Extracting Benefit from Harm: Using Malware Pollution to Analyze the Impact of Political and Geophysical Events on the Internet

Alberto Dainotti
University of Napoli Federico II
alberto@unina.it

Roman Ammann
Auckland University of Technology
roman@xxo.ch

Emile Aben
RIPE NCC
emile.aben@ripe.net

Kimberly C. Claffy
CAIDA/UCSD
kc@caida.org

ABSTRACT

Unsolicited one-way Internet traffic, also called *Internet background radiation (IBR)*, has been used for years to study malicious activity on the Internet, including worms, DoS attacks, and scanning address space looking for vulnerabilities to exploit. We show how such traffic can also be used to analyze macroscopic Internet events that are unrelated to malware. We examine two phenomena: country-level censorship of Internet communications described in recent work [17], and natural disasters (two recent earthquakes). We introduce a new metric of local IBR activity based on the number of unique IP addresses per hour contributing to IBR. The advantage of this metric is that it is not affected by bursts of traffic from a few hosts. Although we have only scratched the surface, we are convinced that IBR traffic is an important building block for comprehensive monitoring, analysis, and possibly even detection of events unrelated to the IBR itself. In particular, IBR offers the opportunity to monitor the impact of events such as natural disasters on network infrastructure, and in particular reveals a view of events that is complementary to many existing measurement platforms based on (BGP) control-plane views or targeted active ICMP probing.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network Monitoring;
C.2.5 [**Local and Wide-Area Networks**]: Internet

General Terms

Measurement, Performance

Keywords

Outages, Censorship, Earthquakes, Network Telescope, Malware

1. INTRODUCTION

Internet background radiation (IBR) [35, 44] is unsolicited one-way network traffic sent to random IP addresses as a result of inadvertent or malicious software behavior, e.g., worms, scanning, misconfigurations. Such a vast number of computers generate such background radiation, mostly unbeknownst to their users, that the resulting traffic aggregate has proven a useful source of data for observing characteristics of the malware itself [16, 32, 33] not revealed by other types of data. In this work we show that IBR traffic can also yield insight into other macroscopic events severe enough to affect Internet connectivity, but unrelated to the malware itself. In particular we show how political events such as large-scale Internet censorship by nation-states, or natural disasters such as earthquakes, can drastically affect the IBR coming from a region, and in doing so illuminate some characteristics of the events and their impact on communications capabilities.

There are three primary causes of IBR traffic: (i) *backscatter* from spoofed denial-of-service (DoS) attacks, (ii) scans, or (iii) bugs and misconfiguration [35]. A DoS attack attempts to overwhelm a victim with traffic or transaction requests in order to reduce or prevent his ability to serve legitimate requests. When the source IP addresses in attacking packets are randomly spoofed, the response packets (e.g. *SYN-ACK* TCP packets in reply to *SYNs* from the attacker) are sent back to the spoofed addresses, producing *backscatter*, which will be captured by telescopes [33] that happen to contain (and thus observe traffic to) those spoofed addresses.

Automated (e.g. from worms) or manually initiated random scanning of address space in search of victims is another component of IBR captured by network telescopes [16, 32].

Misconfiguration of systems can also induce IBR traffic, for example by setting the wrong IP address on a DNS or proxy server. Bugs in network applications and router firmware and software e.g., getting byte ordering wrong, can assign incorrect network addresses to a device, triggering unsolicited traffic in response.

Researchers observe IBR traffic using *network telescopes*, often called *darknets* if the IP addresses are not assigned to devices. Commercial as well as academic research groups have used darknets to support security-related data analysis [3, 21, 44].

In this work we analyze data captured by the UCSD network telescope, which attracts IBR for a /8 network, or 1/256th of the IPv4 address space (about 16.7 million IP addresses) [7]. As of August 2011, the UCSD telescope captures approximately 2-10 GB traffic (mostly packet headers with no payload) every hour. The data is stored for further analysis and protected sharing.

Section 2 provides an overview of research using Internet data sources to analyze macroscopic events. Section 3 summarizes our previous analysis [17] of two country-wide Internet censorship episodes, in Libya and Egypt in early 2011. Section 4 introduces the use of IBR traffic to study two earthquakes that struck New Zealand and Japan in February and March 2011, both strong enough to damage power and communications infrastructure, significantly reducing the number of computers near the epicenters that could access

the Internet, thus reducing IBR traffic. Section 5 discusses limitations of the method and proposes future directions.

2. RELATED WORK

Events in the physical world, as distinct as earthquakes and revolutions, can have observable macroscopic effects on the Internet. This section surveys commonly used methods to study such effects.

Border Gateway Protocol (BGP). In BGP-based Internet interdomain routing, Autonomous Systems (ASes) announce and withdraw reachability information for their network prefixes to the rest of the Internet. Major changes in the number of prefixes announced or withdrawn from a group of ASes, for instance a country-based grouping, are readily detectable from publicly available BGP data. Since these major changes propagate across the global Internet, it is not necessary to capture observations physically close to an event to detect it using BGP. Several research groups have used BGP as the primary data source in analyzing major events, including Hurricane Katrina [28], the 2011 Japan earthquake [13], the Georgia-Russia conflict [45], and the 2011 Egyptian [11] and Libyan [14] revolutions. Li and Brooks proposed a BGP-based "Internet seismograph" [28] to quantitatively assess and compare the impacts on normal BGP behavior of disruptive events such as large-scale power outages, undersea cable cuts and worms. Their proposed system monitors the values of a set of distinct BGP attributes over one-minute time bins in BGP messages exchanged on the Internet and detects and characterizes deviations from a profile defined over a reference period.

