I Can Show You the World (of Censorship): Extracting Insights from Censorship Measurement Data Using Statistical Techniques

Anna Crowder
University of Florida
Gainesville, FL
annacrowder@ufl.edu

Daniel Olszewski
University of Florida
Gainesville, FL
dolszewski@ufl.edu

Patrick Traynor
University of Florida
Gainesville, FL
traynor@ufl.edu

Kevin R. B. Butler
University of Florida
Gainesville, FL
butler@ufl.edu

***Abstract*—In response to the growing sophistication of censorship methods deployed by governments worldwide, the existence of open-source censorship measurement platforms has increased. Analyzing censorship data is challenging due to the data's large size, diversity, and variability, requiring a comprehensive understanding of the data collection process and applying established data analysis techniques for thorough information extraction. In this work, we develop a framework that is applicable across all major censorship datasets to continually identify changes in censorship data trends and reveal potentially unreported censorship. Our framework consists of control charts and the Mann-Kendall trend detection test, originating from statistical process control theory, and we implement it on Censored Planet, GFWatch, the Open Observatory of Network Interference (OONI), and Tor data from Russia, Myanmar, China, Iran, Türkiye, and Pakistan from January 2021 through March 2023. Our study confirms results from prior studies and also identifies new events that we validate through media reports. Our correlation analysis reveals minimal similarities between censorship datasets. However, because our framework is applicable across all major censorship datasets, it significantly reduces the manual effort required to employ multiple datasets, which we further demonstrate by applying it to four additional Internet outage-related datasets. Our work thus provides a tool for continuously monitoring censorship activity and acts as a basis for developing more systematic, holistic, and in-depth analysis techniques for censorship data.**

I. Introduction

Censorship of information is a powerful means by which governments control their populations. Because of the crucial role the Internet plays in providing access to materials and communication, it is a common target for censorship. To assess the state of censorship within nations worldwide, researchers have developed measurement platforms [18], [30], [36], [58], [67], [86] to determine the accessibility of information over a variety of network protocols. These platforms vary in their usage modalities, with some initiated by users and others based on probes from local and remote points.

Censorship measurement platforms are a valuable resource, but ongoing problems with the analysis of the data have left their full potential unrealized. Censorship data is difficult to analyze due to its massive size and diversity across datasets. Each platform returns differing types of information, from the countries covered to the protocols and user bases taken into account. Understanding whether information across these platforms can be correlated and fused to provide a more holistic determination of global censorship has not been fully explored.

In this paper, we explore whether consolidating datasets from disparate censorship measurement platforms and consistently applying statistical techniques to detect change in the data allows for more robust detection of unknown censorship events. Our analysis spans more than two years, from January 2021 to March 2023, and incorporates data from Censored Planet (the Satellite, Hyperquack HTTP, and Hyperquack HTTPS datasets) [67], The Open Observatory of Network Interference (OONI) [30], GFWatch [36], and Tor [78], across six countries - Russia, Myanmar, China, Iran, Türkiye, and Pakistan - encompassing over 1.7 billion measurements. We examine the correlation between datasets and censorship signals across platforms and introduce a framework based on scalable techniques grounded in foundational statistical process control theory to identify the presence of potential censorship events. Well-reported events (e.g., the Russian invasion of Ukraine in 2022 [79]) and prior research [61], [70] are used where possible as ground truth of when censorship events occur to validate corresponding signals identified by our framework. We then perform an extensive qualitative analysis of news events in each of the six countries, evaluating 1,650 news articles to determine whether the remaining signals indicate the existence of previously unknown censorship activities. Finally, we perform a comparative analysis of our framework against Censored Planet's *CenDTect* system [81] and demonstrate the flexibility of our framework by applying it to additional network traffic datasets. Our contributions are thus as follows:

- **Statistical Framework Development:** We develop a statistical framework[1] to identify signals of possible unknown censorship events in six censorship measurement datasets across six countries incorporating over 1.7 billion measurements. Our approach is scalable, continuous,

[1] Available at https://github.com/censorship-event-detection/acsac-censorship-event-detection/

consistent across diverse datasets, and statistically appropriate for complex processes represented by noisy data. The framework facilitates the use of multiple datasets and reduces the amount of manual effort required to analyze censorship measurement datasets. By pinpointing specific time frames during which the data changes, our signals limit the amount of data that requires further investigation to confirm censorship events.

- **Cross-Dataset Analysis:** We are the first to measure the overlap between signals identified across the different datasets and implement correlation metrics on pairs of datasets across platforms. We show that minimal overlap or correlation exists between censorship measurement platforms. As a result, we recommend including all censorship measurement datasets in future investigations on censorship events.
- **Qualitative Investigation of Events:** We analyze 1,650 news articles around the associated signals our framework identifies as increasing censorship within the data to verify the possibility of censorship occurring within their time frame. Our analysis shows that all but one of these signals corresponds to at least one major news event.

The remainder of the paper is organized as follows: Section II provides background information on censorship measurement platforms; Section III states our hypothesis; Section IV outlines our framework and experimental design; Section V presents the results of our correlation analysis; Section VI presents the results of our time series analysis; Section VII presents the implications of our qualitative analysis; Section VIII outlines a comparison of our framework against previous work as well as discusses additional datasets and limitations; Section IX brings attention to related work; Section X concludes our work.

II. Measurement Platforms

The prevalence and scale of censorship in use worldwide has led researchers to develop several platforms for measuring censorship. The platforms differ in several design decisions, including which countries are observed, how the data is ethically accessed within those countries, how long the measurements are performed, and what methods of censorship are tested. As a result, each platform provides a different picture of censorship. Table I displays a list of the primary Internet measurement platforms used by the censorship research community. We focus our analysis on the Censored Planet, GFWatch, OONI, and Tor platforms because they are extensive, continuous, and very relevant to censorship activity. We further discuss the reasons for this focus in Section VIII. We now provide a short description of each of the focus platforms.

TABLE I: **A summary of Internet measurement datasets. We analyze the datasets that measure one or more forms of censorship, are open source, and are ongoing. ⋆: Censored Planet dataset. Gray rows: Included in our analysis.**

Datasets	*Measures*	*Open Source*	*Ongoing*
Augur ⋆	TCP/IP	✗	✓
CensorWatch	DNS, HTTP/S	✓	✗
GFWatch	DNS	✓	✓
Hyperquack ⋆	HTTP/S	✓	✓
ICLab	DNS, HTTP	✓	✗
IODA	Internet outages	✓	✓
Kentik	Network Activity	✗	✓
Lantern	Circumvention Activity	✗	✓
OONI	DNS, TCP/IP, HTTP, Users	✓	✓
Psiphon	Network Activity	✓	✓
Satellite ⋆	DNS	✓	✓
Tor	Bridge Reachability, Users	✓	✓
Triplet Censors	DNS	✓	✗

Censored Planet: A censorship measurement platform that uses aspects of the Internet infrastructure to perform remote tests to monitor censorship globally. The platform includes three systems that collect measurements at different levels of the network stack. Satellite detects DNS manipulation by querying open DNS resolvers [63], [76]. Augur measures TCP reachability using TCP/IP side channels [62]. Hyperquack monitors application layer blocking using Echo or Discard servers, as well as web servers monitoring HTTP and HTTPS [68], [82]. As of writing, Satellite and Hyperquack data are publicly available while Augur data is not.

OONI: A free censorship measurement software for users worldwide to test the current state of connectivity in their country from their mobile devices. The results of OONI tests are automatically gathered and published in real time. OONI offers several tests for the user to choose from including a Web Connectivity test [60] which first identifies which DNS resolver the user is communicating with before doing a DNS lookup for the test URL. If the DNS lookup is resolved properly, it is followed by the initiation of a TCP connection. Finally, if the TCP handshake is completed, an HTTP or HTTPS GET request is sent.

GFWatch: A censorship platform that measures the DNS filtering behavior of the Great Firewall of China (GFW). GFWatch first runs queries from two controlled machines in China, before comparing the results to the same queries run from a controlled machine in the US. Because GFWatch runs all tests using controlled machines, the system consistently measures millions of domains every day.

Tor Project: A widely used tool for censorship circumvention and avoiding surveillance online. The Tor dataset includes a variety of metrics on their software including daily number of users, number of servers, amount of traffic, and performance. Additionally, Tor developers are currently building an automated process to test the reachability of Tor *Snowflake* and *obfs4* bridges. Although the Tor bridge reachability data provides an interesting perspective on censorship, it is not extensive enough to be included in our research at this time. Instead we include Tor user data [1] in our analysis excluding Myanmar as Tor does not have data on Myanmar.

III. Hypothesis

The full potential of the large collection of existing censorship measurement data remains largely unexplored. Due to the difficulties of handling the entirety of a single censorship dataset, researchers regularly rely on a subset of the data from a single source within a limited time frame and

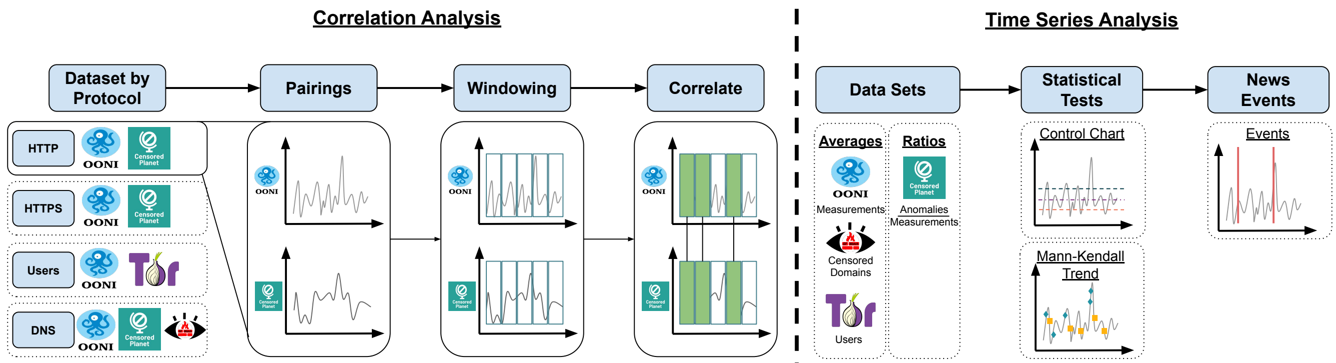


Fig. 1: **The pipeline for the methodology of our correlation analysis and our time series analysis. In the correlation analysis, we take pairs of datasets by protocol, window the datasets, and then run Spearman's rank and Kendall rank correlations. In the time series analysis, we run the control chart and Mann-Kendall trend detection tests on the various datasets. Then, we collect the news articles to identify plausible censorship-causing events.**

investigate a single major event. Furthermore, the methodology employed for data analysis lacks consistency from one project to another. We hypothesize that the systematic application of well-established tools from statistical process control to all censorship measurement data regardless of amount or length of time will enable the identification of signals within the data, uncovering previously undisclosed censorship events. Our investigation of this claim is guided by the following research questions:

RQ1 Do statistical frameworks identify critical changes in censorship measurement data related to censorship activity?
RQ2 To what extent do the different censorship measurement platform datasets correlate?
RQ3 Do news events indicative of possible censorship exist within the time frame of the changes identified by the framework?

