On Privacy and Intersections

Constantinos Patsakis
Department of Informatics, University of Piraeus, Greece
Interdisciplinary Workshop on Data Privacy 2015
Maynooth University

October 7, 2015

Scope of this talk

Discuss about the problems of **Private Equality Testing**, **Private Set Intersection** and **Private Set Similarity**.

Private Equality Testing

PET

Let us assume that we have Alice and Bob who have a value and are willing to share **only one bit of information** whether their values are the same or not.

Private Proximity Testing

PPT

Alice and Bob want to test whether they are in proximity or not.

Private Set Intersection

PSI

Let us assume that we have Alice and Bob who are willing to share their common elements, but **nothing** more with each other.

Private Set Similarity

PSS

Let us assume that we have Alice and Bob are willing to share how many common elements they do have but **nothing** more with each other.

Excesive use of PSI/S

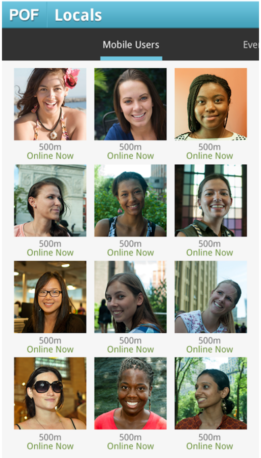
What happens after many queries?

Alice may start comparisons with 000...0 and 100...0 to find Bob's set using PSS

What can we do with them?

- Determine if two people are in the same area or in proximity (more on this next).
- Determine if two people know each other in a OSN.
- Determine if Alice has a bank account in Bank ABC.
- Determine if a suspect is in a flight.
- Determine if a person is in two suspect lists.
- Document similarity
- Biometric authentication (more on this next)
- DNA similarity/queries
- Private distributed Uber-like services (under development)

Location-awareness of OSNs



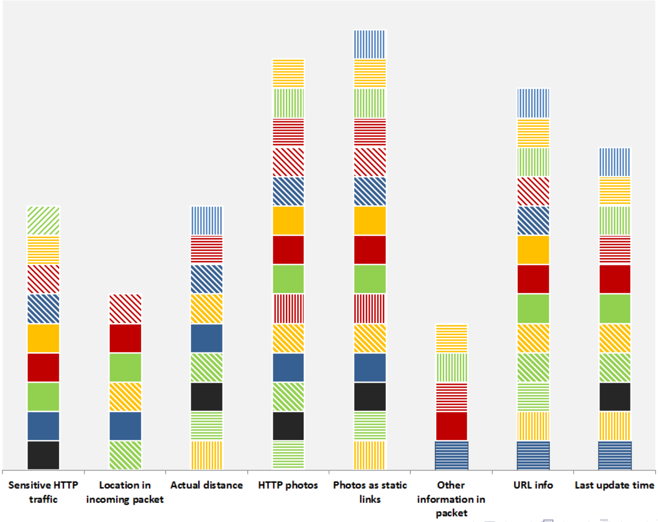
Questions

- How accurate are these distances?
- Do these applications allow location-based attacks?
- Could we locate people from the reported distances? If so, with what accuracy?
- What kind of data are they sending?
- How do they send this data?
- What others can infer?

Numbers...

Application	Version	Installations	Code	Application	Version	Installations	Code
ChatOn	3.0.2	100m-500m		Singles around me	3.3.1	500K-1m	
Grindr	2.0.24	5m-10m		SKOUT	4.4.2	10m-50m	
Hornet	2.0.14	1m-5m		Tagged	7.3.0	10m-50m	
I-Am	3.2	500K-1m		Tango	5.8	100m-500m	
LOVOO	2.5.6	10m-50m		Tinder	4.0.9	10m-50m	
MeetMe	9.2.0	10m-50m		Tingle	1.12	-	
MoMo	5.4	1m-5m		Waplog	3.0.3	5m-10m	
POF	2.40	10m-50m		WeChat	6.0.1	100m-500m	
SayHi	3.6	10m-50m		Zoosk	8.6.21	10m-50m	

Vulnerabilities



Authentication

With passwords we authenticate users using something that they *know*. Another approach is to authenticate users by something that they *are*, something that cannot be forgotten or forged.

