Characterization of Internet Routing Anomalies Through Graph Mining

Pablo Moriano*
School of Informatics and Computing
Indiana University
Bloomington, IN 47408
pmoriano@indiana.edu

Srivatsan Iyer
School of Informatics and Computing
Indiana University
Bloomington, IN 47408
srriyer@indiana.edu

L. Jean Camp
School of Informatics and Computing
Indiana University
Bloomington, IN 47408
ljcamp@indiana.edu

ABSTRACT

Our goal is to contribute to the understanding and detection of control plane anomalies as perturbations in a graph representation of connected autonomous systems (ASes). We reconstructed the autonomous system (AS) level graph for three large-scale routing incidents and evaluated the topological properties of the graphs before, during, and after these events. The three incidents we examined were the Indosat hijack in April 2014, the Telecom Malaysia leak in June 2015, and the Bharti Airtel Ltd. hijack in November 2015. Using observations from the AS graph topology, we illustrate that the incidents were visible as anomalies before they are widely diffused. Topological features in the graph as a whole did not show significant immediate changes over the course of these events. However, significant changes are evident in the average path length and clustering coefficient of the observed graphs when they are decomposed using k-shell decomposition analysis. The k-shell decomposition distinguishes between the core and periphery (also called crust) graphs. In this k-shell decomposition, the core consists of ASes with of at least connectivity k, with the crust consisting of those ASes which have less than k connectivity. While anomalous behavior was not observable in the core graph, the events are immediately apparent on the crust. Specifically, when the AS-level graph is examined using k-shell decomposition, there are topological changes in the crust in path length and clustering measurements. Our explanation is that, in graph theoretical terms, these incidents require the initiators to move closer to the core, away from the periphery, and the concentric impacts of the disturbances are visible as these move across the crust. This technique has potential for early detection of large-scale control-plane anomalies, which could enable quicker mitigation.

KEYWORDS

Internet measurements, BGP, prefix hijacking, route leak, graph mining, k-shell decomposition

1 INTRODUCTION

The Internet's "interconnectedness,"—its most remarkable feature—is among its major vulnerabilities. The Internet is composed of decentralized but coordinating entities known as autonomous systems. Within an autonomous system, routers, IP prefixes (blocks of IP addresses), and routing policies are under common administrative control [17]. Each AS has assigned sets of distinct IP prefixes that can be reachable from other ASes. Communication exchange between different ASes on the Internet is coordinated through the Border Gateway Protocol (BGP) [34]. BGP is used to exchange IP network routes between ASes.

BGP exchanged messages are assumed to be valid by default. This means that the reachability information shared between ASes is assumed to be trusted, and therefore, it is not verified. Although the latest version of the BGP protocol was released in 2006 [35], the design of the protocol itself does not feature an inherited protection mechanism against participants advertising false routes. Specifically, BGP lacks authentication mechanisms for the announcements or origins of IP ranges. BGP also lacks authentication of path announcements. This leaves BGP vulnerable to unintended misconfiguration and malicious attacks [13]. The impact of these misconfigurations or attacks are usually reflected in (i) traffic blocking and (ii) interception. In traffic blocking, the network traffic is directed to the bogus AS, never reaching its legitimate destination [4]. In traffic interception, the bogus AS reroutes traffic for the victim IP prefix and then redirects it to the original origin AS [8]. Prior to that, the traffic might be subject to eavesdropping [3], traffic analysis [39], or tampering [37].

Traditionally, most of the approaches to address the problem of identifying routing anomalies rely on (i) cryptographic implementations to authenticate origin or (ii) anomaly detection techniques. Cryptographic implementations to authenticate origin include RPKI [24] and BGPSec [25], which further offers the ability to cryptographically authenticate the entire path. These techniques are powerful, but there has not been adequate incentive to drive widespread adoption [12]. Cryptographic solutions are expensive. They require changes in the current routing infrastructure, i.e., to widespread changes in the authentication mechanism of the protocol across the entire Internet. Perhaps more importantly, it has been shown than even with their widespread adoption, it will be not possible to avoid certain more subtle types of attacks, such as was used to induce the traffic between network operators in Denver to travel to Iceland before returning to Denver [8].

Anomaly detection techniques rely on conducting measures on the control-plane level (using BGP feeds) or data-plane level (by exploring reachability of IP addresses in suspicious announced routes), or a combination of both. Anomaly detection schemes do not require changes in the protocol itself. They primarily are used in detecting anomalies based on passive or active measurements, i.e., to alert operators to mitigate threats [20, 38, 46]. However, in many of the anomaly detection schemes, the prefix measurements are precomputed and not dynamic. This implies that their underlying mechanisms need to be recomputed if there is a change in the observed routing infrastructure.

*Corresponding author.

