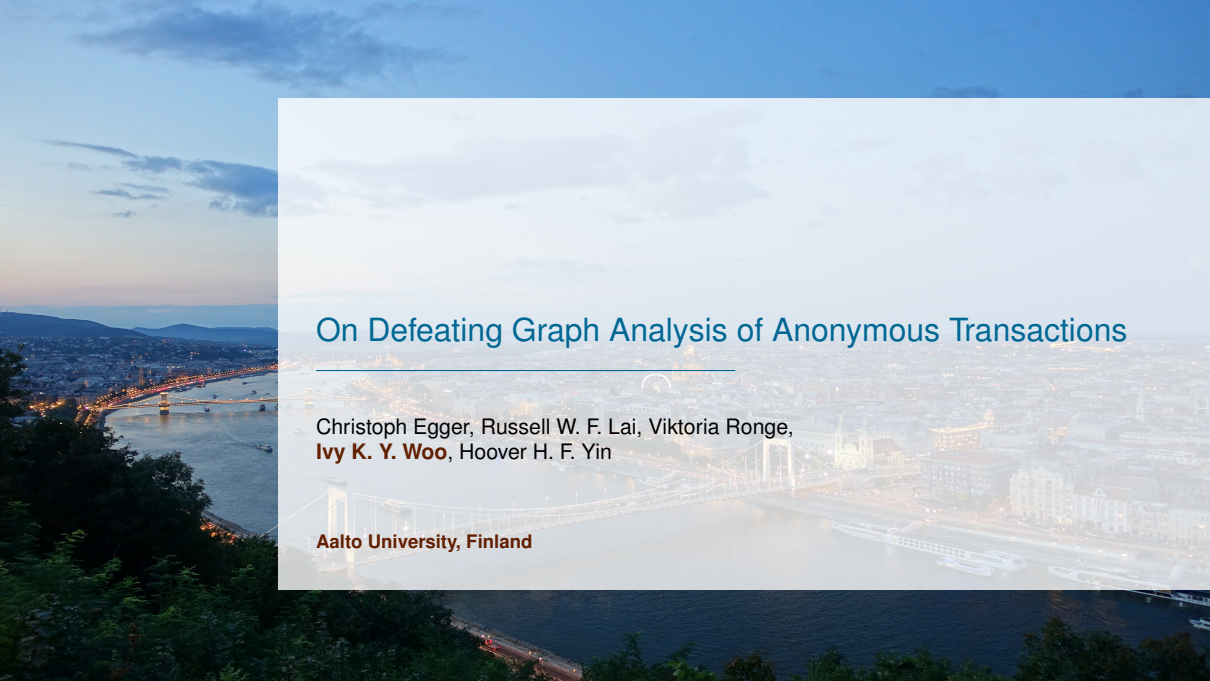


On Defeating Graph Analysis of Anonymous Transactions

Christoph Egger, Russell W. F. Lai, Viktoria Ronge,
Ivy K. Y. Woo, Hoover H. F. Yin

Aalto University, Finland

Content

Content

The Example of Anonymous Cryptocurrencies

Alice

Blockchain/Everyone

Choose decoys .

"I own one among (ring)

tx := which contains 1 coin, and

I'm transferring 1 coin to Bob's ."

$\pi \leftarrow$ Prove(tx)

$\text{tx}, \pi \longrightarrow$

Ring Samplers

- $U :=$ set of all users
- $\mathcal{R} \leftarrow \text{Samp}(i)$:
  PPT algorithm, inputs signer $i \in U$, outputs a ring $\mathcal{R}$ such that $i \in \mathcal{R} \subseteq U$
- Real signer i is hidden within set of users $\mathcal{R}$

Ring Samplers

- $U :=$ set of all users
- $\mathcal{R} \leftarrow \text{Samp}(i)$:
  PPT algorithm, inputs signer $i \in U$, outputs a ring $\mathcal{R}$ such that $i \in \mathcal{R} \subseteq U$
- Real signer i is hidden within set of users $\mathcal{R}$
- For efficiency, typically $|\mathcal{R}| \ll |U|$

Ring Samplers

- $U :=$ set of all users
- $\mathcal{R} \leftarrow \text{Samp}(i)$:
  PPT algorithm, inputs signer $i \in U$, outputs a ring $\mathcal{R}$ such that $i \in \mathcal{R} \subseteq U$
- Real signer i is hidden within set of users $\mathcal{R}$
- For efficiency, typically $|\mathcal{R}| \ll |U|$

Dilemma

Anonymity increases with ring size, while efficiency decreases with ring size.
How to pick a middle ground?

Ring Samplers

- $U :=$ set of all users
- $\mathcal{R} \leftarrow \text{Samp}(i)$:
  PPT algorithm, inputs signer $i \in U$, outputs a ring $\mathcal{R}$ such that $i \in \mathcal{R} \subseteq U$
- Real signer i is hidden within set of users $\mathcal{R}$
- For efficiency, typically $|\mathcal{R}| \ll |U|$

Dilemma

Anonymity increases with ring size, while efficiency decreases with ring size.
How to pick a middle ground?

- But wait, how to quantify “anonymity” in the first place?

Ring Samplers

- $U :=$ set of all users
- $\mathcal{R} \leftarrow \text{Samp}(i)$:
  PPT algorithm, inputs signer $i \in U$, outputs a ring $\mathcal{R}$ such that $i \in \mathcal{R} \subseteq U$
- Real signer i is hidden within set of users $\mathcal{R}$
- For efficiency, typically $|\mathcal{R}| \ll |U|$

Dilemma

Anonymity increases with ring size, while efficiency decreases with ring size.
How to pick a middle ground?

