Kako dokazati, da poznate skrivnost, ne da bi jo razkrili?

Samir Hodžić

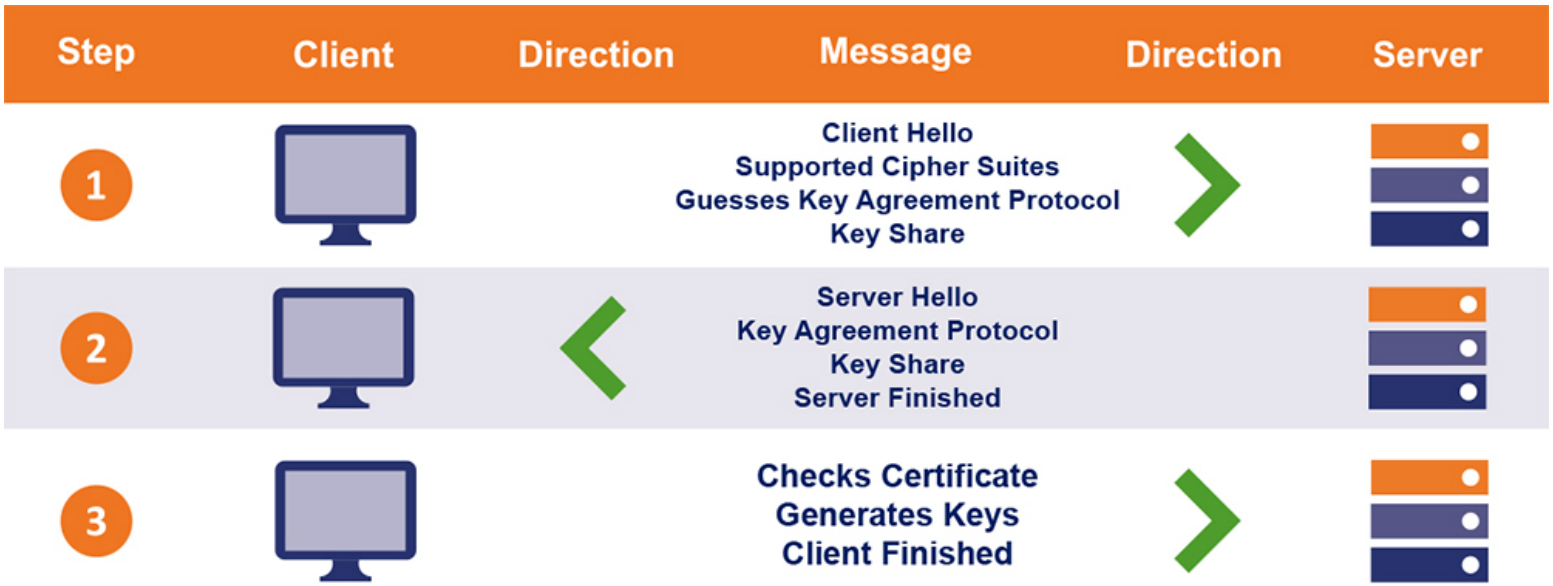
Univerza na Primorskem, FAMNIT

Pregled

- Kriptografski protokoli
- Protokoli, ki se temeljijo na dokazih brez znanja
- Primeri protokolov

Kriptografski protokol?

- **Kriptografski protokol** - omogočajo varno komunikacijo z uporabo kriptografskih metod
- **Primer:** *Transport Layer Security* (TLS) - zagotavlja varno komunikacijo v računalniškem omrežju
  (**Drugi**: IPSec, SSL, SSH, S/MIME, OpenPGP/GnuPG/PGP, Kerberos)



