Multiplicative Data Perturbation for Privacy Preserving Data Mining

Kun Liu

Ph.D. Dissertation Defense

Dept. of Computer Science and Electrical Engineering

University of Maryland, Baltimore County (UMBC)

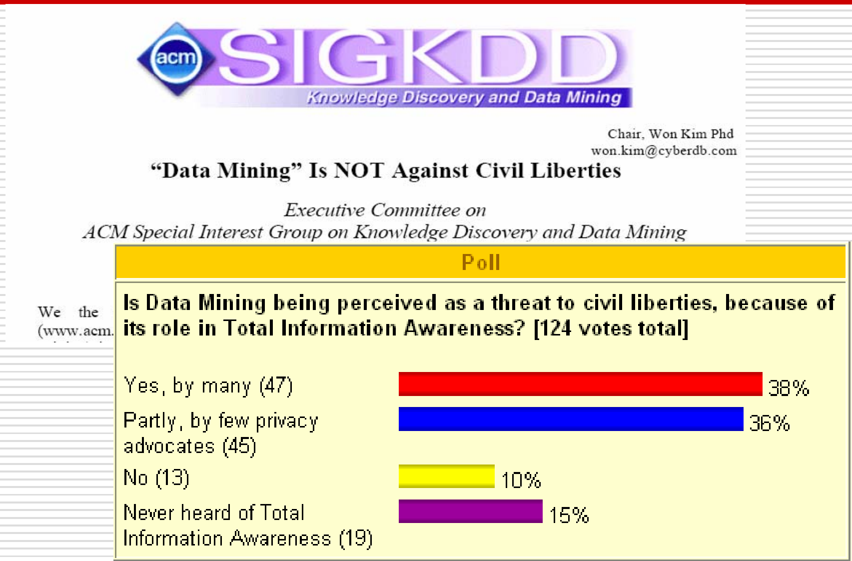
January 15th, 2007

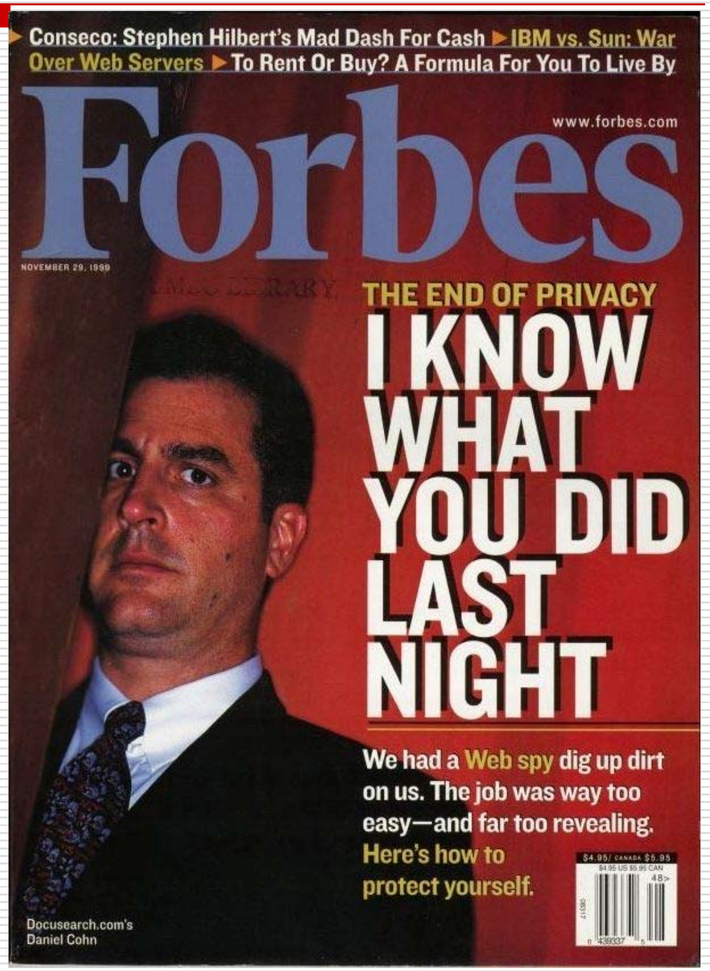
UMBC

AN HONORS UNIVERSITY IN MARYLAND

DIADIC Laboratory

Growing Privacy Concerns





"Detailed information on an individual's credit, health, and financial status, on characteristic purchasing patterns, and on other personal preferences is routinely recorded and analyzed by a variety of governmental and commercial organizations."

- M. J. Cronin, "e-Privacy?" Hoover Digest, 2000.

Privacy-Preserving Data Mining

- “the best (and perhaps only) way to overcome the ‘limitations’ of data mining techniques is to do more research in data mining, including areas like data security and privacy-preserving data mining, which are actually active and growing research areas.”

  - SIGKDD Executive Committee, “‘Data Mining’ Is NOT Against Civil Liberties,” 2003.

- Privacy-preserving data mining is “the study of how to produce valid mining models and patterns without disclosing private information.”

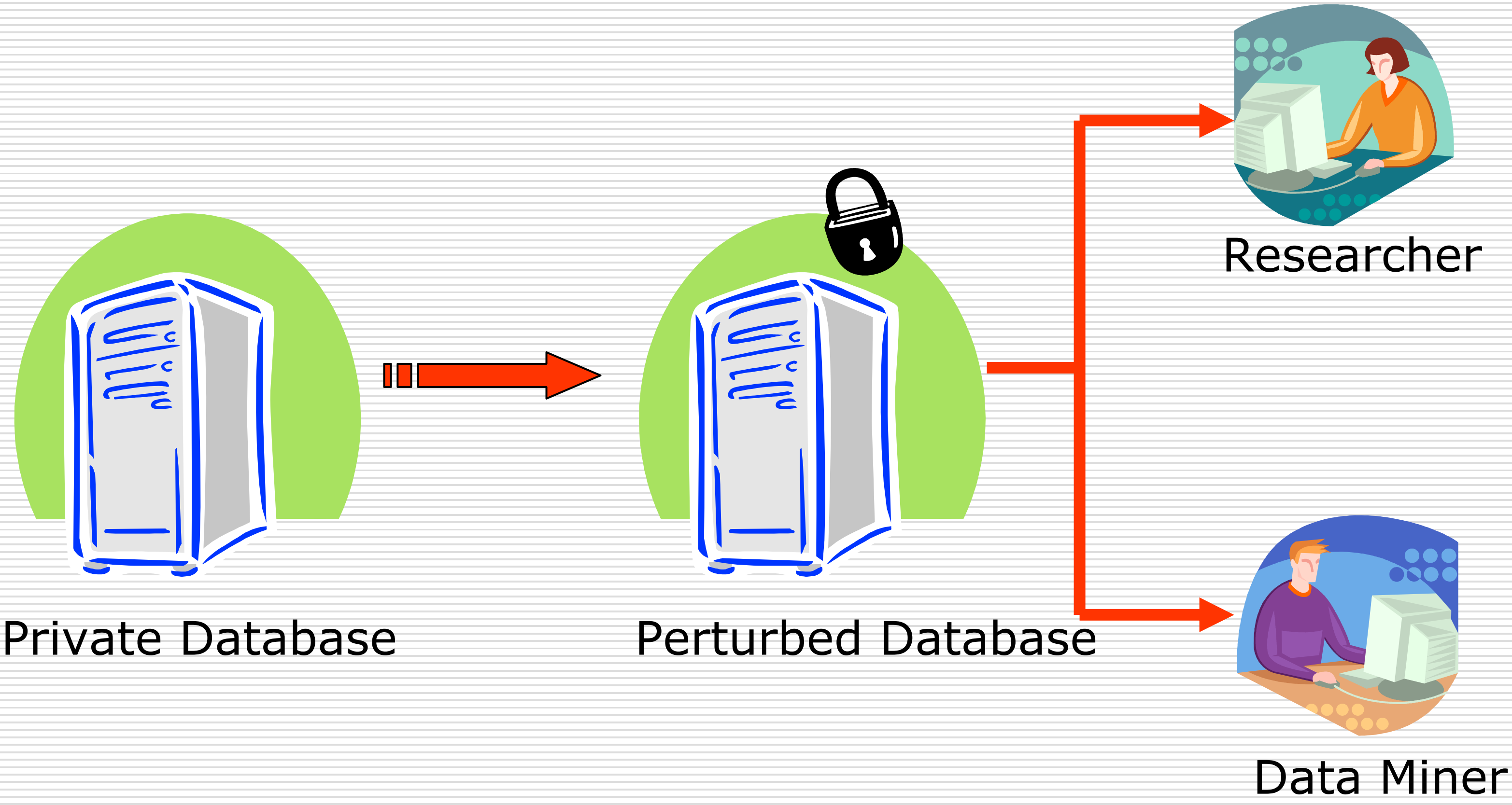
  - F. Giannotti and F. Bonchi, “Privacy Preserving Data Mining,” KDUbiq Summer School, 2006.

Privacy-Preserving Data Mining

- Data Perturbation
  - Hiding private data while mining patterns
- Secure Multi-Party Computation
  - Building a model over multi-party distributed databases without knowing others' inputs
- Knowledge Hiding
  - Hiding sensitive rules/patterns
- Privacy-aware Knowledge Sharing
  - Do the data mining results themselves violate privacy?

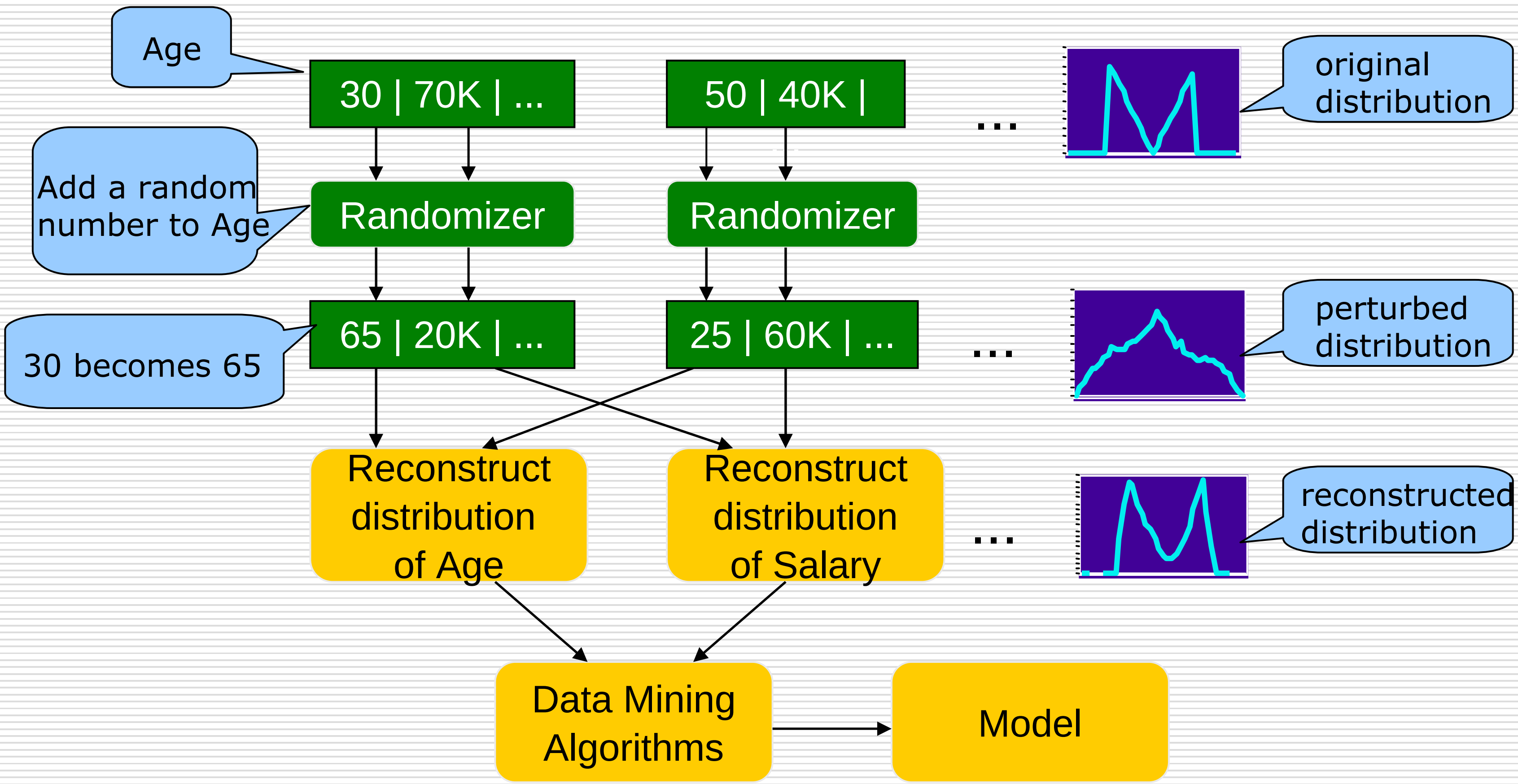
[more]

Data Perturbation



Census Model

Additive Data Perturbation



R. Agrawal and R. Srikant, "Privacy-preserving data mining," ACM SIGMOD, 2000.

Additive vs. Multiplicative Noise

- Additive perturbation is not safe.
  - "in many cases, the original data can be accurately estimated from the perturbed data using a spectral filter that exploits some theoretical properties of random matrices"

    - Kargupta et al., "On the Privacy Preserving Properties of Random Data Perturbation Techniques," ICDM, 2003.
  - Related work: [Huang05], [Guo06], etc.