It is still an open issue to understand and characterize the occurrence of anomalous events on the Internet using the information extracted from the reachability graphs built from BGP announcements. Here we seek to contribute to the understanding and characterization of routing incidents through the use of methods from other fields. Specifically, we use graph mining from the study of human social networks to detect these anomalies. Not only is a graph-theoretical approach suitable for the connection of ASes, this approach can incorporate the dynamic behavior of the observed routing infrastructure. To summarize, we make the following key contributions:

- *Temporal graph analysis framework:* We propose a generic temporal graph analysis framework to model the evolution on the Internet at the AS-level. The proposed framework is based on the idea that the evolution of the Internet can be abstracted as a dynamic system of consecutive graphs—also called graph stream (Section 2.1.3). We use the proposed framework to formalize a set of measurements of the observed graphs at each time instant.
- *Graph mining analytics:* We use graph mining to reveal that some properties of the AS-level graphs are useful for early detection of routing incidents. In particular, we show that centrality, average path length, and clustering measurements are more susceptible to be perturbed when they are analyzed using k-shell decomposition—to decompose graphs between the core and the periphery (Section 2.1.4). Our results suggest that topological signatures from the AS-level graph representation can be used to infer when an anomalous routing event is happening before widespread disruption.
- *Evaluation and case studies:* We study the capabilities of the proposed approach by building AS-level graphs of three different large-scale routing incidents, i.e., Indosat in April 2014, Telecom Malaysia in June 2015, and Bharti Airtel Ltd. in November 2015 (Section 2.1.1). Our work is differentiated from the work in [11, 21] in that we use dynamic update information from the RouteViews project (with granularity every 15 minutes) to reconstruct the network topology at the AS-level and study the robustness of network topological properties, before, during, and after the incident (Section 3). This approach allows for differentiation between normal behavior at the network level and disruption or anomalous changes during the incidents. The three cases we address were easily identified following the large-scale disruption but not before.

The rest of the paper is structured as follows. Section 2 provides a description of the data sources (along with their preprocessing) and the topological properties that were used to evaluate the impact of the routing anomalous incidents that were studied in the AS-level graph topology. Section 3 shows the results of the dynamic network analysis. We placed special emphasis on the characterization of the time periods before, during, and after the specific event, in terms of the impact on the referred structural properties. Section 4 is our discussion of the implications of the results, including addressing the possible implications for implementation that take into account global measures of topological disruption. Section 5 discusses related work in both cryptographic and anomaly detection schemes to prevent, identify, or mitigate routing anomalies. Finally, Section 6 presents concluding remarks and areas for future research.

2 METHODS

Every AS originates the prefixes that have been allocated to it. All ASes can announce prefixes as well as paths. ASes build a graph of interactions with others ASes based on information about reachable paths to IP prefixes. The reachability of these paths is determined through BGP announcements that ASes receive from their neighbors. Gateway routers in the ASes use route updates to modify their routing tables, and these determine how to direct traffic. It has been shown that routing decisions depend mainly on path length (i.e., the number of hops to reach the destination prefix); secondarily on the cost of directing traffic through a specific neighbor based on previously established business contracts; and then on diverse tertiary criteria [13]. The resulting graph is a dynamic system based on protocol incentives and economic constraints with continuous addition and deletion of nodes and edges [31].

Using the graph of ASes, we performed a longitudinal empirical analysis of the network topological measures that correlate with the occurrence of three major BGP disruptions. For each event, we reconstructed the evolving network topology around the date and time of these incidents. This procedure is done to identify statistically significant changes in the graph topology. In Section 2.1, we provided details on the data source and the incidents that we examined for this research.

2.1 Data sources

In this section, we detail the data sources (along with their preprocessing) that were used to perform the analysis which follows. We start by describing the data used for the construction of (*i*) the database of the large-scale routing anomalous events (to establish the ground truth); (*ii*) the AS-level data; (*iii*) the description of the AS-level graph representation; and then end with (*iv*) the definition of the graph topological properties measured in the Internet graphs.

2.1.1 Routing anomalous events. We considered the AS-level graphs for very well-known cases of routing anomalous events that led to large-scale disruptions during the last few years. In particular, we performed an analysis of (*i*) an Indonesian ISP hijack that covered much of the world; (*ii*) an Malaysian ISP that generated global collateral damage by leaking prefixes to large-scale providers; and (*iii*) an Indian ISP that hijacked prefixes of other important Internet players. Note that these anomalous events have been studied and corroborated from different sources. We describe more details about these incidents below. Events are listed in chronological order.

An Indonesian ISP hijacks the world. On April 2, 2014, starting at 18:26 UTC, Indosat (one of the largest telecommunications providers in Indonesia) announced more than 320 000 IP prefixes belonging to other networks. In fact, Indosat announced roughly two-thirds of the entire Internet address space [47]. A large fraction of the hijacked prefixes belonged to Akamai, which is one of the larger Content Delivery Networks. This event lasted for several hours until approximately 21:15 UTC. Traffic continued to be delivered; however, the path of the traffic was significantly perturbed.

Global collateral damage of the Telecom Malaysia leak. On June 12, 2015, starting at 08:43 UTC, Telecom Malaysia announced about 179 000 IP prefixes to Level 3 (the largest crossing AS) [42]. Level 3 accepted these announcements and then propagated the routes to their peers and customers around the word. Because Telecom Malaysia is a customer of Level 3, the routes announced by Telecom Malaysia were identified as a preferred delivery route for Level 3. At around 10:40 UTC, there were slowly observed improvements, and by 11:15 UTC the errors in the RIB began to be resolved. Note this was a leak, so the data were not delivered after being transmitted to Telecom Malaysia.

