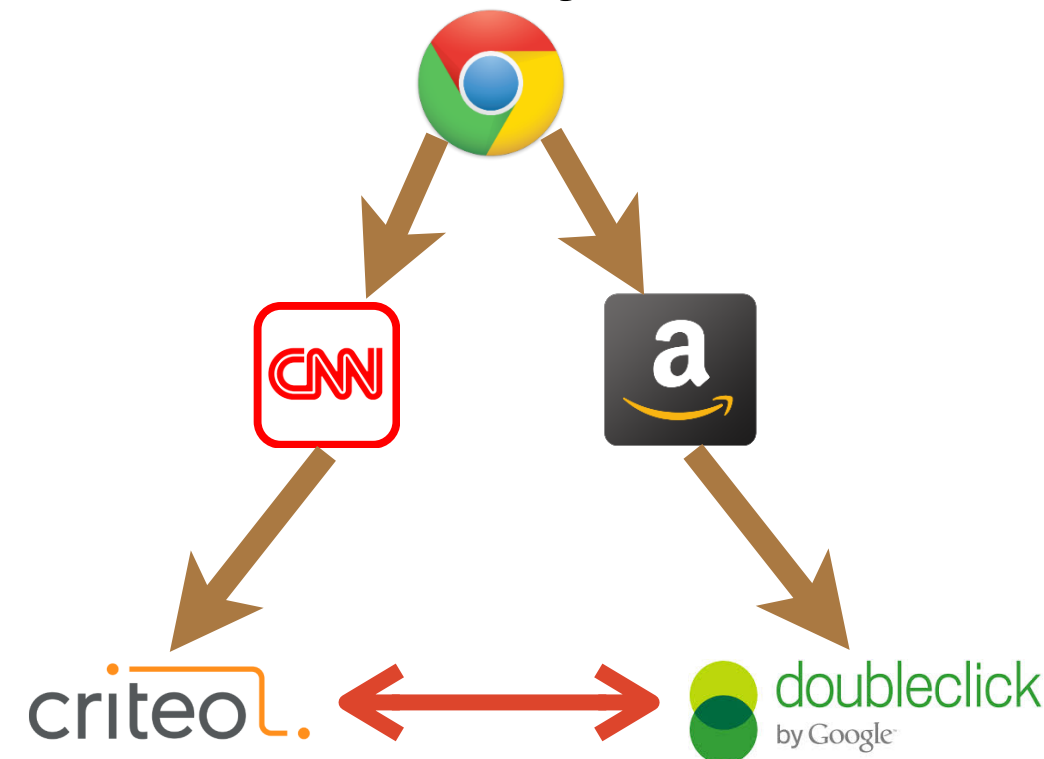


Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

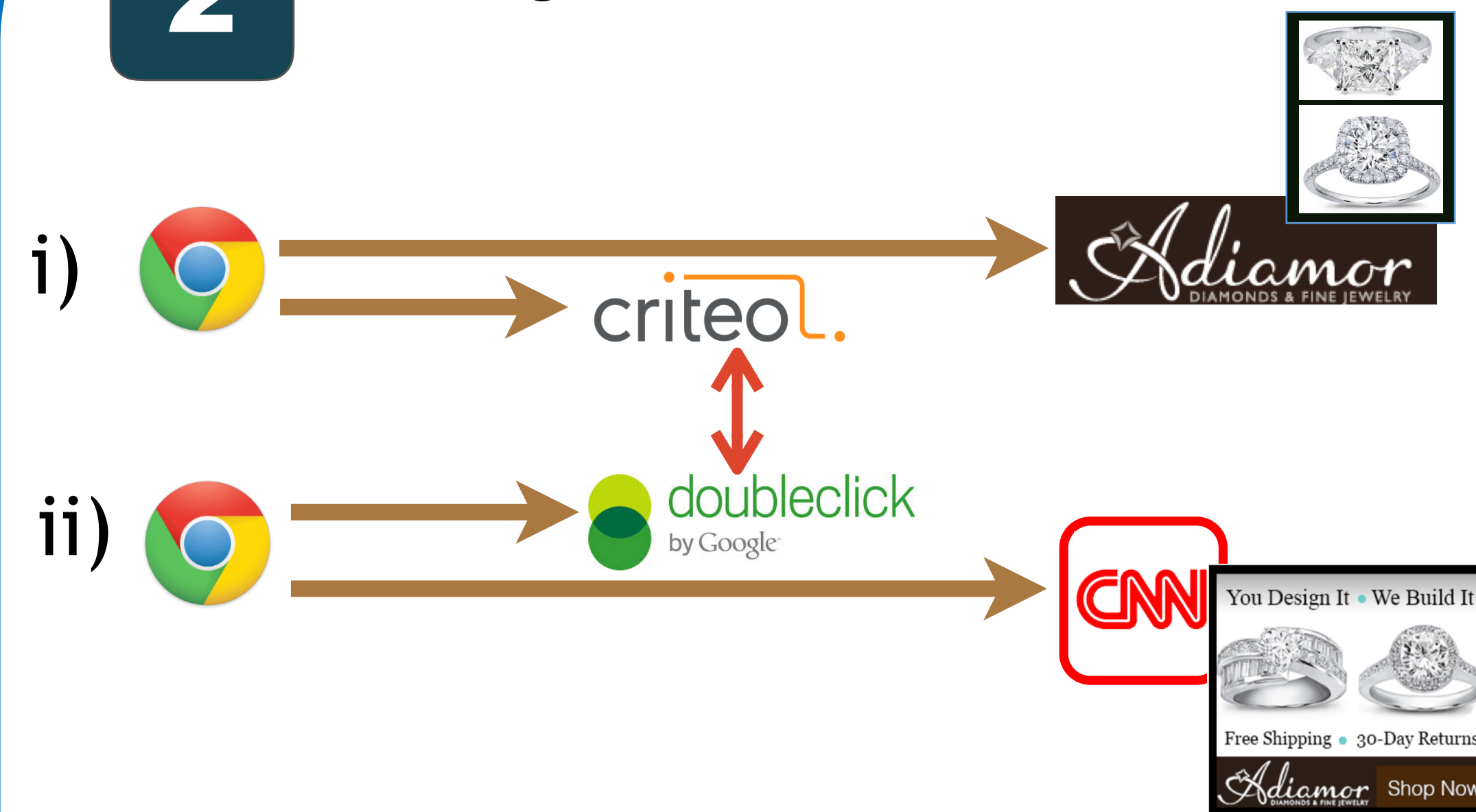
Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, Christo Wilson

1 Problem

- Users are tracked by ad networks over the Web (brown edges).
- Ad networks share that tracking information (red edges) with each other.
- **Goal:** To identify the red edges.

