The Clean Paper

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

Et kryptografisk bevis kan skjule sin hemmelighed ved at gøre den manglende simulator svær at bevise

Nulvidensbeviser lader nogen bevise et udsagn uden at afsløre vidnet. Klassisk teori siger, at dette i fuld forstand ikke kan opnås med én meddelelse, uden betroet opsætning og med perfekt sundhed. Rahul Ilangos artikel får ikke denne umulighed til at forsvinde. Den ændrer målet: I stedet for at kræve, at en simulator virkelig findes, kræver den, at intet valgt bevissystem effektivt kan bevise, at simulatoren ikke findes. Under store antagelser fra beviskompleksitet og kryptografi er det nok til at genvinde falsificerbare, spilbaserede konsekvenser af nulviden, egenskab for egenskab. Pointen er ikke en internetprimitiv, der kan tilsluttes direkte. Det er en bevisteoretisk måde at gøre matematisk ubeviselighed til kryptografisk dække på.

Written by Cinzia Vaglio · Cross-checked by Vera Rubin · Edited by Michele Renda · Figures by Nova Vela · AI-assisted, human-reviewed

Paper: *Gödel in Cryptography: Effectively Zero-Knowledge Proofs for NP with No Interaction, No Setup, and Perfect Soundness*, Rahul Ilango, FOCS 2025 / IACR ePrint 2025/1296 · DOI: 10.1109/FOCS63196.2025.00058

Tricket er ikke at bevise, at hemmeligheden er skjult

Begynd med den enkleste version af nulviden.

Alice vil overbevise Bob om, at en sudoku har en løsning. Hvis hun sender løsningen, bliver Bob overbevist, men opgaven er ødelagt. Det, hun ønsker, er noget mærkeligere: et bevis på, at der findes en løsning, uden at afsløre løsningen.

Det er løftet i et nulvidensbevis. Beviseren (Alice) overbeviser verifikatoren (Bob) om, at et udsagn er sandt, samtidig med at hun ikke afslører andet end udsagnets sandhed.

Problemet er, at dette løfte har en pris. Et almindeligt matematisk bevis har to behagelige egenskaber. Det er **én meddelelse**: Man skriver det ned, afleverer det og går sin vej. Og det har **perfekt sundhed**: Et falsk udsagn har slet ikke noget gyldigt bevis. Klassiske umulighedsresultater siger, at nulviden må opgive begge egenskaber — og ikke blot de to tilsammen; hver af dem er udelukket også alene.

For det første kræver et nulvidensbevis en samtale. Hvis Alice sender en enkelt meddelelse uden en betroet opsætning, der er arrangeret på forhånd, bryder nulvidensgarantien sammen — og det gælder, uanset hvor meget sundhed man er villig til at ofre til gengæld.

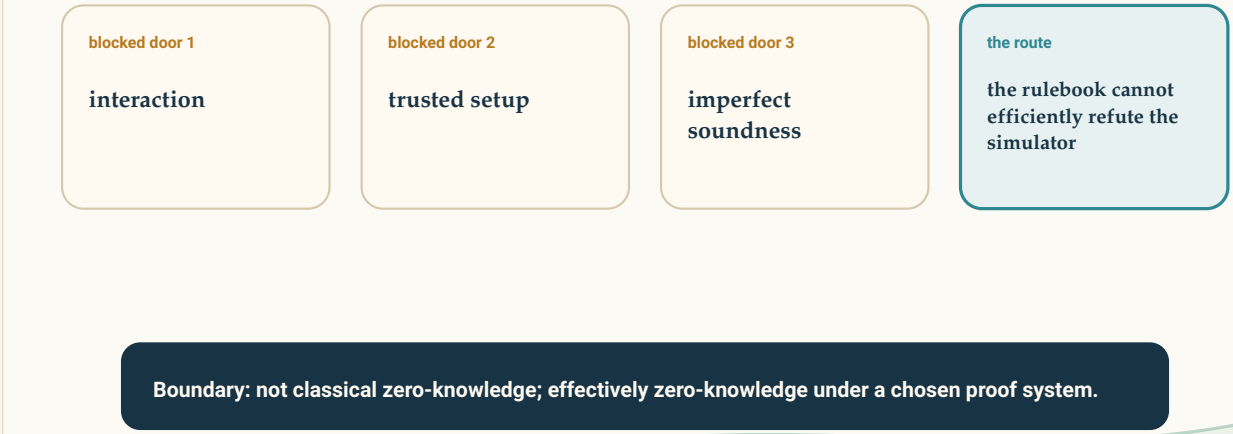
For det andet kræver et nulvidensbevis en lille tolerance for fejl. Det viser sig, at kravet om perfekt sundhed i det stille også ødelægger interaktionen: En verifikator, som aldrig kan narres, uanset hvilke tilfældige valg den træffer, kan lige så godt fastlægge disse valg på forhånd — og når verifikatoren først er forudsigelig, kan Alice besvare alt i en enkelt meddelelse, hvilket netop er det tilfælde, der allerede brød sammen.

Rahul Ilangos artikel handler om en vej uden om denne dobbelte mur. Ikke ved at foregive, at muren ikke findes, og ikke ved at skabe klassisk nulviden i den umulige ramme. Grebet er mere subtilt: at svække betydningen af "afslører intet", men svække den på en måde, der bevarer de sikkerhedsegenskaber, kryptografer faktisk kan teste.

Resultatet kaldes **effektiv nulviden**.

