Improved User-Private Information Retrieval via Finite Geometry

RMIT

Padraig Ó Catháin (WPI)
joint with Oliver W. Gnilke, Marcus Greferath, Camilla Hollanti, Guillermo Nuñez Ponasso, Eric Swartz

7th October 2019

Private Information Retrieval

- I want to download the i^{th} file F_i of a Database
- I do **not** want someone who observes my request or the response from the Database to learn i.

Private Information Retrieval

- I want to download the i^{th} file F_i of a Database
- I do **not** want someone who observes my request or the response from the Database to learn i.
- With a single Database, perfect privacy requires downloading all the files.
- What about multiple Databases?

Private Information Retrieval

- I want to download the i^{th} file F_i of a Database
- I do **not** want someone who observes my request or the response from the Database to learn i.
- With a single Database, perfect privacy requires downloading all the files.
- What about multiple Databases?
- Assume all files are binary, and of equal length. Then request a random linear combination $S = \sum_{j \in J} F_j$ of files from D_1
- Request $S + F_i$ from D_2, and compute the sum of the responses to recover F_i.

Private Information Retrieval

- I want to download the i^{th} file F_i of a Database
- I do **not** want someone who observes my request or the response from the Database to learn i.
- With a single Database, perfect privacy requires downloading all the files.
- What about multiple Databases?
- Assume all files are binary, and of equal length. Then request a random linear combination $S = \sum_{j \in J} F_j$ of files from D_1
- Request $S + F_i$ from D_2, and compute the sum of the responses to recover F_i.
- This works, if an eavesdropper agrees to observe only a single database...

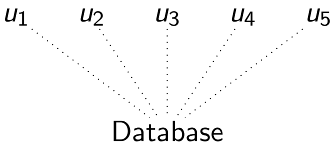




User Private Information Retrieval

Setup

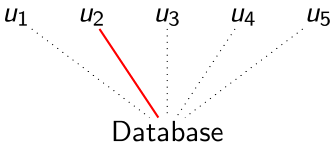
- A set $\mathcal{U}$ of users wants to communicate with an honest-but-curious database



User Private Information Retrieval

Setup

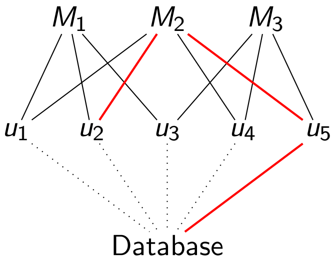
- A set $\mathcal{U}$ of users wants to communicate with an honest-but-curious database
- If the users send their requests directly an observer will be aware of the identity of the user



User Private Information Retrieval

Setup

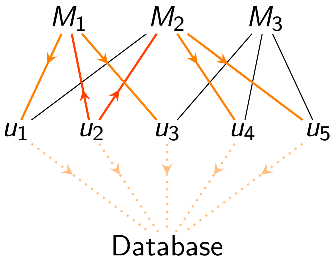
- A set $\mathcal{U}$ of users wants to communicate with an honest-but-curious database
- Therefore the users will forward each others' requests via shared message spaces M_i, that are not visible to outside observers



User Private Information Retrieval

Setup

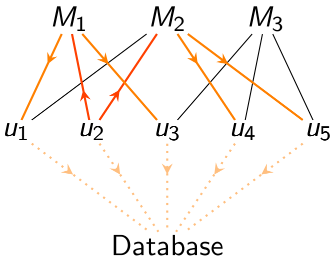
- A set $\mathcal{U}$ of users wants to communicate with an honest-but-curious database
- Therefore the users will forward each others' requests via shared message spaces M_i, that are not visible to outside observers
- If the users choose the proxy uniformly at random from the set of all users, perfect anonymity wrt. the database is achieved



User Private Information Retrieval

Setup

- A set $\mathcal{U}$ of users wants to communicate with an honest-but-curious database
- Therefore the users will forward each others' requests via shared message spaces M_i, that are not visible to outside observers
- If the users choose the proxy uniformly at random from the set of all users, perfect anonymity wrt. the database is achieved
- But what do the other users learn?



