Homomorphic Encryption – are we there yet?

Anirban Basu

KDDI R&D Laboratories, Japan

02 September 2015
Kyushu University

KDDI R&D LABS

At a glance

1. Who am I?
2. Why are we listening to this?
   - The privacy problem and homomorphic encryption
3. Collaborative filtering and privacy
   - SlopeOne predictors for CF
   - Privacy-preserving CF
   - Deployment on SaaS engines (PaaS clouds)
   - PPCF experimental results and inferences
4. Lightweight and practical anonymous message routing
   - Why?
   - The state of the art
   - Sender anonymity through message unlinkability
   - Anonymous messaging experimental validation
5. Epilogue
   - What can we conclude?

KDDI R&D LABS

Profile[1]

- Researcher within the Information Security Group at KDDI R&D Laboratories.
- Post-doc (Tokai University, Japan), 2013: privacy preserving collaborative filtering.
- PhD (University of Sussex, UK), 2010: a reputation framework for computer networks.
- BEng (University of Sussex, UK), 2004: augmented reality visualisation.
- Hobbies: programming, photography, travelling, cycling.
- Home town: Chandannagar, West Bengal, India.

KDDI R&D LABS

[1] See: `http://www.linkedin.com/in/anirbanbasu`.

Bigger the data, worse the privacy

- Big data is getting bigger – the Internet of Things!
- Good news: more data to analyse and build intelligence.
- Bad news: privacy of data is a growing concern.
- One cloud adoption barrier: privacy of data, both individual and organisational.

KDDI R&D LABS

Privacy – what to do?

- Data release through anonymisation, perturbation: privacy and utility do not agree.
- Why do we not encrypt all the data?
  - But computing something meaningful (i.e., data mining) over that encrypted data?
- Homomorphic encryption: compute blindly over encrypted data.
- The elephant in the room: is homomorphic encryption magical or mythical? How far is it from reality?

KDDI R&D LABS

Homomorphic encryption – brief background

- Generally speaking: $f(m_1, m_2) \equiv g(c_1, c_2)$:
  - function f on plaintext messages m_i is equivalent to a function g over ciphertexts of these messages c_i.
- Different classes of homomorphic encryption:
  - additive,
  - multiplicative,
  - somewhat homomorphic, and
  - fully-homomorphic.

KDDI R&D LABS

The magic of additive homomorphic encryption

- $\mathcal{E}(m_1 + m_2) = \mathcal{E}(m_1) \cdot \mathcal{E}(m_2)$
  - The encryption of the sum of two plaintexts is the modular multiplication of their individual ciphertexts.
- $\mathcal{E}(m_1 \cdot m_2) = \mathcal{E}(m_1)^{m_2}$
  - The encryption of the multiplication of two plaintexts is the modular exponentiation of the ciphertext of one by the other plaintext.
- Examples of such cryptosystems: Paillier, Elliptic Curve ElGamal, Damgård-Jurik.

KDDI R&D LABS

Is this practical?

- Fully-homomorphic encryption is still somewhat far from realistic applicability.
- Practice: partially homomorphic encryption and a mixture of various other encryption techniques for specific application scenarios.
- Two application scenarios: (privacy preserving) collaborative filtering (2013) and anonymous message routing (2015).

KDDI R&D LABS

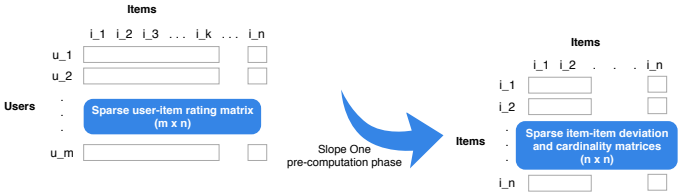
Recommendation and collaborative filtering

- Collaborative filtering (CF) employs opinions of the community.
- Problem is with the privacy of rating data.

KDDI R&D LABS

SlopeOne CF

- Precomputation: Δ: item-item deviation matrix; ϕ: item-item cardinality matrix.



