

Measuring the Internet during Covid-19 to Evaluate Work-from-Home

Xiao Song

University of Southern California
songxiao@usc.edu

John Heidemann

University of Southern California
johnh@isi.edu

ABSTRACT

Covid-19 has radically changed our lives, with many governments and businesses mandating work-from-home and remote teaching. However, work-from-home policy is not always known globally, and even when enacted, compliance can vary. These uncertainties suggest a need to *measure* work-from-home and confirm actual policy implementation. Work-from-home changes which networks people use during the day. Some networks reflect the presence of computers with more public IP addresses that respond to pings. Our insight is that we can detect workday networks by their diurnal and weekly patterns of IP address response, then detect work-from-home when those patterns change. We develop new algorithms to extract these patterns from existing, continuous, global scans of most of the responsive IPv4 Internet (about 5.1M/24 blocks). Applications of these new algorithms to existing data allow us to study the period before and after the spread of Covid-19 to see the global reaction. We demonstrate this algorithm in networks with known ground truth, evaluate the data reconstruction and algorithm design choices with studies of real-world data, and validate our approach testing random samples against news reports. We use our algorithms to identify changes that are consistent with Covid-19 actions, and also identify a government-mandated Internet shutdown that was not Covid-related. Our results show the opportunity to use Internet observation to infer real-world behavior and policies.

KEYWORDS

Internet Measurement, Active Probing, Covid-19, Work-from-Home

1 INTRODUCTION

In 2020, the Covid-19 pandemic has completely changed our lives. Media reports show millions of people are forced to

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Under Review (submission), Nov. 2021, Virtual Location

© 2021 Copyright held by the owner/author(s).

stay at home, with some working from home and others becoming unemployed [19].

However reactions to Covid-19 are not always certain. Public reports of Covid-19 responses (such as stay-at-home) are not always timely—the policy in some countries may be unclear abroad, and its implementation may be uneven. Even when a government establishes a policy, the people may embrace or reject that policy, so actual participation may diverge [37]. For these reasons we would like to confirm actual stay-at-home practice.

The Internet has changed the ways in which we interact with society, and Covid-19 has increased our Internet use [45]. As one example, early in the pandemic streaming media providers reduced video quality to manage increased use [23]. Facebook reported a sharp 20% increase in traffic in late March after widespread stay-at-home [21]. Researchers saw a similar 15–20% increase in traffic at IXPs in mid-March [38]. Changes depend on the perspective of the observer: mobile (cellular) networks show a 25% drop in traffic and a decrease in user mobility as people stay at home [47]. While how the Internet changes varies, these changes hold the potential to allow us to actually *observe* use and compliance with stay-at-home orders, if they can be interpreted correctly.

In this paper we look at *Internet address use as an indication of stay-at-home*. Many people use mobile devices (laptops, tablets, smartphones) on the Internet at the home and office, and these devices acquire IP addresses when they are using the Internet. Often these addresses are dynamically allocated using protocols such as DHCP. In many networks today, these dynamic allocations use private address space [53], in some cases these changes in use are visible in the public IPv4 addresses space [25, 49, 50, 52, 63].

Our premise is that we can see these changes in public IPv4 address usage, and that it can be interpreted to indicate stay-at-home behavior. We expect to see increased address use at home networks and decreased use of office and school networks. We hope that this information can provide information about stay-at-home activity in places where policies are unknown and about compliance with stay-at-home in places where compliance is uncertain.

We explore this question by developing new algorithms that can track changes in address usage through new analysis

of ongoing active scanning for Internet outage detection [51]. This approach has the advantage of broad, long-term, fine-grained coverage—measurements go to 4 to 5 million /24 IPv4 networks, have taken place consistently since 2014, and it provides new data about individual addresses every 11 minutes. Reusing existing data is necessary to capture changes that happened when Covid-19 spread globally in early 2020. However, re-analysis of existing data requires care, since probing rates vary and the probing mechanism is tuned for other purposes.

The first contribution of our work is to define a new algorithm that identifies changes in network usage of each /24 IPv4 block, a signal that correlates with work-from-home (§2). We examine the active state of /24 blocks by counting the number of active IPv4 addresses—those that reply to an ICMP echo request. We find *change-sensitive* network blocks, where the number of active IPv4 addresses reflect the number of individuals actively using the network in that location, and we see noticeable diurnal changes. Our algorithm identifies these blocks, determines long-term trends that may indicate changes in usage, and detects changes in these trends that correlate with work-from-home.

Our second contribution is to validate this algorithm against real-world data. We learn from known cases of Covid-related work-from-home where we know ground truth (§3). We systematically evaluate how the design decisions made in our algorithm (§4). Finally, we validate detection of inferred work-from-home (§4.5) and demonstrate that our approach discovers Covid-related changes (§4.6), using random samples tested against public news reports.

Our final contribution is to apply our algorithm to study 447k change-sensitive blocks worldwide over the first six month of 2020 (§5). We evaluate known real-world events, like changes in Wuhan in January 2020, and we discover Covid-related events unknown to us in India and the Philippines. Our algorithm also identifies non-Covid-related Internet shutdowns in India in February 2020.

Our source data is already available at no cost, and we plan to make the data resulting from our work and our analysis software available with the publication of our paper.

Although we examine IP addresses to observe behavior resulting from individual actions, we use IP addresses in the aggregate and do not relate them to specific individuals. Our work has therefore been classified as non-human-subjects research by our university’s IRB (USC #UP-20-00909)..

2 METHODOLOGY

To detect changes in Internet use and infer work-from-home, we begin with address scans, then filter and analyze the data with the following steps:

- (1) Probe IP addresses for activity
- (2) Identify active addresses by accumulating partial scans

- (3) Identify change-sensitive blocks
- (4) De-trend address usage
- (5) Detect changes in usage
- (6) Combining data from multiple observers

2.1 Probing IP Addresses For Activity

Our approach observes active probing of the IPv4 address space. We require that many active IP addresses be probed, that it reports both positive and negative or non-replies, and that each address should be reprobbed frequently enough so that we can detect diurnal changes in address response.

2.1.1 Potential Data Sources. In principle, any frequent public scan of the IP space can serve as input, such as data from USC’s IPv4 census and surveys [43] or Trinocular outage scans [51], CAIDA’s Archipelago [26], ZMap [35], or Censys’s scans [34]. Of these, we currently focus on Trinocular because it is constantly updating information about millions of IPv4 blocks every 11 minutes, and it provides data taken 24x7 since November 2013. Critically to our work, it includes data from months before and after Covid-19 swept the globe. While it provides complete coverage, it adds some addresses incrementally (from 1 to 16 addresses) following a known, deterministic pattern, each 11 minutes [20]. In §2.2 we discuss how we assemble these observations into a complete picture.

Alternative datasets include USC surveys, ZMap, CAIDA Archipelago, and Census. USC Internet surveys cover all IP addresses in each /24 block, but only for about 40k blocks for two-week periods. We use survey data for validation in §4.2. ZMap has run quick, complete scans on occasion, but it reports only positive replies and it stateless probing requires that we infer negative results. Archipelago covers all routed /24s blocks, but it is too slow for our purposes, with teams 17-18 probers covering all routed blocks every 2-3 days. Censys scans emphasize daily services and certificates on hosts and not changes in reachability over a single day. Both are potential input sources, but bulk availability of timely data is unclear.

Another option is to deploy our own probing. While it would be ideal to optimize probing for our purposes, our desire to see what happened before and as Covid-19 began in 2020 requires that we use existing datasets to perform the analysis. In addition, high-rate probing imposes a small but non-zero cost on the world’s networks, motivating us to maximize reuse where possible. Reuse also avoids duplication of existing opt-out mechanisms and abuse handling, a noticeable operational cost of sustained probing.

2.1.2 Datasets Used In This Paper. This paper uses the datasets listed in Table 1. We obtained this data as collected and provided at no cost to researchers. We plan to make the results of our analysis available on similar terms.

abbr.	dataset name	start	duration
2019q4-w	internet_outage_adaptive_a38w-20191001	2019-10-01	12 weeks
2020q1-w	internet_outage_adaptive_a39w-20200101	2020-01-01	12
2020q2-w	internet_outage_adaptive_a40w-20200401	2020-04-01	12
2020q1-j	internet_outage_adaptive_a39j-20200101	2020-01-01	12
2020q2-j	internet_outage_adaptive_a40j-20200401	2020-04-01	12
2020q1-n	internet_outage_adaptive_a39n-20200101	2020-01-01	12
2020q2-n	internet_outage_adaptive_a40n-20200401	2020-04-01	12
2020q1-e	internet_outage_adaptive_a39e-20200101	2020-01-01	12
2020q2-e	internet_outage_adaptive_a40e-20200401	2020-04-01	12
2020q1-g	internet_outage_adaptive_a39g-20200101	2020-01-01	12
2020q2-g	internet_outage_adaptive_a40g-20200401	2020-04-01	12
2020it89-w	internet_address_survey_reprobing_it89w-20200219	2020-02-19	2

Table 1: Datasets used in this paper. w: ISI-West in Los Angeles; j: Japan data from Keio University (SJF campus) near Tokyo; n: Netherlands data from SurfNet; e: ISI-East data from near Washington, DC; g: Greek data from Athens University of Economics in Business.

2.2 Identifying Active Addresses

As described above, we reuse existing data to provide visibility into the Internet before and at the onset of Covid-19, and to avoid duplicating existing measurements. We reuse data from Trinocular [51]. Trinocular scans the visible Internet incrementally, so we need to reconstruct the state of the Internet from these incremental results.

