Don't Forget the User: It's Time to Rethink Network Measurements

Aryan Taneja*[1], Rahul Bothra*[1,2], Debopam Bhattacherjee[1], Rohan Gandhi[1], Venkata N. Padmanabhan[1], Ranjita Bhagwan[1,†], Nagarajan Natarajan[1], Saikat Guha[1], Ross Cutler[3]
[1]Microsoft Research India, [2]UIUC, [3]Microsoft

ABSTRACT

Network measurement has long focused on the bits and bytes — low-level network metrics such as latency and throughput, which have the advantage of being objective and directly characterizing the performance of the network. We argue that users provide a rich and largely untapped source of *implicit* as well as *explicit* signals that could complement and expand the coverage of traditional methods. Implicit feedback leverages user actions to indirectly infer the network performance and the resulting quality of user experience. Explicit feedback leverages user input, typically provided offline, to expand the reach of network measurement, especially for newer ones.

We analyze two scenarios: capturing implicit feedback through user actions from a large-scale conferencing service – MS Teams and gathering explicit feedback via social media posts pertaining to the SpaceX Starlink Low Earth Orbit (LEO) satellite network undergoing deployment. We believe our techniques complement the traditional measurement methods and open up a broad set of research directions, ranging from rethinking measurement tools to designing user-centric networked systems and applications.

CCS CONCEPTS

• **Networks** → *Network measurement*; *Network performance analysis*;

KEYWORDS

Network Measurement, User Feedback, User Engagement

ACM Reference Format:
Aryan Taneja*[1], Rahul Bothra*[1,2], Debopam Bhattacherjee[1], Rohan Gandhi[1], Venkata N. Padmanabhan[1], Ranjita Bhagwan[1,†], Nagarajan Natarajan[1], Saikat Guha[1], Ross Cutler[3] . 2023. Don't Forget the User: It's Time to Rethink Network Measurements. In *The 22nd ACM Workshop on Hot Topics in Networks (HotNets '23), November 28–29, 2023, Cambridge, MA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3626111.3630095

*equal contribution | †presently at Google.

HotNets '23, November 28–29, 2023, Cambridge, MA, USA

ACM ISBN 979-8-4007-0415-4/23/11…$15.00
https://doi.org/10.1145/3626111.3630095

1 INTRODUCTION

The goal of network measurement is to characterize how and how well a network is performing. Traditionally, such measurement has focused on low-level metrics such as network latency and throughput. While being objective and insightful, this approach suffers from two broad limitations.

First, performing such low-level measurements typically requires a point of presence in the network of interest, i.e., one or more computers under the control of the experimenter that are in the desired network locations. While there might be the opportunity to piggyback on existing large-scale commercial services (e.g., [15]), this might be out of the reach of the typical researcher. So, the coverage of the measurements is often limited to a modest size, spanning the nodes that the experimenter is able to recruit or perhaps has access to via a testbed (e.g., [17]).

Second, low-level measurements do not directly capture the user experience, which is key to understanding the relative importance of various low-level metrics. For example, is the latency low enough that there is little benefit in terms of user experience to optimize latency further? To address this limitation, there has been work on gathering user feedback (e.g., call quality surveys at the end of Skype calls) and then learning a model to predict the user experience based on the low-level metrics [23, 48, 61]. A key challenge, however, is that seeking such user feedback is an imposition on users, so it is done infrequently to keep the overhead low, e.g., only for a fraction of calls and that too only at the end of the calls.

We argue that the key to overcoming these limitations lies with the users and in leveraging the rich user feedback, both implicit and explicit, that is available "for free".

Implicit feedback in the form of user actions might be indicative of the user experience. For instance, a user experiencing high latency during an audio-video call might stay on mute or might prefer to turn off their video. While such implicit signals might be plentiful, e.g., available throughout the user session, the connection between such signals and the user experience might be loose, so assessing the latter based on the former would be a key challenge.

Explicit feedback in the form of social media posts is another source of signal. Unlike the call ratings splash screens, which are often viewed as an imposition and simply ignored by users, social media posts arise out of the users' own volition. Such posts can shed light on the performance of networks that might otherwise be inaccessible. A key challenge, however, is extracting useful information from loosely structured or even unstructured data.

To give these ideas concrete shape, we discuss and present preliminary results from two settings: quantifying the correlations between network performance, implicit user actions, and explicit user feedback or MOS (Mean Opinion Score) for a large-scale video conferencing application, MS Teams (§3), and characterizing the performance of the relatively nascent Starlink LEO satellite network based on Reddit posts (§4). Our results point to the promise of tapping implicit or explicit user feedback – implicit user actions could be used as proxies for MS Teams' heavily sampled user feedback. Aggregated insights on Starlink user sentiment from Reddit could be used as inputs for future network provisioning and optimization strategies.

These specific examples point to the broader opportunity (§5) of and the challenges entailed with these approaches, which we discuss in §6.

2 RELATED WORK

Measuring and modeling the impact of the network on application performance and user experience via both active and passive experiments has generated a lot of research interest; techniques include surveys [19, 45, 73], laboratory experiments [7, 14, 18, 46], crowd-sourcing [26], A/B testing [13, 41], and passive analyses [21, 23, 53, 61, 74]. There has been past work on identifying the metrics for user experience ([11, 28, 37]). There has been past work [8, 61] on understanding user engagement in on-demand video streaming. Unfortunately, such works only focus on a single user action – early abandonment. A recent work [47] analyzes Zoom traffic traces to deduce application-level metrics while another work [43] explores aggregated viewership-based analytics (e.g., flash crowd versus QoE) for video delivery services. In contrast, we use non-intrusive strategies to (a) quantify the impact of network performance on *user engagement* and (b) to find correlations between user actions and *explicit feedback* for video conferencing.

