Efficient Private Set Intersection for a Decentralised Web of Trust

Álvaro García-Recuero

October 31, 2017

"Privacy-preserving protocols for the WWW in the age of mass surveillance and adversarial learning"

Why is that?

£2.30
Saturday 22.06.13
Published in London and Manchester
guardian.co.uk

the guardian
weekend edition

How GCHQ watches your every move

Exclusive
Operation Tempora revealed

Internet traffic and calls tapped from fibre optic cables

Information shared with American spy agency

Britain's spy agency GCHQ has secretly gained access to the network of cables which carry the world's phone calls, internet traffic and has started to ... vast streams of sensitive ... tions which it is sharing ...

deemed legal, even though the warrant system was supposed to limit interception to a specified range of targets.

The existence of the programme has been disclosed in documents shown to the Guardian by the NSA whistleblower Edward Snowden as part of his attempt to expose what he has called "the largest programme of suspicionless surveillance in human history". "It's not just a US problem. The UK has a huge dog in this fight," Snowden, who was reportedly charged with espionage by US authorities last night, told the Guardian. "They [GCHQ] are worse than the US."

But a source with knowledge of intelligence argued last night that the data was collected legally under a system of safeguards, and had provided material that had led to significant breakthroughs in detecting and preventing serious crime.

Britain's technical capacity to tap into the cables that carry the world's communications – referred to in the documents as special source exploitation – has made GCHQ an intelligence superpower.

By 2010, two years after the project was first trialled, it was able to boast it had the "biggest internet access" of any member of the Five Eyes electronic eavesdropping alliance, comprising the US, UK, Canada, Australia and New Zealand.

UK officials could also claim GCHQ "produces larger amounts of metadata than NSA". (Metadata describes basic information on who has been contacted by whom, without detailing the content.)

By May last year 300 analysts ... GCHQ, and 250 from the NSA ...

ing as more cables are tapped and GCHQ data storage facilities in the UK and abroad are expanded with the aim of processing terabits (thousands of gigabits) of data at a time. For the 2 billion users of the world wide web, Tempora represents a window on to their everyday lives, sucking up every form of communication from the fibre-optic cables that ring the world.

The NSA has meanwhile opened a second window, in the form of the Prism operation, revealed earlier this month by the Guardian, from which it secured access to the internal systems of global companies that service the internet.

The GCHQ mass tapping operation has been built up over five years by attaching intercept probes to transatlantic fibre-optic cables where they land on British shores carrying data to western Europe from telephone exchanges and internet servers in north America.

This was done under secret agreements with commercial companies, described in one document as "intercept partners".

The papers seen by the Guardian suggest some companies have been paid for the cost of their co-operation and GCHQ went to great lengths to keep their names secret. They were assigned "sensitive

Continued on page 7 »

Inside »

How GCHQ tackled its own ... the ...

Strong and Malicious

- Mass-surveillance AND personal data collection by third-parties on the WWW are a real threat to liberal societies and citizens![a].

[a] https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers

Countermeasures

- A truly decentralised WWW will require the network to provide privacy and trust by design.

Strong and Malicious

- Mass-surveillance AND personal data collection by third-parties on the WWW are a real threat to liberal societies and citizens![a].

[a] https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers

Countermeasures

- A truly decentralised WWW will require the network to provide privacy and trust by design.

How safe is Big Data?

Adversarial learning

Manipulating or inserting corrupted samples in the dataset to obtain a desired outcome (e.g., financial credit score in OSNs).

De-anonymisation

Possible to use external data sources to re-identify users and their preferences.

Privacy breaches

WoT[a] extension collecting users' metadata in the browser.

[a] https://en.wikipedia.org/wiki/WOT_Services

How safe is Big Data?

Adversarial learning

Manipulating or inserting corrupted samples in the dataset to obtain a desired outcome (e.g., financial credit score in OSNs).

De-anonymisation

Possible to use external data sources to re-identify users and their preferences.

Privacy breaches

WoT[a] extension collecting users' metadata in the browser.

[a] https://en.wikipedia.org/wiki/WOT_Services

How safe is Big Data?

Adversarial learning

Manipulating or inserting corrupted samples in the dataset to obtain a desired outcome (e.g., financial credit score in OSNs).