IV. Design

To develop a general procedure for identifying changes in censorship data, we perform a time series analysis across six open-source censorship datasets using a statistically based framework. We compare whether there is consistency between the identified signals found in each dataset. Moreover, we complete a correlation analysis on each pair of datasets grouped by the protocol they use to measure censorship. Finally, we perform a qualitative analysis of news articles within the time frame of the signals produced by our framework to confirm that the signals likely correspond to a censorship event. Our full methodology is outlined in Figure 1.

In this section, we describe our methodology for how the data is processed, the correlation tests implemented on each pair of datasets, the design of our framework, and our qualitative analysis for identifying ground truth around those signals.

A. Processing Datasets

Censored Planet and OONI publish their raw data while GFWatch publishes processed data that only includes the confirmed censored domains, not the entirety of domains measured by GFWatch. Consequently, Censored Planet and OONI datasets require further processing that GFWatch does not. Because the Tor data we use is based on user count and not censorship of domains, it also does not require processing. When handling the raw Censored Planet and OONI data, we follow the best practices outlined by each platform in their documentation [52], [59], [69], [80]. The processed data contains each measurement labeled as either anomalous (potential censorship) or not anomalous.

OONI provides source code and a tutorial for processing their raw data [59] which includes automated and manual exploration steps. Within the automated section of the code, the response of each query in the measurement is compared to the response of a corresponding control measurement. This comparison relies on a collection of heuristics derived from the specific level of the network stack being tested by the current query. If there is no match between the measurement and the control at each level, the measurement is labeled as an anomaly. The manual exploration phase is dedicated to identifying and eliminating false positives based on a predefined set of criteria. For instance, a response to a measurement containing the name of a well-known Content Delivery Network (CDN) in the Autonomous System (AS) field is deemed indicative of a false positive. Invalid measurements arise when a control measurement fails to execute or is unavailable for a particular query. These instances are typically attributed to innocuous network errors and consequently excluded from our analysis.

Censored Planet developed a data analysis pipeline [69] that takes in the raw Censored Planet data and adds metadata such as domain category, geolocation, AS information, IP organization, HTTP body, and TLS certificate data. Next, the pipeline compares each measurement's responses to a control measurement's response. Additionally, each response is checked against a blockpage fingerprint database and a non-censorship cases fingerprint database. Finally, each measurement is labeled based on the outcome with `expected` corresponding to no signs of censorship and all other labels corresponding to different types of anomalies (e.g., potential censorship). While the Hyperquack data is ready for analysis after the pipeline, Satellite requires an additional round of processing.

With the growing popularity of CDNs and content localization, DNS manipulation is increasingly difficult to detect. The reliability of control DNS queries, which were previously relied upon to yield expected responses, is no longer guaranteed. As a result, they cannot be used as a comparison tool for test DNS queries. In response to this issue, Censored Planet deprecated all Satellite data that was collected before July 2022 as well as the associated methodology for labeling Satellite data as anomalous or not. They developed a new methodology to reduce the number of false positives that were occurring which employs TLS certificates for labeling Satellite data and is implemented on the processed data [80]. The source code for this novel technique is publicly available [52], and we incorporate it into our processing phase. We also limit the Satellite data in our analysis to data collected in July 2022 or after.

OONI and Censored Planet data have characteristics that impact the metric we used for each platform during our correlation and time series analysis. Within our analysis, the number of measurements collected by OONI per day is highly correlated with the number of anomalies in those measurements, because OONI is a user-based data collection process (i.e., users run OONI more often when they are experiencing censorship). This relationship implies that when a signal appears in the number of measurements, it will likely appear in the number of anomalies. Therefore, the **OONI metric = number of measurements per day**. We also apply our experiments on the number of anomalous OONI measurements separated by the type of blocking observed to catch the few instances where the correlation of measurements to anomalies is small. The Censored Planet data does not have the same procedural relationship between the number of measurements and the number of anomalies because their data collection process is partially automated. Instead, the number of measurements collected per day by Censored Planet is dependent on how many vantage points are used that day. Thus, the **Censored Planet metric = ratio of anomalies to measurements**. The **GFWatch metric = number of censored domains** and the **Tor metric = the number of users per day** because those are the only metrics available in the GFWatch and Tor data.

B. Techniques for Correlation Analysis

To undertake a comparative analysis of the datasets, we first separate them based on the individual protocols being measured for censorship. We separate the OONI measurements according to the specific protocol (e.g., DNS, HTTP, TLS) where an anomaly is observed. Subsequently, we compare these measurements against the corresponding dataset from Censored Planet, GFWatch, or Tor.

Our comparison uses both Kendall rank correlation coefficient [65] and Spearman's rank correlation coefficient [37]. Both methods measure the strength and direction of the monotonic relationship between two datasets based on ranks.

Kendall and Spearman are less restrictive than other correlation metrics as they make no assumption about normality and measure for a monotonic relationship instead of a linear one. As a result, they are less sensitive to outliers and noise which are common characteristics of censorship measurement data.

We pre-process two datasets to ensure they have the same number of data points. OONI, GFWatch, and Tor collect measurements daily while Censored Planet collects measurements a few times a week. Thus, if OONI and Satellite are being compared, the OONI data from days when Satellite does not have data are averaged together to line up with Satellite. Next, the two datasets are split into overlapping rolling windows ($n = 14$), and the correlation coefficient is calculated for each window. This window sample size was chosen a prior by considering the margin of error (E) for each correlation coefficient (ρ) based on calculations of confidence intervals (CI) [57], [72]. Our concern was to find a window size n that results in a CI that does not contain zero when $\rho >= |0.5|$ because a CI that does not contain zero indicates a significant $\rho \neq 0$. Because $CI = \rho \pm E$, we determined what window size was necessary to achieve $E < 0.5$. After calculating the correlation coefficients for each window, we keep the coefficients where the CI does not contain zero, indicating a significant correlation not equal to zero.

TABLE II: **Signal Identification Guidelines: We define what activity constitutes a signal for each technique and specify the number of data points required to confirm a signal**

Technique	Activity	Data points
Control Chart	Outside the control limits	2 out of 3
Mann-Kendall	Increasing or decreasing trend	2 out of 3
Both	Both activities	1+

C. Techniques for Time Series Analysis

Because censorship measurement data does not have an established practice for detecting change, our analysis is grounded in approaches employed in other fields. We treat censorship measurement data as an ongoing process defined as a system or set of conditions that behave in a standard way based on their results [84, p.3-4], and a signal occurs when the standard behavior changes. *Control charts*, used in statistical process control, are well-established temporal data analysis tools for detecting outliers, shifts in average performance, or changes in variability and are employed in a wide variety of fields from manufacturing [20] to finance [28] to healthcare [77]. Nonparametric trend tests, which are distribution-free statistical tests, are also widely used with temporal data, with examples occurring from medicine [24] to finance [19] to environmentalism [50]. The *Mann-Kendall* (MK) trend test [49] signals the presence of a monotonic (not necessarily linear) trend, as opposed to a shift in process. The two methods in combination improve the detection ability of the system. The decision process for choosing when to investigate the control chart and Mann-Kendall signals is shown in Table II.

Control chart - The appropriate control chart to use for each dataset is dependent on its metric. While the OONI, GFWatch, and Tor metrics are counts (the number of OONI measurements, censored domains measured by GFWatch, and Tor users per day) the Censored Planet metric (anomalies to measurements per day) is a ratio. Additionally, the control chart requires an assumption of the normal distribution of the data. Because the OONI, GFWatch, and Tor metrics could not be assumed to be normally distributed, we rely on the Central Limit Theorem [55, p. 89-90] which states that sample averages approach normality in their distribution even when the underlying distribution is not normal. The use of averages smooths out variability occurring in the data. Thus, the OONI, GFWatch, and Tor data are plotted on X-bar control charts [55, p. 259-261] using non-overlapping averages of $n = 5$. We chose a sample size of $n = 5$ because it accomplishes the smoothing needed but allows us to retain as much information as possible from the dataset. The Censored Planet metric is plotted on an Individual Measures control chart [55, p. 267-268], which assumes data is normally distributed. Proportions are assumed to approach normality in their distribution when n is large, which is the case for the Censored Planet metric.

The X-bar chart and the Individual Measures chart are both Shewhart style control charts [55, p. 193]. In this style of control chart, data is plotted across time with three key lines giving reference to performance. The central line is typically the process average, and an upper control limit (UCL) and lower control limit (LCL) provide bounds such that a point falling outside of these limits is considered a low-probability event. The UCL and LCL are set based on three standard deviations from the average. To create the control charts, we first determine a baseline within the data where the metric is operating consistently, before any obvious shifts up or down, and use the the baseline to calculate the UCL, Mean, and LCL. The baseline becomes the standard for what is currently expected in the metric's process. Our baseline is calculated from 20 consecutive data points with no MK trend, which is greater than the minimum required for creating control limits [55, p. 206]. The UCL and LCL for the OONI, GFWatch, and Tor X-bar control charts are shown in Equations 1. $\bar{\bar{x}}$ is the average of the non-overlapping averages, $\bar{s}$ is the average of the standard deviations, and c_4 is a constant used as part of an unbiased estimator of σ [55, p. 259-261].

$$UCL = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}}, \quad Mean = \bar{\bar{x}}, \quad LCL = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}} \quad (1)$$

The limits for the Censored Planet Individuals control chart are shown in Equations 2. $\bar{x}$ is the average of the ratios, $\overline{MR}$ is the average moving range where the moving range is the absolute value of the difference between consecutive data points. d_2 is the constant used as an estimator of σ when the MR is calculated from two consecutive data points which is the case for the Individuals control chart [55, p. 267-268].

$$UCL = \bar{x} + 3\frac{\overline{MR}}{d_2}, \quad Mean = \bar{x}, \quad LCL = \bar{x} - 3\frac{\overline{MR}}{d_2} \quad (2)$$

We use the control chart to identify signals for a potential shift in the level of metric performance. Under the assumption of a consistent process with the metric, the probability of a point falling above the UCL (or below the LCL) is 0.00135 [84, p. 180-181]. While this probability is already low, due to the variability inherent in the data, we minimize the risk of false positives by investigating a signal only when 2 out of 3 metric points go beyond the control limits. The probability of 2 out 3 points beyond the control limits in a process operating steadily is 0.000005 [84, p. 180-181]. If the data leaves the control limits for five consecutive data points, we establish a new baseline and then calculate new control chart limits.