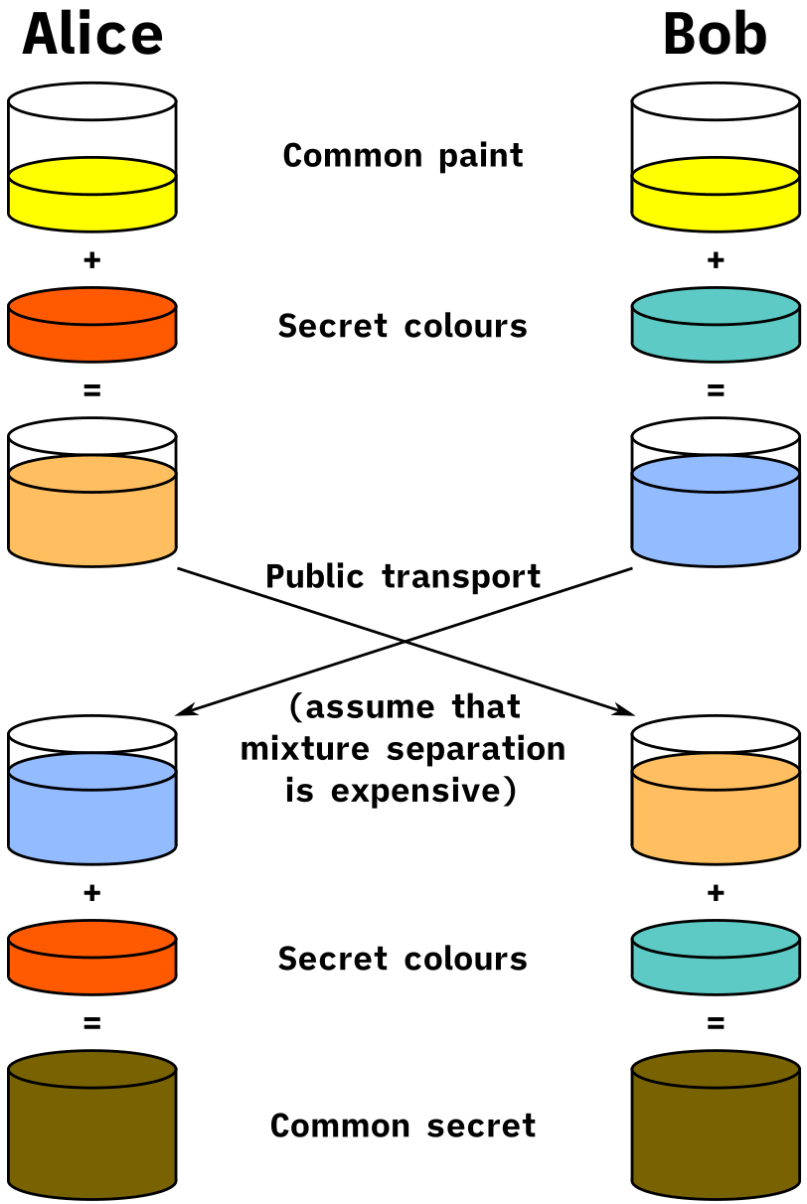
Kriptografski protokol ima običajno vsaj nekatere od teh lastnosti:

- **Izmenjava ključev**
- **Prepoznavanje/avtentikacija entitete** (Ali se Alice res pogovarja z Bobom?)
- **Vzpostavitev metod za šifriranje in avtentikacijo sporočil**
- **Metode brez zavrnitve** (Alice pošlje sporočilo Bobu. Alice kasneje ne more zanikati da je poslala to sporočilo, in Bob ne more zanikati da je prejel sporočilo.)
- **Metode deljenja skrivnosti**
- **Varno večstransko računanje**

Izmenjava ključev

- **Protokoli za izmenjavo ključev** - vzpostavitev skupnega ključa, ki ga lahko uporabita za šifriranje ali podpisovanje podatkov
- Protokoli, kjer **obe strani** vplivata na končni izpeljani ključ, so edini način za izvajanje popolne skrivnosti
- **Primer:** Diffie-Hellman key exchange protocol (1976)
- Whitfield Diffie and Martin Hellman - **Turing Award 2015**, za temeljne prispevke k moderni kriptografiji

Diffie-Hellman protokol (ilustracija procesa)



Diffie-Hellman protokol - matematični opis

$$(g^a)^b = (g^b)^a = g^{ab} \pmod{p}$$

To everybody: We choose g = 23 and p = 9719

A to B: here's $g^a = 7053$

B to A: here's $g^b = 4800$

PRIVATE
Alan:
$(g^b)^a = 4800^a$
$= 1195$

PRIVATE
Betty:
$(g^a)^b = 7053^b$
$= 1195$

Težave?