Large scale BGP hijack in India. On November 6, 2015, starting at 05:52 UTC, Bharti Airtel Ltd., claimed the ownership of about 16 123 IP prefixes. These prefixes corresponded to more than 2000 unique ASes [41]. This event became widespread because two large ASes (e.g., Cogent Communications and GlobeNet Cabos Submarinos S.A.) accepted and propagated these routes to their peers and customers. Legitimate owners of the prefixes included Akamai, Tata Communications, and Apple Inc. This event lasted until approximately 14:40 UTC.

We summarize the details of the incidents used in this study in Table 1.

Table 1: Summary of large-scale routing incidents.

Incident	Date	Start time	End time	Duration
Indosat	2014-04-02	18:26 UTC	21:15 UTC	≈ 2.9 h
Telecom Malaysia	2015-06-12	08:43 UTC	11:15 UTC	≈ 2.7 h
Bharti Airtel Ltd.	2015-11-06	05:52 UTC	14:40 UTC	≈ 8.9 h

2.1.2 BGP data. We collected BGP measurement data using BGPStream[1]. BGPStream provides an open-source software framework for the analysis of historical and real-time BGP data [30]. To do so, BGPStream extracts data directly from route collectors. A route collector is a host running a collector process. The collector emulates a router that establishes BGP peering seasons with real BGP routers. These collection points are known as vantage points (VPs, hereafter).

A BGP router maintains its reachability information in the Routing Information Base (RIB). Together, the VPs ideally provide a list of paths between Autonomous System Numbers (ASNs) of every reachable network. The collection points aggregate data including update messages that reflect routes being added or deleted from the RIB. Update messages contain fine granular information about routing dynamics [26], i.e., changes in the paths. By sampling changes in the routing table of the VPs, collectors can reconstruct RIBs from their peering routers. This constitutes a partial view of the Internet at the AS-level, i.e., an undirected graph in which vertices are ASes and the edges are routing links between them.

There are two popular projects running route collectors processes, RouteViews [27] and RIPE RIS [28]. They make dumps available in public archives. At the time of this writing, they operate 19 and 17 collectors, respectively, which peer with hundreds of VPs respectively [15]. RouteViews and RIPE RIS collect a RIB dump every two hours and eight hours, and update dumps every 15 and five minutes, respectively.

BGPStream allows the setup of different parameters in the data collection process. In particular, it is possible to manipulate the start and end date of data collection, and the specific project running route collector processes, among other variables. We collected approximately two days of observations around the start date of each of the incidents in Table 1—to gather observations before, during, and after the selected events. The purpose of collecting data over this time period is to be able to distinguish between regular and anomalous behavior. In this analysis, we only collected BGP measurements from the RouteViews project. Previous research has shown that there is a considerable overlap between the measurements from RouteViews and RIPE RIS projects [6].

2.1.3 AS-level graph representation. We aggregated every possible path in the RIBs among collectors (at a certain time) to build snapshots of the Internet topology. The resulting graph topology was constructed from observed paths derived from BGP updates. For repeated paths among the different collectors, we only considered one instance of the path. Specifically we modeled the AS-level topologies as graphs with the following considerations.

Consider the sequence of n intervals $A = \{A_1, A_2, \ldots, A_n\} = \{A_t\}_{t=1}^{n}$, where

1. $A_t = [a_t, a'_t)$ for all $t < n$ and $A_n = [a_n, a'_n]$ for $t = n$; and
2. $a_t < a'_t = a_{t+1}$ for all t;

An interval represents a fixed-length unit of time, i.e., the granularity at which BGP updates dumps are collected. In this case, it is 15 minutes, based on the data available from the RouteViews project. Note that dump times are synchronized among the collectors for each experiment discussed in this paper. Condition (1) implies that all intervals are left-closed and right-open (except the last one which includes a'_n). That is, if events at time t are part of one sequence, then only events later than t are part of the next sequence. It guarantees that the sequence of intervals is disjoint. Condition (2) implies that intervals are non-empty. Note that a'_t and a_{t+1} represent the time instants of a transition between intervals. For any interval A_t, the right endpoint a'_t corresponds to the left endpoint of the interval A_{t+1}. Together with Condition (1), Condition (2) guarantees that the union of all intervals $\bigcup_{t=1}^{n} A_t = [a_1, a'_n]$ is a closed interval.