De-anonymisation

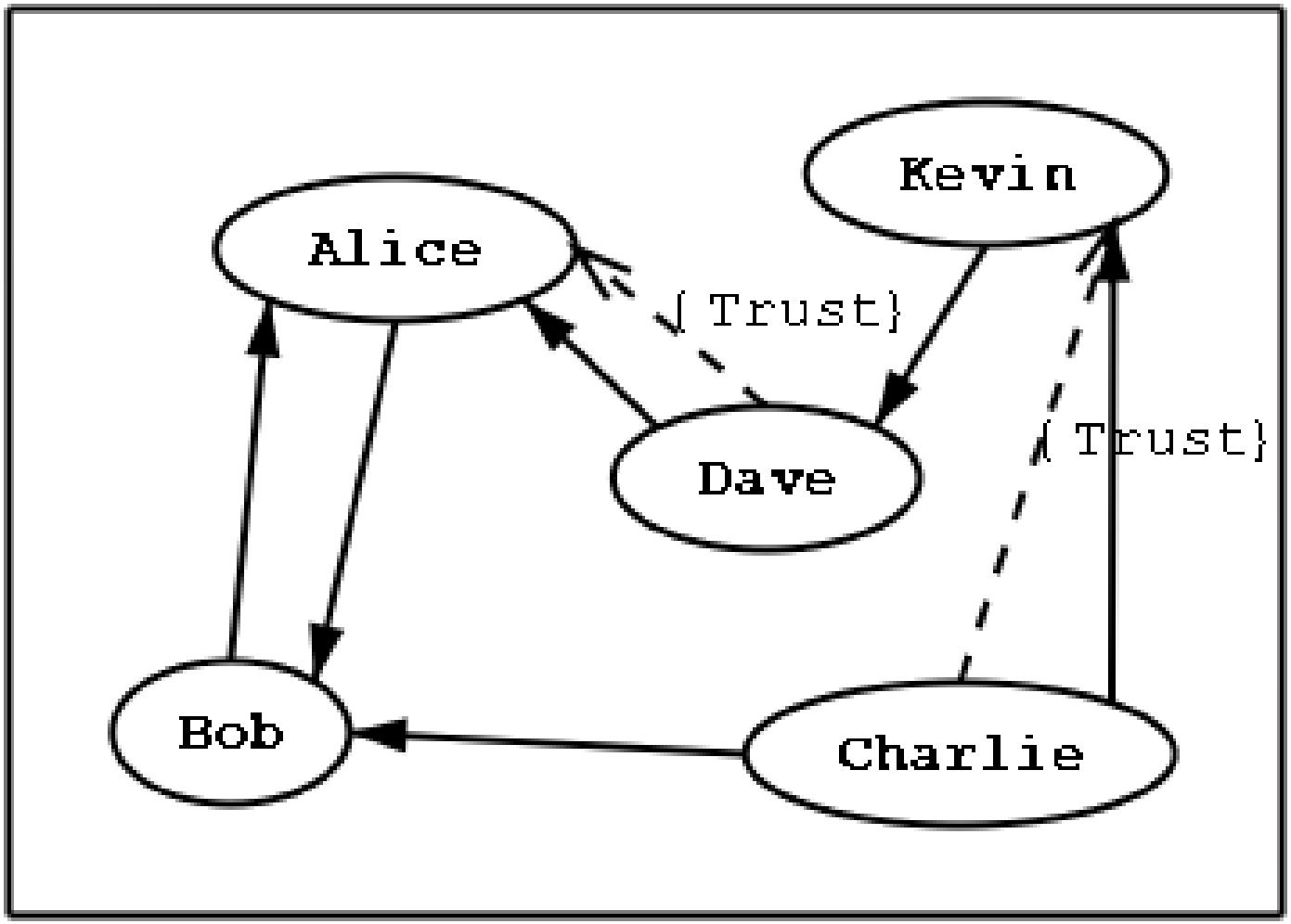
Possible to use external data sources to re-identify users and their preferences.

Privacy breaches

WoT[a] extension collecting users' metadata in the browser.

[a] https://en.wikipedia.org/wiki/WOT_Services

What is the Web-of-Trust about?



An example of the web of trust model

What is decentralised PSI useful for?

Trust for a non-public Web-of-Trust

- We should be able to establish trust without a centralised Certification Authority (CA).

Going Decentralised

- The user should able to establish direct trust with its peers, similarly to what happens with PGP, GnuPG and others, but without exposing who the signers are, etc.

Why is it desirable?

- Centralised data silos prone to privacy breach, e.g., third-party apps as the WoT plugin.
- Governments and powerful authorities, e.g., NSA, GCHQ.

What is decentralised PSI useful for?

Trust for a non-public Web-of-Trust

- We should be able to establish trust without a centralised Certification Authority (CA).

Going Decentralised

- The user should able to establish direct trust with its peers, similarly to what happens with PGP, GnuPG and others, but without exposing who the signers are, etc.

Why is it desirable?

- Centralised data silos prone to privacy breach, e.g., third-party apps as the WoT plugin.
- Governments and powerful authorities, e.g., NSA, GCHQ.

What is decentralised PSI useful for?

Trust for a non-public Web-of-Trust

- We should be able to establish trust without a centralised Certification Authority (CA).

Going Decentralised

- The user should able to establish direct trust with its peers, similarly to what happens with PGP, GnuPG and others, but without exposing who the signers are, etc.

Why is it desirable?

- Centralised data silos prone to privacy breach, e.g., third-party apps as the WoT plugin.
- Governments and powerful authorities, e.g., NSA, GCHQ.

Abusing the WWW

Definition

Modeling Abuse

- Deny
- Deceive
- Degrade
- Disrupt

Government Communications Headquarters

Defining Deceive

Modeling Abuse

- Supplanting a known user identity (impersonation) for influencing other users behaviour and activities, including assuming false identities (but not pseudonyms).
- SYLVESTER: framework for automated interaction & alias management in Online Social Networks.
- UNDERPASS Change outcome of online polls.
- SCRAPHEAP CHALLENGE: perfect spoofing of emails from Blackberry targets.
- BURLESQUE: capability to send spoofed SMS text messages.

Defining Degrade

Modeling Abuse

- Disclosing personal and private data of others without their approval as to harm their public image or reputation.
- BIRDSTRIKE is a Twitter monitoring and profile data collection tool.
- SPRING BISHOP: finds private photographs of targets in Facebook.

Defining Deny

Modeling Abuse

- Encouraging self-harm to other users, promoting violence (direct or indirect), terrorism or similar activities.
- CLEAN SWEEP: masquerades Facebook wall posts for individuals or entire countries, effectively denying access to information (censorship).
- ROLLING THUNDER: distributed denial of service using P2P.

Defining Disrupt

Modeling Abuse

- Distracting provocations, denial-of-service, flooding with messages, promote abuse.
- BIRDSONG: automated posting of Twitter updates.
- CANNONBALL: capability to send repeated text messages to a single target.
- PITBULL: enabling large scale delivery of a tailored message to users of instant messaging services.