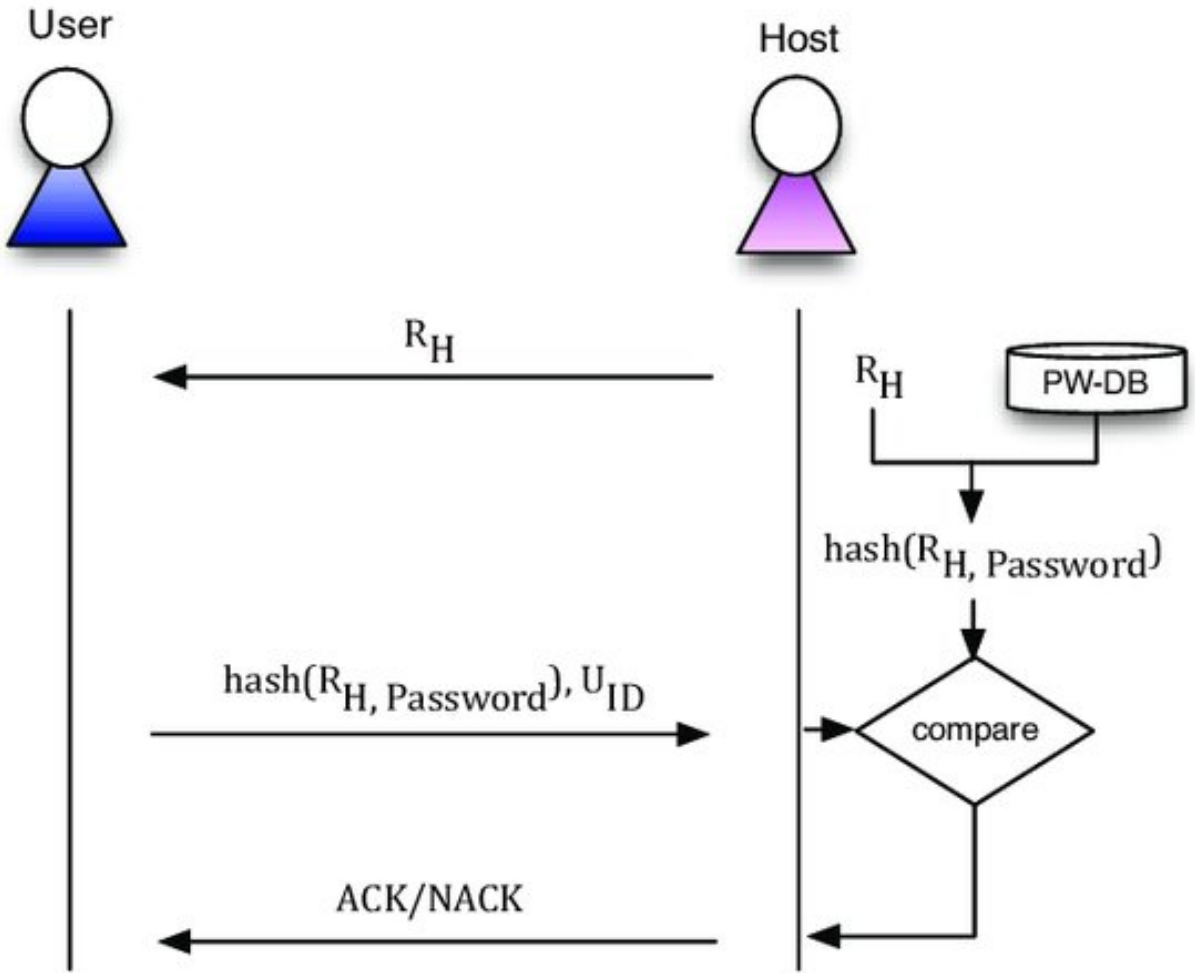
Številni protokoli uporabljajo Diffie-Hellmanov algoritem za izboljšanje varnosti: Secure Shell (SSH), TLS, Public Key Infrastructure (PKI), Internet Protocol Security (IPSec)...

Omejitve DH-algoritma:

- Algoritem se lahko uporablja samo za simetrično izmenjavo ključev
- Pomanjkanje postopka avtentikacije
- Ker ni vključena avtentikacija, je algoritem ranljiv za t. i. napad "človeka v sredini"
- Ker je računsko intenziven, je drag v smislu časa delovanja CPU - S tem algoritmom ni mogoče izvesti šifriranja informacij
- Digitalnega podpisa ni mogoče podpisati z Diffie-Hellmanovim algoritmom

Avtentikacija bazirana na izzivu

- **Postopek preverjanja identitete** - ena stranka predstavi vprašanje (izziv), in druga stranka mora zagotoviti veljaven odgovor
- **Primeri:** Sistemi pametnih kartic, CAPTCHA, biometrični sistemi...
- **Naključno generirana informacija** (ki je del izziva) pri vsaki izmenjavi (in kjer je odgovor drugačen od izziva) ščiti pred možnost napada ponovnega predvajanja (človeka v sredini)


User
Host
R_H
R_H
PW-DB
$hash(R_{H,\ Password})$
$hash(R_{H,\ Password}), U_{ID}$
compare
ACK/NACK

(Spet) Težave?