In addition, we let $\mathcal{H} = \{1, 2, \ldots, N\}$ be the set of nodes (e.g., set of ASes). Then, $\mathcal{V}(t) \subseteq \mathcal{H}$ is the subset of nodes which interact (i.e., which have an identified path during interval $A_t = [a_t, a'_t)$). Let $\mathcal{E}(t) = \{e_{ij}(t) : i, j \in \mathcal{H}\}$ be an adjacency matrix of edges $e_{ij}(t)$ that captures the existence of a routing link between node i and node j during interval A_t. Let the graph $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$ represents an undirected graph that captures all interactions that occur from endpoints a_t to a'_t, $t \in \{1, 2, \ldots, n\}$. The total number of nodes of nodes in $\mathcal{G}(t)$ is $N(t) = |\mathcal{V}(t)|$. The sequence $\{\mathcal{G}(t)\}_{t=1}^{n}$ denotes the graph series G.

Let $P_{ij}(t)$ be a set of paths between nodes i and j at time interval t. Each path $p_{ij}(t) \in P_{ij}(t)$ is a sequence of edges made from nodes from $\mathcal{V}(t)$ that forms an undirected path that starts at i and finishes at j. For example, $\{(i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k)\}$ is a path where $i_1 = i$ and $i_k = j$, and each node in the sequence $i_1, \ldots, i_k$ is

[1] Available at https://bgpstream.caida.org/

distinct. Let $p^*_{ij}(t)$ be the shortest (geodesic) path between nodes i and j at time interval t. Let $d^*_{ij}(t)$ be the length of the shortest path between nodes i and j, i.e., $|p^*_{ij}(t)|$.

Finally, let $q_i(t)$ represent the set of neighbors of node i at time t, i.e., $q_i(t) = \{j : e_{ij}(t) \in \mathcal{E}(t)\}$. Then, the degree of node i is $|q_i(t)|$ and the total number of edges of $\mathcal{G}(t)$ is $E(t) = \sum_{i \in \mathcal{V}(t)} |q_i(t)|$.

Using these conditions and this approach, we constructed graphs as each time period. That allowed us to measure the topological properties of the graph as the anomalies were introduced to the network, diffused, and then removed.

2.1.4 Topological properties. We measured a set of 10 graph topological properties to study their correlation with very well-known cases of large-scale Internet disruptions. The selected topological measures are the most relevant for the understanding of the structure and function of the Internet according to [10, 16]. We grouped these properties in three different categories: centrality, path length, and community structure measures.

Table 2 summarizes each topological category and their relationships with the structure and function of the Internet. We discuss the details about these categories below.

Table 2: Summary of topological properties.

Property	Internet effect
Degree centrality	ASes connectivity (importance)
Average path length	Routing efficiency
Clustering	Peering structure (alternate routes)

Centrality measures. Node centrality reveals the importance of a node in the graph. In the context of the Internet, it has been shown that these measures are relevant to understand ISP regulation [19] and the robustness of the Internet [9]. We defined each of the centrality measures used below.

Number of nodes: This measure corresponds to the number of unique nodes in the graph at a certain time. In this paper, we have reported the number of nodes as $N(t)$.

Number of edges: This measure corresponds to the number of unique edges in the graph at a certain time. In this paper, we have reported the number of edges as $E(t)$.

Maximum degree: The degree of a node is the number of edges attached to it. In this paper, we report the maximum degree across the nodes in the graph as $\max\{|q_i(t)| \ \forall i \in \mathcal{V}(t)\}$.

Path length. The average path length relates to the number of hops between nodes for every possible pair. In the context of the Internet, it is an important property to study because it relates to the efficient routing of packets. Although it is known that not all packets travel through the shortest path given commercial agreements, routing efficiency is ultimately influenced by shortest path measures. Here, we used the average path length, i.e., the mean of the shortest paths between each pair of nodes in the graph. The shortest path between two nodes belonging to different components is said to be infinite. More formally, it is defined as $L(t) = \frac{1}{N(t)(N(t)-1)} \sum_{i,j \in \mathcal{V}(t)} d^*_{ij}(t)$.

Community structure. Community measures relate to the likelihood of finding a group of nodes that are able to form substructures. In the context of the Internet, it has been shown that community structure is key to understand the tiered structure of the Internet [44] and its resilience to both random or targeted removal of nodes [9]. We defined each of the community structure measures used below.

Average clustering coefficient: The local clustering coefficient of a node $i \in \mathcal{V}(t)$ quantifies how close its neighbors are to be connected as well. In particular, the local clustering of node i is defined as $\gamma_i(t) = \frac{2|e_{jk}: j,k \in q_i(t),\ e_{jk}=1|}{|q_i(t)|(|q_i(t)|-1)}$. In this paper, we reported on the average clustering coefficient defined as $\bar{C}(t) = \frac{1}{N(t)} \sum_{i \in \mathcal{V}(t)} \gamma_i(t)$.

Components: A component is a subgraph in which any two pair of nodes are connected to each other by at least a path. In this paper, we reported on the average size of the components.

Within each category, we have also made use of group techniques on the vertices to decompose the graphs into shells—informed by previous empirical observations about the actual structure of the Internet [2]. In particular, we tangentially applied k-shell decomposition to analyze properties from each of the previous described categories.

k-shell decomposition. The k-shell decomposition fragments a graph between core and crust subgraphs respectively. The k-shell of a node (also called coreness) is a measure of the centrality of the node with respect to its neighbors. This decomposition technique is an iterative process. It starts from degree $k = 1$, and in every step, nodes with similar degree are removed until the nucleus of the graph is revealed—the maximal k that keeps $N(t)$ larger than zero. The details about this process at time instant t are described below.

