Federated Infrastructure: Usage, Patterns, and Insights from "The People's Network"

Dhananjay Jagtap*
dtjagtap@eng.ucsd.edu
UC San Diego

Alex Yen*
alyen@ucsd.edu
UC San Diego

Huanlei Wu
huw012@ucsd.edu
UC San Diego

Aaron Schulman
schulman@cs.ucsd.edu
UC San Diego

Pat Pannuto
ppannuto@ucsd.edu
UC San Diego

ABSTRACT

In this paper, we provide the first broad measurement study of the operation, adoption, performance, and efficacy of Helium. The Helium network aims to provide low-power, wide-area network wireless coverage for Internet of Things-class devices. In contrast to traditional infrastructure, "hotspots" (base stations) are owned and operated by individuals who are paid by the network for providing coverage and are paid directly by users for ferrying data.

As of May, 2021, Helium has over 40,000 active hotspots with 1,000 new hotspots coming online every day. This deployment is decentralized – 84% of users own at most three hotspots. Some support infrastructure remains highly centralized, however, with over 99% of data traffic routed through one cloud endpoint and multiple cities in which all hotspots rely on one ISP for backhaul. Helium is largely speculative today with more hotspot activity than user activity. Crowdsourced, incentive-guided infrastructure deployment largely works but shows evidence of gamification and apathy. As Helium lacks clear, radio-oriented coverage maps, we develop and test coverage models based on network incentives. Finally, empirical testing with IoT devices finds basic success, but uncovers numerous reliability issues.

CCS CONCEPTS

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

KEYWORDS

Decentralized wireless, LoRa, Internet of Things, Helium, LPWAN

ACM Reference Format:
Dhananjay Jagtap, Alex Yen, Huanlei Wu, Aaron Schulman, and Pat Pannuto. 2021. Federated Infrastructure: Usage, Patterns, and Insights from "The People's Network". In *ACM Internet Measurement Conference (IMC '21), November 2–4, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3487552.3487846

1 INTRODUCTION

Core to the success of the Smart City will be its supporting infrastructure. While the edge devices are ready, there is not yet a widely-deployed supporting communication infrastructure suited to their needs. Embedded systems are defined by the walls, sidewalks, and windows into which they are literally embedded. For this network, scale will come not from millions of devices sending millions of bytes but billions of devices sending thousands of bytes.

As we pack more wireless devices into the same 3D space, we need more supporting infrastructure. The need for small cells for IoT-class devices was predicted and quantified by Mark Weiser [23] and recently refined by Ghena et al. for the modern IoT [6]. Pragmatically, this means the rollout of infrastructure capable of supporting IoT-class devices across wide geographic areas with density proportionate to the number of deployed nodes.

On the surface, this problem does not sound different from traditional cellular systems. However, with the emergence of the Helium network, we see the emergence of a new approach to deploying this infrastructure. Rather than relying on one or a few large entities to deploy many million IoT base stations, Helium invites many million individuals to deploy one or two base stations.

Helium released their first infrastructure product, a LoRa gateway called the Helium Hotspot, in the fall of 2019. Today, the Helium company does not manufacture any hotspot hardware. Rather, five (and growing) third-party companies have Helium-compliant product lines, and over 1,000 Helium hotspots now come online every day. These hotspots are purchased by individual investors, which distributes the requisite capital, and risk, widely.

To manage and administer this decentralized network, Helium built a custom blockchain. The use of a blockchain enables a microtransaction model not otherwise easily accessible. Each packet for an IoT devices costs just $0.00001 USD, which can be paid directly to the infrastructure owner. Interestingly, the development of a new cryptocurrency, and its (initial) ability to mint coins, further enables Helium to "pay" infrastructure operators without needing to immediately raise fiat. If Helium is successful, payment for use of the network will raise the value of the new currency enough[1] for hotspot owners to recoup investment and turn a profit.

Now is a very interesting time to study the Helium network. The network is big enough to see trends but also still early enough to make changes. The infrastructure is beginning to see exponential

[*] Authors contributed equally to this paper.

[1] In practice, speculation has done this already. Hotspots pay for themselves in a few weeks, but we do not view the current valuation of the HNT token as sustainable if the paying user base does not grow as well.

IMC '21, November 2–4, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-9129-0/21/11.
https://doi.org/10.1145/3487552.3487846

growth but some of the associated challenges of scale as well. There remains very little data traffic, although we do see steady growth over the course of our analysis. In short, this is a new model for deploying and managing wireless infrastructure, and now that it is beginning to see real adoption, we aim to study how well it works.

As the first study of Helium, the research questions we ask are exploratory: What does a decentralized network topology look like, and how does it change over time? Who is using the Helium network, and how much do they use it? How does this decentralized network sustain itself? How well does the network perform and how reliable is it?

To start, we look at what we can learn from analyzing transactions on the blockchain. We are most interested in understanding this new model of distributed wireless infrastructure. Thus we focus in Section 4 on hotspots, where we investigate what we can learn of the physical deployment, the people and organizations who choose to invest and deploy, and how the infrastructure evolves over time. In Section 5, although there is only a small user base, we see what we can learn how users use Helium today and what we may be able to learn about users when adoption grows.

One of Helium's major selling points is that it is a "decentralized wireless network." Helium relies on existing wireline providers for hotspot backhaul, who it may ultimately compete with on hotspot deployments. Hotspots are deployed by non-experts, which may place them behind NATs, firewalls, or other uncontrolled middleboxes. In Section 6, we investigate how the connectivity of uncontrolled, crowdsourced infrastructure may affect the robustness and performance of Helium.

While hotspot deployment is uncontrolled, it is not without direction. Helium employs a governance-by-incentive model—i.e., those who deploy honestly and in the best interest of the network should reap the most rewards. In Section 7, we look into how well some of the incentive-based rules are, or are not, working. We investigate case studies of incentives and provide insights for possible future refinement of both incentive enforcement and design.

Finally, Section 8 asks the empirical question, how well does Helium work as a wireless network? In particular, we investigate how well a Helium user might expect the network to work in practice—i.e., where is there supposed to be network coverage, and how well does the network perform in the covered areas? Helium's incentives are designed to promote wide-area coverage, so we use these incentives as a baseline to develop coverage models. We then measure Helium's raw performance as well as its expected performance based on the incentive-derived models. While Helium today is perhaps adequate for best-effort service, the current infrastructure is not yet reliable. It delivers only around 70% packets in a controlled, best-case scenario and provides unpredictable geo-spatial coverage, even in areas with dense hotspot deployments.

2 HELIUM – QUICK OVERVIEW

Helium is a new wireless provider. The initial focus of Helium is on building crowdsourced hotspot infrastructure to provide wide-area LoRa coverage. LoRa is one of several new radio technologies competing to provide coverage for low-power, low-bandwidth edge devices (IoT-class devices). In contrast to traditional wireless infrastructure, Helium does not own the deployed LoRa base stations. Rather, individuals purchase, deploy, and maintain Helium hotspots in exchange for payment for coverage and for the data that their hotspot ferries. The Helium company built and sold the initial hotspots, maintains firmware for new third-party, mass-produced hotspots, develops supporting cloud infrastructure, and develops the blockchain that underpins the network.

We focus on Helium's LoRa network. However, Helium has announced an expansion to 5G [7]. In Section 9.2 we consider how the results of this study reflect on the future of Helium more broadly.

2.1 Helium from a Basic User's Perspective

We first explain how Helium works for end users. To deploy a LoRa device that uses the Helium network, such as a sensor, a user first registers a new application with the Helium Console (a Helium-provided cloud service that acts as a LoRaWAN router plus Helium wallet). Users could deploy their own Console equivalent, but as we will show later in our analysis, the vast majority do not and simply rely on the Helium Console. Next, users deposit money in their Console account to pay for future data. Then, they must register a new device with the Console, which gives configuration parameters for the device's networking stack (blindly copied `#define`s prepended to a Helium library). After this, the device is ready to deploy: it can now send and receive packets from the Helium network. In the field, packets sent by devices are received by hotspots, who forward their data to the Console, which pays hotspots and provides packet payloads to application users via HTTP (or numerous other means).

2.2 Components of the Helium Network

Next, we look "under the hood" to see how Helium actually works. Summarized in Figure 1, the Helium network is built on top of the the LoRaWAN network architecture. LoRaWAN is a cloud-based protocol that routes LoRa packets between wireless devices and cloud services. The primary difference from LoRaWAN is that Helium makes it possible to have crowdsourced hotspots to bridge cloud services with the LoRa wireless network; LoRaWAN requires application owners to deploy and operate their own hotspots. Helium adds routing by overloading identifiers in LoRaMAC that normally identify the device and its owning application.

The following describes each component of the Helium network:

Routers are cloud servers which manage the LoRaWAN protocol across devices and applications. They are responsible for authenticating devices and receiving messages from devices. Hotspots find Helium-compliant routers by looking up device owners using packet metadata and a filter list in the Helium blockchain (in contrast to standard LoRaWAN, where gateways have one, statically configured router). Helium routers must also negotiate with hotspots to pay for data. This is done via the `state_channel` mechanism described later. Anyone can host their own Helium compatible LoRaWAN router by purchasing device and application identifiers via a transaction on the blockchain.