Re-analysis of Trinocular begins by determining the state of all addresses in the block by accumulating data over multiple Trinocular rounds. We assume addresses do not change state until they are re-scanned. We accumulate addresses until we have observed every active address in the block. At that point, we begin updating our prior estimate as new observations arrive. Thus we generate new estimates of the number of active addresses every 11 minutes, but each estimate reflects data from multiple prior rounds.

The exact time needed to report on all addresses in a block varies as function of how full the block appears to be (Trinocular scans only addresses that have been active in the last two years, the *ever active* addressees) and how many addresses currently respond (Trinocular stops scanning each round when it gets enough responses). In §4.1 we examine how long this takes in practice for our goal, building on prior work that reconstructed block state for a subset of blocks [20].

The idea of our algorithm is to track the state of IP addresses (either up, down, or unknown) of /24 blocks for every 11 minutes round, keep updating the active states when the state of IP addresses are changed at this round, and count the number of active addresses of /24 blocks at specific time when all addresses in the ever-active list have been probed.

Once we have obtained the active states of all addresses that are in pre-generated ever-active IP addresses list, meaning no address is marked as unknown, we then output the

number of currently active addresses for each /24 block. The corresponding timestamp is the round in which all addresses in the ever-active list have a determined active state. We regard all previous round as a full round.

In the following round, as long as one address changes its active state (from up to down or from down to up), we will output a new number of active addresses at that timestamp.

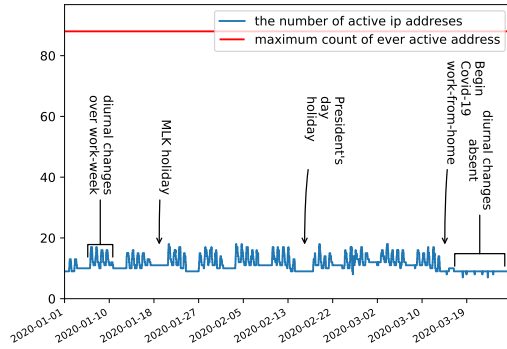
2.3 Identifying Change-sensitive Blocks

To detect work-from-home we must have blocks where address changes reflect people’s daily schedules. We call such blocks *change sensitive*, and identify them by two characteristics: first, they show a regular, *diurnal* pattern; second, they need to show enough *swing* (change) in the number of active addresses between day and night that we can detect changes in use with confidence.

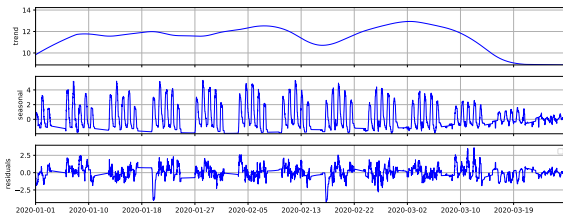
Figure 1a shows an example change-sensitive block. This graph shows the number of active addresses in this block over three months (the blue line). We usually see five clear bumps corresponding to the work week, followed by two days of flat behavior. This block is at USC, so we confirm these changes correspond to laptops using DHCP-assigned IP addresses from discussions with the network operators. In addition, these three months included two 3-day weekends both of which show up with 4 (not 5) workdays.

Our approach must look for blocks that reveal people’s daily work schedules. To identify these blocks, we examine all ping-responsive blocks and look for blocks showing *diurnal* changes that are large enough to reliably detect. In §4.3 we show that between 168k and 450k blocks meet these two requirements, depending on how much and how long we collect data.

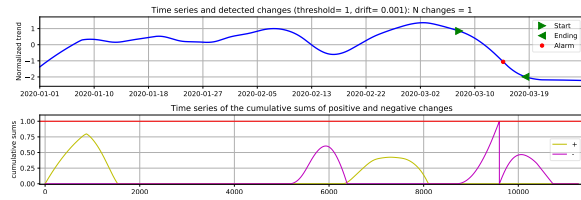
Diurnal blocks: We identify diurnal blocks by taking the FFT of the active-address time series and looking for energy



(a) Active addresses over 3 months (input data).



(b) STL components: trend, seasonal, and residual.



(c) STL trend with detections (top) and CUSUM sums (bottom).

Figure 1: A laboratory (128.9.144.0/24) illustrating address usage changes due to work-from-home.

in frequencies corresponding to 24 hours (or harmonics of that frequency). This approach follows that developed in prior work [52], which showed that IP addresses often reflect diurnal behavior, particularly in Asia, South America, and Eastern Europe.

Sudden absence of the diurnal pattern indicates a network change consistent with work-from-home. The block of Figure 1a is diurnal from 2020-01-01 to 2020-03-15 and becomes flat after 2020-03-15, corresponding to the start of work-from-home. Because we look for absence of diurnal changes in previously diurnal blocks, we must establish a pre-Covid baseline. In §4.3 we show that the first month of 2020 provides a good baseline.

Persistent daily swing: We look for blocks that have a “significant” daily swing in addresses so we are not making

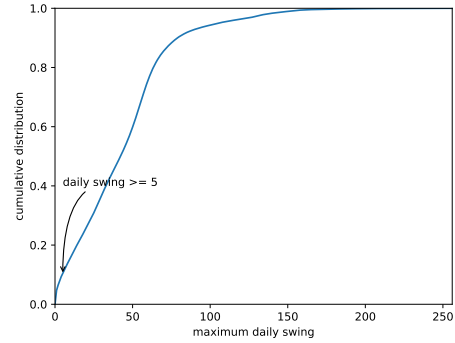


Figure 2: Cumulative distribution of maximum daily swing of all diurnal blocks in 2020q1.

decisions of just one or two addresses. We define the daily swing as the difference between minimum and maximum number of active addresses over the midnight-to-midnight UTC 24-hour window .

We define significant swing as a change of 5 or more addresses, based on evaluation of the data. (Too large a threshold will reduce the number of accepted blocks, but too small makes the algorithm vulnerable to noise.)

Figure 2 shows that there are around 90% diurnal blocks with maximum daily swing greater than 5.

Although we require a daily swing, many blocks (like Figure 1a) show use primarily during the work week and not on weekends and holidays. We therefore require blocks to have a significant daily swing for at least 4 of 7 consecutive days. We consider it as one seasonal week. We filter blocks that have daily swing greater than 5 and at least one seasonal week.

2.4 De-trending Address Usage

We look for diurnal blocks with swings to find blocks that reflect the network usage at work, but daily fluctuations make it difficult to detect changes in use. We expect work-from-home will either remove the diurnal swing, as occurs in Figure 1a after 2020-03-15, or the overall use (and possibly also the size of the swing) will fall as fewer people come into work. Either way, we need to track the mean behavior over the course of a full 24 hours, assuming we “average in” any daily swing.

We track the underlying baseline by applying a standard seasonality model to the data that decomposes the signal into a baseline convolved with a daily and possibly weekly signal. We considered two models: the “naive” seasonality model [57] and the Seasonal-Trend decomposition using LOESS (STL) [30, 57]. Although both are similar, we adopted the STL for our work after comparing the two and finding it slightly more robust.

Figure 1b shows the decomposition from our sample block (Figure 1a) into trend, seasonal, and residual components. We use the trend (top of Figure 1b) to detect changes that occur independent of the diurnal swing (seasonal) and noise (residual).

2.5 Detecting Changes in Usage

Finally, using the STL trend as an estimate of the baseline, we can look for changes in that baseline.

We apply a standard change-point detection algorithm, CUSUM [33, 42]. CUSUM looks for changes in the baseline flags when the upward or downward trend begins, and identifies the specific point in time with largest changes.

Figure 1c shows an example of the STL trend indicating detected change. The bottom graph shows cumulative increase and decrease (purple and yellow colors), and we highlight start and end of changes on the trend line.

This graph shows the change point detected by our algorithm for 128.9.144.0/24. This is a change-sensitive block, with diurnal behavior and a consistent address swing. We can observe the diurnal swing goes away around 2020-03-15 from both Figure 1a and the upper bar chart. The upper bar chart shows the time series of the number of active IP addresses of 128.9.144.0/24 over three months. The middle bar chart shows the trend of the time series. The red dot represents the time when a change point is claimed by the algorithm. The bottom bar chart is the cumulative sums of the normalized trend over time. The change point detected around 2020-03-15 reflects the ground truth that work-from-home begins at ISI/USC.

We assume that there are more work blocks which are change-sensitive than home blocks, so we consider a downward change to indicate increased work-from-home, with a much later upward change to signal a return-to-work. This assumption is true if home networks are likely to either use a single, always-on IP address, as is the case in the U.S. and western Europe, or that home networks are not using public IP addresses, as is the case in some countries like Iran.

Internet outages also appear as a downward change followed by an upward change. We assume closely timed down and upward changes indicate outages since most outages are short, while Covid-induced work-from-home lasted weeks or months.

2.6 Using Multiple Observers

Our description of estimating block usage in §2.2 describes accumulating addresses from a single observer. However, typically Trinocular has multiple observers in different geographic locations (to provide different perspectives). Each observer has the same list of destinations, but they run independently. Therefore, they are almost always out of phase with each other.

We can strengthen our results by combining data from multiple observers, either to accumulate observations rapidly, or to validate results of one against the others.

To provide more complete data, we can combine multiple results into a single, more timely set of observations. Because Trinocular stops probing as soon as it gets a positive response, blocks with many positive responses are probed more slowly, requiring nearly two days to probe all 256 address in the theoretical worst case. To observe diurnal behavior, we need to scan all addresses at least twice per day. Since observers are independent, combining multiple observers can reduce time to scan the full block.