Step 1: Compute the adjacency matrix $\mathcal{E}(t)$ and identify nodes with degree $k = 1$, i.e., $\{i \in \mathcal{V}(t) : |q_i(t)| = 1\}$.

Step 2: Remove all nodes with degree equals to k, i.e., $\{i \in \mathcal{V}(k) : |q_i(k)| = k\}$. This results in a pruned adjacency matrix $\mathcal{E}'(t)$.

Step 3: Compute the degree of each node from the remaining set of nodes. If there are nodes with degree equals to k, step 2 is repeated—producing a new adjacency matrix $\mathcal{E}'(t)$. Otherwise, return to step 1 with an increased value of $k = k + 1$ and $\mathcal{E}(t) = \mathcal{E}'(t)$.

At any given time t, the k-shell is made of all removed nodes (and their respective edges) in a given degree (step) k. The k-shell decomposition reveals hierarchies of ASes. The subgraph that is generated from the accumulation of all removed nodes, i.e., in all previous $k-1$ shells, is called the k-crust [23]. In the context of the Internet, the k-crust reveals the periphery of the AS-level graph. The subgraph that is formed from the remaining graph at any given step k is called the k-core—the maximum subgraph with minimum degree at least k. In general, the nucleus (the k-core graph with the largest possible k) of the Internet has been studied with the aim of understanding the evolution of the Internet [45]. It is worth noting that at the end of every k step, new k-shell, k-crust, and k-core subgraphs are produced. More details on this decomposition method can be found in [33].

Figure 1 shows the representation of an Internet map at the AS-level—as of the end of the second week of May 2017. The visualization algorithm uses weekly updated data from the RouteViews project. A node's annotations correspond to the country at which

ASes are registered. The AS that is tagged in Figure 1 corresponds to Indiana University (ASN 87)—based on the origin IP address at which this plot was generated. In addition, the left scale represents the degree of the nodes (node sizes are scaled with respect to their degree). The right scale represents the k-shell index of the nodes (hotter colors represent ASes that are more towards the core). This map is generated according to the algorithm described in [1].

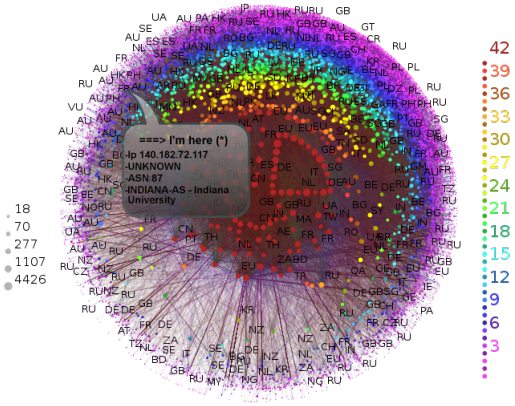

Figure 1: Graphical representation of the AS-level graph.

3 RESULTS

Our results indicate that the anomalous routing events are correlated with changes in the graph topology generated from the BGP measurements. The main idea is to examine the structure of the network before, during, and after the three large-scale anomalies. Topological changes in the network at the AS-level will allow us to test the hypothesis as to whether dynamic transitions in the network structure can be used as a detection signature of anomalous Internet routing events.

We report the results based on the analysis of three large-scale routing incidents using BGP updates from BGPStream. The details about each incident and the data collection are described in Section 2. In this section, we have summarized the properties of each reported time series based on the category to which these belong and the corresponding anomalous routing incident. Remember that we focus on properties that are related to the structure and the function of the Internet. These properties produce a single deterministic number when evaluated and are not distributions.

We used the same visualization conventions for every plot. First, we used open circles when plotting the raw data. Recall the raw data corresponds to the empirical measures for each graph topological property. Second, we used moving average to smooth out short-term fluctuations and highlight long-term trends in the time series. Solid lines represent the moving average when using a centered window of 10 observations around each data point. Third, we computed the moving standard deviation using similar parameters as used for the moving average. We highlighted the region comprehended between ± one standard deviation for each data point in faint red. Finally, we highlighted in yellow the beginning and end times of each of the incidents described in Table 1. We marked minor ticks in the time axis to represent intervals of 15 minutes, i.e., the time between consecutive BGP updates from the RouteViews project.

3.1 Centrality measures

An Indonesian ISP hijacking the world. Here, we report on the results of centrality measures which illustrate the prominence of ASes. Figure 2 shows the maximum degree of each graph snapshot during the observation period. From this plot, it is possible to infer that the only significant changes in this measure are for the graphs captured at April 2, 2014, at 12:00 and April 3, 2014, at 6:00. These discontinuities are consistent with the ones illustrated in Figures 17 and 18 (see Appendix).

Figure 3 shows the number of nodes at various k-levels of crust graphs, i.e., $k = 1, 10$, and the maximum k possible—the one that encloses the biggest crust of the Internet. As we might expect, it is not possible to observe significant changes regarding the total number of nodes in this time series. This suggest that the periphery graph keeps almost the same with respect to the number of ASes.

