# Inhoud | Contents

## Artikelen | Features

### Wiskunde in de cryptografie
*Onder redactie van Bart Mennink en Marc Stevens*



**Column | Casper sees a chance**



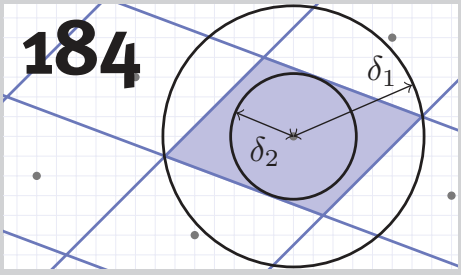
**Onderwijs | Education**



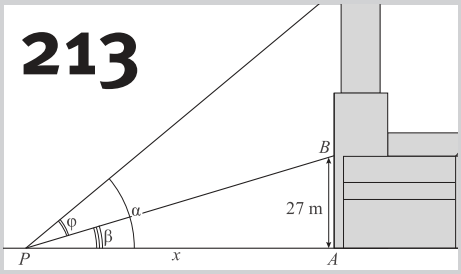
## Rubrieken | Departments



**Uitgelicht**




Léo Ducas behandelt rooster-gebaseerde versleuteling en de invloed van kwantumcomputers hierop.




Wim Caspers bespreekt het examen vwo wiskunde B.

## Colofon

Het Nieuw Archief voor Wiskunde is een uitgave van het Koninklijk Wiskundig Genootschap en verschijnt vier maal per jaar. Het tijdschrift richt zich op eenieder die zich beroepsmatig met wiskunde bezighoudt, als academisch of industrieel onderzoeker, student, leraar, journalist of beleidsmaker. Het stelt zich ten doel te berichten over ontwikkelingen in de wiskunde in het algemeen en in de Nederlandse wiskunde in het bijzonder.

ISSN 0028-9825

Adres
Nieuw Archief voor Wiskunde
Centrum Wiskunde & Informatica
Postbus 94079, 1090 GB Amsterdam
tel. 020-5924288
redactie@nieuwarchief.nl
www.nieuwarchief.nl

Hoofdredactie
Robbert Fokkink (r.j.fokkink@tudelft.nl)
Jan van Neerven (j.m.a.m.vanneerven@tudelft.nl)

Eindredactie/Bladmanager
Fred Drissen (fred@nieuwarchief.nl)

Redactie
Hans Cuypers (TU/e), Jan Draisma (Universität Bern en TU/e), Geertje Hek (UvA), Jan Hogendijk (UU), Teun Koetsier (VU), Sjoerd Rienstra (TU/e), Nicolaos Starreveld (UvA), Hans Sterk (TU/e), Mark Timmer (UT)

Problemenrubriek
problems@nieuwarchief.nl

Bureauredactie
Claartje Ottens

Illustraties
Ryu Tajiri

Vormgeving
Susanne Laws (omslag, begeleiding binnenwerk)
Kitty Molenaar (ontwerp)

Druk
Veldhuis Media

Advertenties
advertenties@nieuwarchief.nl

---

Koninklijk Wiskundig Genootschap
www.wiskgenoot.nl

Platform Wiskunde Nederland
www.platformwiskunde.nl

European Mathematical Society
www.euro-math-soc.eu

International Mathematical Union
www.mathunion.org

## Koninklijk Wiskundig Genootschap

Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaren van de wiskunde. In 1778 opgericht onder de zinspreuk: 'Een onvermoeide arbeid komt alles te boven' is het 's werelds oudste nationale wiskundegenootschap. Leden ontvangen het Nieuw Archief voor Wiskunde als onderdeel van hun lidmaatschap.

**Bestuur**
Erik van den Ban (UU, voorzitter), Joke Blom (CWI, penningmeester), Sonja Cox (UvA, secretaris), Theo van den Bogaart (HU), Marije Elkenbracht (ABN AMRO), André Ran (VU), Herman te Riele (CWI), Mark Veraar (TUD), Jan Wiegerinck (UvA)

**Secretariaat**
Joke Blom, CWI
Postbus 94079, 1090 GB Amsterdam
wiskgenoot@wiskgenoot.nl
www.wiskgenoot.nl

**Ledenadministratie KWG / Abonnementen**
Postbus 1064, 7940 KB Meppel
admin@wiskgenoot.nl, tel. 085-0160252

**Contributie**
De contributie voor gewone leden bedraagt in 2016 € 90 per jaar. Voor gepensioneerden en werklozen € 45. Voor studenten, aio's en oio's van een Nederlandse universiteit of hbo-opleiding € 35. Studenten en pas afgestudeerden kunnen een jaar lang gratis lid worden. Voor leden van de wiskundige vakverenigingen VvS+OR, NVvW en NVORWO geldt het gereduceerd tarief van € 70. Leden die het Nieuw Archief voor Wiskunde niet willen ontvangen, betalen slechts € 50 contributie. Losse nummers zijn te bestellen voor € 20. Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor € 100 (Nederland), € 110 (Europa) of € 112,50 (buiten Europa).

**Members of foreign societies**
Members of *SIAM*, the *Société Mathématique de France*, the *Gesellschaft für Angewandte Mathematik und Mechanik*, the *Deutsche Mathematiker-Vereinigung*, and of the *American*, *Australian*, *Belgian*, *Indian*, *London*, *Spanish*, and *South-African Mathematical Societies* living outside of the Netherlands pay € 70 instead of the full membership fee of € 90 a year. Conversely, these foreign societies, as well as the VvS+OR and the NVORWO, have reduced membership fees for KWG members. A postage charge of € 10 is added for members living abroad but in Europe, and a charge of € 12,50 for members living outside of Europe.

**Exchange subscriptions**
Koninklijk Wiskundig Genootschap Library
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB Amsterdam
KWG-exchange@cwi.nl

## Informatie voor auteurs

Kopij
Het Nieuw Archief voor Wiskunde is een gerediger tijdschrift. Kopij dient per e-mail te worden aangeboden aan de hoofdredactie. Uitgebreide informatie en auteursinstructies kunt u vinden op www.nieuwarchief.nl.

Volgende nummers

| *nummer* | *maand verschijnen* | *deadline kopij* |
|---|---|---|
| 18(4) | december 2017 | verstreken |
| 19(1) | maart 2018 | 01-11-2017 |
| 19(2) | juni 2018 | 01-02-2018 |
| 19(3) | september 2018 | 01-05-2018 |

## Instituutsleden

De publicaties van het Koninklijk Wiskundig Genootschap worden mede mogelijk gemaakt door de bijdragen van de volgende instituten.

- Centrum Wiskunde & Informatica
- Rijksuniversiteit Groningen
- Universiteit van Amsterdam
- Universiteit Leiden
- Vrije Universiteit
- Universiteit Utrecht
- Technische Universiteit Eindhoven
- Eurandom
- Technische Universiteit Delft
- Universiteit Twente
- Radboud Universiteit Nijmegen
- Freudenthal Instituut

## Op het omslag

Een klassieke vorm van 'versleuteling'. Dit themanummer gaat over 'Wiskunde in de cryptografie'. De moderne cryptografie bestudeerd een grote verscheidenheid aan problemen die veiligheid tegen kwaadaardige entiteiten vergen, en probeert die op te lossen met efficiënte algoritmen.

# Wiskunde in de cryptografie

Dit themanummer van het *Nieuw Archief voor Wiskunde* staat in het teken van de wiskunde achter cryptografie. Alhoewel van oudsher methoden voor de geheimhouding van berichten tegen derde partijen en voor de authenticiteit van berichten de meest belangrijke voorbeelden van cryptografie zijn, bestudeert de moderne cryptografie een grotere verscheidenheid aan problemen die veiligheid tegen kwaadaardige entiteiten vergen, en probeert die op te lossen met efficiënte algoritmen.

Een dergelijk voorbeeld is *secure computation*, waarbij een groep personen berekeningen wil uitvoeren over ieders geheime invoer waarbij iedereen slechts de uitkomst leert, zeg de som van individuele stemmen. Hoewel iedereen samenwerkt om deze uitkomst te berekenen, leert niemand enig andere informatie dan de uitkomst, zelfs indien een beperkte deelgroep kwaadaardig samenwerkt.

De meest klassieke richting binnen de cryptografie is symmetrische cryptografie. Het idee binnen symmetrische cryptografie is dat entiteiten een geheime sleutel delen en deze gebruiken voor communicatie. De symmetrische versleutelingsalgoritmen zijn met afstand de meest efficiënte binnen de cryptografie. Veiligheid van symmetrische cryptografische systemen wordt tegenwoordig ondersteund met generieke veiligheidsbewijzen en met cryptanalyse, dat zwakheden in de veiligheid van dergelijke oplossingen probeert aan te tonen. Dit nummer bevat drie artikelen over moderne symmetrische cryptografie. Joan Daemen geeft een uiteenzetting van permutatie-gebaseerde cryptografie, en hoe deze vernieuwde richting de toekomst van symmetrische cryptografie kan waarborgen. Bart Mennink behandelt het aspect van veiligheid van versleutelingsalgoritmen en de relatie tussen cryptografie en de verjaardagsparadox. Berry Schoenmakers legt uit hoe je optimaal een hash chain kunt gebruiken voor authenticatie.

Sinds de verschijning van het Turing-award winnende resultaat van Whitfield Diffie en Martin Hellman uit 1976 vormt publieke-sleutelcryptografie een belangrijke richting binnen de cryptografie. Het basisidee binnen deze richting is dat iedere geheime sleutel gepaard gaat met een publieke (niet-geheime) sleutel: een buitenstaander kan de publieke sleutel gebruiken om data te versleutelen op dusdanige wijze dat alleen de eigenaar van de geheime sleutel de gegevens kan ontsleutelen. Robert Granger beschrijft het probleem van discrete logaritmes, legt het belang van deze voor de cryptografie uit, en geeft een uiteenzetting van enkele recente doorbraken en records in het berekenen van discrete logaritmes. Eric Verheul en Bart Jacobs tonen aan hoe de klassieke ElGamal-publieke-sleutelversleuteling gebruikt kan worden voor privacybeschermende versleuteling en authenticatie, en leggen uit hoe dit protocol gebruikt gaat worden in het Nederlandse eID-systeem.

De — momenteel nog theoretische — kwantumcomputers bieden veel nieuwe mogelijkheden voor cryptanalyse. Alhoewel symmetrische cryptografie enigszins weerbaar is tegen kwantumcomputers (in veel gevallen voldoet het om de sleutellengte te verdubbelen), hebben kwantumcomputers de potentie om vernietigende aanvallen op de voornaamste schema's voor publieke-sleutelcryptografie te realiseren: een algoritme van Shor uit 1994 kan bijvoorbeeld het discrete-logaritmeprobleem in polynomiale tijd oplossen. Léo Ducas behandelt bepaalde vormen van rooster-gebaseerde versleuteling, en wat voor invloed kwantumcomputers hierop kunnen hebben.

Een alternatieve richting is om kwantumtechnieken effectief te gebruiken om cryptografie mee te *bedrijven*. Christian Schaffner geeft een toegankelijke uiteenzetting over wat kwantumcryptografie behelst en hoe het veilige communicatie in de aanwezigheid van kwantumaanvallers mogelijk maakt. David Elkouss behandelt vormen van kwantumsleuteluitwisseling waarvan de veiligheid niet afhangt van de gebruikte kwantumapparatuur. Matthijs Coster bekijkt supersinguliere elliptische krommen, en hoe deze gebruikt kunnen worden voor sleuteluitwisseling.

**Bart Mennink**, *Radboud Universiteit Nijmegen, en CWI, Amsterdam*
**Marc Stevens**, *CWI, Amsterdam*
gastredacteuren

# Agenda

| Upcoming Events

## september 2017

4–8, 11–15 en 18–22 september 2017

❑ **Stochastic Activity Month**

Tijdens de Stochastic Activity Month (SAM) zullen er drie workshops zijn, beginnend met de Young European Probabilists workshop (YEP).

***plaats*** *Eurandom, Eindhoven*

***info*** *eurandom.nl*

11–15 september 2017

❑ **Randomness and Graphs: Processes and Structures**

Een workshop over random graphs, discrete processen en algoritmen op grafen.

***plaats*** *Eurandom, Eindhoven*

***info*** *eurandom.nl*

12–15 september 2017

❑ **Statistical Theory and the Real World**

Een workshop georganiseerd door Giulia Cereda, Peter Grünwald en Aad van der Vaart.

***plaats*** *Lorentz Center@Oort, Leiden*

***info*** *lorentzcenter.nl*

15 september 2017

❑ **Finale Nederlandse Wiskunde Olympiade**

De prijswinnaars van deze finale worden uitgenodigd om deel te nemen aan een speciaal trainingsprogramma voor de internationale wiskundewedstrijden.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *wiskundeolympiade.nl*

18–22 september 2017

❑ **Community Detection and Network Reconstruction**

Een workshop georganiseerd door Luva Avena (Leiden), Jop Briët (Amsterdam, CWI), Rui Castro (Eindhoven) en Diego Garlaschelli (Leiden).

***plaats*** *Eurandom, Eindhoven*

***info*** *eurandom.nl*

19–20 september 2017

❑ **CWI-INRIA workshop**

Het CWI en INRIA zijn in 2016 een samenwerking aangegaan onder de naam CWI-Inria International Lab (IIL). De IIL steunt wetenschappelijk onderzoek en internationale samenwerking.

***plaats*** *CWI, Amsterdam*

***info*** *project.inria.fr/inriacwi*

29 september 2017

❑ **Wiskundetoernooi**

Wiskundetoernooi voor teams van middelbare scholieren. Een team bestaat uit 4 of 5 leerlingen uit 5–6 vwo en een begeleider. Thema dit jaar is 'Foutverbeterende codes'.

***plaats*** *Radboud Universiteit, Nijmegen*

***info*** *ru.nl/wiskundetoernooi*

30 september 2017

❑ **Junior Wiskunde Olympiade**

Jaarlijkse landelijke wiskundewedstrijd voor leerlingen uit de onderbouw van havo en vwo.

***plaats*** *Vrije Universiteit, Amsterdam*

***info*** *wiskundeolympiade.nl*

## oktober 2017

4–6 oktober 2017

❑ **42ste Woudschoten Conferentie**

Een jaarlijkse conferentie van de Nederlands-Vlaamse Numerieke Analyse-groepen, georganiseerd door de Werkgemeenschap Scientific Computing (WSC).

***plaats*** *Woudschoten, Zeist*

***info*** *wsc.project.cwi.nl/events*

7 oktober 2017

❑ **CWI Open dag**

Het CWI doet mee aan de Amsterdam Science Park Open Dag tijdens het Weekend van de Wetenschap.

***plaats*** *CWI, Amsterdam*

***info*** *cwi.nl*

7 oktober 2017

❑ **Het meten van de wereld**

De Werkgroep Geschiedenis organiseert het 23ste symposium getiteld 'Het meten van de wereld' over de geschiedenis van de wiskunde bij het meten van aarde en heelal.

***plaats*** *Academiegebouw, Utrecht*

***info*** *nvvw.nl*

23–27 oktober 2017

❑ **On Growth and Form**

Een workshop in het kader van het honderdjarig bestaan van de eerste druk van het wereldberoemde boek *On Growth and Form* van D' Arcy Thompson.

***plaats*** *Lorentz Center@Snellius, Leiden*

***info*** *lorentzcenter.nl*

*Suggesties voor deze agenda zijn welkom.*
*Redacteur: Nicolaos Starreveld*
*agenda@nieuwarchief.nl*

## november 2017

4 november 2017
❑ **Jaarvergadering NVvW**
Jaarlijkse vergadering en studiedag van de Nederlandse Vereniging van Wiskundeleraren.
***plaats*** *Ichthus College, Veenendaal*
***info*** *nvvw.nl*

10 november 2017
❑ **Prijsuitreiking van de Nederlandse Wiskunde Olympiade**
De prijsuitreiking van de finale van de Nederlandse Wiskunde Olympiade.
***plaats*** *Eindhoven*
***info*** *wiskundeolympiade.nl*

17 november 2017
❑ **Voorronde Wiskunde A-lympiade**
Voorronde van de jaarlijkse competitie voor leerlingen uit de bovenbouw havo/vwo, die wiskunde A in hun pakket hebben.
***plaats*** *eigen school*
***info*** *www.fi.uu.nl/alympiade*

17 november 2017
❑ **Wiskunde B-dag**
De Wiskunde B-dag is een 1-daagse opdracht voor op school voor teams uit 5-havo en 5/6-vwo met wiskunde B in hun profiel. De leerlingen werken de hele dag aan een wiskundig probleem.
***plaats*** *eigen school*
***info*** *uu.nl/onderwijs/wiskunde-b-dag*

## december 2017

4–8 december 2017
❑ **Random Dynamical Systems**
Een workshop mede georganiseerd door Ale Jan Homburg (Amsterdam).
***plaats*** *Lorentz Center@Oort, Leiden*
***info*** *lorentzcenter.nl*

11–15 december 2017
❑ **Future and Emerging Mathematical Technologies in Europe**
Een workshop mede georganiseerd door Wil Schilders (Eindhoven), Evgeny Verbitskiy (Leiden) en Kees Vuik (Delft).
***plaats*** *Lorentz Center@Oort, Leiden*
***info*** *lorentzcenter.nl*

11–15 december 2017
❑ **YEQT XI 'Winter school on energy systems'**
Elfde editie van de Young European Queueing Theorists (YEQT) workshop.
***plaats*** *Eurandom, Eindhoven*
***info*** *eurandom.nl*

## januari 2018

13 januari 2018
❑ **Wintersymposium KWG 2018**
Jaarlijks symposium van het KWG.
***plaats*** *Academiegebouw, Utrecht*
***info*** *wiskgenoot.nl*

17–19 januari 2018
❑ **LNMB-conferentie**
43ste conferentie over de wiskunde van besliskunde georganiseerd door het Landelijk Netwerk Mathematische Besliskunde (LNMB).
***plaats*** *Congrescentrum De Werelt, Lunteren*
***info*** *www.lnmb.nl*

22 januari–1 februari 2018
❑ **Nederlandse Wiskunde Olympiade**
De eerste ronde van de jaarlijkse landelijke wiskundewedstrijd voor leerlingen uit de onderbouw van havo en vwo.
***plaats*** *eigen school*
***info*** *wiskundeolympiade.nl*

22–24 januari 2018
❑ **17th Winter School Mathematical Finance**
Winterschool over financiële wiskunde met onder andere minicursussen van Emmanuel Gobet (Palaiseau) en Sebastian Jaimungal (Toronto).
***plaats*** *Congrescentrum De Werelt, Lunteren*
***info*** *staff.fnwi.uva.nl/p.j.c.spreij/winterschool/winterschool.html*

25–26 januari 2018
❑ **36ste Panama-conferentie**
De jaarlijkse Panama-conferentie richt zich op alle aspecten van leren en onderwijzen van rekenen-wiskunde.
***plaats*** *Congrescentrum Koningshof, Veldhoven*
***info*** *panamaconferentie.sites.uu.nl*

29 januari–2 februari
❑ **Study Group Mathematics with Industry**
Jaarlijkse studiegroep waarin wiskunde en industrie samen problemen oplossen, dit jaar georganiseerd door TU/e.
***plaats*** *Technische Universiteit Eindhoven*
***info*** *swi-wiskunde.nl*

## februari 2018

7 februari 2018
❑ **Onderbouw Wiskunde Dag**
Activiteit van het Freudenthal Instituut. Leerlingen van 3 havo/vwo werken in teams gedurende een dag aan een 'grote' wiskundige (denk)opdracht waarin probleemoplossen centraal staat.
***plaats*** *eigen school*
***info*** *uu.nl/onderwijs/onderbouwwiskundedag*

7–8 februari 2018
❑ **Nationale Wiskunde Dagen**
24ste editie van de tweedaagse conferentie voor wiskundeleraren.
***plaats*** *Noordwijkerhout*
***info*** *uu.nl/nationale-wiskunde-dagen*

## april 2018

3–4 april 2018
❑ **Nederlands Mathematisch Congres**
Groots opgezet jaarlijks congres, dit jaar voor het eerst in Veldhoven. Er wordt een aantrekkelijk programma geboden voor heel wiskundig Nederland.
***plaats*** *Veldhoven*
***info*** *wiskgenoot.nl*

# Nieuws

| News

*Deze rubriek is een kroniek van wiskundige activiteiten in Nederland. Toekomstige activiteiten worden aangekondigd en van voorbije activiteiten wordt verslag gedaan. Wilt u uw aankondiging of verslag in deze rubriek geplaatst zien? Stuur ons dan uw bijdrage, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.*
*Redacteur: Nicolaos Starreveld*
*nieuws@nieuwarchief.nl*

**Maryam Mirzakhani overleden**

De Iraanse professor Maryam Mirkzakhani (Stanford University) kreeg in 2014 als eerste vrouw ooit de Fieldsmedaille toebedeeld. Zij kreeg de prijs voor haar bijdragen in onder andere topologie, Riemann-oppervlakken en dynamica. Helaas is ze na een lange strijd tegen borstkanker op 15 juli 2017 overleden. Maryam Mirzakhani kreeg als talentvolle tiener nationale bekendheid nadat ze op de Internationale Wiskunde Olympiade van 1994 en 1995 een gouden medaille won. In zijn blog schrijft Terence Tao een eerbetoon aan Mirzakhani. Hij herinnert zich hun eerste ontmoeting. Het was in 2010 tijdens een serie lezingen aan Stanford University, de eerste over Perelmans bewijs van het vermoeden van Poincaré en de tweede over de theorie van stochastische matrices. “Er zat een jonge vrouw op de eerste rij die na afloop van beide lezingen inzichtelijke vragen stelde; ik ontdekte pas later dat het Mirzakhani was. Ik kan me helaas niet herinneren wat de vragen precies waren, maar ik kan me wel herinneren dat ze een mooie interpretatie van beide onderwerpen kon geven met het gebruik van dynamische systemen.” Ze leverde onder andere enkele belangrijke bijdragen aan de theorie van moduliruimtes van Riemann-oppervlakken. *diverse bronnen*

Foto: Courtesy Stanford News Service

**Schots eredoctoraat voor TU/e-hoogleraar Onno Boxma**

De Heriot-Watt University in Edinburgh (Schotland) heeft woensdag 21 juni een eredoctoraat (Doctorate of Science) verleend aan prof. dr. ir. Onno Boxma, hoogleraar stochastische besliskunde aan de Faculteit Wiskunde en Informatica van de Technische Universiteit Eindhoven. Prof. Sergey Foss, hoogleraar toegepaste kansrekening aan de Heriot-Watt University en de huidige hoofdredacteur van het tijdschrift *Queueing Systems*, is de erepromotor van Boxma. Onno Boxma (1952) ontvangt het eredoctoraat voor zijn bijdragen aan de kansrekening en besliskunde in het algemeen, en de wachtrijtheorie in het bijzonder.

Boxma: “Dit eredoctoraat zie ik ook als een blijk van erkenning voor onze sectie Stochastiek. Mede dankzij het instituut Eurandom neemt de Eindhovense stochastiek tegenwoordig internationaal een zeer vooraanstaande plaats in.” Boxma werkt al jaren samen met onderzoekers van de Heriot-Watt University: “Zij doen veel onderzoek naar verzekeringswiskunde en wachtrijtheorie, wat

juist mijn voornaamste onderzoeksgebieden zijn." De afgelopen jaren ontving hij meerdere blijken van erkenning, waaronder een eredoctoraat van de Universiteit van Haifa (2009), de ACM SIGMETRICS Achievement Award (2011), en de Arne Jensen Lifetime Award (2014). *Wiskunde PersDienst*

**Quantum Software Consortium ontvangt 18,8 miljoen euro**

Het Ministerie van Onderwijs, Cultuur en Wetenschap heeft Zwaartekracht-financiering toegekend aan grootschalig onderzoek naar quantumsoftware. Met de toekenning van 18,8 miljoen euro kunnen onderzoekers van QuSoft, CWI, Universiteit Leiden, QuTech, TU Delft, UvA en VU de komende tien jaar excellente wetenschappelijke onderzoeksprogramma's opzetten. Minister Bussemaker stelt in totaal 112,8 miljoen euro beschikbaar in het Zwaartekrachtprogramma, met als doel Nederlands onderzoek dat tot de wereldtop behoort te stimuleren. Harry Buhrman (directeur QuSoft, CWI-groepsleider en faculteitshoogleraar aan de Universiteit van Amsterdam) is de hoofdaanvrager van het gehonoreerde zwaartekrachtvoorstel. *nwo.nl*

**Achievement Award voor TU/e-hoogleraar Sem Borst**

Prof. dr. ir. Sem Borst ontving in juni de ACM SIGMETRICS Achievement Award. Deze prestigieuze onderscheiding wordt jaarlijks uitgereikt door de Association for Computing Machinery (ACM) aan een wetenschapper voor zijn of haar fundamentele bijdragen aan de prestatie-analyse van computer- en communicatiesystemen. Borst ontvangt de prijs voor zijn invloedrijke en gewaardeerde bijdragen op het gebied van stochastische modellen voor draadloze netwerken, asymptotische analyse van zwaarstaartige wachtrijsystemen, gedistribueerde algoritmen voor dynamische capaciteitsallocatie, en efficiënte toewijzingsstrategieën in grootschalige cloudsystemen en datacenter-netwerken.

Sem Borst is sinds 1998 deeltijdhoogleraar stochastische besliskunde aan de Faculteit Wiskunde en Informatica van de TU/e. Daarnaast is hij werkzaam voor Nokia Bell Labs in Murray Hill in de Verenigde Staten. *Wiskunde PersDienst*

**Onderwijs 2032 en curriculum.nu**

Het basisonderwijs en voortgezet onderwijs bereidt leerlingen voor op de toekomst. Een curriculum is de landelijk verplichte inhoud van een opleiding: de vakken en de inhoud van die vakken. Het huidige curriculum stamt uit 2006 en er is een lange discussie gaande om dat te hervormen. Ouders, leraren, scholieren, wetenschappers en politieke vertegenwoordigers zijn betrokken in deze discussie.

Tussen november 2014 en april 2017 is door het hele onderwijsveld nagedacht over het onderwijs van de toekomst. In januari 2016 verscheen het eindadvies 'Ons Onderwijs 2032'. De Onderwijscoöperatie en een breed samengestelde regiegroep onderzochten of en hoe het eindadvies haalbaar en toepasbaar is. Die conclusies zijn in november 2016 aangeboden aan de staatssecretaris. Op 20 april 2017 keurde de Tweede Kamer de curriculumherziening goed. Er zullen ontwikkelteams gevormd worden, totaal negen, en ze zullen hun ideeën bij scholen toetsen. Rekenen en wiskunde is één van de vakgebieden waarop ze gaan experimenteren. Medio september 2017 kunnen scholen zich aanmelden als ontwikkelschool en leraren samen met de schoolleiders kunnen meedoen in een ontwikkelteam. De ontwikkelteams gaan aan de slag in 2018. *curriculum.nu*

**Niet-transitieve dobbelstenen**

Stel dat je vier dobbelstenen hebt, noem ze $A_1$, $A_2$, $A_3$ en $A_4$ met verschillende getallen op de zes zijden. Een voorbeeld is $A_1 = (4,4,4,4,0,0)$, $A_2 = (3,3,3,3,3,3)$, $A_3 = (6,6,2,2,2,2)$ en $A_4 = (5,5,5,1,1,1)$. We zeggen dat een dobbelsteen wint van een andere als de kans dat de eerste een groter getal levert groter is dan $\frac{1}{2}$. In het bovenstaande voorbeeld, bedacht door Brandley Effron in de jaren zestig, zien we dat $A_i$ wint van $A_{i+1}$ met kans $\frac{2}{3}$, maar deze relatie is niet transitief. Gegeven dat $A_1$ wint van $A_2$ en $A_2$ van $A_3$, dan wint $A_1$ van $A_3$ met kans $\frac{1}{3}$. Recent werd dit probleem in een algemener kader geplaatst door Conrey, Gabbard, Grant, Liu en Morrison. Ze definiëren een stochastische $n$-zijdige dobbelsteen als een niet-dalende stochastische rij van $n$ getallen gekozen tussen $1$ en $n$ die optellen tot $n \cdot \frac{n+1}{2}$. Net zoals in het eenvoudige voorbeeld van Effron zeggen we dat $A$ wint van $B$ als het meer waarschijnlijk is dat $A$ een groter getal dan $B$ oplevert. Vervolgens, hebben ze computersimulaties uitgevoerd voor het geval van drie dobbelstenen, noem ze $A$, $B$ en $C$, en hebben ze tot hun verbazing gevonden dat de kans dat $A$ wint van $C$ gegeven dat $A$ wint van $B$ en $B$ van $C$ voor $n$ groot genoeg gelijk aan $\frac{1}{2}$ is! De conclusie is dat de twee geconditioneerde relaties geen informatie voor de relatie tussen $A$ en $C$ kunnen leveren! Dit resultaat werd bewezen (vermoedelijk, het is nog niet officieel gepubliceerd) door Fieldsmedaillewinnaar Timothy Gowers. Zijn artikel is op zijn blog te vinden onder de titel 'The probability that a random triple of dice is transitive'. *gowers.wordpress.com*

**Nederlandse scholier wint goud**

Bij de Internationale Wiskunde Olympiade in Rio de Janeiro heeft Gabriël Visser (19) uit Spijkenisse een gouden medaille behaald. Hij loste drie van de zes opgaven volledig op en een vierde bijna. Hiermee bemachtigde hij bij deze meest prestigieuze wiskundewedstrijd voor middelbare scholieren een plek in de top vijftig van de wereld. Ook de andere Nederlandse deelnemers zetten goede prestaties neer: Matthijs van der Poel (16) uit IJsselstein en Levi van de Pol (15) uit Veenendaal wonnen beiden een zilveren medaille en Ward van der Schoot (18) uit Breda behaalde brons. *wiskundeolympiade.nl*

**Wiskundige Cédric Villani stapt over naar de politiek**

Cédric Villani is een wiskundige die in 2010 een Fieldsmedaille ontving voor zijn werk over de Boltzmann-vergelijking en Landaudemping. In 2010 koos Villani voor een bestuurlijke positie als directeur van het Henri Poincaré-instituut in Parijs met als gevolg dat hij nog maar weinig tijd aan onderzoek kon besteden. Villani werd gefascineerd door de politieke beweging van Emanuel Macron en besloot het pad van de politiek verder te volgen. Bij de parlementsverkiezingen van juni 2017 was hij kandidaat voor La Republique En Marche en behaalde een zetel in het Franse parlement. *cedricvillani.org*

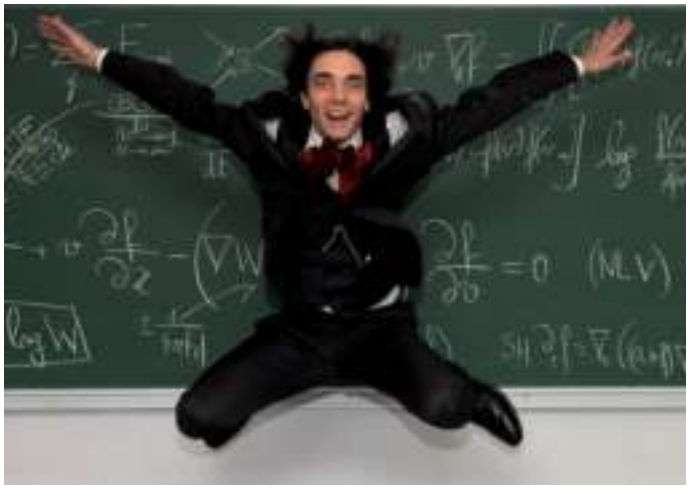


Foto: Cédric Villani

**Forensisch onderzoek in Nederland**

Wiskundigen Jeannette Leegwater (UvA en NFI), Charlotte Vlek (RUG) en Jacob de Zoete (oud-promovendus UvA) hebben een kort stuk geschreven over hun onderzoek in de forensisch wiskunde.

In het forensisch vakgebied komt het vaak voor dat meerdere bewijsstukken geëvalueerd dienen te worden. De makkelijkste manier om dit te doen is om voor elk bewijsstuk apart een bewijswaarde te rapporteren. Echter, als men dit doet, moet men er wel van overtuigd zijn dat de individuele bewijsstukken optimaal gecombineerd worden door de rechter. Specifiek voor gevallen waarin de bewijsstukken van hetzelfde type zijn (bijvoorbeeld twee schoensporen), is het normaal gesproken niet het geval dat de bewijsstukken conditioneel onafhankelijk zijn en is een gecombineerde evaluatie nodig om misverstanden te voorkomen. Om te bepalen hoe de verschillende bewijsstukken elkaar en de kansen behorende bij de gepresenteerde proposities beinvloeden, is het noodzakelijk om verschillende scenario's te bekijken. Bewijsstukken die relevant zijn voor hetzelfde vraagstuk (bijvoorbeeld wie was er op het plaats delict) zijn gemakkelijker te combineren dan bewijsstukken die relevant zijn voor verschillende vragen (bijvoorbeeld wie was er op het plaats delict en wat heeft er plaatsgevonden op het plaats delict). Een rechter moet vervolgens een beargumenteerde afweging maken tussen de verschillende scenario's op basis van het bewijs. Voor het afwegen van verschillende vormen van bewijsmateriaal, van getuigenverklaringen tot statistische informatie over DNA-sporen, worden Bayesiaanse netwerken gebruikt. Zo'n netwerk laat zien welke gebeurtenissen zich binnen een scenario afspelen, waarna het mogelijk is om uit te rekenen hoe groot de kans is dat dit scenario zich heeft afgespeeld. *J. Leegwater, C. Vlek en J. De Zoete*

**CWI en AMC willen medische beelden beter matchen**

Het Centrum Wiskunde & Informatica (CWI) en het AMC Amsterdam starten een onderzoeksproject naar het koppelen van verschillende medische beelden. Daarvoor ontvangen ze 1,4 miljoen euro binnen NWO's Open Technologie Programma. Het combineren van verschillende typen beelden kan soms belangrijke inzichten opleveren in een ziekte. Het kan bijvoorbeeld voorkomen dat medici een mri-beeld met een later gemaakte ct-scan willen vergelijken om te kijken hoe een aandoening zich ontwikkelt, maar die twee methoden laten verschillende dingen zien en de patiënt ligt meestal in een andere houding. Het is daarom lastig om de beeldtypen met elkaar te vergelijken. Binnen het project wordt daarom gewerkt aan software die de verschillende beelden met elkaar kan correleren. Een van de grote uitdagingen daarbij is om de algoritmes robuust te maken voor grote verstoringen, bijvoorbeeld nadat een tumor is verwijderd. Verder is het de bedoeling om software te ontwikkelen die daadwerkelijk bruikbaar is in de klinische praktijk. De focus ligt daarbij op radiotherapie, maar in principe moeten de resultaten ook bruikbaar zijn voor andere domeinen. *cwi.nl*

**Extreme-waardetheorie voorspelt veel vaker hoogwater in 2050**

Een extreem hoge waterstand aan de kust, die nu maar eens in de vijftig jaar voorkomt, zou in 2050 in de tropen elk jaar voor kunnen komen. Een verrassende uitkomst, aangezien de zeespiegelstijging dan maximaal een centimeter of twintig zal bedragen. Deze voorspelling volgt uit berekeningen op grond van de Generalised Extreme Value-verdeling. Voorspellingen van hoogwaterstanden zijn van groot belang om kwetsbare gedeeltes van de kust klimaatbestendig te maken. *nemokennislink.nl*

## Koninklijk Wiskundig Genootschap

❑ **Vacature hoofdredacteur Pythagoras**

Het tijdschrift *Pythagoras* zoekt per 1 oktober 2017 een nieuwe hoofdredacteur (m/v), die verantwoordelijk zal zijn voor de inhoud van dit blad. Het gaat om een betaalde functie voor één dag per week. Voor meer informatie zie https://wiskgenoot.nl/nieuws/gezocht-hoofdredacteur-pythagoras.

❑ **NMC 2018**

Op 3 en 4 april 2018 vindt het 54ste Nederlands Mathematisch Congres plaats in Veldhoven. In lijn met het 'Deltaplan Wiskunde' wordt dit een groots opgezet NMC, dat een aantrekkelijk programma moet bieden voor heel wiskundig Nederland. De programmacommissie, onder leiding van Mark Peletier, is momenteel hard aan het werk om hieraan vorm te geven.

Recent verschenen:

❑ **Epsilon Uitgaven** (www.epsilon-uitgaven.nl)

*89. De kunst van het hoofdrekenen*, Willem Bouman, € 25, 2017.

Joan Daemen
*Digital Security Groep*
*Radboud Universiteit Nijmegen*
*joan@cs.ru.nl*

# Symmetrische cryptografie 2.0

**Joan Daemen is hoogleraar in de Digital Security-groep van de Radboud Universiteit, Nijmegen en verder werkzaam voor STMicroelectronics. Daemen is gespecialiseerd in symmetrische cryptografie, mede-uitvinder van versleutelingsstandaard AES en hashstandaard SHA-3, en andere cryptografische schema's. In dit artikel geeft Daemen een toegankelijke uiteenzetting over symmetrische cryptografie en de nieuwe richting die het vakgebied is ingeslagen.**

**Doelstelling**

Alice en Bob willen communiceren in omstandigheden waar er personen of organisaties zijn die er belang bij hebben om de inhoud van hun communicatie te kennen of te manipuleren. Alice en Bob willen de communicatie dus beveiligen in de zin dat nieuwsgierige neuzen zoals Facebook, Google of de NSA niet kunnen meeluisteren en dat Alice en Bob kunnen zien als er met hun communicatie wordt geknoeid.

Voorbeelden van communicatie tussen personen zijn e-mail, chatsessies, telefoongesprekken of Skypesessies en dergelijke. Maar het is ook belangrijk communicatie te beveiligen tussen computerprocessen zoals tussen een betaalterminal en een centraal banksysteem of een virtueel private netwerk-verbinding. Andere voorbeelden zijn bescherming van gegevens in langetermijngeheugen zoals harde schijven van laptops, smartphones of tablets of gegevens waarvan de opslag wordt uitbesteed aan dienstverleners zoals Dropbox.

De veiligheid die we willen garanderen omvat doorgaans twee aspecten. Het eerste aspect is geheimhouding, waarmee we bedoelen dat alleen de rechtmatige personen toegang hebben tot de inhoud. Het tweede aspect is integriteit, ook wel *waarmerken* genoemd, waarmee we bedoelen dat de rechtmatige personen kunnen zien als er met de inhoud van de communicatie of opslag is geknoeid.

Er zijn nog andere aspecten van communicatie en opslag die kunnen worden beveiligd. Zo kan men proberen ervoor te zorgen dat opslag of communicatie gegarandeerd is. Een ander aspect is het proberen te verbergen van het bestaan van communicatie tussen twee partijen, of het feit dat er op een langetermijngeheugenmedium gegevens zijn opgeslagen. Dit zijn maar een paar voorbeelden.

In dit artikel beperken we ons echter tot het beschermen van de geheimhouding en het waarmerken van communicatie en opslag.

**Asymmetrische cryptografie**

De techniek om geheimhouding te bewerkstelligen is *versleuteling*. Voor de gegevens worden verzonden of opgeslagen, worden ze versleuteld en na ontvangst of ophalen worden ze ontsleuteld. Voor ontsleuteling is een geheime sleutel nodig, die alleen in het bezit is van de rechtmatige ontvanger. Zo'n sleutel is zoals een geheim wachtwoord, maar niet noodzakelijk gemakkelijk te onthouden. Voor versleuteling is er in principe geen geheime sleutel nodig, maar dit vereist wel een openbare sleutel die verbonden is met de geheime sleutel van de ontvanger.

De techniek om berichten of dossiers te waarmerken wordt (elektronische) handtekening genoemd. Hierbij wordt aan het bericht of dossier een bitrij toegevoegd, handtekening genoemd. Het aanmaken van zo een handtekening vereist een geheime sleutel, die alleen in het bezit is van degene die de handtekening zet. De verificatie van een handtekening kan met een openbare sleutel die gelinkt is aan de geheime sleutel waarmee het dossier getekend werd.

De zojuist besproken technieken zijn voorbeelden van asymmetrische cryptografie, zo genoemd omdat de twee bewerkingen verschillende sleutels vereisen. Asymmetrische cryptografie heeft het voordeel dat de eigenaar van een geheime sleutel, meestal private sleutel genoemd, deze met niemand moet delen en dat de openbare sleutel publiek mag worden gemaakt. Een nadeel is dat de private sleutel in theorie uit de openbare sleutel kan berekend worden. De stille hoop is dat dit echter zoveel rekenwerk zou kosten dat het onbetaalbaar is. Dit laatste kunnen we jammer genoeg nooit zeker weten en boven asymmetrische cryptografische schema's hangt dan ook altijd een zwaard van Damocles: ze kunnen elk ogenblik gebroken worden. Om toch enige betrouwbaarheid te geven, wordt er geavanceerde wiskunde ingeschakeld zoals getaltheorie en elliptische krommen. De redenering is dan dat je het schema alleen kan breken als er een historisch moeilijk wiskundig probleem is opgelost. Een voorbeeld van zo'n probleem is het ontbinden van zeer grote getallen in priemfactoren. Op die manier staat asymmetrische cryptografie op de schouders van wiskundige reuzen zoals Fermat, Euler, Gauss en Galois, wat ze een grote geloofwaardigheid verleent. Een nadeel van deze sterk wiskundige fundamenten is dat asymmetrische cryptografie zeer reken-

intensief en volumineus is. Alsof dat niet genoeg is, maakt de laatste tijd het spook van kwantumrekenen opgang. Kwantumrekenen is een hypothetische vorm van rekenen die heel veel bewerkingen tegelijk, *in superpositie*, kan doen en in staat zou zijn historisch moeilijke problemen efficiënt op te lossen. Kwantumrekenen is momenteel nog science fiction en even hypothetisch als economisch rendabele kernfusie, maar creëert toch een hoop onrust. Een positief bijverschijnsel van de kwantumcomputerdreiging is dat er veel fondsen worden vrijmaakt voor cryptografisch onderzoek, met name naar zogenaamd kwantumveilige asymmetrische cryptografie.

**Symmetrische cryptografie**

In een andere tak van de cryptografie gebruikt men dezelfde sleutel voor versleuteling en ontsleuteling. Ook het zetten van een handtekening en het verifiëren ervan vereisen dezelfde geheime sleutel. Daarom wordt deze tak *symmetrische cryptografie* genoemd. Symmetrische cryptografie heeft niet de besproken nadelen van asymmetrische cryptografie en is grootte-orden minder rekenintensief voor dezelfde beveiligingsgraad. Het heeft echter het nadeel dat bij verzending van gegevens de zender een geheime sleutel moet delen met de ontvanger. Hiervoor zijn veel oplossingen. Zo kunnen twee personen met het oog op toekomstige communicatie een geheime sleutel afspreken als ze elkaar ergens ontmoeten. Als dit niet praktisch is kunnen ze bijvoorbeeld asymmetrische cryptografie gebruiken voor het opzetten van symmetrische sleutels. Dit laatste is zinvol want zo beperken ze het gebruik van de trage cryptografie voor het opzetten van een korte sleutel en schakelen ze de snelle cryptografie in voor het beveiligen van de eigenlijke data. Dit artikel gaat over technieken in symmetrische cryptografie en we gaan ervan uit dat er een geheime sleutel beschikbaar is. Het genereren, verdelen en uit gebruik nemen van sleutels is een discipline op zich, die sleutelbeheer wordt genoemd. Dit is een interessant domein en in praktijk van veel groter belang voor (on)-veiligheid dan cryptografie, maar dat is voor een andere keer.

Symmetrische cryptografie beslaat het ontwerpen en ontleden van schema's die willekeurige hoeveelheden gegevens aankunnen. We spreken van schema's voor versleuteling en digitale handtekening. Die laatsten worden gewoonlijk MACs genoemd, wat staat voor *message authentication codes*. De laatste tijd zijn daar ook hybride schema's bijgekomen die versleuteling en waarmerken met elkaar combineren. Deze schema's vormen, samen met het (cryptografisch) *prakken* (hashing in Engels), de hoofdtaken van symmetrische cryptografie. Een prakfunctie zet een bericht van willekeurige lengte om tot een reeks tekens die er volledig willekeurig uitzien. Cryptografisch prakken heeft vele toepassingen waaronder identificatie en asymmetrische handtekeningen, maar dit beschrijven zou ons te ver leiden.