- Standard SlopeOne based prediction: $r_{u,x}$, the rating from user u on item x

$$r_{u,x} = \frac{\sum_{a|a\neq x}(\delta_{x,a} + r_{u,a})\phi_{x,a}}{\sum_{a|a\neq x}\phi_{x,a}} = \frac{\sum_{a|a\neq x}(\Delta_{x,a} + r_{u,a}\phi_{x,a})}{\sum_{a|a\neq x}\phi_{x,a}}$$

KDDI R&D LABS

Privacy-preserving collaborative filtering (PPCF)

- The unencrypted SlopeOne based prediction:

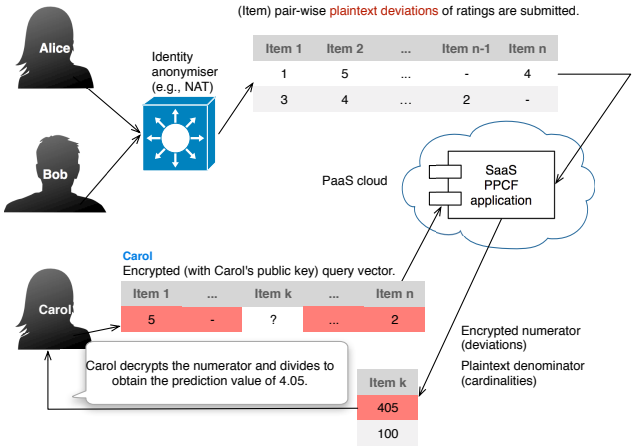
$$r_{u,x} = \frac{\sum_{a|a \neq x}(\Delta_{x,a} + r_{u,a}\phi_{x,a})}{\sum_{a|a \neq x}\phi_{x,a}}$$

- Over an additively homomorphic encrypted domain:

$$r_{u,x} = \frac{\mathcal{D}(\mathcal{E}(\sum_{a|a \neq x}\Delta_{x,a})\prod_{a|a \neq x}(\mathcal{E}(r_{u,a})^{\phi_{x,a}}))}{\sum_{a|a \neq x}\phi_{x,a}}$$

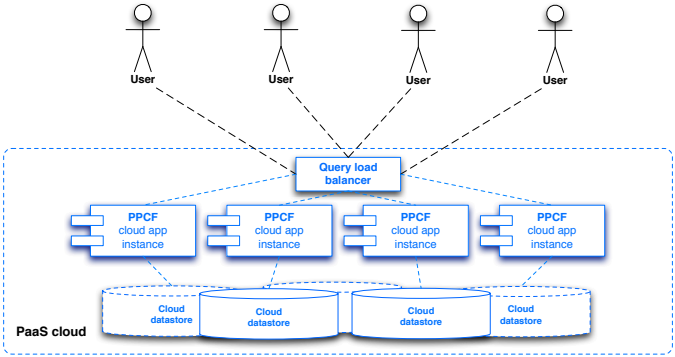
KDDI R&D LABS

PPCF on the cloud



KDDI R&D LABS

PPCF deployment scenario as a SaaS



KDDI R&D LABS

App Engine (GAE/J) versus Elastic Beanstalk (EBS)

	GAE/J	EBS
Java software stack	Limited	Full
Scalability	Very high	High but pricey
Unit performance	Average	Configurable
Data storage	Distributed BigTable, SQL	Distributed SimpleDB, SQL
Vendor lock-in	Yes, partially	No
Free quota	Daily	One-off, first year
Frontend access	HTTP, SPDY (SSL)	HTTP, HTTPS

KDDI R&D LABS

Performance test

- Speed of query processing.
- Varying length of the query vector.
- Varying concurrent user requests.
- Single-threaded or multi-threaded query vector processing.

KDDI R&D LABS

Google App Engine setting

- Instance class: F4 (2400MHz, 512MB RAM).
- Maximum idle instances: automatic.
- Maximum pending latency: 10ms.
- Datastore: master-slave, not high-replication.

KDDI R&D LABS

Amazon Elastic Beanstalk setting

- Instance class: t1.micro EC2 instances (min: 1, max: 8).
- Load balancer increase (by one instance) trigger: over 70% CPU utilization in 1 minute.
- Load balancer decrease (by one instance) trigger: below 40% CPU utilization in 1 minute.
- Datastore: MySQL RDBMS on t1.micro EC2 instance.