- But wait, how to quantify "anonymity" in the first place?
  (Not simply $1/|\mathcal{R}|$, since users could have different probability to be the real signer.)

Entropy-based Anonymity [Ronge et al. (2021)]

- Let $\mathcal{S}$ be signer distribution, $\mathcal{R} := \text{Samp}(\mathcal{S})$ be ring sampled.
  Anonymity of a ring sampler measured by conditional min-entropy:

$$H_\infty(\mathcal{S}|\mathcal{R}) = -\lg(\text{Guess}(\mathcal{S}|\mathcal{R})),$$

  $\text{Guess}(\mathcal{S}|\mathcal{R}) :=$ guessing probability of $\mathcal{S}$ conditioned on $\mathcal{R}$, i.e. upper bound on probability that adversary can guess $\mathcal{S}$ correctly given a sample of $\mathcal{R}$

Entropy-based Anonymity [Ronge et al. (2021)]

- Let $\mathcal{S}$ be signer distribution, $\mathcal{R} := \text{Samp}(\mathcal{S})$ be ring sampled.
  Anonymity of a ring sampler measured by conditional min-entropy:

$$H_\infty(\mathcal{S}|\mathcal{R}) = -\lg(\text{Guess}(\mathcal{S}|\mathcal{R})),$$

  $\text{Guess}(\mathcal{S}|\mathcal{R}) :=$ guessing probability of $\mathcal{S}$ conditioned on $\mathcal{R}$, i.e. upper bound on probability that adversary can guess $\mathcal{S}$ correctly given a sample of $\mathcal{R}$

- Worst-case measure of amount of information (in bits) in $\mathcal{S}$ conditioned on $\mathcal{R}$
- Higher $H_\infty(\mathcal{S}|\mathcal{R}) \Leftrightarrow$ More difficult to guess the real signer given the sampled ring

Entropy-based Anonymity [Ronge et al. (2021)]

- Let $\mathcal{S}$ be signer distribution, $\mathcal{R} := \text{Samp}(\mathcal{S})$ be ring sampled.
  Anonymity of a ring sampler measured by conditional min-entropy:

$$H_\infty(\mathcal{S}|\mathcal{R}) = -\lg(\text{Guess}(\mathcal{S}|\mathcal{R})),$$

  $\text{Guess}(\mathcal{S}|\mathcal{R}) :=$ guessing probability of $\mathcal{S}$ conditioned on $\mathcal{R}$, i.e. upper bound on probability that adversary can guess $\mathcal{S}$ correctly given a sample of $\mathcal{R}$

- Worst-case measure of amount of information (in bits) in $\mathcal{S}$ conditioned on $\mathcal{R}$
- Higher $H_\infty(\mathcal{S}|\mathcal{R}) \Leftrightarrow$ More difficult to guess the real signer given the sampled ring
- Can be viewed as “local” anonymity of ring samplers
  - Does not take into account how other users sample their rings

Some Classes of Ring Samplers

1. Uniform Samplers
   - Sample k decoys uniformly randomly from set of all users
   - Do not take signer distribution $\mathcal{S}$ into account
   - Poor anonymity

Some Classes of Ring Samplers

1. Uniform Samplers
   - Sample k decoys uniformly randomly from set of all users
   - Do not take signer distribution $\mathcal{S}$ into account
   - Poor anonymity
2. Mimicking Samplers
   - Sample k decoys according to signer distribution $\mathcal{S}$
   - Near-optimal anonymity
   - Problem: requires to know signer distribution, which is hard in reality

Some Classes of Ring Samplers

3. Partitioning Samplers
   - Partition the set of all users into chunks, sample k decoys uniformly randomly from the chunk that the signer belongs to
   - Near-optimal anonymity, assuming users in the same chunk have similar prob. to sign

Some Classes of Ring Samplers

3. Partitioning Samplers
   - Partition the set of all users into chunks, sample k decoys uniformly randomly from the chunk that the signer belongs to
   - Near-optimal anonymity, assuming users in the same chunk have similar prob. to sign

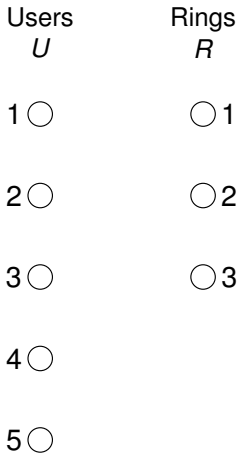
Question

How about “global” anonymity?
How well can partitioning samplers resist “global” attacks (for certain ring size), e.g. attacks based on graph-analysing ring memberships of all transactions?

Content

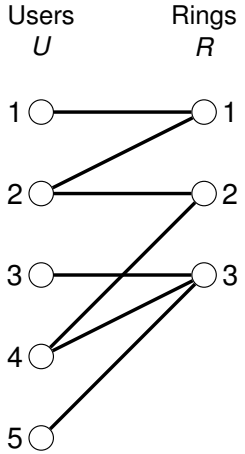
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings



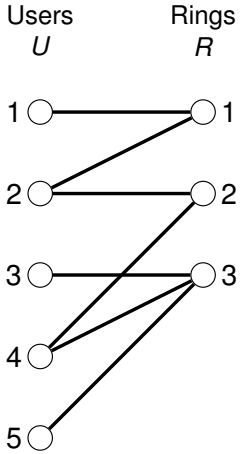
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j



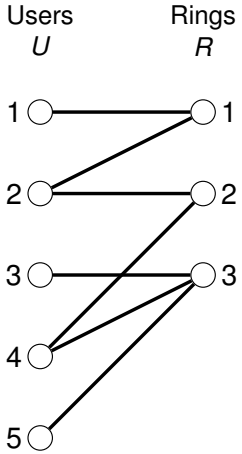
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j
- Exists at least 1 maximum matching
- Wlog, assume $(i, i) \in E$ for all $i \in |R|$