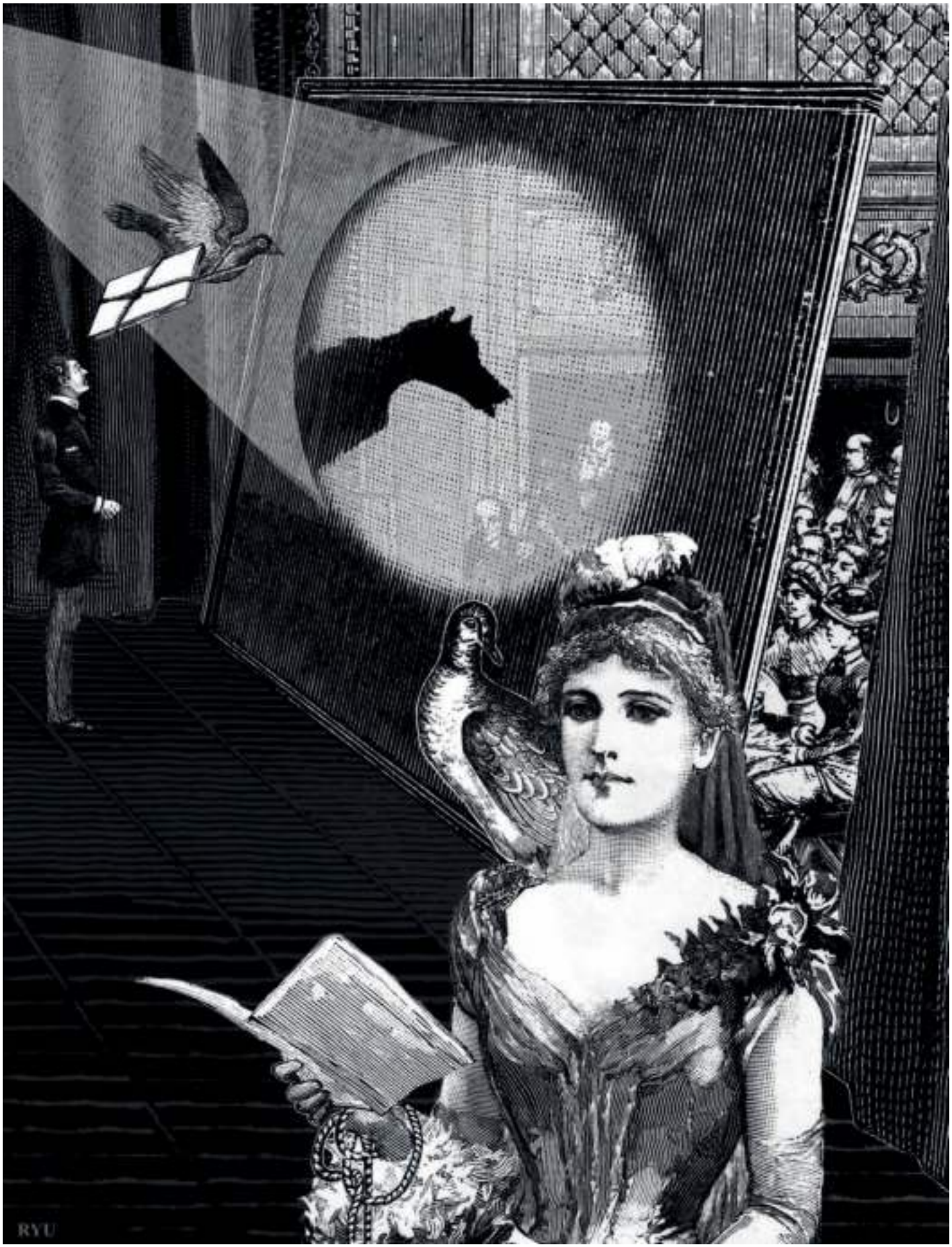
**Symmetrische cryptografie, oude stijl**

Een moeilijkheid van cryptografisch ontwerp is schema's te bouwen die willekeurige hoeveelheden gegevens aankunnen. In de jaren zeventig is de aanpak gelanceerd die tot op heden symmetrische cryptografie domineert. Deze aanpak bestaat erin om een schema te bouwen dat berichten van een bepaalde lengte kan versleutelen tot cryptogrammen met dezelfde lengte. Zo'n schema noemen we een *blokcijfer* en het blokcijfer waar het allemaal mee begon was de zogenaamde Data Encryption Standard (DES) [10]. DES kan berichten van $64$-bit (de zogenaamde bloklengte) versleutelen en het resultaat terug ontsleutelen, allebei onder een geheime sleutel van $56$ bits. Mathematisch zou je kunnen zeggen dat DES bestaat uit een verzameling van $2^{56}$ $64$-bit permutaties en dat de sleutel er één van selecteert. De veiligheid die gewoonlijk van een blokcijfer verwacht wordt is het volgende: een aanvaller die de sleutel niet kent kan het blokcijfer niet onderscheiden van een willekeurig permutatie. Dit heet pseudorandom permutation (PRP) veiligheid.

Door de tijd is het inzicht gegroeid dat het onmogelijk is om een blokcijfer te bouwen waarvoor we wiskundig kunnen bewijzen dat het PRP-veilig is. Hoe kunnen we dan vertrouwen hebben in een blokcijfer? Vertrouwen is gebaseerd op openbaar kritisch onderzoek. DES was niet zo'n slecht ontwerp maar het kon beter. Na de publicatie van DES is het openbaar ontwerp en onderzoek van blokcijfers dan ook een academische bezigheid geworden. Het schrijven van artikelen over breken en verbeteren van blokcijfers was een manier om een academische loopbaan op te bouwen. Deze goedaardige wapenwedloop leidde tot steeds betere blokcijfers, met als hoogtepunt de verkiezing van Rijndael [8] als Advanced Encryption Standard (AES) in 2000 [11]. AES heeft een bloklengte van $128$ bits en ondersteunt sleutels van $128$, $192$ en $256$ bits. We kunnen nog steeds niet bewijzen dat AES PRP-veilig is, maar het feit dat tot op heden niemand erin is geslaagd om AES te breken ondanks verwoede pogingen van vele experts maakt het zeer onwaarschijnlijk dat AES nog gebroken zal worden.

We kunnen nu berichten van $64$ bits versleutelen met DES en berichten van $128$ bits met AES, maar wat doen we met berichten die een andere lengte hebben? Hiervoor is in de loop der jaren een arsenaal aan technieken voorgesteld om versleutelingsschema's te bouwen met een blokcijfer als bouwblok. Deze worden werkingsmodes voor versleuteling genoemd. Ook een MAC-berekening kan men doen op basis van blokcijfers door middel van werkingsmodes voor waarmerken. Meer recent maken werkingsmodes opgang die zowel versleutelen als waarmerken. In tegenstelling tot blokcijfers kan men voor deze werkingsmodes wel bewijzen dat ze aan bepaalde veiligheidsvereisten voldoen, op voorwaarde dat het gebruikte blokcijfer PRP-veilig is. Zelfs cryptografisch prakken wordt gedaan met blokcijfers, alhoewel daar PRP-veiligheid niet voldoende is. Zo komt het dat bijna alle standaarden in symmetrische cryptografie tot voor kort gebaseerd waren op blokcijfers en hun werkingsmodes.

Voor de leek mag deze aanpak dan al naïef lijken, vele cryptografische experts zweren erbij en zijn ervan overtuigd dat dit de juiste weg is. Het blokcijfermodel is recentelijk zelfs nog *verfijnd* door toevoeging van een bijkomende parameter naast de sleutel: een zogenaamde *aanpasser*. Feit is dat blokcijfers zijn ontworpen om berichten van een bepaalde lengte efficiënt te versleutelen en ontsleutelen en dat ontsleuteling in de meerderheid van de werkingsmodes helemaal niet gebruikt wordt. Het is zelfs zo dat in de meeste werkingsmodes de omkeerbaarheid van blokcijfers als een beperking werkt voor de veiligheid: de zogenaamde verjaardagslimiet (zie ook het artikel van Bart Mennink in dit verband). Het lijkt er dan ook op dat er ruimte is voor een meer doordachte manier om aan symmetrische cryptografie te doen.




Illustratie: Ryu Tajiri

Die is er wel degelijk en ik noem het graag symmetrische crypto 2.0.

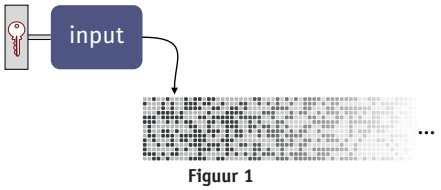
**Symmetrische cryptografie 2.0**

In symmetrische cryptografie 2.0 worden de twee rollen tot nu toe vervuld door blokcijfers verdeeld over twee andere functies. De eerste zijn zogenaamde onvoorspelbare functies, die zullen dienen om cryptografische functies mee te definiëren door er werkingsmodes op toe te passen. De tweede zijn cryptografische permutaties, die we zullen bouwen en dan gebruiken in constructies om er de eerder vermelde onvoorspelbare functies mee te bouwen.

*Onvoorspelbare functies*

In plaats van blokcijfers, stellen we hier een ander type functie centraal, die we *onvoorspelbare functie* (OVF) noemen. Dit is een functie die op basis van een korte sleutel en een argument van willekeurige lengte, een resultaat teruggeeft van onbeperkte lengte. In Figuur 1 illustreren we onze OVF op schematische wijze.

De veiligheidsdoelstelling voor een OVF heet *pseudo-willeurige-functie-veiligheid* (PRF) en dit is het volgende. Voor een tegenstander die de sleutel niet kent, ziet het resultaat van een OVF er volledig willekeurig uit, met als enige beperking dat de OVF met dezelfde argumenten hetzelfde resultaat teruggeeft. Voor verschillende argumenten daarentegen geeft een OVF totaal ongerelateerde resultaten terug, ook al verschillen die argumenten maar heel weinig. We zullen later bespreken hoe we zo'n OVF kunnen bouwen, maar eerst gaan we eens kijken wat we ermee kunnen doen, en verfijnen daarbij het concept.



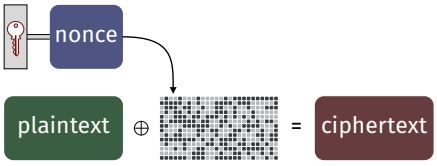
Figuur 1

*Versleuteling*

Er is een methode voor versleuteling die zeer veilig is: het zogenaamde *one-time pad*. Hierbij gebruiken zender en ontvanger een sleutel die even lang is als het bericht. We noemen dit een *sleutelstroom*. De zender telt deze sleutelstroom bit-per-bit op bij het bericht en de ontvanger trekt deze sleutel er weer van af. Voor een tegenstander die geen informatie heeft over de sleutel, met andere woorden, voor wie de sleutel er volledig willekeurig uitziet, geeft het cryptogram geen informatie over de inhoud van het versleutelde bericht. Dit kan men eenvoudig bewijzen, in de veronderstelling dat er geen bits van de sleutelstroom gebruikt worden voor de versleuteling van meerdere berichten.

Het grote nadeel van de one-time pad is dat zender en ontvanger een sleutel moeten delen die even lang is als de totale te verwachten communicatie en dit is onpraktisch voor de meeste toepassingen. We wensen de communicatie te beveiligen met een korte sleutel, en dit is waar de OVF om de hoek komt piepen. Zender en ontvanger gebruiken een OVF om een sleutelstroom te genereren met een gedeelde sleutel en argument en gebruiken deze voor one-time pad-versleuteling van het bericht. We beelden de versleuteling af in Figuur 2.

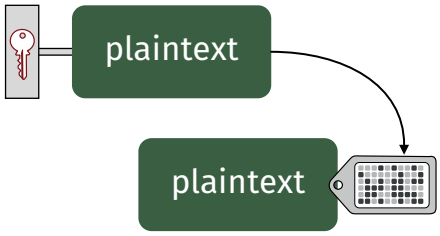
Deze methode is veilig als de OVF veilig is en als het argument voor elk bericht uniek is. Met andere woorden, het argument moet een zogenaamde *nonce* zijn. In de meeste toepassingen is er een nonce voorhanden. Bij e-mail kan men bijvoorbeeld de datum en tijd van verzending gebruiken. Voor dossiers op een computer kan men de naam en de plaats in de folder gebruiken.



Figuur 2

*MAC-berekening*

Een MAC-functie berekent een korte tag op basis van een sleutel en bericht van willekeurige lengte. De veiligheid van een MAC-functie bestaat erin dat het voor een tegenstander moeilijk moet zijn om *vervalsingen* te doen. Een vervalsing bestaat erin een ontvanger een bericht met tag te laten aanvaarden waarbij een tegenstander de tag heeft berekend in plaats van de legitieme afzender. Meer concreet, als de tag een lengte heeft van $n$ bits, mag de kans op succes bij elke poging tot vervalsing maar $2^{-n}$ zijn. Dit is het geval als de tags voor verschillende berichten er volledig willekeurig uitzien. Een OVF leent zich dus uitstekend voor het gebruik als MAC-functie. Voor het berekenen van een $n$-bit tag gebruiken we de OVF met als argument het bericht en kappen we het resultaat af na $n$ bits, zoals afgebeeld in Figuur 3.

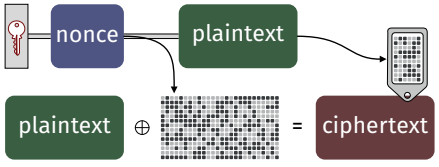


Figuur 3

*Combinatie van versleuteling en MAC*

In de meeste gevallen willen we graag berichten beschermen tegen zowel afluisteren als vervalsing. Dit kunnen we eenvoudig doen door het bericht te versleutelen en een tag te berekenen over het cryptogram. Beide bewerkingen kunnen we weer doen met onze OVF. Eerst berekent de zender een sleutelstroom op basis van een nonce en telt die op bij het bericht, resulterend in de cijfertekst. Dan berekent hij de tag op basis van deze nonce en de cijfertekst. Dit beelden we af in Figuur 4.

Dus het cryptogram bestaat uit een nonce, cijfertekst en tag. Bij ontvangst berekent de ontvanger de verwachte tag opnieuw op basis van de nonce en cijfertekst. Als de ontvangen en berekende tag niet gelijk zijn, verwerpt hij het cryptogram. Als ze wel gelijk zijn, berekent hij de sleutelstroom op basis van de ontvangen nonce en ontsleutelt de cijfertekst door de sleutelstroom erbij op te tellen.



Figuur 4

*OVF met de oplopende eigenschap*

We zien dus dat zender en ontvanger beiden twee OVF-berekeningen doen: één keer met als argument de nonce en één keer de nonce en het bericht. Het zou interessant zijn als de bewerking op de nonce gedeeld moest kunnen worden tussen de twee OVF-berekeningen. Hier kunnen we rekening mee houden als we onze OVF gaan bouwen. Meer concreet, we willen dat het argument van onze OVF bestaat uit een serie bitrijen in plaats van één enkele. We schrijven series van bitrijen met het symbool $\circ$ dat staat voor *volgend op*. Dus $B \circ A$ betekent: eerst bitrij $A$ en dan bitrij $B$.

Twee series bitrijen zijn gelijk enkel als ze evenveel bitrijen bevatten en als de bitrijen één voor één aan elkaar gelijk zijn. De OVF toegepast op twee verschillende bitrijen $B \circ A$ en $D \circ C$ geeft dus verschillende resultaten, zelfs al is $A \,|\, B = C \,|\, D$, waarbij $|$ staat voor concatenatie.

In twee OVF-berekeningen met dezelfde sleutel en waarbij de eerste bitrijen gelijk zijn, kunnen deze geabsorbeerd worden in een enkele berekening. Een OVF met deze eigenschap noemen we *oplopend*. Vanaf hier gaan we ervan uit dat OVFs oplopend zijn. Een toepassing waar de oplopende eigenschap interessant is, is bij de beveiliging van een communicatiekanaal.

*Beveiliging van een communicatiekanaal*

Dikwijls willen we niet gewoon afzonderlijke berichten beschermen tegen afluisteren en vervalsing, maar een stroom berichten, of zelfs geluid- of beeldcommunicatie. We geven hier aan hoe we dat kunnen doen met een oplopende OVF. Om het eenvoudig te houden gaan we in ons voorbeeld uit van communicatie in één richting, maar het kan gemakkelijk veralgemeend worden.

Het protocol beveiligt een stroom berichten waarbij elk bericht bestaat uit klaartekst en randgegevens. We willen beiden waarmerken maar alleen de klaartekst beschermen tegen afluisteren. We spreken van een sessie. De zender zet elk bericht om in een cryptogram, bestaande uit de cijfertekst (versleutelde klaartekst), de (onversleutelde) randgegevens en een tag. De tag wordt niet alleen berekend op het cryptogram en randgegevens van het huidige bericht, maar ook op alle voorgaande berichten.

Dit gebeurt als volgt. Tijdens de sessie houden zender en ontvanger een *geschiedenis* bij die bestaat uit de serie cijfertekst en randgegevens bitrijen. Als de zender een nieuw bericht wil versleutelen, genereert hij de sleutelstroom door de OVF toe te passen op de geschiedenis op dat moment. Hij voegt dan de cijfertekst en de randgegevens toe aan de geschiedenis en berekent de tag door de OVF toe te passen op de geschiedenis.

Voor de veiligheid is het essentieel dat voor elke OVF-berekening het argument een andere waarde heeft. Uit de beschrijving hierboven zie je dat de tag van een cryptogram en de sleutelstroom van het volgende cryptogram worden berekend op basis van dezelfde geschiedenis. Om die reden neemt het protocol de tag als de eerste $n$ bits van het OVF-resultaat en de sleutelstroom vertrekkende van bit $n$. Voor elke tagberekening verlengt de geschiedenis en dus heeft in elke sessie elke OVF-berekening een ander argument. Tussen verschillende sessies met dezelfde sleutel kan de eenmaligheid van het OVF-argument bewerkstelligd worden met een nonce. Zender en ontvanger starten de geschiedenis door er een nonce in te schrijven.

Gedurende de sessie wordt de geschiedenis steeds langer, maar in elke OVF-berekening is de geschiedenis gelijk aan de vorige berekening met twee bitrijen daaraan toegevoegd. Dankzij de oplopende eigenschap hoeft de OVF alleen maar deze laatste twee bitrijen te absorberen.

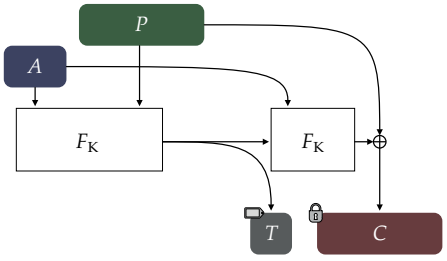
Voor de berekening van de tag moet de geschiedenis op eenduidige manier de serie cijfertekst en randgegevens coderen. Met andere woorden, uit de geschiedenis moet deze serie kunnen worden gereconstrueerd. Dit is geen probleem als elk bericht een klaartekst en randgegevens heeft, want de geschiedenis is simpelweg de opeenvolging van cijferteksten afgewisseld met randgegevens. Wanneer we ook berichten ondersteunen die alleen uit klaartekst of alleen uit randgegevens bestaan, is dit niet meer het geval. Dit kan echter gemakkelijk door cijfertekst en randgegevens te markeren met een bijkomende bit aan het einde zodat de resulterende bitrijen zich in verschillende domeinen bevinden. Vooraleer ze de geschiedenis vervoegen, markeren we dus elke cijfertekst met een bit $0$ en elke randgegevens bitrij met een bit $1$.

*Versleuteling met ingebouwde nonce*

Sommige ontwikkelaars hebben de grootste problemen in de wereld om voor een nonce te zorgen. Nonces gebaseerd op datum en tijd vereisen een correct werkende klok en meestal hebben ze er geen die ze kunnen of willen vertrouwen. Nonces gebaseerd op serienummers vereisen dat de zender over midden- of langetermijngeheugen beschikt en daar hebben ze ook helemaal geen vertrouwen in. Een andere oplossing, het gebruik van willekeurige bitrijen als nonces, heeft dan weer het nadeel dat we door de verjaardagsparadox relatief lange nonces moeten genereren en versturen. Bovendien is het genereren van willekeurige bitrijen iets waar deze mensen maar niet in slagen. Voor zulke mensen zou het goed zijn als we een bericht veilig kunnen versleutelen en waarmerken zonder nood aan een nonce.

We gaan weer berichten beveiligen bestaande uit klaartekst en randgegevens, en deze keer zonder te rekenen op een nonce. Dit impliceert een beperking: als twee berichten hetzelfde zijn, geeft versleuteling met dezelfde sleutel twee dezelfde cryptogrammen. Maar zo snel als twee berichten verschillen, willen we dat de cryptogrammen ook verschillen. Met name, we willen dat de sleutelstromen voor de versleuteling van de klaarteksten verschillen.

Een werkingsmode die dit realiseert heet kunstmatig-beginwaarde, oftewel synthetic initial value (SIV) [14]. Onze OVF leent zich hier perfect voor en dit leidt tot een heel simpele berekening. Eerst berekenen we de tag door de OVF te berekenen over randgegevens $(A)$ gevolgd door de klaartekst $(P)$. Dan berekenen we de sleutelstroom door de OVF te berekenen over randgegevens gevolgd door de tag $(T)$. Dit geeft geen $100$ procent garantie dat het OVF-argument verschilt voor elke sleutelstroom. Met name, het gaat fout als twee berichten dezelfde randgegevens hebben, én dezelfde tag. Maar we kunnen de kans dat dit gebeurt zo klein maken als we zelf willen door de tag lang genoeg te kiezen. De SIV-mode van een OVF ziet uit zoals in Figuur 5.




**Figuur 5**

**Permutaties**

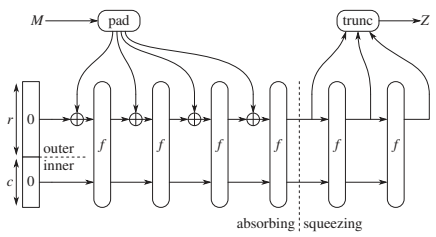
Één pijler van symmetrische cryptografie 2.0 is de vervanging van het blokcijfer door een OVF. De tweede pijler is het bouwblok waar we onze OVFs op gaan baseren: de permutatie. Een cryptografische permutatie beeldt een argument van $b$ bits af op een resultaat van $b$ bits, en dit op een omkeerbare wijze. Je zou het kunnen zien als een blokcijfer, maar er zijn wel belangrijke verschillen. Ten eerste heeft een permutatie geen sleutelargument. Ten tweede hebben we OVF-constructies voor ogen waar geen

nood is aan de berekening van de inverse permutatie, dus de inverse hoeft niet noodzakelijk efficiënt te zijn.

Voor het ontwerpen van permutaties kunnen we de ervaring gebruiken die we hebben opgedaan met blokcijfers. We herhalen gewoon een eenvoudige rondefunctie een aantal keer, dit afgewisseld met het optellen van rondeconstanten. De rondefunctie kunnen we samenstellen als de opeenvolging van een niet-lineaire laag, een menglaag en een dispersielaag, waarbij we die lagen kiezen en samenstellen volgens één of andere ontwerpstrategie. Een populaire ontwerpstrategie is de *strategie van het brede spoor*. Voor het kiezen van het aantal ronden moeten we kijken naar de OVF-constructie waarin de permutatie gebruikt wordt. In principe moeten we het aantal ronden zo kiezen dat de OVF een goede veiligheidsmarge geeft. Met andere woorden, we kijken naar het grootste aantal ronden dat in de context van een OVF gebroken kan worden, en tellen er een aantal ronden bij op. Natuurlijk hangt het vertrouwen dat men kan hebben in een permutatie af van het begrip van zijn potentiële zwakheden. Dit begrip hangt af van hoe overtuigend de auteurs het ontwerp kunnen verantwoorden en hun analyse, maar ook van de cryptanalyse door derden. Hoe meer helderheid en cryptanalyse, hoe groter het vertrouwen. We zullen nu twee voorbeelden geven van OVF-constructies die gebaseerd zijn op permutaties. Voor meer achtergrond over het ontwerp van permutaties verwijzen we naar het laatste hoofdstuk van [3] en [4].

## De spons

De spons is een constructie voor het bouwen van cryptografische prakfuncties op basis van een permutatie [1,2]. Sinds haar introductie in 2007 is zij bezig aan een steile opgang, niet in het minst om het feit dat zij aan de basis ligt van de SHA-3 standaard [12]. Een andere troef is haar elegante structuur, zie Figuur 6.




**Figuur 6**

De spons verdeelt de $b$ bits van de toestand in een *binnengedeelte*, bestaande uit *c* bits, en een *buitengedeelte*, bestaande uit *r* bits. De haalbare beveiligingsgraad hangt af van $c$: hoe kleiner, hoe zwakker. Dit bepaalt de efficiëntie, want $r = b - c$ en $r$ is het aantal bits die de spons kan verwerken per uitvoering van de permutatie $f$.

Terwijl prakfuncties vroeger altijd een resultaat hadden van vaste lengte, kan de spons een resultaat van willekeurig lengte aan. Dit komt tot uiting in de SHA-3 standaard die een nieuwe term introduceert voor zo'n type functie: een eXtendable Output Function (XOF), en ook twee dergelijke functies definieert: SHAKE128 en SHAKE256.

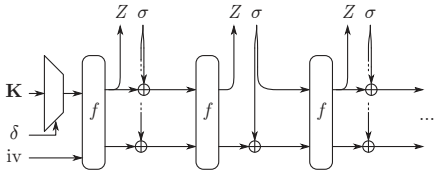
### *OVF op basis van een SHAKE-functie*

De SHAKE-functies hebben slechts één argument en zijn dus niet onmiddellijk bruikbaar als OVF. Maar het volstaat om een injectieve codering van een sleutel en een serie bitrijen naar één enkele bitrij te bouwen om er een OVF van te maken. Met injectief bedoelen we dat het mogelijk is om de sleutel en de serie bitrijen terug op te bouwen alleen maar op basis van het resultaat. Voorbeelden van zulke codering kan men vinden in [13]. Bovendien is de spons oplopend. Een enkele oogopslag volstaat om te zien dat dit het geval is: de spons comprimeert alles wat ze absorbeert in een $b$-bit toestand.

De veiligheidsconcepten voor een XOF en een OVF liggen dicht bij elkaar. Met name, een zwakheid in de OVF impliceert een zwakheid van de onderliggende XOF. Dus vertrouwen in de XOF gaat over op de OVF. Maar de vereisten voor een XOF zijn eigenlijk veel strenger dan voor een OVF: een XOF moet weerstand bieden aan tegenstanders zonder dat er een geheime sleutel is, terwijl voor een OVF er wel zo'n sleutel is. Dit leidt ertoe dat een OVF op basis van een XOF suboptimaal is.

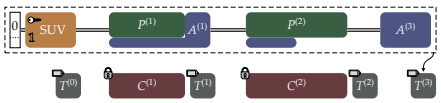
### *Gesleutelde duplex*

Na een turbulente periode waarin varianten van de spons werden voorgesteld en bestudeerd, is er een constructie naar boven gekomen die zeer goed geschikt is om als OVF te gebruiken: de gesleutelde duplex [9]. Net als de spons werkt die op een $b$-bit toestand. In het begin wordt een sleutel en een beginwaarde (IV) in de toestand geschreven. Dan kunnen er zogenaamde *duplex-iteraties* worden gedaan, waarbij telkens een bitrij $\sigma$ van maximaal $b$ bits wordt geladen (in plaats van $r$ bits zoals in de spons) en er een resultaat $Z$ van maximaal $r$ bits terugkomt, zie Figuur 7.




**Figuur 7**

Men kan hierop een OVF bouwen door middel van een eenvoudige codering. Door bijvoorbeeld elke bitrij $\sigma$ te beperken tot maximaal $b-1$ bits en hieraan één bit $1$ op het einde toe te voegen, hebben we een OVF, maar met de beperking dat het argument bestaat uit een serie bitrijen van elk maximaal $b-1$. Men kan er ook rechtstreeks modes op bouwen zoals bijvoorbeeld de communicatiekanaalbeveiligingsmode Motorist [6], die we schematisch illustreren in Figuur 8.



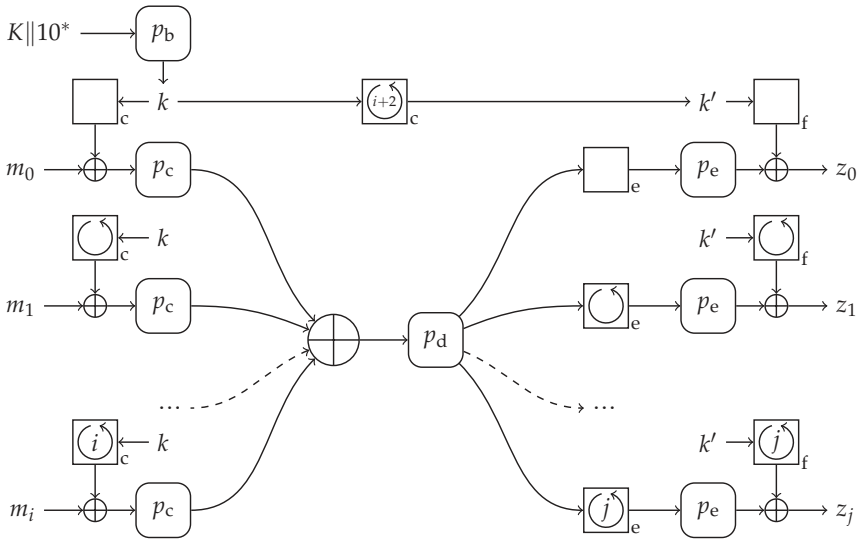

**Figuur 8**

Men kan zich afvragen: kan de gesleutelde duplex-constructie wel veilig zijn, en zo ja, wat is dan de veiligheidsgraad die je ermee kan krijgen? Een antwoord op deze vraag wordt gegeven door zogenaamde onderscheidingsbewijzen. Zulke bewijzen zeggen iets over de veiligheid van een constructie in de veronderstelling dat die gebruikmaakt van een willekeurige permutatie. In zo'n bewijs heeft een hypothetische tegenstander toegang tot een systeem dat ofwel de constructie is met een geheime sleutel, ofwel een ideale functie. A priori weet ze niet met welk van de twee systemen ze spreekt en ze mag dit proberen te achterhalen door aan het systeem het resultaat te vragen voor argumenten van haar keuze. Op het einde van het experiment moet ze gokken welke van de twee het is.

In het bewijs berekenen we dan een bovengrens voor de kans op succes dat ze de juiste keuze maakt. Deze bovengrens is een uitdrukking waarin typisch het totaal aantal bits in de argumenten en resultaten voorkomen en het aantal berekeningen uitgevoerd door de tegenstander en parameters van de constructie zoals de capaciteit $c$. Het is deze uitdrukking die aangeeft hoe hoog de potentiële veiligheidsgraad van de constructie is. Als de kans op succes pas afwijkt van

$1/2$ (kans op succes bij een pure gok) voor een aantal berekeningen tot $2^s$ en voor een aantal bevraagde bits tot $2^s$, dan spreken we van een veiligheidsgraad van $s$ bits.

Als we de constructie gebruiken met een concrete permutatie, is de onderscheidingslimiet niet langer van toepassing. Hij zegt namelijk alleen maar iets over de gezondheid van de constructie en maakt abstractie van de onderliggende permutatie. We gaan er stilzwijgend van uit dat een concrete permutatie in combinatie met de constructie bijvoorbeeld een veilige OVF oplevert. Of dit zo is, moet worden bevestigd door openbare cryptanalyse. Dit is vergelijkbaar met de PRP-veiligheid van blokcijfers. Als we bijvoorbeeld AES gebruiken in een of andere mode die PRP-veiligheid vereist, vertrouwen we stilzwijgend dat AES PRP-veilig is. Dit vertrouwen is niet gebaseerd op een of ander bewijs, maar op openbare cryptanalyse. Als we een permutatie, bijvoorbeeld Keccak-$f$, gebruiken in een of andere mode die vereist dat keyed duplex PRF-secure is, dan vertrouwen we dat Keccak-$f$ veilig is in deze context. Bij het kiezen van het aantal ronden van een permutatie kunnen we verschillende keuzes maken. Als de doelstelling is dat de veiligheidsgraad van de functie niet lager is dan de veiligheidsgraad van de constructie met een willekeurige permutatie, spreken we van *hermetische* ontwerpstrategie. Dit is bijvoorbeeld het geval bij ons vercijferingsschema Keyak [6]. Maar we kunnen ook tolereren dat de veiligheidsgraad zakt en dit compenseren door hogere waarden voor de parameters in de constructie te kiezen. Zo heeft ons vercijferingsschema Ketje [5] een hogere capaciteit dan strikt nodig voor wat het behoort te doen.




**Figuur 9**

## Farfalle

De gesleutelde duplex levert een efficiënte oplopende OVF op die zeer weinig geheugen vereist maar één nadeel heeft: ze is strikt serieel en kan niet optimaal gebruikmaken van beschikbare rekenkracht in moderne processoren. Om die reden zou het interessant zijn om een meer parallelle OVF te bouwen. We zijn hieraan begonnen en het resultaat is Farfalle [7], zoals hier geïllustreerd in Figuur 9.

Farfalle heeft een compressiefase (links) en een expansiefase (rechts) en maakt gebruik van verschillende permutaties $P_F$, $P_f$ en $P_A$. Het idee is dat deze permutaties dezelfde rondefunctie hebben en enkel verschillen in het aantal ronden. Verder maakt het gebruik van zogenaamde rolfuncties voor het laten variëren van maskers en de interne toestand. Dit zijn zeer lichtgewicht functies vergelijkbaar met lineaire terugkoppelingsregisters. Door de seriële schakeling kan het aantal ronden sterk gereduceerd worden in vergelijking met permutaties gebruikt in duplex voor dezelfde veiligheidsgraad. Bovendien zijn zowel de compressiefase en expansiefase zeer parallelliseerbaar en is ze tegelijk oplopend.

## Besluit

Symmetrische cryptografie 2.0 vormt een interessant alternatief voor blokcijfers en zijn modes.

**Dankwoord**

Ik dank Gilles Van Assche voor het gebruik van zijn illustraties.

## Referenties

1 G. Bertoni, J. Daemen, M. Peeters en G. Van Assche, Sponge functions, *Ecrypt Hash Workshop 2007*, May 2007, http://sponge.noekeon.org.

2 G. Bertoni, J. Daemen, M. Peeters en G. Van Assche, On the indifferentiability of the sponge construction, in: N. P. Smart, ed., *Advances in Cryptology–Eurocrypt 2008*, Lecture Notes in Computer Science, Vol. 4965, Springer, 2008, pp. 181–197.

3 G. Bertoni, J. Daemen, M. Peeters en G. Van Assche, Cryptographic sponge functions, January 2011, http://sponge.noekeon.org/.

4 G. Bertoni, J. Daemen, M. Peeters en G. Van Assche, The KECCAK reference, January 2011, http://keccak.noekeon.org.

5 G. Bertoni, J. Daemen, M. Peeters, G. Van Assche en R. Van Keer, CAESAR submission: KETJE v2, September 2016, http://ketje.noekeon.org.

6 G. Bertoni, J. Daemen, M. Peeters, G. Van Assche en R. Van Keer, CAESAR submission: KEYAK v2, document version 2.2, September 2016, http://keyak.noekeon.org.

7 G. Bertoni, J. Daemen, M. Peeters, G. Van Assche en R. Van Keer, Farfalle: parallel permutation-based cryptography, IACR Cryptology ePrint Archive, Report 2016/1188, http://eprint.iacr.org/2016/1188.

8 J. Daemen en V. Rijmen, *The Design of Rijndael–AES, the Advanced Encryption Standard*, Springer, 2002.

9 J. Daemen, B. Mennink en G. Van Assche, Full-state keyed duplex with built-in multi-user support, IACR Cryptology ePrint Archive, Report 2017/498, http://eprint.iacr.org/2017/498.

10 NIST, Federal information processing standard 46, Data Encryption Standard (DES), October 1999.

11 NIST, Federal information processing standard 197, Advanced Encryption Standard (AES), November 2001.

12 NIST, Federal information processing standard 202, SHA-3 standard: Permutation-based hash and extendable-output functions, August 2015, http://dx.doi.org/10.6028/NIST.FIPS.202.

13 NIST, NIST special publication 800-185, SHA-3 derived functions: cSHAKE, KMAC, TupleHash and ParallelHash (draft), August 2016, http://csrc.nist.gov/publications/drafts/800-185/sp800_185_draft.pdf.

14 P. Rogaway en T. Shrimpton, A provable-security treatment of the keywrap problem, in: S. Vaudenay, ed., *Eurocrypt*, Lecture Notes in Computer Science, Vol. 4004, Springer, 2006, pp. 373–390.

Eric Verheul
*Digital Security Group*
*Radboud University Nijmegen*
*eric.verheul@keycontrols.nl*

Bart Jacobs
*Digital Security Group*
*Radboud University Nijmegen*
*bart@cs.ru.nl*

# Polymorphic encryption and pseudonymisation in identity management and medical research

**Eric Verheul and Bart Jacobs are professors at the Digital Security Group at Radboud University, Nijmegen. Verheul is specialized in information security, and particularly interested in online privacy and security management. Jacobs is specialized in software security, and has made various contributions in the direction of privacy and identity management. This paper sketches how the classical ElGamal public key encryption system can be used in a novel way to provide new privacy protection mechanisms, notably in identity management and in medical research.**

In 2014 the first author identified a paradox in the foreseen Dutch eID scheme [3, 5]. This scheme allows citizens to authenticate to governmental organisations through private parties, like banks or telecom providers. For functional reasons, these parties would need to provide a national citizen identifier called 'BSN' to such governmental organisations. However, Dutch privacy regulation precludes private parties from processing the BSN. This led to the following question: is it possible to store the BSN in some encrypted form at an authentication provider such that it can later be transformed into a form decipherable by, and only by, the intended governmental organisation? During transformation the BSN should not temporarily emerge in the clear at the authentication provider, so that a common decrypt-encrypt transformation would not be suitable. Also, in the end, only the intended governmental organisation should be able to decrypt the BSN. A solution to this problem was needed. But the obvious approach to provide each governmental organisation with the same secret key for decryption would undermine the required level of security.

The second author encountered a similar challenge in healthcare and medical research. Here, different parties (doctors, researchers) wish to investigate patient data from various sources. To avoid cumbersome bilateral exchanges, a central repository is required. As in the eID case, privacy regulation mandates that these data be stored in encrypted form. Hence also in this case it would be desirable if the encrypted data (ciphertext) is transformable to a form that is locally decipherable for the different parties.

Both used cases gave rise to the development of Polymorphic Encryption and Pseudonymisation, abbreviated as PEP. With the similar techniques of polymorphic encryption and polymorphic pseudonymisation new security and privacy guarantees can be given which are essential in areas such as privacy-friendly identity management, (personalised) healthcare, medical data collection via self-measurement apps, and more generally in the internet of things and in data analytics.

The key ideas of polymorphic encryption are:

1. Personal data can be encrypted in a 'polymorphic' manner and stored at a central party in such a way that the central storage facility cannot get access. Crucially, there is no need to fix a priori who can decrypt the data later, so that the data can immediately be protected at the source.
2. Later on it can be decided who can decrypt the data, via some transformation of the encrypted data (ciphtertext) which makes it locally decryptable via locally different (diversified) cryptographic keys. This decision will be made on the basis of a policy, in which the data subject should play a key role.
3. This transformation of encrypted data can be performed by a trusted party in a blind manner, without seeing the content; the resulting transformed ciphertext is transformed into locally decryptable ciphertext, for a specific other party.

In an eID scheme the encrypted data can be a national citizen identifier (like BSN) stored at an authentication provider; it can be decided later which government organisation gets access to it, after a citizen's login request. In healthcare, a PEP-enabled measurement device — operated by a doctor or by a user himself — can immediately encrypt the data; the user can decide later that, for instance, doctors $X$, $Y$, $Z$ may at some stage decrypt and use the data in their diagnosis, or medical research groups $A$, $B$, $C$ may use it for their investigations, or third parties $U$, $V$, $W$ may use it for additional services, et cetera.

This PEP technology can provide the necessary security and privacy infrastructure for big data analytics, where data comes from various sources, like in the internet of things. People can entrust their data in polymorphically encrypted form, and each time decide later to make (parts of) it available (decipherable) for specific parties, for specific analysis purposes. In

this way users remain in control, and can monitor which parts of their data are used where, by whom, and for which purposes.

The polymorphic encryption infrastructure can be supplemented with a pseudonymisation infrastructure which is also polymorphic, and guarantees that each individual will automatically have different pseudonyms at different parties.

This paper provides an introduction to Polymorphic Encryption and Pseudonymisation (PEP), focusing on identity management and healthcare as two application areas.

The PEP framework is currently being implemented in the Dutch eID scheme. The framework is also elaborated into an open design and open source (prototype) implementation at Radboud University in Nijmegen. The technology will be used and tested in a real-life Parkinson research project at the Radboud University Medical Center.

**ElGamal revisited**

The expression 'ElGamal' is used for one of the first asymmetric, public key crypto algorithms, named after its inventor [1]. It can be used both for encryption and for digital signatures. Here we only use the encryption version. This section recalls the basic definitions and results, assuming familiarity only with elementary group theory. In particular, it describes three operations on ElGamal ciphertexts that form the basis for PEP. Indeed, the PEP functionality exploits the 'malleability' of ElGamal encryption, see Lemma 1 below.

ElGamal works in a cyclic group. In practice we shall use (prime order subgroups of) elliptic curves [4] as groups, involving addition of points on a curve, and so we prefer additive notation for a group $\mathbb{G} = (\mathbb{G}, +, 0)$. Let $\mathbb{G}$ be a group of prime order $q$ and let $G \in \mathbb{G}$ be a fixed generator. This means that $q$ is the least non-zero natural number with $q \cdot G = 0$ and that each element $H \in \mathbb{G}$ can be written as $H = k \cdot G$ for a unique $k \in \{0, 1, \ldots, q-1\}$. The latter set is the carrier of the field $\mathbb{F}_q$ of size $q$, which is how we shall write it from now on. With $\mathbb{F}_q^*$ we denote the non-zero elements of the field, i.e. its multiplicative group. A randomly selected element in a set is denoted by $\in_R$.

The security of ElGamal encryption depends on the hardness of the discrete logarithm (DL) problem in the group. The DL problem says: given $n \cdot G \in \mathbb{G}$, for some number $n \in_R \mathbb{F}_q$, then it is computationally infeasible to find $n$ in polynomial time in $\log_2(q)$, i.e. in the number of bits in the binary representation of $q$. A suitable instance of $\mathbb{G}$ is the (largest prime order subgroup of the) Montgomery Elliptic Curve Curve25519 (see https://cr.yp.to/ecdh.html for more information), offering $128$ bits of security, or the Brainpool320r1 curve (see http://www.ecc-brainpool.org) offering $160$ bits of security. The latter curve is currently also used in European electronic passports, including the Dutch ones.

We recall the basics of ElGamal encryption.

*Private key*. The private key $y$ of a user is a random element in $\mathbb{F}_q^*$, which is kept secret by the owner.

*Public key*. The public key $Y \in \mathbb{G}$ is the group element $Y = y \cdot G \in \mathbb{G}$. Due to the DL problem, $y$ cannot (feasibly) be obtained from $Y$ and $G$. This value $Y$ is assumed to be known to everyone.

*Encryption*. Let $M \in \mathbb{G}$ be a message that we wish to encrypt, with public key $Y$. ElGamal encryption is 'randomised' or 'probabilistic': it uses randomness in each encryption so that encrypting the same message twice gives different ciphertexts, with high probability. We choose a non-zero $r \in_R \mathbb{F}_q$ and encrypt $M$ as the pair of group elements:

$$\langle r \cdot G, M + r \cdot Y \rangle. \tag{1}$$

We recall that a fresh (new) random number $r$ should be used for each encryption.

*Decryption*. Let a ciphertext pair $\langle B, C \rangle \in \mathbb{G} \times \mathbb{G}$ be given corresponding to the public key $Y = y \cdot G$. The ElGamal decryption of $\langle B, C \rangle$ is the group element:

$$C - y \cdot B. \tag{2}$$

(We use the letters $B$ for blinding and $C$ for cipher.) One can easily verify correctness, i.e. that decryption returns the original message $M$. Security is based on the DL problem.

*Notation*. We shall write $\mathcal{EG}$ for the ElGamal encryption function, but with a minor twist. We define:

$$\mathcal{EG}(r, M, Y) = \langle r \cdot G, M + r \cdot Y, Y \rangle. \tag{3}$$

As before $r$ is the random number that needs to be different each time. Notice that the function $\mathcal{EG}$ produces a 3-tuple in (3), instead of a 2-tuple in (1): its type is $\mathcal{EG} : \mathbb{F}_q \times \mathbb{G} \times \mathbb{G} \to \mathbb{G} \times \mathbb{G} \times \mathbb{G}$. This is purely for administrative reasons: it makes it easier to formulate the results in Lemma 1 below. We do not use a special function or notation for ElGamal decryption.

We now describe the three homomorphic properties of ElGamal that form the basis of PEP. They are used in the operations of *re-randomising*, *re-keying*, and *re-shuffling* that act on ciphertexts.

**Lemma 1.** *In the notation introduced above we define three functions $\mathcal{RR}$, $\mathcal{RK}$, $\mathcal{RS}$ each with type:*

$$\mathbb{G}^3 \times \mathbb{F}_q^* \to \mathbb{G}^3$$

*and describe their properties.*

(a) *The* re-randomisation *of a triple $\langle B, C, Y \rangle \in \mathbb{G}^3$ with $s \in \mathbb{F}_q^*$ is defined via the function:*

$$\mathcal{RR}(\langle B, C, Y \rangle, s) \overset{\text{def}}{=} \langle s \cdot G + B, s \cdot Y + C, Y \rangle. \tag{4}$$

*If the input $\langle B, C, Y \rangle$ is an ElGamal ciphertext, then so is the output:*

$$\mathcal{RR}(\mathcal{EG}(R, M, Y), s) = \mathcal{EG}(s + r, M, Y). \tag{5}$$

*This ciphertext decrypts to the original message $M$ via the original private key $y$.*

(b) *The* re-keying *with $k \in \mathbb{F}_q^*$ is defined via the function:*

$$\mathcal{RK}(\langle B, C, Y \rangle, k) \overset{\text{def}}{=} \langle \tfrac{1}{k} \cdot B, C, k \cdot Y \rangle, \tag{6}$$

*where $\frac{1}{k}$ is the multiplicative inverse of $k$ in the field $\mathbb{F}_q$. We then have:*

$$\mathcal{RK}(\mathcal{EG}(r, M, Y), k) = \mathcal{EG}(\tfrac{r}{k}, M, k \cdot Y). \tag{7}$$

*This ciphertext decrypts to the original message $M$ via a different private key $k \cdot y$.*

(c) *The* re-shuffling *with $n \in \mathbb{F}_q^*$ is defined as a function:*

$$\mathcal{RS}(\langle B, C, Y \rangle, n) \overset{\text{def}}{=} \langle n \cdot B, n \cdot C, Y \rangle. \tag{8}$$

*Then:*

$$\mathcal{RS}(\mathcal{EG}(r, M, Y), n) = \mathcal{EG}(n \cdot r, n \cdot M, Y). \tag{9}$$

*Hence in this case we can decrypt with the original private key to a re-shuffled message $n \cdot M$.*

*Proof*. All results are obtained by easy calculations. As an illustration we prove that

equation (5) holds: re-randomisation (4) on an ElGamal encryption yields a new ElGamal encryption of the same message with the same public key, but with random number $s+r$, since:

$$\begin{aligned}&\mathcal{RR}(\mathcal{EG}(r,M,Y),s)\\&\overset{(1)}{=}\mathcal{RR}(\langle r\cdot G, r\cdot Y+M, Y\rangle, s)\\&\overset{(4)}{=}\langle s\cdot G+r\cdot G, s\cdot Y+r\cdot Y+M, Y\rangle\\&=\langle (s+r)\cdot G, (s+r)\cdot Y+M, Y\rangle\\&=\mathcal{EG}(s+r,M,Y).\end{aligned}$$ □

The purpose of re-randomisation in the first part of Lemma 1 is to create a copy of an ElGamal encryption that is unlinkable to the original. The obtained unlinkablity is equivalent to a mathematical problem called the Decision Diffie–Hellman problem in $\mathbb{G}$ which is believed to be hard in the elliptic curve groups mentioned earlier. This problem can be formulated as: given $H\in_R\mathbb{G}$ and the quadruple $(G,H,a\cdot G,b\cdot H)$ for $a,b\in_R\mathbb{F}_q^*$ decide if $a=b$. Compare [2, Theorem 10.20]. Sometimes we shall combine the re-keying and re-shuffling operations. The next result tells that the order of such combinations does not matter.