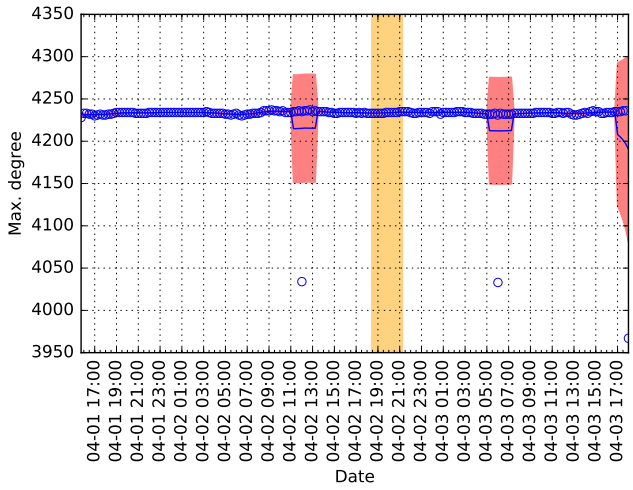

Figure 2: Maximum degree Indonesia event.

Global collateral damage of Telecom Malaysia leak. For this incident, Figure 4 shows the maximum degree. As can be seen, the only significant variations for this property occurs for observations derived at June 11, 2015, at 10:00, and 18:00; June 15, 2014, at 00:00, and 18:00; and June 13, 2015, at 00:00, and 8:00. Similarly, for the crust graphs, we computed the number of nodes as is shown in Figure 5. We did not observe significant variations in the number of nodes that is in the largest crust.

Large scale BGP hijack in India. For the Indian incident, we tracked of the same properties we did for the Indonesian and Malaysian incidents. In particular, Figure 6 shows the maximum degree for each graph snapshot during the observation period. As in the previous cases, there are some discontinuities in the time series. The discontinuities are evident in November 5, 2015, at 10:00, 12:00, 14:00, 16:00, 18:00, 20:00, 22:00; and November 6, 2015, at

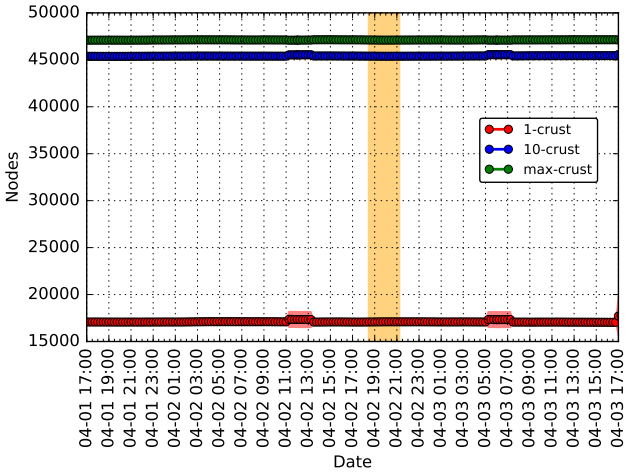

Figure 3: Nodes per crust Indonesia event.

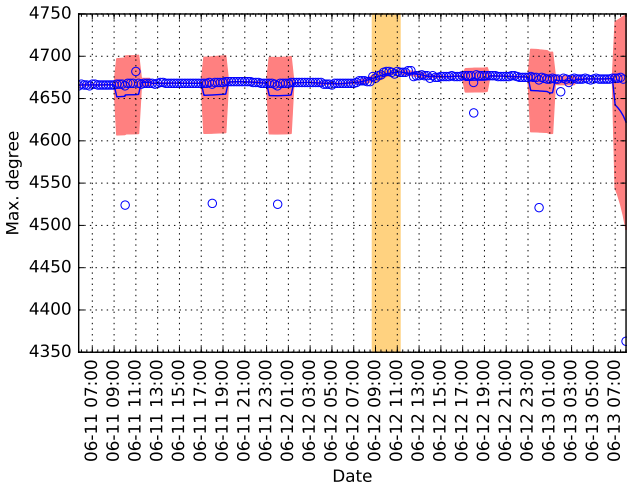

Figure 4: Maximum degree Malaysia event.

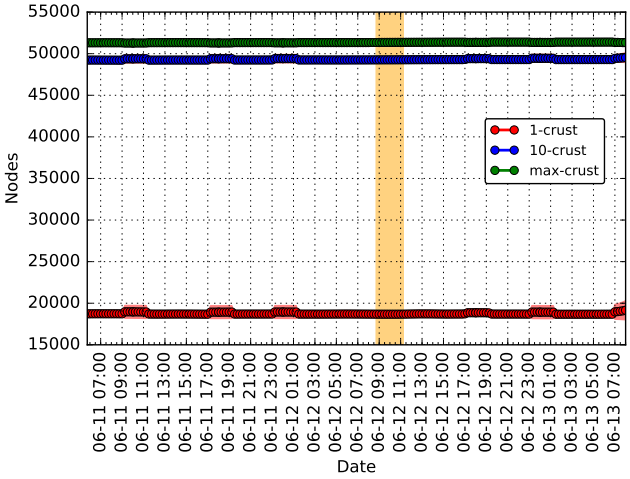

Figure 5: Nodes per crust Malaysia event.

