FairTest: A Service for Uncovering Privacy Bugs in Data Driven Applications

V. Atlidakis and Derek X. He

Motivation

- Data-driven applications collect and process vast amounts of data
- Little has been done to allow developers understand the use of those data
- Lack of transparency may lead to privacy implications and discrimination treatment of users

Motivation

- Websites vary prices based on certain user’s information, such as user’s location
- Residents of a borough consistently receive higher prices than residents of another.
- This may lead into situations discriminatory behavior based on sensitive attributes (e.g., race or income)

Motivation

Staples example: Differential treatment of users

- Discovered by Wall Street Journal investigators in 2012 [1]
- Staples Inc. displayed different prices to people after estimating their locations
  - Consider the person's distance from a rival brick-and-mortar store.
  - If rival stores within 20 miles or so, showed a discounted price.
  - Areas that tended to see discounted prices had higher avg. income than areas that tended to see higher price

Related Work

- Define statistical parity of individuals within a set
- Study fairness on the treatment of individuals
  - [1] Dwork et al: “Fairness Through Awareness”

    Two individuals who are similar with respect to a particular task should be classified similarly
  - [2] Friedler et al: “Certifying and removing disparate impact”

    Removing Disparate impact and target on neutral selection of individuals

Statistical Parity

For two sets of users S, S’ and an output O, the users of set S are equally probable with the users of set T to see output O.

Equivalently:

$$|Pr\{O \mid x \in S\} - Pr\{O \mid x \in S'\}| \leq \varepsilon \quad (1).$$

Applying Statistical Parity

An example with Discriminatory behavior

Population	#Members	Price		Statistical Parity (for high price)
		Low	High	
A	30	10	20	0.198 = |20/30 - 33/70|
B	30	16	14	0.09
C	40	21	19	0.09
Total	100	47	53	-

Relaxing Statistical Parity: Introducing Business Necessity

- There are cases of reasonable discrimination, a.k.a. business necessity or requirement
- Let users be discriminated on two sets: users in R meet a requirement and users in R' not.
- Apply statistical parity separately on R and R'

$$|P\{O \mid x \in S \cap R\} - P\{O \mid x \in T \cap R\}| \leq \varepsilon \quad (1),$$

$$|Pr\{O \mid x \in S \cap R'\} - Pr\{O \mid x \in T \cap R'\}| \leq \varepsilon \quad (2)$$

Applying Relaxed Statistical Parity

An example with business necessity

Credit History	Loan Type (Population A) Payday	Personal	Total	Loan Type (Population A) Payday	Personal	Total
YES	5	15	20	15	40	55
NO	80	0	80	45	0	45
Total	85	15	100	60	40	100

Naive approach:

- 25% more users of A than B receive Payday loans.
- Statistical parity condition $|Pr\{O \mid x \in S\} - Pr\{O \mid x \in S'\}|$ yields: $|85/100 - 60/100| = 0.25$

Applying Relaxed Statistical Parity

An example with business necessity (cont'd)

Credit History	Loan Type (Population A) Payday	Personal	Total	Loan Type (Population A) Payday	Personal	Total
YES	5	15	20	15	40	55
NO	80	0	80	45	0	45
Total	85	15	100	60	40	100

Relaxed statistical parity approach:

