CenAlert: Amplifying User Voices to Rally Censorship Investigation

Aaron Ortwein, Anna Ablove, Armin Huremagic, Luqin Chang, Vinicius Fortuna†, and Roya Ensafi
University of Michigan, {aortwein, aablove, agix, luchang, ensafi}@umich.edu
†*Google Jigsaw, fortuna@google.com*

***Abstract*—The commoditization of deep packet inspection technology has eased the deployment of Internet censorship, even in countries typically considered open. While the Internet freedom community has been vital to increasing transparency, its coverage relies heavily on NGOs and activists, whose ability to relay user reports of censorship is increasingly threatened by mounting risks, anti-NGO legislation, and foreign funding cuts. To bridge this widening gap, we examine whether other data sources can be repurposed to amplify user voices and rally censorship investigation.**

We leverage Google Trends data to build *CenAlert*, a user-driven alert system that detects spikes in search interest for circumvention tools. Each spike is scored with an *impact factor* quantifying increases in circumvention tool demand and enabling prioritization of response efforts. We demonstrate the effectiveness and practicality of *CenAlert* across 76 censoring countries over 14 years. Of 269 selected spikes, 191 are explainable, including 153 as censorship events. Notably, 68 spikes correspond to censorship events not previously reported by the community. To facilitate integration with observatories, *CenAlert* is open-source and provides a dashboard, API, and Slack webhook for notification, visualization, and analysis of potential censorship events. Ultimately, *CenAlert* offers a novel solution to long-standing challenges faced by the Internet freedom community.

1. Introduction

The commoditization of deep packet inspection technology [78], [100], [112] is accelerating the proliferation of Internet censorship beyond traditional offenders, contributing to the global decline in Internet freedom [2]. As censorship spreads, it is also evolving: governments are increasingly deploying new techniques such as blocking circumvention protocols [115], throttling Internet speeds [117], and demanding server-side takedowns [13]. The alarming scale and diversity of censorship present a formidable challenge for oversight and accountability. The Internet freedom community has been critical in bringing transparency to where, when, and how censorship occurs and in enacting meaningful change. For example, observatories such as the Open Observatory of Network Interference (OONI), Censored Planet, Internet Outage Detection and Analysis (IODA), and NetBlocks have documented censorship during politically sensitive events, across new countries, and in both novel (e.g., large-scale TLS interception in Kazakhstan) and extreme forms (e.g., total Internet shutdowns) [19], [83], [86], [87]. Civil society organizations such as Access Now, Internet Society Pulse, Freedom House, and European Digital Rights have furthered these monitoring efforts by corroborating reports from local partners and news media with observatory data. The resulting awareness and understanding of censorship events have informed journalism, supported activism, and driven policy responses to defend Internet freedom [5], [64].

The Internet freedom community relies heavily on alert channels to determine where to focus its limited time, personnel, and funding. These alerts come primarily from on-the-ground activists and non-governmental organizations (NGOs) that relay user reports of censorship [54], limiting coverage to regions where connections with such organizations have been established. Beyond the inherent risks to activists, the operation of many NGOs is increasingly under threat. Several countries have passed anti-NGO legislation [10], [41], and the United States—one of the key funders of pro-democracy organizations worldwide—has effectively dismantled both the United States Agency for International Development (USAID) and the Bureau of Democracy, Human Rights, and Labor (DRL), together with the overwhelming majority of their foreign programs [67].

With traditional alert channels eroding, we need new ways to hear users' voices, especially in countries with limited coverage. Given these constraints, we seek to answer two key questions: (1) Are there existing data sources that can be repurposed for supporting anti-censorship efforts? (2) Can this data be leveraged efficiently and reliably as an alert channel to reduce the burdens on local NGOs and activists? The business models of "surveillance giants"—unfortunately built on collecting vast amounts of user data for advertising and tracking—have created potential data sources that reflect users' voices and could be repurposed to support anti-censorship efforts [11], [12]. For instance, the Google Transparency Report and Cloudflare Radar have indicated Internet shutdowns based on sharp drops in usage of their services [6]. These examples underscore the promise of new platforms that leverage industry data for public-interest monitoring.

In this work, we present *CenAlert*, a system that leverages Google Trends data to detect potential changes in censorship practices. Google Trends [47] is a public source of search data spanning back to 2004 and is widely used for marketing purposes [48]. For a given search topic, country, and time range, Google Trends returns a time series of search interest. *CenAlert* detects spikes in circumvention-

related searches, indicating changes in users' expectations or experiences of Internet restrictions. However, using this data introduces several challenges: (1) *variability*, where identical queries at different times may yield different results; (2) *normalization*, where values are relative proportions of the time series maximum; (3) *resolution constraints*, where fine-grained (e.g., daily) data is only available for short time ranges; and (4) *sparsity*, where the time series may contain a high proportion of zeros. These challenges limit its utility for the Internet freedom community, which needs reliable and timely alerts to allocate its limited resources effectively.

CenAlert aggregates multiple downloads of each query to stabilize results, and jointly addresses normalization and resolution constraints by scaling and stitching short windows into a single time series. It then applies statistical techniques to detect spikes in circumvention-related searches. To improve its robustness across diverse time series patterns both within and across countries, *CenAlert* dynamically selects between anomaly detection algorithms based on the sparsity of recent data and uses parameters tuned on a per-country basis. Each spike is assigned an *impact factor*, which quantifies users' desire to access unavailable content and can help the Internet freedom community prioritize response efforts.

We demonstrate the utility of *CenAlert* by evaluating its spike detection over 14 years and across 76 countries previously reported to restrict Internet access. Our results show that *CenAlert* is highly effective and easily integrable, with an annual spike volume well within the capacity of existing observatories; even the highest-volume regions receive an average of just 1.79 alerts per month. We investigate a total of 269 spikes, including the 100 highest impact spikes and all spikes for nine countries considered "Not Free" by Freedom House. We explain 191 of them, with 153 corresponding to censorship events. Notably, 68 spikes were censorship events not previously reported by the Internet freedom community but were manually verified via news reports or social media posts. These include detecting the blocking of Facebook during civil war in Libya, WhatsApp during presidential term limit protests in Togo, VoIP services in Qatar and UAE, VPN protocols in Myanmar, and Roblox and Wattpad in Vietnam. Despite all being high-impact censorship events, they have not been studied in depth.

Because *CenAlert* measures changes in user behavior driven by experiences or expectations of Internet restrictions, it is best suited to detect censorship events that affect large populations over sustained periods. The above examples involved widely used and hard-to-replace services such as social media, entertainment, and circumvention tools. By contrast, subnational, short-lived, or narrowly targeted (e.g., disproportionately affecting marginalized or minority populations) censorship events tend to have little impact on national-level search patterns and are therefore outside of *CenAlert*'s scope.

Overall, the Internet freedom community's ability to address complex challenges—monitoring censorship in overlooked countries, investigating diverse tactics, and tracking the global export of censorship technology—is under severe threat. Over the past decade, support from organizations such as DRL and the Open Technology Fund has enabled the community to play a central role in facilitating Internet access worldwide, especially during crises when access was urgently needed. For example, in countries like Iran, Russia, and Turkmenistan, the connectivity of millions of users has often depended on researchers, circumvention tool providers, and journalists rapidly assessing and responding to emerging censorship events, supporting activists and upholding democratic values in the face of an otherwise insurmountable digital divide.

CenAlert provides a novel and user-driven solution to reduce the burden on traditional alert channels. It is open-source, with a dashboard, API, and Slack webhook to facilitate notification, visualization, and analysis of potential censorship events globally. *CenAlert* serves as a valuable complement to existing monitoring efforts, offering insight into where and when new censorship practices are most impactful for those affected, and ultimately assisting the Internet freedom community in optimizing the use of its limited resources.

2. Background and Related Work

2.1. Current Monitoring and Detection Efforts

Measuring Censorship. Existing censorship measurement efforts are divided between organizations with different but complementary perspectives on the accessibility of Internet resources. The Open Observatory of Network Interference (OONI) [38] crowd-sources censorship measurements from consenting volunteers, sequentially testing for blocking of specific resources at multiple network layers, from DNS manipulation to application-layer interference. Censored Planet also measures filtering, but conducts remote measurements to infrastructural hosts over six different Internet protocols (DNS, TCP, Echo, Discard, HTTP, and HTTPS) [24], [99].

Other organizations focus on network shutdowns. Internet Outage Detection and Analysis (IODA) [55], identifies shutdowns by monitoring simultaneous drops in BGP prefix advertisements made by Internet routers, unsolicited traffic collected by a network telescope, and active /24 blocks as determined by ICMP probing. The Google Transparency Report [49], Cloudflare Radar [28], and Mozilla telemetry data [39] can also measure network shutdowns by showing anomalous drops in traffic to their servers.

Additionally, circumvention tool providers such as Tor and Psiphon maintain user metrics [85], [105], which often react to censorship events. Rapid growth in circumvention tool usage can indicate new censorship policies, while sharp decreases can suggest either that the circumvention tool itself is the target of censorship or that Internet connectivity has been cut.

Verifying and Documenting Censorship Events. Other organizations verify censorship events using multiple sources. Access Now [4] leads a global coalition of civil society organizations in running the Shutdown Tracker

Optimization Project (STOP), which documents censorship events affecting multi-way communication platforms (e.g., social media). Events are verified using measurement data and qualitative reporting from over 25 organizations and are published annually. Internet Society Pulse [97] similarly aggregates several data sources but publishes censorship events as soon as they are verified. Both organizations provide analysis of each censorship event, including its cause, geographic scope (e.g., national or regional), type (e.g., filtering, throttling, or network shutdown), and supporting measurement data or reporting.

Challenges of Censorship Measurement. Despite this diverse ecosystem of censorship measurement and monitoring organizations, its collective ability to detect and verify censorship events is still limited by coverage and resources.

For instance, organizations conducting active measurements can struggle to obtain representative vantage points. Remote measurement organizations send probes to infrastructural servers to comply with ethical guidelines, but non-residential machines may be subject to different censorship policies than end users [88]; in-country volunteers can test from residential networks, but measurements may be sporadic, and recruitment can be challenging given the potential risks of censorship measurement [81]. Even with wide geographic coverage, capturing censorship events requires frequent testing and continual maintenance of locally tailored test lists.

Existing passive measurement organizations are often limited in either scale or the diversity of interference they can detect. Organizations that specialize in detecting network shutdowns by monitoring changes in traffic can more easily operate at a global scale but lack visibility into fine-grained blocking policies. Circumvention tool metrics can give insight into multiple types of censorship, but only if censorship events affect the use of that specific tool.

Furthermore, verifying anomalous measurements is a manual and resource-intensive process, as it requires finding relevant reports or syncing and corroborating with other censorship measurement and civil society organizations [6], [54]. Unfortunately, this laborious process can leave many censorship events unreported.

2.2. Time Series-Based Censorship Detection

Researchers have developed various time series-based techniques to detect Internet censorship by identifying patterns and anomalies in network behavior. These methods generally fall into two broad categories: those that track service usage metrics and those that monitor lower-level network activity.

Prior studies have shown that monitoring the usage of certain Internet services can correspond to reported censorship events. Early work by Danezis [30] modeled a country's Tor usage with a ratio of the current day versus the past week and calculated whether this ratio falls within 99.99% of the normal distribution of 50 other countries' usage ratios; otherwise, it is an anomaly. Similarly, Wright et al. [114] and Kargar et al. [56] analyzed Tor and Psiphon usage metrics, respectively, showing that sudden increases often reflect blocking of social media and messaging platforms, while sharp decreases correspond to blocking of the circumvention tool itself. Filastò et al. [39] conducted case studies in 2021 for Myanmar, Uganda, Belarus, and Iran, highlighting that an absence of Mozilla telemetry data corresponded with Internet shutdowns.

Beyond usage metrics, Farkas [37] proposed a method based on CUSUM (Cumulative Sum) charts to detect subtle performance anomalies in Measurement Lab data, identifying Internet throttling during events like elections. Finally, Chocolatine [50] applies seasonal ARIMA modeling to unsolicited traffic (Internet Background Radiation) to detect connectivity loss resulting from outages or deliberate Internet shutdowns.

