ABSTRACT
As the Internet matures, policy questions loom larger in its operation. When should an ISP, city, or government invest in infrastructure? How do their policies affect use? In this work, we develop a new approach to evaluate how policies, economic conditions and technology correlates with Internet use around the world. First, we develop an adaptive and accurate approach to estimate block availability, the fraction of active IP addresses in each /24 block over short timescales (every 11 minutes). Our estimator provides a new lens to interpret data taken from existing long-term outage measurements, thus requiring no additional traffic. (If new collection was required, it would be lightweight, since on average, outage detection requires less than 20 probes per hour per /24 block; less than 1% of background radiation.) Second, we show that spectral analysis of this measure can identify diurnal usage: blocks where addresses are regularly used during part of the day and idle in other times. Finally, we analyze data for the entire responsive Internet (3.7M /24 blocks) over 35 days. These global observations show when and where the Internet sleeps—networks are mostly always-on in the US and Western Europe, and diurnal in much of Asia, South America, and Eastern Europe. ANOVA (Analysis of Variance) testing shows that diurnal networks correlate negatively with country GDP and electrical consumption, quantifying that national policies and economics relate to networks.

Categories and Subject Descriptors
C.2.3 [Computer-Communication Networks]: Network Operations—Network Monitoring; C.2.5 [Computer-Communication Networks]: Local and Wide-Area Networks—Internet; C.4 [Performance of Systems]: Reliability, availability, and serviceability

General Terms
Measurement

Keywords
Internet; Internet reliability; diurnal network usage

1. INTRODUCTION
The Internet is a large and complex combination of technology, policy, and human behavior. Separating these effects is difficult, yet businesses, governments, and individuals must make decisions every day, even with incomplete and contradictory information.

One metric of Internet maturity is network availability and the shift from on-demand on always-on network availability. Increasing use of broadband networks brings always-on networks. In fact, always-on is part of the U.S. Federal Communication Commission’s definition of broadband [9]. Always-on networks have practical differences on how networks are used—the ease of not having to turn on a computer and connect is one enabler of the shift of home telephone and entertainment to IP. A similar shift is seen in the deployment of high-speed mobile phones, where always-on 3G and 4G networks are fast enough to take on data-intensive tasks like social media and streaming video. For many, an always-on phones is their primary interface to social media.

Government policy makers and ISPs evaluate how the choices they make affect network usage. Until recently, tools to study the results of policy choices have been limited, with telephone-based surveys reaching a few thousand of randomly chosen people (for example, [40]), and in-home surveys reaching a few hundred, statically-chosen volunteers. While important, these studies consider a very sparse sample of individuals. A broader perspective would better quantify how policies and technologies affect Internet usage, and it is necessary to compare global differences and understand natural experiments that arise when different countries adopt different policies. For example, we know that some ISP’s use network blocks on a diurnal basis because of economic constraints (saving power bills), or because of different access technologies (such as dial-up use with mandatory dynamic addressing [22]); in some countries these policies dominate use. To move beyond these anecdotes we would like to systematically explore what factors contribute different network usage.

To reach this goal, the first contribution of our work is to develop a new method to estimate block availability, the fraction of addresses in each /24 block that are in use and will reply to an ICMP probe. While the data collection is not new work, our analysis (§ 2) is completely new. We
provide a new analytic tool to interpret existing long-term measurement studies of network outages [31]. We take this approach to leverage existing data, and because these outage measurements have already been shown to impose minimal cost (in probing traffic) on the subject networks. On average, outage detection requires less than 20 probes per hour, per /24, less than 1% of typically Internet background radiation observed in 2010 [39]. Even if our analysis required new data collection, its impact would be minimal.

Second, we show how this measurement can be interpreted to detect diurnal blocks. We use spectral analysis to identify blocks whose usage shows large daily changes. We also show that analysis of the phase of these changes relates to the longitude of blocks. We develop these two methods (§ 2) and validate them with controlled experiments (§ 3).

Our final contribution is to use these methods to study both diurnal behavior of 3.7M responsive blocks (each a /24 IPv4 prefix that replies to active probes) over more than one month (§ 4). While it is well known that network usage varies, our analysis provides the first long-term, wide-scale evaluation of diurnal use of the global Internet. Using ANOVA testing, we correlate these direct observations with economic and technical factors to examine how external factors and national constraints relate to Internet use, showing correlations by country and link technologies, as well as by economic factors (GDP and electricity consumption) (§ 5). Although we do not directly study policies of individual states, our methodology and data will enable future policy studies.

All data from this paper is currently available at no cost to other researchers [37].

2. METHODOLOGY

Our analysis begins with low-rate estimation of block-level availability (§ 2.1), allowing detection of diurnal blocks (§ 2.2). We then relate these measures to geolocation, organization, and link technologies (§ 2.3).

2.1 Dynamic Tracking of Block Availability

Our first new method estimates block availability [10], the probability that addresses in the block will respond to ICMP probes. We study the network in /24 blocks, since prior work has shown they are often homogeneous in use [3], and that larger aggregates are too coarse [31]. Although we track /24 blocks, we can detect diurnal activity that affects tens of addresses in a full block (§ 3.2.4). (As with other active probing, we cannot detect activity behind NATs or firewalls.)

While prior work defined availability $A$ over the whole block, here we estimate three variants: short- and long-term estimates of the availability of the ever-active IP addresses ($A_o$ and $A_l$), and an operational ($A_s$) value that is consistently designed to avoid overestimating $A$.

2.1.1 Problem Statement

Our goal is to derive an estimate $\hat{A}$ from analysis of probe data already taken as part of outage detection in Trinocular [31]. Trinocular already minimizes probes, with each observer sending on average to less than 1% of background radiation as of 2010 [39]. We wish to estimate $A$ with no additional probes beyond what outage detection already requires.

We define the ground truth $A$ as the fraction of addresses that respond when all our probed. While easy to define, a full measurement of ground truth is too expensive for operation.