- How about multiplicative noise?
  - Has not been carefully studied.
  - Topic of this dissertation.

Primary Contributions

- We examined the effectiveness of exact Euclidean distance preserving data perturbation, and developed three attack techniques.
  - K. Liu, C. Giannella, and H. Kargupta, “An attacker's view of distance preserving maps for privacy preserving data mining,” *10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'06)*, 2006.

- We proposed a random projection-based approximate distance preserving perturbation as a possible remedy, and analyzed its privacy issues.
  - K. Liu, H. Kargupta, and J. Ryan, “Random projection-based multiplicative perturbation for privacy preserving distributed data mining,” *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 18(1), 2006.

Roadmap

- Traditional Multiplicative Noise
- Distance Preserving Data Perturbation
  - Fundamental Properties
  - Known Input-Output Attack
  - Know Sample Attack
  - Independent Signal Attack
- Random Projection-based Perturbation
  - Fundamental Properties
  - Bayes Privacy Model
  - Attacks Revisit
- Conclusion and Future Work

Traditional Multiplicative Noise

Private Database X

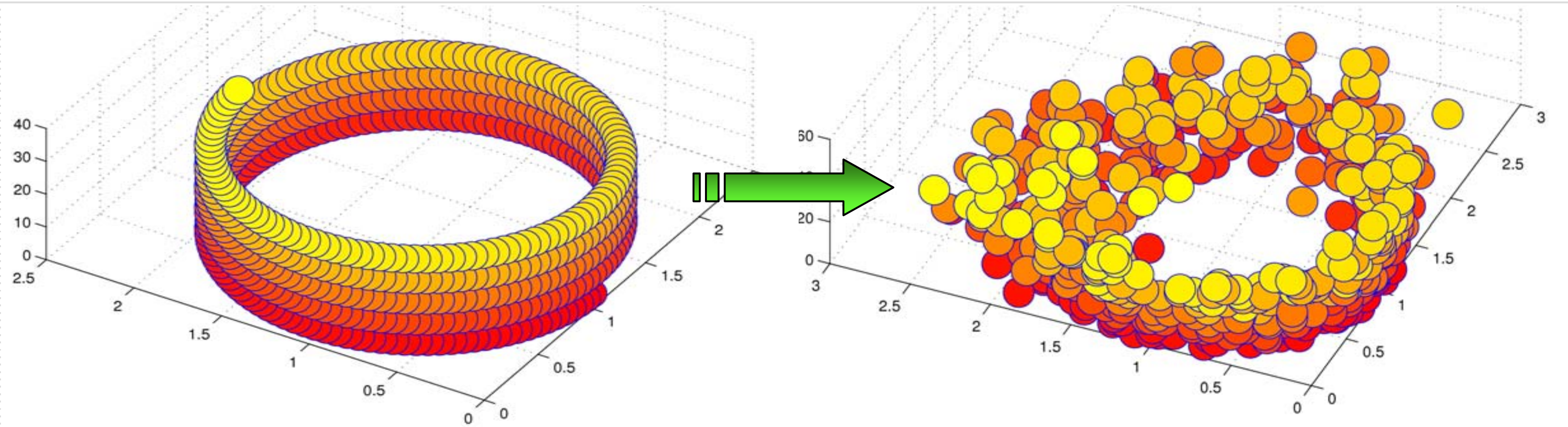
ID	1001	1002
Wages	98,563	83,821
Rent	1,889	1,324
Tax	2,899	2,578

Perturbed Database Y

ID	1001	1002
Wages	116,166	85,396
Rent	1,878	1,381
Tax	2,964	2,135

2,899 * 1.0224 = 2,964

$y_{ij} = x_{ij} \times r_{ij},$ where x_{ij} is the private data, $r_{ij} \sim N(1, \sigma)$ [Kim03].



Traditional Multiplicative Noise

- Mechanism
  - Each data element/cell is randomized independently by multiplying a random number.
- Pros
  - Summary statistics (*e.g.*, mean, variance) can be estimated from the perturbed data.
  - Effective if data disseminator only wants minor perturbation
  - Popular in the statistics community.
- Cons
  - Equivalent to additive perturbation after a logarithmic operation. Vulnerable to attacks designed for additive noise.
  - Not preserving Euclidean distance; not suitable for many data mining tasks.

Roadmap

- Traditional Multiplicative Noise
- Distance Preserving Data Perturbation
  - Fundamental Properties
  - Known Input-Output Attack
  - Know Sample Attack
  - Independent Signal Attack
- Random Projection-based Perturbation
  - Fundamental Properties
  - Bayes Privacy Model
  - Attacks Revisit
- Conclusion and Future Work

Distance Preserving Perturbation

- Dist. preserving perturbation

  $T:\mathbb{R}^n \to \mathbb{R}^n$ if $\forall x, y \in \mathbb{R}^n,\ \| x-y \| = \| T(x)-T(y) \|$

- Dist. preserving perturbation is equivalent to

  $x \in \mathbb{R}^n \to Mx+v,$ for $M \in \mathrm{O}_n$ and $v \in \mathbb{R}^n,$

  where O_n is the set of all $n \times n$ orthogonal matrices.

- Dist. preserving perturbation with origin fixed

  $x \in \mathbb{R}^n \to Mx,$ where $M \in \mathrm{O}_n \leftrightarrow$ Orthogonal Transformation

  Our Focus

Distance Preserving Perturbation

ID	1001	1002
Wages	-26,326	-22,613
Rent	-94,502	-80,324
Tax	10,151	8,432

=

-0.2507	0.4556	-0.8542
-0.9653	-0.0514	0.2559
0.0726	0.8887	0.4527

×

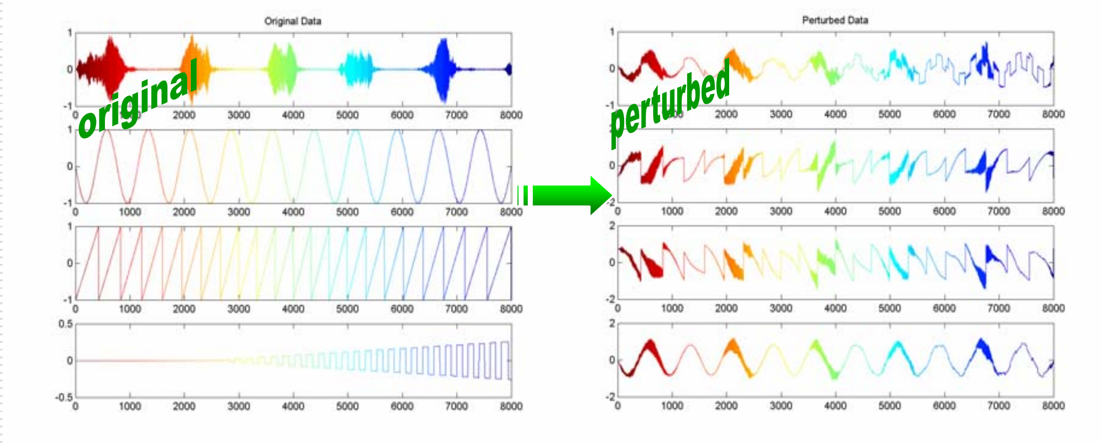
ID	1001	1002
Wages	98,563	83,821
Rent	1,889	1,324
Tax	2,899	2,578

Y M X

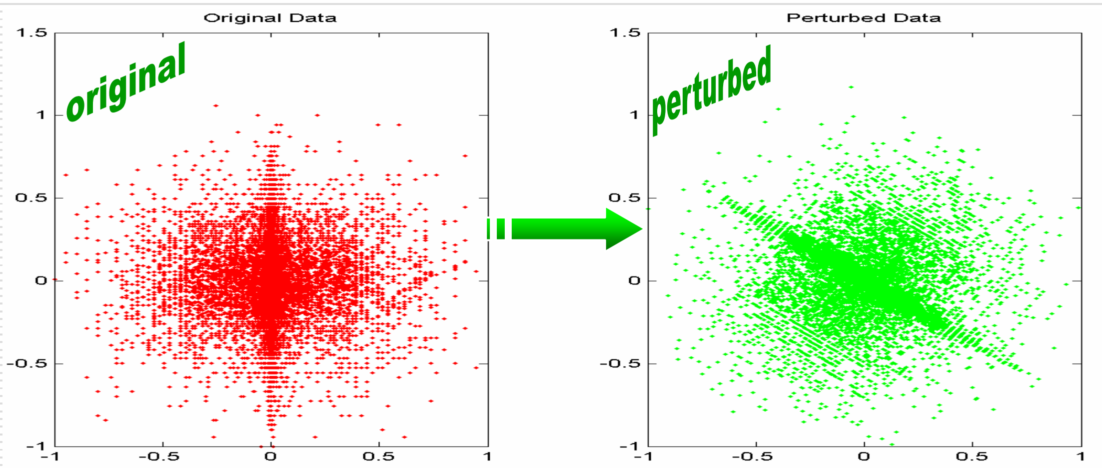
- Perturbation Model $Y = MX$
  - X: original private data with each column a record
  - Y: perturbed data
  - M: orthogonal perturbation matrix
- Perturbed data produces exactly the same data mining results
  - Clustering [Oliveira04], Classification [Chen05]
  - Other related: [Mukherjee06], etc.

Distance Preserving Perturbation

Attributes



Records



Is Dist. Preserving Perturbation Secure?

- Known Input-Output Attack: attacker knows some collection of linearly independent private data records and their corresponding perturbed version.

- Known Sample Attack: attacker has a collection of independent data samples from the same distribution the original data was drawn.

- Independent Signals Attack: all data attributes are non-Gaussian and statistically independent

Privacy Breach

- ε-Privacy Breach

  For any $\varepsilon > 0$, we say that an ε-privacy breach occurs if

  $$\| \hat{x} - x_i \| \leq \| x_i \| \varepsilon$$

  where $\hat{x}$ is the attacker's estimate of x_i, the i^{th} data tuple in X.

- Probability of ε-Privacy Breach

  $$\rho(x_i, \varepsilon) = \mathrm{Prob}\{ \| \hat{x} - x_i \| \leq \| x_i \| \varepsilon \}$$

  the probability that an ε-privacy breach occurs.

Known Input-Output Attack

$$\boxed{[Y_{n\times k} \quad Y_{n\times(m-k)}]} = M_{n\times n}[\boxed{X_{n\times k}} \quad X_{n\times(m-k)}]$$

KNOWN

- Assumption (can be relaxed): rank(X_{nxk})=k
- If k=n:
  - $M = Y_{n\times k} X^{-1}{}_{n\times k}, \; X_{n\times(m-k)} = M^T Y_{n\times(m-k)}$
  - Probability of privacy breach $\rho(x_i, \varepsilon) = 1$ for $\varepsilon = 0$ and any i.
  - The attacker has a perfect recovery of the private data.
- If k<n, what is going to happen?