Sentiment analysis, which is used for detecting sentiments in underlying text, has been exhaustively applied to a multitude of use cases – understanding product feedback and preferences [4, 42], and predicting trends and outcomes of large-scale events [9, 36, 44]. Leveraging social media to understand sentiments has been used to understand/detect mobile network performance [27, 58] and demands [80], service availability and failures [49, 72], and attacks and security vulnerabilities [2, 16, 38, 59, 60, 66]. SpaceX Starlink is a nascent LEO satellite network with stated [68] goals to offer global low-latency broadband connectivity. While things are in a state of flux (currently under deployment), it is important to keep track of user sentiment, which, as we will see later, depends on a complex calculus of deployment cadence, footprint expansion, and user adoption. Hence, we mine a social network in this very new context and leverage cutting-edge language and vision tools to extract feedback and aggregated sentiment insights useful for the ISP.

3 USER ACTIONS TELL A STORY

In this section, we detail how user actions could offer *implicit* signals about the user experience. We analyzed ~X (between 150 and 200[1]) million calls spanning Jan-Apr, 2022 (older data due to business sensitivity; without loss of generality) in our large-scale video conferencing service, MS Teams. We show (a) network conditions and in-session user actions (user engagement) are correlated, and (b) these implicit user signals are correlated well with MOS. Importantly, while MOS is available for only a subset of calls, user signals are prevalent for all calls, thus allowing stakeholders to exhaustively consume user feedback hidden in those signals.

3.1 Methodology

We first define the terms and metrics used in our analysis.

Network condition metrics: The client running on the user-end of MS Teams gathers network latency, packet loss percent, jitter, and available bandwidth information every 5 seconds. When the user session ends, each client computes the mean, median, and 95^{th} percentile (P95) value for each of these metrics per session. In this paper, we report results using the mean but similar trends hold for P95 values as well.

User engagement metrics: We define the following user engagement (action) metrics: (a) `Presence`: The client records the user session duration. The presence denotes a user's session duration as a fraction (%) of the median session duration across all users in the call. We use the median instead of maximum (i.e., the total call duration) as a baseline since it is robust to outliers (users that stay on the call well past the time others have left). `Presence` is capped at 100. (b) `Cam On`: Fraction (%) of the user session for which a user has their camera on. (c) `Mic On`: Fraction (%) of the user session for which a user has their microphone on.

User feedback metrics: MS Teams requests a subset of users to submit *explicit feedback* at the end of sessions – a rating between 1 (worst) and 5 (best). Such feedback across users is aggregated and averaged to compute the MOS. Such feedback is only provided for a small fraction (between 0.1% and 1%)[1] of sessions.

Call dataset: To tackle confounders, we study only enterprise calls during business hours (9 AM - 8 PM EST) on weekdays with 3+ participants, all in the US.

3.2 Networks impact user engagement

First, we show how network conditions impact user engagement metrics across the ~X (between 150 and 200[1]) million calls analyzed. While evaluating one network condition metric, we try to analyze the calls where other metrics are roughly constant (latency between 0 - 40 ms, loss rate between 0 - 0.2%, jitter between 0 - 5 ms, and bandwidth between 3 - 4 Mbps).

Network latency: In Fig. 1 (left), as the mean network latency increases from 0 to 300 ms, both `Presence` and `Cam On` reduce by ~20% while `Mic On` reduces by more than 25%. Also, the slope of the `Mic On` plot is steeper until 150 ms

[1] Actual value hidden due to business sensitivity.

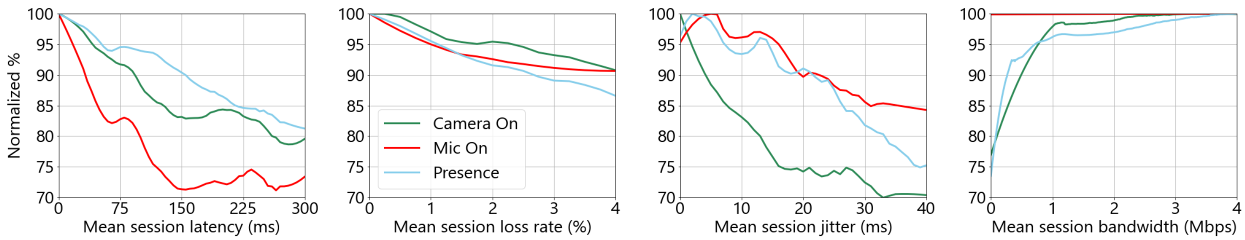


Figure 1: User engagement changes with network latency (left), packet loss (middle-left), network jitter (middle-right), and bandwidth (right). Engagement metrics and network performance metrics are computed per session.

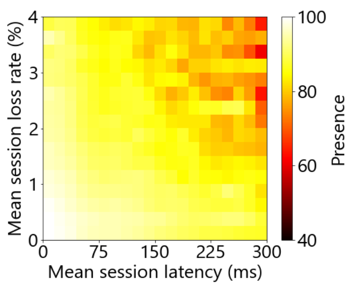


Figure 2: High network latency and high packet loss together have a compounding impact on `Presence`.

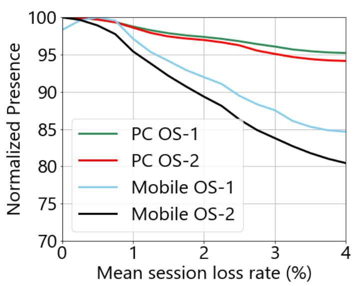


Figure 3: The platform type impacts user sensitivity to network loss rate.

after which it plateaus out. This indicates that at higher network latencies, MS Teams users mute themselves more often and might be muting themselves as the means of first resort, before taking more drastic steps such as turning the camera off or dropping off the call.