2.3. Google Trends

Google Trends is a public source of anonymized, aggregated Google search data dating back to 2004. It accepts the following inputs: a search query (e.g., vpn) or topic (e.g., <Virtual Private Network>), an aggregation of related queries across languages and spellings; a country or subregion, a first-level political division like a state or province; and a time window. Its primary output is a time series representing a random sample of all searches for the query or topic over the location and time window. The time series is normalized relative to its maximum and scaled between $[0, 100]$. Depending on the size of the time window, the granularity of the time series ranges from hourly, daily, weekly, and monthly data points. Google Trends also outputs a list of first-level political subdivisions (if the location is a country) or cities (if the location is a subregion) ordered by search volume, as well as lists of related search queries and topics that are also trending.

Google Trends has gained popularity across research disciplines, with applications ranging from public health surveillance to economic forecasting and behavioral analytics. In epidemiology, it has been used to monitor interest in disease outbreaks such as Ebola and influenza, with several studies identifying strong correlations between search spikes and case counts [8], [21], [80]. In economics, it has been shown to improve near real-time forecasting of unemployment claims, consumer behavior, and automobile sales [22], [27]. Conservation and social science researchers have also used it to understand interest in environmental topics [77] and to forecast protests [107].

3. *CenAlert* Methodology

In this section, we describe the design of *CenAlert*, a system which detects spikes in demand for censorship circumvention technologies using Google Trends data. *CenAlert* has three core components, shown in Figure 1. First, it collects and processes raw search volume data from Google Trends on a daily basis (Section 3.1). It then performs anomaly detection on the processed time series

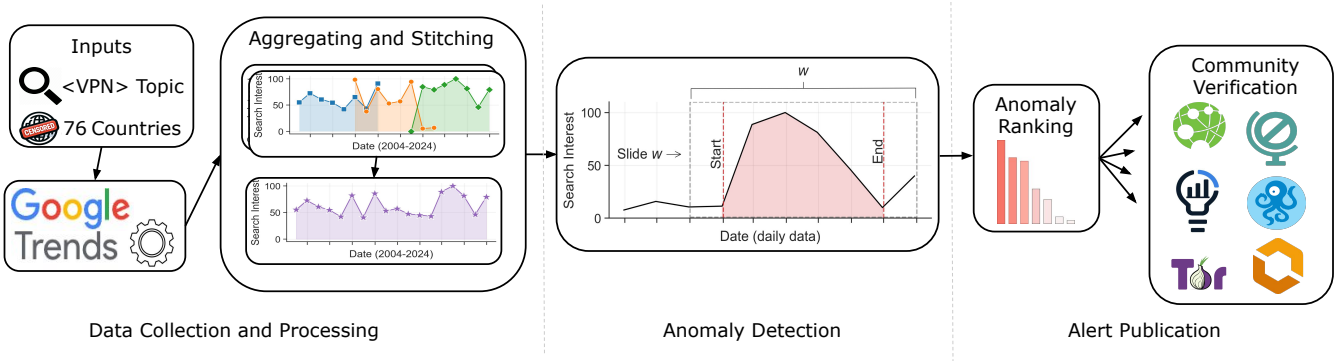

Figure 1: ***CenAlert* System Overview.** *CenAlert* monitors Google Trends data for a circumvention-related term and a set of countries. It first aggregates multiple samples of the most recent data to mitigate sampling variability, then stitches the result together with historical data to enable long-term comparisons of search interest. Next, it applies anomaly detection to detect search spikes. Spikes are scored with an *impact factor*, which quantifies the change in demand for circumvention tools. Alerts of these search spikes will ultimately be published to the Internet freedom community for further investigation.

(Section 3.2), using carefully tuned parameters (Section 3.3), and assigns to each spike an *impact factor*, which can be used to prioritize investigation efforts. Finally, it outputs spikes for further investigation by the Internet freedom community.

3.1. Processing Raw Google Trends Data

Google Trends data is available via a public dashboard. Each day, *CenAlert* downloads the latest circumvention-related search data for every country monitored. However, the raw data is unsuitable for spike detection because it presents three key challenges: *variability*, *normalization*, and *resolution limits*. We now discuss these challenges and our solutions to them in further detail.

Variability. The relative search volumes returned by Google Trends are derived from a random sample of all searches for a term or topic over a given geographic location and time period. The underlying samples change over time, so the same query made at two different times can yield different results. Sampling variability is small for popular search terms or topics but large for unpopular ones.

Previous works commonly mitigate sampling variability by averaging multiple downloads of each time series together, although the exact number of downloads used varies widely [22], [58], [107]. Cebrián and Domenech offer a formula for the optimal number of downloads to achieve a desired margin of error from the true population relative search volumes [23]. However, even for a "medium popularity" search term, achieving 1% error would require averaging 1,000 downloads. Data collection at this volume—especially when extending to multiple geographic regions—is intractable because Google Trends rate limits requests.

When using Google Trends data to forecast other time series such as epidemiological or economic trends, precisely reconstructing population relative search volumes across the entire time series is of utmost importance. However, for detecting search spikes, it is more important to converge on the presence and values of *anomalies*, eliminating just enough sampling variability to avoid false positives. Our insight is that meaningful search spikes will be visible in *every* download. We therefore require consistency across downloads: if a date has zero search volume in any download, we set its value to zero in the aggregated time series. For dates that meet this criterion, we average their demand values across all downloads.

Normalization and Resolution Limits. The challenge of collecting Google Trends data is compounded by normalization of each time series relative to its maximum and the availability of daily data only for time frames of up to 270 days. The former prevents direct comparison of relative search volumes across windows, while the latter prevents downloading long spans of fine-resolution data in one request. However, for Google Trends data—and any spikes derived from it—to be comparable longitudinally, the short time frames downloaded each day must be *stitched* into a single time series, where every point is normalized relative to the all-time maximum.

Existing approaches for reconstructing long-term daily time series all fundamentally rely on the relationship between overlapping time frames [33], [58]. The same date may have different relative search volumes across overlapping time frames that are each normalized to different maxima. However, because the underlying absolute search volume is the same, the ratio between corresponding points in the overlap gives a scaling factor that aligns the normalization of one window with that of the other.

Consider two time series T and w, where T is the stitched historical daily series and w is an aggregated window whose last point is the search volume c for the current day. T and w share an overlapping period δ. We compute the median ratio of points in δ—excluding zero, infinite, or undefined ratios—using it to rescale c before appending the result to T:

$$T' = T||med(\frac{T_\delta}{w_\delta})c \quad (1)$$

However, this approach fails when a gap in data exceeds the duration of δ. In such cases, we leverage coarse-resolution (i.e., weekly or monthly) data. We first download the coarse-resolution time series spanning both T and $T||c$. We then compute the median ratio over the overlapping coarse-resolution period and scale c by the result.

Finally, because this stitching process does not guarantee that c remains in the interval $[0, 100]$ after scaling, we apply min-max normalization to rescale the entire time series. Ensuring a consistent scale enables comparison of relative search volumes—and spikes derived from them—across countries.

3.2. Anomaly Detection

We next perform anomaly detection on the processed time series to identify spikes in relative search volumes.

3.2.1. Needs and Challenges. Regardless of the anomaly detection approach, it must satisfy a few key properties.

Spike Detection. There are many different types of anomalies, and numerous algorithms have been proposed to detect them. Consistent with previous work leveraging user-driven signals to detect potential censorship events [52], [56], [70], [84], [92], we focus on spikes in demand for circumvention tools.

Unsupervised. Spikes in Google Trends data are not labeled, and manual labeling is both costly and error-prone [116]. While prior work has attempted to guide anomaly detection using known events [106], there is no one-to-one correspondence with spikes in Google Trends, as event lists are derived from external sources not influenced by national-level search patterns. Even if every event did map directly to a spike, censorship is difficult to document comprehensively because blocking is often opaque. Relying on partial lists would bias detection towards known events and overlook novel ones. As such, the anomaly detection algorithm must not require existing labeled training data.

Early Detection. Spikes should be reported as soon as they are detected, enabling rapid response by the Internet freedom community. Anomaly detection must therefore operate online, using only past data to determine whether the current day's search volume constitutes or is part of a spike.

Adaptability. Due to the evolution of search term popularity, growing adoption of both the Internet and Google search, and improvements to Google Trends over time, the time series are highly non-stationary. Accordingly, we favor algorithms that can adapt to shifting time series patterns, as well as those that make minimal assumptions about the underlying data distribution.

Robust to Diverse Spikes. Even when focusing solely on search spikes, there are many shapes to capture: additive outliers, which affect a single point; level shifts, abrupt and permanent jumps to a new mean; and transient changes, temporary level shifts that gradually decay back to normal levels [26]. Not all anomaly detection approaches handle these equally well. For example, the Tor metrics anomaly detection system [30] is unlikely to consider the entirety of multi-day spikes as anomalous because every point is independently evaluated as a spike. In contrast, the CUSUM-based approach used by Farkas [37] to detect anomalies in M-Lab Network Diagnostic Test data is designed specifically for sustained spikes rather than single-day ones. Our approach should be adept at detecting both single-point and sustained spikes.

Robust to Intermittent Data. Anomaly detection algorithms require a reference of normal behavior. However, because searches are user-initiated, consistent demand is not guaranteed. When time series contain many zeros, any non-zero value can appear anomalous, even if it is not especially meaningful. Detecting censorship events in low-activity time series has long been challenging, and prior work has excluded them from analysis [114]. Nevertheless, given that capturing these events remains a priority, our anomaly detection should instead manage sparse subsequences so genuine spikes are not missed.

Efficient Parameter Tuning. Anomaly detection algorithms often require parameter tuning to control their behavior. Given the limited resources of the Internet freedom community, it is important to balance limiting the number of spikes that require investigation with maximizing the early detection of significant spikes. Finding acceptable—let alone optimal—parameters can be challenging: the search space quickly grows with each additional parameter, and the objectives are in tension, requiring a trade-off between effort and coverage. Parameter tuning must therefore effectively balance these objectives.

3.2.2. Anomaly Detection Algorithms. At a high level, the anomaly detection algorithm provides a scoring function that quantifies how different (e.g., in terms of probability or distance) the current day's search volume is from a sliding window of w previous observations. If the score exceeds a predefined threshold, an anomaly is detected. In designing *CenAlert*, we consider four unsupervised, distribution-free anomaly detection algorithms for detecting spikes in time series. We evaluate these algorithms in Section 4.

1. Z-Score. Z-score-based anomaly detection identifies anomalies based on how many standard deviations a value lies away from the mean of the sliding window w. For normally distributed data, the Empirical Rule gives that 68%, 95%, and 99.7% of data lies within one, two, and three standard deviations of the mean, respectively. However, the assumption of normality may not hold in practice. Chebyshev's Inequality applies to arbitrary distributions, stating the probability of a value lying k standard deviations from the mean is at most $\frac{1}{k^2}$ [9]. This upper bound may be loose. For example, with normally distributed data and $k = 3$, Chebyshev's Inequality gives a probability of $\frac{1}{9}$, while

the Empirical Rule gives 0.003. We therefore use separate thresholds depending on whether w is normally distributed, as determined by the Shapiro-Wilk test ($\alpha = 0.05$).

2. Median Method. The Median Method [15], previously shown to perform well in detecting extreme values (spikes) [96], forecasts the next value using the median of previous values and the median of their first differences (i.e., the differences between consecutive values). Specifically, for a sliding window w of size 2κ, the prediction is calculated as the median of w plus κ times the median of the first differences of w. A value is flagged as an anomaly if its deviation from the prediction exceeds a threshold τ.

3. Isolation Forest. Isolation Forest [61] constructs an ensemble of binary trees known as Isolation Trees. Each tree recursively partitions the sliding window w at random split values until each point is isolated in its own leaf node. Intuitively, anomalies tend to be isolated in fewer partitions because they are rare and far away from normal data, resulting in shorter paths to their corresponding leaves. The anomaly score is derived from the average path length across all trees and normalized to the range $[0, 1]$, with shorter paths mapping to scores closer to 1.