Known Input-Output Attack

$$\boxed{[Y_{n\times k} \quad Y_{n\times(m-k)}]} = M_{n\times n}[\boxed{X_{n\times k}} \quad X_{n\times(m-k)}]$$

KNOWN

- If k<n, any matrix $\hat{M}$ in the set

  $$\Omega = \{\hat{M} \in O_n : \hat{M} X_{n\times k} = Y_{n\times k}\}$$

  can be the original perturbation matrix $M_{n\times n}$, where is O_n is the set of all nxn orthogonal matrices. [more]

- The attacker chooses one uniformly from Ω as an estimation of $M_{n\times n}$, uses that to recover other private data, and computes the probability of privacy breach. [more]

Known Input-Output Attack

- Probability of Privacy Breach

$$\rho(x_i,\varepsilon) = \text{Prob}\{\|\hat{x} - x_i\| \ \le \ \|x_i\|\varepsilon\}$$

$$= \text{Prob}\{\|\hat{M}Mx_i - x_i\| \ \le \ \|x_i\|\varepsilon\}$$

$$= \begin{cases} \dfrac{1}{\pi} 2\arcsin\left(\dfrac{\|x_i\|\varepsilon}{2d(x_i, X_{n\times k})}\right) & \text{if } \|x_i\|\varepsilon < 2d(x_i, X_{n\times k})\ ; \\ 1 & \text{otherwise.} \end{cases}$$

where $d(x_i, X_{n\times k})$ is the distance of x_i from the column space of $\mathrm{X}_{n\times k}$,

and $\hat{M}$ is uniformly chosen from $\Omega = \{\hat{M} \in \mathrm{O}_n : MX_{n\times k} = Y_{n\times k}\}$. [more]

Known Input-Output Attack Example

	X_1	X_2	X_3
Private Data X:	25.0000	30.0000	45.0000
	75.0000	90.0000	105.0000

	Y_1	Y_2	Y_3
Perturbed Data Y:	-42.0198	-50.4237	-68.5443
	66.9652	80.3582	91.3875

X_1->Y_1 KNOWN

→UNKNOWN

- The distance of X_2 from the column space of X_1 is 0, therefore $\rho(x_2,\varepsilon)=1$ for any ε.
- The distance of X_3 from the column space of X_1 is 9.4868, therefore $\rho(x_3,\varepsilon)=\frac{1}{\pi}2\arcsin\left(\frac{\|x_3\|\varepsilon}{2\times 9.4868}\right)$, e.g. $\rho(x_3,0.01)=3.84\%$.

Known Sample Attack

- Assumptions
  - Each data record arose as an independent sample from some unknown distribution
  - The attacker has a collection of samples independently chosen from the same distribution
  - The covariance of the distribution has all distinct eigenvalues (holds true in most practical situations [Jolliffe02]).
- Attack Technique
  - Exploring the relationship between the principal eigenvectors of the original data and the principal eigenvectors of the perturbed data.

Known Sample Attack

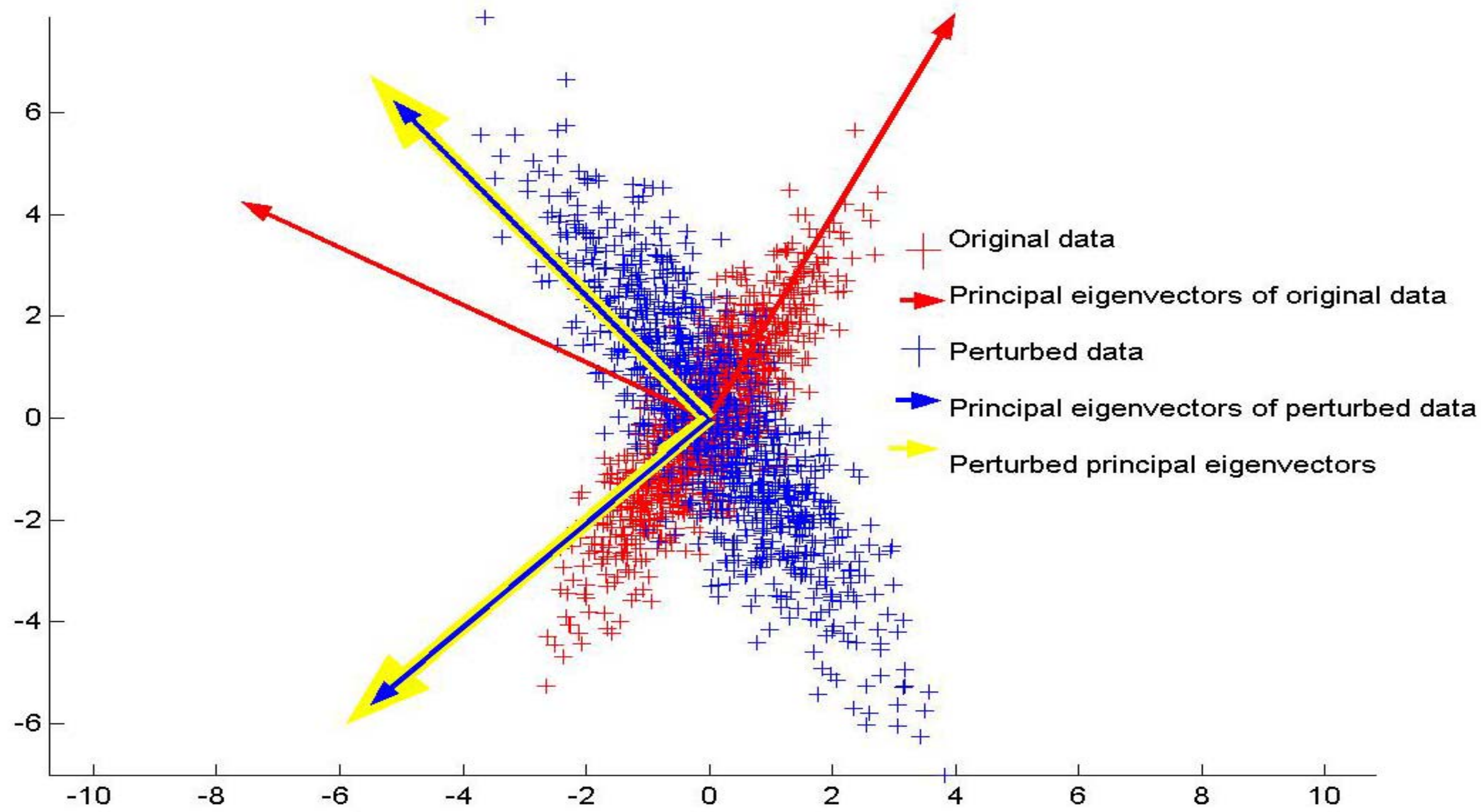


Fig. Relationship between original and perturbed principal eigenvectors. [more]

Known Sample Attack Experiments

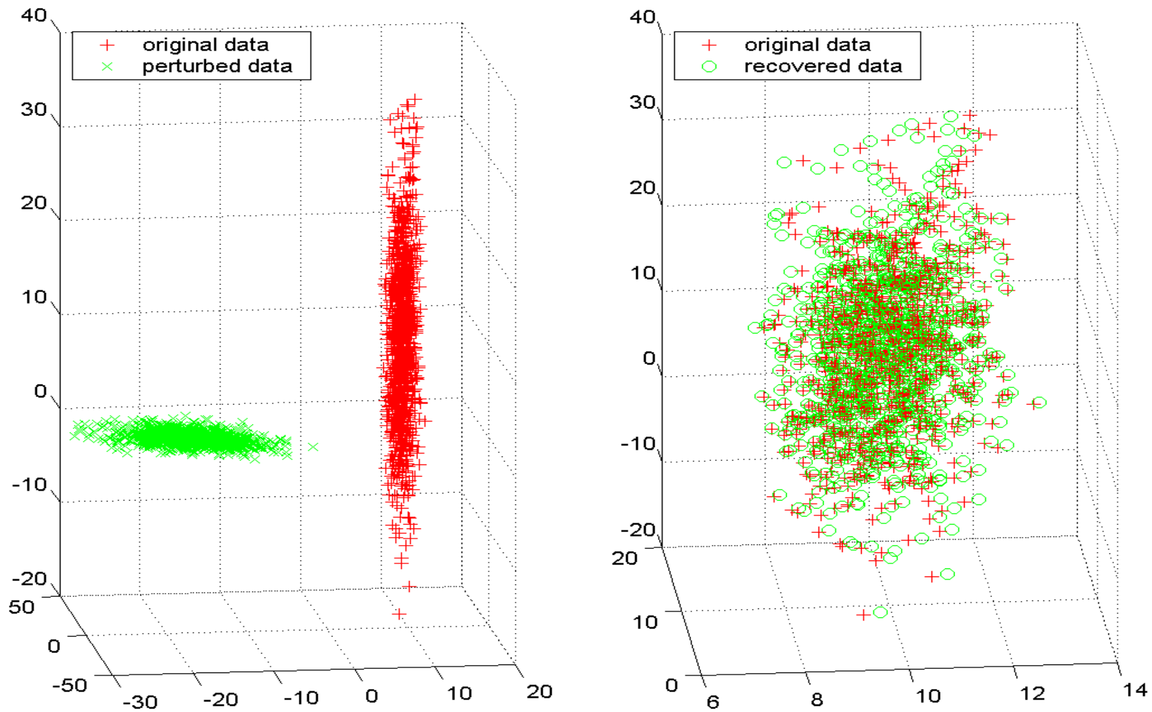


Fig. Known sample attack for 3D Gaussian data with 10,000 private tuples. The attacker has 2% samples from the same distribution. The average relative error of the recovered data is 0.0265 (2.65%). [more]

Known Sample Attack Experiments

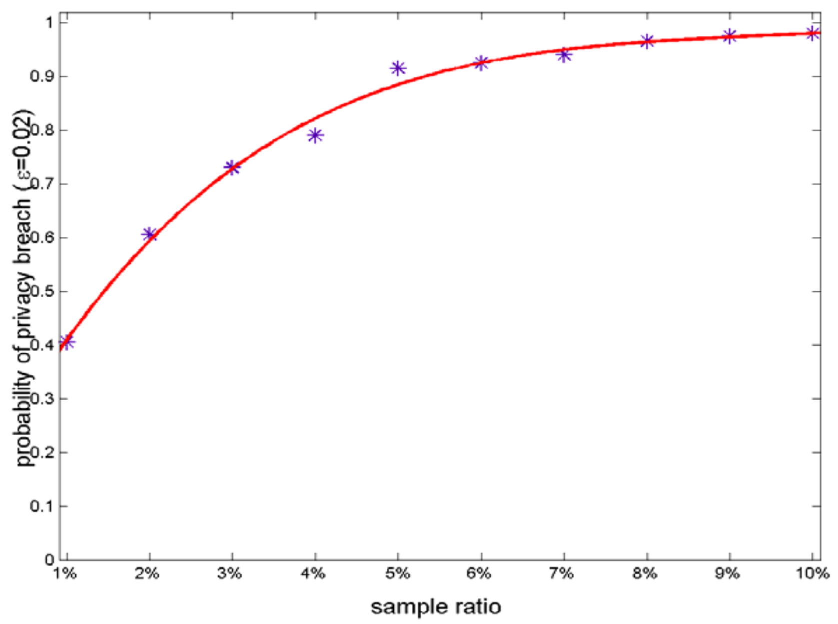


Fig. Probability of privacy breach w.r.t. attacker's sample size. The relative error bound ε is fixed to be 0.02. (3D Gaussian data with 10,000 private tuples.)

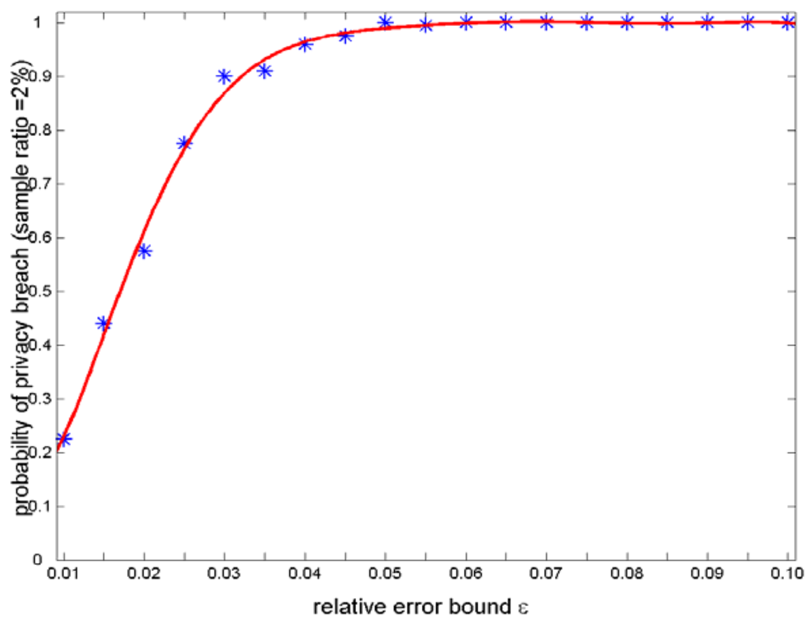


Fig. Probability of privacy breach w.r.t. the relative error bound ε. The sample ratio is fixed to be 2%. (3D Gaussian data with 10,000 private tuples.)