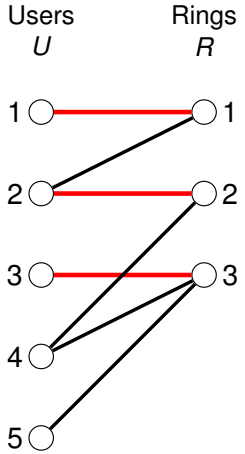
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j
- Exists at least 1 maximum matching
- Wlog, assume $(i, i) \in E$ for all $i \in |R|$
- Possible signer-signature assignments = Maximum matchings of G



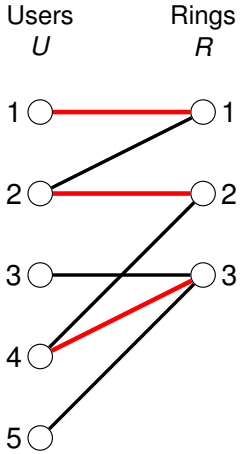
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j
- Exists at least 1 maximum matching
- Wlog, assume $(i, i) \in E$ for all $i \in |R|$
- Possible signer-signature assignments = Maximum matchings of G



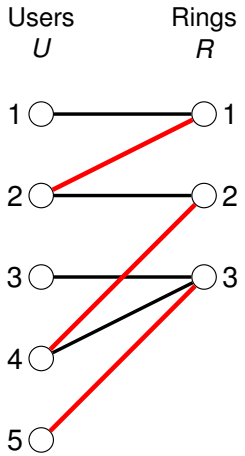
Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j
- Exists at least 1 maximum matching
- Wlog, assume $(i, i) \in E$ for all $i \in |R|$
- Possible signer-signature assignments = Maximum matchings of G



Transaction Graph

- Bipartite graph $G = (U, R, E)$
- Nodes U: set of users
- Nodes $R \subseteq U$: set of rings
- Edges E: ring membership, $(i, j) =$ user i in ring j
- Exists at least 1 maximum matching
- Wlog, assume $(i, i) \in E$ for all $i \in |R|$
- Possible signer-signature assignments = Maximum matchings of G



Modelling Attack by Security Game

$$\boxed{\begin{array}{l}\mathsf{Exp}_{\mathcal{A},\mathsf{Samp}}(U,|R|)\\ \hline (G=(U,R,E),M)\leftarrow \mathcal{G}^{\mathsf{Samp}}(U,|R|)\\ (u^*,r^*)\leftarrow \mathcal{A}(G)\\ \textbf{return } ((u^*,r^*)\in M)\end{array}}$$

- $U =$ set of users, $R =$ set of rings, $G =$ transaction graph,
  $M =$ maximum matching in G representing the true signer-signature assignment
- $\mathcal{G}^{\mathsf{Samp}}$: takes ring sampler Samp as oracle and samples (G, M)
- Given G, adversary $\mathcal{A}$ wins if it outputs an edge in M

Modelling Attack by Security Game

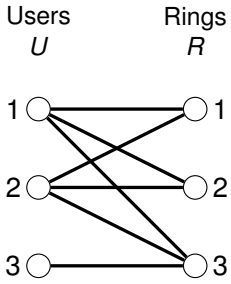
$$\boxed{\begin{array}{l}\mathsf{Exp}_{\mathcal{A},\mathsf{Samp}}(U,|R|)\\ \hline (G=(U,R,E),M)\leftarrow \mathcal{G}^{\mathsf{Samp}}(U,|R|)\\ (u^*,r^*)\leftarrow \mathcal{A}(G)\\ \textbf{return } ((u^*,r^*)\in M)\end{array}}$$

- $U =$ set of users, $R =$ set of rings, $G =$ transaction graph,
  $M =$ maximum matching in G representing the true signer-signature assignment
- $\mathcal{G}^{\mathsf{Samp}}$: takes ring sampler Samp as oracle and samples (G, M)
- Given G, adversary $\mathcal{A}$ wins if it outputs an edge in M
- Lower $\Pr\left[\mathsf{Exp}_{\mathcal{A},\mathsf{Samp}}(U,|R|)\right] \Leftrightarrow$ Higher anonymity under graph-based attacks

What can Adversary do with G?

Given transaction graph $G = (U, R, E)$,

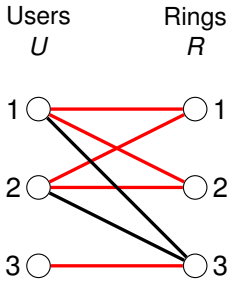
- Dulmage-Mandelsohn(DM) decomposition
  - Core $(G) = (U, R, E')$, where $E' \subseteq E$ is union of all maximum matchings



What can Adversary do with G?

Given transaction graph $G = (U, R, E)$,

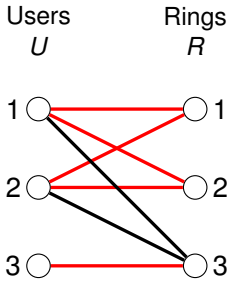
- Dulmage-Mandelsohn(DM) decomposition
  - Core $(G) = (U, R, E')$, where $E' \subseteq E$ is union of all maximum matchings



What can Adversary do with G?