There are several challenges that make $A$ estimation difficult to add to outage detection. First, sampling in outage detection is biased. The goal of outage detection is to make the minimum number of probes to reduce stress on the target, so a few or even one positive response is usually sufficient to terminate probing. Probes are thus biased in favor of positive responses.

Second, outage detection itself depends on $\hat{A}$; we require an operational $\hat{A}_o$ estimate that is be too high. When $\hat{A}_o > A$, a few negative probes are a stronger indication that the block is down and Trinocular will produce false outages.

Two additional challenges are that observations are quantized, and that our initial $A$ estimate may be quite inaccurate. Because we take from 1 to 15 observations per round, precision of a new $A$ value is at most $\pm 0.07$ and could be as coarse as 0.5 or 1. Short-term estimates therefore show significant jitter. Our initial estimates for $A$ are based on historical data over several years. They may be off significantly if block usage has changed.

2.1.2 Approach

Each of our estimates use exponentially weighted moving average (EWMA) to smooth raw observations.

For each block, in each round of adaptive probing, we observe $p$ positive responses and $t$ total responses.

We first model short-term EWMA of positive and negative responses to quickly adapt to real values.

$\hat{p}_s = \alpha_s p + (1 - \alpha_s)\hat{p}$

$\hat{i}_s = \alpha_s t + (1 - \alpha_s)\hat{i}$

We use a gain of $\alpha_s = 0.1$ to quickly adapt to changes, and calculate the short-term availability $\hat{A}_s$:

$\hat{A}_s = \frac{\hat{p}_s}{\hat{i}_s}$

It is important that we estimate $p$ and $t$ separately. Because $A$ is their ratio, applying EWMA directly to observations of $A$ skews the data, for the same reason that one must use geometric mean to estimate normalized results [17]. (Our experiments in dataset AL219 use a variation on this approach, where short-term availability $\hat{A}'_s$ is tracked directly with EWMA, rather than tracking $\hat{p}_s$ and $\hat{i}_s$ separately. This method consistently over-estimates $\hat{A}$. While we accommo-date for this variation in $\hat{A}_o$ and verify that it does not affect periodicity in $\hat{A}_s$, our current approach provides tighter results.)

An example of $\hat{A}_s$ is shown as a green line in Figures 1 through 3. Compared to the true $A$ taken from complete survey data (the black line), we see that $\hat{A}_s$ is quite noisy (see Figure 1). However, this rapidly changing estimate is useful to understand diurnal block changes (§ 2.2).

We model long-term estimates of availability ($\hat{A}_l$) with the same process using a gain $\alpha_l = 0.01$.

Our goal in a conservative $\hat{A}_o$ is that it should not exceed the true value. To reflect this goal, we evaluate the absolute deviation of each sample from the estimate and then intentionally underestimate the operational value by a half of this amount. We model a deviation term $d_i$, and use it to compute a more conservative $\hat{A}_o$ (note that we limit $\hat{A}_o$ to a minimum of 0.1; very small values will result in excessive
probes and Trinocular policy limits probes per round to do no harm):
\[ \hat{d}_t = a_0 \hat{A}_t - p/t + (1 - a_0) \hat{d} \]
\[ \hat{A}_t = \max(\hat{A}_t - \frac{1}{2} \hat{d}_t, 0.1) \]

Our use of Trinocular works out well, since it provides lightweight probing. Once we accounting for Trinocular’s policy of stopping on first positive response, its policy of walking all responsive addresses in a pseudorandom order is ideal for analysis of diurnal blocks.

2.2 Detecting Diurnal Blocks

Block availability relates to the number of active IPs in the block. Blocks that change diurnally will show regular daily changes to A as computers come and go. To identify diurnal blocks we estimate A as a timeseries of \( \hat{A}_t \) observations taken each 11 minutes (following [10]). We then take the Fast-Fourier Transform of this timeseries to determine periodicity and phase.

Data cleaning: Spectral analysis typically requires an evenly sampled timeseries. We map our probing output to an 11-minute timeseries, but is not perfectly aligned with 11 minute rounds, sometimes we seeing missing or duplicate observations in a round (about 5% of rounds). Like previous work [30], we correct these by extrapolating single missing estimates, and trusting most recent observation when duplicates occur.

We trim our timeseries to start and end near midnight UTC. This approach ties phase to physical time, and it also reduces noise in FFT analysis of diurnal frequencies.

Detecting diurnal blocks: We use a Discrete Fast-Fourier transform to identify periodicity typical of diurnal use. Given a timeseries \( A = a_n, m \in [0, n - 1] \) of n input samples, we compute its \( k \in [0, n - 1] \) frequency-domain components as:
\[ \alpha_k = \sum_{m=0}^{n-1} a_m e^{-2\pi imk/n} \]

The magnitude of \( \alpha_k \) shows the strength kth frequency component (sub-signal) of \( a_n \). With our 11-minute (\( R = 660 \) s) sampling period, \( \alpha_0 \) shows the frequency at \( k/(Rn) \) Hz. To identify diurnal patterns for an experiment lasting \( N_d \) days we look in the \( k = N_d \) component of \( \alpha_k \). To account for noise, we also consider the adjacent \( k = N_d + 1 \).

Diurnal activity in a block appears as strength in \( \alpha_k \) corresponding to 24 hour intervals (and multiples of this frequency). For a strictly diurnal block, the strongest frequency occurs at 1 cycle per day (when \( k = N_d \)), and its strength is at least twice the next strongest, non-harmonic frequency, and greater than all harmonics.

We also consider relaxed diurnal blocks, where the strongest frequency is at 1 cycle per day or the first harmonic, without comparison to the second strongest frequency.

Block phase: In addition to the amplitude of the FFT (the “strength” of a given periodicity), spectral analysis also provides the phase (when the period occurs relative to the start of measurements). Phase is shown in the angle of the complex-valued FFT coefficients.