Abuse detection

Abuse ground truth

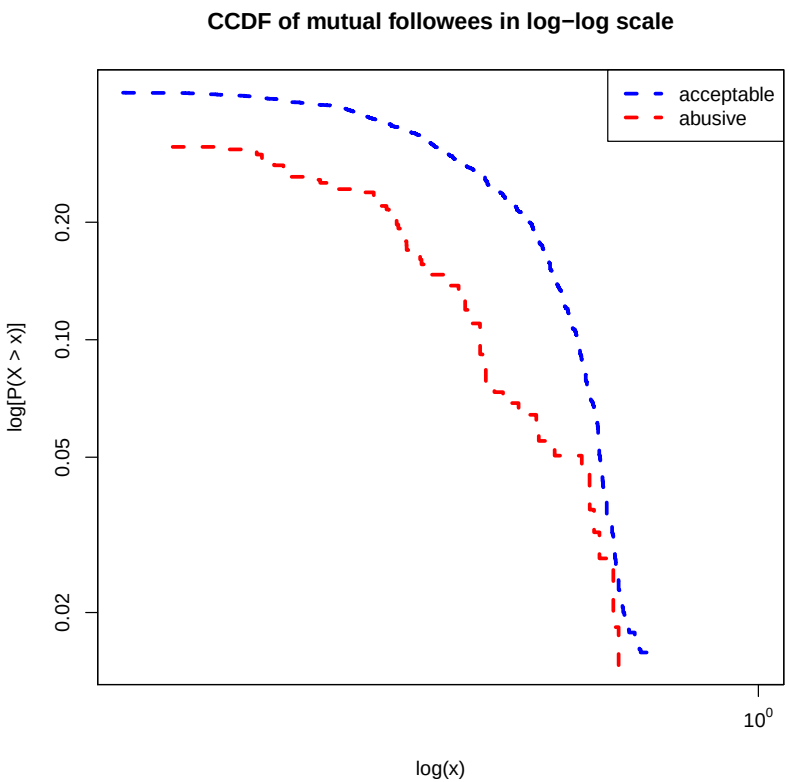
Trollslayer tool

```
d888888b d8888b.  .d88b.  db      db      .d8888. db       .d8b.  db    db d88888b d8888b.
`~~88~~' 88  `8D .8P  Y8. 88      88      88'  YP 88      d8' `8b `8b  d8' 88'     88  `8D
   88    88oobY' 88    88 88      88      `8bo.   88      88ooo88  `8bd8'  88ooooo 88oobY'
   88    88`8b   88    88 88      88        `Y8b. 88      88~~~88    88    88~~~~~ 88`8b
   88    88 `88. `8b  d8' 88booo. 88booo. db   8D 88booo. 88   88    88    88.     88 `88.
   YP    88   YD  `Y88P'  Y88888P Y88888P `8888Y' Y88888P YP   YP    YP    Y88888P 88   YD


To mark a tweet as abuse, we ask you to read the JTRIG techniques for online HUMINT Operations.


#### JTRIG 4 D's: Deny, Disrupt, Degrade or Deceive:


- Deny: encouraging self-harm to others users, promoting violence (direct or indirect), terrorism or
similar activities.

- Disrupt: distracting provocations, denial-of-service, flooding with messages, promote abuse.

- Degrade: disclosing personal and private data of others without their approval as to harm their public
image/reputation.

- Deceive: supplanting a known user identity (impersonation) for influencing other users behavior and
activities, including assuming false identities (but not pseudonyms).


Abusive tweet matching Deny

Tweet:   I retract my awful statement of #XXXX people with batman/anime/Sin City avatars deserve death.
        I really meant "frozen in time forever".


Please enter your id below, choose something unique and that you can remember (annotations are grouped by id):
If you have already annotated data, please reuse your unique identifier to continue annotations
To exit: Ctrl + C
```

[1] github.com/algarecu/trollslayer

Mutual Subscriptions

Feature analysis



- |Subscription ∩ Subscription| CCDF shows less overlap among subscriptions of author of abusive messages and subscriptions of potential victim.
- Privacy: it needs a protocol to protect metadata.
- Security? Hard to prevent increase in overlap of subscriptions of potential victim if that is public information.

Straw-man version

Privacy Protocol

Problem: Alice wants to compute $n := |\mathcal{L}_A \cap \mathcal{L}_B|$

Suppose each user has a private key c_i and the corresponding public key is $C_i := g^{c_i}$ where g is some generator

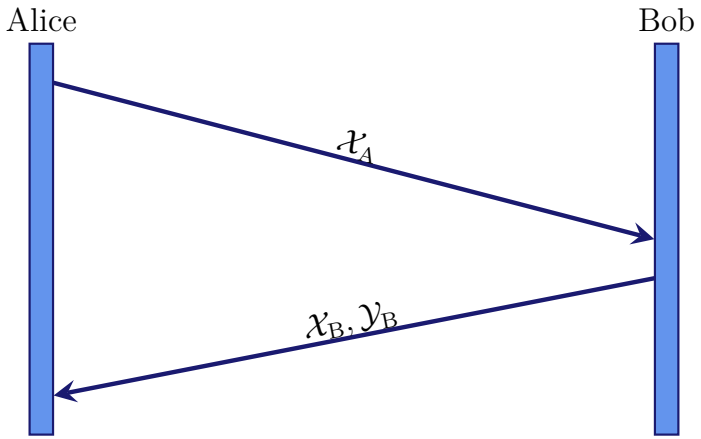
The set up is as follows:

- $\mathcal{L}_A$: set of public keys representing Alice's subscriptions
- $\mathcal{L}_B$: set of public keys representing Bob's subscriptions
- Alice picks an ephemeral private scalar $t_A \in \mathbb{Z}/p\mathbb{Z}$ (set of scalars used for a D-H exchange).
- Bob picks an ephemeral private scalar $t_B \in \mathbb{Z}/p\mathbb{Z}$ (set of scalars used for a D-H exchange).

