#### Optimal Verification of Operations on Dynamic Sets

Charalampos Papamanthou, UC Berkeley Roberto Tamassia, Brown University Nikos Triandopoulos, RSA Labs & BU



**CRYPTO 2011** 08/15/11

# Data in the cloud

- Data privacy
  - Server wants to learn our data
  - Can we enable the server use encrypted data in a meaningful way?
  - Computing on encrypted data
- Data and computations integrity
  - Server wants to tamper with our data
  - Are answers to queries the same as if the data were locally stored?
  - Authenticated data structures
  - Verifiable delegation of computation



### Verifying outsourced computation

Gmail	Search M	tail Search the Web Could a Mar	1 Settings 1 Older, version 1 Help 1 Sign.
Compose Mall	Read items from any RSS or Alom feed righ	chere. Customize Clos	Grad Tar 🗶 💌
indexs (7.640	Archive   Report Spam   Delete	More Actors + Refresh	1-50 of 1643 (Hold) (Hold)
Elacted 🛱	Select All None Read Unread Starved U		
Chata Ch	Catarior58	determine time - Just sharing pictures of balan. May	the set there a second set of the 14
Sent Mail	C Acema Balai	yoo - yoos share hav goes it as wanted to say in	
Staffa (14)	E Lesle Cator (2)	Join my network on Linkedin - Krain Rivers, a Colley	
M.Mat	Catanonda	the subject - five found a few jobs for me but nothing	
Sparts Staats	California (5)	New stuff needs your critique - Reduce the size of the	
19830	Daniel Chen	Merry Christmas and Happy New Year - Illion overyo	
Contacta	C David Carmona	Join my network on Linkedin - David Carmona, a Cla	
* Chat	T Karin Andrada	Join my network on Linkedin - Karty Andrade, a Cole	
bearch, add, or invite	C Quiu com	Welcome to Guru com - Duru com i Welcome to Guru	
· State 5	The Atlantitudes Alumni	The Art institutes Alumni News Letter - This message	
Satistation here 17	C Denis Osmanbegovic	resume - here's stume de a beder thend, newshound	
e falloutir11	Elexis, me (4)	the subject - http://users.heerournal.com/ buildcore/	
Working for the man	alex park (2)	Travavideo Christmas Party - Pri - \$10.00 is a suppr	and a second second
Ryan Pollock	C Dechen Wangdi	Check out my Facebook profile - lacebook Dechen W	
Anthony Phusekowsk	Cacanion66	Ped: FW: Got my tree up at last night Check out A	
a bhutanboy@gmail.or	randrade@gmail.com	TV Corporate Website Copy - fve shared a document	
ù Dev	F Rebecca Petter	Fail 2007 Portfolio Show - Hey Shame, This is Berlin,	
u Hark Duderstadt	C Dessart Donna	The Hotest Stop on the Linel - Dear A/CASF Aument	
a honor gunday	C des and dot	Evite: Reminder for A Surprise Birthday Party for Eug	
u Illepooky u Ricardo Andrade	C Sharon Kalmer	Fwd. COO1 needs people talented and preferably	demented people - Arts 11/19/07
u Sara C	C Southwest Airlines	Tickedess Travel Passenger linerary - Southwest Air	Snee Travel Rinerary The 11/16/07
Cuters V All Control	Linkedin Updates	Linkedin Network Updates, 11/17/2007 - Linkedin Nil	TWORK UPDATES PM 19/7/02
Property - Operation	me. Elexis (3)	N - holy trap. Greyhound's relinuteurs. It's ponna soal	me almost \$40 to go to 2 11/15/07
Labels	A-Jay Jimenez (5)	Come to my next event : Praggle Rock! This Friday, N	low, 16 2007 1 - Blasse thi 15/12/07
tmp_acables.(33)	(T me, Sara (2)	Hi sara chun - Cooce Petities, On Nov 8, 2007 11:00	AM, Shane S. Kalukrick 15/1507
Editabels	Cacamiondill, me (4)	whes - I don't know. I think fit be prety miserable oth-	er way. On Nov 9, 2007   11/9/07
v invite a triend	Exets.me (7)	(no subject) - you always let me win, that's no fun. On	Nov 9, 2007 2:50 PM, 8 15607
a state a state of	The Assessed Barbert	an example the mental second second second	And the second second second

- Conjunctive queries
  - Emails that have the terms "Brown" and "Berkeley"
- Disjunctive queries
  - Emails that have the terms "thesis" or "publication"
- All these queries boil down to set operations!

