Indigenous Internet: Nuances of Native American Internet Use

Esther Showalter*, Nicole Moghaddas*, Morgan Vigil-Hayes†, Ellen Zegura‡, Elizabeth Belding*
*Dept. of Computer Science, UC Santa Barbara; †School of Informatics, Computing, and Cybersystems, Northern Arizona University; ‡School of Computer Science, Georgia Institute of Technology
ehs,nmoghaddas,ebelding@cs.ucsb.edu,morgan.vigil-hayes@nau.edu,ewz@cc.gatech.edu

ABSTRACT

We investigate Internet traffic logs on rural Native American reservations in California served by a tribally-owned Internet provider. The anonymous user browsing traffic and device preferences create unique failure patterns that reveal where connectivity is not well-served by standard networking technology. Geographical and cultural marginalization has lead to distinctive Internet usage when connectivity is available. For instance, residents commonly visit websites that are not as popular in the wider United States; and mobile devices dominate web request traffic, often with content-heavy video and media downloads, despite sustaining a higher rate of failure than desktop devices. This statistical analysis of passive measurements avoids institutional and cultural biases and ensures continuing research will contribute to a decolonizing narrative of Native American informational practices and values. Based on our analysis, we propose follow-up research angles to better understand the technological and social drivers behind these findings in order to improve user experience in this, and similar, networks.

1 INTRODUCTION

Between 2014 and 2016, urban areas of the United States (U.S.) saw the penetration of high speed cellular service rise from 81.9% to 90.5%. During the same time, penetration throughout tribal areas remained a flat 64% while anecdotal reports put this number much lower [5]. Residents of U.S. tribal lands cannot assume constant connectivity nor the proximity of online content. In response to highly fluctuating connectivity conditions, along with the geographical and cultural isolation of many Indigenous American reservations, tribal communities develop nuances in Internet use that differentiate them from the more general American population [9, 10]. Understanding the differences that emerge from low-bandwidth, dispersed Internet use can help local Internet providers broaden Internet accessibility on tribal lands and increase the quality of use for America's most rural residents.

In this project we analyze anonymous network traffic collected within the Tribal Digital Village network (TDV)[1], a tribally-owned Internet service provider (ISP) founded in 2001 that brings access to homes, educational centers, and community hubs across 17 tribes in 13 Native American reservations in eastern San Diego county, California. Using traces collected over five weeks in February and March 2018 from seven distinct reservations, we characterize the way residents within these communities prefer to go online. This ongoing study has revealed new insights into the quality of connectivity experienced by TDV users by identifying access and usage patterns that may not be well-served by the standard network equipment configuration. Our findings include:

- The most popular websites visited by TDV users do not align with the most popular websites across the U.S.
- Mobile devices interact with more video and image download sites, while fixed devices are more likely to interact with adware.
- Most connection failures are due to the originating device canceling the request.
- Mobile devices experience more failures than desktop devices across all brands.

2 RELATED WORK

The path to effective information and communication technologies (ICT) on sovereign lands within the U.S. depends on the determination of Native nations' leaders to take control of the digital information flow surrounding indigenous social and political goals [4]. Indeed, the developers of TDV overcame many unusual geographic, economic, technological, and political struggles to take the initiative for their own Internet access and still continue the fight for reliable connectivity in the face of public disinterest [7].

Studies of Internet use in other marginalized communities include [2], [6], and [3]. Six weeks of web traffic from a primary school outside of Bangalore, India was assessed to identify constraints on browsing performance in [2]. The authors then employed a local caching strategy that more than doubled the speed at which the students could browse. In Zambia, traffic captures on a rural wireless community Internet service combined with user interviews revealed possibilities for low-bandwidth resiliency network design suggestions that better serve the social needs of the community [6]. In South Australia, an empirical study of online interaction and digital information practices of an indigenous group was conducted through extensive interviews with a broad range of demographic participants [3]. The findings helped to structure community training videos around accepted traditional knowledge and drive subsequent studies of Internet application adaptation.