User Private Information Retrieval

Behaviour of the users

- Swanson and Stinson proved that user u_i has perfect secrecy with respect to outside observers if and only if u_i selects proxies uniformly at random from all of $\mathcal{U}$ (including u_i).

User Private Information Retrieval

Behaviour of the users

- Swanson and Stinson proved that user u_i has perfect secrecy with respect to outside observers if and only if u_i selects proxies uniformly at random from all of $\mathcal{U}$ (including u_i).
- All eavesdroppers will be considered honest-but-curious: they forward messages and follow instructions in the same way as non-eavesdroppers, but they remember queries they have seen, and may communicate these to other eavesdroppers.

User Private Information Retrieval

Behaviour of the users

- Swanson and Stinson proved that user u_i has perfect secrecy with respect to outside observers if and only if u_i selects proxies uniformly at random from all of $\mathcal{U}$ (including u_i).
- All eavesdroppers will be considered honest-but-curious: they forward messages and follow instructions in the same way as non-eavesdroppers, but they remember queries they have seen, and may communicate these to other eavesdroppers.
- In earlier works the requirement that every pair of users share at exactly one message space has been made: PBD

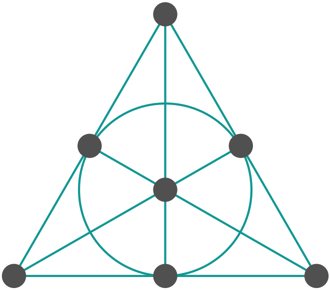
User Private Information Retrieval

Behaviour of the users

- Swanson and Stinson proved that user u_i has perfect secrecy with respect to outside observers if and only if u_i selects proxies uniformly at random from all of $\mathcal{U}$ (including u_i).
- All eavesdroppers will be considered honest-but-curious: they forward messages and follow instructions in the same way as non-eavesdroppers, but they remember queries they have seen, and may communicate these to other eavesdroppers.
- In earlier works the requirement that every pair of users share at exactly one message space has been made: PBD
- If all message spaces are the same size, and their number is minimized: projective plane

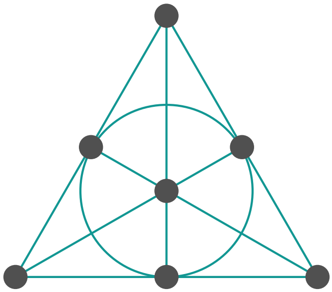
Projective planes

- Every pair of points determine a unique line.
- Every pair of lines intersect in a unique point.
- There exist at least four points no three collinear.



Projective planes

- Every pair of points determine a unique line.
- Every pair of lines intersect in a unique point.
- There exist at least four points no three collinear.

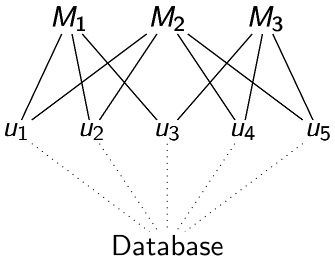


- Let V be a three dimensional vector space over field k.
- 1-d subspaces are *projective points*.
- 2-d subspaces are *projective lines*.

Linked Queries

Setup

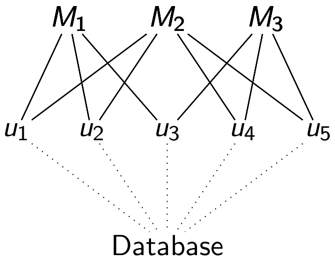
- Queries can be linked by their content, e.g. obscure topics



Linked Queries

Setup

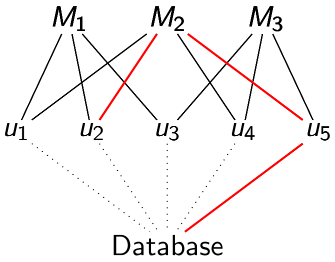
- Queries can be linked by their content, e.g. obscure topics
- Or by meta-content like user behaviour, timing, headers, etc.



