en het is daarom niet toegestaan om vanuit deze computers een verbinding met het internet op te bouwen, zonder dat deze verbinding via de servers van het iRN wordt omgeleid. Dit is een veiligheidsmaatregel die eraan bijdraagt, dat er geen kwaadwillige software op de computers komt te staan. Dit kan worden gecontroleerd, omdat men precies weet waar de clients staan en er een fysieke controle over deze bestaat. De software op de servers dwingt verder af, dat geen verbinding met het openbare internet kan worden opgebouwd, zonder dat een dossier is geopend waaraan het internetverkeer wordt toegevoegd. Het is dus zelfs via de servers van het iRN niet mogelijk om de clients voor iets anders dan de werkzaamheden waarvoor deze bedoeld zijn te gebruiken.

Alle binnen het netwerk gebruikte technologieën en software oplossingen zijn open source, wat de werkwijze van het systeem voor iedereen transparant en toetsbaar maakt.

Figuur 2.1 geeft een overzicht van de gebruikte infrastructuur weer. Deze figuur beschrijft de verbinding tussen de clients en servers en de verbinding van de servers met het internet.



Figuur 2.1: Infrastructuur van het iRN

2.2 Cryptografie

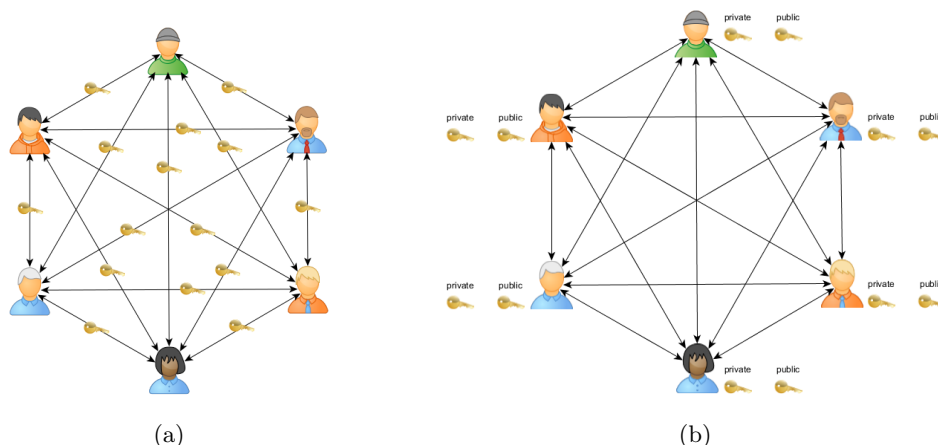
Cryptografie kan worden gebruikt om geheimen uit te wisselen en op te slaan, zonder dat onbevoegden deze geheimen kunnen lezen. Op deze manier kunnen cryptografische methoden worden toegepast om verschillende veiligheidsdoelen, zoals confidentialiteit, integriteit en/of authenticiteit te bereiken [2]. Het veiligheidsdoel dat in deze bachelorscriptie het meest aan bod zal komen is de confidentialiteit van de opgeslagen gegevens. Dit betekent, dat niemand de data mag lezen die hier geen permissie voor heeft.

Cryptografische algoritmen kunnen worden opgesplitst in twee categorieën, symmetrische en asymmetrische (public key) algoritmen. Het verschil tussen deze twee soorten van algoritmen is, dat in symmetrische algoritmen dezelfde sleutels voor ver- en ontsleuteling worden gebruikt, terwijl in asymmetrische algoritmen altijd twee verschillende sleutels voor encryptie en decryptie worden gebruikt. Het is ook mogelijk dat in symmetrische algoritmen verschillende sleutels worden gebruikt voor deze twee taken, maar in dat geval is de ene sleutel makkelijk te berekenen op basis van de andere. In asymmetrische algoritmen is het echter nooit mogelijk om de sleutel voor de ontsleuteling van de gegevens uit de sleutel voor de versleuteling te achterhalen of andersom [2, 3].

Het idee bij de communicatie met symmetrische cryptografie algoritmen is, dat alleen de communicerende partijen de gemeenschappelijke sleutel kennen en dus de berichten kunnen ontsleutelen. Het idee achter public key cryptografie is echter, dat iedere deelnemer een eigen key paar heeft waarvan de ene sleutel voor de versleuteling (public key) en de andere voor de ontsleuteling (private key) van de berichten wordt gebruikt. De private key

is geheim en alleen in het bezit van de ontvanger van een bericht. Op die manier kan iedereen een bericht voor deze persoon versleutelen door gebruik te maken van zijn public key, maar niemand behalve de ontvanger zelf dit bericht weer ontsleutelen.