A proof without leaking the witness

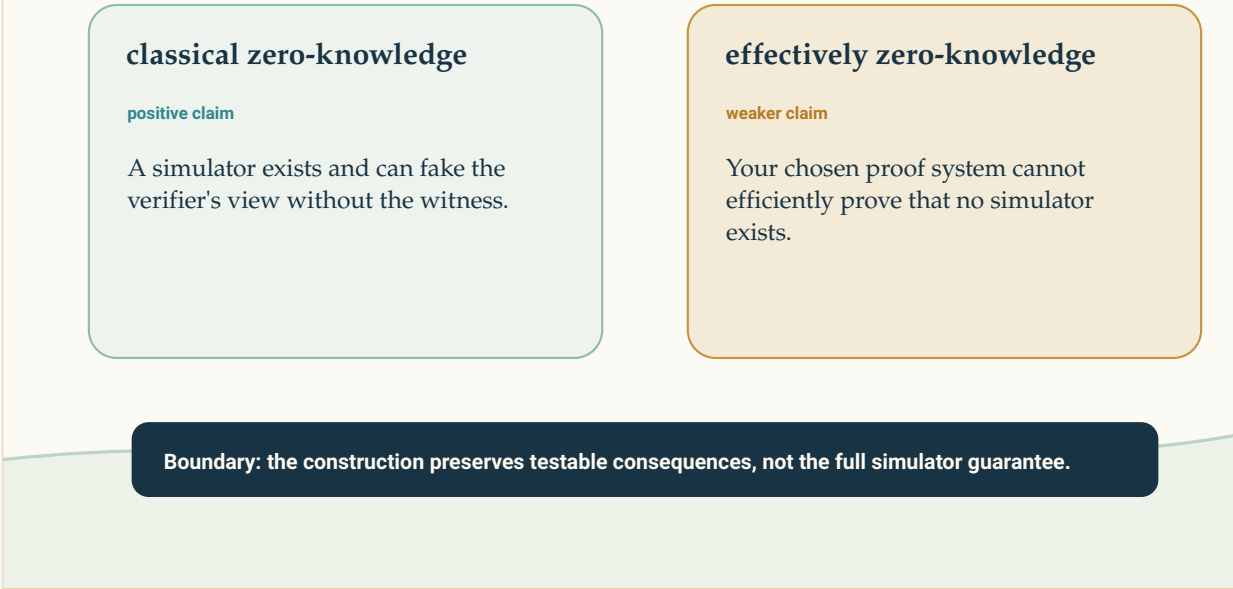
The paper keeps the ordinary-proof shape by weakening what privacy must prove.



Nulviden er blokeret ved tre døre — interaktion, betroet opsætning og uperfekt sundhed. Ilangos konstruktion smutter gennem en anden: Regelsættet kan ikke effektivt gendrive simulatoren.

Classical privacy vs effective privacy

The relaxation is the point: the paper changes what the privacy claim has to establish.



Klassisk nulviden spørger, om en simulator findes; ”effektiv nulviden” spørger kun, om det valgte regelsæt effektivt kan bevise, at ingen findes. Det svagere spørgsmål er det, der lader konstruktionen bevare én

meddelelse, ingen opsætning og perfekt sundhed.

Den gamle test: En simulator findes

Den klassiske måde at formalisere nulviden på bruger en fiktiv hjælper, som kaldes en **simulator**.

Ideen er denne: Forestil dig Jane, som **ikke** kender Alices hemmelighed. Hvis Jane helt på egen hånd kan fremstille beviser, der ser ud præcis som de beviser, Bob ville have modtaget fra Alice, så har Alices beviser ikke lært Bob noget nyt. Jane kunne allerede efterligne oplevelsen uden Alices hemmelighed.

Klassisk nulviden kræver altså en virkelig simulator. Der skal findes en effektiv algoritme, som kan fremstille beviser, der ser ægte ud, uden at kende hemmeligheden — *vidnet*, i fagsproget; for sudoku er vidnet ganske enkelt det løste gitter.

Den definition er stærk, men det er også netop dér, den gamle umulighed bider. Her er intuitionen. Et virkelig ikke-interaktivt bevis er blot en streng. Når Bob først har denne streng, kan han vise den til en anden: Han har fået evnen til at bevise udsagnet over for andre, hvilket allerede lyder som mere end ”intet”. De klassiske teoremer skærper denne intuition til umulighederne ovenfor.

