Modeling and Generation of Realistic Network Activity

Stefan Tschimben
University of Colorado Boulder
stefan.tschimben@colorado.edu

Isabella Bates
University of Colorado Boulder
isabella.bates@colorado.edu

James H. Curry
University of Colorado Boulder
james.h.curry@colorado.edu

Keith D. Gremban
University of Colorado Boulder
keith.gremban@colorado.edu

Alexandra Siegel
University of Colorado Boulder
alexandra.siegel@colorado.edu

***Abstract*—The growing quantity of wireless network activity generated every second of every day creates challenges for network operators, such as detecting anomalies and providing sufficient capacity. This same network activity also creates opportunities for Smart and Connected Systems (SCSs) to adapt to changing population dynamics, detect and proactively adapt to unexpected events such as public safety threats, traffic jams, or adverse weather events, for example. The GHOST project is researching the challenges of modeling, analyzing, and generating patterns of network activity. The GHOST project has demonstrated that Nonnegative Matrix Factorization (NMF) provides a robust mechanism for modeling network activity patterns that can be used to generate realistic network activity. The GHOST team has further demonstrated the capability for injecting programmed activity patterns into a live, functioning wireless network.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Every day, 2.45 billion pieces of content are shared on Facebook, nearly 500 million Tweets are shared on Twitter, and 3.5 billion Snaps are shared on Snapchat [1]. The daily amount of data generated by social media alone amounts to 41.72 million terabytes. Internet traffic conducted on mobile phones worldwide has risen from 0.7% in 2009 to over 50% of all traffic by 2023, further accelerated by 5G [2]. Africa leads mobile traffic with over 70% of web page views generated by mobile devices. In 2017 Cisco estimated that average network traffic per capita would reach 49.8 GB per month by 2022 with global IP network traffic reaching 332,7 Exabytes per month [3].

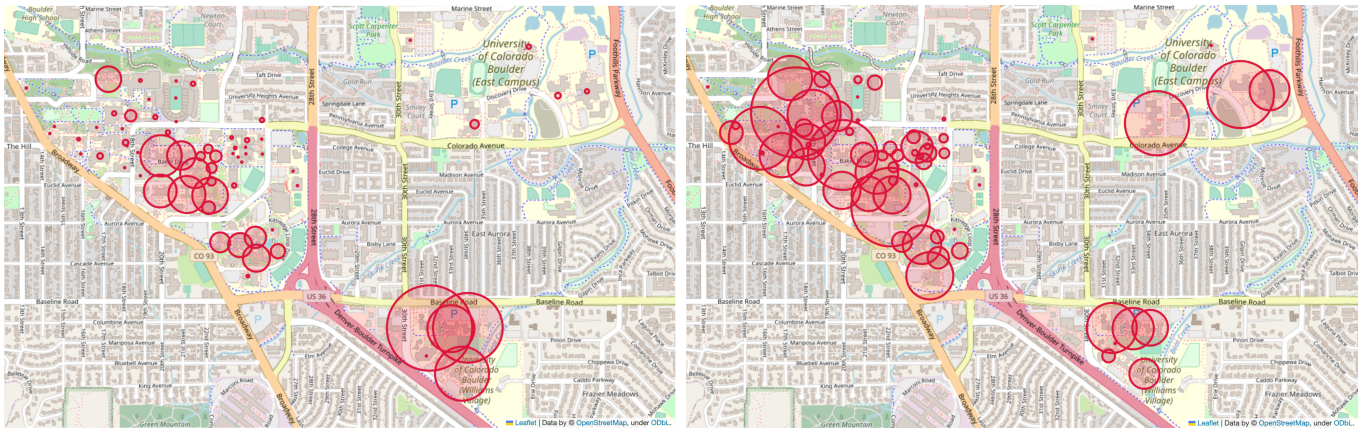
This large quantity of data not only creates a challenge for network operators to detect anomalies and provide sufficient capacity but also creates opportunities for Smart and Connected Systems (SCSs) to adapt to changing population dynamics. NIST defines a SCS as a collection of interrelated systems that can collect and analyze data to make decisions with and without human interaction [4]. Without human interaction and access to personal identifiable information (PII), SCSs have the capability to use the monitored network activity to adapt to regular patterns of usage, or detect and predict significant and potentially challenging or threatening events while also preserving user anonymity. Examples of detectable events include large gatherings like celebrations of sports victories; protests with the potential to become riots; public safety threats; traffic jams; or adverse weather events.