Alternatively, separate observers can be treated as equivalent but independent, validating their results against each other. We know that different vantage points sometimes see different replies [43, 62]. We can therefore compare results of observations from different sites to confirm that our results hold and avoid any vantage-point-specific bias.

We follow the first approach, combining results from multiple observers, since the additional information improves reconstruction of the true signal (see §4.2) and increases our coverage (§5). This choice brings the time to scanning a full block under 12 hours, a requirement to observe diurnal effects.

3 CASE STUDIES

Before validating our approach we examine several blocks from the University of Southern California where we can confirm changes with the network operators.

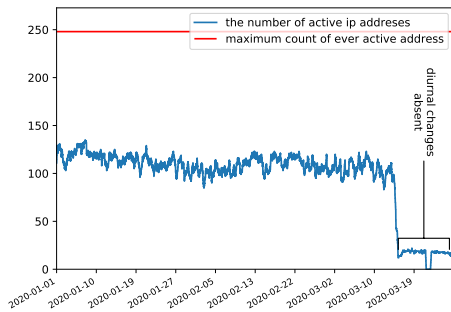
We know that USC switched to online classes on 2020-03-11. Spring recess started on 2020-03-16, with many students leaving campus. Classes resumed the following week (2020-03-23) with distance learning and work-from-home continuing.

3.1 Address Dynamics in a Research Lab

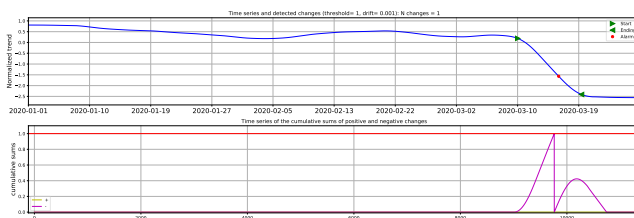
First we examine a /24 address block (128.9.144.0/24 allocated to a research lab at USC/ISI with a mix of researchers and students, using dataset 2020q1-w (Figure 1). Network operators confirm that these addresses are statically assigned to individual computers.

Some addresses are allocated to laptops that people bring in to work, and each of these laptops is assigned a unique IP addresses with DHCP. Other addresses in the block are for office computers that are always on. Figure 1a shows the number of active addresses in this block over 2020q1.

We see between 8 and 18 addresses are active, with a baseline of 10 active addresses and workday “bumps” of 7 or 8 addresses. These bumps do not appear on weekends or holidays (2020-01-19 and 2020-02-15 are U.S. national holidays). These results confirm our ability to see work-from-office our in address space data.



(a) Active addresses over 3 months (input data).



(b) STL trend with annotations (top) and CUSUM sums (bottom).

Figure 3: A VPN block (128.125.52.0/24) and detection.

We see that the diurnal bumps go away starting 2020-03-15 with the start of work-from-home—the laptops are now using addresses in home networks, not ones in this address block. In addition, we see the baseline drops from 10 to 7, suggesting some always-on computers were shut down.

Our change-point detection in Figure 1c shows three changes in usage: an increase at the beginning of January as researchers return to work after winter break, a slight dip around 2020-02-13 when several always-on computers are turned off and there is a 3-day weekend. Finally, there is a large change around 2020-03-15 corresponding to the start of work-from-home.

This block illustrates the diurnal pattern with significant daily swing that we look for, and shows how change-detection identifying the start of work-from-home.

3.2 A Pre-Covid VPN

We next consider a /24 block (128.125.52.0/24) that is part of USC’s VPN (Figure 3). We initially determined it was VPN based on reverse DNS addresses, and then later confirmed this use with USC network operators.

Figure 3a shows the number of active addresses over time. We see that after 10 weeks of steady use, the address use drops off significantly, just as work-from-home begins. This outcome seems backwards from what one would expect—VPN use should go up with work-from-home. USC network

operators explained that they shifted the campus VPN to a newer, larger address space because of an anticipated increase in use. Thus use of this address space went down because use shifted to another block.

Figure 3b confirms that this change in use is detected by our change-point detection. We observe the number of active IP addresses has a significant drop around 2020-03-15 based on Figure 3a and the upper bar chart of Figure 3b. The change point detected around 2020-03-15 reflects ground truth that work-from-home begins at USC.

Our detection algorithms find it, although they cannot (by themselves) show that these users moved to a different address block.

4 VALIDATING DESIGN CHOICES

We next discuss design choices, evaluating them using our datasets: Do we track block state quickly enough to see diurnal changes? How accurate is the reconstruction? How many blocks do we see, and where are they?

4.1 Block Refresh Rate

We examine how quickly we complete scans of each /24 block, a *full block scan*. Prior work has suggested sparsely occupied blocks are fully scanned in about two hours in the worst case [20], but that analysis covered only the subset of all blocks that were intermittently responsive. Here we care about change-sensitive blocks, a different subset, prompting a new analysis.

Prior work placed bounds on the time in the context of the Full-Block Scanning algorithm to address false outages in sparse blocks [20]. This analysis showed 17 rounds (3.1 hours) as an upper bound for scanning sparse blocks, based on 15 probes per each round, over 11 minutes, and 256 addresses to cover in the block. This analysis uses 15 probes per round (the maximum) because it assumes blocks are sparse and so most addresses do not respond. A more general worst case will consider blocks where all 256 addresses always respond, so only one address is probed per round and a full scan requires 256 rounds (1.8 days). Such blocks are not change-sensitive, but this rare worst case suggests we need to look at block scanning duration empirically.

We computed time to scan all ever-responsive addresses in each block across all blocks in 2020q1 for three cases: when scanned from single observer, from three observers, or from five observers (as per §2.6). Figure 4 shows these results. We see that about 80% change-sensitive blocks can be fully scanned in 8 hours or less when data is combined from five observers (the left-most dark blue curve, compared to the left dashed line at 8 hours). By contract, 8 hours provides about 60% of blocks with three observers, and about 35% with one. Given 12 hours, 5 observers cover almost all blocks (95%), compared to 80% with three and 50% with one. We

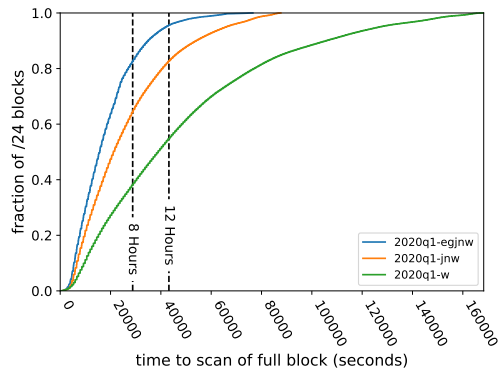


Figure 4: Cumulative distribution over all blocks of time to complete a scan of all known active addressees, varying number of VPs: 1 (bottom, right line, VP w), 3 (middle line, VPs j, n, and w), 5 (top, left line, VPs e, g, j, n, and w).

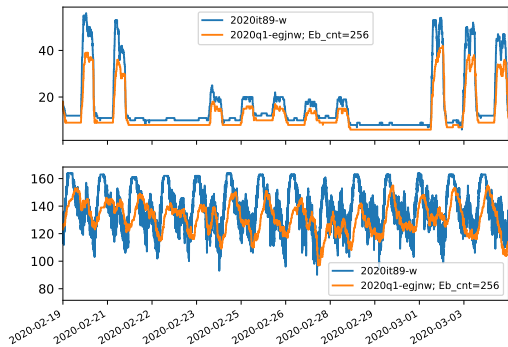


Figure 5: Comparing two reconstructed /24 blocks with their ground truth.

conclude that multiple observers are important to ensure we see diurnal changes.

4.2 Reconstruction Quality

Time to refresh is one factor that affects reconstruction of the signal from observations. We examine reconstruction accuracy by comparing reconstruction to ground truth from a complete scan. For ground truth we use freely available Internet address surveys. A survey covers about 2% of the responsive blocks in the IPv4 Internet, with data for every address in each block every 11 minutes, over two weeks. Here we compare the 5,440 blocks that overlap survey 2020it89-w (Table 1) and the change-sensitive blocks reconstructed from five observers (2020q1-egjnw).

4.2.1 Reconstructed blocks. We begin by examining sample blocks in detail. We generate reconstructions of all 5,440 overlapping blocks and examine time series of number of active addresses for a random sample of 50 blocks. Two

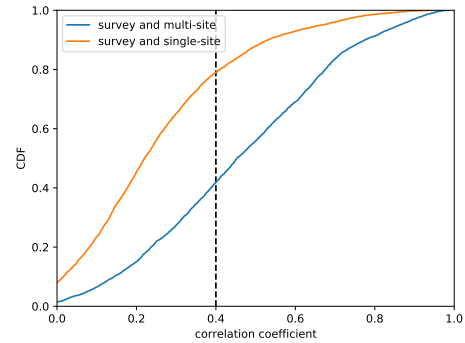


Figure 6: Cumulative distribution of the correlation coefficient over the reconstructed time series and the ground truth.

representative examples are shown in Figure 5, with the dark blue line showing ground truth from the survey, and the light orange line showing our reconstruction.

The top graph in Figure 5 shows a good reconstruction—the correlation coefficient is 0.89 and block is fully scanned in 3300 s (just less than an hour). We see that the shape of the reconstruction matches very well, with the orange line having the same daily peaks indicating working hours, sharp changes, and flat behavior between workdays.

