Bart Mennink
Radboud Universiteit Nijmegen, en
Centrum Wiskunde & Informatica, Amsterdam
b.mennink@cs.ru.nl

De verjaardagsparadox in de cryptografie

Bart Mennink is een NWO Veni-laureaat werkzaam aan de Digital Security-groep van de Radboud Universiteit, Nijmegen. Mennink is gespecialiseerd in theoretische veiligheidsbewijzen. In dit artikel behandelt hij een wiskundig aspect binnen veiligheidsbewijzen, namelijk de verjaardagsparadox.

De verjaardagsparadox is zonder twijfel een van de bekendste paradoxen in de kansrekening: voor een groep van 23 mensen is de kans dat er twee mensen op dezelfde dag jarig zijn meer dan 50 procent (onder de aanname dat geboortedata uniform verdeeld zijn). De verjaardagsparadox speelt op veel plaatsen een rol binnen de cryptografie, en er bestaan aanvallen op cryptografische schema's die effectief gebruikmaken van de verjaardagsparadox.

Blokcijfers

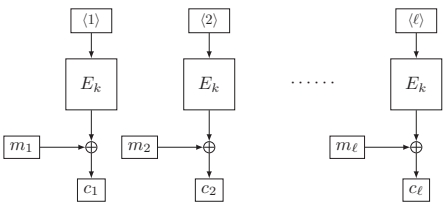
Het leeuwendeel van hedendaagse versleuteling vindt plaats door middel van 'blokcijfers'. Een blokcijfer $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ is een familie van permutaties over verzameling $\mathcal{M}$ geïndexeerd met een sleutel uit een sleutelverzameling $\mathcal{K}$. Dat wil zeggen, voor iedere $k \in \mathcal{K}$ is de functie $E(k,\cdot)$ een permutatie op $\mathcal{M}$. De functie E is niet geheim, en een aanvaller kent het algoritme. Echter, zodra een gebruiker volledig willekeurig een sleutel $k \in \mathcal{K}$ selecteert, dan zou het voor deze aanvaller met beperkte rekenkracht moeilijk moeten zijn om $E(k,\cdot)$ te onderscheiden van een volledig willekeurige permutatie op $\mathcal{M}$. Het bekendste voorbeeld van een blokcijfer is de 'Advanced Encryption Standard' (AES) van Daemen en Rijmen [5], en er wordt verondersteld dat AES inderdaad deze eigenschap heeft. Merk op dat, om deze eigenschap te hebben, $\mathcal{K}$ significant groter moet zijn dan de hoeveelheid evaluaties (sleutelgissingen) die een aanvaller kan maken.

AES verwerkt berichten van grootte 128 bits, oftewel $\mathcal{M} = \{0,1\}^{128}$. Om versleuteling van berichten van willekeurige lengte mogelijk te maken, wordt zo'n blokcijfer meestal in een zekere werkingsmode uitgevoerd. De bekendste werkingsmode is de 'tellermode' (Engels: 'counter mode'): voor een bericht $m = m_1 \cdots m_l$ bestaande uit $l \geq 1$ blokken van 128 bits, wordt de cijfertekst $c = c_1 \cdots c_l$ berekend als

$$c_i = E(k, \langle i \rangle) \oplus m_i \quad \text{voor } i = 1,\ldots,l, \qquad (1)$$

waarbij $\langle i \rangle$ de encoding van i als een bitstring van lengte 128 en $\oplus$ de bitsgewijze exclusieve disjunctie is (zie Figuur 1).

De kwaliteit van een dergelijke werkingsmode wordt doorgaans gemeten in hoeverre deze zich 'gedraagt' als een volledig willekeurige functie. Met andere woorden, idealiter is c moeilijk te onderscheiden van een uniform verdeelde string c uit $\{0,1\}^{128\ell}$. Hier wringt de schoen: als het bericht m uit l identieke blokken bestaat (ofwel, $m_1 = m_2 = \cdots = m_l$) dan zullen voor tellermode de cijferteksten $c_1,\ldots,c_l$ allemaal verschillend zijn (omdat $E(k,\cdot)$ een permutatie is), terwijl voor een uniform verdeelde cijfertekst c botsingen $c_i = c_j$ tussen de individuele blokken kunnen voorkomen. Counter mode kan zodoende worden onderscheiden van volledig willekeurig, op voorwaarde dat genoeg cijfertekstblokken beschikbaar zijn. Iets gedetailleerder: als een aanvaller ongeveer 2^{64} bericht-cijfertekstblokken leert, kan hij met grote succeskans correct gokken of deze data middels tellermode versleuteld zijn of middels een perfect veilige functie. Hij kan tellermode dus onderscheiden van willekeurig, en dus breken! Feitelijk wordt hier de verjaardagsparadox toegepast op een jaar met 2^{128} dagen, en waarbij het aantal mensen correspondeert met het aantal beschikbare cijfertekstblokken.