Biometric authentication

- Iris
- Retina
- Fingerprint
- Face
- Vains
- Gait
- Ear
- Palm
- ...

Drawbacks

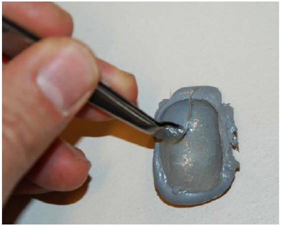
They are not exact

Regardless of the underlying data, every measurement is not exactly the same as the one registered.

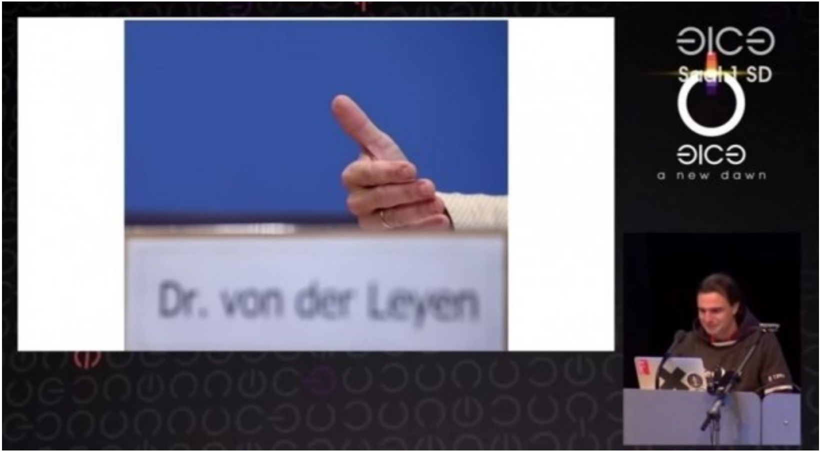
They are permanent

While one could easily pick another password if it has been compromised, what should a user do if her biometrics are lost?

Copy+Paste

Copy+Paste



Copy+Paste

> *"Research is needed in order to explore whether it is possible to use other biometric data (potentially already used in another context and in another domain) than fingerprint, iris or facial picture to store in the e-Passport chip, which would guarantee the same or higher level of security, but would be more accurate and could be retrieved in a more efficient manner than in the case of the conventionally used biometric data types."*

From BES-06-2015 H2020 Border crossing points topic 2: Exploring new modalities in biometric-based border checks

The need

We need protocols which do *not* leak any information about the biometrics, or at least minimize the exposure.

The need – refined

We need protocols which do *not* leak any information when exchanging sensitive information about individuals, or at least minimize the exposure.

Related work

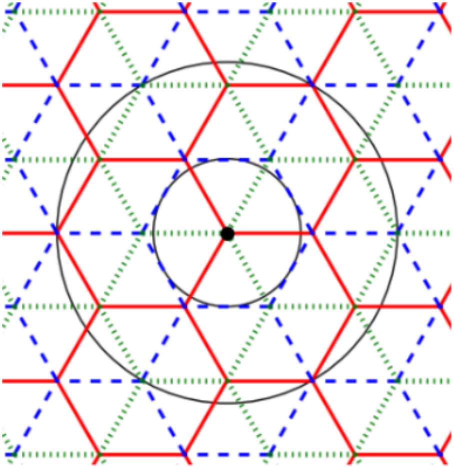
The PET problem was first introduced by Huberman, Franklin, and Hogg in 1999. Their proposed solution used the DH key agreement scheme.

Why don't you simply use hashes?

Related work (cont)

The PSI problem was introduced by Freedman, Nissim, and Pinkas in 2004. The proposed solution used polynomial interpolation. Currently, there are many approaches using e.g. Yao's garbled circuits, Bloom filters and Oblivious Transfer. However, one work that must be mentioned is the work of De Cristofaro who reduced the complexity of the problem to linear.

The Narayanan et al. grid



Narayanan et al. protocol

Alice picks a random r and calculates $C_A = (g_r, h^{a+r})$ and sends it to Bob. On receiving $C_A = (g_1, g_2)$, Bob picks two random numbers s and t and computes: $C_B = (g_1^s g^t, g_2^s h^{(t-sb)})$ and sends it to Alice. When Alice receives $C_B = (u_1, u_2)$, she computes $m = u_2 u_1^{-x}$. If $m = 1$, Alice knows that she is in proximity to Bob, otherwise, she cannot deduce anything more about Bob's whereabouts.