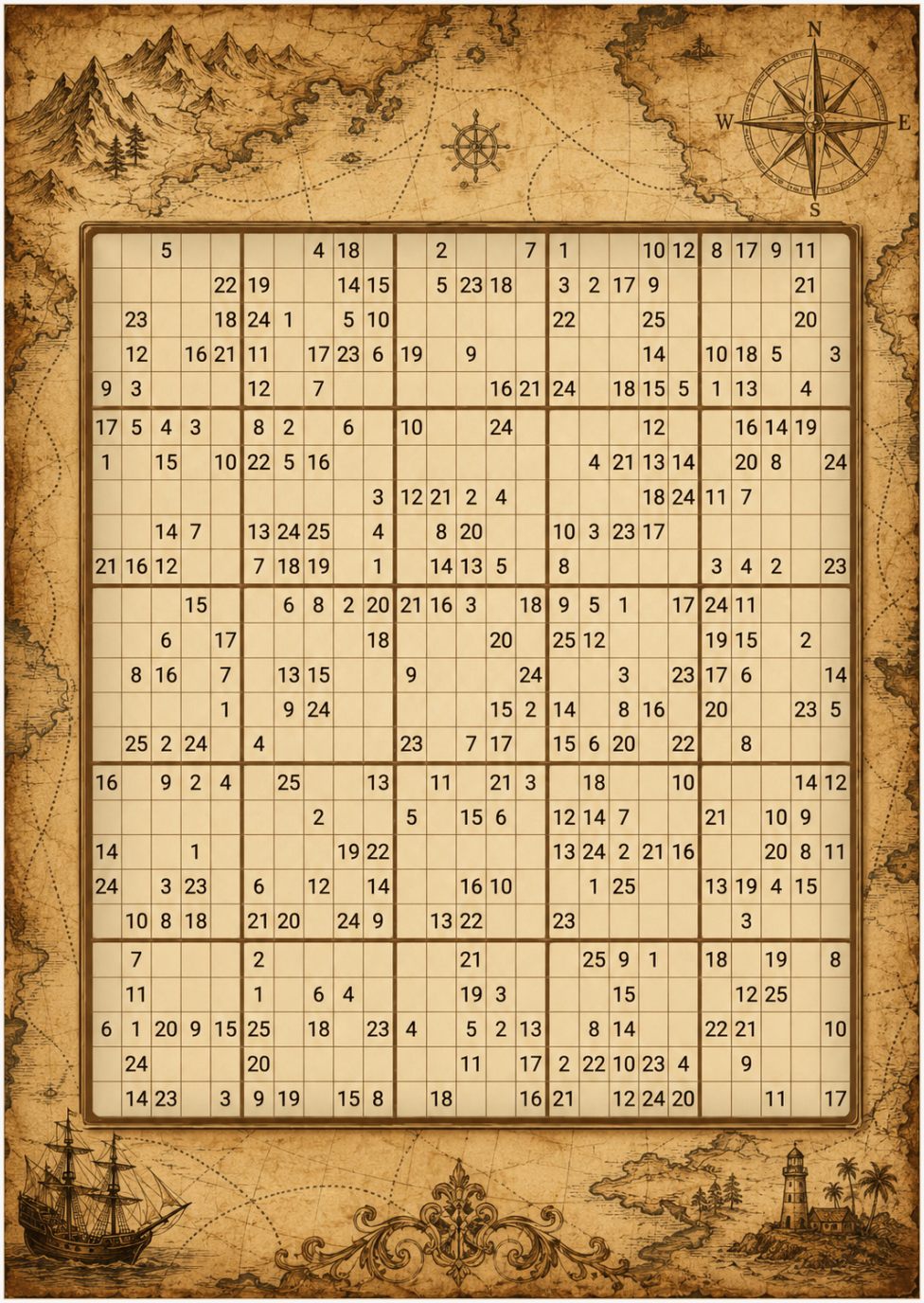
> **De tre egenskaber, denne artikel insisterer på**
>
> Artiklens titel nævner tre begrænsninger:
> **Ingen interaktion:** Alice sender én bevisstreng. Der er ingen protokol med meddelelser frem og tilbage.
> **Ingen opsætning:** Alice og Bob er ikke afhængige af en betroet fælles referencestreng eller anden på forhånd arrangeret offentlig tilfældighed. Mange systemer, der kaldes ”ikke-interaktiv nulviden”, er stadig afhængige af opsætning; denne artikel mener slet ingen opsætning.
> **Perfekt sundhed:** Et falsk udsagn har intet gyldigt bevis. Ikke ”accepteres næsten aldrig”; der findes intet gyldigt bevis.
> Disse tre egenskaber er præcis dem, almindelig skriftlig matematik har — og som forklaret ovenfor kan klassisk nulviden ikke bevare dem.

En mega-sudokuversion af forskellen

Her er en bevidst forenklet måde at få en fornemmelse af forskellen på.

Brug ikke en almindelig 9 gange 9-sudoku til den seriøse del af analogien. Den er for lille og for endelig: En computer kan ganske enkelt løse den eller bevise, at den ikke har nogen løsning. Forestil dig i stedet en familie af **MegaSudoku(n)**-opgaver. Skaler den sædvanlige regel op: Vælg en blokstørrelse

n, lad *N* = *n*^2, og byg et *N* gange *N*-gitter opdelt i *n* gange *n*-blokke med *N* symboler. Almindelig sudoku er blot det lille tilfælde *n* = 3, *N* = 9: et 9 gange 9-gitter, 3 gange 3-blokke og ni symboler. Historien om beviskompleksitet begynder først, når *n* får lov til at vokse, og når gitteret kan rumme ekstra gadgets, der får det til at opføre sig som en SAT-formel forklædt som en sudoku. En SAT-formel er blot en liste med ja/nej-betingelser: Kan man tildele variablerne sand/falsk-værdier, så hver betingelse er opfyldt?



En 25x25-sudoku: Dens regler kan kontrolleres uden at afsløre det færdige gitter — en visuel stedfortræder for et bevis, der verificerer en skjult løsning, vidnet.

Sudoku og SAT: den samme opgave i to forklædninger

Påstanden om, at en sudoku kan "opføre sig som en SAT-formel", er ikke en metafor. Oversættelsen går begge veje, og den nemme retning kan skrives fuldt ud.

Fra sudoku til SAT. SAT taler kun sandt/falsk, så giv den én boolesk variabel for hver tripel (række, kolonne, værdi): *x(r,c,v)* betyder "cellen i række *r*, kolonne *c* indeholder værdien *v*". En 4 gange 4-sudoku (2 gange 2-blokke, værdierne 1–4) kræver 4·4·4 = 64 variabler; den klassiske 9 gange 9 kræver 729. Hver sudokuregel bliver derefter til en samling klausuler. (En klausul er en ELLER-forbindelse af variabler eller deres negationer; hele formlen er OG-forbindelsen af alle dens klausuler.)

Hver celle indeholder mindst én værdi — én klausul pr. celle:

$$x(1,1,1) \lor x(1,1,2) \lor x(1,1,3) \lor x(1,1,4)$$

Hver celle indeholder højst én værdi — en "ikke begge"-klausul for hvert par af værdier:

$$\neg x(1,1,1) \lor \neg x(1,1,2) \quad \neg x(1,1,1) \lor \neg x(1,1,3) \quad \text{... og så videre for alle seks par.}$$

Hver række indeholder hver værdi — for række 1 og værdien 3: mindst én gang,

$$x(1,1,3) \lor x(1,2,3) \lor x(1,3,3) \lor x(1,4,3)$$

og højst én gang: $\neg x(1,1,3) \lor \neg x(1,2,3)$, og så videre for hvert par af celler i rækken.