Privacy Protocol: straw-man version

$$\mathcal{X}_A := \left\{ C^{t_A} \;\middle|\; C \in \mathcal{L}_A \right\}$$

$$\mathcal{Y}_A := \left\{ \hat{C}^{t_A} \;\middle|\; \hat{C} \in \mathcal{X}_B \right\} = \left\{ C^{t_A \cdot t_B} \;\middle|\; C \in \mathcal{L}_A \right\}$$



$$\mathcal{X}_B := \left\{ C^{t_B} \;\middle|\; C \in \mathcal{L}_B \right\}$$

$$\mathcal{Y}_B := \left\{ \overline{C}^{t_B} \;\middle|\; \overline{C} \in \mathcal{X}_A \right\} = \left\{ C^{t_B \cdot t_A} \;\middle|\; C \in \mathcal{L}_B \right\}$$

Alice can get $|\mathcal{Y}_A \cap \mathcal{Y}_B|$ within linear cost

Straw-man Protocol 1

Attack 1

- Attack 1: insertion of sock-puppet accounts to infer size of the potential's victim contact list.
- Solution: defeat it with shuffling of contact list before sending it to other party.

Straw-man Protocol 1

Attack 2

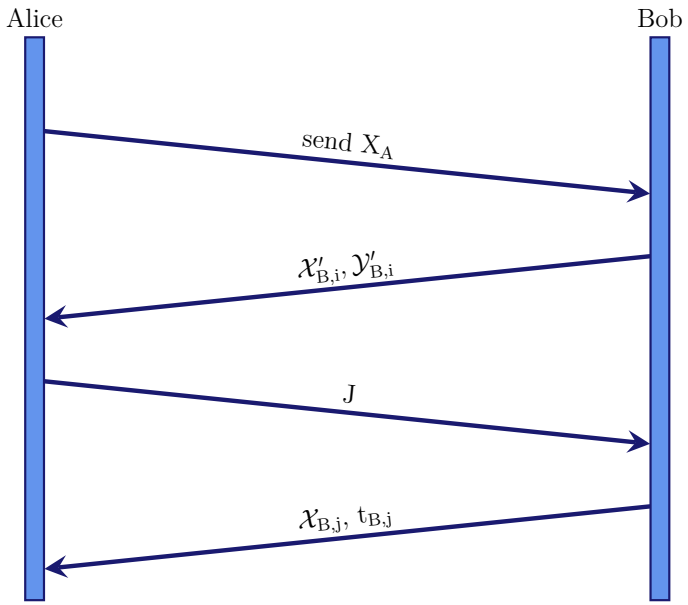
- Attack 2: insertion of sock-puppet account with a marker in the potential perpetrator list allows to infer set membership in potential victim's list (identifying pair of elements).

- Solution: hash the commitments of reblinded contact list in the reply to potential perpetrator.

- Assume a fixed system security parameter $\kappa \geq 1$

- For any list or set Z, define $Z' := \{h(x)|x \in Z\}$, e.g., $\mathcal{X}'_{B,i}$: hashing each element $\in X_B$

Protocol 1: Cut & choose version

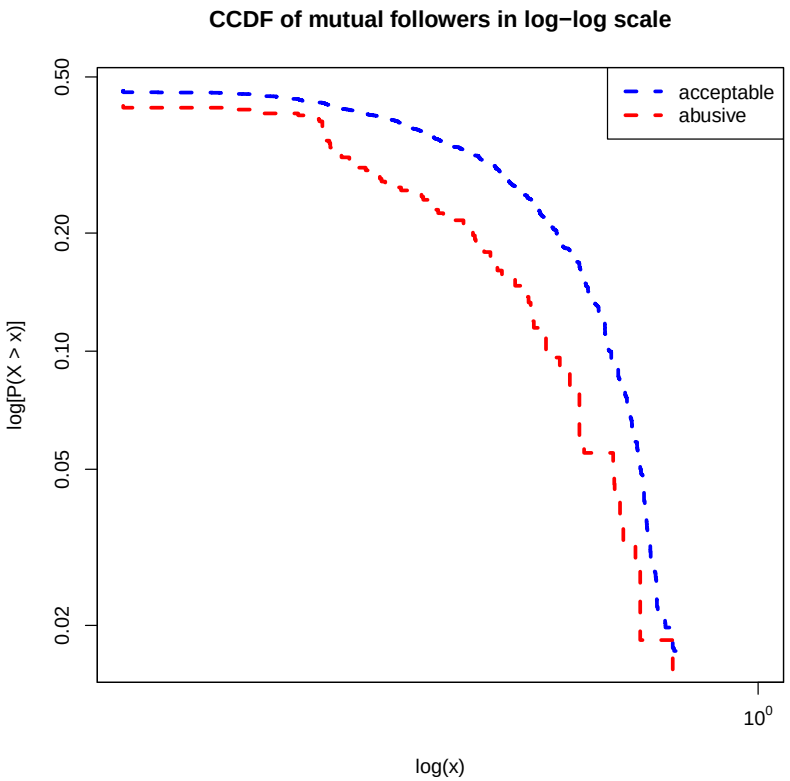


1. Alice sends:
   $\mathcal{X}_A :=$
   $\text{sort}\left[C^{t_A} \mid C \in \mathcal{A} \right]$
2. Bob responds with commitments:
   $\mathcal{X}'_{B,i}, \mathcal{Y}'_{B,i} \quad \text{for } i \in 1, \ldots, \kappa$
3. Alice picks a non-empty random subset $J \subseteq \{1, \ldots, \kappa\}$ and sends it to Bob.
4. Bob replies with $\mathcal{X}_{B,j}$ for $j \in J$, $t_{B,j}$ for $j \notin J$