[more]

Independent Signals Attack

- Basic Independent Component Analysis Model

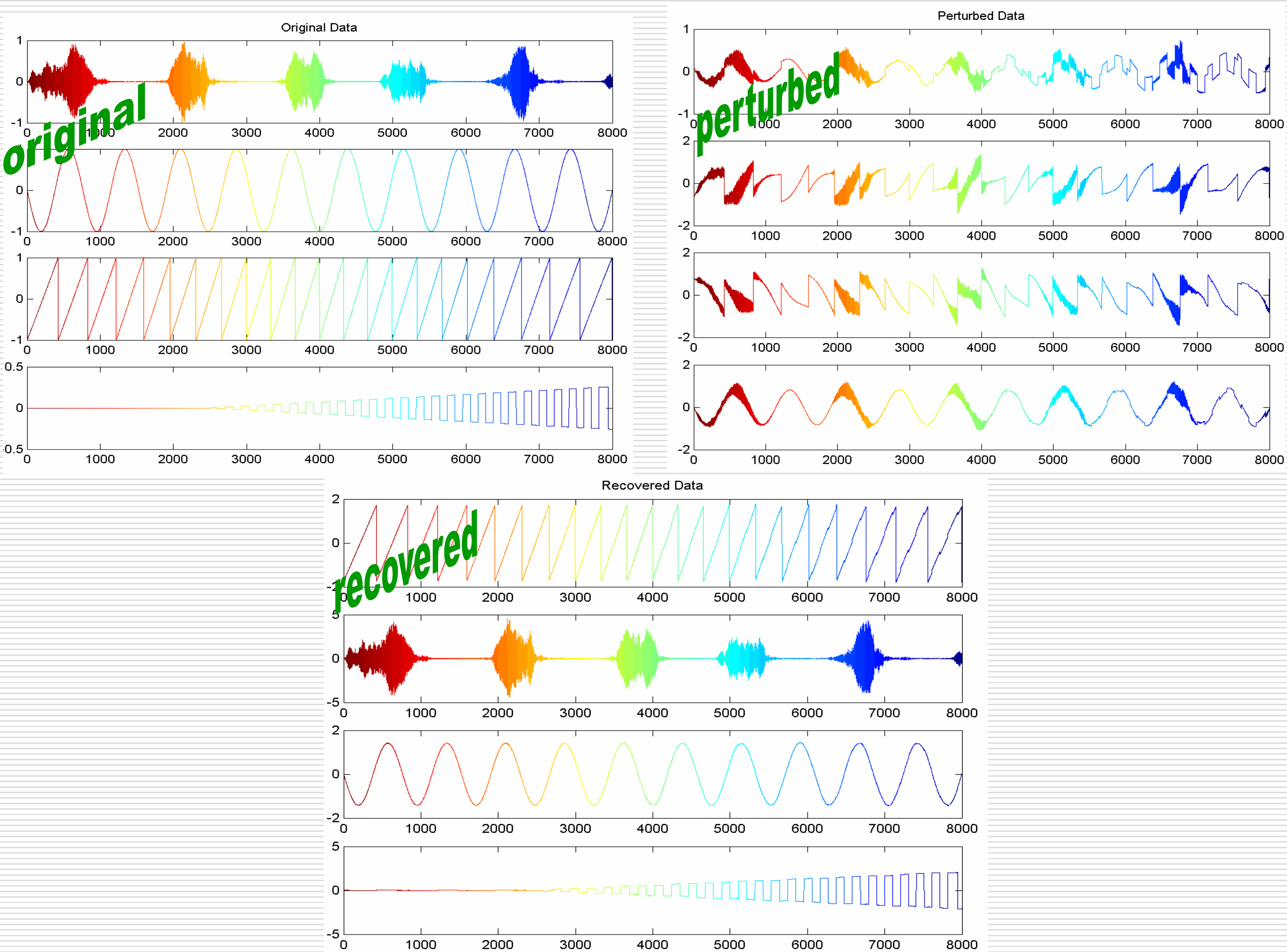
$$\boxed{Y_{k\times m}} = \boxed{M_{k\times n} X_{n\times m}} = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \dots & \dots & \dots & \dots \\ m_{k1} & m_{k2} & \dots & m_{kn} \end{bmatrix} \begin{bmatrix} x_1(t_1) & x_1(t_2) & \dots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \dots & x_2(t_m) \\ \dots & \dots & \dots & x_3(t_m) \\ x_n(t_1) & x_n(t_2) & \dots & x_n(t_m) \end{bmatrix}$$

KNOWN UNKNOWN

- Objective: recover the original signals X from only the observed mixtures Y.
- Requirements
  - Source signals are statistically independent
  - All signals must be non-Gaussian with exception of one
  - k ⩾ n
  - Matrix M must be of full column rank

[more]

Independent Signals Attack Experiments


Original Data
original
Perturbed Data
perturbed
Recovered Data
recovered

Distance Preserving Perturbation Summary

- Mechanism
  - Whole data set is perturbed by multiplying an orthogonal matrix.
- Pros
  - Perturbed data preserves Euclidean distance.
  - Many data mining algorithms can be applied to the perturbed data and produce exactly the same results as if applied to the original data.
- Cons
  - Vulnerable to known input-output attack
  - Vulnerable to known sample attack
  - Vulnerable to independent signals attack

Roadmap

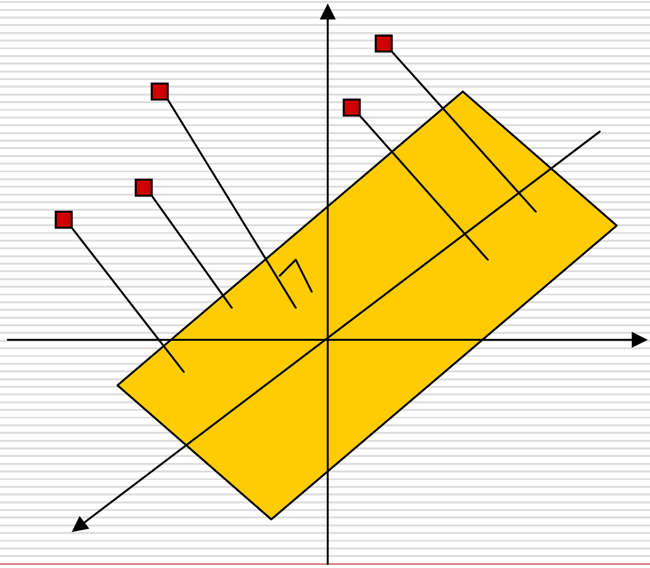
- Traditional Multiplicative Noise
- Distance Preserving Data Perturbation
  - Fundamental Properties
  - Known Input-Output Attack
  - Know Sample Attack
  - Independent Signal Attack
- Random Projection-based Perturbation
  - Fundamental Properties
  - Bayes Privacy Model
  - Attacks Revisit
- Conclusion and Future Work

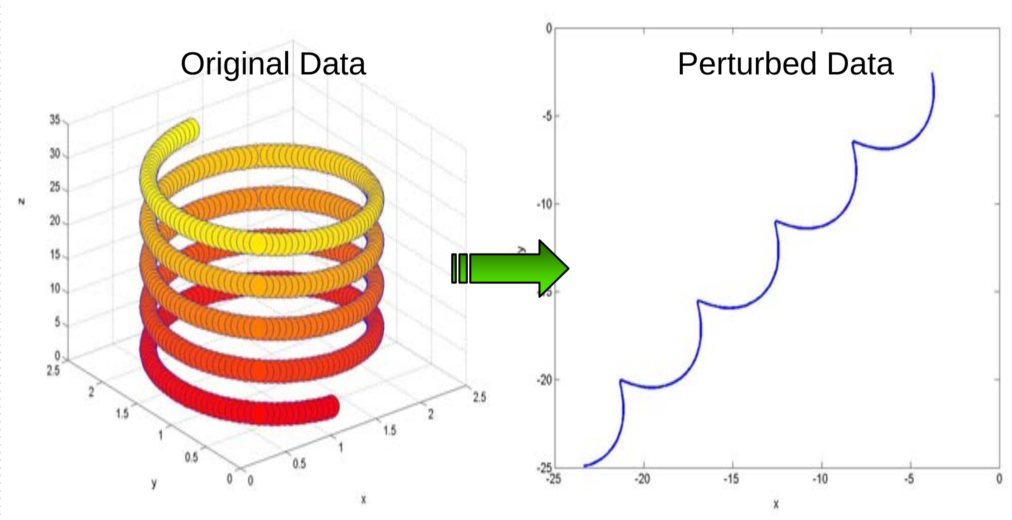
Random Projection

- Basic Model

$$u = \frac{1}{\sqrt{k}\sigma_r} R_{k\times m} x_{m\times 1}, \text{ and } v = \frac{1}{\sqrt{k}\sigma_r} R_{k\times m} y_{m\times 1},$$

$$\text{where } k < m \text{ and } r_{ij} \text{ is i.i.d.} \sim N(0, \sigma_r).$$





Random Projection

- Preserving Inner Product

$$E[u^T v - x^T y] = 0 \text{ and } Var[u^T v - x^T y] = \frac{1}{k}(\sum_i x_i^2 \sum_i y_i^2 + (\sum_i x_i y_i)^2).$$

→The distortion produced by random projection is zero on the average, and its variance is inversely proportional to k, dimension of new space.

[more]

- Preserving Euclidean Distance

$$\Pr\{(1-\eta)\|x-y\|^2 \le \|u-v\|^2 \le (1+\eta)\|x-y\|^2\} = \int_{k(1-\eta)}^{k(1+\eta)} f(t;k)dt,\ \eta>0$$

where $f(t;k)$ is the p.d.f. of chi-square distribution with k degrees of freedom.

→The probability that relative error is bounded with in (1 ± η) increases proportionally with k.

[more]

Bayes Privacy Model

- Primitives
  - Let x be the private data and y the perturbed one.
  - Attacker's Prior Belief: $f(x)$
  - Attacker's Additional Background Knowledge: θ
  - Attacker's Posteriori Belief: $f(x \mid y, \theta)$
- Information Non-Disclosure Principle
  - The perturbed data should provide the attacker with little additional information beyond the attacker's prior belief and other background knowledge.
- Example
  - (ρ_1, ρ_2)-privacy [Evfimevski03] happens when

    $f(x) < \rho_1$ and $f(x \mid y, \theta) > \rho_2$ OR $f(x) > 1-\rho_1$ and $f(x \mid y, \theta) < 1-\rho_2$

Maximum a Posteriori Probability (MAP) Estimate

- (ρ_1, ρ_2,)-privacy works only for discrete data. It assumes statistically independent inputs and outputs, and requires transition probability explicitly defined. Not appropriate for multiplicative perturbation.

- We propose a *maximum a posteriori probability (MAP) estimate*-based approach
  - 1. $\hat{x}_{MAP} = \arg\max_{x} f(x \mid y, \theta)$
  - 2. $\hat{x}_{MAP}$ is compared with x_i to see whether any extra information is disclosed, *e.g.*, $\| \hat{x}_{MAP} - x_i \| \leq \| x_i \| \varepsilon.$

Why Maximum a Posteriori Probability (MAP) Estimate

- It is closely related to maximum a posteriori probability hypothesis testing. [more]
- It considers both prior and posterior belief. In the absence of a priori knowledge, MAP estimate becomes maximum likelihood estimate (MLE).
- It often produces estimates with errors that are not much higher than the minimum mean square error.
- It is relatively easy to derive the conditional p.d.f. in the multiplicative data perturbation scenario.