Figuur 1 Counter mode.

Sweet32-verjaardagsaanval

In bovenstaand geval hebben we nog altijd ongeveer 2^{64} cijfertekstblokken nodig om de werkingsmode van willekeurig te onderscheiden, en er zijn weinig gevallen waarbij een aanvaller daadwerkelijk zo veel blokken leert. Er bestaan echter blokcijfers met een kleinere berichtenruimte. De 'Data Encryption Standard' (DES), Triple-DES en Blowfish zijn voorbeelden van blokcijfers die een blokgrootte van 64 bits hebben. Als zo'n 64-bits blokcijfer wordt gebruikt, dan heeft bovenstaande aanval slechts ongeveer 2^{32} cijfertekstblokken nodig. Dit maakt bovenstaande aanval praktisch uitvoerbaar!

In 2016 introduceerden Bhargavan en Leurent [3] de zogeheten 'Sweet32-aanval' op TLS (het protocol dat websites beveiligt) en OpenVPN (een protocol dat kan worden gebruikt om VPN-verbindingen tot stand te brengen). Zij maakten gebruik van het feit dat (ten tijde van het onderzoek) Blowfish het standaardblokcijfer binnen OpenVPN was en het antieke Triple-DES nog altijd ondersteund werd door TLS, en wisten op vernuftige wijze de verjaardagsparadox toe te passen op de zogenaamde 'cijfertekstverketening' (Engels: 'cipher block chaining') werkingsmode.

We gaan cijfertekstverketening niet in detail uitleggen, maar de kern van de aanval bestaat eruit dat als er een botsing tussen twee cijferteksten $c_i = c_j$ is, de bitsgewijze exclusieve disjunctie $m_i \oplus m_j$ afgeleid kan worden. Dankzij deze eigenschap en het feit dat een bericht doorgaans een zekere hoeveelheid aan redundantie bevat, kan berichtdata worden afgeleid (de auteurs hebben in hun aanval nog een aantal andere technische moeilijkheden moeten overwinnen, die we voor het gemak even negeren). De aanval van Bhargavan en Leurent heeft ongeveer 785 GB aan getransporteerde data nodig om een HTTPS-verbinding te breken. Hun aanval heeft ertoe geleid dat meerdere bedrijven, waaronder eBay, de veiligheid van hun websites aangescherpt hebben.

De verjaardagsparadox verslaan

In sommige gevallen kan het probleem met de verjaardagsparadox gemakkelijk verholpen worden: bijvoorbeeld, voor tellermode met 128-bits AES is het voldoende om na ongeveer 2^{40} blokversleutelingen de sleutel te verversen: door met een nieuwe sleutel te werken wordt de aanvaller als het ware gereset. Voor algemene oplossingen zijn er zogeheten 'beyond birthday bound' veilige modes: schema's die veilig zijn zelfs als het aantal evaluaties $2^{n/2}$, waarbij n de blokgrootte is, overstijgt. Binnen de cryptografie vormt de richting van, vrij vertaald, 'nog-lang-niet-jarig-oplossingen' (als een aanvaller met zo'n schema te maken krijgt, dan is hij nog lang niet jarig), een onderzoeksrichting op zich.

Een bekende nog-lang-niet-jarig-oplossing is de zogenaamde 'som van permutaties', een functie die een $(n-1)$-bit waarde x naar een n-bit waarde transformeert middels twee oproepen naar het onderliggend blokcijfer:

$$F(k,x) = E(k,0\|x) \oplus E(k,1\|x), \qquad (2)$$

waarbij $\|$ concatenatie en $\oplus$ de bitsgewijze exclusieve disjunctie is. Na een reeks publicaties over deze constructie door Bellare e.a. [1, 2] en Lucks [8], slaagde Patarin [11, 13] erin om te bewijzen dat F niet onderscheidbaar is van een volledig willekeurige functie zolang een aanvaller ten hoogste $2^n/67$ evaluaties leert. Dit is een veel hogere grens dan $2^{n/2}$.