Our methodology assesses the digital information practices of indigenous Americans within the TDV network to discover patterns of use that will drive culturally accurate research questions toward a true, well-told, and decolonizing narrative of Native American informational practices and values [8].

[1] https://sctdv.net

ICTD, January 2019, Ahmedabad, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

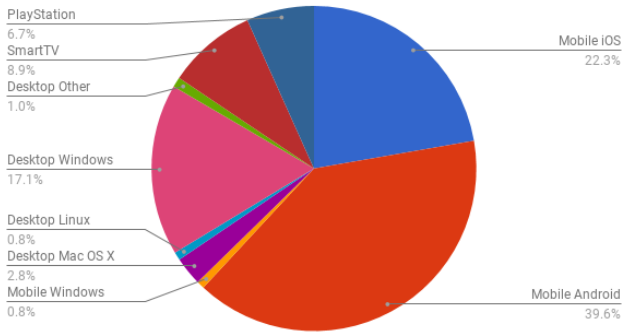


Figure 1: Percentage of network traffic by device type.

3 NETWORK & DATA COLLECTION

TDV maintains a backhaul of microwave links over 50 miles across the back country mountains of San Diego County. The backhaul terminates at TDV offices in Pala, where a fiber connection from AT&T provides Internet access at 500 megabits per second (Mbps). 200 Mbps are reserved in Pala to serve the residents and to provide operational support. The remaining 300 Mbps serve the reservations down the backhaul. Relay links over unlicensed 2.4 gigahertz (GHz) and 5 GHz frequencies run the last miles from towers along the backhaul to individual homes and community centers on reservations. From these, broadcast Wi-Fi provides access points to user end devices.

In collaboration with TDV and with IRB approval, our research team has collected anonymous network traffic traces since 2014. Traffic is mirrored from the gateway router at TDV offices and condensed by Bro Network Security Monitor software into connection-level logs to conserve storage space. Bro logs provide flow-level records such as bytes transferred, traffic protocol, user agent strings when web request traffic is seen, as well as indicators of whether a transmission control protocol (TCP) connection saw a successful setup and tear down. Individual user privacy is maintained by an access point network address translation (NAT) protocol for each household, as well as a consistent anonymization scheme run on the Internet protocol (IP) identifiers in the logs before encrypted transfer across the Internet to the analysis repository. For this study we analyze over 224 million TCP connection logs representing network usage between February and March of 2018 from seven reservations situated one and two hops from the Internet gateway along the backbone of microwave towers.

4 FINDINGS

4.1 Device Identification

We interpret user agent strings in web requests to identify nine categories of devices that together generate almost 75% of the total network traffic. The remaining traffic connections either use protocols that do not provide a user agent string, or offered a string that was too vague to decode to a specific brand and model. User agent strings are recorded in hyper text transfer protocol (HTTP) request logs, along with the corresponding upstream and downstream byte count. Unless otherwise noted, the results presented in this paper involve traffic generated by identifiable devices, that is, devices that undertake HTTP requests, and whose user agents are sufficiently recognizable by free Python libraries.

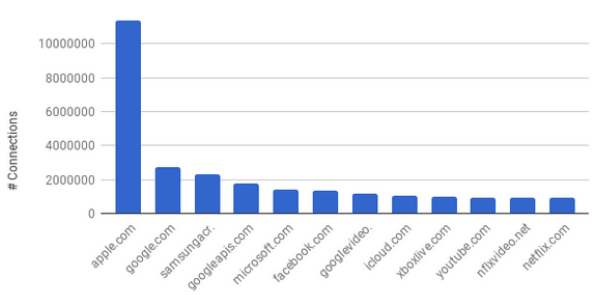
Figure 1 shows that 62.7% of all connections are initiated from mobile devices. Overall, Android devices make the most connections (39.6%), followed by iOS handheld devices (22.3%), with Windows desktops (17.1%) third. U.S. market share percentages over three months encompassing the period of this study show the same top three types of devices are popular across the country, but in opposite order: Windows desktop devices held the greatest percentage of market share at 37%, followed by iOS at 29%, and Android at 20%[2]. Although not perfectly equatable quantities, these measurements offer a rough estimate of the relative prevalence of these devices, suggesting that the TDV user base has a larger mobile device population compared to the U.S. average. This trend has been observed elsewhere throughout the world in rural, less connected populations [1]. In April of 2018, about 40% of all U.S. Internet traffic was served to mobile devices[3]. In the following section, we examine the amount of traffic in gigabytes (GB) sent over this network to discover which devices tend to visit which websites most frequently.