Maximum a Posteriori Probability (MAP) Estimate

- Assumption I: Attackers' best knowledge of f(x) is it is uniformly distributed over an multi-dimensional interval.
- Assumption II: Attacker has no other background knowledge, *i.e.*, $\theta = \varnothing$.
- MAP Estimate: $$\hat{x}_{MAP} = \arg\max_x f(x \mid y, \theta)$$ $$= \arg\max_x \frac{k^{1/2}}{(2\pi x^T x)^{k/2}} \exp\left(-\frac{k y^T y}{2 x^T x}\right),$$ $$\text{where } x \in \mathbb{R}^n \text{ and } y \in \mathbb{R}^k.$$
- Solution: Any $\hat{x}$ in the interval that satisfies $\hat{x}^T \hat{x} = y^T y$.
- Conclusion: MAP does not offer attacker more info than what has been implied by properties of random projection itself.

Privacy /Accuracy Control

- Random Projection

$$u = \frac{1}{\sqrt{k}\sigma_r} R_{k\times m} x_{m\times 1}, \text{ and } v = \frac{1}{\sqrt{k}\sigma_r} R_{k\times m} y_{m\times 1},$$

$$\text{where } k < m \text{ and } r_{ij} \text{ is i.i.d.} \sim N(0, \sigma_r).$$

- Accuracy

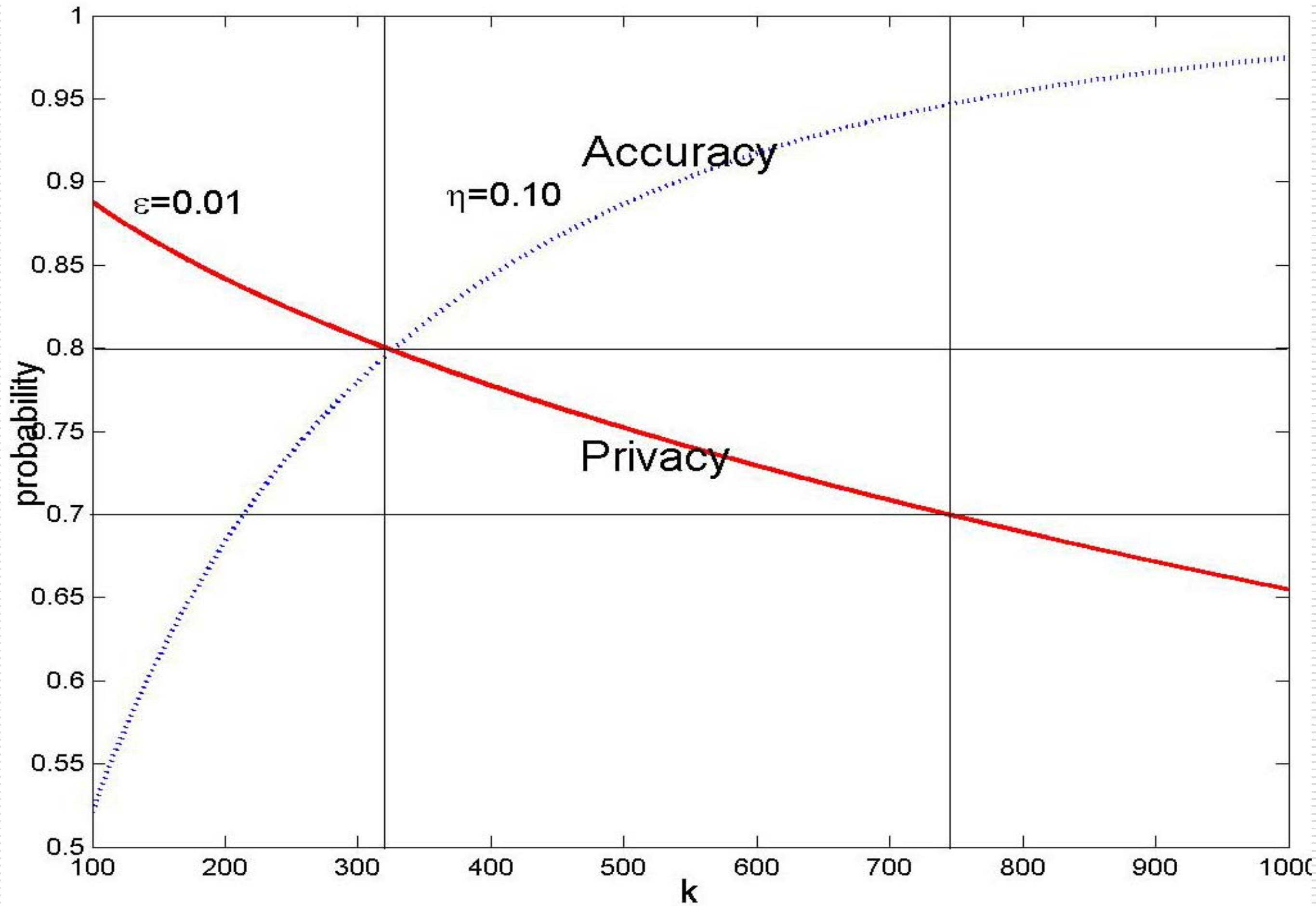
$$\Pr\{(1-\eta)\|x-y\|^2 \leq \|u-v\|^2 \leq (1+\eta)\|x-y\|^2\} = \int_{k(1-\eta)}^{k(1+\eta)} f(t;k)dt,\ \eta > 0.$$

- $\neg$ ε-Privacy Breach

$$\Pr\{\|\hat{x}_{MAP} - x\| > \|x\|\varepsilon\} = \int_{-\infty}^{k(1-\varepsilon)^2} f(t;k)dt + \int_{k(1+\varepsilon)^2}^{+\infty} f(t;k)dt.$$

Here f(t;k) is the p.d.f. of chi-square distribution with k degrees of freedom.

Privacy/Accuracy Control



Roadmap

- Traditional Multiplicative Noise
- Distance Preserving Data Perturbation
  - Fundamental Properties
  - Known Input-Output Attack
  - Know Sample Attack
  - Independent Signal Attack
- Random Projection-based Perturbation
  - Fundamental Properties
  - Bayes Privacy Model
  - Attacks Revisit
- Conclusion and Future Work

Independent Signals Attack

$$Y_{k\times m} = \frac{1}{\sqrt{k}\sigma_r} R_{k\times n} X_{n\times m}$$

- When k<n, at most (k-1) source signals can be separated out [Cao96].

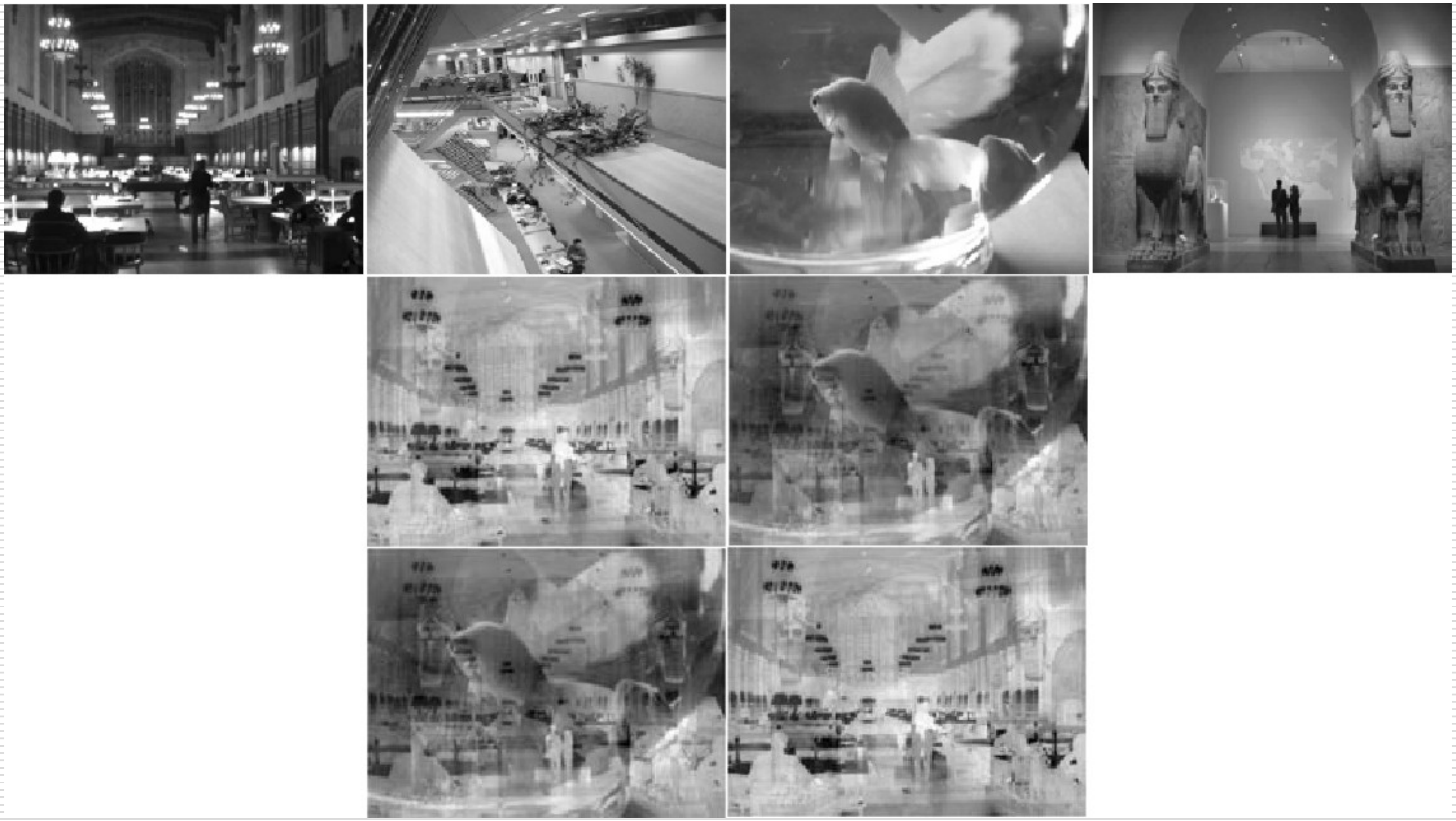
- With probability one, linear ICA can't separate out any of the original signals for any (k x n) (k⩽n/2, n ⩾ 2) random matrix with i.i.d. entries chosen from continuous distribution [Liu06a].