**Lemma 2.** *The re-keying and re-shuffling operations $\mathcal{RK}$ and $\mathcal{RS}$ from Lemma 1 commute. Explicitly:*

$$\mathcal{RS}(\mathcal{RK}(\langle B,C,Y\rangle,k),n) = \mathcal{RK}(\mathcal{RS}(\langle B,C,Y\rangle,n),k).$$

*Proof.* This follows from an easy calculation:

$$\begin{aligned}&\mathcal{RS}(\mathcal{RK}(\langle B,C,Y\rangle,k),n)\\&=\mathcal{RS}(\langle \tfrac{1}{k}\cdot B, C, k\cdot Y\rangle, n)\\&=\langle n\cdot(\tfrac{1}{k}\cdot B), n\cdot C, k\cdot Y\rangle\\&=\langle \tfrac{1}{k}\cdot(n\cdot B), n\cdot C, k\cdot Y\rangle\\&=\mathcal{RK}(\langle n\cdot B, n\cdot C, Y\rangle, k)\\&=\mathcal{RK}(\mathcal{RS}(\langle B,C,Y\rangle,n),k).\end{aligned}$$ □

Based on the above lemma we can combine re-keying and re-shuffling into a single function $\mathcal{RKS}:\mathbb{G}^3\times(\mathbb{F}_q^*)^2\to\mathbb{G}^3$ by:

$$\begin{aligned}&\mathcal{RKS}(\langle B,C,Y\rangle,k,n)\\&=\langle \tfrac{n}{k}\cdot B, n\cdot C, k\cdot Y\rangle.\end{aligned}\qquad(10)$$

**Polymorphic encryption**

From now on we assume that there is a system-wide fixed group $\mathbb{G}$ with generator $G\in\mathbb{G}$ of prime order $q$. Also, some trusted party has generated a master private key $y\in_R\mathbb{F}_q^*$, with corresponding public key $Y=y\cdot G$. The private $y$ is securely stored, for instance in a hardware security module (HSM). It will not be used for decryption, but only for generating other, derived private keys.

Given certain (personal) data $D$, anyone can form what we call the *polymorphic encryption* of $D$, of the form:

$$\mathcal{EG}(r,D,Y)\quad\text{where } r\in_R\mathbb{F}_q^*.\qquad(11)$$

This means that any *data source* can encrypt data, using the master public key $Y$, and a self-chosen random number $r$. Our aim is to transform this ciphertext in such a way that dedicated parties can decrypt it.

We consider a collection of *service providers* $S_j$, for some finite index sets of $j$'s. In order to perform their services, they need to get access to (parts of) the polymorphically encrypted data. In an identity management context this could be a governmental organisation and in an healthcare context this could be a doctor or a medical researcher. Note that in the first context a data source can polymorphically encrypt any data and not only the BSN.

In our setup, there is for each service provider $S_j$ a secret number $s_j\in_R\mathbb{F}_q^*$ that is only known to a trusted party called the *transformer*. The service provider obtains a private key $y_j\in\mathbb{F}_q^*$ which has the form $y_j=s_j\cdot y$, where $y$ is the master private key, mentioned earlier. The corresponding public key $Y_j$ of $S_j$ is then equal to $s_j\cdot Y$, where $Y$ is the master public key. Indeed:

$$y_j\cdot G=(s_j\cdot y)\cdot G=s_j\cdot(y\cdot G)=s_j\cdot Y=Y_j.$$

Given some ciphertext $\langle B,C,Y\rangle$ arising as in (11), the transformer can turn it into a ciphertext that can be decrypted by a given service provider $S_j$. This is done via re-keying with the secret factor $s_j$, as in:

$$\begin{aligned}\mathcal{RK}(\langle B,C,Y\rangle,s_j)&=\langle \tfrac{B}{s_j}, C, s_j\cdot Y\rangle\\&=\langle \tfrac{B}{s_j}, c, Y_j\rangle.\end{aligned}$$

As we have seen in equation (7), via such re-keying, any data $D$ that is polymorphically encrypted with the master public key $Y$, becomes encrypted with the public key $Y_j$ of service provider $S_j$. Hence this service provider can, after this intervention of the transformer, decrypt the data.

We remark that the transformer should also apply re-randomisation on either the input or output of the transformation to avoid linkability issues. Notice that the security of the system rests on having two separate trusted parties, one holding the master private key $y$, and one 'transformer' holding the key factors $s_j$ for each service provider $S_j$. If these two trusted parties collude, the system breaks down. Notice that the transformer manipulates ciphertexts, but cannot see the content. This is a very powerful and useful feature that we will further discuss in the upcoming sections.

**Polymorphic pseudonymisation**

In some cases service providers do not only want access to personal data but also want to have a persistent identifier related to the person to which the data pertains. That is, for the same person this identifier is the same over all sources providing data. In an identity management context this could be a webshop that is not allowed to process the BSN but needs a persistent identifier to give clients access to their own accounts. Different webshops should get different identifiers for the same client, so that they cannot combine their records — simply based on the identifier. Researchers in a healthcare context typically are not allowed to process the BSN either, but need to be able to link the data from various sources to the same individual.

To facilitate these requirements PEP supports service provider specific pseudonyms that can accompany the data. To this end, we assume that the data sources also have access to a 'global' personal identification number $Id$ of the person to which it relates. In a Dutch setting one can think of (some hash of) the earlier mentioned BSN.

In the previous section we assumed a master public key $Y$ and a transformer which holds for each service provider $S_j$ a secret key factor $s_j$. We now assume that there exists another master public key $Z=z\cdot G$ and that the transformer has for each $S_j$ secret key factors $t_j$ similar to $s_j$ and additional 'pseudonym' factors $u_j$. Thus, these $t_j$ and $z$ play the same role as $s_j$ and $y$. All these factors $s_j$, $u_j$, $t_j$ are random but fixed.

For the actual usage of these pseudonyms, the transformer plays an important

role again. Suppose service provider $S_j$ also wants access to a pseudonym related to the person with identity $Id$. Then the data source first embeds $Id$ into the group $\mathbb{G}$ through an (one-way) embedding $\mathcal{I}(\cdot)$. Then the data source polymorphically encrypts $\mathcal{I}(Id)$ using public key $Z$. This results in the *polymorphic pseudonym* $\mathcal{EG}(r,\mathcal{I}(Id),Z)$, and sends this together with the index $j$ to the transformer. The transformer looks up the key factor $t_j$ and the pseudonym factor $u_j$ for service provider $S_j$, and performs both re-keying (with $t_j$) and re-shuffling with $u_j$, written as $\mathcal{RKS}$ in (10). This gives:

$$\begin{aligned}\mathcal{RKS}(\mathcal{EG}(r,\mathcal{I}(Id),Z),t_j,u_j)\\ = \mathcal{EG}(\tfrac{r}{t_j},u_j\cdot\mathcal{I}(Id),t_j\cdot Z)\\ = \mathcal{EG}(\tfrac{r}{t_j},u_j\cdot\mathcal{I}(Id),Z_j)\,.\end{aligned}$$

The result is the encrypted local pseudonym of the form $u_j\cdot\mathcal{I}(Id)$, which can be decrypted by $S_j$. Notice that the transformer learns nothing, except that someone is accessing service provider $S_j$. Each time this process is run, it produces the same local pseudonym $u_j\cdot\mathcal{I}(Id)$ at service provider $S_j$, and a different local pseudonym $u_k\cdot\mathcal{I}(Id)$ at a different service provider $S_k$. Pseudonym unlinkability is guaranteed through the hardness of the Decision Diffie–Hellman problem. As remarked earlier, the transformer should apply re-randomisation on either the input or output of the transformation to avoid linkability issues.

**PEP in the Dutch eID scheme**

In the projected Dutch eID scheme a central government organisation called *BSN Linking* (BSN-l) plays the role of data source discussed in the previous two sections. The transforming role is played by (private) parties performing authentication for the government. As part of user (citizen) registration, authentication providers provide BSN-l with information uniquely identifying the citizen, e.g. first and last name, date of birth et cetera. BSN-l then looks up the citizen and its BSN. The BSN-l forms both a Polymorphic Identity (PI) and a Polymorphic Pseudonym (PP). The PI is simply a polymorphic encryption $\mathcal{EG}(r,BSN,Y)$ of the BSN. The Polymorphic Pseudonym is a polymorphic encryption of the form $\mathcal{EG}(s,\mathcal{I}(BSN),Y)$. The embedded BSN, i.e. $\mathcal{I}(BSN)$, of the previous section is based on a keyed hash function (HMAC). Although the embedded value should never be accessible outside BSN-l, the keyed hash ensures that the BSN cannot be derived from it. Both PI and PP are then sent to the requesting authentication provider and stored in a client database. A citizen can register at multiple authentication providers, for instance in order to have a back-up authentication mechanism.

If registration was successful, the authentication provider supplies the citizen with a (strong) means of authentication, linked to the PI/PP pair. This, for instance, could be a smart card, a challenge/response token or an authentication app on a mobile device. If a citizen wants to login to a web service, he is re-directed to an authentication provider of his choosing — where he has been registered already. By use of the authentication means, the citizen can be linked to its PI/PP in the client database of the authentication provider. If the web service is allowed to use the BSN, the authentication provider then blindly turns the PI to an Encrypted Identity holding the BSN using equation (7). The Encrypted Identity is then sent to the organisation who can decipher the BSN from it. If the organization is not allowed to use the BSN, the authentication provider selects the PP and blindly turns this to Encrypted Pseudonym via (10). This is then sent to the organisation who can decipher a local pseudonym from it.

Dutch governmental organisations are allowed by law to use the BSN, but only if strictly necessary. If a pseudonym suffices, then that should be used. This is also known as the *data minimisation principle* stipulated in European privacy regulations. Hence in the case of governmental organisations there is a choice, namely to use the Polymorphic Identity (PI) — for BSN — or the Polymorphic Pseudonym (PP) — for a pseudonym — at the authentication provider. The status controller introduced below is a first example of a governmental service based on pseudonyms instead of BSNs.

To facilitate these transformations, the authentication providers are given secret factors $s_j$, $t_j$, $u_j$ by the government which need to be stored in a hardware security module. Notice that in the polymorphic setup, authentication providers do not get access to citizen BSNs solving the paradox from the introduction. Actually, the polymorphic setup can also solve another paradox. In the setup indicated above, authentication providers know both the identities of citizens and the service providers that they want to login to. There are many cases where just registering that a user accessed a specific service can constitute a breach of privacy. Such cases include a user retrieving his results of a medical test or a user having an online psychiatric consultation. This issue becomes even more manifest if one is using private organisations that also provide other services. As an illustration, suppose one is regularly logging into an online consultation for alcoholics through a bank acting as authentication provider. How comfortable would one then be to apply for a mortgage or a car insurance application at that bank?

We note that privacy regulations mandate that authentication providers need user consent to send information to the governmental organisation. So not supplying the authentication provider with the identity of the governmental organisation is not suitable. In the projected Dutch eID scheme [3] such 'privacy hotspot' issues are procedurally mitigated: authentication providers are required to separate their registrations holding identifying user data, e.g. name, address, et cetera, from registrations holding usage data, i.e. authentication transactions. Note that the polymorphic setup conveniently caters for this as authentication providers can store transactions under a local pseudonym.

This lead to the question if this separation can be technically enforced: is it possible that an authentication provider authenticates a user for an organisation without knowing the identity of the user? This is paradoxical as the authentication provider is required to identify the user and to personally provide him with means of authentication. This paradox can also be solved through the polymorphic setup via a personal PEP-enabled smart card. This is actually being developed for the public authentication provider (DigiD) in the Dutch eID scheme. For this version of the eID system the PI and PP are not (only) stored by the authentication provider, but also on a contactless Dutch identity card, or driver's license card (hereafter simply called eID card). During authentication DigiD reads the PI and PP from the eID card whereby the card re-randomises these first. DigiD is then able to do the transformation for a governmental organisation but cannot determine the identity of the citizen. Actually,

if the same citizen would authenticate one second later, DigiD would not even be able to determine this. This is due to the hardness of the Diffie–Hellman Decision problem mentioned earlier. To determine if the card has not been revoked, e.g. after loss or theft, DigiD also uses the polymorphic setup. DigiD forms an encrypted pseudonym for a so-called status controller and requests the status of the card by sending this to the controller. The status of the card is maintained by the issuer of the card which is also provided an encrypted pseudonym during production of the card. See also [5].

To allow his eID card to be read by DigiD, the user needs to connect a contactless card reader to his computer or to use a mobile device (smartphone) supporting Near Field Communication (NFC). The eID card is heavily based on electronic passport technology; in effect the PI/PP on the card are protected as fingerprints on passports. Through this technology it is also arranged that the user is technically in control of whether his card is providing both a PI and a PP to DigiD or only a PP. This allows for applications such as referenda where the pseudonym enforces that a citizen can cast his ballot only once but where BSN usage would violate ballot secrecy.

**PEP for medical research**

The second application area for PEP that we briefly elaborate on is medical research. In addition to traditional 'one-time' medical data sources, like an ECG or MRI scan, researchers nowadays like to have continuous, real-time access to patient data, for instance via various wearable monitors and activity trackers. This presents challenges for protected data management.

Via polymorphic encryption each data item $D$ can be stored securely at some storage facility as $\mathcal{EG}(r, D, Y)$. Each device can itself compute such encryptions for the data that it generates, since all that is needed is the public key $Y$. As before, a 'transformer' can keep key factors for each participating researcher or doctor, and re-key the data so that it can be decrypted by that particular party.

In addition, via polymorphic pseudonyms it can be achieved that different researchers receive different pseudonyms for the same patient. This makes it hard to combine data, for instance after data loss or theft. Again, the pseudonymisation factors need to be associated with each participating party, known by the transformer, who can then re-shuffle and re-key in order to form local, decryptable pseudonyms.

By also providing local database pseudonyms to the researchers in encrypted form, researchers can put their findings back into the database, so that it can become visible for other participating researchers. Use of re-randomisation of encrypted pseudonyms avoids linkability issues. Even though these researchers all have different polymorphic pseudonyms, the enriched data that they put back will end up with the right individuals.

The PEP technology will be used for the first time in 2017 on a larger scale in a medical research project on Parkinson's disease (see http://www.parkinsonopmaat.nl), set up jointly by the Radboud University Medical Center and by Verily, the life science branch of the Google group, now called Alphabet. This study will involve 650 patients, who will be monitored for three years. Verily will provide wearable devices for this purpose. Radboud's computer security group, to which the authors belong, contributes with an implementation of the PEP technology (see http://pep.cs.ru.nl for more information; the PEP development is funded by the province of Gelderland).

Apart from the basic functionality sketched above, the PEP implementation provides authentication and authorisation for the various participants. The study is open to other (international) research groups. They will first have to submit a research plan to a supervisory body, and, after approval, will get the appropriate (derived) cryptographic keys, in order to access parts of the collected data that are relevant for their research. In this set-up they will also get their own (polymorphic) pseudonyms for the patients involved.

Thus, the PEP infrastructure functions as a secure database with encryption and pseudonymisation. This sounds ideal, but there are also some restrictions. We mention the two most prominent ones.

1. It always remains possible to de-pseudonymise patients via the contents of the data, especially with rare symptoms, or by combining the data with other sources. PEP will not protect against this. In the Parkinson study such de-pseudonymisation is simply forbidden by contract.
2. When data is stored in encrypted form, searching in the stored data, in order to select specific parts, is not possible. (There are advanced cryptographic techniques for searching in encrypted data, but they have not been integrated (yet) with PEP.) In principle, a researcher will have to download all the data, decrypt locally, and then search and select. This problem is alleviated by storing the encrypted data together with unencrypted meta-data. These meta-data can be used for selection, for instance based on content or dates.

Once the PEP software is sufficiently tested and stable, it will be made available as open source.

**References**

1 T. ElGamal, A Public Key Cryptosystem and a Signature scheme Based on Discrete Logarithms, *IEEE Transactions on Information Theory* 31(4) (1985), 469–472.

2 J. Katz and Y. Lindell, *Introduction to Modern Cryptography*, CRC Press, 2008.

3 Ministry of the Interior and Kingdom Relations, *Uniforme Set van Eisen*, version 1.0, 15-12-2016.

4 D. Hankerson, A. Menezes and S. Vanstone, *Guide to Elliptic Curve Cryptography*, Springer, 2004.

5 E.R. Verheul, The polymorphic eIDAS token, *Keesing Journal of Documents & Identity*, February 2017.

Christian Schaffner
*QuSoft and ILLC*
*University of Amsterdam*
*c.schaffner@uva.nl*

# Recent developments in quantum cryptography

**Christian Schaffner is a NWO Vidi Laureate and assistant professor at the ILLC at UvA, where he works on quantum cryptography. Quantum cryptography is the art and science of exploiting quantum mechanical effects in order to perform cryptographic tasks. Recent progress in building quantum computers has led to new opportunities for cryptography, but also endangers existing cryptographic schemes. In this article Schaffner surveys both aspects of this double-edged sword.**

## Quantum supremacy

We stand at a special moment in time, where the promise of a quantum computer and quantum internet seems to be within reach. Large-scale efforts around the world by academic and industrial players including many heavyweights such as Google, IBM, Microsoft and Intel are in progress to develop quantum technologies. The first experiments demonstrating 'quantum supremacy' are expected to be carried out possibly as early as at the end of 2017. Such experiments will demonstrate that a quantum computer can solve some (probably uninteresting) problem faster than any classical supercomputer. In order to outperform today's biggest classical computers, a quantum computer has to work with at least $40$ to $50$ *quantum bits* (also called *qubits*). In contrast to a classical bit which can take on values either $0$ or $1$, a quantum bit can be in an arbitrary superposition between $0$ and $1$. Mathematically, the (pure) state of a single qubit can be expressed as unit vector in a two-dimensional complex Hilbert space $\mathcal{H}$, i.e.

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$$
$$\text{with } \alpha,\beta \in \mathbb{C} \text{ and } |\alpha|^2 + |\beta|^2 = 1$$

where the two states $|0\rangle, |1\rangle$ form an orthonormal basis of $\mathcal{H}$, and $\alpha$ and $\beta$ are called amplitudes of the state. Measuring the quantum state $|\phi\rangle$ in basis $|0\rangle, |1\rangle$ yields a classical bit $b$ as outcome. Outcome $b=0$ is obtained with probability $|\alpha|^2$ and outcome $b=1$ is obtained with probability $|\beta|^2$.

Quantum mechanics dictates that the state space of $n$ qubits is given by the $n$-fold tensor product $\mathcal{H}^{\otimes n}$. Remarkably, this space has dimension $2^n$ which is exponential in the number of qubits! Therefore, the state of a $40$-qubit quantum computer can be described by $2^{40}$ complex amplitudes, and a quantum computation on such a device can be classically simulated by tracking the changes of all these amplitudes during the computation. The exponential size of the state space explains why it becomes intractable to classically simulate arbitrary quantum computations on, say, more than $50$ qubits.

One of the most promising proposals by the Google group [4] for an experiment which demonstrates the superiority of a quantum computer over all classical computers is to perform the following quantum computation: initialize a two-dimensional grid of $7$ by $7$ qubits in a fixed state, and apply a number of random $2$-qubit quantum gates on any two neighboring qubits. The squared amplitudes of the quantum state after the application of the quantum circuit define a probability distribution, and repeated application of the same quantum procedure allows to sample from this distribution by simply measuring the final quantum state. There is complexity-theoretic evidence that it is difficult to sample from the same distribution on a classical computer [1, 4].

Clearly, such a sampling problem is not of great practical interest. However, it is well-known since the late eighties that large-scale quantum computers can solve some important problems far more efficiently than classical computers. The most well-known example is the problem of finding the prime factorisation of an integer $N \in \mathbb{N}$. For large problem sizes, the fastest classical algorithms have an expected running time $O(e^{(\log N)^{1/3}})$ which is subexponential in the size of the integer $N$ to be factored. In contrast, a quantum algorithm discovered by Peter Shor requires only $O((\log N)^2)$ quantum operations to solve the problem. This time complexity is polynomial in the size of $N$ and therefore exponentially faster than the best-known classical algorithm.

The main high-level idea of Shor's quantum algorithm is to reduce the factoring problem to the problem of period-finding, which can be solved efficiently on a quantum computer by exploiting the power of the quantum Fourier transform. It turns out that also other cryptographically relevant problems such as the discrete-logarithm problem in finite fields can be reduced to period-finding, and hence fall to Shor's quantum algorithm as well, see the box on the next page.

**Discrete logarithms**

For a finite group $G$ of order $q$, generated by an element $g \in G$, the *discrete-logarithm problem* is defined as follows: Given a uniformly random group element $h \in G$, find $x$ such that $g^x = h$. For a generic group $G$ (i.e. if one can only use the group operation, and no additional mathematical structure of $G$), one can prove that any classical algorithm must perform $\Omega(\sqrt{q})$ group operations to compute discrete logarithms [18]. Again, Shor's algorithm can solve the problem even in a generic group using a number of quantum operations which is polynomial in $\log(q)$. Popular choices of the finite group $G$ in cryptography are $\mathbb{Z}_p^*$, the multiplicative group of integers modulo a large prime $p$ (with some additional requirements), or the elliptic-curve groups over finite fields. The advantage of elliptic curves over $\mathbb{Z}_p^*$ is that instance sizes can be much smaller while keeping the conjectured difficulty of the discrete-logarithm problem at the same level.

**Quantum-safe cryptography**

Cryptography is of utmost importance in today's digital world. Cryptographic systems that use classical communication are widely deployed and form the cornerstone of digital technologies, ranging from the internet to mobile phones. If present-day cryptographic systems were broken, we could no longer securely perform online banking, use mobile devices or the internet of things, our medical data would no longer be protected, et cetera. Almost all of the currently employed public-key cryptography is based on either the factoring or on the discrete-logarithm problem. Therefore, these systems can all be broken using Shor's quantum algorithm. An example is the widely deployed RSA cryptosystem [15]. Forging RSA signatures on operating-system updates would allow an attacker to gain control of billions of computers around the world.

Crucially, the security of such cryptographic systems can be broken *retroactively* by an attacker who obtains a large-scale quantum computer in the future. Data encrypted today can thus lose its confidentiality once a quantum computer becomes available. For the protection of state secrets or medical data, new forms of quantum-safe cryptographic protection have to be employed *already today* in order to guarantee the confidentiality of the data for the coming decades.

At the moment, large parts of the international cryptographic research community are searching for new public-key schemes based on mathematical problems which are believed to be hard also for quantum computers. The US National Institute for Standards in Technology (NIST) has initiated a project to determine a new standard for such quantum-safe scheme, see [14]. (The term 'post-quantum cryptography' is quite well-established for these quantum-safe schemes, but chosen somewhat unfortunately, because the research area is concerned with cryptography which is still secure *at the beginning* and not after the end of the era of large-scale quantum computers.)

The main contenders in terms of assumptions for classical quantum-safe cryptography are the following:

- lattice-based cryptography (see article by Léo Ducas in this issue),
- multivariate cryptography,
- hash-based cryptography,
- code-based cryptography,
- supersingular elliptic curve isogeny cryptography (see article by Mathijs Coster in this issue).

In order to evaluate the proposed schemes and assumptions, an extensive amount of quantum cryptoanalysis will be performed over the coming years. The challenge is that expertise from very different research fields such as cryptography, mathematics and quantum computing is required. The Quantum Software Consortium, a collaboration of researchers from CWI, QuSoft, TU Delft and Leiden university have recently obtained a large gravitation grant from NWO (18.8 million Euro for ten years). As this consortium unites experts from all aforementioned fields, it is natural that a considerable part of the grant will be spent on the quantum cryptanalysis of candidate schemes for quantum-safe cryptography. A thorough understanding of the capabilities of current quantum algorithms and quantum computers is essential to determine the correct parameters for which the above hardness assumptions hold.

It is important to note that the use of quantum technology cannot only be used to attack existing classical schemes, but on the contrary, it has been realized already in the late sixties by Stephen Wiesner [19] that the use of quantum communication can lead to new cryptographic applications. The most famous example is Quantum Key Distribution (QKD) [3, 12] which allows two honest parties Alice and Bob to establish a classical secret key even in the presence of an eavesdropper Eve who has complete control over the quantum communication between Alice and Bob, and who can eavesdrop on (but not modify) the classical communication. It can be proven in an information-theoretic sense that no matter how much computational power an eavesdropper has, her knowledge about the secret key obtained by Alice and Bob can be made arbitrarily small. It is quite remarkable that such an information-theoretical key establishment is possible using quantum communication, because for purely classical communication, it is excluded by Shannon's impossibility result for perfectly secure encryption [16].

On a higher level, quantum key distribution is yet another form of quantum-safe cryptography which can be compared to the other proposals above. The main advantage of QKD is the information-theoretic security of the established keys, and the fact that an attacker cannot retroactively break the security of QKD. An attacker has to act at the moment when the key-establishment protocol is performed, so large-scale quantum computers developed in the future do not change the security of keys established today. The main disadvantages of QKD are the high deployment costs (because quantum communication is required) and the current distance limit of about $150$ kilometers between Alice and Bob (which implies that for larger distances, a multi-hop connection is required where the middle nodes have to be trusted).

**Quantum fully-homomorphic encryption**

In recent research [11], we were able to develop a new method for securely delegating a quantum computation. In fact, we developed the quantum analogue of what is classically known as fully-homomorphic encryption. Homomorphic encryption is a special form of encryption and is best motivated by the scenario of secure cloud computing, where a user Alice would like to outsource a difficult or tedious computation to an untrusted computing cloud.

For instance, Alice would like to tag a lot of holiday pictures, but instead of doing this tedious job manually, she would like to delegate the task to an advanced image-recognition algorithm in the cloud. As Alice does not trust the cloud with all her private images, she can use homomorphic encryption to encrypt her pictures and send them to the cloud, which cannot see the contents due to the secrecy of the encryption but which can nevertheless run its tagging algorithms on the encrypted images due to the homomorphic property of the encryption. As a result, the cloud returns encrypted tags of Alice's pictures, and Alice is the only person who can decrypt them. For a long time, cryptographers thought it was impossible to construct such homomorphic encryption systems, until Craig Gentry showed in a seminal article in 2009 how to do this [13]. This breakthrough sparked an enormous amount of cryptographic research on the topic and led in recent years to another hot topic of cryptographic research on (software) obfuscation.

Given the recent progress building a quantum computer, it is very natural to consider the scenario of a quantum-computing cloud. In fact, it is rather likely that quantum computing will only be available at a few quantum-computing facilities in the world. As a concrete example, IBM is already providing access to a mini quantum-computation cloud by allowing the general public to run algorithms on their 5-qubit computer. Hence, we were wondering whether it is possible to construct homomorphic encryption schemes for quantum data. Previous results in this direction had serious drawbacks and allowed for homomorphic evaluations of only a very limited set of operations. Our recent result [11] provides quantum encryptions that allow homomorphic evaluations of arbitrary efficient quantum circuits.

Illustration: Tremani, TU Delft

**Figure 1** In early stages of the quantum age, quantum computing will only be available in specialized facilities. Secure delegated quantum computation allows to outsource a quantum computation to a server in the quantum-computing cloud without revealing the inputs.

Our quantum fully-homomorphic encryption scheme offers a round-optimal solution to the problem of *secure delegated quantum computation*, see Figure 1. Pioneering work of Childs [10] and Arrighi and Salvail [2] studied this problem for the first time. The first practical and universal protocol for private delegated quantum computation, called 'universal blind quantum computation', was given by Broadbent, Fitzsimons and Kashefi [5]. In this protocol, the client only needs to be able to prepare random single-qubit auxiliary states, but does not require quantum memory nor a quantum processor. Via a classical interaction phase, the client remotely drives a quantum computation of her choice, such that the quantum server cannot learn any information about the computation that is performed — with only the client learning the output. Our new scheme from [11] reduces the number of rounds of interaction between client and server, at the expense of some extra demands on the quantum capabilities of the client.

Currently, we are working on extending our scheme so that the client can classically verify whether the server has carried out the correct quantum computation. Such a property seems crucial for using it as a building block to obtain more advanced cryptographic primitives such as zero-knowledge proofs [8], quantum one-time programs [6], or quantum obfuscation.

## Conclusion

These are exciting times for the active research field of quantum cryptography, as quantum computers are becoming a reality in the near future. Many challenges await us in finding alternatives to the currently broken classical public-key cryptosystems, and in further exploiting the power of quantum computing to extend cryptographic schemes into the quantum domain.

**Acknowledgments**

Small parts of this article have appeared previously in a survey article co-authored by Anne Broadbent [9], where we introduce quantum cryptography to classical cryptographers. I would like to thank Anne Broadbent for her kind permission to reuse some parts here. I am also grateful for comments on this article from Anne Broadbent and Ronald de Wolf.

## References

1 S. Aaronson and L. Chen, Complexity-theoretic foundations of quantum supremacy experiments, arxiv:1612.05903.

2 P. Arrighi and L. Salvail, Blind quantum computation. *Int. J. Quantum Inf.* 4 (2006), 883–898.

3 C.H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, *International Conference on Computers, Systems and Signal Processing*, 1984, pp. 175–179.

4 S. Boixo, S.V. Isakov, V.N. Smelyanskiy, et al., Characterizing quantum supremacy in near-term devices, arxiv:1608.00263.

5 A. Broadbent, J. Fitzsimons and E. Kashe, Universal blind quantum computation, in *FOCS 2009*, pp. 517–526.

6 A. Broadbent, G. Gutoski and D. Stebila, Quantum one-time programs, in *CRYPTO 2013*, pp. 344–360.

7 A. Broadbent and S. Jeery, Quantum homomorphic encryption for circuits of low T-gate complexity, in *CRYPTO 2015*, pp. 609–629.

8 A. Broadbent, Z. Ji, F. Song and J. Watrous, Zero-knowledge proof systems for QMA, in *FOCS 2016*, IEEE, 2016, pp. 31–40.

9 A. Broadbent and C. Schaner, Quantum cryptography beyond quantum key distribution, *Designs, Codes and Cryptography* 78 (2016), 351–382.

10 A. Childs, Secure assisted quantum computation, *Quantum Information and Computation* 5 (2005), 456–466.

11 Y. Dulek, C. Schaner and F. Speelman, Quantum homomorphic encryption for polynomial-sized circuits, in *CRYPTO 2016*, pp. 3–32.

12 A.K. Ekert, Quantum cryptography based on bell's theorem, *Physical Review Letters* 67 (1991), 661–663.

13 C. Gentry, Fully homomorphic encryption using ideal lattices, in *STOC 2009*, Vol. 9, pp. 169–178.

14 National Institute of Standards and Technology, Post-Quantum crypto Project, http://csrc.nist.gov/groups/ST/post-quantum-crypto.

15 R.L. Rivest, A. Shamir and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, *Communications of the ACM* 21(2) (1978), 120–126.

16 C.E. Shannon, Communication theory of secrecy systems, *The Bell System Technical Journal* 28(4) (1949), 656–715.

17 P.W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *35th Annual Symposium on Foundations of Computer Science – FOCS 1994*, IEEE, 1994, pp. 124–134.

18 Victor Shoup, Lower bounds for discrete logarithms and related problems, in *EUROCRYPT 1997*.

19 S. Wiesner, Conjugate coding. *SIGACT News* 15(1) (1983), 78-88, originally written in 1968 but unpublished.

Robert Granger
*School of Computer and Communication Sciences*
*École polytechnique fédérale de Lausanne*
*robert.granger@epfl.ch*

# Indiscreet discrete logarithms

**In 2013 and 2014 a revolution took place in the understanding of the discrete logarithm problem (DLP) in finite fields of small characteristic. Consequently, many cryptosystems based on cryptographic pairings were rendered completely insecure, which serves as a valuable reminder that long-studied so-called hard problems may turn out to be far easier than initially believed. In this article, Robert Granger gives an overview of the surprisingly simple ideas behind some of the breakthroughs and the many computational records that have so far resulted from them.**

By way of motivation, we begin with the landmark work of Diffie and Hellman, who in 1976 introduced the following very well known key-agreement protocol, which allows two parties — referred to as Alice and Bob — to agree a shared secret key over an insecure channel which can then be used for secure communications between them [13]. To do so, Alice and Bob agree in advance on two public system parameters: $p$ a prime integer and $g$ a primitive root modulo $p$. We denote by $\mathbb{F}_p$ the finite field of $p$ elements, represented as usual as the quotient $\mathbb{Z}/p\mathbb{Z}$ with coset representatives always in $\{0,\ldots,p-1\}$; $g$ thus generates the multiplicative group $\mathbb{F}_p^\times$. To establish a shared key, Alice picks a secret integer $a \in \{1,\ldots,p-1\}$, computes $g^a$ and sends this to Bob. Likewise, Bob picks a secret integer $b \in \{1,\ldots,p-1\}$, computes $g^b$ and sends this to Alice. Using their respective secrets Alice and Bob can both compute the shared key

$$g^{ab} = (g^b)^a = (g^a)^b.$$

In order for this key to be secure, it is necessary that it is hard to compute $g^{ab}$ from the public information $(g,g^a,g^b)$, for some notion of 'hard' that is discussed later on. The task of doing so is known as the *Diffie–Hellman problem* (DHP). One way to solve the DHP is to recover $a$ from $(g,g^a)$ and then compute the shared key as Alice does (or equivalently recover $b$ from $(g,g^b)$ and compute the shared key as Bob does). Hence, it is also necessary that the problem of recovering $a$ from $(g,g^a)$ — known as the *discrete logarithm problem* (DLP), since it is the inverse of exponentiation — is hard to solve too. Note that finite cyclic groups other than $\mathbb{F}_p^\times$ may also be used to instantiate this protocol. We therefore formalise the DLP more generally with the following.

**Definition 1.** Given a finite cyclic group $(G,\cdot)$, a generator $g \in G$ and another group element $h \in G$, the DLP is the problem of finding an integer $t$ such that $h = g^t$. The integer $t$ — denoted by $\log_g h$ — is uniquely determined modulo the group order and is called the *discrete logarithm* of $h$ with respect to the base $g$.

Although the DHP is clearly reducible to the DLP, in the sense that an algorithm to solve the latter provides an algorithm to solve the former, it is not known whether the converse holds in general; there are however several positive results in this direction [7, 39, 40]. Since there are no known algorithms which solve the DHP directly, research on the hardness of the DHP has focused almost entirely on the hardness of the DLP, which explains its cryptographic importance. We note that while necessary, the hardness of the DHP is by no means sufficient to ensure the security of the protocol, since several other issues must also be addressed, see [50, Chapter 18], for example.

The above key-agreement protocol can easily be extended to a public-key encryption scheme, in which Alice can send a message $m$ to Bob over an insecure channel, without having first agreed on a secret key with him [14]. In particular, Bob chooses a key-pair $(b,g^b)$ consisting of a private key $b$, which is kept secret, and a public key $g^b$, which is published. To encrypt a message $m \in \{1,\ldots,p-1\}$ to Bob, Alice chooses a random $a \in \{1,\ldots,p-1\}$ and sends to him $(c_1,c_2) = (g^a, m(g^b)^a)$. Bob decrypts by computing $m = c_2/c_1^b$, using his private key $b$. One can also obtain a digital signature scheme in a similar manner [14] and there are a large number of variations and cryptosystems with more complex properties, all of which rely on the hardness of the DLP in one form or another. Hence, having groups in which the DLP is hard is essential.

## Pairing-based cryptography

An interesting family of protocols which are pertinent to this story are those which arose from the invention of pairing-based cryptography in 2000, allowing cryptographic functionalities such as identity-based non-interactive key distribution [47], one-round tripartite key-agreement [25] and identity-based encryption [8] (and later several hundred others). All of these rely on the existence of certain non-degenerate efficiently computable bilinear maps, known as pairings. Such a map has the form

$$e : G_1 \times G_2 \to G_3,$$

where $G_1$ and $G_2$ are abelian groups of exponent $l \in \mathbb{N}$, which by convention are written in additive notation with identity

element $0$, and $G_3$ is a cyclic group of order $l$, written in multiplicative notation with identity element $1$.

The non-degeneracy condition is that for all $P \in G_1 \backslash \{0\}$ there is a $Q \in G_2$ such that $e(P,Q) \neq 1$, and for all $Q \in G_2 \backslash \{0\}$ there is a $P \in G_1$ such that $e(P,Q) \neq 1$. The bilinearity condition is that for all $P, P' \in G_1$ and all $Q, Q' \in G_2$ one has

$$e(P+P',Q) = e(P,Q)\,e(P',Q),$$
$$e(P,Q+Q') = e(P,Q)\,e(P,Q'),$$

which implies that for all $P \in G_1$, all $Q \in G_2$ and all $a, b \in \mathbb{Z}$ one has

$$e(aP,bQ) = e(P,Q)^{ab}. \tag{1}$$

Although such maps arise naturally from the Tate and Weil pairings on arbitrary abelian varieties over local or finite fields, for efficiency purposes they are usually instantiated using elliptic curves over finite fields. In this case, for the Tate pairing [16] we have $G_1 = E(\mathbb{F}_q)[l]$, the group of $l$-torsion points on an elliptic curve $E$ defined over $\mathbb{F}_q$, with $q = p^r$ and $l$ coprime to $p$. $G_3$ is the group $\mu_l$ of $l$-th roots of unity in $\overline{\mathbb{F}}_q$, which embeds into $\mathbb{F}_{q^n}^{\times}$, with $n$ the order of $q$ modulo $l$, also known as the embedding degree. Finally, $G_2$ is the quotient group $E(\mathbb{F}_{q^n})[l]/lE(\mathbb{F}_{q^n})[l]$, whose coset representatives we do not describe here. For the definition of the pairing itself and other technical conditions we refer the interested reader to [5, Chapter 9], which contains a comprehensive introduction to the area.

The property (1) was originally exploited using the linearity of the pairing in the first argument only, in order to transfer a DLP from an elliptic curve group to a DLP in an extension of the underlying base field [41]. Indeed, if the input DLP in $G_1$ is $(P, aP)$, one selects an appropriate $Q \in G_2$ such that $e(P,Q) \neq 1$ — which exists by the non-degeneracy condition — and computes $g = e(P,Q)$ and $g^a = e(aP,Q)$. The *raison d'être* of this transfer is that in general the best algorithms for solving the finite field DLP have lower complexity than the best algorithms for solving the elliptic curve DLP, so even though the inputs to each problem have different sizes, namely $n \log_2 q$ and $\log_2 q$, if the embedding degree is small enough then the transferred DLP will be easier to solve.

However, it is the full bilinearity that enables interesting cryptographic applications. Such applications require that the embedding degree is small enough that the pairing itself is efficiently computable, but large enough that the DLP in $\mathbb{F}_{q^n}^{\times}$ is hard. Therefore, while as with the key-agreement protocol each pairing-based protocol comes with a set of problems other than the DLP which must be hard in order for it to be secure, all are vulnerable to developments in discrete logarithm algorithms for the finite field DLP.

**Hardness of the DLP**

Let $(G, \cdot)$ be a finite cyclic group of known order $N$ in which the group operation is assumed to be computable in unit time, and let $g$ be a generator of $G$. First observe that exponentiation in $G$ can be performed in time polynomial in the bitlength of $N$, i.e., $\lceil \log_2 N \rceil$, since one can take the binary expansion of an exponent $e$ and compute $g^e$ via the square-and-multiply algorithm or its variants. Hence, for a group $G$ to be useful for discrete logarithm-based cryptography, discrete logarithms should not be computable in polynomial time, and should preferably be *much* harder to compute.

Note that there are groups in which the DLP is easy. For instance, if $G = (\mathbb{Z}/N\mathbb{Z}, +)$ and $1 \leq g \leq N-1$ is coprime to $N$, then $g$ is a generator and 'exponentiation by $e$' is just $eg \pmod N$. Thus the discrete logarithm of any element $h$ is just $hg^{-1} \pmod N$. As all cyclic groups of order $N$ are isomorphic to $(\mathbb{Z}/N\mathbb{Z}, +)$, one might expect the DLP in such groups to be easy. However, since computing the image of such an isomorphism requires solving a DLP with respect to a generator, this need not be so. Indeed, the representation of group elements may obscure the cyclic structure to varying degrees, which dictates the apparent hardness of the DLP in each group.

If one insists upon not exploiting any information regarding the representation of group elements, i.e., the worst case from a cryptanalytic perspective, then the DLP in such groups can be analysed using the *generic group model*. This model stipulates that group elements are represented using random encodings, while the group operation is performed using an oracle which takes as input the encodings of two elements and outputs the encoding of the group operation applied to the input elements. In this case it can be shown that the DLP requires $\Omega(\sqrt{N})$ oracle calls in order to be solved with high probability, i.e., at least $c\sqrt{N}$ for some constant $c > 0$, for $N$ sufficiently large [43, 49]. Since this is exponential in the size of the problem, namely $\log N$, we see that in the ideal case from a cryptography perspective, the DLP is exponentially hard. Of course, the DLP can be no harder than exponential since a naive enumeration of powers of the generator will solve it too. Note that a square root complexity can be achieved using a standard time-space trade-off known as Baby Step/Giant Step, or a memory efficient version based on random walks, due to Pollard [46]. Further note that if the prime factorisation of $N$ is known then the DLP can always be reduced to a set of DLPs in prime order subgroups, by projecting into them via exponentiation by their cofactors, using a form of Hensel lifting for prime-power order subgroups, and applying the Chinese remainder theorem [45]. So in practice the DLP can be assumed to be in a group of prime order. These results imply that in order to solve the DLP in a time faster than exponential, one *must* exploit representational properties of elements of the group. In some scenarios this is possible, as for multiplicative groups of finite fields for example, while in others it is apparently not, as for the group of $\mathbb{F}_p$-rational points on a suitably chosen elliptic curve. The latter explains the popularity of elliptic curve cryptography, first proposed in 1985 independently by Miller [42] and Koblitz [33], which essentially achieves optimal security per bit as a result.

In general, the hardness of a DLP is measured by the complexity of the fastest algorithm known to solve it. The following function is often used in this regard:

$$L_N(\alpha, c) = \exp\big((c + o(1))(\log N)^{\alpha}(\log\log N)^{1-\alpha}\big),$$

where $\alpha \in [0,1]$, $c > 0$, $\log$ denotes the natural logarithm and the $o(1)$ denotes a function that tends to zero as $N \to \infty$. Observe that $L_N(0,c) = (\log N)^{c+o(1)}$, which thus represents algorithms which run in polynomial time, while $L_N(1,c) = N^{c+o(1)}$ represents algorithms which run in exponential time. For $0 < \alpha < 1$, the function $L_N(\alpha, c)$ represents algorithms which are said to run in *subexponential* time. As is customary we often omit the subscript $N$ and the constant $c$ when convenient.

The first subexponential algorithm for the finite field DLP was shown by Adleman in 1979 to have *heuristic* complexity $L(1/2)$ [1], see the next section. It is termed heu-

ristic because the analysis relied on unproven assumptions. In 1984 Coppersmith proposed the first (again, heuristic) $L(1/3)$ algorithm for the DLP in binary fields, i.e., in $\mathbb{F}_{2^n}$ [11], which generalises to arbitrary families of extension fields $\mathbb{F}_{q^n}$ with a fixed base field, also referred to as small characteristic fields. The later development of the number field sieve [37] and the function field sieve [2,3,27,28] led to heuristic $L(1/3)$ algorithms for all finite fields [20,29]. Between 1984 and 2013, no algorithms went below the $L(1/3)$ barrier, although the far less important constant $c$ was occasionally lowered for some $(q,n)$ families. It thus seemed plausible that this was the natural complexity of the finite field DLP, and cryptographers were fairly confident that it would not be broken any time soon, excepting of course the possible development of a large-scale quantum computer, which threatens all DLPs as well as the integer factorisation problem, thanks to Shor's algorithm from 1994 [48].

Each of the subexponential algorithms exploits the property that field elements, when viewed as elements in the parent ring, can be factored into a product of irreducible elements. Thanks to this property a very natural and broadly applicable framework first discussed by Kraitchik in the 1920s [34,35] can be applied, namely, *index calculus*, which we now introduce.

**Index calculus**

The term index calculus — which literally means 'calculating the index' — originates from the at least two-centuries-old name used by Gauss for the discrete logarithm of an integer modulo $p$ relative to a primitive root, namely, the index [17, art. 57–60]. For a group $G$ written multiplicatively and a generator $g$, let $h \in G$ be an element whose discrete logarithm with respect to $g$ is to be computed. The index calculus method consists of the following three steps.