Als we nu terugkeren naar tellermode, vergelijking (1), en in plaats van een simpel blokcijfer de som van permutaties gebruiken, verkrijgen we (zie Figuur 2)

$$c_i = E(k,0\|\langle i \rangle) \oplus E(k,1\|\langle i \rangle) \oplus m_i \quad \text{voor } i = 1,\ldots,l, \qquad (3)$$

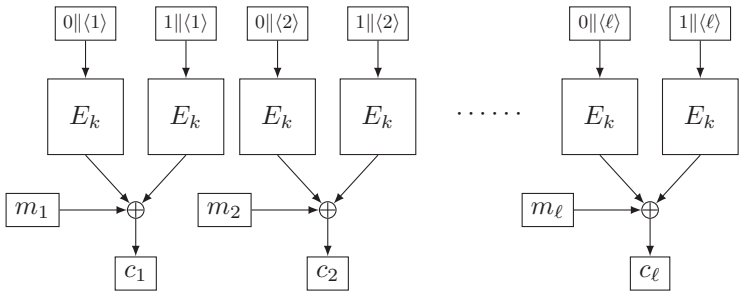
waarbij $\langle i \rangle$ in dit geval de encoding van i als een bitstring van lengte 127 noteert. Deze constructie is veilig zolang het aantal evaluaties onder $2^n/67$ blijft. Dit is een significante verbetering ten opzichte van de birthday bound $2^{n/2}$ voor (1). Feitelijk hebben we ervoor gezorgd dat botsingen in tellermode nu met (bijna) dezelfde distributie voorkomen als in een volledig willekeurige functie. De veiligheidswinst is echter niet gratis: de constructie in vergelijking (3) is tweemaal zo duur als de constructie in vergelijking (1).

De constructie in vergelijking (3) is in feite een speciaal geval van de zogeheten 'CENC'-werkingsmode van Iwata uit 2006 [6]. CENC kan uitgerekt worden: door de uitvoerwaarden van $E(k,0\|\cdot)$ te 'hergebruiken' voor meerdere datablokken, waarbij $E(k,1\|\cdot)$ wél voor ieder datablok een nieuwe invoer krijgt, kan CENC geoptimaliseerd worden zonder veel aan veiligheid te hoeven inboeten. Technisch gezien bekijken we de constructie waarbij de $E(k,0\|\cdot)$-waarde om de w versleutelde blokken ververst wordt:

$$c_i = E(k,0\|\langle\lceil i/w \rceil\rangle) \oplus E(k,1\|\langle i \rangle) \oplus m_i \quad \text{voor } i = 1,\ldots,l. \qquad (4)$$

Iwata, Mennink, en Vizár [7] hebben recentelijk bewezen dat deze constructie veilig is zolang het aantal evaluaties ten hoogste ongeveer $2^n/w$ is. Oftewel: als w groot wordt, zal de efficiëntie van de constructie die van de originele tellermode benaderen ($w+1$ blokcijferevaluaties om w datablokken te versleutelen) zonder enig significant veiligheidsverlies.

Een saillant detail is dat Iwata e.a. [7] niet écht een bewijs hebben geleverd dat CENC deze veiligheidsgrens bereikt, maar simpelweg opgemerkt dat het een direc-



Figuur 2 Counter mode gebaseerd op de som van permutaties.

te consequentie is van een generalisatie van Patarins resultaat voor de som van permutaties [13]. Dit resultaat was binnen het vakgebied onopgemerkt gebleven tot de publicatie van Iwata e.a., en Mennink en Neves [9] hebben dit resultaat recentelijk gemoderniseerd, gegeneraliseerd, en gebruikt om de veiligheid van andere nog-lang-niet-jarig-oplossingen te bewijzen. In essentie is dit resultaat een geïsoleerd combinatorisch probleem dat door Patarin de 'spiegelstelling' is genoemd.

Spiegelstelling

We gaan kijken naar vergelijkingen over n-bits onbekenden, voor zekere n. Beschouw $r \geq 1$ onbekenden $\{p_1,\ldots,p_r\}$ en een systeem van $q \geq 1$ lineaire vergelijkingen van de vorm

$$\begin{aligned} p_{a_1} \oplus p_{b_1} &= y_1, \\ p_{a_2} \oplus p_{b_2} &= y_2, \\ &\vdots \\ p_{a_q} \oplus p_{b_q} &= y_q, \end{aligned} \tag{5}$$

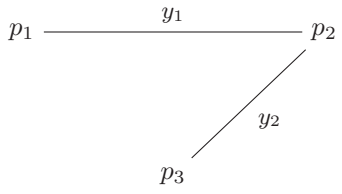
waarbij de indexering van de onbekenden impliciet gebeurt middels een surjectieve afbeelding

$$\varphi : \{a_1,b_1,\ldots,a_q,b_q\} \to \{1,\ldots,r\}$$

(dat wil zeggen, als $\varphi(a_1) = \varphi(a_2) = 1$, dan wordt met p_{a_1} en p_{a_2} de onbekende p_1 bedoeld). Het algemene doel van de spiegelstelling is om een ondergrens te vinden voor het aantal mogelijke oplossingen voor $\{p_1,\ldots,p_r\}$ zodanig dat de p_i's onderling verschillend zijn. Het probleem klinkt simpel, maar de gewenste ondergrens is sterk afhankelijk van de waarden $y_1,\ldots,y_q$, en de surjectie φ.

