De mate van anonimiteit in mixnetwerken gebruikmakend van redundante berichten

Scriptie

Voor de opleiding

MSc. Software Engineering

Universiteit van Amsterdam

Afstudeerdocent: **Prof.Dr. D.J.N. van Eijck**

Uitgevoerd aan

Information Communication Theory Group

Supervisor: **Dr.Ir. J.C.A. van der Lubbe**

Begeleider: **Ir. K. Cartrysse**

Raimondo Faustinelli

studentnummer 0399809

Voorwoord

De 1-jarige masteropleiding Software Engineering, waaraan de Universiteit van Amsterdam (UvA), de Hogeschool van Amsterdam en de Vrije Universiteit samenwerken, wordt afgesloten met een drie maanden durende masterproject. De student heeft bij het vinden naar een afstudeerplek enigszins vertraging opgelopen, maar is door MSc. Marnix Dekker (werkzaam als PhD. TNO - Telecom, NIS Network and Information Security) uiteindelijk verwezen naar Dr. ir. Jan van der Lubbe. Hij doceert de vakken "Cryptography" en "Omgaan met onzekerheid" aan de TU Delft en leidt onderzoeken voor het thema "Security and Privacy", dat onderdeel is van de Information & Communication Theory Group. Naast de heer Van der Lubbe is Ir. Kathy Cartrysse als PhD werkzaam. Beide hebben in een ontmoeting met de student tezamen een project opgesteld met als titel "De mate van anonimiteit in mixnetwerken, gebruikmakend van nepberichten" en hebben de student de kans aangeboden om dit project ten uitvoer te brengen. Voor deze kans wil de student beiden danken.

Vanuit de UvA heeft de student zijn begeleider Prof. dr. Jan van Eijck toegewezen gekregen. Hij is werkzaam voor het Centrum voor Wiskunde en Informatica te Amsterdam als projectleider van Interactive Software Development and Renovation. De student wil hem danken voor zijn hulp en tips tijdens het masterproject en korte maar bijzondere lessen aan de UvA. Uiteraard gaat hierbij ook een dank uit naar alle andere docenten die colleges hebben gegeven en andere activiteiten hebben georganiseerd tijdens de opleiding aan de UvA.

Tijdens het masterproject heeft begeleidster Cartrysse enorm geholpen met de informatietheorie en de discussies rondom het masterproject. Het zoeken naar een maat van anonimiteit was geen eenvoudige zaak geworden zonder het (zelfstudie) boek "Informatietheorie", geschreven door o.a. de heer Van der Lubbe. De student wil hen beiden hiervoor nogmaals bijzonder danken.

Verder wil de student voor dit werk zijn ouders, Marian en Rino, zijn familie (met name Diny die op het moment van het schrijven van dit werk een moeilijke tijd doormaakt), en vrienden ontzettend bedanken voor hun steun, geloof en liefde.

Raimondo Marco Faustinelli
Amsterdam, augustus 2005

Samenvatting

De mix is een router dat tot doel heeft verkeersanalyse moeilijker te maken. Zij moet dus de relatie tussen haar input met haar output verborgen houden en realiseert dit door ontvangen data te vertragen, te vercijferen en tenslotte te mixen. Door het gebruik van de mix worden zowel zendersanonimiteit, het is onzeker wie de zender is van een gegeven bericht uit een groep mogelijke zenders, als ontvangersanonimiteit, de onzekerheid over wie de ontvanger is van een gegeven bericht uit een groep mogelijke ontvangers, gewaarborgd.

De onzekerheid uit de informatietheorie is een maat om de relatie tussen de input en de output van de mix te kwantificeren. Hoe groter de onzekerheid wordt, des te groter wordt de anonimiteit. Door normalisatie van de onzekerheid met de output of met de input, krijgen we onze maat respectievelijk de zendersanonimiteit en de ontvangersanonimiteit.

Nepberichten hebben als taak het werk van verkeersanalyse door een aanvaller nog moeilijker te maken en zijn nauwelijks te onderscheiden van echte berichten. De nepberichten die door de mixen zijn gegenereerd kunnen op twee manieren worden ingezet: nepberichten worden tijdens het versturen van haar berichten gemixt met nepberichten (genaamd Flush Nepberichten Methode, afgekort tot FNM) en nepberichten kunnen direct in de pool van de mix geplaatst worden (genaamd Pool Nepberichten Methode, afgekort tot PNM).

Voor het berekenen van de anonimiteit in een mixnetwerk wordt voor elke mix een deel van de totale anonimiteit berekend.

Inhoudsopgave

Hoofdstuk 1

Inleiding

Deze scriptie behandelt het masterproject "*De mate van anonimiteit in mixnetwerken gebruikmakend van redundante berichten*" die is uitgevoerd aan de TUDelft en in opdracht van de Universiteit van Amsterdam. We zullen onze maat voor de anonimiteit in een eigen model definiëren. Hierin speelt de mix een centrale rol. Een mix heeft als taak de relatie tussen ingaande en uitgaande berichten/datapackets te verbergen [Cha81]. Daardoor kan een mix voor anonieme communicatie zorgen. Een netwerk van meer dan één mix, wordt een mixnetwerk genoemd. Naast normale berichten worden ook nepberichten in een mixnetwerk geïntroduceerd. Deze berichten, die nauwelijks verschillen met normale berichten, dragen erbij toe dat een bepaald anonimiteitsniveau behaald wordt.

We praten over zendersanonimiteit wanneer we door een mix verzonden bericht in ons bezit hebben en we afvragen wie de zender van dat bericht is. Zo kunnen we ook praten over ontvangersanonimiteit wanneer we een bericht afkomstig van een zender in ons bezit hebben en we afvragen wie de ontvanger van dat bericht zal worden.

Dit masterproject is geïnspireerd door een open probleem uit [DP04b]: er is meer onderzoek nodig naar het effect van nepberichten in een mixnetwerk. Dit onderzoek is nodig om een beter beleid voor nepberichten te kunnen vinden.

1.1 Probleemdefinitie

Het Freedom-net[Dai96] is een mixnetwerk waarmee anoniem datapackets kunnen worden verstuurd. Om de anonimiteit te verhogen versturen mixen in het Freedom-net nepberichten naar elkaar. Daardoor wordt het analyseren van datacommunicatie door een buitenstaander (bijvoorbeeld een aanvaller) moeilijker gemaakt.

Met aanval[Dai96] op het PipeNet-protocol is geprobeerd om getransporteerde data van een zender te volgen door een verbinding van de aanvaller via een deel van het netwerk naar de aanvaller terug vol te laten stromen met eigen data tot de bandbreedte limiet bereikt wordt. Vervolgens wordt het nepberichtenverkeer stop gezet en kan een aanvaller de data van een willekeurige zender volgen. Alleen data die in het aangevallen gedeelte stroomt, kan door de aanvaller gevolgd worden.

De volgende vraag is te formuleren om een balans te kunnen vinden voor anonimiteit en het gebruik van nepberichten: welke maat voor een mix en voor een mixnetwerk kan gebruikt worden om de anonimiteit uit te drukken voor zowel met als zonder nepberichten gebruik.

1.2 Onderzoek

Om een ideaal beleid voor nepberichten generatie te vinden, wordt een onderzoek verricht waarin een viertal onderzoeksvragen beantwoord moeten worden. Het zoeken naar deze antwoorden leidt tot een beleid voor nepberichten. De vragen die gesteld zijn voor het onderzoek worden nu behandeld.

1.2.1 Onderzoeksvragen

De vier onderzoeksvragen voor dit onderzoek zijn:

1. *Hoe kunnen we de anonimiteit meten in een mixnetwerk, zonder nepberichten?* Bij deze vraag wordt gekeken welke participanten bestaan in een wereld waar de mix een centrale rol speelt[1]. We zoeken voor elke participant een overeenkomst met een functie of object in de informatietheorie. Het begrip en het kwantificeren van de anonimiteit worden voor deze vraag geïntroduceerd.

2. *Hoe kunnen we de anonimiteit meten in een mixnetwerk, met nepberichten?* Na de introductie van de maat anonimiteit, wordt bepaald hoe de anonimiteit berekend kan worden indien nepberichten worden gebruikt.

3. *Hoe kunnen we het anonimiteitsniveau zelf bepalen in een mixnetwerk, gebruikmakend van nepberichten?* Als we de anonimiteit kunnen berekenen met en zonder nepberichten, dan wordt de volgende stap: is het mogelijk voor een zender die anoniem een bericht wenst te versturen een bepaald anonimiteitsniveau in mixnetwerk garanderen? Dit houdt in dat een zender in staat moet zijn om zelf een anonimiteitsniveau te 'regelen', waarvoor (indien nodig) nepberichten verstuurd worden.

4. *Hoe kan een mixnode een bepaald anonimiteitsniveau garanderen, al dan niet gebruikmakend van nepberichten, waarbij het netwerk niet overbelast raakt?* Bij deze laatste vraag is de mixnode verantwoordelijk een bepaald anonimiteitsniveau te garanderen, waarvoor (indien noodzakelijk) nepberichten verstuurd moeten worden. Dit doet een mixnode op een dergelijke wijze dat het netwerk niet overbelast raakt.

Om deze vier onderzoeksvragen te kunnen beantwoorden, wordt in de volgende sectie uitgelegd hoe we dit gaan aanpakken. Overigens worden onderzoeksvragen nummer één en twee als minimum eis voor dit project gesteld en worden deze vragen in deze scriptie behandeld.

1.2.2 Aanpak

Het masterproject is gestart met een literatuuronderzoek en is het eerste onderdeel van de aanpak om de onderzoeksvragen te kunnen beantwoorden. De studie richt zich met name op het uitdrukken van de (digitale) anonimiteit met behulp van de informatietheorie. Hiermee wordt een model beschreven waarin de mix een centrale rol speelt en uiteindelijk de anonimiteit berekend kan worden. Na de literatuurstudie is een onderzoek verricht naar het vinden van een maat voor de anonimiteit. Het onderzoek is vervolgens afgesloten met het maken van voorbeelden en bijbehorende berekeningen.

[1] Voorbeelden hiervan zijn zender, ontvanger, passieve en actieve aanvaller, echte en nepbericht, mixnode en communicatieverbinding.

1.3 Gerelateerd werk

In [DSCP02] wordt de anonimiteitsgraad van een anonimiteitssysteem gedefinieerd. Als definitie gebruiken ze het entropieverschil van de maximale en de door een aanvaller berekende entropie. Dit verschil geeft de hoeveelheid informatie waarover een aanvaller beschikt om bij een mix een ingaand bericht te laten corresponderen met een uitgaand bericht.

De anonimiteitsgraad is toegepast op een aantal anonimiteitssystemen, waaronder een mixnetwerk. De berekening van de anonimiteitsgraad in het voorbeeld met een mixnetwerk is vereenvoudigd doordat het netwerk als één informatiebron is aanschouwd. Elke mix in dat netwerk is niet als een (unieke) informatiebron aangegeven. In ons model worden de mixen wel als unieke informatiebronnen gezien. Welke type mixen er zijn, worden in sectie 2.6 beschreven.

In [SD02] wordt een drietal definities gegeven waarvoor in de eerste de onzekerheid berekend wordt over een bericht die bij een gebruiker met een ontvangende of versturende rol hoort. De wijze waarop dit berekend wordt, wordt in de tweede definitie gegeven: de berekening gaat op dezelfde wijze als Shannon's marginale informatiemaat berekening [Sha01].

Met de laatste definitie kan voor alleen pool mixen en voor thresholdmixen (indien de pool 0 is) de zendersanonimiteit voor een bepaalde ronde berekend worden. Bij zendersanonimiteit wordt de identiteit van de zender verborgen gehouden: bij een mix bestaat de onzekerheid over welk bericht uit de groep ingaande berichten bij een bepaald uitgaand bericht hoort. De ronde is het moment waarop een aantal berichten worden ontvangen, een aantal in de mix verblijven en een aantal berichten de mix verlaten.

Tevens wordt een formule voorgesteld waarmee de totale entropie van een netwerk met mixen berekend kan worden. Hierin is bijvoorbeeld niet opgenomen het stuk dat beschrijft van zender naar mixnetwerk en van mixnetwerk naar ontvanger.

In [DP04a] worden berekeningen voor de zenders- en de ontvangersanonimiteit, voorgesteld waarbij met en zonder nepberichten worden ingezet. De berekeningen hebben alleen betrekking op de mix zelf: de deterministische en binomiale mix. De conclusie van de schrijvers omtrent de nepberichten is dat het nut van het nepberichten gebruik alleen toegevoegde waarde heeft voor de ontvangers- en niet voor de zendersanonimiteit. We onderzoeken of dit in ons model ook naar voren komt. Een belangrijk gegeven uit dit artikel is het feit dat het plaatsen van nepberichten bij de output een verhoogde anonimiteit zorgt dan nepberichten te plaatsen in de mix zelf.i

Onze maat voor de zenders- en de ontvangersanonimiteit wordt uitgedrukt in een waarde tussen nul en één. Dit is een waarde die voor het brede publiek toegankelijker is dan de entropie in de literatuur, zoals die wordt gebruikt in [SD02] en [DP04a]. Terwijl de anonimiteit in [SD02, DP04a] voor maximale anonimiteit een waarde groter is dan nul en voor minimale anonimiteit nul is, hebben wij gekozen voor de waarde één indien de anonimiteit maximaal is en voor de waarde nul indien de anonimiteit minimaal is.

In [DSCP02] wordt de anonimiteitsgraad weliswaar uitgedrukt in een waarde tussen 0 en 1, maar wordt toegepast voor één anonimiteitssysteem zonder rekening te houden met de mixparameters en ronden van de individuele mixen. In onze berekeningen wordt met beide factoren rekening gehouden.

Hoofdstuk 2

Model

In ons model wordt alleen gekeken naar een mixnetwerk waarmee anoniem berichten verstuurd kunnen worden. Om dit te realiseren kunnen berichten voor een bepaalde periode in de mix blijven. Een mixnetwerk kan ook voor bijna realtime toepassingen worden gebruikt zoals web-browsing en e-voting. De bijna realtime toepassingen vallen buiten de scope van ons onderzoek.

Welke participanten en welke acties worden uitgevoerd, worden in dit hoofdstuk beschreven.

2.1 Bericht

Een e-mail of een datapacket dat in ons model anoniem wordt verstuurd, wordt een bericht genoemd. Net als een e-mail, bevat een bericht een adres en soms een antwoordadres. Wil een zender geheel anoniem blijven dan wordt aan het bericht geen antwoordadres toegevoegd. Indien de zender een antwoord wenst te ontvangen op voorwaarde dat hij/zij anoniem blijft, dan wordt in het bericht een pseudoniem antwoordadres gebruikt. Voor de berekening van de anonimiteit in ons model worden de adressen van een bericht buiten beschouwing gelaten.

De berekening van de anonimiteit wordt immers op netwerk- en niet op binair niveau uitgevoerd. We nemen aan dat onze berichten één lengte hebben en de inhoud van onze berichten geen rol speelt in onze berekeningen.

In het model en in de rest van deze scriptie wordt voortaan de term bericht gebruikt in plaats van e-mail of datapacket. De berichten die we in ons model beschrijven, zijn vercijferde berichten.

2.2 Nepbericht

Een nepbericht[1] is vergelijkbaar met een bericht uit de vorige sectie. Het enige verschil tussen een nepbericht en een *echt* bericht is dat het als doel heeft om de anonimiteit in een mixnetwerk te garanderen. De ontvanger van een nepbericht kan met de inhoud niets doen, omdat (zoals de naam het aangeeft) het bericht *nep* is. Het nepbericht in ons model is een vercijferd nepbericht.

2.3 Zender

Iemand die in ons model een bericht verstuurt, wordt de zender genoemd. De zender heeft als functie berichten te genereren en te versturen. Een bericht dat door een zender wordt verstuurd, wordt altijd (in ons model) naar één persoon verstuurd en niet zoals in de praktijk mogelijk is

[1] We gebruiken het Nederlandse woord nepbericht in plaats van het Engelse woord 'dummy'.

naar meerdere personen. Tevens zendt een zender maar één bericht. Dit wordt zo gedefinieerd om het overzicht en de berekening eenvoudig te houden.

De zender kan in de werkelijke wereld vergeleken worden met een persoon die een e-mail verstuurt of een applicatie die automatisch een e-mail verstuurt. Voordat de zender het bericht stuurt, definieert de zender het pad voor het bericht. Het bericht passeert de mixen die in het pad beschreven staan.