Given transaction graph $G = (U, R, E)$,

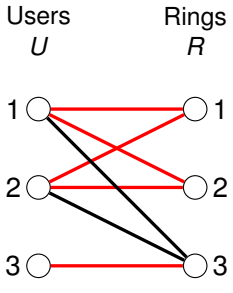
- Dulmage-Mandelsohn(DM) decomposition
  - Core $(G) = (U, R, E')$, where $E' \subseteq E$ is union of all maximum matchings
  - Runs in linear time of numbers of nodes and edges, given one maximum matching is known



What can Adversary do with G?

Given transaction graph $G = (U, R, E)$,

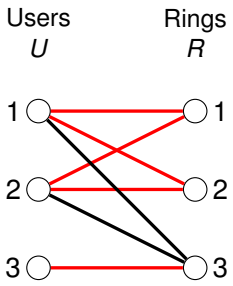
- Dulmage-Mandelsohn(DM) decomposition
  - Core $(G) = (U, R, E')$, where $E' \subseteq E$ is union of all maximum matchings
  - Runs in linear time of numbers of nodes and edges, given one maximum matching is known
- Deanonymisation attack
  - Rules out edges not in Core (G)



What can Adversary do with G?

Given transaction graph $G = (U, R, E)$,

- Dulmage-Mandelsohn(DM) decomposition
  - Core $(G) = (U, R, E')$, where $E' \subseteq E$ is union of all maximum matchings
  - Runs in linear time of numbers of nodes and edges, given one maximum matching is known
- Deanonymisation attack
  - Rules out edges not in Core (G)
  - $G \neq$ Core $(G) \Rightarrow$ Decreased anonymity



Upper-bounding $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$

- A graph-analysing adversary can exclude edges in G not in $\mathrm{Core}(G)$

Question 1

Relation between $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}[G \neq \mathrm{Core}(G)]$ and $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$?

Upper-bounding $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$

- A graph-analysing adversary can exclude edges in G not in $\mathrm{Core}\,(G)$

Question 1

Relation between $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right]$ and $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$?

- Trivial attack: to deanonymise a signer by random guessing, winning prob. $= \frac{1}{k+1}$

Upper-bounding $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$

- A graph-analysing adversary can exclude edges in G not in $\mathrm{Core}\,(G)$

Question 1

Relation between $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right]$ and $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$?

- Trivial attack: to deanonymise a signer by random guessing, winning prob. $= \frac{1}{k+1}$
- We proved this is optimal attack when $G = \mathrm{Core}\,(G)$ and with partitioning samplers:

$$\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right] \leq \Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right] + \frac{1}{k+1}$$

Upper-bounding $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$

- A graph-analysing adversary can exclude edges in G not in $\mathrm{Core}\,(G)$

Question 1

Relation between $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G\neq\mathrm{Core}\,(G)\right]$ and $\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]$?

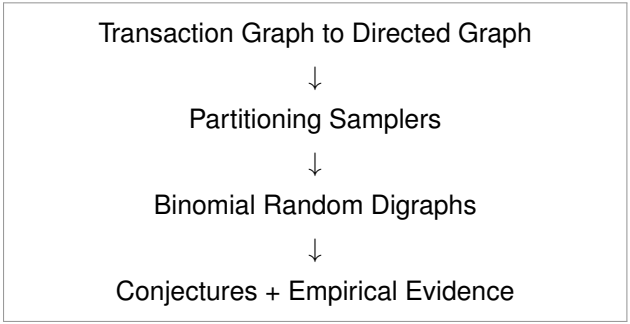
- Trivial attack: to deanonymise a signer by random guessing, winning prob. $=\frac{1}{k+1}$
- We proved this is optimal attack when $G=\mathrm{Core}\,(G)$ and with partitioning samplers:

$$\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right]\leq\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G\neq\mathrm{Core}\,(G)\right]+\frac{1}{k+1}$$

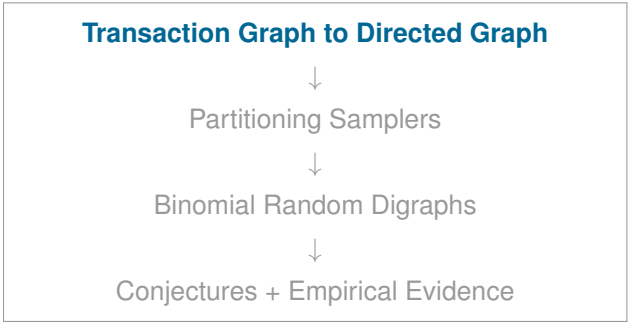
Question 2

How to upper-bound $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G\neq\mathrm{Core}\,(G)\right]$ for $\mathrm{Samp}=$ partitioning sampler?

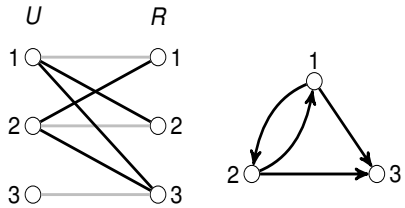
Roadmap


Transaction Graph to Directed Graph
Partitioning Samplers
Binomial Random Digraphs
Conjectures + Empirical Evidence

Roadmap


Transaction Graph to Directed Graph
Partitioning Samplers
Binomial Random Digraphs
Conjectures + Empirical Evidence

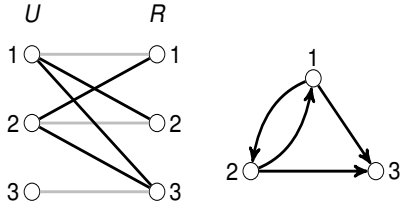
Induced Directed Graph



- Transaction graph $G = (U, R, E)$
- Induced digraph of G:
  $\mathrm{id}(G) = (U, F)$ where $F = \{(i, j) : (i, j) \in E, i \neq j\}$
- Biadjacency matrix of $G \approx$ Adjacency matrix of $\mathrm{id}(G)$