Simpel geval 1

Beschouw een simpel geval van drie onbekenden $\{p_1,p_2,p_3\}$ en twee vergelijkingen $p_1 \oplus p_2 = y_1$ en $p_2 \oplus p_3 = y_2$. We kunnen dit systeem van vergelijkingen visualiseren middels een graaf bestaande uit drie punten en twee lijnen gelabeld met de gekende waarden y_1 en y_2 (zie Figuur 3).



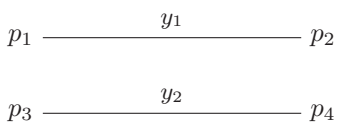
Figuur 3 Een systeem van twee vergelijkingen over drie onbekenden.

Het doel is om een ondergrens te vinden voor het aantal mogelijke oplossingen, waarbij p_1,p_2,p_3 onderling verschillend moeten zijn.

- Als $y_1 = 0$, dan is er geen oplossing voor het systeem: de eerste vergelijking impliceert immers dat we $p_1 = p_2$ moeten hebben. Net zo zijn er geen oplossingen als $y_2 = 0$ (omdat dit $p_2 = p_3$ impliceert) en als $y_1 = y_2$ (omdat dit $p_1 = p_3$ impliceert). In alle drie de gevallen bevat de graaf een pad waarvan de labels naar 0 optellen.
- Als $y_1,y_2 \neq 0$ en $y_1 \neq y_2$, is het aantal mogelijke oplossingen gemakkelijk vast te stellen: we hebben 2^n mogelijke waarden voor p_1. Zodra p_1 vastligt, dan ligt $p_2 = y_1 \oplus p_1$ en $p_3 = y_2 \oplus p_2 = y_2 \oplus y_1 \oplus p_1$ vast. In dit geval hebben we dus *exact* 2^n oplossingen.

Simpel geval 2

We kunnen nu een iets moeilijker geval bekijken, namelijk van vier onbekenden $\{p_1,p_2,p_3,p_4\}$ en twee vergelijkingen $p_1 \oplus p_2 = y_1$ en $p_3 \oplus p_4 = y_2$. Dit systeem kan gevisualiseerd worden middels de graaf in Figuur 4.



Figuur 4 Een systeem van twee vergelijkingen over vier onbekenden.

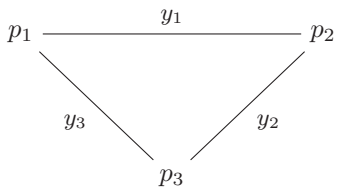
Net zoals in het vorige voorbeeld heeft dit systeem geen oplossing als $y_1 = 0$ of $y_2 = 0$. Als $y_1,y_2 \neq 0$ kunnen we op naïeve wijze het aantal oplossingen gaan tellen: we hebben 2^n mogelijke waarden voor p_1, en zodra p_1 vastligt, dan ligt $p_2 = y_1 \oplus p_1$ ook vast. De twee vergelijkingen zijn, in tegenstelling tot het vorige voorbeeld, niet gekoppeld, en de waarden voor p_3 en p_4 liggen niet vast. We weten echter dat onze keuze voor p_3 en p_4 moet voldoen aan de beperking dat $p_3 \notin \{p_1,p_2\}$ en $p_4 \notin \{p_1,p_2\}$. Vanwege de vergelijking $p_3 \oplus p_4 = y_2$, moet onze keuze voor p_3 dus voldoen aan

$$p_3 \notin \{p_1,p_2,y_2 \oplus p_1,y_2 \oplus p_2\},$$

hetgeen betekent dat we *tenminste* $2^n - 4$ mogelijke keuzes hebben voor p_3 (en iedere keuze legt p_4 vast). In totaal hebben we dus tenminste $2^n(2^n - 4)$ mogelijke oplossingen $\{p_1,p_2,p_3,p_4\}$.

Simpel geval 3

Een derde voorbeeld dat een andere beperking laat zien is gegeven in Figuur 5. In dit voorbeeld beschouwen we een systeem van drie onbekenden $\{p_1,p_2,p_3\}$ en drie vergelijkingen $p_1 \oplus p_2 = y_1$, $p_2 \oplus p_3 = y_2$ en $p_1 \oplus p_3 = y_3$.



Figuur 5 Een systeem van drie vergelijkingen over drie onbekenden.

We bekijken weer alleen het geval dat de y_i's ongelijk aan 0 en onderling verschillend zijn. We kunnen nu een onderscheid maken tussen twee gevallen:

- $y_1 \oplus y_2 \oplus y_3 = 0$. In dit geval bevat het systeem een overbodige vergelijking. We kunnen deze vergelijking (zonder beperking der algemeenheid de derde) weglaten en we zijn terug bij het eerste voorbeeld.
- $y_1 \oplus y_2 \oplus y_3 \neq 0$. In dit geval heeft het systeem duidelijk geen oplossingen: als we de drie vergelijkingen bij elkaar optellen vinden we $0 = y_1 \oplus y_2 \oplus y_3$.