We examine phase in the 1-cycle-per-day component, and consider it valid only for strong or relaxed diurnal blocks. (For non-diurnal blocks, phase of this component is effectively random.)

Phases range from \([-\pi, \pi]\], with the exact value depending on when a measurement begins. We consider relative phases and compare them to physical location (§5.2); future work may tie phase to time-of-day.

Data appropriateness: FFT over data too short or non-stationary can distort analysis of periodic behavior. Our datasets are sufficiently long (two or more weeks) to capture many diurnal periods.

We verified our data is roughly stationary for our measurement periods by doing a linear fit of A over the observation and confirming slopes are near-zero. We confirmed with all 29,001 blocks in Survey \( S_{51w} \) that about 80.3% of these blocks are stationary, with a slope equivalent to less than 1 address changes per day.

2.3 Other Network Factors: Geolocation, Organizations, and Link Types

We analyze network behavior in /24 blocks, but to put these in context, we relate diurnal behavior to other factors: physical location, logical location (the organization operating the address), and last-mile link technology.

2.3.1 Geolocation

Different countries and regions have different policies and economics for networking. To understand these effects we use IP geolocation to get city-level physical locations. While there have been many approaches to IP geolocation (see for example [13,28]), we use MaxMind’s city-level database [25] as a free and widely used source. Although this source is not the most accurate (claimed accuracy is 40 km), it is sufficient to demonstrate at least country-level correlation.

We map each /24 block to location. Although sometimes blocks span multiple locations, such instances are fairly rare. Our analysis of existing per-IP data [11] shows about 93% of /24 blocks have homogeneous locations, and our MaxMind dataset specifies location with at most /24 precision.

2.3.2 Organizations

To understand how the policies of different organizations affect how they use IP addresses, we map IP addresses to Autonomous System (AS) numbers, and ASes to organizations. (Ideally we would subdivide some ASes, perhaps into commercial and home networks, but we know of no such classification.)

We use Team Cymru’s IP to AS number mapping [36]. We map each /24 to an AS based on its .0 address. Dividing ASes across /24s is rare, with only 0.005% of blocks in Team Cymru’s data showing differing .0 and .128 ASes. Their data provides AS numbers and names for 99.41% of /24 blocks.

We map ASes to organizations using prior work that uses WHOIS and string-based clustering [4]. For a given organization or ISP \( P \) (for example, Time Warner Cable), we first use keyword matching (ex. “Time Warner”) to find relevant clusters, then find all ASes within same cluster(s). Finally, for all ASes within \( P \), we join with IP/AS mapping and find all relevant IP blocks for \( P \).

The above method assumes good accuracy of IP/AS and AS/organization mappings. Future work may consider how accuracy of these mappings affects our correlations, or to compare the behavior of different ASes within the same organization.
2.3.3 Link Technology

We define the link type as the technology connecting the final hop to the destination. Identification of link types for blocks is difficult because such information is not readily available, and because in principle different parts of a /24 block could be connected differently (for example, a router at .1 and the rest of the addresses with dial-up). Previous work has shown one can often infer link types from reverse domain names [4,16,27,33]. We build Thunderping’s idea [33], but cannot reuse their data directly because their classification was done manually and only for the United States; we wish to study millions of international blocks. Other data is also sparse or old, so we identify block-level link types as follows.

First, we look up the reverse domain name of each address in each analyzable block. We then use string matching against pre-defined keywords to non-exclusively identify features (corresponding to link types, see § 5.5) for each address. For example, a reverse name of dhcp-dialup-001.example.com is marked as both DHCP and dial-up. Each block is therefore defined by a vector of 256 addresses, each with zero or more features. We suppress minor features in each block by filtering out features that are less than \( \frac{1}{2} \) of the most frequent feature. Finally, we label the block with all remaining features that have non-zero counts.

We consider 16 keywords (sta, dyn, srv, rtr*, gw*, dhcp, ppp, dsl, dial, cable, ded*, res, client*, sql*, wireless*, wifi*). Of these, we discard the seven marked with asterisks because they are dominant in less than 1000 blocks. These keywords are all based on English; future work should explore non-English networking keywords and explore potential under-representation in countries where English is not the dominant language. We believe this approach works reasonable for ISPs serving homes; it is not designed to classify commercial providers (clouds or CDNs).

For our 3.7M blocks in dataset \( A_{12w} \), and find that 46.3% of these blocks have some feature (4x more than Thunderping), and 11.4% have multiple features.

2.4 Factorial Analysis with ANOVA

To quantify the correlation of policies, places, and technologies on Internet usage, we need to weigh the factors that relate to network outcomes such as diurnal address usage and network reliability. While anecdotal and personal evidence suggests that individuals in some countries turn off computers at night to save power, we wish to measure these effects systematically.

To discover correlations between a range of possible factors and network outcomes we use analysis of variance or ANOVA, a form of statistical hypothesis testing. It tests the probability (p-value) of a factor or factors (separately or in combination) being correlated with an observation. A high p-value indicates that the factor has little correlation with outcome, while a low p-value indicates that correlation of the factor and outcome are unlikely to have been by chance alone. In practice, p-values less than a threshold (0.05 or 0.01) are typically considered statistically significant, because a low p-value (with high probability) rejects the null-hypothesis: that the input variable(s) and output variable are unrelated.

A full discussion of ANOVA is found in many textbooks (for example, [26]). We use the open source ANOVA package in \( R \) for our analysis.

2.5 Datasets

Our work draws on two active measurements: IPv4 surveys and Trinocular outage measurements.

To validate our results with complete information we re-evaluate Internet IPv4 surveys [10]. These datasets have probes to every address in about 2% of IPv4 /24 blocks, taken every 11 minutes for 2 weeks. With complete data for all IPs in the block, surveys provide ground truth. We primarily use Survey \( S_{21w} \), starting 2012-11-16. To detect trends over the last three years (Figure 11), we analyze historical surveys taken since 2010 (Survey \( S_{31w} \) [37]).