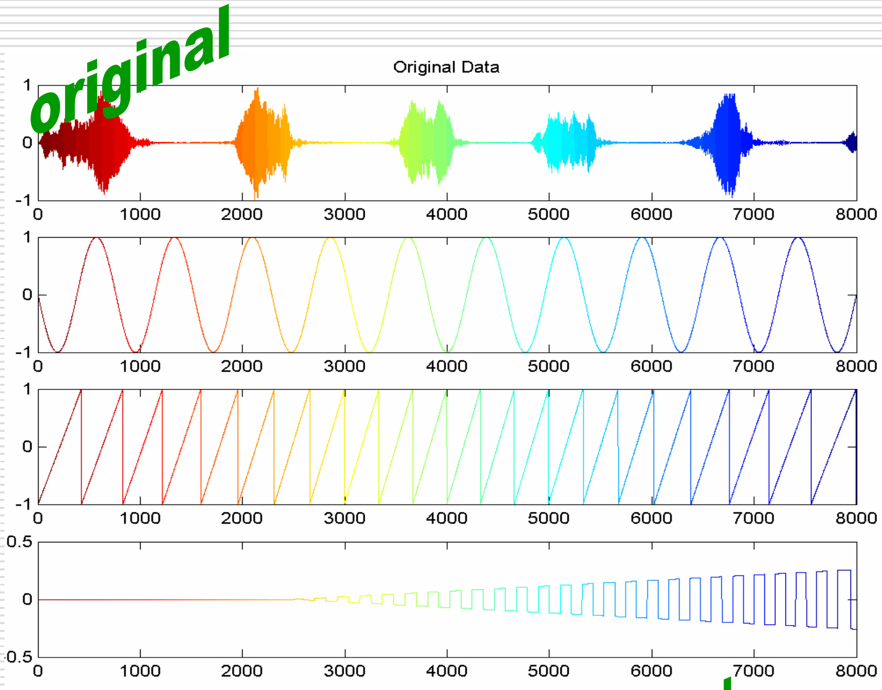
Independent Signals Attack Experiments

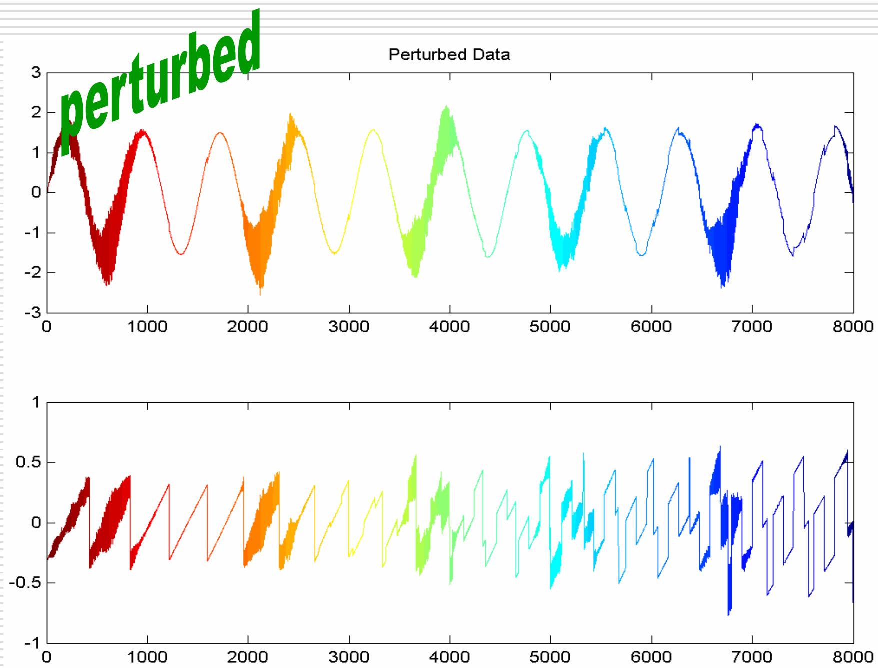
original

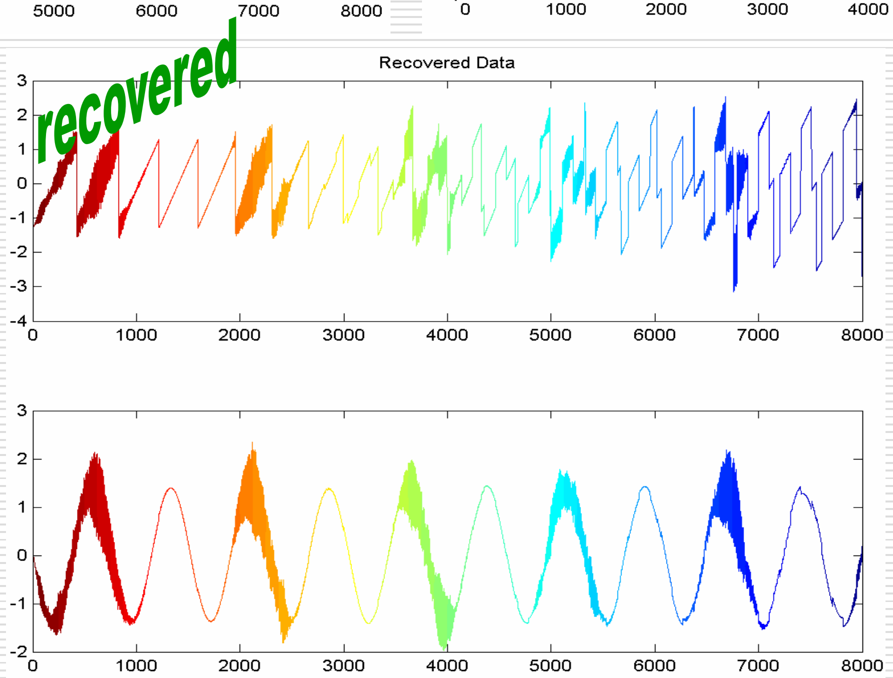
perturbed

recovered




original
Original Data


perturbed
Perturbed Data


recovered
Recovered Data

Known Sample Attack

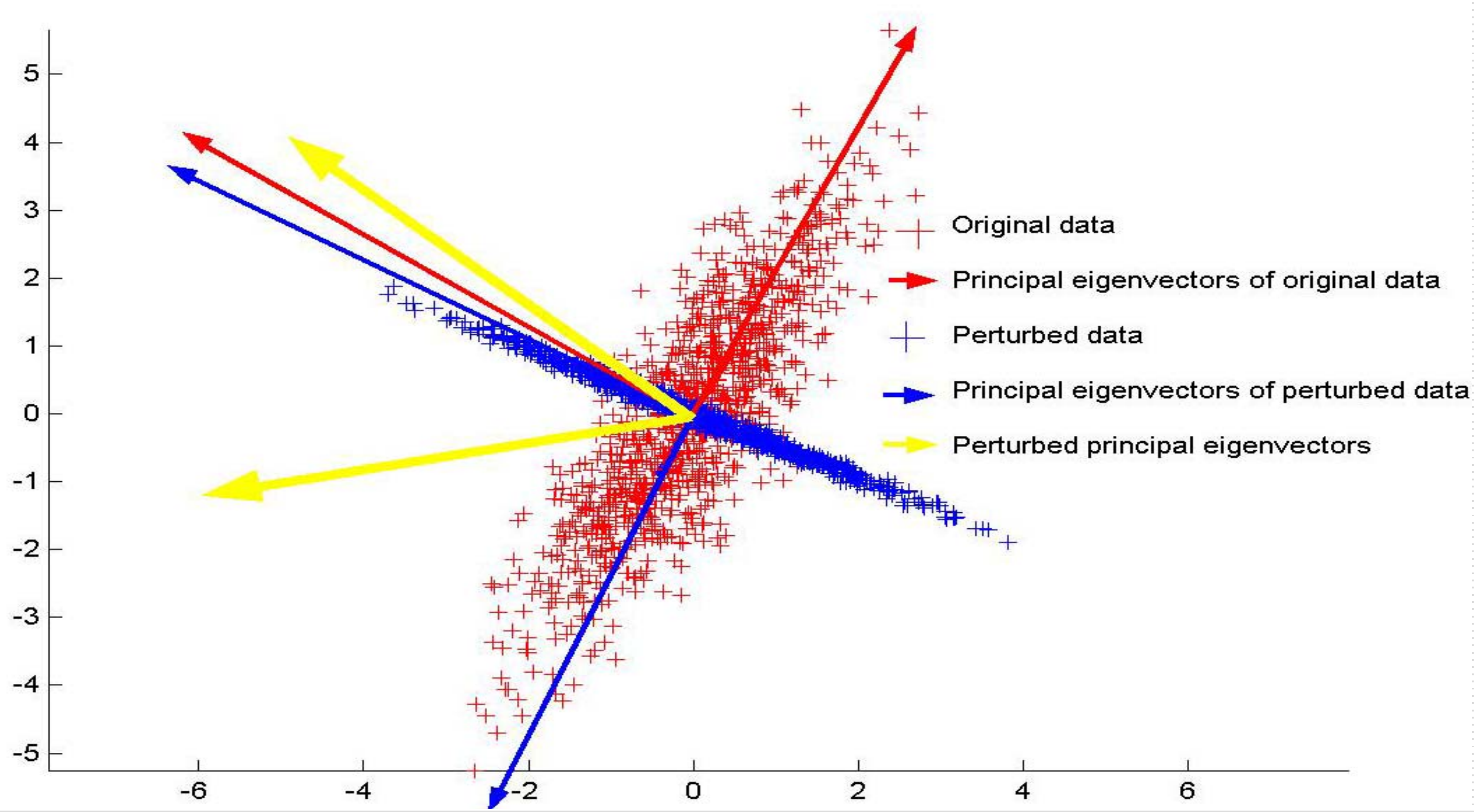


Fig. Relationship between original and perturbed principal eigenvectors.

Known Input-Output Attack

$$\boxed{[Y_{k\times p} \quad Y_{k\times(m-p)}]} = \frac{1}{\sqrt{k}\sigma_r} R_{k\times n} [\boxed{X_{n\times p}} \quad X_{n\times(m-p)}]$$

KNOWN

- If p=n and rank(X_{nxp}) = p, R can be recovered, but still it is an under-determined system of linear equations.

- MAP estimate shows that relative error decreases as known input-output pairs increases; relative error increases as k decreases.

[more]

Random Projection-based Perturbation Summary

- Mechanism
  - Data is projected to a lower dimensional random space.
- Pros
  - From the perspective of MAP estimate, random projection does not disclose more information than what have been implied by the distance preservation properties.
  - It offers better privacy protection than orthogonal transformation-based distance preserving perturbation.
- Cons
  - Perturbed data approximately preserves Euclidean distance, therefore little loss in accuracy.

Conclusions

- Traditional Multiplicative Data Perturbation

- Distance Preserving Data Perturbation
  - Known Input-Output Attack (linear algebra, statistics)
  - Known Sample Attack (PCA)
  - Independent Signals Attack (ICA)

- Random projection-based Data Perturbation
  - Accuracy Analysis
  - Bayes Privacy Model
  - Maximum a Posteriori Probability (MAP) Estimate
  - Privacy/Accuracy Control
  - Attack Analysis

- Privacy Issues Are Intrinsically Complex
  - Need joined efforts from researchers, engineers, sociologists, legal experts, policy makers…

Future Work

- A game theoretic framework for large scale distributed privacy preserving data mining
  - Distributed and ubiquitous computing becomes popular
  - Some participants cooperative and honest, some malicious
  - Computation in such environment is more like a game
  - Necessary to develop a game theoretic framework

- Combination of cryptographic techniques and perturbation techniques
  - Cryptographic techniques offers strong privacy guarantee, but with high communication and computation cost
  - Perturbation provides statistically weaker privacy protection, but more efficient
  - Would be ideal to combine them to achieve both efficiency and privacy

References

[Kargupta03] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques," in *Proceedings of the IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[Huang05] Z. Huang, W. Du, and B. Chen, "Deriving private information from randomized data," in *Proceedings of the 2005 ACM SIGMOD Conference (SIGMOD'05)*, Baltimore, MD, June 2005, pp. 37-48.

[Guo06] S. Guo, X. Wu, and Y. Li, "On the lower bound of reconstruction error for spectral filtering based privacy preserving data mining," in *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'06)*, Berlin, Germany, 2006.

[Kim03] J. J. Kim and W. E. Winkler, "Multiplicative noise for masking continuous data," Statistical Research Division, U. S. Bureau of the Census, Washington D.C., Tech. Rep. Statistics #2004-01, April 2003.

[Liu06a] K. Liu, H. Kargupta, and J. Ryan, "Random projection-based multiplicative data perturbation for privacy preserving distributed data mining," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 18, no. 1, pp. 92–106, January 2006.

[Liu06b] K. Liu, C. Giannella, and H. Kargupta, "An attacker's view of distance preserving maps for privacy preserving data mining," in Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'06), Berlin, Germany, pp. 297-308, 2006.

[Mukherjee06] S. Mukherjee, Z. Chen, and A. Gangopadhyay, "A privacy preserving technique for Euclidean distance-based mining algorithms using Fourier-related transforms," *The VLDB Journal*, p. to appear, 2006.

[Chen05] K. Chen and L. Liu, "Privacy preserving data classification with rotation perturbation," in *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05)*, Houston, TX, pp. 589–592, November 2005.

References

[Oliveira04] S. R. M. Oliveira and O. R. Zaïane, "Privacy preservation when sharing data for clustering," in *Proceedings of the International Workshop on Secure Data Management in a Connected World*, Toronto, Canada, pp. 67–82, August 2004.

[Jolliffe02] I. T. Jolliffe, Principal Component Analysis, 2nd ed., ser. Springer Series in Statistics. Springer, 2002.

[Evfimevski03] A. Evfimevski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proceedings of the ACM SIGMOD Conference (SIGMOD'03)*, San Diego, CA, June 2003.

[Cao96] X.-R. Cao and R.-W Liu, "A General Approach to Blind Source Separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 562-571, 1996.

