Configurations and Peer-to-Peer Privacy

Maria Bras-Amorós,
in collaboration with Josep Domingo-Ferrer, Qianhong Wu, Jesús Manjón, Klara Stokes and Marcus Greferath

Universitat Rovira i Virgili, Tarragona, Catalonia

Interdisciplinary Workshop on Data Privacy
Maynooth, September 28, 2015

Contents

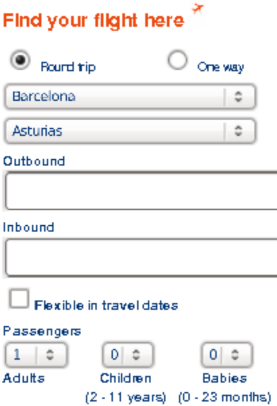
Privacy of queries in front of a data base

Privacy of queries in front of a data base

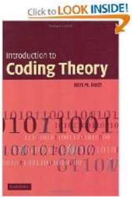
Privacy when performing queries to a data base has two faces:

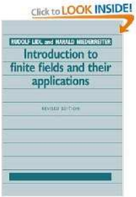
1. Privacy of the query itself
   $>$ Private Information Retrieval (PIR)
2. Privacy of the user identity or profile
   $>$ User Private Information Retrieval (UPIR)

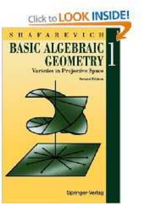
Private Information Retrieval (PIR)



User Private Information Retrieval (UPIR)

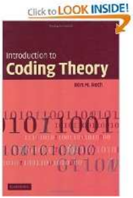


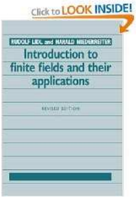


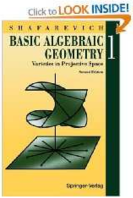


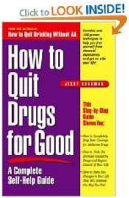
User Private Information Retrieval (UPIR)

amazon.com





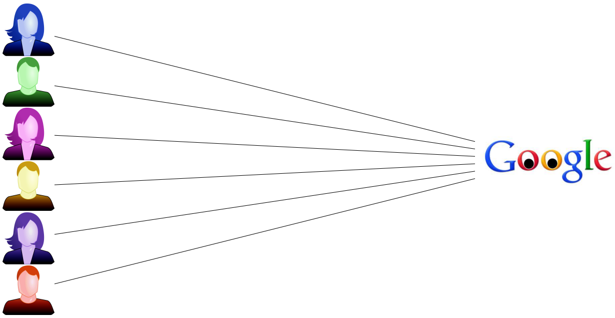




Peer-to-Peer User Private Information Retrieval (P2P UPIR)

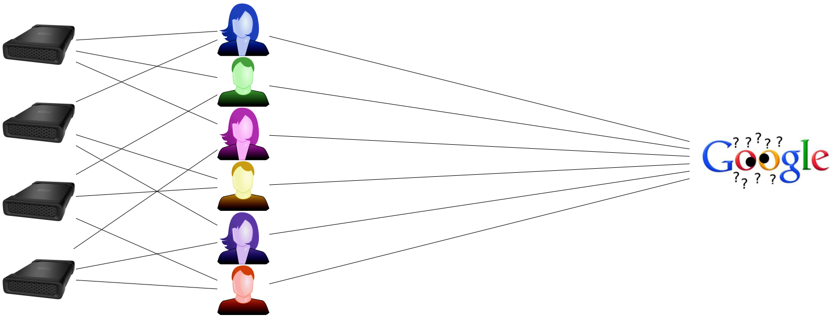
Peer-to-Peer User Private Information Retrieval (P2P UPIR)

J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, J. Manjón *User-Private Information Retrieval Based on a Peer-to-Peer Community*, 2009



Peer-to-Peer User Private Information Retrieval (P2P UPIR)

J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, J. Manjón *User-Private Information Retrieval Based on a Peer-to-Peer Community*, 2009



Peer-to-Peer User Private Information Retrieval (UPIR)

Peer-to-peer UPIR

- Each user shares one or more communication spaces (memory + cryptographic key) with other users.
- Users submit queries on behalf of other users.