Active Probing. Active probing using ICMP echo-requests, or traceroute-like measurements, is a means of troubleshooting IP-layer connectivity problems, and is sometimes used to support BGP-based analyses [11]. Analyses based solely on active probing data, for instance the PingER Project's [42] analysis of performance during the 2011 Japan earthquake [10], can also offer insights into connectivity, in case the probing happens to cross infrastructure affected by the disruption. Several research groups use active probing to study the global Internet topology [8, 29, 38, 40], but these data sources typically probe too slowly or from too few vantage points to reconstruct a comprehensive timeline of a macrosopic event. The Hubble system [20] was a prototype that demonstrated the potential of combining event detection based on BGP data and active measurements of router-infrastructure, to identifying how many prefixes are reachable from some vantages and not others, how often these problems occur, and how long they persist.

Passive Traffic Measurement. Traffic data could also inform analyses of macroscopic Internet events, although so little traffic data is available to researchers that it is unlikely to capture macroscopic events of interest. As aggregation points of many different traffic sources, Internet Exchange Points (IXPs) are more likely to see macroscopic events than single backbone links, but typically publish only coarse-grained traffic numbers, which at most are useful to augment other methods and data sources, such as in [13]. In order to detect and analyze Internet-wide disruptions, data collection must occur in the core of many major ASes, such as in Arbor Networks's commercial traffic collection infrastructure [27] which Arbor has used to analyze national Internet disruptions in Myanmar [23] and Georgia [34] as well as the recent censorship episodes that we studied in Egypt and Libya [22, 24, 25].

Google Services. The near-omnipresence of Google, or perhaps more accurately, of its user base, makes it an interesting source of data on Internet-wide disruptions. Cowie [14] uses the Google Transparency Report [18] to show distinct drops in the use of Google services at the beginning of the conflict in Libya. In a later analysis he shows that the same approach does not support analysis of Internet usage during the 2011 Japan earthquake [13]. The first limitation is that Google's data describes a subset of user behavior, which only indirectly relates to the state of the network. For example, if Internet links are physically severed, people behind these links will not be able to use Internet services, but other people in the earthquake-affected areas that still have connectivity might increase their use of the Internet to find information about the event or to communicate with friends and relatives. The second limitation is the coarse granularity used in data curation. At the moment no raw data is available, only country-level aggregated statistics, which does not support independent analysis of events within a specific region of a country. Finally, use of Google services is not globally ubiquitous, and sometimes is blocked or censored. These arguments hold for data from other large-scale distributed service providers such as Twitter or Facebook.

Peer-to-Peer Traffic. Peer-to-peer (P2P) networks are another source of data to study Internet-wide disruptions. Bischof and Otto used the BitTorrent network, specifically data from a plugin of the Vuze BitTorrent client, to analyze the disconnection of Egypt and Libya from the Internet [5] and the 2011 Japan earthquake [6]. Assuming a certain dispersion and activity of active P2P clients, this approach is promising, although its strength depends on the popularity of the protocol or application under examination, in particular its usage during the event under study. In the case of P2P, the P2P activity of the end users (installing the client, downloading or seeding content) is imperative, and not all parts of the world use the same P2P networks. Since IBR traffic is mostly generated by malware, our method does not depend on end user actions or even awareness that the IBR traffic is being generated.

Internet Background Radiation. In [9], Casado et al. proposed for the first time the use of IBR for "opportunistic measurement". They argued that IBR is a viable and unique source of Internet measurement data, and showed that by analyzing specific categories of IBR they could infer several properties (e.g. access link bandwidth, NAT usage) of the infected hosts generating such traffic.

In [2] and [17] we used IBR as a novel method to analyze large-scale Internet disruption. In [17] we combined observations of IBR, BGP updates, and active measurements from the Ark Project [8], and showed how they complement each other in our analyses of political censorship episodes. In particular the IBR traffic data allowed us to detect data-plane disruptions such as packet filtering even when there was no disruption of the (BGP) control plane, which revealed additional insight into how these governments implemented their censorship policies. In this work we explore a generalization of this idea as the basis for a methodology for analyzing geophysical events affecting communications infrastructure, using case studies of two recent severe earthquakes. We introduce a new metric of regional activity based on the number of unique IP addresses per hour contributing to IBR, which is not affected by bursts of traffic from a few hosts, and provides a lightweight efficient indicator of regional Internet service disruptions.

3. CENSORSHIP

On January 27, 2011 around 22:34 UTC, several sources reported the withdrawal of almost all routes to Egyptian networks in observable global Internet routing tables [4, 11, 26]. The outage lasted for more than five days, during which additional routes, i.e., active BGP IPv4 prefixes, in Egypt were withdrawn. As newspapers and telecom operators reported [37], the disruption of Internet connectivity was ordered by the Egyptian government in response to the protests in the country, which eventually led to the resignment of President Hosni Mubarak.

A few days later, similar protests erupted in Libya. On the night

of February 18 the government imposed a block on all Internet access until morning, and repeated it the next day. On the 3rd of March, Internet access was disabled again for nearly four days.

In [17] we analyzed these events using three different types of data available to researchers: (i) BGP updates collected by Route-Views [43] and RIPE NCC's Routing Information Service (RIS) [39]; (ii) active traceroute probing from Ark [8]; and (iii) IBR from the UCSD network telescope. For the IBR traffic data, we first isolated all the traffic from IP addresses that geolocated to Egypt and Libya for a period of time including the outages. For IP geolocation we used two databases: the AfriNIC Regional Internet Registry [1] and the MaxMind GeoLite Country database [31].

