Under the Spotlight: Web Tracking in Indian Partisan News Websites

Vibhor Agarwal[1], Yash Vekaria[1], Pushkal Agarwal[2], Sangeeta Mahapatra[3], Shounak Set[2], Sakthi Balan Muthiah[1], Nishanth Sastry[4], Nicolas Kourtellis[5]

[1] The LNM Institute of Information Technology, India, [2] King's College London, UK, [3] German Institute for Global and Area Studies, Germany, [4] University of Surrey, UK, [5] Telefonica Research, Spain

Introduction & Motivation

- Majority of the privacy studies are confined to the western world. India, a large market which covers 20% of the world's population, is not studied yet.
- Absence of privacy laws in India allows third-parties to track users intensely to show targeted ads.
- Targeted ads can be used to change the political views and behaviour of the users.
- The paper performs, to the best of our knowledge, the first comprehensive analysis on the Indian online news media with respect to tracking and partisanship.
- We build a dataset of 103 Indian news websites, labelled by two experts, as Left, Right, or Centre websites.
- Measuring the invasive and advanced tracking strategies demonstrates the dire need of privacy laws in India to protect users from being tracked by "The Wild Tracking East".

Methodology
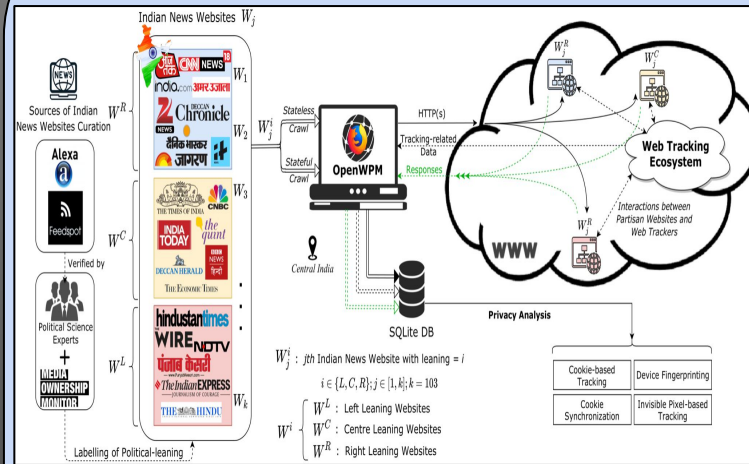
- The dataset of 103 Indian news websites is labelled as Left, Right, or Centre websites by two experts.
- To collect websites' data, **OpenWPM** is used for crawling.
- We study and compare user tracking in 103 news websites with the following metrics:
    - **Cookie-based tracking**
    - **Cookie Synchronization**
    - **Device Fingerprinting**
    - **Invisible Pixel-based tracking**

Results: User Tracking vs. Partisanship

- The 103 news websites studied have more than **100K cookies**, for an average of over 100 cookies per website.
- Several websites like *Sandesh.com* have **~1400 cookies** set by this first-party and its third-parties.
- Left and Centre-leaning websites tend to serve **more** cookies than Right.
- **Desktop versions** of websites set **more** cookies than their **mobile versions**, with interesting exceptions.
- Around 18% of all distinct TPs, and 25% of all distinct FPs in our data are involved in **cookie synchronization**.
- Over 25% of news websites use **device fingerprinting**, which is invisible to the user and invasive to their online privacy.
- Around 25.7% of Left, 23.7% of Right, and 17.9% of Centre websites employ different **fingerprinting scripts** to track users.
- More than **2.5K invisible (1x1 pixel) images** (i.e., 23% of all sent images) are detected on news website homepages.
- **Invisible pixel-based tracking** is employed **more** by Centre, followed by Left, and then the Right websites.