Google

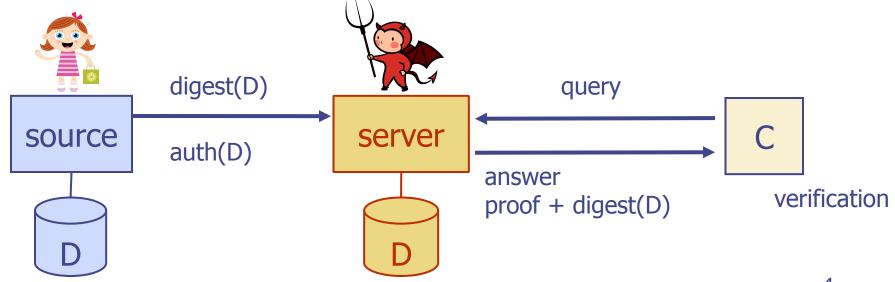
### Authenticated data structures model

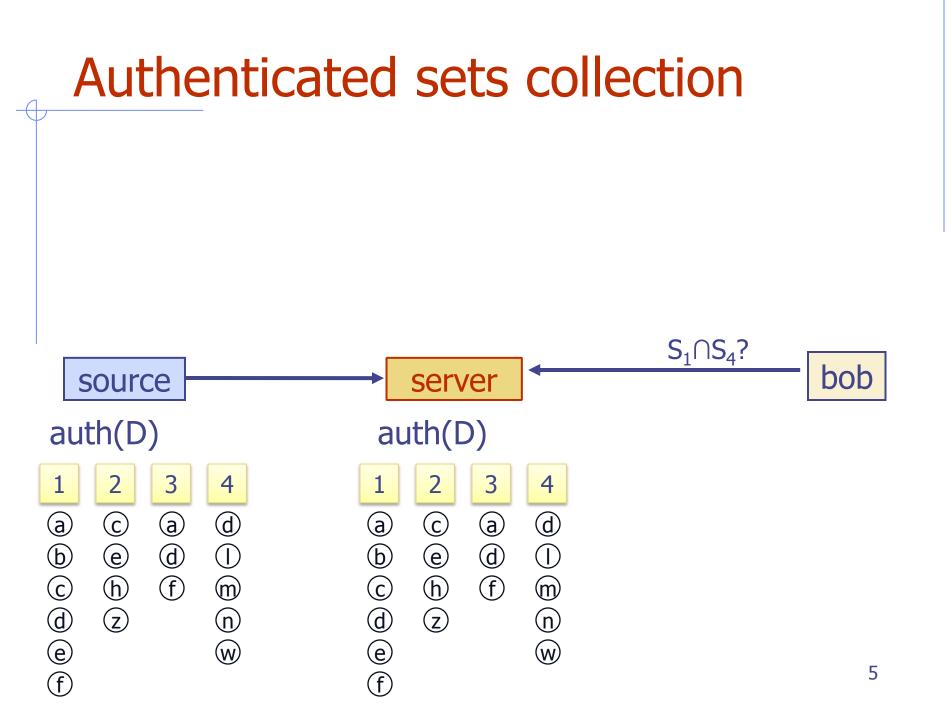
### Complexity

- Update at source and server
- Query at server
- Verification at client
- Size of proof
- Space

### Security

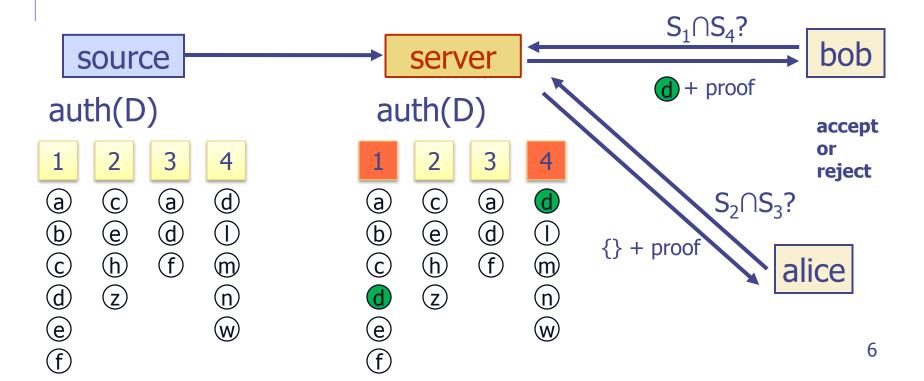
- A poly-bounded adversary cannot construct invalid proofs except with negligible probability
- Need for computational assumptions