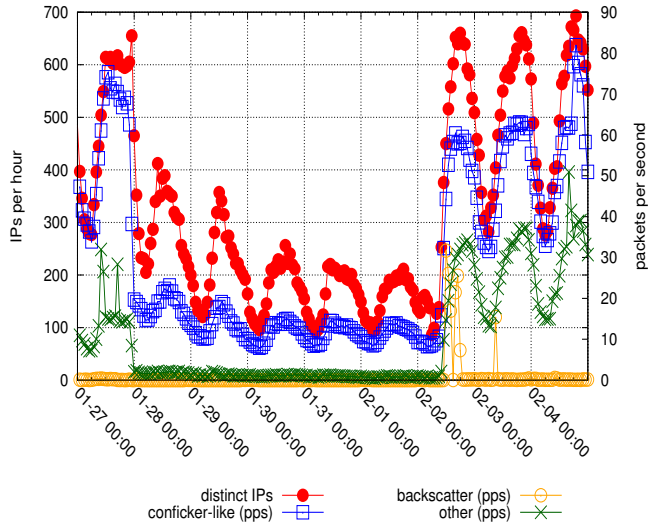


Figure 1: Unsolicited traffic from IPs geolocated in Egypt to UCSD's network telescope: number of distinct source IP addresses observed every hour and packet rate per category (*Conficker-like, backscatter, other*). The large drop in both distinct IP addresses and packets reaching the telescope corresponds to the Egyptian outage. The sudden peaks in backscatter traffic are due to large denial-of-service attacks each to (and thus causing backscatter from) a single victim. Such spikes substantially increase the overall packet rate but not the rate of distinct IP addresses.

IBR traffic observed at a darknet from a specific area can vary significantly based on the different contributing behaviors. Figure 1 shows two metrics of traffic reaching the UCSD darknet for a nine-day period covering the January episode of censorship in Egypt: the number of distinct IP addresses that geolocated to Egypt and reached the darknet (left y-axis), and the packet rate from those IP addresses geolocated in Egypt, separated into three categories (y-axis). The Conficker-like traffic category refers to TCP SYN packets with destination port 445 and packet size 48 bytes, presuming that these packets are generated by systems infected by the Conficker worm [36]. In contrast, backscatter traffic is generated in response to spoofed-IP-address denial of service attacks. We see significant levels of such traffic only when a large attack targets a victim in the address space we observe, which happened during our study interval for Egypt. Figure 1 shows three large denial of service attacks between 2-4 February 2011. The third category is "other" IBR traffic, mostly due to scanning activity, sometimes carried out by botnet-infected computers [17].

To mitigate the impact of one or few hosts dramatically affecting the unsolicited traffic rate, which is particularly a concern when we are examining a small subset of IBR traffic, e.g., from a small country, we used the rate of distinct IP source addresses. Figure 1 confirms that this metric is not significantly affected by spikes in the backscatter and "other" traffic categories. The number of distinct IPs per hour suddenly decreased after 22:00 UTC on January 27, and returned to pre-outage levels after 10:00 UTC on February 2, consistent with other reports of when the outage started and ended [12,24]. Some IBR traffic remained even during the outage for two reasons: (i) some network prefixes were not withdrawn and (ii) if not caused by the data-plane going down, a BGP withdrawal may only affect inbound connectivity, outbound packets can still be sent if a network uses default routing for upstream connectivity.

The gradual decrease in the rates of both unique IP addresses and packets during the outage is due to the progressive withdrawal of more BGP routes that during the first day were kept reachable [11].

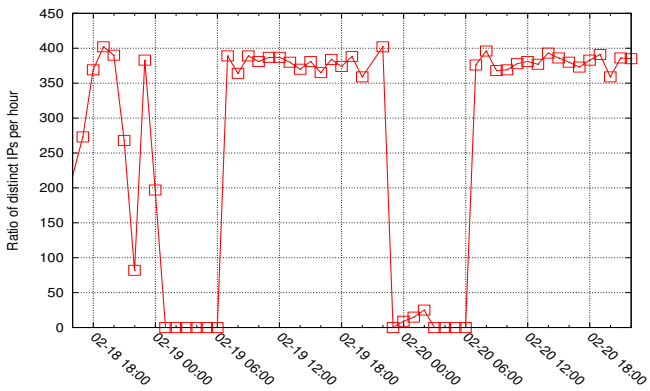


Figure 2: Unsolicited traffic to UCSD's network telescope originating from Libya during the first two Libyan outages. The rate of distinct IPs per hour drops to approximately zero during both outages. The small peaks during the second outage are due to a different blocking technique that was put in place (packet filtering) which kept communication alive for a few networks.

In Libya, three different outages have been identified and publicly recognized [14], all reflected in variations in the IBR IP-address rate observed by the UCSD network telescope. Figure 2 shows a view from the telescope of the first two Libyan outages. Normal values for Libyan IBR to this telescope are around 380 distinct IPs per hour, while during both outages the IP rate dropped to almost zero, consistent with other reports [14,15,17]. The analysis we conducted in [17] revealed that the second outage was implemented in a first phase through BGP withdrawals, and then, while BGP routes were restored, by deploying packet filtering at a country-level granularity. This insight demonstrated the potential of network telescopes to detect large-scale connectivity disruptions that are not detectable by monitoring BGP connectivity.

4. EARTHQUAKES

In this section we look at the earthquakes that struck Christchurch, New Zealand on 22 February 2011 and Tohoku, Japan on 11 March 2011. These examples illustrate how IBR reflects the impacts of such natural disasters on network infrastructure. We introduce a lightweight but effective metric to analyze such impacts. Table 1 lists key characteristics of the two earthquakes.

We use the MaxMind GeoLite City database [31] to calculate the *great-circle distance* [41] from a given network to the epicenters of the earthquakes. Table 2 shows the number of IP hosts geolocated to within increasing radii of each epicenters. The two most striking contrasts between the two earthquake epicenters are: (1) how much further the Tohoku epicenter (which was in the Pacific ocean) was from any significant population of IP addresses (100 km); and (2)

	Christchurch - NZ	Tohoku - JP
Magnitude	6.1	9.0
Latitude	43.583 °S	38.322 °N
Longitude	172.701 °E	142.369 °E
Date UTC	21.02.2011, 23:51:42 UTC	11.03.2011, 05:46:23 UTC
Date Localtime	22.02.2011, 12:51:42 PM	11.03.2011, 02:46:23 PM

Table 1: Key characteristics of the two earthquakes. Christchurch is on the southern island of New Zealand. Tohoku is in the northeastern region of Honshu, the largest island of Japan.