When doing change-detection, this allows us to observe both daily and weekly trends. While the reconstruction is consistently lower than truth, that does not affect change detection. We believe this underestimate is because Trinocular stops probing when it gets a positive reply, so during the work-day it is slow to discover all of the active addresses.

The lower figure shows a more challenging block. This block is heavily used (120 to 160 active addresses), and it sees a large daily shift, with consistent changes every day of the week. Here the correlation coefficient is 0.40, a worse fit than our first example. Because many addresses are active, a full scan requires around 8 hours. For this block, the reconstruction (the light yellow line) “spreads out” some of the true behavior of the block, flattening the peaks and raising the valleys. Again, this reconstruction error is because the block has so many active addresses, Trinocular adds less information per round over the course of the day. Reconstruction is poorer for this block than for the prior example block, because each Trinocular round terminates after finding one active address. Even though the reconstruction is not perfect, both blocks’ reconstruction and ground truth are identified as change-sensitive. It is still successfully handled by our algorithms, but more error might cause us to misclassify the block as not change-sensitive.

Dataset:	2020it89 -w	2020q1 -w*	2020q1 -egjnw*	2020m1 -egjnw*
duration (weeks)	2	12	12	4
completeness (sites)	full	1 site (*intersected with it89-w)	...5 sites
responsive				32,437
not diurnal	25,170	30,137	29,493	29,049
diurnal	7,257	2,300	2,944	3,388
narrow swing	15,104	12,597	11,112	11,112
wide swing	17,333	19,840	21,325	21,325
not change sensitive	26,997	30,630	30,434	29,890
change sensitive	5,440	1,807	2,003	2,547

Table 2: The number of each types of /24 blocks detected by Trinocular reconstruction and Internet address surveys.

These block show reconstruction quality varies by block, with larger, denser blocks resulting in longer scan times and poorer reconstruction. However, they also show that even with only moderately good reconstruction, like the second example in Figure 5, we still classify the block as change-sensitive and are able to detect changes. We conclude that even though Trinocular is specifically made for outage detection, the reconstruction is often good enough for our detection purposes.

4.2.2 Quantifying the fit. To quantify goodness-of-fit, we compute the correlation coefficient for 5,440 change sensitive blocks in both datasets. Figure 6 shows the CDF of correlation coefficients for these blocks, comparing reconstruction with one site (the top, left yellow line) to five sites (the right, darker blue line). A higher correlation coefficient indicates a more accurate reconstruction, and we see that the median block has a correlation coefficient of only 0.22 when observed from a single site, but 0.45, when results of all five sites are combined.

4.2.3 Algorithm success: Diurnal blocks, large daily swing, and change detection. Although the 0.45 correlation coefficient with five sites is much better than 0.22, it is imperfect. However, our goal is that the reconstruction supports change detection, perfection is not required. (Both reconstructions in Figure 5 support change detection, even though the bottom block has a correlation coefficient of only 0.40.)

We measure how well our algorithm (change detection) and its components (diurnal and wide swing) compare in Table 2. We compare four cases: two weeks of full data (2020it89-w), one observer for a quarter (2020q1-w), five observers for a quarter (2020q1-egjnw), and five observers for a month (2020m1-egjnw). The 32k responsive blocks are probed by both Trinocular observers and Internet address surveys.

Table 2 shows that for the same /24 blocks, the number of change sensitive blocks and their components vary depending on the duration and completeness.

For diurnal, Internet address surveys detects the most diurnal blocks, which is around 22%, whereas Trinocular detects 7.1%, 9.1%, and 10.5% for one observer for a quarter, five observers for a quarter, and five observers for a month, respectively. The best reconstruction (2020m1-egjnw, one month) captures around half diurnal blocks in Internet address surveys.

Two factors influence the number of diurnal blocks: reconstruction and duration. We already saw (§4.2) that more observers provide a better reconstruction, so we see somewhat more diurnal blocks with 5 observers than with just 1 (comparing 2020q1-egjnw to 2020q1-w). However, a longer duration *also* matters—a diurnal block must change every day, but with longer duration there may be more days where the block is flat because the underlying block usage changes. We see this effect comparing 2020q1-egjnwj with 2020m1-egjnwj, where the longer duration (12 weeks in q1 vs. 4 in m1) give the block more days to not show change.

The effect of observation duration is a concern when looking for *end* of work-from-home in late 2020 and in 2021—we expect networks to change due to work-from-home. However, if we *detect* that change because a diurnal block ceases to be diurnal, we may be unable to detect the block’s recovery since our algorithm will discard it as no longer diurnal.

We see that reconstruction successfully identifies blocks with daily swing. In Table 2, 53.4% responsive blocks have wide daily swing in surveys, and we see slightly more in reconstruction—61.1%, 65.7%, and 65.7% for one observer for a quarter, five observers for a quarter, and five observers for a month, respectively. All reconstructions observe about the same number of daily-swing blocks because blocks with swing are well covered since they draw many probes. (Recall Trinocular probes until it finds an active addresses, and in these blocks, many addresses are non-responsive at night.)

Change sensitive blocks must be diurnal and show a large daily swing. We know that diurnal detection is more difficult in reconstruction, and we see this challenge reflected in smaller numbers of change-sensitive blocks with reconstruction: 5.5%, 6.1%, and 7.8% for one observer for a quarter, five observers for a quarter, and five observers for a month, compared to 16.7% in ground truth.

We conclude our use of reconstruction decreases sensitivity, capturing only about half the change sensitive blocks we should see if we probed every address always (compare 2020m1-egjnw to 2020it89-w). However, adaptive probing is *required* to make probing tractable, since probing every address every 11 minutes is technically possible but would draw too many abuse complaints to be sustainable. It is also required to allow re-examination of historical data before

and during the onset of Covid. We therefore reflect this incomplete sensitivity as the necessary cost of reusing existing data.

Based on our findings that longer duration become less sensitive, and since we know the Internet changed in response to Covid-19 work-from-home, we identify change-sensitive blocks based on the 2020m1-egjnw dataset for evaluation across all of 2020 (Table 3).

4.3 How Many Change-Sensitive Blocks?

Our algorithm detects changes in change-sensitive blocks—blocks with a noticeable diurnal swing. Such blocks exist when ISPs assign IP addresses to users during the workday and release them during the night (or vice versa). We next examine two questions: How many change-sensitive blocks are there? Does that number change over time? We also revisit the effect of observation duration on number of change-sensitive blocks.

Decrease over time: The left three columns of Table 3 show three quarters of data from 2019q4 to 2020q2. Each quarter uses a 12-week observation. We see the number of change-sensitive blocks decreases somewhat over this period: from 370k to 327k to 275k. We expect some of the decrease from 2020q1 to 2020q2 may reflect Covid-19 work-from-home as people move from universities and workplaces with more public IP addresses to homes where always-on modems shield their workday status.

Turnover: There is a fairly large rate of turnover in change-sensitive blocks. We see that in comparing the first six months of 2020 (2020h1-w) with each three-month quarter (2020q1-w and 2020q2-w). The number of change sensitive blocks for 2020h1-w is the intersection of the two quarters, and with 169k blocks, it is only 53% or 61% of each quarter. We believe this drop is because when blocks are not consistently diurnal for the entire observation period we classify them as non-diurnal. We see a similar drop to 199k when merging 2019q4 and 2020q1, consistent with duration as the primary factor.

Input targets: A complicating factor in this analysis is that the underlying target list changes over time: in each quarter, the target list is updated to reflect currently responsive blocks, and in 2020q1 the target list was expanded from 4.0M blocks to 5.2M to take advantage of algorithm changes that correctly handle sparse blocks [20] (compare the number of responsive blocks in 2019q4-w and 2020q1-w). This expansion only modestly increases the number of change sensitive blocks (it adds 11,547 or 3.6% to 2020q1-w, and 16,852 or 3.8% to 2020m1-egjnw) because only a few of the newly added blocks are diurnal.

We conclude that the set of change sensitive blocks change over time. It is due to the Covid pandemic, as discussed in §4.2.3.

Measurement Duration: We previously observed that longer periods decrease the number of diurnal blocks when we observed for longer periods. We see that effect again here, comparing 2020m1-w to 2020q1-w to 2020h1-w. As discussed in §4.2.3, to factor out this change we detect change-sensitive blocks based on 2020m1-egjnw and apply that to all of 2020h1-egjnw for our data in §5. (The longer duration in 2020h1-egjnw would reduce the number of change sensitive blocks by 42%, in part because of the very changes in address use we are working to detect.)

4.4 Where are Change-Sensitive Blocks?

Prior analysis of diurnal blocks has shown their frequency varies by country [52]. Countries have different amounts of IPv4 address space, and they also adopt different telecommunications and cultural policies about keeping devices “always-on”. Change sensitive blocks occur when devices are directly attached to the public Internet, so we do not see change-sensitive blocks in ISPs where devices use private address space behind an always-on router on the public IPv4 Internet.

Figure 7 shows all change-sensitive blocks we see in the 2020m1-egjnw dataset. We geolocate all change-sensitive blocks (with Maxmind GeoCities Light) and count blocks in each 2x2 degree latitude/longitude grid, showing the number of blocks as the area of a red circle in each grid cell.

We see that the best coverage is in East Asia (China, Korea, and Japan), with moderate coverage in Brazil and North Africa, and sparse coverage in the United States, Europe, and India. Coverage reflects the intersection of where IPv4 addresses are allocated and where users of those IP addresses turn off devices at night. Relative sparse results from the U.S. and Europe reflect widespread use of always-on home routers. Although these occupy a large number of public IPv4 addresses, their 24x7 operation means they are not diurnal and so do not show when they are actually in use. Users may be turning off devices at home, however, those devices are mostly hidden behind these always-on routers with Network Address Translation. Instead, diurnal blocks in the U.S. and Europe often indicate universities or other organizations where users occupy public IP addresses during the work week.