Desirable properties on the distribution of users and communication spaces

- All users have the same number of communication spaces
- All communication spaces are shared by the same number of users
- Two different users share at most one communication space

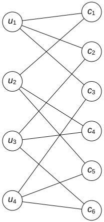
Combinatorial configurations

Combinatorial configurations

A (v, b, r, k)-configuration is a connected bipartite graph with

- v vertices on the left and b vertices on the right
- constant degree r for all vertices on the left and constant degree k for all vertices on the right
- no cycle of length 4.

Example of a $(4, 6, 3, 2)$-configuration



Combinatorial configurations

Projective planes

Given a finite field $\mathbb{F}_q$ consider all points

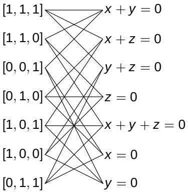
$$\{[a:b:c] : (a,b,c) \in \mathbb{F}_q^3 \setminus \{(0,0,0)\}\}/_{([a:b:c]=[a':b':c'] \Longleftrightarrow \frac{a}{a'}=\frac{b}{b'}=\frac{c}{c'})}$$

and all lines

$$\{Ax + By + Cz = 0 : A, B, C \in \mathbb{F}_q\}.$$

Each line contains $q+1$ points and each point is contained in $q+1$ lines, with no two lines meeting in more than one point.
This is a $(q^2+q+1, q^2+q+1, q+1, q+1)$-configuration.

Projective plane over $\mathbb{F}_2$



Problems related to combinatorial configurations with applications to P2P-UPIR

1. What are the optimal configurations for P2P-UPIR?

2. Although having such a simple definition, very little is known about the existence and construction of combinatorial configurations.
   What can we say?

3. What configurations prevent collusion attacks?
   What can we say about their existence and construction?

Problems related to combinatorial configurations with applications to P2P-UPIR

1. What are the optimal configurations for P2P-UPIR?

   K. Stokes, M. Bras-Amorós: Optimal Configurations for Peer-to-Peer User-Private Information Retrieval, Computers and Mathematics with Applications, Elsevier, vol. 59, n. 4, pp. 1568-1577, February 2010. ISSN: 0898-1221.

2. Although having such a simple definition, very little is known about the existence and construction of combinatorial configurations.
   What can we say?

   M. Bras-Amorós, K. Stokes: The Semigroup of Combinatorial Configurations, Semigroup Forum, Springer, vol. 84, n. 1, pp. 91-96, January 2012. ISSN: 0037-1912.

3. What configurations prevent collusion attacks?
   What can we say about their existence and construction?

   K. Stokes, M. Bras-Amorós: Associating a Numerical Semigroup to the Triangle-Free Configurations, Advances in Mathematics of Communication, American Institute of Mathematical Sciences, vol. 5, n. 2, pp. 351-371, May 2011. ISSN: 1930-5346

Optimal configurations for P2P-UPIR

Optimal configurations for P2P-UPIR

Suppose there are n_u users, n_c communication spaces, with each user having access to d_u communication spaces and each communication space assigned to d_c different users.

Privacy in front of the database

The query profile of u_i is diffused among the $d_u(d_c - 1)$ users sharing keys with u_i. Thus privacy in front of the database increases with $d_u(d_c - 1)$.

Lemma

$d_u(d_c - 1) \leqslant n_u - 1$

$>$ Optimal configurations satisfy $d_u(d_c - 1) = n_u - 1$

Optimal configurations for P2P-UPIR

Lemma

If $d_u(d_c - 1) = n_u - 1$ then $d_c \leqslant d_u$.

Optimal configurations should have minimal n_c and maximal n_u. Since $n_c/n_u = d_u/d_c$, we need to choose $d_c = d_u$ (and so $n_u = n_c$).

Theorem

Optimal configurations for Peer-to-Peer UPIR are the projective planes, that is, the symmetric configurations with $d_u = d_c = d$ and $n_u = n_c = n = d^2 - d + 1$.