2.4 Ontvanger

Een persoon of een applicatie die een bericht ontvangt, wordt de ontvanger genoemd. De ontvanger is in ons model alleen verantwoordelijk voor het ontvangen van berichten.

2.5 Aanvaller

De aanvaller in ons model is een persoon of systeem dat dataverkeer observeert en een verkeersanalyse verricht. Bij verkeersanalyse wordt onderzocht wie-met-wie communiceert. Vanwege het observatie gedrag wordt deze aanvaller ook wel een passieve aanvaller genoemd. Op basis van de analyseresultaten kan de aanvaller een bericht laten corresponderen met een gebruiker. Welke gebruiker en met welke rol (ontvanger/zender) bij een bepaald bericht hoort, is afhankelijk van welke kansverdeling de aanvaller heeft gemaakt voor een groep gebruikers.

De aanvaller in ons model heeft de capaciteit om alle communicatieverbindingen te observeren. Dit wil zeggen de observatie zijn mogelijk tussen:

- zender & mix;
- mixen onderling;
- ontvanger & mix.

Daardoor weet de aanvaller altijd het aantal berichten dat door een mix wordt ontvangen en het aantal berichten dat door een mix wordt verstuurd. Daarnaast is het type mix bij de aanvaller bekend.

In tegenstelling tot een passieve aanval kan een aanvaller ook een actieve aanval verrichten, waarbij een aanvaller getransporteerde data kan tegenhouden, wijzigen en verwijderen. De actieve aanvaller wordt in ons model buiten beschouwing gelaten om ons model en onze berekeningen eenvoudig te houden.

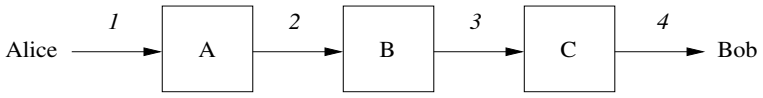
2.6 Mixen

David Chaum introduceerde in 1981 met [Cha81] de mix waarmee vandaag de dag onder andere anonieme e-mails kan worden verstuurd. Chaum's mix biedt bescherming tegen verkeersanalyse en gebruikt hiervoor public key cryptografie.

2.6.1 Mix werking

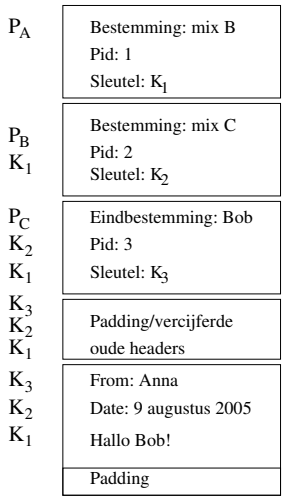
We gaan de werking van een praktische mix, genaamd MixMaster, behandelen[Cot]. Dit systeem is bedoeld om anoniem postings te doen in nieuwsgroepen en e-mails te versturen. De werking leggen we uit aan de hand van een voorbeeldsituatie die in figuur 2.1 wordt afgebeeld.

Een zender genaamd Alice wenst anoniem een e-mail te versturen naar ontvanger Bob, waarvoor drie mixen worden gebruikt.



Figuur 2.1: Alice stuurt Bob via drie mixen een e-mail.

Alice kiest vooraf een pad met mixen waarlangs haar e-mail anoniem wordt verstuurd. Dit doet haar e-mail, die de publieke sleutels van elke mix ophaalt. De publieke sleutel van elke mix wordt gebruikt om voor elke mix zijn header te vercijferen. We presenteren in figuur 2.2 hoe een datapacket voor MixMaster is opgebouwd.



Figuur 2.2: Opbouw datapacket van MixMaster.

Links van de delen van het datapacket staan de sleutels die gebruikt zijn om dat deel te vercijferen. De volgorde waarin de sleutels voor vercijfering zijn gebruikt, lezen we van boven naar beneden.

De header, dat vercijferd is met publieke sleutel P_i waarbij i de mix de eigenaar is van de publieke sleutel, bevat informatie zoals: het adres van de volgende mix, een packet ID en een symmetrische sleutel. Het packet ID wordt gebruikt ter controle of het datapacket reeds eerder is verstuurd. Indien het datapacket eerder is verstuurd, dan wordt hij verwijderd. De symmetrische sleutel K_i, waarbij i het sleutelnummer voorstelt, wordt door de mix gebruikt om zowel de headers die na zijn eigen header komen als de payload te ontcijferen. De payload bevat het bericht dat naar Bob wordt verstuurd.

Om ervoor te zorgen dat de grootte van het datapacket gedurende het transport gelijk blijft, zal elke mix zijn eigen header (vóór ontcijfering) eraf snijden en een kopie ervan achter de achterste header voor een mix aanplakken.

Elk transport is genummerd dat aangegeven is boven de communicatiepijlen in figuur 2.1. We gebruiken de volgende notaties met bijbehorende betekenissen om de inhoud van elk transport later te kunnen beschrijven:

- $P_{mixnaam}$: de publieke sleutel van de mix met de naam *mixnaam* voor asymmetrische vercijfering;
- $eP_{mixnaam}()$: asymmetrische vercijfering met publieke sleutel $P_{mixnaam}$;
- $K_{sleutelnummer}$: de geheime sleutel met nummer *sleutelnummer* voor symmetrische ver-/ontcijfering;

- $EK_{sleutelnummer}()$: symmetrische vercijfering met geheime sleutel $K_{sleutelnummer}$;
- $DK_{sleutelnummer}()$: symmetrische ontcijfering met geheime sleutel $K_{sleutelnummer}$
- $PID_{headernummer}$: het packet ID van *headernummer*;
- M: het bericht van Alice;
- A, B, C: de mixen;
- $||$: concatenatie.

Voor transport 1, in figuur 2.1, verricht de e-mail client van Alice een aantal verwerkingen om het datapacket naar het mixnetwerk te kunnen versturen. De eerste header van het datapacket bevat het adres van mix B, een packet ID PID_1 en geheime sleutel K_1. Al deze informatie wordt vercijferd met de publieke sleutel P_A. Mix A is de enige die deze header kan ontcijferen en wel met zijn private sleutel. De tweede header wordt vercijferd met geheime sleutel K_1 die zich in de eerste header bevindt en bevat (vercijferd met publieke sleutel P_B) het adres van mix C, packet ID PID_2 en geheime sleutel K_2. De derde header bevat de bestemming van bericht M, namelijk Bob, packet ID PID_3 en geheime sleutel K_3. Deze informatie wordt vercijferd eerst met publieke sleutel P_C, vervolgens met geheime sleutel K_2 en tenslotte met geheime sleutel K_1. Het einde van het datapacket is de payload en bevat bericht M dat vercijferd wordt met geheime sleutel K_3, K_2 en K_1 (in deze volgorde). We komen tot de volgende notatie voor transport 1:

$$eP_A(B, PID_1, K_1)||EK_1(eP_B(C, PID_2, K_2))||EK_1(EK_2(eP_C(Bob, PID_3, K_3))) \\ ||EK_1(EK_2(EK_3(M))).$$

Mix A snijdt bij ontvangst de eerste header af en plakt een kopie ervan aan het einde van alle headers. Vervolgens ontcijfert mix A met zijn private sleutel de afgesneden header en leest naar welke mix het datapacket moet doorsturen (=mix B). Mix A controleert packet ID PID_1 of dit datapacket reeds eerder is verstuurd. Na de controle gebruikt mix A geheime sleutel K_1 om de andere headers te ontcijferen en de laatste header (= eigen header) te vercijferen. Dit resulteert in de volgende notatie van het datapacket voor transport 2:

$$eP_B(C, PID_2, K_2)||EK_2(eP_C(Bob, PID_3, K_3))||EK_1(eP_A(B, PID_1, K_1))||EK_2(EK_3(M)).$$

Als het datapacket bij mix B arriveert, dan snijdt de mix de eerste header eraf, kopieert hem en plakt de kopie achter de laatste header. Mix B ontcijfert de afgesneden header met zijn private sleutel. Mix B leest vervolgens het adres van mix C, controleert daarna packet ID PID_2 en ontcijfert tenslotte met behulp van geheime sleutel K_2 de overige delen van het datapacket. De laatste header wordt vercijferd met K_2. Dit leidt tot de volgende notatie van het datapacket voor transport 3:

$$eP_C(Bob, PID_3, K_3)||EK_2(EK_1(eP_A(B, PID_1, K_1)))||EK_2(eP_B(C, PID_2, K_2))||EK_3(M).$$

Mix C ontcijfert als laatste mix de eerste header en leest dat bericht M naar de eindbestemming verstuurd moet worden, namelijk naar Bob. Voordat dit kan gebeuren controleert ook mix C packet ID PID_3 en ontcijfert de mix met geheime sleutel K_3 de payload. Dit heeft als resultaat dat tijdens transport 4 bericht M naar Bob wordt verstuurd.

We hebben hier een voorbeeld gegeven waarin een bericht M in één datapacket past. Het is ook mogelijk om een bericht die groter is dan de afgesproken payload grootte te versturen. In dat geval zal bericht M verdeeld worden over een aantal datapackets en de header voor de

laatste mix van elk datapacket bevat een message ID. Met het message ID kunnen de headers van andere datapackets herkend worden, die delen van bericht M bevatten. Indien alle delen verzameld zijn door de laatste mix uit het pad, wordt het complete bericht (na reconstructie) naar de eindbestemming verstuurd.

Dit voorbeeld bevat drie mixen maar er kan bijvoorbeeld ook minder mixen worden gebruikt. In dat geval wordt achter de headers willekeurige data opgevuld. Dit opvullen wordt net zolang gedaan totdat het datapacket een afgesproken grootte krijgt. De payload, de headers en het datapacket zelf hebben een vaste grootte. Daardoor is het mogelijk dat ook de payload met willekeurige data opgevuld wordt.

2.6.2 Mix requirements

Bij het implementeren van Chaum's mix concept voor een bepaalde toepassing, vinden een tweetal soorten bewerkingen op de berichten plaats: qua volgorde en uiterlijk. Indien bij de implementatie aan beide aspecten gedacht wordt, wordt op deze wijze verkeersanalyse moeilijker gemaakt.

De volgende *volgorde*-veranderingen zijn op te noemen:

- **Het mixen van berichten:** De volgorde waarin de berichten bij de mix zijn gearriveerd, is bij het verlaten van de mix (in een willekeurige volgorde) gewijzigd. Berichten kunnen moeilijk worden gevolgd indien berichten in geen enkel vaste volgorde door het netwerk stromen.
- **Het tijdelijk plaatsen van berichten in het geheugen:** Een aantal berichten die voor een bepaalde periode in de mix blijven, worden met ontvangende berichten gemixt. Bij anonieme e-mail toepassingen kunnen e-mails van enkele minuten tot enkele dagen in het geheugen van een mix blijven. Door het mixen van berichten in het geheugen, stuurt een mix de berichten in een willekeurige volgorde.

Uiterlijke veranderingen die berichten ondergaan, zijn de volgende:

- **Het vercijferen van berichten:** Elke vercijfering van data leidt tot een wijziging van informatie: de (inhoudelijke) boodschap blijft ongewijzigd, maar alleen de drager van deze boodschap verandert. Bij ontcijfering wijzigt de informatie naar oorspronkelijke staat. Elk bericht dat telkens een mix passeert, wijzigt de informatie. Berichten die na elke passage van een mix qua uiterlijk wijzigen en waarvan de boodschap onleesbaar behouden is, kunnen moeilijk worden geïdentificeerd op hun inhoud en uiterlijke kenmerken.
- **Alle berichten hebben één lengte:** Alle berichten die in het mixnetwerk stromen, hebben één lengte. Voor de payload, voor elke header en voor het bericht zelf hebben elk één afgesproken lengte en een afgesproken locatie in het bericht. Daardoor worden de lege locaties opgevuld met willekeurige data. Dit wordt ook wel 'padding' genoemd. Padding vindt ook plaats tijdens het vercijferen van de boodschap zelf, omdat voor het vercijferen van een boodschap een vaste lengte vereist is.

In ons model wordt verondersteld dat berichten een volgordeverandering ondergaan. Berichten die vercijferd een afgesproken lengte krijgen, worden in ons model als reeds uitgevoerde veranderingen beschouwd.

Op basis van de werking en de requirements zijn een aantal type mixen te definiëren. De volgende 6 mixen worden nu beschreven.

2.6.3 Thresholdmix

De thresholdmix [SDS02] is een mix dat berichten flusht[2] pas als t berichten zijn ontvangen.

2.6.4 Timemix

Deze mix [SDS02] flusht alle berichten na elke periode T. Indien de mix geen berichten heeft ontvangen, dan flusht hij niets.

2.6.5 Thresholdpoolmix

De thresholdpoolmix [SDS02] flusht nadat t berichten zijn ontvangen. Deze mixt houdt pool p berichten en flusht in totaal $(t - p)$ berichten. Deze mix heeft de capaciteit om berichten een aantal ronden in de mix te houden.

2.6.6 Timepoolmix

De timepoolmix [SDS02] flusht na elke periode T. Het aantal geflushte berichten is het aantal berichten in de mix minus pool p. De timepoolmix kan p berichten in de mix houden voor een aantal ronden.

2.6.7 Time dynamic-pool mix

Deze mix die ook wel cottrellmix of mixmaster[SDS02] genoemd wordt, is afhankelijk van de parameters T, t, f, en p, die de volgende betekenissen hebben:

- p: minimum aantal berichten die in de mix verblijven;
- f: fractie (die een waarde heeft tussen 0 en 1) van totaal aantal ontvangen berichten;
- t: het totaal aantal berichten in de mix;
- T: elke periode T flusht de mix.

Ook deze mix heeft de mogelijkheid om minimaal p berichten voor een aantal ronden in de mix te houden. Het aantal berichten dat geflusht wordt, wordt berekend met de volgende pseudo-code:

$$max(min(f * t, t - p), 0)$$

2.6.8 Binomialmix

Het aantal berichten dat door de binomialmix [DS03] geflusht wordt, wordt bepaald door het opgooien van een munt voor elk bericht in de mix. Afhankelijk van het resultaat bij het opgooien, mag een bericht bijvoorbeeld bij munt de mix verlaten en anders (bij kop) in de mix blijven. Het aantal geflushte berichten volgt een binomiale verdeling en de mix flusht na elke periode T. De binomialmix beschikt de capaciteit om berichten voor een aantal ronden in de mix te houden.

[2]Met deze term wordt het versturen bedoeld. Het is een functie die wordt aangeroepen als data uit een buffer, waarin tijdelijk data is opgeslagen, weg wordt geschreven naar een disk of (in dit geval) naar een communicatie-medium.

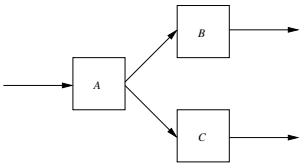
2.7 Mixnetwerk

In theorie is één mix met alle eigenschappen voldoende om verkeersanalyse te voorkomen, maar in de praktijk wordt gekozen om meerdere mixen in een netwerk te plaatsen. Op die manier wordt voorkomen dat één mix teveel vertrouwen krijgt en/of een 'single point of failure' wordt.

Mixnodes kunnen in twee type mixnetwerken worden geïmplementeerd: cascade- en freeroute-mixnetwerk [DS02, BPS00].

2.7.1 Cascade

In een cascade-mixnetwerk zijn mixen vergelijkbaar met een waterval, met elkaar verbonden. Een voorbeeld wordt in figuur 2.3 weergegeven.

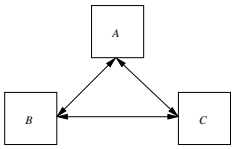


Figuur 2.3: Voorbeeld van een cascade-mixnetwerk.

Het pad waarmee een bericht door een dergelijk mixnetwerk stroomt, ligt gedeeltelijk vast. Dit komt omdat de mixen in een bepaalde communicatierichting zijn geconfigureerd. Een zender is daardoor gedwongen een pad te kiezen die uit bestaande communicatieverbindingen bestaat. Zo kan volgens figuur 2.3 een zender een pad kiezen waarin een bericht langs A naar B gaat. Deze cascade laat zien dat een bericht niet van B naar C kan gaan.