4. Local Outlier Factor. Local Outlier Factor [20] evaluates whether a point p is an outlier by comparing its local density to that of its nearest neighbors in the sliding window w. The local density of p is defined as the inverse of the average reachability distance from its k nearest neighbors. The reachability distance of p with respect to a neighbor o is the maximum of the distance between them and the distance from o to its own kth nearest neighbor. A high reachability distance suggests that p—and potentially its neighbors—are relatively isolated. The local outlier factor of p is the average ratio of its neighbors' local densities to its own. Points with a local outlier factor significantly greater than 1 are considered anomalies.

Croston's Method for Intermittent Time Series. None of the above algorithms are suitable for intermittent time series because sparse non-zero values are trivially considered anomalies. Conversely, ignoring zeros entirely does not account for the distance between non-zero values. To detect intermittency, we follow the demand pattern categorization proposed by Syntetos et al. [102], which classifies a time series as erratic, lumpy, intermittent, or smooth based on two metrics: the squared coefficient of variation CV^2 and the average demand interval (the average duration between non-zero values) p. These categories are defined by the thresholds $CV^2 = 0.49$ and $p = 1.32$. We focus only on the average demand interval; if it exceeds 1.32 over a sliding window of the previous w non-zero values, then the time series is considered intermittent.

We use Croston's method, which is designed for forecasting intermittent time series [29]. Croston's method maintains separate estimates of the demand size (non-zero values) $\hat{z}$ and the demand interval (the number of time steps between non-zero values) $\hat{p}$. At each non-zero value z, which occurs q time steps after the previous non-zero value, Croston's method updates both estimates using simple exponential smoothing with smoothing factor α (which is recommended to be small). The forecast $\hat{y}$ is given by the demand size estimate divided by the demand interval estimate:

$$\hat{z}_t = (1 - \alpha)\hat{z}_{t-1} + \alpha z_t \tag{2a}$$

$$\hat{p}_t = (1 - \alpha)\hat{p}_{t-1} + \alpha q_t \tag{2b}$$

$$\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t} \tag{2c}$$

Croston's method is known to over-forecast [101]. We address this by using the Syntetos-Boylan Approximation, which multiplies the forecast $\hat{y}$ by the correction factor $1 - \frac{\alpha}{2}$. If the difference between the point and the forecast exceeds a predefined threshold, the point is considered an anomaly.

3.2.3. End of Anomalies. After detecting the start of an anomaly, we also need to explicitly detect its end. The significance of an anomaly is determined not only by its magnitude, but also by its duration. Accurately assessing the overall impact of anomalies is essential for helping resource-constrained observatories and journalists effectively prioritize which ones warrant further attention. Additionally, anomalies need to be fully removed from the time series to prevent them from distorting future anomaly detection.

In an online anomaly detection setting, the behavior of the time series after the start of a spike is initially unknown. We first assume the time series is locally stationary—that is, it will eventually return to the pre-anomaly baseline. We score each new point with respect to the pre-anomaly sliding window w and consider the anomaly over if the score falls below the detection threshold. However, the time series may never return to the old baseline, instead stabilizing at a new normal. To handle this case, we use an *efficiency ratio*, which has been used in contexts such as finance and animal path tracking to distinguish directional trends from random fluctuations [17], [35]. The efficiency ratio is calculated between two points T_a and T_b as the net change divided by the sum of all absolute changes over the interval:

$$\textit{Efficiency Ratio} = \frac{T_b^{'} - T_a^{'}}{\sum_{i=a+1}^{b} |T_i^{'} - T_{i-1}^{'}|} \tag{3}$$

We calculate the efficiency ratio between the day before the anomaly started and the current day. Immediately after a spike is detected, the efficiency ratio is 1, reflecting a perfect trend without fluctuations. However, the efficiency ratio will approach 0 in one of two cases. First, as the time series returns closer to the old baseline, the net change will decrease. Second, over time, the movement from fluctuations will outpace the net change (which is bounded by the time series maximum). The anomaly ends when the efficiency ratio falls below a near-zero threshold. This single metric can therefore capture a variety of time series behaviors after the initial spike.

Impact Factor. Having determined the bounds of an anomaly, we can fully measure its significance. We define the *impact factor* of an anomaly as the cumulative area between the time series T and the raw detection threshold t (i.e., the minimum value the time series must reach to be considered an anomaly) from the start of the anomaly s to the end e:

$$\textit{Impact Factor} = \sum_{i=s}^{e} (T_i^{'} - t) \tag{4}$$

Higher impact factors indicate substantial and/or sustained deviations from the detection threshold. Unlike the anomaly score used to detect the start of anomalies, calculated with respect to a local sliding window w, the impact factor gives a globally comparable measure of significance.

For some anomaly detection algorithms, such as Isolation Forest and Local Outlier Factor, it is difficult to analytically invert the anomaly score threshold to calculate t because anomaly scores are not derived from a parametric model of the data, but rather from its underlying structure. For these algorithms, we approximate t by scoring a range of candidate values with respect to w and selecting the one whose score is closest to the anomaly score threshold.

3.2.4. Imputation. As mentioned previously, after detecting the end of an anomaly, it must be removed from the time series. Otherwise, high-magnitude and/or long-running anomalies can inflate the detection threshold and prevent the detection of future anomalies. The effects of anomalies on the detection threshold can be mitigated by either ignoring the anomalous points entirely or imputing (e.g., replacing or smoothing) them with less extreme values.

When an anomaly is detected with Croston's method, its effect on future forecasts is naturally dampened by the small smoothing factor α, so no further action is necessary. Otherwise, we insert the anomalous points into the sliding window w (with post-insertion mean μ and standard deviation σ), estimate the post-anomaly mean μ' and standard deviation σ', and rescale the window to match this distribution:

$$\frac{(w - \mu)\sigma'}{\sigma} + \mu' \tag{5}$$

When the time series returns to the pre-anomaly baseline, we estimate that μ' and σ' are unchanged from the pre-insertion mean and standard deviation of w. When it stabilizes at a new baseline according to the efficiency ratio, we estimate that μ' is the first value following the anomaly and that σ' is unchanged from the pre-insertion standard deviation. Applying rescaling to the entire sliding window rather than only the anomaly is particularly important in the latter case. If we limited rescaling to only the anomaly, then the sliding window would still contain an abrupt baseline shift at the start of the anomaly. The data before the shift would unduly lower the detection threshold for future points, causing normal values to be considered anomalous.

3.3. Parameter Tuning

Our anomaly detection approach requires tuning several parameters: the sliding window size w, the detection thresholds for both the primary anomaly detection algorithm and Croston's method, and the efficiency ratio threshold. Ideally, if ground truth labels for spikes in Google Trends data were available, we could tune these parameters to maximize true positives while minimizing false positives and negatives. However, in an unsupervised setting, we can only control the aggressiveness of detection. Given the resource constraints of observatories and journalists, we aim to minimize the effort required to investigate search spikes while maximizing the detection of visible spikes.

We therefore frame parameter tuning as a multi-objective optimization problem that seeks to balance two or more conflicting objectives. We leverage evolutionary optimization, which applies standard genetic algorithm operators (i.e., selection, crossover, mutation and elitism) to iteratively improve a population of randomly initialized solutions until a termination condition is met (e.g., a fixed number of iterations, a time limit, *etc.*) [31]. While genetic algorithms do not guarantee a global optimum, they tend to find high-quality solutions and are significantly more feasible than an exhaustive search.

In our case, each solution is a complete parameter set. We evaluate each solution on both objectives, using the number of spikes detected by *CenAlert* as a proxy for investigation effort and their cumulative impact factor as a measure of visibility. A solution is considered non-dominated (or Pareto-optimal) if no other solution performs better on both objectives [31]. The set of objective values corresponding to non-dominated solutions—known as the Pareto front—represents the optimal trade-offs between minimizing investigation effort and maximizing spike visibility. Moving along the Pareto front necessarily improves one objective at the cost of the other. Selecting a single preferred solution from the Pareto front ultimately depends on the priorities and constraints of each observatory. Due to the diversity in time series across countries, no single set of parameters performs well universally. We therefore run parameter tuning on a per-country basis.

4. Evaluation

4.1. Running *CenAlert*

Evaluating *CenAlert* requires concrete choices for several system parameters, including the countries, topic, and time frame for which to source data, the number of downloads to mitigate variance, and the anomaly detection algorithm and its associated parameters.

Country Selection. We limited our evaluation of *CenAlert* to countries known to implement censorship. We first compile datasets of known censorship events from four organizations: Access Now, Pulse, OONI, NetBlocks. Access Now and Pulse maintain structured datasets (i.e.,

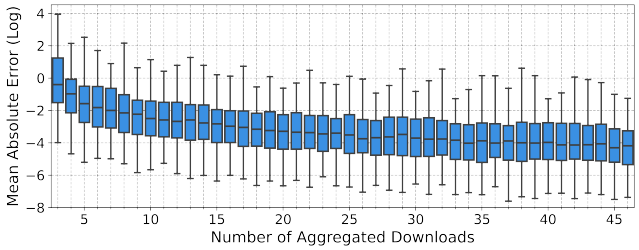

Figure 2: **Number of Downloads to Aggregate.** We stitch time series from iteratively aggregated downloads of Google Trends data. Here, we show the the natural log of the mean absolute error (MAE) between successive iterations. Across all countries, the MAE eventually becomes negligible.

	Anom Overlap	Prop Agree	Mean	Median
Z-Score	**0.79**	**0.87**	**0.01**	**0.005**
Median Method	0.68	0.56	0.008	0.005
Isolation Forest	0.74	0.88	0.011	0.007
Local Outlier Factor	0.74	0.66	0.008	0.005

TABLE 1: **Algorithm Agreement.** For each algorithm, we calculate the proportion of its anomalies also detected by at least one other algorithm (Anom Overlap) and the proportion of all inter-algorithm agreements in which it participates (Prop Agree). We select Z-Score as the overall best performer. We also show the mean and median anomaly rates. For Z-Score, these are 0.01 (51.1 days per country) and 0.005 (25.5 days), respectively.

the #KeepItOn STOP database [6] and Pulse Shutdowns Tracker [53]) that can be automatically downloaded. For OONI and NetBlocks, we manually read their reports [76], [82] and recorded the date, geographic scope, and targets of mentioned censorship events. Because circumvention tools require Internet connectivity, we removed censorship events where indiscriminate bandwidth throttling or total network shutdowns denied Internet access entirely. We selected the 76 countries with at least one blocking event.

Time Period. While Google Trends data spans back to 2004, we restricted our analysis to January 2011 through December 2024. Notably, the Arab Spring in 2011 marked a turning point in the global use of Internet censorship, extending beyond traditional offenders like China and Iran, and catalyzed the development of dedicated censorship measurement observatories [64].

Topic Selection. *CenAlert* fundamentally relies on the assumption that the chosen topic is a reasonable proxy for censorship events. To test this, we evaluated control topics unrelated to censorship and found no strong association with censorship events (see Appendix A.1). We considered several topics related to censorship circumvention, including <Virtual Private Network>, <Proxy Server>, <Internet Censorship Circumvention>, <Tor>, and <Psiphon>. We ultimately selected the <Virtual Private Network> topic due to the widespread popularity of VPNs as a circumvention tool and because alternative topics suffer from lower data quality, including limited availability across countries and extreme sparsity (see Appendix A.2).

Number of Downloads. We collected 45 downloads of each time series. Before stitching them together to reconstruct daily data from 2011 through 2024, we iteratively aggregated increasing numbers of downloads. Figure 2 shows the distribution, across all countries, of the mean absolute error (MAE) on nonzero points between each successive pair of stitched time series. When using 45 downloads, the MAE is negligible, with a maximum of 0.28.

Algorithm Parameters. For each country and algorithm, we run parameter tuning for 5,000 iterations. We select the knee point of the Pareto front, where marginal gains in one objective significantly cost the other. To more reliably identify the knee point, we first fit a smooth continuous function to the discrete Pareto front. However, curve fitting requires assumptions about the general shape of the data. Ideally, each additional spike yields diminishing returns in visibility, as high impact spikes are typically detected first. Additionally, in practice, multi-objective optimization *minimizes* objective functions, so maximizing cumulative impact factor is equivalent to minimizing its negative value; as a result, the cumulative impact factor *decreases* with the number of detected spikes. Together, these considerations suggest that the Pareto front should roughly follow a convex decreasing shape.