The University of Colorado Boulder (UCB) is a medium-scale SCS. UCB enrolls over 36,000 students, has over 16,000 faculty and staff, and operates its own police force, campus-wide Wi-Fi network, and soon its own private 5G network. Figure 1 demonstrates Wi-Fi activity on the UCB campus during a weekday at two different times and highlights some of the patterns that can be discovered in network activity. The size of the circles represent the number of active WiFi connections in each campus building and illustrates the shift in WiFi activity from dormitories in the morning at the lower right of each frame to academic buildings at mid-day in the upper left and upper right of each frame.

Figure 1 highlights how network activity can reveal significant information about the behavior of a population or an organization. It is possible to not only uniquely identify individual users, but entire groups of users using the metadata generated by their devices and their online behavior in combination with the generated network traffic. Recent examples of using such individual and group metadata include the identification of Russian officers using their cellular connection during the war in Ukraine [6], tracking the movement of protests using social media [7], revealing the exact location of Polish troops by using dating apps on their phones [8], drawing GPS maps of US military bases due to troops using Fitbits [9], and AI being used to monitor student protests on social media [10].

These long-term observations of network activity can be used to create patterns of life that can predict regular behavior as well as detect changes in behavior and can be used to as input to models seeking to generate realistic network activity. Conversely, generating realistic traffic is not that simple. Not only does the artificial network traffic need to be responsive

This research was supported in part by the National Science Foundation (NSF) and the Office of the Under Secretary of Defense for Research and Engineering (OUSD(RE)) under NSF Award No. 2226426 as part of the NSF Convergence Accelerator Track G: Securely Operating Through 5G Infrastructure Program. Student support was provided in part from Sandia National Laboratory under Purchase Order 2234364.



(a) User count in the morning.

(b) User count at noon.

Fig. 1: Wi-Fi user count showing student movement from dormitories to academic buildings [5].

to network conditions, but the traffic generator also needs to capture realistic packet interactions. Most traffic generation is intended for capacity planning, bandwidth measurements, and simulations, where mostly aggregate characteristics matter. Components such as accurate packet inter-arrival rates, packet size distributions, broken packets, and port distributions are of less importance [11]. Without modeling user behavior and realistic network activity first, it is impossible to capture the realistic effects in the generated traffic.

Therefore, the objective of this research is to first model, analyze, and generate realistic network activity corresponding to medium to large-scale human behavior in natural environments. Real user behavior needs to be modeled first to identify baseline activity, regular patterns, and typical events, which can then be used to detect and predict anomalies that could indicate threats or significant changes in behavior. Finally, network activity patterns are generated that mimic and are indistinguishable from real large-scale network activity patterns. The results will provide a proof-of-concept for other organizations and SCSs to monitor network activity for threats and dynamically optimize system parameters. Additionally, realistic activity patterns can be used to drive system architecture and design as well as upgrades to meet changing demands.

II. Background

This research continues research from 2020 that began with a study of network activity on the UCB campus Wi-Fi using anonymized data [5]. At UCB, all Wi-Fi access points (APs) are geospatially registered and report statistics such as the number of connected devices, occupied bandwidth, and requested destination addresses. Before being made available for research, all data is converted to a summary format of user counts that does not contain any information that could be used to identify or track individuals. Still, the number of Wi-Fi connections at various locations is closely correlated to population density. Hence, variations in the number of connections reveal population behavior over time.

[12] explored the use of non-negative matrix factorization (NMF) for modeling network activity. Using Wi-Fi data from the UCB campus library, they noted a diurnal pattern of activity that also distinguished weekends from weekdays. They modeled the pattern using a four component NMF decomposition, and showed that the general pattern of behavior could be reconstructed from the NMF decomposition.

Patterns of life in general consists of three components [13]: (1) the event describing a significant change in state; (2) an activity describing what an individual or a group is doing; and (3) a sequence of observable actions establishing a behavior. Together, they establish a pattern that can be considered normality. Having access to such a pattern of normality enables an SCS to understand how and when people engage in commerce, when activity is shifted to recreation, or when and where religious activities are conducted. However, defining such a pattern is not without its challenges. While some research has attempted to measure communications traffic patterns using machine learning, most of it is focused on short-term predictions such as using Hidden Markov Models and Long Short Term Memory models to detect anomalies [14] or using network flow information to predict network traffic bursts [15]. More recently, the use of transformers has gained in popularity in various research areas, but has not found many applications in pattern of life analysis yet and is beyond the scope of this research. Additionally, most ordinary neural networks do not predict network activity well and most traffic analysis tools like Wireshark or Netflow don't have the necessary power to handle large-scale network activity [16].