	Christchurch - NZ		Tohoku - JP	
Distance (Km)	Networks	IP Addresses	Networks	IP Addresses
< 5	1	255	0	0
< 10	283	662,665	0	0
< 20	292	732,032	0	0
< 40	299	734,488	0	0
< 80	309	738,062	5	91
< 100	310	738,317	58	42,734
< 200	348	769,936	1,352	1,691,560
< 300	425	828,315	3,953	4,266,264
< 400	1,531	3,918,964	16,182	63,637,753
< 500	1,721	4,171,527	41,522	155,093,650

Table 2: Networks and IP addresses within a given distance to the epicenters.

how many more IP addresses are within 500 km of Tohoku's epicenter (which includes Tokyo), consistent with the orders of magnitude larger population in Japan.

In search of a metric to express the effects of the disasters on nearby Internet infrastructure, we computed the number of distinct source IP addresses per hour seen by the telescope over two contiguous 24-hour periods before and after earthquake. We define $I_{\Delta t_i}$ the number of distinct source IP addresses seen by the telescope over the interval Δt_i, where $\Delta t_1, ..., \Delta t_n$ are 1-hour time slots following the event and $\Delta t_{-1}, ..., \Delta t_{-n}$ are those preceding it. We then define the ratio θ as in Eq. 1,

$$\theta = \frac{\sum_{i=-1}^{-24} I_{\Delta t_i}}{\sum_{j=1}^{24} I_{\Delta t_j}} \quad (1)$$

which provides an indicator of how many IP addresses, in the geographical area from which we observe IBR, likely lost connectivity to the Internet following the earthquake. We consider 24-hour periods in order to capture the phenomena over a full 1-day cycle: IBR follows diurnal patterns of human activity, being mostly generated by (infected) users' PCs [17].

To estimate the maximum radius ρ_{max} of impact of the earthquake on Internet connectivity, in Figure 3 we plot a histogram of θ values calculated for network prefixes (address ranges) geolocated at different distances from the epicenter of Tohoku's earthquake, from 0 to $500km$ in bins of $1km$ each. Values of θ around 1 indicate no substantial change in the number of unique IP addresses observed in IBR before and after the event. Figure 3 shows that there is a significant reduction in the number of IP addresses observed before and after the earthquake, i.e., θ is significantly above 1, for address ranges up to $304km$ from the epicenter, where $\theta = 9.3$. We consider the distance from the epicenter where this ratio drops below a threshold, as highlighted in Figure 3, to be rough estimate of the maximum radius of impact of the earthquake on network connectivity, ρ_{max}.

Each bin plotted may represent a different number of network prefixes and IP addresses, and the θ ratio is less meaningful if too few IP addresses are in the numerator. In order to avoid overinterpreting such bins, we only count (plot data for) bins from which the telescope observed at least 1 IP per hour in the 24-hour period preceding the earthquake. Figure 3 shows that some networks look less affected by the earthquake, which could be true or could reflect errors in the geolocation mappings we used.

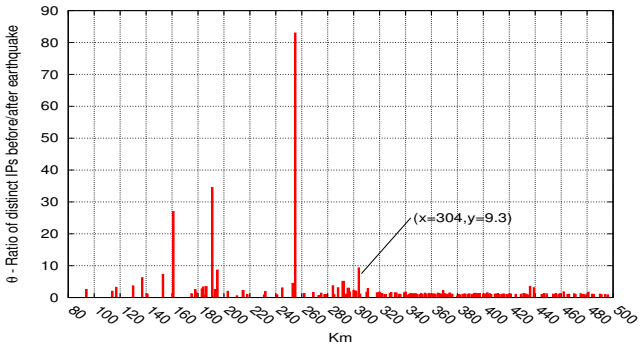


Figure 3: Impact of Tohoku's earthquake on network connectivity: distribution of θ for networks at varying distance, in bins of 1km each, across a range of 0 to 500km. Values of θ around 1 indicate no substantial change in the amount of distinct IPs observed in IBR. Plotting the data this way allow us to roughly estimate the maximum radius ρ_{max} of impact of the earthquake on network connectivity (annotated in figure).

Figure 4 shows the same diagram for the Christchurch earthquake, where a significant value of θ is observable up to $\rho_{max} = 20km$ from the epicenter (θ = 2.4). Figures 5(a) and 5(b) map the

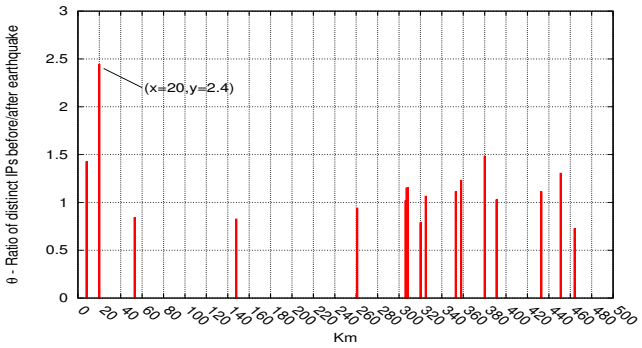
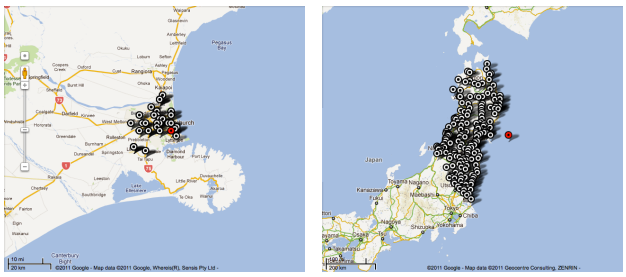


Figure 4: Impact of Christchurch's earthquake on network connectivity: histogram of θ for networks at varying distance, in bins of 1km each, across a range of 0 to 500km. This metric suggests a maximum radius of impact ρ_{max} of $20km$, annotated in the graph.

proximity of the networks within the maximum distance ρ_{max} from the epicenters for both earthquakes.

