Synthoid: Endpoint User Profile Control

Marcel Flores and Aleksandar Kuzmanovic
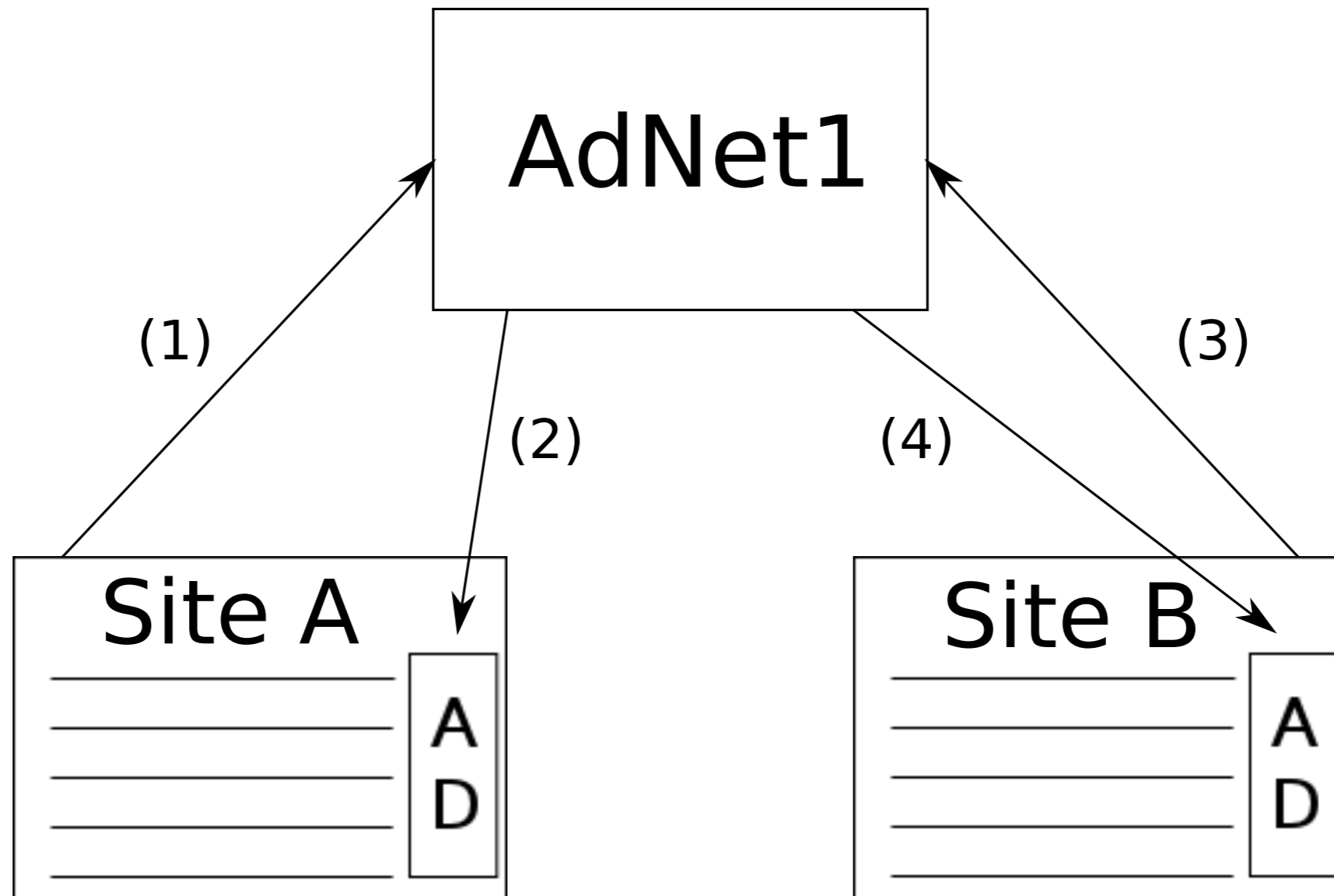
WI 2014

Tracking Background

- Large scale advertising offers fresh vantage point on user behavior.