We therefore build on the results of [5] and [12] to construct realistic models of large-scale and long-term network activity using Wi-Fi user counts spanning July 2019 to the beginning of Covid in March 2020. Those models are then used to compute expected network activity and implement software on network devices to generate actual network activity. In the next section, we describe the methodology used, including details on NMF, the underlying data set, and details on how the generated

model could be used to replicate the user connection count with a single device.

III. Methodology

A. Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) factors a matrix D into the product of two lower rank matrices, W and H. An important limitation is that the input matrix D must be completely non-negative. Let the data matrix D be an $m \times n$ matrix where D_{ij} is greater than or equal to 0. NMF decomposes D into two matrices such that:

$$D \approx W \cdot H \tag{1}$$

Every element of matrices W_{ij} and H_{ij} are greater than or equal to 0, where the so called inner dimension k satisfies $k < min\,(x, y)$. W is an $m \times k$ basis matrix that holds the patterns, or clusters, discovered from the dataset. H is a $k \times n$ coefficient matrix that holds the corresponding weight, or importance, of each pattern in W.

$W \cdot H$ is an approximation of D. To optimize the approximation, an algorithm from Python's scikit-learn package, known as the "multiplicative update rule", can be used. This algorithm calculates W and H by measuring the error between the matrix D and the product of its factors W and H, on the basis of Euclidean distance:

$$||D - WH|| \tag{2}$$

The Euclidean distance is non-increasing under the update rules. Using this, two equations can be derived to update matrices W and H.

To update the W matrix:

$$W_{ic} \leftarrow W_{ic} \frac{\left(W^T D\right)_{ic}}{\left(W^T W D\right)_{ic}} \tag{3}$$

To update the H matrix:

$$H_{cj} \leftarrow H_{cj} \frac{\left(H^T D\right)_{cj}}{\left(H^T H D\right)_{cj}} \tag{4}$$

The matrices W and H are first initialized with random values. Using equations 2, 3, and 4, values of W and H are then simultaneously updated and the algorithm is run iteratively until we find values for W and H that minimize the cost function (see Equation 2). In other words, the process of computing new matrices for W and H, and the resulting error using Equation 2, is repeated until W and H converge. Thus, the multiplicative update rules will modify the initial values of W and H until the approximation error converges or until the product of them approaches D.

NMF has a rich history dating back to the early 1960s in the fields of analytical chemistry and remote sensing [17]. The work done in the 60s was not directly referred to as NMF, although it was equivalent. In fact, NMF did not appear in its present form until 1994 when a paper was published on positive matrix factorization [18]. The true explosion of the topic can be directly traced to a Nature article by Lee and Seung in 1999 when data compression and feature extraction were shown to be intrinsic properties of NMF [19].

Using NMF to model group behavior based on comprehensive and anonymized network data from the UCB campus Wi-Fi network has several advantages: data often represent the integrated results of interrelated variables acting together, which could also be replaced by a lower-dimensional representation [20]. Additionally, a lot of real data are often non-negative. Therefore, to make data more intuitive and to eliminate conflicts with the data's underlying physical reality, low rank data should also be positive. NMF's non-negative constraint is achieved by not allowing W and H to have an arbitrary sign. This constraint results in NMF only allowing additive and not subtractive combinations, leading to a parts-based representation with localized features that better represent intuitive notions [19].

NMF also has some downsides: due to its non-convexity, NMF can only guarantee locally optimal solutions. Additionally, NMF has a convergence issue by requiring that the inner dimension k must be provided. This method results in different values of k leading to different factorization results and a lack of unique solutions [20], [21]. However, despite its downsides, NMF is highly versatile making it possible to be used with many different types of applications and better suited to model long-term trends. Different from other dimensionality reduction algorithms like principal component analysis (PCA) and singular value decomposition (SVD), NMF has the advantage of creating sparse and easily interpretable features as its parts-based representation is more intuitive than those achieved by PCA and SVD, which take holistic approaches to dimensionality reduction. This advantage can be further strengthened by the use of k-means to intialize NMF [21], [22]. In the end, the intuitive and realistic models created by NMF could in a further step be used as an input to generative models, such as GANs, which then create new network traffic or activity based on the model created by NMF.