2.7.2 Freeroute

In het freeroute-mixnetwerk hebben alle mixen met elkaar een communicatieverbinding. In figuur 2.4 wordt een voorbeeld hiervan weergegeven.



Figuur 2.4: Voorbeeld van een freeroute-mixnetwerk.

In tegenstelling tot het cascade-mixnetwerk heeft de zender een vrijere keuze in het selecteren van een pad.

Hoofdstuk 3

Anonimiteit

3.1 Definitie

De definitie voor digitale anonimiteit die in de literatuur vaak wordt gebruikt is die van Pfitzmann en Köhntopp uit [PK00]:

Anonymity is the state of being not identifiable within a set of subjects,
the anonymity set.

De anonymity set bestaat uit twee of meer objecten die een bepaald kenmerk hebben of een bepaalde actie hebben uitgevoerd. Indien de anonymity set 1 object bevat dan is het duidelijk dat dit object tot een bepaald kenmerk hoort of actie heeft uitgevoerd: er bestaat geen anonimiteit voor dat object.

We kunnen een onderscheid maken tussen ontvangers- en zendersanonimiteit [PK00]. Bij *ontvangersanonimiteit* bestaat de anonymity set uit een aantal mogelijke ontvangers van een verzonden bericht. Zo bestaat bij *zendersanonimiteit* de anonymity set uit een aantal mogelijke zenders van een ontvangen bericht. De mix zorgt ervoor dat de kans voor elke ontvanger/zender uit de anonymity set uniform verdeeld is.

In [PK00] wordt verteld dat de anonimiteit groter wordt, naarmate de anonymity set groter wordt en de kansverdeling over deze set uniform is verdeeld. Indien een mix een maximale anonimiteit wil bereiken dan moet een mix dus een grotere groep mogelijke zenders/ontvangers realiseren met elk gelijke kans om de werkelijke zender of ontvanger te zijn. De mogelijkheid om dit te realiseren is door gebruik te maken van nepberichten.

Naast het vergroten van de anonimiteit zorgen nepberichten ook voor *unobservability* [PK00]. Dit houdt in voor zenders unobservability dat het niet merkbaar is of een zender een bericht heeft verzonden. Zo is het bij ontvangers unobservability niet merkbaar of een ontvanger daadwerkelijk een bericht heeft ontvangen.

Het verschil tussen anonimiteit en unobservability is dat anonimiteit de informatie verborgen houdt over welk object uit een groep (mogelijke) objecten een bepaald kenmerk heeft of actie uitvoert. Bij unobservability moet de informatie verborgen gehouden waarbij het merkbaar is een object een bepaalde kenmerk heeft of actie uitvoert.

Het volgende voorbeeld maakt het wat duidelijker: er bestaat geen anonimiteit en geen unobservability indien een zender, als enige, een bericht verstuurt. Verstuurt een zender nepberichten dan bestaat er nog steeds geen anonimiteit omdat met zekerheid is vast te stellen dat deze ene persoon berichten verstuurt: hetzij echte hetzij nepberichten. Maar er bestaat wel unobservability omdat de zender nepberichten verstuurt waarbij het niet mogelijk is te achterhalen

of de zender daadwerkelijk een echt bericht heeft verstuurd[1]. Er bestaat zowel anonimiteit als unobservability indien er een tweede zender aanwezig is met dezelfde capaciteiten als de eerste zender. Ten eerste is de groep zenders vergroot: een onderschept bericht is nu te linken naar twee zenders. Tenslotte is onbekend wie daadwerkelijk een echt bericht heeft verzonden.

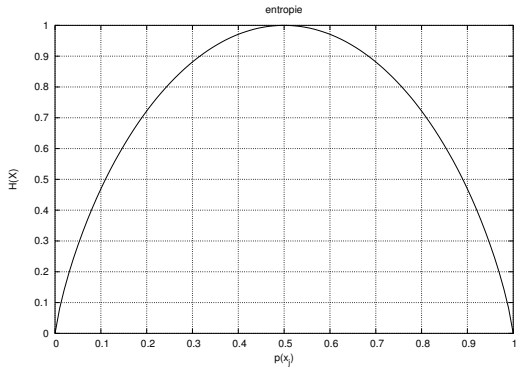
3.1.1 Entropie

We zullen nu aan de hand van de bovengenoemde zenders- en ontvangersanonimiteit een maat zoeken die in de informatietheorie voorkomt die ook voor het brede publiek een betekenis heeft. In deze sectie zullen we ons alleen focussen op de zendersanonimiteit.

Bij mixsystemen willen we vermijden dat een aanvaller kan voorspellen wanneer welk bericht het systeem verlaat. De aanvaller is daardoor onzeker geworden over de in- en de output. De mate van onzekerheid wordt uitgedrukt in Shannon's **entropie**. Deze maat is in [Sha01] geïntroduceerd als marginale informatiemaat: $H(X)$. Hieronder wordt de marginale informatiemaat van Shannon uitgedrukt voor X:

$$H(X) = -\sum_{j=1}^{a} p(x_j) \cdot \log(p(x_j))$$

De Shannon's entropie geeft de onzekerheid aan welk bericht door een informatiesysteem ontvangen wordt. In figuur 3.1 wordt $H(X)$ tegenover $p(x_j)$ gezet voor een informatiesysteem, zoals de mix, die twee berichten kan ontvangen.



Figuur 3.1: De entropie voor een informatiesysteem dat twee berichten ontvangt.

We lezen het volgende van figuur 3.1 af: er bestaat geen enkele onzekerheid ($H(X) = 0$) als de kans dat één van de twee berichten optreedt gelijk is aan nul. Dit komt omdat het andere bericht optreedt met een kans gelijk aan 1. Er bestaat maximale onzekerheid ($H(X) = 1$) indien beide berichten met elk een kans van een $\frac{1}{2}$ optreedt. We weten niet of bericht nummer 1 of bericht nummer 2 optreedt.

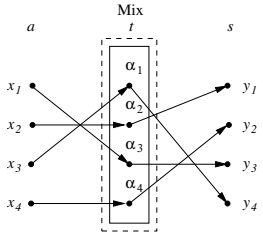
3.2 De anonimiteit exclusief nepberichten

Een ronde r van een mix is de periode waarop de mix een aantal berichten ontvangt, in het geheugen plaatst en flusht. We gebruiken de ronden om later een uitspraak te kunnen doen over scenario's waarin de mix een aantal berichten in de mix houdt.

[1] Dit kan alleen als de inhoud van een bericht ook is vercijferd.

Berichten die door de mix worden ontvangen, krijgen het symbool x_j. Berichten die door de mix worden geflusht, hebben het symbool y_l. De kans dat bericht x_j optreedt wordt genoteerd als $p(x_j)$. Op gelijke wijze wordt het optreden van bericht y_l genoteerd als $p(y_l)$.

Het aantal berichten dat door een mix wordt ontvangen, wordt aangegeven met a. Zo gelden voor t als het totaal aantal berichten in de mix en s als het aantal berichten dat wordt verzonden. In figuur 3.2 wordt een voorbeeld van een thresholdmix gegeven.



Figuur 3.2: Een thresholdmix, die vier berichten ontvangt, een threshold van vier heeft en vier berichten flusht.

Indien de kansen van optreden van zowel X als Y *onafhankelijk* zijn, dan geldt voor de gezamenlijke kans: $p(x_j, y_l) = p(x_j) \cdot p(y_l)$. Bestaat er een y_l die volledig *afhankelijk* is van een x_j, dan geldt: $p(x_j, y_l) = p(x_j)$. Indien we de kans willen berekenen waarbij een informatie reeds in bezit is, dan gebruiken we de conditionele kans. Bijvoorbeeld $p(y_l/x_j) = \frac{p(x_j,y_l)}{p(x_j)}$ of $p(x_j/y_l) = \frac{p(x_j,y_l)}{p(y_l)}$. Dan geeft aan $q(y_l/x_j)$ de kans over y_l indien x_j gegeven is. Er geldt dan voor $p(x_j, y_l) = p(y_l) \cdot p(x_j/y_l) = p(x_j) \cdot p(y_l/x_j)$.

De kans dat een bericht in de mix verblijft, indien in ronde k het bericht wordt ontvangen en in ronde r wordt geflusht, wordt genoteerd als $P(k, r)$.

We hebben een maximale anonimiteit over een bericht x_j indien $p(x_j)$ uniform verdeeld is. Daar komen we achter als we bericht y_l in bezit hebben. Daarom willen we de onzekerheid over X hebben indien Y gegeven is: $H(X/Y)$. Met andere woorden: $H(X/Y)$ is maximaal als deze gelijk is aan $H(X)$, want in $H(X)$ vertegenwoordigt de maximum entropie. We hebben zojuist in woorden beschreven hoe we aan de zendersanonimiteit komen en wanneer deze maximaal is. We stellen daarom de volgende vergelijking voor om onze zendersanonimiteit ZA uit te drukken:

$$ZA = \frac{H(X/Y)}{H(X)} = \frac{-\sum_{j=1}^{a}\sum_{l=1}^{s} p(y_l) \cdot p(x_j/y_l) \cdot \log(p(x_j/y_l))}{-\sum_{j=1}^{a} p(x_j) \cdot \log(p(x_j))} . \tag{3.1}$$

Voor de waarde van de zendersanonimiteit geldt $0 \leq ZA \leq 1$. De minimum waarde is gelijk aan nul en de maximum waarde is gelijk aan één. Indien $H(X)$ gelijk aan nul is dan hebben we geen zendersanonimiteit, omdat een aanvaller met zekerheid een optredende bericht x_j kan identificeren. Dit geldt ook voor $H(X/Y)$. In ons model en onze berekeningen nemen we aan dat $H(X) = 0$ en/of $H(X/Y) = 0$ niet zal voorkomen.

Analoog aan de zendersanonimiteit is eenvoudig de ontvangersanonimiteit te beschrijven: hoe groot is de onzekerheid over geflushte berichten Y als gegeven zijn de berichten X die door de mix zijn ontvangen. De anonimiteit over Y is maximaal indien de kansen dat de geflushte berichten optreden uniform verdeeld zijn. We stellen daarom de volgende vergelijking voor om

de ontvangersanonimiteit te kunnen berekenen:

$$OA = \frac{H(Y/X)}{H(Y)} = \frac{-\sum_{j=1}^{a}\sum_{l=1}^{s} p(x_j)\cdot p(y_l/x_j)\cdot \log(p(y_l/x_j))}{-\sum_{j=1}^{s} p(y_l)\cdot \log(p(y_l))} \ . \tag{3.2}$$

Ook de ontvangersanonimiteit is maximaal als de conditionele entropie even groot is als $H(Y)$. Voor de waarde van de ontvangersanonimiteit geldt $0 \leq OA \leq 1$. De ontvangersanonimiteit is minimaal als het gelijk is aan nul en maximaal als het gelijk is aan één. Indien $H(Y)$ gelijk aan nul is, dan hebben we geen ontvangersanonimiteit, omdat een aanvaller met zekerheid een optredende bericht y_l kan identificeren. Ook voor de ontvangersanonimiteit nemen we aan dat de entropie over Y en de conditionele entropie niet nul kunnen worden. Dit geldt natuurlijk ook voor $H(Y/X)$.

3.2.1 Thresholdmix

Vergelijkingen

We werken nu een voorbeeld met een thresholdmix uit. Deze mix heeft een threshold van t en buiten de mix staan a zenders en s ontvangers. Het aantal zenders en ontvangers komen overeen met ontvangen en geflushte berichten. Elke zender en ontvanger zendt en ontvangt slechts één bericht. Daardoor geldt $t = a = s$. Dit voorbeeld wordt in figuur 3.2 afgebeeld, waarin x_j een bericht dat afkomstig is van een zender en y_l een bericht dat door een ontvanger wordt ontvangen. Voor dit voorbeeld geldt $t = a = s = 4$. Elke punt in de mix representeert een locatie voor een bericht in het geheugen van de mix.

We zullen eerst de kansen $p(x_j)$, $p(y_l)$, $p(y_l/x_j)$ en $p(x_j/y_l)$ moeten uitdrukken, om vervolgens de zenders- en de ontvangersanonimiteit te kunnen berekenen.

We beschrijven nu de kans dat een locatie in het geheugen van de mix gevuld wordt met een bericht afkomstig uit de groep X. Deze kans, $p(\alpha_o)$[2], is afhankelijk van het optreden van een bericht uit de groep X. Vervolgens moeten we de kans beschrijven dat bericht x_j naar locatie α_o in het geheugen gaat. We komen daardoor tot de volgende kans voor het vullen van locatie α_o:

$$p(\alpha_o) = \sum_{j=1}^{a} p(x_j)\cdot p(\alpha_o/x_j) \ . \tag{3.3}$$

De kans voor elk opgetreden bericht y_l is uniform verdeeld. Er geldt:

$$\sum_{l=1}^{s} p(y_l) = 1 \ . \tag{3.4}$$

De kans dat gegeven bericht x_j in y_l aankomt, $p(y_l/x_j)$, wordt beschreven met de kansen:

- $p(\alpha_o/x_j)$: de kans dat bericht x_j locatie α_o in het geheugen van de mix vult;
- $p(y_l/\alpha_o)$: de kans dat een bericht vanuit locatie α_o naar ontvanger y_l gaat.

[2] We gebruiken hier de letter '*o*' in plaats van het cijfer nul

We komen daardoor voor $p(y_l/x_j)$ tot de volgende vergelijking:

$$p(y_l/x_j) = \sum_{o=1}^{t} p(\alpha_o/x_j) \cdot p(y_l/\alpha_o) \ . \tag{3.5}$$

Voor het definiëren van $p(x_j/y_l)$, maken we gebruik van de omkeerregel van Bayes. We komen dan tot de volgende kans:

$$p(x_j/y_l) = \frac{p(y_l/x_j) \cdot p(x_j)}{p(y_l)} \ . \tag{3.6}$$

Rekenvoorbeeld - zendersanonimiteit

We hebben hier de gegevens over de situatie zoals die in figuur 3.2 is afgebeeld:

- de threshold is 4;
- er zijn 4 zenders;
- er zijn 4 ontvangers;
- de 4 berichten die ontvangen zijn, worden na het ontcijferen en mixen direct geflusht.

We beginnen met de kans dat bericht x_j verzonden wordt door een zender. Deze kans is uniform verdeeld en hiervoor geldt: $\frac{1}{\#zenders}$. Dan is de kans $p(x_j)$ eenvoudig te bepalen:

$$p(x_j) = \tfrac{1}{4} \ .$$

De kans dat bericht x_j de locatie in α_o opvult, is uniform verdeeld en gelijk aan:

$$p(\alpha_o/x_j) = \tfrac{1}{4} \ .$$

Deze kans is afhankelijk van de t locaties in het geheugen van de mix. De kans dat een bericht vanuit locatie α_o naar y_l wordt verstuurd, is gelijk aan:

$$p(y_l/\alpha_o) = \tfrac{1}{4} \ .$$

We weten dat er vier berichten worden geflusht en we nemen aan dat de kansen over de geflushte berichten ook uniform verdeeld zijn. Daardoor komen we voor $p(y_l)$ tot de volgende waarde:

$$p(y_1) = p(y_2) = p(y_3) = p(y_4) = \tfrac{1}{4} \ .$$

Voor $p(y_l/x_j)$ nemen we als voorbeeld het berekenen van $p(y_1/x_2)$. Omdat deze kans ook uniform verdeeld is, kunnen $p(y_1/x_2)$ gebruiken voor elke $p(y_l/x_j)$. Voor deze berekening maken we gebruik van vergelijking (3.5).

$$\begin{aligned} p(y_1/x_2) = & p(\alpha_1/x_2) \cdot p(y_1/\alpha_1) + p(\alpha_2/x_2) \cdot p(y_1/\alpha_2) + \\ & p(\alpha_3/x_2) \cdot p(y_1/\alpha_3) + p(\alpha_4/x_2) \cdot p(y_1/\alpha_4) \ . \end{aligned}$$

Ook dit kunnen we eenvoudig schrijven als:

$$p(y_1/x_2) = p(y_l/x_j) = (\tfrac{1}{4} \cdot \tfrac{1}{4}) \cdot 4 = \tfrac{1}{4} \ .$$

Hierin gaat l van 1 naar 4 en gaat j ook van 1 naar 4. Nu we $p(y_l/x_j)$ hebben, kunnen met vergelijking (3.6) $p(x_j/y_l)$ berekenen:

$$p(x_j/y_l) = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{4} \ .$$

Met behulp van vergelijking (3.1) kunnen we de zendersanonimiteit berekenen:

$$ZA = \frac{-\sum_{j=1}^{4}\sum_{l=1}^{4}\frac{1}{4}\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)}{-\sum_{j=1}^{4}\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)} = \frac{-4\cdot 4\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)}{-4\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)} = 1\ .$$

Deze thresholdmix levert in dit voorbeeld een maximale zendersanonimiteit.