Mann-Kendall (MK) trend detection test - The MK indicates a monotonic trend in the metric. It can support a shift identified by the control chart, and highlight trend signals that the control chart missed. The test is performed on overlapping rolling windows ($n = 10$) of the metric values across the full-time range. The sign of the difference between all possible pairs of values within each window is determined and a value is assigned (+1 for positive, -1 for negative, 0 for no difference) [49]. The sum, S, of all the signed values is used to determine the direction of the trend and as the statistic in determining significance. $S > 0$ indicates upward trend and $S < 0$ indicates downward trend. The statistical significance of the trend is determined based on the null hypothesis $S = 0$ with α-level = 0.05. Similar to the control chart, a signal is identified when 2 out of 3 MK windows show. Our guidelines for signal identification are shown in Table II.

V. Correlation Results

To determine whether censorship measurement platforms observe the same censorship events, we examine both the degree of signal overlap between datasets and the results of our correlation analysis. If the results are comparable across datasets, only one may be required when studying censorship activity. Conversely, if the results differ significantly, each dataset could be measuring a unique censorship phenomenon.

After applying our statistical framework to all of the datasets, we observe that the intersection of signals appearing at the same time across different datasets is limited. We calculate the Jaccard similarity coefficient (i.e. $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$) when signals occur for each pair of datasets across all six countries. Figure 2 shows a heatmap of the average coefficient results. The coefficients for each pair are all < 0.2 revealing that the overlap of signals between datasets is low. Overlap primarily occurs between datasets collected daily (GFWatch) or those reliant on users (Tor, OONI) because those datasets tend to generate longer MK signals and therefore have more signals to overlap with. The prolonged MK signals stem from the dynamic nature of GFWatch, OONI, and Tor, with the latter two undergoing frequent changes due to their user-based nature while the cause behind the periodic fluctuations experienced by GFWatch remains less evident. Still, the overlap between these three datasets is not substantial.

When we applied the correlation metrics to each pair of datasets, the average fraction of windows displaying any form

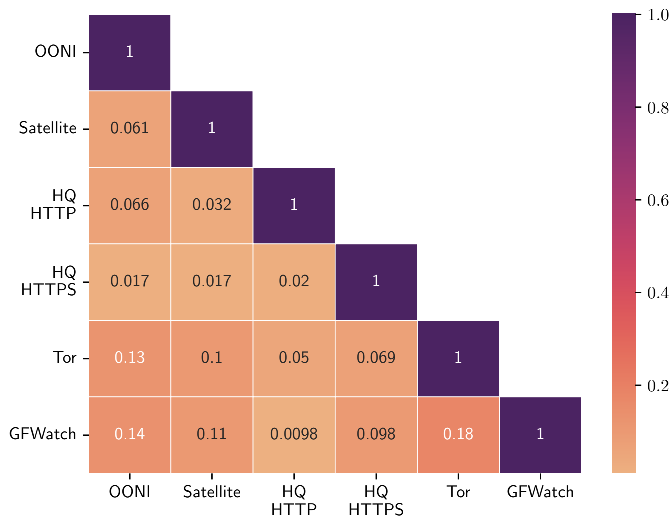


Fig. 2: **Heatmap of the average Jaccard coefficient for the overlap of signals identified in each dataset. The coefficient for all pairs is quite low, indicating that minimal overlap exists and the use of multiple datasets is essential.**

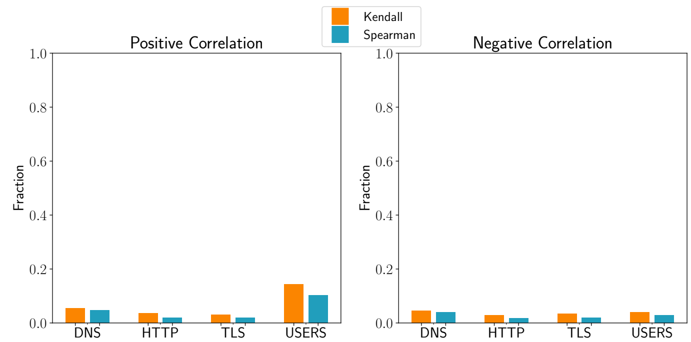


Fig. 3: **Fraction of windows with positive or negative correlation averaged across all the countries. The amount of correlation between the datasets is very low throughout highlighting the importance of using multiple datasets.**

of correlation (positive or negative) across all six countries and protocols was quite low. Furthermore, the proportions of positive and negative windows were comparable for all protocols. One minor exception was the correlation between the number of users in Tor and OONI datasets, where the fraction of positive correlation windows stood out. The average proportion of positive correlation between the user data exhibited the highest overall average fraction at 0.144 for Kendall and 0.103 for Spearman. Figure 3 shows the average correlation levels observed across each protocol. We provide full correlation analysis figures in Appendix A.

The lack of overlapping signals and correlation among the datasets suggests that the variations between censorship measurement platforms are consequential. These platforms differ widely from the specific domains and Autonomous Systems (AS) that are measured, to the methodologies employed for data collection (remote, user-based, direct), the frequency of data collection, and the data labeling approaches. Given the inherent challenge of determining the ground truth regarding censorship events, each censorship dataset potentially holds valuable insights, warranting a comprehensive investigation of the signals generated by each. Our framework serves to facilitate the use of multiple datasets by highlighting specific periods of change, limiting the data points of interest, and reducing the manual effort required for thorough investigation, thereby enhancing the capacity for leveraging the diverse information offered by these datasets to uncover unexplored censorship activity.

VI. Time Series Results

Using our statistical framework, we detect periods of new censorship activity in six datasets across six countries over 820 days, encompassing over 1.7 billion measurements. The number of identified signals per dataset and country ranges from 2 to 14, averaging only 20% of the data points in each dataset. These signals transform the data into manageable segments for manual investigation which is required to confirm the newly identified censorship events due to the lack of ground truth around censorship practices.

In this section, we first discuss a detailed example of our methods on Iran's OONI data. We then highlight signals that align with prior research findings and corroborate known censorship events to validate the ability of our framework across the six countries where we focus our analysis. Finally, we unveil previously unreported instances of heightened censorship.

A. Iran - Detailed Example

The raw OONI measurements from Iran shown in Figure 4 exhibit a consistent increase in measurements since early 2021; however, that alone does not inherently provide meaningful information. Upon implementing the control chart and MK test (Figure 4 below the raw data), consequential signals emerge. Horizontal lines represent the UCL, Mean, and LCL, and data point colors and shapes denote the MK test results for the window of data points starting at that point. The first signal is observed in April 2021 when three data points move above the UCL. Later, at the beginning of June 2021, we observe four windows (1 window = 10 data points) with an increasing MK trend, starting at the location of the blue diamonds. Subsequently, the data points move above the UCL again. When either method generates a signal, it provides confidence that a noteworthy shift has transpired, thereby serving as a starting point for further examination. When both methods produce signals within a similar time frame as is the case in June 2021, our confidence in the occurrence of a change is reinforced. After the data exceeds the limits for 5 data points in a row (e.g., $t_{i:i+5} > UCL$ or $t_{i:i+5} < LCL$), the UCL, Mean, and LCL are recalculated. This requires 20 data points exhibiting no MK trend, a condition not satisfied by the remaining Iran OONI data employed in our analysis. We simplify the signals into a summary graph depicted below the control chart and MK trends in Figure 4. We provide the full figures of the signals for all focus countries in Appendix B.

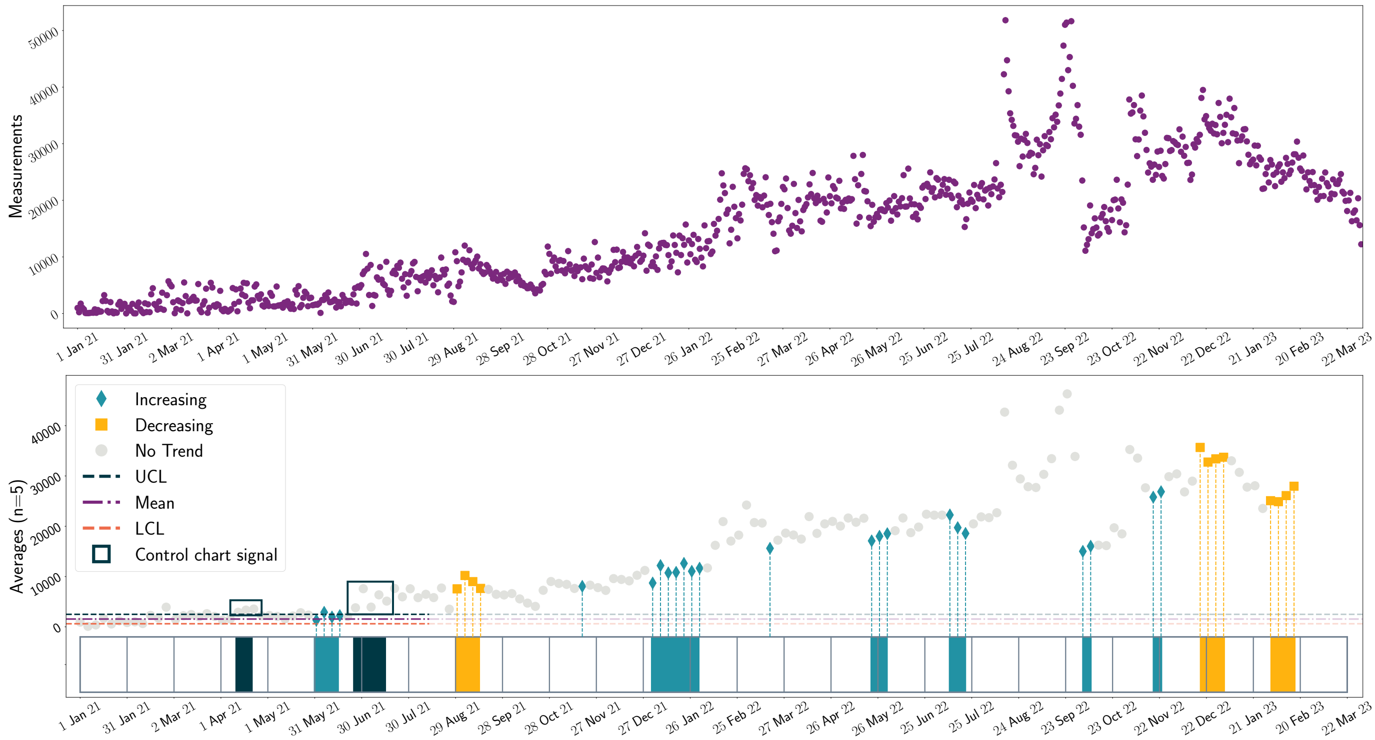


Fig. 4: **The top figure shows the number of measurements collected per day by OONI from Iran. The data is steadily increasing from June 2021 through December 2022. The bottom shows the same OONI data as non-overlapping averages with our framework applied to it. The horizontal lines are the control chart UCL, Mean, and LCL. The color and shape of the data points correspond to the Mann-Kendall (MK) trend test result for the window that starts at that data point (blue diamonds = upward trend, yellow squares = downward trend, and gray circles = no trend). The control chart signals when the data goes above the limits as highlighted by the blue boxes. The MK signals whenever a trend exists. Underneath the data, we simplify the data into a summary of the control chart and MK signals identified in the same OONI data.**