1. *Relation generation:* Choose a subset $\mathcal{F} \subset G$ which we call the factor base, and find multiplicative relations between its elements. Observe that each such relation provides a linear equation in the logarithms of the factor base elements with respect to any generator, modulo $|G|$.
2. *Linear algebra:* Once at least $|\mathcal{F}|$ linearly independent equations between logarithms of elements of $\mathcal{F}$ have been generated, obtain these logarithms by solving the corresponding linear system.
3. *Individual logarithms:* Find an expression for $h$ as a product of factor base elements, for example by computing $hg^e$ for random $e$ until this factors completely over $\mathcal{F}$, from which one can easily deduce $\log_g h$.

The elements of $\mathcal{F}$ are usually chosen to be the set of 'prime' elements whose 'norm' is less than some bound (for some notions of prime and norm), since such a choice generates the maximum number of elements of $G$ amongst all sets of the same cardinality. How steps (1) and (3) are performed in practice depends very much on the group in question and the ingenuity of the cryptanalyst. In order to illustrate the method we now present a very simple example.

**Example.** Let $p = 1009$. Then $g = 11$ is a generator of $G = \mathbb{F}_p^\times$. Let $\mathcal{F} = \{2,3,5,7\}$. Relations are obtained by computing $g^e \bmod p$ for random $e \in \{1,\ldots,p-1\}$ and then using trial division to check whether this integer is a product of the primes in $\mathcal{F}$. The following relations were quickly obtained:

$$\begin{aligned} 11^{796} \bmod p &= 15 = 3 \cdot 5, \\ 11^{678} \bmod p &= 315 = 3^2 \cdot 5 \cdot 7, \\ 11^{992} \bmod p &= 63 = 3^2 \cdot 7, \\ 11^{572} \bmod p &= 10 = 2 \cdot 5. \end{aligned}$$

Note that the factorisations occur in the parent ring $\mathbb{Z}$ of $\mathbb{F}_p \cong \mathbb{Z}/p\mathbb{Z}$. These relations yield the following equations mod $p-1$:

$$\begin{aligned} 796 &\equiv \log_{11} 3 + \log_{11} 5, \\ 678 &\equiv 2\log_{11} 3 + \log_{11} 5 + \log_{11} 7, \\ 992 &\equiv 2\log_{11} 3 + \log_{11} 7, \\ 572 &\equiv \log_{11} 2 + \log_{11} 5. \end{aligned}$$

Writing this linear system in matrix form we have:

$$\begin{bmatrix} 796 \\ 678 \\ 992 \\ 572 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \log_{11} 2 \\ \log_{11} 3 \\ \log_{11} 5 \\ \log_{11} 7 \end{bmatrix}.$$

The matrix is invertible mod $p-1$ and solving the system yields the solutions: $\log_{11} 2 = 886$, $\log_{11} 3 = 102$, $\log_{11} 5 = 694$ and $\log_{11} 7 = 788$, as one can easily verify.

For an individual logarithm, the first non-trivial case is $h = 13$, for which $\log_{11} 13$ can be computed as follows. Testing random $e$, we quickly find that

$$\begin{aligned} hg^e &= 13 \cdot 11^{53} \bmod p \\ &= 720 = 2^4 \cdot 3^2 \cdot 5, \end{aligned}$$

and hence

$$\begin{aligned} \log_{11} 13 &= (4\log_{11} 2 + 2\log_{11} 3 \\ &\quad + \log_{11} 5 - 53) \bmod p - 1 \\ &= 357. \end{aligned}$$

A basic question arising from this approach is how large should the factor base be in order to optimise the running time, as $p \to \infty$? This depends on the density of smooth numbers, whose definition we now recall.

**Definition 2.** A positive integer is said to be $B$-smooth if all of its prime divisors are at most $B$.

The following result on the asymptotic density of smooth numbers amongst the integers is due to Canfield, Erdős and Pomerance [9].

**Theorem 1.** *A uniformly random integer in $\{1,\ldots,M\}$ is $B$-smooth with probability*

$$P = u^{-u(1+o(1))}, \text{ where } u = \frac{\log M}{\log B},$$

*provided that $3 \leq u \leq (1-\varepsilon)\frac{\log M}{\log\log M}$ for some $\varepsilon > 0$.*

Clearly, the larger the factor base then the higher the probability that a uniformly random element of $\mathbb{F}_p^\times$, viewed as an integer, is smooth with respect to the factor base. However, one then requires more relations. On the other hand, a smaller factor base means fewer relations are needed, but each is harder to find. Theorem 1 indicates how to optimise this trade-off; in particular, using the very aptly defined function $L_N(\alpha,c)$ it implies the following.

**Corollary 1.** *Let $M = L_N(2\alpha,\mu)$ and $B = L_N(\alpha,\beta)$. Then the expected number of trials until a uniformly random number in $\{1,\ldots,M\}$ is $B$-smooth is $L_N\left(\alpha,\frac{\alpha\mu}{\beta}\right)$.*

For the $\mathbb{F}_p^\times$ index calculus algorithm we have $M = N = L(1,1)$. Corollary 1 implies that we should set the smoothness bound to be $B = L(1/2,\beta)$ for some unknown $\beta > 0$. Since we need about $|F| \approx B/\log B \leq B$ relations, the estimated running time is

$$L(\tfrac{1}{2},\beta)\cdot L(\tfrac{1}{2},\tfrac{1}{2\beta}) = L(\tfrac{1}{2},\beta+\tfrac{1}{2\beta}).$$

This complexity is minimised for $\beta = 1/\sqrt{2}$, resulting in a running time of $L(1/2,\sqrt{2})$ for step 1. For step 2, as the matrices generated are incredibly sparse, i.e., have very few non-zero entries, by using either Lanczos' algorithm [36] or Wiedemann's algorithm [52], the complexity is about $B^2 = L(1/2,2\beta) = L(1/2,\sqrt{2})$ as well. Step 3 is obviously of lower complexity since only one relation is needed.

For fixed $q$ and $n \to \infty$, one needs to employ a different, but equally natural notion of smoothness in order to apply an analogous algorithm to solve the DLP in $\mathbb{F}_{q^n}^{\times}$; note that the norm of an element in this scenario is its degree.

**Definition 3.** An element in $\mathbb{F}_q[x]$ of positive degree is said to be $b$-smooth if all of its irreducible factors are of degree at most $b$.

The following result on the asymptotic density of smooth polynomials amongst those of the same degree is due to Odlyzko [44] and Lovorn [38].

**Theorem 2.** *A uniformly random polynomial $f \in \mathbb{F}_q[X]$ of degree $m$ is $b$-smooth with probability $P = u^{-u(1+o(1))}$, where $u = \frac{m}{b}$, provided that $m^{1/100} \leq b \leq m^{99/100}$.*

With this notion of smoothness and a corollary to Theorem 2 analogous to Corollary 1 but with $N$ now $q^n$ rather than $p$, the algorithm given for $\mathbb{F}_p^{\times}$ applies to $\mathbb{F}_{q^n}^{\times}$ *mutatis mutandis* and one can show that it also has a running time of $L(1/2,\sqrt{2})$. Note that as described above the algorithm is heuristic since there is no guarantee that the relations generated produce a linear system of full rank. However, it can be made rigorous by using an elementary argument due to Enge and Gaudry [15].

**Obtaining faster algorithms**

The previous analysis demonstrates that when elements of the field in question are generated uniformly at random, an $L(1/2)$ complexity is optimal for the index calculus algorithm employed. To obtain algorithms of better complexity, there are (at least) two approaches that one could attempt to employ.

The first approach is to generate relations between elements of smaller norm than before, and for complexity analysis purposes assume that any such subset of elements has the same smoothness density as uniformly random elements of that norm, which is a *smoothness heuristic*. This is precisely what the $L(1/3)$ algorithms do. In particular, elements of norm $\log(L(2/3))$ are produced and the factor base consists of elements of norm $\log(L(1/3))$. Applying Corollary 1 for the integers (or its analogue for polynomials) once again means that the running times of steps 1 and 2 are both $L(1/3)$. Since the factor base is now smaller, in order to obtain an $L(1/3)$ complexity for step 3 one needs to employ a *descent* strategy. A descent begins by expressing the target element as a product of elements of lower norm, mod $p$ or mod the field-defining polynomial, rather than in $\mathbb{N}$ or the polynomial ring, respectively. When such an expression has been obtained we say that the element has been *eliminated*, since one need no longer compute its logarithm directly; only the logarithms of the elements in the obtained product are needed. By iteratively eliminating all of the elements featured in the product one obtains expressions for the target element of lower and lower norm, until finally one has an expression over the factor base, from which one can easily deduce the target logarithm. Subject to the above smoothness heuristic, techniques to do this have an $L(1/3)$ complexity, but with a smaller $c$ than for steps 1 and 2.

The second approach — which seems not to have even been appreciated as a possibility prior to 2013, perhaps due to the desire to assume the above smoothness heuristic for the sake of the complexity analysis — is to generate relations between elements which have *higher smoothness probabilities* than uniformly random elements of the same norm. While no method is known for achieving this over the integers — and thus for the DLP in $\mathbb{F}_p^{\times}$ — the breakthrough results from 2013 onwards all came about because two ways to do this usefully for polynomial rings — and thus for the DLP in small characteristic extension fields — were independently discovered at essentially the same time. The first was due to Göloğlu, Granger, McGuire and Zumbrägel (referred to hereafter as GGMZ) [18], while the second was due to Joux [26]. Although the ingredients of the two methods are somewhat different, they may be viewed as being essentially isomorphic. Both methods produce relations for a factor base whose size is only *polynomial* in the bitlength of the field in question, in *polynomial time*, as the smoothness probability is exponentially larger than before; indeed, such relations are smooth by construction. Since the factor base in this scenario is very small, usually consisting of degree one elements over a suitable base field, step 3 must now descend further which leads to complexities worse than $L(1/3)$ when using the old, or 'classical' techniques, with the complexity being highest for degree two elimination. Hence, in order to make these insights fully applicable, new descent strategies were also needed.

GGMZ proposed a polynomial-time method for eliminating degree two elements on the fly, i.e., on an element by element basis as required, while Joux proposed a polynomial time method for computing the logarithms of degree two elements in batches, as well as a technique to eliminate elements of very small degree, which leads to a heuristic $L(1/4+o(1))$ algorithm. The two degree two elimination methods led respectively to two very different *quasi-polynomial time* algorithms for solving the DLP in small characteristic extension fields, this complexity arising from the descent step. In particular, Joux's approach led to the first such algorithm in mid 2013 and is due to Barbulescu, Gaudry, Joux and Thomé [4], while Göloğlu et al.'s approach led to the second in early 2014 and is due to Granger, Kleinjung and Zumbrägel (referred to hereafter as GKZ), appearing in the preprint [22] and its final version [23]. Since the GKZ algorithm is somewhat simpler and far more practical than the former, and is rigorously proven for an infinite family of extensions of every base field, we detail only this one in this article.

**The GKZ algorithm**

Unlike in the $L(1/2)$ algorithms, how the target field is represented is now of paramount importance. The GKZ algorithm applies to fields of the form $\mathbb{F}_{q^{kn}}$, with $18 \leq k = o(\log q)$ and $n \approx q$ (see Theorem 5 for additional technical conditions). Any small characteristic field $\mathbb{F}_{q'^n}$ can be embedded into such a field by setting $q = q'^{\lceil \log_{q'} n \rceil}$, thereby increasing the extension degree by a factor of $k\lceil \log_{q'} n \rceil$. The use of such an embedding does not significantly affect the

algorithm's resulting complexity. The field setup used in [23] can be either from [18] (with a small modification from [21]), or from [26], both of which may be seen in the context of the Joux–Lercier doubly-rational function field sieve variant [28].

Let $h_0, h_1 \in \mathbb{F}_{q^k}[X]$ be coprime and of degree $d_h \leq 2$, such that $h_1(X)X^q - h_0(X) \equiv 0 \pmod{I}$ for an irreducible degree $n$ polynomial $I \in \mathbb{F}_{q^k}[X]$. Heuristically, such $h_0, h_1$ can always be found. Let $x$ be a root of $I$ in $\mathbb{F}_{q^{kn}}$, so that $\mathbb{F}_{q^{kn}} = \mathbb{F}_{q^k}(x)$. Observe that by the choice of field-defining polynomial we have $x^q = h_0(x)/h_1(x)$. The factor base is:

$$\mathcal{F} = \{f \in \mathbb{F}_{q^k}[X] \,|\, \deg(f) \leq 1\} \cup \{h_1(x)\}.$$

Let $g \in \mathbb{F}_{q^{kn}}^{\times}$, let $h \in \langle g \rangle$, let $h = g^t$ with the integer $t$ to be computed and let $N = q^{kn} - 1$ be the order of the multiplicative group of $\mathbb{F}_{q^{kn}}$. Thanks to a small adaptation of the argument given by Diem [12], which is an adaptation of that given by Enge and Gaudry [15], one does not need to compute the logarithms of the factor base elements, as we now sketch. Let $F = |\mathcal{F}|$ and let the elements of $\mathcal{F}$ be $f_1, \ldots, f_F$. One constructs a matrix $R = (r_{i,j}) \in (\mathbb{Z}/N\mathbb{Z})^{(F+1)\times F}$ and column vectors $\alpha, \beta \in (\mathbb{Z}/N\mathbb{Z})^{F+1}$ as follows. For each $i$ with $1 \leq i \leq F+1$ choose $\alpha_i, \beta_i \in \mathbb{Z}/N\mathbb{Z}$ uniformly and independently at random and apply the to-be-explained randomised descent algorithm to $g^{\alpha_i} h^{\beta_i}$ to express this as

$$g^{\alpha_i} h^{\beta_i} \equiv \prod_{j=1}^{F} f_j^{r_{i,j}} \pmod{I}. \tag{2}$$

One then computes a lower row echelon form $R'$ of $R$ by using invertible row transformations and applies these row transformations to $\alpha$ and $\beta$, resulting in vectors $\alpha'$ and $\beta'$ respectively. Since the first row of $R'$ vanishes, we have $g^{\alpha'_1} h^{\beta'_1} = 1$ and hence $\alpha'_1 + t\beta'_1 \equiv 0 \pmod{N}$. If $\gcd(\beta'_1, N) = 1$ then one can invert $\beta'_1$ to compute $t$. One can prove that $\beta'_1$ is uniformly distributed in $\mathbb{Z}/N\mathbb{Z}$ (cf. [23, Lemma 2.1]) and so the algorithm succeeds with (the very high) probability $\phi(N)/N$; if it does not then one simply repeats the algorithm until it is successful.

We therefore need only describe the descent procedure for carrying out (2), which need only be applied $F+1$ times, i.e., a polynomial number of times. This depends on the recursive application of degree two elimination, as we now explain.

**The descent**

First note that any element in $\mathbb{F}_{q^{kn}}^{\times}$ can be lifted to an irreducible element of degree $2^e$ in $\mathbb{F}_{q^k}[X]$, provided that $2^e > 4n$, thanks to a Dirichlet-type theorem due to Wan [51, Theorem 5.1], so one applies this to each featured $g^{\alpha_i} h^{\beta_i}$ before descending to the factor base. Second, we claim the following.

**Proposition 1.** *Let $Q \in \mathbb{F}_{q^k}[X]$ be an irreducible polynomial of degree $2d \geq 2$. Then $Q$ can be expressed mod $I$ as a product of at most $q+2$ irreducible polynomials of degree dividing $d$, in time $\mathrm{poly}(q,d)$.*

To see that this implies a quasi-polynomial time algorithm, observe that if $Q$ is irreducible of degree $2^e$, then one application of Proposition 1 leads to at most $q+2$ irreducibles of degree dividing $2^{e-1}$. Applying it to each of these leads to at most $(q+2)^2$ irreducibles of degree dividing $2^{e-2}$. Recursively lowering the degrees in this way eventually leads to at most $(q+2)^e$ degree one polynomials and takes time at most $(q+2)^e \mathrm{poly}(q) = (q+2)^{\log_2 n} \mathrm{poly}(q)$ to compute, as $d$ is at most $2^{e-1} \approx n \approx q$. The running time for the algorithm is therefore $q^{\log_2 n + O(k)}$, which is quasi-polynomial in $q^{kn}$ as claimed.

We now show that in order to prove Proposition 1 it is sufficient for there to be an efficient elimination method for irreducible degree two polynomials in $\mathbb{F}_{q^{kd}}[X]$, expressing each as a product of at most $q+2$ linear polynomials mod $I$. Let $Q$ be as in Proposition 1. Observe that over the degree $d$ extension $\mathbb{F}_{q^{kd}}$ the polynomial $Q$ factors into $d$ irreducible quadratics $Q_1 \cdot \cdots \cdot Q_d$. Applying the hypothesised degree two elimination method to any one of these quadratics — say $Q_1$ — expresses it as a product of at most $q+2$ linear polynomials over $\mathbb{F}_{q^{kd}}$ mod $I$. Then applying the norm map with respect to the extension $\mathbb{F}_{q^{kd}}/\mathbb{F}_{q^k}$ to both sides of the expression maps $Q_1$ back to $Q$ and the linear polynomials to powers of irreducible polynomials of degree dividing $d$ (the degree depending on the base field of the respective constant terms), which thus eliminates $Q$ as per the proposition.

We now describe such a degree two elimination method which first featured in [18]. Let $Q_1 \in \mathbb{F}_{q^{kd}}[X]$ be an irreducible quadratic to be eliminated mod $I$. For $a, b, c \in \mathbb{F}_{q^{kd}}$ consider the following equivalence mod $I$:

$$\begin{aligned} X^{q+1} + aX^q + bX + c \\ \equiv \frac{1}{h_1(X)} (Xh_0(X) + ah_0(X) \\ + bXh_1(X) + ch_1(X)). \end{aligned} \tag{3}$$

Denote the left-hand side and the numerator of the right-hand side of (3) by $L(X)$ and $R(X)$, respectively. The condition $Q_1 \mid R(X)$ can be expressed as

$$b = u_0 a + v_0, \quad c = u_1 a + v_1, \tag{4}$$

for some $u_i, v_i \in \mathbb{F}_{q^{kd}}$ (at least in general; some degenerate cases are easily obviated, see [23, Section 3.1]). Note that since $d_h \leq 2$, the cofactor of $Q_1$ in $R(X)$ has degree at most one and so no smoothness heuristics are required as long as $L(X)$ splits completely over $\mathbb{F}_{q^{kd}}$. Crucially, although the degree of $L(X)$ is $q+1$, such polynomials split completely over $\mathbb{F}_{q^{kd}}$ with probability $\approx 1/q^3$, which is exponentially larger than the $1/(q+1)!$ one expects for uniformly random polynomials of this degree. Indeed, for $kd \geq 3$ if $ab \neq c$ and $b \neq a^q$, $L(X)$ may be transformed (up to a scalar) into

$$\begin{aligned} &F_B(\overline{X}) = \overline{X}^{q+1} + B\overline{X} + B, \\ &\text{with } B = \frac{(b - a^q)^{q+1}}{(c - ab)^q}, \end{aligned} \tag{5}$$

via $X = \frac{c-ab}{b-a^q}\overline{X} - a$. $L(X)$ splits whenever $F_B$ splits and the transformation from $\overline{X}$ to $X$ is valid. The following theorem is due to Bluher [6].

**Theorem 3.** *The number of elements $B \in \mathbb{F}_{q^{kd}}^{\times}$ such that the polynomial $F_B(\overline{X}) \in \mathbb{F}_{q^{kd}}[\overline{X}]$ splits completely over $\mathbb{F}_{q^{kd}}$ equals*

$$\begin{aligned} &\frac{q^{kd-1} - 1}{q^2 - 1} \quad \textit{if } kd \textit{ is odd}, \\ &\frac{q^{kd-1} - q}{q^2 - 1} \quad \textit{if } kd \textit{ is even}. \end{aligned}$$

In both cases the number of such $B$ is $\approx q^{kd-3}$. Let $\mathcal{B}$ be the set of all $B \in \mathbb{F}_{q^{kd}}^{\times}$ such that $F_B$ splits completely over $\mathbb{F}_{q^{kd}}$. Since for any $B \in \mathcal{B}$ one can freely choose $a$ and any $b \neq a^q$, while the expression for $B$ in (5) determines $c$ uniquely, there are $\approx q^{3kd-3}$ such $L(X)$ which split completely over $\mathbb{F}_{q^{kd}}$, which explains the $1/q^3$ splitting probability. One way to intuit why this number is so large is that the subset of polynomials of the form $L(X)$ which split completely over $\mathbb{F}_{q^{kd}}$ may be seen to arise from taking the homogeneous eval-

uation of Möbius transformations of $X$ in $X^q - X = \prod_{\alpha \in \mathbb{F}_q}(X-\alpha)$. In particular, for $a',b',c',d' \in \mathbb{F}_{q^{kd}}$ with $a'd'-b'c' \neq 0$ one has

$$\begin{aligned}(c'X+d')^{q+1}&\left(\left(\frac{a'X+b'}{c'X+d'}\right)^q - \frac{a'X+b'}{c'X+d'}\right)\\ &= (a'X+b')^q(c'X+d')\\ &\quad -(a'X+b')(c'X+d')^q \qquad (6)\\ &= (c'X+d')\\ &\quad \times \prod_{\alpha\in\mathbb{F}_q}(a'X+b'-\alpha(c'X+d')), \qquad (7)\end{aligned}$$

where (6) is of the same form as $L(X)$ (up to a scalar) and (7) is a product of linear polynomials. Indeed, this is precisely how Joux approached obtaining such $L(X)$ [26, Section 4.2]. Joux also showed that the number of such polynomials is

$$\begin{aligned}&|\operatorname{PGL}_2(\mathbb{F}_{q^{kd}})|/|\operatorname{PGL}_2(\mathbb{F}_q)|\\ &\quad = (q^{3kd}-q^{kd})/(q^3-q) \approx q^{3kd-3},\end{aligned}$$

broadly matching the number arising from the approach already described.

The following theorem due to Helleseth and Kholosha [24, Theroem 5] (generalised to arbitrary characteristic) characterises the set $\mathcal{B}$.

**Theorem 4.**

$$\mathcal{B} = \left\{ \frac{(z^{q^2}-z)^{q+1}}{(z^q-z)^{q^2+1}} \,\middle|\, z \in \mathbb{F}_{q^{kd}} \backslash \mathbb{F}_{q^2} \right\}.$$

Combining Theorem 4 with the expression for $B$ in (5) and the expressions for $b$ and $c$ in (4), to eliminate $Q_1$ one needs to find an $(A,Z) \in \mathbb{F}_{q^{kd}} \times (\mathbb{F}_{q^{kd}} \backslash \mathbb{F}_{q^2})$ satisfying

$C/\mathbb{F}_{q^{kd}}$:

$$\begin{aligned}&(Z^{q^2}-Z)^{q+1}(-u_1A^2+(v_1-u_0)A+v_0)^q\\ &-(Z^q-Z)^{q^2+1}(-A^q+u_0+Au_1)^{q+1}=0.\end{aligned}$$

That there are sufficiently many points on $C$ was proven in [23] by analysing the action of $\operatorname{PGL}_2(\mathbb{F}_q)$ on $Z$ in order to prove that there is an absolutely irreducible factor of $C$, and then applying the Weil bound. One also needs to consider so-called *descent traps*, which are elements that divide $h_1(X)X^{q^{kd+1}} - h_0(X)$ for $d \geq 0$ which can not be eliminated in the above manner and so must be avoided during the descent. Computing points on $C$ is efficient since one can take any $Z \in \mathbb{F}_{q^{kd}} \backslash \mathbb{F}_{q^2}$ which gives a polynomial in $A$, whose $\mathbb{F}_{q^{kd}}$ roots can be computed by taking the greatest common denominator with $A^{q^{kd}} - A$, for instance, which takes time polynomial in $\log q^{kn}$. The above algorithm and considerations lead to the following [23, Theorem 1.2].

**Theorem 5.** *Given a prime power $q > 61$ that is not a power of 4, an integer $k \geq 18$, coprime polynomials $h_0, h_1 \in \mathbb{F}_{q^k}[X]$ of degree at most two and an irreducible degree $n$ factor $I$ of $h_1X^q - h_0$, the DLP in $\mathbb{F}_{q^{kn}} \cong \mathbb{F}_{q^k}[X]/(I)$ can be solved in expected time*

$$q^{\log_2 n + O(k)}.$$

Thanks to Kummer theory, such $h_1, h_0$ are known to exist when $n = q-1$, which gives the following easy corollary when $m = ik(p^i-1)$ [23,Theorem 1.1].

**Theorem 6.** *For every prime $p$ there exist infinitely many explicit extension fields $\mathbb{F}_{p^m}$ in which the DLP can be solved in expected quasi-polynomial time*

$$\exp\big((1/\log 2 + o(1))(\log m)^2\big).$$

One may also replace the prime $p$ in Theorem 6 by a (fixed) prime power $p^r$ by setting $k = 18r$. Proving the existence of $h_0, h_1$ for general extension degrees as per Theorem 5 seems to be a hard problem, even though in practice it is very easy to find such polynomials and heuristically is almost certain.

**Computational records and impact**

Since early 2013 several computational records have been set using the techniques from [18, 19, 21, 23, 26, 30], which have dwarfed previous records, demonstrating categorically their superiority. Moreover, such large scale computations also help to inform one of potential pitfalls (cf. the traps noted in [21, Remark 1]), and can also lead to theoretical insights that give rise to novel or improved algorithms.

In practice, since the running time is dominated by the descent one first computes the logarithms of the factor base elements (cf. [18, Section 3] and [26, Section 4.2]), so that only one descent is needed. A descent usually consists of: several classical elimination steps; the GKZ elimination for irreducibles of small even degree (for which $kd \geq 4$ suffices in practice) and Joux's elimination method for irreducibles of small odd degree [26]; and finally degree two elimination, either from GGMZ or [26]. The crossover points between these techniques should be determined using a dynamic programming bottom-up approach [21].

Table 1 contains a selection of discrete logarithm computations in finite fields. All details may be found in [10]. At the time

| Bitlength | Charact. | Kummer | Who and when | Complexity |
|---|---|---|---|---|
| 127 | 2 | no | Coppersmith, 1984 | $L(1/3, [1.526, 1.587])$ |
| 401 | 2 | no | Gordon and McCurley, 1992 | $L(1/3, [1.526, 1.587])$ |
| 521 | 2 | no | Joux and Lercier, 2001 | $L(1/3, 1.526)$ |
| 607 | 2 | no | Thomé, 2002 | $L(1/3, [1.526, 1.587])$ |
| 613 | 2 | no | Joux and Lercier, 2005 | $L(1/3, 1.526)$ |
| 556 | medium | yes | Joux and Lercier, 2006 | $L(1/3, 1.442)$ |
| 676 | 3 | no | Hayashi et al., 2010 | $L(1/3, 1.442)$ |
| 923 | 3 | no | Hayashi et al., 2012 | $L(1/3, 1.442)$ |
| 1175 | medium | yes | Joux, 24 December 2012 | $L(1/3, 1.260)$ |
| 1425 | medium | yes | Joux, 6 January 2013 | $L(1/3, 1.260)$ |
| 1778 | 2 | yes | Joux, 11 February 2013 | $L(1/4 + o(1))$ |
| 1971 | 2 | yes | GGMZ, 19 February 2013 | $L(1/3, 0.763)$ |
| 4080 | 2 | yes | Joux, 22 March 2013 | $L(1/4 + o(1))$ |
| 6120 | 2 | yes | GGMZ, 11 April 2013 | $L(1/4)$ |
| 6168 | 2 | yes | Joux, 21 May 2013 | $L(1/4 + o(1))$ |
| 1303 | 3 | no | AMOR, 27 January 2014 | $L(1/4 + o(1))$ |
| 4404 | 2 | no | GKZ, 30 January 2014 | $L(1/4 + o(1))$ |
| 9234 | 2 | yes | GKZ, 31 January 2014 | $L(1/4 + o(1))$ |
| 3796 | 3 | no | Joux and Pierrot, 15 September 2014 | $L([o(1), 1/4 + o(1)])$ |
| 1279 | 2 | no | Kleinjung, 17 October 2014 | $L([o(1), 1/4 + o(1)])$ |
| 4841 | 3 | no | Adj et al., 18 July 2016 | $L([o(1), 1/4 + o(1)])$ |

**Table 1** A selection of discrete logarithm computations in finite fields.

of writing the largest example DLP to have been solved was in the field of $2^{9234}$ elements, which took $\approx 45$ core years of computation. The impact on cryptography can be seen from the solution of DLPs in the fields of bitlength $4404$ and $4841$, which both arise from what were designed to be industry-standard $128$-bit secure supersingular curves. These took $\approx 5$ and $\approx 200$ core years, respectively. Note that these fields can not be represented by a Kummer extension; fields which admit such a representation make the computation much easier due to the presence of factor base automorphisms and other descent advantages, making them ideal for setting records. Since parameters would have to increase significantly to counter the quasi-polynomial time algorithms, thus making the cryptosystems inefficient, and since further algorithmic developments in this area should be expected, small characteristic supersingular curves (or those with low embedding degree) should be considered completely insecure for pairing-based cryptography.

In summary, we see that long-studied so-called hard problems can suddenly become easy with the right ideas, while basing cryptography on unproven computational assumptions is inherently risky in general. Whether the prime field DLP or the integer factorisation problem will remain hard remains to be seen: the current record bitlength for both of these problems is $768$, which took $\approx 5300$ [32] and $\approx 1700$ [31] core years, respectively. However, the ideas behind the breakthroughs do not seem to be extendable to these scenarios since there is no analogue of the tremendously useful polynomial $X^q - X$ around which one can build such an algorithm.

**Acknowledgements**
The author would like to thank Arjen Lenstra for his useful comments.

**References**

1 L. M. Adleman, A subexponential algorithm for the discrete logarithm problem with applications to cryptography, *20th Annual Symposium on Foundations of Computer Science,* IEEE, 1979, pp. 55–60.

2 L. M. Adleman, The function field sieve, *Algorithmic Number Theory*, Springer, 1994, pp. 108–121.

3 L. M. Adleman, M.-D.A. Huang, Function field sieve method for discrete logarithms over finite fields, *Inform. and Comput.* 151(1) (1999), 5–16.

4 R. Barbulescu, P. Gaudry, A. Joux and E. Thomé, A heuristic quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic, *Advances in Cryptology–EUROCRYPT 2014*, LNCS 8441, Springer, 2014, pp. 1–16.

5 I. F. Blake, G. Seroussi, N. P. Smart, *Advances in Elliptic Curve Cryptography*, Cambridge University Press, 2005.

6 A. W. Bluher, On $x^{q+1} + ax + b$, *Finite Fields Appl.* 10(3) (2004), 285–305.

7 B. den Boer, Diffie–Hellman is as strong as discrete log for certain primes, *Proceedings on Advances in Cryptology–CRYPTO '88*, Springer, 1990, pp. 530–539.

8 D. Boneh, M. Franklin, Identity-based encryption from the Weil pairing, *Advances in Cryptology–CRYPTO 2001*, LNCS 2139, Springer, 2001, pp. 213–229.

9 E. R. Canfield, P. Erdős and C. Pomerance, On a problem of Oppenheim concerning 'factorisatio numerorum', *J. Number Theory*, 17(1) (1983), 1–28.

10 Computations of discrete logarithms sorted by date, https://members.loria.fr/LGremy/dldb/index.html.

11 D. Coppersmith, Fast evaluation of logarithms in fields of characteristic two, *IEEE Trans. Inform. Theory* 30(4) (1984), 587–594.

12 C. Diem, On the discrete logarithm problem in elliptic curves, *Compositio Math.* 147 (2011), 75–104.

13 W. Diffie and M. E. Hellman, New directions in cryptography, *IEEE Trans. Inform. Theory*, 22(6) (1976), 644–654.

14 T. ElGamal, A public key cryptosystem and a signature scheme based on discrete logarithms, *Advances in Cryptology–CRYPTO '84*, LNCS 196, Springer, 1985, pp. 10–18.

15 A. Enge and P. Gaudry, A general framework for subexponential discrete logarithm algorithms, *Acta Arithmetica* 102 (2002), 83–103.

16 S. D. Galbraith, K. Harrison and D. Soldera, Implementing the Tate Pairing, *Proceedings of the 5th International Symposium on Algorithmic Number Theory, ANTS-V*, Springer, 2002, pp. 324–337.

17 C. F. Gauß, *Disquisitiones Arithmeticae*, Leipzig, 1801. Translated by A. A. Clarke. Yale University Press, 1965.

18 F. Göloğlu, R. Granger, G. McGuire and J. Zumbrägel, On the function field sieve and the impact of higher splitting probabilities: application to discrete logarithms in $\mathbb{F}_{2^{1971}}$ and $\mathbb{F}_{2^{3164}}$, *Advances in Cryptology–CRYPTO 2013*, LNCS 8043, Springer, 2013, pp. 109–128.

19 F. Göloğlu, R. Granger, G. McGuire and J. Zumbrägel, Solving a 6120-bit DLP on a desktop computer, *Selected Areas in Cryptography–SAC 2013*, LNCS 8282, Springer, 2014, pp. 136–152.

20 D. M. Gordon, Discrete logarithms in $GF(p)$ using the number field sieve, *SIAM J. Discrete Math.* 6(1) (1993), 124–138.

21 R. Granger, T. Kleinjung and J. Zumbrägel, Breaking '128-bit secure' supersingular binary curves, *Advances in Cryptology–CRYPTO 2014*, LNCS 8617, Springer, 2014, pp. 126–145.

22 R. Granger, T. Kleinjung and J. Zumbrägel, On the powers of $2$, *IACR Cryptology ePrint Archive* (2014), eprint.iacr.org/2014/300.

23 R. Granger, T. Kleinjung and J. Zumbrägel, On the discrete logarithm problem in finite fields of fixed characteristic, to appear in *Transactions of the AMS*.

24 T. Helleseth and A. Kholosha, $x^{2^l+1} + x + a$ and related affine polynomials over $GF(2^k)$' *Cryptogr. Commun.* 2(1) (2010), 85–109.

25 A. Joux, A one round protocol for tripartite Diffie–Hellman, *Algorithmic Number Theory*, Springer, 2000, pp. 385–393.

26 A. Joux, A new index calculus algorithm with complexity $L(1/4 + o(1))$ in small characteristic, *Selected Areas in Cryptography–SAC 2013*, LNCS 8282, Springer, 2014, pp. 355–379.

27 A. Joux and R. Lercier, The function field sieve is quite special, *Algorithmic Number Theory*, Springer, 2002, pp. 431–445.

28 A. Joux and R. Lercier, The function field sieve in the medium prime case, *Advances in Cryptology–EUROCRYPT 2006*, LNCS 4117, Springer, 2006, pp. 254–270.

29 A. Joux, R. Lercier, N. Smart and F. Vercauteren, The number field sieve in the medium prime case, *Advances in Cryptology–CRYPTO 2006,* Springer, 2006, pp.326–344.

30 A. Joux and C. Pierrot, Improving the polynomial time precomputation of Frobenius representation discrete logarithm algorithms, *Advances in Cryptology–ASIACRYPT 2014*, LNCS 8873, Springer, 2014, pp. 378–397.

31 T. Kleinjung, K. Aoki, J. Franke, A. K. Lenstra, E. Thomé, J. W. Bos, P. Gaudry, A. Kruppa, P. L. Montgomery, D. A. Osvik, H. J. J. te Riele, A. Timofeev and P. Zimmermann, Factorization of a 768-Bit RSA Modulus, *Advances in Cryptology–CRYPTO 2010*, LNCS 6223, Springer, 2010, pp. 333–350.

32 T. Kleinjung, C. Diem, A. K. Lenstra, C. Priplata and C. Stahlke, Computation of a 768-bit Prime Field Discrete Logarithm, *Advances in Cryptology–EUROCRYPT 2017*, LNCS 10210, Springer, 2017, pp. 333–350.

33 N. Koblitz, Elliptic Curve Cryptosystems, *Mathematics of Computation* 48 (1987), 203–209.

34 M. Kraitchik, *Théorie des nombres*, Vol. 1, Gauthiers-Villars, 1922.

35 M. Kraitchik, *Recherches sur la théorie des nombres*, Gauthiers-Villars, 1924.

36 C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Research Nat. Bur. Standards* 45 (1950), 255–282.

37 A.K. Lenstra and H.W. Lenstra, Jr (eds.), *The Number Field Sieve*, Springer, 1993.

38 R. Lovorn, *Rigorous Subexponential Algorithms for Discrete Logarithms over Finite Fields*, Ph.D. thesis, University of Georgia, 1992.

39 U.M. Maurer, Towards the equivalence of breaking the Diffie–Hellman protocol and computing discrete logarithms, *Advances in Cryptology–CRYPTO '94*, LNCS 839, Springer, 1994, pp. 271–281.

40 U.M. Maurer and S. Wolf, Diffie–Hellman Oracles, *Advances in Cryptology–CRYPTO '96*, LNCS 1109, Springer, 1996, pp. 268–282.

41 A. Menezes, S. Vanstone and T. Okamoto, Reducing elliptic curve logarithms to logarithms in a finite field, *Proceedings of the Twenty-third Annual ACM Symposium on Theory of Computing (STOC '91)*, ACM, 1991, pp. 80–89.

42 V. Miller, Uses of Elliptic Curves in Cryptography, *Advances in Cryptology–CRYPTO 1985*, LNCS 218, Springer, 1985, 417–426.

43 V.I. Nechaev, On the complexity of a deterministic algorithm for a discrete logarithm, *Mat. Zametki* 55(2) (1994), 91–101, 189.

44 A.M. Odlyzko, Discrete logarithms in finite fields and their cryptographic significance, *Advances in Cryptology–CRYPTO '84*, LNCS 209, Springer, 1985, pp. 224–314.

45 S.C. Pohlig and M.E. Hellman, An improved algorithm for computing logarithms over $GF(p)$ and its cryptographic significance (Corresp.), *IEEE Trans. Inform. Theory* 24(1) (1978), 106–110.

46 J.M. Pollard, Monte Carlo methods for index computation (mod $p$), *Math. Comp.* 32(143) (1978), 918–924.

47 R. Sakai, K. Ohgishi and M. Kasahara, Cryptosystems based on pairing, *Symposium on Cryptography and Information Security*, Okinawa, Japan, 2000, pp. 26–28.

48 P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Computing* 26(5) (1997), 1484–1509.

49 V. Shoup, Lower bounds for discrete logarithms and related problems, *Advances in Cryptology–EUROCRYPT '97*, LNCS 1223, Springer, 1997, pp. 256–266.

50 N.P. Smart, *Cryptography Made Simple*, Springer, 2016.

51 D. Wan, Generators and irreducible polynomials over finite fields, *Math. Comp.* 66(219) (1997), 1195–1212.

52 D.H. Wiedemann, Solving sparse linear equations over finite fields, *IEEE Trans. Inform. Theory* 32 (1986), 54–62.

Léo Ducas
*Cryptology Group*
*Centrum voor Wiskunde & Informatica*
*ducas@cwi.nl*

# Advances on quantum cryptanalysis of ideal lattices

**Léo Ducas is a NWO Veni Laureate working in CWI's Cryptology Group on lattice-based cryptology. He co-authored the efficient quantum-safe key exchange algorithm 'New Hope' which was awarded the Facebook Internet Defense Award. Although using lattices with additional structure can lead to more efficient cryptographic algorithms, in this article Ducas explains how lattices that have too much additional structure may be insecure due to efficient quantum attacks.**

The problem of finding a shortest vector of a Euclidean lattice (the shortest vector problem, or SVP) is a central hard problem in complexity theory. Approximated versions of this problem (e.g. $\alpha$-SVP, the problem of finding a vector at most $\alpha$ times longer than the shortest one) have become the theoretical foundation for many cryptographic constructions. Indeed, lattice-based cryptography typically benefits from *worst-case hardness* [1, 14, 18]: it is sufficient that there exists *some* lattices in which finding short vectors is hard for those cryptosystems to be secure. Among several advantages, lattice-based cryptography is also praised for its apparent resistance to quantum algorithms, unlike the current public-key schemes based on factoring or discrete logarithm.

The main drawback of lattice-based cryptography is its large memory and bandwidth footprints: a lattice is represented by a basis, i.e. an $n \times n$ matrix for a dimension $n$ of several hundreds. For efficiency reasons, it is tempting to rely on structured lattices, such as lattices generated by a circulant matrix. The earliest example of such a cryptosystem is the NTRUEncrypt proposal from Hoffstein et al. [9] from 1998. Algebraically, those lattices can be viewed as ideals or modules over cyclotomic number fields.

Nevertheless, there is no guarantee that hard lattice problems remain hard on particular classes of structured lattices, and indeed, a series of results [4–8] have lead to new quantum algorithms solving certain ideal lattice problems. To the best of our knowledge, the same problems remain hard over arbitrary lattices, even with a quantum computer. More precisely, for certain sub-exponential approximation factors $\alpha$, $\alpha$-SVP on ideal lattices admit a polynomial-time algorithm, as depicted in Figure 1. In this survey, we give an overview of the techniques that have lead to these results.

The first quantum attack on certain ideal lattices of cyclotomic fields was sketched by Campbell, Groves and Shefferd [5], and applies to a few schemes, in particular to one of the first Fully-Homomorphic Encryption schemes [17]. Yet those broken schemes were based on ad-hoc problems that do not benefit from worse-case hardness.

The first step of this attack does not actually solve a lattice problem: it does not provide guarantees about the shortness of

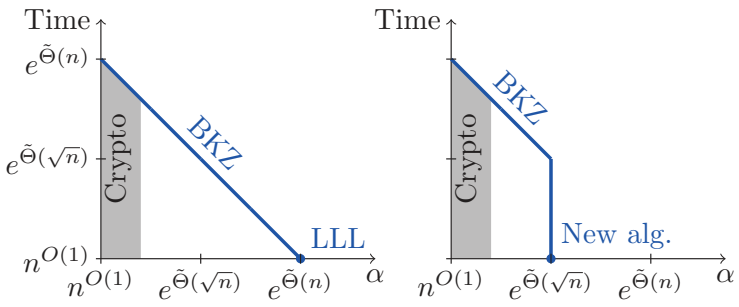



**Figure 1** Best known quantum algorithm for general $\alpha$-SVP (left), and for $\alpha$-SVP in cyclotomic ideal lattices (right).

the solution. Namely, hinted by recent results of Eisenträger, Hallgren, Kitaev and Song [8], it is conjectured in [5] that the Principal Ideal Problem (finding a generator of a given principal ideal) could be solved in quantum polynomial time. This was soon confirmed by the work of Biasse and Song [4]. The second step was also only conjectured to be correct, but could easily be checked in practice. Precisely, taking logarithms, finding a short generator can be phrased as a lattice problem in a fixed lattice (Dirichlet's unit lattice), for which we know a seemingly good basis. A detailed geometric analysis of the cyclotomic units [6] confirmed that conjecture, using tools from analytic number theory.

While this initial attack concerned a particular distribution of principal ideal lattices, the work of [6] also considers what can be done in the worst-case: using similar algorithms, one can always recover a generator longer than the shortest vector by a factor at most $\alpha = \exp(\tilde{O}(\sqrt{n}))$. This constitutes a first worst-case hardness gap between generic lattices and structured ones. The gap was widened in a follow-up result of Cramer, Ducas and Wesolowski [7], showing how to extend these algorithms to non-principal ideals. Naturally, one would look for an ideal $\mathfrak{ab} \subset \mathfrak{a}$ which is a multiple of $\mathfrak{a}$, that is principal, and with a small relative index $\#(\mathfrak{a}/\mathfrak{ab})$. Again, this problem can be translated to a lattice problem in a fixed lattice, namely the lattice underlying Stickelberger's class group annihilation theorem [19].

### Lattices and computational problems

We recall that a lattice is a discrete subgroup of the vector space $\mathbb{R}^n$, equipped with its canonical Euclidean norm denoted $\|\cdot\|$. The minimal distance of a lattice $\Lambda$ is defined by $\lambda_1(\Lambda) = \min_{\mathbf{x} \in \Lambda \setminus \{0\}} \|x\|$.

Our main goal is to solve the following problem in a particular class of lattices.

**Definition.** The Short Vector Problem with approximation factor $\alpha$ ($\alpha$-SVP) is defined as:

- Given a basis $\mathbf{B}$ of a lattice $\Lambda \subset \mathbb{R}^n$,
- Find $\mathbf{v} \in \Lambda \setminus \{0\}$ such that $\|v\| \leq \alpha \cdot \lambda_1(\Lambda)$.

For our purpose, we will also consider two related problems, namely the approximate Close Vector Problem ($\delta$-CVP), and the Bounded Distance Decoding problem ($\delta$-BDD). For convenience, we will define them with respect to an absolute distance $\delta$, rather than an approximation factor $\alpha$.

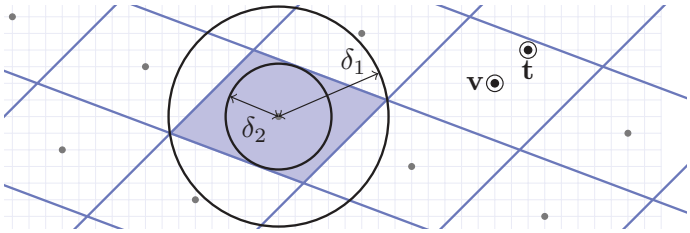



**Figure 2** Rounding with a good basis.

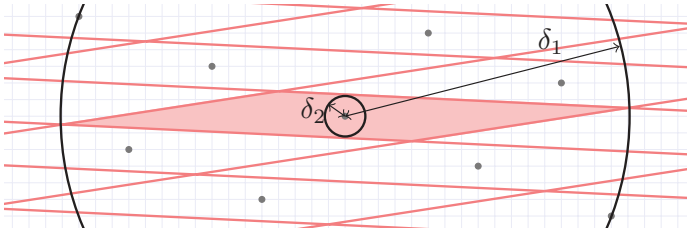



**Figure 3** Rounding with a bad basis.

**Definition.** The Close Vector Problem up to distance $\delta$ ($\delta$-CVP) is defined as:

- Given a basis $\mathbf{B}$ of a lattice $\Lambda \subset \mathbb{R}^n$,
- Given a target $\mathbf{t} \in \mathbb{R}^n$,
- Find $\mathbf{v} \in \Lambda \setminus \{0\}$ such that $\|v - t\| \leq \delta$,

where $\delta$ is large enough so that a solution exists for any target $\mathbf{t}$ (namely $\delta$ is larger than the *covering radius* of $\Lambda$).