I Have a Dream

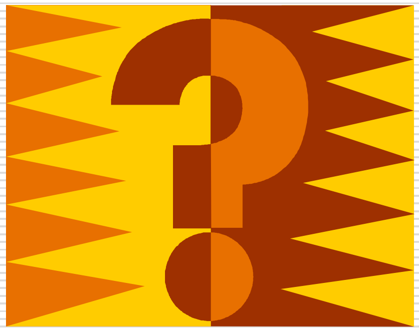
"I have a dream that one day this nation will rise up and live out the true meaning of its creed: 'We hold these truths to be self-evident, that all men are created equal.'"

- Martin Luther King, Jr., 1963.

"I have a dream that one day I will get a Ph.D. degree."

- Kun Liu, when he was a kid.

Thank you and Questions



Backup Slides

Overview of PPDM backup

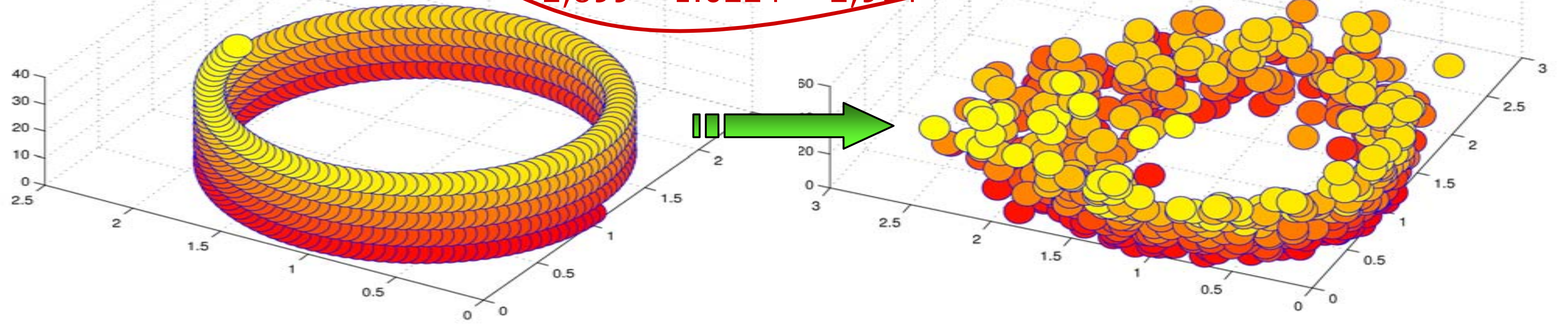
- Data Hiding
  - Data Perturbation
    - Value Distortion
      - Additive Perturbation
      - Multiplicative Perturbation
      - Data Microaggregation
      - Data Anonymization
      - Data Swapping
      - Other Techniques
    - Probability Distribution
      - Sampling Method
      - Analytical Method
  - Secure Multi-Party Computation
  - Distributed Data Mining
- Rule Hiding
  - Association Rule Hiding
  - Classification Rule Hiding

Traditional Multiplicative Noise backup

ID	1001	1002
Wages	98,563	83,821
Rent	1,889	1,324
Tax	2,899	2,578

ID	1001	1002
Wages	116,166	85,396
Rent	1,878	1,381
Tax	2,964	2,135

2,899 * 1.0224 = 2,964



- Properties:
  - $y_{ij} = x_{ij} \times r_{ij},$ where x_{ij} is the private data, $r_{ij} \sim N(1, \sigma)$ [Kim03].
  - Each data element randomized independently.
  - Original Mean and variance can be estimated from perturbed data.
  - Equivalent to additive perturbation after a logarithmic operation.
  - Not preserve Euclidean distance.

Knowledge Hiding backup

- What is disclosed?
  - the data (modified somehow)
- What is hidden?
  - some “sensitive” knowledge (*i.e.* secret rules/patterns)
- How?
  - usually by means of data sanitization. The data which we are going to disclose is modified, in such a way that the sensitive knowledge can no longer be inferred, while the original database is modified as less as possible.

Privacy-aware Knowledge Sharing backup

- What is disclosed?
  - the intentional knowledge (*i.e.*, rules , patterns, models)
- What is hidden?
  - the source data
- The central question
  - Do the data mining results themselves violate privacy

Privacy-aware Knowledge Sharing backup

Age = 27, Zip = 15254, Christian->American
(sup_count = 758, confidence = 99.8%)

Age = 27, Zip = 15254->American
(sup_count = 1518, confidence = 99.9%)

sup_count (27, 15254, Christian) = 758/.998 = 759.5
sup_count(27, 15254, Christian, ¬ American) = 759.5*0.002 = 1.519

sup_count(27, 15254) = 1518/0.999 = 1519.5
sup_count(27, 15254, ¬ American) = 1519.5 * 0.001 = 1.5195

Age = 27, Postcode = 45254, ¬ American->Christian
(sup_count ≈ 1.5, confidence ≈ 100.0%)

This information refers to my France neighbor…. he is Christian!

Known Input-Output Attack backup

$$[\boxed{Y_{n\times k} \quad Y_{n\times(m-k)}}] = M_{n\times n}[\boxed{X_{n\times k}} \quad X_{n\times(m-k)}]$$

KNOWN

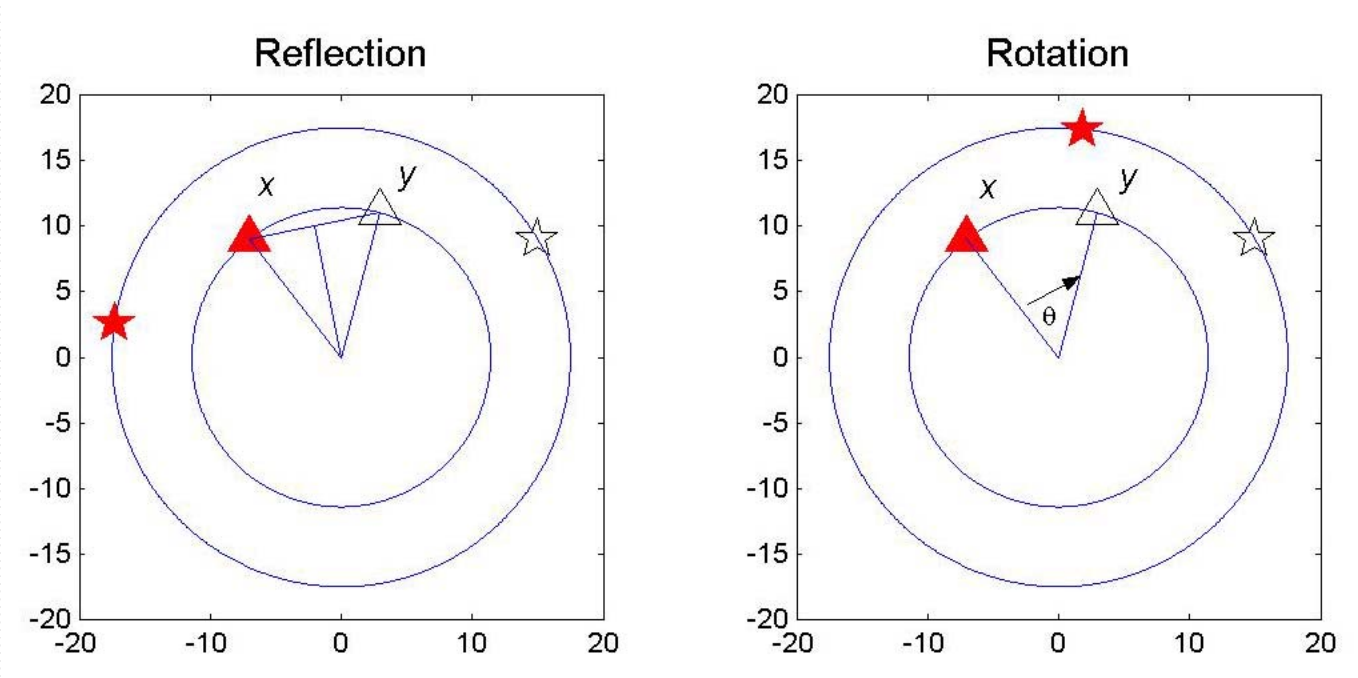
- Closed-form Expression of Ω

$$\Omega = \{\hat{M} \in O_n : MX_{n\times k} = Y_{n\times k}\}$$

$$= \{M(U_k U_k^T + U_{n-k} P U_{n-k}^T) : \ \forall P \in O_{n-k}\},$$

where U_k is the orthonormal basis for the column space of $X_{n\times k}$, U_{n-k} is the orthonormal basis for the orthogonal complement of the column space of $X_{n\times k}$.

Known Input-Output Attack backup



Special case in 2D space: when k = 1 and n = 2.

The attacker can't distinguish rotation and reflection.

Known Input-Output Attack backup

- Properties of the Probability of Privacy Breach
  - Attacker can compute the probability of privacy breach for a given private record and a relative error bound ε .
  - The larger the ε , the higher the probability of privacy breach.
  - The closer the private record is to the column space of the known records, the higher the probability of privacy breach.
  - The distance $d(x_i, X_{n \times k})$ can be computed from the perturbed data.

Known Sample Attack backup

- The principal eigenvectors of the original data have experienced the same distance preserving perturbation as the data itself.

  Let $Y = MX$, we have $Z_Y = MZ_X D$,

  where Z_Y is the eigenvector matrix of the covariance of Y;

  Z_X is the eigenvector matrix of the covariance of X;

  and D is a diagonal matrix with each entry on the diagonal ± 1.

- Z_Y can be computed from the perturbed data, Z_X can be estimated from the sample data. (See dissertation for choice of D, details omitted.)
- Attacker uses Z_X, Z_Y and D to recover M, and therefore X.

Known Sample Attack Experiments backup

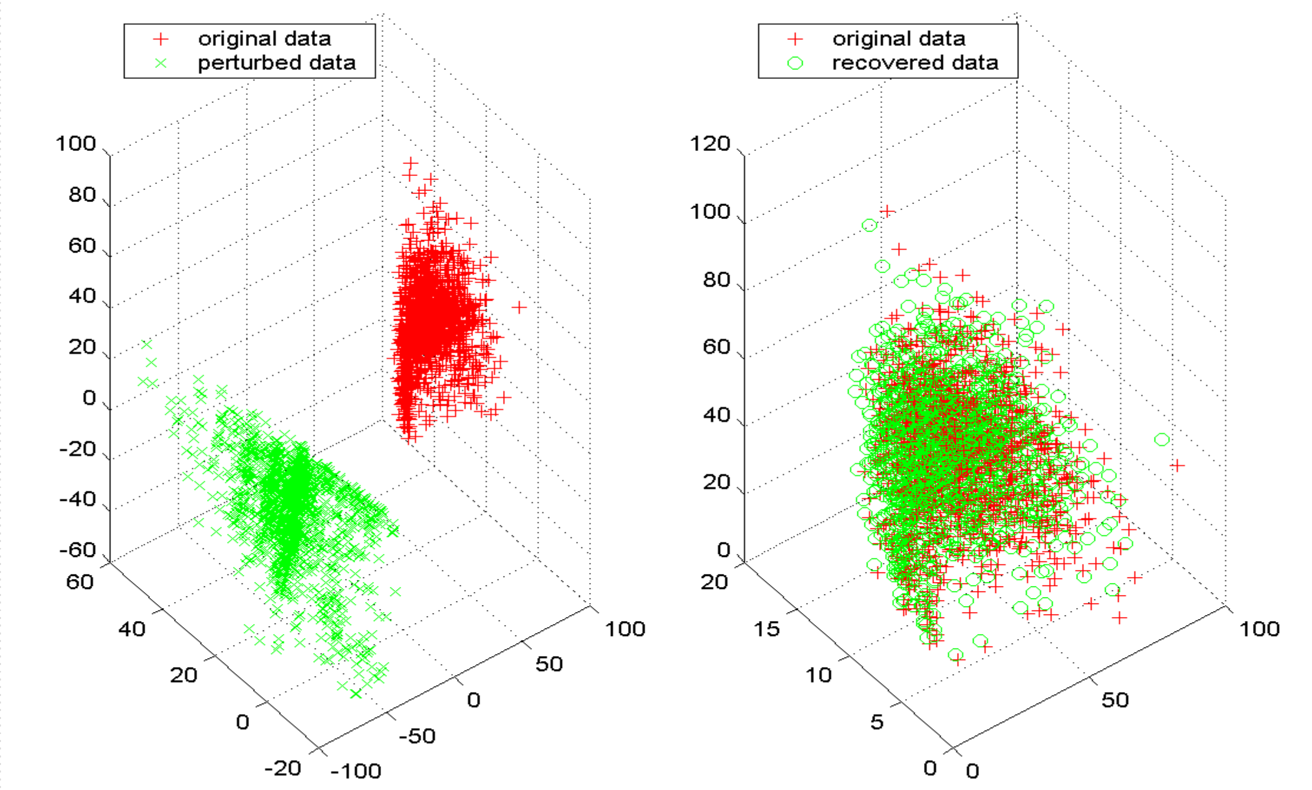


Fig. Known sample attack for Adult data with 32,561 private tuples. The attacker has 2% samples from the same distribution. The average relative error of the recovered data is 0.1081 (10.81%).