Edge Devices are the LoRa-enabled end wireless devices. They can be embedded in products such as sensors, collars, or tags and can be used for a variety of applications including environmental, pet, or asset monitoring. Devices are pre-provisioned with a Device End User Identifier (EUI), an Application EUI, and an App key. These are used during Over The Air Activation (OTAA), part of LoRaWAN,

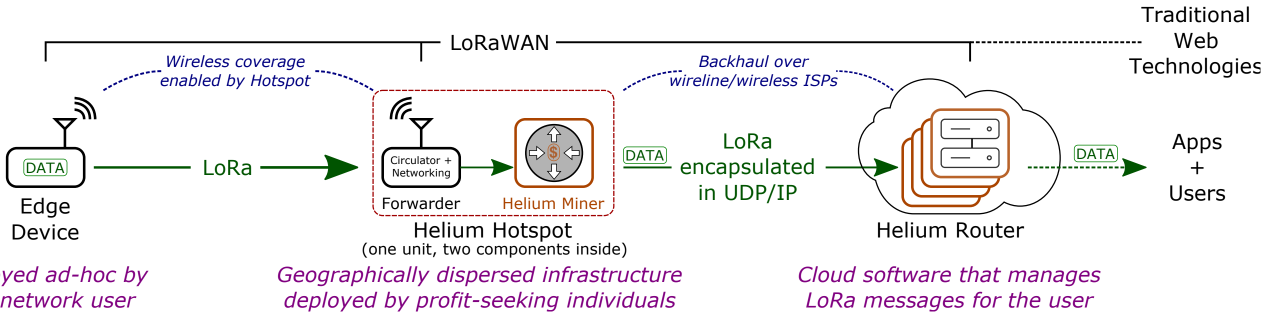

LoRaWAN
Traditional Web Technologies
Wireless coverage enabled by Hotspot
Backhaul over wireline/wireless ISPs
DATA
Edge Device
LoRa
Circulator + Networking
Forwarder
Helium Miner
Helium Hotspot
(one unit, two components inside)
DATA
LoRa encapsulated in UDP/IP
Helium Router
DATA
Apps + Users
Deployed ad-hoc by any network user
Geographically dispersed infrastructure deployed by profit-seeking individuals
Cloud software that manages LoRa messages for the user

Figure 1: Helium Overview. Edge devices broadcast packets over the LoRa PHY. One or more nearby infrastructure nodes—hotspots—recover this packet. Helium Routers pay hotspots to release packets from edge devices they wish to recover data from. Routers must be reliably available as the LoRa protocol requires that any downlink responses from the router arrive at the edge device within one second.

to authenticate to a LoRaWAN Router (i.e., a cloud service). Helium differs from LoRaWAN in that EUIs are first used to look up Helium Organizationally Unique Identifiers (OUIs). These determine which router to authenticate to.

Packet Forwarders are LoRa wireless modules in hotspots. These use the industry-standard Semtech packet forwarder [16] to relay edge device radio packets to and from the co-resident miner in the hotspot (in regular LoRaWAN, these are forwarded directly to the owning router; the miner is new from Helium). They are generally made up of a LoRa concentrator, which is a radio transceiver capable of operating on multiple parallel subbands and a supporting processor which runs the packet forwarder.

Miners transmit encapsulated LoRa packet payloads to and from Helium routers and maintain the Helium blockchain. A miner is usually a small embedded Linux device.[2] Miners receive Helium tokens (HNT) as rewards for device data transit, network coverage validation, and blockchain consensus activities. They send traffic to LoRaWAN routers on the internet at large using any UDP/IP capable backhaul-like wireline (e.g., Cable and DSL) or wireless (e.g., LTE) Internet access networks.

Hotspots are physical boxes with a packet forwarder and a miner. Hotspots make up the majority of deployed Helium infrastructure to date.[3] In principle, forwarders and miners could be separated, as the connection is simply an IP link. In practice, they are co-located to overcome a short-term engineering limitation:

> "The protocol between the gateway and the server is purposefully very basic and for demonstration purpose only, or for use on private and reliable networks. There is no authentication of the gateway or the server, and the acknowledges are only used for network quality assessment, not to correct UDP datagrams losses (no retries)."[4]

Helium Console – The Console is a cloud service provided by Helium which acts as a Helium router. Instead of buying an OUI and running their own router, users deploying end devices can opt to use the Console. At this time, the Console charges users wholesale for data communicated with sensors and has no overhead cost.

Validators are a new class of Helium node, currently in testing, that will take over blockchain maintenance events from miners. The proposal (HIP25) to introduce validators was ratified in January 2021. Validators solve well-known scalability and robustness challenges for blockchains and use well-known staking techniques. As validators were not yet live during the period under study (through May 2021), we do not consider them extensively. We note validators here as they appear as special-case miners on the blockchain and will influence some of our analysis of deployed hotspots later. Our focus in this study is primarily Helium as a wireless network, as opposed to the maintenance and operation of the Helium blockchain.

Handlers is a new term we define to describe individuals who own and maintain Helium hotspots. In contrast to users, who send and receive data using the network, handlers and their hotspots do not generate any data; they simply forward it and receive payment. It is possible for individual entities (e.g. the authors of this paper) to act as both users of the network and handlers who help deploy the network (i.e., we own two hotspots and around a dozen devices).

2.3 Proof of Coverage

To be useful as a wireless carrier, Helium needs to provide geographically wide-area coverage. As a decentralized wireless network not owned and guaranteed by any one entity, there needs to be a mechanism for the network to prove that a hotspot is providing coverage. Helium uses an algorithm called Proof of Coverage (PoC) to prove the location of a hotspot. At random, one hotspot will act as a *challenger* and will randomly select another hotspot, referred to as either the *challengee* or *transmitter*, to send a wireless packet with an encrypted secret. Any hotspot that is in wireless range can *witness* this packet by reporting its contents to the challenger. Hotspot challenges are not geographically coordinated and can be acted on any other hotspot in the world. They do not target and prove any specific region has coverage, rather they stochastically validate every node in the network's coverage over time.

Each of the challenger, challengee, and witnesses receive HNT rewards after the PoC event. Challenger rewards are fixed (the reward is in creating and administering the challenge), but challengees' and witnesses' rewards scale with the quality of coverage (in simple terms here, more witnesses are better). As we will show, there have been and still are numerous attempts to cheat this metric, and it continues to evolve.

[2] e.g. in the Rak Wireless v1.5 hotspot, the miner is an off-the-shelf Raspberry Pi 4 Model B – see https://www.youtube.com/watch?v=BUrAb2GoHhc for more detail.
[3] Very early, Helium permitted a small batch of "DIY Hardware." New hotspots must be from certified vendors to earn rewards on the network.
[4] From https://github.com/Lora-net/packet_forwarder/blob/master/PROTOCOL.TXT

When deployments are sparse, i.e., if a hotspot cannot "see" any other hotspots, then hotspots can only earn PoC rewards for challenge construction. This aims to incentivize hotspot density such that there are not holes or gaps in coverage; however, there also is decaying of rewards if hotspots are too dense. Qualitatively, this seems to work, as improving coverage rewards is often discussed on Helium forums and message boards. We discuss these incentives and coverage quantitatively in Sections 7 and 8.

2.4 HNT and DC

For general purpose users of the network, Helium aims to provide a stable pricing model. Miners are rewarded with newly minted Helium tokens (HNT). HNT is a traditional cryptocurrency, whose value has ranged from $8.32–19.70 USD in the month of May, 2021. Users planning deployments require more stable pricing—transit costs cannot unexpectedly double overnight. For this, Helium introduces Data Credits (DC), whose value is fixed at $0.00001 USD per 1 DC. All payments are done using fixed-cost DC.

Together, DC and HNT create a "burn-and-mint equilibrium." In the long run, the intent is to stabilize the price of HNT and tether it to network usage. In the short term, DC provide a stable, deterministic payment model for network users, while network operators receive incentives in the more speculative form of HNT. Notably, this model enables the Helium network to "pay" handlers without incurring traditional capital outlay. The details of this new crypto-economic model are only provided for background, analysis is beyond the scope of this paper; we only study the network infrastructure-related aspects of Helium.

3 DATA SOURCES & METHODOLOGY

The Helium network is a dynamically growing network composed of handlers that deploy hotspots, (ideally) many users that deploy edge devices, and a modest number of advanced users that deploy their own routers (cloud endpoints). While the distributed nature makes direct inspection of network activity challenging, the pay-per-access design means that most transactions are ultimately recorded in a publicly accessible ledger: the Helium blockchain.

Most of our analysis stems from an examination of the history of all transactions on the blockchain. While anyone can download and parse the blockchain—easiest done by running the miner Docker container locally—most of our analysis takes advantage of a replica of the blockchain continuously extracted, transformed, and loaded into a database by the Decentralized Wireless Alliance (DeWi). This database also monitors the Helium p2p network, which we also use for our analyses. Details on the DeWi database are available on their Discord.[5] We spot-verified several of our own transactions and our hotspots' p2p records to ensure they appear correctly in this database. We also perform some controlled measurements in Section 8 to measure the actual performance of the network.

The Helium blockchain is a fast-moving target. New blocks are minted every 60 s, and as of this writing, roughly 1,000 new hotspots are being added daily. Unless otherwise noted, measurements in this paper reflect the state of the network as of late May, 2021.

There are 20 native Helium blockchain transactions [8]. The transactions most relevant to our analysis are as follows:

add_gateway adds a new hotspot to the network. It includes the hotspot ID, owner ID, location, and time when it was added.

assert_location allows an established hotspot to change its location. To discourage frequent moving of hotspots, this transaction carries a 1,000,000 DC fee ($10 USD).

PoC_request/PoC_receipt is created every PoC challenge and indicates the work done to validate network coverage.

state_channel_open creates a sidechain[6] for a router to use to receive packets. This stakes DC in advance, which allows packet delivery without waiting for individual payment transactions.

state_channel_close settles payments. Spent DC are burned, miners that ferried data receive HNT proportional to total network traffic, and unspent DC are returned to the router.

To start, we can look at this blockchain data as a whole for initial insights and investigatory guidance. We see the first real entry to the blockchain was recorded on July 29, 2019. The vast majority of transactions recorded on the blockchain consist of PoC requests and receipts. Out of 59,092,640 total transactions, 58,619,153 are carried out only to provide proof for the network accuracy and validity. Since the inception of the Helium network, approximately 99.2% of all blockchain transactions are PoC transactions.

4 WHAT DOES "DECENTRALIZED WIRELESS" LOOK LIKE IN PRACTICE?

We begin our analysis by looking at hotspots, which make up the majority of the deployed infrastructure of Helium. We are interested in understanding where hotspots are deployed, who is deploying them, and what happens to deployed hotspots over time.

4.1 Where are Hotspots?

Whenever a hotspot's location is first published to the Helium network or changed, Helium records a transaction that contains the hotspot location encoded via the H3 geospatial indexing system [19]. The H3 encoding system encodes locations to hexagonal areas on a map; if a hotspot lies within borders of the hexagon, the hotspot's latitudinal and longitudinal coordinates are mapped to that specific H3 identity. H3 hexes allow for different "resolutions," which encode location with varying precision. Hotspot locations are stored at resolution 12, which are hexagons with an average edge length of 9.4 m and average area of 3.1 m^2; for our location analysis, we assume all hotspots are located at the center of their hex.[7] We decode the H3 location to retrieve hotspot latitude and longitude, which we use in all subsequent analyses.