Cut & choose version of Protocol 1: Verification

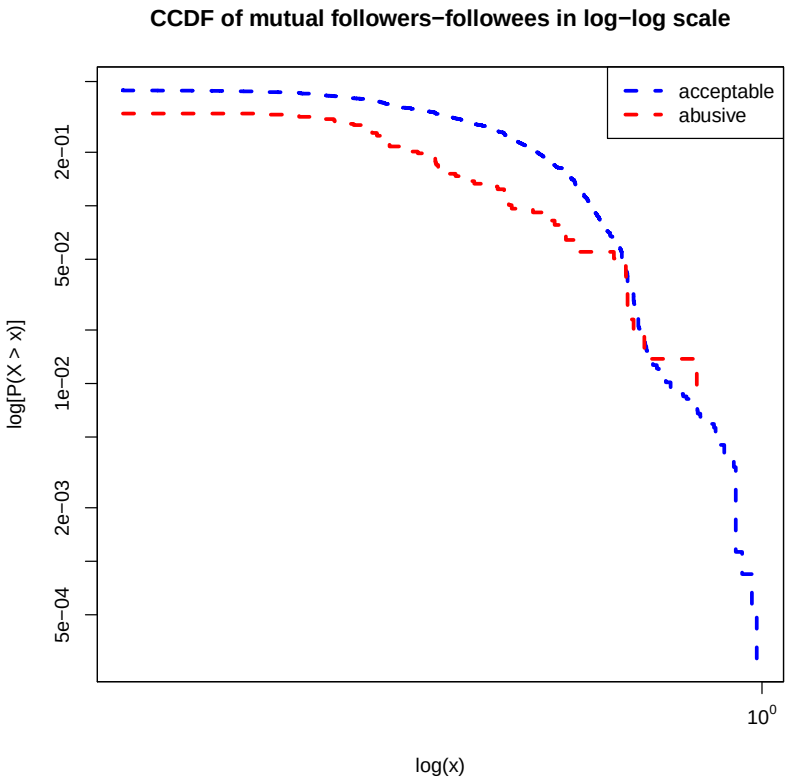
- For $j \notin J$, Alice checks the $t_{B,j}$ matches the commitment $\mathcal{Y}'_{B,j}$.
- For $j \in J$ Alice checks the commitment to $\mathcal{X}_{B,j}$ and computes: $\mathcal{Y}_{A,j} := \left\{ \hat{C}^{t_A} \;\middle|\; \hat{C} \in \mathcal{X}_{B,j} \right\}$
- Finally, Alice computes: $n = |\mathcal{Y}'_{A,j} \cap \mathcal{Y}'_{B,j}|$.
- Alice checks that n values for all $j \in J$, agree.

Privacy Analysis of PSI features



- $|\text{Subscriber} \cap \text{Subscriber}|$ CCDF shows that authors of abusive messages are less likely to have common subscribers.
- Security? Hard to prevent fake subscribers.
- Privacy? Yes, Protocol 1.

Privacy Analysis of PSI features



- CCDF of $|\text{Subscriber}^{\text{s}} \cap \text{Subscription}^{\text{r}}|$ shows less overlap among subscriptions of authors of abusive messages and subscriptions of the potential victims.
- Security? Assume more difficult for an adversary to increase feature overlap.
- Privacy? Yes, our Protocol with BLS signatures.

Protocol 2: PSI with Subscriber Signatures

- Assume Subscribers are willing to sign they are subscribed.
- Subscribers provide the signatures and not a certification authority.
- BLS signatures are compatible with our blinding, so we integrate them with our cut & choose version of the protocol.

Detailed protocol is in the paper.

What is Protocol 2 useful for?

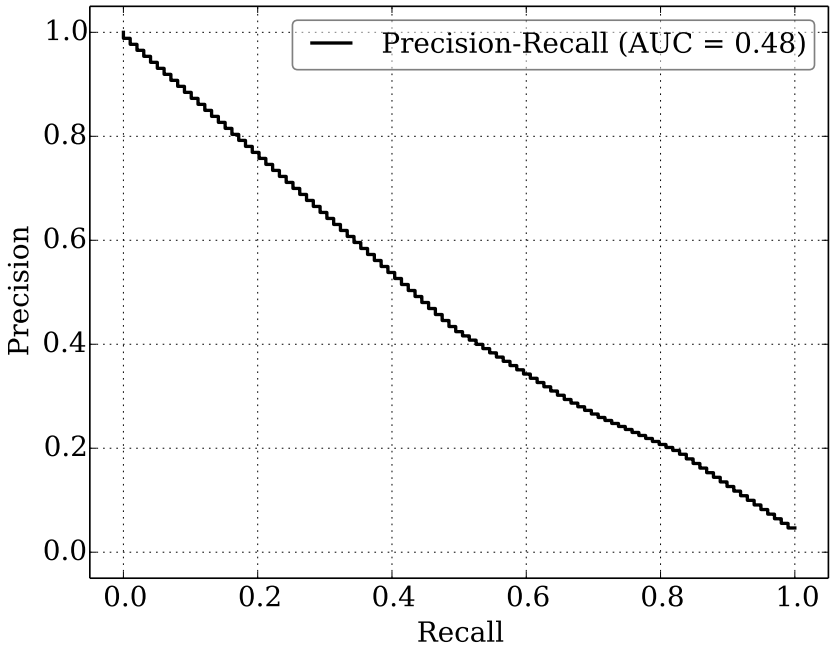
- Prove overlap of subscribers without reveling their identity.
- Key authentication in non-public Web-of-Trust (1-hop only).
- Unlike PSI-CA from De Cristofaro (2016), no need for a CA!