We also briefly consider whether `Cam On` has an impact on latency, say because the video stream congests the network. However, we found that the latency does *not* increase with `Cam On` (Fig. omitted for brevity), suggesting that the causation runs the other way — an increased latency causes users to turn off their cameras.

Packet loss: Surprisingly, the impact of loss on user engagement, as seen in Fig. 1 (middle-left) is substantially weaker – as the mean packet loss rate increases from 0 to 2%, `Mic On`, `Cam On`, and `Presence` drop by less than 10%. The key reason is that MS Teams is able to effectively mitigate the packet loss using application layer safeguards. The effects of network packet loss are small even when the loss rate is as high as 2% – rare [39] in the Internet.

Jitter: High network jitter, unlike loss, significantly impacts `Cam On`. From Fig. 1 (middle-right), we observe that at 10 ms of jitter, `Cam On` drops by more than 15%.

Bandwidth: Fig. 1 (right) shows that MS Teams is not too bandwidth hungry. Even at a mean session bandwidth of 1 Mbps, all the engagement metrics are within 5% of the best achievable (at ~4 Mbps). Also, `Mic On` does not correlate with bandwidth (see red line overlapping at y-axis = 100) – audio streams require bandwidth multiple orders of magnitude lower than what typical broadband Internet connectivity offers today. In the span of bandwidth represented in our data, we did not observe the user's muting actions being impacted by bandwidth.

Interestingly, depending on the aspect of network conditions that are degraded, users take different actions to varying degrees. For example, when the mean network latency is high, users tend to turn off the audio more frequently as the lag hinders the rapid turn-taking called for in an interactive dialogue. At very high packet loss of 3% or more, on the other hand, the chance of a user dropping off increases significantly (by more than 10%), presumably because the audio (and video) quality becomes unacceptably poor. Note that while the plots in Fig. 1 are uneven due to unknown confounders (we did tackle many of them, as already discussed), only the broad trends are relevant in this context.

Compounding impact of network attributes:

We study the additive impact of network performance metrics on user actions during meetings. While the individual impact of high network latency and packet loss is substantial on user engagement, the compounding effect is even higher. As shown in Fig.2, We found that user `Presence` percentage could dip by as much as ~50% for certain combinations of latency and loss relative to the best value across all such combinations. Such combinations do occur on the Internet, as is evident from our data.

Platform matters: Different platforms (PC/mobile, operating system, etc.) have different impacts on user sensitivity to network performance. Fig. 3 shows how `Presence` changes with loss rate for 4 different platforms. Users joining calls on their mobile devices tend to drop off sooner at the same mean network latency than users on PCs. Also, user sensitivity varies with different operating systems. This is, in hindsight, intuitive as user expectations are different on different platforms. For example, users joining work meetings from mobile devices might be less engaged. Also, the application-level optimizations could be different on different platforms depending on CPU and other resource availability impacting the engagement of the user.

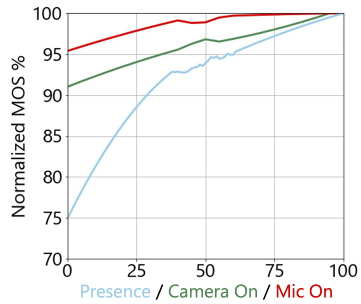


Figure 4: User engagement (X-axis; normalized) correlates with explicit user feedback or MOS.

3.3 User engagement drives experience

Lastly, we evaluate the impact of user engagement metrics on MOS (captured for a subset of calls). The user engagement metrics correlate well with the MOS as seen in Fig. 4. For example, as the users' `Presence` in a meeting increases, the MOS scores improve as well. While `Presence` shows the strongest correlation with MOS, `Cam On` and `Mic On` also show similar trends.

While MOS scores are sampled and delayed, these correlations show that user engagement could be considered as early and more readily available indication of call quality. This finding motivates us to use such user engagement metrics, as an alternative to MOS. §5 discusses our vision using such engagement metrics.

4 CHASING USER EXPERIENCE ON SOCIAL MEDIA

Shifting gears, now we demonstrate how explicit user feedback, shared offline on social media, could augment how ISP networks collect insights on user experience. This is especially important in the context of new networks being rolled out, like SpaceX Starlink LEO network, which aims to offer Internet connectivity globally using thousands of low-flying satellites. As things are in a state of flux, with satellite deployments happening in batches frequently and SpaceX aggressively expanding its coverage footprint, it is critical for them to gauge user sentiment at scale to plan such deployments and service expansion better.

4.1 User sentiment on Starlink

We start by identifying user sentiment on social media. To do so, we analyze user posts on Reddit on `r/Starlink` related to their experience with the SpaceX Starlink network. `r/Starlink` has managed to draw significant participation from enthusiasts, early adopters, and others. There are 372 posts/week (average) on this subreddit. The number of upvotes and comments, which are also strong signals of user activity, are 8,190 and 5,702 per week (average) respectively.

Methodology: For each day between Jan'21 and Dec'22, we analyze the sentiment of individual Reddit posts on `r/Starlink` (text) using Azure's Cognitive Services [5]. We also tie the sentiment to the publicly available announcements. The sentiment analysis service assigns three different scores – positive, negative, and neutral – to each piece of text (posts in this case), which add up to 1. We count the number of posts with strong positive (≥0.7) or negative (≥0.7) scores per day. For each day, we: (a) generate word clouds from all posts published (using NLTK [10]), and (b) discover relevant news articles by searching online for the keywords (top 3 uni-grams from word clouds), with the search query appended with 'Starlink', for the custom date. This pipeline enables the framework to annotate sentiment peaks with news that drive those peaks.

