Trustworthy and Private Computing in the Cloud

Charalampos (Babis) Papamanthou University of Maryland cpap@umd.edu



Cloud computing today



Are there any threats?

- Cloud providers are untrusted Hotmail Data Loss Reveals
 - Can lose data
 - Can return corrupted results
 - Can leak information

... we will have no liability to you for any unauthorized access or use, corruption, deletion, destruction or loss of any of your content or applications...

Amazon web service customer agreement http://aws.amazon.com/agreement/



Cloud storage service provider <u>Dropbox</u> has <u>acknowledged</u> that a file

A configuration error in Microsoft's Business Productivity

While catching up on my reading, I came across an intere-

Dropbox confirms data leak

BPOS: a data leak in Microsoft's cloud December 28th, 2010 - 09:10 am ET by J. G.

01 August 2012, 12:39

Do people care?

- Customers are paying for the services
 - They want reliable storage
 - They want correctness guarantees
 - They want to keep their privacy

...58% of the public and 86% of business leaders are excited about the possibilities of cloud computing. But more than 90% of them are worried about security, availability, and privacy of their data as it rests in the cloud...

Microsoft survey in 2010 http://news.cnet.com/8301-1009_3-10437844-83.html

What do I do?

- Enable people to use the cloud safely
- Verifiability in the cloud
 - Verify that the cloud did the work correctly
- Privacy in the cloud
 - Use the cloud in a privacy-preserving manner

efficient both in theory and in practice

provably secure

no assumptions at the server



Roadmap

Motivation

Verifiability in the cloud

- Storage applications [STORAGESS08, CCS09, CCS13]
- Information retrieval applications [CRYPTO11, VLDB12]

Privacy in the cloud

Searching encrypted data [CCS12, FC13]

Research agenda and vision



- How to verify that a file has not been corrupted?
 - Keep a hash (i.e., checksum) locally for each file
 - Download: recompute and check
 - Upload: compute and store new hash

A more efficient protocol



- Hashing over a skip list and store only the roothash
 - Download: Fetch O(log n) hashes
 - Upload: An interactive protocol
- Technical contributions
 - Parallel O(log n) update complexity

Implementation with Amazon S3

- Extended an S3 client app (JetS3t Cockpit)
- Used Amazon EC2 and Amazon S3



Performance

- Stored 80,000 1KB files on Amazon S3
- Used 1,000 1KB files as workload
- Negligible overhead



Other considerations on storage

How can you verify **all the files more efficiently**?

 Dynamic proofs of retrievability [CCS09, CCS13]



What happens after a verification fails?

Provable data recovery [in progress]



Querying a database in the cloud

SELECT * FROM EMP, DEPT WHERE EMP.eID = DEPT.dID

equi-join queriesnot-in queries

elD	name	dID	division
1	Alex	1	Sales
22	James	21	Advertising
34	George	22	IT
		76	Sales
		88	Sales