Observing how hotspot locations change gives insight into whether the policy and rewards set out by the Helium network improve real-world coverage. In Figure 2, we look at the absolute number of moves per hotspot and find that the vast majority of hotspots either do not move at all or move no more than two times. The Helium

[5] discord.com/channels/404106811252408320/769659586205712424/786375046930104341

[6] See Singh et al. for background and detailed discussion of sidechains [17]

[7] One may note from these averages that H3 hexes are not regular hexagons. This is a result of H3's use of gnomonic projection. For resolution 12 hexes, the min and max area are 1.9 m^2 and 3.7 m^2 respectively. In practice, hotspot locations are set using manually entered GPS fixes from a "set pin" interface on a companion phone app. Resolution 12 hexes are likely more precise than these fixes for the majority of use cases. As our analyses are primarily concerned with distances of several hundred meters or more, a few meters imprecision in hotspot location is not significant.

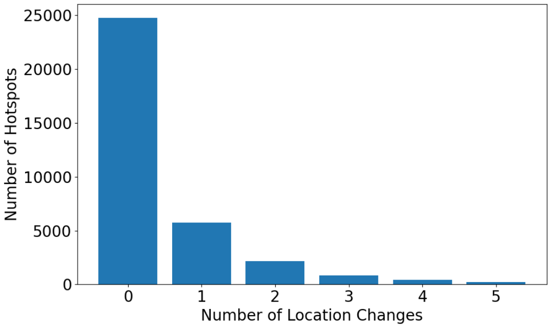
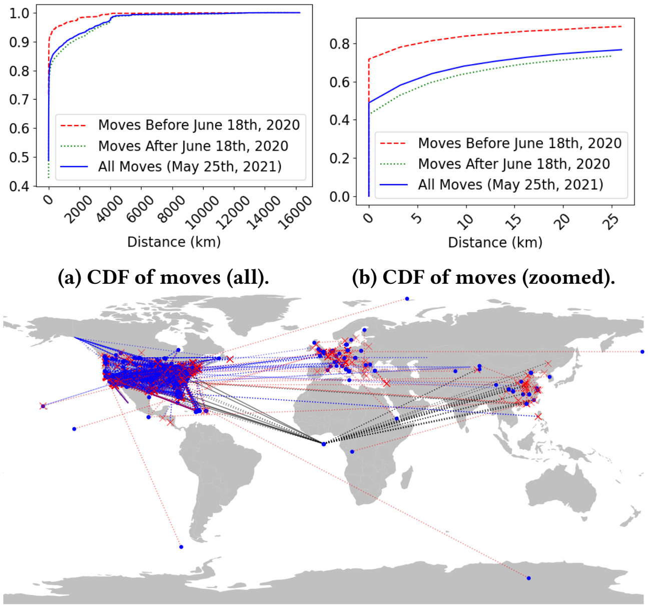


Figure 2: Location changes per hotspot. 71.9% of hotspots never move once deployed. 94.8% of hotspots do not move more than two times, and only 1% of hotspots move more than five times.



(a) CDF of moves (all).

(b) CDF of moves (zoomed).

(c) Hotspots move from blue circles to red "X" marks.

Figure 3: CDF of all moves and map of hotspot location changes greater than 500 km. Figure 3c shows every location change from initial to final destination distances that are greater than 500 km away. Blue dotted lines indicate moves that departed the US; red dotted lines indicate location initialization outside of the US. Black dotted lines indicate moves from the initial location at (0,0) to anywhere else in the world, and black solid lines indicate a hotspot whose location changed to (0,0).

network permits hotspots to move up to two times for “free” (the Helium company pays the `assert_location` fee). One explanation for the large number of “first-move” events may be new hotspot owners first testing the hotspot in an easily accessible location and then moving the hotspot to a permanent deployment location.

To test this hypothesis, we study all hotspot moves. In Figure 3, we see two broad categories of moves. The first are a large number of short-distance moves, likely indicative of a local test-then-deploy scenario. The second are a large number of long-distance moves. We visualize these longer moves in Figure 3c.

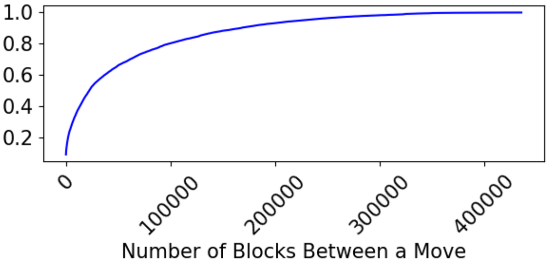


Figure 4: CDF of block intervals between hotspot relocations. Blocks are minted roughly once a minute.

We see two classes of long-distance movement. The first are hotspots that assert their location with a default (0,0) latitude and longitude—the large cluster in the ocean just below West Africa—and then later move to their true location. Hotspots occasionally move to (0,0) but do not stay there; currently, aside from hotspots who initialized at (0,0), there are no online hotspots that have moved to and remain at (0,0). We found that 331 (89%) of (0,0) location assertions were first time assertions, which suggests that most of the initial assertions at (0,0) were accidental (e.g. no GPS fix); the total number of assertions at (0,0) is 372. The remaining 41 assertions at (0,0) were attempted relocations. This could have also been accidental, a test out of curiosity, people attempting to game the system by clustering hotspots at a fake location, or possibly Helium developers testing validator nodes (which appear in the blockchain as hotspots that never transmit packets).

The second trend is a non-trivial “flow” of hotspots from the US to international destinations, particularly Europe. We attribute these moves to resale of hotspots, and the heavy US-export to Helium’s initial sales restriction to only the US market.

We also studied the typical timing interval between a hotspot’s relocation. We plotted a CDF regarding the number of blocks between a hotspot’s relocation in Figure 4 to observe the time interval between location assertions; one block approximately corresponds to one minute. 17.9% of hotspot relocations occur within a day, 35.8% of relocations occur within a week, and 63.2% of relocations occur within a month. The remaining 36.8% of hotspot relocations have a timing interval of more than a month.

Finally, in the analysis of hotspot moves, we identify a small number of outliers. For example, one hotspot moved twenty times. It is possible this is an individual with multiple residences bringing their hotspot along with them, perhaps in service of providing local coverage for their own edge devices. However, we also note that the owner of this hotspot has well over 800,000 HNT, possibly posing as a business entity or developer account.

4.2 How Fast is Helium Growing?

Figure 5 shows the cumulative and daily number of hotspots that were added to the Helium network. For coverage discussions, we refine this count and draw a distinction between “connected” and “online” hotspots. A connected hotspot is one that has *ever* connected to the Helium network (this is the number displayed on

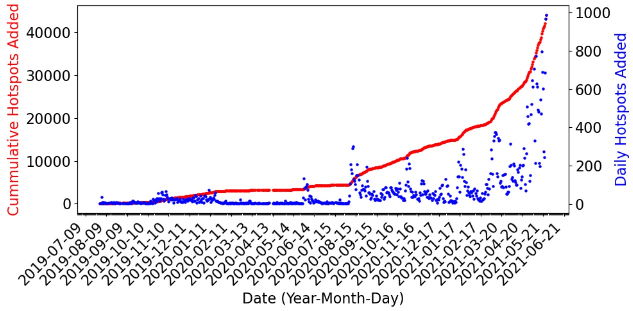


Figure 5: Helium network cumulative and daily growth. Qualitatively, growth seems mostly limited by hotspot availability. New production runs ('batches') are quickly placed into service.

various Helium status pages), while an online hotspot is one that is still connected and active (defined as fully synced and participating in PoC challenges). On March 7th, 2021, the total number of connected hotspots was about 20,000, however, only approximately 16,000 hotspots were online. On May 26th, 2021, the total number of connected hotspots was about 44,000, with 34,000 hotspots online.

We are also interested in understanding where new hotspots are deployed. On March 7th, 2021, we found approximately 15,000 online hotspots located within the United States and 1,000 online hotspots located outside of the United States. On May 26th, 2021, we found approximately 20,000 online hotspots in the US and about 14,000 outside of the US. While international coverage is growing rapidly, the network launched in the US in summer 2019 and was not available outside the US until summer 2020.

4.3 Who is Deploying Hotspots?

Every hotspot has a designated owner, or more precisely, a wallet that receives the rewards earned by the hotspot. We found an exponential decay relationship between the number of hotspots a single owner owns and the number of owners; approximately 5,700 owners (62.1%) own only one hotspot, about 1,300 owners (14.6%) own two hotspots, about 600 owners (7%) own three hotspots. There are about 9,000 unique owners total, of which 83.7% own 3 or fewer hotspots and 10.3% own 5 or more. As of May, 2021, the maximum number of hotspots owned by a single owner is 1,903. Contrast this with March, 2021, when the maximum was 160. The majority of hotspots are located in the US, and the second most common region is western Europe. While there are a modest number of large-scale owners, it is fair to claim that ownership (by unique wallets at least) of the Helium hotspot infrastructure today is decentralized.

We next investigate several of the larger hotspot pools. A common inference from HNT balances over time is that owners which are using Helium in service of a real-world, end application engage in a large number of data transactions and have thousands to tens of thousands of HNT in their account. In contrast, owners which do not take part in data transactions generally have low HNT in their accounts as they frequently encash their HNT. These owners appear to be using the network as means for gaining profit rather than supporting their own edge devices. Next, we look at a few examples of each of these classes of owner.

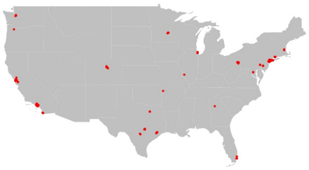


Figure 6: Hotspot distribution of one larger owner.

4.3.1 Bulk Owners – Commercial. We start from large owners on the Helium blockchain and describe how we work backwards to identify the owning company. We show how this procedure may be used to identify entities using Helium infrastructure solely with information available on the public blockchain or other easily-accessible public sources.[8] A similar approach can identify unannounced users of the network. We wish to emphasize that Helium does not explicitly provide receiver or transmitter anonymity. However, the design of the network also does not explicitly identify the users that are performing transactions. We show how easily publicly announced application traffic can be identified in the blockchain. This may be problematic for applications intended to be private.