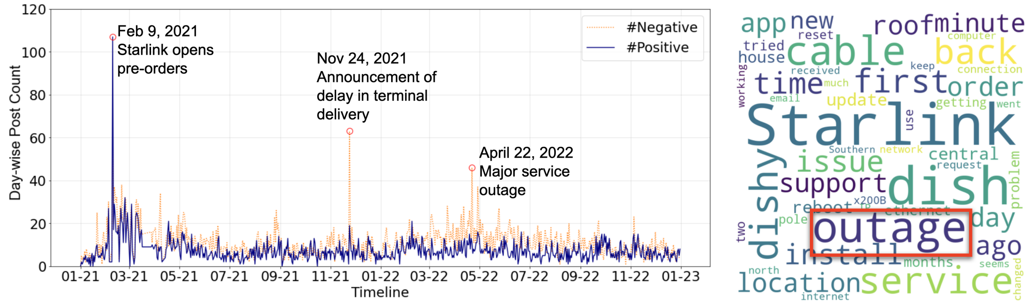
The top three sentiment peaks, as shown in Fig. 5(a), correspond to events of three distinct flavors. On 9^{th} Feb'21, Redditors showed strong positive sentiment towards Starlink opening up pre-ordering of user terminals in the US, Canada, and UK [62]. On 24^{th} Nov'21, SpaceX's email [32] to pre-order customers on delay in terminal delivery led to a negative sentiment peak. Lastly, we found that the third highest peak (22^{nd} Apr'22) is driven by negative sentiment. For this peak, the third most common word in the generated word cloud (see Fig. 5(b)) is *outage*. Interestingly, we could not find any relevant news on an outage for this date, although Redditors from 14 different countries (including ~190 reports from the US) confirmed an outage online! This motivated us to dive deeper into understanding user feedback on outages on this public forum.

To analyze the posts around outages, we first built a dictionary (a manual tedious process at the moment, scanning such posts and online articles on network outages) with keywords related to outages and filtered the Reddit threads containing them. Fig.6 plots the day-wise occurrences of these keywords in these filtered Reddit threads. Note that these occurrences are only counted if the user sentiment attached to them was *negative* to avoid false positives. 7^{th} Jan'22 and 30^{th} Aug'22 have the largest spikes of such keywords in our dataset and correspond to reported outages [34, 40]. Interestingly, there are numerous shorter peaks in Fig.6 over time which correspond to local transient outages. Most of these outages are *not* publicly reported. While Ookla's Downdetector [54] only logs large-scale incidents, in these early days of LEO deployment, it is critical to understand transient small-scale outages too which might be occurring at places due to a complex mix of satellite and earth geometry, weather events, GEO-arc avoidance, deployment planning issues, etc.

We highlight here that we were also able to detect Redditors discussing the *roaming* feature of Starlink almost ~2 weeks before [76, 77] Elon Musk (CEO, SpaceX) announced it on Twitter [51] (and ~3 months before the Starlink public notification [35]) using a systematic pipeline which mines popular discussions (using upvotes and comment numbers). The most common words were *'roaming'* and *'roaming enabled'* in these discussions with a positive user sentiment attached to these threads.

4.2 Following the Shifting Fulcrum

Next, we focus on another illustration where social media posts provide strong hints about the network conditions (downlink speeds in this case). We also show that users slowly get *conditioned* to networks with the perception changing over time for the same conditions.



(a) (b)

Figure 5: (a) Sentiment peaks could be often tied to events of interest that led to these peaks. (b) word cloud for the 3^{rd} highest peak (22^{nd} Apr'22) observed during and after a large-scale service outage otherwise unreported online.

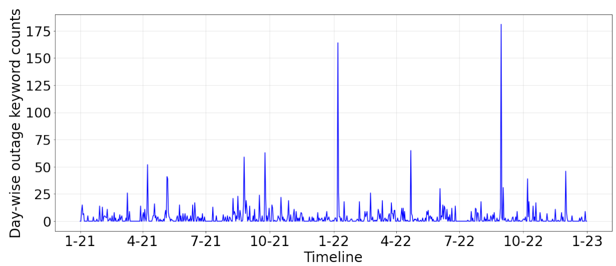


Figure 6: While a few larger outages sparked a lot of discussions on `r/Starlink`, outages with smaller impacts are quite frequent. Threads with positive or neutral sentiments have been filtered out.

To do so, we gather screenshots (or links to them) of network performance test reports from Redditors on `r/Starlink`. The test report screenshots are across test providers like Ookla [55], Fast (powered by Netflix) [22], Starlink itself, and others. We extract uplink speed, downlink speed, latency information, etc. using Azure's Optical Character Recognition (OCR) [6]. We identify ~1750 reports of Starlink speed-tests being shared publicly between Jan, 2021 and Dec, 2022.

Fig. 7 shows the change in observed downlink speeds with time. For each month, we plot the median speeds across all shared screenshots of Starlink speed tests. We annotate the observed speeds with the number of Starlink launches [1, 30, 78, 79] and also the reported number of Starlink users (whenever public information is available) [24, 33, 50, 52, 63–65, 67, 69, 70]. We also plot the monthly median downlink speeds with 95% and 90% of the monthly speed data picked uniformly at random – the plots closely follow each other showing that the observed medians are considerably stable.