### Queries on sets

- m: number of sets (e.g., m = 4)
- M: sum of sizes of **all** the sets (e.g., M = 6 + 4 + 3 + 5 = 18)
- t: number of queried sets (e.g., t = 2)
- $\delta$ : number of elements contained in the **answer** (e.g.,  $\delta = 1$ )
- n: the sum of sizes of the queried sets (e.g., n = 6 + 5 = 11)



# Related work and comparison

- Optimal proof size and verification time: O(δ)
- Linear space: O(m + M)
- Efficient queries and updates
- Performance comparison for the intersection of c = O(1) sets

	space	query	proof	assumption
D+04 YP09	m + M	n + log m	n + log m	Generic CR
M+04	m + M	n	n	Strong RSA
PT04	m <sup>c</sup>	1	δ	Discrete log
PTT10	m + M	n log <sup>3</sup> n + m <sup>ε</sup> log m	δ	Bilinear q- strong DH

# Our solution: Sets and polynomials

- Set X with n elements  $X = \{x_1, ..., x_n\}$
- Set Z is the intersection of X and Y
- The intersection of X and Y is empty, i.e.,
   X ∩ Y = Ø

Polynomial X(s) in Zp

 $X(s) = (s+x_1)...(s+x_n)$ 

- Polynomial Z(s) is the GCD of X(s) and Y(s)
- X(s) and Y(s) have GCD equal to 1, i.e., gcd(X(s),Y(s)) = 1

#### $\Leftrightarrow$

 There are polynomials P(s) and Q(s) such that P(S)X(s) + Q(s)Y(s) = 1

### Cryptographic tools we use

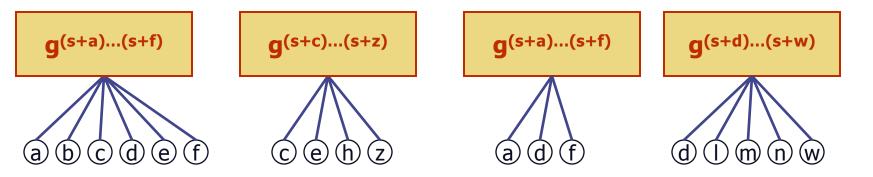
- Two multiplicative groups G and T of prime order p
- g is a generator of G
- A bilinear map e(.,.) from G to T such that
  - e(g<sup>a</sup>,g<sup>b</sup>) = e(g,g)<sup>ab</sup> for all a,b in Zp
  - e(g,g) generates T
- Bilinear q-strong Diffie Hellman Assumption
  - Pick a random s in Zp
  - s is the trapdoor
  - Compute g<sup>s</sup>, g<sup>s<sup>2</sup></sup>, g<sup>s<sup>3</sup></sup>,..., g<sup>s<sup>q</sup></sup>
  - The public key pk are the values g<sup>s</sup>, g<sup>s<sup>2</sup></sup>, g<sup>s<sup>3</sup></sup>,..., g<sup>s<sup>q</sup></sup>
  - The probability that a PPT Adv can find an a in Zp and output the tuple (a,e(g,g)<sup>1/(s+a)</sup>) is negligible

# Bilinear-map accumulator

- G and T of order p have a map e(.,.)
- X={x,y,z,r} in Z<sub>p</sub>
- Base  $g \in G$ , generator of G
- Secret  $s \in Z_p$
- Digest
  - $D = g^{(x+s)(y+s)(z+s)(r+s)}$
- Witness for x
  - $W_x = g^{(y+s)(z+s)(r+s)}$
- Verification
  - $e(D,g) = e(W_x,g^{(x+s)})?$
- Security: q-strong Diffie-Hellman assumption
- [Nguyen (05)]