While plotting the θ distributions can help identify the largest region in which network connectivity was directly affected by the earthquake, we would also like to quantify the impact. Figure 6 plots θ for *all* the networks within a given range (from zero to the value on the x axis), to better reflect the overall impact of the earthquake on the region. We call θ_{max} the largest value of θ observed – together with the distance at which it is observed, θ_{max} represents the *magnitude* of the impact of the earthquake on nearby infrastructure, as observed by the network telescope.

Figure 6 shows that the highest impact of Christchurch's earthquake occurred within 6 kilometers of the epicenter with $\theta_{max} =$



(a) Christchurch (b) Tohoku

Figure 5: Networks selected within the estimated maximum radius of impact of the earthquake (20km for Christchurch and 304km for Tohoku). We based our geolocation on the publicly available MaxMind GeoLite Country database.

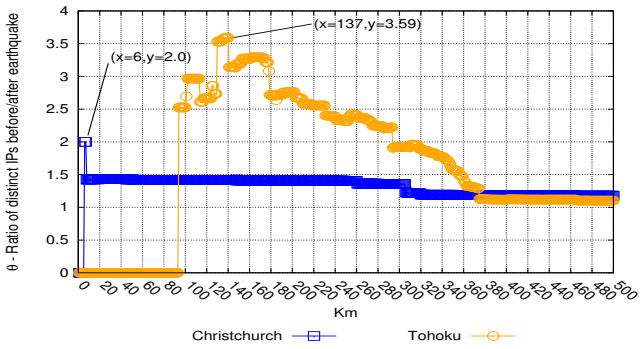


Figure 6: Measuring the impact of the earthquake on network connectivity as seen by the telescope: value of θ for all networks within a given range from the epicenter. The peak value θ_{max} reached by θ can be considered the *magnitude* of the impact.

2.00. The second highest value of θ, 1.434, is reached at 20km. The Tohoku earthquake induced a θ_{max} of 3.59 at a distance of 137 kilometers from its epicenter, consistent with the stronger magnitude of Tohoku's earthquake (see Table 1) and news reports regarding its impact on buildings and power infrastructure. Table 3 summarizes these indicators found for both earthquakes.

	Christchurch	Tohoku
Magnitude (θ_{max})	2 at $6km$	3.59 at $137km$
Radius (ρ_{max})	$20km$	$304km$

Table 3: Indicators of earthquakes' impact on network connectivity as observed by the UCSD network telescope.

IBR traffic also reveals insight into the evolution of the earthquake's impact on network connectivity. Figure 7 plots the number of distinct source IPs per hour of packets reaching the telescope from networks within the $\rho_{max} = 20\ km$ radius from the epicenter of Christchurch's earthquake. All times are in UTC. The time range starts approximately one week before the earthquake and ends two weeks after. We would not expect the IBR traffic to drop to zero, for two reasons. First, not all networks are necessarily disabled by the earthquake. Second, the geolocation database services we use are not 100% accurate.

For a few days before the event, peaks are always above 140 unique IP addresses per hour (IPs/hour) on weekdays, sometimes above 160 IPs/hour. In the 24 hours after the earthquake, the rate drops, with a peak slightly above 100 IPs/hour. The IPs/hour rate climbs slowly, reaching pre-event levels only after a week, which correlates with the restoration of power in the Christchurch area [30].

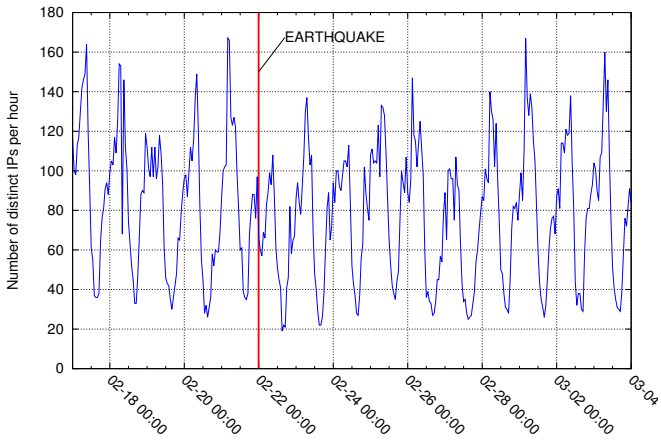


Figure 7: Rate of unique source IP addresses found in unsolicited traffic reaching the UCSD network telescope from networks geolocated within a $\rho_{max} = 20km$ range from the Christchurch earthquake epicenter. The rate of distinct IPs per hour drops immediately after the earthquake. Peaks before the earthquake were above 140-160 IPs/hour on weekdays (weekend is on 19-20 February), while the first peak after the earthquake is slightly above 100 IPs/hour. Levels remain lower for several days, consistent with the slow restoration of power in the area.

Figure 8 plots the same graph for IBR traffic associated with the Tohoku earthquake, within a maximum distance $\rho_{max} = 304\ km$ from the epicenter. The much steeper drop in the number of unique IPs per hour sending IBR traffic is consistent with the Tohoku earthquake's much larger magnitude than that of the Christchurch earthquake. In the days after the event the IBR traffic starts to pick up again, but does not reach the levels from before the event during the analyzed time interval, also consistent with the dramatic and lasting impact of the Tohoku earthquake on Northern Japan.