Finding Core (G)

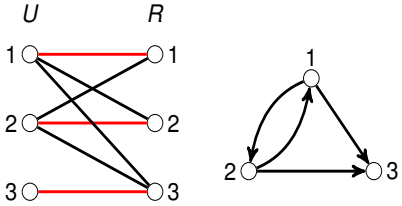
Case 1: $|U| = |R|$ (balanced)



Finding Core (G)

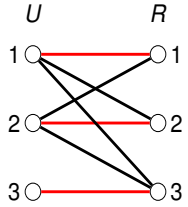
Case 1: $|U| = |R|$ (balanced)

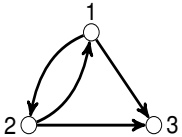
- (i, i) is an edge of Core (G)



Finding Core (G)

Case 1: $|U| = |R|$ (balanced)



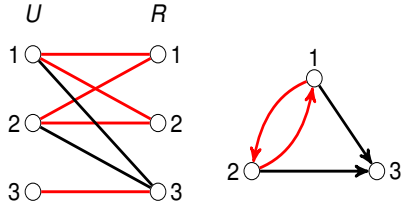


- (i, i) is an edge of Core (G)
- Tassa (2014): For all $i \neq j$,

$$(i, j) \text{ is an edge of a strongly connected component in id } (G)$$
$$\Leftrightarrow (i, j) \text{ is an edge of Core } (G)$$

Finding Core (G)

Case 1: $|U| = |R|$ (balanced)

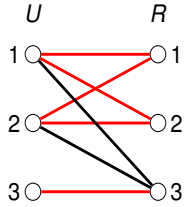


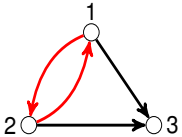
- (i, i) is an edge of Core (G)
- Tassa (2014): For all $i \neq j$,

$$(i, j) \text{ is an edge of a strongly connected component in id } (G)$$
$$\Leftrightarrow (i, j) \text{ is an edge of Core } (G)$$

Finding Core (G)

Case 1: $|U| = |R|$ (balanced)





- (i, i) is an edge of Core (G)
- Tassa (2014): For all $i \neq j$,

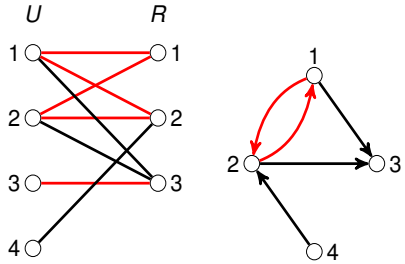
$$(i, j) \text{ is an edge of a strongly connected component in id } (G)$$
$$\Leftrightarrow (i, j) \text{ is an edge of Core } (G)$$

$\Rightarrow$ If id (G) is strongly connected (S.C.), then $G = \text{Core } (G)$

$\Rightarrow \Pr_{G \leftarrow \mathcal{G}} [G \neq \text{Core } (G)] \leq \Pr_{G \leftarrow \mathcal{G}} [\text{id } (G) \text{ not S.C.}]$

Finding Core (G)

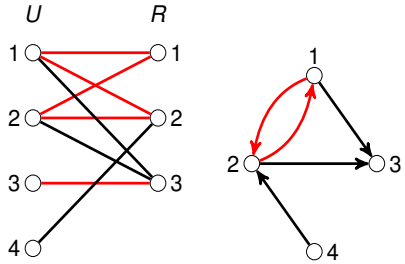
Case 2: $|U| > |R|$ (imbalanced)



Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

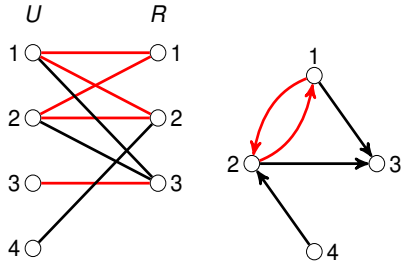
- Lower nodes = $\{i : i \in U, i > |R|\}$



Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

- Lower nodes = $\{i : i \in U, i > |R|\}$
- All edges found in case 1 ✓



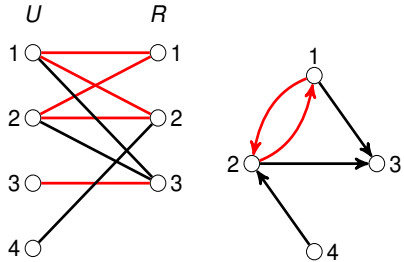
Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

- Lower nodes = $\{i : i \in U, i > |R|\}$
- All edges found in case 1 ✓
- Tassa (2014): Additionally,

(i, j) can be reached from a lower node in id (G)

$\Rightarrow (i, j)$ is an edge of Core (G)



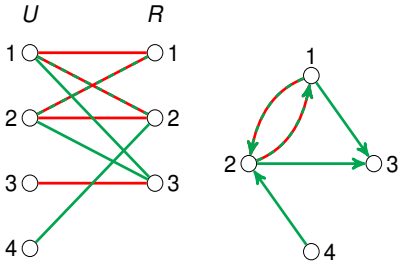
Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

- Lower nodes = $\{i : i \in U, i > |R|\}$
- All edges found in case 1 ✓
- Tassa (2014): Additionally,

(i, j) can be reached from a lower node in id (G)

$\Rightarrow (i, j)$ is an edge of Core (G)