B. Corroboration of Known Censorship

Russia - During Russia's invasion of Ukraine in March 2022 control of information within Russia was a significant concern. The Russian government employed censorship measures and implemented restrictions on "fake news" to influence the information available to its citizens [79]. In their study on the state of the network in Russia following the invasion, Ramesh et al. [70] manually observed that while the overall number of measurements collected by OONI remained relatively stable, the occurrence of anomalies increased significantly. Figure 5 shows the signals found by our framework around the time of the invasion. The total number of measurements collected by OONI does not yield a discernible signal, but signals appear in the OONI DNS, HTTP, and TLS anomaly datasets in February 2022, where the MK test exhibits an increasing trend in several windows. These signals provide statistically based corroboration for the findings of Ramesh et al. within OONI. They also reported increased censorship of specific domains within popular ASes based on the Hyperquack HTTPS data. Our analysis does not detect a signal within this dataset which may be a consequence of our framework considering the entirety of the dataset.

Myanmar - In February 2021, Myanmar experienced a military coup that resulted in the seizure of power and a temporary shutdown of the Internet [4]. In subsequent months the military exerted control over the Internet through nightly shutdowns and the blocking of social media platforms [73]. When Padmanabhan et al. [61] examined this event in detail, they used data from OONI for 25 popular websites spanning from February 1 to April 30 in 2021. They found that immediately following the coup, there was a greater number of OONI measurements compared to the subsequent months. The findings from our analysis of the entire OONI dataset for this time frame align with these results and are shown in Figure 6. Specifically, our analysis reveals a single window at the beginning of the OONI data with an increasing trend according to the MK test; if our dataset extended further back in time, we might have identified additional windows with an increasing trend. Our methodology also uncovers an OONI signal in early February with five windows exhibiting a decreasing MK trend, which confirms that the surge observed was temporary. The other datasets included in our analysis provide additional evidence that there was a brief increase in censorship activity around the coup as Hyperquack HTTPS experiences an increasing signal before the coup, and Hyperquack HTTP shows a decreasing signal following the coup.

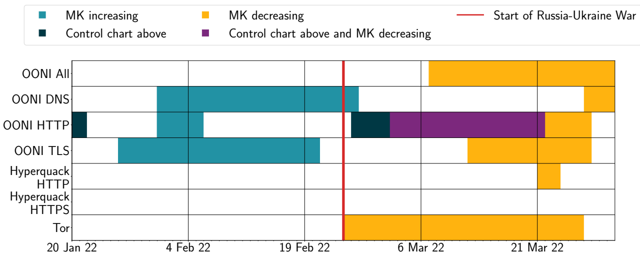


Fig. 5: **Russia - Signals during the Russian invasion of Ukraine. The OONI measurements remain stable leading up to the invasion, but the OONI anomalies increase, aligning with previous results [70].**

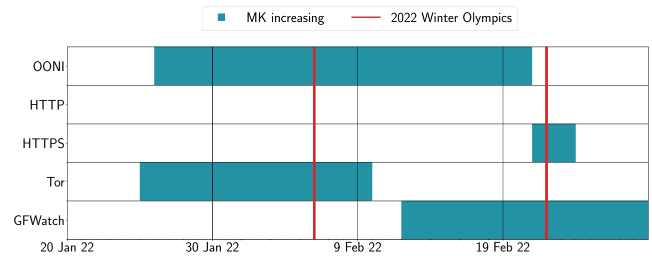


Fig. 7: **China - Signals during the 2022 Winter Olympics in Beijing. Excluding Hyperquack HTTP, each dataset experienced an increase in censorship during the Olympic games.**

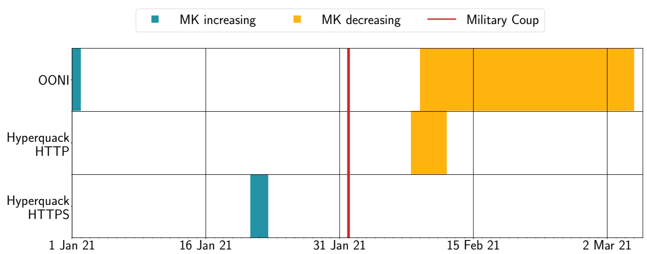


Fig. 6: **Myanmar - Signals during the 2021 military coup. The OONI and Hyperquack HTTPS data increased leading up to the coup, but OONI shortly returned to lower levels.**

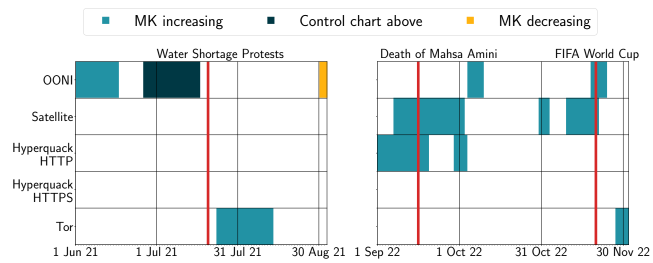


Fig. 8: **Iran - Signals during protests in 2021 and 2022. OONI increased leading up to the water shortage protests, and Tor increased after they began. OONI and Satellite increased at the start of the 2022 protests and the World Cup. Hyperquack HTTP also increased at the start of the 2022 protests.**

China - When the 2022 Winter Olympics in Beijing were approaching, there arose a heightened sense of apprehension among athletes, attendees, and journalists regarding their ability to access the Internet while in China. Journalists, in particular, expressed concerns about potential limitations that could hinder their freedom to openly report on the games [48]. Although the organizing committee assured unrestricted Internet access for all participants, similar promises were made before the 2008 Summer Olympics in Beijing, but substantial restrictions were in use during this event [22]. Media accounts from the recent 2022 Winter Games shed light on the pervasive extent of China's censorship apparatus during the event [10]. Figure 7 shows the results of our analysis and reveals signals indicating OONI and Tor experienced an upsurge in censorship leading up to the games while Hyperquack HTTPS and GFWatch experienced an increase during the games.

C. Identification of Potential Censorship

Iran - Following the death of Mahsa Amini, which occurred as a result of her arrest by the Iranian morality police due to a violation of their dress code, widespread protests erupted throughout Iran in September 2022 [47], [54]. In an apparent attempt to contain the uprisings, the government resorted to censorship measures, implementing blocks on popular social media platforms such as WhatsApp, Skype, and Instagram [94]. With our framework, we find corroborating evidence in support of these reports. Specifically, both the Satellite and Hyperquack HTTP data reveal signals (shown in Figure 8) indicating increased censorship starting in September 2022, characterized by multiple windows exhibiting upward trends according to the MK test. Subsequently, OONI data also manifests a similar upward signal. A few months later, during the FIFA World Cup in November 2022, the Iranian players elevated the protests onto the global stage by refraining from singing the national anthem [27]. Coinciding with this event, both Satellite and OONI once again indicate an increase in censorship.

Over a year earlier our framework identified similarly increasing signals in the OONI and Tor. At this time Iran was facing the worst water shortage in 50 years which sparked protests. During these protests, at least three people were shot, and over 102 people had been arrested [7], [56]. While there were no media reports on censorship during these protests, because of Iran's history of censoring communication during periods of unrest, the employment of censorship during this time is possible. The signals detected during the 2021 and 2022 protests provide an important starting place for further investigation into how Iran employs censorship during protests and how their censorship has changed over time. The signals during the 2021 protest can also be seen in Figure 8.

Türkiye - In March 2021, the OONI measurements data revealed an increasing signal indicated by the MK test as seen in Figure 9. While there were no official reports of censorship around this time, the increase coincided with an important development: Turkish President Recep Tayyip

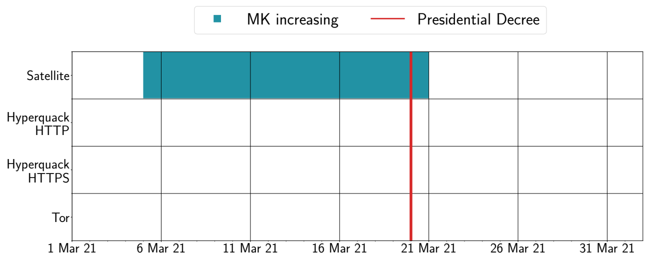


Fig. 9: **Türkiye - A summary of signals around the announcement of Türkiye's withdrawal from the Istanbul Convention and the resulting protest. OONI measurements experienced an increase in censorship.**

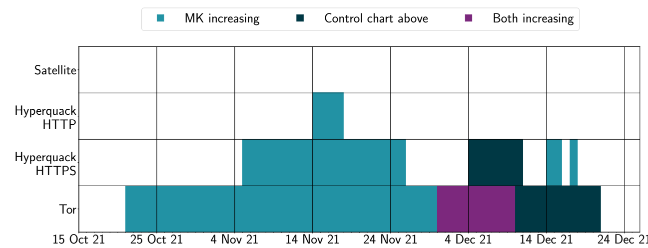


Fig. 10: **Pakistan - A summary of the large concentration of signals across multiple datasets in October and November 2021 revealing an important time for further investigation into Pakistan's censorship practices.**

Erdogan's announcement of Türkiye's withdrawal from the Istanbul convention [74]. The Istanbul convention serves as a crucial international treaty in the fight against violence targeting women, and Erdogan's decision sparked widespread protests across the country. Given President Erdogan's history of censoring any content critical of his administration [39], along with the substantial negative public reaction following his announcement, the signals identified by our framework in the OONI measurements provide strong evidence of heightened efforts to control or restrict the flow of information during this period.