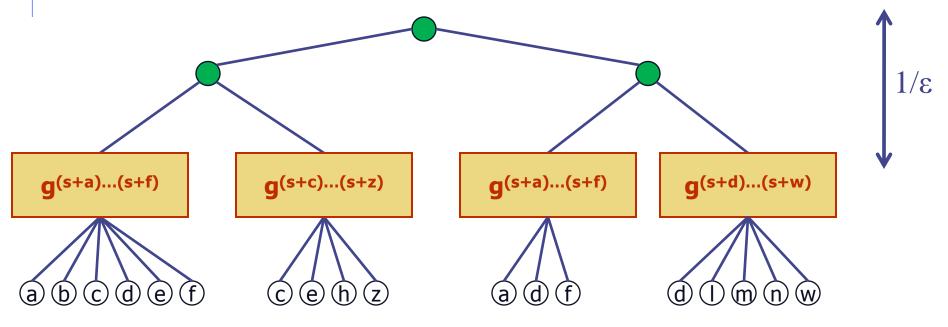
### Our construction

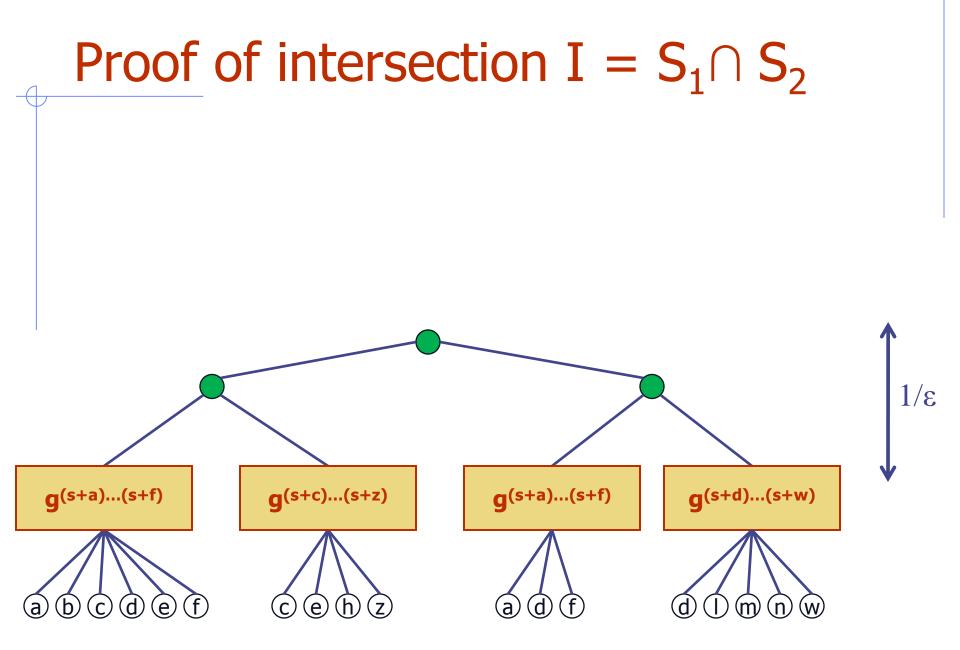
• Compute the accumulation value for every set