We therefore attempt to fit one of five candidate convex decreasing functions—exponential decay, power law decay, reciprocal, negative logarithm, and inverse square root—to the Pareto front and use the coefficient of determination (i.e., R^2) to assess goodness of fit. If no candidate has a sufficiently high R^2 value (we use 0.95, close to the maximum R^2 value of 1), we assume the Pareto front is not convex decreasing across its entire domain and instead fit a high-degree polynomial. We then use the Kneedle algorithm [95] to find the first knee point on the curve evaluated only at the objective values on the Pareto front. If there is no knee point, we conservatively default to the solution that minimizes the number of spikes.

Algorithm Selection. Effort and visibility are measures of algorithm behavior, but neither is a sufficient measure of algorithm *quality*; two algorithms can detect similar numbers of spikes with comparable impact factors while flagging entirely different time ranges. However, evaluating the quality of unsupervised algorithms is challenging because we lack ground truth labels for when spikes occur. We therefore assume that spikes independently flagged by multiple, methodologically diverse anomaly detection algorithms are more likely to warrant further investigation, while those flagged by only a single algorithm may reflect oversensitivity.

We used the per-country parameters tuned for each algorithm to generate its final set of detected spikes. We then excluded spikes detected by Croston's method and evaluated each algorithm on two metrics: (1) the proportion of its spikes also detected by at least one other algorithm,

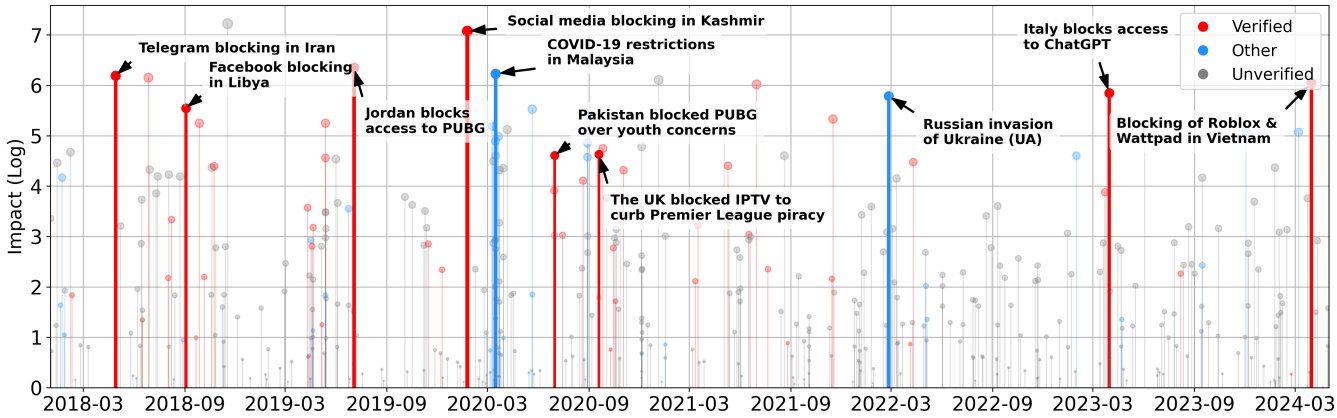

Figure 3: **Subset of Spikes Detected from 2018 to 2024.** We show manually verified censorship events (*Verified*), explainable events that are not censorship (*Other*), and unverified events that we either did not examine or for which we could not determine a cause (*Unverified*). We label several spikes among the 100 highest impact, finding diverse causes ranging from blocking of social media and messenger applications in Iran, Libya, and India to blocking of video games in Jordan and Vietnam.

and (2) the proportion of inter-algorithm agreements (i.e., spikes detected by at least two algorithms) it participates in—indicating whether it misses spikes consistently detected by other algorithms. We say multiple algorithms agree on a spike if each identifies it in the same country on the same day. Table 1 shows the results for each algorithm. We ultimately selected Z-Score as the underlying spike detection algorithm for *CenAlert*, as it had the highest proportion of spikes also detected by at least one other algorithm (79%) and ranked a close second in the number of inter-algorithm agreements it participates in (87%).

4.2. Verifying Detected Spikes

We leverage *community-reported* censorship events retrieved from the four aforementioned sources to automatically match search spikes to them. We attribute a spike to censorship if its start date falls within close proximity to the start date of a known event. We tolerate a difference of up to six days for practical considerations, such as uncertainty in the exact start date of some events, inconsistencies in the timezone used when reporting, delays between legal or court mandates of blocking and its technical implementation, or late-night blocking not widely realized until the next day.

However, as discussed earlier, there is no comprehensive ground truth list of censorship events that we can automatically match against. Thus, for spikes that do not coincide with any community-reported events, we perform a best-effort manual verification process to determine their likely cause. We first examine the rising topics and queries during its date range and check for mentions of censorship or censorship circumvention (e.g., Block, Internet censorship circumvention), common targets of blocking (e.g., Facebook, Telegram), or politically sensitive events when censorship tends to increase (e.g., Election contest) [93].

We then use these terms to guide online searches for news articles or social media posts that explicitly mention that certain platforms or protocols were blocked. If we cannot confirm censorship, we investigate alternative explanations, including politically sensitive events where users may expect censorship but none is reported to occur, or Internet disruptions such as geoblocking or server outages.

5. Results

We demonstrate the utility of *CenAlert* to the Internet freedom community by investigating high impact spikes, all spikes in highly censoring countries, and the total volume of spikes to be verified.

5.1. Highest Impact Spikes

We examine the 100 highest impact spikes across all countries considered, finding that 58 countries are represented and 91 spikes are explainable. Each *explainable spike* is attributed to a known censorship event reported by at least one of the four aforementioned monitoring organizations, a censorship event we manually confirmed from other sources, or a non-censorship event that drives VPN interest (e.g., political events, georestrictions, *etc.*). Figure 3 highlights several manually verified spikes. A list of these spikes and their explanations is available in Table 4 in the Appendix.

Censorship Events. The majority (76) of the highest impact spikes coincide with censorship events, of which 52 were previously reported and 24 were manually verified. In many cases, blocking is politically motivated. In India, *CenAlert* captured a spike in January 2020 following the end of a six-month total shutdown in Jammu and Kashmir, which has been heavily affected by Internet shutdowns since India revoked its autonomy. Although connectivity was restored, officials allowed access to just 301 domains [121].

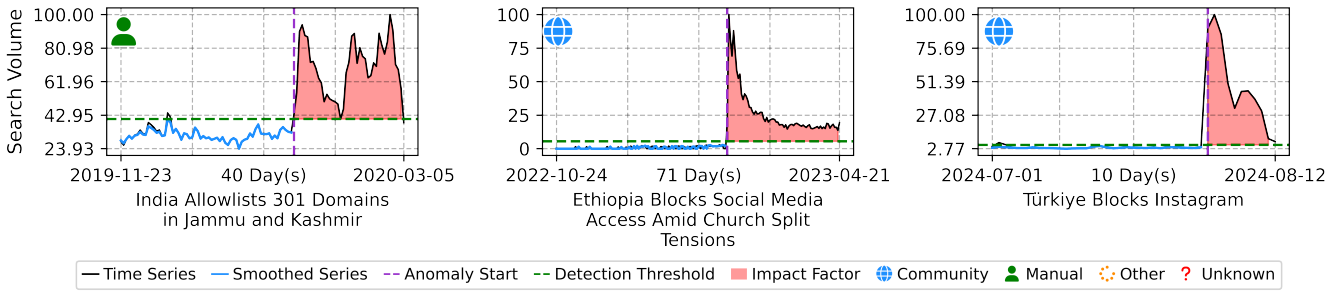

Figure 4: **Notable Search Spikes Detected by *CenAlert*.** India restored minimal Internet access in Jammu and Kashmir following a six-month total shutdown, allowing access to just 301 domains [121]. Ethiopia blocked social media amid a rift within the Orthodox Church following accusations of ethnic discrimination and calls for protests [120]. Türkiye blocked Instagram for failing to remove illegal content [43].

In Ethiopia, *CenAlert* detected a spike in February 2023 coinciding with social media blocking to quell unrest over accusations of ethnic discrimination within the Orthodox Church [120]. In Türkiye, *CenAlert* identified a spike in August 2024 in response to the blocking of Instagram. While officials later claimed that the platform violated Turkish law, reports suggested that the move was triggered by Instagram's removal of condolence messages for the recently assassinated leader of Hamas [43]. We show *CenAlert*'s detection of these three events in Figure 4. We also observe blocking for other reasons. For example, Skype was blocked in the United Arab Emirates in December 2017 because VoIP services can only be provided by licensed telecommunications providers [51], while the online game PUBG was banned in both Jordan and Pakistan for its violent content [14], [103].

Alternative Explanations. Beyond censorship events, we attribute a number of alternative explanations to 15 of the remaining spikes. In 5 cases, the spikes come at the onset of COVID-19 lockdowns. Increased reliance on the Internet for remote work and entertainment heightened the demand for VPNs [68]. In 3 cases, spikes immediately followed the announcement of legislation users believed threatened their security and privacy online. The highest impact example occurred in Hong Kong in May 2020, when China proposed a national security law criminalizing secession and subversion, among other acts, to end anti-extradition bill protests [16]. We also observe 3 cases of users attempting to access georestricted content, as well as 1 regional outage.

Finally, we see 3 spikes coinciding with politically sensitive events—including war and protests—that are common causes of censorship, although we could not confirm blocking by the relevant country at the start of the spike. One spike occurs for Ukraine the day after Russia's invasion, which was accompanied by cyberattacks and destruction of Internet infrastructure [75], as well as intensified censorship in Crimea, subject to Russian Internet regulations since 2014 [40]. Regardless of whether censorship occurred during these events, it is still valuable for the Internet freedom community to monitor them to ensure transparency and accountability [5].

Unexplainable Spikes. Of the 9 unexplainable spikes, over half occur in countries known for extreme censorship and/or suppression of independent journalism, including Turkmenistan (3), Myanmar (1), and Uzbekistan (1) [2], [79]. These spikes may represent censorship events that did not receive media attention.

5.2. Highly Censoring Countries

We investigate how VPN search spikes correlate with censorship events in nine countries designated as "Not Free" in the Freedom on the Net 2024 report [2] that have the highest numbers of recorded censorship events by observatories. Figure 6 shows the results of our verification. In total, we attribute explanations to 126 spikes, of which 101 correspond to community-reported or manually verified censorship events. Across these countries, the vast majority (96% on average) of the cumulative impact factor is explainable, mostly as censorship (91.2% on average).

Censorship in Egypt appears to receive limited international attention, as we manually verified four major blocking events from local media reports or social media posts: disruptions to VoIP services in April 2017 [34], blocking of pornography websites in December 2017 and October 2018 [1], [109], and blocking of EgyBest and other popular piracy websites in May 2019 [7]. The fact that, after manual verification, most of the cumulative impact factor for Egypt is explainable as censorship highlights *CenAlert*'s potential to increase coverage of censorship reporting by the Internet freedom community.

Azerbaijan is the only country for which all spikes detected by *CenAlert* are explainable as censorship, although it also has the fewest spikes overall. In both cases, spikes coincide with community-reported social media blocking events in September 2020 and September 2023, which occurred amid territorial conflicts with Armenia [44], [118].