Beide technieken hebben hun eigen voor- en nadelen. Als er in een groep van k mensen ieder tweetal veilig wil communiceren, zo zijn er met een symmetrische crypto-algoritme $k \cdot (k-1)/2$ verschillende sleutels nodig. Dit is een sleutel per verbinding. Met een asymmetrische crypto-algoritme zijn dit er maar k privé-/public-key paren, dus 2 sleutels per deelnemer. Voor de veilige communicatie binnen een groep van 10 mensen zijn er dus $10 \cdot (10-1)/2 = 45$ verschillende sleutels nodig als men gebruik maakt van een symmetrische algoritme en maar 10 verschillende sleutel paren als men gebruik maakt van een asymmetrische algoritme. Dit verschil wordt in figuur 2.2 verduidelijkt. Echter zijn de gebruikte sleutels voor asymmetrische algoritmen veel groter dan die voor symmetrische algoritmen. De sleutels die voor een symmetrische versleuteling worden gebruikt zijn typisch 256-bit, terwijl de sleutels van asymmetrische algoritmen typisch 2048-bits groot zijn. Ook al bestaan er asymmetrische algoritmen diens keys minder groot kunnen zijn, zoals elliptic curve cryptografie, zo is de benodigde tijd voor ver- en ontsleuteling veel langer als men gebruikt maakt van asymmetrische algoritmen [2].



Figuur 2.2: Illustratie van de hoeveelheden van benodigde sleutels voor een vertrouwelijke communicatie tussen 6 verschillende gebruikers met symmetrische (a) en asymmetrische (b) versleuteling.

Om de nadelen van beide technieken te omzeilen wordt in de praktijk vaak een combinatie van beide technieken toegepast. Hierbij wordt eerst asymmetrische cryptografie gebruikt om een symmetrische sleutel af te spreken en het echte bericht achteraf met de afgesproken sleutel symmetrisch versleuteld verstuurt [2]. De afgesproken sleutel wordt hierbij meestal voor iedere sessie vernieuwd om de veiligheid te verhogen.

Hoofdstuk 3

Onderzoek

In dit hoofdstuk ga ik mijn onderzoek over de vertrouwelijkheid van opgeslagen data in het iRN verder toelichten. Dit omvat zowel de onderzoeksvraag, als de voorwaarden waaraan een vertrouwelijk systeem moet voldoen en aanvalscenarios, waarmee kwaadwilligen zouden kunnen proberen om het systeem te omzeilen om aan gevoelige informatie te komen.

3.1 Onderzoeksvraag

De onderzoeksvraag die ik met deze scriptie wil beantwoorden is:

Hoe kunnen de gegevens in een dossier in het iRN op een vertrouwelijke manier worden opgeslagen, zodat alleen diegenen erbij kunnen die aan de bijhorende zaak werken?

Om deze vraag te kunnen beantwoorden moeten de antwoorden op de volgende deelvragen worden beantwoord:

- Welke versleutelingsalgoritmen zijn er al en hoe kunnen deze worden ingezet?
- Wat zijn de eigenschappen van deze algoritmen?
- Kan uit de bestaande algoritmen een nieuwe algoritme met de gewenste eigenschappen worden afgeleid of samengesteld?
- Op welke manieren zouden kwaadwilligen kunnen proberen om het systeem te omzeilen, oftewel wat zijn de aanvallersmodellen?
- Hoe kan de beschikbaarheid van de gegevens worden garandeerd, ook bij verlies van de sleutel of vertrek van een agent?

Deze vragen moeten worden beantwoord, omdat het iRN op dit moment nog geen mechanisme bevat om de vertrouwelijkheid van de opgeslagen gegevens

te garanderen. Er bestaan wel mechanismes, die de integriteit en authenticiteit van de gegevens toetsbaar maken, maar er werd nog geen aandacht aan privacy besteed. Op dit moment is het verzamelde internetverkeer als onvercijferd text opgeslagen en kan dus iedereen die op een of ander manier toegang tot de database verkrijgt alle dossiers lezen. Dit is een ongewenste eigenschap omdat kwaadwilligen op deze manier heel veel gevoelige gegevens kunnen stelen. Het antwoord op de onderzoeksvraag zou dus een bijdrage leveren om het iRN veiliger te maken door te garanderen dat de opgeslagen gegevens vertrouwelijk zijn en niemand de gegevens kan lezen zonder dat deze persoon hiervoor de toestemming van de eigenaar van het dossier heeft.