We know that projective planes exist whenever $d - 1$ is a prime power. It is not known in general if there exists a projective plane for a general d.

Numerical semigroup of configurable tuples

Configurable tuples

Definition

We say that the tuple (v, b, r, k) is configurable if a (v, b, r, k) configuration exists.

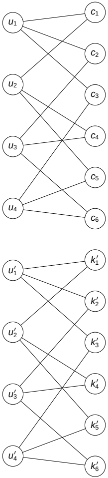
Lemma

If (v, b, r, k) is configurable then

$$vr = bk.$$

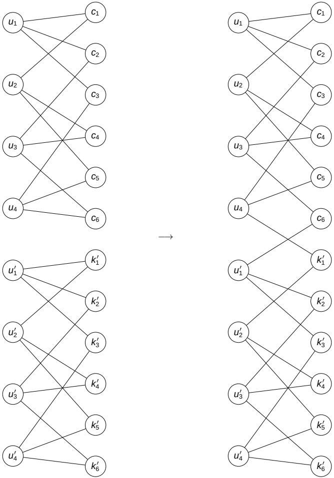
Merging configurations

Two configurations with the same constant degrees r and k can be merged as illustrated.



Merging configurations

Two configurations with the same constant degrees r and k can be merged as illustrated.



Merging configurations

Consequently,

$$\left.\begin{array}{ll} (v,b,r,k) & \text{configurable} \\ (v',b',r,k) & \text{configurable} \end{array}\right\} \Longrightarrow (v+v',b+b',r,k) \text{ configurable.}$$

Lemma

Fix r,k. If (v,b,r,k) is configurable then there exists $d \in \mathbb{Z}$ such that $(v,b,r,k) = (d\frac{k}{\gcd(r,k)}, d\frac{r}{\gcd(r,k)}, r, k)$.

Definition

$$D_{r,k} = \{d \in \mathbb{Z} : (d\frac{k}{\gcd(r,k)}, d\frac{r}{\gcd(r,k)}, r, k) \text{ is configurable}\}.$$

Lemma

- $0 \in D_{r,k}$,
- $d, d' \in D_{r,k} \Longrightarrow d + d' \in D_{r,k}$.

Numerical semigroups

Definition

A numerical semigroup is a set $S \subseteq \mathbb{N}_0$ such that

- $0 \in S$
- $s, s' \in S \implies s + s' \in S$
- $\#(\mathbb{N}_0 \setminus S) < \infty$.

Numerical semigroups

Example

The set S of amounts of money that can be obtained from an ideal cash point satisfies

- $0 \in S$
- $s, s' \in S \Longrightarrow s + s' \in S$

If we divide these amounts by 10 (or 100, or 1000) then

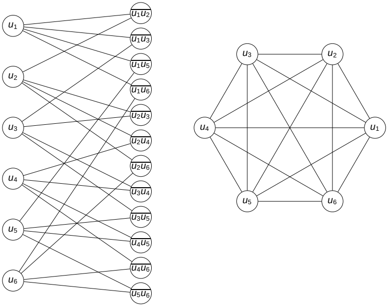
- $\#(\mathbb{N}_0 \setminus S) < \infty$.

$D_{r,2}$

There is a natural bijection

$(v, b, r, 2)$-configurations $\longleftrightarrow$ r-regular connected graphs with v vertices and b edges.

Two vertices in the graph share an edge if and only if the corresponding nodes in the configuration share a neighbor and viceversa.



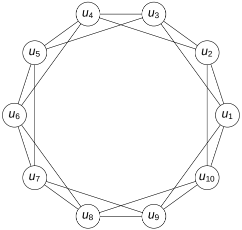
$D_{r,2}$

Lemma

Let r be an even positive integer. A connected r-regular graph with v vertices exists if and only if $v \geqslant r + 1$.

Proof.