Rekenvoorbeeld - ontvangersanonimiteit

We kunnen nadat we de zendersanonimiteit hebben berekend, eenvoudig de ontvangersanonimiteit OA berekenen. We gebruiken hiervoor vergelijking (3.2). De gegevens die we voor de ZA hebben berekend kunnen we nu hiervoor gebruiken:

- $p(y_l) = \frac{1}{4}$;
- $p(x_j) = \frac{1}{4}$;
- $p(y_l/x_j) = \frac{1}{4}$.

Dus OA voor dit voorbeeld kan als volgt worden berekend:

$$OA = \frac{-\sum_{j=1}^{4}\sum_{l=1}^{4}\frac{1}{4}\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)}{-\sum_{l=1}^{4}\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)} = \frac{-4\cdot 4\cdot\frac{1}{4}\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)}{-4\cdot\frac{1}{4}\cdot\log\left(\frac{1}{4}\right)} = 1\ .$$

Ook met dit voorbeeld levert de thresholdmix een maximale ontvangersanonimiteit.

Vereenvoudiging van ZA en OA

We geven hier met behulp van de mixparameters vergelijkingen, waarmee de zendersanonimiteit en de ontvangersanonimiteit berekend kunnen worden. De t representeert hier de threshold. We nemen aan dat de berichten die door een mix worden gemixt uniform verdeeld zijn.

$$p(x_j) = \frac{1}{t} \tag{3.7}$$

$$p(y_l/x_j) = \frac{1}{t}\cdot\frac{1}{t}\cdot t = \frac{1}{t} \tag{3.8}$$

$$p(y_l) = \frac{1}{t} \tag{3.9}$$

$$p(x_j/y_l) = \frac{p(y_l/x_j)\cdot p(x_j)}{p(y_l)} = \frac{\frac{1}{t}\cdot\frac{1}{t}}{\frac{1}{t}} = \frac{1}{t} \tag{3.10}$$

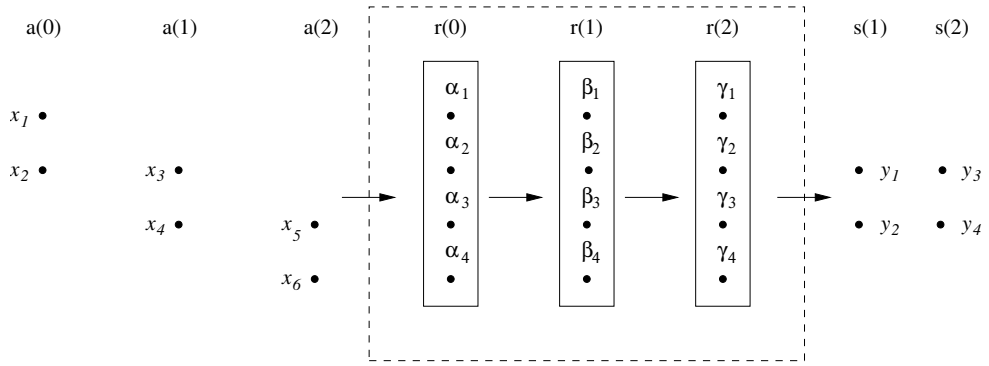
$$ZA = \frac{-t\cdot t\cdot\frac{1}{t}\cdot\frac{1}{t}\cdot\log\left(\frac{1}{t}\right)}{-t\cdot\frac{1}{t}\cdot\log\left(\frac{1}{t}\right)} = 1 \tag{3.11}$$

$$OA = \frac{-t\cdot t\cdot\frac{1}{t}\cdot\frac{1}{t}\cdot\log\left(\frac{1}{t}\right)}{-t\cdot\frac{1}{t}\cdot\log\left(\frac{1}{t}\right)} = 1 \tag{3.12}$$

3.2.2 Thresholdpoolmix

We zullen eerst een aantal voorbeelden geven en we komen tenslotte met formules voor ZA en OA die afgeleid zijn van de voorbeelden voor de thresholdmix en de thresholdpoolmix.

De parameters van deze mix zijn: threshold t is gelijk aan vier en pool p is gelijk aan twee. Voor de eenvoud van onze berekeningen zullen deze parameters gedurende de ronden onveranderd blijven. Een voorbeeld hiervan wordt in figuur 3.3 afgebeeld.



Figuur 3.3: Een thresholdpoolmix die twee ronden flusht.

We geven met $a(i)$ aan welke zenders in ronde i aanwezig zijn. Zo zien we dat voor $a(0)$ er twee berichten zijn die afkomstig zijn van twee zenders, namelijk x_1 en x_2. In het figuur geven we een ronde aan in de mix met $r(i)$. Zo heeft de mix in $r(0)$ (ronde nul) vier locaties in de mix, elk aangegeven met een α. In de volgende ronde ($r(1)$) heeft de mix hetzelfde aantal locaties, maar dit keer aangegeven met β's. We vinden rechts de ontvangers. Als we het over deze groep hebben, dan gebruiken hiervoor het symbool s. Dit geldt voor het moment indien een aanvaller het systeem observeert. Tenslotte als we het hebben over ronde nul, dan spreken van de initialisatie ronde.

Vergelijkingen

We behandelen hieronder een aantal kleine situaties met kleine getallen en we gebruiken hiervoor het voorbeeld dat in figuur 3.3 is gepresenteerd.

Tijdens de initialisatie ronde zenden twee zenders berichten x_1 en x_2 naar de mix. De mix vult met de twee berichten, twee van de vier lege locaties. Op dat moment zal de mix niet flushen omdat de threshold nog niet is bereikt. We kunnen dus met zekerheid zeggen dat de twee berichten van de initialisatie ronde naar ronde 1 in de mix zullen blijven. Stel dat bericht x_1 locatie α_1 vult en bericht x_2 locatie α_2 vult, dan geldt voor de kans om naar de volgende ronde te gaan:

$$p(\beta_1/\alpha_1) = p(\beta_2/\alpha_2) = 1.$$

In ronde 1 komen berichten x_3 en x_4 aan bij de mix. Deze berichten proberen elk een locatie in de mix te veroveren. Helaas zijn voor deze berichten niet mogelijk om locaties β_1 en β_2 te vullen. Daardoor geldt:

$$p(\beta_1/x_3) = p(\beta_2/x_3) = p(\beta_1/x_3) = p(\beta_2/x_4) = 0.$$

Met andere woorden voor zowel bericht x_3 als bericht x_4 geldt de kans dat een locatie β gevuld kan worden, is gelijk aan $\frac{1}{2}$. Dit komt omdat er twee van de vier locaties vrij zijn. Tevens geldt er een kans dat een bericht één van deze lege locaties vult. Deze kans is ook gelijk aan $\frac{1}{2}$. Uiteindelijk geldt er:

$$p(\beta_3/x_3) = p(\beta_4/x_4) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}.$$

De mix bevat nu vier berichten, dan is het tijd om te gaan flushen. 'Wij' kiezen voor dat moment om zowel het bericht in locatie β_3 als het bericht in locatie β_4 te laten flushen. Zo gaat bijvoorbeeld β_3 naar y_1 en gaat β_4 naar y_2. De kansen $p(y_1/\beta_3)$ en $p(y_2/\beta_4)$ zijn beide uniform verdeeld. De kans dat er een bericht de mix kan verlaten is gelijk aan een $\frac{1}{2}$ ($=\frac{2}{4}$). Tevens komt een kans erbij die aan geeft naar welke ontvanger mogelijkerwijs het bericht verstuurd wordt. Deze kans heeft een grootte van een $\frac{1}{2}$, omdat we twee ontvangers hebben. Daardoor geldt:

$$p(y_1/\beta_3) = p(y_2/\beta_4) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}.$$

Stel dat het bericht in locatie β_3 een ronde in de mix verblijft dan is de kans dat een bericht van β_3 naar γ_3 gaat gelijk aan $\frac{1}{2}$. De reden hiervoor is dat bericht β_3 niet geflusht wordt. De kans dat een bericht geflusht wordt is gelijk aan een $\frac{1}{2}$, en daardoor is de kans dat een bericht *in* de mix verblijft gelijk aan $(1 - \frac{1}{2})$.

Omdat we een aantal voorbeelden hebben gegeven, kunnen we nu vergelijkingen opstellen voor berichten die in ronde k bij de mix arriveren en in ronde r de mix verlaten.

Situatie: $k = 0$, $r = 1$

We nemen aan dat $p(x_j)$ uniform is verdeeld en er geldt:

$$\sum_{j=1}^{a} p(x_j) = 1$$

De kans dat bericht x_1 optreedt is gelijk aan $p(x_1) = \frac{1}{4}$. Dit komt omdat er in ronde 0 twee berichten arriveren en in ronde 1 ook twee berichten arriveren. We willen de kans $p(y_1/x_1)$ uitdrukken. Er geldt voor deze kans:

$$\begin{aligned} p(y_1/x_1) =& p(\alpha_1/x_1) \cdot p(\beta_1/\alpha_1) \cdot p(y_1/\beta_1) + p(\alpha_2/x_1) \cdot p(\beta_2/\alpha_2) \cdot p(y_1/\beta_2) + \\ & p(\alpha_3/x_1) \cdot p(\beta_3/\alpha_3) \cdot p(y_1/\beta_3) + p(\alpha_4/x_1) \cdot p(\beta_4/\alpha_4) \cdot p(y_1/\beta_4) \ . \end{aligned}$$

Dit wil zeggen dat bijvoorbeeld bericht x_1 naar locatie α_1 gaat. Vervolgens blijft het bericht in de mix van de initialisatie ronde naar de eerste ronde, aangegeven met $p(\beta_1/\alpha_1)$. Tenslotte zal het bericht in de eerste ronde vanuit locatie β_1 in bericht y_1 eindigen.

De kans dat bericht x_1 een locatie α_o vult, is gelijk aan:

$$p(\alpha_1/x_1) = p(\alpha_2/x_1) = p(\alpha_3/x_1) = p(\alpha_4/x_1) = \tfrac{4}{4} \cdot \tfrac{1}{4} = \tfrac{1}{4} \ .$$

Met zekerheid kunnen we de kans $p(\beta_o/\alpha_o)$ de waarde 1 geven, omdat de threshold nog niet is bereikt en dus mag de mix niet flushen. Er geldt indien een α_o gevuld is:

$$p(\beta_1/\alpha_1) = p(\beta_2/\alpha_2) = p(\beta_3/\alpha_3) = p(\beta_4/\alpha_4) = 1 \ .$$

Vanuit een locatie β_o kan het bericht of in de mix blijven of de mix verlaten. Tevens kan het bericht of naar ontvanger y_1 of naar y_2 gaan. We komen daardoor tot de volgende uitdrukking:

$$p(y_1/\beta_1) = p(y_1/\beta_2) = p(y_1/\beta_3) = p(y_1/\beta_4) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ .$$

We komen voor $p(y_1/x_1)$ tot het volgende:

$$p(y_1/x_1) = (\tfrac{1}{4} \cdot 1 \cdot \tfrac{1}{4}) \cdot 4 = \tfrac{1}{4} \ .$$

Voor $p(y_l)$ geldt:

$$\sum_{l=1}^{s} p(y_l) = 1$$

We nemen aan dat deze kans uniform verdeeld is. Daardoor komen we tot het volgende voor $p(y_1)$:

$$p(y_1) = p(y_2) = \tfrac{1}{2} \ .$$

We kunnen tenslotte $p(x_1/y_1)$ volgens de omkeerregel berekenen:

$$p(x_1/y_1) = \frac{p(y_1/x_1) \cdot p(x_1)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{8}$$

We weten dat $p(y_1/x_1)$ is gelijk aan $p(y_1/x_2)$ en dat deze kansen uniform verdeeld zijn. We hebben de volgende uitdrukkingen:

$$p(y_1/x_1) = p(y_2/x_1) = p(y_1/x_2) = p(y_2/x_2) = \tfrac{1}{4}$$

En voor $p(x_j/y_l)$ waarbij j van 1 naar 4 loopt en l van 1 naar 2 loopt:

$$p(x_j/y_l) = \tfrac{1}{8}$$

Voor de berekening van $p(y_1/x_3)$ verwijzen we je naar de volgende paragraaf. Maar voor hier wordt aangenomen dat deze berekeningen reeds uitgevoerd en juist zijn. Het volgende geldt:

$$\begin{aligned} p(y_1/x_3) =& p(y_2/x_3) = p(y_1/x_4) = p(y_2/x_4) = \tfrac{1}{4} \\ p(x_3/y_1) =& p(x_3/y_2) = p(x_4/y_1) = p(x_4/y_2) = \tfrac{1}{8} \end{aligned}$$

We kunnen nu eenvoudig de zendersanonimiteit ZA en ontvangersanonimiteit OA berekenen:

$$\begin{aligned} H(X/Y) =& -\sum_{j=1}^{a}\sum_{l=1}^{s} p(y_l) \cdot p(x_j/y_l) \cdot \log(p(x_j/y_l)) = -\sum_{j=1}^{4}\sum_{l=1}^{2} \frac{1}{2} \cdot \frac{1}{8} \cdot \log\left(\frac{1}{8}\right) \\ =& -4 \cdot 2 \cdot \tfrac{1}{2} \cdot \tfrac{1}{8} \cdot \log\left(\tfrac{1}{8}\right) = 1\tfrac{1}{2} \\ H(X) =& -\sum_{j=1}^{a} p(x_j) \cdot \log(p(x_j)) = -\sum_{j=1}^{4} \frac{1}{4} \cdot \log\left(\frac{1}{4}\right) = -\log\left(\frac{1}{4}\right) = 2 \\ ZA =& \frac{H(X/Y)}{H(X)} = \frac{3}{4} \ . \end{aligned}$$

$$\begin{aligned} H(Y/X) =& -\sum_{j=1}^{a}\sum_{l=1}^{s} p(x_j) \cdot p(y_l/x_j) \cdot \log(p(y_l/x_j)) = -\sum_{j=1}^{4}\sum_{l=1}^{2} \frac{1}{4} \cdot \frac{1}{4} \cdot \log\left(\frac{1}{4}\right) \\ =& -4 \cdot 2 \cdot \tfrac{1}{4} \cdot \tfrac{1}{4} \cdot \log\left(\tfrac{1}{4}\right) = 1 \\ H(Y) =& -\sum_{l=1}^{s} p(y_l) \cdot \log(p(y_l)) = -\sum_{j=1}^{2} \frac{1}{2} \cdot \log\left(\frac{1}{2}\right) = -\log\left(\frac{1}{2}\right) = 1 \\ OA =& \frac{H(Y/X)}{H(Y)} = 1 \ . \end{aligned}$$

We hebben hier te maken met een zendersanonimiteit van een $\frac{3}{4}$. De mix zorgt ervoor dat één bericht niet van de vier berichten kan zijn $(1 - \frac{1}{4})$.

We hebben een maximale ontvangersanonimiteit en dit is te verklaren door het feit dat de onzekerheid is gelijk aan het te behalen entropie maximum over de geflushte berichten: de kansen over Y zijn inderdaad uniform verdeeld als we vier berichten naar de mix versturen.

