The Case For a WYSIWYG Mobile Browsing Panel

Markus Mobius
Microsoft Research

Tanya Rosenblat
University of Michigan

Raj Bhargava
Microsoft Research

Michael Butler
Microsoft Research

November 2, 2018

Abstract

Mobile devices are quickly becoming the main gateway to news consumption for Americans. We estimate that between 41-56% of overall news consumption (across both desktop and mobile devices) is conducted through mobile news and social media applications rather than through browsers. This app traffic is difficult to measure using current methods that intercept traffic between mobile devices and the internet due to apps caching data and in some cases hosting publisher content on their own servers entirely (such as Facebook Instant Articles). We argue that there is a need to develop a new type of mobile browsing panel for social scientists that overcomes the limitations of existing techniques. We propose a system that instead captures news consumption at the *screen level* by intercepting a user's interaction with news content on *any* app. This WYSIWYG ("what you see is what you get") approach to news consumption also captures further user-level metadata and behaviors that are typically only available through eye-tracking studies such as scrolling through news snippets and headlines without clicking or otherwise interacting with them.

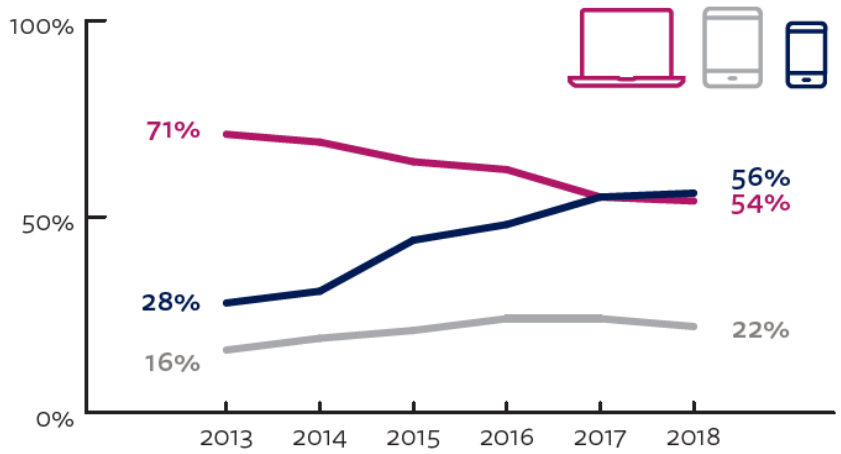
1 Introduction

The internet has fundamentally changed the news industry. Until the 1980s most medium-sized towns had local newspapers and most Americans could choose from at most three evening network news broadcasts, each presenting a seemingly neutral take on the daily news (Duca and Saving, 2017). Today, news distribution is no longer limited by geography, local newspapers are dying, and there are a large number of national online publishers and TV channels catering to every political taste. This has raised concerns that people on either side of the political spectrum not only interpret facts differently but listen to different facts altogether. Hence, there is a renewed interest in measuring how people consume news.

Measuring news consumption is relatively easy on desktop browsers because content is loaded whenever a user interacts with a page (such as clicking on a link). However, passive traffic collection on mobile devices is much harder because a lot of news content is consumed in mobile apps – both first-party publisher apps (such as the Wall Street Journal or the New York Times apps) and third-party apps such as Facebook. These apps often cache data that might only weakly correlate with a user's actual consumption. Third-party apps can also host publisher content directly (such as Facebook Instant Articles) and the corresponding publisher URLs will not show up in a user's traffic stream.

This mobile measurement gap is already large and growing. In 2017, 85% of U.S. adult accessed news on a mobile device at least once[1]. Figure 1 illustrates that in 2018, for the first time, U.S. news audiences are more likely to use mobile phones to regularly consume news than they are laptop and desktop devices.

Figure 1: Devices used for primary news access (2013-2018)



Source: Newman et al. (2018; accessed November 2, 2018, page 113). News consumption on laptop/desktop/smartphone is indicated by purple/gray/blue line.

The increase in mobile news consumption has been driven by older Americans as well as

[1]See Lu (June 12, 2017; accessed November 2, 2018)

those living in low-income households, as consumption statistics for these aforementioned groups begin to match those of younger and wealthier audiences, respectively.[2] Thus in the past few years, digital news browsing has become increasingly synonymous with mobile news browsing.

We propose a new approach to passive data collection on mobile devices that we call the WYSIWYG ("what you see is what you get") approach to mobile user panels. Our method aims to look over the shoulder of users and captures how they interact with news content on their mobile screens. It is agnostic to the type of app that runs in the foreground: it could be a browser, a publisher's first-party app, or the Facebook app. Our approach not only addresses the key shortcoming of current passive data collection on mobile devices but also allows us to measure new metadata on how people interact with news content that is typically only available in eye-tracking studies.