3.2 Aanvalscenarios

Er zijn enkele aanvalscenarios denkbaar tegen welke het systeem beschermd moet worden. De belangrijkste eigenschap van het systeem is, dat niemand in staat mag zijn om informatie te verkrijgen die niet voor hem bedoeld is, dus het garanderen van de vertrouwelijkheid van de in de dossiers opgeslagen data. Dit geldt zowel voor medewerkers die toegang tot het netwerk, als ook voor externe partijen die mogelijk proberen het systeem te kraken om aan informatie te komen. Medewerkers van het iRN kunnen verder worden onderverdeeld in ambtenaren die toegang tot het netwerk hebben en systeembeheerders die bovendien ook nog toegang tot de database hebben. Er moeten zowel de technische aspecten, zoals de algemene veiligheid van het systeem, als ook de menselijke aspecten, zoals het doorgeven van informatie aan derden, worden bekeken. De twee manieren waarop data kan lekken zijn dus:

- **Data lekken in de database:** De database, de hierin opgeslagen gegevens en de sleutels kunnen lekken. Daarom moet ervoor worden gezorgd, dat een aanvaller de opgeslagen gegevens niet kan ontsleutelen, zelfs als deze de hele database heeft. Om dit te kunnen garanderen zou een niet-centrale opslag kunnen worden gebruikt, bijvoorbeeld door bepaalde sleutels voor de ontsleuteling niet in de database zelf maar op een andere plek op te slaan. Hierdoor zou het systeem ook meteen zijn beschermd tegen aanvallen vanuit de systeembeheerders, die toegang tot de database hebben.
- **Data lekken door medewerkers:** Mensen zijn altijd een groot veiligheidsrisico. Op het moment dat iemand toegang tot data heeft en deze data heeft gezien zou hij deze data zonder problemen door kunnen geven aan derden, mocht hij dat willen. Men zou dus iedere medewerker moeten vertrouwen, dat deze geen data doorgeeft, maar hoe kan men weten wie men kan vertrouwen? Dit is een niet te beantwoorden

vraag en juist in een netwerk waar bijna 10.000 mensen betrokken zijn onmogelijk te controleren. Daarom moet ervoor worden gezorgd, dat de rollen van de ambtenaren duidelijk van elkaar zijn gescheiden, zodat zelfs als dit gebeurt niemand méér gegevens kan verzamelen om door te geven dan die hij voor zijn eigen werk echt nodig heeft, om op die manier de schade die een onbetrouwbare medewerker kan veroorzaken te beperken. Verder mag geen enkele persoon toegang tot alle opgeslagen gegevens hebben, omdat dit een te groot risico zou zijn voor de vertrouwelijkheid van de opgeslagen data. Als er een enkele persoon toegang tot alle gegevens heeft, kan deze persoon deze gegevens gewild of ongewild doorgeven en kan een aanvaller op die manier alle informatie uit het netwerk verkrijgen.

3.3 Voorwaarden

Het systeem moet aan meerdere eigenschappen voldoen:

- Iedere agent heeft alleen maar toegang tot de data van de zaken waar hij zichzelf mee bezighoudt.
- Er zijn mogelijk meerdere personen die toegang tot dezelfde gegevens hebben.
- Als er een nieuwe persoon begint aan een zaak te werken, moet deze toegang tot de bestaande gegevens kunnen krijgen.
- Permissies om gegevens te lezen moeten makkelijk kunnen worden teruggetrokken.
- De gegevens over een zaak moeten altijd beschikbaar blijven, ook als een agent weggaat of zijn sleutel verliest.

3.4 Notaties

In deze scriptie zullen enkele notaties worden gebruikt om verschillende dingen af te korten. Deze notaties worden in deze sectie toegelicht en zijn noodzakelijk om de rest van de scriptie goed te kunnen begrijpen.

3.4.1 Versleuteling

Zoals al aangeduid sectie 2.2 wordt er verschil gemaakt tussen symmetrische en asymmetrische cryptografie. Dit verschil is ook in de notaties te herkennen.

Symmetrische versleuteling

Symmetrisch versleutelde berichten zullen met $K\{\text{bericht}\}$ worden genoteerd, waarbij K de symmetrische sleutel voor de encryptie is. Omdat ontsleuteling in symmetrische algoritmen op dezelfde manier werkt als versleuteling, maar de sleutel kan verschillen, wordt de ontsleuteling met de inverse van de sleutel K genoteerd. Deze inverse K^{-1} is de symmetrische sleutel voor de decryptie. De ontsleuteling van een met een symmetrisch algoritme versleuteld bericht zal dus met $K^{-1}\{K\{\text{bericht}\}\}$ worden genoteerd.

Asymmetrische versleuteling

Asymmetrisch versleutelde berichten zullen daarentegen met $\{\text{bericht}\}_{K_e}$ worden genoteerd. Hierbij is K_e de publieke sleutel die voor de versleuteling werd gebruikt. De ontsleuteling van een met een asymmetrische algoritme versleuteld bericht zal ik met $[\{\text{bericht}\}_{K_e}]_{K_d}$ noteren, waarbij K_d de private key is. Hierbij worden niet alleen maar verschillende sleutels voor ver- en ontsleuteling gebruikt, maar ook verschillende haakjes in de notatie. Dit verschil is er om ook in de notatie duidelijk te maken, dat de decryptie in asymmetrische algoritmen fundamenteel anders werkt dan de encryptie.