We believe the heavier coverage in East Asia reflects local ISP policies, with most users using dynamically assigned, public IP addresses.

4.5 Algorithm Validation from a Random Sample

We next validate the correctness of our algorithm by examining a random sample of events and comparing them to ground truth news about Covid-related changes. We validate the correctness of our algorithm by showing an event detected in our change-detection algorithm corresponds to

Dataset:	2019q4-w	2020q1-w	2020q2-w	2020h1-w	2020m1-w	2020h1-egjnw	2020m1-egjnw
duration (weeks)	12			24	4	24	4
completeness (sites)	1					5	
allocated blocks	14,483,456						
not routed	3,388,236	3,361,864	3,333,669	3,333,669	3,361,864	3,333,669	3,361,864
routed blocks	11,095,220	11,121,592	11,149,787	11,149,787	11,121,592	11,149,787	11,121,592
not responsive	7,049,754	5,922,911	5,925,238	6,024,576	5,922,911	6,024,576	5,922,911
responsive	4,045,466	5,198,681	5,224,549	5,125,211	5,198,681	5,125,211	5,198,681
not diurnal	3,631,272	4,799,382	4,849,579	4,877,967	4,796,117	4,889,070	4,647,861
diurnal	414,194	399,299	374,970	247,244	402,564	236,141	550,820
narrow swing	1,375,566	2,170,901	1,700,045	2,269,473	2,977,067	1,825,465	1,656,565
wide swing	2,669,900	3,027,780	3,524,504	2,855,738	2,221,614	3,299,746	3,542,116
not change sensitive	3,675,118	4,880,856	4,948,888	4,956,244	4,888,659	4,937,024	4,751,291
change sensitive	370,348	317,825	275,661	168,967	310,022	188,187	447,390

Table 3: Blocks before and after filtering (in /24s). Change-sensitive is interpreted as /24 blocks that are diurnal and with wide swing. Allocated addresses from IPv4 Address Space Registry [44]; Routing data from Routeviews [15–17].

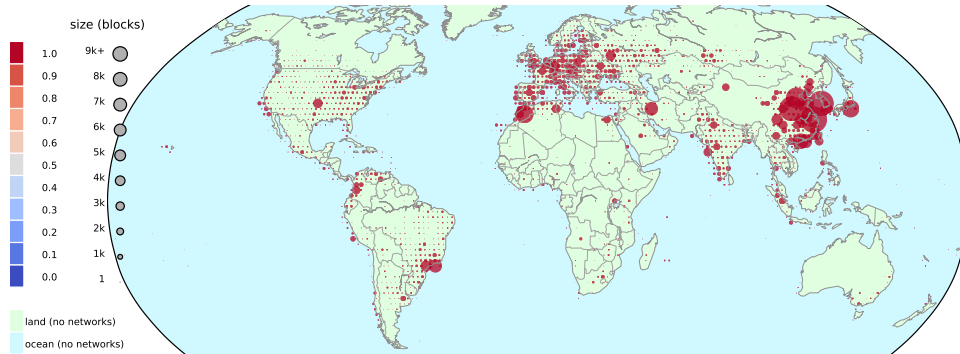


Figure 7: Change-sensitive blocks by geolocation. Circles show the number of blocks in each 2x2 degree geographic grid. Dataset: 2020m1.

a Covid-related switch to work-from-home. We also show sometimes Internet changes correspond to non-Covid-related shutdowns (we examine one event in detail in §5.2). We consider non-Covid shutdowns true positive – if we can verify them. Although some of them might be related to ISP maintenance [54], if they cannot be verified, we consider them as noise (false positive). We claim to detect all types of Internet shutdown, some of which are not Covid-related, not *only* Covid-related shutdowns.

While we can evaluate the correctness of the algorithm (are the changes it sees consistent with news reports), we cannot evaluate *completeness*. Our algorithm applies to only change-sensitive blocks, and we know that most ISP customers in countries such as the U.S. and in western Europe use always-on modems, decreasing our ability to see work-from-home. In addition, not *every* change-sensitive block is necessarily affected by work-from-home, since some blocks may show behavior at essential businesses. We therefore do

not claim completeness. Nevertheless, we believe the ability to see Covid-related events around the globe correctly is useful even if we cannot guarantee full coverage.

Validation begins by selecting 50 random blocks with CUSUM-detected from all blocks that see changes. While our case studies in §5 examine specific locations where we know events happen, our random selection here provides an unbiased evaluation of the algorithm.

We then geolocate each block (all are geolocatable in our sample). We see the block are global, in 19 different countries or regions. Locations follow Figure 7, with 22 in China, 5 in Russia, 4 in Malaysia, 3 in India, and 2 each in Brazil and Hong Kong (SAR)

To avoid transient behavior at the beginning and end of our data we discard changes seen in the first week of January. We use global Covid-19 lockdown dates from multiple media sources [1–5, 7–10, 12, 14, 29, 31, 48, 59, 60]. We find 44 blocks with lockdown dates in this quarter, discarding 6 blocks for

Russia and Singapore that have lockdown not visible in this quarter (March 30 and April 7).

Our algorithm detects changes in 14 blocks (31%) of these 44 blocks. In 13 of the 44 (29%) we confirm a CUSUM-detected change in the raw data within 4 days of the reported Covid-19 lockdown date. (We discard one case since visual inspection suggests a network outage and not a Covid-related change, although it is at the right time.) These cases represent true positives. We confirm that all blocks see a change in the raw data, although for one block the algorithm detect a change but visually it is not obvious.

We also find 6 blocks that are missed by our algorithm but could have been found; these represent false negatives that we could possibly find by tuning our detection parameters. Visually examination of raw data from all blocks shows 5 additional blocks with visual changes missed by the algorithm. The sixth is a block in China that just missed our 4-day window.

For 9 blocks our algorithm detects changes at dates other than the Covid-related days, and we confirm those changes with manual examination of the raw data. Networks change for many reasons, including ISP maintenance [54], and our algorithm detects these types of reconfiguration. These blocks show correct detection of network changes, and they are legitimate events. However, they are “noise” relative to Covid work-from-home, and need to be filtered out as described below.

Our algorithm detects no changes in the rest of blocks that do not show changes in the raw data.

Overall, we show that *our algorithm is successful at finding Covid-related changes*—we confirm this result in §5 where we show events that it discovers which we can confirm in news reports. However, Covid-caused changes do not always affect *all* blocks in a region. While the signal we watch for in blocks often changes in response to work-from-home, work-from-home does not change *all* blocks, so the sensitivity of this signal is moderate (29%, 13 of 44 blocks). Finally, other events, such as ISP maintenance, also trigger our algorithm. Fortunately, while such false positives occur, ISP maintenance is not usually correlated geographically. In §5.2 we show an example of a large non-Covid event, that is a legitimate Internet lockdown, but due to other political causes.

The implication of this validation is that our algorithm functions best as a *detector for unknown Covid-19 work from home* and for *verification of the implementation of work-from-home*. In both cases, they should be accompanied by manual validation. Our results suggest that our algorithm is too imprecise to be suitable for unsupervised classification, and correlation across multiple networks and localized geography should be used to filter out the scattered false positives

that will occur due to ISP maintenance. We use these principles in the next section to investigate events our algorithm discovers.

4.6 Validating Discoverability

§4.5 suggests our algorithm is imperfect at detection, but useful for discovery. We show discovery with selected examples in §5, but those examples are selected because they stand out on our data. To *validate* the ability of the algorithm to assist discovery, we next select two random locations from all grid cells with change-sensitive blocks and look at the raw data.

United Arab Emirates: We randomly select grid cell 24N,54E in United Arab Emirates, then select 25 of the 188 change-sensitive blocks in this region. This country started a sterilization campaign on 2020-03-22 and then had a night curfew on 2020-03-26 [13].

As before, we validate all blocks by comparing detection dates to news reports and checking against raw data. We find 11 blocks of the 25 showing changes near the lockdown date, within 4-day duration (true positives), and 4 blocks with changes at other dates (false positives). The peak of Covid-related events is strong (35% of blocks), allowing false positives to be discarded. We see 6 blocks showing no change (true negatives if we accept that our algorithm does not apply to all blocks). Finally, 4 blocks show changes in the raw data, but fail to be detected (false negative).

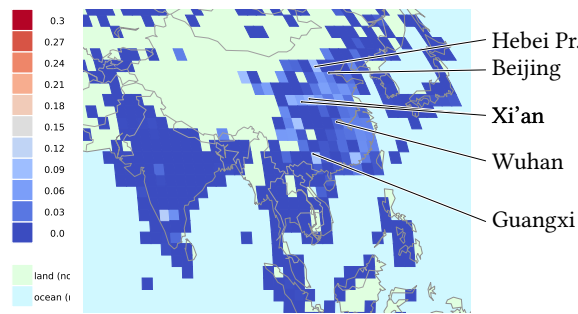
Slovenia: The second grid cell we randomly select is 46N,14E in Slovenia. We then examine 25 of the 735 change-sensitive blocks in the region. Slovenia first closed all educational institutions on 2020-03-16 and then placed more suspension later [11].