B. Dataset and Data Pre-processing

The data used in this study was provided by the UCB Office of Information Technologies (OIT). To calculate the total number of active users in each building, the data from multiple Wi-Fi access points (APs) within the building was combined. The APs at UCB are set up in a configuration where several APs provide coverage for a large area, and devices can roam between them effortlessly. As a result, each device is counted by only one AP, ensuring that duplicate counts for a building are avoided.

By collecting connected device counts throughout the day, time series data is obtained that will be analyzed using NMF. To protect privacy, OIT removed all identifying information from the data before releasing it, leaving only the active user counts. This anonymity is one of the appealing aspects of this dataset, as it minimizes privacy concerns for network users and does not violate any privacy rights.

Unfortunately, the sampling frequency of the UCB Wi-Fi network system is rather irregular. Without any control

over when device counts are reported, the intervals at which user counts are sampled vary irregularly over time and space, usually ranging from 3 to 12 minutes. Additionally, irregular outages in the data reporting, lasting for hours or even days, can lead to unfortunate gaps in the data. To ensure consistent interpretation of the data, a standard sampling frequency is preferred before applying any matrix factorization. To improve data consistency, linear interpolation was performed on the dataset, to yield a ten-minute interpolation of the data.

Section 4 of this paper will illustrate the analysis of data from UCB's Norlin Library, Williams Village, and the Ann and H.J. Smead Aerospace Engineering building.

C. Matrix Interpolation

With the ten-minute interpolation, 144 data points are acquired per day. Each day is treated as a sample and arranged chronologically as columns in the data matrix D. As a result, the factors W and H possess straightforward interpretations. Specifically, the columns of W represent the activity patterns, while the rows of H represent the weights assigned to the activity patterns in the original data. It is crucial to have the columns of the data matrix aligned with the same time points throughout the day to ensure reliable comparisons. Without interpolation, the columns would likely have different dimensions.

This embedding of data assists in constraining the factorization process and enhances NMF's capability to generate patterns that can be physically interpreted. Given any day in the training data, we can reconstruct its pattern by combining the columns of W, each of which represents a pattern of network activity, in a linear manner, with the coefficients (weights) represented by a corresponding column of H for the desired day. Additionally, it is noteworthy that we can assess the significance of each column of W on a specific day by examining the weights within the H matrix.

The model created using NMF can then be used to generate realistic network activity by adding user counts to the UCB network in specific locations mimicking real user behavior.

D. Multiple Device Identities

While normally each Wi-Fi connected device is represented by a single network interface controller (NIC) and a single media access control (MAC) address, it would be unfeasible to carry around hundreds of devices to recreate the observed movement of user counts or create unexpected, but realistic looking movement in unusual location.[1] Instead of pulling a cartload of devices across campus, a single Raspberry Pi 4 together with a single USB Wi-Fi adapter is used to simulate multiple devices authenticated and connecting to UCB access points (AP). For this to work, it is essential that the Wi-Fi adapter supports monitor mode, which is why the CanaKit Raspberry Pi WiFi Wireless Adapter with the Ralink 5370 chipset was used [24]. While it does support monitor mode and a wide range of operating systems, it is limited to the 2.4 GHz Wi-Fi spectrum. With the CanaKit adapter, the Raspberry Pi OS has access to capturing all Wi-Fi traffic. Scapy, an interactive packet manipulation library written in Python, is then used to manipulate the captured packets.

To make one Raspberry Pi appear as multiple devices on a Wi-Fi AP, the CanaKit adapter is first put into monitor mode and the Wi-Fi interface is reconfigured in Raspberry Pi OS to use the target AP's control channel. The control channel can be either determined by finding the channel on which the AP sends out its beacons or by using Scapy to transmit probe requests and capturing the probe response. The Raspberry Pi is also pre-loaded with a file containing a list of valid first 3 MAC address octets. 48-bit MAC addresses, usually represented by 6 octets in the form of hexadecimal 2-digit numbers, consist of 2 parts: the organizationally unique identifier (OUI) and the NIC identifier. The first 3 octets are randomly pulled from a list of valid OUIs to prevent access points from rejecting an unknown OUI. The second pair representing the NIC on the other hand is completely randomly generated.