We are able to attribute a cause to the most spikes for Türkiye, where 25 of 28 explainable spikes coincide with censorship events. In many cases, Türkiye blocked prominent social media platforms to suppress coverage during politically sensitive events. For example, we capture the

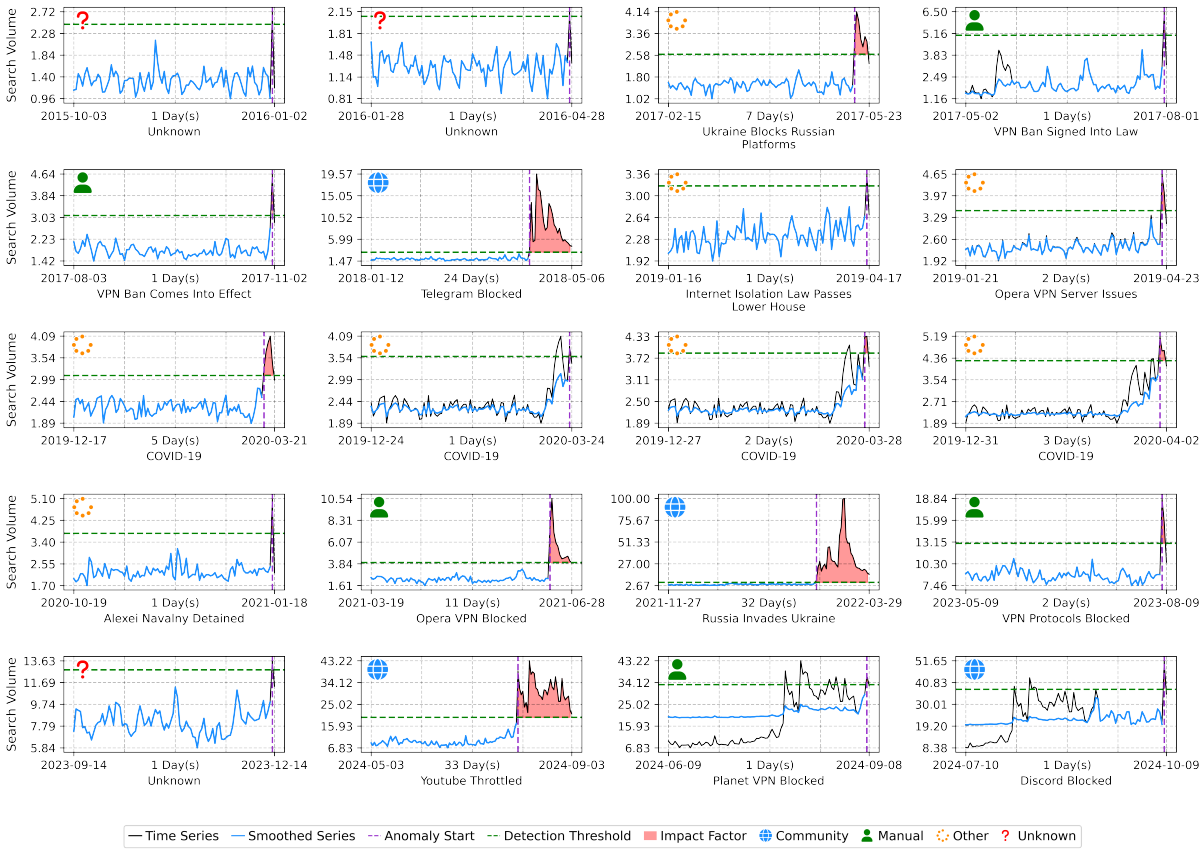

Figure 5: **All Detected Search Spikes in Russia.** We detect notable events such as the blocking of Telegram in April 2018 [62], Russia's invasion of Ukraine in February 2022—which was accompanied by both censorship and geoblocking [87]—and the throttling of YouTube in August 2024 [66].

blocking of Twitter in March 2014 for failing to remove links to audio recordings alleging corruption among officials [89], the blocking of social media in November 2016 amid demonstrations over the arrest of opposition party leaders [108], and the blocking of social media in February 2020 following an attack on Turkish soldiers in Syria [74].

We also achieve explainability for over 10 spikes in each of Ethiopia (13), Iran (12), Kazakhstan (12), Pakistan (23), and Russia (17). While spikes in Ethiopia and Pakistan largely reflect social media blocking, many of those in Iran, Kazakhstan, and Russia have more unique causes and are perhaps more difficult for active measurement to comprehensively capture. For instance, in Kazkahstan, we find two spikes aligned with blocking of regionally popular mobile applications: InDriver, a taxi service, in April 2016 [42] and GetContact, an app that identifies (spam) callers, in February 2018 [59]. In both Russia and Iran, we observe multiple spikes when circumvention tools were restricted. Iran blocked the PPTP VPN protocol in September 2011 [18], VPNs not registered with the government in March 2013 [71], and Psiphon in February 2014 [45]. Similarly, Russia blocked VPNs not registered with the government in November 2017 [104], Opera VPN and VyprVPN in June 2021 [113], and common VPN protocols (e.g., IPsec, OpenVPN, and Wireguard) in August 2023 [110].

5.3. *CenAlert* for Global Monitoring

We evaluate *CenAlert* on two crucial qualities of a practical detection system: early detection of potential censorship events and manageability of alert volume.

Early Detection. We first consider when *CenAlert* detects the 191 explainable spikes evaluated in Sections 5.1 and 5.2. In 108 cases, detection occurred the same day as the underlying event. In 54 others, the spike is detected either the day before (15) or the day after (39). Interestingly, *CenAlert* detects a couple of spikes where users reacted to potential or planned blocking multiple days in advance. For example, we capture a spike in Uganda that started four days before and peaked on the day that ISPs blocked social media platforms to enforce a tax on their use in July 2018 [119]. While prior work has criticized Google Trends data for being primarily influenced by media coverage [25], we find that spikes in circumvention-related searches align with meaningful events.

Manageability. Next, we consider the manageability of using *CenAlert* for organizations operating at global

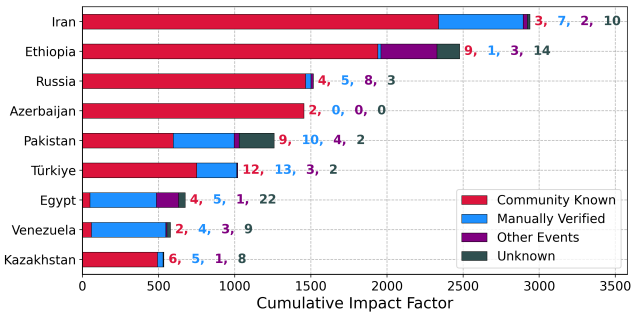

Figure 6: **Explainable Impact in Restrictive Countries.** Half of the spikes and the vast majority of their cumulative impact factor for highly restrictive countries are explainable as censorship events, either previously documented by observatories or manually verified.

and regional scales. The organizations from which we sourced community-reported censorship events all monitor at a global scale. From 2020 to 2024, *CenAlert* outputs fewer than 100 spikes per year on average. This is well within what the Internet freedom community can handle. For comparison, from 2016 to 2023, Access Now recorded an average of 169 and maximum of 253 censorship events per year, while Pulse recorded an average of 132 and maximum of 200 censorship events per year from 2019 to 2024.

Because censorship investigation efforts are often prompted by local NGOs, we also investigate the regional effort required to respond to *CenAlert*. We adopt the regions used by Access Now's #KeepItOn STOP database: Africa, Asia-Pacific (APAC), Eastern Europe and Central Asia (EECA), Europe, Latin America and the Caribbean (LAC), and the Middle East and North Africa (MENA). Five countries without censorship events in the STOP database were manually assigned to the appropriate region.

The number of spikes output by *CenAlert* for each region over progressively longer intervals (each ending on December 31, 2024) is shown in Table 2. The most spikes are output for the APAC and MENA regions in all intervals. These two regions contain several countries that consistently rank among the most restrictive, including 4 (Azerbaijan, Iran, Türkiye, and Pakistan) of the 9 examined in Section 5.2, as well as Myanmar and India [2], [93]. Over the past two years, organizations in these regions would have monitored a monthly average of just 1.79 spikes. All six regions registered at least 5 spikes within the past 6 months and at least 10 spikes within the past 2 years. Overall, the number of spikes to investigate across all regions is low.

6. Scope of *CenAlert*

CenAlert and existing censorship monitoring efforts fundamentally differ in scope. *CenAlert* does not measure censorship directly—that is, by detecting the censorship response within network traffic—but rather *changes in user behavior driven by experiences or expectations of Internet restrictions*. *CenAlert* best detects the onset of nationwide blocking of popular, irreplaceable services, which in turn impact broad segments of the population.

	Africa	APAC	Europe	EECA	LAC	MENA
Last 1 Month	2	1	0	1	0	1
Last 3 Months	4	6	1	4	0	4
Last 6 Months	5	13	6	8	6	9
Last 12 Months	10	25	9	12	9	21
Last 24 Months	21	43	14	26	12	43

TABLE 2: **Spike Volume by Region.** We show the breakdown of spikes output by *CenAlert* per region for several intervals ending on December 31, 2024. Organizations operating in the APAC and MENA regions would monitor the most spikes.

Accordingly, *CenAlert* is unlikely to detect blocking that disproportionately impacts marginalized or minority groups, such as community-reported events where blocked websites were LGBT-inclusive (15) or focused on human rights (5). Similarly, due to their relatively limited use on a national scale, *CenAlert* rarely detects censorship targeting specific news media (36) and political criticism websites (17).

To further understand the limits of *CenAlert*'s detection, we carefully investigated all community reported events. We prioritized the analysis of social media censorship, which accounts for nearly two-thirds of all community-reported events, to reason about *CenAlert*'s limitations even when widely used services are blocked.

Blocking confined to subnational regions may affect too few people to influence nation-level search patterns. Of the 43 community-reported regional blocking events, 69.8% (30) were undetected, over half of which (20) occurred in India, where blocking is often limited to smaller subdivisions such as districts or cities. In contrast, 87.1% of detected community-reported events involved national-level blocking.

Blocking may be too short-lived for users to notice or react. Of 83 events whose longest uninterrupted block lasted less than a day, 84.3% (70) were undetected. Venezuela is the largest contributor, with 30 events from 2019 to 2020 where social media and streaming platforms were blocked for 12 minutes to 20 hours during public appearances of opposition figure Juan Guaidó [72]. Conversely, 87.1% of detected community-reported events last multiple days.

Frequently recurring or waves of censorship events may not produce distinct spikes for each restriction. Recurring events may diminish circumvention interest if users have already installed circumvention tools, while waves of related events may produce a single large spike that encompasses multiple restrictions. We found 17 undetected community-reported events linked to earlier ones that were detected. In Jordan, Facebook livestreaming was blocked weekly during anti-austerity protests beginning in November 2018 [63] and during Teachers' Union protests beginning in July 2020 [73]; *CenAlert* detected only the first two instances of the former and the first of the latter. In Senegal, several social media platforms were blocked on June 1, 2023 amid protests, followed a few days later by a TikTok ban [60];

CenAlert detected a single large spike starting at the initial restrictions and encompassing TikTok's blocking.

Finally, the availability of alternative platforms may reduce motivation to circumvent censorship. For example, *CenAlert* does not detect Iran's blocking of Signal in January 2021 [69] or Russia's blocking of Viber in December 2024 [91]. However, Telegram is the dominant messaging app in Iran [69], while WhatsApp and Telegram lead in Russia [3].

7. Discussion

***CenAlert* for Free Countries.** In countries with minimal censorship, circumvention-related spikes are seemingly unlikely to be caused by website blocking, but rather reflect other events affecting VPN use. To further characterize these events, we chose 16 countries considered "Free" in the 2024 Freedom on the Net Report and were not among the 76 selected in Section 4.1, and ran *CenAlert* on time series stitched from 25 downloads of each window. Because there does not exist a comprehensive list of non-censorship events that could drive VPN interest, we performed a best-effort manual investigation following the procedure in Section 4.2.

CenAlert detects blocking of piracy websites in both Australia and the Netherlands [36], [90], as well as government seizure of piracy websites in France [94]. Furthermore, some of the highest-impact explainable spikes for multiple countries are associated with measures enabling or implying increased surveillance. For example, the highest-impact spike in the United States coincides with a congressional vote to repeal privacy protections preventing ISPs from selling user data to advertisers without consent [98]. In Canada, the second-largest spike (after COVID-19) coincides with the introduction of the "notice-and-notice" policy, which compels ISPs to forward notices of copyright infringement from content creators to end users [32]. Even if not censorship, these cases follow the same pattern observed in Sections 5.1 and 5.2, where high-impact spikes often coincide with government regulation of the Internet.

Resilience of *CenAlert*. Like any system, the utility of *CenAlert* is limited by the quality of its underlying data. We already use Croston's method to detect spikes in sparse data. A more fundamental threat would arise if Google search itself were blocked. However, Google Trends appears to be resilient even when direct access to Google search is disrupted. For example, during Russia's Telegram ban in April 2018—which extended to 19 million IP addresses, including some used by Google search [62]—we still detect the third-highest impact spike in Russia. Similarly, we detect a spike in Kazakhstan in February 2020, despite blocking of Google services [111].