Of the 25 randomly selected blocks, we find 7 blocks of the 25 showing changes near 2020-03-16 (true positives), with 6 changes at other dates (false positives). We see 10 blocks showing no change (true negative) and 2 blocks showing changes in the raw data, but failing to be detected (false negative). As with United Arab Emirates, the peak of changes associated Covid work-from-home is obvious and supports filtering false positives (Appendix A).

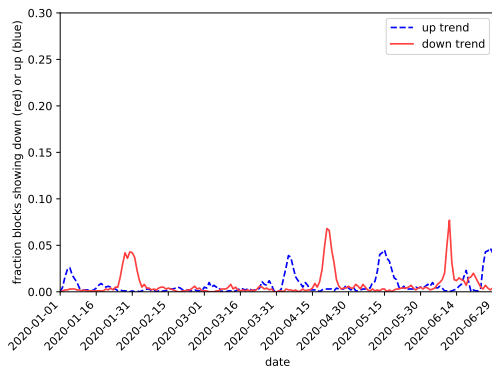
Discussion: These two examples confirm our prior results: correlated changes that relate to Covid-related lockdown show up as peaks in our algorithm. We see a number of false positives, but they could be filtered because they occur at different times. Also, a number of blocks show no change.

5 RESULTS: REAL WORLD EVENTS

Finally, we show that our analysis can confirm and discover changes in Internet usage that correspond to Covid-19 work-from-home events. We look at real-world events in three different countries. We begin with China in January, where our results confirm changes in Internet usage that correspond



(a) A heatmap of 2x2 degree cells on 2020-01-27.



(b) Block changes over time for the Wuhan, China grid cell.

Figure 8: Fraction of block changes in China.

to the earliest Covid-19 outbreaks and lockdown. We then look at India and the Philippines later in the quarter—we discover these through our algorithms, then confirm they correspond to Covid-19 lockdown.

For each case, we show our world-map with a 2x2 degree geographic grid and the number of changed blocks as a heatmap. The color indicates the proportion of detected change points in the 2x2 degree geographic grid on this day, from dark blue to white to red. We also show the line chart of the number of changes over time for a given region.

5.1 2020-01-27 in China

According to media reports, Wuhan went on lockdown on 2020-01-23 [12]. We show in Figure 8 change-sensitive blocks in China with the whole quarter for Wuhan (30E,114N) and all of China on January 27.

Figure 8a shows our 2x2 degree grid of changes on 2020-01-27. The large number of light blue and pink red cells shows that there is a large network change around this day.

Figure 8b counts up the number of downward detection we observe across all blocks for each day over six months. It shows that there is a peak around 2020-01-27, indicating the

network usage drops significantly on that date—about 5% of change-sensitive blocks show reduced usage.

While Wuhan peaks with 5% of the blocks out on January 27, several other grids in China showed larger changes on that day. The pink grid cell at 24N,108E (16%) is in Guangxi Province. It sees a larger change in part because it has fewer change-sensitive blocks: 177 vs. 2,923 in Wuhan’s grid cell.

We also see other changes around China at the same time. Two white squares correspond to Xi’an (8% change in 2,290 change-sensitive-blocks at grid 34N,108E, just northwest of Wuhan) and suburbs (9% change in a 140 change-sensitive blocks at grid 34N,110E). We also see changes in Beijing, the light-blue grid cell at (6% changes with 13,399 change-sensitive blocks at 38N,116E), and more dramatic changes to its northwest in Hebei province (11% with 57 change-sensitive blocks at 40N,114E).

These examples show two things: first, our approach can confirm changes in Internet use in Wuhan, and we see similar changes at the same time in other large cities. Second, one must consider block density to put changes in context. We see the largest fraction of changes occur in sparser areas (for example, Hebei’s 57 change-sensitive blocks), but we also see large quantitative changes in dense areas (the 6% change in Beijing corresponds to 13,399 blocks).

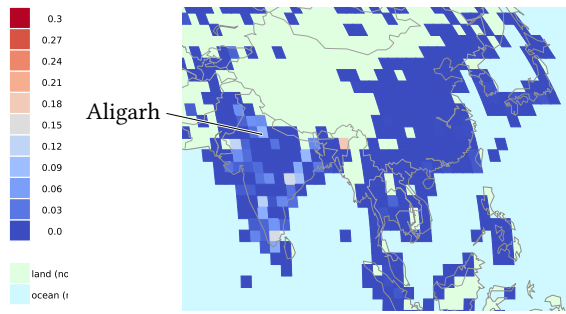
5.2 2020-02-28 and 2020-03-23 in India

We next examine India on 2020-02-28 and then three weeks later in March. We examined this date in India because it showed a large fraction changed-blocks as we examined our data, so it demonstrates an event about which we were unaware that was discovered by our algorithms. We will show that what it finds both Covid-related work-from-home and non-Covid Internet changes, both of which correspond to news reports.

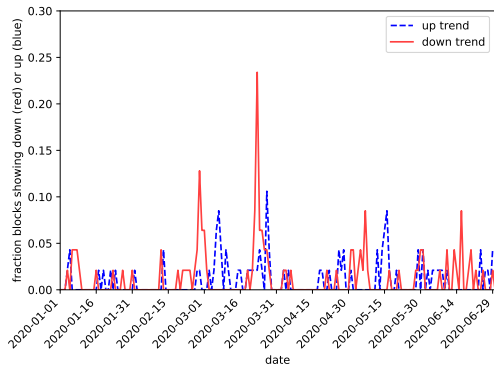
Beginning with 2020-02-28, Figure 9a shows a large network change, and Figure 9b shows the 2x2 degree grid cell at Aligarh in Uttar Pradesh (28N,78E, with 67 change-sensitive blocks) over time. This cell a 13% of change-sensitive blocks reduce usage.

News reports describe violence that broke out in Aligarh on 2020-02-23 as protests to changes to a citizenship law. The state shut down the mobile internet and suspended it for six hours [24], and the report shows this shutdown was extended until February 28 in Aligarh [6]. Although the event is not Covid-19-related, it is a change in Internet usage in response to a different type of social change. We therefore consider it a true positive event in detecting Internet change.

We return to India on March 23, three weeks later. (Data is in Appendix B due to space limits.) We see New Delhi (grid cell 28N,76E, with 1,899 change-sensitive blocks) over time, with a large drop (12%) in usage on the 23rd. We also see a



(a) A heatmap of 2x2 degree cells on 2020-02-28.



(b) Block changes over time for the Aligarh, Uttar Pradesh, India grid cell.

Figure 9: Non-Covid block changes in India.

larger drop (17%) in Rajkot in western India (the pink grid cell at 22N,70E with 98 change-sensitive blocks).

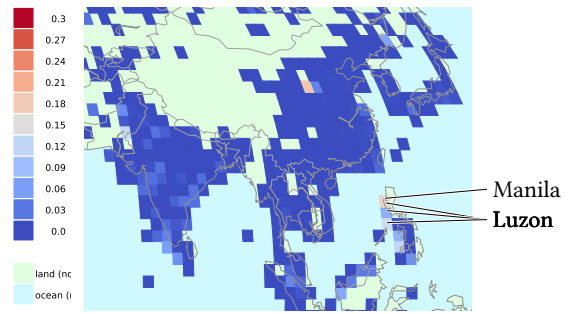
These events correspond to the first Covid-related curfew in India, the Janata curfew on 2020-03-22 [4] and a nationwide lockdown on 2020-03-24.

These examples show that the largest relative changes occur in less dense blocks (like Rajkot), but we also see large absolute numbers of changes in more populated blocks (like New Delhi). They also show that our algorithms are sensitive to work-from-home changes in India, in part because the country has a large population of Internet users who are directly connect to the public Internet without using an always-on router.

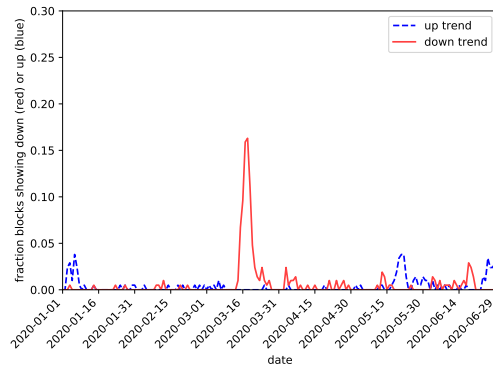
5.3 2020-03-19 in the Philippines

Finally, we turn to the Philippines in March. As with India, we discovered this event from our data, and it shows what we believe was Covid-related work-from-home in the Philippines.

Figure 10a shows our 2x2 degree grid of changes on 2020-03-19. We see a light red cell in Manila (at 14N,120E) with a sudden large drop in network usage—about 24% of the 344



(a) A heatmap of 2x2 degree cells on 2020-03-19.



(b) Block changes over time for the Manila, Luzon, the Philippines grid cell.

Figure 10: Block changes in the Philippines.

change-sensitive blocks, as confirmed in Figure 10b. This date is shortly after Manila’s lockdown on 2020-03-15 [61]. This lockdown was extended to the island of Luzon on 2020-03-17.

We consider this event to be a true positive of Covid-19-induced work-from home, detected by our algorithm. It confirms our ability to discover changes that correspond to Covid-related events.

6 CONCLUSIONS

This paper has explored how we can interpret responsiveness of individual IPv4 addresses to detect Internet changes, such as those that occur as a result of work-from-home. Our algorithms allow us to reconstruct diurnal trends from existing data collected to detect Internet outages, then detect changes in daily Internet use.

We validated these algorithms through comparison of random samples of detected events (§4.5), then applied them to 447k /24 blocks over 6 months of 2020 to examine known events in China and to discover events in India and the Philippines (§5).