4.2 Popular Web Domains

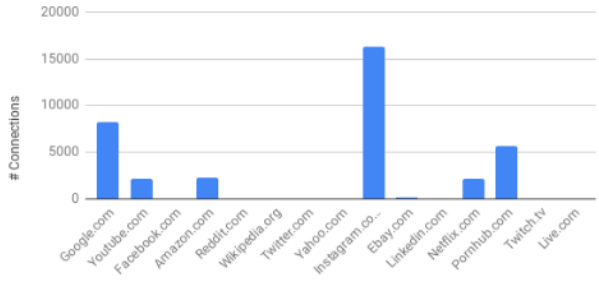
We first rank the web domains visited by TDV users during the two months of this study by the amount of web connection flows in Figure 2a. Rankings are similar when measured by number of bytes uploaded and downloaded from each site. However, application connections such as Instagram tend to transfer a greater amount of bytes (either uploading or downloading) for a single connection request than non-application connections. The most trafficked websites measured this way vary significantly from those of the wider U.S. during a similar time period of 2018. Alexa ranks site popularity with an average of page views, time spent on site, links from external sites, and the percentage of traffic that was directed to a site from a search engine[4]. Although these metrics are not directly comparable, they give an initial estimate of the differences in domain popularity between this minority group and the wider U.S. Figure 2b shows the amount of traffic in GB generated by TDV users to top sites in the U.S. ranked by Alexa. Social media (Instagram) and web search tools (Google) are present in the top 15 sites, along with advertisement sites, video content streaming, and gaming.

Figure 3a ranks the top URLs visited by mobile devices, while Figure 3b does the same for fixed devices. On mobile, Apple.com generates more than twice as much traffic than the next site, an Instagram content delivery network (CDN). The Instagram CDN similarly generates more than twice as much traffic as the next site, lkqd.net, which appears to host advertisement services. After the first two top sites, no other site generated more than 6 GB total over the course of the study. These remaining top sites consist of image hosting, video and music streaming, CDNs, and more sites that appear to be adware. On fixed devices, however, advertisement sites such as adsafeprotected.com and atemda.com generate the greatest amount of traffic, but all top sites generated totals of under

[2] http://gs.statcounter.com/os-market-share/all/united-states-of-america
[3] https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions
[4] https://www.alexa.com/topsites/countries/U.S.

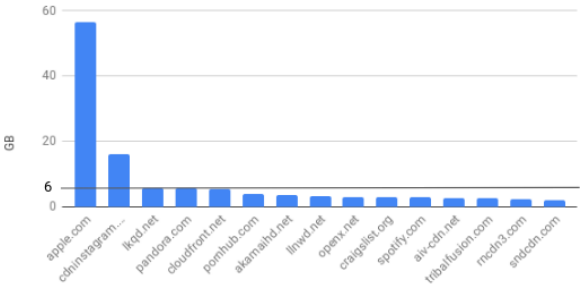


(a) Domains visited by TDV users, ranked in order of number of connection flows.

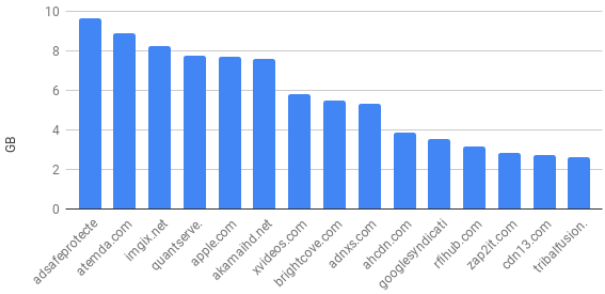


(b) Amount of TDV traffic to domains popular in the U.S. during the time frame of this study, ranked by Alexa.