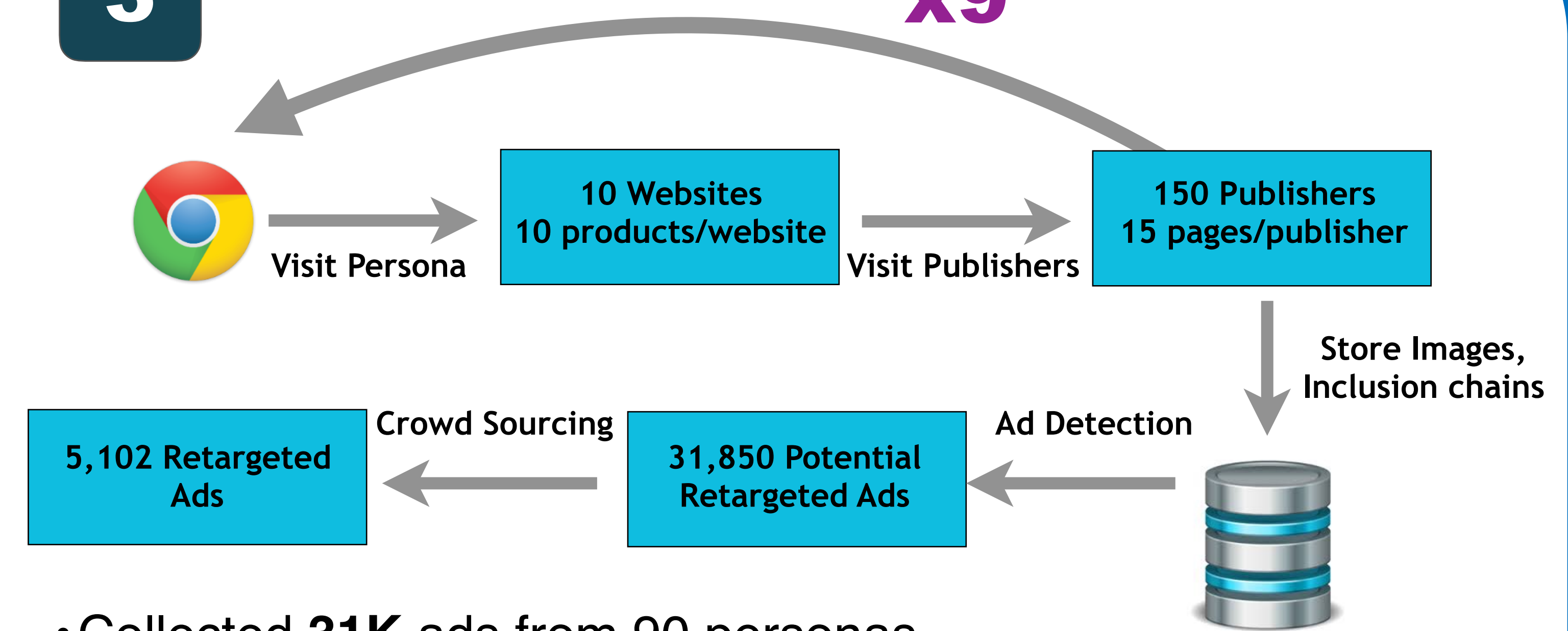


2 Retargeted Ads



Key Idea: Use retargeted ads to conduct controlled experiments.

3 Data Collection



- Collected **31K** ads from 90 personas.
- Labeled all ads using AMT workers.
- **5102** are unique retargeted ads.

4 Detecting Information Sharing

Classification	Shopper Side	In Between	Publisher Side
Trivial Matching	 $\wedge \text{shop} \rightarrow \cdot \rightarrow a \rightarrow \cdot * \$$		 $\wedge \text{pub} \rightarrow a \$$
Cookie Matching	 $\wedge \text{shop} \rightarrow \cdot \rightarrow a \rightarrow \cdot * \$$	 $\wedge * \rightarrow \cdot \rightarrow e \rightarrow a \rightarrow \cdot * \$$	 $\wedge \text{pub} \rightarrow \cdot \rightarrow e \rightarrow a \$$
Latent Matching (Server Side)	 $\wedge \text{shop} \rightarrow [^ea] \$$		 $\wedge \text{pub} \rightarrow \cdot \rightarrow e \rightarrow a \$$

5 Classification Results

- Cookie Matching is most common.
- We identified **200** pairs of collaborating ad networks.
- **31%** of them cannot be detected using prior detection methods.

Type	Raw		Clustered	
	Chains	%	Chains	%
Direct (Trivial) Match	1770	5	8449	24
Forward Cookie Match	24575	69	25873	73
Backward Cookie Match	19388	55	24494	70
Indirect Match	2492	7	178	1
Latent Match	5362	15	343	1
No Match	775	2	183	1

6 Future Work

- Many people use ad blockers to avoid creepy ads.
- Sadly, ad blockers kill revenue for small publishers.
- We are working on a way through which one can see clean ads while avoiding such information sharing.