We observe that, between Jan and Sep'21, the median downlink speeds increased in general. SpaceX made 14 launches with ~ 60 satellites per launch during this period and the number of users increased from 10K (in Feb) to 90K (in Aug). Further, between Sep'21 and Dec'22, there has been an almost steady decrease in observed speeds although SpaceX launched batches of Starlink satellites 37 times. Note, however, that the number of reported Starlink users increased from 90K to 1M+ during the same period, resulting in a significant increase in downlink demand.

Between Jun and Aug'21, 21K new users started using Starlink with no new launches happening. This is reflected in the sharp decrease in median speeds during the period. Beyond Sep'21, reported bandwidths have decreased almost steadily given the large increase in demand as Starlink service expanded to various countries across the globe.

The wheel of time: With time, the perception of users on network performance changes. In Fig. 7 we also plot (green, dashed) the strong positive sentiment of users on downlink speeds. To do this, we analyze the sentiment of posts (text content) that share Starlink speed-test reports using Azure's Cognitive Services. We identify posts with strong positive (≥0.7) and negative (≥0.7) scores. We define the normalized strong *positive sentiment score (Pos)* as the ratio of total strong positive posts and total (strong positive and negative) posts in a month thus filtering out edge cases when identifying the sentiment is hard.

We observe that *Pos* broadly follows the observed downlink speed trends, but there are interesting exceptions. E.g., while downlink speed is higher in Dec'21 than Apr'21, *Pos* is drastically lower for Dec'21. We believe this is because user sentiment is, in general, a reflection of both short-term and long-term conditioning – users get acclimatized to their current network conditions and give negative sentiment for any degradation in network conditions even if such conditions are better than the past. The exact inverse of this trend is also visible – the downlink speeds decrease between Mar'22 and Dec'22 while the *Pos* improves over time. This demonstrates users getting conditioned to lower speeds, but not necessarily attachment and loyalty to the ISP.

Social media insights are not real-time. Hence, they could only complement (not replace) network measurements by offering signals of user dis/satisfaction that, in turn, might hint at some systematic/broad issues related to deployment, provisioning, configuration, etc.

5 TOWARD USER SIGNALS AS-A-SERVICE

While in the last two sections, we discussed two deliberately disparate user signals, one *implicit* and the other

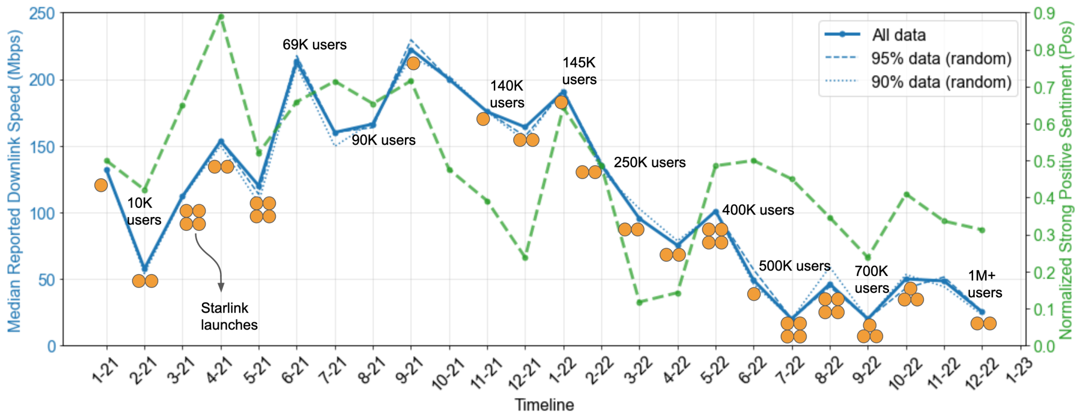


Figure 7: Downlink speeds on Starlink evolve with more launches and customers. User sentiment largely follows the observed speeds. The plot is annotated with the reported number of Starlink users.

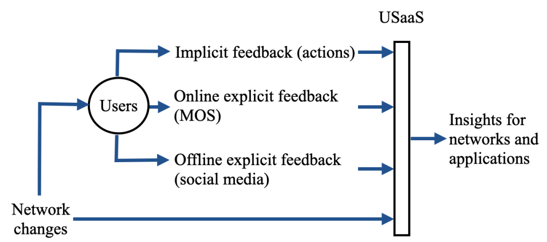


Figure 8: User Signals as-a-Service: Network changes (performance, provisioning, routing, etc.) lead to various implicit and explicit user signals.

explicit, here we discuss our broader vision of User Signals as-a-Service or USaaS that helps both network and networked service providers to consume deep insights on user experience. Network (or network performance) changes might result in users taking actions (§3), having a different experience (MOS), or sharing offline feedback. USaaS collects such user feedback, both online and offline, finds correlations, and shares useful user-centric insights back. The queries could take as input the network/service under consideration, network performance metrics and possible user actions of interest, application QoE metrics, etc. The goal of such a service is not to be an alternative to network diagnostic tools but to augment them by bringing users into the loop.

If SpaceX Starlink, for example, wants to understand how users on their network are perceiving the MS Teams experience, USaaS could filter online user actions and MOS on MS Teams pertaining to Starlink and the offline feedback on the same on social media [3, 20, 29, 31, 57, 71, 81]. Note that privately captured user actions and publicly available social media posts are complimentary to each other. User actions could be used to corroborate the user posts on social media. We do not endorse social media tracking, rather we believe the social media user feedback insights should be aggregated to get useful insights.

Leveraging LLMs and AI/ML USaaS could incorporate generative AI models[12, 56, 75] in its pipeline, effectively summarizing and quantifying contextual user feedback while removing harmful/biased content. Azure's Cognitive Services [5], for example, offers both pre-trained and custom AI models with access control for different stakeholders. We are currently also using AI/ML techniques to predict MOS scores from user engagement and network conditions for MS Teams (omitted for brevity).