Kolonner og blokke — identiske samlinger; kun gruppen af celler ændres. For blokken øverst til venstre og værdien 2:

$$x(1,1,2) \lor x(1,2,2) \lor x(2,1,2) \lor x(2,2,2)$$

plus de parvise "ikke begge"-klausuler.

De trykte ledetråde — den enkleste del: Hver ledetråd er en klausul med en enkelt variabel. Et trykt 3-tal i øverste venstre hjørne bliver til klausulen

$$x(1,1,3)$$

OG-forbindelsen af alt dette er satisfierbar præcis, når sudokuen har en løsning — og en satisfierende tildeling *er* løsningen: Aflæs, hvilke x(r,c,v) der er sande, og udfyld gitteret. For en 9 gange 9 bliver det til 729 variabler og nogle tusinde klausuler, som en moderne SAT-løser klarer på millisekunder. Bemærk ledetrådsklausulen x(1,1,3): Den siger "denne celle er præcis 3", ikke "alle disse celler er forskellige" — den samme asymmetri, der vil kræve det ekstra trick for ledetrådsceller i protokolnoten længere nede.

Fra SAT til sudoku. Artiklen har brug for den modsatte, sværere retning: Givet en *vilkårlig* SAT-formel skal man bygge en mega-sudoku, der har en løsning præcis, når formlen har det. Sudokus egne regler kan kun sige "alle disse celler er forskellige", så vilkårlige logiske betingelser skal *bygges* — og det er netop, hvad gadgets er. En gadget er en lille præfabrikeret klynge af celler, én pr. klausul i formlen, hvor udpegede celler spiller variablernes rolle (det symbol, de indeholder, koder sandt eller falsk), og klyngens interne betingelser er konstrueret sådan, at dens eneste lovlige udfyldninger svarer til tildelinger, der opfylder klausulen. Dette

er standardhåndværk fra beviser for NP-fuldstændighed; for generaliseret sudoku blev det udført af Yato og Seta i 2003.
Tilsammen siger de to retninger, at N gange N-sudoku og SAT er det samme problem i forskellige forklædninger. Det er dette, der giver denne tekst — og forskningsartiklen — grundlag for at fortælle en historie om hele NP ved hjælp af gitre og symboler.

Vidnet er stadig let at forestille sig. Alice kender en fuldstændig, gyldig udfyldning af mega-sudokuen. Bob vil overbevises om, at en sådan udfyldning findes, men Alice vil ikke afsløre den. Hvis hun sender hele udfyldningen, bliver Bob overbevist, men hemmeligheden er væk.

I den klassiske nulvidensversion interagerer Alice og Bob. En gammeldags tankemodel bruger tildækkede brikker. Alice skjuler det løste gitter, omdøber i hemmelighed symbolerne før hver runde og lader Bob inspicere én tilfældigt valgt lokal betingelse: en række, en kolonne, en blok eller en gadget. Hvis de åbnede celler viser indbyrdes forskellige symboler, får Bob større tiltro. Så dækkes alt til igen, og symbolerne får helt nye navne. (Der er en komplikation: Opgavens givne ledetråde kræver et ekstra trick, fordi omdøbningen også skjuler dem. Noten nedenfor forklarer, hvordan de klassiske protokoller løser dette; legetøjsbilledet er tilstrækkeligt til det følgende.)

Sådan håndterer de klassiske protokoller egentlig ledetrådscellerne

Omdøbningstricket har en blind vinkel. Reglerne for rækker, kolonner og blokke siger alle "disse celler er alle forskellige", og *alle forskellige* overlever enhver omdøbning af symbolerne. Men en ledetråd siger "denne celle indeholder præcis 5", og efter omdøbningen ser Bob kun $\sigma(5)$ — et maskeret symbol — uden at kende omdøbningen σ. Han kan ikke kontrollere noget. Hvis dette ikke blev rettet, kunne Alice bevise, at *et eller andet* gyldigt gitter findes, mens hun fuldstændig ignorerede de trykte ledetråde, hvilket intet beviser om *denne* opgave. Den klassiske litteratur har to standardløsninger.
Paletten. Føj én ekstra række med N celler til det skjulte gitter — en palet, som Alice udfylder med symbolerne 1…N i en fast, offentlig rækkefølge og derefter omdøber sammen med alt andet, så den indeholder $\sigma(1)…\sigma(N)$. Bobs tilfældige udfordring har nu en ekstra mulighed. Ud over at vælge en række, kolonne, blok eller gadget, der skal åbnes, kan han vælge **paletten plus én ledetrådscelle**. Alice afdækker begge; paletten afslører rundens omdøbning, og Bob kontrollerer, at ledetrådscellen viser præcis den omdøbte version af den trykte ledetråd. Dette forbliver nulviden, fordi Bob kun lærer σ — som trækkes på ny i hver runde og er værdiløs i sig selv — samt værdien af en celle, han allerede kendte fra opgaven. Intet om de hemmelige celler lækker, og en simulator kan efterligne visningen ved at trække en tilfældig σ. Protokollen er sund, fordi en Alice, der snyder, bliver opdaget med en fast sandsynlighed pr. runde, og runderne gentages, indtil tvivlen er negligerbar.
Kompilering af ledetrådene væk. En mere strukturel variant fjerner den særlige udfordring i stedet for at tilføje den. I stedet for at *verificere* ledetrådsværdien tvinger man den frem med ulighedsbetingelser: Forbind ledetrådscellen med hver paletcelle bortset fra den, der bærer dens egen værdi — "forskellig fra $\sigma(1)$, forskellig fra $\sigma(2)$, …, forskellig fra alt undtagen $\sigma(5)$". Det eneste symbol, cellen lovligt kan indeholde, er ledetrådens. Hver betingelse er nu igen af

typen ”disse to er forskellige” — invariant under omdøbning og kontrollerbar præcis som en række. Det er den samme manøvre, der bruges til på forhånd farvede knuder i den klassiske protokol for graffarvning, og det er ånden i ordet *gadgets* ovenfor: I billedet, hvor MegaSudoku er SAT, kompileres ledetrådene til ulighedsgadgets ligesom alle andre betingelser.