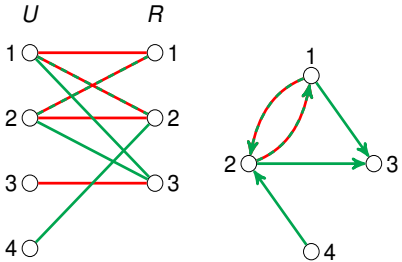


Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

- Lower nodes = $\{i : i \in U, i > |R|\}$
- All edges found in case 1 ✓
- Tassa (2014): Additionally,

$$(i, j) \text{ can be reached from a lower node in id } (G)$$
$$\Rightarrow (i, j) \text{ is an edge of Core } (G)$$



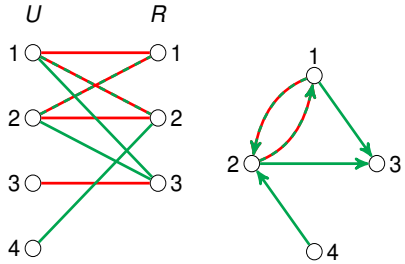
- Intuitively, the existence of lower nodes can significantly increases Pr $[G = \text{Core } (G)]$

Finding Core (G)

Case 2: $|U| > |R|$ (imbalanced)

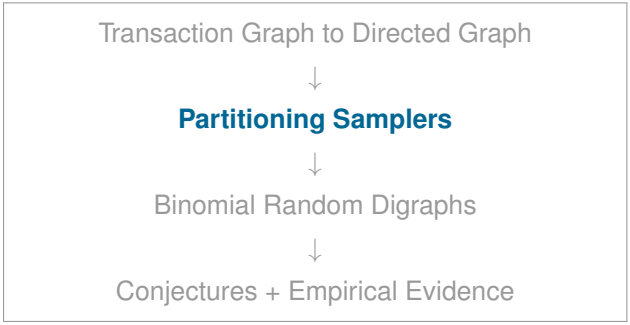
- Lower nodes = $\{i : i \in U, i > |R|\}$
- All edges found in case 1 ✓
- Tassa (2014): Additionally,

$$(i, j) \text{ can be reached from a lower node in id } (G)$$
$$\Rightarrow (i, j) \text{ is an edge of Core } (G)$$



- Intuitively, the existence of lower nodes can significantly increases $\Pr[G = \text{Core}(G)]$
- We proved that $\Pr[G \neq \text{Core}(G)]$ is greatest when $|R| = |U|$ (for fixed $|U|$)

Roadmap


Transaction Graph to Directed Graph
Partitioning Samplers
Binomial Random Digraphs
Conjectures + Empirical Evidence

Induced Digraphs of Partitioning Samplers

- Recall: sample k decoys uniformly randomly within a chunk
- $G :=$ Transaction graph of a chunk, number of users $= n$
  - Each of the n nodes of $\mathrm{id}(G)$ have in-degree k
  - k incoming nodes are sampled uniformly randomly from the other $n - 1$ nodes

Induced Digraphs of Partitioning Samplers

- Recall: sample k decoys uniformly randomly within a chunk
- $G :=$ Transaction graph of a chunk, number of users $= n$
  - Each of the n nodes of $\mathrm{id}(G)$ have in-degree k
  - k incoming nodes are sampled uniformly randomly from the other $n - 1$ nodes
  - k-in-degree regular random digraphs with n nodes $=: \mathcal{G}^{\mathtt{reg}}$

Induced Digraphs of Partitioning Samplers

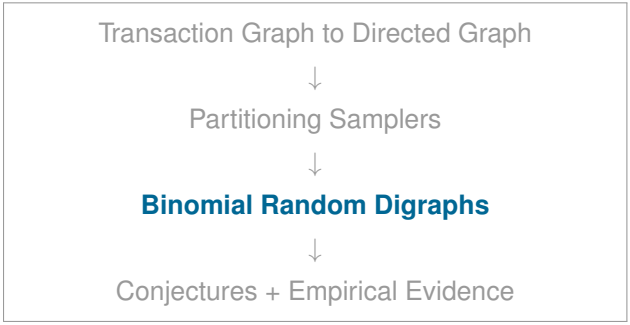
- Recall: sample k decoys uniformly randomly within a chunk
- $G :=$ Transaction graph of a chunk, number of users $= n$
  - Each of the n nodes of $\mathrm{id}(G)$ have in-degree k
  - k incoming nodes are sampled uniformly randomly from the other $n - 1$ nodes
  - k-in-degree regular random digraphs with n nodes $=: \mathcal{G}^{\mathtt{reg}}$
- Probability being strongly connected?

Induced Digraphs of Partitioning Samplers

- Recall: sample k decoys uniformly randomly within a chunk
- $G :=$ Transaction graph of a chunk, number of users $= n$
  - Each of the n nodes of $\mathrm{id}(G)$ have in-degree k
  - k incoming nodes are sampled uniformly randomly from the other $n-1$ nodes
  - k-in-degree regular random digraphs with n nodes $=: \mathcal{G}^{\mathtt{reg}}$
- Probability being strongly connected?

  $\rightarrow$ The Scottish Book (problem 38): Open problem for over 40 years...?

Roadmap


Transaction Graph to Directed Graph
Partitioning Samplers
Binomial Random Digraphs
Conjectures + Empirical Evidence

Binomial Random Digraphs $=: \mathcal{G}^{\mathtt{bin}}$

- Random digraph with n nodes, each directed edge exists with prob. p independently
- For each node, in-degree follows $Bin(n-1, p)$, with mean $(n-1)p$

Binomial Random Digraphs $=: \mathcal{G}^{\mathtt{bin}}$

- Random digraph with n nodes, each directed edge exists with prob. p independently
- For each node, in-degree follows $Bin(n-1, p)$, with mean $(n-1)p$
- Induced digraph by the following ring sampler:
  - n users, each of the other $n-1$ users chosen as decoy with prob. p independently

Binomial Random Digraphs $=: \mathcal{G}^{\mathtt{bin}}$

- Random digraph with n nodes, each directed edge exists with prob. p independently
- For each node, in-degree follows $Bin(n-1, p)$, with mean $(n-1)p$
- Induced digraph by the following ring sampler:
  - n users, each of the other $n-1$ users chosen as decoy with prob. p independently
- Probability being strongly connected?