Known Sample Attack Experiments backup

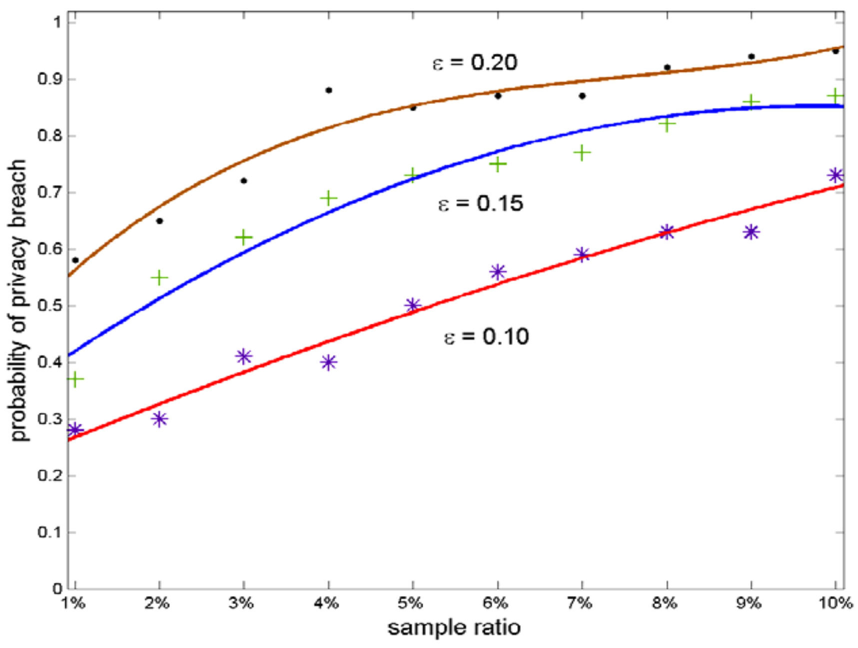


Fig. Probability of privacy breach w.r.t. attacker's sample size. The relative error bound ε changes from 0.10 to 0.20. (Adult data with 32,561 private tuples)

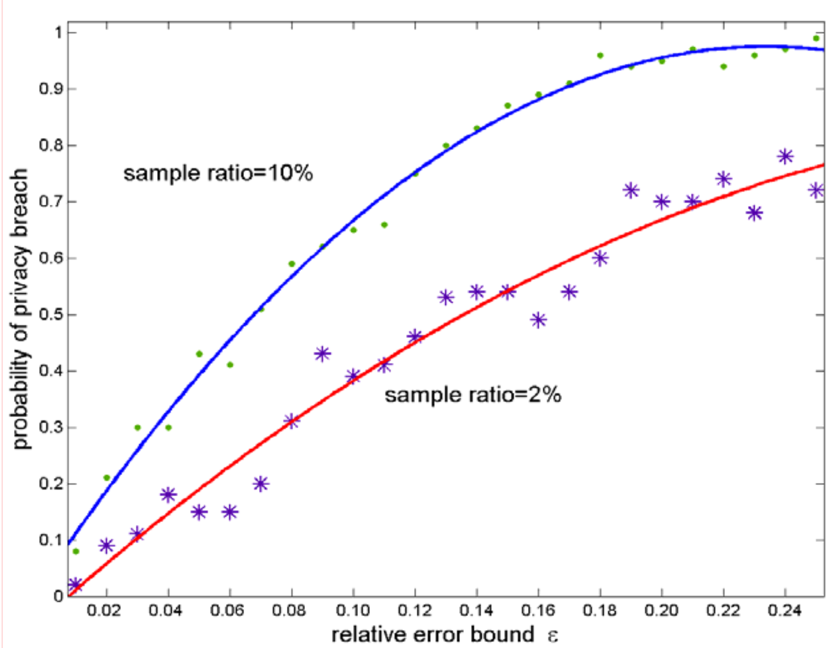


Fig. Probability of privacy breach w.r.t. the relative error bound ε . The sample ratio is fixed to be 2% and 10%. (Adult data with 32,561 private tuples.)

Effectiveness of Known Sample Attack backup

- Covariance Estimation Quality
  - Larger sample size gives attacker better recovery
  - Robust covariance estimator helps to downweight the influence of outliers
- p.d.f. of the Data
  - The greater the difference between any pair of eigenvalues of the covariance, the higher the probability of privacy breach
- More details can be found in the dissertation.

Independent Component Analysis

- Basic Model $Y_{k\times m} = M_{k\times n} X_{n\times m} = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \dots & \dots & \dots & \dots \\ m_{k1} & m_{k2} & \dots & m_{kn} \end{bmatrix} \begin{bmatrix} x_1(t_1) & x_1(t_2) & \dots & x_1(t_m) \\ x_2(t_1) & x_2(t_2) & \dots & x_2(t_m) \\ \dots & \dots & \dots & x_3(t_m) \\ x_n(t_1) & x_n(t_2) & \dots & x_n(t_m) \end{bmatrix}$
- ICA Estimation
  - To find a matrix W such that WY = WMX = X
- Nongaussian is Independent
  - Central limit theory - sum of random variables has a distribution closer to Gaussian than any of the original random variables.
  - ICA looks for a W that maximizes the nongaussianity of WY.
- Measures of Nongaussianity
  - Kurtosis: $\mathrm{kurt}(x) = \mathrm{E}[x^4] - 3\mathrm{E}^2[x^2]$
  - Negentropy: $\mathrm{J}(x) = \mathrm{H}(x_{gaussian}) - \mathrm{H}(x)$
  - Mutual information: $\mathrm{I}(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n} \mathrm{H}(x_i) - \mathrm{H}(x)$

Random Projection backup

Relative Errors in Computing the Inner Product of Two Attributes

k	Mean(%)	Var(%)	Min(%)	Max(%)
100(1%)	**9.91**	0.41	0.07	23.47
500(5%)	**5.84**	0.25	0.12	18.41
1000(10%)	**2.94**	0.05	0.03	7.53
2000(20%)	**2.69**	0.04	0.01	7.00
3000(30%)	**1.81**	0.03	0.27	6.32

Relative Errors in Computing the Euclidean Distance of the Two Attributes

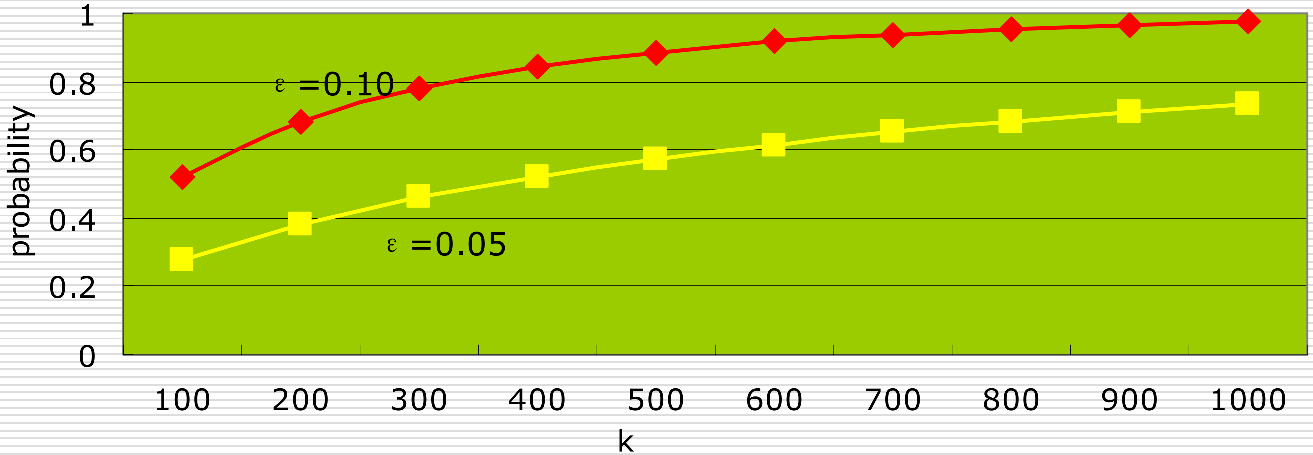
k	Mean(%)	Var(%)	Min(%)	Max(%)
100(1%)	**10.44**	0.67	1.51	32.58
500(5%)	**4.97**	0.29	0.23	18.32
1000(10%)	**2.70**	0.05	0.11	7.21
2000(20%)	**2.59**	0.03	0.31	6.90
3000(30%)	**1.80**	0.01	0.61	3.91

Adult data from UCI Repository. The first 10,000 elements of attributes *fnlwgt* and *education-num*.

Random Projection backup

$$\Pr\{(1-\eta)\|x-y\|^2 \le \|u-v\|^2 \le (1+\eta)\|x-y\|^2\} = \int_{k(1-\eta)}^{k(1+\eta)} f(t;k)dt,\ \eta>0$$

where $f(t;k)$ is the p.d.f. of chi-square distribution with k degrees of freedom.



The probability of the accuracy of random projection w.r.t. k and ε. Each entry of the random matrix is i.i.d., chosen from a Gaussian distribution with mean zero and constant variance.

The probability that relative error is bounded with in (1± η) increases proportionally with k.

Maximum a Posteriori (MAP) Test backup

- Given a binary hypothesis testing experiment with outcome *s*, the following rule leads to the lowest possible value of P_{ERROR}:

$$s \in A_0 \text{ if } \mathrm{Prob}\{H_0 \mid s\} \geq \mathrm{Prob}\{H_1 | s\}; \;\; s \in A_1 \text{ otherwise.}$$

- Here $P_{ERROR} = \mathrm{Prob}\{A_1 \mid H_0\}\ \mathrm{Prob}\{\mathrm{H}_0\} + \mathrm{Prob}\{\mathrm{A}_0 | \mathrm{H}_1\}\ \mathrm{Prob}\{\mathrm{H}_1\}.$

- The test design divides *S* into two sets, A_0 and $A_1 = A_0^c$. If the outcome s is in A_0, the conclusion is *accept* H_0. Otherwise, the conclusion is *accept* H_1.

MAP Known Input-Output Attack backup

$$\boxed{[Y_{k\times p} \quad Y_{k\times(m-p)}]} = \frac{1}{\sqrt{k}\sigma_r} R_{k\times n} [\boxed{X_{n\times p}} \quad X_{n\times(m-p)}]$$

KNOWN

- Assumption I: Attackers' best knowledge of f(X) is it is uniform.
- Assumption II: Attacker has no other background knowledge, *i.e.*, $\theta = \varnothing$.

$$\hat{x}_{MAP} = \arg\max_x f(\mathbf{x} = x \mid \frac{1}{\sqrt{k}}\mathbf{R}\mathbf{x} = y, \frac{1}{\sqrt{k}}\mathbf{R}X_p = Y_p) \ ,$$

$$= \arg\max_x \ (2\pi)^{-\frac{1}{2}k(p+1)} \det\left(\frac{1}{k}\overline{X}^T\overline{X}\right)^{-\frac{1}{2}k} etr\left(-\frac{1}{2}\overline{Y}(\frac{1}{k}\overline{X}^T\overline{X})^{-1}\overline{Y}^T\right),$$

$$where\ r_{ij} \sim N(0,1),\ \overline{X} = [x\ X_p], \overline{Y} = [y\ Y_p],\ \overline{X}\ has\ full\ column\ rank.$$