Searching my email inbox

imail <u>Calendar</u> Docum	ents Photos Groups Web more	 I Settings Older version 	ion theig Sign.:
Gmail	Search	Mail Search the Web Contra Mice	
ompose Mail	Read items from any RSS or Atom feed	Ight here. Customize Clips	Omal Tip < >
box.(744)	Archive Report Spam Delete	More Actions + Battesh 1-50 of 1643	Older: Oldest.
arred 🛱	Salect All None Read Lineard Starrey	Linsterved	
20 P	Caracitati	chaining time - has sharing pictures of health. Marcony of hous a secondarial a	Ø. Jan. 18
ent Mail	Concerna Balai	who, upon shane. Internet is les wanted in say i seally anotheriate a hairing	Jan 10
afta (14)	Contracting Contract (1)	Internet and and a linkedia . Matter Rivers a College as a Transider Studi	Jan 9
Mar	Catalog (a)	In a biarth. No found a few into for me had adding for Part can be hold using	Jan B
900 C		New shift and a second time . Reduce the size of the hit and in the second	Jaco d
80	The Desired Charge	New sun needs your cheque - Needuce the size of the hit and its line spacing.	1206.07
stacts	Caniel Crem	Interry Conserves and Happy New Year - Inter everyone a werry Conserves an	100007
Chat	Find Camora	Join my nestoric of Linkelle. Knis technics, a Calestrate at Transides me	10/09/07
anth, add, or invite	C Concerne	Weisere in Orec and Orec and Villainers in Orec and	12/10/07
Shana S	The Art half has done	The Art both the Alumei Meyer I after. This measure is hains sent in you have	12/10/07
Set status here V		The Art manufes Autom News Later - this message is being sent or pot deal	A shares
Selloutik11	T A Daris conscious	hearte - nere s raume de a deder meno, newsnound, and know-n-an with Tag	1279907
Working for the man	Money we(e)	pro subject - impuration involution complete rest and interest for	12/74/07
Ryan Pollock	arex park (2)	transvideo Christmas Party - PTI - \$10.00 is a suggested amount. Prease grin	12/10/07
Anthony Prusakowsk	Dechen Wangol	Check out my Pacebook prote - tocebook Dechen Wango Dechen has: 67 P	12/10/07
bhutanbov@gmail.or	Cacanionoe	Pwd: Pwc. dot my tee up at last night Unlost out AUL Money & Priance's lig	9120007
Dev	randradeggmail.com	TV Corporate Website Copy - I've shared a document with you called "TV Cor	12/3/07
Hank Duderstadt	Rebecca Pelber	Fall 2007 Portfolio Show - Hey Shane, This is Becky, I don't know if you remer	12/2/07
honor gunday	Dessart, Donna	The Hodest Stop on the Line! - Dear A/CASF Alumni: We condulity invite you to	9112807
lispooky	C des and dot	Evite: Reminder for A Surprise Birthday Party for Eugenet - evite H scicicicija	11/22/07
Ricardo Andrade	Sharon Kalber	Pwd: COG1 needs people talented and preferably demented people A re-	11/19/07
Sara C	Dourwest Artines	tickebess travel Passenger sherafy - Southwest Artines Travel Energy This	11/16/07
ations V Add Contact	C Linkedin Updates	Linkedin Network Updates, 11/17/2007 - Linkedin NETWORK UPDATES Prof	11/1/07
shale	me, Elexis (3)	N - hoty orap. Greyhound's indicutous. It's gonna cost me almost \$40 to go to 8	11/15/07
and bird (13)	A-Jay Jimenez (5)	Come to my next event :: Fraggie Rock! This Friday, Nov, 16 2007 1 - Suppo thi	11/12/07
Editoria (California)	me, Sara (2)	Hi sara chun - Cocca Pebbles. On Nov 8, 2007 11:50 AM, Shane S. «alickrick	11/11/07
EDC MORT	Cacanion66, me (4)	xmas - I don't know. I think I'l be pretty miserable either way. On Nov 9, 2007 :	11/6/07
Invite a friend	Elexis, me (7)	(no subject) - you always let me win, that's no fun. On Nov 9, 2007 2:50 PM, E	11/9/07



- keyword search
- temporal queries



My contribution

in theory

	server comp.	proof	client verif.	update
CRYPTO11	$n \log^2 n$	δ	δ	1
generic circuits*	>N	1	δ	N/A

n: intersected sets size δ : answer size *N*: collection size

in practice: $n = 10,000, \delta = 100, N = n$					
	server comp.	proof	client verif.	update	
VLDB12	1.7 s	5 KB	20 ms	0.2 ms	
generic circuits**	37 hours	< 40 KB	< 1 ms	N/A	

*[Gennaro et al., EUROCRYPT13]

**[Setty et al., EUROSYS13]

Implementation & experiments

• C++

- NTL [1] and LiDIA [2] libraries for arithmetic ops
- DCLXVI [3] library for group operations
- Synthetic data
- Wall street journal corpus
 - Articles between 1987-1992
 - 173K articles &135K terms b



- [2] cdc.informatik.tu-darmstadt.de/TI/LiDIA/
- [3] cryptojedi.org/crypto/



Time to compute proof (server) 3 Two terms Answer mostly 2.5 between 0-200 time (sec) 2 articles Scales even for 1.5 popular keywords 1 Parallelized 0.5 when possible 0 2000 8000 10000 12000 14000 16000 0 4000 6000 total size of intersected sets

Time to verify proof (client)



- Proportional to the size of the answer
- Order of milliseconds

How about privacy?



- How can you search your encrypted files?
 - Not feasible in principle
 - Encrypt with fully-homomorphic encryption (FHE)?
 - Not very practical

Searchable encryption



- My First dynamic efficient scheme, [CCS12]
 - Privately indexes keywords, not only files
 - Efficient system implementation
- My work: First parallel scheme, [FC13]
 - Uses a tree-based approach

Research agenda

Practical protocols

- Construct customized schemes, e.g., geometric/text processing queries
- Consider highly dynamic/streaming data

Secure cloud computing in the age of big data

- Take advantage of multicore architectures, e.g., build parallel algorithms
- Implement on hardware

Reconsider the model

- Relax assumptions, e.g., more efficiency with trusted hardware
- Consider multi-client settings

Theoretical directions

- Cryptographic verification via certification algorithms
- Explore other crypto primitives (e.g., lattice-based cryptography)

Thank you!

cpap@umd.edu