KDDI R&D LABS

The datasets used

	Jester	MovieLens 100K
Users	73,421	943
Items	100	1,682
Range	$[-10.00 \quad 10.00]$	$\{1, 2, 3, 4, 5\}$
Ratings	4,100,000	100,000
Rating density	55.8%	6.3%
Min. rating	-9.95	1
Max. rating	10.0	5
Rating mean	0.744	3.539
Data points[2]	4,950	983,206
Density	100%	69.5%

KDDI R&D LABS

[2] "Data points" and "Density" refer to Slope One data points and their density.

Graph legend

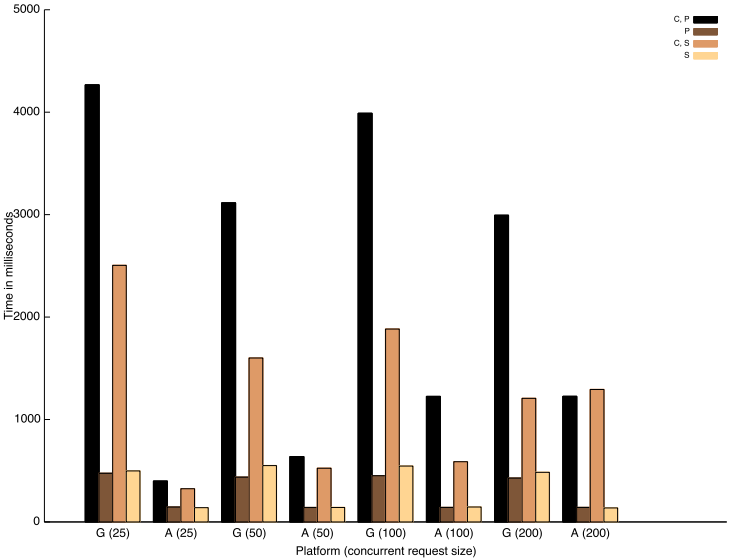
Heads up for the experiment categories

S: single-threaded, single query, P: multi-threaded, single query; C, S: single-threaded, concurrent query and C, P: multi-threaded, concurrent query.

KDDI R&D LABS

Jester dataset (1024-bits)

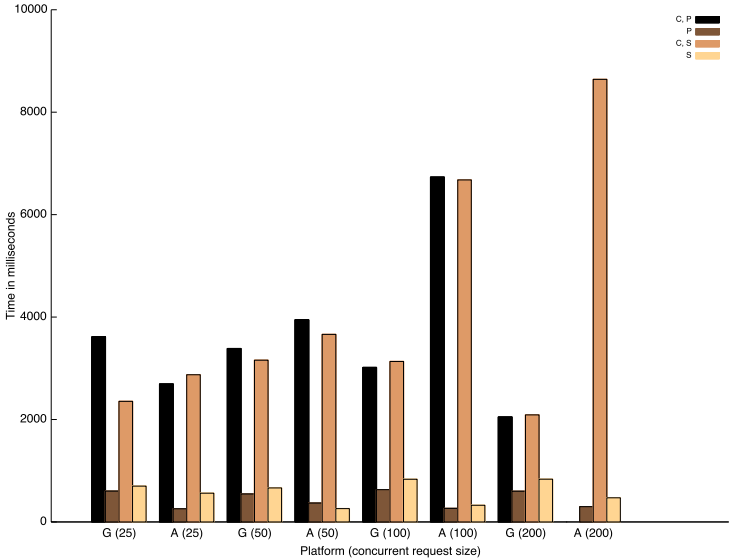
- Example query vector size: 20.



KDDI R&D LABS

MovieLens dataset (1024-bits)

- Example query vector size: 50. EBS partial failure: datastore scalability.



KDDI R&D LABS

General observations

	GAE/J[3]	EBS[4]
Response to short bursty load	Very fast	Slow
Response to steady load	Steady	Steady
Parallel query vector processing	Good	Not necessarily good
Configurability	Limited	High
Ease of deployment	High	Moderately difficult
Running cost	Low, can be capped	High

- Google App Engine is better suited for the type of application and deployment setup we had.
- GAE/J is better with applications that receive high user requests but take relatively short time to process each request.