8. Conclusion

In this work, we investigate the potential of Google Trends to aid Internet freedom efforts. We build *CenAlert*, a user-driven alerting system that identifies spikes in censorship-related searches (e.g., VPNs). Our evaluation of *CenAlert* across 76 countries from 2011 to 2024 demonstrates that it is effective and practical. Considering both high-impact spikes and all spikes in highly restrictive countries, we find that most are explainable, primarily as censorship events. *CenAlert* detects these spikes early, and it maintains manageable alert volumes both globally and regionally. Our work highlights the value of industry data in supporting the Internet freedom community, and we hope it inspires further efforts to repurpose other forms of telemetry data for censorship event detection. We make *CenAlert* accessible to the community along with a dashboard, API, and Slack webhook for notification, visualization, and analysis of censorship events. Our public dashboard is available for use at https://dashboard.censoredplanet.org/cenalert.html.

We have open-sourced two versions of *CenAlert*: the implementation used for evaluating historical Google Trends data in this work (available at https://github.com/censoredplanet/cenalert-paper) and the implementation that underpins our dashboard (available at https://github.com/censoredplanet/cenalert).

9. Acknowledgements

The authors would like to thank Hieu Le and Brennen Daudlin for their contributions to this project, as well as Sudeepti Rao for her work on the design and development of *CenAlert*'s public dashboard. We also appreciate the constructive feedback provided by the anonymous reviewers. This material is based upon work supported by the National Science Foundation under Grant Numbers CNS-2237552 and CNS-2452883.

Ethics Considerations

CenAlert is a passive monitoring system built on top of Google Trends data, which spans back to 2004 and has been publicly available since 2008 [47]. Search data is reported only for popular terms, is anonymized and aggregated to protect user privacy [46], and has been widely used in prior research across multiple disciplines [8], [21], [22], [27], [77], [80], [107]. The Google Trends team is aware of our use of their data for *CenAlert*. Additionally, we have presented this work to many organizations within the Internet freedom community—including OONI, IODA, and Access Now—and have incorporated feedback to meet their needs.

References

[1] Internet Megathread. https://www.reddit.com/r/Egypt/comments/96hn1m/internet_megathread/, 2018.

[2] Freedom On The Net 2024, 2024. https://freedomhouse.org/sites/default/files/2024-10/FREEDOM-ON-THE-NET-2024-DIGITAL-BOOKLET.pdf.

[3] Mediascope: WhatsApp remains the most popular messenger in the Russian Federation. https://tass.ru/ekonomika/22390143, 2024.

[4] Access Now. https://www.accessnow.org/.

[5] Access Now. 2025 Elections AND Internet Shutdowns Watch. https://www.accessnow.org/campaign/2025-elections-and-internet-shutdowns-watch/, 2025.

[6] AccessNow. Shutdown Tracker Optimization Project. https://www.accessnow.org/guide/shutdown-tracker-optimization-project/.

[7] Al-Masry Al-Youm. EgyBest shuts down amidst purge of piracy websites. https://cloudflare.egyptindependent.com/egybest-shuts-down-amidst-purge-of-piracy-websites/, 2019.

[8] Cristiano Alicino, Nicola Luigi Bragazzi, Valeria Faccio, Daniela Amicizia, Donatella Panatto, Roberto Gasparini, Giancarlo Icardi, and Andrea Orsi. Assessing ebola-related web search behaviour: insights and implications from a study of google trends-based query volumes. *Infectious diseases of poverty*, 2015.

[9] Brett G Amidan, Thomas A Ferryman, and Scott K Cooley. Data outlier detection using the Chebyshev theorem. In *2005 IEEE Aerospace Conference*, 2005.

[10] Amnesty International. Laws Designed to Silence: The Global Crackdown on Civil Society Organizations. https://www.amnestyusa.org/wp-content/uploads/2019/02/Laws-designed-to-silence_final_web-version.pdf, 2019.

[11] Amnesty International. Surveillance giants: How the business model of Google and Facebook threatens human rights. https://www.amnesty.org/en/documents/pol30/1404/2019/en/, 2019.

[12] Amnesty International. What is Big Tech's surveillance-based business model? https://www.amnesty.org/en/latest/campaigns/2022/02/what-is-big-techs-surveillance-based-business-model/, 2022.

[13] AppCensorship. Russian Government Escalates War on VPNs and Censorship Circumvention Tools. https://appcensorship.org/news/russian-government-escalates-war-on-vpns-and-censorship-circumvention-tools, 2025.

[14] Ary News. Ban to remain enforced over PUBG in Pakistan: PTA, 2020. https://arynews.tv/ban-remain-enforced-over-pubg-pakistan-pta/.

[15] Sabyasachi Basu and Martin Meckesheimer. Automatic outlier detection for time series: an application to sensor data. *Knowledge and Information Systems*, 2007.

[16] BBC. The Hong Kong protests explained in 100 and 500 words, 2019. https://www.bbc.com/news/world-asia-china-49317695.

[17] Simon Benhamou. How to reliably estimate the tortuosity of an animal's path:: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, 2004.

[18] bestvpnservice.com. Reports: Iran Blocked PPTP and L2TP VPN Protocols. SSTP VPN is Unstable. https://web.archive.org/web/20111001041241/bestvpnservice.com/blog/reports-iran-blocked-pptp-and-l2tp-vpn-protocols-sstp-vpn-is-instable, 2011.

[19] Zachary S. Bischof, Kennedy Pitcher, Esteban Carisimo, Amanda Meng, Rafael Bezerra Nunes, Ramakrishna Padmanabhan, Margaret E. Roberts, Alex C. Snoeren, and Alberto Dainotti. Destination Unreachable: Characterizing Internet Outages and Shutdowns. In *ACM SIGCOMM*, 2023.

[20] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 2000.

[21] Herman Anthony Carneiro and Eleftherios Mylonakis. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases*, 2009.

[22] Yan Carrière-Swallow and Felipe Labbé. Nowcasting with google trends in an emerging market. *Journal of Forecasting*, 2013.

[23] Eduardo Cebrián and Josep Domenech. Addressing google trends inconsistencies. *Technological Forecasting and Social Change*, 2024.

[24] Censored Planet. Censored planet. https://censoredplanet.org/.

[25] Gianfranco Cervellin, Ivan Comelli, and Giuseppe Lippi. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *Journal of epidemiology and global health*, 2017.

[26] Chung Chen and Lon-Mu Liu. Forecasting time series with outliers. *Journal of Forecasting*, 1993.

[27] Hyunyoung Choi and Hal Varian. Predicting the present with google trends. *Economic record*, 2012.

[28] Cloudflare. Cloudflare radar, 2020. https://radar.cloudflare.com/.

[29] J. D. Croston. Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly (1970-1977)*, 1972.

[30] George Danezis. An anomaly-based censorship-detection system for tor. https://research.torproject.org/techreports/detector-2011-09-09.pdf.

[31] Kalyanmoy Deb. Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, 2011.

[32] Ernesto Van der Sar. Canadian ISPs and VPNs Now Have to Alert Pirating Customers. https://torrentfreak.com/canadian-isps-vpns-now-alert-pirating-customers-150102/, 2015.

[33] Vera Z Eichenauer, Ronald Indergand, Isabel Z Martínez, and Christoph Sax. Obtaining consistent time series from Google Trends. *Economic Inquiry*, 2022.

[34] Mohamed Alaa El-Din. NTRA denies blocking VOIP services in Egypt. https://www.dailynewsegypt.com/2017/04/22/ntra-denies-blocking-voip-services-egypt/, 2017.

[35] Craig A. Ellis and Simon A. Parbery. Is smarter better? A comparison of adaptive, and simple moving average trading strategies. *Research in International Business and Finance*, 2005.

[36] European Digital Rights. Dutch Internet providers forced to block The Pirate Bay. https://edri.org/our-work/edrigramnumber10-1dutch-isps-block-piratebay/, 2012.

[37] Kinga Farkas. CUSUM Anomaly Detection. https://www.measurementlab.net/publications, 2016.

[38] Arturo Filastò and Jacob Appelbaum. OONI: Open Observatory of Network Interference. In *FOCI*, 2012.

[39] Arturo Filastò, Maria Xynou, Ramakrishna Padmanabhan, and Alberto Dainotti. Investigating internet shutdowns through mozilla telemetry, 2021. https://ooni.org/post/2021-investigating-internet-shutdowns-mozilla-telemetry/.

[40] Romain Fontugne, Ksenia Ermoshina, and Emile Aben. The Internet in Crimea: a Case Study on Routing Interregnum. In *2020 IFIP Networking Conference (Networking)*, 2020.

[41] Freedom House. The Spread of Anti-NGO Measures in Africa: Freedoms Under Threat. https://freedomhouse.org/report/special-report/2019/spread-anti-ngo-measures-africa-freedoms-under-threat, 2019.

[42] Venus Gaifootdinova. In Indriver ready to challenge blocking in Kazakhstan through court. https://forbes.kz/articles/indriver_gotova_osporit_v_sude_blokirovku_v_kazahstane, 2016.

[43] Arzu Geybullayeva. Turkey blocks access to Instagram. https://globalvoices.org/2024/08/02/turkey-blocks-access-to-instagram/.

[44] Arzu Geybullayeva, Maria Xynou, and Arturo Filastò. Media censorship in Azerbaijan through the lens of network measurement. https://ooni.org/post/2021-azerbaijan/, 2021.

[45] Nariman Gharib. [liberationtech] Many VPNs and Psiphon are currently blocked in Iran right now. https://liberationtech.stanford.narkive.com/DdXOvpcv/many-vpns-and-psiphon-are-currently-blocked-in-iran-right-now, 2014.

[46] Google. FAQ about Google Trends data. https://support.google.com/trends/answer/4365533.

[47] Google. Google trends. https://trends.google.com/trends/.

[48] Google. Google trends for marketers in a dynamic environment. https://support.google.com/google-ads/answer/9817630.

[49] Google. Traffic and disruptions to google, 2022. https://transparencyreport.google.com/traffic/overview.

[50] Andreas Guillot, Romain Fontugne, Philipp Winter, Pascal Merindol, Alistair King, Alberto Dainotti, and Cristel Pelsser. Chocolatine: Outage detection for internet background radiation. In *TMA*, 2019.

[51] Gulf Business. UAE blocks Skype amid internet calling uncertainty, 2017. https://gulfbusiness.com/uae-blocks-skype-amid-internet-calling-uncertainty/.

[52] William R Hobbs and Margaret E Roberts. How sudden censorship can increase access to information. *American Political Science Review*, 2018.

[53] Internet Society Pulse. Global Internet Shutdowns. https://pulse.internetsociety.org/shutdowns.

[54] IODA. Internet freedom ioda user guide + internet shutdown rapid response protocol. https://ioda.inetintel.cc.gatech.edu/reports/ioda-user-guide-rapid-response-protocol/, 2024.

[55] IODA. IODA — Monitor Macroscopic Internet Outages in Near Real-Time, 2025. https://ioda.inetintel.cc.gatech.edu/.

[56] Simin Kargar, Keith McManamen, and Jacob Klein. Early Detection of Censorship Events with Psiphon Network Data. In *The Eighteenth International Conference on Networks*, 2019.

[57] Juliana Kim. ChatGPT is temporarily banned in Italy amid an investigation into data collection. https://www.npr.org/2023/03/31/1167491843/chatgpt-italy-ban-openai-data-collection-ai, 2023.

[58] Ege Cem Kirci, Martin Vahlensieck, and Laurent Vanbever. "Is my Internet down?": Sifting through User-Affecting Outages with Google Trends. In *IMC*, 2022.

[59] Almaz Kumenov. Kazakhstan Bans Snooping App Over Privacy Concerns. https://eurasianet.org/kazakhstan-bans-snooping-app-over-privacy-concerns, 2018.

[60] Laura Schwartz-Henderson, David Belson, Zach Rosson, Felicia Anthonio, Maria Xynou, Arturo Filastò. Senegal: Social media blocks and network outages amid political unrest. https://ooni.org/post/2023-senegal-social-media-blocks/, 2023.

[61] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation Forest. In *2008 Eighth IEEE International Conference on Data Mining*, 2008.