The approach of this paper is as follows. Section 2 discusses the technical shortcomings of existing passive news traffic measurements on mobile devices. Section 3 reviews existing data sources and how they have been used in recent research. Section 4 explains a system that implement WYSIWYG approach to user panels.

2 Challenges in Passive Mobile News Data Collection

Before the extreme shift towards mobile news consumption, researchers could passively track the full scope of online news consumption using browser logs. Browser manufacturers such as Google and Microsoft natively collect such data for consenting users who, e.g., install their respective browser toolbars or use certain features of their software. Moreover, market research companies such as Nielsen and comScore have collected browser logs from representative user samples for many years.

A browser log has a simple structure: it consists of a sequence of URLs with time stamps that corresponds to the way the user navigates the web. For example, consider a user who visits the New York Time homepage and from there clicks on the business landing page:

- The browser log adds a row with URL `https://www.nytimes.com` and the current time.
- When she then clicks on the link to the business landing page of the New York Times a new row with URL `https://www.nytimes.com/section/business` is added.
- The time difference between these two consecutive time stamps allows us to infer the user's dwell time on the home page.[3]
- The browser log entry immediately preceding the visit to the NYT homepage helps us

[2] See Lu (June 12, 2017; accessed November 2, 2018)

[3] Many browsing logs, such as the Microsoft toolbar log, also keep track of browser tabs and include referrer URLs which allow researchers to reconstruct non-linear browsing behaviors (i.e., where a user switches between tabs while browsing).

determine whether the news visit started through direct navigation, a news aggregator such as Google News or Yahoo News, a search engine, or social media.[4]

However, browser logs are insufficient to capture mobile news consumption because a large share of news consumption is happening outside of mobile browsers through mobile applications. First, most publishers offer their own native apps that allow for offline reading and easy subscription management. Second, news aggregators such as Google News, Apple News, and Android's Google Feed offer their own third-party apps. Finally, third-party social media apps often provide their own news delivery mechanisms, such as Facebook Instant Articles, that offer faster load times to publishers while keeping users engaged inside their own applications.

In the following, we explain why mobile browsing traffic is harder to measure than desktop browsing traffic, why app traffic is less useful even when measured correctly, and why our failure to measure app traffic matters.

2.1 Measuring Mobile Browsing Traffic

On desktops, browsing traffic is either measured natively by the vendor or through extensions and plugins (such as the Google or Bing toolbar). In contrast, mobile browsers generally do not allow extensions or plugins for performance reasons.

This requires us to instead measure browsing traffic on the device level. The current default technique is to ask the user to install a VPN (virtual private network) app and a custom CA (certificate authority). We will refer to this technical solution as the VPN/CA method from now on. The VPN intercepts and eavesdrops on all traffic between the device and the internet. This use of VPN capabilities built into mobile operating systems is non-standard because a normal VPN does not eavesdrop and only intercepts traffic to encrypt and tunnel it safely to some secure network (such as company's intranet).

A VPN alone will observe all the URLs in intercepted HTTP traffic but only the domains for intercepted HTTPS traffic.[5] However, all major news publishers in the US already use HTTPS. In order to capture individual URLs transmitted over HTTPS the mobile user also has to install a custom CA to their certificate store. This essentially allows the VPN app to act as an active "man-in-the-middle" and eavesdrop on the browser-publisher communication. The VPN creates two separate sessions: one HTTPS session between the browser and the VPN and another HTTPS session between the VPN and the publisher (such as the New York Times). Using the custom CA, the VPN app creates, on the fly, a valid SSL certificate[6]

[4] A common way to directly navigate to the NYT is to through a "navigational search" where the user submits a query such as "New York Times" and the search engine return the homepage as the first result. Typical navigational queries are treated like bookmarks as direct navigation.

[5] For example, all visits to the New York Times would show up as `https://www.nytimes.com`.

[6] Every domain that serves HTTPS traffic needs an SSL certificate to verify the domain's identity to the browser. The VPN app is creating what's called a self-signed certificate which a browser would typically reject *unless* the creator of the certificate has certificate authority.

which makes the browser trust the VPN app. This allows the VPN app to collect detailed URL data from HTTPS traffic beyond the domain.