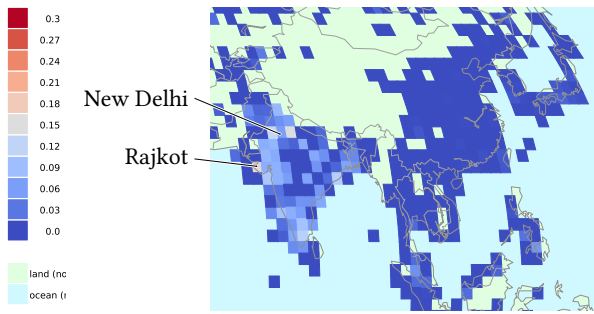
ACKNOWLEDGMENTS

This work is partially supported by the project “Measuring the Internet during Novel Coronavirus to Evaluate Quarantine (RAPID-MINCEQ)” (NSF 2028279), and John Heidemann’s work is partially supported by the project “CNS Core: Small: Event Identification and Evaluation of Internet Outages (EIEIO)” (CNS-2007106). We thank Guillermo Baltra for prototyping analysis of Trinocular addresses across a block, his involvement in early versions of this work, and his comments on the work. We thank Yuri Pradkin for collecting the Trinocular data that we use in our analysis.

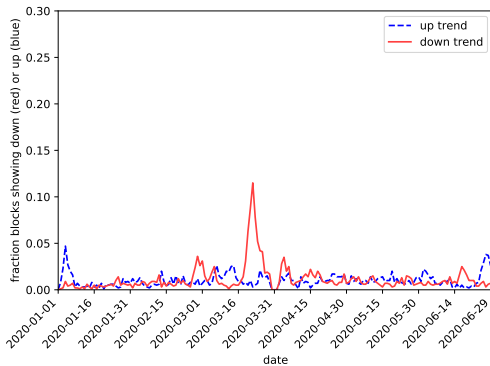
REFERENCES

- [1] 2020. Belgium enters lockdown over coronavirus crisis – in pictures. <https://www.theguardian.com/world/gallery/2020/mar/18/belgium-enters-lockdown-over-coronavirus-crisis-in-pictures>.
- [2] 2020. Boris Johnson orders UK lockdown to be enforced by police. <https://www.theguardian.com/world/2020/mar/23/boris-johnson-orders-uk-lockdown-to-be-enforced-by-police>.
- [3] 2020. Coronavirus: What are the lockdown measures across Europe? <https://www.dw.com/en/coronavirus-what-are-the-lockdown-measures-across-europe/a-529051371>.
- [4] 2020. COVID-19: Lockdown across India, in line with WHO guidance. <https://news.un.org/en/story/2020/03/1060132>.
- [5] 2020. Covid-19: Movement Control Order imposed with only essential sectors operating. <https://www.nst.com.my/news/nation/2020/03/575177/covid-19-movement-control-order-imposed-only-essential-sectors-operating>.
- [6] 2020. Internet suspension extended for another 24-hour in Aligarh. <https://www.aninews.in/news/national/general-news/internet-suspension-extended-for-another-24-hour-in-aligarh20200226230151/>.
- [7] 2020. Iran: Nationwide lockdown implemented as over 11,300 COVID-19 cases confirmed March 13 /update 12. <https://www.garda.com/crisis24/news-alerts/322811/iran-nationwide-lockdown-implemented-as-over-11300-covid-19-cases-confirmed-march-13-update-12>.
- [8] 2020. Lockdown was started in Freiburg, Baden-Württemberg and Bavaria on 20 March 2020. Two days later, it was expanded to the whole of Germany. <https://pbs.twimg.com/media/ETjc738WoAI5zAt?format=jpg&name=large>.
- [9] 2020. Moscow goes into lockdown, urges other regions to take steps to slow coronavirus. <https://www.france24.com/en/20200330-moscow-goes-into-lockdown-urges-other-russian-regions-to-take-measures-to-curb-coronavirus-spread>.
- [10] 2020. Pedro Sánchez anuncia el estado de alarma para frenar el coronavirus 24 horas antes de aprobarlo. <https://www.elmundo.es/espana/2020/03/13/5e6b844e21efa0dd258b45a5.html>.
- [11] 2020. Staršem 50 odstotkov plače, varstvo za nujno potrebne poklice. <https://www.rtvsl.si/zdravje/novi-koronavirus/starsem-50-odstotkov-place-varstvo-za-nujno-potrebne-poklice/516908>.
- [12] 2020. Timeline: China’s COVID-19 outbreak and lockdown of Wuhan. <https://abcnews.go.com/Health/wireStory/timeline-chinas-covid-19-outbreak-lockdown-wuhan-75421357>.
- [13] 2020. UAE: Three-day lockdown scheduled March 26-29 /update 16. <https://www.garda.com/crisis24/news-alerts/326651/uae-three-day-lockdown-scheduled-march-26-29-update-16>.
- [14] 2020. Venezuela’s to implement nationwide quarantine as coronavirus cases rise to 33. <https://www.reuters.com/article/us-health-coronavirus-venezuela-maduro/venezuelas-to-implement-nationwide-quarantine-as-coronavirus-cases-rise-to-33-idUSKBN214015>.
- [15] APNIC. 2020. Routing Table Report - Japan view. <https://mailman.apnic.net/mailling-lists/bgp-stats/archive/2019/10/msg00001.html>
- [16] APNIC. 2020. Routing Table Report - Japan view. <https://mailman.apnic.net/mailling-lists/bgp-stats/archive/2020/01/msg00001.html>
- [17] APNIC. 2020. Routing Table Report - Japan view. <https://mailman.apnic.net/mailling-lists/bgp-stats/archive/2020/04/msg00001.html>
- [18] Roy Arends. 2020. Analysis of the Effects of COVID-19-Related Lockdowns on IMRS Traffic. (Apr. 15 2020). <https://www.icann.org/en/system/files/files/octo-008-15apr20-en.pdf>
- [19] ChaeWon Baek, Peter B McCrory, Todd Messer, and Preston Mui. 2020. Unemployment effects of stay-at-home orders: Evidence from high frequency claims data. *Review of Economics and Statistics* (2020), 1–72.
- [20] Guillermo Baltra and John Heidemann. 2020. Improving Coverage of Internet Outage Detection in Sparse Blocks. In *Proceedings of the Passive and Active Measurement Conference*. Springer, Eugene, Oregon, USA.
- [21] Timm Böttger, Ghida Ibrahim, and Ben Vallis. 2020. How the Internet reacted to Covid-19—A perspective from Facebook’s Edge Network. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Pittsburgh, PA, USA, 34–41. <https://doi.org/10.1145/3419394.3423621>
- [22] Timm Böttger, Ghida Ibrahim, and Ben Vallis. 2020. How the Internet reacted to Covid-19: A perspective from Facebook’s Edge Network. In *Proceedings of the ACM Internet Measurement Conference*. 34–41.
- [23] Jon Brodtkin. 2020. Netflix, YouTube cut video quality in Europe after pressure from EU official. *Ars Technica* <https://arstechnica.com/tech-policy/2020/03/netflix-and-youtube-cut-streaming-quality-in-europe-to-handle-pandemic/>
- [24] Zee Media Bureau. 2020. Anti-CAA protest: Internet services suspended in Aligarh after protesters-police clash. <https://zeenews.india.com/india/anti-kaa-protest-internet-services-suspended-in-aligarh-after-protesters-police-clash-2265646.html>.
- [25] Xue Cai and John Heidemann. 2010. Understanding Block-level Address Usage in the Visible Internet. In *Proceedings of the ACM SIGCOMM Conference*. ACM, New Delhi, India, 99–110. <https://doi.org/10.1145/1851182.1851196>
- [26] CAIDA. 2007. Archipelago (Ark) Measurement Infrastructure. website <https://www.caida.org/projects/ark/>
- [27] CAIDA. 2012. Network Telescope. https://www.caida.org/projects/network_telescope/
- [28] Massimo Candela, Valerio Luconi, and Alessio Vecchio. 2020. Impact of the COVID-19 pandemic on the Internet latency: A large-scale study. *Computer Networks* 182 (2020), 107495.
- [29] Tony Cheung, Natalie Wong, and Chan Ho-him. 2020. Coronavirus: ‘little, if any, possibility’ Hong Kong schools resume fully on April 20, Lam says. <https://www.scmp.com/news/hong-kong/education/article/3075521/coronavirus-little-if-any-possibility-hong-kong-schools>.
- [30] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. 1990. STL: A seasonal-trend decomposition. *Journal of Official Statistics* 6, 1 (1990), 3–73.
- [31] Anthony Cuthbertson. 2020. Coronavirus: France imposes 15-day lockdown and mobilises 100,000 police to enforce restrictions. <https://www.independent.co.uk/news/world/europe/coronavirus-france-lockdown-cases-update-covid-19-macron-a9405136.html>.
- [32] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. 2011. Analysis of country-wide internet outages caused by censorship. In *Proceedings of*