[62] Ingrid Lunden. Google confirms some of its own services are now getting blocked in Russia over the Telegram ban. https://techcrunch.com/2018/04/22/google-confirms-some-of-its-own-services-are-now-getting-blocked-in-russia-over-the-telegram-ban/.

[63] Issa Mahasneh and Simone Basso. Jordan: Measuring Facebook live-streaming interference during protests. https://ooni.org/post/jordan-measuring-facebook-interference/, 2019.

[64] Maria Xynou. Year in Review: OONI in 2023, 2023. https://ooni.org/post/2023-year-in-review/.

[65] Andy Maxwell. High Court Grants Premier League IPTV Blocking Order, Fans Beg For More Legal Options. https://torrentfreak.com/high-court-grants-premier-league-iptv-blocking-order-fans-beg-for-more-legal-options-200903/, 2020.

[66] Meduza. 'No decent alternative' As the Kremlin cracks down on YouTube, Russians cancel Internet contracts and organize protests. https://meduza.io/en/feature/2024/08/13/no-decent-alternative, 2024.

[67] Maureen Meyer. CState Department's Restructuring and Proposed Budget Cuts Roll Back U.S. Role in Promoting Democracy and Human Rights Globally; Congress Has a Chance to Fix This. https://www.wola.org/analysis/state-department-budget-cuts-us-democracy-human-rights/, 2025.

[68] Simon Migliano. VPN Demand: Global Statistics in 2020. https://www.top10vpn.com/research/vpn-demand-statistics-2020/, 2020.

[69] Maziar Motamedi. Iran blocks Signal messaging app after WhatsApp exodus. https://www.aljazeera.com/news/2021/1/26/iran-blocks-signal-messaging-app-after-whatsapp-exodus, 2021.

[70] Zubair Nabi. Censorship is futile. *arXiv preprint arXiv:1411.0225*, 2014.

[71] Garrett Nada. Iran Blocks Bypass of Internet Filter. https://web.archive.org/web/20241024201812/https://iranprimer.usip.org/blog/2013/mar/11/iran-blocks-bypass-internet-filter, 2013.

[72] NetBlocks. Streaming services restricted in Venezuela as Guaidó holds press conference in Caracas. https://netblocks.org/reports/streaming-services-restricted-in-venezuela-as-guaido-holds-press-conference-in-caracas-xyMGQGAZ, 2019.

[73] NetBlocks. Facebook Live streams restricted in Jordan during Teachers' Syndicate protests. https://netblocks.org/reports/facebook-live-streams-restricted-in-jordan-during-teachers-syndicate-protests-XB7K1xB7, 2020.

[74] NetBlocks. Social media blocked in Turkey as Idlib military crisis escalates. https://netblocks.org/reports/social-media-blocked-in-turkey-as-idlib-military-crisis-escalates-r8VWGXA5, 2020.

[75] NetBlocks. Internet disruptions registered as Russia moves in on Ukraine. https://netblocks.org/reports/internet-disruptions-registered-as-russia-moves-in-on-ukraine-W80p4k8K, 2022.

[76] NetBlocks. NetBlocks Reports. https://netblocks.org/reports, 2025.

[77] Le TP Nghiem, Sarah K Papworth, Felix KS Lim, and Luis R Carrasco. Analysis of the capacity of google trends to measure interest in conservation topics and the role of online news. *PloS one*, 2016.

[78] Helmi Noman and Jillian C. York. West Censoring East - The Use of Western Technologies by Middle East Censors. https://opennet.net/sites/opennet.net/files/ONI_WestCensoringEast.pdf, 2011.

[79] Sadia Nourin, Van Tran, Xi Jiang, Kevin Bock, Nick Feamster, Nguyen Phong Hoang, and Dave Levin. Measuring and Evading Turkmenistan's Internet Censorship: A Case Study in Large-Scale Measurements of a Low-Penetration Country. In *WWW 2023*, 2023.

[80] Sudhakar V Nuti, Brian Wayda, Isuru Ranasinghe, Sisi Wang, Rachel P Dreyer, Serene I Chen, and Karthik Murugiah. The use of google trends in health care research: a systematic review. *PloS one*, 2014.

[81] OONI. Risks: Things you should know before using ooniprobe. https://ooni.org/about/risks.

[82] OONI. https://ooni.org/reports/, 2025.

[83] Ramakrishna Padmanabhan, Arturo Filastò, Maria Xynou, Ram Sundara Raman, Kennedy Middleton, Mingwei Zhang, Doug Madory, Molly Roberts, and Alberto Dainotti. A multi-perspective view of internet censorship in myanmar. In *FOCI*, 2021.

[84] Proton VPN. Proton VPN Observatory (2025). https://protonvpn.com/internet-censorship-observatory.

[85] Psiphon. Psiphon Data Engine, 2025. https://psix.ca/d/Vcvqvj6Wk.

[86] Ram Sundara Raman, Leonid Evdokimov, Eric Wustrow, J. Alex Halderman, and Roya Ensafi. Investigating large scale https interception in kazakhstan. In *IMC*, 2020.

[87] Reethika Ramesh, Ram Sundara Raman, Apurva Virkud, Alexandra Dirksen, Armin Huremagic, David Fifield, Dirk Rodenburg, Rod Hynes, Doug Madory, and Roya Ensafi. Network Responses to Russia's Invasion of Ukraine in 2022: A Cautionary Tale for Internet Freedom. In *USENIX Security Symposium*, 2023.

[88] Reethika Ramesh, Ram Sundara Raman, Matthew Bernhard, Victor Ongkowijaya, Leonid Evdokimov, Annie Edmundson, Steve Sprecher, Muhammad Ikram, and Roya Ensafi. Decentralized Control: A Case Study of Russia. In *NDSS*, 2020.

[89] Kevin Rawlinson. Turkey blocks use of Twitter after prime minister attacks social media site. https://www.theguardian.com/world/2014/mar/21/turkey-blocks-twitter-prime-minister, 2014.

[90] Claire Reilly. Australian ISPs ordered to block The Pirate Bay by year's end. https://www.cnet.com/tech/services-and-software/australian-isps-block-the-pirate-bay-rights-holders-pay-village-roadshow-foxtel-telstra-tpg/, 2016.

[91] Roskomnadzor. Roskomnadzor has restricted access to Viber. https://t.me/rkn_tg/1306, 2024.

[92] Roskomsvoboda. VPN in Russia: from blocking services to blocking protocols. https://roskomsvoboda.org/uploads/en__vpn_in_russia__from_blocking_services_to_blocking_protocols.pdf, 2023.

[93] Zach Rosson, Felicia Anthonio, and Carolyn Tackett. Emboldened offenders, endangered communities: internet shutdowns in 2024. https://www.accessnow.org/wp-content/uploads/2025/02/KeepItOn-2024-Internet-Shutdowns-Annual-Report.pdf, 2025.

[94] Daniel Sanchez. French Authorities Shut Down Two Top French Piracy Sites. https://www.digitalmusicnews.com/2016/11/29/french-authorities-close-piracy-sites/, 2016.

[95] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *ICDCS Workshops*, 2011.

[96] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. Anomaly detection in time series: a comprehensive evaluation. *VLDB*, 2022.

[97] Internet Society, 2025. https://pulse.internetsociety.org/.

[98] Olivia Solon. Trump poised to sign away privacy protections for internet users. https://www.theguardian.com/technology/2017/mar/28/privacy-protection-sell-web-browsing-history-data, 2017.

[99] Ram Sundara Raman, Prerana Shenoy, Katharina Kohls, and Roya Ensafi. Censored Planet: An Internet-Wide, Longitudinal Censorship Observatory. In *CCS*, 2020.

[100] Ram Sundara Raman, Adrian Stoll, Jakub Dalek, Reethika Ramesh, Will Scott, and Roya Ensafi. Measuring the Deployment of Network Censorship Filters at Global Scale. In *NDSS*, 2020.

[101] Aris A Syntetos and John E Boylan. The accuracy of intermittent demand estimates. *International Journal of forecasting*, 2005.

[102] Aris A Syntetos, John E Boylan, and JD Croston. On the categorization of demand patterns. *Journal of the operational research society*, 2005.

[103] The Jordan Times. PUBG ban: Players groan, parents rejoice. https://jordantimes.com/news/local/pubg-ban-players-groan-parents-rejoice.

[104] The Moscow Times. Russian Law Banning VPNs Comes Into Effect. https://www.themoscowtimes.com/2017/11/01/russian-law-banning-anonymous-online-surfing-comes-into-effect-a59434, 2017.

[105] The Tor Project. Tor Metrics, 2025. https://metrics.torproject.org/.

[106] Joshua J Thompson, Robert L Wilby, Tom Matthews, and Conor Murphy. The utility of Google Trends as a tool for evaluating flooding in data-scarce places. *Area*, 2022.

[107] Joan C Timoneda and Erik Wibbels. Spikes and variance: Using google trends to detect and forecast protests. *Political Analysis*, 2022.

[108] TurkeyBlocks. Facebook, Twitter, YouTube and WhatsApp shutdown in Turkey. https://turkeyblocks.org/2016/11/04/social-media-shutdown-turkey/, 2016.

[109] UnpopularOpinion1122. Porn Blocked in Egypt? https://www.reddit.com/r/Egypt/comments/7mjpzh/porn_blocked_in_egypt/, 2017.

[110] ValdikSS. Blocking VPN protocols on TSPU (08/05/2023 - xx.xx.202x). https://ntc.party/t/vpn-05082023-xxxx202x, 2023.

[111] Vlast. Kazakhstan has limited access to Google services. https://vlast.kz/novosti/37259-v-kazahstane-ogranicen-dostup-k-servisam-google.html, 2020.

[112] Valentin Weber. The Worldwide Web of Chinese and Russian Information Controls. https://public.opentech.fund/documents/English_Weber_WWW_of_Information_Controls_Final.pdf/, 2019.

[113] wkrp. Roskomnadzor's plans to block various VPN services. https://github.com/net4people/bbs/issues/76, 2021.

[114] Joss Wright, Alexander Darer, and Oliver Farnan. On Identifying Anomalies in Tor Usage with Applications in Detecting Internet Censorship. In *Proceedings of the 10th ACM Conference on Web Science*, 2018.

[115] Mingshi Wu, Jackson Sippe, Danesh Sivakumar, Jack Burg, Peter Anderson, Xiaokang Wang, Kevin Bock, Amir Houmansadr, Dave Levin, and Eric Wustrow. How the Great Firewall of China Detects and Blocks Fully Encrypted Traffic. In *USENIX Security 23*, 2023.

[116] Renjie Wu and Eamonn Keogh. Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[117] Diwen Xue, Reethika Ramesh, Valdik S S, Leonid Evdokimov, Andrey Viktorov, Arham Jain, Eric Wustrow, Simone Basso, and Roya Ensafi. Throttling twitter: an emerging censorship technique in russia. In *IMC*, 2021.

[118] Maria Xynou. Azerbaijan blocked TikTok and Google Play Store amid military offensive in Nagorno-Karabakh. https://explorer.ooni.org/en/findings/67768606801, 2023.

[119] Maria Xynou, Leonid Evdokimov, Arturo Filastò, DefendDefenders, and POLLICY. Uganda's Social Media Tax through the lens of network measurements. https://ooni.org/post/uganda-social-media-tax, 2018.

[120] Maria Xynou and Arturo Filastò. Ethiopia: Ongoing blocking of social media. https://ooni.org/post/2023-ethiopia-blocks-social-media/, 2023.

[121] Safwat Zargar. The internet is painfully slow in Kashmir, but users have found a way to access restricted websites. https://scroll.in/article/951519/the-internet-is-painfully-slow-in-kashmir-but-users-have-found-a-way-to-access-restricted-websites.

Appendix A. Supplemental Data

A.1. Non-Circumvention Topics

Because *CenAlert* is topic-agnostic, its value to the Internet freedom community depends on <Virtual Private Network> being a valid proxy for censorship events. While our evaluation shows that many spikes align closely with such events, a potential concern is that this relationship may stem from unrelated factors affecting VPN use that frequently occur at the same time. To test this, we ran *CenAlert* on two negative-control topics, <Soccer> and <Film>, which are plausibly related to VPN use (e.g., for circumventing georestrictions) but not necessarily primary drivers of censorship. For each topic, we ran *CenAlert* on time series stitched from 25 downloads of each window and applied the automatic matching process against both the community-verified events from Section 4.1 and the manually verified censorship events from Section 4.2.