The VPN/CA method is used by the YouGov Pulse panel and appears to be also used by Nielsen and comScore. However, this interception technique has important limitations. First, none of the browser metadata can be captured (such as tab identifiers). Second, the method cannot distinguish between true page loads and caching where the browser prefetches resources from the server. Third, the VPN will typically capture any app traffic and not just mobile browsing. Without further information, such as measuring the running foreground and background apps on a mobile device, it is difficult to attribute traffic uniquely to the mobile browser.

2.2 Measuring App Traffic

Measuring traffic from mobile apps poses additional challenges. First, it is impractical to ask app vendors to natively measure consumption and then aggregate the data across users for the purpose of creating mobile news consumption logs. While there are only four major browser vendors (Chrome, Safari, Firefox and Edge/IE), there are thousands of first-party and third-party news apps. Even if all of them would collect user data in a common format it would be impossible to aggregate on the user level without a common user or device ID. Hence while initiatives such as Facebook's *Social Science One* are valuable, they cannot substitute for browser logs because we are unable to observe news read outside the Facebook mobile app.

Of course, we can still measure app traffic the same way we measure mobile browser traffic on the device level with the VPN/CA method. However, this type of data is even more limited when generated by apps compared to the mobile browser. In particular, mobile apps preload and cache lots of content in regular intervals in order to improve the user experience and enable offline news consumption. Preloading gives rise to *asynchronicity*: the timestamp on pre-loaded articles corresponds to when the articles were fetched and cached rather than when they were consumed. It also gives rise to *overloading* where much (or even most) of the retrieved content is never read by the user.

Finally, mobile apps often employ *hosting* mechanisms such as Facebook Instant Articles that host news articles directly on their server rather than having to fetch them from the publisher's server. Thus the publisher and/or URL of a consumed news article might not appear in the data at all.

According to our estimates, the amount of news traffic across both desktop and mobile devices that is conducted through mobile apps (and hence difficult to measure) is between 41% and 56% of total news consumption. Table 1 shows a breakdown of news browsing across desktop and mobile browser, mobile news apps and third party apps. We distinguish between three scenarios depending on whether news visits on third party social media apps redirect to the browser (in which case we can include them in the browser traffic) or whether they stay within social media apps (such as Instant Articles on Facebook) and thus cannot be observed. We explain our methodology in Appendix A-1.

Table 1: Percent of time spent consuming news by platform

Desktop Browser	Mobile Browser	News Apps	Social Media Apps
News visits initiated on social media apps: all within the app			
26%	18%	31%	25%
News visits initiated on social media apps: 50% within the app			
30%	20%	35%	15%
News visits initiated on social media apps: all within the browser			
35%	24%	41%	0%

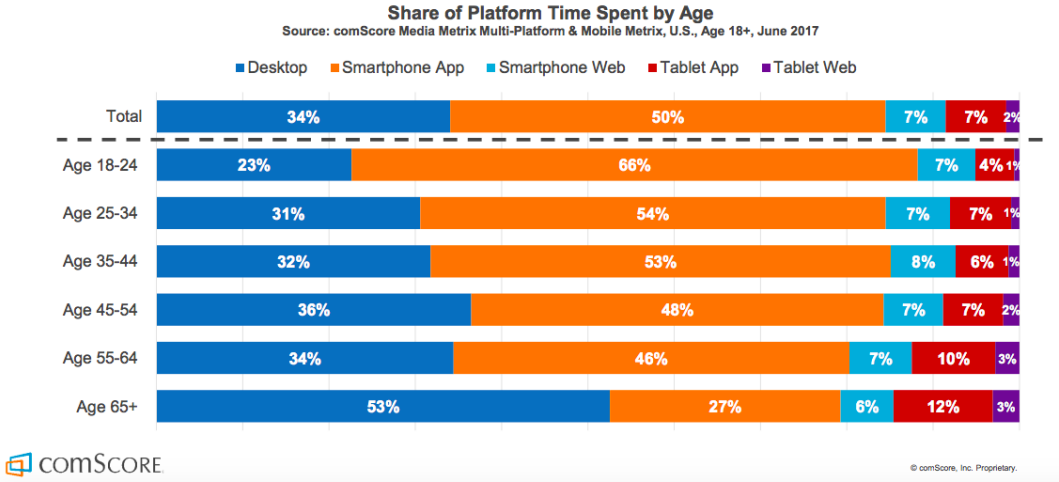
We distinguish between the extremes where (1) news consumption on social media apps stays completely within the apps and (2) all visits are referred to the browser. Since we do not know the actual proportion of visits referred to browsers, we also consider an intermediate scenario.

2.3 Extrapolating from Mobile Browsing Traffic

If browsing data captured a representative sample of news consumption, then there would be less concern about the app traffic that is difficult to measure; we could simply extrapolate from the browsing data. However, mobile browsing data does not appear to be representative of overall news consumption.