Our new estimators to derive short-timescale block availability and diurnality are based on data from Internet outage observation. We use outage data collected from Trinocular [31]. Our analysis here uses the \( A_{12w} \) dataset, taken from Los Angeles, California, USA [38] and covering 3.7M Internet edge network blocks over 35 days, starting at 2013-04-24 17:18 UTC. We observe similar results from two other sites in Fort Collins, Colorado, USA, and Keio, Japan.

These datasets are publicly available [37]. We add new public datasets for link technology (§ 2.3.3) and our new availability and diurnal analysis.

3. VALIDATION

To validate our new measurement approaches we re-evaluate Internet surveys (summarized above § 2.5). We add controlled simulations of diurnal blocks to systematically investigate our sensitivity to variation in diurnal behavior. We also evaluate diurnal detection against blocks at USC that are or perhaps should be diurnal.

3.1 Tracking Block Availability

We first validate that we accurately estimate availability (\( \hat{A} \)), since detection of diurnal blocks uses this estimate. We show our measure with examples and evaluation of all 29k blocks in \( S_{51w} \).

3.1.1 Sample Blocks

To illustrate estimation of \( \hat{A} \), we first compare \( \hat{A} \) to true \( A \) for several representative sample blocks drawn from Survey \( S_{51w} \) (Figures 1 to 3). The survey provides complete data for all 256 addresses in the block for two weeks. We use the Trinocular algorithm to downsample this data, considering from 1 to 15 samples per round (shown as red dots on the graphs). We apply Trinocular’s adaptive sampling to consider from 1 to 15 samples per 11 minute round. We then compare short-term and operational estimates (\( \hat{A} \) and \( \hat{A}_o \), the green and dark blue lines) and compare them to the true value (\( A \), the black line) measured from complete data. Full data is shown in the image above each graph, with light green dots showing responses and black showing non-responses. We have examined hundreds of blocks; the examples we show here represent different conditions.

Figure 1 shows a block that is sparse, with only a few responding addresses (42 active over time), but relatively high availability (\( A = 0.735 \)). The middle graph shows that the short-term estimate is noisy around the true value (the jagged green \( \hat{A}_s \) line, compared to the steady black \( A \) line). We also see the operational value never overestimates once it leaves its inaccurate initial value (\( \hat{A}_o < A \), the dark blue line is lower than the black line). Finally, we detect an outage at round 957.
Figure 1: Block 1.9.21/24 from $S_{51w}$, a sparse but high availability block (42 every responding addresses, $A = 0.735$). Top: raw survey data: green is response, black is no response. Middle: compares true $A$ (black, upper line with minimal variation), short-term estimate $\hat{A}_s$ (light green, upper line with larger variation), operational estimate $\hat{A}_o$ (dotted dark blue, lower line), and report the number probes per round (red dots at bottom and right scale). An outage occurs at round 957 (the magenta line). Probes-per-round (red dots, mostly along the bottom) use the right scale. Bottom: amplitude of FFT ($N_d = 14$) of $\hat{A}_s$, with no strong diurnal signals.

Figure 2: Block 93.208.233/24 from $S_{51w}$, $|E(b)| = 245$, $A = 0.598$. Diurnal block. Lines described in Figure 1. Bottom: amplitude of FFT ($N_d = 14$) of $\hat{A}_s$ showing strong diurnal amplitude.

Figure 3: Block 27.186.9/24 from $S_{51w}$, $|E(b)| = 256$, $A = 0.191$. Diurnal block. Lines described in Figure 1. Bottom: amplitude of FFT ($N_d = 14$) of $\hat{A}_s$, showing strong diurnal amplitude.

3.1.2 Comparison in Full Survey Blocks

To generalize from these examples we next systematically compare all 29k /24 blocks in Survey $S_{51w}$. We apply our availability tracking algorithm with sparse sampling and estimation to each block and compare to full data as ground truth.

Figure 4 compares our short-term estimates ($\hat{A}_s$) against ground truth ($A$) as a density plot. We normalize the density with the product of number of blocks and number of rounds. The dense cluster near the “perfect correlation” of $x=y$ line shows this correlation visually. Over the density plot, we plot the median (white line) and quartiles for all values in each group of 0.1 true $A$ (white boxes), confirming our estimator is unbiased.

Figure 5 shows the same visualization for operational $\hat{A}_o$. Our goal is that the operational value not overestimate. This data shows our operational $\hat{A}_o$ values are almost always unbiased.
3.2 Diurnal Blocks

We next consider sample diurnal blocks, then systematically explore how different factors detect of diurnallness.

3.2.1 Sample Blocks

We first consider what spectral analysis (§3.2) reports about our sample blocks. In Figure 1, a few addresses are always used, and in Figure 2 all addresses are used intermittently. Both blocks show a flat spectrum, as illustrated by the bottom graph in Figure 1 (the second example is similar but omitted for space).

By contrast, the daily pattern in Figure 3 is visible in the raw data (top) and observed $A_i$ (middle). It produces a strong peak corresponding to daily activity (bottom graph, (94% of the time) under true $A$. Since we do not probe very-sparse blocks, we omit cases where $A_o < 0.1$.

Finally, we quantify the accuracy of our estimates by computing the correlation coefficient over true block availability $A$ and estimated values. We find good correlation between true $A$ value and estimated $A_i$ with an overall correlation coefficient of 0.95685.

3.2.2 Controlled Diurnal Blocks with Simulation

We next turn to simulation to systematically study how easily we can detect diurnallness in the face of different kinds of variation and noise.

We simulate one /24 block (256 addresses) to provide exact ground truth against which to test our algorithm. Our simulation evaluates responses every 11-minutes for 4 weeks. In that block, 50 addresses are stable and always responding, and $n_d = 100$ addresses are diurnal, and the remaining addresses are not active. Diurnal addresses are responsive for 8 hours and down for 16 hours each day. Each diurnal address $i$ turns on at a certain time during the day, the phase $\phi_i$.