Den fysiske protokol. Kortprotokollen for sudoku i den virkelige verden (Gradwohl, Naor, Pinkas og Rothblum, 2007) bruger slet ingen omdøbning og afgør ledetrådene, før skjulningen overhovedet begynder. For hver celle lægger Alice tre identiske kort med cellens værdi — med billedsiden nedad for hemmelige celler, men **med billedsiden opad for ledetrådsceller**, så Bob med egne øjne ser, at ledetrådene overholdes, før kortene vendes. Derefter kommer ét kort fra hver celle i rækkens bunke, ét i kolonnens og ét i blokkens; hver bunke blandes og vises, og Bob kontrollerer, at den indeholder alle N symboler. Blandingen ødelægger positionsinformationen (det er nulviden), men ledetrådene var allerede slået fast, da kortene blev lagt.

Uanset metoden er læren den samme, som denne tekst hele tiden vender tilbage til: En nulvidensprotokol er et omhyggeligt regnskab over, *hvilke* fakta der overlever skjulningen. Omdøbning bevarer ”alle forskellige” og udvisker ”er lig med 5” — derfor skal ”er lig med 5” smugles tilbage på andre måder.

Det er ikke protokollen i forskningsartiklen. Det er tankemodellen for klassisk nulviden:

- Alice og Bob sender meddelelser frem og tilbage.
- Bob vælger tilfældige kontroller.
- Alice afslører kun lokal konsistens, ikke hele løsningen.
- Beviset for privatliv virker ved at vise, at Bobs visning kunne være blevet genereret uden Alices hemmelige løsning.

Klassisk nulviden er altså bygget op omkring en positiv kendsgerning:

> En simulator findes virkelig.

Fjern nu de behagelige dele. Alice sender én bevisstreng og går sin vej. Der er ingen betroet opsætning, ingen fælles tilfældig streng, som er forberedt på forhånd, og Bob må aldrig acceptere en falsk opgave. Det er den ramme, klassisk nulviden ikke kan overleve i.

Der er brug for én figur mere før tricket. Fastlæg et **regelsæt**: et formelt bevissystem i logikerens forstand — et fast sæt aksiomer plus mekaniske regler til at kontrollere skrevne matematiske beviser. ZFC, matematikkens standardaksiomer, er det kanoniske eksempel. Alt herfra formuleres relativt til et regelsæt, der er valgt på forhånd, og valget er fleksibelt: Konstruktionen virker for ethvert regelsæt, man fastlægger, inklusive ZFC.

(En bemærkning om ord, lånt fra selve forskningsartiklen: ”Bevissystem” betyder her altid dette regelsæt — det formelle system, der kontrollerer matematiske beviser — aldrig de meddelelser, Alice sender. Alices og Bobs maskineri kaldes ”beviseren og verifikatoren”.)

Versionen i Gödels ånd bevarer historien om mega-sudoku, men ændrer beviset.

Vælg et andet betingelsessystem af samme viste størrelse, og kald det **D**. I historien er S og D to MegaSudoku(n)-opgaver i samme format. Bag kulisserne kan D være begyndt som en vanskelig logisk formel af en anden størrelse; om nødvendigt kan den fyldes ud med uskadelige dummybetingelser, så den passer i det samme gitter. D er bygget ud fra en logisk formel, som faktisk er **usatisfierbar**: Der findes ingen mulig værditildeling, som gør alle dens betingelser sande, ligesom en ødelagt opgave ikke har noget lovligt færdigudfyldt gitter. Et simpelt eksempel ville være en formel, der kræver både ”X er sand” og ”X er falsk”. D har altså ingen gyldig udfyldning.

Men D må ikke være en ødelagt opgave, der er *let at afsløre*. Det simple eksempel ovenfor fejler her: Ethvert regelsæt gendriver ”X og ikke-X” på én linje. D skal være falsk på en måde, som det valgte regelsæt ikke kan bekræfte med et kort argument. Hvis regelsættet kunne gendrive D med et kort bevis, ville historien nedenfor bryde sammen: Den alternative vej, der kunne have frembragt beviser uden Alices hemmelighed, kunne formelt udelukkes, og privatlivsgarantien ville forsvinde med den. Derfor vælges D fra en familie, som det fastlagte regelsæt ikke effektivt kan gendrive: Inden for dette regelsæt findes der intet kort bevis på, at D ikke har nogen løsning.

Alices bevis i én meddelelse handler derefter om et enten/eller-udsagn:

> enten har den virkelige mega-sudoku S en løsning, eller også har lokkemidlet D en løsning.

Dette er den logiske forbindelse. D genereres **ikke** på en magisk måde, der gør S sand. Beviset argumenterer ikke ”D har ingen løsning, derfor har S en løsning”. Det beviser disjunktionen **S eller D**. Perfekt sundhed siger, at en falsk disjunktion ikke kan have et gyldigt bevis. Eftersom D er falsk i virkeligheden — den har ingen løsning — kan disjunktionen kun være sand, hvis S er sand. Hvis beviset accepteres, må S altså have en løsning. Lokkemidlet kan ikke gøre en falsk S sand.

Men for den nulvidenslignende del: Spørg, hvad der ville ske, hvis D *faktisk* havde en løsning. Denne lokkemiddelløsning ville fungere som et alternativt vidne. Den ville lade nogen fremstille beviser uden at kende Alices virkelige mega-sudokuløsning — med andre ord en simulator. I virkeligheden har D ingen løsning, så denne vej til en simulator er lukket. Pointen er, at regelsættet ikke effektivt kan bevise, at den er lukket.