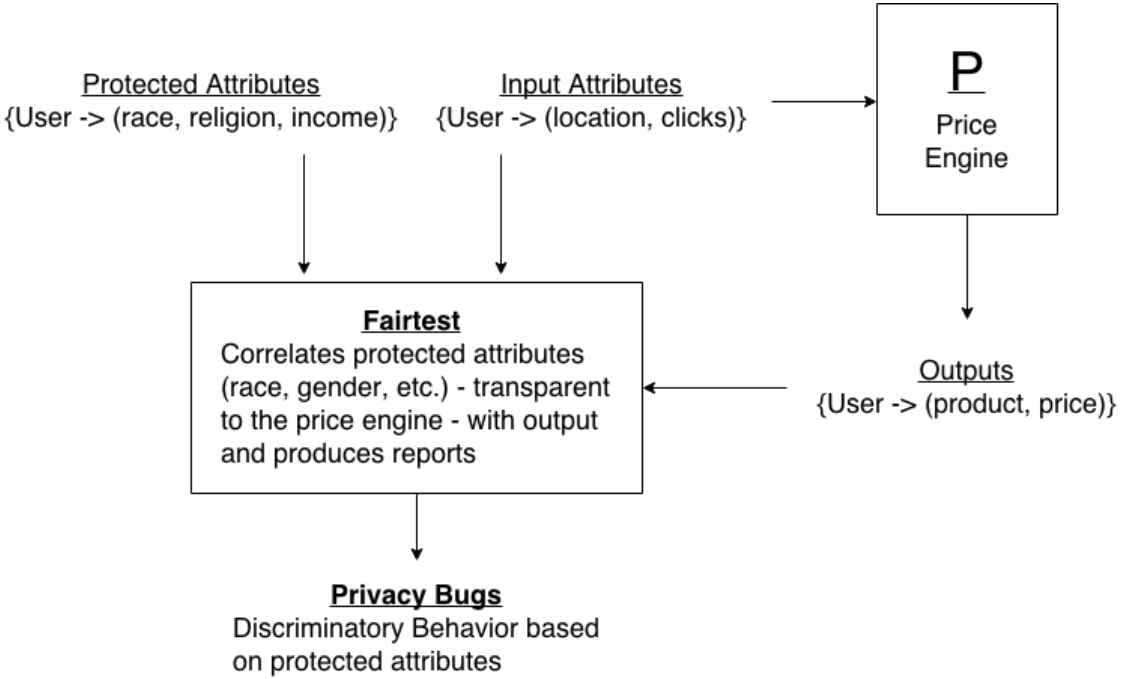
- $|P\{O \mid x \in S \cap R\} - P\{O \mid x \in T \cap R\}| \leq \varepsilon$ (1)
  $|P\{O \mid x \in S \cap R'\} - P\{O \mid x \in T \cap R'\}| \leq \varepsilon$ (2)

- Condition (1) yields: |5/20 - 15/55| = 0.02
  Condition (2) yields: |80/80 - 45/55| = 0

Fairtest

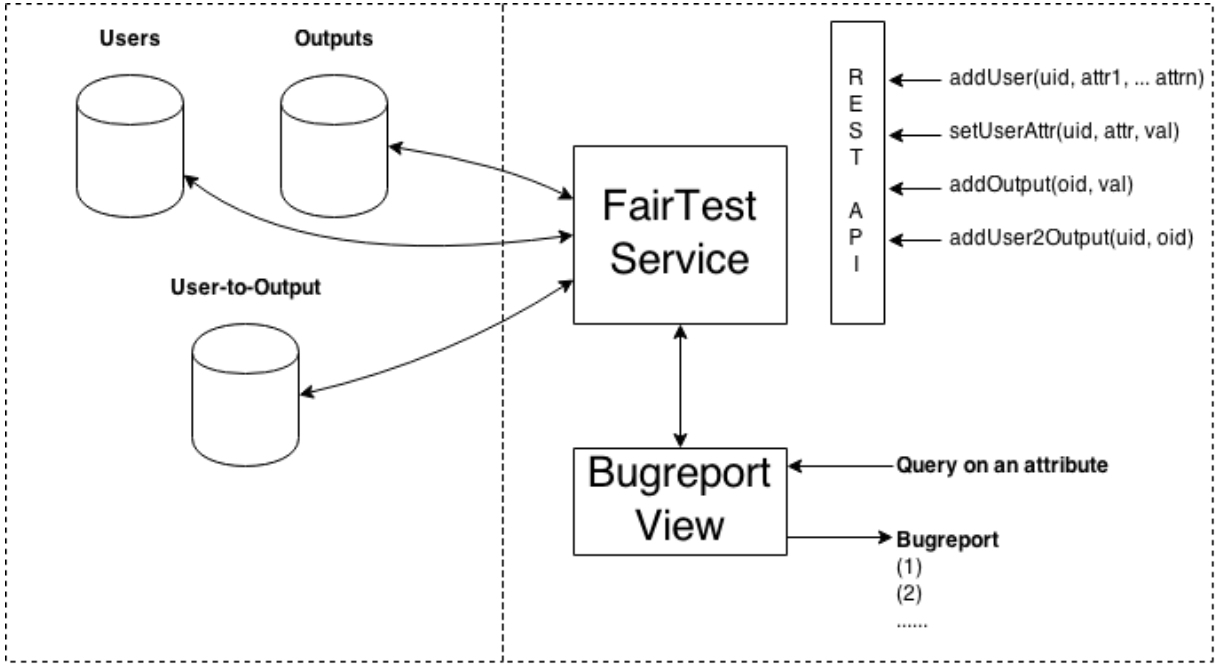
- Service for data-driven applications, such as online stores
- Uncover privacy bugs that lead to discriminatory behavior of populations
- Find correlation between output and “protected attributes”
- Applications use it through a REST-ful API