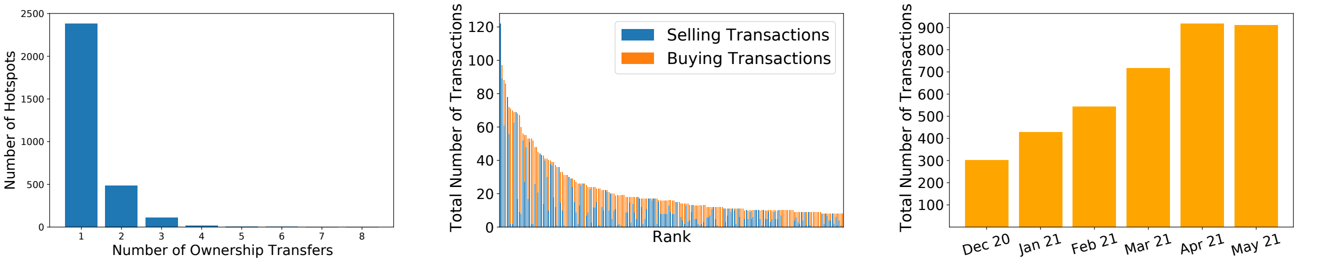
Careband is a small startup in Chicago which specializes in developing wander-management wearables to detect patient movement, especially for those who have dementia. Their main office is located in 222 West Merchandise Mart Plaza. We looked at hotspots within that area and found one owner ID which owns 25 hotspots mostly in and around in Chicago city and some individual hotspots in rest of the United States. We believe these hotspots are owned by Careband or provide coverage to Careband customers.

nowi is another startup that uses the Helium network to support water monitoring systems for multi-family property owners. They have one testimonial from Edworks LLC Property management. We found out they are registered in Stonington, Connecticut. There are 19 owners that collectively own 61 hotspots in Stonington as of September 17, 2021. Out of them, 9 owners in Stonington own multiple hotspots and regularly sends at least tens of data packets every couple of of hours. This indicates that there are devices communicating their data over the Helium network.

4.3.2 Bulk Owners – Mining Pools. We believe that owners which own multiple hotspots and carry out frequent data transactions are service providers who manage these hotspots to provide coverage to their products. Other owners, such as the one in Figure 6, own multiple hotspots which were geographically distributed but do not engage in data transactions. We believe these owners are running hotspots to earn coverage rewards.

Ownership clusters in cities are popular. In one case, two owners own 144 hotspots and 136 hotspots in the Denver, Colorado area. The hotspots appear to be distributed evenly around the city. Deploying hotspots too geospatially close to each other reduces reward benefits; hotspots yield better mining efficiency when placed reasonably far away from each other. This incentive of improved rewards from improved network coverage appears to be important to these types of owners.

[8] The examples we present here have publicly announced their use of Helium via the Helium blog [9, 10], which we believe mitigates any potential harm from using them.



(a) Ownership transfers per hotspot. **(b) Number of hotspots transferred per owner.** **(c) Hotspot transfer transactions over time.**

Figure 7: Resale market analysis. About 95.4% of the total hotspots transferred do not change owners more than 2 times as seen in (a). This shows that once hotspots are transferred they do not have a tendency to change ownership. In (b), we show the 200 owners which have participated in the most hotspot transfers either purchasing or selling them. These owners account for about 10% of the total owners who have participated in carrying out these transactions. Over time there is growth in the number of owners participating in these hotspot transfer transactions as seen in (c). There were a total of 3,819 such transactions over a span of 6 months.

4.3.3 Resale Market. Enthusiasm for Helium, coupled with the global electronics shortage [20] as well as the general manufacturing latency of new-to-market products, has resulted in a shortage of hotspots. The original Helium hotspot sold for $500 USD and newer, Helium-sanctioned third-party hotspots cost around $300-400 USD. An informal survey of the resale market finds a median price of $989 USD for a Helium hotspot among the top twenty eBay listings (min: $405, max: $6,500).

Some of the resale market is new, unopened hotspots. We cannot track sales of those hotspots. However, Helium also supports a `transfer_hotspot` transaction, which allows one user to sell an established hotspot to another. Figure 7 analyzes hotspots and owners involved in resale transactions. About 8.6% of the total hotspots deployed are transferred to another owner. Over 95.8% of hotspot transfer transactions transfer 0 DC between buyer and seller, which suggests that the majority of resale payments take place using an off-chain marketplace, such as eBay; although it is not clear what ensures sellers actually relinquish ownership of the hotspot to buyers in such markets.

5 HOW MUCH IS BUILT ON HELIUM TODAY?

Next, we seek to understand what types of users and applications are running on the Helium network today. Our analysis reveals that currently Helium remains highly speculative, with more handlers deploying hotspots than users using Helium to ferry data. This is perhaps not surprising. The design of Helium’s coverage-based reward model is to break into the chicken-and-egg problem of ‘no users to pay for infrastructure’ and ‘no infrastructure to support possible users.’ Still, we find a small but steady growth in what appears to be real-world application traffic, that suggests that ‘if [Helium] can build it, they will come.’

5.1 How does payment-for-data actually work?

Data transfer transactions are not recorded immediately and directly to the main Helium blockchain. Instead, short-lived “state channels” aggregate batches of packets. The purpose of state channels is to permit fast and scalable payment for individual packet transfers, a critical facet of the Helium microtransaction model.

To receive data, routers must first open a state channel. A state channel open transaction stakes DC to pay for packets that may eventually be transferred and sets a deadline some number of blocks[9] after which the channel will be closed. Hotspots that receive a wireless packet from an edge device use the metadata in the packet to look up the owning router and send an offer to the router to buy the packet; the offer includes packet metadata but not yet payload contents. A router purchases a packet by sending back a signed offer to buy, at which point the hotspot releases the packet.

Routers are responsible for closing state channels after they expire. The state channel close transaction (should) include every offer to buy a packet made by the router. If a router signed an offer but never received the packet, it omits that offer from the close transaction. When a hotspot that did send data is left out of a close transaction, there is a 10-block grace period for the hotspot to a submit signed demand that amends the closing. If a hotspot lies about sending data, routers have no recourse but to add the hotspot to a blocklist and not make future offers to purchase its packets. As the value of individual data transactions is quite low and the duration of state channels is short, this discovery of malicious hotspots does not pose significant economic risk to routers.

For our analysis of data transfer behavior, then, we are limited by the resolution of state channel transactions. State channel duration is decided by each router independently. As we explain in the next section, however, nearly all traffic to date is sent on state channels with a 240-block (roughly 2 hour) duration, which is the granularity with which our subsequent traffic analyses are able to operate.

One additional thing to note here is that it is possible for multiple hotspots to receive the same packet. While there is sufficient information in the packet metadata that a router can identify duplicate receptions, it can still choose to buy as many copies of a packet as it wishes. Observing payment flow then is a measure of data transferred between hotspots and routers, which may overestimate the actual flow of data from edge devices, depending on how often routers choose to purchase duplicate copies.

[9] While not stated in any documentation, the blockchain implementation (github.com/helium/blockchain-core/blob/9011de7537ecfd737074b85b7b16e7d8e1ceef00/src/transactions/v1/blockchain_txn_state_channel_open_v1.erl#L208) limits this to a minimum of 10 blocks (~10 min) and a maximum of one week (exact block count is derived from current block time).

5.2 Who is running Helium routers?

Setting up a router and payment processor requires non-trivial technical expertise. Routers must be continuously online and responsive. The LoRaMAC between edge device and gateway has two acknowledgment windows, at precisely 1 s and 2 s after a packet transmission. The LoRaWAN protocol dictates that routers are responsible for sending acknowledgments if requested (as it is the router which much choose which gateway should send the acknowledgment packet if multiple gateways hear the original uplink packet from the device). Thus the cloud service must (1) learn of a proffered packet, (2) return a signed commitment to pay, (3) receive payload data, (4) generate an acknowledgment, and (5) send a signed commitment to pay for acknowledgment to a hotspot in under 1 s (or, with less reliability 2 s) for each data packet.

As of May 2021, there are only ten OUIs registered. OUI 1 and OUI 2 are registered to the Helium company. Of all state channel open/close transactions, 81.18% belong to OUI 1 and OUI 2.

As a (currently) free service, the Helium company provides the Helium Console, which is both a Helium router as well as an interface for provisioning and managing devices. The Console includes numerous integrations that allow data collected from sensors to flow to other services such as cloud database providers and mapping systems. Console users are required to buy DC for their devices, but this is purchased and used at-cost.

From an analysis perspective, this monopolistic router limits direct insight into application users, as all data transfer payments are from the Helium Console OUI, rather than from individual users or applications. To fund user accounts with DC on the Console, users can either burn their own HNT with the Console wallet as the destination—a transaction which is visible per-user to our analysis—or they can make a credit card purchase, in which case the Console will acquire and burn HNT using its own account. In practice, DC are so cheap that funding events are rare. Indeed, long before beginning this study, our research group made a one-off $10 USD purchase of DC (which is the minimum purchase amount permitted by the Console) in November 2020 to support another experiment. As of this writing, we have used less than 15% of this purchase, despite regular use of the network by multiple research projects and extensive use during our experiments for this work.

5.3 How much actual data is sent over Helium?

With the caveats on resolution outlined, namely state channel batching, opaque duplicate purchases, and a centralized OUI operator, we next look to analyze the behavior of data on the Helium network.

5.3.1 Device Data Traffic. We observe the trends of the device data traffic since 2019 to identify how much of the network is used for data transfers. Figure 8 gives a macro view of data activity. Most of data transfers in the earlier blocks were carried out by the Helium Router. However, the data transfers carried out by third party routers have recently started to increase, evidence of increased usage of the Helium network by end applications.

5.3.2 Fake Data & Arbitrage. One exceptional situation occurred on August 12, 2020, which is the date that DC payments first went live on the network. We see a sharp rise in the data transfer between August 12, 2020 and Sep 6, 2020 and a sudden drop after that.

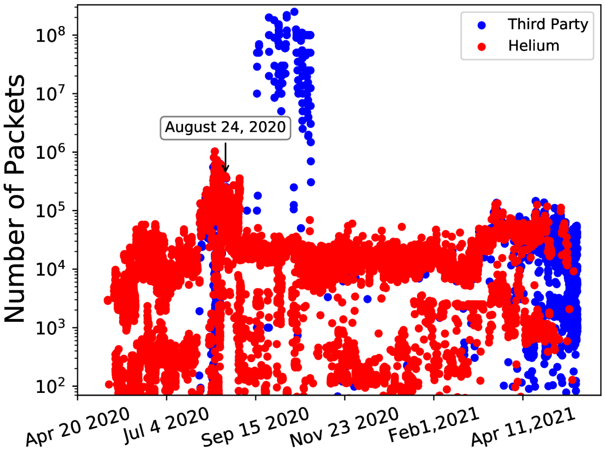


Figure 8: Packets transfer analysis. This shows the number of packets paid for with each state channel closing transaction, sorted by block. The primary trend is data paid for by OUI 1 and OUI 2, the Helium Console, which closes a state channel roughly every 120 blocks. At around 1 min/block, aggregate user traffic is approaching 14 packets/second across the whole network.