D har altså to opgaver. For **sundheden** er D falsk, så et gyldigt bevis for ”S eller D” fremtvinger S. For **effektiv nulviden** er D vanskelig at gendrive, og derfor kan regelsættet ikke hurtigt udelukke lokkemiddelvejen, som ville have gjort simulering mulig.

Sikkerhedstesten er dermed ikke længere:

> Kan vi bevise, at en simulator virkelig findes?

Den bliver:

> Kan dit regelsæt effektivt bevise, at simulatoren er umulig?

Hvis svaret er nej, følger noget overraskende stærkt: Enhver sikkerhedsgaranti, der (a) kan observeres ved at køre en test og (b) bevisligt følger — inden for dette regelsæt — af eksistensen af

en simulator, gælder faktisk. Et vellykket angreb på en af dem ville i sig selv udgøre den manglende korte gendrivelse, og den manglende korte gendrivelse findes ikke. Det er den "effektive" del af effektiv nulviden.

Kontrasten i klasseværelset er altså:

Klassisk nulviden: Beviserne er sikre, fordi en simulator findes.

Effektiv nulviden i Gödels ånd: Beviserne behandles som sikre for observerbare sikkerhedstest, fordi regelsættet ikke effektivt kan bevise, at simulatoren er umulig.

Det andet udsagn er svagere. Det er også derfor, artiklen kan bevare de tre egenskaber, der fik den klassiske version til at bryde sammen: én meddelelse, ingen opsætning og perfekt sundhed.

Den nye test: Man kan ikke bevise, at simulatoren mangler

Ilangos lempelse ændrer spørgsmålet.

Klassisk nulviden spørger:

> Findes der en simulator?

Effektiv nulviden spørger om noget svagere:

> Kan det valgte regelsæt effektivt bevise, at ingen simulator findes?

Det kan lyde som en teknisk undvigemanøvre, men det er kerneideen. Konstruktionen befinder sig i en mærkelig tilstand: En simulator findes faktisk ikke — artiklen siger dette udtrykkeligt — men det regelsæt, man har fastlagt, kan ikke effektivt bevise, at den ikke findes. Hvis enhver dårlig konsekvens, man bekymrer sig om, ville kræve en sådan gendrivelse, opfører systemet sig stadig som nulviden med hensyn til disse konsekvenser.

Det er her, Gödel kommer ind. Ikke som pynt og ikke som "Gödel gør krypto sikkert". Forbindelsen er bevisteoretisk. Et regelsæt kaldes **optimalt**, hvis det i en præcis forstand er det bedst mulige: Når et hvilket som helst regelsæt kan gendrive en formel af den relevante slags med et kort bevis, kan det optimale regelsæt også gøre det, med et bevis, der højst er polynomisk længere. Krajíček og Pudlák fremsatte i 1989 formodningen om, at **intet optimalt bevissystem findes**: Uanset hvilket regelsæt man fastlægger, beviser et andet regelsæt en eller anden familie af sande udsagn langt mere kortfattet. Dette er en af de centrale åbne formodninger i beviskompleksitet, og det er den endelige, kompleksitetsteoretiske slægtning til Gödels ufuldstændighedssætning: Nogle sande udsagn har intet kort bevis i det regelsæt, man fastlagde — ikke fordi de principielt er ubeviselige, men fordi ethvert fastlagt regelsæt efterlader nogle korte sandheder uden korte beviser.

Artiklen antager denne formodning (i en lidt stærkere "uendeligt ofte"-form, som er standard, når formodninger bruges kryptografisk). Udbyttet, gennem et teorem af Krajíček og Pudlák, er konkret:

For hvert regelsæt findes der en sekvens af formler, som virkelig er usatisfierbare, og som regelsættet ikke kan gendrive med korte beviser — og som en effektiv algoritme, helt afgørende, kan *generere*. Den sidste egenskab, uniformitet, er det, der forvandler hele ideen fra et eksistensudsagn til en faktisk algoritme, Alice kan køre: Hendes lokkemidler D kommer fra et samlebånd, ikke ud af den blå luft.

Det kryptografiske greb er at sætte denne mangel på beviskraft i arbejde.

Hvad konstruktionen gør

Her er konstruktionen i artiklen, reduceret til sin form.

Fastlæg et regelsæt — for eksempel ZFC. Under antagelsen fra beviskompleksitet findes der en effektivt genererbar sekvens af formler, som faktisk er usatisfierbare, men regelsættet har intet kort bevis for, at de er usatisfierbare.

Byg nu et bevis i én meddelelse med denne form:

> enten er det virkelige udsagn satisfierbart, eller også er denne særlige vanskelige formel satisfierbar.

Den særlige vanskelige formel er ikke satisfierbar. Hvis det underliggende bevismaskineri har perfekt sundhed, betyder accept af meddelelsen derfor stadig, at det virkelige udsagn er sandt. Det giver perfekt sundhed.

Men for den nulvidenslignende sikkerhed kan man forestille sig, at den særlige vanskelige formel *var* satisfierbar. Så kunne dens vidne bruges til at simulere beviser uden at kende det virkelige vidne. Formlen er ikke satisfierbar i virkeligheden — men regelsættet kan ikke effektivt bevise det. Derfor kan det ikke effektivt bevise, at simulatoren er umulig.

Det er hængslet. Systemet skjuler ikke hemmeligheden ved at frembringe en klassisk simulator. For en stor klasse af observerbare sikkerhedstest skjuler det hemmeligheden bag regelsættets manglende evne til at bekræfte, at simulatoren mangler.