[3] Google App Engine.
[4] Amazon Elastic Beanstalk.

KDDI R&D LABS

Query processing time estimation

- Estimated query size for a 30s turn-around time, single-threaded processing and with only one query at a time.
- Google App Engine is generally slower but performs better with concurrent loads.

	GAE/J	EBS	Theoretical sizes[5]
Encrypted query vector size	1376 items	3274 items	–
HTTP POST size (numeric IDs[6])	698KB	1.624MB	520n bytes
HTTP POST size (string IDs)	731KB	1.698MB	544n bytes

[5] For Paillier 2048-bits, n query items; ignoring other POST overheads.

[6] Excludes overhead of JSON packaging.

KDDI R&D LABS

Intermission

KDDI R&D LABS

Anonymous communication

- Encourages free speech: no fear of reprisal.
- End-to-end encrypted messaging is *not* anonymous communication in the context of this talk.

KDDI R&D LABS

Motivating use cases

- Anonymous opinions.
  - Present: insufficient ‘anonymity’ guarantees in existing survey systems, e.g., Survey Monkey.
  - Future: anonymise survey participants.
- Anonymous micro-blogging.
  - Present: micro-blogging platforms, e.g., Twitter, identify bloggers, or re-posters.
  - Future: anonymise micro-bloggers.
- Limit participation to specific groups with private information retrieval, blind signatures.
- Another use case: anonymously posting data to a public cloud-based classifier.

KDDI R&D LABS

But there are tools that already do this

- Tor – the well-known anonymous network.
- Generalising: high-latency systems (mix networks) and low-latency systems (e.g., onion routing).
- Specialised configurations/permissions, e.g., opening ports through the firewall.
- Pre-existing paths in Tor, potentially breakable[7].
- Recall the adjectives for the title of this work: *lightweight* and *practical*?

KDDI R&D LABS

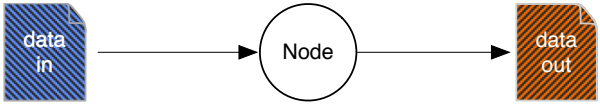
[7] See: `https://blog.torproject.org/blog/one-cell-enough`.

Why are we different?

- We are proposing an anonymous messaging scheme that:
  - provides sender anonymity (*not* recipient anonymity);
  - works without any specialised network configurations – pure HTTP(S), HTML and Javascript;
  - works with a public untrusted cloud – our router (!);
  - preserves secrecy of the message; and
  - works even with some dishonest participants.

KDDI R&D LABS

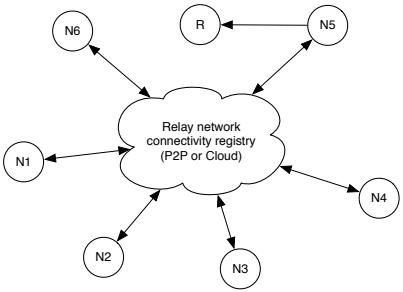
The crux of message unlinkability



- If the ingress and egress messages look indistinguishable then it is hard to tell (traffic analysis aside!) if a message going into a node is the one coming out.
- Have nodes to forward messages around before sending it to the final recipient.

KDDI R&D LABS

Message forwarding network



- Public recipient, R. Public untrusted router – the cloud or a P2P network.
- Any node n_i can either forward a message or send it to R.
- Example: R thinks that the message is from n_5 but it could be from any other node.

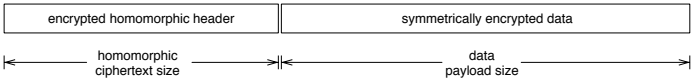
KDDI R&D LABS

How is it done?