Binomial Random Digraphs $=: \mathcal{G}^{\mathtt{bin}}$

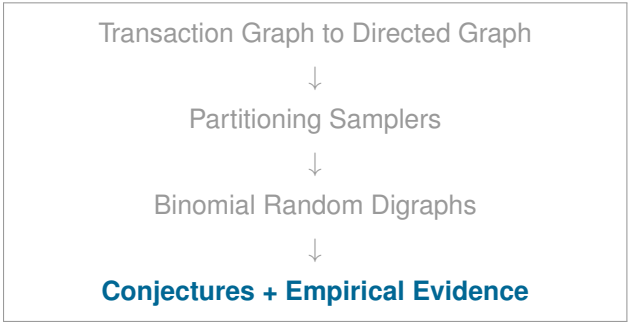
- Random digraph with n nodes, each directed edge exists with prob. p independently
- For each node, in-degree follows $Bin(n-1, p)$, with mean $(n-1)p$
- Induced digraph by the following ring sampler:
  - n users, each of the other $n-1$ users chosen as decoy with prob. p independently
- Probability being strongly connected?

  Graham and Pike (2008):

  If $p = p(n) = \frac{\ln n + c}{n}$ for some constant $c \in \mathbb{R}$,

$$\lim_{n \to \infty} \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] = 1 - e^{-2e^{-c}}$$

Roadmap


Transaction Graph to Directed Graph
Partitioning Samplers
Binomial Random Digraphs
Conjectures + Empirical Evidence

Conjecture 1 – Relating $\mathcal{G}^{\mathtt{reg}}$ and $\mathcal{G}^{\mathtt{bin}}$

$\mathcal{G}^{\mathtt{reg}}$: Fix in-degree $= k$ for each node

$\mathcal{G}^{\mathtt{bin}}$: Let $p = \frac{k}{n-1}$, i.e. fix *expected* in-degree $= k$ for each node

Conjecture 1 – Relating $\mathcal{G}^{\mathtt{reg}}$ and $\mathcal{G}^{\mathtt{bin}}$

$\mathcal{G}^{\mathtt{reg}}$: Fix in-degree $= k$ for each node

$\mathcal{G}^{\mathtt{bin}}$: Let $p = \frac{k}{n-1}$, i.e. fix *expected* in-degree $= k$ for each node

- $\Pr_{G \leftarrow \mathcal{G}^{\mathtt{reg}}}[G \text{ not S.C.}] \leq \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}}[G \text{ not S.C.}]$

Conjecture 1 – Relating $\mathcal{G}^{\mathtt{reg}}$ and $\mathcal{G}^{\mathtt{bin}}$

$\mathcal{G}^{\mathtt{reg}}$: Fix in-degree $= k$ for each node

$\mathcal{G}^{\mathtt{bin}}$: Let $p = \frac{k}{n-1}$, i.e. fix *expected* in-degree $= k$ for each node

- $\Pr_{G \leftarrow \mathcal{G}^{\mathtt{reg}}}[G \text{ not S.C.}] \leq \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}}[G \text{ not S.C.}]$
- Intuitively true:
  - For graphs from $\mathcal{G}^{\mathtt{reg}}$, each node must be weakly connected to k other nodes
  - For graphs from $\mathcal{G}^{\mathtt{bin}}$, nodes may be weakly connected to fewer nodes

Conjecture 1 – Relating $\mathcal{G}^{\mathtt{reg}}$ and $\mathcal{G}^{\mathtt{bin}}$

$\mathcal{G}^{\mathtt{reg}}$: Fix in-degree $= k$ for each node

$\mathcal{G}^{\mathtt{bin}}$: Let $p = \frac{k}{n-1}$, i.e. fix *expected* in-degree $= k$ for each node

- $\Pr_{G \leftarrow \mathcal{G}^{\mathtt{reg}}}[G \text{ not S.C.}] \leq \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}}[G \text{ not S.C.}]$
- Intuitively true:
  - For graphs from $\mathcal{G}^{\mathtt{reg}}$, each node must be weakly connected to k other nodes
  - For graphs from $\mathcal{G}^{\mathtt{bin}}$, nodes may be weakly connected to fewer nodes
- Holds for all $n \geq 16$ for all k's tested

Conjecture 2 – Upper Bound for Sampling from $\mathcal{G}^{\mathtt{bin}}$

- Recap: If $p = p(n) = \frac{\ln n + c}{n}$ for some $c \in \mathbb{R}$,

$$\lim_{n \to \infty} \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] = 1 - e^{-2e^{-c}}$$

Conjecture 2 – Upper Bound for Sampling from $\mathcal{G}^{\mathtt{bin}}$

- Recap: If $p = p(n) = \frac{\ln n + c}{n}$ for some $c \in \mathbb{R}$,

$$\lim_{n \to \infty} \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] = 1 - e^{-2e^{-c}}$$

- Write $c = pn - \ln n$ and $p = \frac{k}{n-1}$

Conjecture 2 – Upper Bound for Sampling from $\mathcal{G}^{\mathtt{bin}}$

- Recap: If $p = p(n) = \frac{\ln n + c}{n}$ for some $c \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] = 1 - e^{-2e^{-c}}$$