Hoofdstuk 4

Gerelateerd onderzoek

Op het onderwerp van deze scriptie, dus het toegepaste sleutelbeheer om de vertrouwelijkheid van opgeslagen gegevens te garanderen, is er naar mijn weten geen wetenschappelijk onderzoek gedaan. Sleutelbeheer en encryptie van vertrouwelijke gegevens wordt meestal ad hoc gedaan en hoewel er veel technieken bekend zijn die in verschillende blog-posts en dergelijke op het internet worden uitgelegd, ontbreekt er een wetenschappelijke onderbouwing voor de correctheid van deze technieken. Al deze bekende technieken maken gebruik van cryptografie, waar wel veel onderzoek naar is gedaan. Door dit onderzoek zijn er verschillende cryptografische algoritmen met verschillende eigenschappen bekend, waarvan vervolgens gebruik gemaakt kan worden.

Er is echter wel onderzoek gedaan naar privacy issues en hun oplossingen in andere toepassingen die hieraan gerelateerd kunnen worden, zoals bijvoorbeeld cloud computing. Het doel van cloud computing is, alle gegevens en programma's van een gebruiker altijd toegankelijk te maken vanaf alle apparaten. Hiervoor worden zowel het opslaan van gegevens als het processen van deze gegevens niet op de lokale machine van de gebruiker gedaan, maar op de servers van externe partijen die diverse cloud computing diensten aanbieden. Een soortgelijke toepassing vinden we ook in het iRN, waarin het internetverkeer van de gebruiker wordt omgeleid via hun servers en hier opgeslagen en geprocessed wordt.

Een van de grootste problemen in de toepassing van cloud computing is het vertrouwen, dat de gebruiker in de aanbieder van de gebruikte dienst moet hebben [4]. De gebruiker moet erop vertrouwen, dat de aanbieder de gegevens en processen op een vertrouwelijke manier opslaan en uitvoeren. De onderzoekers Itani, Kayssi en Chehab [4] stellen daarom een third-party oplossing voor, waarbij zowel de klant als ook de aanbieder van de cloud computing service deze partij vertrouwd. Deze partij kan dan voor de goede encryptie van de gegevens zorgen en op die manier hun vertrouwelijkheid garanderen. De third-party partij biedt hierbij de privacy van de gebruiker als een service aan aan zijn klanten. Dit principe noemen ze privacy as

a service (PasS). Deze punt speelt in deze scriptie echter geen rol, omdat zowel clients als server door de medewerkers van het iRN worden beheert en er dus geen sprake kan zijn van niet vertrouwen in de eigenaar of beheerder van de servers. Echter moet er wel een manier worden gevonden om de data op de goede manier te beschermen.

Een ander gerelateerd onderwerp is het veilige communiceren binnen een groep van personen. In het iRN moeten meerdere mensen toegang tot bepaalde dossiers kunnen verkrijgen. Daarom is het belangrijk, dat de gegevens dusdanig versleuteld zijn, dat verschillende personen toegang hebben tot verschillende dossiers en er ook een mogelijkheid bestaat om dossiers voor meerdere personen tegelijk toegankelijk te maken. Deze groepen van gebruikers die toegang tot bepaalde gegevens hebben moeten makkelijk te bewerken zijn, zonder iedere keer dat er een verandering in een groep plaatsvindt een nieuwe sleutel afgesproken dient te worden en de data met de nieuwe sleutel herversleuteld moet worden. Herversleutelen van de data is geen gewenste oplossing omdat het opnieuw versleutelen van de gegevens een vrij dure operatie is en dit de servers onnodig zou belasten. Dit is vergelijkbaar met de encryptie van groep chats, waarbij de gebruikers binnen een groep allemaal toegang tot de verzonden berichten hebben, maar als er een nieuwe persoon bij komt, deze alleen maar toegang tot de nieuwe berichten mag hebben en zodra een persoon die groep verlaat mag deze persoon geen toegang meer hebben tot nieuwe berichten. Dit is vergelijkbaar met groepen van personen die toegang hebben tot een dossier en dit dossier mogen lezen. Het fundamentele verschil tussen deze twee onderwerpen is, dat de gegevens binnen het iRN in dossiers zijn opgeslagen en iemand met toegang tot het dossier het hele dossier mag bekijken totdat deze permissie weer wordt teruggetrokken. Kumari et al. [5] beschrijven in hun paper een manier om de encryptie van chat berichten binnen een groep mogelijk te maken, zonder data opnieuw te moeten versleutelen op het moment dat de groep veranderd. Hun basis idee komt erop neer om iedere keer dat de groep wijzigt ook de sleutel voor de encryptie te wijzigen en de nieuwe berichten met een andere sleutel te versleutelen. Dit is echter niet gewenst binnen het iRN, omdat de toegang binnen het iRN per dossier, vergelijkbaar met een gesprek, wordt toegelaten en niet per bericht zoals dit in chat beveiliging het geval is.