- Alter every ingress encrypted (with recipient's public key) message at node n_i to generate the egress message for node n_{i+1} as $\mathcal{E}(m)_{n_{i+1}} = \mathcal{E}(m)_{n_i} \cdot \mathcal{E}(0)$.
- Forward the egress message with probability p_f or send it to the final recipient with probability $1 - p_f$.
- Recipient: $m = \mathcal{D}(\mathcal{E}(m)_{n_k})$.
- Ensure that the messages are of the same size, e.g., $|m| = 2048$ bits.

KDDI R&D LABS

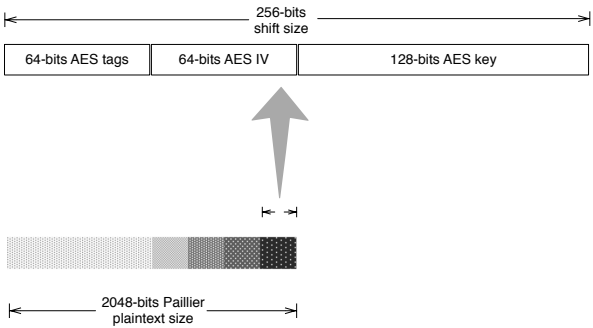
Large size of m – hybrid encryption?



- Apply symmetric key encryption on the message.
- Store random symmetric keys in a homomorphic header.
- Recipient decrypts message in multiple rounds of symmetric key decryption with keys obtained from the header.
- Will need to break messages apart and pad to maintain fixed sizes.
- Limitation: forwarding hop count.

KDDI R&D LABS

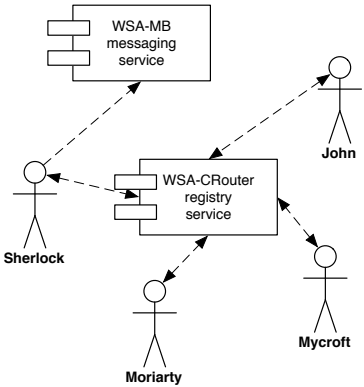
Homomorphic header: an example with AES



- At any node n_i, egress message header:
  $\mathcal{E}(h_{n_{i+1}}) = \mathcal{E}(h_{n_i})^{2^{|k|+|p|}} \cdot \mathcal{E}(k_{n_i}||p_{n_i})$.
- Keys added with left shifts by $|k| + |p|$, but not many shifts before information is lost.

KDDI R&D LABS

Real world: Sherlock's secret message



- Moriarty and his network: who was the actual sender?

KDDI R&D LABS

Experimental validation

- HTML5 and Javascript client.
- Google App Engine (Java) cloud-based apps (F2 instance class: 1200MHz CPU, 256MB RAM).
- Demo: who wants to be Sherlock?

The public cloud-based router.

A public message board.

KDDI R&D LABS

Client side sender and forwarder nodes

- (mobile) Chrome 35.0.1916.38/iOS 7.1.1, iPhone 5S, 4G network;
- (desktop) IE 10.0.9200.16899S/Windows 8.0, 3GHz Intel Core i7 processor, 16GB RAM, 1Gbps wired network;
- (desktop) Firefox 29.0/Ubuntu Linux 14.10, 3GHz Intel Core i7 Extreme processor, 16GB RAM, 1Gbps wired network; and
- (laptop) Chrome 35.0.1916.114/Mac OS X 10.9, Macbook Air, 1.8GHz Intel Core i7 processor, 8GB RAM, 54Mbps IEEE 802.11g network.

KDDI R&D LABS

Client side performance

Platform	Mean forwarding time
(mobile) Chrome/iOS	70.9s
(desktop) IE/Windows	17.2s
(laptop) Chrome/OS X	4.06s
(desktop) Firefox/Linux	1.89s

- Bottleneck is the performance of Paillier in Javascript.
- A randomised time delay may actually help against traffic analysis attacks.
- Good news: high-performance lattice crypto in Javascript.

KDDI R&D LABS

Concluding remarks – are we there yet?

- Practical applications of partial homomorphic encryption.
- Cloud-based classifiers: collaborative filtering (this talk), support vector machines, decision trees.
- Anonymous messaging routing.
- Short-term future: partial homomorphic encryption and various encryption techniques.

KDDI R&D LABS

Thank you for your time!

Any questions?

KDDI R&D LABS