Fairtest - Overview



Fairtest Data Flow

Fairtest - Architecture



Evaluation

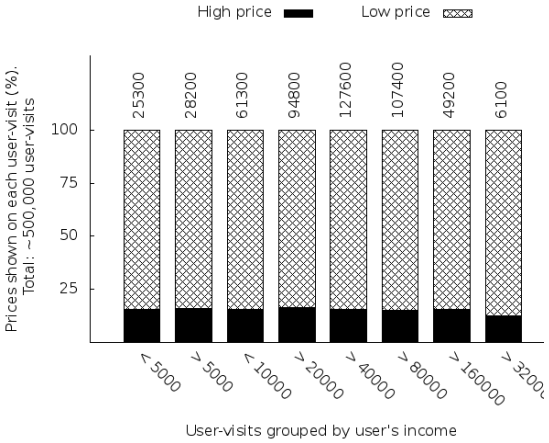
- 5,000 randomly generated users
  - Using U.S. census data to simulate the actual demographics in and between zip codes
- A sample price engine
  - Biased on proximity of user's location with certain actual locations:
    - The business (e.g. Staples stores)
    - Competitors (e.g. Office Depot stores)
- Measurements
  - Deployed Fairtest to the price engine, which returns a "low" or "high" price with a probability set according to the location data

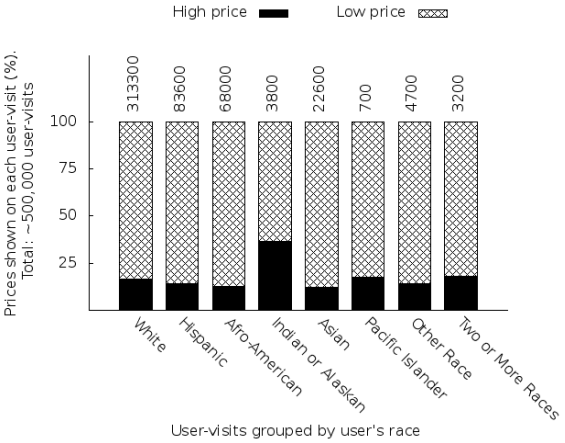
Results

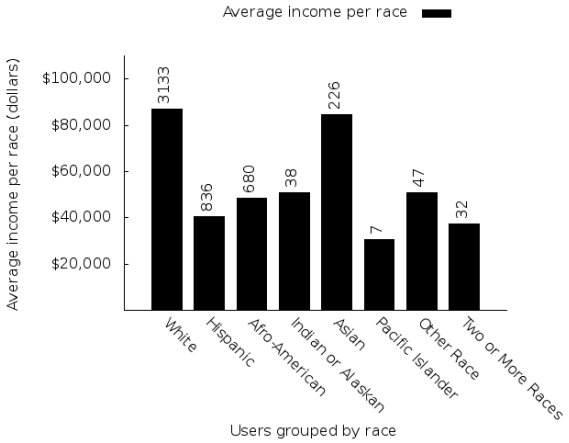
Figures: Statistical parity and its dependency on user's location. Shows the dependency of statistical parity, i.e., number of samples that violate condition 1, as a function of (a) user's income, (b) user's race, and (c) user's sex.

Conclusion: Statistical parity on user-races correlates with the dependency of the price engine on user's location, but not on user-income and user-sex.

Results (cont'd)







Left, Center: Proportions of groups that see the high versus low price. It is significant in the test data result that Indian or Alaskan population see high prices significantly more. Also persons with >USD $320000 income are the least likely group to receive high price.

Right: U. S. average income based on race. Despite Indian or Alaskan population receives higher price, they have relatively lower average income.

Conclusions & Future Work

- We built and evaluated a prototype of FairTest: an service to uncover and report discriminatory behavior of populations on certain protected attributes

---

- Repeat experiments at larger scale in terms of users (~100,000 users)
- Study different price engines
- Apply mixed models [3] to evaluate statistical parity

References

[1] Dwork et al: “Fairness Through Awareness”

[2] Friedler et al: “Certifying and removing disparate impact”

[3] D. Hsu et al: Learning Mixtures of Spherical Gaussians: Moment Methods and Spectral Decompositions

Backup Slide 1