With the interface configured and a list of valid random MAC addresses generated, the process begins by using Scapy to capture a beacon frame and reading the frame's "supported rates" field as the AP expects the correct supported rates in any subsequent requests. The extracted supported rates are then inserted into authentication and association requests, which are then transmitted in the same order. The AP then replies with an association response containing either a success message or the reason for the association failure.

At this point, each randomly generated MAC address is associated with the AP and a network operator can see that all of these devices are in the process of connecting to the AP. For the network operator the process however is not completed until each MAC address is associated with an IP address. Two methods can be used to complete this final step, depending on the AP's configuration: (1) if the AP accepts static IP address, a DHCP inform packet can be used to inform the AP of the specific MAC address' selected IP address. While this can admittedly lead to collisions if a selected IP address has already been taken by a different device on the network, the DHCP inform process is much simpler than the second method and can simply be repeated with a different IP address until a free one has been found. (2) If the AP is configured to not allow static IP addresses, a DHCP discover message has to be sent first, followed by a capturing a DHCP offer response and sending the DHCP request. For method (2) 2 requests have to be sent (DHCP discover and DHCP request) and 2 responses have to be captured (DHCP offer and the acknowledgment) per MAC address.

Once this step has been completed for each MAC address, each ghost device is associated in the network with an IP address and appears to the network operator as connected to the network and as a physical device that can be tracked. At this point, each MAC and IP address combination can be used to transmit packets, access the network, and make it look like multiple devices are using the network from this location.

[1] Although, something similar has been demonstrated by an artist in Berlin using a cartload of phones to appear as a traffic cluster on Google Maps, causing Google Maps to divert road users [23].

IV. Results

[12] and [5] demonstrated that a particular pattern of network activity could be modeled and accurately reconstructed using NMF. Our GHOST project demonstrated that NMF decompositions could be used to generate scalable patterns of network activity that could be additively injected into an existing pattern to alter its appearance. We call the injected pattern GHOST traffic.

Figure 2 illustrates the basic GHOST concept. The figure shows in blue the original Wi-Fi activity for a randomly selected Saturday at the UCB library. The red line is a scaled NMF model for a typical weekday at the UCB library. The purple line shows the user count activity that would be observed by the network operator by injecting GHOST device activity into the recordings, effectively changing the Saturday pattern in shape and scale.

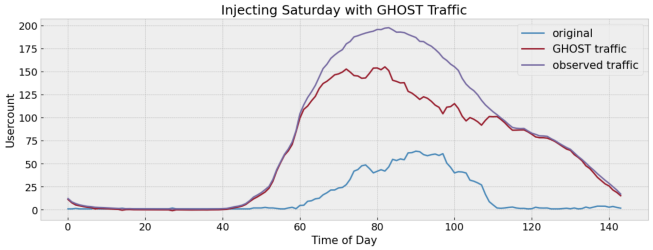


Fig. 2: *Injecting GHOST traffic developed from an NMF model changes the observed pattern of network activity.*

Figure 3 shows how GHOST models can be used to generate realistic network activity on a larger scale by injecting GHOST devices into recordings of Wi-Fi activity. In the figure, the blue line shows the original Wi-Fi activity over an entire week from Sunday to the next Saturday at the UCB library. The figure clearly shows diurnal activity peaks for each day of the week. Weekend activity is clearly different from weekday activity. The red lines show the injected activity data for Sunday and Saturday derived from the NMF model of a weekday. The purple line shows the activity that would be observed by the network operator. Injected GHOST activity has effectively made weekend activity match weekday activity.

Figure 4 demonstrates the view of a network operator during tests injecting GHOST device activity onto a real network and access count data collected by the UCB Wi-Fi network. The experiment was conducted on a Thursday afternoon on the third floor of the Ann and H.J. Smead Aerospace Engineering building. As previous models highlighted, Thursday already shows declining activity. Additionally, the third floor of the Aerospace Engineering building has less traffic than for example the first floor, resulting in a fairly flat activity curve. However, the data received from the UCB Wi-Fi system clearly shows that the Raspberry Pi system was able to make it look like much more activity was observed than on a typical Thursday afternoon and night. Note that the baseline of 2 devices accessing the Wi-Fi network at all times is caused by nearby vending machines.