NTRU

A lattice-based encryption algorithm, very good homomorphic properties (Semi Homomorphic Encryption), extremely fast and secure, even from quantum attacks.

NTRU: Setup

We then select two random polynomials f and g with small coefficients (-1, 0 and 1). We also require f to be invertible in $\mathbb{Z}_q[x]/(x^N - 1)$ and $\mathbb{Z}_p[x]/(x^N - 1)$, and we denote these inverses f_q and f_p respectively. The public key h is defined as $h = pgf_q$, while f and f_p are the private key.

NTRU: parameters

Level(bits)	p	q	n	D_1	D_2	D_3	D_g	D_m
128	3	2048	439	9	8	5	146	112
192	3	2048	593	10	10	8	197	158
256	3	2048	743	11	11	15	247	204

NTRU: Encrypt/Decrypt

Encrypt

We map the message to a polynomial m with small coefficients and pick a random "small" polynomial r, and send the message $c = hr + m \in \mathbb{Z}_q[x]/(x^N - 1)$.

Decrypt

The recipient multiplies it with f and rearranges the coefficients to reside within $[-q/2, q/2]$ and reduces it modulo p. Finally, she multiplies the result with f_p.

Privacy-Preserving Biometric Authentication

Many approaches such as:

- Blanton et al. [1] exploit the homomorphic properties of the encryption method of Damgard et al. [2].
- Shahandashti et al. [3] the Paillier homomorphic scheme for private fingerprint matching.

These methods are too slow and consume a lot of bandwidth.

Blundo et al. sampling

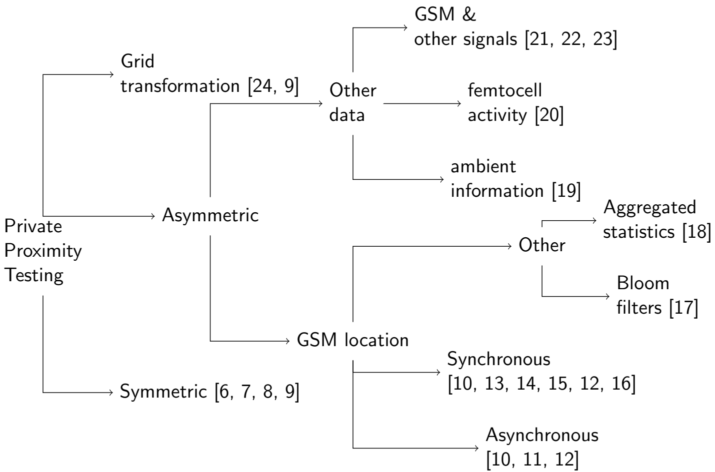
Blundo et al. [4] proposed a probabilistic protocol for the privacy-preserving evaluation of "sample set similarity". They use MinHash to sample each set, and perform the protocol of De Cristofaro et al. [5] to determine the *cardinality* of the common elements of both sets.

From PET to PPT

Private Equality Testing (PET)

- Alice and Bob that want to reveal **only a single** bit of information: whether they have the same secret value or not.
- PET is a potential building block for a PPT protocol. Locations are represented by geographical cells.
- Then each cell is mapped to a unique value in a finite set (a unique id for each "cell").
- Problem: What happens if users **are in proximity** but they reside at the edges of **neighboring cells**?

A roadmap of PPT protocols



Setup

- We use the idea of overlapping grids of Narayanan et al. [10], to reduce PPT to PET.
- For simplicity we describe the procedure in one grid. In each grid, each cell is marked as ℓ_i.
- The scope of the protocol is to determine whether two users, Alice and Bob are in the same cell or not.
- The set of all possible cells is denoted as $\mathcal{L}$, so $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_k\}$ where $|\mathcal{L}| \leq \mathcal{O}(2^{32})$.

A factoring based protocol

Alice (the initiator of the protocol) has published $n_A = p_A \cdot q_A$ (where p_A, q_A are kept private) and she currently uses the private key pair $d_A (\equiv l_A)$ and e_A. In the same way, Bob has published $n_B = p_B \cdot q_B$ and he currently uses the private key pair $d_B (\equiv l_B)$ and e_B.