Hvad artiklen hævder

Hovedteoremet kommer i flere lag. Kerneresultatet er dette:

Under en kryptografisk standardantagelse — eksistensen af **ikke-interaktive vidneuskelnelige beviser**, velstuderede objekter, der følger af flere etablerede antagelsespakker — og under den beviskompleksitetsteoretiske formodning om, at **intet (uendeligt ofte) optimalt bevissystem findes**, konstruerer artiklen for hvert valg af regelsæt en beviser og verifikator med én meddelelse for NP/SAT, med **perfekt sundhed** og uden opsætning, som er effektivt nulviden relativt til dette regelsæt. (NP/SAT er den sædvanlige ”sværeste fællesnævner” for opgavelignende problemer;

mega-sudoku er én forklædning, det bærer.)

For den bredere påstand om at bevare falsificerbare sikkerhedsegenskaber tilføjer artiklen endnu en standardantagelse, derandomiseringsopfattelsen **P = BPP** (groft sagt: Tilfældighed giver ikke algoritmer nogen væsentlig ekstra kraft).

Oversat fra teoremsprog:

- Beviset er én meddelelse.
- Der er ingen betroet opsætning.
- Falske udsagn kan ikke bevises.
- Beviseren er ikke klassisk nulviden — den har ingen simulator.
- Men enhver falsificerbar, spilbaseret sikkerhedskonsekvens af klassisk nulviden kan opnås i denne ramme.

”Falsificerbar” er vigtigt. Det betyder, at en sikkerhedsfejl kan testes ved at køre en modstander i et spil. Mange kryptografiske sikkerhedsdefinitioner har denne form: Kan modstanderen skelne mellem to krypteringer, invertere en funktion, genskabe et vidne eller vinde et bestemt eksperiment? Teoremet giver en beviser for hver falsificerbar egenskab, én ad gangen. En enkelt beviser, der har *alle* falsificerbare egenskaber på én gang, er sandsynligvis umulig — det gamle genbrugsangreb (”Bob kan vise beviset til andre”) er selv en falsificerbar egenskab, og den svigter faktisk her. Artiklens forslag er, at en enkelt beviser plausibelt kan dække alle *naturlige* falsificerbare egenskaber — dem, der faktisk forekommer i kryptografisk praksis — men denne del er et betinget teorem, der hviler på et uformelt begreb om ”naturlig”, plus en udtrykkelig formodning. Garantien er rettet mod observerbare fejl, ikke mod enhver filosofisk eller simuleringsbaseret betydning af hemmeligholdelse.

Én konkret korollar er værd at nævne: Konstruktionen giver de første ikke-interaktive *vidneskjulende* beviser med en uniform beviser — ”et bevis for en opgave hjælper dig ikke med at finde dens løsning”, uden interaktion og uden opsætning — et beskedent klingende objekt, der havde modstået konstruktion i årtier.

Hvad dette ikke siger

Dette er afsnittet, der holder teksten ærlig.

Det siger **ikke**, at de gamle umulighedsteoremer tog fejl. Konstruktionen undgår dem ved at ændre definitionen.

Det giver **ikke** almindelig, klassisk nulviden uden interaktion, uden opsætning og med perfekt sundhed. Artiklen siger udtrykkeligt, at den konstruerede beviser ikke har nogen simulator.

Det betyder **ikke**, at beviset ikke kan genbruges. Et bevis i én meddelelse kan stadig vises til en anden; artiklen bevarer ikke egenskaber af typen benægtelighed. (Ikke-interaktiv nulviden med betroet opsætning har den samme begrænsning.)

Det betyder **ikke**, at dette er en praktisk protokol, der er klar til ibrugtagning. Dette er kompleksitet-

steori og kryptografisk grundforskning. Resultatet afhænger af store antagelser fra beviskompleksitet og kryptografi, og konstruktionen handler om, hvad der principielt er muligt.

Det gør **ikke** "Gödel" til en magisk sikkerhedsprimitiv. Gödel-forbindelsen går gennem bevissystemer, optimale bevissystemer og endelige analoger til ufuldstændighed. Den brugbare intuition er ikke "ufuldstændighed beskytter din adgangskode". Den er: Hvis et regelsæt ikke effektivt kan bevise, at en simulator er umulig, kan angreb, der ville kræve et sådant bevis, blokeres på sikkerhedsdefinitionernes niveau.

Hvorfor det alligevel er interessant

Kryptografi omdanner ofte sværhed til sikkerhed. Faktorisering er svært, så antagelser af RSA-typen bliver nyttige. Gitterproblemer er svære, så gitterkryptografi bliver nyttig. Her er sværheden mere fremmedartet: ikke "svært at beregne en hemmelighed", men "svært at bevise, at et bestemt bevisobjekt ikke kan findes".

Det er derfor, artiklen føles usædvanlig. Den behandler aksiomer og regelsæt næsten som kryptografiske ressourcer. Den sædvanlige umulighed siger, at der er en spænding mellem sundhed og simulering. Ilangos greb er at placere spændingen bag et bevisteoretisk forhæng: Simulatoren mangler, men det formelle system kan ikke effektivt afsløre dette fravær.

For en læser er det overraskende ikke, at dette vil erstatte nutidens nulvidenssystemer. Det vil det sandsynligvis ikke, i hvert fald ikke direkte. Det overraskende er, at en begrænsning fra matematisk logik kan bruges konstruktivt: ikke blot som en mur, men som en form for dække.

Hvor stærk er evidensen?

Dette er en teoremartikel, så "evidens" betyder noget andet end i en artikel om biologi eller astronomi. Spørgsmålet er ikke, om et eksperiment er blevet gentaget. Spørgsmålet er, om definitionerne, antagelserne og beviskæden understøtter påstanden.

Beviset er formelt, og artiklen er tydelig om sine antagelser. Antagelserne er ikke tilfældige. Ikke-interaktive vidneuskelnelige beviser er standardobjekter i kryptografi og følger af flere etablerede antagelsespakker. Formodningen om, at intet optimalt bevissystem findes, er en central formodning i beviskompleksitet. P = BPP er en standardopfattelse om derandomisering, som kun bruges til det bredere teorem om falsificerbare egenskaber.