According to the comScore U.S. 2017 Mobile App report, 18-24 year olds spend 66% of their time online on apps, while adults over the age of 65 only spend 27% of their time online on apps. Figure 2 illustrates this breakdown by age.

Figure 2: Share of platform time spent by age



Source: comScore (2018; accessed October 4, 2018)

As such, browsing data captures a potentially non-representative sample of users and of news

consumption behavior. This is further supported by findings that younger people tend to read more news on their mobile devices.[7] Overall time spent on apps, then, is driven by younger populations and will grow over time, indicating an even greater need for the data collection methods we propose.

News consumption behavior on apps is significantly different to consumption on browsers. For example, the rise of app notifications and aggregator alerts changes the way people consume news on apps. In the U.S., around 20% of smartphone owners say they have seen a news alert within the past week.[8] Alerts increasingly both provide information and act as a gateway to further news consumption; this behavior cannot be observed or understood through traditional methods.

Research including both eye-tracking and traffic analysis methods has indicated that browsing behavior differs between news apps and mobile browsers (Dunaway et al., 2018). For example, attention to news is higher in apps than when news is browsed in mobile browsers; however, there is also a breadth vs depth trade-off in that the number of users reached via app is only a fraction of the users that news media reach via mobile browsers. Data on the differences in browsing, and their implication for overall news consumption, cannot be gleaned from current methods.

3 Existing Data Sources and Recent Research

In this section we review existing data sources for news consumption, how they treat mobile news consumption, and how this data has been used in recent research.

3.1 Desktop Browsing Data

Some studies ignore mobile browsing altogether and simply focus on desktop browsing logs. This approach can be appropriate for event analyses or to document general patterns that we expect to stay consistent across demographic groups.

Among browser vendors, only Microsoft appears to allow researchers to use their data for academic studies. Athey et al. (2017) use Microsoft Toolbar data to analyze the shutdown of Google News in Spain in December 2014 after a change in copyright law. Athey et al. (2017)'s goal was to understand whether the news consumption of former users of Google News tended to increase (Google News induced users to leave its platform) or decrease (it might have become more difficult to discover news). Even though the toolbar sample was highly non-representative the sign of the overall effect of the Google News shutdown could plausibly be similar across demographics. Flaxman et al. (2016) also use Microsoft toolbar data to understand to what extent left-leaning (right-leaning) users consume right-leaning

[7] See Lu (June 12, 2017; accessed November 2, 2018).

[8] See Newman et al. (2018; accessed November 2, 2018, page 15)

(left-leaning) content. Again, it could be argued that evidence for or against "echo chambers" should apply across various demographic groups.

Representative user samples comprising several hundred thousand desktop users are maintained by comScore, Alexa (owned by Amazon), Nielsen, YouGov and SurveySavvyConnect (Luth Research). While these companies are secretive about their precise methods they appear to use a combination of browser extensions (Alexa) and device-level traffic measurement using the VPN/CA approach discussed previously.[9] Gentzkow and Shapiro (2011) analyze a similar question as Flaxman et al. (2016) but use comScore desktop browser logs.

3.2 Surveys

Surveys are used by many studies to measure media consumption across all devices (including mobile); comScore, Nielsen and YouGov complement their user panel data with regular surveys.

Every year, the Reuters Institute for the Study of Journalism at the University of Oxford publishes among the most comprehensive studies of digital news consumption.[10] The latest such report, the Reuters Institute's 2018 Digital News Report, collected information from over 74,000 online news consumers in 37 countries. The Digital News Report uses a combination of online surveys and face-to-face focus groups to establish its findings. Allcott and Gentzkow (2017) collect their own survey data to study the prevalence of "fake news" before the 2016 U.S. presidential election. These authors provide benchmarks of the rate of exposure (i.e. articles seen) to fake news and attempt to study inference about true versus false news headlines.[11] Guess et al. (2018), meanwhile, study similar questions as Allcott and Gentzkow (2017) and complement passive data from YouGov's Pulse panel with data from their own survey administered to the same users.

While surveys are helpful to capture broad consumption patterns such as the relative importance of mobile and desktop/laptop devices for a user's news consumption, they cannot replace passive data collection when it comes to questions about sites visited, amount of content consumed, overall time spent, or other behaviors that are difficult for respondents to recall.

3.3 First-Party App Data

Data can be collected by first-party news publishers directly. Parse.ly conducts web analytics for websites (comparable to Google Analytics) and provides free reports on Internet-wide trends through its Currents service. For a given word or phrase, Currents provides the

[9] Even though device level traffic measurement has the same shortcomings on desktop as on mobile, programs can in principle embed themselves much more deeply in the desktop operating system and hence can be designed to only intercept browser traffic. Moreover, the browser is the main app for news consumption on the desktop.