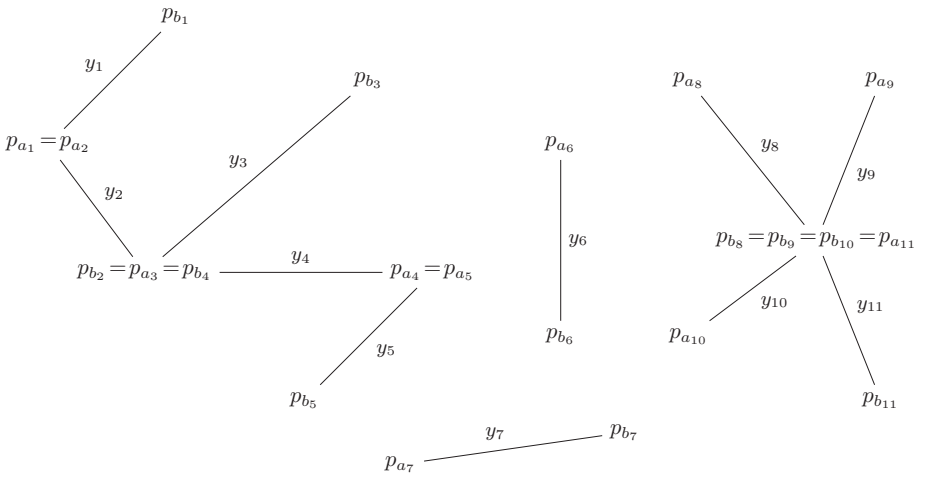
Algemene stelling

De drie voorbeelden geven een idee van wat er mis kan gaan met het systeem van vergelijkingen, en hoe we een ondergrens kunnen vinden in het geval er geen problemen zijn. We kunnen het algemene geval, het systeem van vergelijkingen (5), op soortgelijke wijze representeren middels een graaf bestaande uit r punten en q lijnen gelabeld met de waarden y_i. Zie Figuur 6 voor een willekeurig voorbeeld bestaande uit 15 punten en 11 lijnen.

In 2010 heeft Patarin [11,13] bewezen dat als het systeem van vergelijkingen (dat wil zeggen, de graaf die het systeem representeert) (i) geen cirkel bevat en (ii) geen pad bevat waarvan de labels optellen naar 0, het aantal mogelijke oplossingen voor $\{p_1,\ldots,p_r\}$ zodanig dat de p_i's onderling verschillend zijn ten minste

$$\frac{2^n(2^n-1)\cdots(2^n-r+1)}{2^{nq}}$$

is. De berekening geldt op voorwaarde dat n groot genoeg is en de graaf geen 'te grote' boom bevat, iets wat we in deze informele behandeling voor het gemak even



Figuur 6 Een systeem van 11 vergelijkingen over 15 onbekenden.

vergeten. De afleiding van het resultaat is zeer technisch van aard; we verwijzen naar [9] voor een gedetailleerde versie van de stelling. (Patarins krachtige stelling is algemener dan hier beschreven, zie ook [9].)

Toepassingen

De relatie tussen de spiegelstelling enerzijds en de som van permutaties en CENC anderzijds is, op het eerste oog, ver te zoeken. De relatie is, helaas, ook op het tweede oog ver te zoeken. Pas vanaf het derde oog, namelijk als we gaan kijken naar hoe veiligheidsbewijzen werken, wordt het verband duidelijk.

In voorgaande cryptografische schema's maakten we gebruik van een blokcijfer E met een geheime sleutel k. Onder de aanname dat E een goed blokcijfer is, zoals AES, kunnen we de functie $E(k,\cdot)$ modelleren als een volledig willekeurige permutatie π. Vanaf nu bekijken we een algemene functie

$$F^{\pi}(x) = \pi(f_1(x)) \oplus \pi(f_2(x)), \qquad (6)$$

waarbij f_1 en f_2 zekere functies zijn. Als we bijvoorbeeld $f_1(x) = 0\|x$ en $f_2(x) = 1\|x$ kiezen, verkrijgen we de som van permutaties van vergelijking (2). De interne constructie in CENC van vergelijking (4) correspondeert met $f_1(x) = 0\|\lceil x/w \rceil$ en $f_2(x) = 1\|x$.