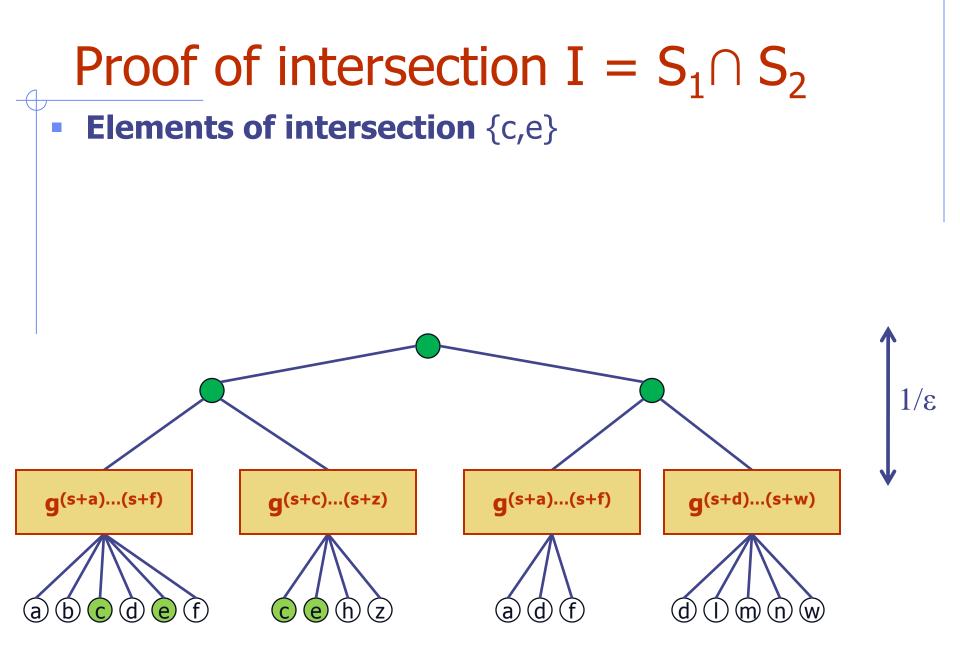


### Our construction

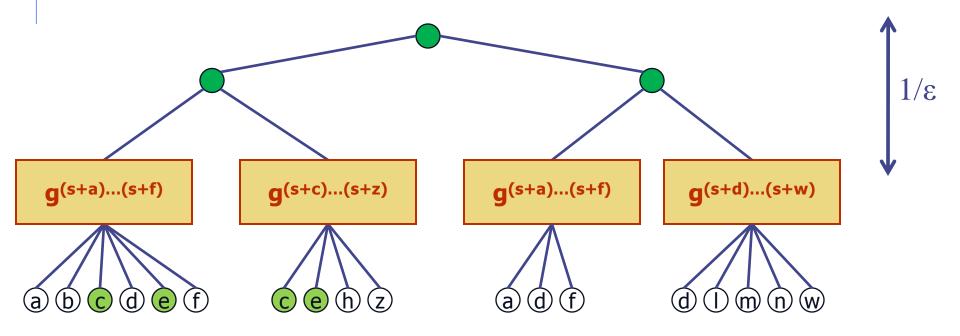
- Compute the accumulation value for every set
- Build an accumulation tree on top [CCS 2008]
  - O(1/ ε) levels and O(m<sup>ε</sup>) internal degree
  - O(m<sup>ε</sup>logm) query, O(1) update and O(1) proof
- The accumulation values protect the integrity of the set elements
- The accumulation tree protects the integrity of the acc. values