[10] See Newman et al. (2018; accessed November 2, 2018).

[11] Allcott and Gentzkow (2017) complement their analysis with desktop browsing data from Alexa.

number of news articles published over the day related to that phrase, the total user views for that phrase, links to the related news articles, viewership by location, percent of views by platform (mobile, tablet, computer), and percent of views by access method (social media, web browser, text/email referral, news landing page). A paid subscription to Currents grants access to the data. Currents has a few limitations: Parse.ly publishes data by phrase, not by news article; the only user demographic data Parse.ly collects is location; Parse.ly only collects data from news sites that have installed the Parse.ly data collector; and it is unclear how many news domains actually use Parse.ly.

A general limitation of data collected by first-parties is that such data cannot be aggregated across apps due to a lack of a common user identifier.

3.4 Third-Party App Data

Social Science One is a new partnership between academics and Facebook which provides secure access to Facebook user data for academic research. The first data set released contains all URLs shared on Facebook broken down by various demographics and by device (computer or mobile) starting January 1st, 2017. To analyze the data, researchers submit a request for proposal; if accepted, researchers must send their code to Facebook to run in-house.

The data has similar limitation as first-party app data in it cannot be linked to consumption data from mobile browsing and news and other social media apps. However, *Social Science One* does capture almost one-third of referral traffic to newspapers.

3.5 Device-level VPN/CA Data

Device-level traffic data using the VPN/CA method described in Section 2.1 is the default traffic collection method for mobile devices. Mobile panels are maintained by comScore (called *Mobile Metrix*), Nielsen, YouGov (called *YouGov Pulse*) and Luth Research (*SurveySavvyConnect*). comScore can observe the behavior of panel members across desktop and mobile devices and can "deduplicate" users who access a publisher through several devices.

Nelson and Lei (2018) use comScore mobile and desktop data to measure how Americans consume news by platform. Guess et al. (2018) use the YouGov Pulse panel to analyze fake news consumption by U.S. users prior to the 2016 presidential election.

As discussed in Section 2.2, capturing non-browsing app data through the VPN/CA method is unreliable due to incorrect time stamps (asynchronicity), overloading, and the fact that content hosted by third parties such as Facebook Instant Articles remains unrecorded.

4 Proposal for a New Mobile Panel

We have argued in Section 2 that current techniques to passively capture news consumption mis-measure about half of all consumption, namely news that is consumed through first-party mobile publisher apps and third-party mobile apps such as Facebook. The problem lies with the default VPN/CA method that has been used to passively intercept traffic data between the mobile device and publishers. In Section 3 we discusssed that collecting data on the publisher level or conducting user surveys is valuable but cannot substitute for passive data collection.

In this Section we propose a new approach to passively track news consumption on mobile devices which is based on a very simple idea: since we are interested what people consume and click on their mobile screens, why not capture this information and actions *directly* rather than the traffic that these actions generate? We call this new the WYSIWYG ("what you see is what you get") approach to mobile news consumption data collection.

4.1 Description of WYSIWYG Mobile Panel

Figure 3 explains how our proposed WYSIWYG mobile panel will work. There are three major components: there is an application (pink boxes) that users are expected to download and will monitors a user's interactions with her screen. There is a server backend that processes possible news interactions (blue boxes) and a database that stores verified news interactions (green box). In the following, we explain each component in detail.

Mobile app. We recruit participants to our panel and ask them to download our app. The app asks for user's informed consent and installs a background service that constantly runs on the user's phone. The app also exposes a settings screen that allows users to view server data that has been collected at any time, leave the panel at any time, and request deletion of their data at any time.