- Trackers can measure users across sites,
  - Construct interest profiles for users.
  - Deliver of targeted ads.

Tracking Background

Existing Approaches

- Block or disrupt the ad interaction

- Privacy preserving infrastructures

- Do Not Track, Opt-out mechanisms

Synthoid

- Return power over user profiles to the user.

  - No cooperation from trackers.

- Control the signal that advertisers measure:

  - Provide synthetic signal.

  - Consistently and regularly visit sites of specific topics which include tracking ads.

Goals

- Influence the user's advertising profile.

- Hide a user's behavior amongst synthetic interests chosen by the user.

- Do so generically for all trackers and tracking methods.

Synthoid

- User specifies a set of topics.

- Synthoid browses websites of these topics,

  - Performs usual cookie transaction.

  - Ad loads inform trackers of topics of visited sites.

Browsing

- Want to generate meaningful traffic:

  - Draws sites from Open Directory

  - Human-like diurnal behavior

  - Loads a site, follows 4 links

- Can be entirely configured by users.

- Directly uses the user's browser via Selenium

Synthoid
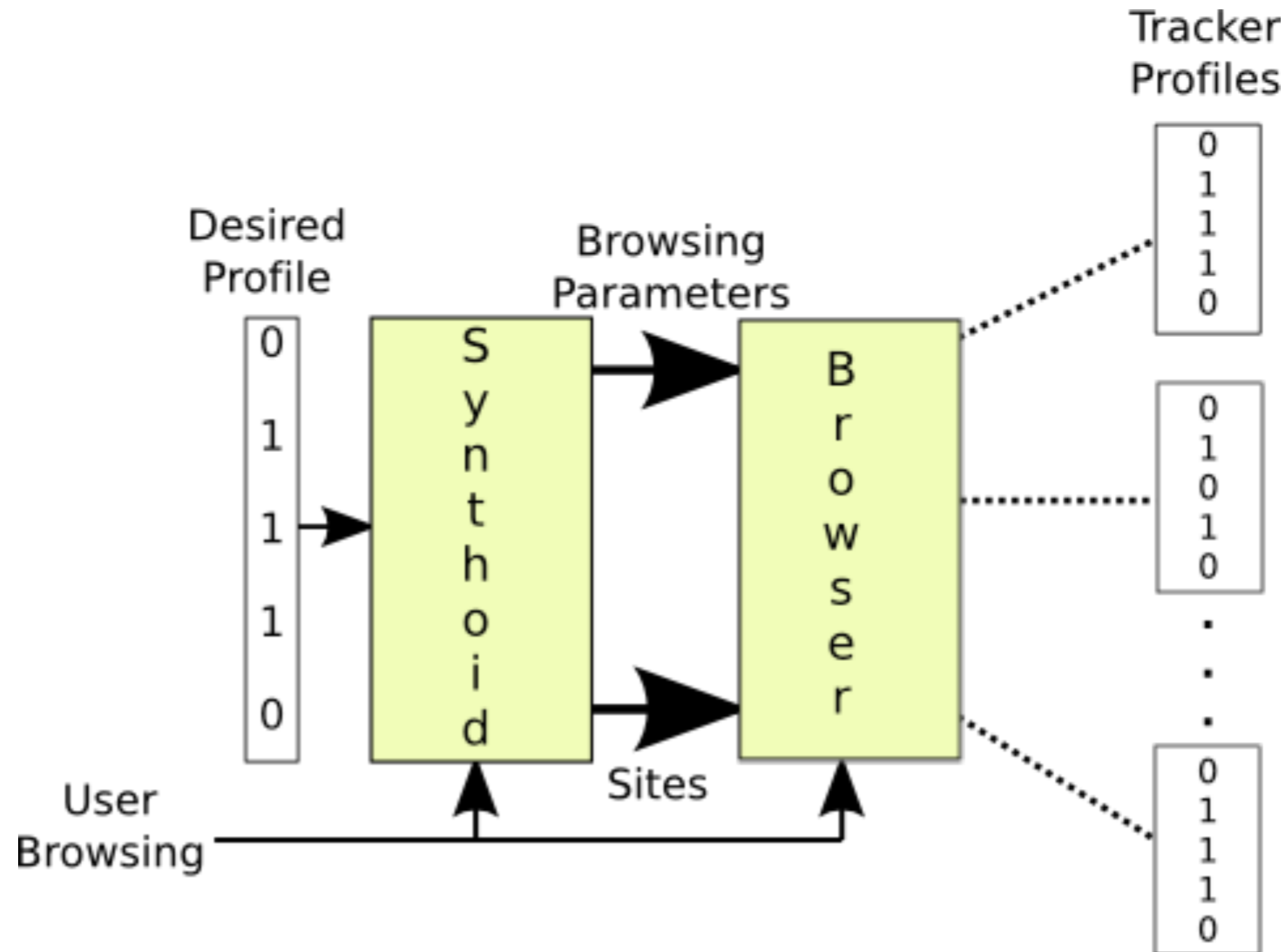
Tracker Feedback

- Require feedback to measure our performance.

  - DoubleClick, Yahoo, BlueKai make profiles available.

- We select DoubleClick.

  - Largest and most influential

Scoring System

- Consider vector space where each dimension is a topic.

- Generate vector from observed profile:

  - 1 if topic-dimension present,

  - 0 otherwise.

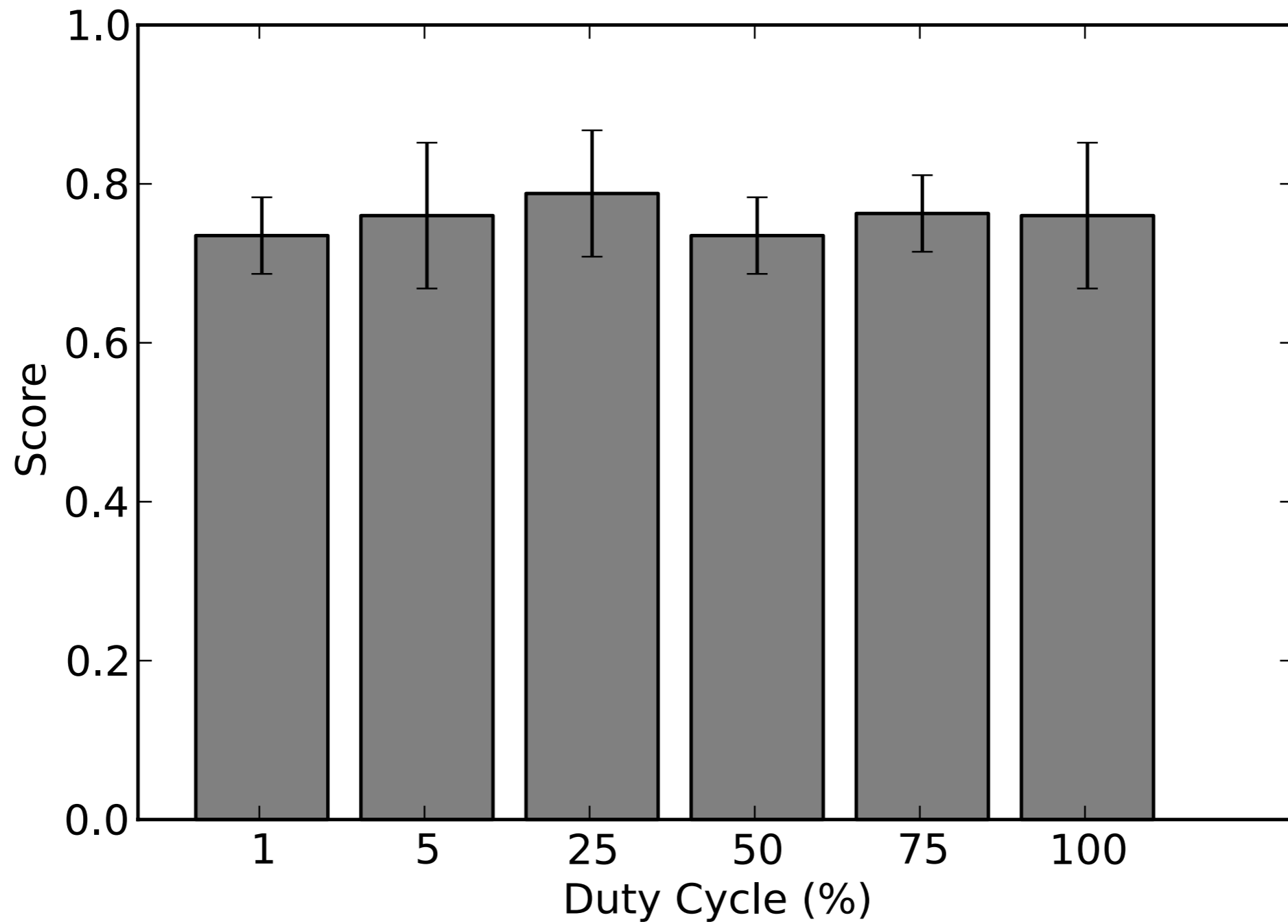- Compute cosine similarity with unit vector.