Hoofdstuk 5

Mogelijke Oplossing

In dit hoofdstuk ga ik een methode beschrijven, waarmee de gegevens op de servers van het iRN opgeslagen kunnen worden, zonder de privacy van de verdachten te bedreigen. Naast het beschrijven van deze oplossing, zal ik aantonen, dat deze manier van data opslag aan alle in sectie 3.3 genoemde voorwaarden voldoet. Verder zal ik op enkele veiligheidsrisico's ingaan die ook met deze oplossing nog bestaan en voorstellen maken om (een deel van) deze problemen op te lossen.

5.1 Algemene beschrijving

Een voor de hand liggende mogelijkheid de vertrouwelijkheid van de opgeslagen gegevens te garanderen is, de gegevens per dossier te versleutelen en ervoor te zorgen, dat alleen personen die de permissie hebben om een dossier te lezen en te bewerken in staat zijn om de data op de juiste manier te ont-respectievelijk versleutelen. Dit komt erop neer, de sleutels voor de dossiers op een manier te verspreiden, dat alleen maar bevoegde personen de sleutel kunnen verkrijgen om de data te lezen of te bewerken.

Om de data te versleutelen, kan voor ieder dossier een unieke, symmetrische key aan worden gemaakt. Alle gegevens binnen een dossier worden vervolgens met de bijhorende symmetrische sleutel versleuteld. Deze symmetrische sleutel zal in het vervolg K_{D_i} worden genoemd, waarbij i een identifier voor het dossier is. Een symmetrische algoritme is het meest geschikt voor de versleuteling van de data binnen het iRN omdat het om grote hoeveelheden data gaat en deze efficiënter geïmplementeerd kunnen worden dan asymmetrische algoritmen [2].

Voor de confidentiele verspreiding van de sleutels heeft ieder ambtenaar een eigen public/private keypaar. De public key van een gebruiker zal $K_{U_i,e}$ worden genoemd en de private key zal als $K_{U_i,d}$ worden genoteerd. Voor beide sleutels geldt, dat i een identifier voor hun eigenaar is. Om een ambtenaar g toegang tot de sleutel van een dossier d en daarmee toegang tot

het bijhorende dossier te geven, wordt de sleutel van het dossier K_{D_d} met de public key van deze ambtenaar $K_{U_{ge}}$ versleuteld en de resulterende string $\{K_{D_d}\}_{K_{U_{ge}}}$ vervolgens in de database opgeslagen.

De gebruiker kan dan zijn private key $K_{U_{gd}}$ gebruiken om de sleutel van het dossier te ontsleutelen en is vervolgens in staat om nieuwe gegevens aan het dossier toe te voegen of de inverse van de key $K_{D_d}^{-1}$ te berekenen en te gebruiken om de opgeslagen gegevens te lezen.

Welke algoritmes precies gebruikt worden om de dossiers en de sleutels ervan te versleutelen is niet van belang, als deze maar veilig zijn en de versleuteling niet efficiënt gekraakt kan worden. Mogelijkheden voor algoritmes die gebruikt kunnen worden zijn bijvoorbeeld de Advanced Encryption Standard (AES) voor de symmetrische versleuteling van de dossiers en RSA of El Gamal voor de versleuteling van de dossiersleutels.

5.1.1 Backup van sleutels

Om geen gegevens kwijt te raken als een agent uit dienst gaat of zijn private sleutel verliest, moet er een backup worden ingevoerd, waarmee alle gegevens oftewel hun sleutels ook zonder de private keys van de agenten achterhaald kunnen worden. Er mag echter geen master key in het systeem zitten, waarmee alle gegevens ontsleuteld kunnen worden omdat dit een te groot risico zou zijn voor de vertrouwelijkheid van de opgeslagen data. Verder mag er geen enkele persoon toegang tot alle gegevens verkrijgen omdat deze dan ook een soort master key zou hebben.