Prior to this date, data transfer was free, and mining rewards that would have been allocated to data transfer were instead allocated to PoC activity. When DC payments and miner data transfer rewards first went live, there was no cap on the reward one received for network data transfers. Every epoch, 32.5% of newly minted HNT was divided among hotspots that ferried data, in proportion to the amount of data they carried – essentially, more data transfers would fetch you more rewards. Recall, however, that the cost of data is fixed, creating an arbitrage opportunity among DC/USD and HNT. Users were thus gaming the network by spamming packets to devices they owned to increase their shared of mined HNT. The arbitrage was stopped on August 24, 2020 with the implementation of HIP 10 [12], though it took slightly longer for the spam packets to fall off the network. This event remains the largest sustained volume of data traffic carried by the Helium network to date.

6 META-INFRASTRUCTURE

In this section, we take a more holistic view of the Helium network and networking at large. We are interested to understand what infrastructure the Helium infrastructure relies on, and whether there are hidden points of centralization in this otherwise decentralized network. We find that despite the wide array of individuals deploying hotspots, the Helium network has potential choke points.

6.1 What ISPs do Hotspots Rely On?

One such example of a choke point is a region's reliance on ISPs. As all Helium hotspots are currently miners, they all participate in one, large p2p network.[10] We use the zannotate utility [4] together with

[10] We note that there is a limited window of time for this analysis. With the impending launch of validator nodes, hotspots will have the option to convert to so-called "light" nodes. Only the validators will maintain a fully connected p2p graph, and thus only they will have access to the network information of some hotspots in the future.

Table 1: Top 15 ISPs used for hotspot backhaul.

	ISP	Number of Hotspots			
1	Spectrum	2497	**9**	Sky UK	199
2	Comcast	1922	**10**	Telefonica	199
3	Verizon	1590	**11**	CenturyLink	188
4	Cablevision	450	**12**	TELUS	185
5	AT&T	338	**13**	RCN	154
6	Virgin Media	333	**14**	Frontier	146
7	Cox	314	**15**	Google Fiber	142
8	Level 3	202			

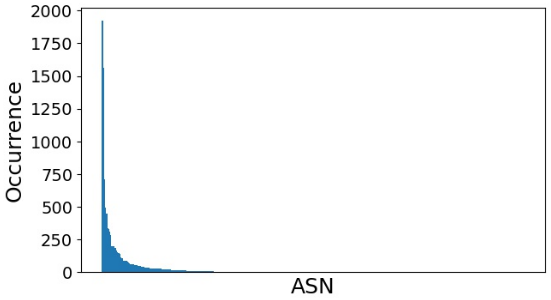
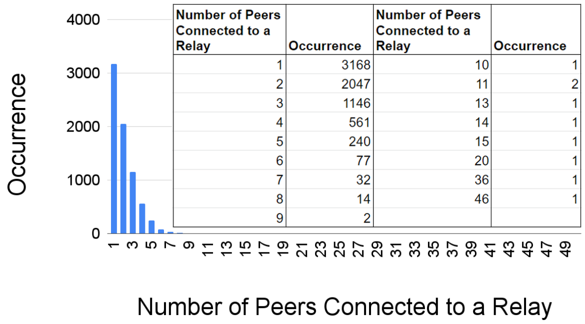


Figure 9: Distribution of ASNs for hotspots (with public IPs). Sorted by the number of hotspots per ASN, the overwhelming majority of hotspots hang off of just a few networks, although there is a very long tail of ASNs with just one or two hotspots.

Route Views data to identify the ASN of all non-relayed hotspots (i.e. hotspots with public IP addresses) connected to the p2p network.

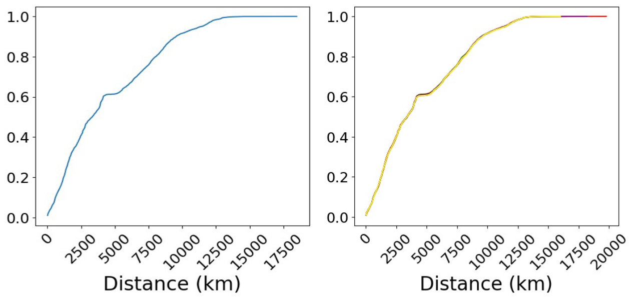
In total, Helium hotspots are deployed in 454 ASNs. Figure 9 shows the complete distribution of these ASNs and Table 1 shows the top 15 ISPs derived from the ASNs using CAIDA’s as2org dataset [2]. We find that the most widely-used ISP is Spectrum with 2,497 hotspots, and the second most used ISP is Comcast, but it hosts significantly fewer hotspots. Verizon comes in a close third. Most hotspots are on Verizon’s wireline network, but surprisingly 30 of the 1,590 hotspots are backhauled through Verizon wireless. We also discovered that there are hotspots that use a cloud provider, such as Digital Ocean (72 hotspots) and Amazon (44 hotspots) rather than a last-mile provider. We believe these hotspots are validators, as they appear as hotspots on the blockchain.

Another statistic we looked at was the percentage of unique ASNs within a city. The number of locally unique ASNs is important as relying on one ISP could cause a regional outage if that ISP goes out. In total, there are 3,958 cities with at least one hotspot. Out of those cities, there is a total of 1,588 cities that relied on only one ASN, with 414 of those cities having at least 2 hotspots. These included cities such as Palma, Spain (with a total of 76 hotspots), Mesa, Arizona (13 hotspots), and Rome, Italy (12 hotspots). An example of an outage that may have had a large impact on Helium was the 2020 Spectrum outage in Los Angeles [1]. For a few hours, Spectrum customers across the city lost Internet access. This could have taken down 291 out of the 333 hotspots (87%) in Los Angeles.



Number of Peers Connected to a Relay	Occurrence	Number of Peers Connected to a Relay	Occurrence
1	3168	10	1
2	2047	11	2
3	1146	13	1
4	561	14	1
5	240	15	1
6	77	20	1
7	32	36	1
8	14	46	1
9	2		

Figure 10: Relay nodes with *n* peer nodes. Most hotspots relay only a few nodes. The cause of high-relay nodes is unknown.



(a) Actual. **(b) 5 random assignments.**

Figure 11: Relay to peer node distance, actual & simulation. Peers choose relays randomly, without geospatial consideration.

6.2 Relay Analysis

One side-effect of adoption by individuals and smaller operators is that many hotspots are on network connections, such as residential home networks, that do not provide public IPs to all devices. Depending on NAT (or firewall) configuration, hotspots may not be able to accept inbound connections. `libp2p`, which Helium uses to form its network, addresses this with “Circuit Relays” [13].

When a hotspot cannot directly communicate, it opens a persistent connection with another hotspot on a less restrictive network to relay messages and data. Peerbook entries are formatted in two ways: `/p2p/relay_node_hash/p2p-circuit/p2p/peer_node_hash` for hotspots who rely on a relay node and `/ip4/ipv4_address/tcp/port` for hotspots that have public IPs and accessible ports. Using this information, we are able to study relay prevalence and behavior.

First, we are surprised by the prevalence of relays. Of the 27,281 hotpots with non-empty listening addresses, 55.48%—more than half the network!—are relayed. This heavy reliance on relay nodes increases the meta-infrastructure risks identified in the previous section, as relayed nodes are beholden to their relaying device.

Next, Figure 10 looks at how relays are distributed among hotspots. While the majority of relaying nodes support just one or two peer nodes, there are a few who relay as many as 46 other nodes. We are unable to determine why these nodes relay such a large number of peers. One hypothesis we investigated was whether these are the hard-coded seed peers[11] that ship with the hotspot firmware image, but the high count relay nodes do not match these IPs.

Our next hypothesis is that hotspots bias towards a geographically nearby peer. Such a design could be problematic for local robustness: if many geospatially clustered nodes rely on the same relay peer, then coverage reliability for that whole area would fall to the reliability of the relaying node. At the same time, ignoring location also can create problems for a globally distributed network, particularly one which requires multiple round trip communications in under 1 s to support LoRaMAC acknowledgments.

We use the asserted location data of each hotspot to compute the distance between each peer and its relay node and graph the distribution as a CDF, shown in Figure 11a. The majority of distances are below 5,000 km. The minimum distance is 0.46 km and the maximum is 18,491.10 km. While this suggests peer selection is random, given the non-uniform geospatial distribution of hotspots, distance alone is insufficient. We next take the list of all relays and relayed nodes and run multiple trials which randomize the assignment of peers to relays, as shown in Figure 11b. With this analysis, we are confident that the Helium network does in fact assign peers randomly to relay nodes.

7 GOVERNANCE BY INCENTIVE

Because the Helium network is decentralized, it cannot directly affect change on the deployed infrastructure. Instead, the network uses economic incentives to motivate changes in user behavior. While these are relatively stable, "Helium Improvement Proposals," or HIPs, can change the rules of the Helium blockchain.[12] In principle, HIPs create economic incentives for hotspot owners to change their behavior. This section looks at examples of the efficacy and inefficacy of government by incentive on the Helium network.

7.1 Case Study 1: Silent Movers

As mentioned in Section 2.3, to test location, any hotspot can send a challenge every 480 blocks to any other hotspot to request that the "challengee" hotspot prove its location. We wanted to identify any potential challengees with supposed witnesses that are physically impossible. To do so, we matched hotspots' asserted location to the location of where they witnessed for another hotspot. While there were not many unique offenders, there was one common offender whose asserted location was across the country from its witness location. We will refer to this hotspot as Joyful Pink Skunk.[13]

Joyful Pink Skunk's last `assert_location` transaction was on April 11, 2021, when it reasserted its location from the state of Florida to the state of Pennsylvania. We confirm this transition to be honest as the next time the network selected Joyful Pink Skunk as a challengee, it was witnessed by hotspots located in Pennsylvania. Starting May 2, 2021, hotspots in New York, NY and Brooklyn, NY became valid witnesses for the Joyful Pink Skunk. At the time of writing this paper, it still has not reasserted its location, and it witnesses hotspots in the state of New York.

Joyful Pink Skunk never reasserted its location when it moved to New York. Normally, hotspots should be incentivized to update their location to earn PoC rewards. Yet, from the challenge receipts, Joyful Pink Skunk is receiving HNT regardless of whether its current position matches its last asserted location.

Moreover, hotspots do not have to provide an accurate location at all. This is evident through the hotspot Striped Yellow Bird[13] whose only `assert_location` puts it in Spokane, Washington, but all of its challenge receipts place it in San Francisco, California. Nevertheless, it is still rewarded HNT for providing coverage in an area that is about 1,150 km away from its purported location.