- **Statični izzivi** - Skrito vprašanje se ne spreminja prepogosto
- **Izbira gesla** - uporabniki uporabljajo in reciklirajo svoja gesla v več digitalnih računih
- Spet možnost napada "človeka v sredini" (zbiranje dovolj delnih informacij o skritoj informaciji)
- Poslana informacija (odgovor izziva) lahko vsebuje nekaj, kar je vezano za uporabnikovo skrivnost

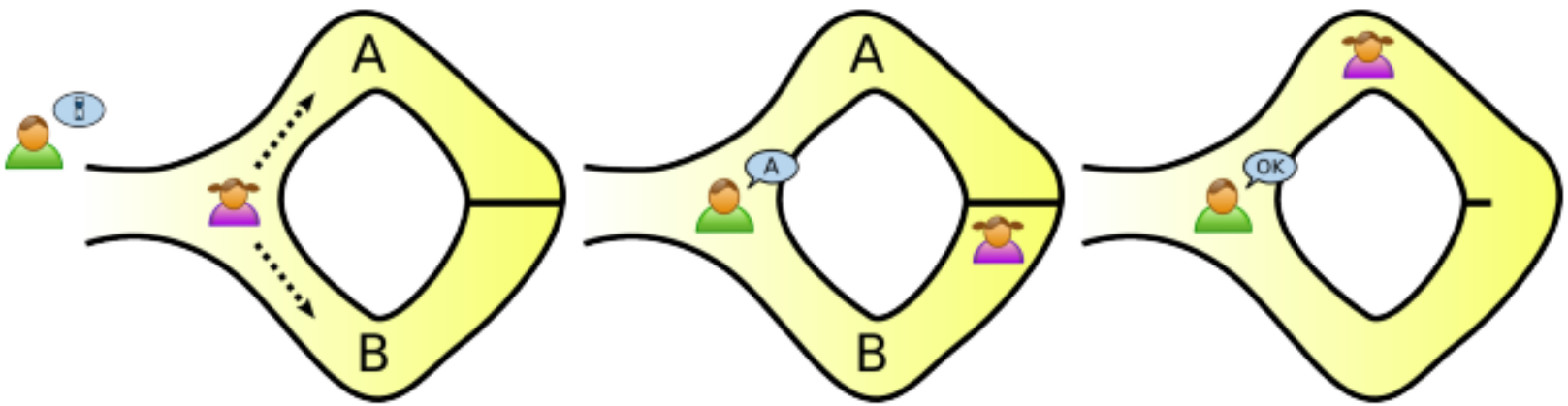
Dokazi brez znanja (Zero-knowledge proofs)

- **Dokazevalec (prover)** bo poskušal dokazati poznavanje določene skrivnosti **preveritelju (verifier)**
- **Cilj** - dokazovanje poznavanja določenega dejstva, brez razkrivanja samega dejstva ali kakršnekoli dodatne informacije
- **Tehnika je bila predstavljen** 1985 - Shafi Goldwasser, Silvio Micali, in Charles Rackoff ("The Knowledge Complexity of Interactive Proof-Systems")

Micali Goldwasser Rackoff

Primer 1 - Ali Babina jama

- Primer predstavlja temeljno idejo dokazov brez znanja - (1989, Jean-Jacques Quisquater et al.)
- Ali Babina jama vsebuje skrivna vrata, ki se lahko odprejo s skrivnim geslom
- Peggy pozna geslo za vrata in želi prepričati Victorja da ona to ve, vendar noče, da bi Victor vedel kakšno je geslo



Delujeta na naslednji način:

- Peggy gre nakljućen krak jame. Victor stoji zunaj jame in ne ve v kateri Petra bo šla
- Victor pride v jamo in pove nakljućen krak jame (levi ali desni), v katerega bi morala priti Peggy
- Če Peggy pozna skrivno geslo, lahko vsakič pride na pravi krak jame (ker po potrebi uporabi geslo da odpre vrata)
- Če Peggy ne ve gesla, ima 50-odstotno možnost, da bo na začetku šla v napačen krak, in ker ne more mimo skrivnih vrat, Victor jo lahko razotkri

To ponovita tolikokrat, kolikor je potrebno da prepričata Victorja.