Linked Queries

Setup

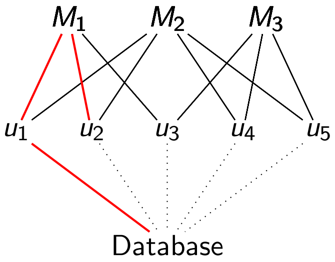
- Queries can be linked by their content, e.g. obscure topics
- Or by meta-content like user behaviour, timing, headers, etc.
- Collecting enough of these queries could identify a user within the network as the source of such requests and hence compromise her anonymity.



Linked Queries

Setup

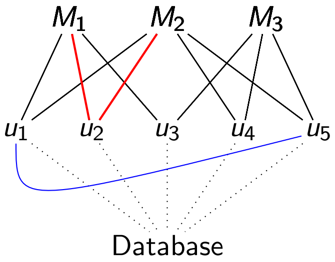
- Queries can be linked by their content, e.g. obscure topics
- Or by meta-content like user behaviour, timing, headers, etc.
- Collecting enough of these queries could identify a user within the network as the source of such requests and hence compromise her anonymity.



Linked Queries

Setup

- Queries can be linked by their content, e.g. obscure topics
- Or by meta-content like user behaviour, timing, headers, etc.
- Collecting enough of these queries could identify a user within the network as the source of such requests and hence compromise her anonymity.
- **Intersection attack!**



Privacy and Pseudonymity

- What is a good measure of privacy?
- Let $\mathcal{C}$ be a coalition of conspirators.
- Say that users u and v are **pseudonymous** if for any possible query observed by $c \in \mathcal{C}$ we have

$$\frac{\mathbb{P}(u \text{ sent } Q \mid c \text{ observed } Q)}{\mathbb{P}(u \text{ sent } Q)} = \frac{\mathbb{P}(v \text{ sent } Q \mid c \text{ observed } Q)}{\mathbb{P}(v \text{ sent } Q)}$$

- A family of UPIR systems is **secure** against coalitions of size t, if for any $\mathcal{C}$ of at most t users, the probability that two users chosen uniformly at random are pseudonymous tends to 1 as the number of users tends to ∞.

Proejctive planes are always bad

- Suppose that every pair of users share a message space, and that users always send messages via shortest paths.

Proejctive planes are always bad

- Suppose that every pair of users share a message space, and that users always send messages via shortest paths.
- Why? What are the pseudonymity classes with respect to user c?

Proejctive planes are always bad

- Suppose that every pair of users share a message space, and that users always send messages via shortest paths.
- Why? What are the pseudonymity classes with respect to user c?
- If $c, u_1 \in M_1$ and $u_2 \notin M_1$ then u_1 and u_2 are not pseudonymous.

Proejctive planes are always bad

- Suppose that every pair of users share a message space, and that users always send messages via shortest paths.
- Why? What are the pseudonymity classes with respect to user c?
- If $c, u_1 \in M_1$ and $u_2 \notin M_1$ then u_1 and u_2 are not pseudonymous.
- If message spaces have size k, pseudonymity classes have size at most $k - 1$.

Proejctive planes are always bad

- Suppose that every pair of users share a message space, and that users always send messages via shortest paths.
- Why? What are the pseudonymity classes with respect to user c?
- If $c, u_1 \in M_1$ and $u_2 \notin M_1$ then u_1 and u_2 are not pseudonymous.
- If message spaces have size k, pseudonymity classes have size at most $k - 1$.
- If c can also observe messages addressed to other users, all other users can be identified.

Formal(ish) Protocol

- Each user has a public key and a private key.

Formal(ish) Protocol

- Each user has a public key and a private key.
- When u wants to submit a query through a proxy v, she chooses a shortest path $[u, M_1, u_1, M_2, u_2, \ldots, M_t, u_t, M_{t+1}, v]$ to v, and a private key ψ.

Formal(ish) Protocol

- Each user has a public key and a private key.
- When u wants to submit a query through a proxy v, she chooses a shortest path $[u, M_1, u_1, M_2, u_2, \ldots, M_t, u_t, M_{t+1}, v]$ to v, and a private key ψ.
- u writes to M_1 the message

$$[(\phi_1(u_1, M_2, \phi_2(u_2, \ldots, M_n, \phi_v(v) \ldots))), \phi_v(Q), \phi_v(\psi)]$$

- In every step user u_i will decrypt the content in M_i with her private key, and writes the next message to M_{i+1}.