Takeaway: If location is not properly considered in the rewarding process, hotspot owners have little to no incentive to keep their location accurate. As we see in Section 8, inaccurate locations impede coverage modeling. The $40 USD cost to re-assert location is designed to promote stable spatio-temporal coverage by deterring hotspot moves, but it does little if owners can skip reporting moves.

7.2 Case Study 2: Lying Witnesses

FCC regulations limit transmitters to +36 dBm EIRP. Yet some witnesses claim an RSSI as high as 1,041,313,293 dBm (presumably either from a buggy radio driver or a misguided attempt to earn more rewards for witnessing "well"). While this value is easily dismissed, it exemplifies that the current PoC model relies on witnesses reporting their RSSI truthfully, while RSSI is easily forged. Colluding, modestly geospatially clustered nodes could easily gossip challengee secrets to increase the number of challenges (plausibly!) "witnessed," and in turn the gossip clique all earn more rewards.

The blockchain implementation has checks that attempt to use RSSI to establish whether a witness is "valid" (and should thus receive PoC reward payment). Real-world RSSI can exhibit enormous variation [5, 18], however, which inevitably makes such heuristics brittle. Ultimately, there are misaligned incentives here. The network wants witnesses to honestly report RSSI to better estimate coverage, while witnesses want to report whatever RSSIs maximize their likelihood of being rewarded for witnessing (independent of whether they are an honest recipient of the challengee's packet).

Takeaway: RSSI is an unreliable, imprecise, and unstandardized measure. Tying reward payments to it will only incentivize gaming the metric. Users with uncharacteristic, but honest RSSIs will be frustrated by unfairly lost revenue and expert manipulators (with access to the cheating detection algorithm running on the public blockchain) will always be able to defeat heuristics.

8 EMPIRICAL TESTING

For our final measurements, we ask the question: (how well) does Helium actually work? We find that while we can deploy devices and recover data, there are significant limitations today in the reliability of Helium. One of the largest challenges for persons considering the own deployment is the absence of a meaningful coverage model – will Helium cover *my* system? We use extant blockchain incentives to derive implicit coverage models, but find these are quite imperfect, which may imply that our models are too simplistic, that current incentives do not sufficiently promote meaningful coverage, or some mixture of both.

[11] `/ip4/35.166.211.46/tcp/2154,/ip4/44.236.95.167/tcp/2154` from https://github.com/helium/router/blob/master/.env-template#L2

[12] For details on the HIP process, see: https://github.com/helium/HIP.

[13] Name anonymized to protect hotspot identity.

8.1 Basic functionality

For a first test, we consider a best-case scenario. We own an (unmodified) original-batch Helium hotspot, which is attached to our campus backhaul network (on a subnet that grants public IPs and access to arbitrary ports). We provision a ST B-L072Z-LRWAN1 LoRaWAN development board, which we choose as it was the first development board explicitly supported by Helium, and it remains at the top of the list of platforms on the Helium Quickstart Guide.[14]

We load a basic app on the device which sends an incrementing counter. The app is a free-running send, which attempts to send another packet immediately after the prior packet response.[15] We run this app for about 24 hours and see a packet reception ratio of 68.61%. We see occasional outages in the network of around 2 hours where no data reaches the cloud. But in between these outages, we see almost all the packets transmitted making through. This experiment was carried out between 18 May, 2021 and 19 May, 2021. During this same time a new firmware was released [11] which is possibly why we see the network outages leading to a lower PRR.

To try to remove this firmware confound, we re-run this experiment in September 2021. We also relocate our sensor to a residential neighborhood with a much greater density of hotspots (location detail in Appendix, Figure 16). Despite these changes, we still see unreliable performance, with an overall PRR of 73.2% across three trials. There are no significant gaps. 83.5% of missed packets are single-misses (i.e. packets before and after were received), 92.2% are single- or double-misses, and the longest sequential run is a single instance of 34 consecutive missed packets.

8.2 Coverage

At the end of the day, the most important question for a wireless infrastructure provider is the quality and availability of service. The Helium network is expanding quickly. As of this writing, Helium is averaging an addition of 1,000 new hotspots per day [14] (a claim we verified in Figure 5). For our studies of coverage, we consider the state of the network as of May 23, 2021.

8.2.1 Coverage Models. The blockchain records hotspot locations, but LoRa is intended as a “long-range” wireless technology. This means we need to develop a model to go from a list of hotspots to expected geospatial coverage. For these coverage analyses, we focus on the United States, as Helium’s initial launch was restricted to US-only territories. While the network is seeing rapid international expansion, the US remains its most established market, which makes it a better representation of current best-case capabilities.

From `helium.com`: Helium provides a “Coverage Map” at explorer.helium.com/coverage. Figure 12a is a screenshot from that website. Green dots represent online hotspots while red dots represent offline hotspots. While the map is a good representation of hotspot locations, this view of coverage can be misleading as the dots indicating hotspots always render at the same size, and thus individual hotspots appear to cover more and more area as the map is zoomed out and the landmass underneath a hotspot-dot grows.

Density Incentive: HIP 15 specifies that hotspots within 300 meters of each other cannot act as a witness for one another. This is to promote wide-area coverage by discouraging hotspots from clustering too closely together. The implication then is that a hotspot should be able to provide coverage to any device within a 300 m radius. We graph the Helium coverage of the contiguous US using this 300 m radius model in Figure 12b. Compared to the Helium coverage map in Figure 12a, the area of coverage for the 300 m radius approach is barely perceivable. The total percentage of contiguous US landmass covered by the 300 meter radius model is 0.09295%.

Witnesses: As a reminder, witnesses are hotspots that report challenge packets transmitted by a challengee to confirm the challengee’s location. There are two types of witnesses: a valid witness and an invalid witness. A witness is marked valid unless it is deemed invalid by satisfying one of the following criteria:

- is too close to the challengee (<300m)
- has too high of an RSSI (several heuristics)
- has too low of an RSSI (several heuristics)
- is pentagonally distorted (rare artifact of H3 distance)
- claims capture on the wrong channel (impossible)

While our 300 m radius coverage map provides a better understanding of the actual coverage, we believe it to be too conservative. To alleviate this fact, we use the valid witnesses of a challengee to infer an empirical measure of coverage. For each challenge, we draw a convex hull around the challengee and its valid witnesses and assume coverage of the interior of this hull. We overlay this coverage model on top of a contiguous US map in Figure 12c.

One problem that becomes evident from the convex hull model is that some of the “valid” witnesses should in reality not be valid (indeed, debugging this model led to the examples for Section 7.1). To craft a more realistic estimate of coverage, we can remove questionable witnesses by using a more realistic max distance. Murata, a top LoRa radio vendor, suggests that the realistic range is, “…more than 10 km, between 15 to 20 km” [15].[16] We look at the distribution of distances for all purportedly valid witnesses in Figure 13. For our revised convex hull model, we choose a generous 25 km cutoff, after which we reject “valid” witnesses and exclude them from the hulls. Figure 12d shows this revised model, with coverage now covering about 0.5723% of the total contiguous US landmass.

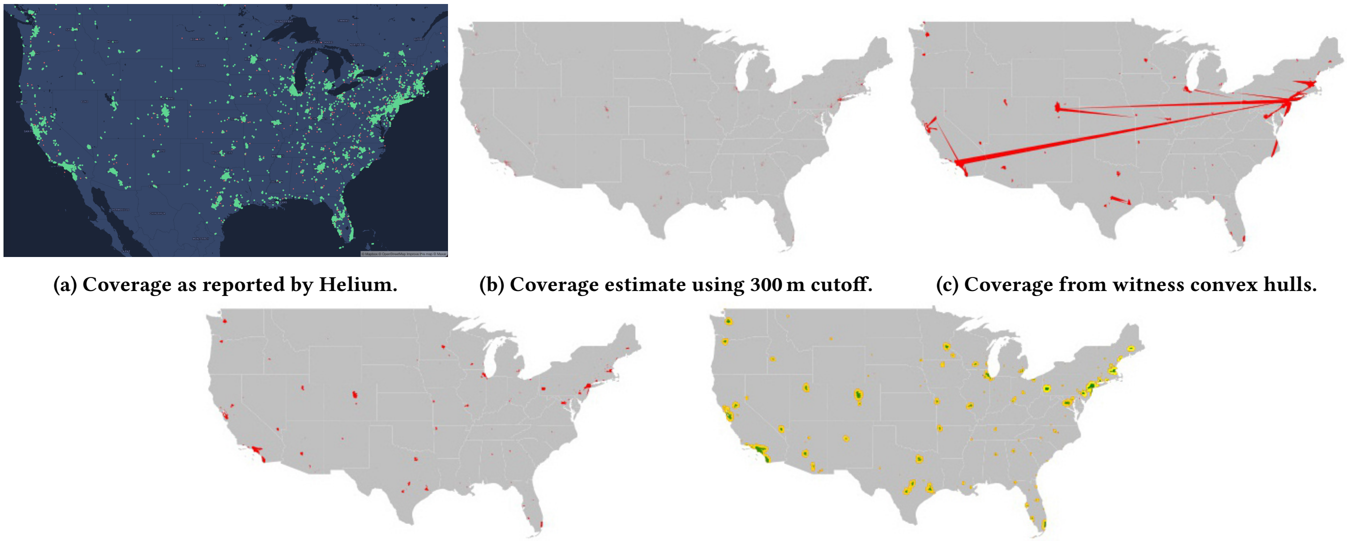
Witness RSSI: While more generous, the basic convex hull model is still too conservative. In particular, it does not factor in coverage by the hotspots that make up the exterior vertices of the convex hulls. We revise the model one final time by including a radial vertex hotspot coverage model and an RSSI coverage model.

The radial coverage by the vertices is simple: we find the distance from the vertex witness to the challengee and use the result as the radial coverage for that witness. As for the RSSI model, we take the vertex witness RSSI and “grow” the witness–challengee radius using the standard free space path loss model, $d = 10^{\frac{w-s}{20}}$ where d is the distance to lengthen the radius, w is the witness’s RSSI value, and s is the sensitivity of the device hoping for coverage. We set s to be a constant -134 dBm as that is the receiver sensitivity of the recommended ST LoRa hardware platform.

[14] docs.helium.com/use-the-network/devices/development/quickstart-guides/

[15] LoRaMAC has two acknowledgment windows, one second and two seconds after transmit. If every packet were ACK’d on the first try, this app would send one packet per second; if no packet were ever ACK’d, it would send one packet every two seconds.