Figure 2: Most visited web domains.



(a) Top sites by total traffic from mobile devices. After Apple.com and Instagram.com, all other sites generated less than 6 GB total over the course of the study.



(b) Top sites by total traffic from fixed devices. These include laptops but exclude gaming devices or smart TVs. All sites generated less than 10 GB over the course of this study.

Figure 3: Top sites accessed by different device types.

10 GB throughout the entire analysis period. These findings demonstrate that TDV users prefer mobile devices even for content-heavy activities such as music and video streaming or image viewing.

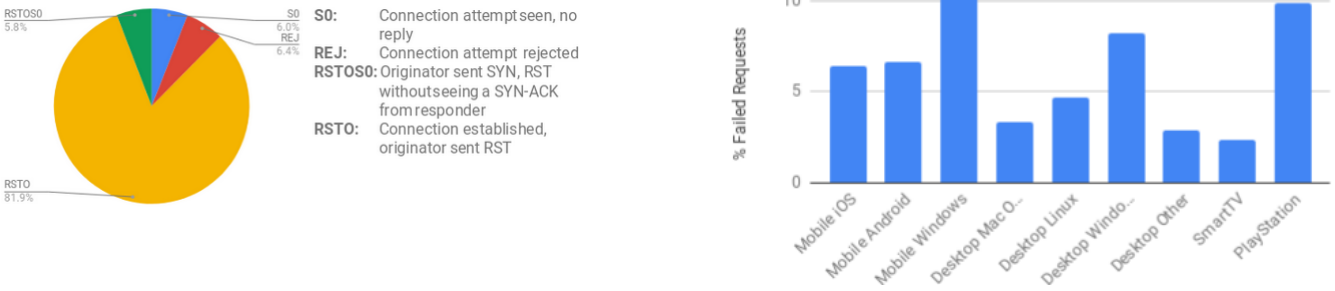
4.3 Connection Failure Behavior

A web connection can produce errors under multiple conditions. We define failure as any of four distinct TCP behaviors, described in Figure 4a by the Bro log status code recorded for each connection at the time of capture. Vigil et al. found that in packet samples of video downloads collected in 2014, connection failures caused by reset from the initiator (RST0) accounted for up to 60% of failures experienced by the most popular video hosting sites: Youtube, Netflix, and Instagram [9]. Through data presented in Figure 4a, we find that over 80% of failed flows (across all locations and all web services) failed in the same manner. Failure to reply can be caused by packet loss, timeouts, or other factors. The next highest reason for failure, accounting for just under 25% of all failures, is when a packet with the reset flag (RST) set was sent by the originator after the connection was established and was currently in use. While it might be normal operation with certain browser configurations for the server to use a RST to clean up an idle connection instead of a FIN, this more likely signifies an application disconnect. It may also represent the user manually canceling a page load if the process was taking too long. Future work will consider the differing TCP versions used by each operating system in the network for a more granular analysis of this failure.

Figure 4b depicts the percentage of failed connections experienced by each device type. While mobile devices appear to experience more failure than fixed devices, we notice that they also initiate more connections and are more likely to compete for bandwidth. The overall percentage of connection failure experienced by users on each reservation is shown in Figure 5. Included in this statistic are all connections from known devices as well as those with unrecognized user agent strings, which appear to fail much more frequently than identifiable devices and drive the total percentages up. However, in data collected on TDV traffic across similar locations in 2014, packet retransmission rates were recorded between 23% and 25% [9]. We see a slight improvement in that all reservations experienced between 19% and 23% connection failure four years later. This is still a high amount of disruption and will be investigated further by refining user agent identification algorithms.

5 CONCLUSION

The growth of indigenous cultural identity must drive the development of Internet access of tribal lands. Though several initiatives have studied the ways indigenous people around the world use Internet access, only a few have attempted traffic analysis of network



(a) Failure types and Bro connection status code.