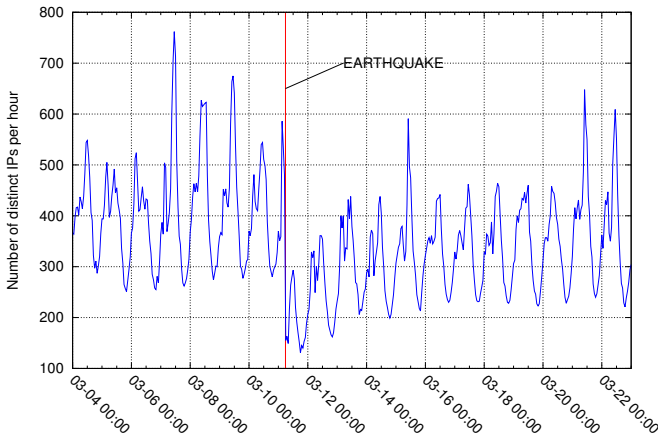


Figure 8: Rate of unique source IP addresses found in unsolicited traffic reaching the UCSD network telescope from networks geolocated within $\rho_{max} = 304km$ of the Tohoku earthquake epicenter. The rate of distinct IPs per hour shows a considerable drop after the earthquake which does not return to previous levels even after several days.

Figures 7 and 8 show that the rate of unique IP addresses per

hour observed by the telescope matches the dynamics of the earthquakes, reflecting their impact on network connectivity. In order to further confirm that the variations in rate of unique IP addresses are anomalous compared to IBR behavior typically observed by the telescope, we plot θ over a longer time frame (two months) surrounding the earthquake using two sliding 24-hour windows before and after each point plotted. Figure 9 plots a two-month period of θ values for networks within a $\rho_{max} = 20\ km$ range of the Christchurch earthquake's epicenter. Normally, values of θ stay within an envelope [0.7 , 1.3], but the value of θ breaks out above the 1.3 upper bound exactly when the earthquake hits. Another lower spike shortly after the earthquake may have been due to blackouts caused by attempts to restore electricity. The corresponding drop is also visible, although less obvious, in Figure 7. The coincidence of the spike in θ with the earthquake suggests the utility of θ as a meaningful indicator of disruption to network infrastructure.

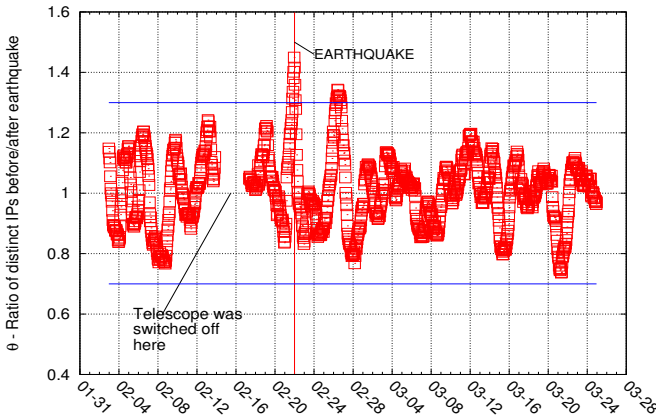


Figure 9: Ratio of number of unique IP addresses reaching the UCSD darknet in two successive 24-hour periods (before vs after the given data point) from networks within a $\rho_{max} = 20\ km$ range from the Christchurch earthquake's epicenter. We plot this θ value over this two-month period, using two sliding 24-hour windows before and after each point plotted. The anomalous breakout above the 1.3 upper bound of θ coincides with the earthquake strike, and represents a large (about 30%) drop in the number of IP addresses reaching the UCSD darknet. The corresponding drop is also visible, although less obvious, in Figure 7.

Figure 10 shows the same diagram for Tohoku using its $\rho_{max} = 304\ km$: there is a similar jump far above a θ ratio of 1.3 exactly when the earthquake strikes. Although one could select an upper bound for θ that is lower than 1.3 for the Tohoku earthquake, since its normal operating range of θ is narrower than for the Christchurch earthquake, the most important point is that the data represent further evidence that θ serves as a useful metric for assessing disruption to network infrastructure.

5. DISCUSSION AND CONCLUSIONS

Ironically, the insidious pervasive reach of malware and misconfiguration on the Internet enables detection and analysis of macroscopic changes in Internet behavior through the observation of Internet background radiation (IBR), or unsolicited one-way *data plane* traffic on the Internet. This traffic has grown to such significant continuous levels that instrumentation capturing a large enough aggregate of such traffic can serve as a lens to illuminate different types of macroscopic events and their impact on communications capabilities. For technical as well as political and economic reasons, such traffic is most easily observable using a large number of routed

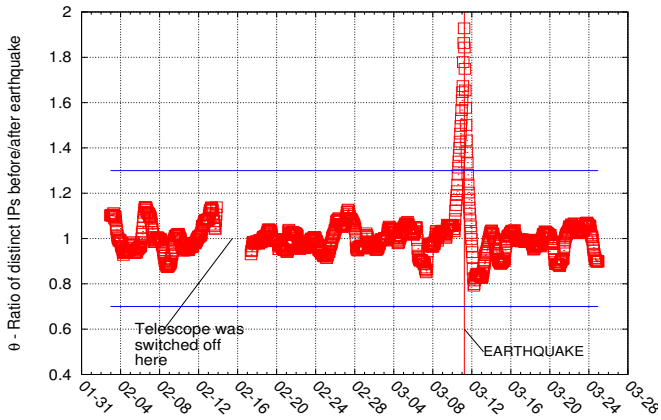


Figure 10: Ratio of number of IP addresses reaching the UCSD darknet in two successive 24-hour periods (before vs after the given data point) from networks within a $\rho_{max} = 304\ km$ range from the Tohoku earthquake's epicenter. Although we use a different distance threshold than for the θ values in the Christchurch plot in Figure 9, there is a similar breakout above a θ ratio of 1.3 exactly when the earthquake strikes.

but unassigned IPv4 addresses, i.e., a darknet, where there is no need to filter out legitimate bidirectional traffic. By geolocating the source IP addresses of traffic destined to the darknet addresses, we can identify when sizeable geographic regions appear to have lost connectivity. Country-level disruptions appear particularly prominently in the data analysis since geolocating IP addresses to countries is more accurate than finer-grained geolocation, e.g, to cities. The ubiquitous presence of this pollution in the data plane also allows us to infer events, such as packet-filtering-based censorship, not observable in other types of data, e.g., BGP. We used four case studies from 2011 to test our approach: two episodes of broad-scale country-level politically motivated censorship, and two high magnitude earthquakes.