Formal(ish) Protocol

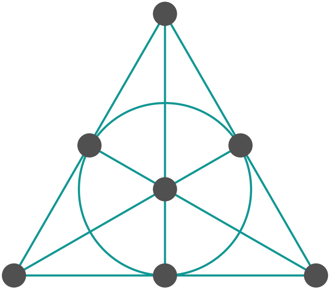
- Each user has a public key and a private key.
- When u wants to submit a query through a proxy v, she chooses a shortest path $[u, M_1, u_1, M_2, u_2, \ldots, M_t, u_t, M_{t+1}, v]$ to v, and a private key ψ.
- u writes to M_1 the message

$$[(\phi_1(u_1, M_2, \phi_2(u_2, \ldots, M_n, \phi_v(v) \ldots))), \phi_v(Q), \phi_v(\psi)]$$

- In every step user u_i will decrypt the content in M_i with her private key, and writes the next message to M_{i+1}.
- The proxy will evaluate the query, and encrypt the response R using u's private key ψ.
- Each user u_i seeing the response in M_{i+1} copies it to M_i.

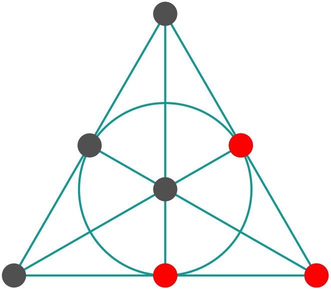
The encrypted projective plane is still bad

- Assume a UPIR scheme based on a projective plane



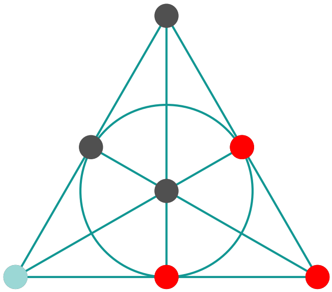
The encrypted projective plane is still bad

- Assume a UPIR scheme based on a projective plane and a coalition of three eavesdroppers in general position.



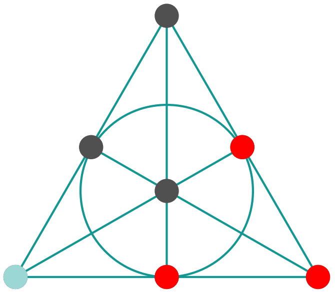
The encrypted projective plane is still bad

- Assume a UPIR scheme based on a projective plane and a coalition of three eavesdroppers in general position.
- Any user shares exactly one message space with any eavesdropper



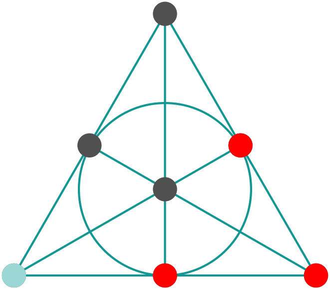
The encrypted projective plane is still bad

- Assume a UPIR scheme based on a projective plane and a coalition of three eavesdroppers in general position.
- Any user shares exactly one message space with any eavesdropper and at least two distinct message spaces with the coalition.



The encrypted projective plane is still bad

- Assume a UPIR scheme based on a projective plane and a coalition of three eavesdroppers in general position.
- Any user shares exactly one message space with any eavesdropper and at least two distinct message spaces with the coalition.
- As soon as the user chooses two eavesdroppers in different message spaces as a proxy, they can identify him as the single intersection of their message spaces.

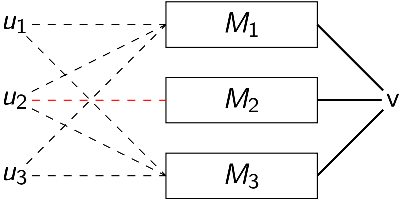


Information leaking

- Queries are indistinguishable for the users u_i on the path $[u, u_1, u_2, \ldots u_t, v]$.
- Only the proxy v learns the content of the query.
- Only v can identify linked queries. What can v learn about u?