De functie F^{π} wordt 'goed' geacht als haar uitvoeren ongeveer dezelfde distributie hebben als een volledig willekeurige functie ρ: als we een aanvaller toegang geven tot *ofwel* F^{π} *ofwel* ρ, moet het moeilijk zijn voor deze aanvaller om te kunnen raden tot welk orakel hij toegang heeft, zelfs als het aantal evaluaties dat hij leert, q, groot wordt. Feitelijk willen we dat de statistische afstand tussen de twee orakels, $\Delta(F^{\pi};\rho)$, verwaarloosbaar klein is (afhankelijk van de waarden q en n).

Er zijn verschillende manieren om aan te tonen dat deze afstand daadwerkelijk klein is, en we zullen gebruik maken van Patarins 'H-coëfficiënt-techniek' (de reden voor de aanwezigheid van de letter H is enigszins obscuur). Merk op dat $q \geq 1$ evaluaties van het systeem F^{π} of ρ samengevat kunnen worden in een tupel $\tau = \{(x_1,y_1),\ldots,(x_q,y_q)\}$ (waarbij, zonder beperking der algemeenheid, $x_i \neq x_j$ voor alle $i \neq j$). We kunnen voorts de set $\mathcal{T}$ definiëren van alle mogelijke tupels τ die een aanvaller theoretisch gezien zou kunnen verkrijgen. Uiteindelijk kunnen we een willekeurige partitie $\mathcal{T} = \mathcal{T}_{\text{goed}} \cup \mathcal{T}_{\text{slecht}}$ van deze verzameling beschouwen. De H-coëfficiënt-techniek toont nu het volgende aan (de afleiding is basiskansrekening, zie [4, 12]): als er δ,ε bestaan zodanig dat voor de slechte tupels,

$$\Pr(\rho \text{ genereert een tupel in } \mathcal{T}_{\text{slecht}}) \leq \delta, \quad (7)$$

en voor alle goede tupels $\tau \in \mathcal{T}_{\text{goed}}$,

$$\frac{\Pr(F^{\pi} \text{ genereert tupel } \tau)}{\Pr(\rho \text{ genereert tupel } \tau)} \geq 1-\varepsilon, \quad (8)$$

dan is $\Delta(F^{\pi};\rho) \leq \delta + \varepsilon$. F^{π} wordt veilig geacht als $\delta + \varepsilon \ll 1$ is. Als bijvoorbeeld $\delta + \varepsilon \approx q^2/2^n$, dan betekent dit dat een aanvaller ongeveer $q \approx 2^{n/2}$ evaluaties van de constructie moet leren om haar te kunnen onderscheiden van willekeurig.

Als $\tau = \{(x_1,y_1),\ldots,(x_q,y_q)\}$ een *gegeven* lijst van tupels van lengte q is, dan geldt voor de noemer in vergelijking (8) dat

$$\Pr(\rho \text{ genereert tupel } \tau) = 1/2^{nq}, \quad (9)$$

omdat de aanvaller $x_1,\ldots,x_q$ kan kiezen en de waarden $y_1,\ldots,y_q$ gegenereerd worden door de volledig willekeurige functie ρ. Voor de teller in vergelijking (8) kunnen we opmerken dat

$$\begin{aligned}\Pr(F^{\pi} &\text{ genereert tupel } \tau) \\ &= \frac{\left|\{\pi \mid F^{\pi}(x_i) = y_i\,(\forall i)\}\right|}{\left|\{\pi\}\right|} \\ &= \frac{\left|\{\pi \mid F^{\pi}(x_i) = y_i\,(\forall i)\}\right|}{2^n!} \\ &=: \frac{\Xi(\tau)}{2^n!}. \end{aligned} \quad (10)$$

Terug naar (7) en (8): het is nu onze taak om een geschikte partitie van $\mathcal{T}$ en zo klein mogelijke δ,ε te vinden, waarbij ε moet voldoen aan

$$\frac{\Xi(\tau)\,2^{nq}}{2^n!} \geq 1-\varepsilon. \quad (11)$$

Het van onder begrenzen van $\Xi(\tau)$ is, omdat $\tau = \{(x_1,y_1),\ldots,(x_q,y_q)\}$ een vastgelegde tupel is, een combinatorisch probleem, en Patarins spiegelstelling geeft een zekere ondergrens [13].

Inderdaad, ieder tupel (x_i,y_i) correspondeert met twee evaluaties van π, namelijk $\pi(f_1(x_i)) =: p_{a_i}$ en $\pi(f_2(x_i)) =: p_{b_i}$, en de gehele lijst τ definieert aldus q vergelijkingen van de vorm (5). Mogelijke relaties tussen de p_{a_i},p_{b_i} zijn afhankelijk van de functies f_1 en f_2.