Pakistan - During October and November 2021, multiple datasets exhibited a series of signals indicating a noticeable increase in censorship within Pakistan. Figure 10 shows that both Hyperquack HTTPS and Tor displayed numerous windows with an increasing MK trend, soon followed by the data surpassing the UCL. In addition, Hyperquack HTTP also exhibited a few windows with an increasing MK trend during this period. The convergence of these signals within a relatively close timeframe makes this period particularly noteworthy. Several significant events were unfolding in Pakistan during this time, potentially contributing to the observed increase in censorship. Pakistan was dealing with an influx of thousands of refugees from Afghanistan due to the Taliban's takeover [34]. Additionally, the country was engaged in conflicts with a different faction of the Taliban within its borders [35]. Furthermore, this timeframe preceded the removal of Imran Khan as Prime Minister by several months [12], and the signals in the data could be attributed to the government under Imran Khan restricting Internet access in the lead-up to the vote. Given the diverse range and abundance of signals observed during this period, it merits further investigation and analysis to uncover the underlying factors contributing to the surge in censorship.

VII. Qualitative Analysis

A. Ground Truth Analysis

While we cannot know the ground truth for censorship signals, we explore the news events around each of the identified increasing signals to corroborate the possible existence of a censorship event and further validate the effectiveness of our framework. For each country, we find the date range for each increasing signal and take the top ten English news articles from a search engine within each date range. We then classify each article as breaking news (i.e., news-worthy events) or opinion pieces (e.g., op-eds, editorials, etc.). Each article is classified by two reviewers, and any disparity is discussed until agreement. A measure for inter-rater reliability is Cohen's Kappa coefficient, which balances the rate the reviewers agree against the random chance that they agree. Cohen's Kappa of $\kappa > 0.7$ is considered a strong agreement [53], and we calculated $\kappa = 0.90$. After we identify each event, we discuss the possible motivation for the increasing censorship signal based on the event. We discuss the top events for each increasing signal that are most likely to prompt the countries to employ censorship. We find the events previously discussed in Section VI appear in our qualitative analysis. Further, we identify several types of events that correlate to other identified signals and find that all but one signal from China included plausible censorship events. We analyze 1,650 news articles to find current events surrounding each signal to show that our platform correlates to plausible unknown censorship events. The signals discussed in this section are shown in the figures in Appendix B.

B. Results

Russia - When searching the top ten articles for each increasing censorship signal for Russia, we find that of the 26 signals we found across all of the datasets, each signal had at least two articles that discussed breaking news. We find that twelve signals relate to developments in the Ukraine War, four signals relate to negotiations with the United States over Britney Griner's release and five relate to cutting off gas exportation to Europe. As discussed in Section VI-B, there is an abundance of evidence that Russia censored traffic during the Ukraine War to shape the public narrative. It follows that there is a precedence for Russia to censor during international events.

A total of 179 of the 260 news articles included breaking events. The largest source of news events is the Ukrainian War with 66 of the 260 articles (25.4%) relating to the war. These events included direct developments in the war, military strikes against Ukraine, invasion preparation, losses in Ukraine, the grain agreement between Russia and Ukraine, and restrictions due to the invasion. Smaller but notable events include foreign investigations into local Russia's actions (3%), political statements made by President Vladimir Putin (10%) (e.g., the constitutional changes that would allow him to maintain power until 2036 [41]), protests (1.1%), Brittany Griner's release (4%), Navalny's imprisonment (3.8%), and COVID-19 related policies (4.2%). Each of these events faced significant dissent from the population and as such are plausible censorship events.

Myanmar - We find a total of 19 increasing signals for Myanmar. Although only 90 of the 190 articles associated with the increasing signals are events, we find motivating events for each signal. Corroborating our previous analysis, we identify the Myanmar military coup in February 2021. The rest of the signals surround the humanitarian crisis the military coup has created. For example, in December of 2021 our framework identifies an increasing signal in both OONI and Hyperquck HTTP. The military committed atrocities during this time including massacring 35 people [8], killing and burning 13 villagers [5], and blocking humanitarian aid from reaching millions of displaced citizens [3].

Further increasing signals surrounded protests, rebel actions, executions, and the sentencing of Aung San Suu Kyi. These events occurred during an ongoing coup during which the government would likely seek to censor communication to control dissent.

China - There are 35 increasing censorship signals across all of the datasets for China. Although we only classify 109 of the 350 articles as events, we find motivating events for all but one signal. For example, previous work [26] shows that there is significant evidence that China employed vast censorship during COVID-19 due to their unpopular Zero Covid Policy, and 9 of the 35 increasing signals contain breaking news related to this policy. In March of 2022, China had its worst outbreak of COVID-19 [85], and we find an increasing signal in Hyperquack HTTPS and GFWatch. Similarly, in November of 2022, our framework finds increasing signals in GFWatch, OONI, and Hyperquack HTTPS. During that time, China saw further rising cases of COVID-19 that led to historic protests [42], [44]. Thus, the signals we find coincide during times of public unrest, and the events are possible events that would incite censorship. The rest of the increasing signals appeared around events such as US-China relations, climate-related disasters, and new government restrictions.

Iran - For the 20 increasing signals, we find 150 of the 200 articles are breaking news articles. As discussed in Section VI, the water shortage protests in July 2021, the Mahsa Amini protests in September 2022, and the demonstrations at the World Cup in December of 2022 account for 9 of the increasing signals we find. While 25% of the news articles relate to these events, there are several other motivating events we find. Specifically, our framework identifies multiple signals related to the arrests of activists. Zaghari-Ratcliffe was charged for a second time only a week after serving a 5-year prison sentence [2], and three Iranian filmmakers were arrested with charges relating to "attempting to 'inflame and disrupt the psychological safety of the community'" [43]. The remaining increasing signals center around events such as national disasters (e.g., a building collapsing in Abadan [33]).

Türkiye - When investigating the 31 increasing signals for Türkiye, we find that there are 167 of 310 (53.8%) articles that contained breaking news. Seven signals from OONI and Hyperquack HTTPS center around the devastating earthquake that rocked Türkiye in early 2023. Over 380,000 people were left homeless by the earthquake [75] with a death toll surpassing 37,000. President Erdogan faced criticism over the response as evidence surfaced that building codes were ignored [38], funds for emergencies were spent on non-emergency infrastructure [75], and laws allowed builders to bypass inspections [40]. While the earthquake accounts for the most increasing signals, others coincide with the Russia-Ukraine War, Türkiye's inflation, and restrictions of free speech on social media platforms. The restrictions inhibiting free speech came during a time of significant concerns about free speech in Türkiye [14]. Each of the events represents significant public dissent, and as such is indicative of plausible censorship.

Pakistan - Our framework identifies 32 increasing signals for Pakistan, and we find 148 of the 320 news articles are breaking news. The signals coincided with events relating to the Taliban (e.g., the Taliban ending a ceasefire [17]), arresting of notable figures (e.g., Idris Khattak sentenced to 14 years of imprisonment [6]), and natural disasters (e.g., floods causing over $40 billion in damage [16]). In April of 2022, Prime Minister Imran Khan was ousted and opposition leader Shahbaz Sharif was voted in. We observe two OONI increasing signals related to this. One which leads into April, and the other starts right as Shahbaz Sharif is elected and continues to the end of May.

After looking at 1,650 news articles, our framework identifies changes in the censorship measurement data that corresponds to plausible censorship events for the countries considered. All but one signal coincides with an event that would elicit censorship to control the public narrative. Often the events relate to protests, war, or dissent of policy, and interestingly, we find that three countries censor traffic during natural disasters.

Although we cannot know the ground truth, we find examples in the news around each signal that motivate the possibility of censorship activity. Thus, we provide new insights into censorship events through the use of statistically-based analysis methods.

VIII. Discussion

A. Comparison to Previous Techniques

In this section, we motivate the validity of our statistical framework by comparing it against previous work. Tsai et al. [81] developed *CenDTect*, a system for identifying new censorship events and characterizing blocking rules. *CenDTect* processes Censored Planet data by creating decision trees for each domain then clustering the trees into censorship events using iterative parallel Density-Based Spatial Clustering of Applications with Noise (DBSCAN). *CenDTect* identifies censorship events based on the appearance of new clusters above a set threshold.

CenDTect detected 11 new censorship events from Hyperquack HTTP and HTTPS data during 2022. Tsai et al. validated 7 of the 11 censorship events using news media, similar to our validation procedure, and OONI data. We apply our framework to data from Hyperquack HTTP, Hyperquack HTTPS, OONI, and Tor over the same countries and time frame. *CenDTect* indicates at which AS the censorship events occur so we applied our framework to the data at the country level and data from those specific ASes. The signals located by our framework are shown in Table III. We identify an increasing signal in at least one of our datasets at the same time as each of the events identified by *CenDTect*.

To evaluate the false negative rate of *CenDTect* on reported censorship events, Tsai et al. gathered a list of 38 Potential Censorship Events (PCEL) from various media sources. Through unspecified manual analysis, they identified 12 of the events in the Censored Planet data and verified that the remaining 26 events do not exist within the Censored Planet data. Our framework locates 18 of the 38 PCEL events within the OONI, Tor, or Censored Planet data. Of the events our framework did not detect, 3 were absent from both Censored Planet and OONI measurements. Additionally, 9 of the undetected events lasted less than 2 days. Since our framework requires that at least 2 out of 3 data points show signal activity to confirm a signal (as outlined in Table II), it is unable to identify censorship events that last less than 2 days.

CenDTect offers valuable insights into blocking rules using Censored Planet data. However, its limitations prevent it from applying to other censorship measurement datasets, resulting in incomplete coverage of censorship events. *CenDTect* forms domain clusters based on the location of vantage points and blocking methods, information that is not present in Tor or GFWatch data. While OONI data does include this information, the domains measured by OONI change daily. For example, within the Iran OONI raw measurements, each domain was measured on average only 40% of the days with some domains only being measured one day during the entire time frame of our study. Because *CenDTect* creates decision trees based on the measured domains, this frequent change in the domain measurement list would cause the system to create new clusters too frequently, leading to a high false positive rate. In contrast, our framework calculates signals based on the metrics (e.g., censored domains, number of measurements, proportion of anomalies) available within the censorship measurement dataset, making it adaptable to various datasets and ensuring comprehensive coverage of censorship events.