Privacy-preserving features

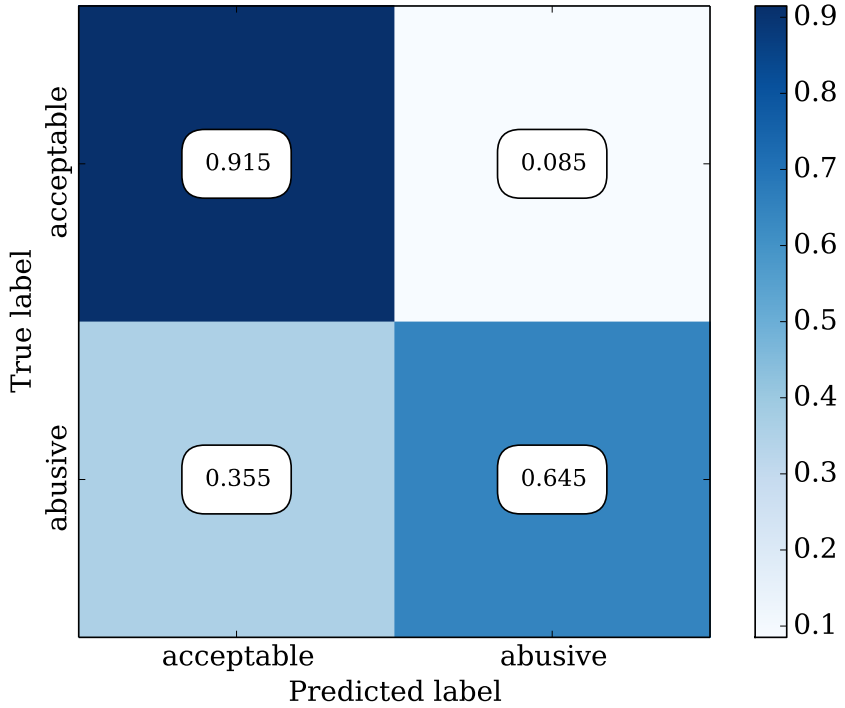
	Feature	Falsification/Adaptation	Crypto helps?
5.1	# lists	trivial	n/a
	# subscriptions	trivial	n/a
	$\frac{\#\text{ subscriptions}}{\text{age}}$	trivial	n/a
	$\frac{\#\text{subscriptions}}{\#\text{subscribers}}$	trivial	n/a
5.2	# mentions	costly	n/a
	# hashtags	costly	n/a
	$\frac{\#\text{ mentions}}{\text{age}}$	costly	yes
	$\frac{\#\text{ mentions}}{\#\text{ messages}}$	costly	n/a
5.3	message invasive	hard	n/a
5.4	$\frac{\#\text{ messages}}{\text{age}}$	costly	yes
	# retweets	costly	n/a
	# favorited messages	costly	n/a
5.5	age of account	hard	yes
5.6	# subscribers	possible	minimally
	$\frac{\#\text{ subscribers}}{\text{age}}$	possible	minimally
5.7	subscription $\cap$ subscription	costly	w. privacy
5.8	subscriber $\cap$ subscriber	possible	w. privacy
5.9	$\text{subscriber}^{\text{s}} \cap \text{subscription}^{\text{r}}$	very hard	yes
	$\text{subscription}^{\text{s}} \cap \text{subscriber}^{\text{r}}$	possible	w. privacy

Decision Trees with privacy

Objective function is to maximize AUC under the P-R curve.

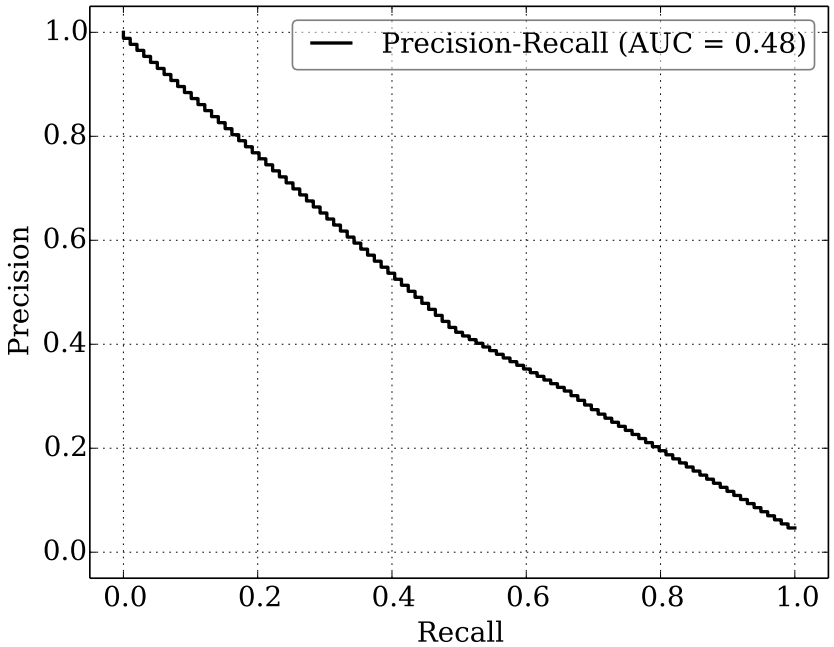


Objective function is to minimize FP and FN rates.