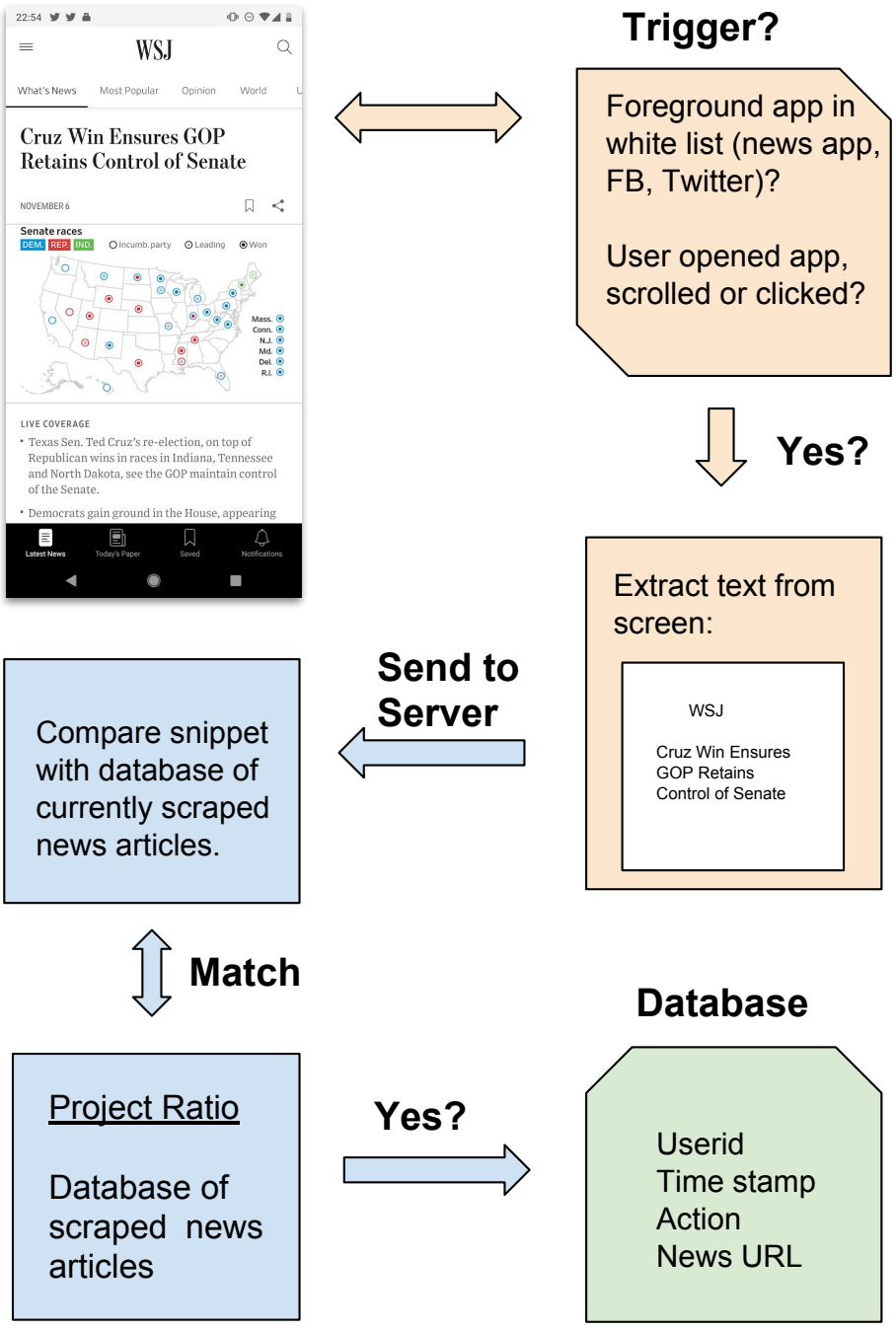
We focus on Android users as the Android OS exposes the accessibility services API,[12] allowing our background service to (a) check which apps are running in the foreground and (b) intercept user actions such as a clicks or swipes. Our service is triggered if the following two conditions are true:

- A first-party news app or third-part app such as Facebook or Twitter are running in the foreground and the screen is on.
- The app either opened a new screen, loaded 10 seconds ago, or the user interacted with the app through a gesture such as a click or between swipes (so reader might be expected to read text on screen).

If these conditions are true, our service captures a text view of the current window and adds a timestamp as well as the conditions that triggered this event (foreground app and

[12] A well known app on the Google Play Store that relies on this functionality is LastPass by LogMeIn which keeps tracks of users' passwords.

Figure 3: System for capturing news consumption by monitoring user screen



Our proposed mobile panel application resides on panel participants' phones and is activated only when users interact with a whitelist of first- and third-party applications (e.g., major news apps, Facebook, Google News, etc.) and their notifications. When a participant interacts with a whitelisted app, our application captures the content of the screen as a text snippet as well as the specific action performed by the user. Data corresponding to the action including {user ID, foreground app, user action , text snippet, timestamp} are then sent to a server. The server matches the text snippet to a pre-scraped set of URLs. Upon a successful matching, {user ID, foreground app, user action, URL, timestamp} are written to a database, thus capturing the specific news article and all associated user-level actions.

user action). Since we only capture text data each such event takes only a small amount of storage space. This information is sent at appropriate times to our backend (for example, when the device is connected to wi-fi or when charging).

