On the Computational Complexities of Three Privacy Measures for Large Networks Under Active Attack

Bhaskar DasGupta

Department of Computer Science

University of Illinois at Chicago

Chicago, IL 60607, USA

dasgupta@cs.uic.edu

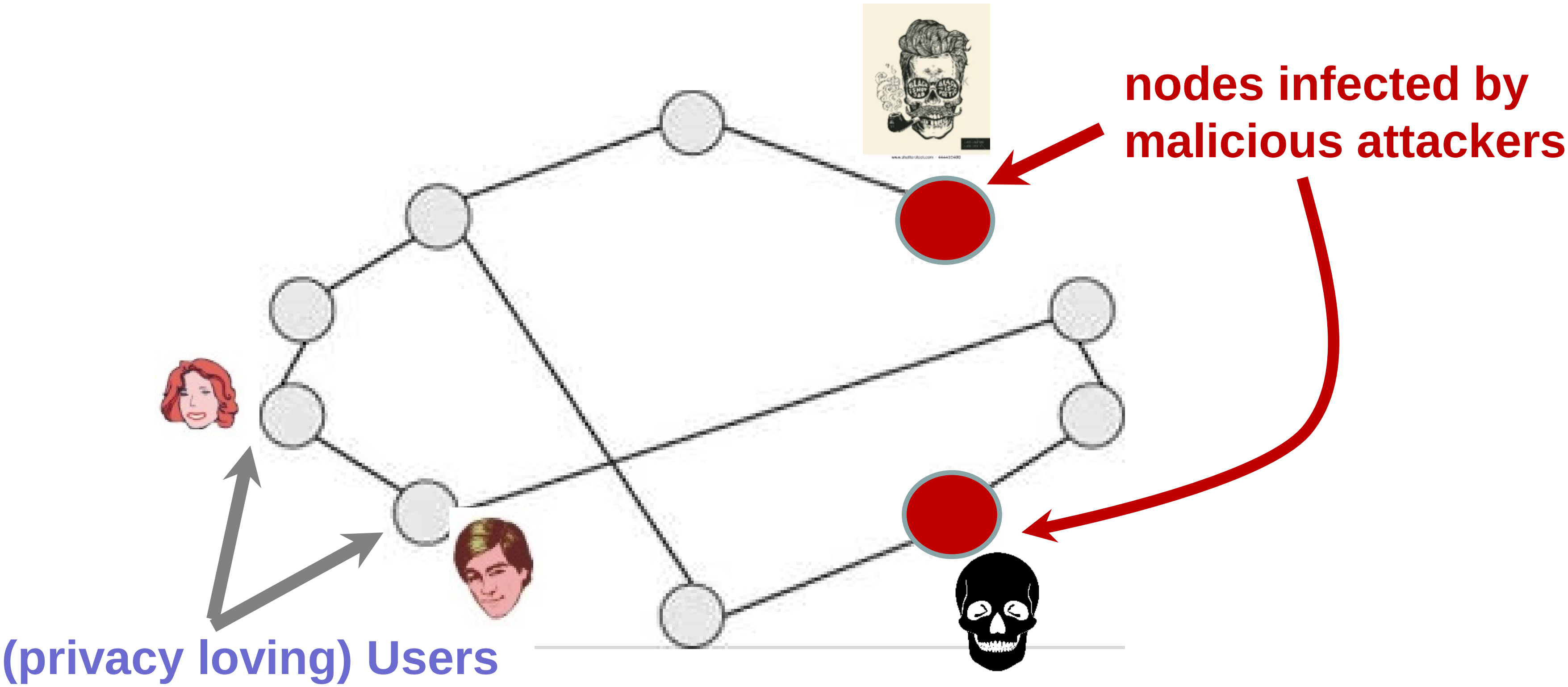
http://www.cs.uic.edu/~dasgupta

Bases on joint work with

T. Chatterjee, N. Mobasheri, V. Srinivasan and I. Yero

Supported by NSF grant IIS-1160995

Network Privacy Under Active Attack



malicious attackers are interested in sensitive attributes such as

- **node degrees**
- **inter-node distances**
- **connecitivity of network**

(k, ℓ)-anonymity (Trujillo-Rasua and Yero, 2016)

- ℓ is the maximum number of attacker nodes
  - (*e. g.*, estimated through statistical methods)

- k is a number indicating a privacy threshold
  - prevent adversary from "identifying individuals" with probability higher than $1/k$

identifying the "relevant attribute"

(for this talk)
distance vector
from attacked nodes

k-antiresolving set and (k, ℓ)-anonymity

Illustration for $k = 2, \ell = 5$

Undirected graph $G = (V, E)$

$V = \{v_1, v_2, \ldots, v_{12}\}$

	v_1	v_2	v_3	v_4	v_5
v_6	3	1	1	2	5
v_7	3	1	1	2	5
v_8	3	1	1	2	5
v_9	1	2	1	2	4
v_{10}	1	2	1	2	4
v_{11}	1	2	1	2	1
v_{12}	1	2	1	2	1

dist_{v_6,v_3}

(length of a shortest path between v_6 and v_3)

$\mathcal{S} = \{v_1, v_2, v_3, v_4, v_5\}$

is a 2-antiresolving set of size 5

(k-antiresolving set may not exist for some k)

k-metric antidimension $\mathrm{adim}_k(G)$

minimum cardinality of any k-antiresolving set

Related Prior Concepts

- Metric dimension (also called landmarks)

  Distance vectors must be *mutually non-identical*

  [Harary & Melter; 1976] [Khuller, Raghavachari & Rosenfeld; 1996] [Hauptmann, Schmied & Viehmann; 2012]