Artiklen argumenterer også for, at antagelserne er den rette pris, ikke et vilkårligt stillads: Den beviser en omvendt påstand, som viser, at de i det væsentlige er nødvendige — hvis konstruktioner som denne overhovedet findes, skal ikke-interaktive vidneuskelnelige beviser findes, og (givet almindelige envejsfunktioner) kan intet optimalt bevissystem findes. Og antagelserne er "vind-vind": At gendrive en af dem ville i sig selv være en skelsættende opdagelse i beviskompleksitet, kryptografi eller kompleksitetsteori.

Men fordi resultatet er betinget, er tilliden til det også betinget. Hvis disse antagelser svigter, ændres fortolkningen af teoremet. Og selv hvis antagelserne holder, er garantien ikke fuld klassisk nulviden; den er artiklens svækkede, bevisteoretiske version.

Den rette tillid er derfor høj til, at artiklen etablerer et sammenhængende, betinget mulighedsresultat; moderat til, at dens antagelser beskriver den kryptografiske verden, vi faktisk lever i; og lav med hensyn til umiddelbare praktiske konsekvenser.

Hvorfor det har betydning

Artiklen åbner en vej, der skulle være lukket.

Klassisk teori siger: Fuld nulviden kan ikke være én meddelelse uden opsætning og kan ikke have perfekt sundhed. Ilangos artikel siger: Hvis vi spørger efter de konsekvenser af nulviden, der kan testes i sikkerhedsspil, og hvis vi lader sikkerhedsdefinitionen afhænge af, hvad et regelsæt kan eller ikke effektivt kan gendrive, så kan meget af den nyttige adfærd genvindes — med én meddelelse, ingen opsætning og perfekt sundhed.

Det er ikke en lille definitionsændring. Det er en anden måde at tænke på kryptografiske garantier. I stedet for kun at spørge, hvad der findes, kan man spørge, hvad ens regelsæt kan udelukke. I stedet for at behandle ubeviselighed som en filosofisk gene kan man bruge den som struktur.

Den praktiske verden ændrer sig måske ikke i morgen. Men det gør det begrebsmæssige kort. Der findes nu en formel betydning, hvor ”ingen kan effektivt bevise, at hemmeligheden lækkede” kan være stærkt nok til at genvinde mange af de spilbaserede beskyttelser, vi ønskede fra ”hemmeligheden lækkede ikke”.

Det er derfor, Gödel hører hjemme i titlen.

Kort fortalt

Nulvidensbeviser lader en beviser overbevise en verifikator om, at et udsagn er sandt, uden at afsløre vidnet. Klassiske umulighedsresultater siger, at nulviden ikke kan presses ind i én meddelelse uden opsætning og ikke kan have perfekt sundhed. Rahul Ilangos artikel gendriver ikke disse umuligheder. Den definerer et svagere begreb, effektiv nulviden: I stedet for at kræve, at en simulator virkelig findes, kræver det, at et valgt bevissystem — et formelt regelsæt som ZFC — ikke effektivt kan bevise, at ingen simulator findes. Under store antagelser fra kryptografi (ikke-interaktive vidneuskelnelige beviser) og beviskompleksitet (intet optimalt bevissystem findes) konstruerer artiklen bevisere med én meddelelse for NP/SAT, uden opsætning og med perfekt sundhed, som opnår de falsificerbare, spilbaserede konsekvenser af nulviden egenskab for egenskab. En enkelt beviser, der dækker alle ”naturlige” sådanne egenskaber, er en yderligere, delvis formodningsbaseret udvidelse — og det er sandsynligvis umuligt at dække bogstaveligt talt enhver falsificerbar egenskab, fordi beviser stadig kan genbruges. Resultatet er teoretisk og betinget, ikke en ibrugtaget primitiv, men det viser en ny

måde at bruge bevisteoretisk ubeviselighed som en kryptografisk ressource på.

Uden omsvøb

Hvad artiklen viser: Under de angivne antagelser kan man bygge bevisere for NP/SAT med én meddelelse, uden opsætning og med perfekt sundhed, som er effektivt nulviden relativt til ethvert valgt bevissystem, og som opnår hver falsificerbar, spilbaseret konsekvens af klassisk nulviden.

Hvad der er plausibelt, men ikke bevist ubetinget: At de nødvendige antagelser fra beviskompleksitet og kryptografi holder. De er seriøse, velstuderede antagelser — og artiklen viser, at de i det væsentlige både er nødvendige og tilstrækkelige — men de er stadig antagelser.

Hvad den ikke viser: Klassisk nulviden uden interaktion, uden opsætning og med perfekt sundhed; et praktisk system klar til ibrugtagning; benægtelighed eller at beviser ikke kan genbruges; eller at Gödels ufuldstændighedssætning i sig selv sikrer kryptografi.

Vigtigste begrænsninger: Garantien er en lempelse af nulviden; den bredeste version afhænger af flere antagelser; påstandene om én universel beviser er stadig delvis formodningsbaserede; og resultatet er primært grundlæggende teori.

Hvor stor tillid bør en almindelig læser have? Høj tillid til, at dette er et vigtigt, betinget teoriresultat, hvis definitionerne accepteres. Moderat tillid til, at antagelserne indfanger virkeligheden. Lav tillid til umiddelbar praktisk ibrugtagning. Den sikre konklusion er: Artiklen bryder ikke umulighedsresultaterne for nulviden; den finder en ny bevisteoretisk vej uden om de dele af dem, der har betydning for mange sikkerhedsspil.

Kilder

Ilango, IACR ePrint 2025/1296
`https://eprint.iacr.org/2025/1296`

FOCS 2025, DOI 10.1109/FOCS63196.2025.00058
`https://doi.org/10.1109/FOCS63196.2025.00058`