Backend and Database. The backend receives users events including user ID. Importantly, we never store these events for privacy reasons. Instead, we *match* the text snippet to a database of news articles (headlines and URLs) that we concurrently scrape for publishers. In particular, Project Ratio at the Harmony Institute collects this data.

Every event with a headline that can be matched to a publisher/URL in our Project Ratio data is called a *match* and is written to a database in the form:

{user ID,foreground app, user action, URL, timestamp}

Events that cannot be matched are discarded and stored as failures in the database *without* the text snippet. This ensures that we do not inadvertently store any non-news events such as users messaging to their friends.[13]

Our matching algorithm ensures that the sensitive text snippet data is only stored on user devices and briefly in the memory of the backend server during the matching process but never in any backend database.

4.2 Discussion

Some caveats about our proposed WYSIWYG panel are in order. The quality of the data depends on having access to a high quality repository of news articles that can be used to match to snippets. The system might also need to be adjusted in cases where hosted content is customized for the third-party app (for example, if the New York Times creates custom headlines for Instant Articles on Facebook).

On the other side, if good matching can be achieved, the WYSIWYG panel overcomes all of the key limitations of the VPN/CA method. It is not affected by data caching and can capture any news data in any app including third-party hosted content such as Facebook Instant Articles.

Moreover, our method can capture many interactions that could never be captured in a browser log such as which areas of a landing page a user pays attention to before even clicking the article. After a user opens a specific article, we can capture whether the user reads the entire article or only the first paragraph. This type of data could previously only be recorded from eye-tracking studies.

5 Conclusion

We argue that current passive data collection processes of online news consumption either miss or mis-measure about half of all news consumption, namely news that is consumed

[13]For debugging purposes, we do want to store the foreground app of any failed event.

on mobile first-party mobile publisher apps and third-party mobile social media apps. This problem is likely only to get worse as people increasingly consume news on their mobile devices.

Our WYSIWYG approach to mobile user panels not only addresses the key shortcoming of current passive data collection on mobile devices but also allows us to measure new metadata on how people interact with news content that is typically only available in eye-tracking studies.

References

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *The Journal of Economic Perspectives*, 2017, *31* (2), 211–235.

Athey, Susan, Markus Mobius, and Jeno Pal, "The Impact of Aggregators on Internet News Consumption," January 2017. working paper.

comScore, *The 2017 U.S. Mobile App Report*, comScore, 2018; accessed October 4, 2018. `https://static.poder360.com.br/2017/08/2017USMobileAppReport-1.pdf`.

Duca, John V. and Jason L. Saving, "Income Inequality, Media Fragmentation, and Increased Political Polarization," *Contemporary Economic Policy*, 2017, *35* (2), 392–413.

Dunaway, Johanna, Kathleen Searles, Mingxiao Sui, and Newly Paul, "News Attention in a Mobile Era," *Journal of Computer-Mediated Communication*, 2018, *23* (2), 107–124.

Flaxman, Seth, Sharad Goel, and Justin M. Rao, "Filter Bubbles, Echo Chambers, and Online News Consumption," *Public Opinion Quarterly*, 2016, *80* (S1), 298–320.

Gentzkow, Matthew and Jesse M. Shapiro, "Ideological Segregation Online and Offline *," *The Quarterly Journal of Economics*, 2011, *126* (4), 1799–1839.

Guess, Andrew, Brendan Nyhan, and Jason Reifler, "Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign," January 2018. working paper.

Lu, Kristine, *Growth in mobile news use driven by older adults*, Pew Resarch Center, June 12, 2017; accessed November 2, 2018. `http://www.pewresearch.org/fact-tank/2017/06/12/growth-in-mobile-news-use-driven-by-older-adults/`.

Nelson, Jacob L. and Ryan F. Lei, "The Effect of Digital Platforms on News Audience Behavior," *Digital Journalism*, 2018, *6* (5), 619–633.

Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus Kleis Nielsen, *Reuters Institute - Digital News Report 2018*, Reuters Institute, 2018; accessed November 2, 2018. `http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf`.

Parse.ly, *Explore traffic source trends for digital publishers*, Parse.ly, 2018; accessed October 4, 2018. https://www.parse.ly/resources/data-studies/referrer-dashboard/.