(b) Failed connections by device type.

Figure 4: Failure characteristics.

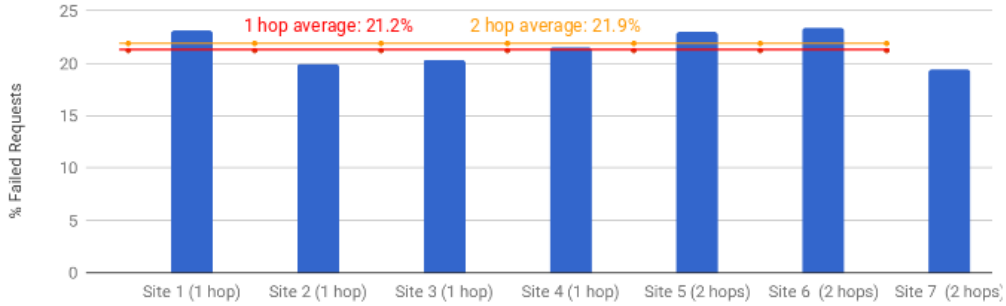


Figure 5: Connection failures by reservation. "Hop" indicates the number of relay towers traveled from the Internet gateway along the backhaul before branching off on unlicensed spectrum. Note averages include connections from devices with indistinguishable user agent strings.

use to support user behavior. TDV is a notable network in that it serves remote people groups with a high bandwidth broadband connection. We have observed traffic patterns in this network over two months to discover nuances in preferred Internet use. Our analysis has produced initial findings of unique on-line usage that has driven actionable questions for future research in browsing preferences, device connectivity, and causes for failure. We will build on these findings to develop network optimizations that increase the quality of connection experience for American Indians living on reservations served by this and other networks.

6 ACKNOWLEDGEMENTS

This work was funded by grants NSF-1563436 and NSF-1637265. We thank the users and administrators of TDV for their assistance in this research.

REFERENCES

[1] Asheeta Bhavnani, Rowena Won-Wai Chiu, Subramaniam Janakiram, Peter Silarszky, and Deepak Bhatia. 2008. The role of mobile phones in sustainable rural poverty reduction. *retrieved November* 22 (2008), 2008.

[2] Jay Chen, David Hutchful, William Thies, and Lakshminarayanan Subramanian. 2011. Analyzing and Accelerating Web Access in a School in Peri-urban India. In *Proceedings of the 20th International Conference Companion on World Wide Web (WWW '11)*. ACM, New York, NY, USA, 443–452. https://doi.org/10.1145/1963192.1963358

[3] Jia Du, Jelina Haines, Qiaoling Sun, Helen Partridge, and Dandan Ma. 2015. Understanding Indigenous People's Information Practices and Internet use: A Ngarrindjeri perspective *(PACIS, 2015)*.

[4] Marisa Elena Duarte. 2017. *Network Sovereignty: Building the Internet Across Indian Country*. University of Washington Press.

[5] FCC. 2018. Broadband Deployment Report. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2018-broadband-deployment-report

[6] David L. Johnson, Veljko Pejovic, Elizabeth M. Belding, and Gertjan van Stam. 2011. Traffic Characterization and Internet Usage in Rural Africa. In *Proceedings of the 20th International Conference Companion on World Wide Web (WWW '11)*. ACM, New York, NY, USA, 493–502. https://doi.org/10.1145/1963192.1963363

[7] Christian Sandvig. 2012. Connection at Ewiiaapaayp Mountain. *Routledge* (2012).

[8] Linda Tuhiwai Smith. 2013. *Decolonizing Methodologies: Research and Indigenous Peoples*. Zed Books, Ltd.

[9] Morgan Vigil, Matthew Rantanen, and Elizabeth Belding. 2015. A First Look at Tribal Web Traffic. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. 1155–1165.

[10] Mike P. Wittie, Veljko Pejovic, Lara Deek, Kevin C. Almeroth, and Ben Y. Zhao. 2010. Exploiting Locality of Interest in Online Social Networks. In *ACM CoNEXT*.