By definition, any r-regular graph must have a number of vertices at least $r + 1$. Conversely, suppose $v \geqslant r + 1$. Consider a set of vertices $u_1, \ldots, u_v$. Put an edge between u_i and u_j, with $i \leqslant j$, if $j - i \leqslant r/2$ or $i + v - j \leqslant r/2$. This gives a connected r-regular graph with v vertices. □



$D_{r,2}$

Corollary

If r is an even *positive integer then*

$$D_{r,2} = < r+1, r+2, \ldots, 2r+1 > .$$

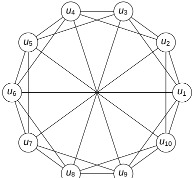
$D_{r,2}$

Lemma

Let r be an odd *positive integer. A connected r-regular graph with v vertices exists if and only if v is even and $v \geqslant r+1$.*

Proof.

By definition, any r-regular graph must have a number of vertices at least $r+1$. Now, since the number of edges is $rv/2$ this means that rv must be even and since r is odd v must be even.
Conversely, suppose v is even and $v \geqslant r+1$. Consider a set of vertices $u_1, \ldots, u_v$. Put an edge between u_i and u_j, with $i \leqslant j$, if $j-i \leqslant (r-1)/2$ or $i+v-j \leqslant (r-1)/2$. Put edges between u_i and $u_{i+v/2}$ for i from 1 to $v/2$. □



$D_{r,2}$

Corollary

If r is an odd *positive integer then*

$$D_{r,2} = < \frac{r+1}{2}, \frac{r+1}{2}+1, \frac{r+1}{2}+2, \ldots, r > .$$

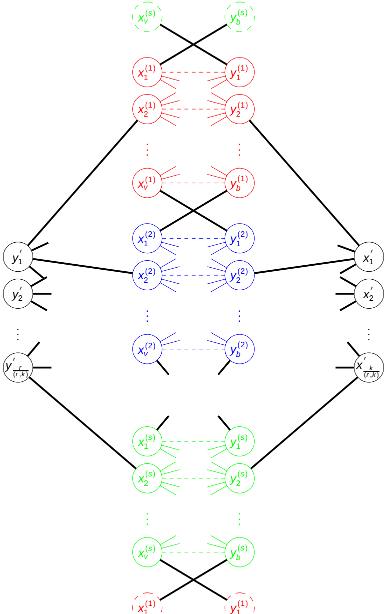
$D_{r,k}$ is always a numerical semigroup

Theorem

For any $r, k \in \mathbb{Z}$, with $r, k > 1$, $D_{r,k}$ is a numerical semigroup.

Steps of the proof:

1. There exists at least one non-zero integer $m \in D_{r,k}$ for all r, k.
2. Suppose we have a (v, b, r, k)-configuration with $r, k \geqslant 2$. There exist three edges in the configuration such that the six ends are all different.
3. Suppose $m \in D_{r,k}$. Merge $s = \frac{rk}{\gcd(r,k)}$ copies of an m-configuration with $\frac{k}{\gcd(r,k)}$ further vertices of degree r and $\frac{r}{\gcd(r,k)}$ vertices of degree k.


$x_v^{(s)}$
$y_b^{(s)}$
$x_1^{(1)}$
$y_1^{(1)}$
$x_2^{(1)}$
$y_2^{(1)}$
$x_v^{(1)}$
$y_b^{(1)}$
$x_1^{(2)}$
$y_1^{(2)}$
$x_2^{(2)}$
$y_2^{(2)}$
$x_v^{(2)}$
$y_b^{(2)}$
y'_1
y'_2
$y'_{\frac{r}{(r,k)}}$
x'_1
x'_2
$x'_{\frac{k}{(r,k)}}$
$x_1^{(s)}$
$y_1^{(s)}$
$x_2^{(s)}$
$y_2^{(s)}$
$x_v^{(s)}$
$y_b^{(s)}$
$x_1^{(1)}$
$y_1^{(1)}$

$D_{r,k}$ is always a numerical semigroup

4. We obtain a new configuration with parameters $(sv + k/\gcd(r,k), sb + r/\gcd(r,k), r, k) = (smk/\gcd(r,k) + k/\gcd(r,k), smr/\gcd(r,k) + r/\gcd(r,k), r, k) = ((sm+1)k/\gcd(r,k), (sm+1)r/\gcd(r,k), r, k)$ and so $sm + 1 \in D_{r,k}$.
5. m, $sm + 1$ coprime and m, $sm + 1 \in D_{r,k} \Rightarrow D_{r,k}$ is a numerical semigroup.