Če bo Victor zadovoljen z 1/1024 možnosti da Peggy ne pozna gesla, potrebujejo 10 ponovitev ($2^{10} = 1024$)

Primer 2 - Dve žogici in barvno slepi prijatelj

- Predstavljajte si, da je vaš prijatelj rdeče-zelen daltonist (vi pa niste) in imate dve žogici: eno rdečo in eno zeleno, sicer pa identični
- Vašemu prijatelju se zdita popolnoma enaki in je skeptičen, da se dejansko razlikujeta
- Želite mu dokazati, da sta v resnici različno obarvani (in nič drugega)
- **Dokaz:** Prijatelju date dve žogici in on ju položi za hrbet
- Nato vzame eno od žogic in jo pokaže izza hrbta ter jo pokaže
- Potem spet položi za hrbet in se odloči razkriti samo eno od dveh žogic, pri čemer izbere eno od obeh naključno z enako verjetnostjo

- Vprašal vas bo: Ali sem zamenjal žogo?
- Celoten postopek se ponovi tolikokrat kot je potrebno
- Če pogledate njihove barve, lahko seveda z gotovostjo ugotovite, ali jih je zamenjal ali ne
- Po drugi strani, če bi bili enake barve in zato nerazločljivi, nikakor ne bi mogli pravilno uganiti z verjetnostjo, višjo od 50%
- Verjetnost da bi vam naključno uspelo identificirati eno zamenjavo žogic je 50%, in verjetnost da bi vam naključno uspelo pri vseh zamenjavah, je skoraj nič

Primer 3 - Sudoku

- Sudoku je uganka, ki vključuje mrežo kvadratov 9×9, kjer lahko vsak kvadrat vsebuje številke od 1 do 9
- Če želite rešiti uganko, morate izpolniti vse celice tako, da vsaka vrstica in stolpec vsebuje številke od 1 do 9 brez ponovitev
- Poleg tega ima mreža devet razdelkov 3×3, od katerih vsak mora vsebovati številke od 1 do 9

	2		6		8			
5	8				9	7		
				4				
3	7					5		
6								4
		8					1	3
				2				
		9	8				3	6
			3				9	

1	2	3	6	7	8	9	4	5
5	8	4	2	3	9	7	6	1
9	6	7	1	4	5	3	2	8
3	7	2	4	6	1	5	8	9
6	9	1	5	8	3	2	7	4
4	5	8	7	9	2	6	1	3
8	3	6	9	2	4	1	5	7
2	1	9	8	5	7	4	3	6
7	4	5	3	1	6	8	9	2

Dokaz brez znanja - Sudoku

Alice želi Bobu dokazati da ima rešitev za uganko Sudoku, vendar ji Bob ne verjame. **Koraki dokaza (za eno ponovitev):**

- Alice ustvari permutacijo vnosev: $1 \to 3$, $2 \to 1, \ldots$
- Poleg tega generira naključno zaporedje bajtov (nonce) za vsako sudoku celico

1	→	3
5	→	2
9	→	7
3	→	9
6	→	4
4	→	5
8	→	6
2	→	1
7	→	8

c1	aa	f3	4d	b8	e6	de	ad	c0
b2	13	4a	10	9f	e0	be	ef	ff
3f	2b	0d	3a	d8	af	9a	e1	3e
c3	a6	f1	d5	4a	39	2a	65	7f
0e	52	12	20	63	9f	ec	89	de
4e	2b	6e	8e	7f	12	40	fa	a1
0a	9c	6f	4f	0b	5	36	25	f0
19	0d	74	65	b1	bf	5e	a9	c8
8f	bb	2c	9e	ea	4a	a0	b3	14

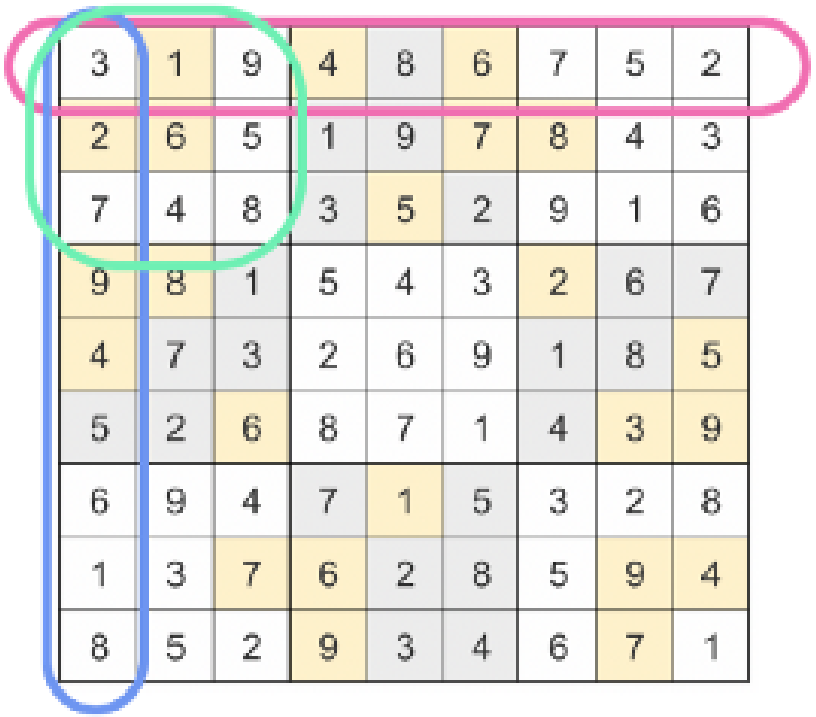
- Uporabi preslikavo na številke svoje sudoku rešitve, da dobi maskirano rešitev
- **Opomba:** Permutiranjem vrednosti številke še vedno pojavijo samo enkrat, saj gre za preslikavo ena proti ena