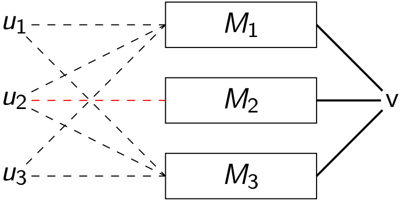
Information leaking

- Queries are indistinguishable for the users u_i on the path $[u, u_1, u_2, \ldots u_t, v]$.
- Only the proxy v learns the content of the query.
- Only v can identify linked queries. What can v learn about u?
- Only the set of message spaces containing v which lie on some geodesic $[u, v]$. So u_1 and u_3 are pseudonymous wrt v.



Information leaking

- Queries are indistinguishable for the users u_i on the path $[u, u_1, u_2, \ldots u_t, v]$.
- Only the proxy v learns the content of the query.
- Only v can identify linked queries. What can v learn about u?
- Only the set of message spaces containing v which lie on some geodesic $[u, v]$. So u_1 and u_3 are pseudonymous wrt v.



- **So we should build a protocol where all users at distance ≥ 2 from v write to every message space containing v.**

Generalised quadrangles

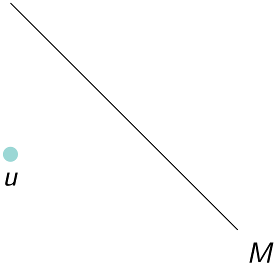
Generalized Quadrangles

A generalised quadrangle is a partial linear space in which lines have size $t+1$, and every point meets $s+1$ lines, and which satisfies the **GQ axiom**: For every point, line pair $[u, M]$ such that u is not contained in M, there exists a unique point u_1 in M which is incident with x.

Generalised quadrangles

Generalized Quadrangles

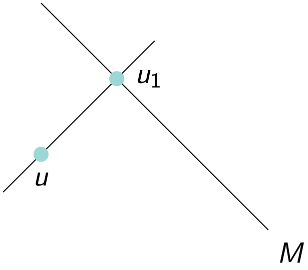
A generalised quadrangle is a partial linear space in which lines have size $t+1$, and every point meets $s+1$ lines, and which satisfies the **GQ axiom**: For every point, line pair $[u, M]$ such that u is not contained in M, there exists a unique point u_1 in M which is incident with x.



Generalised quadrangles

Generalized Quadrangles

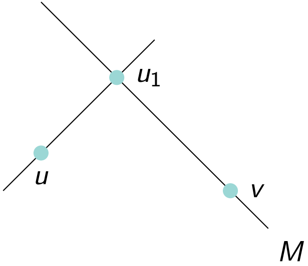
A generalised quadrangle is a partial linear space in which lines have size $t+1$, and every point meets $s+1$ lines, and which satisfies the **GQ axiom**: For every point, line pair $[u, M]$ such that u is not contained in M, there exists a unique point u_1 in M which is incident with x.



Generalised quadrangles

Generalized Quadrangles

A generalised quadrangle is a partial linear space in which lines have size $t+1$, and every point meets $s+1$ lines, and which satisfies the **GQ axiom**: For every point, line pair $[u, M]$ such that u is not contained in M, there exists a unique point u_1 in M which is incident with x.

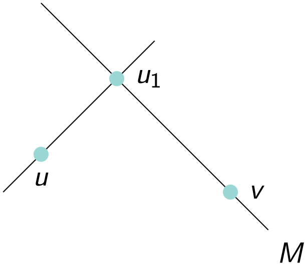


- Let u and v be users sharing no message space. Let M be a message space containing v.

Generalised quadrangles

Generalized Quadrangles

A generalised quadrangle is a partial linear space in which lines have size $t+1$, and every point meets $s+1$ lines, and which satisfies the **GQ axiom**: For every point, line pair $[u, M]$ such that u is not contained in M, there exists a unique point u_1 in M which is incident with x.



- Let u and v be users sharing no message space. Let M be a message space containing v.
- There exists a unique user $u_1 \in M$ and a unique message space which contains u and u_1.