Om dit voor elkaar te krijgen kan secret sharing worden gebruikt. Een mogelijke uitwerking kan gebruik maken van het systeem van Pedersen [6]. Hij stelt in zijn artikel een manier voor om een geheim te distribueren over n personen, zodat minimaal k personen nodig zijn om dit geheim te ontsleutelen. Hiervan kan ook in deze context gebruik gemaakt worden door iedere dossiersleutel nog een keer met dit schema te versleutelen en op te slaan. Daardoor zijn de opgeslagen gegevens nog steeds leesbaar als een sleutel verloren gaat, omdat k van deze n personen de dossiersleutel samen kunnen ontsleutelen. Welke aantallen voor k en n nodig zijn om de data zo goed mogelijk te beschermen zonder het systeem onbruikbaar te maken moet in verder onderzoek worden onderzocht.

5.2 Database

De database zou in dit systeem uit vier hoofdcomponenten bestaan. Deze componenten zijn:

1. **Dossiers:** In deze tabel zijn alle dossiers opgeslagen inclusief de bijhorende data. Hierbij is alle data versleuteld met een symmetrische key K_{D_i} , waarbij i een identifier van het dossier representeert.

2. **Gebruikers:** In deze tabel zijn alle gebruikers opgeslagen met hun public keys $K_{U_{ie}}$, waarbij i een identifier voor de gebruiker is.
3. **Toegangsrechten:** In deze tabel is opgeslagen welke gebruiker toegang tot welke dossiers heeft. Bij iedere record wordt ook de symmetrische sleutel K_{D_i} van het dossier versleuteld met de public key van de gebruiker $K_{U_{ie}}$ bijgehouden.
4. **Sleutelbackup:** Deze component bestaat echter uit twee tabellen. In een tabel zijn de met Pedersens [6] schema versleutelde dossiersleutels opgeslagen en in de andere is opgeslagen welke gebruikers in de groep zitten om één van deze sleutels te ontsleutelen.

Tabel 5.1 geeft de eerste drie componenten van deze database-indeling met een voorbeeldpopulatie weer. In dit voorbeeld zijn er twee dossiers en twee gebruikers. De gebruiker met id 1 heeft toegang tot beide dossiers, terwijl de andere gebruiker slechts toegang tot het dossier met id 1 heeft.

Tabel 5.1: Database indeling met voorbeeld populatie

(a) Dossiers		(b) Gebruikers	
dossier_id	data	gebruiker_id	public key
1	$K_{D_1}\{\text{data}\}$	1	$K_{U_{e_1}}$
2	$K_{D_2}\{\text{data}\}$	2	$K_{U_{e_2}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

(c) Toegang		
dossier_id	gebruiker_id	key
1	1	$\{K_{D_1}\}_{K_{U_{e_1}}}$
1	2	$\{K_{D_1}\}_{K_{U_{e_2}}}$
2	1	$\{K_{D_2}\}_{K_{U_{e_1}}}$
$\vdots$	$\vdots$	$\vdots$

5.3 Operaties op het dataset

In deze sectie ga ik de belangrijkste operaties op het dataset nader toelichten. Hierbij is U een willekeurige gebruiker en D een dossier waarop U toegang heeft.

5.3.1 Lezen van dossiers

Mocht een gebruiker nu de data van een dossier waar hij toegang tot heeft willen lezen, wordt de bijhorende record in de tabel met toegangsrechten opgezocht en de symmetrische sleutel ontsleuteld met behulp van de private key K_{U_d} van de gebruiker. Nu is de symmetrische key van het dossier K_D bekend, omdat $[\{K_D\}_{K_{U_e}}]_{K_{U_d}} = K_D$ geldt. Omdat voor de versleuteling van de data een symmetrische algoritme werd gebruikt, kan de inverse K_D^{-1} van deze key worden berekend en vervolgens worden gebruikt om de data van het dossier te ontsleutelen omdat $K_D^{-1}\{K_D\{data\}\} = data$ geldt.

5.3.2 Bewerken van dossiers

Het bewerken van een dossier, dus het toevoegen of verwijderen van gegevens kan op een soortgelijke manier als het lezen van de opgeslagen data. Doordat de bevoegde ambtenaar met behulp van zijn eigen private key de symmetrische sleutel K_D van het dossier kan achterhalen en zijn inverse K_D^{-1} kan berekenen, is hij in staat om de opgeslagen data op de zelfde manier te ontsleutelen als voor het lezen van de data. Vervolgens kan de data worden bewerkt en achteraf opnieuw worden versleuteld met de sleutel van het dossier K_D . Tot slot wordt de nieuwe versleutelde data $K_D\{data\}$ weer in de database opgeslagen.