**Definition.** The Bounded Distance Decoding Problem up to distance $\delta$ ($\delta$-BDD) is defined as:

- Given a basis $\mathbf{B}$ of a lattice $\Lambda \subset \mathbb{R}^n$,
- Given a target $\mathbf{t} \in \mathbb{R}^n$ at distance at most $\delta$ from $\Lambda$,
- Find $\mathbf{v} \in \Lambda \setminus \{0\}$ such that $\|v - t\| \leq \delta$,

where $\delta$ is small enough so that at most one solution exists for any target $\mathbf{t}$ (namely $\delta < \lambda_1(\Lambda)/2$).

Both problems are somehow dual, in particular $\delta$-CVP gets easier as $\delta$ increases, while $\delta$-BDD gets easier as $\delta$ decreases. A very simple and efficient algorithm for those problems is given by a simple coordinate-wise rounding:

$$\mathbf{v} = \mathbf{B} \cdot \lfloor \mathbf{B}^{-1} \cdot t \rceil.$$

This algorithm induces a parallelepipedic tiling of the space as depicted in Figure 2, where the shape of the tile $P$ is given by the basis $\mathbf{B}$:

$$P = \mathbf{B} \cdot \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]^n = \left\{ \sum \mathbf{x}_i b_i \mid x_i \in \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \right\}.$$

This fast algorithm solves $\delta_1$-CVP (respectively $\delta_2$-BDD) for radii defined by the smallest enclosing sphere of (respectively largest enclosed sphere) of $P$. More precisely:

$$\delta_1 = \tfrac{1}{2} \max \left\| \sum \pm \mathbf{b}_i \right\| \leq \tfrac{1}{2} \sum \left\| \mathbf{b}_i \right\|,$$

$$\delta_2 = \tfrac{1}{2} \min \left\| \mathbf{b}_i^\vee \right\|,$$

where $\mathbf{b}_i^\vee$ denotes the $i$-th vector of the dual basis $(\mathbf{B}^T)^{-1}$. We see that the ability to solve these problems highly depends on the *quality* of the basis $\mathbf{B}$. For comparison, we picture what happens with a *bad* basis of the same lattice in Figure 3: the CVP and BDD radii get worse.

*Lattice-based cryptography*

The gap between what can be done with *good* and *bad* bases is what gives rise to public key cryptography: the bad basis will be used as a public key (allowing to generate noisy lattice points as ciphertexts), while the good basis is kept secret (allowing to solve BDD for decryption). The secret-key owner is able to construct a good basis only because he controls the construction of the lattice.

In this brief overview, we explained why CVP and BDD are useful problems in cryptography, but it turns out that the core problem is SVP. For example, solving SVP a few times allows to construct a basis with small vectors, allowing in turn to solve CVP. And the converse is also true for certain variants of CVP and BDD, as demonstrated by the worst-case to average-case reductions of Ajtai and others [1, 14, 18]. These converse results allow to *prove* that breaking certain cryptosystems is at least as hard as solving $\alpha$-SVP for some approximation factor, typically polynomially large in the dimension $\alpha = n^{O(1)}$.

The hardness of $\alpha$-SVP decrease with growing approximation factor $\alpha$. For small $\alpha = O(1)$, this problem is known to be NP-hard [12], unfortunately it seems impossible to base cryptosystems on $\alpha$-SVP with such a small approximation factor. The best known algorithms for $\alpha$-SVP for polynomial approximation factors $\alpha = n^{O(1)}$ in unstructured lattices require time exponential in $n$. The conjecture that it cannot be done much faster implies that lattice-based cryptosystems are unbreakable in an asymptotic sense. More generally, the best algorithms to solve $\exp(n^c)$-SVP is BKZ [15], a generalization of the Lenstra–Lenstra–Lovàsz algorithm (LLL) [11], and runs in time $\exp(\tilde{\Theta}(n^{1-c}))$, as depicted in Figure 1.

*Cyclotomic ideal lattices*

Consider the $m$-th cyclotomic number field $K = \mathbb{Q}(\zeta)$, where $\zeta$ denotes a formal $m$-th primitive root of unity. Its ring of integer is known to be $\mathcal{O}_K = \mathbb{Z}[\zeta]$. The number field $K$ is equipped with $n = \phi(m)$ complex embeddings, sending $\zeta$ to each of the primitive $m$-th roots of unity in $\mathbb{C}$: $\psi_i : \zeta \mapsto \omega^i$ for each $i \in (\mathbb{Z}/m\mathbb{Z})^\times$, where $\omega = \exp(2\pi/m)$.

An (integral) ideal $\mathfrak{I} \subset \mathcal{O}_K$ is an additive subgroup of $\mathcal{O}_K$ also closed under multiplication by elements of $\mathcal{O}_K$. An ideal may be viewed as a euclidean lattice via the Minkowski embedding:

$$\psi : x \in K \mapsto (\psi_1(x), \ldots, \psi_{m-1}(x)) \in \mathbb{H} = \mathbb{C}^n \simeq \mathbb{R}^{2n}.$$

Each embedding $\psi_i$ is a field morphism; in particular $\psi$ is linear, and multiplication in $K$ corresponds to component-wise multiplication in $\mathbb{H}$.

## Quantum algorithms and HSP

In 1994, Shor [16] formulated a factorization algorithm that would run in polynomial time on a quantum computer. Shor's algorithm exploits the properties of quantum mechanics to efficiently find the period of the function:

$$f : x \in \mathbb{Z} \mapsto a^x \bmod N$$

which reveals the order $r$ of $a \in (\mathbb{Z}/N\mathbb{Z})^\times$. Unless $a^{r/2} = -1 \bmod N$, the quantities $\gcd(a^{r/2}+1, N)$ and $\gcd(a^{r/2}-1, N)$ will provide non-trivial factors of $N$. Very similar ideas also allow to solve the discrete logarithm problem over any cyclic group $G$. In a world with large general-purpose quantum computers, all the public key cryptographic schemes deployed nowadays would become insecure, as they are all based on factoring and discrete logarithm problems.

This motivates the development of cryptosystems based on other mathematical problems, such as lattice-based schemes. This also calls for a better understanding of the power of quantum computers, especially with respect to Shor's idea of period finding. This is generalized as the following problem.

**Definition.** The Hidden Subgroup Problem (HSP) over the Abelian group $G$ is defined as:

- Given an efficient quantum computable function $f : G \to S$, which is exactly $H$-periodic for a subgroup $H \subset G$,
- Find the hidden subgroup $H$,

where $S$ denotes the set of quantum states.

This problem admits an efficient quantum algorithm in many cases. For example, Shor's algorithm is an instance of the HSP algorithm where $G = \mathbb{Z}$, $H = r\mathbb{Z}$, and where the function $f$ is injective modulo the period $r$. In this particular case, $f$ produces a classical result, that is trivially encoded as a quantum state. More generally, quantum algorithms for HSP are now known for larger Abelian groups such as $\mathbb{Z}^n$, and even $\mathbb{R}^n$ [8] with some technical restriction on $f$.

*Quantum encodings of lattices*

As we will see in the next section, many problems in number theory can be phrased as HSP using a function $f$ producing lattices rather than quantum states: $f : G \to \mathcal{L}_V$ where $\mathcal{L} = \{L \mid L \subset V \text{ is a lattice}\}$. To apply the known HSP algorithm (using $R \circ f$) one therefore needs to be able to compute canonical representation for lattices $R : \mathcal{L}_V \to S$, a task that is not always so easy.

For integer lattices $L \subset \mathbb{Z}^d$, such a representation is provided by the Hermite Normal Form, which is computable in classical polynomial time: the representation $R$ is classical. When $L \subset \mathbb{R}^d$ is a lattice of small dimension $d$, one can also compute a normal form, for example using an Hermite–Korkin–Zolotarev (HKZ) reduced basis. Again, this representation of $L$ is purely classical. But as dimension grows, HKZ reduction becomes exponentially hard.

This issue was circumvented by Eisenträger et al. [8], this time by resorting to quantum representations. While we do not know of an efficiently computable normal form for non-integer large dimensional lattices, we do know how to sample according to a canonical distribution over a lattice, with an efficient classical probabilistic algorithm. First, one would start by finding a weakly reduced basis with the LLL algorithm [11], followed by Klein's sampling algorithm [10], which produces a wide discrete Gaussian distribution supported by the lattice $L$.

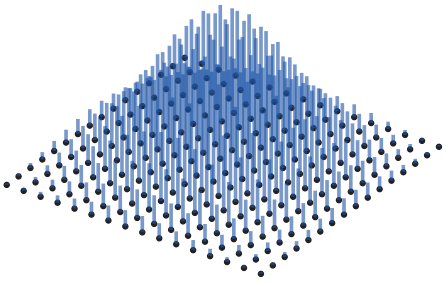


**Figure 4** Discrete Gaussian distribution over a lattice.

This probabilistic algorithm producing a classical distribution can be adapted [8, 14] to a quantum algorithm producing the corresponding quantum state $S$, namely $S$ is a weighted quantum superposition of all lattice points:

$$R : \mathcal{L} \to S,$$
$$R(L) = \sum_{\mathbf{x} \in L} \exp\left(\frac{-\|\mathbf{x}\|^2}{2\sigma^2}\right) \cdot |\, \mathbf{x} \,\rangle .$$

The above quantum superposition is infinite, but can be tail-cut considering the rapid decay of Gaussian distributions. Extra effort is also needed to discretize $\mathbb{R}^d$ and represent each point $\mathbf{x} \in \mathbb{R}^d$ using a finite amount of qubits (see 'straddle encoding' in [8]).

## HSPs in number theory

In this section, we describe algorithms [4, 5, 8] that apply to any number field $K$, given its ring of integers $\mathcal{O}_K$. Let us start by recalling the definitions of ideals of $K$:

**Definition.** An *integral ideal* $\mathfrak{I}$ of $K$ is an additive subgroup $\mathfrak{I} \subset \mathcal{O}_K$ closed by multiplication by elements of $\mathcal{O}_K$:

$$a \in \mathfrak{I}, x \in \mathcal{O}_K \to ax \in \mathfrak{I}.$$

A *fractional ideal* $\mathfrak{f}$ of $K$ is a set of the form $\mathfrak{f} = \frac{1}{z}\mathfrak{I}$ for some non-zero integer $z \in \mathbb{Z}$, and some integral ideal $\mathfrak{I} \subset \mathcal{O}_K$.

An ideal $\mathfrak{f} \subset K$ is said *principal* if it is generated by a single element, i.e. if $\mathfrak{f} = g\mathcal{O}_K = \{gx \mid x \in \mathcal{O}_K\}$ for some $g \in K$; such a $g$ is called a *generator* of $\mathfrak{f}$.

Principal ideals have multiple generators, more precisely, $g$ and $g' \in K$ generate the same ideal if and only if $u = g/g'$ is a unit of $O_K$. The multiplicative group of units is denoted $O_K^\times = \{u \in O_K \mid u^{-1} \in O_K\}$.

Recall that ideals can be multiplied:

$$\mathfrak{a} \cdot \mathfrak{b} = \left\{\sum a_i b_i \mid a_i \in \mathfrak{a}, b_i \in \mathfrak{b}\right\},$$

which makes the set $\mathcal{F}_K$ of fractional ideals an Abelian group. The set of $\mathcal{P}_K \subset \mathcal{F}_K$ of principal ideals form a subgroup of $\mathcal{F}_K$.

With those definitions, we can already consider two important computational problems in number theory.

**Definition.** The Unit Group Problem (UGP) consists in:

- Given a number field $K$ and its ring of integers $O_K$,
- Find a finite set of units $u_1, \ldots, u_d \in O_K^\times$ that generate $O_K^\times$.

**Definition.** The Principal Ideal Problem (PIP) consists in:

- Given a number field $K$ and its ring of integers $O_K$,
- Given a principal ideal $\mathfrak{I} \subset K$,
- Find a generator $g$ of $\mathfrak{I}$.

Both problems can be viewed as particular cases of the more general problem of computing the group of $S$-units for a well chosen set $S$ of prime ideals, as done in the paper of Biasse and Song [4].

We start by phrasing the Unit Group Problem as a multiplicative Hidden Subgroup Problem. Note that $u \in K$ is a unit of $O_K$ if and only if $uO_K = O_K$, and more generally $gO_K = g'O_K$ if and only if $u = g/g'$ is a unit of $O_K$. This means that the function:

$$f_{\mathsf{UGP}} : K^\times \to \mathcal{P}_K$$
$$x \mapsto x \cdot O_K$$

is (multiplicatively) $O_K^\times$-periodic, and one easily checks that it is injective modulo $O_K^\times$ as well. The images of such a function are ideals, and can therefore be viewed as lattices. Using the strategy described in the previous section, one can efficiently construct a canonical quantum representation of these lattices.

A lot of technicalities remain to implement this approach. In particular, the domain of the function $f$ described above is the multiplicative group $K^\times$, while one wishes to apply the known HSP algorithm over the vector space $\mathbb{R}^n$. This issue is essentially dealt with by resorting to the well known logarithmic embeddings:

$$\mathrm{Log} : K^\times \to \mathbb{R}^n$$
$$x \mapsto (\log \mid \psi_1(x) \mid, \ldots, \log \mid \psi_{m-1}(x) \mid).$$

In more details, define $\mathrm{Exp} : \mathbb{C}^n \to \mathbb{H}$ by coordinate-wise application of $\exp : \mathbb{C} \to \mathbb{C}^\times$. Up to an appropriate quotient on the domain, one can set

$$f_{\mathsf{UGP}} : \mathbb{C}^n \to \mathcal{L}$$
$$\mathbf{y} \mapsto \mathrm{Exp}(\mathbf{y}) \odot \psi(O_K)$$

and recovers $\mathrm{Log} O_K^\times$ from the period of $f_{\mathsf{UGP}}$. Geometrically, $f_{\mathsf{UGP}}(\mathbf{y})$ is a deformation the of the lattice $\psi(O_K)$, where the $i$-th coordinate axis of $\mathbb{H} = \mathbb{C}^n$ has been stretched by the complex factor $\exp(y_i)$; this deformation leaves $\psi(O_K)$ invariant precisely when $\mathrm{Exp}(\mathbf{y})$ equals to $\psi(u)$ for some unit $u \in O_K^\times$.

While this strategy seems simple, proving its correctness requires an in-depth analysis of the metric properties of $R \circ f_{\mathsf{UGP}}$: Lipschitz continuity and some strong form of injectivity. We refer to the original article for more details [8].

Finally, we sketch a (over-)simplified strategy to generalize the above to the Principal Ideal Problem. Given a principal ideal $\mathfrak{I}$, one extends the function $f$ to:

$$f_{\mathsf{PIP}(\mathfrak{I})} : (\mathbf{y}, i) \mapsto \mathrm{Exp}(\mathbf{y}) \odot \psi(\mathfrak{I}^i).$$

The periods of this function contains the extension of the previous, namely, it is $(\mathrm{Log} O_K^\times, 0)$ periodic; but it is also $(\mathrm{Log} g, -1)$ periodic for any generator $g$ of $\mathfrak{I}$, as $f(\mathrm{Log} g, -1) = g \cdot \mathfrak{I}^{-1} = O_K$. With this function $f_{\mathsf{PIP}(\mathfrak{I})}$, the quantum algorithm for Hidden Subgroup Problem of [8] allows not only to recover the unit group, but also a generator of the principal ideal $\mathfrak{I}$. Again, much care is required to ensure that this strategy will indeed work, see [4].

### Short generators of principal ideals

So far, we have been concerned with problems that were purely of number theoretic nature, in the sense that the solutions to UGP and PIP have no guarantees in term of size. In this section we explain how one can recover a short generator of a principal ideal from an arbitrary generator in the particular case of cyclotomic number fields.

**Definition.** The Short Generator Problem $\alpha$-SGP consists in:

- Given an element $h \in K^\times$, generating an ideal $\mathfrak{I} = hO_K$
- Find a generator $g \in K^\times$ of $\mathfrak{I}$ of small Euclidean length: $\| g \| \leq \alpha \cdot \lambda_1(\mathfrak{I})$.

Remembering that $h$ and $g$ generate the same ideal if and only if $g = hu$ for some unit $u \in O_K$, the idea consists in rephrasing this problem as a close vector problem using the logarithmic embedding: indeed, we have that $\mathrm{Log} g = \mathrm{Log} h + \mathrm{Log} u$ must belong to the lattice coset $\log h + \mathrm{Log} O_K^\times$. Furthermore, up to appropriate rescaling, it can be proved that the length of $g$ is related to the length of its logarithmic embedding $\mathrm{Log} g$. Minimizing $g$ can therefore be rephrased as finding a unit $u$ such that $\mathrm{Log} u$ is close to $-\mathrm{Log} h$, in other words solving a Close Vector Problem over Dirichlet's logarithmic unit lattice $\mathrm{Log} O_K^\times$.

Moreover, in certain cryptosystems, we have additional constraints on the ideal $\mathfrak{I}$, ensuring that a unusually short generator $g$ exists (which is used as the secret key). This suggested that the Close Vector Problem may actually become a BDD problem [5]. And indeed, experiments confirmed that this BDD is easily solved in practice. Running such experiments requires knowing the group of units $O_K^\times$. Fortunately, in the case of cyclotomic number fields, there are some well known units — that very often generate the whole group $O_K^\times$ — namely, the cyclotomic units:

$$\{\zeta\} \cup \left\{ u_i = \frac{1-\zeta^i}{1-\zeta} \mid i \in (\mathbb{Z}/m\mathbb{Z})^\times \right\}.$$

*Geometric analysis of the cyclotomic units*

The fact that the attack works in practice suggests that the matrix $\mathbf{U} = (\mathrm{Log} u_i)_{i \in (\mathbb{Z}/m\mathbb{Z})^\times}$ forms a good basis for BDD. Yet it is not so straightforward to prove it: recalling the first section, one needs to show that the *dual* vectors $\mathbf{u}_i^\vee$ are short. To proceed with the analysis, Cramer et al. [6] instead considered the related matrix $\mathbf{M} = (\mathrm{Log}(1 - \zeta^{i^{-1}}))_i$, where

$$\mathbf{M}_{i,j} = \log \mid \psi_j(1 - \zeta^{i^{-1}}) \mid = \log \mid 1 - \omega^{ji^{-1}} \mid$$

for indices $i, j$ running over the group $G = (\mathbb{Z}/m\mathbb{Z})^\times$. Since this matrix is $G$-circulant, it can therefore be explicitly diagonalized, and a lower-bound on the diagonal coefficients will provide an upper-bound

on the length of the dual vectors $\|m_i^\vee\|$. The eigenvalue $\lambda_\chi$ associated to the character $\chi : G \to \mathbb{C}$ of $G$ is given by:

$$\lambda_\chi = \sum_{i \in G} \chi(i) \log |1 - \omega^i| .$$

Using classical techniques of analytic number theory (in this case, the Taylor series of $\log$, and separation of Gauss sums), the above formula can be massaged to

$$\lambda_\chi = \sqrt{f_\chi} \cdot L(\chi, 1),$$

where $f_\chi$ is the conductor of $\chi$, and $L$ denotes Dirichlet's $L$-series. Lower bounds on $L$-series at $1$ have a very long history and play a crucial role in the study of the distribution of prime numbers. For example, Landau proved that $L(\chi,1) \geq 1/O(\log f_\chi)$ for non-quadratic characters. With more effort, Cramer et al. [6] conclude on an upper bound on $\|m_i^\vee\|$ and then on $\|u_i^\vee\|$.

*Extension to the worst-case*

In addition, the article [6] also covers the performance of this strategy in the worst-case, that is when there is no guarantee of existence of a particularly short generator $g$, by quantifying how good is the basis $\mathbf{U}$ to solve CVP. This analysis is somewhat easier as it concerns the length of the primal vectors, whose length can be bounded using the following finite integral:

$$\int_0^1 (\log |1 - \exp(2\imath\pi x)|)^2 dx < \infty.$$

Further efforts lead to a classical probabilistic polynomial time algorithm that solves $\alpha$-SGP for sub-exponential $\alpha = \exp(\tilde{O}(\sqrt{n}))$. Combined with the previous algorithm of Biasse and Song for PIP [4], this provides a solution to $\alpha$-SVP in quantum polynomial time over principal ideal lattices in the worst case, outperforming the best known generic algorithms LLL and BKZ.

Finally, it is also shown in [6] that this result is roughly optimal: there exist many ideals for which the shortest generator $g$ is much larger than the shortest vector, by a factor $\exp(\tilde{O}(\sqrt{n}))$. This is established by a lower bound on the covering radius of the lattice $\mathrm{Log}\,\mathcal{O}_K^\times$. Lowering the SVP approximation factor reachable in polynomial time will necessarily require algorithms that are not limited to finding generators.

## Short vectors in arbitrary ideals

Considering that the previous result only applies to principal ideals, which form a very small fraction of all ideals, it is unclear whether this previous result has an impact on lattice-based cryptography beyond the few aforementioned atypical cryptosystems [5, 17]. Indeed, most schemes are instead based on worst-case problems, and are not affected by the presence of a small fraction of weak ideal lattices.

The obstacle ahead to attack non-principal ideals is the class group, the quotient of all ideals by the principal ones:

$$\mathrm{Cl}_K = \mathcal{F}_K / \mathcal{P}_K.$$

The class of an ideal $\mathfrak{a} \subset K$ in this quotient is denoted $[\mathfrak{a}] \in \mathrm{Cl}_K$, and the neutral element is $[\mathcal{O}_K]$ (the class of principal ideals). This quotient is always finite, and in the case of cyclotomic number fields, it has size about $\#\mathrm{Cl}_K = 2^{\Theta(n \log n)}$: the fraction of principal ideals is super-exponentially small.

To generalize the previous result to non-principal ideals, the natural strategy consists in trying to find sub-ideals that are principal. More formally, Cramer, Ducas and Wesolowski [7] define the following problem:

**Definition.** The Close Principal Multiple problem with approximation factor $F$ ($F$-CPM) is defined as:

- Given an ideal $\mathfrak{a} \subset K$,
- Find an *integral* ideal $\mathfrak{b}$ such that $\mathfrak{ab}$ is principal (i.e. $[\mathfrak{ab}] = [\mathcal{O}_K]$) and such that $\mathfrak{b}$ is a dense ideal: $N\mathfrak{b} \leq F$,

where $N\mathfrak{b} := \#(\mathcal{O}_K/\mathfrak{b})$ denotes the algebraic norm (i.e. the sparsity) of the ideal $\mathfrak{b}$.

Combining algorithm for $F$-CPM with the previous algorithms provides solution to $\alpha$-SVP over non-principal ideals within approximation factor:

$$\alpha = F^{1/n} \cdot \exp(\tilde{O}(\sqrt{n})) .$$

In this section, we will sketch how this CPM problem was solved for $F = \exp(\tilde{O}(n^{3/2}))$, which leads to a similar SVP approximation factor $\alpha = \exp(\tilde{O}(\sqrt{n}))$ as in the principal case. Consider a factor basis of prime ideals $\mathfrak{B} = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_d\}$ that are dense ($N\mathfrak{p}_i \leq n^{O(1)}$) and the morphism:

$$\phi : \mathbb{Z}^d \to \mathrm{Cl}_K, \quad \phi(\mathbf{e}) = \left[\prod \mathfrak{p}_i^{e_i}\right].$$

Assuming that $\phi$ is surjective, we can rephrase CPM as a CVP problem in the lattice of class relations $\Lambda = \ker \phi$. Indeed, consider $\mathbf{v} \in \phi^{-1}([\mathfrak{a}])$, and $\mathbf{c} \in \Lambda$ such that $\mathbf{c} - \mathbf{v}$ is small, and all positive. Then $\mathfrak{b} = \prod \mathfrak{p}_i^{c_i - v_i}$ will be an appropriate solution to CPM, up to a factor $F = \prod (N\mathfrak{p}_i)^{v_i - c_i} = n^{O(\|\mathbf{v} - \mathbf{c}\|_1)}$ where $\|x\|_1 = \sum |x_i|$ denotes the $\ell_1$-norm of $\mathbf{x}$.

In general, there is no reason why CVP should be easy in the lattice $\Lambda$, which is not even explicitly known. Yet, by choosing an appropriate factor basis, namely, a basis composed of all the Galois conjugates of a single ideal, one can on the contrary get a very explicit description of the lattice $\Lambda$ thanks to the classical theorem of Stickelberger [19]. It turns out that one may easily explicitly construct a short basis of $\Lambda$. Again, this overview is highly simplified, and hides several technicalities, see [7].

## Conclusion and open questions

There remain serious obstacles for this approach to attack ideal lattice-based cryptosystems. First the approximation factor $\alpha = \exp(\tilde{O}(\sqrt{n}))$ is too large to affect cryptographic schemes. Second, these algorithms are limited to ideal lattices (i.e. module lattices of rank $1$), while most cryptosystems in fact use module lattices of rank $2$ or more.

Nevertheless, these recent works questioned our understanding of the hardness of lattice problems when using special classes of lattices. We now know of a specialized algorithm for relevant classes of structured lattices that outperforms generic ones (see Figure 1). Alternatives to cyclotomics ideal lattices are already being studied [2, 3, 13] from various point of view: complexity theory, concrete cryptographic design and cryptanalysis.

There are many cases where the volume of the log-unit lattice (the regulator) and the lattice of class relation (the class number) is well understood. We have seen here that in the case of cyclotomic number field, much more can be said about those lattices (known good bases, covering radius,...). Generalizing this geometric analysis to other number fields seems to be an interesting mathematical problem, with potential cryptanalytic implications.

**Acknowledgments**

The author wishes to express his gratitude to Koen de Boer, Fang Song and Benjamin Wesolowski for their precious comments on drafts of this article.

## References


1 M. Ajtai, Generating hard instances of the short basis problem, *ICALP 1999*.

2 J. Bauch, D. J. Bernstein, H. de Valence, T. Lange and C. van Vredendaal, Short generators without quantum computers: The case of multiquadratics, *Eurocrypt 2017*.

3 D. J. Bernstein, C. Chuengsatiansup, T. Lange and C. van Vredendaal, NTRU Prime, Preprint 2016.

4 J.-F. Biasse and F. Song, A polynomial time quantum algorithm for computing class groups and solving the principal ideal problem in arbitrary degree number fields, *SODA 2016*.

5 Peter Campbell, Michael Groves and Dan Shepherd, Soliloquy: A cautionary tale, *ETSI 2nd Quantum-Safe Crypto Workshop*, 2014.

6 R. Cramer, L. Ducas, C. Peikert and O. Regev, Recovering short generators of principal ideals in cyclotomic rings, *Eurocrypt 2016*.

7 R. Cramer, L. Ducas and B. Wesolowski, Short Stickelberger class relations and application to ideal-SVP, *Eurocrypt 2017*.

8 K. Eisenträger, S. Hallgren, A. Kitaev and F. Song, A quantum algorithm for computing the unit group of an arbitrary degree number field, *STOC 2014*.

9 J. Hoffstein, J. Pipher and J. H. Silverman, NTRUSIGN: Digital signatures using the NTRU lattice, *CT-RSA 2003*.

10 P. Klein, Finding the closest lattice vector when it's unusually close, *SODA 2000*.

11 A. K. Lenstra, H. W. Lenstra and L. Lovàsz, Factoring polynomials with rational coefficients, *Mathematische Annalen* 261(4) (1982).

12 M. Daniele, The shortest vector in a lattice is hard to approximate to within some constant, *SIAM Journal on Computing* 30(6) (2001).

13 C. Peikert, O. Regev and N. Stephens-Davidowitz, Pseudorandomness of Ring-LWE for any ring and modulus, *STOC 2017*.

14 O. Regev, On lattices, learning with errors, random linear codes, and cryptography, *STOC 2005*.

15 C.-P. Schnorr, A hierarchy of polynomial time lattice basis reduction algorithms, *Theoretical Computer Science* 53(2) (1987).

16 P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, *FOCS 1994*.

17 N. P. Smart and F. Vercauteren, Fully homomorphic encryption with relatively small key and ciphertext sizes, *PKC 2010*.

18 D. Stehlé, R. Steinfeld, K. Tanaka and K. Xagawa, Efficient public key encryption based on ideal lattices, *Asiacrypt 2009*.

19 L. C. Washington, *Introduction to Cyclotomic Fields*, Springer, 1997.

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**

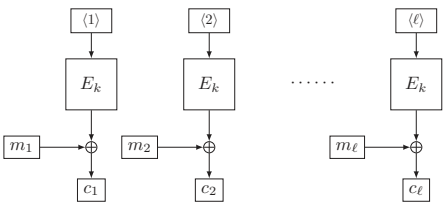
Het leeuwendeel van hedendaagse versleuteling vindt plaats door middel van 'blokcijfers'. Een blokcijfer $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ is een familie van permutaties over verzameling $\mathcal{M}$ geïndexeerd met een sleutel uit een sleutelverzameling $\mathcal{K}$. Dat wil zeggen, 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)

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

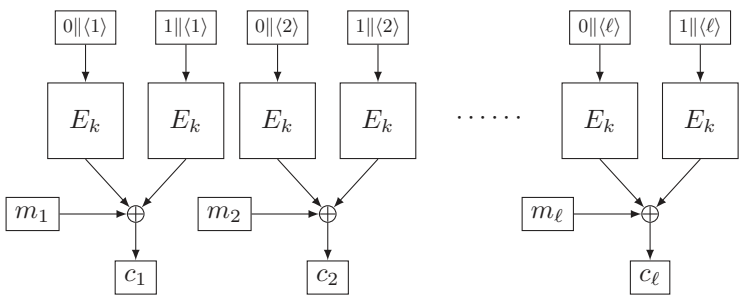
waarbij $\langle i \rangle$ in dit geval de encoding van $i$ als een bitstring van lengte $127$ noteert. Deze constructie is veilig zolang het aantal evaluaties onder $2^n/67$ blijft. Dit is een significante verbetering ten opzichte van de birthday bound $2^{n/2}$ voor (1). Feitelijk hebben we ervoor 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

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

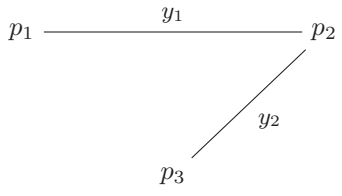
waarbij de indexering van de onbekenden impliciet gebeurt middels een surjectieve afbeelding

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

(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 $\varphi$.

*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).



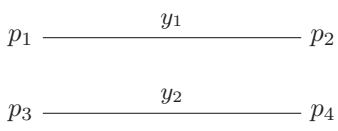

**Figuur 3** Een systeem van twee vergelijkingen over drie onbekenden.

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.

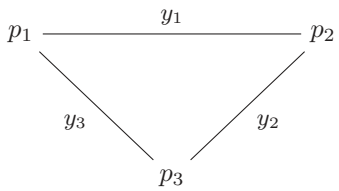
Net zoals in het vorige voorbeeld heeft dit systeem geen oplossing als $y_1 = 0$ of $y_2 = 0$. Als $y_1,y_2 \neq 0$ kunnen we op naïeve wijze het aantal oplossingen gaan tellen: we hebben $2^n$ mogelijke waarden voor $p_1$, en zodra $p_1$ vastligt, dan ligt $p_2 = y_1 \oplus p_1$ ook vast. De twee vergelijkingen 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 $0 = y_1 \oplus y_2 \oplus y_3$.

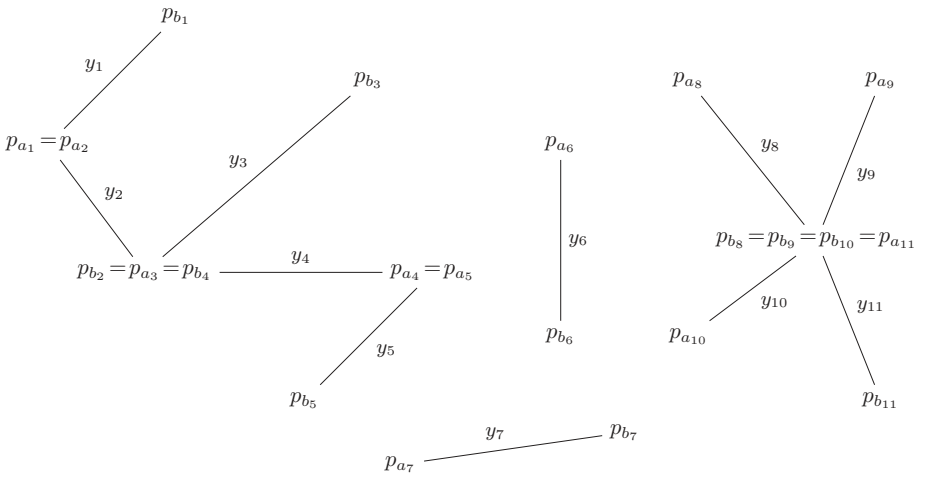
*Algemene stelling*

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 $\pi$. 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 $F^{\pi}$ wordt 'goed' geacht als haar uitvoeren ongeveer dezelfde distributie hebben als een volledig willekeurige functie $\rho$: als we een aanvaller toegang geven tot *ofwel* $F^{\pi}$ *ofwel* $\rho$, 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 $F^{\pi}$ of $\rho$ 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 $\tau$ 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-coëfficiënttechniek toont nu het volgende aan (de afleiding is basiskansrekening, zie [4,12]): als er $\delta,\varepsilon$ 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 $\Delta(F^{\pi};\rho) \leq \delta + \varepsilon$. $F^{\pi}$ 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 $\rho$. Voor de teller in vergelijking (8) kunnen we opmerken dat

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

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

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

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 $\pi$, namelijk $\pi(f_1(x_i)) =: p_{a_i}$ en $\pi(f_2(x_i)) =: p_{b_i}$, en de gehele lijst $\tau$ 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 $\varepsilon$ 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 $\pi$ 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 $q\ll 2^n$.

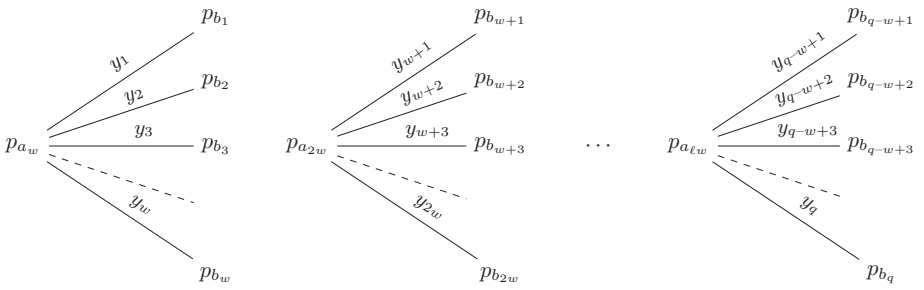
*Constructie in CENC in vergelijking (4)*

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).

David Elkouss
*QuTech*
*TU Delft*
*d.elkousscoronas@tudelft.nl*

# Key distribution and the CHSH game

**David Elkouss is assistant professor at QuTech at TU Delft, where he works, among others, on quantum key distribution. In this article, Elkouss reviews the relation between the violation of a Bell inequality and device independent quantum key distribution, by departing from the so-called CHSH inequality and linking it with a protocol for key distribution.**

Quantum key distribution (QKD) [3, 8] allows distributing information-theoretically secure keys to two distant parties. The only required assumptions are the validity of quantum theory and a characterization of the devices used for implementing the key distribution protocol. Hence, whenever the characterization is not precise enough, it can be exploited via side-channel attacks that do not break the distribution protocol but profit from the device imperfections. An alternative paradigm is device-independent quantum key distribution (diQKD) [8]. In diQKD the required assumptions are milder, i.e. it is enough to assume quantum theory to be correct. However, in contrast with QKD, the implementation of diQKD remains elusive. Recently, the first loophole-free Bell experiments have been implemented. These implementations promise to bring diQKD close to reality.

Here, we begin from basic principles and review the relation between the violation of a Bell inequality and the feasibility of diQKD. The structure of the article is as follows: We first describe the setting of the CHSH inequality as a non-local game [5], and how the winning probability achievable by Alice and Bob is limited if they play with a local strategy. Then we introduce the formalism of quantum information and show that with quantum resources Alice and Bob can achieve a larger winning probability. In a real experiment, one necessarily only encounters a finite number of samples, hence the winning probability can never be estimated with certainty. Thereafter, we argue how it is still possible to make a rigorous statement about the type of strategy played by Alice and Bob. Finally, we link winning probabilities beyond the local strategy limit with the distribution of secret keys.

## The CHSH game and its classical value

In this section, we describe a non-local game and derive the maximum winning probability for classical or local players. This game was first introduced by John Clauser, Michael Horne, Abner Shimony, and Richard Holt in [7] and thereafter has been known as the CHSH game.

First, let us introduce the scenario. The CHSH game is an example of a bipartite non-local game. That is, a game played by two parties or players. We call them Alice and Bob. Both are located in two separated locations such that during each round of the game they cannot exchange information. This locality restriction might be enforced because they are spatially separated or because we assume that we have well characterized their laboratories and can conclude that they do not leak during the game execution. For a review on non-local games see [5].

Alice and Bob have a binary-input binary-output box. It receives respectively $x, y \in \{0,1\}$ and outputs $a, b \in \{0,1\}$, see Figure 1. We can imagine a box with two buttons and whenever the player pushes a button the box outputs a bit. The rules of the CHSH game are as follows. The inputs $x$, $y$ are chosen uniformly at random and Alice and Bob are free to choose any strategy for their boxes to output the values $a$, $b$. They win the game whenever $x \cdot y = a \oplus b$, where $a \oplus b$ indicates $a + b \bmod 2$, and they lose otherwise.

Let us first focus on local or classical strategies. We call a strategy local or classical if there exists some shared source of randomness (possibly trivial) between the players such that

$$\Pr(a,b \mid x,y,\lambda) = \Pr(a \mid x,\lambda)\Pr(b \mid y,\lambda)$$

where $\Pr(a,b \mid x,y,\lambda)$ is the probability that the boxes output $a$, $b$ given $x$, $y$ and $\lambda$, the outcome of the shared randomness.

Before we analyze the maximum winning probability, we may wonder what is the role of randomness. A strategy is de-

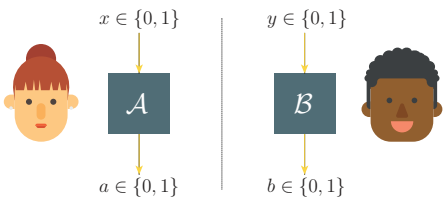



**Figure 1** In the CHSH game, two space-like separated parties, that we call Alice and Bob receive two binary inputs $x, y \in \{0,1\}$ and produce two binary outputs $a, b \in \{0,1\}$. Alice and Bob win the game if $x \cdot y = a \oplus b$, where $a \oplus b = a + b \bmod 2$.

terministic if $\Pr(a,b\,|\,x,y)$ takes only values zero or one. Any non-deterministic local strategy can be implemented with shared randomness in such a way that $\Pr(a,b\,|\,x,y,\lambda)$ takes only values zero or one. Hence, the winning probability can be written in the form

$$p_{\text{win}} = \sum_{\lambda} \Pr(\lambda) \sum_{x,y} \tfrac{1}{4} \sum_{a,b:x\cdot y = a\oplus b} \Pr(a,b\,|\,x,y,\lambda)$$

where $\Pr(a,b\,|\,x,y,\lambda)$ is deterministic. Can non-deterministic strategies help Alice and Bob achieve better winning probabilities than deterministic strategies? It is easy to see that for any probabilistic strategy we can upper bound the winning probability by:

$$p_{\text{win}} \leq \max_{\lambda} \sum_{x,y} \tfrac{1}{4} \sum_{a,b:x\cdot y = a\oplus b} \Pr(a,b\,|\,x,y,\lambda)$$

where the right-hand side is the winning probability of a deterministic strategy.

Now we can investigate what is the maximum value achievable with a local strategy. From the argument above, we only need to consider deterministic strategies. Alice and Bob have only four different choices for choosing their outputs. Either they fix the output to zero or one, either they output the input or they flip it. Hence, combining Alice and Bob choices there are only sixteen different deterministic strategies. For instance, let us suppose that Alice chooses $a = x$ and Bob chooses $b = 0$. Then, they win whenever the input is $(x,y) = (0,0)$, $(x,y) = (0,1)$ and $(x,y) = (1,1)$ but they lose when $(x,y) = (1,0)$. Since all inputs are chosen with equal probability, they win with probability $p_{\text{win}} = 0.75$. Can they do any better? It is a matter of checking the remaining fifteen combinations of strategies, but it turns out that it is not possible. The best winning probability that can be achieved by a local strategy is three over four. We call this best probability the classical value of the game and we denote it by $p^*_{\text{win}}$.

**The quantum value of CHSH**

Let us do a detour to introduce some rudiments of quantum information. This is necessary in order to talk about the quantum value of the CHSH game. We first see how to represent quantum states and then we discuss a class of measurements known as projective measurements. We point the interested reader to [11] for an in-depth introduction to quantum information.

A qubit represents a two-level quantum mechanical system, for instance the polarization of a photon. We can describe the state of the two-level system by a vector in $\mathbb{C}^2$, that we can write as

$$\begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} = \alpha_0\begin{pmatrix}1\\0\end{pmatrix} + \alpha_1\begin{pmatrix}0\\1\end{pmatrix}$$

with $\alpha_0, \alpha_1 \in \mathbb{C}$ and such that $|\,\alpha_0\,|^2 + |\,\alpha_1\,|^2 = 1$. The reasons for restricting to unit vectors will become clear later.

The so-called ket notation introduced by Dirac allows for a more compact description of quantum systems. We can rewrite the state of a qubit $|\,\phi\,\rangle = \alpha_0\,|\,0\,\rangle + \alpha_1\,|\,1\,\rangle$ where

$$|\,0\,\rangle = \begin{pmatrix}1\\0\end{pmatrix} \quad \text{and} \quad |\,1\,\rangle = \begin{pmatrix}0\\1\end{pmatrix}.$$

Of course, there is nothing particular about the canonical or computational basis. For any orthonormal basis given by $v$ and $v^\perp$, we can find the complex coefficients $\beta_v, \beta_{v^\perp}$ such that $|\,\phi\,\rangle = \beta_v\,|\,v\,\rangle + \beta_{v^\perp}\,|\,v^\perp\,\rangle$.

A two-qubit system can be described by a unit vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$. We can write its state using ket notation in an analogous way to one qubit systems. Let

$$|\,0\,\rangle_A\,|\,0\,\rangle_B = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad |\,0\,\rangle_A\,|\,1\,\rangle_B = \begin{pmatrix}0\\1\\0\\0\end{pmatrix},$$

$$|\,1\,\rangle_A\,|\,0\,\rangle_B = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad |\,1\,\rangle_A\,|\,1\,\rangle_B = \begin{pmatrix}0\\0\\0\\1\end{pmatrix},$$

be the ket representation of each vector of the computational basis. Then an arbitrary state is of the form

$$|\,\phi\,\rangle_{AB} = \sum_{i,j\in\{0,1\}} \alpha_{ij}\,|\,i\,\rangle_A\,|\,j\,\rangle_B.$$

Now, let us assume that we have two one-qubit systems $A$ and $B$ with state $|\,\phi\,\rangle_A = \alpha_0\,|\,0\,\rangle + \alpha_1\,|\,1\,\rangle$ and $|\,\psi\,\rangle_B = \beta_0\,|\,0\,\rangle + \beta_1\,|\,1\,\rangle$. We can describe the joint composite system, $|\,\omega\,\rangle_{AB}$, by taking the tensor product of $|\,\phi\,\rangle_A$ and $|\,\psi\,\rangle_B$:

$$\begin{aligned}|\,\omega\,\rangle_{AB} &= |\,\phi\,\rangle_A \otimes |\,\psi\,\rangle_B \\ &= \begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} \otimes \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} \\ &= \begin{pmatrix}\alpha_0\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_0\\ \alpha_1\beta_1\end{pmatrix} \\ &= \sum_{i,j\in\{0,1\}} \alpha_i\beta_j\,|\,i\,\rangle_A\,|\,j\,\rangle_B.\end{aligned}$$

The natural question is if all two qubit states can be written in this form. The answer is no, but whenever this is the case and we can write the state of a two-qubit state as the tensor product of two one-qubit states we say that the state is a product state. Otherwise, we say that the state is *entangled*. A very important consequence of this definition is that it is not possible to describe the individual qubit systems of an entangled state with a ket. The individual qubit systems of an entangled state do not have a definite state. Alternatively, if a qubit system has a definite state (can be described by a ket) it can not be part of an entangled state. We call a system with a definite state a *pure* state.

The next step in our brief introduction to quantum information is measurement. We only consider a subclass of measurements called rank-one projective measurements. These measurements are specified by a measurement basis and can be understood as reading in which element of the basis the state is. The outcome of such a measurement is probabilistic. Let an arbitrary $d$-level quantum system be given by

$$|\,\phi\,\rangle = \sum_{i=0}^{d-1} \alpha_i\,|\,u\,\rangle_i$$

and let us suppose that we measure this state in the basis given by $\{|\,u\,\rangle_i\}_{i=0}^{d-1}$. Then, the probability that we obtain measurement outcome $i$ is $|\,\alpha_i\,|^2$. The restriction to unit vectors guarantees that the sum over all possible outcomes $\sum_i |\,\alpha_i\,|^2 = 1$ as expected.