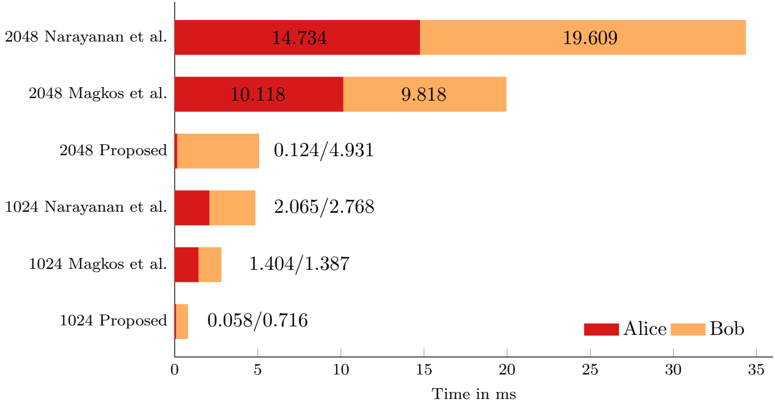
A factoring based protocol (cont.)

Step 1: Alice picks a random integer r_A of high entropy (say, 1024 bit) and computes: $c_A = ({r_A}^{l_A} \bmod n_B)^{e_A} \bmod n_A$ and sends it to Bob.

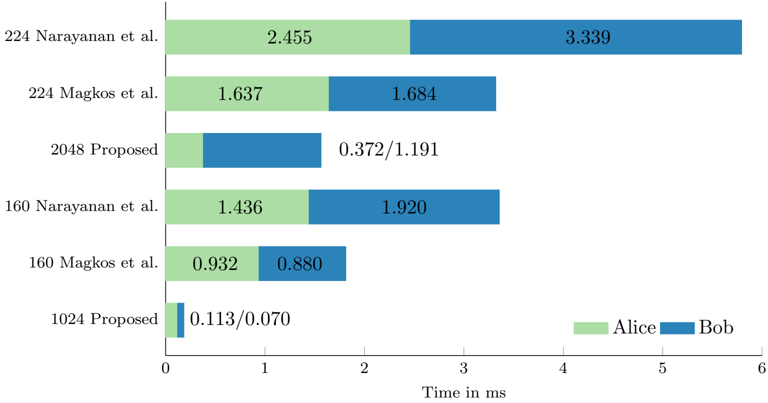
Step 2: Bob computes $x = ({c_A}^{l_B} \bmod n_A)^{e_B} \bmod n_B$ and $c_B = H(x)$ and sends c_B to Alice.

Step 3: Alice checks whether $c_B \stackrel{?}{=} H(r_A \bmod n_B \bmod n_A)$. If the equality holds, Alice is convinced that $l_A = l_B$. If not, Alice learns nothing about Bob's private input.

Efficiency


2048 Narayanan et al.
14.734
19.609
2048 Magkos et al.
10.118
9.818
2048 Proposed
0.124/4.931
1024 Narayanan et al.
2.065/2.768
1024 Magkos et al.
1.404/1.387
1024 Proposed
0.058/0.716
Alice
Bob
0
5
10
15
20
25
30
35
Time in ms

Efficiency with EC



Communication cost

	Alice	Bob
Narayanan et al.	1024	2048
Magkos et al.	1024	1280
Proposed protocol	1024	256

Description

Let Alice be located in ℓ_A and Bob in ℓ_B.

Step 1: Alice sends the message $c_A = rh + \ell_A$ to Bob, where r is a random invertible polynomial in $\mathbb{Z}_q[x]/(x^N - 1)$.

Step 2: Bob picks a random polynomial ρ with small coefficients and sends Alice $c_B = \rho(c_A - \ell_B)$.

Step 3: Alice receives it and checks whether $r^{-1}c_B$ decrypts to zero.

Correctness

Assume that $l_A = l_B$. Let m denote the expected encrypted message.

- In step 2, Bob computes:

$$c_B \equiv \rho(c_A - \ell_B) \equiv rh\rho$$

  that he will sent to Alice.
- When Alice in step 3 decrypts:

$$r^{-1}c_B = r^{-1}rh\rho \equiv h\rho$$

  the result will be 0, otherwise it will be a random polynomial.