- the 2011 ACM SIGCOMM conference on Internet measurement conference*. 1–18.
- [33] M. Duarte. 2020. detecta: A Python module to detect events in data. <https://github.com/demotu/detecta>.
- [34] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *Proceedings of the ACM Conference on Computer and Communications Security*. ACM, Denver, CO, USA, 542–553. <https://doi.org/10.1145/2810103.2813703>
- [35] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *22nd USENIX Security Symposium (USENIX Security 13)*. USENIX Association, Washington, D.C., 605–620. <https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/durumeric>
- [36] Maria Effenberger, Andreas Kronbichler, Jae Il Shin, Gert Mayer, Herbert Tilg, and Paul Perco. 2020. Association of the COVID-19 pandemic with internet search volumes: a Google TrendsTM analysis. *International Journal of Infectious Diseases* 95 (2020), 192–197.
- [37] Reid J. Epstein and Kay Nolan. 2020. A Few Thousand Protest Stay-at-Home Order at Wisconsin State Capitol. *New York Times* (Apr. 24 2020). <https://www.nytimes.com/2020/04/24/us/politics/coronavirus-protests-madison-wisconsin.html>
- [38] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, and Daniel Wagner. 2020. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Pittsburgh, PA, USA, 1–18. <https://doi.org/10.1145/3419394.3423658>
- [39] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, et al. 2021. A year in lockdown: how the waves of COVID-19 impact internet traffic. *Commun. ACM* 64, 7 (2021), 101–108.
- [40] Robert Graham, Paul McMillan, and Dan Tentler. 2014. Mass Scanning the Internet. Presentation at Defcon 22. <https://defcon.org/images/defcon-22/dc-22-presentations/Graham-McMillan-Tentler/DEFCON-22-Graham-McMillan-Tentler-Massscanning-the-Internet.pdf>
- [41] Andreas Guillot, Romain Fontugne, Philipp Winter, Pascal Merindol, Alistair King, Alberto Dainotti, and Cristel Pelsser. 2019. Chocolate: Outage detection for internet background radiation. In *2019 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 1–8.
- [42] Fredrik Gustafsson and Fredrik Gustafsson. 2000. *Adaptive Filtering and Change Detection*. Vol. 1. John Wiley & Sons, Inc.
- [43] John Heidemann, Yuri Pradkin, Ramesh Govindan, Christos Papadopoulos, Genevieve Bartlett, and Joseph Bannister. 2008. Census and Survey of the Visible Internet. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Vouliagmeni, Greece, 169–182. <https://doi.org/10.1145/1452520.1452542>
- [44] IANA. 2021. IPv4 Address Space Registry. <https://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml>
- [45] Cecilia Kang, Davey Alba, and Adam Satariano. 2020. Surging Traffic Is Slowing Down Our Internet. *New York Times* (Mar. 26 2020). <https://www.nytimes.com/2020/03/26/business/coronavirus-internet-traffic-speed.html>
- [46] Ethan Katz-Bassett, Harsha V Madhyastha, John P John, Arvind Krishnamurthy, David Wetherall, and Thomas E Anderson. 2008. Studying Black Holes in the Internet with Hubble. In *NSDI*, Vol. 8. 247–262.
- [47] Andra Lutu, Diego Perino, Marcelo Bagnulo, Enrique Frias-Martinez, and Javad Khangoslar. 2020. A Characterization of the COVID-19 Pandemic Impact on a Mobile Network Operator Traffic. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Pittsburgh, PA, USA, 19–33. <https://doi.org/10.1145/3419394.3423655>
- [48] Por Valéria Martins. 2020. Casos de pacientes com coronavírus sobe para 197 em SC e governo prorroga quarentena. <https://g1.globo.com/sc/santa-catarina/noticia/2020/03/29/casos-de-pacientes-com-coronavirus-sobe-para-197-em-sc-e-governo-prorroga-quarentena.ghtml>.
- [49] Giovane C. M. Moura, Carlos Gañán a Qasim Lone, Payam Poursaied, Hadi Asghari, and Michel van Eeten. 2015. How Dynamic is the ISPs Address Space? Towards Internet-Wide DHCP Churn Estimation. In *Proceedings of the IFIP Networking*. IFIP. <https://doi.org/10.1109/IFIPNetworking.2015.7145335>
- [50] Ramakrishna Padmanabhan, Amogh Dhamdhere, Emile Aben, kc claffy, and Neil Spring. 2016. Reasons Dynamic Addresses Change. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Santa Monica, CA, USA, 183–198. <https://doi.org/10.1145/2987443.2987461>
- [51] Lin Quan, John Heidemann, and Yuri Pradkin. 2013. Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *Proceedings of the ACM SIGCOMM Conference*. ACM, Hong Kong, China, 255–266. <https://doi.org/10.1145/2486001.2486017>
- [52] Lin Quan, John Heidemann, and Yuri Pradkin. 2014. When the Internet Sleeps: Correlating Diurnal Networks With External Factors. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Vancouver, BC, Canada, 87–100. <https://doi.org/10.1145/2663716.2663721>
- [53] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. 1996. *Address Allocation for Private Internets*. RFC 1918. Internet Request For Comments. <ftp://ftp.rfc-editor.org/in-notes/rfc1918.txt>
- [54] Philipp Richter, Ramakrishna Padmanabhan, Neil Spring, Arthur Berger, and David Clark. 2018. Advancing the Art of Internet Edge Outage Detection. In *Proceedings of the ACM Internet Measurement Conference* (johnh: pafile). ACM, Boston, Massachusetts, USA, 350–363. <https://doi.org/10.1145/3278532.3278563>
- [55] Philipp Richter, Ramakrishna Padmanabhan, Neil Spring, Arthur Berger, and David Clark. 2018. Advancing the art of internet edge outage detection. In *Proceedings of the Internet Measurement Conference 2018*. 350–363.
- [56] Aaron Schulman and Neil Spring. 2011. Pingin’ in the Rain. In *Proceedings of the ACM Internet Measurement Conference* (johnh: pafile). ACM, Berlin, Germany, 19–25. <https://doi.org/10.1145/2068816.2068819>
- [57] Skipper Seabold and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference*, Vol. 57. Austin, TX, 61.
- [58] Anant Shah, Romain Fontugne, Emile Aben, Cristel Pelsser, and Randy Bush. 2017. Disco: Fast, good, and cheap outage detection. In *2017 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 1–9.
- [59] Erbil Sulaymaniyah. 2020. raq’s KRG eases coronavirus lockdown after two months. <https://www.aa.com.tr/en/middle-east/iraqs-krgeases-coronavirus-lockdown-after-two-months/1838231>.
- [60] Eric Sylvers and Giovanni Legorano. 2020. As Virus Spreads, Italy Locks Down Country. <https://www.wsj.com/articles/italy-bolsters-quarantine-checks-after-initial-lockdown-confusion-11583756737>.
- [61] Rambo Talabong. 2020. Metro Manila to be placed on lockdown due to coronavirus outbreak. <https://www.rappler.com/nation/metro-manila-placed-on-lockdown-coronavirus-outbreak>.
- [62] Gerry Wan, Liz Izhikevich, David Adrian, Katsunari Yoshioka, Ralph Holz, Christian Rossow, and Zakir Durumeric. 2020. On the Origin of Scanning: The Impact of Location on Internet-Wide Scans. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Pittsburgh, PA, USA, 662–679. <https://doi.org/10.1145/3419394.3424214>
- [63] Yinglian Xie, Fang Yu, Kannan Achan, Eliot Gillum, Moises Goldszmidt, and Ted Wobber. 2007. How Dynamic are IP Addresses?. In *Proceedings of the ACM SIGCOMM Conference*. ACM, Kyoto, Japan, 301–312. <https://doi.org/10.1145/1282380.1282415>



(a) A heatmap of 2x2 degree cells on 2020-03-23.



(b) Block changes over time for the New Delhi, India grid cell.

Figure 13: Fraction of block changes in India with Covid-related events on 2020-03-23.

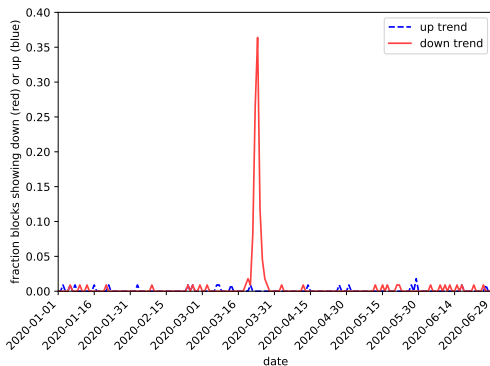


Figure 11: The fraction of blocks that decrease (light red) or increase (dark blue) usage at 24N,54E in United Arab Emirates for 2020q1 and q2.

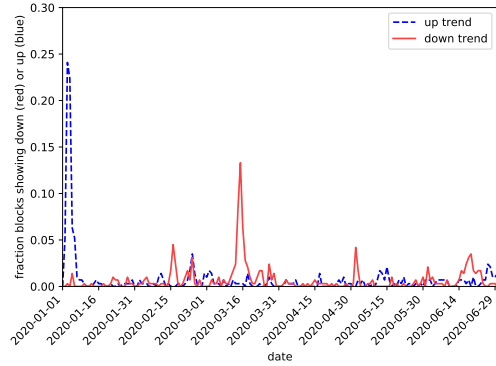


Figure 12: The fraction of blocks that decrease (light red) or increase (dark blue) usage at 46N,14E in Slovenia for 2020q1 and q2.

A VALIDATING DISCOVERABILITY: ADDITIONAL DETAILS

Here we provide graphs documenting our random samples used in validating discoverability in §4.6.

Figure 11 shows the large change near the curfew date, with a 36% of blocks showing a decrease in use on 2020-03-24. This example from UAE confirms our ability to discover Covid-related changes.

Again, the Covid-related changes stand out from noise in Figure 12, with a 14% of blocks showing a decrease in use. This example from Slovenia confirms our approach.

B 2020-03-23 IN INDIA

Figure 13 provides the graphs supporting the Covid-related changes we saw in India on 2020-03-23. We discuss these events in §5.2