Near example

- Let V be a four dimensional vector space over a field k.
- Define the *points* of $\mathcal{Q}$ to be 2-d subspaces of V.
- Say that two points are *collinear* if they intersect in a 1-d subspace.
- A *line* is a set of mutually collinear points, consisting of all points containing a fixed 1-d subspace.
- If $P = \langle e_1, e_2 \rangle$ and ℓ is the line defined by $\langle e_3 \rangle$ then there are multiple points on ℓ incidence with P, $\langle e_1, e_3 \rangle$ and $\langle e_2, e_3 \rangle$, for example. (This is not a GQ).
- In fact, one can obtain a generalised quadrangle by keeping only points and lines which are identically zero under a quadratic form.

What is a GQ anyway?

- The isotropic points and lines of a nondegenerate quadratic form of projective index 1.

What is a GQ anyway?

- The isotropic points and lines of a nondegenerate quadratic form of projective index 1.
- Let V be a four dimensional vector space, and consider the form $Q(v) = v_1 v_2 + v_3 v_4 = 0$ on V.
- Observe that $Q(\alpha v) = \alpha^2 Q(v)$, so the zero-set of Q is a union of lines through 0. Call these lines the **points** of our GQ.

What is a GQ anyway?

- The isotropic points and lines of a nondegenerate quadratic form of projective index 1.
- Let V be a four dimensional vector space, and consider the form $Q(v) = v_1 v_2 + v_3 v_4 = 0$ on V.
- Observe that $Q(\alpha v) = \alpha^2 Q(v)$, so the zero-set of Q is a union of lines through 0. Call these lines the **points** of our GQ.
- Observe that Q contains many two dimensional subspaces: e.g. the set of points of the form $[0, x, 0, y]$, call such a space a **line** of the GQ.

What is a GQ anyway?

- The isotropic points and lines of a nondegenerate quadratic form of projective index 1.
- Let V be a four dimensional vector space, and consider the form $Q(v) = v_1 v_2 + v_3 v_4 = 0$ on V.
- Observe that $Q(\alpha v) = \alpha^2 Q(v)$, so the zero-set of Q is a union of lines through 0. Call these lines the **points** of our GQ.
- Observe that Q contains many two dimensional subspaces: e.g. the set of points of the form $[0, x, 0, y]$, call such a space a **line** of the GQ.
- To check: over $\mathbb{F}_q$, every line contains $q + 1$ points, every point is contained in $q + 1$ lines. And the GQ-axiom.

Lemma

In an encrypted GQ-UPIR scheme, suppose u chooses v as a proxy with $d(u, v) = 2$, and chooses a geodesic to v uniformly at random. Then v is equally likely to observe the request in any message space to which she has access.

Proof.

By hypothesis, u and v do not share a line. Let M be a line through u: then there exists a unique line through v meeting M by the GQ-axiom. The number of lines through a point is $s + 1$, and a GQ contains no triangles. So every line through u meets a unique line through v. So if u chooses uniformly at random from the geodesics to v, then v is equally likely to observe the request in any message space to which he has access. □

Any two users at distance two from v are pseudonymous with respect to v.

The main result

- A generalised quadrangle has order (s, t), if $s + 1$ points are incident with a given line and $t + 1$ lines are incident with a given point.

The main result

- A generalised quadrangle has order (s, t), if $s + 1$ points are incident with a given line and $t + 1$ lines are incident with a given point.
- If the order of a GQ is (s, t) then it has $(s + 1)(st + 1)$ points, $s(t + 1)$ at distance 1 and s^2t at distance 2.
- Higman: $s < t^2$ and $t \leq s^2$.

The main result

- A generalised quadrangle has order (s, t), if $s + 1$ points are incident with a given line and $t + 1$ lines are incident with a given point.
- If the order of a GQ is (s, t) then it has $(s + 1)(st + 1)$ points, $s(t + 1)$ at distance 1 and s^2t at distance 2.
- Higman: $s < t^2$ and $t \leq s^2$.
- The neighbourhood of v contains $O(st)$ users, while the number of users at distance 2 is $O(st^2)$.
- Users at distance 2 from every member of a coalition remain mutually anonymous: if $|\mathcal{C}| = o(t)$, then 'most' users remain at distance 2.