6 FUTURE WORK

We are currently exploring and soliciting community feedback on future directions:

Are networks to blame always?: In §3.2, we found that network conditions have a profound impact on user actions. However, there could be confounders that need to be taken care of while correlating network performance with user actions. While Fig. 3 already sheds light on the impact of user platform on user actions, we also found meeting size (number of participants) and long-term conditioning (exposure to network conditions could set expectations) to have (relatively weaker) impact on user actions. An effective USaaS should take into account all such confounders.

Traffic engineering & network planning opportunities: Both service and network providers could proactively act based on USaaS output. If call latency, for example, is the discerning factor affecting user experience on MS Teams, could network resource allocation be tuned online to cater to the demand? Also, could SpaceX change Starlink deployment plans (which LEO satellite shell to deploy next) given the current deployment, footprint, and user sentiment?

The social network bias: Social media is known to have its own bias (users reporting only good/bad things, over-enthusiasm, bias due to socio-demographics [25], etc.). USaaS aims to address such bias by leveraging multi-modal insights (like online user signals, MOS) and aggregation of data across online (social) media.

7 CONCLUSION

We demonstrate examples of using rich user feedback, both implicit and explicit, to generate useful insights for networks and networked services – such as network provisioning and traffic engineering. We then propose a generalized framework, User-Signals-as-a-Service, which consumes such signals, and correlates them with network (performance) changes thus complementing the existing network measurement techniques.

Privacy & ethics: We do not use any PII in our analyses.

REFERENCES

[1] 2023. satellitemap.space. https://satellitemap.space/starlink/launches.html. (2023). [Online; accessed 29-June-2023].

[2] Sumaya Al-Qasem, Ahmad T Al-Hammouri, et al. 2013. Leveraging online social networks for a real-time malware alerting system. In *IEEE Conference on LCN.*

[3] Reddit User alwaysrealapril. 2022. Starlink kept dropping Google Meeting. https://www.reddit.com//r/Starlink/comments/tq909p/starlink_kept_dropping_google_meeting/. (2022). [Online; accessed 29-June-2023].

[4] Minara P Anto, Mejo Antony, KM Muhsina, Nivya Johny, Vinay James, and Aswathy Wilson. 2016. Product rating using sentiment analysis. In *IEEE ICEEOT.*

[5] Azure. 2023. Azure Cognitive Services. https://learn.microsoft.com/en-us/azure/cognitive-services/. (2023). [Online; accessed 29-June-2023].

[6] Azure. 2023. OCR cognitive skill. https://docs.microsoft.com/en-us/azure/search/cognitive-search-skill-ocr. (2023). [Online; accessed 29-June-2023].

[7] Leif Azzopardi, Diane Kelly, and Kathy Brennan. 2013. How query cost affects search behavior. In *ACM SIGIR conference on Research and development in information retrieval.*

[8] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang. 2013. Developing a predictive model of quality of experience for internet video. *ACM SIGCOMM CCR.*

[9] Adam Bermingham and Alan Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In *SAAIP.*

[10] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[11] Ilker Nadi Bozkurt, Anthony Aguirre, Balakrishnan Chandrasekaran, P. Brighten Godfrey, Gregory Laughlin, Bruce Maggs, and Ankit Singla. 2017. Why is the Internet so slow?!. In *Springer PAM.*

[12] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems.*

[13] Jake Brutlag. 2009. Speed Matters for Google Web Search. http://goo.gl/vJq1lx. (2009). [Online; accessed 29-June-2023].

[14] Jake D Brutlag, Hilary Hutchinson, and Maria Stone. 2008. User preference and search engine latency. (2008).

[15] Matt Calder, Ryan Gao, Manuel Schröder, Ryan Stewart, Jitendra Padhye, Ratul Mahajan, Ganesh Ananthanarayanan, and Ethan Katz-Bassett. 2018. Odin: Microsoft's Scalable Fault-Tolerant CDN Measurement System.. In *Usenix NSDI.*

[16] Nathanael Chambers, Ben Fry, and James McMasters. 2018. Detecting denial-of-service attacks from social media text: Applying nlp to computer security. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1.*

[17] Brent Chun, David Culler, Timothy Roscoe, Andy Bavier, Larry Peterson, Mike Wawrzoniak, and Mic Bowman. 2003. Planetlab: an overlay testbed for broad-coverage services. *ACM SIGCOMM CCR.*

[18] Anita Crescenzi, Diane Kelly, and Leif Azzopardi. 2015. Time pressure and system delays in information search. In *ACM SIGIR Conference on Research and Development in Information Retrieval.*

[19] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting effectiveness and inclusiveness in remote collaboration. *Proceedings of the ACM on Human-Computer Interaction.*

[20] Reddit User DaveTV-71. 2022. MS Teams on StarLink. https://www.reddit.com//r/Starlink/comments/tixtjc/ms_teams_on_starlink/. (2022). [Online; accessed 29-June-2023].

[21] Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. 2011. Understanding the impact of video quality on user engagement. *ACM SIGCOMM CCR.*

[22] FAST.com. 2023. Internet Speed Test. https://fast.com/. (2023). [Online; accessed 29-June-2023].

[23] Alessandro Finamore, Marco Mellia, Maurizio M Munafo, Ruben Torres, and Sanjay G Rao. 2011. Youtube everywhere: Impact of device and infrastructure synergies on user experience. In *ACM IMC.*

[24] Chris Forrester. 2022. Starlink has 700,000 subs. https://advanced-television.com/2022/09/19/starlink-has-700000-subs/. (2022). [Online; accessed 29-June-2023].

