

Goal

Develop **risk assessment methods** to identify and classify **privacy leaks** in Google's Ads Data Hub.



ID	Age	Gender	Country	Language	C-ID	Minimal Sample Uniques (MSUs)
1	20-29	Male	India	English	1	(Age, Gender, Country, Language)
2	20-29	Male	India	Hindi	1	(Age, Gender, Language), (Gender, Country, Language)
3	20-29	Female	India	Hindi	1	(Age, Gender, Language), (Gender, Country, Language)
4	20-29	Male	USA	English	1	(Age, Country), (Country, Language)
5	20-29	Female	India	English	1	(Gender, Language)
6	30-39	Male	India	English	1	(Age, Country), (Age, Language)
7	30-39	Female	USA	Hindi	1	(Age, Gender), (Gender, Country)
8	30-39	Male	USA	Hindi	1	(Age, Gender, Country), (Age, Gender, Language), (Gender, Country, Language)
9	20-29	Male	China	English	1	(Country)

Table 1: Example of an underlying dataset and the computed Minimal Sample Uniques.

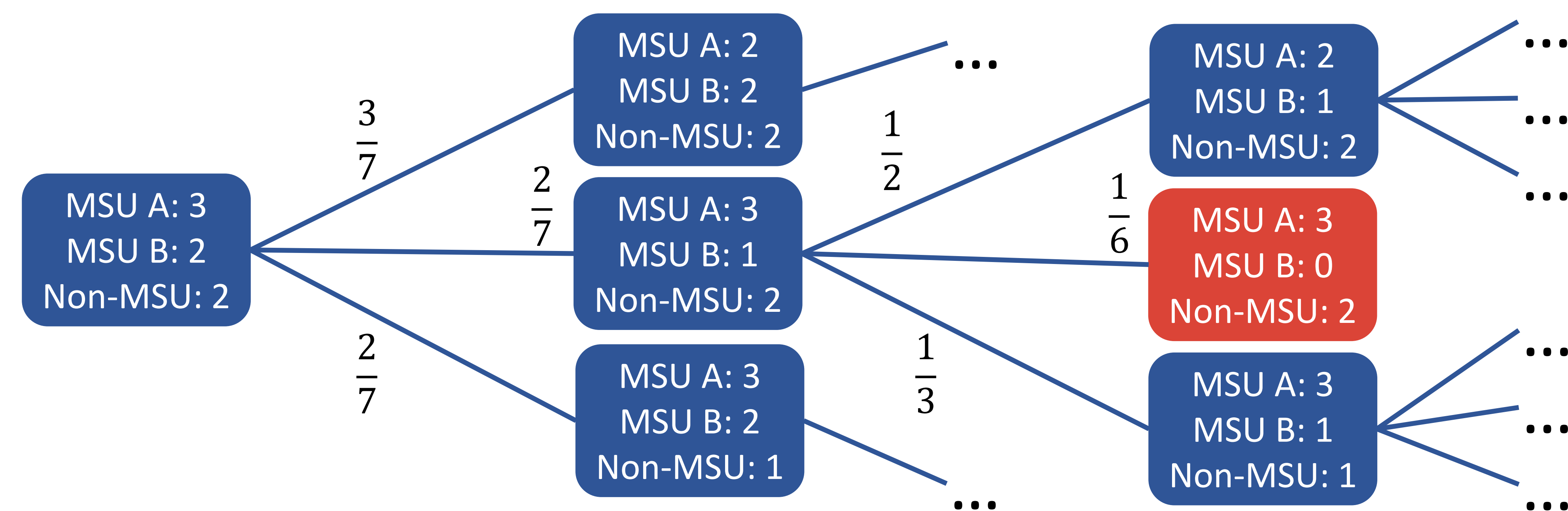
PIRATE Score

Probabilistic Identification Risk and Attacker Threat Estimate Score

The expected number of attributes that have to be revealed in order to complete a Minimal Sample Unique (MSU) for a row, that is:

$$\sum_{t \in T} Depth(t) \cdot \prod_{p \in P(t)} p$$

where T is the set of all terminal nodes and $P(t)$ is a list of the probabilities from the root node to t in a probability tree diagram.



Results

Attribute	Trial 1	Trial 2	Trial 3	Trial 4
Domain	49.2%			
Campaign	3.4%			
Browser	3.8%			10.9%
Language	43.6%		51.1%	45.3%
Gender		29.5%	7.0%	5.4%
Age		34.8%	15.1%	12.6%
Country		0%	0%	0%
Region		0%	6.2%	6.0%
Metro		35.9%	20.7%	19.8%

Table 2: Column contributions for the attributes in each trial.

Trial	Number of attributes	Rows with MSUs	Average PIRATE Score
1	4	54.5%	2.41
2	5	0.4%	4.42
3	6	77.5%	4.41
4	7	85.5%	4.99

Table 3: Average PIRATE Score on test sets.

Conclusion

We developed and tested a risk measure that quantifies the risk of divulging users' information. Our PIRATE score is independent of the size of the underlying dataset, so it can be compared across different queries.

Acknowledgements

Many thanks to Scott Schneider and Vardan Akopian from Google and Susana Serna, Bao Wang and the staff from IPAM for their support.