  Similar in flavor to general set cover problem

- Strong metric dimension

  Constrained distance vectors

  [Oellermann & Peters-Fransen; 2012] [DasGupta & Mobasheri; 2017]

  Similar in flavor to the node cover problem

Other known privacy computational models and concepts

- Multi-party communication context
  - [Yao, 1979], [Kushilevitz, 1992]
- Geometric notions of privacy
  - [Feigenbaum, Jaggard, Schapira, 2010],
    [Comi, DasGupta, Schapira, Srinivasan, 2012]
- Information-theoretic
  - [Bar-Yehuda, Chor, Kushilevitz, Orlitsky, 1993]
- Differential privacy (database retrieval context)
  - [Dwork, 2006]
- Anonymization approach (like this talk)
  - [Backstrom, Dwork, Kleinberg, 2007]

Problem 1 (metric anti-dimension or ADIM)

Find a k-antiresolving set $\mathcal{S}$ of nodes that maximizes k

Intuitively, it sets an absolute bound $1/k$ on the privacy violation probability of an adversary assuming that the adversary can use any number of attacker nodes

In practice, however, the number of attacker nodes employed by the adversary may be limited

This leads us to Problem 2

Problem 2 ($k_{\geqslant}$-metric antidimension or $\mathrm{ADIM}_{\geqslant k}$)

Given k, find a k'-antiresolving node set $\mathcal{S}$ such that

- **$k' \geqslant k$, and**
- **$|\mathcal{S}|$ is minimized**

n is number of nodes

Our Results for Problems 1 and 2

Theorem 1

(a) Both ADIM and ADIM$_{\geqslant k}$ can be solved in $O(n^4)$ time.

(b) Both ADIM and ADIM$_{\geqslant k}$ can also be solved in $O\left(\frac{n^4 \log n}{k}\right)$ time "with high probability"

(*i.e.*, with a probability of at least $1 - n^{-c}$ for some constant $c > 0$)

Remark

The randomized algorithm in (b) runs faster that the deterministic algorithm in (a) provided $k = \omega(\log n)$

Trade-off: (k, ℓ)-anonymity vs. (k', ℓ')-anonymity
$k' > k, \ell' < \ell$

(k', ℓ')-anonymity has smaller privacy violation probability $1/k'$ but can only tolerate infection of fewer number ℓ' of nodes

This leads us to Problem 3

Problem 3 ($k_{=}$-metric antidimension or $\text{ADIM}_{=k}$)

Given k, find a k-antiresolving node set $\mathcal{S}$ that minimizes $|\mathcal{S}|$

n is number of nodes

Our Results for Problems 3

Theorem 2

(a) $\textsc{Adim}_{=k}$ **is NP-complete for any** k **in the range** $1 \leqslant k \leqslant n^{\varepsilon}$ **where** $0 \leqslant \varepsilon < \frac{1}{2}$ **is any arbitrary constant**

even if the diameter of the input graph is 2

(b) Assuming $\mathbf{NP} \not\subseteq \mathbf{DTIME}\left(n^{\log\log n}\right)$**, there exists a universal constant** $\delta > 0$ **such that**
$\textsc{Adim}_{=k}$ **does not admit a** $\left(\frac{1}{\delta}\ln n\right)$**-approximation for any integer** k **in the range** $1 \leqslant k \leqslant n^{\varepsilon}$ **for any constant** $0 \leqslant \varepsilon < \frac{1}{2}$

even if the diameter of the input graph is 2

(c) If $k = n - c$ **for some constant** c **then** $\textsc{Adim}_{=k}$ **can be solved in polynomial time**

Our Results for Problems 3

Remarks on Theorem 2

(i) The result in (b) provides a much stronger inapproximability result compared to that in (a) at the expense of a slightly weaker complexity-theoretic assumption

($i.e.$, NP $\not\subseteq$ DTIME $(n^{\log\log n})$ vs. P $\neq$ NP)

(ii) For $k = 1$, the inapproximability ratio in (a) is asymptotically optimal up to a constant factor

because of the $(1 + \ln(n-1))$-approximation of $\text{A}\small{\text{DIM}}_{=1}$ in Theorem 3(a)

to be discussed next

n is number of nodes

Our Results for Problems 3 (continued)

k=1

Theorem 3

(a) $\textsc{Adim}_{=1}$ **admits a** $(1 + \ln(n-1))$**-approximation in** $O(n^3)$ **time**

(b) If G **has at least one node of degree** 1 **then** $\textsc{Adim}_{=1}$ **can be solved in** $O(n^3)$ **time**

(c) If G **does not contain a cycle of** 4 **edges then** $\textsc{Adim}_{=1}$ **can be solved in** $O(n^3)$ **time**

Some Future Research Questions

- **Is it possible to design a non-trivial approximation algorithm for $\textsc{Adim}_{=k}$ for $k > 1$?**

  We conjecture that a $O(\log n)$ -approximation is possible for $\textsc{Adim}_{=k}$ for every fixed k

- **We provided logarithmic inapproximability result for $\textsc{Adim}_{=k}$ for every k *roughly* up to $\sqrt{n}$. Can this approximability result be further improved when k is not a constant ?**

  We conjecture that the inapproximability factor can be further improved to $\Omega\left(n^{\varepsilon}\right)$ for some constant $0 < \varepsilon < 1$ when k is around $\sqrt{n}$.

- **How about attributes other than distance vectors ?**

"But before we move on, allow me to belabor the point even further..."