B. Other Datasets

To demonstrate the flexibility of our framework, we applied it to additional datasets from The Internet Outage Detection and Analysis (IODA) project [25] that gather information on network activity beyond censorship. IODA collects four

TABLE III: **Results of our framework applied to data at the country level and the AS level over the same time frame and locations where *CenDTect* identified new censorship events. Validated indicates which events were validated by Censored Planet using media or OONI data (⋆). The remaining columns indicate the signals found by our framework - Blue ↑: increasing signal; Yellow ↓: decreasing signal; ∅: no available data; x: no signal**

		HQ HTTP		HQ HTTPS		OONI		Tor
Country	*Validated*	*Country*	*AS*	*Country*	*AS*	*Country*	*AS*	*Country*
Nepal		↑	x	↑	↑	x	∅	x
Venezuela		x	x	x	x	↑	↓	↑
Russia	[9]⋆	x	↑	x	x	↑	↑	x
Sri Lanka	[13]	x	↑	x	x	↑	↑	↑
Estonia	[15]	↓	x	x	x	↓	x	↑
Zimbabwe	⋆	↑	x	x	x	x	x	↓
Uganda		x	↑	x	x	x	x	↓
Burkina Faso		x	↑	↑	x	x	↑	↓
Zambia	[88]⋆	x	x	x	↑	x	x	x
Armenia	[9]⋆	x	↑	x	x	↑	∅	x
Iran	[11]⋆	↑	↑	x	↓	↑	↑	x

datasets with the following data points: (1) IPv4 and IPv6 prefixes visible to the majority of Border Gateway Protocol (BGP) peers calculated from Route Views data, (2) unique source IPs sending traffic to the Merit Network Telescope, (3) active network blocks based on custom active probing, and (4) visits to a specific Google product reported by the Google Transparency Report. On average 70% of the signals we identify in the IODA datasets overlap with at least one signal previously identified in the censorship measurement datasets revealing that major censorship events can disrupt broad network activity. Our framework works on the diverse IODA datasets regardless of their different data points because it uses historical data within each dataset to determine the signal threshold for that dataset. The full IODA signals are provided in Appendix C.

We do not incorporate other network traffic datasets such as Kentik [46], Lantern [51], and Psiphon [64] into our research because the data is not currently accessible. However, our framework can be applied to these datasets. Additionally, some platforms that provide open-source censorship measurement data lack a sufficient amount for our analysis. ICLab [58] has been previously compared to Censored Planet and OONI, but it stopped collecting measurements in 2020. Triplet Censors [18], which focused on measuring DNS injection by the Great Firewall, also halted data collection in 2020 and was superseded by GFWatch. CensorWatch [45], a mobile application measuring censorship in India, is in its early stages, but the authors have not provided any data beyond their initial study in 2020. As a result, there is currently not enough historical data for our framework to develop a detection threshold.

C. Limitations

Because our framework captures changes in censorship measurement data, signals that do not align directly with reported censorship events may instead indicate a modification in the platform's data collection process or an adjustment in the platform's user base. We identified one such instance in the Hyperquack HTTP data from Türkiye. Starting in December 2021, the data surpassed the UCL and oscillated around the

UCL until April 2022. Further investigation revealed that both the start and the end of this lengthy and unusual signal resulted from Censored Planet twice altering the number of measurements collected for this dataset in Türkiye. While we are unaware of any signals that are due to an increase in OONI or Tor users outside of censorship, these signals could additionally exist. Although these types of signals differ from censorship-related ones, they remain valuable as they contribute to understanding a dataset's structure, which is crucial in data analysis. It also reinforces the concept that the signals identified by our automated framework are a starting point for the investigation of potential censorship events and a way to reduce the amount of manual work required for censorship measurement data exploration.

IX. Related Work

OONI and Censored Planet are often employed individually to comprehend instances of censorship. OONI produces a range of reports on global censorship from this data. In 2023 alone, they reported on censorship in Brazil [32], Ethiopia [90], Kazakhstan [31], Azerbaijan [93], Russia [71], Pakistan [91], and Türkiye [92]. OONI often assumes the role of an initial informant on the state of censorship following significant global events (e.g., Russia's invasion of Ukraine [89] and the recent protests in Iran [21]). While OONI provides many high-level reports, the Censored Planet group performs comprehensive case studies, exploring specific occurrences of censorship in Russia [86], [87], Kazakhstan [66], and worldwide COVID-19 censorship [83]. While these approaches report on incidents and countries, our work fundamentally considers a larger and more difficult problem, unifying censorship detection grounded in statistical principles.

Researchers outside of the OONI and Censored Planet groups use these datasets to support new methods of censorship measurement and circumvention. In the research conducted by Bock et al. [23], Censored Planet's Quack data served as the training data for their genetic algorithm, Geneva, which identifies middleboxes engaged in TCP reflected amplification. Other techniques expanded the capabilities of OONI by incorporating a module specifically designed to measure the blocking of HTTP/3 [29].

X. Conclusion

In this work, we introduce a scalable and versatile statistical framework designed to recognize shifts, trends, and outliers in open-source censorship measurement data corresponding to potential censorship events to facilitate thorough investigation of censorship activity. We apply our framework to complete datasets from Censored Planet, OONI, GFWatch, and Tor, conducting an extensive time series and correlation analysis and revealing minimal signal overlap and limited correlation between the datasets, which underscores their diversity. Additionally, we perform a qualitative analysis of news articles corresponding to the time frame of increasing signals identified by our framework, linking these signals to world events as a preliminary step toward deeper investigation. Our methodology exemplifies the proper use of well-established statistical techniques to analyze large-scale measurement data comprehensively, highlighting the importance of integrating multiple datasets to avoid overlooking critical information. We provide[2] the source code for our framework, encouraging its use for comprehensive censorship analysis.

References

[1] "Users – Tor Metrics." [Online]. Available: https://metrics.torproject.org/userstats-relay-country.html

[2] "Iran charges aid worker zaghari-ratcliffe with 'propaganda against the system'," *France 24*, 2021. [Online]. Available: https://www.france24.com/en/middle-east/20210314-iran-charges-aid-worker-zaghari-ratcliffe-with-propaganda-against-the-system

[3] "Myanmar: Junta blocks lifesaving aid," *Human Rights Watch*, 2021. [Online]. Available: https://www.hrw.org/news/2021/12/13/myanmar-junta-blocks-lifesaving-aid

[4] "Myanmar: New internet blackout "heinous and reckless"," *Amnesty International*, 2021. [Online]. Available: https://www.amnesty.org/en/latest/press-release/2021/02/myanmar-new-internet-blackout/

[5] "Myanmar: Soldiers accused of shooting, burning 13 villagers," *BBC*, 2021. [Online]. Available: https://www.bbc.com/news/world-asia-59574528

[6] "Pakistan: Family of enforced disappearance victim idris khattak must be given clarity," *Amnesty International*, 2021. [Online]. Available: https://www.amnesty.org/en/latest/news/2021/12/pakistan-family-of-enforced-disappearance-victim-idris-khattak-must-be-given-clarity/

[7] "Two killed in iran's khuzestan water crisis protests," *Al Jazeera*, 2021. [Online]. Available: https://www.aljazeera.com/news/2021/7/17/one-man-shot-dead-in-irans-khuzestan-water-crisis-protests

[8] "UN 'horrified' by massacre of dozens of civilians in myanmar," *Al Jazeera*, 2021. [Online]. Available: https://www.aljazeera.com/news/2021/12/26/united-nations-horrified-myanmar-massacre-dozens-civilians

[9] "Armenia-Azerbaijan: Almost 100 killed in overnight clashes," *BBC*, 2022.

[10] "China: Censorship Mars Beijing Olympics," 2022. [Online]. Available: https://www.hrw.org/news/2022/02/18/china-censorship-mars-beijing-olympics

[11] "Fury in Iran as young woman dies following morality police arrest," *BBC*, 2022.

[12] "Imran Khan ousted as Pakistan's PM after vote," *BBC News*, 2022. [Online]. Available: https://www.bbc.com/news/world-asia-61055210

[13] "Sri Lanka in crisis · Global Voices," 2022.

[14] "Turkey's parliament adopts media law jailing those spreading 'disinformation'," *Reuter's*, 2022. [Online]. Available: https://www.reuters.com/world/middle-east/law-that-would-jail-those-spreading-disinformation-progresses-turkey-2022-10-13/

[15] "Estonia: Freedom on the Net 2022 Country Report," https://freedomhouse.org/country/estonia/freedom-net/2022, 2023.

[16] M. Ahmed, "Pakistan: World bank estimates floods caused \$40b in damages," *Associated Press*, 2022. [Online]. Available: https://apnews.com/article/floods-pakistan-south-asia-islamabad-25ee9dc0ec7aee6f4f2ef7b557216ee7

[17] ——, "Taliban ends ceasefire with pakistani government, orders new attacks across country," *Associated Press*, 2022. [Online]. Available: https://www.pbs.org/newshour/world/taliban-ends-ceasefire-with-pakistani-government-orders-new-attacks-across-country

[18] Anonymous, A. A. Niaki, N. P. Hoang, P. Gill, and A. Houmansadr, "Triplet Censors: Demystifying Great Firewall's DNS censorship behavior," in *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2020.

[19] H. Arslan, S. Yücel, and S. Ata, "Investigating the Development Trend of Turkish Economy between 2006-2019 through Statistical Methods and Financial Statement Analysis," *International Academic Social Resources Journal*, 2022.

[2] Available at https://github.com/censorship-event-detection/acsac-censorship-event-detection/

[20] N. Bahria, D. I. Harbaoui, A. Chelbi, and H. Bouchriha, "Joint design of control chart, production and maintenance policy for unreliable manufacturing systems," *Journal of Quality in Maintenance Engineering*, 2020.

[21] S. Basso, M. Xynou, A. Filastò, and A. Meng, "Iran blocks social media, app stores and encrypted DNS amid Mahsa Amini protests," 2022. [Online]. Available: https://ooni.org/post/2022-iran-blocks-social-media-mahsa-amini-protests/

[22] D. Batty, "Media face web censorship at Beijing olympics," 2008. [Online]. Available: https://www.theguardian.com/world/2008/jul/30/china.olympicgames2008

[23] K. Bock, G. Naval, K. Reese, and D. Levin, "Even Censors Have a Backup: Examining China's Double HTTPS Censorship Middleboxes," in *Proceedings of the ACM SIGCOMM Workshop on Free and Open Communications on the Internet (FOCI)*, 2021.

[24] D. Braun, E. Braun, V. Chiu, A. E. Burgos, M. Gupta, M. Volodarskiy, and D. Getahun, "Trends in Neonatal Intensive Care Unit Utilization in a Large Integrated Health Care System," *JAMA Network Open*, 2020.

[25] CAIDA, "IODA: Internet outage detection and analysis." [Online]. Available: https://ioda.caida.org

[26] K.-C. Chang, W. R. Hobbs, M. E. Roberts, and Z. C. Steinert-Threlkeld, "Covid-19 increased censorship circumvention and access to sensitive topics in china," *Proceedings of the National Academy of Sciences*, 2022.