$D_{r,k}$ is always a numerical semigroup

Theorem

For any $r, k \in \mathbb{Z}$, with $r, k > 1$,

1. *There exist infinitely many configurable tuples (v, b, r, k);*
2. *There exists at least one configurable tuple (v, b, r, k) with*

$$v \leqslant 2(rk - r - k - 1)(2(rk - r - k) - 10)k$$
$$b \leqslant 2(rk - r - k - 1)(2(rk - r - k) - 10)r;$$

3. *All tuples (v, b, r, k) with $vr = bk$,*
   - $v \geqslant d_0 k / \gcd(r, k)$, *and*
   - $b \geqslant d_0 r / \gcd(r, k)$,

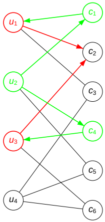
   are configurable for a certain d_0;
4. *If $r, k > 3$ then $d_0 \geqslant rk((4t^2 - 16t)^2 \gcd(r, k) - 4t^2 + 16t)$, where $t = rk - r - k - 1$.*

Collusion-free P2P-UPIR

Configurations for collusion-free P2P-UPIR

Collusion-attack

Two dishonest users connected to an honest user through two different communication spaces, can communicate themselves through a third communication space and infer some joint information.



This can be avoided by avoiding circuits of length 6 in the bipartite graph representing the combinatorial configuration.

The combinatorial configurations with girth larger than 6 are the so-called triangle-free configurations or $(0,1)$-geometries.

Numerical semigroup of triangle-free configurable tuples

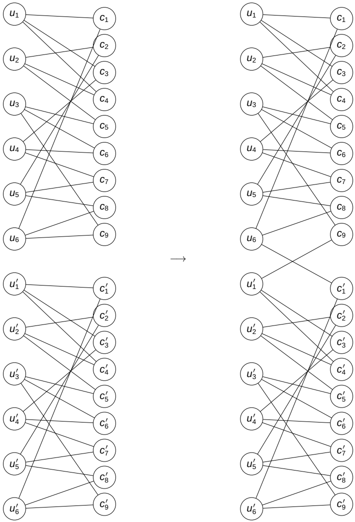
Configurations for collusion-free P2P-UPIR

We say that the tuple (v,b,r,k) is *triangle-free configurable* if a (v,b,r,k) triangle-free configuration exists.

Definition

$$D^{\triangle}_{r,k} = \{d \in \mathbb{Z} : (d\frac{k}{\gcd(r,k)}, d\frac{r}{\gcd(r,k)}, r, k) \text{ is triangle-free configurable}\}.$$

Merging triangle-free configurations



Merging triangle-free configurations

Consequently,

$$\left.\begin{array}{rl} (v,b,r,k) & \text{triangle-free configurable} \\ (v',b',r,k) & \text{triangle-free configurable} \end{array}\right\}$$

$$\Longrightarrow (v+v',b+b',r,k) \text{ triangle-free configurable.}$$

Lemma

- $0 \in D^{\triangle}_{r,k}$,
- $d, d' \in D^{\triangle}_{r,k} \Longrightarrow d + d' \in D^{\triangle}_{r,k}$.

$D^{\triangle}_{r,k}$ is always a numerical semigroup

Theorem

For any $r, k \in \mathbb{Z}$, with $r, k > 1$, $D^{\triangle}_{r,k}$ is a numerical semigroup.

Corollary

For any $r, k \in \mathbb{Z}$, with $r, k > 1$,

1. *There exist infinitely many triangle-free configurable tuples (v, b, r, k);*
2. *All tuples (v, b, r, k) with $vr = bk$,*
   - $v \geqslant d_0 k / \gcd(r, k)$, *and*
   - $b \geqslant d_0 r / \gcd(r, k)$,

   are configurable for a certain d_0;