The main result

- A generalised quadrangle has order (s, t), if $s + 1$ points are incident with a given line and $t + 1$ lines are incident with a given point.
- If the order of a GQ is (s, t) then it has $(s + 1)(st + 1)$ points, $s(t + 1)$ at distance 1 and s^2t at distance 2.
- Higman: $s < t^2$ and $t \leq s^2$.
- The neighbourhood of v contains $O(st)$ users, while the number of users at distance 2 is $O(st^2)$.
- Users at distance 2 from every member of a coalition remain mutually anonymous: if $|\mathcal{C}| = o(t)$, then 'most' users remain at distance 2.
- So the encrypted GQ-UPIR system is secure!

What about the unencrypted case?

- By observing queries, v learns the set of users mutually at distance 1 from u and v: $\mathcal{B}_1(u) \cap \mathcal{B}_1(v)$.
- The set of users pseudonymous with u is $\{u_i \mid \mathcal{B}_1(u_i) \cap \mathcal{B}_1(v) = \mathcal{B}_1(u) \cap \mathcal{B}_1(v)\}$.

What about the unencrypted case?

- By observing queries, v learns the set of users mutually at distance 1 from u and v: $\mathcal{B}_1(u) \cap \mathcal{B}_1(v)$.
- The set of users pseudonymous with u is $\{u_i \mid \mathcal{B}_1(u_i) \cap \mathcal{B}_1(v) = \mathcal{B}_1(u) \cap \mathcal{B}_1(v)\}$.
- This is the **definition** of the *hyperbolic line* through u and v!

What about the unencrypted case?

- By observing queries, v learns the set of users mutually at distance 1 from u and v: $\mathcal{B}_1(u) \cap \mathcal{B}_1(v)$.
- The set of users pseudonymous with u is $\{u_i \mid \mathcal{B}_1(u_i) \cap \mathcal{B}_1(v) = \mathcal{B}_1(u) \cap \mathcal{B}_1(v)\}$.
- This is the **definition** of the *hyperbolic line* through u and v!
- Three users suffice to identify all other users in any unencrypted GQ-UPIR scheme.

What about the unencrypted case?

- By observing queries, v learns the set of users mutually at distance 1 from u and v: $\mathcal{B}_1(u) \cap \mathcal{B}_1(v)$.
- The set of users pseudonymous with u is $\{u_i \mid \mathcal{B}_1(u_i) \cap \mathcal{B}_1(v) = \mathcal{B}_1(u) \cap \mathcal{B}_1(v)\}$.
- This is the **definition** of the *hyperbolic line* through u and v!
- Three users suffice to identify all other users in any unencrypted GQ-UPIR scheme.
- There are seven classical families of GQs, in two of these families hyperbolic lines have size 2: here a single user suffices.

Questions

- GQs are pretty special. What broader class of bipartite graphs give secure UPIR schemes? (Expanders? Graphs of large girth?)
- We know of no secure unencrypted systems. Is it even possible to construct one?
- Could a UPIR system be implemented in some sort of practical way?

References

- J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, and J. Manjon. *User-private information retrieval based on a peer-to-peer community*. Data Knowl. Eng., 68(11):1237–1252, Nov. 2009.
- K. Stokes and M. Bras-Amorós. *Optimal configurations for peer-to-peer user-private information retrieval.* Comput. Math. Appl., 59(4):1568–1577, 2010.
- C. M. Swanson and D. R. Stinson. *Extended combinatorial constructions for peer-to-peer user-private information retrieval.* Adv. Math. Commun., 6(4):479–497, 2012.
- C. M. Swanson and D. R. Stinson. *Extended results on privacy against coalitions of users in user-private information retrieval protocols.* Cryptogr. Commun., 7(4):415–437, 2015.
- Oliver W. Gnilke, Marcus Greferath, Camilla Hollanti, Guillermo Nunez Ponasso, Padraig Ó Catháin, Eric Swartz *Improved User-Private Information Retrieval via Finite Geometry*, arXiv 1707.01551.

Thank You!