Situatie: $k = 1$, $r = 1$

In deze situatie nemen we als voorbeeld dat x_3 in dezelfde ronde van aankomst naar y_1 gaat. Voor $p(y_1/x_3)$ geldt:

$$\begin{aligned} p(y_1/x_3) =& p(\beta_1/x_3) \cdot p(y_1/\beta_1) + p(\beta_2/x_3) \cdot p(y_1/\beta_2) + \\ & p(\beta_3/x_3) \cdot p(y_1/\beta_3) + p(\beta_4/x_3) \cdot p(y_1/\beta_4) \ . \end{aligned}$$

De waarden die bij deze kansen horen, zijn:

$$\begin{aligned} p(\beta_1/x_3) =& p(\beta_2/x_3) = p(\beta_3/x_3) = p(\beta_4/x_3) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \\ p(y_1/\beta_1) =& p(y_1/\beta_2) = p(y_1/\beta_3) = p(y_1/\beta_4) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ . \end{aligned}$$

Daardoor heeft $p(y_1/x_3)$ de volgende waarde:

$$p(y_1/x_3) = (\tfrac{1}{4} \cdot \tfrac{1}{4}) \cdot 4 = \tfrac{1}{4} \ .$$

Voor zowel $p(x_j)$ als $p(y_l)$ gelden:

$$\sum_{j=1}^{a} p(x_j) = 1 \ ,$$

$$\sum_{l=1}^{s} p(y_l) = 1 \ .$$

Zowel $p(x_j)$ als $p(y_l)$ zijn uniform verdeeld. Daardoor hebben we voor $p(y_l)$ de volgende waarde:

$$p(y_1) = p(y_2) = \tfrac{1}{2} \ ,$$

en het optreden van x_j is gelijk aan:

$$p(x_1) = p(x_2) = p(x_3) = p(x_4) = \tfrac{1}{4} \ .$$

Uit de vorige situatie kunnen we berekeningen gebruiken voor de volgende kansen:

$$\begin{aligned} p(\alpha_1/x_1) =& p(\alpha_2/x_1) = p(\alpha_3/x_1) = p(\alpha_4/x_1) = p(\alpha_1/x_2) = p(\alpha_2/x_2) = p(\alpha_3/x_2) = p(\alpha_4/x_2) = \tfrac{1}{4} \\ p(\beta_1/\alpha_1) =& p(\beta_2/\alpha_2) = p(\beta_3/\alpha_3) = p(\beta_4/\alpha_4) = 1 \\ p(\beta_1/x_3) =& p(\beta_2/x_3) = p(\beta_3/x_3) = p(\beta_4/x_3) = \tfrac{1}{4} \\ p(y_1/\beta_1) =& p(y_1/\beta_2) = p(y_1/\beta_3) = p(y_1/\beta_4) = \tfrac{1}{4} \ . \end{aligned}$$

Nu kunnen we $p(x_3/y_1)$ berekenen:

$$p(x_3/y_1) = \frac{p(y_1/x_3) \cdot p(x_3)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{8} \ .$$

De waarden van zowel de zenders- als de ontvangersanonimiteit zijn gelijk uit de vorige situatie. De initialisatie ronde en ronde 1 kunnen als één hele ronde worden aanschouwd.

Situatie: $k = 1$, $r = 2$

Tijdens de initialisatie ronde bevatten twee locaties elk een bericht, afkomstig van x_1 en x_2. Welke locaties precies worden gevuld is natuurlijk onbekend. In deze ronde zal de mix niet flushen en blijven de berichten in de mix. In de volgende ronde arriveren berichten x_3 en x_4 bij de mix.

We gaan nu de kansen $p(x_3)$, $p(y_3/x_3)$, $p(y_3)$ en $p(x_3/y_3)$ berekenen. De kans $p(x_3)$ heeft een waarde van $\frac{1}{6}$, omdat we te maken hebben met in totaal zes berichten die door de mix worden ontvangen. Ook hier wordt aangenomen dat $p(x_j)$ uniform is verdeeld.

Voor de kans dat gegeven x_3 naar y_1 gaat, geldt het volgende:

$$\begin{aligned} p(y_3/x_3) =& p(\beta_1/x_3) \cdot p(\gamma_1/\beta_1) \cdot p(y_3/\gamma_1) + p(\beta_2/x_3) \cdot p(\gamma_2/\beta_2) \cdot p(y_3/\gamma_2) + \\ & p(\beta_3/x_3) \cdot p(\gamma_3/\beta_3) \cdot p(y_3/\gamma_3) + p(\beta_4/x_3) \cdot p(\gamma_4/\beta_4) \cdot p(y_3/\gamma_4) \ . \end{aligned}$$

De kans dat bericht x_3 een locatie kan veroveren in de mix is gelijk aan een $\frac{1}{2}$, omdat we weten dat er slechts twee van de vier locaties vrij zijn. Hierbij komt nog de kans dat het bericht één van deze locaties vult. Die kans is gelijk aan een $\frac{1}{2}$. Dus geldt het volgende:

$$p(\beta_1/x_3) = p(\beta_2/x_3) = p(\beta_3/x_3) = p(\beta_4/x_3) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ .$$

Vervolgens zal bericht x_3 in een locatie β_o een ronde in de mix verblijven. De kans dat het bericht in de mix verblijft is een $(1 - \frac{1}{2})$. Daardoor geldt het volgende:

$$p(\gamma_1/\beta_1) = p(\gamma_2/\beta_2) = p(\gamma_3/\beta_3) = p(\gamma_4/\beta_4) = \tfrac{1}{2} \ .$$

Het bericht x_3 dat zich in een locatie γ_o bevindt, gaat naar y_3. De kans dat dit gebeurt is ten eerste een $\frac{1}{2}$, want het bericht kan misschien in de mix blijven. Vervolgens komt de kans erbij dat het geflushte bericht daadwerkelijk naar y_3 gaat, en niet naar y_4. Deze kans is ook een $\frac{1}{2}$. Dus komen we voor de kans over y_3 gegeven γ_o tot de volgende vergelijking:

$$p(y_3/\gamma_1) = p(y_3/\gamma_2) = p(y_1/\gamma_3) = p(y_3/\gamma_4) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ .$$

We kunnen nu $p(y_3/x_3)$ berekenen:

$$p(y_3/x_3) = (\tfrac{1}{4} \cdot \tfrac{1}{2} \cdot \tfrac{1}{4}) \cdot 4 = \tfrac{1}{8} \ .$$

We gaan nu de kans $p(y_1)$ berekenen. Hiervoor geldt het volgende:

$$\sum_{l=1}^{s} p(y_l) = 1$$

De kans $p(y_l)$ wordt ook hier aangenomen dat hij uniform is verdeeld. Daardoor hebben we tot de volgende kansen:

$$p(y_1) = p(y_2) = p(y_3) = p(y_4) = \tfrac{1}{4}$$

De kans van het optreden van een bericht x_j is nu gelijk aan $\frac{1}{6}$. Het berekenen van de kans dat elke x_j naar een locatie in de mix gaat, hebben we reeds in de voorgaande situaties berekend. Daardoor gelden de volgende kansen:

$$\begin{aligned} p(\alpha_1/x_1) =& p(\alpha_2/x_1) = p(\alpha_3/x_1) = p(\alpha_4/x_1) = p(\alpha_1/x_2) = p(\alpha_2/x_2) = p(\alpha_3/x_2) = p(\alpha_4/x_2) = \tfrac{1}{4} \\ p(\beta_1/x_3) =& p(\beta_2/x_3) = p(\beta_3/x_3) = p(\beta_4/x_3) = p(\beta_1/x_4) = p(\beta_2/x_4) = p(\beta_3/x_4) = p(\beta_4/x_4) = \tfrac{1}{4} \\ p(\gamma_1/x_5) =& p(\gamma_2/x_5) = p(\gamma_3/x_5) = p(\gamma_4/x_5) = p(\gamma_1/x_6) = p(\gamma_2/x_6) = p(\gamma_3/x_6) = p(\gamma_4/x_6) = \tfrac{1}{4} \ . \end{aligned}$$

De initialisatie berichten (x_1 en x_2) blijven met zekerheid een ronde in de mix. Daardoor geldt:

$$p(\beta_1/\alpha_1) = p(\beta_2/\alpha_2) = p(\beta_3/\alpha_3) = p(\beta_4/\alpha_4) = 1 \ .$$

We gaan nu naar de tweede ronde. Voor elk bericht in een locatie β_o naar γ_o geldt de kans:

$$p(\gamma_1/\beta_1) = p(\gamma_2/\beta_2) = p(\gamma_3/\beta_3) = p(\gamma_4/\beta_4) = \tfrac{1}{2} \ .$$

Dit hebben we in voorgaande situaties reeds berekend en uitgelegd. Hetzelfde geldt ook voor het volgende:

$$p(y_3/\gamma_1) = p(y_3/\gamma_2) = p(y_3/\gamma_3) = p(y_3/\gamma_4) = \tfrac{1}{4} \ .$$

Met behulp van de omkeerregel kunnen we nu $p(x_3/y_3)$ berekenen:

$$p(x_3/y_3) = \frac{p(y_3/x_3) \cdot p(x_3)}{p(y_3)} = \frac{\frac{1}{8} \cdot \frac{1}{6}}{\frac{1}{4}} = \frac{1}{12} .$$

We hebben natuurlijk ook bijvoorbeeld de kans $p(x_5/y_3)$ nodig. Die vinden we doordat we $p(y_3/x_5)$ reeds hebben berekend. Er gelden namelijk dezelfde condities als voor $p(y_3/x_3)$ uit de vorige situaties: het bericht verlaat direct de mix bij aankomst ($k = 1$, $r = 1$). Met de omkeerregel vinden $p(x_5/y_3)$:

$$p(x_5/y_3) = \frac{p(y_3/x_5) \cdot p(x_5)}{p(y_3)} = \frac{\frac{1}{4} \cdot \frac{1}{6}}{\frac{1}{4}} = \frac{1}{6}$$

Om $p(x_3/y_1)$ te vinden, hebben we $p(y_1/x_3)$ nodig en die hebben we in de vorige situatie berekend:

$$p(x_3/y_1) = \frac{p(y_1/x_3) \cdot p(x_3)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{6}}{\frac{1}{4}} = \frac{1}{6}$$

Nu kunnen we de zenders- en ontvangersanonimiteit berekenen:

$$\begin{aligned}
H(X/Y) =& -\sum_{j=1}^{6}\sum_{l=1}^{4} p(x_j/y_l) \cdot p(y_l) \cdot \log(p(x_j/y_l)) \\
=& -(\frac{1}{6} \cdot \frac{1}{4} \cdot \log(\frac{1}{6})) \cdot 4 \cdot 2 - (\frac{1}{12} \cdot \frac{1}{4} \cdot \log(\frac{1}{12})) \cdot 4 \cdot 2 - (\frac{1}{6} \cdot \frac{1}{4} \cdot \log(\frac{1}{6})) \cdot 2 \cdot 2 \\
=& 0.862 + 0.598 + 0.431 = 1.89 \\
H(X) =& -\sum_{j=1}^{6} p(x_j) \cdot \log(p(x_j)) = -\log(\frac{1}{6}) = 2.585 \\
ZA =& \frac{1.89}{2.585} = 0.731 .
\end{aligned}$$

$$\begin{aligned}
H(Y/X) =& -\sum_{j=1}^{6}\sum_{l=1}^{4} p(y_l/x_j) \cdot p(x_j) \cdot \log(p(y_l/x_j)) \\
=& -(\frac{1}{4} \cdot \frac{1}{6} \cdot \log(\frac{1}{4})) \cdot 4 \cdot 2 - (\frac{1}{8} \cdot \frac{1}{6} \cdot \log(\frac{1}{8})) \cdot 4 \cdot 2 - (\frac{1}{4} \cdot \frac{1}{6} \cdot \log(\frac{1}{4})) \cdot 2 \cdot 2 \\
=& \tfrac{2}{3} + \tfrac{1}{2} + \tfrac{1}{3} = 1\tfrac{1}{2} \\
H(Y) =& -\log(\frac{1}{4}) = 2 \\
OA =& \frac{1\frac{1}{2}}{2} = \frac{3}{4} .
\end{aligned}$$

Formule: $p(x_j)$

Voor onze berekeningen nemen we aan dat $p(x_j)$ uniform is verdeeld. Voor een thresholdmix geldt voor $p(x_j)$:

$$p(x_j) = \frac{1}{a} \tag{3.13}$$

Waarvoor a het totaal aantal ontvangen berichten geldt. Bij een thresholdpoolmix is het optreden van bericht x_j afhankelijk van de berichten in ronde r, plus de berichten die in voorgaande ronden zijn opgetreden. Daardoor is $p(x_j)$ gelijk aan het volgende:

$$p(x_j) = \frac{1}{\sum_{i=0}^{r} a(i)} . \tag{3.14}$$

Hierin is i de ronde waarin bericht x_j kan opgetreden zijn. Dit kan vanaf de initialisatie ronde zijn tot aan de ronde waarin bericht x_j geflusht wordt, namelijk ronde r. Want het is mogelijk dat bericht x_j in ronde r arriveert en in dezelfde ronde geflusht wordt. We gebruiken voor de eenvoud het aantal ontvangen berichten $a(i)$ in ronde i.

Formule: $p(y_l/x_j)$

De kans dat gegeven een bericht x_j naar y_l gaat, is een sommatie van kansen die naar het t locaties in de mix gaat. Deze sommatie bestaat uit drie delen:

1. een deel dat de kans beschrijft voor een bericht van een zender naar de mix;
2. een deel dat de kans beschrijft voor een bericht in de mix zelf;
3. een deel dat de kans beschrijft voor een bericht van de mix naar een ontvanger.

Een voorbeeld hiervan is de situatie waarbij $k = 1$ en $r = 2$ gelden:

$$p(y_3/x_3) = p(\beta_1/x_3) \cdot p(\gamma_1/\beta_1) \cdot p(y_3/\gamma_1) + \cdots \ .$$

Hierin is $p(\beta_1/x_3)$ dat bericht x_3, afkomstig van een zender, naar locatie β_1 in de mix gaat. De kans $p(\gamma_1/\beta_1)$ beschrijft dat bericht x_3 gedurende een ronde in de mix verblijft. Tenslotte beschrijft $p(y_3/\gamma_1)$ het deel dat bericht x_3 de mix verlaat en naar y_3 gaat.

We kunnen dit in de volgende formule korter en algemener opschrijven, waarvoor geldt dat k de ronde is waarin bericht x_j door de mix is ontvangen en r de ronde is waarin bericht x_j door de mix is geflusht:

$$p(y_l/x_j) = \sum_{i=1}^{t} p(r(k)/x_j) \cdot P(k,r) \cdot p(y_l/r(r)). \tag{3.15}$$

Hierin geldt voor $r(k)$ als een locatie in de mix in ronde k en voor $r(r)$ als een locatie in ronde r.

De kans $p(r(k)/x_j)$ beschrijft hoe groot de kans is dat bericht x_j in locatie $r(k)$ terecht komt. De grootte van deze kans is afhankelijk van twee kansen:

1. hoe groot is de kans dat er lege locaties zijn in de mix, en;
2. hoe groot is de kans dat het bericht één van deze lege locaties verovert.

$P(k,r)$ geeft de kansen aan dat een bericht in een bepaalde locatie gedurende $(r-k)$ ronden in de mix verblijft. Deze kansen wordt als volgt uitgedrukt:

$$P(k,r) = \prod_{j=k}^{r-1} (1 - P(j)) \ . \tag{3.16}$$

Waarvoor $P(j)$ geldt de kans waarop een bericht in ronde j geflusht wordt. Dit wil zeggen in ronde j geldende kans dat een bericht geflusht wordt (en niet $P(k,r)$). We hebben dit reeds laten zien bijvoorbeeld in situatie $k = 1$ en $r = 2$ met de berekening van $p(y_3/x_3)$. We hebben daar bericht x_1 één ronde in de mix laten zitten en we hebben dit aangegeven met de kans:

$$p(\gamma_1/\beta_1) \ .$$

De kans $p(y_l/r(r))$ bestaat uit de volgende twee kansen:

1. de kans dat het bericht vanuit locatie $r(r)$ geflusht wordt, en;
2. de kans dat het bericht vanuit locatie $r(r)$ naar één van de ontvangers verstuurd wordt.

Formule: $p(y_l)$

Voor de kans dat een geflusht bericht y_l optreedt, geldt:

$$\sum_{l=1}^{s} p(y_l) = 1 \ . \tag{3.17}$$

Waarvoor s geldt het totaal aantal berichten dat door deze mix is geflusht. Dit wil zeggen van ronde 1 tot en met ronde r. We nemen tevens aan dat $p(y_l)$ uniform is verdeeld.

Formule: $p(x_j/y_l)$

Deze laatste kans kan als volgt worden uitgedrukt indien voorgaande formules zijn gebruikt:

$$p(x_j/y_l) = \frac{p(y_l/x_j) \cdot p(x_j)}{p(y_l)} \ . \tag{3.18}$$

We hebben hier de omkeerregel van Bayes toegepast, omdat we de kans over x_j willen berekenen waarbij het ontvangen van bericht y_l heeft plaats gevonden. Hiervoor gebruiken we de kans over y_l waarin x_j is reeds opgetreden en de kansen waarin x_j en y_l zijn opgetreden.

