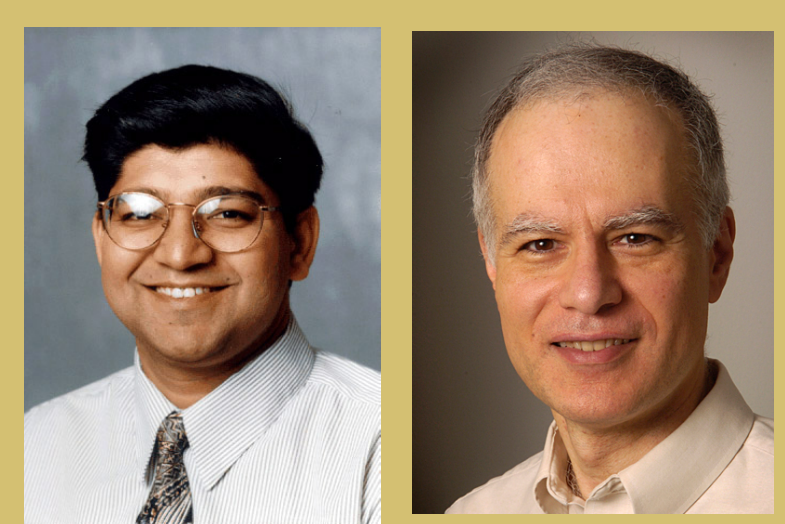


Watermarking Databases

S. Prabhakar, M. J. Atallah, Purdue University {sunil, mja}@cs.purdue.edu



Grant No: IIS-0242421

Problem

The critical need to exert digital rights over data is increasingly felt as ever-increasing amounts of valuable and proprietary data are exchanged with semi-trusted parties such as valid clients. Unlike media objects, databases do not have obvious encoding channels -- thus media watermarking methods are inapplicable.

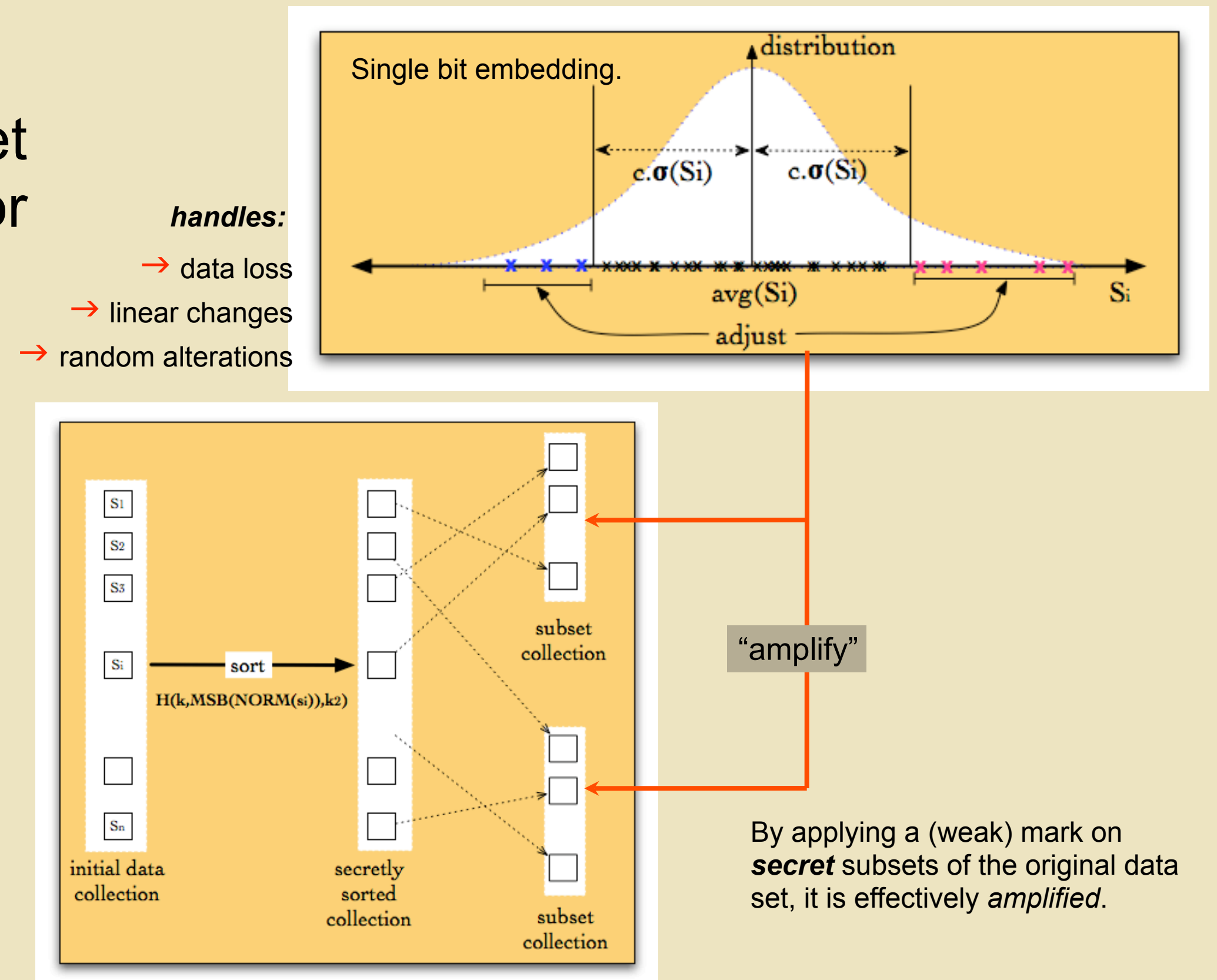
Acceptable change is highly application-dependent. Important attacks include scaling, sorting, sampling, random perturbations, substitutions, and summarizations.