[16] Geography influences this greatly. As example, there appears to be honest hotspots (likely with advanced antennas mounted at high-altitude) in Chicago and western Michigan that witness successfully at ranges of 60-110 km across Lake Michigan.



(a) Coverage as reported by Helium.

(b) Coverage estimate using 300 m cutoff.

(c) Coverage from witness convex hulls.

(d) Coverage from witness convex hulls (removing witnesses that are more than 25 km away from challengee).

(e) Revised convex hull coverage map that factors in the hotspots that make up the vertices of the convex hull and RSSI.

Figure 12: Estimates of Coverage.

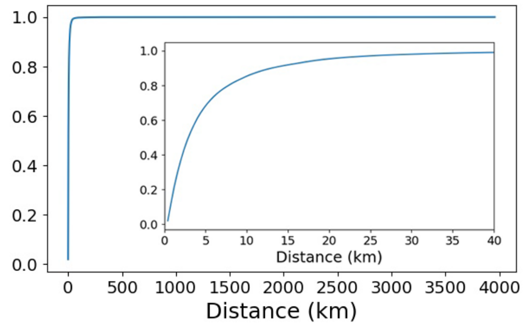


Figure 13: CDF of all the valid witness' distance with an inset CDF of the distance interval from 0 km to 40 km.

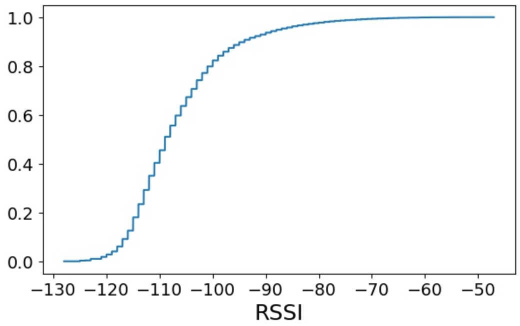


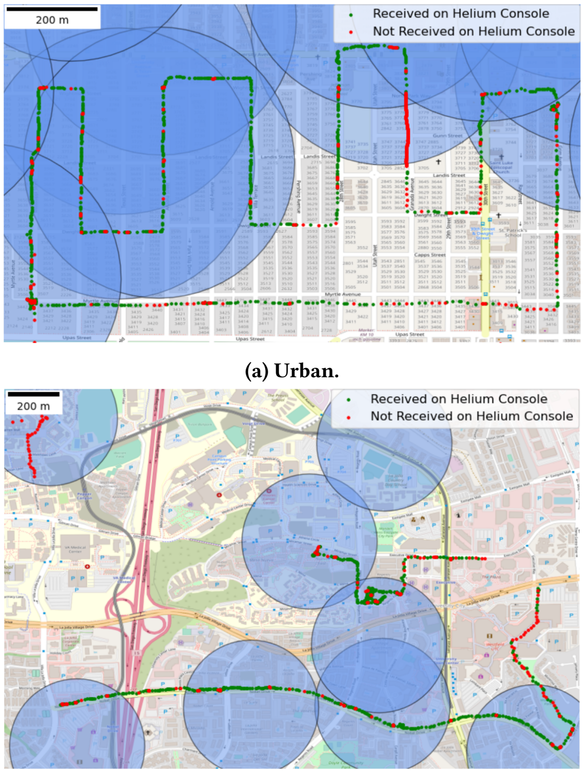
Figure 14: CDF of RSSI values recorded by witnesses during PoC requests from 2021-05-18 to 2021-05-22.

Figure 12e shows the final result of this model. The green regions come from the revised convex hull model. The yellow areas come from the added radial coverage. The almost-invisible red trim around the yellow regions comes from the RSSI coverage. Zoomed out, it is difficult to see the impact of RSSI coverage. Figure 14 shows the distribution of RSSIs reported by witnesses from 2021-05-18 to 2021-05-22. At the median -108 dBm, the RSSI step adds only an additional 20 m of coverage range. With this coverage model, the network achieves 3.3032% coverage of the contiguous US.

8.2.2 Measured Coverage. We run two real-world experiments, with results shown in Figure 15a and Figure 15b, to observe the empirical coverage of hotspots. we plan neighborhood walks through areas with varying hotspot density. While walking, we carry an edge device running the counter app described previously. We add GPS coordinates and a timestamp to the app payload. Packets are also logged to an SD card when sent, to create a record which we can compare to what was received by the cloud. Overall, these walks have a Packet Reception Rate (PRR) of 72.9% and 77.6% respectively.

We investigate whether the HIP 15 promise holds and explains when and where losses occur. Predicting reception when within 300 m of a hotspot is accurate 55.5% of the time, while predicting no reception outside of the radius is accurate for 79.6% of packets.

The LoRa protocol includes an acknowledgment mechanism, where an edge device can request an ACK response from its owning router. At the edge device, failure to receive an ACK is recorded as a NACK. Note that the LoRa PHY is asymmetric; said simply, uplink (edge→gateway) is easier than downlink (gateway→edge) [21]. This means the cloud may record data which the edge device thinks it needs to retransmit. We run statistics regarding ACK and NACK


200 m
Received on Helium Console
Not Received on Helium Console
200 m
Received on Helium Console
Not Received on Helium Console

(a) Urban.

(b) Suburban.

Figure 15: Empirical coverage testing. Green dots represent packets sent by our device and received on the cloud. Red dots are packets that were sent but not received on the cloud. Transparent blue circles show 300 meter hotspot “coverage” radii (hotspots are at the center of the circles). These experiments were carried out by walking outside while attempting to send packets from an edge device to a logging application attached to the Helium Console.

Table 2: LoRa ACK/NACK Validity from Figure 15a

	Packets Sent	Correct ACK	Correct NACK	Incorrect ACK	Incorrect NACK
Count	2393	1106	986	0	301
Percent	100%	46.2%	41.2%	0%	12.6%

Table 3: LoRa ACK/NACK Validity from Figure 15b

	Packets Sent	Correct ACK	Correct NACK	Incorrect ACK	Incorrect NACK
Count	1027	585	237	0	205
Percent	100%	57.0%	23.1%	0%	20.0%

validity as shown in Table 2 and Table 3. We found that there were no false ACK messages and many false NACK messages—packets received on the cloud but recorded as NACKs by the edge device.

There are many factors unknown to us regarding the reliability and coverage of the Helium network. We have only conducted a basic experiment to explore network reliability. Future analyses should deploy their own routers, own more gateways to monitor traffic, and extract more diagnostics from the edge device LoRa stack to enable rich root cause analysis.

9 DISCUSSION

9.1 Legal Considerations

Several of the challenges faced by Helium, such as the high proportion of relayed nodes and the unreliability of hotspot network connections likely stem from hotspots using residential ISPs for backhaul. However, many of these same ISPs are also looking into, or actively rolling out, competing IoT networks [3, 22]. Helium hotspots may not even be permitted on these networks. For instance, Spectrum, the top Helium ISP, may not allow users to host hotspots according to its terms of service:

> Because the Service is for residential use only, any use of the service for non-residential purposes is not permitted and may result in reduction in service, suspension, or termination at the sole discretion of Spectrum. Non-residential purposes include, without limitation, the following: 1. Running any type of server on the system that is not consistent with personal, residential use. This includes but is not limited to FTP, IRC, SMTP, POP, HTTP, SOCS, SQUID, NTP, DNS or any multi-user forums.

Helium hotspots are not hard to detect. They attempt to use a unique port, 44158, and report their IPs to a public database. If Spectrum were to flip the switch and enforce these provisions, at least 17% of the US hotspots would fall offline.

9.2 Looking Beyond LoRa

The initial Helium network targets an unfilled, emerging niche, namely low-power IoT devices. These devices are particularly well-suited to Helium’s micro-transaction model, which enables significantly lower barrier to entry than offerings from traditional service providers. The notion of using residential Internet infrastructure to provide coverage is not new: many service providers today use home routers as APs for wide-area WiFi coverage for their subscribers. MachineQ proposes putting LoRa radios in Comcast Customer Premises Equipment to provide coverage (although *without* paying home users for this geospatial access privilege).

Early on, when competing only in the nascent IoT space, Helium was perhaps not yet a threat. Now, Helium has announced intent to compete in more traditional communication markets, namely 5G. It will be interesting to see how Helium’s meta-infrastructure suppliers respond to this potential business threat and how consumers respond to the opportunity provided by Helium to partake in the profits of the wireless infrastructure ecosystem.

10 CONCLUSION

The Helium network continues to expand rapidly to provide broad connectivity for commodity edge devices. The growth of the network is in part because Helium is providing well-designed infrastructure and incentives. Many individuals are readily purchasing and deploying hotspots to mine HNT—the Helium cryptocurrency—and the incentives drive deployment of base stations that can provide new coverage. While the uncontrolled and unplanned deployment of hotspots allows for efficient growth, it does not necessarily ensure reliable or predictable coverage. Helium continues to propose new policies that attempt to improve coverage and implicitly improve reliability. However, Helium must enforce these policies to ensure network robustness. Also, while Helium network has grown, it still mostly relies on a small number of residential ISPs to backhaul traffic. If those ISPs disallow Helium hotspots on their network, Helium’s viability will be in question.

ACKNOWLEDGEMENTS

We thank our shepherd Jelena Mirkovic and the anonymous reviewers for their comments and support for studying this unconventional infrastructure built on top of the Internet.

REFERENCES

[1] CBSLA Staff. Internet outage hits thousands of southern california spectrum customers. https://losangeles.cbslocal.com/2020/07/30/internet-outage-southern-california-spectrum-customers/, 7 2020. Accessed Sep 2021.

[2] Center for Applied Internet Data Analysis. Mapping autonomous systems to organizations: CAIDA's inference methodology. https://www.caida.org/archive/as2org/, 8 2020. Accessed Sep 2021.

[3] Comcast. MachineQ. https://machineq.com. Accessed: May 2021.

[4] Z. Durumeriç. ZAnnotate. https://github.com/zmap/zannotate/, 2017. Accessed: May 2021.

[5] O. Elijah, S. K. A. Rahim, V. Sittakul, A. M. Al-Samman, M. Cheffena, J. B. Din, and A. R. Tharek. Effect of weather condition on lora iot communication technology in a tropical region: Malaysia. *IEEE Access*, 9:72835–72843, 2021.