Protocol implementation

- We compare our protocol with the Narayanan et al. in an Intel Core i3-2100 CPU(3.1GHz) with 6GB of RAM, running on 64-bit linux using Sage[1].
- The protocol of Narayanan et al. has been implemented over elliptic curves, using Curve25519 [25] for 128-bits of security, and for 192 and 256 bits security we used the curves M-383 and M-511 respectively [26].
- For NTRU we have used the latest parameters proposed by SecurityInnovation [27].

Table: NTRU parameters for different security levels

Level(bits)	**p**	**q**	**n**	D_1	D_2	D_3	D_g	D_m
128	3	2048	439	9	8	5	146	112
192	3	2048	593	10	10	8	197	158
256	3	2048	743	11	11	15	247	204

Computation and communication costs

	Narayanan et al.			Proposed			R
Security	Alice	Bob	Total	Alice	Bob	Total	
128	80.718	99.194	179.912	7.362	1.051	8.413	21
192	102.267	133.873	236.140	10.527	1.518	12.045	19
256	155.329	193.887	349.216	12.733	1.745	14.478	24

Table: Time in ms and Security in bits.

Security	Narayanan et al.	Proposed
128	128	1208
192	192	1630
256	256	2044

Setup

We assume that Alice has created an NTRU key pair, so h is her public key and f, f_p her private.

Both parties split their feature vectors in blocks of length λ, creating k blocks.

Moreover, we assume that both of them know a function $\chi : \{0,1\}^\lambda \rightarrow \mathbb{D}$, where $\mathbb{D}$ contains the polynomials of $\mathbb{Z}_q[x]/(x^N - 1)$ with coefficients -1, 0 and 1. For the sake of simplicity instead of $\chi(m)$ we will write m.

Description

Let α_i and $\beta_i, i \in [1, k]$ denote the blocks of Alice and Bob respectively.

Step 1: Alice sends Bob the message

$$M_A = \{hs_i + \alpha_i\}, \forall i \in [1, k]$$

where s_i are random polynomials in $\mathbb{D}$.

Step 2: Bob computes the vector

$$M_B = \{M_{A_i} - (hs_i' + \beta_i)\}, \forall i \in [1, k]$$

where s_i' are random polynomials in $\mathbb{D}$.
Bob picks a random permutation π and sends Alice $M_B' = \pi(M_B)$.

Step 3: Alice decrypts each $M_{B_i'}$ and computes the weight w_i of each recovered message. If $\sum_{i=1}^{k} w_i < \tau$ then Alice deduces that $d_H(\mathcal{A}, \mathcal{B}) < \tau$.

Protocol implementation

- We compare our protocol with the Narayanan et al. in an Intel Core i3-2100 CPU(3.1GHz) with 6GB of RAM, running on 64-bit linux using Sage[2].
- For NTRU we have used the latest parameters proposed by SecurityInnovation [27].

Table: NTRU parameters for different security levels

Level(bits)	p	q	n	D_1	D_2	D_3	D_g	D_m
128	3	2048	439	9	8	5	146	112
192	3	2048	593	10	10	8	197	158
256	3	2048	743	11	11	15	247	204

[2] sagemath.org

Parameters

Security Level	RSA	NTRU p	NTRU q	NTRU n	NTRU Public key (bits)
128	3072	3	2048	439	4829
192	7680	3	2048	593	6523
256	15360	3	2048	743	8173

Parameters for the most popular security levels (in bits). For RSA the numbers denote the length (in bits) of the underlying modulo field according to NIST [28]. For NTRU, the numbers are precise and recommended by SecurityInnovation (`https://github.com/NTRUOpenSourceProject/ntru-crypto/blob/master/doc/NewParameters.pdf`).

Computation cost

Security	Blundo et al.			Proposed		
	Alice	Bob	Total	Alice	Bob	Total
128	0.024	2.227	2.251	0.187	0.115	0.302
192	0.066	12.352	12.418	0.250	0.153	0.403
256	0.183	59.421	59.605	0.299	0.220	0.519

Communication cost

Security	Blundo et al.	Proposed
128	78.125	75.453
192	190.625	101.922
256	378.125	127.703

Approximate communication cost in KB. Security in bits.