A-1 Methodology for Calculating Time Spent by Platform

We estimate the percentage of observed news data using time spent browsing as our metric. As such, the percent of news data that we currently capture by measuring a user's browser click stream equals:

$$\frac{\text{time spent consuming news on browsers}}{\text{time spent consuming news on browsers and apps}}$$

According to Comscore's 2017 U.S Mobile App Report, online news readers spend 35% of their reading time on their desktop, 24% on their mobile web browser, and 41% on news mobile apps. According to this breakdown, our current data only capture 59% of news consumption on the time spent metric. This measure does not account for news consumption on social media apps like Facebook and Twitter, as comScore does not classify these apps as "news apps". In the following section, we estimate from comScore data the amount of time spent on news consumption within social media apps.

Consider a user who spent 100 minutes consuming news on browsers and on news apps. According to comScore's stats, we would expect that the user spent 41 minutes consuming news on news apps, 24 on mobile browsers, and 35 minutes on desktop browsers. In order to estimate total news consumption, we must estimate the amount of time our user spent consuming news on social media apps.

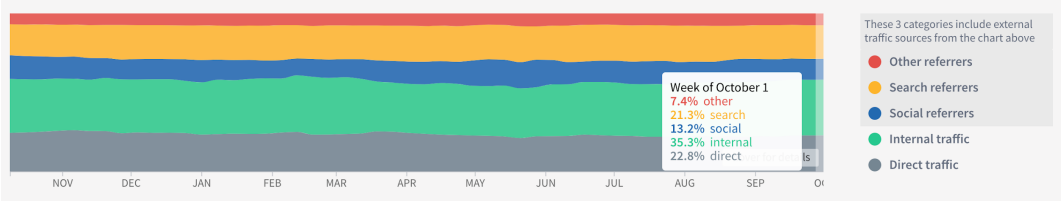
According to Parse.ly, 26% of external referrals to news sites originated from social media (see Figures 4 and 5). This figure, however, underestimates the amount of news consumption that social media initiates. For example, as previously discussed, technologies such as Facebook Instant Articles have allowed people to consume news without leaving the Facebook app and redirecting to news sites. If we assume a linear relationship between articles clicked and time spent reading, then we would expect our user to spend 15 minutes of his 59 browser news consumption minutes reading articles she discovered on social media.

We assume that our user has the same propensity to click articles on social media while on browsers as in apps. According to comScore's 2017 *U.S Mobile App Report*, the average individual spends 70% of her time spent on social media within social media apps as opposed to browsers.[1] Thus our user would spend 230% of the time spent on browser on on social media apps. Given that our user's browser social media resulted in 15 minutes of news consumption minutes and her equal propensity to consume news on social media on browsers and apps, we would expect social media apps to result in 34 minutes of news consumption. Since we only capture data on news consumption via browsers, we only capture data on the 59 minutes our user spent consuming news on browsers and not on her 41 minutes on news media apps or 34 minutes on social media apps. That is, we only collect data on 44% of all news consumption (alternative calculations, where we assume that 100% and 50% of social media app news consumption redirects to mobile browsers, result in figures of 59% and 50%

[1] See comScore (2018; accessed October 4, 2018).

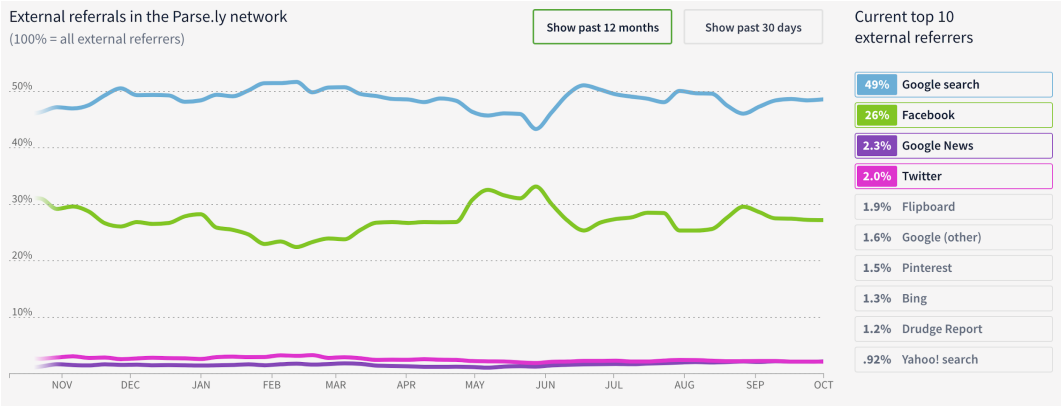
respectively). Generalizing this to the population results in our estimates for the breakdown of news consumption by platform. Refer to Table 1 for estimates.

Figure 4: Referral traffic by category



Source: Parse.ly (2018; accessed October 4, 2018).

Figure 5: Referral traffic by external referrers



Source: Parse.ly (2018; accessed October 4, 2018).