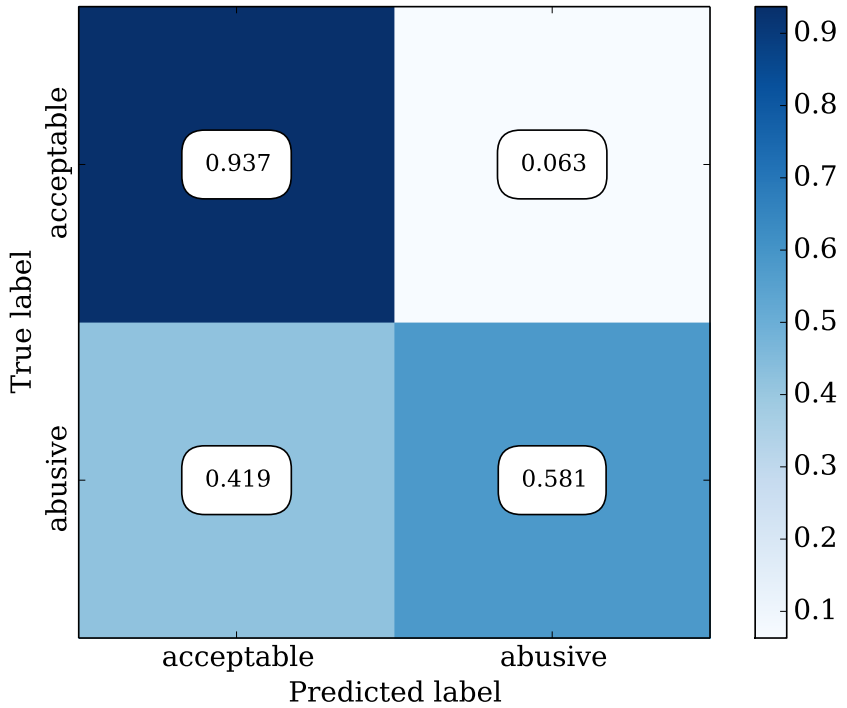


Random Forest with privacy

Objective function is to maximize AUC under the P-R curve.



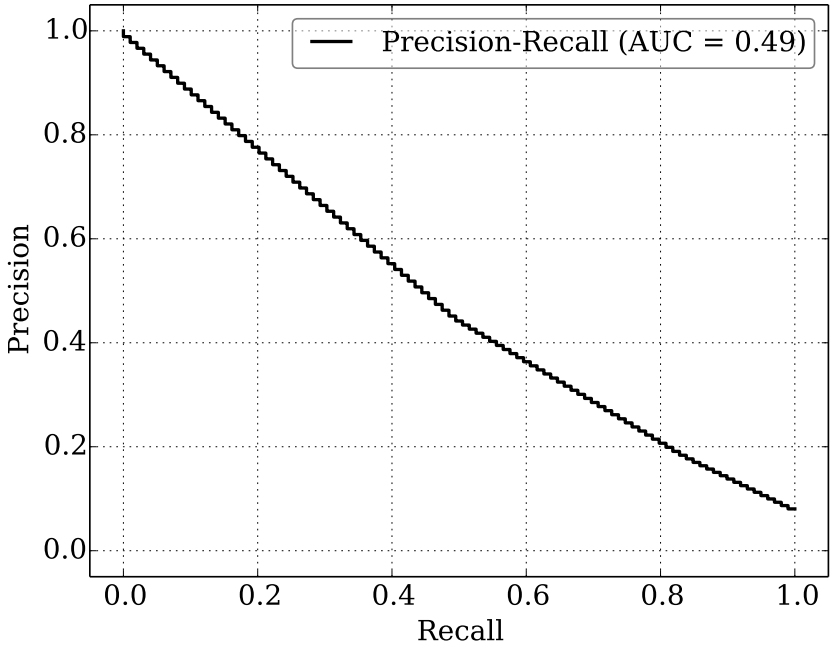
Objective function is to minimize FP and FN rates.



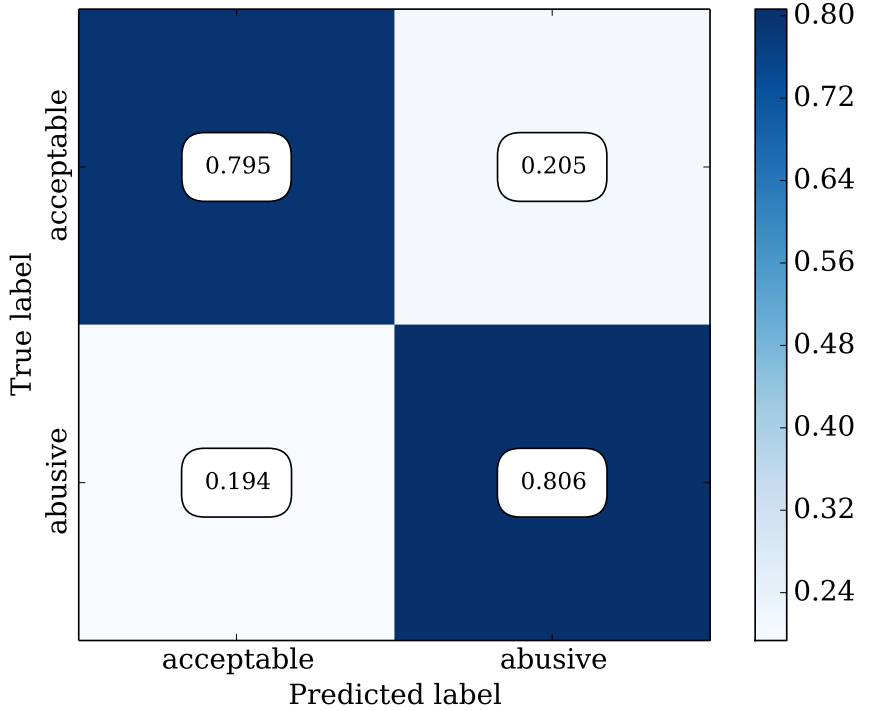
Extra Trees with privacy

Extra Trees is the most balanced among FP and FN, and has the best P-R curve.

Objective function is to maximize AUC under the P-R curve.