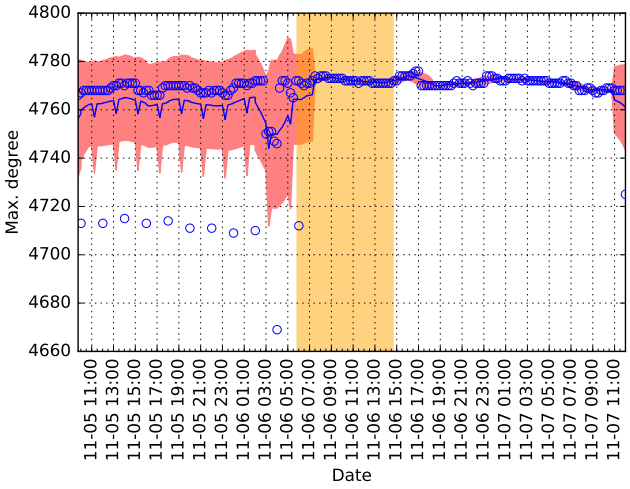

Figure 6: Maximum degree India event.

00:00, 02:00, 04:00, and 06:00. In particular, the discontinuity on November 6, 2015 at 4:00 is more stronger. This suggests that the node with most connections in the graph suddenly decreases its degree—which is remarkable given that it happens almost two hours before the anomaly was reported by other BGP monitoring projects, e.g., BGPmon.net and Dyn Research. Similarly, Figure 7 shows the number of nodes in the crust subgraphs. This measure does not reveal significant changes during the observation period.

3.2 Path length

An Indonesian ISP hijacking the world. Figure 8 shows the average path length for each graph snapshot during the Indonesia incident, i.e., for different values of k in the crust. In particular, we observed that the average path length seems to be altered based on the value of k being analyzed. Although for a value of $k = 10$ this measure tends to decrease in agreement with the discontinuities noticed in Figure 17, for a bigger crust graph—the whole graph without the nucleus of the network—this measure tends to increase. This is an important observation given that these properties are directly related with the routing efficiency and the number of alternative routes to reach different networks. This might be expected in the case of a disruption of service observed in these types of events.

Global collateral damage of Telecom Malaysia leak. Figure 9 shows the average path length measure over different crust subgraphs for the Malaysia incident. As for the Indonesian incident, relative changes in this measure over the crust depends on the k value being analyzed, i.e., the average path length decreased for a value of $k = 10$, but it increased for the the largest generated subgraph.

Large scale BGP hijack in India. Similar to Figures 8 and 9, Figure 10 shows a changing pattern in the average path length depending on the crust being analyzed. These changes correlated with the abrupt discontinuities illustrated in Figure 6.

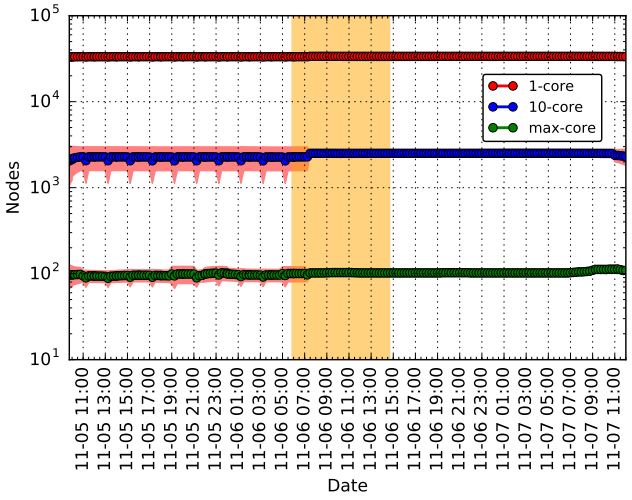

Figure 7: Nodes per crust India event.

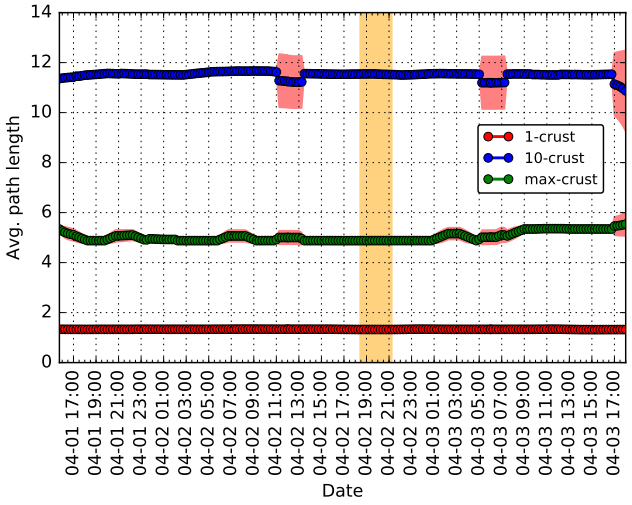

Figure 8: Average path length in the crust Indonesia event.