We evaluate the effects of three kinds of variation in this model. First, we vary phase, selecting $\phi_i$ for each address (once, at simulation start) uniformly from the range $[0, \Phi)$. We also add normally-distributed noise, either adding variance $\sigma_a$ to the start time of the up period, or changing its duration with variance $\sigma_d$. In both cases, noise is drawn each day for each address. In each simulation below, we run 10 batches of experiments, each batch with 100 experiments to report accuracy: percentage of correct detection of diurnal blocks within the 100 experiments. The error bars show median and quartiles in the 10 batches.

For each simulation we observe $A_i$ (§3.2) and diurnal detection (§3.2.2). We detect 100% of these simulated diurnal blocks in the simplest case when there is no noise ($\Phi = \sigma_a = \sigma_d = 0$).

Number of diurnal addresses: We first explore effects of how many addresses are diurnal. In Figure 7, we vary $n_d$, the number of diurnal addresses, from 1 to 100 (or 2% to 67% of responsive addresses). Our detection accuracy improves quickly. After 10 addresses (or 17% of responsive addresses), the accuracy is more than 85%. Examination of several cases shows the misses with $n_d < 10$ are because $n_d$ is only a small fraction of the 50 always-on addresses, and our probing will usually pick an address within the stable ones and stop.

Varying phase: To vary phase we return to $n_d = 100$ but now select $\phi_i$ randomly, picking a value linearly dis-
distributed between 0 and $\Phi$: to decide when it starts during the time of each day. Figure 8 shows a sharp drop when maximum $\Phi$ at 14 hours. This sharp threshold is due to the twice next strongest amplitude requirement of strict diurnality; although variations in phase decay gradually, the signals from different addresses to blur together around this period, as confirmed in simulations where we vary maximum $\Phi$ from 0 to 24 hours (omitted due to space).

**Varying duration:** We vary duration by changing $\sigma_d$ in Figure 9. We vary $\sigma_d$ from 0 to 24 hours to evaluate our sensitivity to variations in up-time. We find that this variance only slightly affects accuracy for large $\sigma_d$ (>10 hours). This is because we synchronize the up periods daily, similar to real world settings where people synchronize clocks and go to work. With a normal distribution, variations in up-time for the sub-signals will cancel each other out over time. Considering that ordinary people’s schedules have variations within only a few hours, this shows that our algorithm works well for wide range of up-times.

After studying sensitivity of $n_d$, $\phi$, $\sigma_d$, we are confident with our availability tracking and diurnal detection algorithms. Because we can track blocks with small $n_d$ (less than 20% of stable addresses), large $\Phi$ (typical human phase is less than 4 hours), and large $\sigma_d$ (typical up time is 6 to 10 hours).

**3.2.3 Comparison in Full Survey Blocks**

We next evaluate diurnal behavior with and without estimated $A$. Following § 3.1.2, we define ground truth as diurnal detected with true availability ($A$) taken from full data in Survey $S_{51w}$; we cannot contact thousands of network operators to get actual ground truth. We compare that to diurnal computed from our lightweight estimate of availability ($\hat{A}$).

The comparison in Table 1 shows that our method has good precision (82.48%); we rarely falsely predict a wrong diurnal block. We also provide good overall accuracy (90.99%), which represents the ability to correctly track both positives and negatives. Our measurement is conservative in detecting diurnality, with a fairly high false negative rate. For the comparisons we make in § 5, this bias seems preferable to false positives.

**3.2.4 Validation with USC Ground Truth**

We complement blind validation and controlled experiments by comparing our results to operator-provided ground truth at USC (we cannot obtain ground truth for thousands of ISPs in the general Internet). We next verify that the di-
urnal blocks we discover there are correct (no false positives) and evaluate completeness (false negatives).

**True Positives:** In $A_{i2w}$ we find four strictly diurnal blocks and 44 relaxed diurnal blocks. We evaluate block usage using both reverse DNS names and discussions with network operators. USC is large and network configuration is distributed and done partially on physical location, so DNS names provide the most precise information; we then confirm our results with operators.

Of the four strictly diurnal blocks, two are wireless blocks with dynamic address assignment. The other two have mixed use, mostly assigned to general use (assigned to buildings or departments). One block includes a range of 16 dynamically-assigned addresses. Examining the additional 40 relaxed diurnal blocks, keyword assignment identifies 20 as dynamic, 4 as wireless, and the other 16 are general use. In addition to hostnames, for 39 blocks (2 strong and 37 relaxed diurnal) we have full raw data from Survey $S_{51w}$.

**False Positives** are blocks detected as diurnal when survey shows consistent availability. Examination of all 39 blocks with full data shows clear diurnal trends in 38 blocks: $A$ changes by at least 10% over multiple days in the two week survey. DNS data for the one block without diurnal trends shows a group of 15 dynamic addresses for mobile wired devices. We believe these addresses were activated in the four months between ground truth ($S_{51w}$) and detection ($A_{i2w}$). We conclude that there are at most 3% false positives for USC.

**False Negatives** are diurnal blocks that we miss. At first glance, labeling only 20 dynamic and 6 wireless blocks as diurnal suggests a high false negative rate. (DNS at USC identify 32 blocks as dynamic and 142 as wireless, types that are often diurnal.)

Examination of these blocks shows that USC’s wireless is heavily overprovisioned, with one wireless address for every student (142 blocks). On average, each block has around ten live addresses at any time (a load confirmed by operators), even though many addresses are used over months of observation. To avoid excessive per-address probing, by default Trinocular removes blocks with fewer than 15 active addresses, and it is difficult to detect diurnal trends in very sparse blocks (for example, see Figure 7). Thus, of the 142 wireless blocks, 119 were excluded from probing. Of the 23 blocks that are probed, we identify 6 as diurnal. We conclude that diurnal detection requires either moderately dense blocks (more than 15 active addresses) or tolerance for higher probing rates than 19 probes per hour per /24. These self-imposed policy constraints (constraining probing) result in false positives for USC’s sparsely operated wireless network.