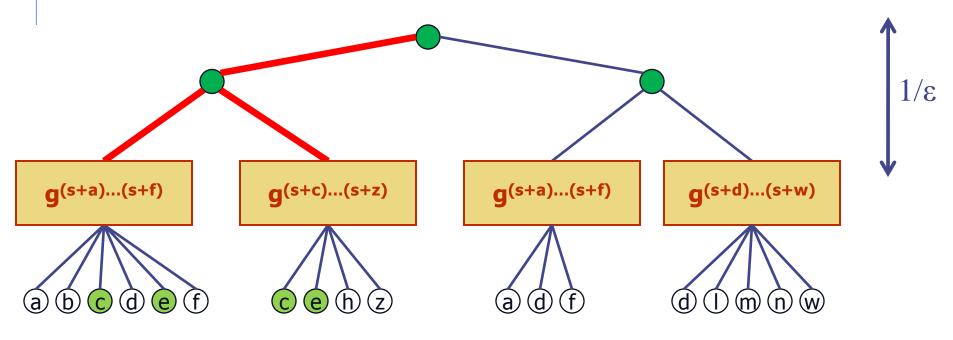


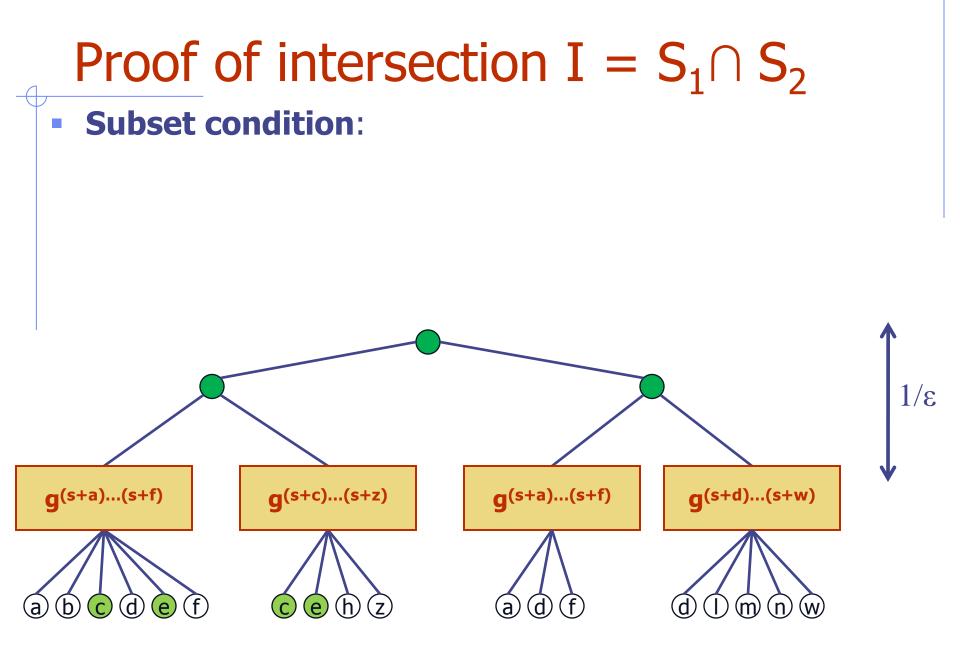


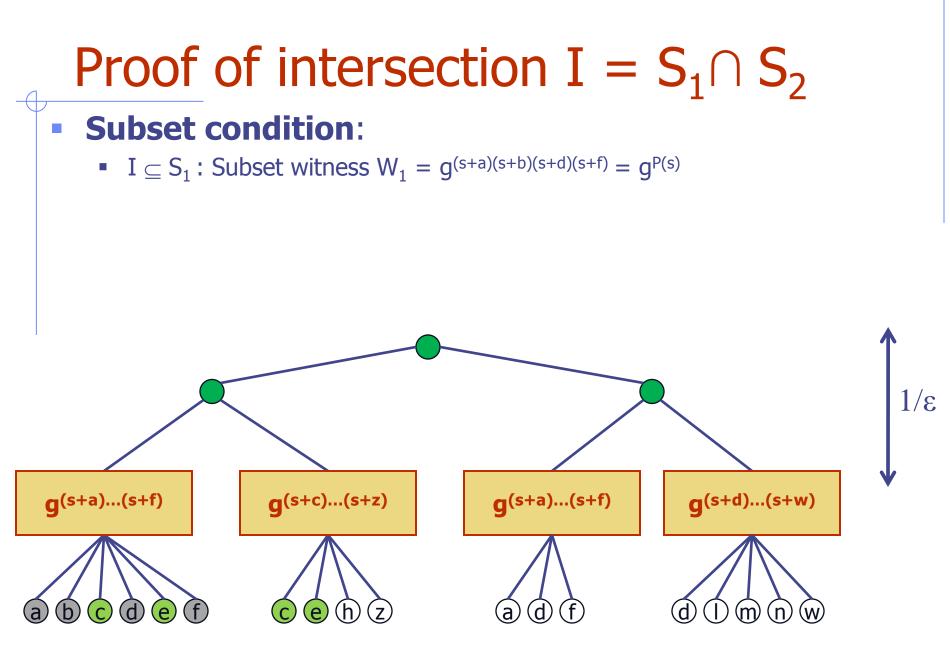
- Proof of accumulation values A<sub>1</sub> and A<sub>2</sub>
- Let Π<sub>1</sub> and Π<sub>2</sub> be such proofs



- Proof of accumulation values A<sub>1</sub> and A<sub>2</sub>
- Let Π<sub>1</sub> and Π<sub>2</sub> be such proofs
  - Values along the path of the tree
  - Construction of proofs: O(m<sup>ε</sup> logm)
  - Size of proofs: O(1)

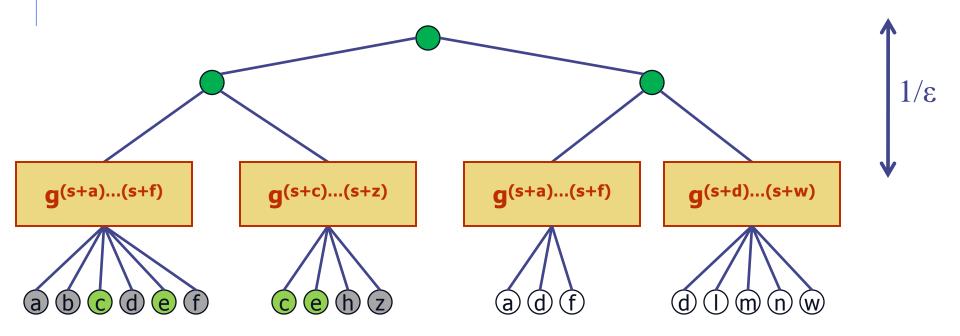






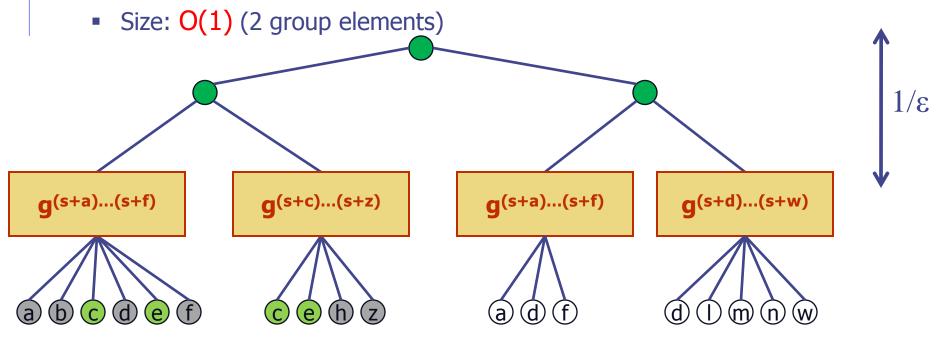
#### Subset condition:

- $I \subseteq S_1$ : Subset witness  $W_1 = g^{(s+a)(s+b)(s+d)(s+f)} = g^{P(s)}$
- $I \subseteq S_2$ : Subset witness  $W_2 = g^{(s+h)(s+z)} = g^{Q(s)}$



#### Subset condition:

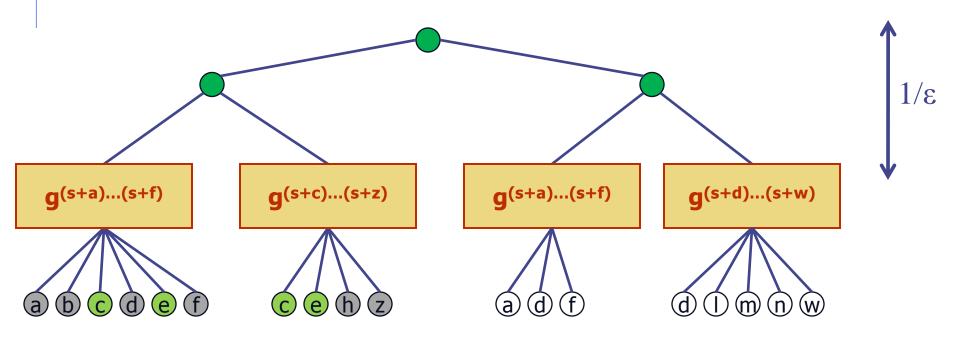
- $I \subseteq S_1$ : Subset witness  $W_1 = g^{(s+a)(s+b)(s+d)(s+f)} = g^{P(s)}$
- $I \subseteq S_2$ : Subset witness  $W_2 = g^{(s+h)(s+z)} = g^{Q(s)}$
- Complexity
  - Construction: O(nlog n) (polynomial interpolation)



#### Proof of intersection $I = S_1 \cap S_2$ **Completeness condition:** • $(S_1 - I) \cap (S_2 - I)$ is empty $1/\epsilon$ **q**(s+c)...(s+z) **q**(s+d)...(s+w) **q**(s+a)...(s+f) **q**(s+a)...(s+f) $\mathbf{d}$ ۰Ĝ e (b) (d) e(h) (a) (a)(m) $(\mathbf{n})$

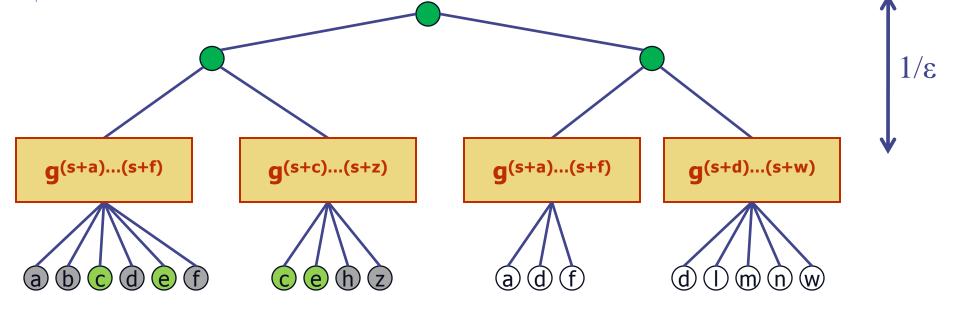
#### Completeness condition:

- $(S_1 I) \cap (S_2 I)$  is empty
- Recall  $W_1 = g^{P(s)}$  and  $W_2 = g^{Q(s)}$



#### Completeness condition:

- $(S_1 I) \cap (S_2 I)$  is empty
- Recall  $W_1 = g^{P(s)}$  and  $W_2 = g^{Q(s)}$
- Completeness witness  $F_1 = g^{A(s)}$  and  $F_2 = g^{B(s)}$
- A(s)P(s)+B(s)Q(s) = 1
- Complexity: O(nlog<sup>2</sup>nlog log n) (ext. Euclidean algorithm)



- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ

- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ
Accumulation values proofs	tm <sup>ɛ</sup> log m	t

- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ
Accumulation values proofs	tm <sup>ɛ</sup> log m	t
Subset witnesses	Nlog N	t

- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ
Accumulation values proofs	tm <sup>ɛ</sup> log m	t
Subset witnesses	Nlog N	t
Completeness witnesses	Nlog <sup>2</sup> Nloglog N	t

- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ
Accumulation values proofs	tm <sup>ɛ</sup> log m	t
Subset witnesses	Nlog N	t
Completeness witnesses	Nlog <sup>2</sup> Nloglog N	t
TOTAL	Nlog <sup>2</sup> Nlog log N + tm <sup>ɛ</sup> log m	t+δ

- **t** sets are intersected and **δ** is the size of the answer
- **N** is the sum of sizes of intersected sets

element of the proof	complexity	size
Intersection elements	Ν	δ
Accumulation values proofs	tm <sup>ɛ</sup> log m	t
Subset witnesses	Nlog N	t
Completeness witnesses	Nlog <sup>2</sup> Nloglog N	t
TOTAL almost optimal	Nlog <sup>2</sup> Nlog log N + tm <sup>ɛ</sup> log m	t+δ

### Size of proof for $X \cap Y$ in practice

X	Y	<b> X</b> ∩Y	KBytes [M+ 04]	KBytes this work
1000	1000	10	3.34	1.73
1000	100	1	1.68	1.55
1000	10	0	1.01	1.53
1000	1	0	0.46	1.53
10000	10000	100	26.88	3.53
10000	1000	10	12.15	1.73
10000	100	1	6.86	1.55
10000	10	0	3.08	1.53
100000	100000	1000	263.25	21.53
100000	10000	100	116.13	3.53
100000	1000	10	63.18	1.73
100000	100	1	26.29	1.55

Thank you!

# **Application: Supporting timestamps**

- For timestamped documents, use segment tree over the time dimension (N timestamps)
- Search interval covered by O(log N) canonical intervals in the segment tree, each corresponding to a set of documents T<sub>i</sub>
- Timestamped keyword search equivalent to O(log N) set intersections
  - T<sub>1</sub> ∩ S<sub>1</sub> ∩ S<sub>2</sub> … ∩ S<sub>t</sub>
  - $T_2 \cap S_1 \cap S_2 \dots \cap S_t$

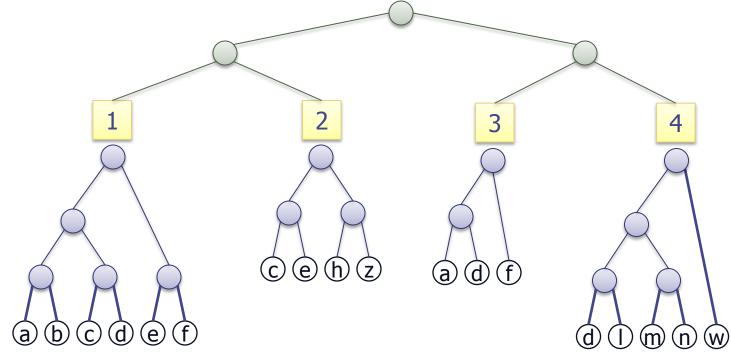
# Verifying outsourced computation

- Computation "on demand"
  - E.g., Google docs

- Find the pattern comput\* in my document
- Is the result correct?
- Need for efficient computations

# First solution: hashing

- [Devanbu et al., Algorithmica 2004; Yang and Papadias, SIGMOD 2009]
- Two-level tree structure and hierarchical cryptographic hashing
- Space: O(m + M), update: O(log m + log n)
- Intersection of two sets: O(n + log m) proof size and verification time
- **Security**: Cryptographic hashing
- Same complexities: Morselli et al., INFOCOM 2004



### Second solution: precomputation

- [Pang and Tan, ICDE 2004]
- Sign the answer to every possible query
- **Space**: O(m<sup>2</sup> + M) for a 2-intersection
- For any possible intersection space is
  O(2<sup>m</sup>)
- Proof size and verification: O(δ)
- **Update**: O(m<sup>2</sup>) for a 2-intersection
- Security: discrete log

Signatures of  $S_1 \cap S_2$   $S_1 \cap S_3$   $S_1 \cap S_4$   $S_2 \cap S_3$   $S_2 \cap S_4$   $S_2 \cap S_4$  $S_3 \cap S_4$ 

. . .