[27] B. Church, "Iran players remain silent during national anthem at World Cup in apparent protest at Iranian regime," 2022. [Online]. Available: https://www.cnn.com/2022/11/21/football/iran-national-anthem-england-2022-world-cup-spt-intl/index.html

[28] R. B. Dull and D. P. Tegarden, "Using control charts to monitor financial reporting of public companies," *International Journal of Accounting Information Systems*, 2004.

[29] K. Elmenhorst, B. Schütz, N. Aschenbruck, and S. Basso, "Web censorship measurements of HTTP/3 over QUIC," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2021.

[30] A. Filastò and J. Appelbaum, "OONI: Open observatory of network interference," in *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2012.

[31] A. Filastò and S. Basso, "Throttling of news media amid Kazakhstan's 2022 presidential election," 2023. [Online]. Available: https://ooni.org/post/2023-throttling-kz-elections/

[32] A. Filastò, M. Xynou, S. Basso, and F. O. D. Busnello, "Brazil: OONI data on the blocking of Telegram," 2023. [Online]. Available: https://ooni.org/post/2023-brazil-telegram/

[33] A. France-Presse, "Death toll from iran tower block collapse rises to 19," *VOA*, 2022. [Online]. Available: https://www.voanews.com/a/death-toll-from-iran-tower-block-collapse-rises-to-19-/6591734.html

[34] A. Gul, "More Than 300,000 Afghans Flee to Pakistan Since Taliban Takeover of Afghanistan," 2021. [Online]. Available: https://www.voanews.com/a/more-than-300-000-afghans-flee-to-pakistan-since-taliban-takeover-of-afghanistan-/6357777.html

[35] A. Hashim, "Pakistani Taliban ends ceasefire, future of peace talks uncertain," 2021. [Online]. Available: https://www.aljazeera.com/news/2021/12/10/pakistan-taliban-ceasefire-peace-talks-ttp

[36] N. P. Hoang, A. A. Niaki, J. Dalek, J. Knockel, P. Lin, B. Marczak, M. Crete-Nishihata, P. Gill, and M. Polychronakis, "How great is the Great Firewall? Measuring China's DNS censorship," in *Proceedings of the USENIX Security Symposium*, 2021.

[37] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*. John Wiley & Sons, 1973, pp. 1991–1992.

[38] J. Horton and W. Armstrong, "Turkey earthquake: Why did so many buildings collapse?" *BBC*, 2023. [Online]. Available: https://www.bbc.com/news/64568826

[39] F. House, "Turkey: Freedom on the Net 2021 Country Report," 2021. [Online]. Available: https://freedomhouse.org/country/turkey/freedom-net/2021

[40] E. Ince and L. Sackur, "As turkey's earthquake death toll grows, so does anger at government," *NBC*, 2023. [Online]. Available: https://www.nbcnews.com/news/world/turkey-earthquakes-death-toll-anger-erdogan-ak-party-amnesty-rcna71400

[41] V. Isachenkov, "Putin signs law allowing him 2 more terms as russia's leader," *PBS*, 2021. [Online]. Available: https://www.pbs.org/newshour/world/putin-signs-law-allowing-him-2-more-terms-as-russias-leader

[42] J. Jett, A. Ling, and H. Lee, "For china, moving away from 'zero-covid' is easier said than done," *NBC*, 2022. [Online]. Available: https://www.nbcnews.com/news/world/china-outbreaks-zero-covid-rcna58015

[43] K. E. D. John, Tara, "Iran has arrested three filmmakers in less than a week," *CNN*, 2022. [Online]. Available: https://www.cnn.com/2022/07/12/middleeast/iran-jafar-panahi-filmmaker-arrest-intl/index.html

[44] D. Kang, "China's xi faces threat from public anger over 'zero covid'," *Associated Press*, 2022. [Online]. Available: https://apnews.com/article/health-china-arrests-beijing-xi-jinping-1596604782d165d268d1dbc137034749

[45] D. Katira, G. Grover, K. Singh, and V. Bansal, "CensorWatch: On the Implementation of Online Censorship in India," in *Proceedings of the Workshop on Free and Open Communications on the Internet (FOCI)*, 2023.

[46] Kentik. [Online]. Available: https://www.kentik.com/

[47] A. Kohli, "Iran Protests Over Mahsa Amini's Death: What to Know," 2022. [Online]. Available: https://time.com/6216513/mahsa-amini-iran-protests-police/

[48] L. Kuo, "Is China ready to host the winter olympics?" *The Washington Post*, 2022. [Online]. Available: https://www.washingtonpost.com/sports/olympics/2022/01/07/china-winter-olympics-beijing-omicron/

[49] P. N. N. Laboratory, "Mann-kendall test for monotonic trend." [Online]. Available: https://vsp.pnnl.gov/help/vsample/design_trend_mann_kendall.htm

[50] G. Lacombe, M. McCartney, and G. Forkuor, "Drying climate in Ghana over the period 1960–2005: Evidence from the resampling-based Mann-Kendall test at local and regional levels," *Hydrological Sciences Journal*, 2012.

[51] Lantern. [Online]. Available: https://getlantern.org/en_US/

[52] S. Laplante and V. Fortuna, "Censored Planet Observatory's Data Analysis Pipeline." [Online]. Available: https://github.com/censoredplanet/censoredplanet-analysis/blob/master/table/queries/derived_satellite_scans.sql

[53] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia Medica*, 2012.

[54] A. Moaveni, "The protests inside iran's girls' schools," *The New Yorker*, August 2023. [Online]. Available: https://www.newyorker.com/magazine/2023/08/14/the-protests-inside-irans-girls-schools

[55] D. C. Montgomery, *Introduction to Statistical Quality Control*. Wiley, 2013.

[56] R. Mostaghim, "At least three people killed in protests over water shortages in iran," *CNN*, 2021. [Online]. Available: https://www.cnn.com/2021/07/25/middleeast/iran-water-shortages-protests-intl/index.html

[57] NCSS, "Confidence Intervals for Kendall's Tau-b Correlation," *PASS Sample Size Software*, 2023.

[58] A. A. Niaki, S. Cho, Z. Weinberg, N. P. Hoang, A. Razaghpanah, N. Christin, and P. Gill, "ICLab: A Global, Longitudinal Internet Censorship Measurement Platform," in *Proceedings of the IEEE Symposium on Security & Privacy (SP)*, 2020.

[59] OONI, "Analyzing raw ooni data, a case study." [Online]. Available: https://ooni.org/notebooks/tutorial-russia-data-analysis-case-study.html

[60] Open Observatory of Network Interference (OONI), "Web connectivity," 2021. [Online]. Available: https://ooni.org/nettest/web-connectivity/

[61] R. Padmanabhan, A. Filastò, M. Xynou, R. S. Raman, K. Middleton, M. Zhang, D. Madory, M. Roberts, and A. Dainotti, "A multi-perspective view of Internet censorship in Myanmar," in *Proceedings of the ACM SIGCOMM Workshop on Free and Open Communications on the Internet (FOCI)*, 2021.

[62] P. Pearce, R. Ensafi, F. Li, N. Feamster, and V. Paxson, "Augur: Internet-wide detection of connectivity disruptions," in *Proceedings of the IEEE Symposium on Security & Privacy (SP)*, 2017.

[63] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, "Global measurement of DNS manipulation," in *Proceedings of the USENIX Security Symposium*, 2017.

[64] Psiphon, "Psiphon data engine dashboard." [Online]. Available: https://psix.ca/d/Vcvqvj6Wk/global-overview?orgId=2

[65] L. Puka, *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, 2011, pp. 713–715.

[66] R. S. Raman, L. Evdokimov, E. Wustrow, J. A. Halderman, and R. Ensafi, "Investigating large scale HTTPS interception in Kazakhstan," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2020.

[67] R. S. Raman, P. Shenoy, K. Kohls, and R. Ensafi, "Censored Planet: An Internet-wide, longitudinal censorship observatory," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2020.

[68] R. S. Raman, A. Stoll, J. Dalek, R. Ramesh, W. Scott, and R. Ensafi, "Measuring the deployment of network censorship filters at global scale," in *Proceedings of the The Internet Society Network and Distributed System Security Symposium (NDSS)*, 2020.

[69] R. S. Raman, A. Virkud, S. Laplante, V. Fortuna, and R. Ensafi, "Advancing the art of censorship data analysis," in *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2023.

[70] R. Ramesh, R. S. Raman, A. Virkud, A. Dirksen, A. Huremagic, D. Fifield, D. Rodenburg, R. Hynes, D. Madory, and R. Ensafi, "Network responses to Russia's invasion of Ukraine in 2022: A cautionary tale for Internet freedom," in *Proceedings of the USENIX Security Symposium*, 2023.

[71] Roskomsvoboda and OONI, "How Internet censorship changed in Russia during the 1st year of military conflict in Ukraine," 2023. [Online]. Available: https://ooni.org/post/2023-russia-a-year-after-the-conflict/

[72] J. Ruscio, "Constructing Confidence Intervals for Spearman's Rank Correlation with Ordinal Data: A Simulation Study Comparing Analytic and Bootstrap Methods," *Journal of Modern Applied Statistical Methods*, 2008.

[73] R. Sandler, "Social Media Blackout: Myanmar Military Government Blocks Twitter, Instagram," 2021. [Online]. Available: https://www.forbes.com/sites/rachelsandler/2021/02/05/social-media-blackout-myanmar-military-government-blocks-twitter-instagram/?sh=40409fec2f05

[74] G. T. Sariyuce, Isil, "Turkey withdraws from Istanbul convention to combat violence against women," 2021. [Online]. Available: https://www.cnn.com/2021/03/20/europe/turkey-convention-violence-women-intl/index.html

[75] R. Schmitz, "As the earthquake death toll soars, so does criticism of turkey's government response," *NPR*, 2023. [Online]. Available: https://www.npr.org/2023/02/09/1155647266/turkey-earthquake-erdogan-government-response-criticism

[76] W. Scott, T. Anderson, T. Kohno, and A. Krishnamurthy, "Satellite: Joint analysis of CDNs and network-level interference," in *Proceedings of the USENIX Annual Technical Conference*, 2016.

[77] G. Suman and D. Prajapati, "Control chart applications in healthcare: A literature review," *International Journal of Metrology and Quality Engineering*, 2018.

[78] Tor, "The Tor Project: Privacy & freedom online." [Online]. Available: https://www.torproject.org

[79] A. Troianovski, "Russia takes censorship to new extremes, stifling war coverage," *The New York Times*, 2022. [Online]. Available: https://www.nytimes.com/2022/03/04/world/europe/russia-censorship-media-crackdown.html

[80] E. Tsai, D. Kumar, R. Sundara Raman, G. Li, Y. Eiger, and R. Ensafi, "CERTainty: Detecting DNS Manipulation at Scale using TLS Certificates," in *Proceedings on Privacy Enhancing Technologies*, 2023.