Sparse blocks cause only false negatives, not false positives. We therefore claim that our Internet-wide measurements are a lower-bound on the fraction of diurnal blocks, and emphasize the correctness of diurnal claims (97% true positives).

**Surprises:** We found two surprising results examining USC. First, while DNS names show a few large, continuous areas that are diurnal (centralized wireless), decentralized address management also results in pockets of dynamic addresses in many /24 blocks. These dynamic addresses result in diurnal trends in otherwise general-use blocks. Second, while we expected diurnal usage of wireless and dynamically assigned addresses, we were surprised by the strong diurnal use in the 16 general-use blocks. We speculate that diurnal use reflect either broad use of laptop computers that are away from USC outside business hours, or a policy of turning off desktop computers at night.

### 3.3 Stability from Different Locations

To confirm that our approach is stable, we collected data from two other sites concurrent with $A_{i2w}$: Colorado State University, and $A_{i2j}$ from Keio University in Japan. We next compare $A_{i2w}$ against these datasets.

For each dataset, we define $d$ as strongly diurnal blocks, $r$ as relaxed diurnal blocks, $e$ as either $d \cup r$, and $N$ as neither. In Table 2 we compare $d$, $e$, and $N$ for $A_{i2w}$ and $A_{i2j}$. The first row shows that of the 345,976 diurnal blocks in $A_{i2w}$, $A_{i2j}$ finds 85% as strongly diurnal and 98.8% as either strong or weakly diurnal, suggesting strong disagreement in only about 1.2% of blocks. We see similar results when we compare $A_{i2w}$ and $A_{i2j}$ after we account for a software bug (see [32]). We conclude that our approach is not sensitive to measurement location.

### 4. DIRECTLY OBSERVED RESULTS

In this section, we show direct observations of diurnal blocks in $A_{i2w}$. We compare these observations with other factors to identify correlations in the next section § 5). We find 412k blocks (11%) that are strictly diurnal in $A_{i2w}$, and 926k blocks (25%) that are relaxed diurnal of the 3.7M blocks we measure. These blocks are distributed all over the address space.

**Daily or other periodicity?** Our test for diurnality looks strictly for regularity at 24 hours, requiring that to be the strongest periodicity in the block. We focus on daily patterns because they are likely to be related to human use of the Internet. There may be other sources of periodicity, for example related to DHCP lease cycles. If dynamic addresses are allocated for some period $p$, and given out sequentially across a region that spans multiple /24 blocks, then those blocks will see usage that changes with period $p$.

To understand the how prominent different periodicities are, Figure 10 shows the distribution of the strongest frequency for each of the 3.7M blocks in $A_{i2w}$. As expected, we observe a strong peak at 1 cycle per day (24 hours), about 25% of all the blocks. We only declare 11% to be strongly diurnal, where the 24-hour amplitude is twice the next strongest frequency; the remaining are relaxed diurnal. The second group (3%) is at about 4.3 cycles per day. This periodicity is a probing artifact, because we restart our probing software every 5.5 hours (4.3 times per day) to recover from possible probe failure. Our measurements starting in 2014-04 ($A_{16fall}$ [37]) use restart times around one week to reduce this effect.

<table>
<thead>
<tr>
<th>$A_{i2j}$</th>
<th>diurnal</th>
<th>either</th>
<th>non-diurnal</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>297,080</td>
<td>341,919</td>
<td>4057</td>
<td>345,976</td>
</tr>
<tr>
<td>$e$</td>
<td>332,250</td>
<td>569,882</td>
<td>106,391</td>
<td>676,273</td>
</tr>
<tr>
<td>$N$</td>
<td>1898</td>
<td>122,173</td>
<td>1,996,653</td>
<td>2,118,826</td>
</tr>
<tr>
<td>all</td>
<td>334,148</td>
<td>692,055</td>
<td>2,103,044</td>
<td>2,795,099</td>
</tr>
</tbody>
</table>

Table 2: Comparing diurnal ($d$), diurnal and relaxed ($e$), and non-diurnal ($N$) for $A_{i2w}$ and $A_{i2j}$. 

---

For each dataset, we define $d$ as strongly diurnal blocks, $r$ as relaxed diurnal blocks, $e$ as either $d \cup r$, and $N$ as neither. In Table 2 we compare $d$, $e$, and $N$ for $A_{i2w}$ and $A_{i2j}$. The first row shows that of the 345,976 diurnal blocks in $A_{i2w}$, $A_{i2j}$ finds 85% as strongly diurnal and 98.8% as either strong or weakly diurnal, suggesting strong disagreement in only about 1.2% of blocks. We see similar results when we compare $A_{i2w}$ and $A_{i2j}$ after we account for a software bug (see [32]). We conclude that our approach is not sensitive to measurement location.
Long-term status: We next consider how block diurnalness changes over time. In Figure 11, we apply our spectral diurnal detection algorithm over more than 3 years of Internet surveys (each a sample of about 2% of IPv4 blocks). We observe that the percentage of diurnal blocks is relatively stable, but that it shows a marked decline since 2012. While we have observed in increasing trend in dynamic address allocation this downward trend may indicate that these dynamic addresses are slowly shifting to always-on activity (even if they move around). These results are generally consistent the 11% diurnal blocks in $A_{12w}$.

5. INDIRECTLY OBSERVED RESULTS

We next compare diurnal blocks to other factors: location, time-of-day, block allocation dates, economic factors, and link technologies.

5.1 Where Does the Internet Sleep?
(Locating diurnal blocks)

Our first question is to identify where diurnal Internet use is common. We study this question using $A_{12w}$ with MaxMind geolocation (succeeding for 3.45M of blocks, 93%). We first put our results in context by showing where the geolocatable blocks are located (Figure 12). In this world map, we count /24s blocks geolocated into a 2x2 degree grid. We see large numbers of addresses in North America and Europe, as well as concentrations in Japan, China, and several other countries. We also see some geolocation anomalies: when MaxMind knows only country, it locates the blocks in the geographic center of the country, falsely placing many networks away from population in Brazil, Russia, and Australia.