- Write $c = pn - \ln n$ and $p = \frac{k}{n-1}$
- Conjectured closed-form upper bound:

$$\Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] \leq 1 - e^{-2e^{\ln n - \frac{k}{n-1} n}}$$

Conjecture 2 – Upper Bound for Sampling from $\mathcal{G}^{\mathtt{bin}}$

- Recap: If $p = p(n) = \frac{\ln n + c}{n}$ for some $c \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] = 1 - e^{-2e^{-c}}$$

- Write $c = pn - \ln n$ and $p = \frac{k}{n-1}$
- Conjectured closed-form upper bound:

$$\Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}] \leq 1 - e^{-2e^{\ln n - \frac{k}{n-1} n}}$$

- Holds for all $n \geq 16$ for all k's tested

Empirical Evidence

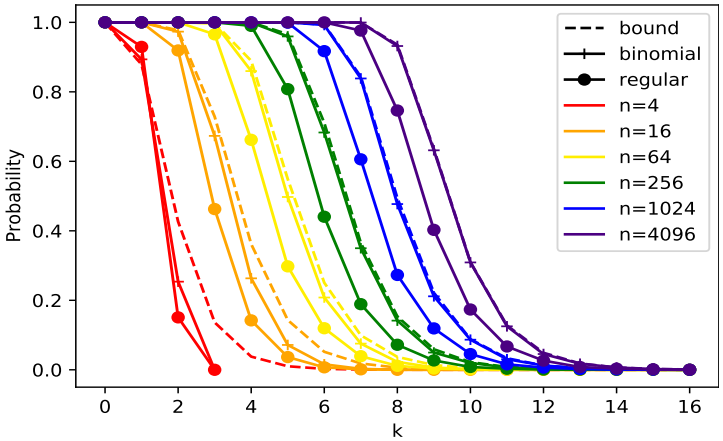


Figure: Pr [G not S.C.] against k under various n.

Chaining Things Together

Consider transaction graph G of a chunk. Let n be chunk size, k be number of decoys.

$$
\begin{aligned}
& \Pr_{G \leftarrow \mathcal{G}^{\mathtt{Samp}}} [G \neq \mathrm{Core}(G)] \quad \text{for any } |R| \leq n \\
\leq & \Pr_{G \leftarrow \mathcal{G}^{\mathtt{Samp}}} [G \neq \mathrm{Core}(G)] \quad \text{for } |R| = n \\
\leq & \Pr_{G \leftarrow \mathcal{G}^{\mathtt{Samp}}} [\mathrm{id}(G) \text{ not S.C.}] \\
= & \Pr_{G \leftarrow \mathcal{G}^{\mathtt{reg}}} [G \text{ not S.C.}] \\
\leq & \Pr_{G \leftarrow \mathcal{G}^{\mathtt{bin}}} [G \text{ not S.C.}], \; p = \tfrac{k}{n-1} \qquad \text{(Conj.1)} \\
\leq & 1 - e^{-2e^{\ln n - \frac{k}{n-1} n}} \qquad \text{(Conj.2)}
\end{aligned}
$$

$\Pr[G \neq \mathrm{Core}(G)]$ for the set of all users: apply union bound.

Implication

- Recall:

$$\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right] \leq \Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\left(G\right)\right] + \frac{1}{k+1}$$

Implication

- Recall:

$$\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right] \leq \Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right] + \frac{1}{k+1}$$

- If $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right] \leq \frac{1}{k+1}$, then a graph-analysing adversary is at most twice as successful as with the trivial attack

Implication

- Recall:
$$\Pr\left[\mathrm{Exp}_{\mathcal{A},\mathrm{Samp}}(U,|R|)\right] \leq \Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right] + \tfrac{1}{k+1}$$
- If $\Pr_{G\leftarrow\mathcal{G}^{\mathrm{Samp}}}\left[G \neq \mathrm{Core}\,(G)\right] \leq \frac{1}{k+1}$, then a graph-analysing adversary is at most twice as successful as with the trivial attack
- If Conjectures 1 and 2 hold, it suffices to set
$$k \geq \ln(2|U|) + \sqrt{2\ln(2|U|)},$$
i.e. set k to be logarithmic in number of users to resist graph-based attacks.

Summary

- Background of ring-signature-based anonymous systems and ring samplers
- Modelling anonymity of ring samplers by transaction graphs and security games
- Conjectures about strong connectivity of random directed graphs
- Provably secure ring size for partitioning samplers to resist graph-based deanonymisation attacks
- E-print: `ia.cr/2022/132`

Ivy Woo
Aalto University, Finland
`ivy.woo@aalto.fi`

Thank You!

References

- Viktoria Ronge, Christoph Egger, Russell W. F. Lai, Dominique Schröder, and Hoover H. F. Yin. Foundations of ring sampling.
  Proceedings on Privacy Enhancing Technologies, 3:265–288, 2021
- Tamir Tassa. Finding all maximally-matchable edges in a bipartite graph.
  Theoretical Computer Science, 423:50–58, 2012
- R. Daniel Mauldin. *The Scottish book: mathematics from the Scottish Café, with selected problems from the new Scottish Book*.
  Birkhäuser, 2015
- Alasdair J. Graham and David A. Pike. A note on thresholds and connectivity in random directed graphs.
  Atl. Electron. J. Math, 3(1):1–5, 2008
- Ilona Palásti. On the strong connectedness of directed random graphs.
  Studia Sci. Math. Hungar, 1:205–214, 1966
- Saravanan Vijayakumaran. Analysis of cryptonote transaction graphs using the dulmage-mendelsohn decomposition.
  Cryptology ePrint Archive, Report 2021/760, 2021