Our preliminary approach has several limitations. First, the reliability of IBR as a data source is influenced by unpredictable events and decisions. Law enforcement or other forces might disable a specific botnet, along with whatever fraction of IBR it was generating. Levels of backscatter and IBR traffic would also be reduced by ISPs deciding to filter packets with spoofed source IP addresses, or traffic considered malicious to its customers or network. Vendor software patches and other software or hardware upgrades can close vulnerabilites to infections, reducing the number of IP addresses emitting IBR or otherwise changing its characeristics. Since major natural disasters might damage PCs to the point of requiring replacements, presumably with new operating systems, a longer term reduction in IBR after an event in a specific area might correlate with the number of PCs disabled by the event. In fact, there is no guarantee that a network sends IBR at all, although the correlation of IBR with sources of human error and mischief, the difficulty of eradicating it, and its persistence over decades are all factors that lend credibility to approaches that take advantage of its omnipresence to infer network conditions.

Second, geolocation of the IP addresses sending traffic to the darknet is critical to the inferences, and the accuracy of geolocation databases is necessarily limited, especially at granularity finer than country boundaries. MaxMind states that their GeoLite City database [31] has an accuracy of 62% on a city level (IP addresses not covered are excluded) for New Zealand and 69% for Japan. CAIDA recently performed a comparison of available geolocation tools and services [19], confirming the existence of considerable disparities

across services, that MaxMind's accuracy is comparable or better than most other commercial services, and the fact that regardless of database, country-level inferences are much more accurate than finer-grained location inferences. We also do not address IP address mobility, another concern that will become more important as more mobile devices harbor malware and generate significant background radiation to the global Internet.

Country-level outages such as those induced by government censorship are stunningly visible using darknet instrumentation, and in conjunction with other sources of data can reveal information about how censorship is being implemented over time [17]. Darknet instrumentation can easily detect signal along well-defined geographic boundaries such as countries.

For natural disasters such as earthquakes, the tools of geologists will likely always be more effective for detecting and measuring geophysical impacts of earthquakes, and we have not used our darknet to implement any early warning system functionality. But given some knowledge of the location of a natural disaster, a darknet can support analysis immediately following, and even during, the event regarding its impact on nearby network infrastructure. To assess the impact of the earthquake around its epicenter, we developed a metric based on the number of unique IP addresses that can reach the darknet, a basic reference point for building quantitative indicators. Our metric compares the ratio of the number of IP addresses located within a given distance from the epicenter of an earthquake reaching the darknet in successive (adjacent) 24-hour periods. Plotted over the time period surrounding the natural disaster event, this ratio reveals clearly when a large fraction of normally background-radiating IP addresses falls silent, i.e., loses connectivity to the global Internet. This metric allows a quantitative assessment of impacts that could be used to compare different events, and in conjunction with geolocation information could also be used to assess the geographical extent of damage to communication infrastructure, identify which networks were affected, and reveal insight into the temporal dynamics of the event, e.g. the persistence of network infrastructure disruption subsequent to the event. As an example, our metric clearly revealed that the Tohoku earthquake had considerably higher impact on nearby (and much further away) network infrastructure than the Christchurch earthquake. We were also able to compare the geographic extent (radius) of the damage, and the restoration time based on when IBR traffic returned as observed by our telescope.

The approach we describe here holds many opportunities for improvement. We would like to visualize at finer granularity the disconnection and restoration of networks over time. For this purpose, one could group the networks into clusters based on lat/long coordinates rather than the simple radius distance from the epicenter which does not differentiate direction.

Our current approach also does not distinguish between end-hosts losing connectivity due to network reachability versus loss of power. Additional sources of data, such as active probing, would further improve our ability to characterize the nature, timing, and effects of outages, especially since only a subset of data sources may be available at any given time. In [17] we used IBR traffic data in conjunction with BGP and traceroute data to extract fine-grained insights into how government censorship policies were being tested and implemented. In this paper we demonstrated that IBR traffic can also be used to observe the impacts of other regionally disruptive events. Although we have only scratched the surface of what is possible with this data, our investigation thus far has convinced us that IBR traffic is an important building block for a more comprehensive methodology for monitoring, analysis, and possibly even detection of events unrelated to the IBR itself.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1059439.

6. REFERENCES

[1] AfriNIC. http://www.afrinic.net.

[2] E. Aben. Unsolicited Internet Traffic from Libya. http://labs.ripe.net/Members/emileaben/unsolicited-internet-traffic-from-libya, March 23 2011.

[3] M. Bailey, E. Cooke, F. Jahanian, N. Provos, K. Rosaen, and D. Watson. Data reduction for the scalable automated analysis of distributed darknet traffic. In *ACM SIGCOMM conference on Internet Measurement*, 2005.

[4] BGPmon. Internet in Egypt offline. http://bgpmon.net/blog/?p=450, January 28 2011.

[5] Z. Bischof and J. Otto. Egypt and Libya Internet disconnections. http://www.aqualab.cs.northwestern.edu/blog/egypt-libya-peers.html, March 10 2011.