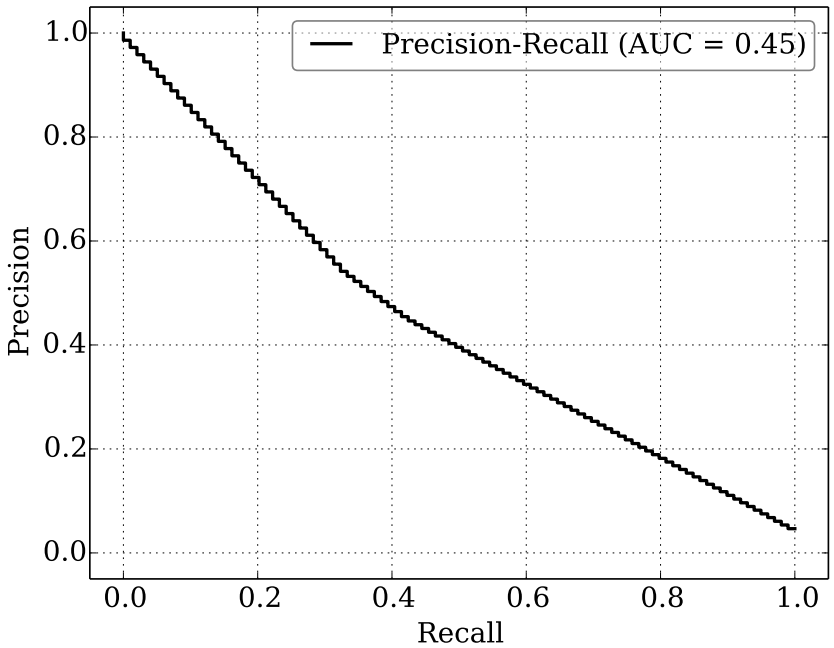
Objective function is to minimize FP and FN rates.



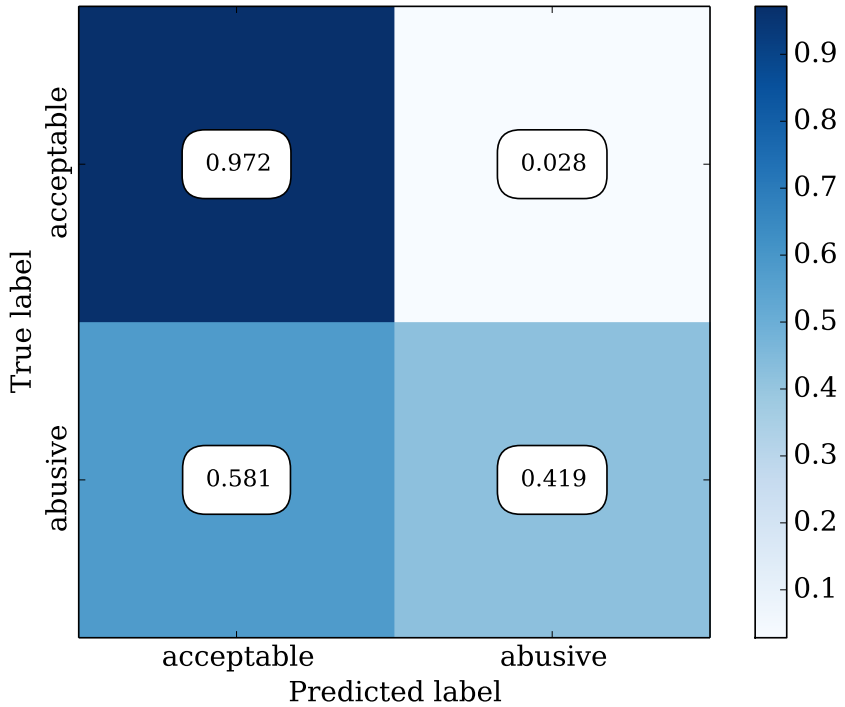
Gradient Boosting with privacy

Gradient Boosting = Gradient Descent + Boosting.

Objective function is to maximize AUC under the P-R curve.



Objective function is to minimize FP and FN rates.



- Method can protect privacy.
- Method can handle adaptive adversaries.
- Using reduced ground truth almost as Human Score.

Data minimisation

MinHashes

Data minimisation

- Intuition: we can use the intersection to estimated the approximated Jaccard index of $\mathcal{J}(\mathcal{A}, \mathcal{B})$ by counting the number of indexes (i) such as that $\mathrm{h}_{\min}^{\mathcal{A}}(\cdot) = \mathrm{h}_{\min}^{\mathcal{B}}(\cdot)$.

- Evaluate approximate, privacy-preserving PSI in terms of:
  (i) computation time
  (ii) accuracy of classification.

The Efficient Privacy-Preserving Protocol

- Approximated Jaccard index estimation with MinHashes reduces computational footprint.
- In addition, Data Minimisation provides our PP Protocol for DOSN just a fingerprint of the one-hop graph metadata.
- Note that even centralised Social Network providers as LinkedIn stop counting contacts after +500 on their site.

Results and performance

Features	Timing (ms)	# of hash. func. (k)	Error bound
All using $\mathcal{J}$ index)	3 018 632.98	–	–
All using approx. $\mathcal{J}$ index)	2 626 971.92	64	$\mathcal{O}(1/\sqrt{\mathrm{k}})$
All using approx. $\mathcal{J}$ index)	2 642 225.02	128	$\mathcal{O}(1/\sqrt{\mathrm{k}})$

- We see a reduction in computation time for the same set sizes (details in ASONAM '17 article)
- Supervised learning results come close to what we obtained using no approximation features with PSI, now the Jaccard index thanks to MinHashes.

Conclusions & Future Work

- Our protocol is resistant against malicious adversaries.
- Data minimisation reduces exposing training process to malicious adversaries tampering training samples.
- Use our protocol to support a decentralised Web-of-Trust that provides trust but also privacy.

Q & A

QUESTIONS?

Contact: algarecu.wordpress.com
Repos: github.com/algarecu

Publications

- Á. García-Recuero Efficient Privacy-Preserving Adversarial Learning in Decentralized Online Social Networks.
  In 2017 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Sydney, Australia.

- Á. García-Recuero, J. Burdges, and C. Grothoff.
  Privacy-preserving abuse detection in future decentralized online social networks.
  In 11th International ESORICS Workshop in Data Privacy Management, DPM 2016. Springer Lecture Notes in Computer Science.