Table 3 lists the countries with the most diurnal blocks.

Table 3: Fraction of diurnal blocks, top 20 countries (with at least 1000 blocks in our study), and United States. Diurnal analysis data is from $A_{12w}$; Geolocation data is from MaxMind [25]; Per-capita GDP data is from CIA world factbook [5].

5.2 When Does the Internet Sleep?
(Phase of Diurnal Blocks)

Amplitude in spectral analysis of $A$ shows which blocks are diurnal, and geolocation showed where these blocks are. We next show that phase in the FFT coefficients shows when these blocks are active.

We next look at the phase of geolocated diurnal blocks to see that it correlates strongly with longitude and therefore timezone. We measure how many blocks occur at each combination of longitude and phase. Since both phase and longitude are continuous, we count similar blocks by quantizing longitude and phase each into 100 bins ($\pm 180^v$).

Since both wrap around the circle, we “unroll” them by plot-
Table 4: Fraction of diurnal blocks grouped by regions. Diurnal analysis data is from $A_{12w}$; Geolocation data is from MaxMind [25].

<table>
<thead>
<tr>
<th>region</th>
<th>blocks (/24s)</th>
<th>frac. diurnal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern America</td>
<td>721716</td>
<td>0.002</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>11255</td>
<td>0.0108</td>
</tr>
<tr>
<td>W. Europe</td>
<td>275224</td>
<td>0.0109</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>133911</td>
<td>0.0131</td>
</tr>
<tr>
<td>Caribbean</td>
<td>2174</td>
<td>0.016</td>
</tr>
<tr>
<td>Oceania</td>
<td>27206</td>
<td>0.0349</td>
</tr>
<tr>
<td>W. Asia</td>
<td>25570</td>
<td>0.0765</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>9984</td>
<td>0.0992</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>134933</td>
<td>0.124</td>
</tr>
<tr>
<td>Central America</td>
<td>44644</td>
<td>0.133</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>146552</td>
<td>0.135</td>
</tr>
<tr>
<td>Southern Asia</td>
<td>44524</td>
<td>0.200</td>
</tr>
<tr>
<td>South America</td>
<td>133493</td>
<td>0.208</td>
</tr>
<tr>
<td>South-Eastern Asia</td>
<td>48885</td>
<td>0.219</td>
</tr>
<tr>
<td>Eastern Asia</td>
<td>757352</td>
<td>0.279</td>
</tr>
<tr>
<td>Central Asia</td>
<td>3832</td>
<td>0.401</td>
</tr>
</tbody>
</table>

5.3 Correlating Block Diurnality and Allocation Date

Policies for IP address allocation have evolved since IPv4 was introduced, with stricter requirements for density and reuse that encourage dynamic allocation through full alloca-

not attempt to align phase zero with longitude zero; their proximity is accidental and calibrating phase with local time of day is future work.

In Figure 14b we repeat this analysis with 500k blocks that are relaxed diurnal and geolocatable. The fit here is not quite as good (correlation coefficient is 0.763), but it is very close. This strong similarity suggests that our strict test for diurnal may be overly conservative.

In both Figure 14 we observe that the correlation between phase and longitude is fairly strong, except for longitudes between 100°E to 140°E. This range corresponds with Mongolia, China and southeast Asia. We believe this variation is primarily because our IP geolocation is not particularly accurate in China, and because it is a geographically wide country that uses a single timezone.

Finally, the relationship between phase and longitude suggests that phase may help geolocate diurnal blocks. Figure 14c shows mean and standard deviations when this mapping is derived from relaxed diurnal data. We see that the quality of the predictor varies significantly: a phase from $-\pi$ to 0 indicates only the hemisphere, while most other phases predict longitude within ±20 degrees.
5.4 Effects of Economic Conditions

We hypothesize that economic factors affect Internet use, since dynamic address assignment and turning devices off at night is more efficient than static or always-on dynamic assignment and always-on devices, and economic factors affect investment in telecommunications infrastructure. While one might suspect that economic factors correlate with diurnal-ness, our evaluation is able to quantify this relationship for the first time.

5.4.1 Correlating Diurnal Blocks with Economics

Why do some countries have a greater fraction of diurnal blocks (§ 5.1)? Cultural differences across countries prompt different rates of savings, and different economic situations leave citizens different levels of disposable income. These differences and anecdotal evidence suggest that people of some countries, such as China and India, to turn off their computers at night to save electricity bills.

Although we cannot directly model cultural differences, we can observe correlations between disposable income (measured in per-capita GDP, taken from the CIA World Factbook [5]) and diurnal network usage. Table 3 previously gave specific values. Figure 16 visualizes this correlation, including the weak linear fit (confidence coefficient: -0.526; such weak fits are common with coarse GDP data and few countries). The top 20 countries with diurnal fractions above 0.15 generally have a per-capita GDP less than USD $15,000, less than one-third of the United States. We conclude that economic factors correlate with differences in Internet use.

5.4.2 Non-GDP Factors: electricity and allocation age

GDP is one of many economic indicators. To understand correlations with other factors we carry out ANOVA factorial analysis (§ 2.4) comparing five factors: per-capita GDP,
Internet users per host, electricity consumption per capita (each from the CIA factbook), and age of first (and mean) block allocation (from IANA) and fraction of diurnal blocks (from our analysis).

Table 5 shows this analysis for all one- and two-factor combinations. A p-value less than 0.05 suggests a strong correlation. We find there are three combinations meet this requirement: per-capita GDP, with p-value of $6.61 \times 10^{-8}$. The second is the combination of per-capita electricity consumption and mean age of allocation, with p-value of 0.001476. The third is mean age of allocation, with a p-value of 0.031354.