A generic attack

Alice performs one execution of the protocol with Bob using firstly the sequence $00\dots000$ and then $10\dots000$. She can tell which one is closest...

A patch

Let $\mathcal{F}(k, x)$ denote a Pseudo Random Function (PRF), where k is the PRF key and x is the point at which the function is evaluated. Bob proposes a random seed s so Alice and Bob compute the following for their sequences: $\mathcal{F}(s, m_i || i) \mod 2, i \in \{1, 2, \ldots, k\}$.

Where to find these results

- C. Patsakis, J. van Rest, M. Choras and M. Bouroche Privacy-Preserving Biometric Authentication and Matching via Lattice-Based Encryption 10th International Workshop on Data Privacy Management (DPM 2015), Vienna, Austria - September 21-22, 2015.
- C. Patsakis, P. Kotzanikolaou and M. Bouroche, Private Proximity Testing on Steroids: An NTRU-based protocol", 11th International Workshop on Security and Trust Management (STM 2015), 21-22 September 2015, Vienna, Austria.
- C. Patsakis, A. Zigomitros, A. Solanas, "Analysis of Privacy and Security Exposure in Mobile Dating Applications", International Conference on Mobile, Secure and Programmable Networking (MSPN'2015), Paris, France, June 15-17, 2015.

Where to find these results(cont)

- C. Patsakis, A. Zigomitros, A. Solanas, "Analysis of Privacy and Security Exposure in Mobile Dating Applications", International Conference on Mobile, Secure and Programmable Networking (MSPN'2015), Paris, France, June 15-17, 2015.
- C. Patsakis, A. Zigomitros, A. Solanas, Privacy-Aware Genome Mining: Server-Assisted Protocols for Private Set Intersection and Pattern Matching", 28th International Conference on Computer-Based Medical Systems, CBMS 2015, 22-25 June, Sao Paulo, Brazil.
- P. Kotzanikolaou, C. Patsakis, E. Magkos, M. Korakakis, "Lightweight Private Proximity Testing for Geospatial Social Networks", Computer Communications, Special Issue on Online Social Networks (Imprint)

Thank you!

Q&A

kpatsak@{unipi.gr/gmail.com/protonmail.com}
www.cs.unipi.gr/kpatsak

Bibliography I

M. Blanton, P. Gasti, Secure and efficient protocols for iris and fingerprint identification, in: Computer Security–ESORICS 2011, Springer, 2011, pp. 190–209.

I. Damgard, M. Geisler, M. Kroigard, Homomorphic encryption and secure comparison, International Journal of Applied Cryptography 1 (1) (2008) 22–31.

S. F. Shahandashti, R. Safavi-Naini, P. Ogunbona, Private fingerprint matching, in: Information Security and Privacy, Springer, 2012, pp. 426–433.

Bibliography II

C. Blundo, E. De Cristofaro, P. Gasti, EsPRESSo: efficient privacy-preserving evaluation of sample set similarity, in: Data Privacy Management and Autonomous Spontaneous Security, Springer, 2013, pp. 89–103.

E. De Cristofaro, P. Gasti, G. Tsudik, Fast and private computation of cardinality of set intersection and union, in: Cryptology and Network Security, Springer, 2012, pp. 218–231.

L. Šikšnys, J. R. Thomsen, S. Šaltenis, M. L. Yiu, O. Andersen, A location privacy aware friend locator, in: Advances in Spatial and Temporal Databases, Springer, 2009, pp. 405–410.

Bibliography III

- L. Siksnys, J. R. Thomsen, S. Saltenis, M. L. Yiu, Private and flexible proximity detection in mobile social networks, in: Mobile Data Management (MDM), 2010 Eleventh International Conference on, IEEE, 2010, pp. 75–84.
- S. Mascetti, D. Freni, C. Bettini, X. S. Wang, S. Jajodia, Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies, The VLDB journal 20 (4) (2011) 541–566.
- K. P. Puttaswamy, S. Wang, T. Steinbauer, D. Agrawal, A. El Abbadi, C. Kruegel, B. Y. Zhao, Preserving location privacy in geosocial applications, Mobile Computing, IEEE Transactions on 13 (1) (2014) 159–173.