Now let us show how entanglement can help Alice and Bob obtain a winning probability in the CHSH game larger than the classical value. For this, let us assume that Alice and Bob share the following quantum state, known as the maximally entangled state, before the beginning of the game:

$$|\,\phi\,\rangle_{AB} = \frac{|\,0\,\rangle_A\,|\,0\,\rangle_B + |\,1\,\rangle_A\,|\,1\,\rangle_B}{\sqrt{2}} \qquad (1)$$

Let us define the following measurement bases:

$$\begin{aligned}A_0 &= \{|\,0\,\rangle, |\,1\,\rangle\}, \\ A_1 &= \left\{\tfrac{|\,0\,\rangle + |\,1\,\rangle}{\sqrt{2}}, \tfrac{|\,0\,\rangle - |\,1\,\rangle}{\sqrt{2}}\right\} \\ B_0 &= \{\cos(\pi/8)\,|\,0\,\rangle + \sin(\pi/8)\,|\,1\,\rangle, \\ &\qquad -\sin(\pi/8)\,|\,0\,\rangle + \cos(\pi/8)\,|\,1\,\rangle\} \text{ and} \\ B_1 &= \{\cos(\pi/8)\,|\,0\,\rangle - \sin(\pi/8)\,|\,1\,\rangle, \\ &\qquad \sin(\pi/8)\,|\,0\,\rangle + \cos(\pi/8)\,|\,1\,\rangle\}.\end{aligned}$$

The strategy of Alice and Bob is, given the

inputs $x$ and $y$, measure in the bases $A_x$ and $B_y$. Let us compute the winning probability with this strategy.

$$\begin{aligned}&\Pr(\text{win} \mid x=0,y=0)\\&\quad=\Pr(a=0,b=0 \mid x=0,y=0)\\&\qquad+\Pr(a=1,b=1 \mid x=0,y=0)\\&\quad=\cos^2\frac{\pi}{8}.\end{aligned}$$

We can also easily verify that

$$\begin{aligned}&\Pr(\text{win} \mid x=0,y=1)\\&\quad=\Pr(\text{win} \mid x=0,y=1)\\&\quad=\Pr(\text{win} \mid x=1,y=0)\\&\quad=\Pr(\text{win} \mid x=1,y=1)\\&\quad=\cos^2\frac{\pi}{8}\approx 0.85.\end{aligned}$$

We do not prove it here, but it turns out that the strategy that we just analyzed is optimal. That is, it is not possible, within the formalism of quantum mechanics, to achieve a larger winning probability [6]. Hence, in particular, there exists no quantum strategy that allows Alice and Bob to win the game with probability one. Moreover, the maximally entangled state is unique in achieving it, this will be crucial for the purpose of key distillation. More precisely, if Alice and Bob achieve the optimal winning probability then they can conclude that up to a local unitary transformation they share a maximally entangled state. This result can be made robust. That is, if the winning probability of CHSH is close to the optimal value, then Alice and Bob can conclude that their state is close to the optimal state [10]. This closeness is quantified according to some distance measure between quantum states that goes beyond the scope of this introduction [11].

### Non-locality with finite data

In the previous sections, we have derived the optimal winning probabilities under the assumption of local or quantum strategies. In particular, we have seen that there are quantum strategies that yield strictly larger winning probabilities than local strategies. This is a fundamental theoretical statement, but by itself, it is not very useful. In a real experiment, one observes a finite sequence of inputs and outputs. With this finite number of runs, it is possible to compute the frequency of wins, but the true probability is beyond reach. Is it then possible to decide that an experimental implementation is governed by a non-local strategy?

The solution is to perform a hypothesis test where the null hypothesis is that the experiment is run by a local strategy. Then the CHSH game is played some predefined number of times $n$ and the number of wins $c$ is recorded. With this value, it is possible to compute the probability of obtaining the same or a larger number of wins under the assumption that the null hypothesis is true. If this probability is below some number decided in advance, the null hypothesis is discarded. Let us now make explicit the computation of this probability. Let $C_i$ be a random variable that takes value one if Alice and Bob win the $i$-th game and zero otherwise. Then conditioned to the occurrence of any sequence of outcomes in the first $i-1$ rounds of CHSH the winning probability remains bounded by the classical value [4]:

$$\Pr(C_i=1 \mid C_1 \ldots C_{i-1}) \leq p^*_{\text{win}}.$$

The intuition behind this statement is that the previous sequence of outcomes can be regarded as an instance of shared randomness. Building on top of this it is possible to show that for all local strategies

$$\Pr(C \geq c) \leq \sum_{k=c}^{n} \binom{n}{k} (p^*_{\text{win}})^k (1-p^*_{\text{win}})^{n-k}, \quad (2)$$

where the right-hand side of Eq. (2) is the tail of a binomial distribution with parameters $n$ and $p^*_{\text{win}}$. This tail, which is the probability that the binomial takes a value equal or larger than $c$ can be computed numerically. In practice, if the winning probability is above the classical value, the right-hand side decreases very fast. For a fixed frequency of wins above $p^*_{\text{win}}$, the tail decreases exponentially fast to zero, see Figure 2. A similar result holds even if the inputs of the CHSH game are not chosen uniformly at random but they have a small bias [4, 9].

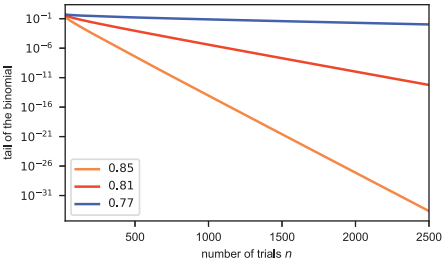



**Figure 2** Evaluation of the right-hand side of Eq. (2) for different values of frequencies of wins above the classical value.

### Key distillation

In this section, we relate the winning probability with the possibility of distilling secret keys. For the purpose of this discussion we say that a protocol produces a perfect key of $n$ bits if it is distributed uniformly at random over the set of strings of length $n$, Alice and Bob share the same string with probability one and the probability that any third party guesses the value of the string is no better than a random guess, i.e. the guessing probability is at most $2^{-n}$. This definition can be made robust by allowing close to uniform distributions in terms of the variational distance and close to random guesses.

Let us assume that Alice and Bob share a maximally entangled state (see Eq. (1)). Then, we know that since it is a pure state, it needs to be in tensor product form with any additional system that belongs to Eve. Moreover, if two states are in product form, then it is not possible to infer information about one by measuring the other. You can convince yourself by writing a product state and computing the joint probability distribution after measuring each state in some basis. Moreover, if Alice and Bob now measure the state in the computational basis they obtain a uniformly random bit and in consequence they share a perfect secret key.

The problem in practice is that it is not possible a priori to know what is the state shared by Alice and Bob without relying on additional assumptions. However, as we saw in the previous section, if Alice and Bob observe a frequency of wins close to the quantum value of CHSH, they can conclude that they share states close to the maximally entangled state. Hence, if Alice and Bob have access to some source that provides them with maximally entangled states they could try to distil a key as follows. In each round, both Alice and Bob decide randomly whether they play the CHSH game or they try to distil key by measuring in the computational basis. After $n$ rounds, they share their random choices and compute the number of wins in the CHSH game. If the number of wins is large enough they can conclude that they shared maximally entangled states with high probability and that the values measured in the distillation rounds constitute a secret key. Of course, making this intuition rigorous is rather challenging, see [1] for a recent proof.

Even if the protocol that we describe below is dubbed device-independent there are some assumptions that need to be made. Let us go over them. First, there is some classical information that Alice and Bob need to exchange during the protocol. This information, although not secret, should arrive undistorted at the other party. For this, Alice and Bob can use an authenticated classical communications channel. Alternatively, they can themselves implement an authenticated channel if they share in advance a small secret key. Second, both Alice and Bob's laboratories are assumed to be secure, i.e. no information leaks during or after the protocol. Note, however, that no assumption is made on the measurement or preparation devices that can be completely uncharacterized. In this sense, the protocol is device-independent.

The following protocol is a version of the Ekert protocol [8] proposed in [12]:

1. Alice generates a uniformly random string of measurement bases $X = (x_0,\ldots,x_{n-1}) \in \{0,1\}^n$ and measures sequentially in the basis given by $A_x$. The device outputs a string of measurement outcomes $A = (a_0,\ldots,a_{n-1}) \in \{0,1\}^n$ .
2. Bob generates a uniformly random string of measurement bases $Y \in \{0,1,2\}^n$ and measures sequentially in the basis given by $B_y$, where $B_2 = A_0$. The device outputs a string of measurement outcomes $B = (b_0,\ldots,b_{n-1}) \in \{0,1\}^n$.
3. Alice and Bob communicate to each other the measurement strings $X$ and $Y$ over the classic communications channel.
4. Alice chooses a random subset of $S \subset \{0,1,\ldots,n-1\}$ of size $|S| = n/2$ and sends $S$ to Bob. Alice and Bob compute the following sets: $T = \{i \in S, y_i \neq 2\}$, $U = \{i \in S, x_i = 0, y_i = 2\}$ and $V = \{i \notin S, x_i = 0, y_i = 2\}$.
5. Alice and Bob compute the number of CHSH wins and mismatches on the sets $T$ and $U$:
$$f_{\text{win}} = \frac{|\{i \in T, x_i y_i = a_i \oplus b_i\}|}{|T|},$$
$$f_{\text{error}} = \frac{|\{i \in U, x_i \neq y_i\}|}{|U|}.$$
If $f_{\text{win}}$ and $1 - f_{\text{error}}$ are not above some predetermined threshold the protocol aborts.
6. Alice and Bob perform a sequence of classical postprocessing steps to distil a secret key [13].

Using this protocol, for $n$ large enough Alice and Bob can distil secret keys at a rate that scales linearly with $n$ and is a function only of the frequencies of wins and mismatches $f_{\text{win}}$ and $f_{\text{error}}$ [1].

## Summary and further reading

We introduced several topics. Let us summarize and point to appropriate sources for further reading. First, we presented the CHSH game and showed that the maximum winning probability with local strategies is $0.75$. In order to discuss the winning probability with quantum resources, we briefly introduced some rudiments of quantum information and showed that the maximally entangled state achieves a strictly larger winning probability than the local value. We point the reader to the review paper [5] for more information on non-local games and to [11] for a thorough introduction to quantum information.

Then, we argued that in a real experiment it is not possible to observe probabilities, but it is still possible to conclude that Alice and Bob are implementing a non-local strategy. Moreover, for a large enough number of games, if the frequency of wins is close to the optimal value it is possible to conclude that Alice and Bob share a state close to the maximally entangled state. We concluded by introducing the concept of secret key and arguing that the maximally entangled state can be used to produce a secret key. Finally, we presented a protocol for key distribution that randomly intercalates rounds of CHSH and key production. We point the reader to [13] for an in-depth discussion of (non-device-independent) QKD and to [1] for a security proof of diQKD.

## References


1 R. Arnon-Friedman, R. Renner and T. Vidick, Simple and tight device-independent security proofs, arXiv:1607.01797, 2016.

2 J.S. Bell, *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy,* Cambridge University Press, 2004.

3 C.H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, *International Conference on Computers, Systems & Signal Processing,* Bangalore, India, 1984.

4 P. Bierhorst, A robust mathematical model for a loophole-free Clauser–Horne experiment, *Journal of Physics A: Mathematical and Theoretical* 48(19) (2015).

5 N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner, Bell nonlocality, *Reviews of Modern Physics* 86(2) (2014).

6 B.S. Cirel'son, Quantum generalizations of Bell's inequality, *Letters in Mathematical Physics* 4(2) (1980).

7 J.F. Clauser, M.A. Horne, A. Shimony and R.A. Holt, Proposed experiment to test local hidden-variable theories, *Physical Review Letters* 23(15) (1969).

8 A.K. Ekert, Quantum cryptography based on Bell's theorem, *Physical Review Letters* 67(6) (1991).

9 D. Elkouss and S. Wehner, (Nearly) optimal P values for all Bell inequalities, *NPJ Quantum Information* 2 (2016), 16026.

10 M. McKague, T.H. Yang and V. Scarani, Robust self-testing of the singlet, *Journal of Physics A: Mathematical and Theoretical,* 45(45) (2012).

11 M.A. Nielsen and I.L. Chuang, *Quantum Information and Quantum Computation,* Cambridge University Press, 2000.

12 S. Pironio, A. Acin, N. Brunner, N. Gisin, S. Massar and V. Scarani, Device-independent quantum key distribution secure against collective attacks, *New Journal of Physics* 11(4) (2009), 045021.

13 M. Tomamichel and A. Leverrier, A rigorous and complete proof of finite key security of quantum key distribution, arXiv:1506.08458, 2015.

14 U. Vazirani and T. Vidick, Fully device-independent quantum key distribution, *Physical Review Letters* 113(14) (2014).

Berry Schoenmakers
*Department of Mathematics & Computer Science*
*Eindhoven University of Technology*
*berry@win.tue.nl*

# Binary pebbling algorithms for in-place reversal of one-way hash chains

**Berry Schoenmakers is associate professor (UHD) in the Coding and Crypto Group at Eindhoven University of Technology. Schoenmakers is known for his work on cryptographic protocols for electronic voting, electronic payment, and secure computation. In this article he discusses an algorithmic problem pertaining to the backward traversal of a one-way hash chain, which has a unique motivation in cryptography. At the core of its recently obtained solution lies an intricate fractal structure which turns out to have a very nice and simple characterization.**

The problem is formulated in terms of a length-preserving one-way function $f$. A concrete example is the classical Davies–Meyer one-way function constructed from a block cipher such as AES:

$$f : \begin{cases} \{0,1\}^{128} \to \{0,1\}^{128} \\ \qquad x \mapsto \mathrm{AES}_x(\mathbf{0}). \end{cases}$$

That is, $f(x)$ is computed as an AES encryption of the trivial all-zero message $\mathbf{0}$ under the key $x$, which can obviously be done efficiently. On the other hand, recovering $x$ from $f(x)$ is tantamount to recovering an AES key given a single plaintext–ciphertext pair, which is assumed to be computationally hard. Therefore, $f$ is called a *one-way function*, as it is easy to evaluate but hard to invert. Block ciphers like AES are normally used for symmetric encryption to provide confidentiality, whereas one-way functions like $f$ are often used for asymmetric authentication, e.g., in the construction of digital signature schemes.

## One-way chains

Back in 1981, Lamport (the 'La" in LaTeX) proposed an elegant asymmetric identification scheme which operates in terms of one-way chains [12]. A *one-way chain* is the sequence formed by the successive iterates of $f$ for a given value. For example, in a client-server setting, the client may apply $f$ four times for a randomly chosen 128-bit seed value $x_0$ to obtain a length-4 chain:

$$x_0 \xrightarrow{f} x_1 \xrightarrow{f} x_2 \xrightarrow{f} x_3 \xrightarrow{f} x_4.$$

Lamport's *identification scheme* then operates as follows. At the start, the client registers itself securely with the server, as a result of which the server associates the endpoint $x_4$ with the client. Depending on the details, this registration step may be rather involved. However, from now on the client may identify itself securely to the server simply by releasing the next preimage on the chain. In the first round of identification, the client releases preimage $x_3$, and the server checks this value by testing if $f(x_3) = x_4$ holds. An eavesdropper obtaining $x_3$ cannot impersonate the client later on because the next round the server will demand a preimage for $x_3$. At this stage, preimage $x_2$ satisfying $f(x_2) = x_3$ is known only to the client; it is not even known to the server yet, which is why the scheme provides *asymmetric* authentication.

One-way chains and variations thereof are often referred to as *hash chains* since cryptographic hash functions such as SHA-256 are commonly used as alternatives for $f$. Hash chains are fundamental to many constructions in cryptography, and even to some forms of cryptanalysis (e.g., rainbow tables). Bitcoin's blockchain [15] is probably the best-known example of a hash chain nowadays — but note that blockchains are costly to generate due to the additional 'proof of work' requirement for the hash values linking successive blocks. Hash chains are also used in digital signature schemes required to be quantum secure, building on work by Merkle from 1979 [14]. Incidentally, Merkle attributes the use of iterated functions to Winternitz. However, Winternitz's idea is to use only one preimage on a length-$n$ chain, basically to securely encode an integer in the

set $\{0,\ldots,n-1\}$, whereas Lamport's idea is to use all of the $n$ preimages. The CAFE phone-tick scheme [2, Section 3.5] (see also [17]) and later micropayment schemes (e.g., PayWord [19]) actually combine these two ideas. In the case of phone-ticks, the caller releases the endpoint of a chain at the start of a call; at each tick, the caller simply releases the next preimage (as in Lamport's scheme) to pay for continuing the call. After the call ends, the phone company only needs to keep the last preimage released by the caller to claim the amount due (as in Winternitz's encoding).

**Security of one-way chains**

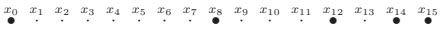
The use of a cryptographic hash function to create a one-way chain is overkill, however. A function like SHA-256 is not just one-way but is also designed to compress bit strings of practically unlimited length, and related to this, SHA-256 is required to be collision-resistant as well. For the security of a one-way chain, $f$ should be one-way, that is, given $y$ in the range of $f$ it must be hard to find any $x$ such that $f(x)=y$. Or rather, as recognized in [13,17], $f$ should necessarily be *one-way on its iterates*, which says that, for a length-$n$ chain, given an $n$th iterate image $y$ (in the range of $f^n$) it must be hard to find any $x$ such that $f(x)=y$.

Viewing $f$ as a random function (as in the random oracle model for hash functions), it follows that finding such a preimage $x$ takes $2^{128}/n$ time approximately. If $n=1$ this is simply the problem of inverting $f$, which can only be solved by making random guesses for $x$; on each attempt one succeeds with probability $1/2^{128}$. For $n>1$, however, one should not guess randomly. First, observe that the set of $n$th iterate images $y$ (range of $f^n$) is much smaller than $\{0,1\}^{128}$. In fact, the expected number of $n$th iterate images $y$ is equal to $(1-\tau_n)2^{128}$, where $\tau_0=0$, $\tau_n=e^{-1+\tau_{n-1}}$ for $n\geq 1$ [4, Theorem 2(v)]. To take advantage of the given that $y$ is not just any image but an $n$th iterate image, start with a random guess $x_0$ and then check if $x_1=f(x_0)$ happens to match $y$. Next, compute $x_2=f(x_1)$ and again test for equality with $y$, and continue to do so until $x_n$ is reached. The overall probability of hitting $y$ and thus obtaining a preimage of $y$ as well works out as $n/2^{128}$ approximately. Hence, even for very long chains of length $n=2^{32}$, say, the security level is still $2^{96}$. See also [8, Theorem 3] for a further analysis.

**Pebbling algorithms**

The above provides a solid basis for Jakobsson's wonderful idea of using efficient pebbling algorithms to make Lamport's scheme practical even for very long chains [11]. Naive implementations would render Lamport's scheme completely impractical: both (i) computing $x_{n-1}=f^{n-1}(x_0)$ to perform the first round of identification, then computing $x_{n-2}=f^{n-2}(x_0)$ from scratch, and so on, and (ii) storing all of $x_0,x_1,\ldots,x_{n-2},x_{n-1}$ to perform each round of identification instantly, are out of the question. The crux of Jakobsson's pebbling algorithm is to achieve a good *space-time trade-off*: for chains of length $n=2^k$, Jakobsson's algorithm stores $O(\log n)$ hash values throughout, and the maximum number of hashes performed in any round of identification is $O(\log n)$ as well.

Each hash value stored is associated with a *pebble*. For a length-16 chain, five pebbles are initially arranged as follows, which is typical of a *binary pebbling algorithm*:




The general pattern is that starting from the rightmost pebble, the distance to the next pebble doubles each time. From this initial arrangement, the first two elements $x_{15}$ and $x_{14}$ of the reverse of $\{x_0,x_1,\ldots,x_{15}\}$ can be output directly. For the third element $x_{13}$ we need to apply $f$ once to recompute it from $x_{12}$. The fourth element $x_{12}$ can be output again without any effort.

To produce $x_{11}$, something interesting happens. Because $f$ is one-way, the only sensible option is to recompute it from $x_8$ as $x_{11}=f^3(x_8)$. But while doing so, the value of $x_{10}=f^2(x_8)$ is also stored for later use. Hence, just before $x_{11}$ is output, the pebbles are arranged as follows:


Proceeding this way and computing outputs just-in-time, the *rushing* binary pebbling algorithm $R_k$ is obtained:

$$R_0(x) = \text{output } x,$$
$$R_k(x) = R_{k-1}(f^{2^{k-1}}(x)); R_{k-1}(x).$$

The reader may check that $R_k(x)$ outputs the sequence

$$f_k^*(x) = \{f^i(x)\}_{i=0}^{2^k-1}$$

in reverse, using $k2^{k-1}$ hashes in total. In addition, the storage requirements are low: $R_k(x)$ needs to store $x$ for the recursive call $R_{k-1}(x)$ later on, which leads to a maximum of $k+1$ values stored (pebbles) at any moment.

The only drawback is that $R_k$ in the worst case requires time exponential in $k$ between producing successive outputs. Removing these slow rounds is exactly what makes the problem non-trivial. That is, we seek a way to reverse $f_k^*(x)$ satisfying the performance constraints of using $O(k)$ storage (pebbles) and using $O(k)$ applications of $f$ (hashes) between producing *any* two successive outputs. To study this problem we introduce a specific framework for binary pebbling algorithms that operate in rounds.

At this point we like to mention that there are many similar notions of 'pebbling' in the literature. In particular, pebbling games (see, e.g., [16]) are somewhat related, and have recently been used in the context of cryptography to prove memory-hardness of certain hash functions [1]. Graph pebbling is another well-known problem (see, e.g., [9]). Reversible computing (see, e.g., [18]) gives rise to even more uses of pebbling (aka 'checkpointing', see below). As discussed in [20], however, the specific worst-case constraint limiting the number of hashes per round is unique to the cryptographic setting, starting with the work in [10,11].

**Framework for binary pebbling**

For $k\geq 0$, $P_k(x)$ will be defined as an algorithm that runs for $2^{k+1}-1$ rounds in total, and outputs $f_k^*(x)$ in reverse in its last $2^k$ rounds. It is essential that we include the initial $2^k-1$ rounds (in which no outputs are produced) as an integral part of pebbler $P_k(x)$, as this allows for a fully recursive definition and analysis of binary pebbling. In fact, in terms of a given *schedule* $T_k=\{t_{k,r}\}_{r=1}^{2^k-1}$, which fixes the number of hashes for each initial round, a *binary pebbler* $P_k(x)$ is completely specified by the recursive definition given in Figure 1. This means, for example, that $P_0(x)$ runs for one round only outputting $y_0=x$ itself, and that $P_1(x)$ will run for three rounds, performing $t_{1,1}=1$ hash in its first round, outputting $y_0=f(x)$ in its second round, and outputting $y_1=x$ in its last round. In general, $P_k(x)$ computes $f^{2^k-1}(x)$ using exactly $2^k-1$ hashes in total in its initial stage, storing only the values $y_k,\ldots,y_0$ along the way. Running pebblers

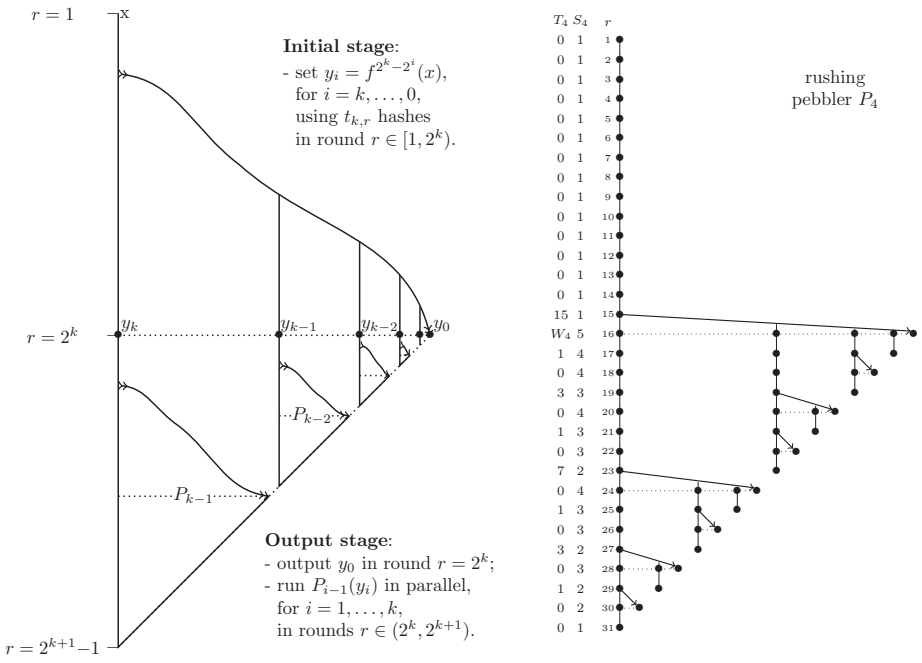



**Figure 1** Binary pebbler $P_k(x)$ for schedule $T_k = \{t_{k,r}\}_{r=1}^{2^k-1}$ satisfying $\sum_{r=1}^{2^k-1} t_{k,r} = 2^k - 1$ (left). Schedule $T_4$, work $W_4$, and storage $S_4$ for rushing pebbler $P_4$ in rounds $r = 1$ to $r = 31$ (right). Bullets represent stored values (pebbles), rightwards arrows represent hashing, vertical lines represent copying.

$P_{k-1}, \ldots, P_0$ in parallel in the output stage means that pebblers take turns to execute for one round each, where the order in which this happens within a round is irrelevant. It is not hard to prove that in every round exactly one of the pebblers running in parallel will be in its first output round, and that the sequence of outputs is always equal to $f_k^*(x)$.

Schedule $T_k$ specifies the number of hashes for the initial stage of $P_k$. To analyze the *work* done by $P_k$ in its output stage, we let sequence $W_k$ of length $2^k - 1$ denote the number of hashes performed by $P_k$ in each of its last $2^k - 1$ rounds — noting that by definition no hashes are performed by $P_k$ in round $2^k$. The following recurrence relation for $W_k$ will be useful throughout:

$$W_0 = \{\}, \quad W_k = T_{k-1} + W_{k-1} \| \{0\} \| W_{k-1},$$

where $T_{k-1} + W_{k-1}$ denotes elementwise addition of $T_{k-1}$ and $W_{k-1}$ and $\|$ concatenation of sequences (+ takes precedence over $\|$).

To analyze the *storage* needed by $P_k$ the number of hash values stored by $P_k$ will be counted for each round. We let sequence $S_k = \{s_{k,r}\}_{r=1}^{2^{k+1}-1}$ denote the total storage used by $P_k$ at the start of each round. For instance, $s_{k,1} = 1$ as $P_k$ only stores $x$ at the start, and $s_{k,2^k} = k + 1$ as $P_k$ stores $y_0, \ldots, y_k$ at the start of round $2^k$ independent of schedule $T_k$.

The rushing pebbler $P_k$ corresponding to $R_k$ introduced above is obtained by taking schedule $T_k$ with $t_{k,2^k-1} = 2^k - 1$ and $t_{k,r} = 0$ elsewhere. Rushing pebbler $P_4$ is illustrated in Figure 1 in our framework for binary pebbling. The storage $S_4$ is minimal throughout, but for the work $W_4$ there are big peaks: e.g., in round 23, in total seven hashes are performed, while the pebbler is idle in all even rounds.

## Towards optimal solution

As it turns out, our framework admits a simple solution obtained by taking schedule $T_k = \{1\}_{r=1}^{2^k-1}$, resulting in the *speed-1* pebbler illustrated in Figure 2. The above recurrence relation for $W_k$ yields $\max(W_k) = k - 1$ for $k \geq 1$, and it can also be shown that $\max(S_k) = \max(k + 1, 2k - 2) = O(k)$. The speed-1 pebbler thus achieves the desired asymptotic bounds. For practical purposes, however, further savings are needed to limit the costs as much as possible. E.g., to enable a lightweight client device to identify itself every half hour for a period of three years using a length-$2^{16}$ chain.

Jakobsson's pebbling algorithm [11] provides a clever way to cut storage $\max(S_k)$ in half essentially. Translated to our framework for binary pebbling, the corresponding schedule $T_k$ is obtained by setting $t_{k,r} = 0$ for $1 \leq r < 2^{k-1}$, $t_{k,r} = 2$ for $2^{k-1} \leq r < 2^k - 1$, and $t_{k,2^k-1} = 1$. The pebbler obtained this way is called the *speed-2* pebbler, illustrated in Figure 2.

Compared to a speed-1 pebbler, the crucial idea of a speed-2 pebbler is to remain idle for the first half of the initial stage — preventing that too many pebblers are active at the same time — and then make up for this by hashing at double speed in the remaining time. It can be proved that $\max(W_k) = k - 1$ for $k \geq 1$ also holds for a speed-2 pebbler, but compared to a speed-1 pebbler storage is now reduced by 50%, achieving $\max(S_k) = k + 1$. For binary pebbling algorithms, storage $S_k$ of up to $k + 1$ hash values is optimal, since this amount of storage is already needed during the first output round $r = 2^k$, for any binary pebbler $P_k$. The interesting question is whether the work $\max(W_k)$ can be reduced any further?

## Optimal binary pebbling

An elementary analysis yields $\max(W_k) \geq \lceil k/2 \rceil$, $k \geq 2$, as lower bound for any binary pebbling algorithm (see [20, Theorem 2]). So, the best that can be achieved is to reduce the maximum number of hashes for any output round to $k/2$ roughly. The problem of optimally efficient hash chain reversal was extensively studied by Coppersmith and Jakobsson [3]. They achieved nearly optimal space-time complexity for a complicated pebbling algorithm using $k + \lceil \log_2(k+1) \rceil$ pebbles and no more than $\lfloor \frac{k}{2} \rfloor$ hashes per round. Hence, an excess storage of approximately $\log_2 k$ hash values compared to optimal binary pebbling.

Fortunately, Yum et al. [21] observed that a greedy implementation of Jakobsson's original pebbling algorithm already achieves the optimal space-time trade-off for binary pebbling. Their idea is to greedily use up a budget of $\lfloor \frac{k}{2} \rfloor$ hashes per round subject to the constraint that no more than about $k$ hash values are stored at any time. The only drawback of the greedy approach is that no apparent structure is revealed.

In contrast, we have found an explicit, essentially unique solution for optimal binary pebbling, which leads to a complete understanding of the problem and paves the way for fully optimized in-place implementations. As a closed formula, the optimal schedule $T_k$ is obtained by setting $t_{k,r} = 0$ for $1 \leq r < 2^{k-1}$, and setting $t_{k,r}$ to

$$\left\lfloor \tfrac{1}{2}((k+r) \bmod 2 + k + 1 - \mathrm{len}((2r) \bmod 2^{\mathrm{len}(2^k - r)})) \right\rfloor$$

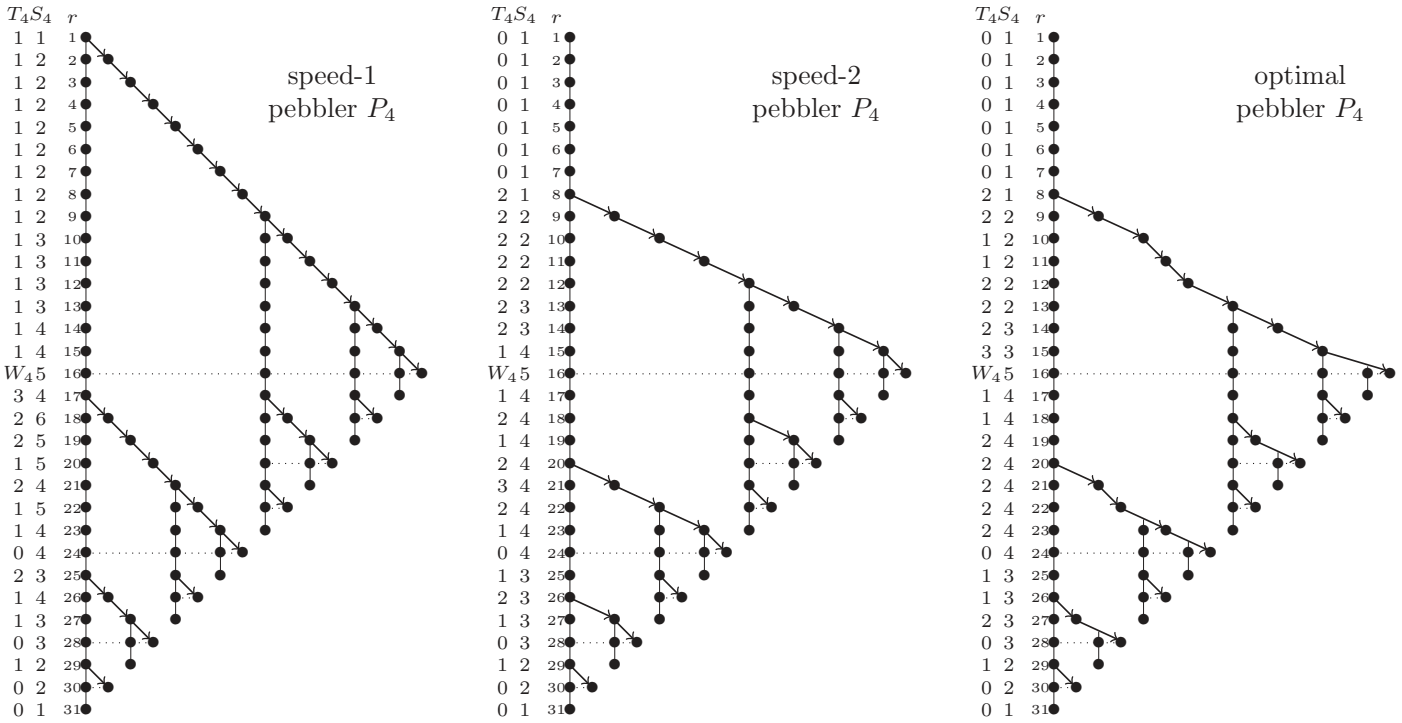



**Figure 2** Schedule $T_4$, work $W_4$, storage $S_4$ for three types of binary pebblers $P_4$ in rounds 1–31

for $2^{k-1} \le r < 2^k$, where $\mathrm{len}(n)$ denotes the bit length of nonnegative integer $n$. Optimal pebbler $P_4$ is illustrated in Figure 2, which uses the following optimal schedules:

$$T_0 = \{\},$$
$$T_1 = \{1\},$$
$$T_2 = \{0,1,2\},$$
$$T_3 = \{0,0,0,2,1,2,2\},$$
$$T_4 = \{0,0,0,0,0,0,0,2,2,1,1,2,2,2,3\}.$$

In general, an optimal pebbler $P_k$ will use up to $\max(W_k) = \lceil k/2 \rceil$ hashes in any output round. For the optimal pebbler $P_4$ in Figure 2, this works out as $\max(W_4) = 2$ hashes, compared to the speed-2 pebbler $P_4$ which needs 3 hashes in output round $r = 21$.

The fractal nature of the optimal schedule $T_k$ is revealed by the recursive characterization in terms of sequences $U_k$, $V_k$ defined in Table 1. These sequences are defined over $\frac{1}{2}\mathbb{Z}$ — rather than over $\mathbb{Z}$ as will ultimately be required for use in a pebbling algorithm. Without rounding of these half-integers, the optimal schedule satisfies the following *key equation* in terms of sequences $U_k$, $V_k$, $k \ge 2$:

$$(U_k \,\|\, V_k) + (\{0\} \,\|\, W_{k-1}) = \{\tfrac{k+1}{2}\}^{2^{k-1}}.$$

This equation basically says that the optimal schedule does not leave any gaps: in each round exactly the maximum number of hashes are performed to meet the lower bound for binary pebbling.

### Efficient in-place implementations

Without strict performance requirements, our framework for binary pebbling allows for relatively straightforward implementations. Figure 4 is showcasing a conceptually simple implementation based on Python generators. For demonstration purposes, we are using MD5 as a 128-bit length-preserving one-way function — MD5 is readily available in Python, also no practical attacks against the one-wayness of MD5 are known to this day.

By exploiting specific properties of the optimal schedule, we will next show how to implement binary pebblers with minimal overhead. In fact, we present *in-place* hash chain reversal algorithms, where the entire state of these algorithms (apart from the hash values) is represented between rounds by a single $k$-bit counter *only*. Below, this is shown for Jakobsson's speed-2 pebblers; refer to [20] for further results.

We use the following terminology to describe the state of a pebbler $P_k$ (which applies to both speed-2 pebblers and optimal pebblers). Pebbler $P_k$ is said to be *idle* if it is in rounds $[1, 2^{k-1})$, *hashing* if it is in rounds $[2^{k-1}, 2^k]$, and *redundant* if it is in rounds $(2^k, 2^{k+1})$. An idle pebbler performs no hashes at all, while a hashing pebbler will perform at least one hash per round, except for round $2^k$ in which $P_k$ outputs its $y_0$ value. The work for a redundant pebbler $P_k$ is taken over by its child pebblers $P_0, \ldots, P_{k-1}$ during its last $2^k - 1$ output rounds.

The important observation is that for each round $r$ the complete state of a pebbler $P_k$ can be deduced quickly from the binary representation of the counter $c = 2^{k+1} - r$, which counts down how many rounds are

| $U_2 = \{\frac{3}{2}\}, U_k = U_{k-1} + \frac{1}{2} \,\|\, \{1\}^{2^{k-3}}$ | | $V_2 = \{\frac{3}{2}\}, V_k = U_{k-1} + \frac{1}{2} \,\|\, V_{k-1} + \frac{1}{2}$ | |
|---|---|---|---|
| $U_3$ | $21$ | $V_3$ | $22$ |
| $U_4$ | $\frac{5}{2}\frac{3}{2}11$ | $V_4$ | $\frac{5}{2}\frac{3}{2}\frac{5}{2}\frac{5}{2}$ |
| $U_5$ | $32\frac{3}{2}\frac{3}{2}1111$ | $V_5$ | $32\frac{3}{2}\frac{3}{2}3233$ |
| $U_6$ | $\frac{7}{2}\frac{5}{2}22\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}11111111$ | $V_6$ | $\frac{7}{2}\frac{5}{2}22\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{7}{2}\frac{5}{2}22\frac{7}{2}\frac{5}{2}\frac{7}{2}\frac{7}{2}$ |
| $U_7$ | $43\frac{5}{2}\frac{5}{2}2222\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}1111111111111111$ | $V_7$ | $43\frac{5}{2}\frac{5}{2}2222\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}\frac{3}{2}43\frac{5}{2}\frac{5}{2}222243\frac{5}{2}\frac{5}{2}4344$ |
| | avg. speed 2 speed 1 | | avg. speed 2 avg. speed 3 |

**Table 1** Recursive definition of optimal schedule $T_k = \{0\}^{2^{k-1}-1} \| U_k \| V_k$ over $\frac{1}{2}\mathbb{Z}$ (no rounding). Explicit formula is in this case given by $T_k = \{0\}^{2^{k-1}-1} \,\|\, \{\frac{1}{2}(k+1-\mathrm{len}((2r) \bmod 2^{\mathrm{len}(2^k-r)}))\}_{r=2^{k-1}}^{2^k-1}$.

still left. This is illustrated in Figure 3 for a speed-2 pebbler $P_k(x)$. The pseudocode shows how to run the pebbler in-place, that is, in such a way that the storage between rounds is limited to a length-$k$ array $z$ of hash values and counter $r$. The information about the states of all pebblers running in parallel is deduced directly from $c$. This information includes which pebblers are present, whether these pebblers are idle or hashing, which hash values have already been computed by a pebbler, and where these are stored in array $z$, et cetera.

The example in Figure 3 shows the details for a $P_9$ pebbler at round $r = 664$. Four child pebblers $P_8, P_6, P_5, P_3$ are running in parallel: $P_8$ is hashing and has entries $z[7,8]$ in use, $P_6$ is idle occupying one entry $z[6]$, $P_5$ is hashing and has entries $z[4,5]$ in use. The $P_3$ pebbler has just reached its first output round occupying four entries $z[0,3]$ and outputs its $y_0$ value stored in $z[0]$. Subsequently, this $P_3$ pebbler becomes redundant and is replaced by its child pebblers $P_2$, $P_1$, $P_0$, which will each use one entry of array $z$. Entry $z[3]$ has been freed, but is immediately used again by the $P_5$ pebbler, which just reached the point where it starts working on its $y_3$ value.

The schedule for a speed-2 pebbler is integrated in the pseudocode of Figure 3. For optimal pebbling, however, we need to evaluate the formula for the optimal schedule to find the exact number of hashes to be performed by each pebbler. An intuitive way to interpret this formula is explained by means of the following example, cf. Figure 3. Consider optimal pebbler $P_9$ at $c = 360$ rounds from the end:

| $c_8$ | $c_7$ | $c_6$ | $c_5$ | $c_4$ | $c_3$ | $c_2$ | $c_1$ | $c_0$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| $P_8^{\text{hashing}}$ | | | $P_5^{\text{hashing}}$ | | | | | |

The formula of Table 1 for the optimal schedule (before rounding) partitions the bits of $c$ into the two colored segments as indicated. The underlying rule is as follows. First, all the hashing pebblers are identified, ignoring the rightmost one: this results in two hashing pebblers $P_8$ and $P_5$ (idle pebbler $P_6$ and the rightmost hashing pebbler $P_3$ are ignored). Then, each of these hashing pebblers $P_i$ gets the segment assigned starting at bit $c_i$ and extending to the right. The number of hashes to be performed by each of these hashing pebblers — as given by the formula of the optimal schedule — exactly matches the length of these segments divided by 2. In case of $P_8$ this works out as $\frac{3}{2}$ hashes, and for $P_5$ we get $\frac{6}{2}$ hashes, hence exactly $\frac{9}{2}$ hashes are used in total for this round.

In general, this rule implies that no more than $\frac{k}{2}$ hashes are performed in any output round of $P_k$. Moreover, this simple rule will orchestrate the entire computation, ensuring that all intermediate hash values are computed right on time — not too late to fail producing an output on time, and not too early, before another free entry in array $z[0,k)$ becomes available. The optimized implementations in [20] are based on this rule.

**Round** $r$:
1: output $z[0]$
2: $c \leftarrow 2^{k+1} - r$
3: $i \leftarrow \text{pop}_0(c)$
4: $z[0,i) \leftarrow z[1,i]$
5: $i \leftarrow i+1$; $c \leftarrow \lfloor c/2 \rfloor$
6: $q \leftarrow i-1$
7: **while** $c \neq 0$ **do**
8: $z[q] \leftarrow f(z[i])$
9: **if** $q \neq 0$ **then** $z[q] \leftarrow f(z[q])$
10: $i \leftarrow i + \text{pop}_0(c) + \text{pop}_1(c)$
11: $q \leftarrow i$

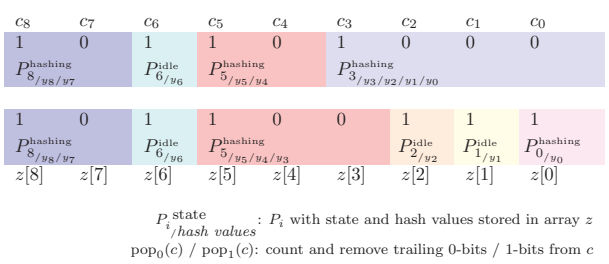



**Figure 3** Pseudocode for in-place speed-2 pebbler $P_k(x)$ at output round $r$, $2^k < r < 2^{k+1}$. Initially, array $z[0,k)$ satisfies $z[i-1] = f^{2^k-2^i}(x)$ for $i = 1,\ldots,k$ (left). Transition of $P_9$ from round $r = 664$ to $r = 665$, hence from $c = 360 = (101101000)_2$ to $c = 359 = (101100111)_2$ (right).

## Lower bound

Optimal binary pebbling achieves a space-time product of $0.50k^2$ for a chain of length $n = 2^k$. In an upcoming paper with Niels de Vreede, we will show how to reduce the space-time product to $0.46k^2$ by means of Fibonacci pebbling and how to reduce this even further down to just $0.37k^2$ by more intricate pebbling algorithms. We note that Coppersmith and Jakobsson [3] gave a lower bound of $0.25k^2$, but whether this bound can be attained is doubtful: the lower bound is derived without taking into account any limits on the number of hashes per output round.

Incidentally, the lower bound of $0.25k^2$ had been found already in a completely different context [7], for a similar problem studied in the area of algorithmic (or, automatic, computational) differentiation [6]. The lower bound applies to the space-time complexity of so-called *checkpointing* for the reverse (or, adjoint, backward) mode of algorithmic differentiation. In contrast to our case, however, there it is even possible to attain the lower bound [5]. The critical difference is that in the setting of algorithmic differentiation the goal is basically to minimize the *total time* for performing this task (or, equivalently, to minimize the *amortized time* per output round). This contrasts sharply with the goal in the cryptographic setting, where we want to minimize the *worst case time* per output round while performing this task.

```
import hashlib, itertools

f = lambda x: hashlib.md5(x).digest()

tR = lambda k,r: 0 if r < 2**k - 1 else 2**k - 1
t1 = lambda k,r: 1
t2 = lambda k,r: 0 if r < 2**(k-1) else 2 if r < 2**k - 1 else 1
tS = lambda k,r: 0 if r < 2**(k-1) else ((k + r) % 2 + k + 1 - ((2 ** r) % (2**(2**k - r).bit_length())).bit_length()) // 2

def P(k,x):
    y = [None] * k + [x]
    i = k; g = 0
    for r in range(1, 2**k):
        for _ in range(t(k,r)):
            z = y[i]
            if g == 0: i -= 1; g = 2**i
            y[i] = f(z)
            g -= 1
        yield
    yield y[0]
    for v in itertools.zip_longest(*(P(i-1, y[i]) for i in range(1, k+1))):
        yield next(filter(None, v))

t = eval(input())
k = int(input())
x = f(b'')
for v in P(k, x):
    if v: print(v.hex())
```

**Figure 4** Python program for recursive binary pebblers without any optimizations, cf. definition of $P_k(x)$ in Figure 1. Inputs: `tR`/`t1`/`t2`/`tS` for rushing/speed-1/speed-2/optimal and nonnegative integer `k`. `P(k,x)` is a Python generator: each evaluation of a `yield` expression corresponds to a round of $P_k(x)$.