1	↦	3
5	↦	2
9	↦	7
3	↦	9
6	↦	4
4	↦	5
8	↦	6
2	↦	1
7	↦	8

1	2	3	6	7	8	9	4	5
5	8	4	2	3	9	7	6	1
9	6	7	1	4	5	3	2	8
3	7	2	4	6	1	5	8	9
6	9	1	5	8	3	2	7	4
4	5	8	7	9	2	6	1	3
8	3	6	9	2	4	1	5	7
2	1	9	8	5	7	4	3	6
7	4	5	3	1	6	8	9	2

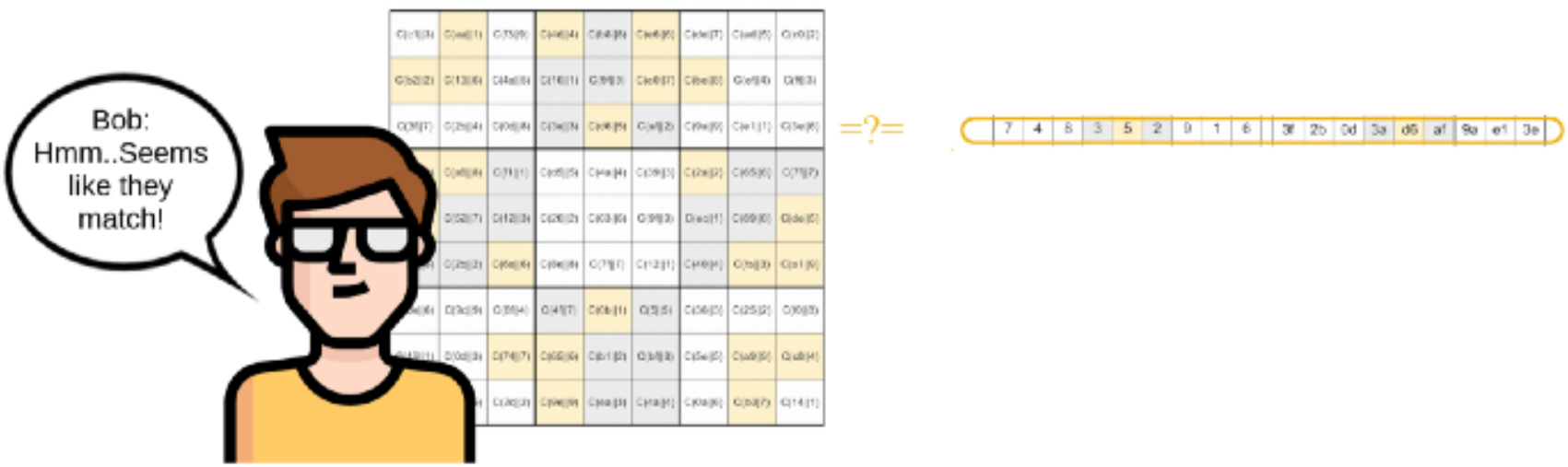
3	1	9	4	8	6	7	5	2
2	6	5	1	9	7	8	4	3
7	4	8	3	5	2	9	1	6
9	8	1	5	4	3	2	6	7
4	7	3	2	6	9	1	8	5
5	2	6	8	7	1	4	3	9
6	9	4	7	1	5	3	2	8
1	3	7	6	2	8	5	9	4
8	5	2	9	3	4	6	7	1

- Alice razdeli maskirano rešitev na nize števil iz vsake vrstice, stolpca, podmreže in niz znanih števil, ki so bile najprej del definicije uganke
- Nato ustvari "zavezo" za maskirano rešitev tako, da *zgosti* vsako celico z ustreznim bajtom, to zavezo pošlje Bobu in prosi Boba, naj izbere vrstico, stolpec, podmrežo ali nabor znanih števil