[25] Eszter Hargittai. 2020. Potential biases in big data: Omitted voices on social media. *Social Science Computer Review.*

[26] Tobias Hoßfeld, Michael Seufert, Matthias Hirth, Thomas Zinner, Phuoc Tran-Gia, and Raimund Schatz. 2011. Quantification of YouTube QoE via crowdsourcing. In *IEEE International Symposium on Multimedia.*

[27] Wenling Hsu, Guy Jacobsen, Yu Jin, and Ann Skudlark. 2011. *Using social media data to understand mobile customer experience and behavior.* European Regional ITS Conference, Budapest: Innovative ICT Applications - Emerging Regulatory, Economic and Policy Issues 52180. https://ideas.repec.org/p/zbw/itse11/52180.html

[28] Daniel Imms. 2014. Speed Index: Measuring Page Load Time a Different Way. https://www.sitepoint.com/speed-index-measuring-page-load-time-different-way/. (2014). [Online; accessed 29-June-2023].

[29] Reddit User jcsqaure1900. 2021. Bittersweet Relationship with Starlink. https://www.reddit.com//r/Starlink/comments/neke1g/bittersweet_relationship_with_starlink/. (2021). [Online; accessed 29-June-2023].

[30] Jonathan's Space Pages. 2023. Starlink Statistics. https://planet4589.org/space/con/star/stats.html. (2023). [Online; accessed 29-June-2023].

[31] Reddit User Jorpops. 2023. Video Conferencing. https://www.reddit.com//r/Starlink/comments/zhtj1w/video_conferencing/. (2023). [Online; accessed 29-June-2023].

[32] Michael Kan. 2021. Starlink Disappoints Pre-Order Customers by Pushing Back Delivery Times. https://tinyurl.com/2p8jenhj. (2021). [Online; accessed 29-June-2023].

[33] Michael Kan. 2022. SpaceX Beats Annual Launch Record As It Preps More Starlink Satellites. https://www.pcmag.com/news/spacex-beats-annual-launch-record-as-it-preps-more-starlink-satellites. (2022). [Online; accessed 29-June-2023].

[34] Michael Kan. 2022. SpaceX's Starlink Suffers Global Outage. https://www.pcmag.com/news/spacexs-starlink-suffers-global-outage. (2022). [Online; accessed 29-June-2023].

[35] Michael Kan. 2022. Starlink Becomes Movable With New ́PortabilitýOption. https://tinyurl.com/26kbj223. (2022). [Online; accessed 29-June-2023].

[36] Vytautas Karalevicius, Niels Degrande, and Jochen De Weerdt. 2018. Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance.*

[37] Conor Kelton, Jihoon Ryoo, Aruna Balasubramanian, and Samir R Das. 2017. Improving user perceived page load times using gaze. In *USENIX NSDI.*

[38] Rupinder Paul Khandpur, Taoran Ji, Steve Jan, Gang Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Crowdsourcing cybersecurity: Cyber attack detection using social media. In *ACM CIKM.*

[39] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, et al. 2017. The quic transport protocol: Design and internet-scale deployment. In *ACM SIGCOMM.*

[40] Tom Li. 2022. Starlink internet is experiencing worldwide service interruptions. https://tinyurl.com/yckemj8b. (2022). [Online; accessed 29-June-2023].

[41] Greg Linden. 2006. Geeking with Greg. http://glinden.blogspot.com/2006/11/marissa-mayer-at-web-20.html. (2006). [Online; accessed 29-June-2023].