Our analysis with ANOVA supports our previous hypothesis that per-capita GDP is the main factor for diurnlness, and suggests that electrical usage and mean allocation age are also correlated. Electricity consumption is correlated with economic activity (and thus GDP), but ANOVA considers each combination of factors. Mean allocation age suggests that newer countries have more diurnal use, as shown in §5.3.

5.4.3 Independence and Relationship to Policies

ANOVA shows correlations between diurnal networks and several individual and pairs of several measurable factors. Although ANOVA is a powerful technique, we must take care that factors not necessarily independent—low GDP may imply less developed infrastructure and less available power. Future work is needed to show a causal relationship between these (or other) factors and network use, and to identify how specific government and telecommunications policies relate to these measurable factors. Policy analysis is typically much more qualitative, since policies vary by country and region, and attribution of specific results to even a single policy requires significant effort (the U.S. Broadband Plan is more than 350 pages [9]). Our work suggests the potential to identify such relationships, is the first to describe specific measurements to assist this direction of study.

5.5 Effects of Access-Link Technology

We next consider correlations between access-link technology and diurnalness. National policies are often set to encourage new link technologies and promote always-on networks. In addition, while such policies and economic conditions are difficult for individuals to influence, users and ISPs often can choose different “last-mile” methods to connect to the Internet.

We infer link type (§2.3.3) for all blocks in $A_{12w}$, classifying 22.4% of those blocks. For these blocks, we correlate link type with diurnlness. Each of these link types have more than 30k blocks; we omit the wireless keyword because we see only 691 wireless/wifi blocks, not enough to provide statistically strong results.

Figure 17 shows the fraction of diurnal networks for each different link type. As expected, dynamic addresses are strongly correlated with diurnal behavior, with dynamic at 19%. Somewhat surprisingly, dialup is not strongly diurnal (< 3%), while dial is more diurnal (11%). These results suggest the importance of measuring network behavior rather than assuming.

5.6 Applications

This paper demonstrates that it is possible to identify diurnal network usage from active measurements, and that diurnal networks correlate with GDP and other factors. A direct application of this work is to use it to explore the outcomes of government and telecommunications policy decisions (§5.4.3). For example, a goal of U.S. broadband deployment is always-on networking [9]; one can evaluate if new network technologies (fiber-to-the-home or cable modems) have are less diurnal (and so more always-on) than older technologies (dial-up, ISDN) by looking for correlations in access-link technology (§5.5). Comparisons of fraction of diurnal networks per country or state can be used to judge the progress of different areas towards the goal of always-on networking.

Information about blocks that are diurnal can also be used to calibrate other measurements. For example, one can scan the IPv4 space in tens of minutes to estimate the availability of each /24 block, but this near-snapshot will be representative only for non-diurnal blocks. For diurnal blocks, one needs several measurements at different times-of-day to determine the range of values.

Finally, information about diurnal networks is one component towards estimating the size of the Internet measured in active, public IP addresses [10].

6. RELATED WORK

ISP comparison: Several third-party systems exist to compare performance of ISPs, including Keynote [19], top ten reviews [19], and the popular speed test tool offered by Ookla [29]. They share a common drawback: lack of scal-
ability. Perhaps closest to our work is Netdiff [24], which measures performance of hundreds of paths for 18 backbone ISPs. They find that ISP performance depends on geographic properties of traffic and popularity of destinations. Our work is more focused on nodes instead of paths, and thus complements Netdiff.

Adaptive and pervasive sampling: Several prior studies have used adaptive probing techniques to find network faults (for example, [2]), and use simulations for validation. Similar to these, we use adaptive sampling inside network blocks to achieve accurate availability information ($A_i$), with low probing cost.

Prior work has explored full enumeration of IPv4 [10]. We use some surveys from this work to validate our sampled results. Although we can do full enumeration quickly, as advocated by others [8], such effort is a poor match for long-term data collection: it is unnecessary and annoys target network operators. Instead, we build directly on the adaptive data collection methods of Trinocular [31], adding new analysis to identify diurnal networks.

Node-based studies: Barford et al. use node-based methods to detect network performance anomalies [1], with clever weighting algorithms to find which links to probe next; they simulate 15 types of topologies based on ns-2. Several other previous studies also use similar strategies, including Hubble [18] and iPlane [23]. Such node-based studies are precise in understanding performance on link paths between the node mesh. However, they usually “only” cover hundreds or thousands of nodes. We complement such systems with a much larger coverage (3.7M /24 blocks), and with measurements on a regular basis every 11 minutes.

Tomography studies: Network tomography is interested in finding the exact location of problems [6,7,12,20,21]. We correct for correlated error by removing outages near our provers. We are interested in tracking block-level network reachability in the whole Internet edge, but without exact localization of outages. Our work and tomography studies can benefit from each other. Our future work could be more precise by localizing problems. Tomography studies can use our system as a pre-step for finding symptoms.

7. CONCLUSIONS

In this paper, we develop an accurate adaptive approach to estimate network block availability over time. Based on this, we perform the first long-term, wide-scale analysis of diurnal network behavior and outages. We also begin to evaluate how policies, economic conditions and technologies affect Internet use in the world. With statistical analysis (ANOVA), we correlate diurnal usage and outages to economic factors such as GDP and power consumption.

Acknowledgments

This research is partially sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, HSARPA, Cyber Security Division, BAA 11-01-RIK and Air Force Research Laboratory, Information Directorate under agreement number FA8750-12-2-0344, and via SPAWAR Systems Center Pacific under Contract No. N66001-13-C-3001. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views contained herein are those of the authors and do not necessarily represent those of DHS or the U.S. Government.

This work was classified by USC’s IRB as non-human subjects research (IIR00001648).

We thank Brian Yamaguchi (USC/ITS) for guidance about USC networks, and Christos Papadopoulos (CSU) and Rod Van Meter, Yohei Kuga, and Midori Kato (WIDE) for supporting data collection from their sites. We thank John Wroclawski, Alberto Dainotti, Georgios Smaragdakis (our shepherd), and the anonymous reviewers for comments on the paper.

8. REFERENCES