[6] B. Ghena, J. Adkins, L. Shangguan, K. Jamieson, P. Levis, and P. Dutta. Challenge: Unlicensed lpwans are not yet the path to ubiquitous connectivity. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[7] A. Haleem. Episode two: The path to 5G. https://blog.helium.com/episode-two-the-path-to-5g-3f704a58661, 4 2020. Accessed May 2021.

[8] Helium. Helium blockchain primitives. https://docs.helium.com/blockchain/blockchain-primitives. Accessed: May 2021.

[9] Helium. New helium partner – careband. https://blog.helium.com/helium-careband-make-covid-19-contact-tracing-affordable-4dcc90eccc4, 2020. Accessed May, 2021.

[10] Helium. New helium partner – nowi. https://blog.helium.com/nowi-is-using-the-peoples-network-to-stop-leaks-and-water-waste-a00bb6aa280f, 2020. Accessed May, 2021.

[11] Helium Engineering Blog. App version 3.2.0. https://engineering.helium.com/2021/05/18/app-version-320.html. Accessed May, 2021.

[12] A. Kumar. Hip10: Usage-based data transfer rewards. https://github.com/helium/HIP/blob/master/0010-usage-based-data-transfer-rewards.md, 2020. Accessed: May 2021.

[13] libp2p. Circuit relay. https://docs.libp2p.io/concepts/circuit-relay/. Accessed: May 2021.

[14] F. Mong. Adding over 1k [helium] hotspots per day. https://twitter.com/fmong/status/1395478353629499393. Frank Mong is the COO of Helium.

[15] Murata. LoRa FAQ. https://www.murata.com/support/faqs/lpwa/lora/hardware/0008. Accessed: May 2021.

[16] Semtech. LoRa network packet forwarder project. https://github.com/Lora-net/packet_forwarder, 2017.

[17] A. Singh, K. Click, R. M. Parizi, Q. Zhang, A. Dehghantanha, and K.-K. R. Choo. Sidechain technologies in blockchain networks: An examination and state-of-the-art review. *Journal of Network and Computer Applications*, 149:102471, 2020.

[18] J. S. Turner, M. Ramli, L. Kamarudin, A. Zakaria, A. Shakaff, D. Ndzi, C. Nor, N. Hassan, and S. Mamduh. The study of human movement effect on signal strength for indoor wsn deployment. In *2013 IEEE Conference on Wireless Sensor (ICWISE)*, pages 30–35, 2013.

[19] Uber. H3.

[20] B. Vakil and T. Linton. Why we're in the midst of a global semiconductor shortage. https://hbr.org/2021/02/why-were-in-the-midst-of-a-global-semiconductor-shortage. Accessed: May 2021.

[21] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovacs, P. Mogensen, and M. Sorensen. Coverage and capacity analysis of sigfox, lora, gprs, and nb-iot. In *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2017.

[22] Verizon. Cat-m1 and NB-IoT solutions. https://www.verizon.com/business/products/internet-of-things/connected-assets/lte-category-m1-narrow-band-technologies/. Access: May 2021.

[23] M. Weiser. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36(7):75–84, July 1993.

A ETHICAL CONSIDERATIONS

This work does not directly interact with human subjects, nor does it make use of network services outside of their intended function. All services used in this work are paid for as intended by the network (using credits earned from operating infrastructure for the network). While we examine case studies of individuals who may be attempting to manipulate or cheat the network, we do not engage in any of these activities ourselves.

All of the analysis in this paper is performed on publicly accessible data. That said, many users and hotspot operators may not be directly aware of the volume or level of detail that is publicly available. We attempt to minimize explicit identification of any individual users in a way that may publicly identify them, or make them easily and immediately available to undue public scrutiny, with the exception of companies who have explicitly and publicly announced their relationship with the Helium corporation and/or Helium network. Where we do perform explicit analysis of public entities, we attempt to show how such mapping of user or operator identity to blockchain transactions might be done by a technically competent user without providing a step-by-step guide.

The nature of the Helium design is such that hotspot owners reveal their location, very often personal residences, to within a few meters. While hotspot owners implicitly consent to the sharing of this information by asserting the hotspot location on the public blockchain, it is unlikely that any one user expects to be singled out in a case study. For this reason, where we do discuss individual device owners, we either omit identifying details unnecessary to the analysis at hand or make use of an appropriate pseudonym.

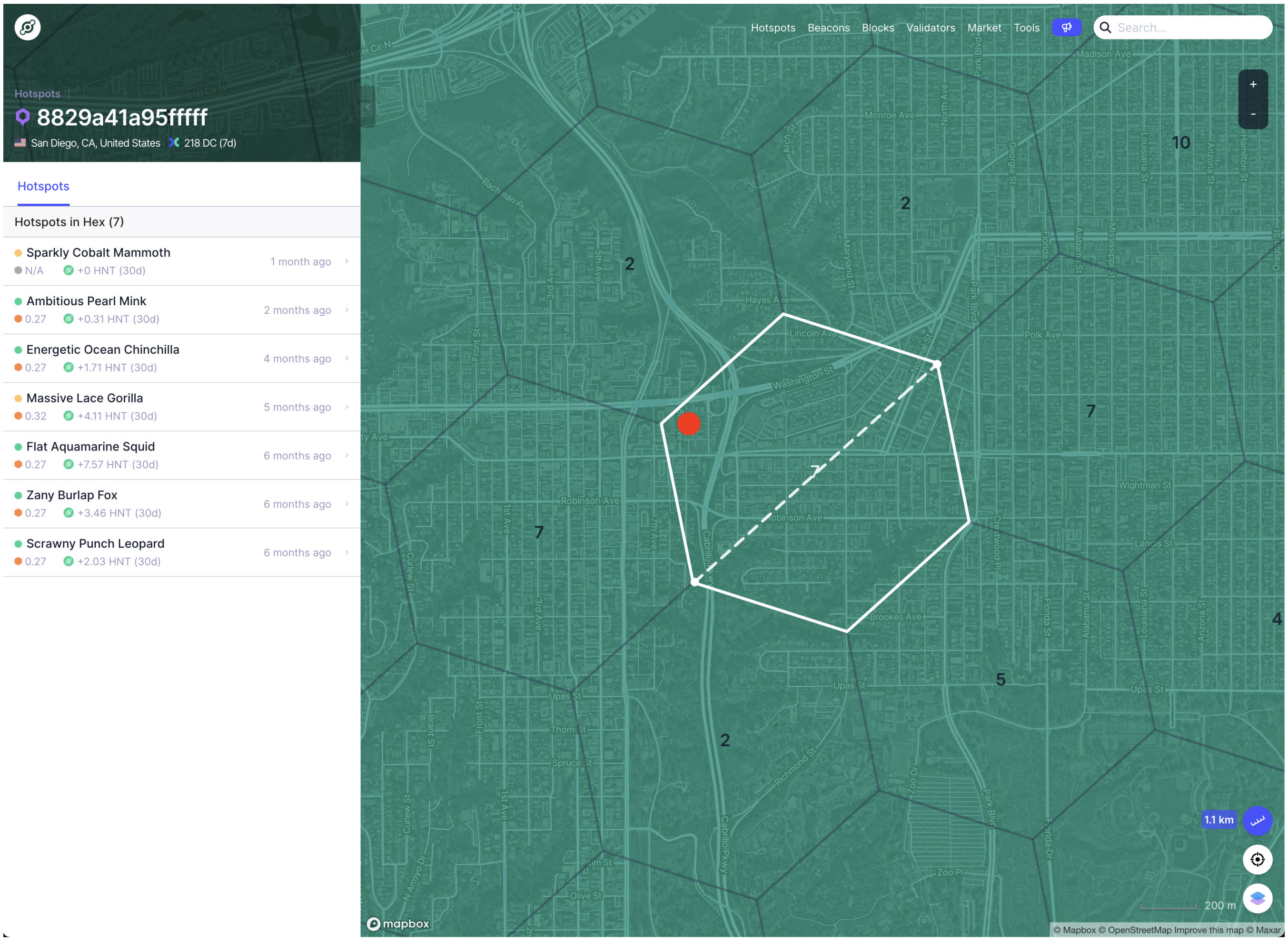
B DISCLOSURE OF POSSIBLE CONFLICTS OF INTEREST

All of the authors of this work are either students or faculty at an academic institution. They have no direct affiliation with the Helium corporation. The lead faculty author met several of the early Helium employees in 2017, but has not interacted with them in any substantial manner since. The Helium corporation has not been involved with this study.

The research group has purchased, deployed, and operates two Helium hotspots. The first hotspot was deployed in winter of 2020, and as a consequence, the research group has acquired several thousand HNT through normal hotspot operations. These tokens have only been used to purchase DC in support of other research activities.

One of the authors owns a modest amount of HNT as a personal investment. This work is not intended to endorse, promote, or in any way influence the value of HNT.

C ADDITIONAL FIGURES



Hotspots

8829a41a95fffff

San Diego, CA, United States 218 DC (7d)

Hotspots

Hotspots in Hex (7)

Sparkly Cobalt Mammoth N/A +0 HNT (30d)	1 month ago	›
Ambitious Pearl Mink 0.27 +0.31 HNT (30d)	2 months ago	›
Energetic Ocean Chinchilla 0.27 +1.71 HNT (30d)	4 months ago	›
Massive Lace Gorilla 0.32 +4.11 HNT (30d)	5 months ago	›
Flat Aquamarine Squid 0.27 +7.57 HNT (30d)	6 months ago	›
Zany Burlap Fox 0.27 +3.46 HNT (30d)	6 months ago	›
Scrawny Punch Leopard 0.27 +2.03 HNT (30d)	6 months ago	›

Figure 16: Annotated view of the Helium Explorer (https://explorer.helium.com/hotspots/hex/8829a41a95fffff). The sensor transmitting packets for the second experiment in Section 8.1 is located at the red dot. The sensor is unencumbered, outdoors, and placed near no metal on a plastic, elevated platform at approximately head-height on a second story balcony in a mixed residential neighborhood immediately surrounded by two-story structures but with three- to five-story structures on the block. The clearest line-of-sight is to the north and east where the balcony faces alleys. Spot-checks on the diagnostics available in the Helium Console reveal that at least six different hotspots ferry data from this sensor over the course of the experiment. Receiving hotspots report an RSSI ranging from -120 to -55 (n.b. the -55 is our owned hotspot, located in the same structure as the sensor; remarkably, it is rarely chosen by the Console, perhaps because this hotspot is on a NAT’d residential connection and is relayed). The best third-party hotspot reports RSSI values around -90.