Som van permutaties

Functie (6) correspondeert met de som van permutaties als we $f_1(x) = 0\|x$ en $f_2(x) = 1\|x$ kiezen. Omdat we (zonder beperking der algemeenheid) $x_i \neq x_j$ voor alle $i \neq j$ hebben, kunnen we concluderen dat $f_1(x_i) \neq f_1(x_j)$ en $f_2(x_i) \neq f_2(x_j)$ voor alle $i \neq j$, en dus dat $p_{a_1},\ldots,p_{a_q},p_{b_1},\ldots,p_{b_q}$ exact $2q$ verschillende onbekenden zijn. Het systeem van vergelijkingen (5) bevat aldus geen cirkel, het bevat alleen maar paden van lengte één. We kunnen de spiegelstelling toepassen *op voorwaarde dat* er geen lijn met label 0 is.

We zeggen derhalve dat een transcript $\tau = \{(x_1,y_1),\ldots,(x_q,y_q)\}$ slecht is als $y_i = 0$ voor $i \in \{1,\ldots,q\}$. De set $\mathcal{T}_{\text{slecht}}$ bestaat uit alle slechte tupels, en we vinden voor vergelijking (7):

$$\begin{aligned}\Pr(\rho \text{ genereert een tupel in } \mathcal{T}_{\text{slecht}}) &\leq q/2^n \\ &=: \delta.\end{aligned}$$

Voor de waarde ε vertrekken we vanuit (11). Patarins spiegelstelling vertelt ons dat het aantal oplossingen voor de p_{a_i} en p_{b_i}

ten minste

$$\frac{2^n(2^n-1)\cdots(2^n-2q+1)}{2^{nq}}$$

bedraagt. Iedere oplossing legt $2q$ invoer–uitvoer-waarden van π vast, en er zijn $(2^n-2q)!$ mogelijke permutaties voor iedere oplossing. Ofwel, $\Xi(\tau)=2^n!/2^{nq}$, en $\varepsilon=0$. We vinden dus dat $\Delta(F^\pi;\rho)\leq q/2^n$, hetgeen impliceert dat de som van permutaties zich als een volledig willekeurige functie gedraagt zolang $q\ll 2^n$.

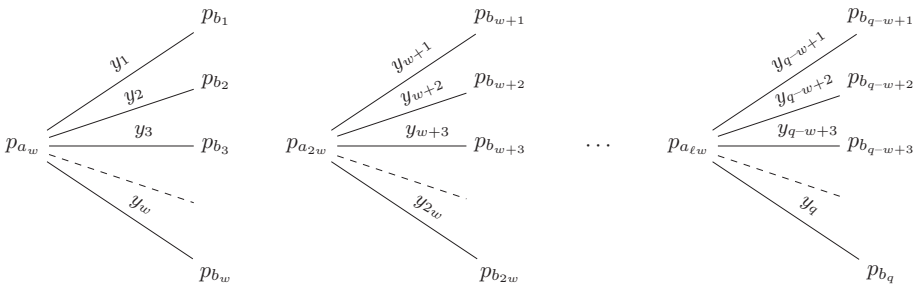
Constructie in CENC in vergelijking (4)

Voor de constructie in CENC in vergelijking (4) is $f_2(x)=1\|x$ en zijn om dezelfde reden $p_{b_1},\ldots,p_{b_q}$ verschillende onbekenden. De functie $f_1(x)=0\|\lceil x/w\rceil$ staat echter botsingen toe, wat wil zeggen dat $\{p_{a_1},\ldots,p_{a_q}\}$ in verzamelingen van ten hoogste w elementen kan worden gepartitioneerd zodanig dat $p_{a_i}=p_{a_j}$ als en alleen als ze in dezelfde set zitten. In het specifieke geval van CENC, waar simpelweg $x_i=i$, hebben we dus

$$\begin{aligned}
&p_{a_1}=p_{a_2}=\cdots=p_{a_w},\\
&p_{a_{w+1}}=p_{a_{w+2}}=\cdots=p_{a_{2w}},\\
&\vdots\\
&p_{a_{(l-1)w+1}}=p_{a_{(l-1)w+2}}=\cdots=p_{a_{q=lw}},
\end{aligned}\tag{12}$$

en $p_{a_w},p_{a_{2w}},\ldots,p_{a_{\ell w}}$ zijn exact l verschillende onbekenden, er voor het gemak van uitgaande dat $q=lw$. De graaf die correspondeert met het systeem is afgebeeld in Figuur 7.

Het is duidelijk dat het systeem van vergelijkingen wederom geen cirkel bevat, maar het bevat wel paden van lengte één én twee. We kunnen de spiegelstelling toepassen *op voorwaarde dat* er geen pad is waarvan de labels optellen naar 0. Met andere woorden, we willen dat $y_i\neq 0$ voor alle i, alsmede dat $y_i\neq y_j$ voor iedere boom in het bos.