We found that neither term exhibits a strong relationship with censorship events. Considering the 100 highest-impact

Topic	Countries With Non-Zero Data	Median Sparsity (2011 - 2024)	Median Sparsity (From First Non-Zero)	Median Sparsity (Last 5 Years)
Virtual Private Network	72	0.73	0.51	0.35
Proxy Server	57	0.92	0.91	0.96
Internet Censorship Circumvention	10	0.99	0.98	0.99
Tor	32	0.99	0.99	1.00
Psiphon	39	0.99	0.99	1.00

TABLE 3: **Data Quality of Circumvention-Related Topics.** We evaluate the data quality of five circumvention-related topics in Google Trends across four metrics: the number of countries with at least one non-zero point in the stitched time series, and, for these countries, the median sparsity over the entire period from January 2011 to December 2024, from the first non-zero value, and over the past five years. <Virtual Private Network> has data in the most countries and shows the least sparsity.

spikes for <Soccer> and <Film>, only three in each case occur near the start of censorship events, compared to 76 for <Virtual Private Network> (Section 5.1). In the same nine highly censoring countries analyzed in Section 5.2, <Soccer> and <Film> spikes coincide with just 5 and 4 events, respectively, compared to 101 for <Virtual Private Network>. For <Soccer>, these matches account for a low proportion of per-country cumulative impact factor (11.5%, averaged over countries with any spikes), while for <Film> the proportion (61.6%) is high only because the total number of spikes in relevant countries is very small. For <Virtual Private Network>, however, 91.2% of the cumulative impact factor can be attributed to censorship events. Together, these results indicate that spikes in the control topics rarely coincide with censorship events, suggesting that <Virtual Private Network> spikes during censorship events are unlikely to be caused by unrelated factors.

A.2. Other Circumvention-Related Topics

Google Trends has several topics related to censorship circumvention beyond <Virtual Private Network>, including <Proxy Server>, <Internet Censorship Circumvention>, <Tor Browser>, and ¡Psiphon¿. Before committing to <Virtual Private Network>, we examined these alternatives for the 76 countries selected in Section 4.1 but found they lack comparable data quality. Table 3 summarizes data availability across four metrics: the number of countries with at least one non-zero point in the stitched time series, and, for these countries, the median sparsity over the entire period from January 2011 to December 2024, from the first non-zero value, and over the past five years. <Virtual Private Network> is available in the most countries and has the lowest sparsity across all periods. The extreme sparsity of other topics—exceeding 95% in recent years—limits the occurrence and reliable detection of spikes both within and across countries.

A.3. Spike Trends Over Time

Figure 7 shows the evolution of the spikes detected by *CenAlert* over time. Both the number of spikes and the number of countries where spikes are detected increase with time, peaking in 2020 alongside the COVID-19 pandemic. These trends reflect a growing global need to circumvent the increasing splintering of the Internet, including, but not limited to, Internet censorship. As highlighted by organizations such as Access Now and Freedom House, global Internet freedom has been declining [2], with both the number of documented censorship events and the countries where they occur steadily increasing over time [93]. In the past 5 years, we observe spikes associated with blocking even in countries such as Italy (which temporarily banned ChatGPT in March 2023 [57]) and the United Kingdom (which blocked pirated sports streams multiple times in September 2020 [65]), which are typically regarded as having a free Internet.

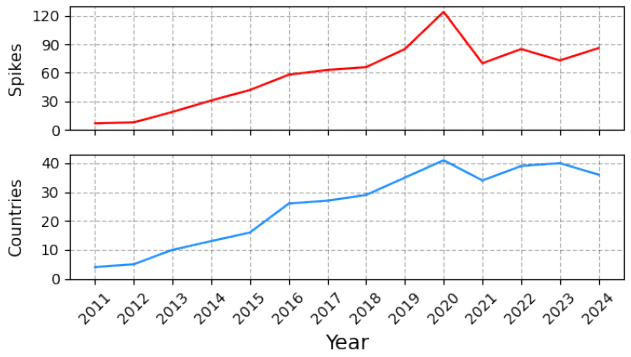

Figure 7: **Trends in Detected Spikes.** Over time, both the number of spikes detected by *CenAlert* as well as the number of countries in which they occur have been steadily increasing, peaking in 2020 due to the COVID-19 pandemic. Countries reflect not only traditional ones, but also relative newcomers to blocking such as Italy and the United Kingdom.

TABLE 4: **A Subset of the 100 Highest Impact Spikes** (in reverse chronological order). Due to space constraints, the remaining spikes and their descriptions can be found at: https://github.com/censoredplanet/cenalert-paper/blob/main/auxiliary_data/top_100_impact_events.csv.

Country	Start Date	Description
PK	2024-11-24	Pakistan has blocked WhatsApp in addition to X, Facebook, Instagram, and Bluesky as authorities tighten security ahead of anti-government protests, while the government has set a November 30 deadline for VPN services to register with the Pakistan Telecommunication Authority. *Verified by: community*
MU	2024-11-01	Mauritius suspended access to social media platforms ahead of parliamentary elections. *Verified by: community*
PK	2024-08-09	Internet services across Pakistan faced disruption for a second consecutive day, with slow speeds and messaging app issues, while the root cause remains undetermined and no official government explanation has been provided. *Verified by: manual*
EC	2024-08-04	A court order required ISPs to block over 180 IP addresses that illegally broadcast Ecuadorian soccer matches. *Verified by: manual*
TR	2024-08-02	Türkiye's communications authority blocked access to Instagram following accusations that the platform censored posts related to Hamas leader Ismail Haniyeh's killing. *Verified by: community*
RU	2024-08-01	Russia has systematically suppressed independent media by blocking at least 279 news media domains, implementing widespread TLS interference through decentralized DPI deployments, and designating organizations as "foreign agents" or "undesirable," resulting in financial challenges, security risks, audience loss, and increased self-censorship among journalists. Russia simultaneously throttled YouTube to unusable speeds before blocking it entirely a week later. *Verified by: community*
BD	2024-07-23	NetBlocks reported a partial restoration of Internet services following a shutdown imposed to quell protests over discriminatory government job quotas, although social media platforms remained blocked. *Verified by: community*
VE	2024-07-21	Ahead of the July 27, 2024 elections, Venezuelan ISPs blocked several media and civil society websites, along with Proton VPN. Days later, Venezuela also blocked X after Elon Musk accused President Nicolás Maduro of election fraud. *Verified by: community*
KE	2024-06-23	During protests in Kenya against the 2024 Finance Bill, authorities disrupted internet access across major networks on June 25, 2024, with connectivity dropping by nearly 40% nationwide, prompting Access Now and the KeepItOn coalition to condemn these actions as violations of fundamental rights. *Verified by: community*
MM	2024-05-30	In Myanmar, since May 30, 2024, a Chinese company reportedly identified as Geedge Networks has implemented VPN blocking technology for the military government, preventing citizens from accessing Facebook and other social media platforms that had been commonly accessed via VPNs since the 2021 military coup. *Verified by: manual*
KG	2024-04-17	Kyrgyzstan is restricting access to TikTok throughout the country, with the State Committee for National Security citing a lack of content censorship that could potentially harm children's development. *Verified by: community*
VN	2024-03-30	Vietnam blocked Roblox and Wattpad, though a major ISP later claimed that Roblox had been blocked accidentally during efforts to restrict illicit content. *Verified by: manual*
ES	2024-03-07	A court ruling in Barcelona allows legal action against individuals who consume pirated football content at home, requiring internet operators to provide user information to LaLiga for potential fines. *Verified by: other*
SA	2023-12-03	Saudi Arabia sparked public debate on social media platform X regarding the legality of using VPN services, with legal and cybersecurity experts offering conflicting interpretations—some maintaining there is no specific text criminalizing VPN use in Saudi law, while others argued it constitutes an illegal access crime under Article 3, Paragraph 3 of the Information Crime System, potentially punishable by one year imprisonment or a fine of 500,000 riyals. *Verified by: other*
NP	2023-11-13	Nepal banned TikTok in November 2023, citing the app's alleged disruption to social harmony and negative impact on family and social structures in the country. *Verified by: community*
TM	2023-06-03	unknown
SN	2023-06-01	During political unrest in Senegal, authorities blocked access to social media platforms including WhatsApp, Telegram, Facebook, Instagram, Twitter, YouTube, and TikTok between June 1-7, 2023, alongside implementing network outages, following protests over opposition leader Ousmane Sonko's sentencing. *Verified by: community*
PK	2023-05-09	Following the arrest of former Pakistani Prime Minister Imran Khan in May 2023, the Pakistani government ordered mobile internet shutdowns and social media blocks, resulting in a 30% decline in national internet traffic and a 60% drop in mobile device usage as citizens increasingly turned to Cloudflare's 1.1.1.1 DNS resolver to circumvent restrictions. *Verified by: community*
IT	2023-03-31	In March 2023, Italy became the first country to temporarily ban ChatGPT amid an investigation into privacy concerns, including a data breach and questionable data collection practices by OpenAI. *Verified by: manual*
SR	2023-02-18	The Suriname capital Paramaribo experienced violent protests against rising costs of living, with demonstrators targeting the National Assembly and looting stores, leading the government to impose a curfew while the main internet provider Telesur cut access to social media networks, an unprecedented censorship action in the country. *Verified by: community*

Appendix B. Meta-Review

The following meta-review was prepared by the program committee for the 2026 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

B.1. Summary

The paper presents *CenAlert* to detect Internet censorship events. *CenAlert* uses spikes in Google searches for VPN-related topics to detect censorship events. Evaluations on past searches and externally documented censorship events show that many spikes correlate with known censorship events and that there appear to many novel censorship events discovered by the tool.

B.2. Scientific Contributions

- Independent Confirmation of Important Results with Limited Prior Research.
- Provides a New Data Set For Public Use.
- Creates a New Tool to Enable Future Science.
- Addresses a Long-Known Issue.
- Provides a Valuable Step Forward in an Established Field.

B.3. Reasons for Acceptance

1) *CenAlert* shows immediate practical benefit in identifying new censorship events in past data.
2) *CenAlert* provides a new method to detect censorship events automatically and on an ongoing basis.

B.4. Noteworthy Concerns

1) *CenAlert* is not useful in settings of pervasive network censorship, including in particular Chinese censorship, because users constantly face some kind of censorship (although the nature may change) and thus are likely to be familiar with VPN workarounds already.
2) *CenAlert* depends on the accessibility of Google search from the censored country.
3) *CenAlert* only detects blocks for which VPNs are a solution.
4) The false/true positives of the tool are not rigorously evaluated.

B.5. Response to the Meta-Review

We provide context addressing several concerns raised in the meta-review.

Pervasive Censorship. We acknowledge that pervasive censorship can be a limitation of *CenAlert*. However, this limitation is only truly demonstrable for China, which is the most extensively studied censorship ecosystem and already well-served by many dedicated measurement projects. *CenAlert* is most valuable in the many countries that lack such coverage. Moreover, Section 5.2 demonstrates that *CenAlert* still detects several censorship events in other highly restrictive countries, including Iran, Russia, and Türkiye. Even if users are familiar with VPNs, modern censorship has evolved to target circumvention tools themselves, forcing users to frequently search for alternatives. We highlight multiple instances of such circumvention tool blocking in Section 5.2.

Google Accessibility. While *CenAlert* might seem wholly ineffective in countries that block Google, Section 7 demonstrates otherwise. During Russia's April 2018 Telegram ban, during which 19 million IP addresses including Google services were blocked, we still detected Russia's third-highest impact spike. We similarly detected a spike in Kazakhstan despite Google services briefly being blocked in February 2020.

Dependence on the Efficacy of VPNs. We emphasize that *CenAlert* detects spikes in VPN *searches*, not the use or efficacy of VPNs in circumventing censorship. This allows it to pick up on censorship events even when VPNs are not a viable technical solution. In cases where the technical implementation of blocking is not yet well understood, or when users are non-technical, VPNs are often the instinctive first choice for attempting to bypass restrictions. As a result, VPN searches can provide a reliable signal of censorship events regardless of whether the tools work in practice.