[42] Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies.*

[43] Antonis Manousis, Harshil Shah, Henry Milner, Yan Li, Hui Zhang, and Vyas Sekar. 2021. The shape of view: an alert system for video viewership anomalies. In *ACM IMC*.
[44] Yogev Matalon, Ofir Magdaci, Adam Almozlino, and Dan Yamin. 2021. Using sentiment analysis to predict opinion inversion in Tweets of political communication. *Scientific reports* (2021).
[45] Marko Matulin, Štefica Mrvelj, Borna Abramović, Tomislav Šoštarić, and Marko Čejvan. 2021. User quality of experience comparison between Skype, Microsoft teams and zoom videoconferencing tools. In *International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures*. Springer.
[46] David Maxwell and Leif Azzopardi. 2014. Stuck in traffic: How temporal delays affect search behaviour. In *Information interaction in context symposium*.
[47] Oliver Michel, Satadal Sengupta, Hyojoon Kim, Ravi Netravali, and Jennifer Rexford. 2022. Enabling passive measurement of zoom performance in production networks. In *ACM IMC*.
[48] Karan Mitra, Arkady Zaslavsky, and Christer Åhlund. 2013. Context-aware QoE modelling, measurement, and prediction in mobile computing systems. *IEEE Transactions on Mobile Computing*.
[49] Marti Motoyama, Brendan Meeder, Kirill Levchenko, Geoffrey M Voelker, and Stefan Savage. 2010. Measuring online service availability using twitter. In *WOSN*.
[50] Elon Musk. 2021. Starlink simultaneously active users just exceeded the strategically important threshold of 69,420 last night! https://twitter.com/elonmusk/status/1408558492009566214. (2021). [Online; accessed 29-June-2023].
[51] Elon Musk. 2022. Mobile roaming enabled. https://twitter.com/elonmusk/status/1499442132402130951. (2022). [Online; accessed 29-June-2023].
[52] Elon Musk. 2022. Over 250k Starlink user terminals. https://twitter.com/elonmusk/status/1493358044989767683. (2022). [Online; accessed 29-June-2023].
[53] Hyunwoo Nam, Kyung-Hwa Kim, and Henning Schulzrinne. 2016. QoE matters more than QoS: Why people stop watching cat videos. In *IEEE INFOCOM*.
[54] Ookla. 2022. Downdetector. https://downdetector.com/. (2022). [Online; accessed 20-June-2022].
[55] Ookla. 2023. Speedtest by Ookla - The Global Broadband Speed Test. https://www.speedtest.net/. (2023). [Online; accessed 29-June-2023].
[56] OpenAI. 2023. GPT-4 Technical Report. (2023). arXiv:cs.CL/2303.08774
[57] Reddit User PopeJohnThompsonII. 2022. Zoom calls and outages. https://www.reddit.com//r/Starlink/comments/v5w6zq/zooms_calls_and_outages_ive_got_a_decent/. (2022). [Online; accessed 29-June-2023].
[58] Tongqing Qiu, Junlan Feng, Zihui Ge, Jia Wang, Jun Xu, and Jennifer Yates. 2010. Listen to me if you can: tracking user experience of mobile network on social media. In *ACM IMC*.
[59] Alan Ritter, Evan Wright, William Casey, and Tom Mitchell. 2015. Weakly supervised extraction of computer security events from twitter. In *ACM WWW*.
[60] Carl Sabottke, Octavian Suciu, and Tudor Dumitraș. 2015. Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting {Real-World} Exploits. In *USENIX Security*.
[61] Muhammad Zubair Shafiq, Jeffrey Erman, Lusheng Ji, Alex X Liu, Jeffrey Pang, and Jia Wang. 2014. Understanding the impact of network dynamics on mobile video user engagement. *ACM SIGMETRICS*.
[62] Michael Sheetz. 2021. SpaceX begins accepting $99 preorders for its Starlink satellite internet service as Musk eyes IPO. https://tinyurl.com/ywy7h424. (2021). [Online; accessed 29-June-2023].
[63] Michael Sheetz. 2021. SpaceX says Starlink has about 90,000 users as the internet service gains subscribers. https://tinyurl.com/5u7vhvnu. (2021). [Online; accessed 29-June-2023].
[64] Michael Sheetz. 2022. SpaceX's Starlink internet service has more than 145,000 users so far. https://tinyurl.com/33p7eexp. (2022). [Online; accessed 29-June-2023].
[65] Michael Sheetz. 2022. SpaceX's Starlink satellite internet surpasses 400,000 subscribers globally. https://tinyurl.com/yc5rhh3n. (2022). [Online; accessed 29-June-2023].
[66] Kai Shu, Amy Sliva, Justin Sampson, and Huan Liu. 2018. Understanding cyber attack behaviors with sentiment information on social media. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer.
[67] SpaceX. 2022. Starlink now has more than 1,000,000 active subscribers. https://twitter.com/SpaceX/status/1604872936976154624. (2022). [Online; accessed 29-June-2023].
[68] SpaceX. 2023. Starlink. https://www.starlink.com/. (2023). [Online; accessed 29-June-2023].
[69] SpaceX. 2023. Thank you to our 1.5M+ customers around the world! https://twitter.com/Starlink/status/1654673695007457280. (2023). [Online; accessed 29-June-2023].
[70] Starlink Services LLC. 2021. FCC Filing - Application for ETC Designation. https://tinyurl.com/msy86n5h. (2021). [Online; accessed 29-June-2023].
[71] Reddit User symphax. 2022. Google Meet? https://www.reddit.com/r/Starlink/comments/nun6gf/google_meet/. (2022). [Online; accessed 29-June-2023].
[72] Kei Takeshita, Masahiro Yokota, and Ken Nishimatsu. 2015. Early network failure detection system by analyzing Twitter data. In *IFIP/IEEE International Symposium on Integrated Network Management (IM)*.
[73] Jaime Teevan, Kevyn Collins-Thompson, Ryen W White, Susan T Dumais, and Yubin Kim. 2013. Slow search: Information retrieval without time constraints. In *Symposium on Human-Computer Interaction and Information Retrieval*.
[74] Parth Thakkar, Rohan Saxena, and Venkata N Padmanabhan. 2021. AutoSens: inferring latency sensitivity of user activity through natural experiments. In *ACM IMC*.
[75] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
[76] Reddit User tuckstruck. 2022. Starlink Roaming is working. https://www.reddit.com/r/Starlink/comments/ssj464/starlink_roaming_is_working/. (2022). [Online; accessed 29-June-2023].
[77] Reddit User tuckstruck. 2022. The roaming function isn't limited by state borders. https://www.reddit.com/r/Starlink/comments/sz3cvd/i_was_wrong_the_roaming_function_isnt_limited_by/. (2022). [Online; accessed 29-June-2023].
[78] Wikipedia. 2023. List of Starlink and Starshield launches. https://en.wikipedia.org/wiki/List_of_Starlink_and_Starshield_launches#Starlink_Launches. (2023). [Online; accessed 29-June-2023].
[79] Wikipedia. 2023. List of Starlink launches. https://en.wikipedia.org/wiki/List_of_Starlink_launches. (2023). [Online; accessed 29-June-2023].
[80] Bowei Yang, Weisi Guo, Bozhong Chen, Guangpu Yang, and Jie Zhang. 2016. Estimating mobile traffic demand using Twitter. *IEEE Wireless Communications Letters*.
[81] Reddit User YellowVivid1477. 2021. Done with the disconnects while on Teams or VPN . https://www.reddit.com//r/Starlink/comments/mn8j3j/done_with_the_disconnects_while_on_teams_or_vpn/. (2021). [Online; accessed 29-June-2023].