[6] Z. Bischof and J. Otto. Japanese earthquake and P2P disruption. http://www.aqualab.cs.northwestern.edu/blog/sendai-earthquake-japan-peers.html, March 17 2011.

[7] CAIDA. UCSD Network Telescope. http://www.caida.org/data/passive/network_telescope.xml, 2010.

[8] CAIDA. Archipelago measurement infrastructure. http://www.caida.org/projects/ark/, July 28 2011.

[9] M. Casado, T. Garfinkel, W. Cui, V. Paxson, and S. Savage. Opportunistic Measurement: Spurious Network Events as a Light in the Darkness. In *ACM Fourth Workshop on Hot Topics in Networks (HotNets-IV)*, New York, NY, USA, 2005. ACM.

[10] L. Cottrell. Internet End-to-End Performance Monitoring - Japanese Earthquake, April 2011. https://confluence.slac.stanford.edu/display/IEPM/Japanese+Earthquake+March+11th%2C+2011.

[11] J. Cowie. Egypt leaves the internet. http://www.renesys.com/blog/2011/01/egypt-leaves-the-internet.shtml, January 27 2011.

[12] J. Cowie. Egypt Returns To The Internet. http://www.renesys.com/blog/2011/02/egypt-returns-to-the-internet.shtml, Feb 2 2011.

[13] J. Cowie. Japan quake. http://www.renesys.com/blog/2011/03/japan-quake.shtml, March 11 2011.

[14] J. Cowie. Libyan disconnect. http://www.renesys.com/blog/2011/02/libyan-disconnect-1.shtml, February 18 2011.

[15] J. Cowie. What Libya Learned from Egypt. http://www.renesys.com/blog/2011/03/what-libya-learned-from-egypt.shtml, March 2011.

[16] A. Dainotti, A. Pescapé, and G. Ventre. Worm traffic analysis and characterization. In *IEEE ICC 2007*, June 2007.

[17] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé. Analysis of Country-wide Internet Outages Caused by Censorship. In *ACM SIGCOMM conference on Internet measurement*, 2011.

[18] Google. Transparency report, 2011. http://www.google.com/transparencyreport/.

[19] B. Huffaker, M. Fomenkov, and K. Claffy. Geocompare: a comparison of public and commercial geolocation databases. Technical report, May 2011. http://www.caida.org/publications/papers/2011/geocompare-tr/.

[20] E. Katz-Bassett, H. V. Madhyastha, J. P. John, A. Krishnamurthy, D. Wetherall, and T. Anderson. Studying Black Holes in the Internet with Hubble. In *USENIX Networked Systems Design & Implementation (NSDI)*, 2008.

[21] P. Kijewski. Automated Extraction of Threat Signatures from Network Flows. *18th Annual FIRST conference*, June 2006.

[22] C. Labovitz. Libya Firewall Begins to Crumble? http://monkey.org/~labovit/blog//viewpage.php?page=libya_firewall_cracks.

[23] C. Labovitz. Attack Severs Burma Internet. http://asert.arbornetworks.com/2010/11/attac-severs-myanmar-internet/, November 3 2010.

[24] C. Labovitz. Egypt Loses the Internet. http://asert.arbornetworks.com/2011/01/egypt-loses-the-internet/, February 2011.

[25] C. Labovitz. Egypt Returns to the Internet. http://asert.arbornetworks.com/2011/02/egypt-returns-to-the-internet/, January 2011.

[26] C. Labovitz. Middle East Internet Scorecard, February 20 2011. http://asert.arbornetworks.com/2011/02/.

[27] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. *SIGCOMM Comput. Commun. Rev.*, 40, August 2010.

[28] J. Li and S. Brooks. I-seismograph: Observing and measuring Internet earthquakes. In *Proc. IEEE INFOCOM*, April 2011.

[29] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An Information Plane for Distributed Services. *OSDI*, 2006.

[30] N. Mathewson. Power restored to most households. http://www.stuff.co.nz/national/christchurch-earthquake/4734825/, March 5 2011.

[31] MaxMind. GeoLite City, April 2011. http://www.maxmind.com/app/geolitecity.

[32] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the Slammer Worm. *IEEE Security and Privacy*, 1(4):33–39, July 2003.

[33] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage. Inferring Internet denial-of-service activity. *ACM Trans. Comput. Syst.*, 24:115–139, May 2006.

[34] J. Nazario. Georgia DDoS Attacks, August 12 2008. http://asert.arbornetworks.com/2008/08/georgia-ddos-attacks-a-quick-summary-of-observations/.

[35] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of internet background radiation. *ACM SIGCOMM conference on Internet measurement*, 2004.

[36] P. Porras, H. Saidi, and V. Yegneswaran. An Analysis of Conficker's Logic and Rendezvous Points, March 2009. http://mtc.sri.com/Conficker/.

[37] M. Richtel. Egypt Cuts Off Most Internet and Cell Service. http://www.nytimes.com/2011/01/29/technology/internet/29cutoff.html, January 28 2011.

[38] RIPE NCC. RIPE Atlas. http://atlas.ripe.net/.

[39] RIPE NCC. Routing Information Service (RIS). http://www.ripe.net/data-tools/stats/ris/.

[40] Y. Shavitt and E. Shir. DIMES: Let the Internet Measure Itself. *ACM SIGCOMM Computer Communication Review*, October 2005.

[41] R. W. Sinnot. Virtues of the Haversine. *Sky and Telescope*, 68:159, 1984.

[42] SLAC National Accelerator Laboratory. The PingER Project. http://www-iepm.slac.stanford.edu/pinger/site.html.

[43] University of Oregon. Routeviews project. http://www.routeviews.org.

[44] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet background radiation revisited. In *ACM SIGCOMM conference on Internet measurement*, 2010.

[45] E. Zmijewski. Georgia clings to the 'net, August 10 2008. http://www.renesys.com/blog/2008/08/georgia-clings-to-the-net.shtml.