Toepassing formules

In de vorige paragraaf hebben we formules gedefinieerd die toegepast kunnen worden voor elke situatie met een mix. Voor deze formules gelden een aantal randvoorwaarden om ze te mogen gebruiken:

- Vergelijking (3.15) kan voor een thresholdmix gebruikt worden, als $P(k, r)$ weggelaten wordt. $P(k, r)$ beschrijft namelijk het moment waarop een bericht één of meer ronden in de mix verblijft. Bij een thresholdmix verblijft een bericht geen enkele ronde in de mix.

- De mixparameters voor een thresholdpoolmix blijven gedurende de ronden onveranderd. Dit wil zeggen de threshold en de pool grootte veranderen niet. Het aantal zenders en ontvangers per ronde veranderen (daardoor) ook niet.

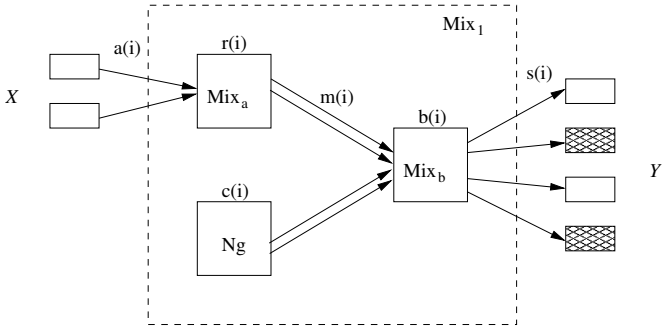
3.3 De anonimiteit inclusief nepberichten

We maken nu gebruik van nepberichten in onze berekeningen voor de zenders- en de ontvangers-anonimiteit. Het doel van de nepberichten is om de relatie tussen de input en de output van de mix nog meer te verbergen.

Er zijn twee methodes om nepberichten toe te voegen aan de echte berichten, namelijk: nepberichten worden toegevoegd aan de geflushte berichten (genaamd Flush Nepberichten Methode), of nepberichten worden bij de andere aanwezige berichten in de mix geplaatst (genaamd Pool Nepberichten Methode). Het voordeel van de laatste methode is dat de aanvaller onzekerheid heeft over de hoeveelheid nepberichten die in de mix zich bevinden. De eerste methode daarentegen verraadt hoeveel nepberichten tussen de geflushte berichten zit. We zouden het de aanvaller moeilijker kunnen maken door bij de FNM het aantal nepberichten per ronde willekeurig te maken, maar voor de eenvoud doen we dit niet in onze berekeningen.

3.3.1 Flush Nepberichten Methode

Een mix Mix_1 gebruikt de FNM waarvan de interne werking in het onderstaande figuur wordt weergegeven.



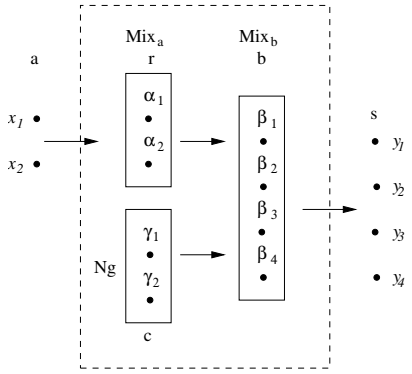
Figuur 3.4: FNM - echte en nepberichten (afkomstig van Mix_a en Ng) worden naar Mix_b gestuurd, die mixt en zonder vertraging alle berichten flusht.

In het figuur staat $a(i)$ voor het aantal ontvangen berichten in ronde i. Zo staan in ronde i, $m(i)$ voor het aantal geflushte berichten door Mix_a, $c(i)$ voor het aantal verzonden nepberichten door de nepberichtengenerator Ng en $s(i)$ voor het aantal geflushte berichten door Mix_b. Er geldt voor deze methode dat $a(i) < s(i)$. Anders zou dit betekenen dat de mix blijft vollopen.

In Mix_1 ontvangt Mix_a berichten van buiten de mix (berichten uit de groep X) en flusht volgens een bepaald algoritme de berichten naar Mix_b. Onafhankelijk van Mix_a produceert en flusht de Ng nepberichten naar Mix_b. Interne Mix_b mixt en flusht als een non-pool mix alle berichten $(m(i) + c(i))$ en zonder vertraging.

3.3.2 Thresholdmix

We zullen in het onderstaande figuur een representatie met punten geven voor een thresholdmix dat overeen komt met het voorbeeld in figuur 3.4.



Figuur 3.5: Een thresholdmix die twee echte berichten met twee nepberichten mixt en flusht.

Hierin vertegenwoordigen x_1 en x_2 de berichten die door de thresholdmix (afgebeeld als een kader met gebroken lijn) worden ontvangen. Alle ontvangen berichten wordt genoteerd met a. Deze berichten veroveren een locatie in Mix_a. Elke locatie wordt gerepresenteerd als een α. Alle locaties in Mix_a wordt aangeduid met een r. Nadat de berichten uit een locatie α worden geflusht naar Mix_b, produceert en flusht Ng alle nepberichten vanuit locatie γ ook naar Mix_b. Mix_b mixt en flusht tenslotte alle berichten. Uit de thresholdmix komen vervolgens bericht y_1 t/m bericht y_4.

We zullen nu een rekenvoorbeeld geven waarin we uitleggen hoe we aan de kansen komen en welke waarden voor dit voorbeeld ingevuld moeten worden. Het gaat hier om een thresholdmix, die een threshold heeft van twee. Het nepberichtenbeleid die voor dit voorbeeld gebruikt wordt bestaat uit het mixen van twee nepberichten met twee echte berichten. Tevens nemen we aan dat de kansen zowel ontvangen berichten als gegenereerde nepberichten uniform verdeeld zijn. Het ontvangen van berichten wordt als een aparte bron gezien en het genereren van de nepberichten wordt ook als een aparte bron. Voor de kans van het optreden van zowel echte berichten als nepberichten geldt het volgende:

$$\sum_{j=1}^{a} p(x_j) = \frac{1}{2}$$
$$\sum_{g=1}^{c} p(\gamma_g) = \frac{1}{2}$$

Waarvoor a staat als het totaal aantal ontvangen berichten. Daardoor kunnen we het volgende zeggen voor het voorbeeld:

$$p(x_1) = p(x_2) = p(\gamma_1) = p(\gamma_2) = \tfrac{1}{4} \ .$$

We berekenen nu hoe groot de kans is als we bericht x_1 reeds hebben en als bericht y_1 geflusht wordt. Het volgende geldt voor $p(y_1/x_1)$:

$$\begin{aligned} p(y_1/x_1) = & p(\alpha_1/x_1) \cdot p(\beta_1/\alpha_1) \cdot p(y_1/\beta_1) + p(\alpha_2/x_1) \cdot p(\beta_1/\alpha_2) \cdot p(y_1/\beta_1) + \\ & p(\alpha_1/x_1) \cdot p(\beta_2/\alpha_1) \cdot p(y_1/\beta_2) + p(\alpha_2/x_1) \cdot p(\beta_2/\alpha_2) \cdot p(y_1/\beta_2) + \\ & p(\alpha_1/x_1) \cdot p(\beta_3/\alpha_1) \cdot p(y_1/\beta_3) + p(\alpha_2/x_1) \cdot p(\beta_3/\alpha_2) \cdot p(y_1/\beta_3) + \\ & p(\alpha_1/x_1) \cdot p(\beta_4/\alpha_1) \cdot p(y_1/\beta_4) + p(\alpha_2/x_1) \cdot p(\beta_4/\alpha_2) \cdot p(y_1/\beta_4) \ . \end{aligned}$$

We leggen nu kort uit hoe groot deze kansen zijn. De andere kansen, worden aangenomen, zijn uniform verdeeld. Daardoor komen we tot de volgende uitdrukking:

$$p(\alpha_1/x_1) = p(\alpha_2/x_1) = \tfrac{1}{2}$$

Bericht x_1 heeft de keuze om in één van de twee locaties van Mix_a terecht te komen. Een bericht die afkomstig is van Mix_a en die naar Mix_b gaat, heeft de keuze om één van de vier locaties te veroveren. Daardoor geldt het volgende:

$$p(\beta_1/\alpha_1) = p(\beta_2/\alpha_1) = p(\beta_3/\alpha_1) = p(\beta_4/\alpha_1) = \tfrac{1}{4} \ .$$

Nadat Mix_b heeft gemixt, zal de mix uiteindelijk vier berichten flushen. Namelijk een bericht vanuit een locatie in Mix_b kan in één van de vier y's eindigen. Daardoor geldt:

$$p(y_1/\beta_1) = p(y_1/\beta_2) = p(y_1/\beta_3) = p(y_1/\beta_4) = \tfrac{1}{4} \ .$$

Het invullen voor de $p(y_1/x_1)$, levert het volgende op:

$$p(y_1/x_1) = (\tfrac{1}{2} \cdot \tfrac{1}{4} \cdot \tfrac{1}{4}) \cdot 2 \cdot 4 = \tfrac{1}{4}$$

Voor het berekenen van $p(y_l)$ geldt:

$$\sum_{l=1}^{s} p(y_l) = 1$$

We nemen aan dat $p(y_l)$ weer uniform is verdeeld. Daardoor komen we voor $p(y_l)$ tot een $\frac{1}{4}$. We zullen nu de kans berekenen waarin nepbericht γ_1 naar y_1 kan gaan:

$$\begin{aligned} p(y_1/\gamma_1) =& p(\beta_1/\gamma_1) \cdot p(y_1/\beta_1) + p(\beta_2/\gamma_1) \cdot p(y_1/\beta_2) + \\ & p(\beta_3/\gamma_1) \cdot p(y_1/\beta_3) + p(\beta_4/\gamma_1) \cdot p(y_1/\beta_4) \end{aligned}$$

Doordat γ_1 kan kiezen 1 van de 2 lege locaties in een halflege Mix_b, geldt voor elke $p(\beta_o/\gamma_1)$ dat deze kans gelijk is aan een $(\frac{1}{2} \cdot \frac{1}{2} =) \frac{1}{4}$. Vervolgens kan γ_1 vanuit een β_o in één van de vier y's eindigen. Na het invullen levert $p(y_1/\gamma_1)$ het volgende op:

$$p(y_1/\gamma_1) = (\tfrac{1}{4} \cdot \tfrac{1}{4}) \cdot 4 = \tfrac{1}{4}$$

Met de omkeerregel berekenen we $p(\gamma_1/y_1)$ en $p(x_1/\gamma_1)$:

$$\begin{aligned} p(\gamma_1/y_1) &= \frac{p(y_1/\gamma_1) \cdot p(\gamma_1)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{4} \\ p(x_1/y_1) &= \frac{p(y_1/x_1) \cdot p(x_1)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{4} \ . \end{aligned}$$

We nemen voor de berekeningen van ZA en OA aan dat $p(y_l/x_j)$ en $p(x_j/y_l)$ uniform zijn verdeeld.

De zendersanonimiteit is als volgt te berekenen:

$$\begin{aligned} H(X/Y) =& -\sum_{j=1}^{2}\sum_{l=1}^{4} p(x_j/y_l) \cdot p(y_l) \cdot \log(p(x_j/y_l)) - \sum_{g=1}^{2}\sum_{l=1}^{4} p(\gamma_g/y_l) \cdot p(y_l) \cdot \log(p(\gamma_g/y_l)) \\ =& -(\frac{1}{4} \cdot \frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 2 \cdot 4 - (\frac{1}{4} \cdot \frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 4 \cdot 2 = 1 + 1 = 2 \\ H(X) =& -\sum_{j=1}^{2} p(x_j) \cdot \log(p(x_j)) - \sum_{g=1}^{2} p(\gamma_g) \cdot \log(p(\gamma_g)) \\ & -(\frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 2 - (\frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 2 = 1 + 1 = 2 \\ ZA =& \frac{H(X/Y)}{H(X)} = 1 \end{aligned}$$

De berekening van de ontvangersanonimiteit levert het volgende op:

$$\begin{aligned} H(Y/X) =& -\sum_{j=1}^{2}\sum_{l=1}^{4} p(y_l/x_j) \cdot p(x_j) \cdot \log(p(y_l/x_j)) \\ =& (-\frac{1}{4} \cdot \frac{1}{4} \cdot \log(\frac{1}{4})) = 1 \\ H(Y) =& -\sum_{l=1} p(y_l) \cdot \log(p(y_l)) \\ & (\frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 4 = 2 \\ OA =& \frac{H(Y/X)}{H(Y)} = \frac{1}{2} \end{aligned}$$

We presenteren hier nieuwe vergelijkingen aan de hand van mixparameters en we gebruiken hiervoor de t als threshold en de d als het aantal nepberichten.

$$p(x_j) = \frac{1}{2 \cdot t} \tag{3.19}$$

$$p(y_l/x_j) = \left(\frac{1}{t} \cdot \frac{1}{t+d} \cdot \frac{1}{t+d}\right) \cdot t \cdot (t+d) = \frac{1}{t+d} \tag{3.20}$$

$$p(y_l) = \frac{1}{t+d} \tag{3.21}$$

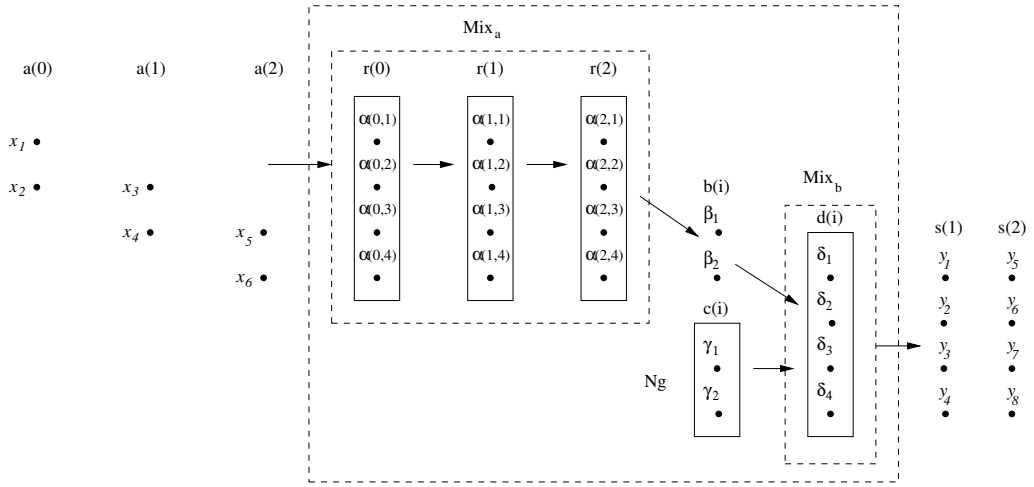
$$p(x_j/y_l) = \frac{p(y_1/x_j) \cdot p(x_j)}{p(y_l)} = \frac{\frac{1}{t+d} \cdot \frac{1}{2 \cdot t}}{\frac{1}{t+d}} = \frac{1}{2 \cdot t} \tag{3.22}$$

$$ZA = \frac{-\frac{1}{2 \cdot t} \cdot \frac{1}{t+d} \cdot \log(\frac{1}{2 \cdot t}) \cdot t \cdot (t+d)}{-\frac{1}{t} \cdot \log(\frac{1}{t}) \cdot t} = \frac{-\frac{1}{2} \cdot \log(\frac{1}{2 \cdot t})}{-\log(\frac{1}{t})} \tag{3.23}$$

$$OA = \frac{-\frac{1}{t+d} \cdot \frac{1}{2 \cdot t} \cdot \log(\frac{1}{t+d}) \cdot (t+d) \cdot t}{-\frac{1}{t+d} \cdot \log(\frac{1}{t+d}) \cdot (t+d)} = \frac{-\frac{1}{2} \cdot \log(\frac{1}{t+d})}{-\log(\frac{1}{t+d})} = \frac{1}{2} \tag{3.24}$$

3.3.3 Thresholdpoolmix

Met figuur 3.6 geven we een representatie waarin de thresholdpoolmix twee ronden draait. Dit figuur kan worden gebruikt voor twee voorbeeldsituaties die later behandeld worden.