5.3.3 Toevoegen van gebruikers aan een dossier

Er is een restrictie op het toevoegen van personen aan een dossier, waardoor alleen diegenen gebruikers die al toegang tot een dossier hebben gebruikers aan dit dossier toe kunnen voegen. Deze restrictie werd geïntroduceerd om te voorkomen dat gebruikers die andere gebruikers aan een dossier toe kunnen voegen in staat zijn om zichzelf aan alle dossiers toe te voegen. Zonder deze restrictie zou het dus mogelijk zijn, dat enkele personen alle gegevens zouden kunnen verkrijgen. Verder is altijd diegene die een dossier heeft aangemaakt verantwoordelijk voor alle gegevens binnen dit dossier en dus ook verantwoordelijk ervoor te bepalen wie toegang tot deze gegevens mag hebben. Om een gebruiker aan een dossier toe te voegen moet er alleen een nieuwe record in de permissie tabel komen te staan. Voor deze record is de juist versleutelde key nodig, dat is de sleutel van het dossier versleuteld met de public key van de nieuwe medewerker. Omdat alleen diegenen gebruikers nieuwe gebruikers aan een dossier toe kunnen voegen, die al toegang tot het dossier hebben, kunnen deze de sleutel van het dossier door ontsleutelen van hun eigen permissie achterhalen en vervolgens versleutelen met de public key van de nieuwe gebruiker. Het resultaat is dan de key die voor het nieuwe toegangsrecht opgeslagen moet worden.

5.3.4 Toegang tot een dossier weigeren

Ook het weigeren van toegang is heel makkelijk door de record in de permissie tabel te wissen. Dit is voldoende, omdat alle versleutelingen op de server van het iRN plaatsvinden en de gebruiker de symmetrische sleutel van het dossier dus nooit in handen krijgt. Zodra de record uit de database is gewist, bestaat er geen mogelijkheid meer voor de gebruiker om de opgeslagen data te lezen.

5.4 Correctheid

De privacy-gevoelige gegevens zijn allemaal versleuteld opgeslagen en alleen bevoegde ambtenaren zijn in staat om deze te ontsleutelen, omdat alleen deze de keys van de dossiers zoals deze in de permissie-tabel zijn opgeslagen kunnen ontsleutelen. Dit omdat hiervoor de private key van de ambtenaar nodig is en deze als het goed is niemand anders heeft. Doordat de sleutels van de dossiers nergens anders zijn opgeslagen bestaat er geen andere mogelijkheid voor kwaadwilligen om aan deze sleutels te komen.

Doordat alle ver- en ontsleutelingen van de dossier keys op de server plaats vinden, krijgt de gebruiker de keys nooit in handen, waardoor de vertrouwelijkheid van de gegevens ook is gegarandeert in het geval dat (delen van) de database lekken omdat de gegevens binnen de database zonder de sleutels niet ontsleuteld kunnen worden.

Ook het backup-mechanisme zorgt ervoor, dat de opgeslagen gegevens vertrouwelijk blijven. Hoewel er daardoor een manier bestaat om de opgeslagen data te ontsleutelen, zonder dat de agent in zijn eentje hiervoor een permissie heeft, is de vertrouwelijk hierdoor niet bedreigt. Dit komt, omdat meerdere agenten samen moeten werken om de dossiersleutel te kunnen achterhalen en het risico van misbruik daardoor verkleind wordt, omdat de agenten elkaar controleren. Verder moeten de personen hiervoor niet voor iedere dossier dezelfde zijn, zodat zelfs als sommige medewerkers proberen het systeem te onderscheppen, maar enkele dossiers gelezen kunnen worden en nooit alle dossiers tegelijk.

5.5 Problemen

Een probleem met deze aanpak is, dat de sleutel van het dossier tijdens het lezen of bewerken van dit dossier onversleuteld in het werkgeheugen van de server zit en een kwaadwilliger op die manier in staat zou kunnen zijn om tijdelijk de sleutel van een dossier te verkrijgen mocht hij het voor elkaar krijgen om kwaadwillige software op de servers te installeren. Een hacker kan met deze architectuur echter altijd alleen maar de keys van actieve dossiers stelen en zal nooit in staat zijn om data uit dossiers te lezen die niet actief

werden gebruikt tijdens zijn aanval. Daardoor wordt de impact van een dergelijke aanval verkleind.