3	1	9	4	8	6	7	5	2
2	6	5	1	9	7	8	4	3
7	4	8	3	5	2	9	1	6
9	8	1	5	4	3	2	6	7
4	7	3	2	6	9	1	8	5
5	2	6	8	7	1	4	3	9
6	9	4	7	1	5	3	2	8
1	3	7	6	2	8	5	9	4
8	5	2	9	3	4	6	7	1

- Bob izbere eno in Alice mu pošlje enkratne in spremenjene številke, ki ustrezajo Bobovi izbiri
- **V primeru**, da je Bob izbral seznam znanih številk, mu Alice pošlje tudi preslikavo ena proti ena, ki jo je ustvarila sprva
- Bob nato preveri, ali se spremenjene vrednosti dejansko pojavijo samo enkrat, in ponovno ustvari zavezo z uporabo nonces, da preveri Alicino zavezo
- **V primeru**, da je Bob izbral seznam znanih številk, preveri tudi, ali so preslikave tudi pravilne



Lastnosti dokaza brez znanja

Splošne lastnosti:

- **Preveritelj se ne more ničesar naučiti iz protokola** (samo tisto kar bi lahko izpeljal iz javne informacije)
- **Dokazovalec ne more goljufati preveritelja (in obratno)** - Verjetnost, da dokazovalec prevara preveritelja se lahko naredi nizko kot je potrebno
- **Preverjevalec se ne more pretvarjati da je dokazovalec tretji osebi**

Posebne lastnosti:

- **Interaktivni sistem dokazovanja** (prejšnji primeri) - Preverjevalec in dokazovalec izmenjujeta več sporočil (izzivi in odgovori)
- **Dokazi so bolj verjetnostni kot absolutni** - Dokaz mora biti le pravilen z omejeno verjetnostjo (blizo 1)

Formalno - Kaj je dokaz brez znanja?

Dokaz je brez znanja, če ima naslednje 3 lastnosti:

- Popolnost (Completeness): Če so informacije resnične, mora dokazovalec prepričati preveritelja da govori resnico
- Ustreznost (Soundness): Če so informacije napačne, ne bi smel dokazovalec prepričati preveritelja da govori resnico
- Ničelno znanje (Zero-knowledge): Metoda mora preveritelju razkriti nič drugega kot to, ali dokazovalec govori resnico ali ne

Načini delovanja dokaza brez znanja:

- **Interaktivni dokazi** - Večkratna izmenjava sporočil
- **Neinteraktivni dokazi** - Dokazilo, ki ga predloži dokazovalec, lahko se preveri samo enkrat in kadar koli

Fiat-Shamir identifikacijski protokol

- Bomo predstavili osnovno različico Fiat-Shamir identifikacijskega protokola (V praksi bi uporabili učinkovitejšo različico)
- Osnovna različica predstavlja splošno idejo o dokazu brez znanja
- **Cilj:** Dokazovalec **A** se želi identificirati z dokazovanjem poznavanja skrivnosti s kateremu koli preveritelju **B**
- Varnost protokola se temelji na težavah pri izračunu kvadratnih korenov po modulu velikih sestavljenih celih števil n neznane faktorizacije

Fiat-Shamir osnovni id. protokol

POVZETEK: A dokazuje poznavanje skrite vrednosti s osebi **B** pri t izvedbah protokola

- **Enkratna postavitev:**
  - Zaupanja vredna entiteta **T** izbere in objavi modul podoben RSA kripto sistemu, $n = pq$, vendar ohrani praštevili p in q kot skrivnost
  - Vsak udeleženec **A** izbere skrivnost s soprosto z n $(1 \leq s \leq n-1)$, in potem izračuna $v = s^2 \ (mod\ n)$ ter registrira v pri **T** kot lastni javni ključ
- **Sporočila protokola:** Vsak od t krogov ima tri sporočila z naslednjim obrazcem:

$$\begin{aligned} \mathbf{A} \rightarrow \mathbf{B}: &\quad x = r^2 \ (mod\ n) \\ \mathbf{A} \leftarrow \mathbf{B}: &\quad e \in \{0,1\} \\ \mathbf{A} \rightarrow \mathbf{B}: &\quad y = r \cdot s^e \ (mod\ n) \end{aligned}$$

- **Dejstvo protokola:** Naslednji koraki se ponovijo t-krat (zaporedoma in neodvisno). **B** sprejme dokaz, če so vsi t krogi uspešni:

  1. **A** izbere naključno (zavezanost) r $(1 \leq r \leq n-1)$, in pošlje $x = r^2 \ (mod\ n)$ osebi B
  2. **B** naključno izbere (izziv) bit $e = 0$ ali $e = 1$ in pošlje e osebi **A**
  3. **A** izračuna in pošlje B (odgovor) y, ali $y = r$ (če je $e = 0$) ali $y = rs \ (mod\ n)$ (če je $e = 1$)
  4. **B** zavrne dokaz, če je $y = 0$, in sicer sprejme po preverjanju $y^2 = x \cdot v^e \ (mod\ n)$.
  5. **Opomba:** Odvisno od e, $y^2 = x$ ali $y^2 = xv \ (mod\ n)$, saj je $v = s^2 \ (mod\ n)$. Za $y = 0$ izključuje možnost $r = 0$.

Pokazani protokol je dokaz brez znanja, ker vsebuje naslednje lastnosti:

- **Popolnost:** Recimo, da ima dokazovalec skrivnost s. Potem lahko preveritelju vedno odgovori pravilno $y = r$ ali $y = rs \ (mod \ n)$. Zato bo pošten preveritelj dokončal vse t iteracij in sprejel dokaz z verjetnostjo 1
- **Ustreznost:** Recimo, da dokazovalec nima skrivnosti s. Potem v katerem koli krogu, lahko zagotovi samo enega od $y = r$ ali $y = rs \ (mod \ n)$. Zato bo pošten preveritelj v vsakem krogu zavrnil odgovor z verjetnostjo $1/2$, kar daje da bo preveritelj zaveden z verjetnosti 2^{-t}
- **Ničelno znanje:** Edina informacija razkrita v vsakem krogu je $x = r^2 \ (mod \ n)$, in ali $y = r$ ali $y = rs \ (mod \ n)$. Takšni pari (x, y) lahko simuliramo tako da naključno izberemo y in nato definiramo $x = y^2$ ali $x = y^2/v$. Takšni pari se računsko ne razlikujejo od interakcije s dokazovalcem

Druge aplikacije:

- **Elektronsko glasovanje** - J. Groth, "Non-interactive zero-knowledge arguments for voting", 2005
- **Jedrska razorožitev** - S. Philippe et al., "A physical zeroknowledge object-comparison system for nuclear warhead verification", 2016
- **Kriptovalute** - Zasebnosti in anonimnosti njihovih transakcij (zk-SNARKs, zk-STARKs, Bulletproofs,...)
- **3. generacija zaščitenega dostopa Wifi (WPA3, Dragonfly protokol**, 2020) - Uporablja dokaze brez znanja da se izogne posredovanju elementov gesla prek omrežja
- **Sistemi za avtentikacijo** (Sigma protokol, 2015)
- ...

Več lastnosti I

- **Matematične osnove:** Protokoli brez znanja temeljijo se na kripto-matematiki, npr. modulo izračun, diskretna matematika, izjemno velika števila (stotine ali na tisoče bitov),...
- **Kriptografska moč** protokolov brez znanja se temelji na nekaj težko rešljivih problemov:
  - Problem reševanja diskretnih logaritmov za velika števila (na stotine/tisoče bitov)
  - Problem vedeti, ali je dano število kvadrat (nekaterega števila) *mod n* ali ne, če ne poznate faktorizacijo od *n*
  - Problem faktorizacije velikih števil, ki so produkti dveh ali več velikih (stotine/tisoče bitov) praštevil
  - Dokazovalec mora biti časovno polinomsko omejen - drugače ne bi bilo preveč smiselno dokazevati svoje znanje

Več lastnosti II

Dokazi brez znanja v primerjavi z drugimi asimetričnimi protokoli:

- **Brez poslabšanja z uporabo:** Protokoli za katere je dokazano da imajo lastnost "brez znanja", ne trpijo poslabšanje varnosti z večkratno uporabo in se upre napadom *izbranega besedila* (chosen-text attacks)
- **Izogibanje šifriranju:** Številne tehnike dokazevanja "brez znanja" se izogibajo uporabi eksplicitnih algoritmov šifriranja (politične prednosti, npr. v zvezi z nadzorom izvoza)
- **Učinkovitost:** Ponavadi dokazi brez znanja imajo običajno višjo komunikacijske in/ali računske "stroške" kot asimetričnimi protokoli
- **Nedokazane predpostavke:** Mnogi protokoli "brez znanja" se zanašajo na nedokazanih predpostavkah kot asimetrične tehnike (npr. nesprejemljivost faktoringa ali kvadratna reziduoznost)

Hvala za vašo pozornost!