Figuur 3.6: De thresholdpoolmix volgens FNM vanaf ronde 0 tot en met ronde 2

Hierin staat $a(i)$ voor de ontvangen berichten, elk aangegeven met x_j in ronde i. Elke locatie j in de mix in ronde i wordt aangegeven met $\alpha(i,j)$. Ronde i met al haar locaties wordt aangeduid met $r(i)$. Elk geflusht bericht vanuit een locatie in de mix gaat naar een β. Alle β's bij elkaar wordt aangeduid met $b(i)$ voor in ronde i. $b(i)$ komt overeen met het aantal geflushte berichten afkomstig van Mix_a. Vervolgens krijgt een bericht β een locatie in Mix_b, aangeduid met een δ. Alle δ's bij elkaar wordt genoteerd als $d(i)$ voor in ronde i. Na het mixen, flusht Mix_b berichten y_1 t/m y_4 in ronde 1 en y_5 t/m y_8 in ronde 2. Alle geflushte berichten wordt genoteerd als $s(i)$ voor in ronde i.

We zullen nu twee situaties uitwerken:

1. berichten die in ronde $k = 1$ bij de mix arriveren en in ronde $r = 1$ de mix verlaten;
2. berichten die in ronde $k = 1$ bij de mix arriveren en in ronde $r = 2$ de mix verlaten.

Situatie: $k = 1$ **en** $r = 1$

We nemen wederom aan dat $p(x_j)$ en $p(\gamma_g)$ uniform verdeeld zijn. Er gelden:

$$\sum_{j=1}^{a} p(x_j) = \frac{1}{2}$$

$$\sum_{g=1}^{c} p(\gamma_g) = \frac{1}{2}$$

We krijgen daardoor:

$$\begin{aligned} p(x_1) =& p(x_2) = p(x_3) = p(x_4) = \tfrac{1}{2} \cdot \tfrac{1}{4} = \tfrac{1}{8} \\ p(\gamma_1) =& p(\gamma_2) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ . \end{aligned}$$

Tevens de kans dat een bericht x_j eindigt in bericht y_l wordt ook aangenomen als uniform verdeeld. De initialisatie berichten (x_1 en x_2) verblijven van de initialisatie ronde naar ronde nummer 1 in de mix. De kans dat dit gebeurt is (met zekerheid) gelijk aan één. Het volstaat dan om alleen bijvoorbeeld $p(y_1/x_3)$ te berekenen, omdat deze kans overeen komt met $p(y_1/x_1)$, $p(y_1/x_2)$ en $p(y_1/x_4)$. Het volgende geldt voor $p(y_1/x_3)$:

$$p(y_1/x_3) = \sum_{o=1}^{r(1)} \sum_{b=1}^{b(1)} \sum_{d=1}^{d(1)} p(\alpha(1,o)/x_3) \cdot p(\beta_b/\alpha(1,o)) \cdot p(\delta_d/\beta_b) \cdot p(y_1/\delta_d) \ .$$

De kans dat bericht x_3 een locatie verovert in de mix is gelijk aan:

$$p(\alpha(1,o)/x_3) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ ,$$

omdat de mix voor een $\frac{1}{2}$ leeg is en de kans dat een lege locatie wordt geselecteerd, is tevens gelijk aan een $\frac{1}{2}$. De kans dat het stuk beschrijft waarin bericht x_3 vanuit zijn locatie in de mix naar output bericht β_b gaat, is gelijk aan:

$$p(\beta_b/\alpha(1,o)) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \ .$$

We komen hierop, omdat de kans dat bericht x_3 geflusht wordt is gelijk aan een $\frac{1}{2}$. En als bericht x_3 geflusht is, is de kans dat dit bericht naar een β gaat ook gelijk aan een $\frac{1}{2}$. We nemen aan dat deze kans uniform is verdeeld.

Indien β_b bij Mix_b arriveert, heeft bericht β_b een kans ($p(\delta_d/\beta_b)$) van een $\frac{1}{4}$ om een lege locatie te veroveren. Ook deze kans is uniform verdeeld. De mix is namelijk op dat moment volledig leeg.

Tenslotte is de kans dat een bericht vanuit Mix_b naar y_1 gaat, is gelijk aan een $\frac{1}{4}$, omdat er in totaal $s(1)$ (=4) berichten geflusht wordt. Uiteindelijk levert voor $p(y_1/x_3)$ de volgende waarde op:

$$p(y_1/x_3) = (\tfrac{1}{4} \cdot \tfrac{1}{4} \cdot \tfrac{1}{4} \cdot \tfrac{1}{4}) \cdot 4 \cdot 2 \cdot 4 = \tfrac{1}{8} \ .$$

We nemen hier wederom aan dat $p(y_l)$ uniform is verdeeld en er geldt voor $p(y_1)$ het volgende:

$$\sum_{l=1}^{s} p(y_1) = 1$$

Waarvoor s het totaal aantal geflushte berichten staat. We krijgen daardoor $p(y_l) = \frac{1}{4}$. We zullen nu $p(y_1/\gamma_g)$ berekenen. Voor $p(y_1/\gamma_1)$ geldt het volgende:

$$p(y_1/\gamma_1) = \sum_{g=1}^{c(1)} p(\delta_d/\gamma_1) \cdot p(y_1/\delta_d)$$

Berichten afkomstig van Mix_a hebben reeds twee locaties in Mix_b veroverd. Dit betekent voor de kans dat nepbericht γ_g een locatie in Mix_b kan veroveren, is gelijk aan een $\frac{1}{2}$ en de kans dat een locatie gevuld wordt, is ook een $\frac{1}{2}$. Dit levert voor $p(\delta_d/\gamma_1)$ de volgende waarde op:

$$p(\delta_d/\gamma_1) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}$$

De berekening voor $p(y_1/\delta_d)$ hebben we reeds gedaan en de grootte van deze kans was gelijk aan $\frac{1}{4}$. Daardoor is $p(y_1/\gamma_1)$ gelijk aan:

$$p(y_1/\gamma_1) = \tfrac{1}{4} \ .$$

We vinden $p(x_3/y_1)$ en $p(\gamma_1/y_1)$ met behulp van de omkeerregel van Bayes:

$$p(x_3/y_1) = \frac{p(y_1/x_3) \cdot p(x_3)}{p(y_1)} = \frac{\frac{1}{8} \cdot \frac{1}{8}}{\frac{1}{4}} = \frac{1}{16}$$
$$p(\gamma_1/y_1) = \frac{p(y_1/\gamma_1) \cdot p(\gamma_1)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{4}$$

We kunnen nu dan onze zenders- en ontvangersanonimiteit berekenen. Voor de zendersanonimiteit ZA krijgen we de volgende waarde:

$$\begin{aligned} H(X/Y) =& -\sum_{j=1}^{4}\sum_{l=1}^{4} p(x_j/y_l) \cdot p(y_l) \cdot \log(p(x_j/y_l)) - \sum_{g=1}^{2}\sum_{l=1}^{4} p(\gamma_g/y_l) \cdot p(y_l) \cdot \log(p(\gamma_g/y_l)) \\ =& -(\frac{1}{16} \cdot \frac{1}{4} \cdot \log(\frac{1}{16})) \cdot 4 \cdot 4 - (\frac{1}{4} \cdot \frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 4 \cdot 2 = 1 + 1 = 2 \\ H(X) =& -\sum_{j=1}^{4} p(x_j) \cdot \log(p(x_j)) - \sum_{g=1}^{2} p(\gamma_g) \cdot \log(p(\gamma_g)) \\ =& -(\frac{1}{8} \cdot \log(\frac{1}{8})) \cdot 4 - (\frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 2 = 1\frac{1}{2} + 1 = 2\frac{1}{2} \\ ZA =& \frac{2}{2\frac{1}{2}} = \frac{4}{5} = 0.8 \ . \end{aligned}$$

$$\begin{aligned} H(Y/X) =& -\sum_{j=1}^{4}\sum_{l=1}^{4} p(y_l/x_j) \cdot p(x_j) \cdot \log(p(y_l/x_j)) \\ =& -(\frac{1}{8} \cdot \frac{1}{8} \cdot \log(\frac{1}{8})) \cdot 4 \cdot 4 = \frac{3}{4} \\ H(Y) =& -\sum_{l=1}^{4} p(y_l) \cdot \log(p(y_l)) \\ =& -(\frac{1}{4} \cdot \log(\frac{1}{4})) \cdot 4 = 2 \\ OA =& \frac{\frac{3}{4}}{2} = \frac{3}{8} = 0.375 \ . \end{aligned}$$

Situatie: $k = 1$ **en** $r = 2$

We krijgen nu te maken met een thresholdpoolmix die twee ronden draait. Gedurende twee ronden krijgen we te maken met als input x_1 t/m x_6 en gesimuleerde input γ_1 t/m γ_4. Daardoor geldt het volgende:

$$p(x_1) = p(x_2) = p(x_3) = p(x_4) = p(x_5) = p(x_6) = \tfrac{1}{2} \cdot \tfrac{1}{6} = \tfrac{1}{12}$$
$$p(\gamma_1) = p(\gamma_2) = p(\gamma_3) = p(\gamma_4) = \tfrac{1}{2} \cdot \tfrac{1}{4} = \tfrac{1}{8}$$

De kans dat x_3 in y_5 is een factor kleiner ten opzichte van de vorige situatie. Voor $p(y_5/x_3)$ geldt:

$$p(y_5/x_3) = \sum_{o=1}^{r(2)}\sum_{b=1}^{b(2)}\sum_{d=1}^{d(2)} p(\alpha(1,o)/x_3) \cdot p(\alpha(2,o)/\alpha(1,o)) \cdot p(\beta_b/\alpha(2,o)) \cdot p(\delta_d/\beta_b) \cdot p(y_5/\delta_d)$$

We gebruiken de uitkomst uit de vorige situatie om hier de nieuwe $p(y_5/x_3)$ te kunnen berekenen:

$$p(y_1/x_3) = \tfrac{1}{8} \cdot \tfrac{1}{2} = \tfrac{1}{16} \ .$$

We hebben zojuist vermenigvuldigd met een $\frac{1}{2}$, omdat de kans dat bericht x_3 een ronde in de mix verblijft is gelijk aan $(1 - \frac{2}{4})$.

We moeten nog $p(y_5/x_5)$ berekenen. Deze kansen zijn gelijk aan $p(y_1/x_3)$ uit de vorige situatie, omdat we met dezelfde deelkansen zitten, namelijk:

- $p(\alpha(2,1)/x_5)$ komt overeen met $p(\alpha(1,1)/x_3)$, omdat $p(\alpha(2,1)/x_5)$ ook bestaat uit een $\frac{1}{2}$ voor de lege locaties en een $\frac{1}{2}$ dat één locatie wordt veroverd.
- $p(\beta_1/\alpha(2,1))$ komt overeen met $p(\beta_1/\alpha(1,1))$, omdat $p(\beta_1/\alpha(2,1))$ ook bestaat uit een $\frac{1}{2}$ dat het bericht uit locatie $\alpha(2,1)$ geflusht wordt en een $\frac{1}{2}$ dat dit bericht naar β_1 gaat.
- De kansen van $b(2)$ via $d(2)$ en uiteindelijk naar $s(2)$, hebben overeenkomstige waarden met de kansen uit $b(1)$ via $d(1)$ en uiteindelijk naar $s(1)$. Er is namelijk in deze route van ronde 1 naar ronde 2niets veranderd.

We komen dan op: $p(y_5/x_5) = \frac{1}{8}$. We berekenen nu $p(y_1)$ uit. Er geldt voor deze kansberekening het volgende:

$$\sum_{l=1}^{s} p(y_l) = 1 \ .$$

Dit levert voor $p(y_l)$ het volgende op:

$$p(y_l) = \tfrac{1}{8} \ .$$

Uiteraard nemen we hier aan dat $p(y_l)$ uniform is verdeeld. We kunnen met behulp van de omkeerregel $p(x_3/y_5)$ berekenen:

$$p(x_3/y_5) = \frac{p(y_5/x_3) \cdot p(x_3)}{p(y_5)} = \frac{\frac{1}{16} \cdot \frac{1}{12}}{\frac{1}{8}} = \frac{1}{24}$$

We berekenen nu nog $p(x_3/y_1)$ uit:

$$p(x_3/y_1) = \frac{p(y_1/x_3) \cdot p(x_3)}{p(y_1)} = \frac{\frac{1}{8} \cdot \frac{1}{12}}{\frac{1}{8}} = \frac{1}{12}$$

Tenslotte zullen we de kans $p(x_5/y_5)$ nog moeten berekenen:

$$p(x_5/y_5) = \frac{p(y_5/x_5) \cdot p(x_5)}{p(y_5)} = \frac{\frac{1}{8} \cdot \frac{1}{12}}{\frac{1}{8}} = \frac{1}{12}$$

Tevens berekenen we $p(\gamma_1/y_1)$ en $p(\gamma_3/y_5)$:

$$p(\gamma_1/y_1) = \frac{p(y_1/\gamma_1) \cdot p(\gamma_1)}{p(y_1)} = \frac{\frac{1}{4} \cdot \frac{1}{8}}{\frac{1}{8}} = \frac{1}{4}$$

$$p(\gamma_3/y_5) = \frac{p(y_5/\gamma_3) \cdot p(\gamma_3)}{p(y_5)} = \frac{\frac{1}{4} \cdot \frac{1}{8}}{\frac{1}{8}} = \frac{1}{4}$$

Dan kunnen we nu de zenders- en de ontvangersanonimiteit berekenen:

$$
\begin{aligned}
H(X/Y) =& -\sum_{j=1}^{4}\sum_{l=1}^{4} p(x_j/y_l)\cdot p(y_l)\cdot \log(p(x_j/y_l)) - \sum_{j=1}^{4}\sum_{l=5}^{8} p(x_j/y_l)\cdot p(y_l)\cdot \log(p(x_j/y_l)) - \\
& \sum_{j=5}^{6}\sum_{l=5}^{8} p(x_j/y_l)\cdot p(y_l)\cdot \log(p(x_j/y_l)) - \sum_{g=1}^{2}\sum_{l=1}^{4} p(\gamma_g/y_l)\cdot p(y_l)\cdot \log(p(\gamma_g/y_l)) - \\
& \sum_{g=3}^{4}\sum_{l=5}^{8} p(\gamma_g/y_l)\cdot p(y_l)\cdot \log(p(\gamma_g/y_l)) \\
=& -(\frac{1}{12}\cdot\frac{1}{8}\cdot\log(\frac{1}{12}))\cdot 4\cdot 4 - (\frac{1}{24}\cdot\frac{1}{8}\cdot\log(\frac{1}{24}))\cdot 4\cdot 4 - (\frac{1}{12}\cdot\frac{1}{8}\cdot\log(\frac{1}{12})) - \\
& (\frac{1}{4}\cdot\frac{1}{8}\cdot\log(\frac{1}{4}))\cdot 2\cdot 4 - (\frac{1}{4}\cdot\frac{1}{8}\cdot\log(\frac{1}{4}))\cdot 2\cdot 4 = 0.597 + 0.382 + 0.299 + \frac{1}{2} + \frac{1}{2} = 2.278 \\
H(X) =& -\sum_{j=1}^{6} p(x_j)\cdot\log(p(x_j)) - \sum_{g=1}^{4} p(\gamma_g)\cdot\log(p(\gamma_g)) \\
=& -(\frac{1}{12}\cdot\log(\frac{1}{12}))\cdot 6 - (\frac{1}{8}\cdot\log(\frac{1}{8}))\cdot 4 = 1.792 + 1\frac{1}{2} = 3.292 \\
ZA =& \frac{H(X/Y)}{H(X)} = \frac{2.278}{3.292} = 0.692
\end{aligned}
$$

$$
\begin{aligned}
H(Y/X) =& -\sum_{j=1}^{4}\sum_{l=1}^{4} p(y_l/x_j)\cdot p(x_j)\cdot \log(p(y_l/x_j)) - \sum_{j=1}^{4}\sum_{l=5}^{8} p(y_l/x_j)\cdot p(x_j)\cdot \log(p(y_l/x_j)) - \\
& \textstyle\sum_{j=5}^{6}\sum_{l=5}^{8} p(y_l/x_j)\cdot p(x_j)\cdot \log(p(y_l/x_j)) - \\
=& -(\frac{1}{8}\cdot\frac{1}{12}\cdot\log(\frac{1}{8}))\cdot 4\cdot 4 - (\frac{1}{16}\cdot\frac{1}{12}\cdot\log(\frac{1}{16}))\cdot 4\cdot 4 - (\frac{1}{8}\cdot\frac{1}{12}\cdot\log(\frac{1}{8}))\cdot 2\cdot 4 \\
=& (\frac{1}{2}+\frac{1}{3}+\frac{1}{4}) = 1\frac{1}{12} \\
H(Y) =& -\sum_{l=1}^{8} p(y_l)\cdot\log(p(y_l)) \\
=& -(\frac{1}{8}\cdot\log(\frac{1}{8}))\cdot 8 = 3 \\
OA =& \frac{H(Y/X)}{H(Y)} = \frac{1\frac{1}{12}}{3} = \frac{13}{36} = 0.361
\end{aligned}
$$

3.3.4 Formules

We zullen nu aan de hand van figuur 3.5 en 3.6 en de berekeningen met de thresholdmix en de thresholdpoolmix formules opstellen. Bij deze formules gelden ronde k als het moment waarop berichten arriveren bij de mix en ronde r als de ronde waarin dezelfde berichten worden geflusht.