Figuur 7 Visualisatie van het systeem van vergelijkingen corresponderende met CENC.

Op dezelfde manier als voor de som van permutaties, zeggen we dat een transcript $\tau=\{(x_1,y_1),\ldots,(x_q,y_q)\}$ slecht is als $y_i=0$ voor $i\in\{1,\ldots,q\}$, of als $(\lceil x_i/w\rceil=\lceil x_j/w\rceil\wedge y_i=y_j)$ voor $i\neq j$. De set $\mathcal{T}_{\text{slecht}}$ bestaat uit alle slechte tupels, en we vinden voor vergelijking (7)

$$\begin{aligned}
&\Pr(\rho\text{ genereert een tupel in }\mathcal{T}_{\text{slecht}})\\
&\quad\leq q/2^n+\binom{w}{2}l/2^n=:\delta.
\end{aligned}$$

We vinden op soortgelijke wijze dat $\varepsilon=0$, en dus dat

$$\begin{aligned}
\Delta(F^\pi;\rho)&\leq q/2^n+\binom{w}{2}l/2^n\\
&\leq q/2^n+wq/2^{n+1}
\end{aligned}$$

(omdat $l=q/w$). Dit impliceert dat CENC zich als een volledig willekeurige functie gedraagt zolang $q\ll 2^n/w$.

Verdere toepassingen

De som van permutaties en CENC zijn simpele toepassingen van de spiegelstelling. Het eerste bewijs van de spiegelstelling stamt uit 2005 [10], met een exacte bound uit 2010 [13], en Patarin heeft het voornamelijk toegepast op de som van permutaties en op zogeheten Feistelschema's. In onze recentelijke modernisatie van de spiegelstelling [9] passen we het verder toe op twee andere constructies om een willekeurige functie op basis van blokcijfers na te bootsen (EDM en EDMD), en gebruiken we de techniek om berichtauthenticatiemethodes optimaal veilig te bewijzen. De veiligheid van al deze schema's volgt uit de spiegelstelling, en het lijkt erop dat deze nog veel meer toepassingen gaat krijgen

Referenties

1 M. Bellare en R. Impagliazzo, A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion, *Cryptology ePrint Archive*, Report 1999/024 (1999).

2 M. Bellare, T. Krovetz en P. Rogaway, Luby–Rackoff backwards: Increasing security by making block ciphers non-invertible, in: K. Nyberg, ed., *EUROCRYPT '98*, LNCS 1403, Springer, 1998, pp. 266–280.

3 K. Bhargavan en G. Leurent, On the practical (in-)security of 64-bit block ciphers: Collision attacks on HTTP over TLS and OpenVPN, in: E.R. Weippl, S. Katzenbeisser, C. Kruegel, A.C. Myers en S. Halevi, eds., *ACM CCS 2016*, ACM, 2016, pp. 456–467.

4 S. Chen en J.P. Steinberger, Tight security bounds for key-alternating ciphers, in: P.Q. Nguyen en E. Oswald, eds., *EUROCRYPT 2014*, LNCS 8441, Springer, 2014, pp. 327–350.

5 J. Daemen en V. Rijmen, The design of Rijndael: AES—The Advanced Encryption Standard, *Information Security and Cryptography*, Springer, 2002.

6 T. Iwata, New blockcipher modes of operation with beyond the birthday bound security, in: M.J.B. Robshaw, ed., *FSE 2006*, LNCS 4047, Springer, 2006, pp. 310–327.

7 T. Iwata, B. Mennink en D. Vizár, CENC is optimally secure, *Cryptology ePrint Archive*, Report 2016/1087 (2016).

8 S. Lucks, The sum of PRPs is a secure PRF, in: B. Preneel, ed., *EUROCRYPT 2000*, LNCS 1807, Springer, 2000, pp. 470–484.

9 B. Mennink en S. Neves, Encrypted Davies–Meyer and its dual: Towards optimal security using mirror theory, in: J. Katz en H. Shacham, eds., *CRYPTO 2017*, LNCS, Springer 2017, te verschijnen.

10 J. Patarin, On linear systems of equations with distinct variables and small block size, in: D. Won en S. Kim, eds., *ICISC 2005*, LNCS 3935, Springer, 2005, pp. 299–321.

11 J. Patarin, A proof of security in $O(2^n)$ for the Xor of two random permutations, in: R. Safavi-Naini, ed., *ICITS 2008*, LNCS 5155, Springer, 2008, pp. 232–248.

12 J. Patarin, The 'coefficients H' technique, in: R.M. Avanzi, L. Keliher en F. Sica, eds., *SAC 2008*, LNCS 5381, Springer, 2008, pp. 328–345.

13 J. Patarin, Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography, *Cryptology ePrint Archive*, Report 2010/287 (2010).