## References


1 J. Alwen, B. Chen, K. Pietrzak, L. Reyzin, and S. Tessaro, Scrypt is maximally memory-hard, *Advances in Cryptology–EUROCRYPT '17*, LNCS 10212, Springer, 2017, pp. 33–62.

2 J.P. Boly, A. Bosselaers, R. Cramer, R. Michelsen, S. Mjølsnes, F. Muller, T. Pedersen, B. Pfitzmann, P. de Rooij, B. Schoenmakers, M. Schunter, L. Vallée and M. Waidner, The ESPRIT project CAFE–High security digital payment systems, *Computer Security–ESORICS 94*, LNCS 875, Springer, 1994, pp. 217–230.

3 D. Coppersmith and M. Jakobsson, Almost optimal hash sequence traversal, *Financial Cryptography 2002*, LNCS 2357, Springer, 2002, 102–119.

4 P. Flajolet and A. Odlyzko, Random mapping statistics, *Advances in Cryptology–EUROCRYPT '89*, LNCS 434, Springer, 1989, pp. 329–354.

5 A. Griewank, Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation, *Optimization Methods and Software* 1(1) (1992), 35–54.

6 A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, 2008, 2nd edition.

7 J. Grimm, L. Potter and N. Rostaing-Schmidt, Optimal time and minimum space-time product for reversing a certain class of programs, *Computational Differentiation–Techniques, Applications, and Tools*, SIAM, 1996, pp. 95–106.

8 J. Håstad and M. Näslund, Key feedback mode: a keystream generator with provable security, *First Modes of Operation Workshop*, *Baltimore, MD, October 2000*, NIST.

9 G. Hurlbert, Recent progress in graph pebbling, *Graph Theory Notes of New York* 49 (2005), 25–37.

10 G. Itkis and L. Reyzin, Forward-secure signatures with optimal signing and verifying, *Advances in Cryptology–CRYPTO '01*, LNCS 2139, Springer, 2001, pp. 332–354.

11 M. Jakobsson, Fractal hash sequence representation and traversal, *Proc. IEEE International Symposium on Information Theory (ISIT '02)*, IEEE, 2002, p. 437. Full version eprint.iacr.org/2002/001.

12 L. Lamport, Password authentication with insecure communication, *Communications of the ACM* 24(11) (1981), 770–772.

13 L. Levin, One-way function and pseudorandom generators, *Proc. 17th Symposium on Theory of Computing (STOC '85)*, ACM, 1985, pp. 363–365.

14 R. Merkle, A digital signature based on a conventional encryption function, *Advances in Cryptology–CRYPTO '87*, LNCS 293, Springer, 1987, 369–378.

15 S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, October 31, 2008, bitcoin.org/bitcoin.pdf.

16 J. Nordström, Pebble games, proof complexity, and time-space trade-offs, *Logical Methods in Computer Science* 9(3:15) (2013), 1–63.

17 T.P. Pedersen, Electronic payments of small amounts, *Security Protocols*, LNCS 1189, Springer, 1996, pp. 59–68.

18 K. Perumalla, *Introduction to Reversible Computing*, CRC Press, 2013.

19 R.L. Rivest and A. Shamir, Payword and micromint: Two simple micropayment schemes, *Security Protocols*, LNCS 1189, Springer, 1996, pp. 69–87.

20 B. Schoenmakers, Explicit optimal binary pebbling for one-way hash chain reversal, *Financial Cryptography 2016*, LNCS 9603, Springer, 2016, pp. 299–320. Sample code in Python, Java, C at www.win.tue.nl/~berry/pebbling.

21 D.H. Yum, J.W. Seo, S. Eom and P.J. Lee, Single-layer fractal hash chain traversal with almost optimal complexity, *Topics in Cryptology–CT-RSA '09*, LNCS 5473, Springer, 2009, pp. 325–339.

Matthijs Coster
*Ministerie van Defensie*
*Den Haag*
*mj.coster@mindef.nl*

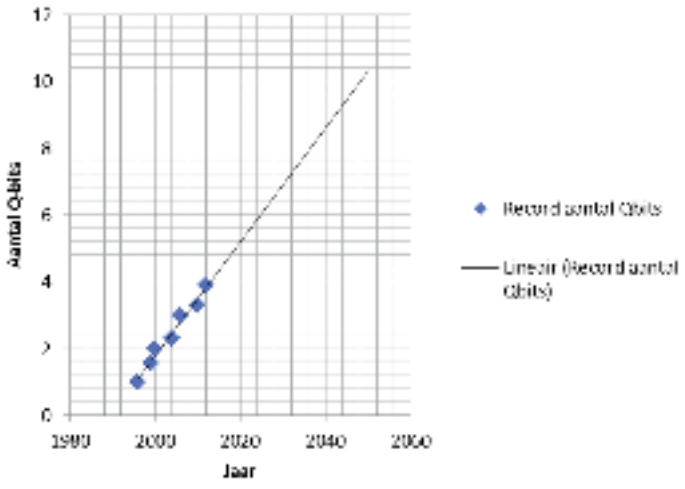
# Isogenieën over supersinguliere elliptische krommen

**Matthijs Coster is een wiskundig cryptologist werkzaam voor de Nederlandse overheid, maar is ook in zijn vrije tijd actief als redacteur voor het wiskundetijdschrift *Pythagoras*. In dit artikel legt Coster uit hoe een kwantumveilig cryptografisch sleuteluitwisselprogramma gebaseerd kan worden op de isogenieën van elliptische krommen.**

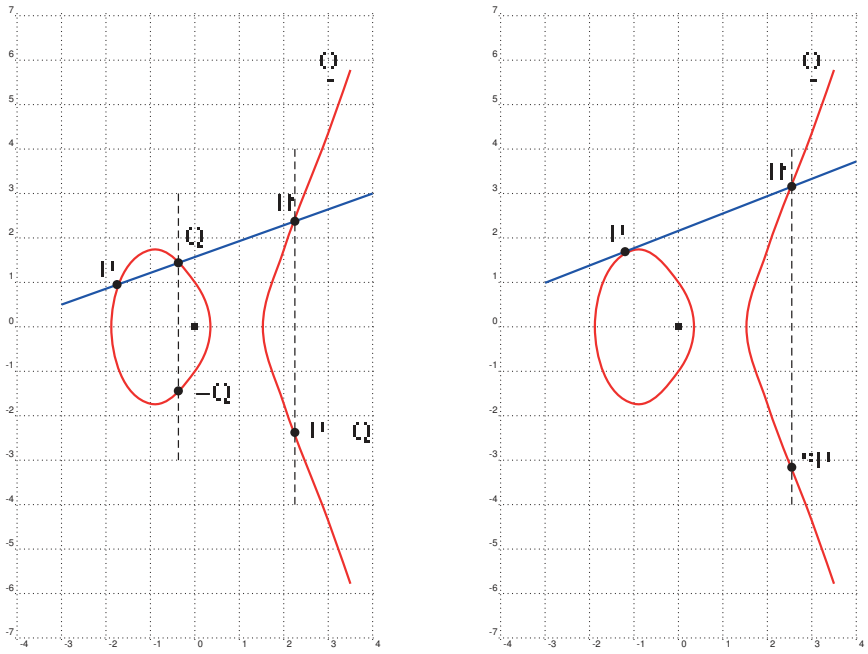
Komen de kwantumcomputers er aan? Afgelopen jaar ben ik intensief bezig geweest met onderzoek naar gevolgen van het gebruik van kwantumcomputers op de veiligheid van cryptografische producten. Bij het beveiligen van verbindingen en vooral data moet je in het hoofd houden dat als iets eenmaal is ontvreemd of geïntercepteerd, dan is dat het geval voor de eeuwigheid. Dat betekent dat als bijvoorbeeld een goed versleutelde laptop wordt ontvreemd, die plotseling met behulp van een kwantumcomputer als nog ontcijferd kan worden (mogelijk jaren na de ontvreemding), dan kan informatie op de laptop als nog worden gelezen. Als zich op de laptop data bevindt die 25 jaar geheim moet blijven, dan moet je er zeker van zijn dat dat de komende 25 jaar ook zo blijft. Er valt op het ogenblik nog niets te zeggen over wanneer de eerste kwantumcomputer in gebruik zal worden genomen. Er wordt van uitgegaan dat dat nog minimaal tien jaar gaat duren, wellicht langer, maar mogelijk is de kwantumcomputer al in gebruik genomen maar hebben we er hier in Nederland geen weet van. In Figuur 1 is een diagram te zien waarin ik een optimale lijn heb getrokken met alle recente doorbraken (jaartallen en logaritme van aantal qbits). Voor het factoriseren van een $1024$-bits RSA-modulus zijn met het algoritme, ontwikkeld door Peter Shor (zie [9]) $2048$ qbits nodig. Volgens het diagram zou dat pas in 2050 worden behaald. Maar een dergelijk diagram zegt natuurlijk heel weinig!

Ook andere voornamelijk publieke-sleuteluitwisselprogramma's kunnen door vergelijkbare algoritmes met een kwantumcomputer worden ontcijferd.

Kwantumcomputers kunnen extra goed worden ingezet op het moment dat er sprake is van een groep met veel structuur. Dat is het geval bij zowel RSA ($\mathbb{Z}/n\mathbb{Z}$) en de discrete logaritme over priemmachten als de discrete logaritme over elliptische krommen. ($\mathcal{E}/\mathcal{F}_q$, met $q$ een (macht van een) priemgetal. Elliptische krommen worden genoteerd met $\mathcal{E}$. In de volgende paragraaf wordt een definitie gegeven.) Op het ogenblik zijn cryptografen op zoek naar nieuwe publieke-sleuteluitwisselprogramma's, die niet een groep met veel structuur ten grondslag hebben, dus de zogenaamde *post-kwantum*-alternatieven van RSA.




**Figuur 1** Aantal qbits uitgezet tegen de tijd. Let op: voor de qbits is een logaritmische schaal gebruikt, dus lees voor $10$: $2^{10} = 1024$ qbits.




**Figuur 2** Optelling van $P$ en $Q$ op een elliptische kromme (links). Scalaire vermenigvuldiging van $P$ met $2$ op een elliptische kromme (rechts).

Onlangs schreven Dan Bernstein en Tanja Lange een overzichtsartikel waarin een aantal alternatieven wordt beschreven (zie [1]). Naast een aantal uitgebreid onderzochte alternatieven zijn de isogenieën over supersinguliere elliptische krommen minder intensief onderzocht. Het idee om isogenieën over elliptische krommen te gebruiken is oorspronkelijk voorgesteld in 2006 door Alexander Rostovtsev en Anton Stolbunov (zie [10]). Het bleek later dat er een aanpassing nodig was door een restrictie op te leggen op de elliptische krommen. Deze restrictie resulteerde in het gebruik van de *supersinguliere elliptische kromme*, maar de motivatie hiervoor noem ik in dit artikel niet. Het voorbeeld dat in dit artikel wordt besproken (conform Luca de Feo, David Jao en Jérôme Plût, zie [5]) bevat een elliptische kromme met $2^{2a}3^{2b}$ punten, met $2^a \approx 3^b \approx 2^{256}$. De discrete logaritme over een dergelijke kromme kan eenvoudig worden gebroken zonder veel rekenkracht. Er wordt vermoed dat isogenieën een vercijferingsschema kunnen opleveren dat lastig te kraken is, maar de motivatie is nog niet overtuigend.

## De basis: de elliptische kromme

Voor een gedetailleerde uiteenzetting over elliptische krommen refereer ik naar Andrew Sutherland [12]. Het artikel van Henk van Tilborg uit 2001 in het *Nieuw Archief voor Wiskunde* vormt een uitstekende basis om kennis te maken met elliptische krommen en toepassingen in de cryptografie (zie [13]).

Ik ga uit van een eindig lichaam $k = \mathcal{F}_q$ (karakteristiek $p > 2$, $q = p^n$) en een elliptische kromme $\mathcal{E}/k$ kan worden gerepresenteerd door de *Weierstrass-vergelijking* $y^2 = x^3 + ax + b$, maar deze representatie kan door een willekeurige worden vervangen (bijvoorbeeld Edwards Curve). Voor het aantal punten ($\#\mathcal{E}$) geldt $|\#\mathcal{E} - (q+1)| \leq 2\sqrt{q}$ (Hasse). Vaak is duidelijk over welk lichaam $k$ de elliptische kromme is gedefinieerd, en kan de index $k$ worden weggelaten. Punten $P$ en $Q$ op de kromme kunnen worden opgeteld en scalair worden vermenigvuldigd (zie Figuur 2). We noteren de scalaire vermenigvuldiging met $nP$. Met $\mathrm{ord}(P)$ (de orde van $P$) wordt het kleinste positieve geheel getal $m$ bedoeld zo dat $mP = O$. Uiteraard is $m$ een deler van $\#\mathcal{E}$.

Vanzelfsprekend bestaat er dan ook een oorsprong, die ik aanduid met $0$ of $O$ en in de literatuur ook wel $\infty$ (*het punt op oneindig*) genoemd wordt; in *projectieve* coördinaten: $(0{:}1{:}0)$.

## Isogenieën tussen de elliptische krommen

Een isogenie is een afbeelding die de onderliggende structuur behoudt (morfisme) tussen twee algebraïsche variëteiten (in ons geval elliptische krommen) die $0$ afbeelden op $0$. (Vergelijk dit met een homomorfisme in de algebra waarbij de eenheid van de ene groep overgaat naar de eenheid van de andere groep.)

Enkele voorbeelden van isogenieën:

- Op $\mathcal{E}$ is het altijd mogelijk om het *automorfisme* $(x,y) \to (x,-y)$ te definiëren. (In feite is dit de afbeelding $P \to -P$.)
- Voor een positief (of negatief) geheel getal $m$ is $[m]$ (de afbeelding $P \to mP$) een *endomorfisme* van $\mathcal{E} \to \mathcal{E}$.
- Het *Frobenius* endomorfisme, meestal genoteerd met $\pi_{\mathcal{E}}$ is de afbeelding $\mathcal{E}$ die $(x,y)$ afbeeldt op $(x^q, y^q)$. Ik ga er niet diep op in, omdat deze afbeelding geen rol speelt bij het hier besproken sleuteluitwisselprogramma.
- *Onjuist voorbeeld:* Als $Q$ een punt is op $\mathcal{E}$, niet de oorsprong, dan is $P \to P + Q$ een voorbeeld van een morfisme, maar niet een homomorfisme, dus geen isogenie.

De isogenie $\varphi$ is in feite een functie $\overline{k} \times \overline{k} \to \overline{k} \times \overline{k}$, waarbij $\overline{k}$ de algebraïsche afsluiting is van $k$. Een typische $\varphi$ heeft het format $\varphi(x,y) = \left(\frac{u(x)}{v(x)}, cy\left(\frac{u(x)}{v(x)}\right)'\right)$ als $\mathcal{E}$ wordt gerepresenteerd als een Weierstrass-vergelijking. (De isogenie heeft in het algemeen het format $\varphi(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$. In de praktijk van dit sleuteluitwisselprogramma kunnen we bovengenoemd simpele format hanteren.) De *kern* van $\varphi$ is $\{P = (x_0, y_0) \mid v(x_0) = 0\} \cup O$. De *graad* van $\varphi$ is $\max\{\deg(u(x)), \deg(v(x))\}$.

Voor ons is de stelling van Tate van belang, die zegt dat als $\mathcal{E}$ en $\mathcal{E}'$ twee elliptische krommen zijn, gedefinieerd over $k$, en bovendien geldt $\#\mathcal{E} = \#\mathcal{E}'$, dan geldt dat er een isogenie bestaat tussen $\mathcal{E}$ en $\mathcal{E}'$. Merk op dat de groepsstructuur niet gelijk hoeft te zijn. Zo kan de ene groep isomorf zijn met $\mathbf{C}_{4N}$ en de andere groep isomorf zijn met $\mathbf{C}_{2N} \times \mathbf{C}_2$. Bovendien zegt deze stelling nog niets over hoe de isogenie eruitziet.

Een andere stelling doet de volgende uitspraak over isogenieën. Laat $\overline{k}$ de algebraïsche afsluiting van $k$ zijn. We bekijken een isogenie $\varphi : \mathcal{E} \to \mathcal{E}'$ over $\overline{k}$. De kern van $\varphi$, de punten op $\mathcal{E}$ die op $0$ van $\mathcal{E}'$ worden afgebeeld, vormt een ondergroep, zeg $G$. De stelling zegt nu dat er voor elke $G \subset \mathcal{E}$ een isogenie $\varphi_G$ en een elliptische kromme $\mathcal{E}'$ bestaan zodanig dat $\varphi_G : \mathcal{E} \to \mathcal{E}'$ kern $G$ heeft. Het blijkt dat de graad van de isogenie $\varphi$ gelijk is aan het aantal elementen van $G$. De kromme $\mathcal{E}'$ wordt ook wel genoteerd als $\mathcal{E}/G$.

Het omgekeerde geldt slechts tot op zekere hoogte. Er kunnen meerdere ellip-

tische krommen $\mathcal{E}'$ bestaan zo dat de kern van de isogenie van $\mathcal{E}$ op deze krommen $\mathcal{E}'$ steeds de groep $G$ is. Stel dat $\mathcal{E}'$ en $\mathcal{E}''$ twee dergelijke elliptische krommen zijn (met isogenieën $\varphi$ en $\phi$), dan geldt echter wel dat er een isomorfisme $\psi$ bestaat, $\psi : \mathcal{E}' \to \mathcal{E}''$. In een plaatje hebben we de volgende situatie:

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\varphi} & \mathcal{E} \\ & \searrow^{\phi} & \downarrow{\scriptstyle \psi} \\ & & \mathcal{E}'' \end{array}$$

$\mathcal{E}'$ en $\mathcal{E}''$ zijn niet identiek maar wel isomorf. Dat houdt in dat er ook een inverse bestaat, ofwel dat er een isogenie tussen $\mathcal{E}'$ en $\mathcal{E}''$ bestaat van graad $1$. En dat betekent dat beide krommen dezelfde $j$-*invariant* hebben. De $j$-invariant kan eenvoudig worden berekend,

$$j(\mathcal{E}) = 1728 \cdot \frac{4a^3}{4a^3 + 27b^2}.$$

Dankzij het werk van Vélu (1965) kan de isogenie $\varphi_G$ expliciet worden berekend (zie bijvoorbeeld Sutherland [12, Lecture 6, p. 6]). Deze berekening is echter vrij lastig. Er zijn verbeteringen bedacht door Kohel [8] en Shumow [11] in hun proefschriften. In het kader op de volgende pagina heb ik de door Shumow opgestelde berekening van de isogenie opgeschreven. Het komt er op neer dat je een groep $G$ formuleert (een verzameling van punten) en vervolgens voor elk van deze punten een bepaalde waarde van een bepaalde functie bij elkaar optelt. Dat betekent dat naarmate $G$ groter wordt, de berekening lastiger wordt.

Een andere waardevolle stelling, waarop het gehele sleuteluitwisselprogramma is gebaseerd, zegt dat er een commutatieve eigenschap bestaat voor isogenieën: $\varphi_{G_A} \circ \varphi_{G_B} \cong \varphi_{G_B} \circ \varphi_{G_A}$. Ofwel, het linker- en rechterlid zijn isomorf. Mogelijk (in de praktijk zeer waarschijnlijk) gaat het om verschillende krommen, maar met gelijke $j$-invariant.

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\varphi_{G_A}} & \mathcal{E}/G_A \\ \downarrow{\scriptstyle \varphi_{G_B}} & & \downarrow{\scriptstyle \psi} \\ \mathcal{E}/G_B & \xrightarrow{\eta} & \mathcal{E}/\langle G_A, G_B \rangle \end{array}$$

Dit diagram blijkt commutatief. Dus door eerst $\mathcal{E}/G_A$ en vervolgens $(\mathcal{E}/G_A)/G_B$ uit te voeren verkrijgen we hetzelfde resultaat als door eerst $\mathcal{E}/G_B$ en daarna $(\mathcal{E}/G_B)/G_A$ uit te voeren.

**Het sleuteluitwisselprogramma**

Het basisidee van het sleuteluitwisselprogramma is dit commutatieve diagram. Het idee is dat Alice een groep $G_A$ van een speciale vorm bepaalt en daaruit $\mathcal{E}/G_A$ berekent. Evenzo berekent Bob $\mathcal{E}/G_B$. Het idee is verder dat het voor iemand die wil afluisteren, laten we zeggen Eve, qua rekenkracht lastig is om uit $\mathcal{E}/G_A$ de groep $G_A$ te reconstrueren, en evenzo $G_B$. Alice en Bob wisselen $\mathcal{E}/G_A$ en $\mathcal{E}/G_B$ uit zonder daarbij respectievelijk $G_A$ en $G_B$ prijs te geven. Vervolgens gaat Bob verder met $\mathcal{E}/G_A$ en bepaalt $(\mathcal{E}/G_A)/G_B$. Evenzo bepaalt Alice $(\mathcal{E}/G_B)/G_A$. Beiden vinden ze $\mathcal{E}/\langle G_A, G_B \rangle$.

Hier komen we tot de kern van het reductieprobleem. De wetenschappers die zich met de isogenie bezighouden gaan ervan uit dat Eve hier een forse uitdaging heeft, maar tot op heden is nog niet bekend tot welk probleem deze problematiek kan worden gereduceerd. (De term *reductieprobleem* wordt gebruikt om te bewijzen of een zekere vercijfering veilig is. Als $\mathcal{B}$ veilig is en je kunt bewijzen dat $\mathcal{A}$ kan worden *gereduceerd* tot $\mathcal{B}$, dan is $\mathcal{A}$ eveneens veilig.)

**Keuze van $p$ en $\mathcal{E}$**

Tot nog toe is er nog niets gezegd over de karakteristiek $p$, en de grootte van het lichaam $\mathcal{F}_q$.

De keuze van het priemgetal is $p = \lambda \cdot \ell_A^a \cdot \ell_B^b - 1$. Hierin zijn $\ell_A$ en $\ell_B$ kleine priemgetallen en $\lambda$ een kleine factor om te zorgen dat $p$ een priemgetal wordt $\equiv 3 \bmod 4$. Verder geldt $\ell_A^a \approx \ell_B^b$. Het lichaam waarover we werken is $\mathcal{F}_{p^2}$. Dit lichaam kan worden gerepresenteerd door $\mathcal{F}_p[x]/(x^2+1) \cong \mathcal{F}_p(i)$. De elliptische kromme die gekozen wordt is $\mathcal{E} : y^2 = x^3 + 1$. Dit is een *supersinguliere* elliptische kromme. Deze $\mathcal{E}$ bevat $(p+1)^2$ punten, en in het bijzonder geldt dat $\mathcal{E} \cong (\mathbb{Z}/\ell_A^a\mathbb{Z})^2 \times (\mathbb{Z}/\ell_B^b\mathbb{Z})^2$.

Conform het voorstel van de auteurs [5], beschouw ik $\ell_A = 2$ en $\ell_B = 3$. Vanaf nu kiezen we $\lambda = 1$. Heel belangrijk is het om kennis te nemen van het feit dat het berekenen van een willekeurige isogenie weliswaar mogelijk is, maar tevens tijdrovend is. Echter door de structuur van de elliptische kromme en enige kennis van $G$ kan de isogenie aanzienlijk eenvoudiger worden berekend. Zonder in details te treden poneer ik hier dat als $\#G = 2^u 3^v$, dan zijn er $u$ isogenieën $\varphi_i$ van graad $2$ en $v$ isogenieën $\phi_j$ van graad $3$, zodanig dat $\varphi_G = \varphi_1 \circ \varphi_2 \circ \varphi_3 \circ \cdots \circ \varphi_u \circ \phi_1 \circ \phi_2 \circ \phi_3 \circ \cdots \circ \phi_v$, hetgeen het rekenwerk aanzienlijk reduceert.

**In concreto**

We laten zien hoe Alice en Bob samen een gemeenschappelijke sleutel kunnen construeren die niet door een buitenstaander gevonden kan worden.

1. Kies een priem $p$ van de vorm $p = 2^a \cdot 3^b - 1$, waarbij $2^a \approx 3^b$, en $p \equiv 3 \bmod 4$.
2. Het lichaam waarover we werken is $\mathcal{F}_{p^2}$. Dit lichaam kan worden gerepresenteerd door $\mathcal{F}_p[x]/(x^2+1) \cong \mathcal{F}_p(i)$.
3. Er geldt $\mathcal{E} \cong (\mathbb{Z}/2^a\mathbb{Z})^2 \times (\mathbb{Z}/3^b\mathbb{Z})^2$. Kies punten $\mathsf{P}_A, \mathsf{Q}_A, \mathsf{P}_B, \mathsf{Q}_B$ zodanig dat $\mathrm{ord}(\mathsf{P}_A) = \mathrm{ord}(\mathsf{Q}_A) = 2^a$, $\mathrm{ord}(\mathsf{P}_B) = \mathrm{ord}(\mathsf{Q}_B) = 3^b$ en $\mathcal{E}$ wordt precies voortgebracht door $\mathsf{P}_A, \mathsf{Q}_A, \mathsf{P}_B$ en $\mathsf{Q}_B$.
4. Alice kiest random $m_A$ en $n_A \in \{1, \ldots, 2^a\}$ en berekent $R_A = m_A \cdot \mathsf{P}_A + n_A \cdot \mathsf{Q}_A$. $G_A$ is de groep die wordt voortgebracht door $R_A$. De groep $G_A$ impliceert de isogenie $\varphi_{G_A}$. Alice berekent $\mathsf{P}_{(A)B} = \varphi_{G_A}(\mathsf{P}_B)$, $\mathsf{Q}_{(A)B} = \varphi_{G_A}(\mathsf{Q}_B)$ en $\mathcal{E}_A \cong \mathcal{E}/G_A$ met behulp van isogenieën. Merk op dat de isogenie $\varphi_{G_A}$ lastig te berekenen is als $G_A$ groot is.
5. Alice maakt $\mathsf{P}_{(A)B}$ en $\mathsf{Q}_{(A)B}$ bekend, ten behoeve van Bob.
6. Bob idem $m_B$, $n_B$, $R_B$, $G_B$, $\mathsf{P}_{(B)A}$, $\mathsf{Q}_{(B)A}$, $\mathcal{E}_B$.
7. Alice ontvangt $\mathsf{P}_{(B)A}$ en $\mathsf{Q}_{(B)A}$ van Bob. Zij vervolgt haar berekening. Ze berekent eerst $R_{(B)A} = m_A \cdot \mathsf{P}_{(B)A} + n_A \cdot \mathsf{Q}_{(B)A}$. Laat $G_{(B)A}$ de groep zijn voortgebracht door $R_{(B)A}$. Deze groep impliceert een isogenie, zeg $\varphi_{G_{(B)A}}$. Vervolgens kan zij $\mathcal{E}_{(B)A} \cong \mathcal{E}_A/G_{(B)A}$ bepalen. Ten slotte bepaalt Alice $j_{(B)A}$.
8. Bob idem $R_{(A)B}$, $G_{(A)B}$, $\varphi_{G_{(A)B}}$, $\mathcal{E}_{(A)B}$, $j_{(A)B}$.
9. Aangezien

   $$\mathcal{E}_{(B)A} \cong \mathcal{E}_A/G_{(B)A} \cong \mathcal{E}_B/G_{(A)B} \cong \mathcal{E}_{(A)B}$$

   volgt $j = j_{(B)A} = j_{(A)B}$.

De in stap 9 gevonden $j$ kan worden gebruikt als een sleutel voor een geheim sleuteluitwisselprogramma.

**Afmetingen sleuteluitwisselprogramma**

In de cryptografie is het gebruikelijk om aan te geven hoe veilig je programma is. Andere cryptografen kunnen dit vergelijken met andere programma's. Het gaat veel te

ver om deze veiligheid nauwkeurig toe te lichten, maar belangrijk is vooral dat je nagaat hoeveel rekenkracht een aanvaller van het systeem nodig zal hebben. $128$ bits houdt in dat de aanvaller hooguit $2^{128}$ mogelijkheden moet nagaan. Voor de isogenieën over supersinguliere elliptische krommen suggereren de auteurs [5] voor een veiligheid van $128$ bits de volgende keuzes van $p$, $a$ en $b$:

- $p = 2^a \cdot 3^b - 1 \approx 2^{512}$.
- In het geval dat $\mathcal{E}$ een supersinguliere elliptische kromme is, geldt $\#\mathcal{E} = 2^{2a} \cdot 3^{2b} \approx 2^{1024}$.
- Het totaal aantal isomorfie-klassen van supersinguliere elliptische krommen is $\frac{p+1}{12} = 2^{a-2} \cdot 3^{b-1} \approx 2^{508}$.

Het begrip kwantumveiligheid is gerelateerd aan de veiligheid nadat ook aanvallers in het bezit zijn van kwantumcomputers. Kunnen deze aanvallers efficiënter een programma aanvallen met een kwantumcomputer? Dit is een uitermate interessant onderwerp, maar ongeschikt om hier even te behandelen.

Voor de kwantumveiligheid van $128$ bits suggereren de auteurs $p \approx 2^{768}$ (en daarmee groeien ook $a$ en $b$). Een sleuteluitwisseling zou kunnen worden uitgevoerd in $50$ ms op een simpele computer.

**Algoritme van Vélu**

We gaan hier uit van de elliptische kromme $\mathcal{E}: y^2 = x^3 + ax + b$. Tevens is er een groep $G \subset \mathcal{E}$ gegeven. We bepalen nu de elliptische kromme $\mathcal{E}': y^2 = x^3 + Ax + B$, zodanig dat $\mathcal{E}' = \mathcal{E}/G$. Als deze isogenie wordt weergegeven met $\phi$, dan wordt tevens het beeld van een punt $P \in \mathcal{E}$, $\phi(P) \in \mathcal{E}'$ bepaald. U zult zien dat het rekenwerk enorm zal toenemen naarmate $|G|$ groeit.

Zij gegeven $G \subset \mathcal{E}$ een groep en $P \in \mathcal{E}$ een punt.

*Stap 1.* We splitsen $G$ op in vier deelverzamelingen:

$$G = \{0\} \cup C_2 \cup R^+ \cup R^-.$$

Hierin is $0$ de oorsprong (punt op oneindig), $C_2$ de $2$-torsiepunten, ofwel punten $T$ met $2T = 0$. Voor $R^+$ en $R^-$ geldt dat de punten in deze verzamelingen tegengestelde $y$-coördinaat hebben. Het maakt niets uit hoe de verdeling wordt gemaakt, want uiteindelijk wordt alleen gebruik gemaakt van $y^2$. Laat $S = C_2 \cup R^+$.

*Stap 2.* Zij $Q = (x_Q, y_Q) \in S$.

$$\begin{aligned} g_Q^x &= 3x_Q^2 + a, \\ g_Q^y &= -2y_Q, \\ u_Q &= \left(g_Q^y\right)^2, \\ v_Q &= \begin{cases} g_Q^x Q \in C_2, \\ 2g_Q^x Q \in R^+. \end{cases} \end{aligned}$$

*Stap 3.* Sommeer voor alle $Q \in S$:

$$\begin{aligned} v &= \sum_{Q \in S} v_Q, \\ w &= \sum_{Q \in S} u_Q + x_Q v_Q. \end{aligned}$$

*Stap 4.* Bepalen van $\mathcal{E}' = \mathcal{E}/G$:

$$\begin{aligned} A &= a - 5v, \\ B &= b - 7w. \end{aligned}$$

Er geldt: $\mathcal{E}': Y^2 = X^3 + AX + B$.

*Stap 5.* Ten slotte bepalen we $\varphi_G(P) = (X, Y)$. Er geldt:

$$\begin{aligned} X &= x + \sum_{Q \in S} \left( \frac{v_Q}{x - x_Q} - \frac{u_Q}{(x - x_Q)^2} \right), \\ Y &= y - \sum_{Q \in S} \left( \frac{2yu_Q}{(x - x_Q)^3} + \frac{v_Q(y - y_Q)}{(x - x_Q)^2} - \frac{g_Q^x g_Q^y}{(x - x_Q)^2} \right). \end{aligned}$$

(bron: proefschrift van Shumow [11, p. 32])
Voor de Sage-gebruiker is het niet nodig om deze formules zelf te implementeren. De $v$ en $w$ in stap 3 zijn te bepalen met respectievelijk `EllipticCurveIsogeny__v` en `EllipticCurveIsogeny__w`.

**Referenties**


1 Daniel Bernstein en Tanja Lange, Post-quantum cryptography – dealing with the fallout of physics success (2017), eprint.iacr.org/2017/314.pdf.

2 Denis Charles, Eyal Goren en Kristin Lauter, cryptographic hash functions from expander graphs (2006), eprint.iacr.org/2006/021.pdf.

3 Christina Delfs en Steven Galbraith, Computing isogenies between supersingular elliptic curves over $\mathcal{F}_p$ (2013), arXiv/1310.7789.

4 Luca de Feo, Cyril Hugounenq, Jérôme Flût, en Éric Schost, Explicit isogenies in quadratic time in any characteristic (2016), arXiv/1603.00711.

5 Luca de Feo, David Jao en Jérôme Flût, Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies, PQCrypto 2011, *Journal of Mathematical Cryptology*, 8(3) (2014) 209–247.

6 Steven Galbraith en Anton Stolbunov, Improved algorithm for the isogeny problem for ordinary elliptic curves (2011), arXiv/1105.6331.

7 David Jao, Isogenies in a quantum world (Slides), ecc2011.loria.fr/slides/jao.pdf.

8 David Kohel, Endomorphism rings of elliptic curves over finite fields (1989), echidna.maths.usyd.edu.au/kohel/pub/thesis.pdf.

9 John Proos en Christof Zalka, Shor's discrete logarithm quantum algorithm for elliptic curves (2003), arXiv.quant-ph/0301141.

10 Alexander Rostovtsev en Anton Stolbunov, Public-key cryptosystem based on isogenies (2006), eprint.iacr.org/2006/145.pdf.

11 Daniel Shumow, *Isogenies of Elliptic Curves, a Computational Approach*, thesis, 2009.

12 Andrew Sutherland, Elliptic Curves, college-dictaten, ocw.mit.edu/courses/mathematics/18-783-elliptic-curves-spring-2015/lecture-notes, 2015.

13 Henk van Tilborg, Elliptic curve cryptosystems; too good to be true?, *Nieuw Archief voor Wiskunde* 5/2(3) (2001), 220–225.

Casper Albers
*Psychometrie & Statistiek*
*Rijksuniversiteit Groningen*
*c.j.albers@rug.nl*

**Column** Casper sees a chance

# The statistician Alan Turing

**Caspers Albers regularly writes a column on everyday statistical topics in this magazine.**

This special issue of *Nieuw Archief voor Wiskunde* is devoted to cryptography and it shouldn't therefore be surprising that I devote my column to Alan Turing, probably the most well-known code breaker. Apart from playing a pivotal role in cryptography as well as being one of the founding fathers of computing science, Turing made some important contributions to the theory of Bayesian statistics.

About ten years ago, I was a research fellow at the Open University in Milton Keynes, England. My room was located in a building with the uninspiring name 'M building' and my fellowship was in the department of 'Mathematics, Statistics and Computing'. Very few people made significant contributions to all three of these fields. Even fewer did that in the vicinity of (what is now) Milton Keynes. Only one person ticked both these boxes and saved millions of lives whilst doing so: Alan Turing. After I left the Open University, they renamed the M building after him (I don't think these two events are related).

Turing is well known for his contributions to cryptography during the Second World War, working with other scientists at Bletchley Park and cracking the German Enigma machine. According to Churchill, their work caused the war to end years earlier, thus saving millions of lives. He's one of the handful of mathematicians to have an Oscar-winning movie (*The Imitation Game*) based on his live. His work on the foundations of computing — most notably his papers describing what is now called the Turing Machine [6] and the Turing Test [7] — has been well documented. His contributions to statistics, however, are less well known, in part because this work consisted of classified documents.

During the war, Turing wrote two reports on probability and statistics [8, 9], but these remained classified by GCHQ until 2012, seventy years later. Both before and after the release of these documents in 2012, various scientists, including Turing's direct colleagues Edward Simpson (from Simpson's paradox) and Jack Good (known for his work on Bayesian analysis), have reflected on his work in statistics [1, 3, 5]. Two of the topics he worked on, Bayes factors and sequential analysis, have become standard equipment in the Bayesian statistician's toolbox.

**Bayes factors**

In null hypothesis significance testing (NHST) the objective is to decide between two competing hypotheses on some parameter of interest. These hypotheses can be simple (e.g. $\mathrm{H}_0\colon \theta = \theta_0$) or composite (e.g. $\mathrm{H}_1\colon \theta \in \Theta_1$, with $\theta_0 \notin \Theta_1$). Jerzy Neyman and Egon Pearson proved in 1933 that the likelihood ratio approach is uniformly most powerful when both hypothesis are simple, $\mathrm{H}_0\colon \theta = \theta_0$ and $\mathrm{H}_1\colon \theta = \theta_1$, but the approach can also be used when the alternative hypothesis is composite. As the name suggests, the likelihood ratio test looks at the ratio of the likelihood under the null distribution and the (maximum) likelihood under the alternative. When this ratio falls below a pre-specified threshold $c$, the null hypothesis is rejected in favour of the alternative. The likelihood function is given by $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$, and the likelihood ratio is

$$\lambda = \frac{L(\theta_0)}{\max_{\theta\in\Theta_1} L(\theta)}.$$

**Example.** Let $\underline{x} = x_1,\ldots,x_n$ be a random sample from the $N(\mu,1)$ distribution. Suppose we are interested in $\mathrm{H}_0\colon \mu = 0$ versus $\mathrm{H}_1\colon \mu \neq 1$. The likelihood function is given by

$$\begin{aligned} L(\mu \mid \underline{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x_i-\mu)^2} \\ &= (2\pi)^{-\frac{n}{2}} \mathrm{e}^{-\frac{n}{2}(\bar{x}-\mu)^2} \times \mathrm{e}^{-\frac{1}{2}\Sigma_{i=1}^{n}(x_i-\bar{x})^2}. \end{aligned}$$

Under $\mathrm{H}_1$, this function is maximised when the maximum likelihood estimator $\hat{\mu} = \bar{x}$ is used. The likelihood ratio can be simplified into

$$\lambda = \frac{L(\theta_0)}{L(\hat{\mu})} = \mathrm{e}^{-\frac{n}{2}(\bar{x}-\mu_0)^2}.$$

Thus, when $n(\bar{x}-\mu_0)$ is larger than a certain value, $\lambda < c$ and the test will reject $\mathrm{H}_0$ in favour of $\mathrm{H}_1$. This frequentist notion of likelihood ratios is turned into a Bayesian one by taking prior information into account. Bayes' Theorem dictates that

$$\underbrace{\frac{\mathrm{P}(\mathrm{H}_0 \mid \underline{x})}{\mathrm{P}(\mathrm{H}_1 \mid \underline{x})}}_{\text{Posterior odds}} = \underbrace{\frac{\mathrm{P}(\underline{x} \mid \mathrm{H}_0)}{\mathrm{P}(\underline{x} \mid \mathrm{H}_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{\mathrm{P}(\mathrm{H}_0)}{\mathrm{P}(\mathrm{H}_1)}}_{\text{Prior odds}}.$$

Thus, the Bayes factor is the ratio of *marginal* likelihoods

$$\mathrm{P}(\underline{x} \mid \mathrm{H}_i) = \int_{\Theta_i} \underbrace{\mathrm{P}(\underline{x} \mid \theta, \mathrm{H}_i)}_{\text{Likelihood}} \times \underbrace{\mathrm{P}(\theta \mid \mathrm{H}_i)}_{\text{Prior}} \,\mathrm{d}\theta \quad (i = 0,1)$$

Photo credits: Antoine taveneaux, Wikimedia Commons.

Statue of Alan Turing at Bletchley Park

and as such quantifies the evidence in the data for $\mathrm{H}_0$ relative to $\mathrm{H}_1$. When both $\mathrm{H}_0$ and $\mathrm{H}_1$ are simple, the Bayes factor approach is identical to the likelihood ratio approach.

In contrast to the frequentist approach, the Bayesian approach is symmetrical: it can be used to reject $\mathrm{H}_0$ in favour of $\mathrm{H}_1$, but it can also be used to reject $\mathrm{H}_1$ in favour of $\mathrm{H}_0$. Furthermore, the ability to incorporate initial knowledge, or even initial guesstimates, through the prior odds, provide important benefits. Under mild conditions, the Bayesian alternative was easier to compute than the frequentist alternative, an important advantage in the era before modern computing.

Bayes factors thus are a method to quantify the (relative) weight of evidence of two competing hypotheses. Together with Jack Good, Turing introduced the terminology 'deciban' for the units in which weight of evidence was measured. The base-$10$ logarithm of the Bayes Factor is measured in decibans. Turing and Good's measure of evidence predates the, now more popular, 'bit' by Shannon and Tukey by about eight years. Bayes factors were already known shortly before the war [2], but according to Good [1], Jeffrey's approach lacked the appealing terminology that Turing's approach did have.

**Sequential analysis**

For Turing, the logarithms of the Bayes Factors were natural ingredients in the sequential analyses. In sequential analysis, the sample size is not fixed in advance: the data are evaluated continuously as they are collected. In a frequentist context, derived by Abraham Wald [10], the usual approach after each new observation is to choose between three alternatives: (1) accept $\mathrm{H}_0$, (2) reject $\mathrm{H}_0$, (3) remain unsure and continue sampling. Deciding between (1), (2) or (3) occurs based on whether a certain likelihood ratio is either below $c_1 < 1$, above $c_2 > 1$ or between $c_1$ and $c_2$.

Independently of Wald and his colleagues at Columbia University, Turing derived a Bayesian reasoning for sequential analysis. By taking the prior odds for observation $k+1$ equal to the posterior odds after observation $k$, and by applying Bayes Factors in each step, he derived a sequential process for updating his degrees of belief. This empirical Bayesian approach is now common within Bayesian statistical research. It has been shown that in many common situations this Bayesian approach coincides with Wald's approach. Even though "Wald gave [sequential analysis] useful applications that were not anticipated by Turing" [1], Turing played an important role in the development of sequential analysis.

He used his sequential conditional probabilities in his work on Banburismus, an automated process to find the most likely settings of the German Enigma machine. This Banburismus was a predecessor of a modern day computing algorithm. In her popular scientific book, McGrayne [4] describes how this system enabled Turing to make a guess of a series of letters in an Enigma message and then to measure his degree of belief in this guess. Subsequently, as additional information came in — e.g. in the form of additionally intercepted Enigma messages — the degree of belief would be updated and the initial guess would be replaced by a guess with a higher degree of belief.

**Polymath**

With the general public — and probably even within mathematicians — Alan Turing is much better known for his other activities than for his statistical work. However, he deserves a place in a line of polymaths, statisticians that are also well known for contributions to other fields such as Ronald Fisher, Francis Galton and Karl Pearson. The impact of Turing's statistical work has been undervalued, in part due to that many of his contributions were designated as classified information and not released until a couple of years ago. It is the task of the 21st century statistician to correct this omission in the recollection of 20th century statistics.

**References**


1 I.J. Good, Studies in the History of Probability and Statistics. XXXVII A.M. Turing's statistical work in World War II. *Biometrika*, 66(2) (1979), 393–396.

2 H. Jeffreys, *Theory of Probability*, Clarendon Press, Oxford, 1939.

3 K.V. Mardia and S. Barry Cooper, Alan Turing and enigmatic statistics, in: S. Barry Cooper and A. Hodges, eds., *The Once and Future Turing – Computing the World,* Cambridge University Press, 2016, Chapter 5.

4 S.B. McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy,* Yale University Press, 2012.

5 E. Simpson, A life in statistics – Edward Simpson: Bayes at Bletchley Park, *Significance*, 18 May 2010.

6 A.M. Turing, On computable numbers, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society* 42(2) (1936).

7 A.M. Turing, Computing machinery and intelligence, *Mind* 59 (1950), 433–460.

8 A.M. Turing, Paper on statistics of repetitions (1941), ArXiv:1505.04715v2.

9 A.M. Turing, The applications of probability to cryptography (1942). ArXiv: 1505/04714v2.

10 A. Wald, Sequential Tests of Statistical Hypotheses, *The Annals of Mathematical Statistics* 16(2) (1945), 117–186.

Jan Beuving
*cabaretier, Zeist*
*janbeuving@gmail.com*

**Het keerpunt van** Maaike Hersevoort

# Wiskunde is niet bedoeld om de wereld beter te begrijpen

**Maaike Hersevoort volgde een TWIN-studie wis- en natuurkunde in Utrecht en belandde uiteindelijk bij het Centraal Bureau voor de Statistiek, eerst in Den Haag en later in Heerlen. Maar vanaf oktober zit ze in Congo-Kinshasa, kortweg Congo, voor Artsen Zonder Grenzen.**

*De eerste vraag die een wiskundige dan natuurlijk stelt is: zijn dat open of gesloten grenzen?*
"Open grenzen natuurlijk! Hoewel: we zijn zónder grenzen, dus het is correcter om te zeggen dat ze niet zijn gedefinieerd."

*Wat ga je doen in Congo?*
"Ik word projectcoördinator. Artsen Zonder Grenzen werkt in projecten, vaak miniziekenhuisjes die een regio bedienen. Ik zeg ziekenhuis, maar het is echt heel basaal. We hebben ook een mobiele post die de afgelegen dorpen bezoekt. Ik krijg één zo'n project onder mijn hoede. Als projectcoördinator ben je hoofd van het project. Je praat dus met de mensen die bij het project betrokken zijn. Dat is maar een beperkt aantal expats zoals ik — het zijn vooral Congolezen. Het is de filosofie van AZG om zo veel mogelijk met lokale mensen te werken. Ik praat ook veel met de mensen die er wonen, om hun behoeften en noden in te schatten. Ik ga nu aan het werk in Zuid-Kivu, dat ligt in het oosten van Congo, in een plaats die Baraka heet. Baraka betekent 'zegen', wat een wrange naam is voor een gebied waar zoveel aan de hand is."