Formule: $p(x_j)$ en $p(\gamma_g)$

Bij een thresholdmix geldt voor $p(x_j)$ waarbij $p(x_j)$ uniform is verdeeld:

$$p(x_j) = \frac{1}{2}\cdot\frac{1}{a}\ . \tag{3.25}$$

Waarvoor a geldt het totaal aantal berichten die door deze mix wordt ontvangen. Dan geldt bij een thresholdmix voor $p(\gamma_g)$ het volgende waarbij $p(\gamma_g)$ uniform is verdeeld:

$$p(\gamma_g) = \frac{1}{2}\cdot\frac{1}{c}\ . \tag{3.26}$$

Voor c geldt het aantal nepberichten die door Ng wordt gegenereerd. Voor een thresholdpoolmix komen we voor a en c tot de volgende vergelijkingen:

$$a = \sum_{i=0}^{r} a(i) \tag{3.27}$$

$$c = \sum_{i=0}^{r} c(i) \tag{3.28}$$

Hierin staat $c(i)$ voor het aantal nepberichten die in ronde i wordt gegenereerd en geflusht. Dan geldt onder de aanname dat $p(x_j)$ en $p(\gamma_g)$ uniform verdeeld zijn:

$$p(x_j) = \frac{1}{2 \cdot a} \tag{3.29}$$

$$p(\gamma_g) = \frac{1}{2 \cdot c} \tag{3.30}$$

Formule: $p(y_l/x_j)$

We zullen nu voor het definiëren van deze kans refereren naar onderdelen uit figuur 3.5 en we doen dit eerst voor de thresholdmix en later met de thresholdpoolmix. De vergelijking voor het berekenen van $p(y_l/x_j)$ is als volgt te schrijven:

$$p(y_l/x_j) = \sum_{o=1}^{r} \sum_{b=1}^{b} p(\alpha_o/x_j) \cdot p(\beta_b/\alpha_o) \cdot p(y_l/\beta_b) \ . \tag{3.31}$$

Deze kans bestaat uit drie delen die elk het volgende beschrijven:

1. $p(\alpha_o/x_j)$ is de kans dat bericht x_j een locatie verovert in Mix_a.
2. $p(\beta_b/\alpha_o)$ is de kans dat bericht x_j vanuit Mix_a een locatie verovert in Mix_b, om daar gemixt te kunnen worden met nepberichten.
3. $p(y_l/\beta_b)$ is tenslotte de kans dat bericht x_j vanuit een locatie in Mix_b geflusht wordt.

Voor de thresholdpoolmix bestaat deze kans uit vier delen. We zullen voor deze kans refereren naar onderdelen in figuur 3.6. We hebben namelijk een deel dat bericht x_j de route neemt van de zender naar de mix, een deel dat bericht x_j een aantal ronden in de mix verblijft, een deel dat bericht x_j gemixt wordt met een aantal nepberichten en een deel dat bericht x_j de route neemt van de mix naar de ontvanger. We komen voor $p(y_l/x_j)$ tot de volgende vergelijking:

$$p(y_l/x_j) = \sum_{o=1}^{r(r)} \sum_{b=1}^{b(r)} \sum_{d=1}^{d(r)} p(\alpha(k,o)/x_j) \cdot P(k,r) \cdot p(\beta_b/\alpha(r,o)) \cdot p(\delta_d/\beta_b) \cdot p(y_l/\delta_d) \ . \tag{3.32}$$

Hierin staan $r(r)$ voor het aantal locaties in de mix in ronde r, $b(r)$ voor het aantal geflushte berichten door Mix_a in ronde r en $d(r)$ voor het aantal locaties in Mix_b in ronde r. De vier delen die in bovenstaande formule zijn opgenomen, zijn:

1. $p(\alpha(k,o)/x_j)$ is de kans dat bericht x_j een locatie verovert in Mix_a.
2. $P(k,r)$ is de kans dat bericht x_j in ronde k bij Mix_a arriveert en tot en met ronde $(r-1)$ in Mix_a verblijft.

3. $p(\beta_b/\alpha(r,o)) \cdot p(\delta_d/\beta_b)$ zijn de kansen dat bericht x_j vanuit locatie $\alpha(r,o)$ in ronde r geflusht wordt naar Mix_b en dat bericht x_j gemixt wordt met nepberichten.

4. $p(y_l/\delta_d)$ is de kans dat bericht x_j tenslotte de mix verlaat.

De kans dat bericht x_j voor $(r-k)$ ronden in de mix verblijft, hebben we genoteerd als $P(k,r)$. Voor deze kans geldt het volgende:

$$P(k,r) = \prod_{j=k}^{r-1} (1 - P(j)) \ . \tag{3.33}$$

Waarvoor $P(j)$ geldt de kans waarop een bericht in ronde j geflusht wordt. Dit wil zeggen in ronde j geldende kans dat een bericht geflusht wordt (en niet $P(k,r)$).

Formule: $p(y_l)$

Een bericht y_l kan of een echt of een nepbericht zijn. De kans dat het een nepbericht is, is gelijk aan $\frac{c(i)}{s(i)}$ voor in ronde i. Voor $p(y_l)$ geldt dat:

$$\sum_{l=1}^{s} p(y_l) = 1 \ . \tag{3.34}$$

Voor s geldt het totaal aantal geflushte berichten door een mix. We nemen hierbij ook aan dat elke $p(y_l)$ uniform verdeeld is.

Formule: $p(x_j/y_l)$

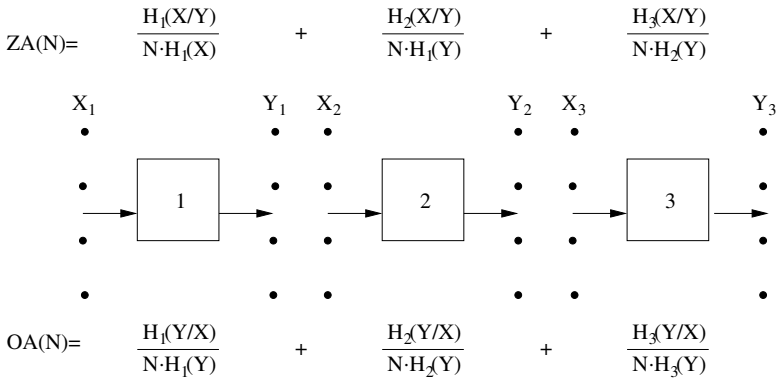
Deze laatste kans kan als volgt worden uitgedrukt indien voorgaande formules zijn gebruikt:

$$p(x_j/y_l) = \frac{p(y_l/x_j) \cdot p(x_j)}{p(y_l)} \ . \tag{3.35}$$

We hebben hier de omkeerregel van Bayes toegepast, omdat we de kans over x_j willen berekenen waarbij het optreden van bericht y_l heeft plaats gevonden. Hiervoor gebruiken we de kans over y_l waarin x_j is reeds opgetreden en de kansen waarin x_j en y_l zijn opgetreden.

3.4 De anonimiteit in een mixnetwerk

Het berekenen van de anonimiteit in een netwerk van mixen, stellen we voor dat elke mix die in het netwerk participeert uniform bijdraagt aan de totale anonimiteit. Dit wil zeggen een mix levert $\frac{1}{N}$ deel aan de totale anonimiteit, indien N het aantal mixnodes is die het netwerk vormen. We hebben in figuur 3.7 een voorbeeld van een cascade netwerk gepresenteerd waarin drie ($N = 3$) mixen participeren.



Figuur 3.7: Een cascade netwerk met drie mixen.

In het bovenstaande figuur wordt boven elke mix $\frac{1}{N}$-de deel van de totale zendersanonimiteit berekend en onder elke mix wordt $\frac{1}{N}$-de deel van de totale ontvangersanonimiteit berekend. Zo wordt voor de eerste mix zijn deel berekend, namelijk de conditionele entropie over X gegeven Y waarvoor geldt $X = X_1$ gegeven $Y = Y_1$. Tevens wordt $H(X)$ berekend waarvoor natuurlijk geldt X_1. We nemen aan dat het optreden van x_j uit $X = X_1$ uniform verdeeld is. Daardoor gelden de volgende kansen voor mix 1:

$$\begin{aligned} p(x_1) &= \frac{1}{X_1} \\ p(y_1) &= p(x_1) \cdot p(y_1/x_1) \\ p(x_1/y_1) &= \frac{p(y_1/x_1) \cdot p(x_1)}{p(y_1)} \end{aligned} \tag{3.36}$$

Hierin geldt voor $p(y_1/x_1)$ de kans dat bericht x_1 afkomstig van een zender naar mix 1 gaat en die een locatie in het geheugen van de mix verovert. Hier komt nog bij de kansen dat het bericht in de mix verblijft gedurende een aantal ronden als we over een pool mix praten. Tenslotte komtde kans er nog bij voor het flushen van bericht x_1 naar bericht y_1. We hebben de $p(y_1/x_1)$ eerder met voorbeelden behandeld. Indien we deze kansen hebben berekend, is het eenvoudig om $p(x_1/y_1)$ te vinden. Die vinden we door de omkeerregel van Bayes te gebruiken.

Berichten die naar Y_1 geflusht zijn, zijn in feite berichten die in X_2 bij mix 2 arriveren. Daarmee willen we zeggen dat het optreden van bericht y_1 overeen komt met het optreden van bericht x_2. We kunnen daardoor het berekenen van $p(y_1)$ gebruiken voor het berekenen van $p(x_2)$. De volgende kansen gelden voor mix 2:

$$\begin{aligned} p(x_2) &= p(y_1) \\ p(y_2) &= p(y_1) \cdot p(y_2/x_2) \\ p(x_2/y_2) &= \frac{p(y_2/x_2) \cdot p(x_2)}{p(y_2)} \end{aligned} \tag{3.37}$$

Net als bij mix 1 worden voor $p(y_2/x_2)$ de kansen berekend van zender naar de mix en het verblijf in de mix. Tenslotte wordt ook $p(x_2/y_2)$ met behulp van de omkeerregel berekend. We

zien dan dat de noemer van de deling voor mix 2, uit $H_1(Y)$ bestaat terwijl daar $H_2(X)$ had moeten staan. We verklaren dit door het feit dat het optreden van bericht y_1 overeenkomt met het optreden van x_2.

Voor de laatste mix in dit netwerk gelden de volgende kansen:

$$\begin{aligned} p(x_3) &= p(y_2) \\ p(y_3) &= p(y_2) \cdot p(y_3/x_3) \\ p(x_3/y_3) &= \frac{p(y_3/x_3) \cdot p(x_3)}{p(y_3)} \end{aligned} \tag{3.38}$$

Ook de noemer van de deling wordt berekend met behulp van van $p(y_2)$ van mix 2, want het optreden van y_2 komt overeen met het optreden van x_3.

Voor het berekenen van de ontvangersanonimiteit gelden dezelfde kansberekeningen per mix. Er geldt wel $p(y_l/x_j)$ in plaats van $p(x_j/y_l)$. We merken op dat het berekenen van $H(Y)$ gebruikt kan worden voor het berekenen van de zendersanonimiteit van de volgende mix.

Hoofdstuk 4

Conclusie

Vanuit de definities voor zenders- en ontvangersanonimiteit hebben we mathematische vergelijkingen opgesteld waarbij kennis uit de informatietheorie is gebruikt. Een maat voor anonimiteit die uit de informatietheorie gehanteerd wordt, is op dit moment de entropie of onzekerheid. We hebben uiteindelijk besloten om de anonimiteit in ons model te laten uitdrukken als een waarde tussen 0 en 1, die respectievelijk voor de maximale en de minimale anonimiteit staan. Deze waarde kan namelijk als een percentage geïnterpreteerd worden.

Dankzij de omkeerregel van Bayes kunnen we nu ook de ontvangersanonimiteit berekenen. Dit is eerst niet mogelijk geweest omdat we anders een uitspraak hebben moeten doen over de toekomstige output.

Om tot onze maat van anonimiteit te komen, hebben we een generieke kans, $P(k, r)$, gedefinieerd die voor een enkele mix opstelling gehanteerd kan worden. De verhouding is alleen te gebruiken voor mixen die een pool bezitten.

We hebben zowel de zenders- als de ontvangersanonimiteit uitgedrukt voor N mixen.

Ook hebben we twee methodes voor het plaatsen van nepberichten beschreven: plaatsing bij de output en plaatsing in de mix zelf. We hebben voor de eerste methode een berekening gegeven en uitgelegd.

Met onze definities kunnen we nu zowel de zenders- als de ontvangersanonimiteit berekenen voor een individuele mix en voor een cascade mixnetwerk. Tevens kunnen we de invloed op de anonimiteit bepalen van nepberichten volgens de Flush Nepberichten Methode.

Bibliografie

[BPS00] Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The disadvantages of free MIX routes and how to overcome them. In Hannes Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 30–45. Springer-Verlag, LNCS 2009, July 2000.
`http://www.tik.ee.ethz.ch/~weiler/lehre/netsec/Unterlagen/anon/disadvan%tages_berthold.pdf`.

[Cha81] David L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Commun. ACM*, 24(2):84–90, 1981.
`http://doi.acm.org/10.1145/358549.358563`.

[Cot] Lance Cottrell. Mixmaster & remailer attacks. Web.
`http://riot.eu.org/anon/doc/remailer-essay.html`.

[Dai96] Wei Dai. Pipenet 1.1. Usenet post, August 1996.
`http://www.eskimo.com/~weidai/pipenet.txt`
`http://www.eskimo.com/~weidai/freedom-attacks.txt`.

[DP04a] Claudia Díaz and Bart Preneel. Reasoning about the anonymity provided by pool mixes that generate dummy traffic. In *Proceedings of 6th Information Hiding Workshop (IH 2004)*, LNCS, Toronto, May 2004.
`http://www.esat.kuleuven.ac.be/~cdiaz/papers/cdiaz_ih.pdf.gz`.

[DP04b] Claudia Díaz and Bart Preneel. Taxonomy of mixes and dummy traffic. In *Proceedings of I-NetSec04: 3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems*, Toulouse, France, August 2004.
`http://www.esat.kuleuven.ac.be/~cdiaz/papers/cdiaz_inetsec.pdf.gz`.

[DS02] Roger Dingledine and Paul Syverson. Reliable mix cascade networks through reputation. In Matt Blaze, editor, *Proceedings of Financial Cryptography (FC '02)*. Springer-Verlag, LNCS 2357, March 2002.
`http://freehaven.net/doc/casc-rep/casc-rep.pdf`.

[DS03] Claudia Díaz and Andrei Serjantov. Generalising mixes. In Roger Dingledine, editor, *Proceedings of Privacy Enhancing Technologies workshop (PET 2003)*. Springer-Verlag, LNCS 2760, March 2003.
`http://www.esat.kuleuven.ac.be/~cdiaz/papers/DS03.ps.gz`.

[DSCP02] Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April

2002.
`http://www.esat.kuleuven.ac.be/~cdiaz/papers/tmAnon.ps.gz`.

[PK00] Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability, and pseudonimity - a proposal for terminology. In Hannes Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 1–9. Springer-Verlag, LNCS 2009, July 2000.

[SD02] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.
`http://www.cl.cam.ac.uk/~aas23/papers_aas/set.ps`.

[SDS02] Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, October 2002.
`http://freehaven.net/doc/batching-taxonomy/taxonomy.pdf`.

[Sha01] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, 2001.
`http://doi.acm.org/10.1145/584091.584093`.