Bibliography IV

A. Narayanan, N. Thiagarajan, M. Lakhani, M. Hamburg, D. Boneh, Location privacy via private proximity testing, in: NDSS, The Internet Society, 2011.
URL http://www.isoc.org/isoc/conferences/ndss/11/

E. Novak, Q. Li, Near-pri: Private, proximity based location sharing, in: INFOCOM, 2014 Proceedings IEEE, IEEE, 2014, pp. 37–45.

G. Zhong, I. Goldberg, U. Hengartner, Louis, lester and pierre: Three protocols for location privacy, in: Privacy Enhancing Technologies, Springer, 2007, pp. 62–76.

Bibliography V

- S. Chatterjee, K. Karabina, A. Menezes, A new protocol for the nearby friend problem, in: Proceedings of the 12th IMA International Conference on Cryptography and Coding, Springer-Verlag, 2009, pp. 236–251.
- J. D. Nielsen, J. I. Pagter, M. B. Stausholm, Location privacy via actively secure private proximity testing, in: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on, IEEE, 2012, pp. 381–386.

Bibliography VI

- E. Magkos, P. Kotzanikolaou, M. Magioladitis, S. Sioutas, V. S. Verykios, Towards secure and practical location privacy through private equality testing, in: Privacy in Statistical Databases, Springer, 2014, pp. 312–325.

- P. Kotzanikolaou, C. Patsakis, E. Magkos, M. Korakakis, Lightweight private proximity testing for geospatial social networks, Computer Communications(Accepted for publication).

Bibliography VII

- P. Palmieri, L. Calderoni, D. Maio, Spatial bloom filters: Enabling privacy in location-aware applications, in: D. Lin, M. Yung, J. Zhou (Eds.), Information Security and Cryptology - 10th International Conference, Inscrypt 2014, Beijing, China, December 13-15, 2014, Revised Selected Papers, Vol. 8957 of Lecture Notes in Computer Science, Springer, 2014, pp. 16–36.
- R. A. Popa, A. J. Blumberg, H. Balakrishnan, F. H. Li, Privacy and accountability for location-based aggregate statistics, in: Proceedings of the 18th ACM conference on Computer and communications security, ACM, 2011, pp. 653–666.

Bibliography VIII

- T. Halevi, D. Ma, N. Saxena, T. Xiang, Secure proximity detection for nfc devices based on ambient sensor data, in: Computer Security–ESORICS 2012, Springer, 2012, pp. 379–396.
- J. Brassil, R. Netravali, S. Haber, P. Manadhata, P. Rao, Authenticating a mobile device's location using voice signatures, in: Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on, IEEE, 2012, pp. 458–465.
- G. Saldamli, R. Chow, H. Jin, B. Knijnenburg, Private proximity testing with an untrusted server, in: Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks, ACM, 2013, pp. 113–118.

Bibliography IX

- Y. Zheng, M. Li, W. Lou, Y. T. Hou, Sharp: Private proximity test and secure handshake with cheat-proof location tags, in: Computer Security–ESORICS 2012, Springer, 2012, pp. 361–378.
- Z. Lin, D. F. Kune, N. Hopper, Efficient private proximity testing with gsm location sketches, in: Financial Cryptography and Data Security, Springer, 2012, pp. 73–88.
- M. Li, H. Zhu, Z. Gao, S. Chen, K. Ren, L. Yu, S. Hu, All your location are belong to us: Breaking mobile social networks for automated user location tracking, arXiv preprint arXiv:1310.2547.

Bibliography X

- D. J. Bernstein, Curve25519: new diffie-hellman speed records, in: Public Key Cryptography-PKC 2006, Springer, 2006, pp. 207–228.
- D. F. Aranha, P. S. L. M. Barreto, G. C. C. F. Pereira, J. E. Ricardini, A note on high-security general-purpose elliptic curves, Cryptology ePrint Archive, Report 2013/647, `http://eprint.iacr.org/` (2013).
- J. Hoffstein, J. Pipher, J. M. Schanck, J. H. Silverman, W. Whyte, Z. Zhang, Choosing parameters for ntruencrypt, Cryptology ePrint Archive, Report 2015/708, `http://eprint.iacr.org/` (2015).

Bibliography XI

E. Barker, Q. Dang, NIST special publication 800-57 part 3: Application-specific key management guidance, NIST Special Publication 800 (57).