Een andere veiligheidsrisico is, dat de gebruikers hun login gegevens aan derden kunnen geven, waardoor deze toegang tot het systeem en een deel van de opgeslagen gegevens kunnen verkrijgen zonder dat dit gewenst is. Om dit probleem te voorkomen zouden smartcards kunnen worden gebruikt om de gebruiker bij het netwerk te authenticeren. Een groot voordeel hiervan is, dat er een fysieke controle bestaat hoe veel smartcards er zijn en wie deze heeft. Daardoor kan de toegang tot bepaalde gegevens niet zo maar aan derden worden gegeven, zonder de smartcard weg te geven. Gewone sleutels kunnen daarentegen wel worden doorgegeven, zonder dat men zelf de sleutel verliest. De smartcards kunnen daarnaast worden gebruikt om de private key van de ambtenaar op te slaan, zodat deze hem niet hoeft te onthouden. Hiermee zou ook de authenticatie van gebruikers meteen een stuk veiliger worden. Dit is op dit moment nog een van de zwakke punten van het iRN volgens onderzoekers van de Radboud Universiteit [7]. Om misbruik van gestolen smartcards te voorkomen kunnen deze met een PIN worden beveiligd. Verder zouden smartcards geblokkeerd kunnen worden op het moment dat een agent geen toegang meer tot het netwerk mag hebben of diefstal of misbruik van een smart card bekend wordt. Op die manier kunnen ook aanvallen van medewerkers enigzins makkelijk beperkt worden.

Het gebruik van smart cards heeft in deze context echter ook nadelen. Zo kan bijvoorbeeld wel de dossiersleutel op de smartcard zelf worden ontsleuteld, maar om meteen het hele dossier op deze card te ontsleutelen is de rekenkracht niet voldoende. Daarom zou er een interface tussen de smart card en de client moeten worden gebouwd, zodat de ontsleuteling van de gegevens in het dossier vervolgens of op de client of op de server kan plaatsvinden. Hiervoor moeten de dossiersleutels echter gecommuniceerd worden via het interface, waardoor een kwaadwillige gebruiker van het systeem deze sleutels zou kunnen af luisteren.

Een laatste probleem met deze aanpak kan zijn, dat er op die manier heel veel verschillende sleutels in het systeem zitten die allemaal beheerd moeten worden (één sleutel per dossier), waarbij deze ook weer meerdere keren opgeslagen moeten worden (één keer per agent die toegang tot het dossier heeft). Dit is echter geen heel groot probleem omdat moderne databases dit aan moeten kunnen.

Hoofdstuk 6

Conclusies

Uit mijn onderzoek blijkt, dat er goede mogelijkheden bestaan om de opgeslagen gegevens binnen het iRN op een privacy-vriendelijke manier op te slaan, zodat geen onbevoegden toegang tot deze gegevens kunnen verkrijgen. Hoewel de beschreven oplossing nog wat zwakke punten bevat, kunnen deze met verder onderzoek eventueel opgelost worden. Er moet dus zeker nog wat wetenschappelijk onderzoek worden gedaan om de zwakke punten van mijn systeem op te lossen en door het gebrek aan wetenschappelijk onderzoek aan sleutelbeheer in het algemeen is dit zeker een goed plan. Volledige veiligheid bestaat natuurlijk niet, maar desondanks kan de voorgestelde manier van versleuteling van de gegevens voor meer vertrouwelijkheid van de gegevens zorgen dan dit nu het geval is.

Deze manier van data opslag daadwerkelijk in de architectuur van het iRN op te nemen is niet heel ingewikkeld, maar voegt wel een nieuwe dimensie van veiligheid toe aan het systeem. De implementatie moet echter heel goed worden getest voordat deze daadwerkelijk wordt gebruikt. Fouten in de implementatie van dit systeem kunnen bijvoorbeeld ertoe leiden, dat gegevens versleuteld worden en niet meer ontsleuteld kunnen worden, waardoor alle gegevens die op dit moment in het netwerk zijn opgeslagen kwijt raken.

Bibliografie

- [1] Nationale Politie. iRN, Mai 2014. <http://www.columbo.nl/>.
- [2] M. Goodrich and R. Tamassia. *Introduction to Computer Security (First Edition)*. Pearson Education Limited, 2014.
- [3] C. H. Meyer. Cryptography-a state of the art review. In *CompEuro '89., 'VLSI and Computer Peripherals. VLSI and Microelectronic Applications in Intelligent Peripherals and their Interconnection Networks'*, *Proceedings.*, pages 4/150–4/154, May 1989.
- [4] W. Itani, A. Kayssi, and A. Chehab. Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures. In *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC), 2009*, pages 711–716, Dec 2009.
- [5] V.V. Kumari, D. V. NagaRaju, K. Soumya, and K. V. S. V. N. Raju. Secure Group key Distribution Using Hybrid Cryptosystem. In *Second International Conference on Machine Learning and Computing (ICMLC), 2010*, pages 188–192, Feb 2010.
- [6] T. P. Pedersen. Non-Interactive and Information-Theoretic Secure Verifiable Secret Sharing. In J. Feigenbaum, editor, *Advances in Cryptology — CRYPTO '91*, volume 576 of *Lecture Notes in Computer Science*, pages 129–140. Springer Berlin Heidelberg, 1992.
- [7] E. Hubbers, B. Jacobs, and R. Wichers Schreur. Evaluatie Internet Recherche Netwerk (IRN). Juni 2012.