Evaluation

- Choose a random sample of 10 topics.

  - Use the same topics for duration of experiments.

- Run Synthoid on a fresh cookie for 7 days.

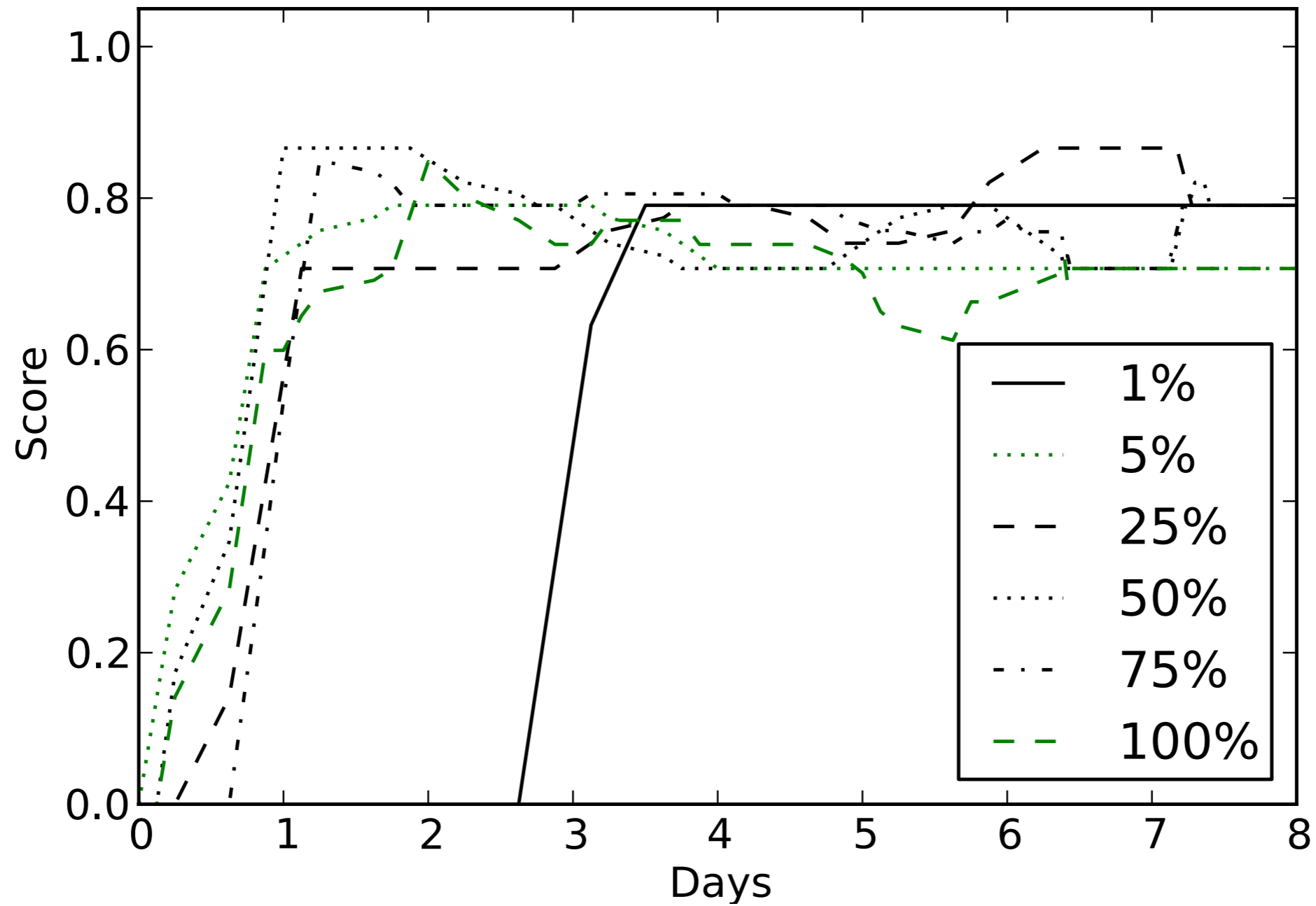- Observe the profile at regular intervals.