*Is het een gevaarlijk land?*
"Vooral voor de mensen daar. De overheid is er niet goed in staat om voor haar inwoners te zorgen. Er zijn veel conflicten tussen groepen onderling. Voor ons valt het mee; wij kunnen uitleggen waarom we daar zijn. Het is heel duidelijk dat wij medische zorg verlenen, vaak aan vrouwen en kinderen. Strijders die een gebied bewaken zien in dat het ook hun vrouwen en kinderen zijn die wij helpen."


Maaike Hersevoort

*Je bent eerder voor AZG in Congo geweest. Wilde je terug?*
"Ik had er niet om gevraagd, maar toen dit aan me werd voorgesteld, wilde ik het graag doen. Congo zit in mijn hart. Het is een prachtig land, qua natuur en omgeving, en de mensen zijn er geweldig. Het is onvoorstelbaar hoe zij — maar ik heb dat in andere landen precies zo meegemaakt — in hun situatie toch zoveel levensvreugde uitstralen. Je moet je voorstellen dat ze vaak geen drinkwater en medicijnen hebben, maar zelfs basisveiligheid is er niet. Als ze 's avonds gaan slapen weten ze niet of dat in een veilig huis is. De gemiddelde levensverwachting in het land is laag, hoewel dat grotendeels komt door de hoge kindersterfte. Ondanks dat stralen ze een enorme kracht uit. Het is iets bizars hoe mensen aan alles kunnen wennen. Ik voerde ooit sollicitatiegesprekken met Congolezen voor een project. Dan vraag je naar een school die niet is afgemaakt, en hoor je 'oh, toen werd mijn dorp aangevallen en moesten we vluchten.' Alsof het de normaalste zaak van de wereld is."

*Heb je altijd al de drang gehad om dit werk te doen?*
"Dan was ik wel wat anders gaan studeren, haha! Maar, hoewel ik het vroeger nooit zo

zou hebben omschreven: ik heb altijd al het geloof gehad dat we als mensen meer op elkaar lijken dan dat we verschillend zijn. Maar waar je wordt geboren maakt zoveel uit voor de kansen die je krijgt. Daardoor ontstaan grote verschillen, en ik wil mij voor die verschillen inzetten."

*Maar je ging dus eerst wis- en natuurkunde doen.*
"Dat heeft me altijd heel erg geïnteresseerd. Hoe zijn dingen ontstaan? Hoe werkt het heelal? Op de middelbare school las ik al heel veel populair-wetenschappelijke boeken. En ik had een wiskundedocent, die overigens ook natuurkunde gaf, die voor belangstellenden één uur in de week speciale relativiteitstheorie gaf. Dat vond ik geweldig. Ik heb lang getwijfeld over een studie theologie — ik ben er ook aan begonnen, in deeltijd. Maar uiteindelijk werd het dus TWIN. Vooral op basis van de natuurkunde, maar ik vond wiskunde ook heel leuk, en al snel eigenlijk interessanter dan natuurkunde."

*Waarom?*
"Natuurkunde is heel leuk omdat je iets leert over de wereld. Althans, dat dacht ik toen nog. Maar het blijft altijd een model. Een model waar je heel veel mee kunt, maar of je ook iets leert over de échte werkelijkheid, weten we niet. En je leert bij natuurkunde niet iets wat echt belangrijk is. Als je dood bent, wil je dan weten hoe de wereld ontstaan is? Of is het belangrijk dat je dan fijne vriendschappen had en een goed mens bent geweest?"

*Maar dat argument kun je tegen de wiskunde ook gebruiken.*
"Ja, maar de wiskunde is niet bedoeld om de wereld beter te begrijpen. Het is een op zichzelf staande wereld, helemaal waar, gebouwd op premisses. Dingen zijn waar of niet waar. Er is geen grijs. Hoewel, ik weet niet of dat helemaal waar is."

*Waarom koos je, met deze voorliefde voor wiskunde, toch voor afstudeeronderzoek in de natuurkunde?*
"Dat had twee redenen. Ten eerste vond ik zuivere wiskunde lastig. Ik kon het wel volgen, maar de volgende stap was moeilijk. Daarnaast had ik mijn klein onderzoek — een onderdeel van de natuurkundeopleiding — gedaan bij Fysica van de Mens. Dat was concreter, en dat vond ik fijn. Hoewel ik mijn afstudeeronderzoek daarna weer aan het CERN deed, bij Hogere Energie Fysica. Ik heb meegewerkt aan het detecteren van vraag-me-niet-meer-welk-deeltje. Maar dat was eigenlijk heel wiskundig werk: ik kreeg een berg data, en daar moest ik de ruis uithalen, en het signaal reconstrueren.

Ik ben nog begonnen aan een promotie bij Fysica van de Mens, maar die heb ik na zes maanden afgebroken. Heel erg diep inzoomen op één heel klein onderwerp ligt me minder. Ik houd van de wijdere blik. Wat al bleek uit mijn vakkenpakket op de middelbare school — naast de bètavakken had ik Latijn, Grieks en Geschiedenis."

*Maar geen Frans dus? Terwijl je dat in Afrika volop spreekt.*
"Dat klopt, ik heb dat later geleerd. Hun Frans is overigens wel anders. Tijdens een eerder AZG-project in de Centraal Afrikaanse Republiek kreeg ik een Franse collega, maar ze kwamen mij vragen wat hij allemaal zei; ze verstonden hem amper. Hij praatte te snel en gebruikte andere woorden."

*Jullie heten Artsen Zonder Grenzen, maar ik neem aan dat grensproblematiek een grote rol speelt?*
"De landsgrenzen niet zozeer, hoewel we natuurlijk veel in gebieden werken waar mensen over zo'n grens zijn gevlucht. De informele grenzen zijn veel belangrijker. Je kijkt vooral naar de actieradius van je eigen project. Soms moet je daarvoor in regio's zijn die door verschillende groepen worden gecontroleerd. Dan moet je onderhandelen."

*Hoe moet ik me dat voorstellen? Spreek jij dan met mannen die het geweer in de hand hebben?*
"Soms staan die geweren wel in dezelfde ruimte, ja."

*Accepteren ze jouw autoriteit dan, als blanke, vrouwelijke projectleider?*
"Ik had verwacht dat dat veel problemen zou geven. Emancipatie staat er in de kinderschoenen, hoewel het langzaam komt. Maar ik heb er nooit last van gehad. Ik had ook wel wrok verwacht jegens blanken, als gevolg van de koloniale tijd, maar ook daarvan merk ik weinig. Misschien omdat we onafhankelijk zijn. De Verenigde Naties zijn in Congo bijvoorbeeld niet heel geliefd. Maar wij zijn niet van overheden afhankelijk voor onze projectgelden."

*Je verlaat voor de tweede keer het CBS. Wat voor werk deed je daar?*
"In Den Haag werkte ik bij beleidsstatistiek. Daar beantwoordden we statistische onderzoeksvragen van ministeries en gemeenten. In mijn tweede periode in Heerlen was ik steekproefdeskundige. Ik maakte de opzet voor enquêtes, en bepaalde hoeveel mensen er geïnterviewd moesten worden voor een representatief resultaat. Ik deed ook optimalisatiewerk — ik bepaalde hoe de interviewers zo efficiënt mogelijk langs alle bezoekadressen konden gaan. Een handelsreizigersprobleem."

*Toch klink je alsof je bij Artsen Zonder Grenzen meer op je plek bent.*
"Ja. Het is de meest uitdagende en afwisselende baan die ik ooit heb gehad. Je ziet heel direct wat je doet: we redden levens. Bovendien is het met veel verschillende culturen — de expats komen overal vandaan — in een ander land. De saamhorigheid is groot. En het is heel analytisch werk. Er zijn heel veel issues, die je heel snel moet analyseren en prioriteren. Een exacte achtergrond is dan heel fijn."

*Is het een roeping?*
"Dat vind ik altijd een moeilijk woord. Alsof het een opoffering is. Zo voelt het helemaal niet. Ik vind het vooral heel fijn om bij te dragen aan een rechtvaardigere wereld — hoewel dat wel grote woorden zijn. Bovendien kan dat hier ook. Je hoeft de grens niet over om grenzeloos goed te doen. Het werk dat ik daar doe is niet belangrijker dan het werk van mensen die in Nederland voor pleegkinderen zorgen, of naar oude buren omzien. Ik doe bijzonder werk, maar het is niet specialer dan wat hier gebeurt.

Goede suggesties voor een Nederlandse wiskundige met een keerpunt in zijn of haar carrière zijn welkom via keerpunt@nieuwarchief.nl.

Wim Caspers
*Lyceum Ypenburg, Den Haag, en*
*Faculteit EWI en Lerarenopleiding, TU Delft*
*w.t.m.caspers@tudelft.nl*

**Onderwijs** Bespreking examen vwo wiskunde B 2017

# Het wiskunde B-examen verandert

**Het wiskunde B-examen van 2017 is de laatste reguliere editie van het oude programma op het vwo — vanaf 2018 gelden er andere eindtermen. Op kleine schaal wordt het nieuwe programma al geruime tijd getoetst op zogeheten pilotscholen. De pilotexamens geven een beeld van wat de leerlingen de komende jaren te wachten staat. Wim Caspers, lerarenopleider en ook zelf docent wiskunde, bespreekt de opvallendste verschillen tussen de oude en de nieuwe examens.**

Wie het reguliere examen van 2017 graag wil bekijken, kan natuurlijk terecht op www.examenblad.nl. Bovendien heeft K.P. Hart het afgelopen centrale examen wiskunde B inclusief het tweede tijdvak uitgebreid besproken in zijn weblog [1] en Joost Hulshof heeft zijn licht laten schijnen over de analyse-opgaven in het examen [2]. In de *WiskundE-brief* nummer 783 is een kleine analyse opgenomen en er wordt gemeld dat het gemiddelde cijfer 7,2 bedroeg, opvallend hoog [3]. Tijd om het oude examen achter ons te laten en de blik te richten op het nieuwe. Dat bepaalt de horde die leerlingen vanaf komend jaar moeten nemen om toegang te krijgen tot een academische technische studie.

## Analyse

De belangrijkste verandering zit hem in de meetkunde, waarover later meer. Op het gebied van de analyse zijn de veranderingen niet zo groot. Nieuwkomers zijn onder andere limieten, inverse functies, scheve asymptoten, meer aandacht voor de absolute-waardefunctie, voor de tangens en voor stelsels vergelijkingen. Daar staat minder aandacht voor onder andere differentiequotiënten, Riemannsommen en cirkelbewegingen tegenover. En in alle nieuwe wiskundeprogramma's (dus ook voor wiskunde A, C en D) spelen wiskundige denkactiviteiten (probleemoplossen, abstraheren, redeneren en bewijzen) een grotere rol.

Van de veranderingen op het gebied van de analyse komt de inverse functie als eerste aan bod in het examen, in de opgave 'Een gebroken functie'. De functie $f$ is gegeven door

$$f(x) = \frac{5}{4x-6}.$$

En de grafiek van $f$ wordt $a$ eenheden naar boven verschoven. Zo ontstaat de grafiek van $g$. (Dan blijkt dat ook naar beneden verschoven mag worden, omdat een negatieve $a$ toegestaan wordt.) Gesteld wordt dat $g$ een inverse functie heeft en dat de grafiek van die inverse functie één verticale asymptoot heeft, net als de grafiek van $g$. Gevraagd wordt de waarden van $a$ te bepalen waarvoor de afstand tussen deze twee verticale asymptoten gelijk is aan $4$. Voor iemand die niet na wil denken is het nog best een klus om de inverse functie en bijbehorende verticale asymptoten te bepalen. Een onderdeel dus waar nadenken loont.

Dat is minder het geval bij het andere nieuwe analyseonderdeel; het vinden van de waarde van $p$ waarvoor de grafiek van $f_p$ een perforatie heeft, waarbij

$$f_p(x) = \frac{px^2+4px+6}{(x^2+1)(x-2)}.$$

Overigens wordt de hernieuwde aandacht voor de absolute-waardefunctie en limieten duidelijk uit de tweede opgave uit het herexamen. Daar wordt de functie met voorschrift

$$f(x) = 4e^{2-x} + |8-4x|$$

met raaklijnen in $P(2,4)$ aan het 'linkerdeel van de grafiek' en het 'rechterdeel van de grafiek' onderzocht. Ook moet de asymptoot van de grafiek gevonden worden.

**Meetkunde**

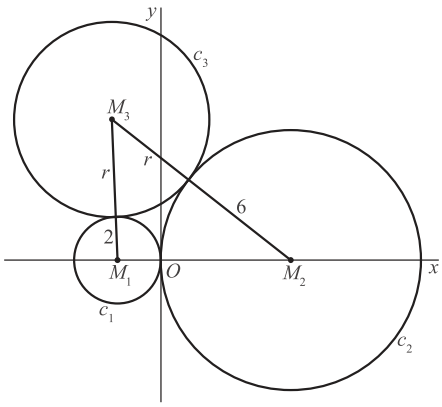
Het aandeel meetkunde is aanzienlijk in het pilotexamen. Zoals gezegd is de verandering in dat domein ook het grootst. Nieuw is de analytische meetkunde (tweedimensionaal). Vectoren, inproduct en cirkelvergelijkingen hebben de plaats ingenomen van de koordenvierhoekstelling en de constante hoek, middelpuntshoek en omtrekshoek. De stelling van Thales, gelijkvormigheid en dergelijke zaken maken nog wel onderdeel uit van het programma.

De meest interessante opgave gaat over twee cirkels $c_1$ en $c_2$ in een assenstelsel, met een derde cirkel $c_3$ die de twee andere raakt (zie Figuur 1). De grootte van $\angle M_1M_2M_3$ is afhankelijk van de straal $r$ van de derde cirkel. Gezocht wordt de limiet waar de grootte van $\angle M_1M_2M_3$ tot nadert als '$r$ onbegrensd toeneemt'. Het is begrijpelijk dat er in het examen een tussenstap wordt ingelast waarin (bijvoorbeeld met de cosinusregel) moet worden aangetoond dat

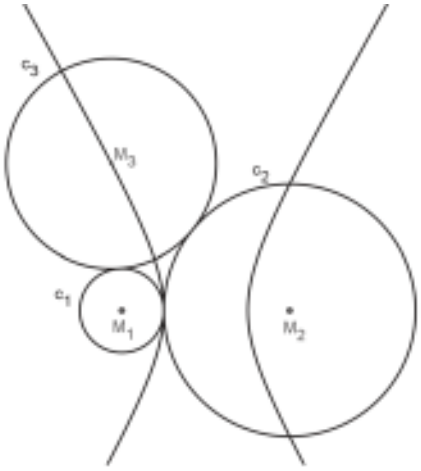
$$\cos(\angle M_1M_2M_3) = \frac{r+12}{2r+12}$$

en dan ligt de limiet voor het grijpen. Een leerling die wiskunde D heeft gevolgd en dus iets van kegelsneden afweet, bedenkt misschien wel dat middelpunt $M_3$ op een hyperbool met brandpunten $M_1$ en $M_2$ ligt, want het verschil van de afstanden $M_1M_3$ en $M_2M_3$ is $4$ (zie Figuur 2). Een vergelijking van die hyperbool in het gegeven assenstelsel is

$$\frac{(x-2)^2}{4} - \frac{y^2}{12} = 1.$$




Figuur 1




Figuur 2

De gevraagde limiet is de grootte van de hoek die een asymptoot van de hyperbool met de lijn $M_1M_2$ maakt.

De laatste vraag over deze situatie gaat over het geval dat de derde cirkel de $x$-as raakt (zie Figuur 3). Het is misschien wel het onderdeel waarin het grootste beroep wordt gedaan op het probleemoplossend vermogen van de leerling. Voornoemde wiskunde D-leerling zou zijn toevlucht kunnen nemen tot het bepalen van het snijpunt van de hyperbool met de parabool met brandpunt $M_1$ en richtlijn $y=-2$ (zie Figuur 4), want de afstand van $M_3$ tot die richtlijn is $r+2$, precies de afstand van $M_3$ tot $M_1$. En $M_3$ ligt ook op de parabool met brandpunt $M_2$ en richtlijn $y=-6$ (zie Figuur 5). Dus oplossen van de vergelijking

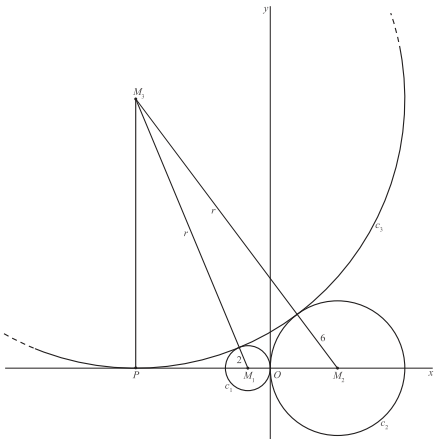
$$\frac{1}{4}(x+2)^2 - 1 = \frac{1}{12}(x-6)^2 - 3$$

is voldoende. Een leerling met alleen wiskunde B in zijn pakket neemt waarschijnlijk zijn toevlucht tot gelijkvormigheid.

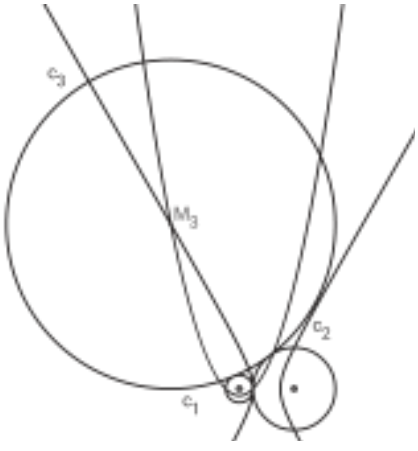
In het oude programma stonden parametervoorstellingen van 'figuren van Lissajous' in het domein goniometrische functies. In het nieuwe examenprogramma maakt de 'baan van een bewegend punt' deel uit van het meetkundedomein. Zodoende worden in het examen vragen gesteld over een punt met bewegingsvergelijkingen

$$\begin{cases} x(t) = \cos(t) + \sin(2t), \\ y(t) = 2\cos(t), \end{cases}$$

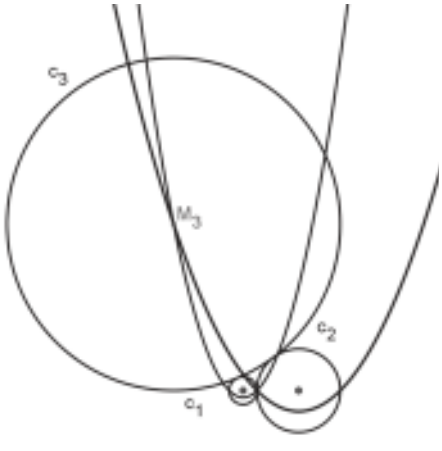
met $t$ in seconden en $x$ en $y$ in meters. Dit levert niet een opgave op met het hoogste gehalte aan benodigde wiskundige denkactiviteiten. Plichtmatig toepassen van gonioformuletjes leidt naar de oplossing.



Figuur 3




Figuur 4



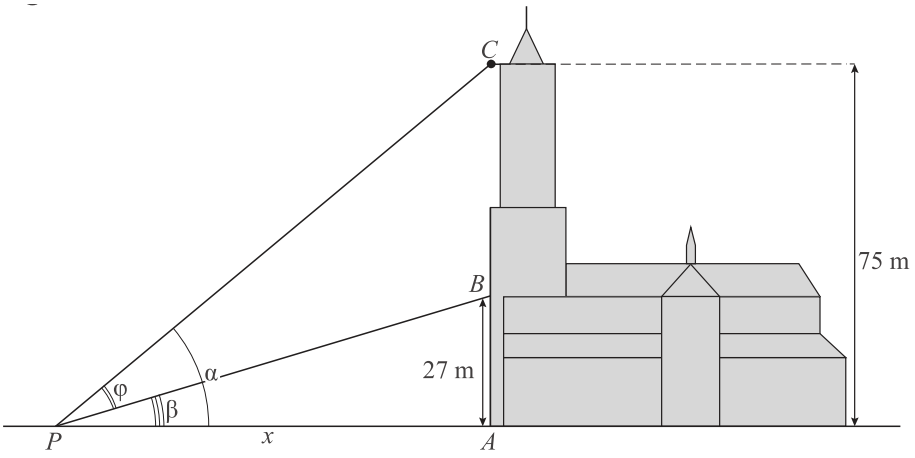

Figuur 5

**Denkactiviteiten**

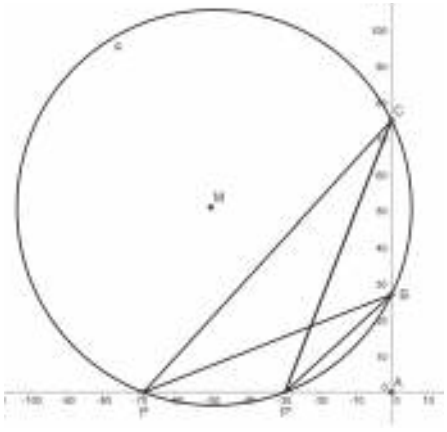
Tot slot een opgave uit het herexamen waar het beroep op wiskundige denkactiviteiten (probleemoplossen, abstraheren, redeneren en bewijzen) ook niet volledig tot zijn recht komt, hoewel denkactiviteiten expliciet vermeld worden in het examenprogramma. In de laatste meetkundeopgave wordt de Eusebiuskerk in Arnhem opgevoerd. Gezocht wordt naar de optimale afstand $x$ waarvoor hoek $\varphi$ zo groot mogelijk is. De opgave leidt de leerling in onderdelen, via een verschilformule voor de tangens, naar de uitdrukking

$$\tan(\varphi) = \frac{48x}{x^2 + 2025}$$

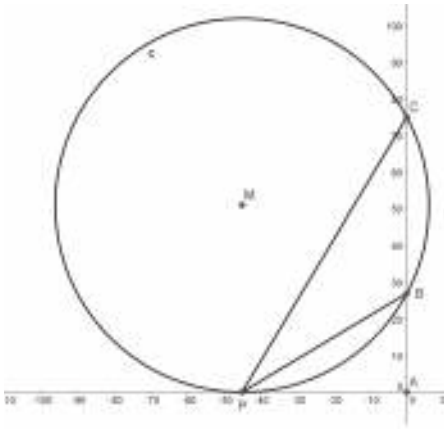
en dan wordt er een maximum gezocht. In de nascholing 'Wiskundige denkactiviteiten' van het vaksteunpunt Delft/Leiden [4] werd een dergelijke kwestie zonder al die tussenstappen voorgelegd aan de deelnemende vo-docenten. Gebruik van een programma als GeoGebra helpt dan om grip te krijgen op de situatie. In Figuur 7 is te zien dat de bedoelde hoek bij $P$ net zo groot is als de hoek bij $P'$. Ertussenin is de hoek groter, hetgeen doet vermoeden dat Figuur 8 de optimale situatie weergeeft. De $x$-coördinaat van $M$ is met wat rekenwerk en Pythagoras terug te voeren op $-\sqrt{y_B \cdot y_C}$. En daarmee komt in het geval van de opgave de optimale $x$ heel mooi uit op $45$. Te mooi om waar te zijn. Het is ook niet waar.




**Figuur 6** Eusebiuskerk in Arnhem.



Figuur 7



Figuur 8

De toren van de Eusebiuskerk is $93$ meter hoog [5], dus $C$ bevindt zich hoogstwaarschijnlijk niet op een hoogte van $75$ meter.

Leerlingen zijn niet bij voorbaat verzot op examenopgaven die een beroep doen op hun wiskundige denkactiviteiten. Misschien moeten we dat ook maar in minder dwingende omstandigheden dan een centraal examen aan de orde laten komen. Het is dus niet zo erg dat het in dit pilotexamen geen grote rol speelt. Het pilotexamen is met name door de analytische meetkunde wel een verademing na een jaar of acht dezelfde examenstof (en eigenlijk verschilde het de jaren ervoor ook niet veel). Inmiddels zit de eerste generatie leerlingen die in groten getale onderworpen gaan worden aan het examen nieuwe stijl in de zesde klas. Het is nog even afwachten of zij zich ook enthousiast door de analytische meetkunde, limieten en dergelijke heen werken om vervolgens net zo hoog te scoren als de voorgaande generaties. Het zou nog wel eens wennen kunnen zijn.

**Referenties**

1 hartkp.weblog.tudelft.nl/2017/05/16/eindexamen-wiskunde-b-2017.
2 www.beteronderwijsnederland.nl/wp-content/uploads/2015/07/wiskundeonderspanning.pdf.
3 www.wiskundebrief.nl.
4 Dank aan Jeroen Spandaw.
5 nl.wikipedia.org/wiki/Sint-Eusebiuskerk.

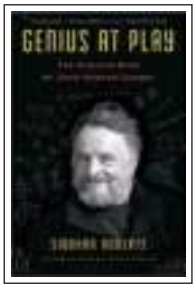
# Boekbesprekingen

| Book Reviews

*Redactie: Hans Cuypers en Hans Sterk*

*Review Editors NAW - MF 7.092*
*Faculteit Wiskunde & Informatica*
*Technische Universiteit Eindhoven*
*Postbus 513*
*5600 MB Eindhoven*

*reviews@nieuwarchief.nl*
*www.win.tue.nl/wgreview*




Siobhan Roberts

**Genius at Play**
**The Curious Mind of John Horton Conway**

*Bloomsbury Publishing PLC, 2015*
*480 p., prijs $ 30.00*
*ISBN 9781620405932*

*Genius at Play* is een biografie van John Conway. Ik verwacht niet dat er snel een Nederlandse vertaling van komt. Niet zozeer omdat er te weinig interesse voor zou zijn, maar omdat het lastig te vertalen zal zijn. Op bijna elke bladzijde kwam ik woorden tegen die ik niet kende, en bij vele ervan hadden ook mijn woordenboeken moeite. Conway is verzot op taal, en dat heeft zijn weg gevonden in deze biografie. De auteur heeft er bovendien een literair werk van gemaakt. Het leest dus niet als wiskunde. Als je daar mee kunt omgaan, dan kan ik het boek van harte aanbevelen.

Zelf heb ik John Conway een paar keer ontmoet, en je kunt hem ook 'meemaken' op YouTube, of beter nog in de interviews die de auteur van deze biografie met steun van de Simons Foundation heeft laten opnemen (Siobhan Roberts, John H. Conway, *Science Lives: John Conway*. Video's van interviews met John Conway, http://www.simonsfoundation.org/science_lives_video/john-conway). Het boek ademt de juiste 'Conway-sfeer' en is derhalve geslaagd.

Siobhan Roberts is bekend geworden door haar biografie van Coxeter, *King of Infinite Space: Donald Coxeter, the Man Who Saved Geometry* (Walker & Company, 2006). In 2003 schreef ze, als journalist in Toronto, het artikel 'Donald Coxeter: The Man Who Saved Geometry' voor *Toronto Life* (http://www.math.toronto.edu/mpugh/Coxeter.pdf) over Coxeter, die later dat jaar overleed. Ze volgde hem naar zijn laatste conferenties, en interviewde daarna ook nog vele wiskundigen om een beter beeld te krijgen van Coxeter voor het schrijven van *King of Infinite Space*. Een van die wiskundigen was Conway. Roberts had al snel in de gaten dat Conway liever over zichzelf praatte dan over Coxeter. Zodoende leerde ze hem beter kennen (en dat Conway een charmeur is zal ook meegespeeld hebben). Conway heeft overigens de conceptversie van *King of Infinite Space* minutieus doorgespit.

In 2007 vatte Roberts het plan op om een biografie van Conway te schrijven, terwijl ze een Director's Visitor was op het Institute for Advanced Study in Princeton, waar ook Conway een aanstelling heeft, op de John von Neumann Chair of Mathematics. Conway heeft zich aanvankelijk flink verzet tegen het idee van een (auto)biografie. Dat heeft vast te maken met zijn ego. In zijn eigen woorden: "I do have a big ego! As I often say, modesty is my only vice. If I weren't so modest, I'd be perfect." Roberts schrijft over zijn weigering: "he was chary, to use a word he likes, about being the subject of a biography. There were too many skeletons in the closet." Maar uiteindelijk geeft hij toch toe.

Uiteraard vind je de dingen die je zoal verwacht in deze biografie. Conway bedacht *The Game of Life* en daar zit een hele geschiedenis aan vast. Van het schaven aan de regels voor leven en dood van de cellen om het spel 'interessant' te houden, tot en met het bewijzen dat *Life* universeel is. Maar Conway wil eigenlijk

niets meer te maken hebben met *Life*, met name omdat zowat iedereen lijkt te denken dat dat het enige is waarmee hij zich heeft beziggehouden. Zijn angst was ook dat een biografie dat beeld zou versterken. Gelukkig doet deze biografie dat niet (en veel later in zijn leven blijkt Conway toch weer vrede gevonden te hebben met zijn verwikkelingen in *Life*).

Natuurlijk is Conway ook bekend van zijn werk aan (eindige enkelvoudige) groepen (denk aan de Conway-groepen en de *ATLAS of Finite Groups*). Ook daar zit een lange geschiedenis aan vast, die de moeite waard is om te kennen (zeker voor beginnende wiskundigen).

Maar Conway heeft veel meer gedaan. Te veel om allemaal in deze recensie op te sommen. Toch wil ik nog wel een paar onderwerpen noemen. Twee onderwerpen, of eigenlijk apparaten, komen ook aan bod in het artikel dat Roberts voor de jaarlijkse Bridges-conferentie op het gebied van wiskunde en kunst schreef in 2015, zie http://archive.bridgesmathart.org/2015/bridges2015-317.html. Daarin licht ze de uitdagingen bij het schrijven van de biografie toe. Het schrijven was een kunstzinnige inspanning van haar kant, maar ze kenmerkt ook Conway als kunstenaar, vanwege zijn originaliteit. Het eerste apparaat is zijn watercomputer, genaamd WINNIE, al weet Conway niet meer waar dat precies voor stond: Water Initiated N... N... I... Engine, iets met Number en/of Numerical. WINNIE was gebouwd met onderdelen van urinalen gebaseerd op het hevelprincipe. Een slimme constructie waarmee binair tellen, optellen en zelfs vermenigvuldigen mogelijk was. Ik was hogelijk verbaasd dat ik daar destijds niets over kon vinden op het internet. Het tweede, evenzeer enigszins vreemde, apparaat is zijn 4D-helm, waarmee hij een tijd door de straten van Cambridge gezworven schijnt te hebben, om te kijken of hij beter gevoel kon krijgen voor de vierde dimensie.

Zijn bijdragen aan *Winning Ways for Your Mathematical Plays* (Elwyn R. Berlekamp, John H. Conway en Richard K. Guy, A. K. Peters Ltd., 2001–2004), vier lijvige delen met analyses van spelletjes, mogen uiteraard niet onvermeld blijven. Conway is een spelletjesfanaat, niet alleen om ze te spelen, maar ook om ze te verzinnen en analyseren. Al vele jaren gaat hij naar wiskundezomerkampen en daagt iedereen uit het tegen hem op te nemen (meestal wint hij, maar de laatste jaren niet meer altijd).

En dan zijn er natuurlijk nog de *surreal numbers*, een voortvloeisel uit zijn werk aan het analyseren van een bepaalde klasse van spelen: in het Engels *partizan games* genoemd, waarbij de zetmogelijkheden voor de twee spelers kunnen verschillen (zoals bij schaken en go), in tegenstelling tot *impartial games*, waarbij de zetmogelijkheden gelijk zijn (zoals bij Nim). Ik heb deze surreal numbers leren kennen via het boek hierover, in dialoogvorm, door Donald Knuth (die ook de naam bedacht). Conway beschouwt dit zelf als zijn grootste wiskundige creatie. En de wiskunde eromheen is nog lang niet af. Een fascinerende wereld.

De informaticus in mij noopt me om ook FRACTRAN te noemen. Dit is een programmeertaal waarbij een programma uit een lijst van (positieve) breuken bestaat. De rekenregel is eenvoudig. De invoer is een positief geheel getal, tevens de startwaarde van het *huidige getal* $n$. Je zoekt de eerste breuk $b$ (van links) die vermenigvuldigd met het huidige getal een geheel getal oplevert. Dit product $nb$ wordt het nieuwe huidige getal, waarmee je weer voor-aan de lijst begint te zoeken. Als zo'n breuk er niet is, dan stopt het programma. De uitvoer wordt afgelezen uit de exponenten in de priemontbinding van het huidige getal. Zo is er een programma bestaande uit veertien breuken dat de rij priemgetallen genereert, dat wil zeggen de priemmachten van twee. Deze programmeertaal is universeel.

Ook zijn *Doomsday*-algoritme komt aan bod. Daarmee kun je, of in elk geval kan Conway, heel snel de dag bepalen waarop een gegeven datum valt. Conway verfijnde zijn algoritme en werd er ongelofelijk snel in. Iets wat hij graag demonstreert.

Conway heeft ook zijn sporen (en een boek) nagelaten op het gebied van symmetrie. Zo is er zijn *orbifold*-notatie voor symmetriegroepen. En hij verzon het woord *metachiral*. Een voorwerp is chiraal als het verschilt van zijn spiegelbeeld. Maar dat kan op twee essentieel verschillende manieren gebeuren. Beschouw de symmetriegroep van het voorwerp en die van het spiegelbeeld. Voor tweedimensionale voorwerpen zijn die twee groepen altijd precies hetzelfde. Maar in de ruimte kunnen ze verschillen, dat wil zeggen dat die groepen uit verschillende transformaties bestaan (de groepen zijn wel isomorf). In dat geval noemt Conway het voorwerp metachiraal. Een oneindige helix (schroeflijn) is een voorbeeld van een metachiraal voorwerp. In de ene groep zitten dan linksdraaiende schroefoperaties (een combinatie van een rotatie en een translatie langs de rotatieas), en in de andere groep zitten rechtsdraaiende schroefoperaties. Dat maakt de (3D-)ruimte zoveel interessanter dan het (2D-)vlak.

Noemde ik zijn werk aan magische vierkanten al? Daar geeft hij graag lezingen over.

Niet wiskundig van aard, maar zijn hartaanval en zijn zelfmoordpoging moeten ook vermeld worden. Deze onderwerpen maken de biografie zeer persoonlijk, en Conway praat er heel open over, net als over religie.

Conways meest recente bevlieging is de *Free Will Theorem*. Kortweg luidt deze stelling: als natuurkundigen een vrije wil hebben bij het uitvoeren van experimenten, dan hebben elementaire deeltjes dat ook. Vrije wil betekent hier dat het gedrag geen deterministische functie is van het verleden. Samen met Simon Kochen heeft hij hier een ingenieus bewijs voor, uitgaande van enkele quantummechanische axioma's. Een gevolg hiervan is echter ook dat willekeur (dat wil zeggen kansrekening) geen goede verklaring biedt voor het gedrag van elementaire deeltjes.

Roberts laat Conway veelvuldig zelf aan het woord, en gebruikt een apart lettertype wanneer ze hem citeert. Dit geeft het gevoel dat Conway rechtstreeks tot de lezer spreekt. Hoewel Conway geen 'grote' prijzen heeft gewonnen, is hij een zeer creatief en exotisch wiskundige met een indrukwekkende staat van dienst. Uit deze biografie zijn ook heel wat wijze lessen te halen, met name voor wiskundigen, maar ook voor docenten:

[Conway citeert eerst Groucho Marx:] *"The secrets to success in life are honesty and sincerity. If you can fake those, then you've got it made."* [En vervolgt dan zelf:] *"And I think that is terribly important in teaching. I would stick enthusiasm in with honesty and sincerity — enthusiasm is very important in teaching. You don't have to fake it. If you actually have it, that's the best thing. But if not, you better fake it."*

*Tom Verhoeff*

Valerio Lucarini, Davide Faranda, et al.

**Extremes and Recurrence in Dynamical Systems**

*Pure and Applied Mathematics: A Wiley Series of Texts, Monographs, and Tracts*
*John Wiley & Sons, 2016*
*312 p., prijs € 106,20*
*ISBN 9781118632192*

Volgens de centrale limietstelling zijn gemiddelden van een rij van onafhankelijke en identiek verdeelde stochasten na een geschikte affiene schaling asymptotisch normaal verdeeld. Een soortgelijke stelling geldt voor de maxima van een rij stochasten in welk geval de asymptotische verdeling wordt gegeven door een Weibull-, Gumbel- of Fréchet-verdeling. Deze stelling speelt een sleutelrol in de statistiek van extreme waarden en heeft allerlei toepassingen in de verzekeringswiskunde. Al sinds een aantal decennia is de theorie uitgebreid naar afhankelijke stochasten, maar een heel recente ontwikkeling is de uitbreiding naar *deterministische* dynamische systemen. Het idee is dat een scalaire tijdreeks voortgebracht door een chaotisch systeem als stochastisch kan worden opgevat mits correlaties voldoende snel afnemen. Lucarini et al. schreven het eerste boek over deze nieuwe ontwikkeling.

Het boek is geschreven door een team van negen vooraanstaande onderzoekers op het gebied van extremen in dynamische systemen. Dat maakt het boek zeer compleet: alle belangrijke ontwikkelingen tot 2015 passeren de revue. Het boek bespreekt uitvoerig onder welke voorwaarden de verdelingen voor extreme waarden in dynamische systemen hetzelfde zijn als voor onafhankelijke stochasten. De wiskundige technieken om deze voorwaarden te verifiëren worden in detail behandeld. Daarnaast worden allerlei aspecten van extremen besproken die specifiek zijn voor dynamische systemen zoals verbanden tussen de meetkunde van attractoren en de parameters van de extreme-waardeverdeling. Een andere interessante toepassing die aan de orde komt is het verband tussen extremen en zogenaamde omslagpunten waarbij de dynamica van een systeem een kwalitatieve verandering ondergaat.

De presentatie vond ik soms wat droog. Bij tijd en wijle heeft het boek een tamelijk opsommend karakter. Meer concrete voorbeelden zouden welkom zijn geweest. De bibliografie is zeer uitgebreid maar helaas niet alfabetisch geordend waardoor het lastig is om artikelen terug te vinden. Verder belooft het boek een appendix met een aantal Matlab scripts ter verduidelijking van de concepten, maar het aantal scripts vond ik vrij mager. Hier had beslist meer in gezeten. Wie zich echter wil bekwamen in extremale statistiek van dynamische systemen kan niet om dit boek heen. Het boek is duidelijk gericht op de lezer met een wiskundige achtergrond: kennis van stochastische processen of ergodentheorie is zeer wenselijk. Het boek is vooral geschikt voor gevorderde werkgroepen en seminars, maar niet voor een cursus. *Alef Sterk*

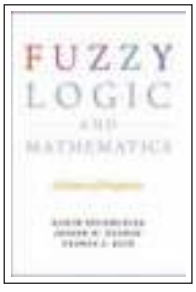
---

*Recent verschenen publicaties. Als u een van deze boeken wilt bespreken of als u suggesties heeft voor andere boeken voor deze rubriek, laat dit dan per e-mail weten aan reviews@nieuwarchief.nl.*

Stefan Müller-Stach (Hrsg.)

**Richard Dedekind**
**Was sind und was sollen die Zahlen? Stetigkeit und Irrationale Zahlen**

*Springer, 2017*
*ISBN 9783662543399*
*springer.com/9783662543382*



Radim Belohlavek, Joseph W. Dauben, George J. Klir

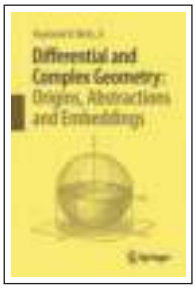
**Fuzzy Logic and Mathematics**
**A Historical Perspective**

*Oxford University Press, 2017*
*ISBN 9780190200015*
*oup.com/academic/product/9780190200015*

Paul Pollack

**A Conversational Introduction to Algebraic Number Theory**
**Arithmetic Beyond Z**

*American Mathematical Society*
*ISBN 9781470436537*
*bookstore.ams.org/stml-84*



Raymond O. Wells, Jr

**Differential and Complex Geometry**
**Origins, Abstractions and Embeddings**

*Springer, 2017*
*ISBN 9783319581842*
*springer.com/9783319581835*

# Problemen

| Problem Section

This Problem Section is open to everyone; everybody is encouraged to send in solutions and propose problems. Group contributions are welcome. For each problem, the most elegant correct solution will be rewarded with a book token worth € 20. (To compete for the book token you should have a postal address in The Netherlands.)
Please send your submission by e-mail (LaTeX is preferred), including your name and address to problems@nieuwarchief.nl.
The deadline for solutions to the problems in this edition is 1 December 2017.

**Problem A**

Let $n$ be a natural number and suppose that $A_1,\ldots,A_n$ are different subsets of $\{1,\ldots,n\}$. Prove that there is a $k \in \{1,\ldots,n\}$ such that $A_1\setminus\{k\},\ldots,A_n\setminus\{k\}$ are different.

**Problem B** (proposed by Hans Zantema)

Let $f,g:\mathbb{N}\to\mathbb{N}$ be strictly increasing functions. Prove that there exists an $n\in\mathbb{N}$ such that $f(g(g(n)))\geq g(f(n))$.

**Problem C** (proposed by René Pannekoek)

Determine all $n\in\mathbb{N}$ such that $2^n-1$ divides $3^n-1$.

**Edition 2017-1** We received solutions from Pieter de Groen (Brussel), Alex Heinis (Amsterdam), Thijmen Krebs (Nootdorp), Hendrik Reuvers (Maastricht), Hans Samuels Brusse (Den Haag), Djurre Tijsma (Zeist), Rob van der Waall (Huizen), Martijn Weterings (Sion) and Hans Zantema (Eindhoven). The book tokens for problems A, B and C go to Rob van der Waall, Hans Samuels Brusse, respectively Hans Zantema.

**Problem 2017-1/A**

Reconstruction. Suppose you get to know the $n$ midpoints of the $n$ edges of a polygon. Can you determine the polygon?

**Solution** Solved by Pieter de Groen, Thijmen Krebs, Hendrik Reuvers, Hans Samuels Brusse, Rob van der Waall, Martijn Weterings and Hans Zantema. Almost all solutions are similar. Rob van der Waall presents a purely geometric construction.
No if $n$ is even. Yes if $n$ is odd. Denote the midpoints by $m_1,\ldots,m_n$. If $n$ is odd, then the alternating sum $m_1-m_2+\cdots+m_n$ is equal to the polygon's initial vertex. All vertices can be reconstructed like this. If $n$ is even, then we may add an arbitrary $v$ to the even vertices and subtract it from the odd vertices without changing the midpoints. We may place the initial vertex at an arbitrary point, and we can construct a polygon by reflections in the midpoints.

**Problem 2017-1/B**

Large is odd. Let $G$ be a graph with vertices $V$ and edges $E$. A subset $U\subset V$ is called *large* if every vertex that is not in $U$ has a neighbor in $U$. Prove that the number of large subsets is odd.

**Solution** We received solutions from Hans Samuels Brusse and Pieter de Groen. Thijmen Krebs noticed that this is a result of Brouwer, Csorba and Schrijver, which can be found

*Redactie:*
*problems@nieuwarchief.nl*
*www.nieuwarchief.nl/problems*

Oplossingen

| Solutions

through www.win.tue.nl/aeb/preprints/domin4a.pdf. Hans Samuels Brusse uses induction on the number of vertices. Pieter de Groen uses induction on the number of edges. The wonderful solution below is due to Brouwer, Csorba and Schrijver.

For every $U \subset V$, let $\bar{U}$ be the subset of vertices in $V \backslash U$ that contain no neighbor in $U$. In particular, $U$ is large if and only if $\bar{U}$ is empty. Consider the family $\mathcal{F}$ of all pairs $(U,W)$ such that $W \subset \bar{U}$. For a fixed $U$, there are $2^{|\bar{U}|}$ such pairs. In particular, for a fixed $U$, there are an odd number of such pairs if and only if $U$ is large. Therefore, the parity of the number of large sets is equal to the parity of $|\mathcal{F}|$. This family is invariant under the involution $(U,W) \mapsto (W,U)$, which has $(\emptyset,\emptyset)$ as a single fixed point. Therefore, the parity of $|\mathcal{F}|$ is odd.

---

**Problem 2017-1/C**

Meanders. Let $n$ be an even number. Consider the integers $1$ to $n$ in the complex plane and connect them by semicircles centered around $\frac{n+1}{2}$ in the upper half plane. Let $n = a + b$ for even numbers $a$ and $b$. Connect the first $a$ integers by semicircles around $\frac{a+1}{2}$ in the lower half plane. Similarly, connect the last $b$ integers by semicircles around $\frac{a+b+1}{2}$ in the lower half plane. The resulting curve, or set of curves, is a meander. For which $a$ and $b$ is it connected?

**Solution** We received solutions from Pieter de Groen, Alex Heinis, Thijmen Krebs, Hendrik Reuvers, Hans Samuels Brusse, Djurre Tijsma and Hans Zantema. All solutions are similar. Alex Heinis remarks that this problem is related to the combinatorics of Hedlund words, as studied in his thesis. Indeed, there is more to this problem. Here we look at meanders for a sum of two even numbers, but the definition extends to sums of more than two numbers. For sums of four or more even numbers it is much harder to decide if the meander is connected. The interested reader should watch Maryam Mirzakhani's Marston Morse lecture https://www.youtube.com/watch?v=mxPE6vYwqLg.

The semicircles connect an even number to an odd number. If we start from an arbitrary even number and we trace an upper semicircle followed by a lower semicircle, then we are back at an even number. More specifically, if we start from $2k$, then we are back at $2k + a$ modulo $n$. The meander is connected if and only if it takes $n/2$ iterations before we are back at $2k$. In other words, the least common multiple of $a$ and $n$ is equal to $na/2$. It is easier to say that the greatest common divisor of $a$ and $n$ is equal to $2$. Since $n = a + b$, the meander is connected if and only if $\gcd(a,b) = 2$.