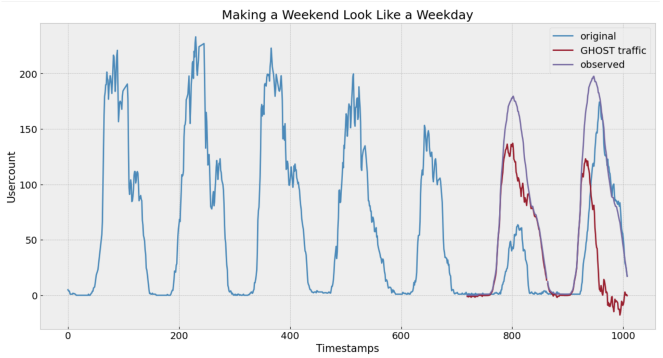


Fig. 3: *Injecting GHOST traffic makes weekend network activity look like weekday activity.*

V. Conclusion and Future Work

Understanding and adapting to changing population dynamics is an important capability for SCSs. Wireless network activity is a data source closely correlated with population dynamics. UCB campus Wi-Fi activity clearly correlates closely with the geospatial movements of students, faculty, and staff over time. The GHOST project, using NMF decompositions of Wi-Fi activity demonstrated that realistic network activity can be injected into observed data streams to alter the observed patterns. Further, the GHOST project demonstrated that a single device can be programmed to appear as multiple devices to the network operator. Integrated, these capabilities are a first step towards enabling SCS owners and operators to identify typical patterns of activity, identify anomalies, and generate realistic activity to inform system design and test / evaluate system performance.

Future research relying on the generated realistic models includes the following:

- Development of a library of parameterized NMF models of typical campus activities, such as: conferences, concerts, sports events, move-in/move-out weeks, semester breaks, summer session, and more. The library will also include models of the transition paths between events.
- Identification and modeling of example anomalous events such as public safety alerts or weather alerts.
- Development of a monitoring capability to detect particular events or anomalies.
- Combine the traffic generation capability with the capability of a single device to present itself as multiple devices.

Acknowledgment

The GHOST project thanks Mr. Glenn Rodriguez of the UCB OIT for providing access to campus Wi-Fi data.

References

[1] P. Taylor. (2022) Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. Online; accessed 05-May-2023. [Online]. Available: https://www.statista.com/statistics/871513/worldwide-data-created/

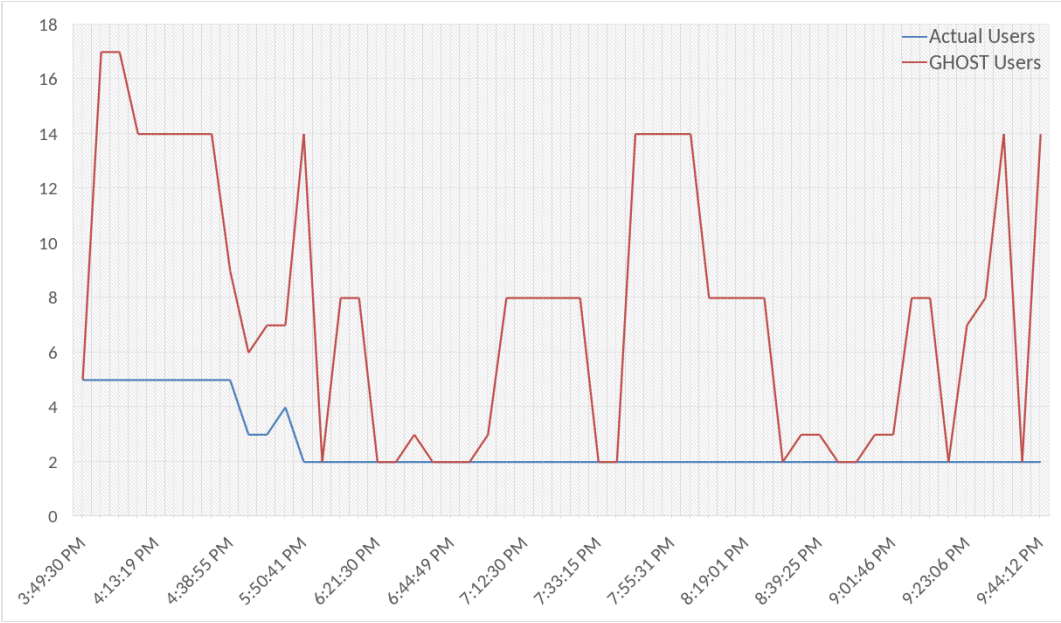


Fig. 4: Access Point inspection highlighting GHOST users generated by a single device.