Volume

- How does changing the total traffic volume of the system affect its ability to imprint a profile?

- Vary duty-cycle from 1% to 100%.

Volume

Volume

Other Analysis

- Size of the pool of sites used

  - Controls number of repeats

- Interference

  - Volume Dependent
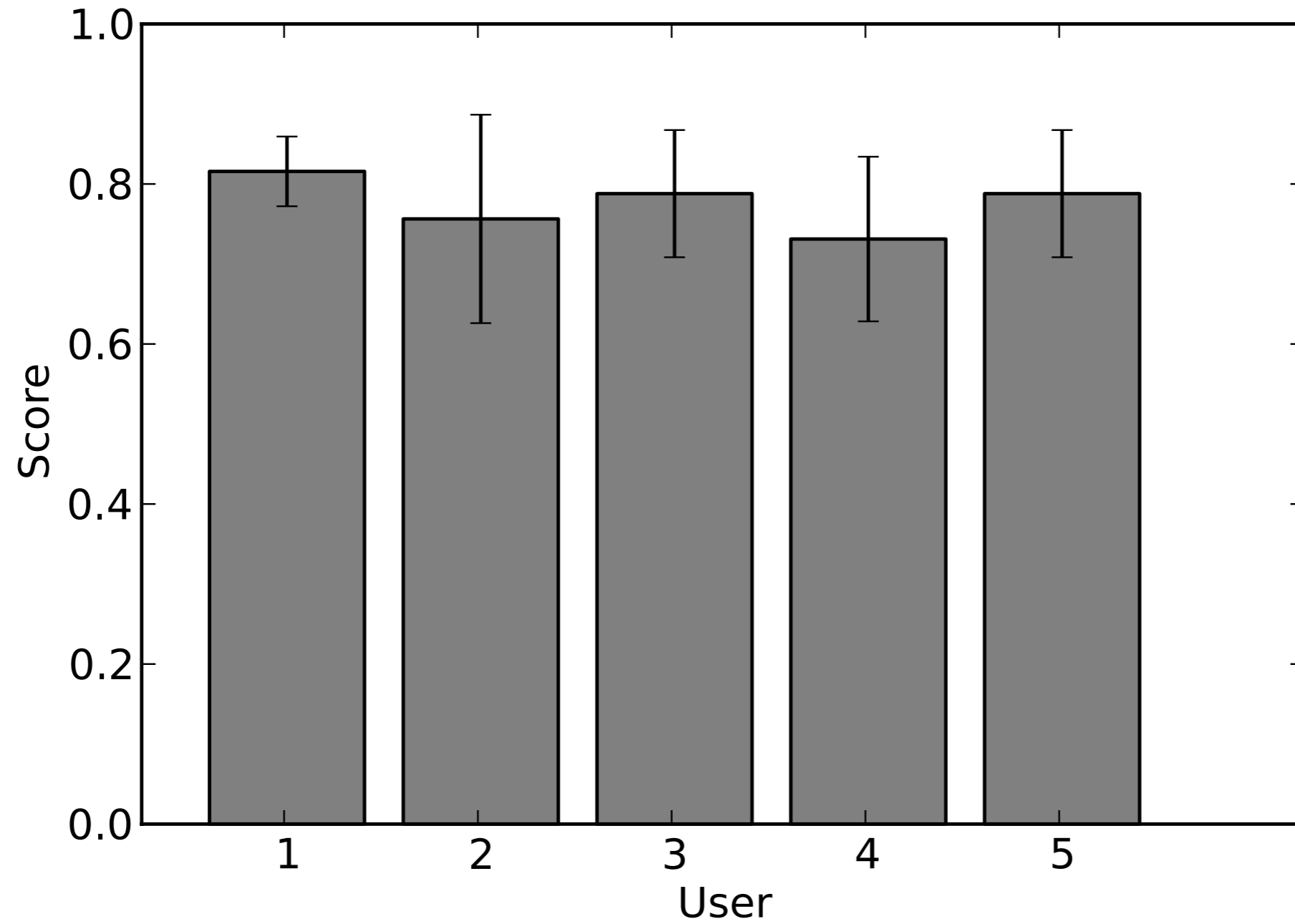
  - Volume Independent

Case Studies

- Collected week long traffic traces from 5 individuals.

- Recreated each trace with Synthoid running at 25% duty-cycle.

- Also ran separate control runs of each human trace.

Case Studies

User	Number of Pages	Unique Domains
1	2019	251
2	559	92
3	1031	186
4	1772	120
5	2369	147

Case Studies

Case Studies

- No overlap between user's control profiles and profiles with Synthoid.

    - Except where desired profile overlapped.

- Original profile was entirely obscured.

Generalizability

- Yahoo - Generally performed well.

  - Had difficulty with certain topics, suggest covers different topics from DoubleClick.

- Blue Kai

  - Much smaller profiles, suggests narrower scope.

  - Still performed well.

Generalizability

- Endpoint design makes it compatible with any trackers it encounters

- Trackers still have a total view of information.

  - Can completely alter profiles.

- Cooperates with fingerprinting techniques, as traffic comes from the user.

Conclusions

- Demonstrated ability of Synthoid to imprint profiles with user preferences.

- Effectively hid user interests with selected topics.

- Demonstrated simultaneous functionality across multiple trackers.

Thank you!

Scoring System

- Consider the cosine similarity of these two vectors:

  - Increased similarity indicates more matching topics (i.e. target matches observations).

- Ignores topics in observed profile not in target profile.

Scoring System

- Build 2 binary vectors

  - Input: each dimension has a value 1

  - Output:

    - 1 if that topic-dimension appeared

    - 0 if it did not

- Score is then the cosine similarity.

Scoring System

Input:
/Art/Movies/Action
/Science/Biology
/Sports/Soccer

Output:
Art - Movies - Martial Arts
Science - Bio - Anatomy
Travel - Destinations - Parks

Topic	Vector
Arts & Entertainment	1
Science	1
Sports	1

Topic	Vector
Arts & Entertainment	1
Science	1
Sports	0