Goal: The project's goals are to develop a set of robust, efficient watermarking techniques for a wide range of structured data sets.

Data *consumer* requirements define distortion metrics and associated bounds. Encoding only guarantees them.

Quality metrics are not hard-coded but rather separated from the watermarking method.

Unlike media watermarking methods, our methods take intended use as an input to the watermarking process.



Approach and Impact

Insight: Encode information in *global numeric properties* of *secret subsets* of the data while continuously evaluating data quality (backtrack if necessary).

- skew numeric distribution
- secrecy of subsets (mark amplification)
- error correction

Impact: Robust watermarking withstands:

- data loss, sampling
- linear changes, scaling
- random alterations
- substitutions

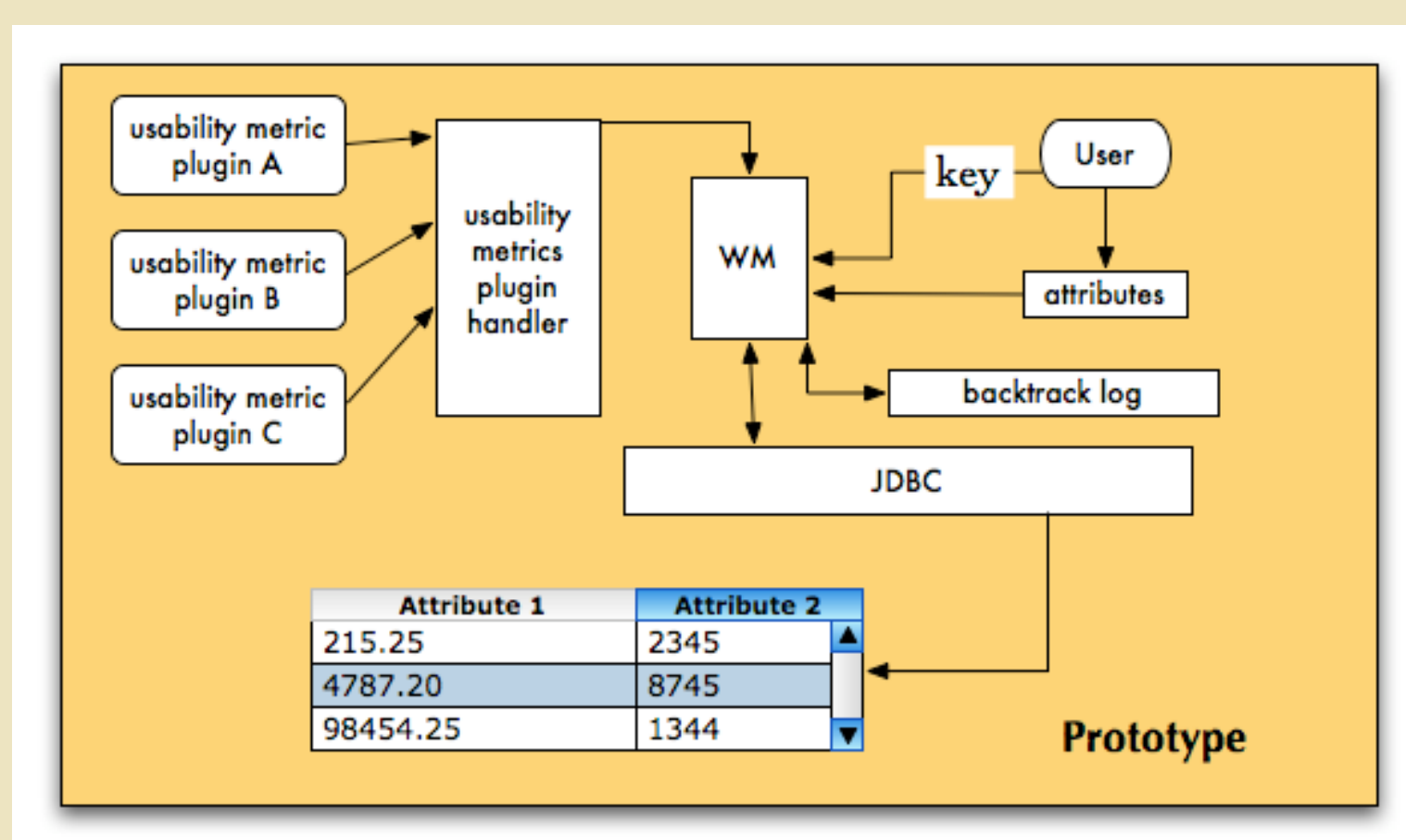
Status

1. Robust watermarking algorithms for:

- *Numeric relational databases* (shown above)
- *Categorical attributes* (embedding in inter-attribute correlation)
- *Streaming data* (embedding in extremities)
- *Semi-structured data* (XML) (embedding in structure and content of the XML document).

2. Prototype implementations (see left)

3. Patent Pending



- multi-threaded
- Java, tested with pgres/file-io/Oracle 9
- different JDBC drivers for different connections
- multiple databases at the same time, parallel watermarking

Selected References

1. **Rights Protection for Discrete Numeric Streams**, R. Sion, M. J. Atallah, S. Prabhakar, *IEEE Transactions on Knowledge and Data Engineering (TKDE)* Vol 18, No. 5 2006
2. **Rights Protection for Categorical Data**, R. Sion, M. J. Atallah, S. Prabhakar, *IEEE Transactions on Knowledge and Data Engineering (TKDE)* Volume 17, No. 7, July 2005
3. **Rights Protection for Relational Data**, R. Sion, M. J. Atallah, S. Prabhakar In *Proceedings of the ACM International Conference on Management of Data SIGMOD 2003*
4. **Resilient Rights Protection for Sensor Streams**, R. Sion, M. J. Atallah, S. Prabhakar, In *Proc. of the International Conference on Very Large Databases (VLDB), 2004.*