[2] T. Bianchi. (2023) Mobile internet traffic as percentage of total web traffic in january 2023, by region. Online; accessed 05-May-2023. [Online]. Available: https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/

[3] (2017) VNI complete forecast highlights. Online; accessed 05-May-2023. [Online]. Available: https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf

[4] N. I. T. Laboratory". (2018) Smart and connected systems. Online; accessed 05-May-2023. [Online]. Available: https://www.nist.gov/programs-projects/smart-and-connected-systems

[5] J. Mcgrath, A. Davis, J. Curry, O. Gartner, G. Rodrigues, S. Spielman, and D. Massey, "Weather of the dorm wifi ecosystem at the university of colorado boulder for fall semester 2019 to spring semester 2020 a case study of wifi and a campus response to the covid-19 perturbation," 2021.

[6] J. Schogol. (2022) Russian troops are proving that cell phones in war zones are a very bad idea. Online; accessed 05-May-2023. [Online]. Available: https://taskandpurpose.com/news/russia-ukraine-cell-phones-track-combat/

[7] S. Biddle. (2020) Police surveilled george floyd protests with help from twitter-affiliated startup dataminr. Online; accessed 05-May-2023. [Online]. Available: https://theintercept.com/2020/07/09/twitter-dataminr-police-spy-surveillance-black-lives-matter-protests/

[8] A. Coakley. (2021) Borderline: Tinder profiles of polish troops appear in belarus. Online; accessed 05-May-2023. [Online]. Available: https://www.independent.co.uk/news/world/europe/belarus-poland-border-tinder-troops-b1957953.html

[9] J. Keller. (2018) Your favorite fitness apps are leaving us military bases abroad exposed. Online; accessed 05-May-2023. [Online]. Available: https://taskandpurpose.com/tech-tactics/strava-military-bases-opsec-map/

[10] A. Douglas Sen and D. Bennett. (2022) Tracked: How colleges use ai to monitor student protests. Online; accessed 05-May-2023. [Online]. Available: https://pulitzercenter.org/stories/tracked-how-colleges-use-ai-monitor-student-protests

[11] K. V. Vishwanath and A. Vahdat, "Swing: Realistic and responsive network traffic generation," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 712–725, 2009.

[12] M. Huffman, A. Davis, J. Park, and J. Curry, "Identifying population movements with non-negative matrix factorization from wi-fi user counts in smart and connected cities," 2021.

[13] R. Craddock, D. Watson, and W. Saunders, "Generic pattern of life and behaviour analysis," in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2016, pp. 152–158.

[14] Y. Peng, T. Feng, C. Yang, C. Leng, L. Jiao, X. Zhu, L. Cao, and R. Li, "Hmm-lstm for proactive traffic prediction in 6g wireless networks," in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, 2021, pp. 544–548.

[15] S. Li, J. Song, L. Xu, Y. Hu, W. Luo, and X. Zhou, "Network traffic prediction based on the feature of newly-generated network flows," in *2022 IFIP Networking Conference (IFIP Networking)*, 2022, pp. 1–8.

[16] T. Han, Y. Zhang, H. Li, X. Zhang, and J. Tao, "Large-scale network traffic analysis and retrieval system using cfs algorithm," in *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2019, pp. 417–421.

[17] N. Gillis, *Nonnegative Matrix Factorization*. Society for Industrial and Applied Mathematics, 2021.

[18] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," in *Fourth International Conference on Statistical Methods for the Environmental Sciences "Environmetrics"*, 1994.

[19] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, no. 401, pp. 788–791, 1999. [Online]. Available: https://www.nature.com/articles/44565

[20] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167947306004191

[21] G. Casalino, "Non-negative factorization methods for extracting semantically relevant features in intelligent data analysis," Ph.D. dissertation, University of Bari Aldo Moro, 05 2015.

[22] N. Gillis, "The why and how of nonnegative matrix factorization," 2014.

[23] A. Hern. (2020) Berlin artist uses 99 phones to trick google into traffic jam alert. Online; accessed 05-May-2023. [Online]. Available: https://www.theguardian.com/technology/2020/feb/03/berlin-artist-uses-99-phones-trick-google-maps-traffic-jam-alert

[24] (2023) Raspberry pi wifi adapter. Online; accessed 05-May-2023. [Online]. Available: https://www.canakit.com/raspberry-pi-wifi.html