[81] E. Tsai, R. S. Raman, A. Prakash, and R. Ensafi, "Modeling and Detecting Internet Censorship Events," in *Proceedings of the The Internet Society Network and Distributed System Security Symposium (NDSS)*, 2024.

[82] B. VanderSloot, A. McDonald, W. Scott, J. A. Halderman, and R. Ensafi, "Quack: Scalable remote measurement of application-layer censorship," in *Proceedings of the USENIX Security Symposium*, 2018.

[83] A. Vyas, R. S. Raman, N. Ceccio, P. M. Lutscher, and R. Ensafi, "Lost in transmission: Investigating filtering of COVID-19 websites," in *Financial Cryptography and Data Security*, 2021.

[84] Western Electric Co., *Statistical Quality Control Handbook*. AT&T, 1956.

[85] H. Wu and D. Kang, "China weighs exit from 'zero covid' and the risks involved," *Associated Press*, 2022. [Online]. Available: https://apnews.com/article/covid-health-business-infectious-diseases-public-health-732de4c302e379ac33c92d20c425ae2e

[86] D. Xue, B. Mixon-Baca, ValdikSS, A. Ablove, B. Kujath, J. R. Crandall, and R. Ensafi, "TSPU: Russia's decentralized censorship system," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2022.

[87] D. Xue, R. Ramesh, ValdikSS, L. Evdokimov, A. Viktorov, A. Jain, E. Wustrow, S. Basso, and R. Ensafi, "Throttling Twitter: An Emerging Censorship Technique in Russia," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2021.

[88] M. Xynou and A. Filastò, "Zambia: Social media blocked amid 2021 general elections," https://ooni.org/post/2021-zambia-social-media-blocks-amid-elections/, 2021.

[89] ——, "New blocks emerge in Russia amid war in Ukraine: An OONI network measurement analysis," 2022. [Online]. Available: https://ooni.org/post/2022-russia-blocks-amid-ru-ua-conflict/

[90] ——, "Ethiopia: Ongoing blocking of social media," 2023. [Online]. Available: https://ooni.org/post/2023-ethiopia-blocks-social-media/

[91] ——, "Pakistan: Blocking of Wikipedia and Deutsche Welle (DW)," 2023. [Online]. Available: https://ooni.org/post/2023-pakistan-blocks-wikipedia-and-dw/

[92] ——, "Turkey: Throttling and DNS blocking of Twitter following deadly earthquake," 2023. [Online]. Available: https://ooni.org/post/2023-turkey-throttling-blocking-twitter/

[93] M. Xynou and A. Geybulla, "OONI measurements show ongoing internet censorship in Azerbaijan," 2023. [Online]. Available: https://ooni.org/post/2023-azerbaijan-internet-censorship/

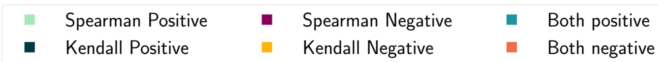
[94] A. Zad, "When Will Iran's Internet Censorship Collapse?" *Slate*, 2022. [Online]. Available: https://slate.com/technology/2022/09/iran-protests-mahsa-amini-internet-censorship.html

APPENDIX

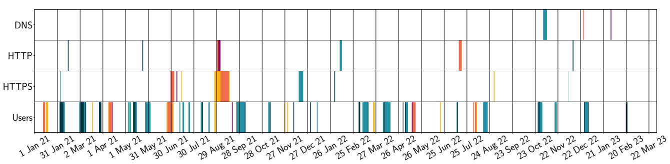
A. Correlation Overlap

The following figures show the windows of time where positive or negative correlation exists between datasets paired by protocol across the entire analysis time frame for each country.

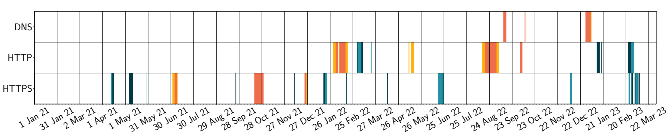
(a) **Correlation Legend**



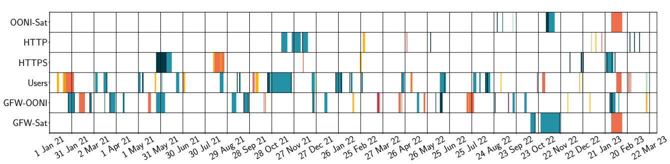
(b) **Russia Correlation Windows**



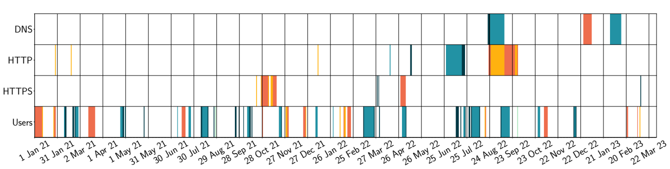
(c) **Myanmar Correlation Windows**



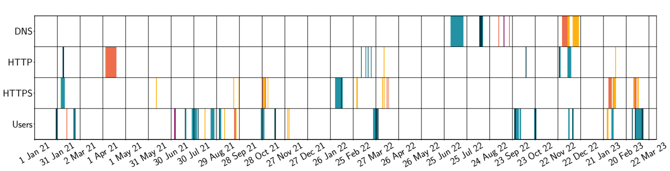
(d) **China Correlation Windows**



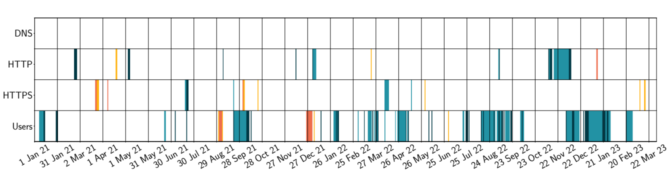
(e) **Iran Correlation Windows**



(f) **Türkiye Correlation Windows**



(g) **Pakistan Correlation Windows**

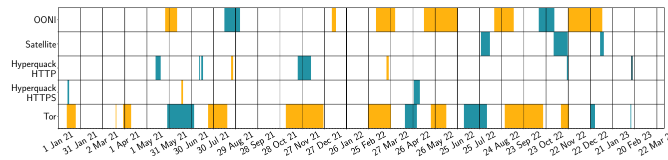


B. Signal Summary

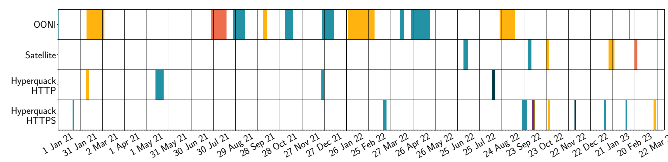
The following is the full version of the figures shown in Section VI. They show a summary of the signals identified across the entire analysis time frame for each country.

(a) **Signals Legend**

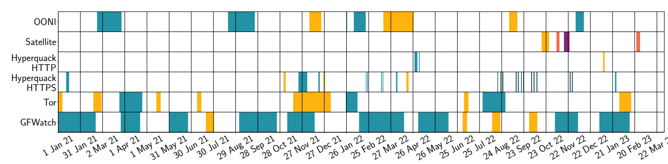
(b) **Russia Signals**



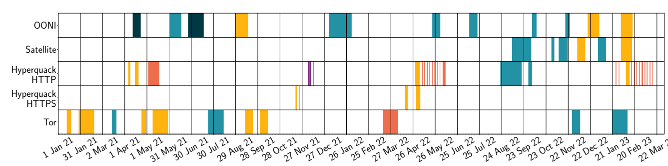
(c) **Myanmar Signals**



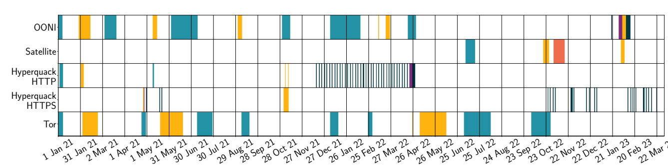
(d) **China Signals**



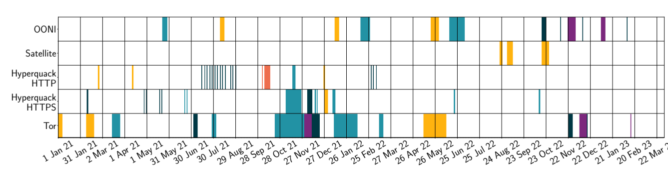
(e) **Iran Signals**



(f) **Türkiye Signals**



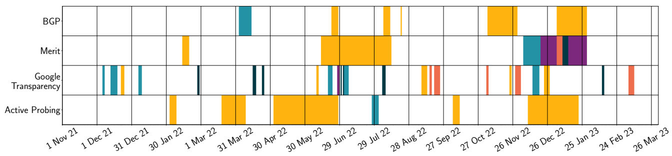
(g) **Pakistan Signals**



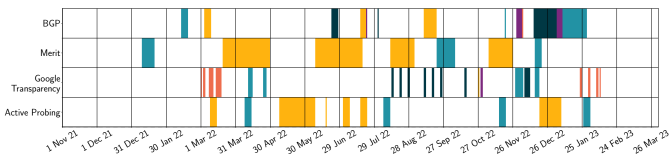
C. IODA Signal Summary

The following is a summary of the signals identified in the IODA datasets as discussed in Section VIII.

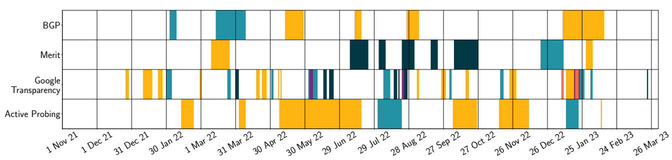
(a) **Russia IODA Signals**



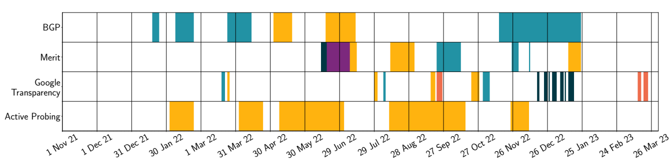
(b) **Myanmar IODA Signals**



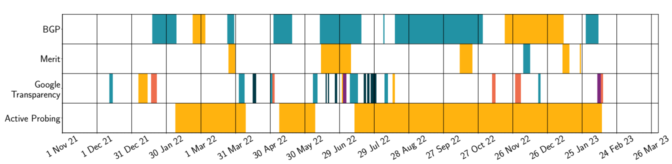
(c) **China IODA Signals**



(d) **Iran IODA Signals**



(e) **Türkiye IODA Signals**



(f) **Pakistan IODA Signals**