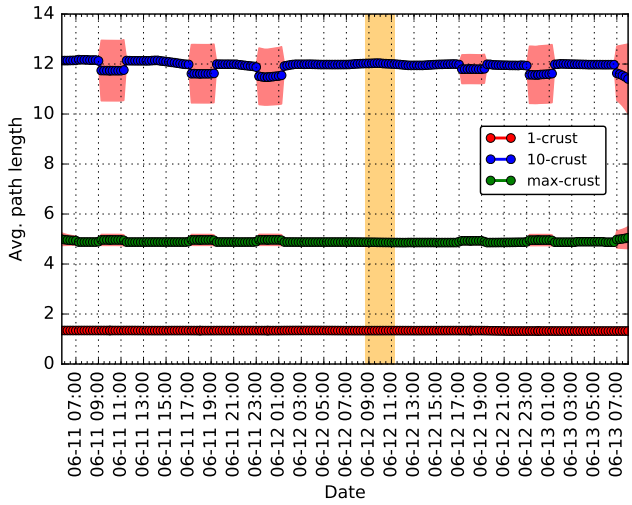

Figure 9: Average path length in the crust Malaysia event.

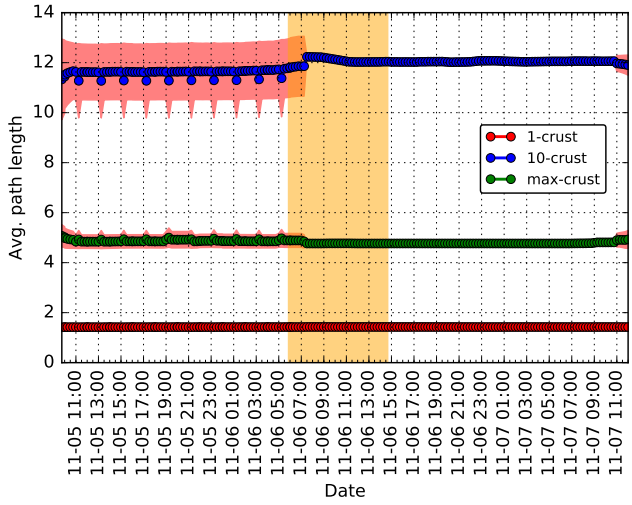

Figure 10: Average path length in the crust India event.

3.3 Community structure

An Indonesian ISP hijacking the world. Community structure measures provide a sense of how clustered are nodes in a graph. This is relevant when studying how disruption in the grouping between ASes can be an indicator of an anomalous event. Figure 26 shows the average clustering coefficient of the graph snapshots during the period of study. This measure seems to be stable during the observation period except for the discontinuities around the same time as we observed before in the centrality and average path length measurements.

We then looked at the average clustering coefficient in the core and crust subgraphs for different values of k in Figures 27 and 11. It is worth nothing that—in advance—of the reported times of the incidents, it is possible to observe some disruptions in the clustering measure for both core and crust subgraphs. More interestingly, Figure 12 shows the average size of the components for the crust subgraphs. There are abrupt changes in the mean size of these components during the observation period, suggesting reallocation of ASes across the multiple shells.

Global collateral damage of Telecom Malaysia leak. For the Malaysian incident, Figure 28, shows the average clustering coefficient for the whole graph—with no k-shell decomposition applied yet. Discontinuities in the signal are observed in correspondence with the same behavior exhibited for other structural properties measured at the general graph, e.g., Figure 20. We also studied the patterns in the number of nodes in the core and crust subgraphs. Figures 29 and 13 shows the variability in these patterns. They seem to coincide with previous illustrated discontinuities for the whole graph snapshots. Figure 14 shows the mean number of nodes in the graph components of the crust subgraphs. Variations in this property are not as much evident as for the case of the Indonesian incident.

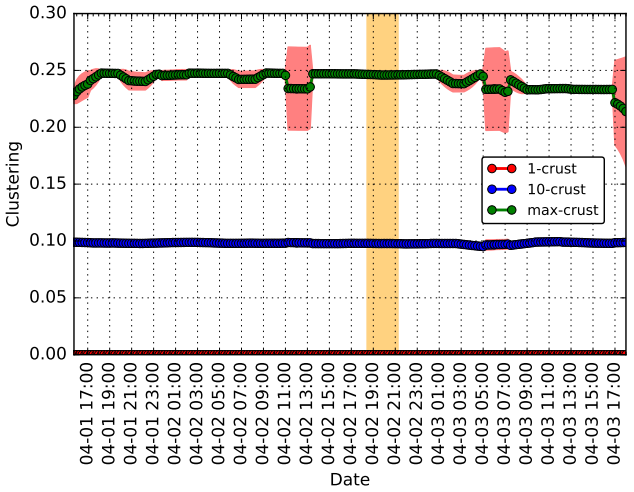

Figure 11: Clustering per crust Indonesia event.

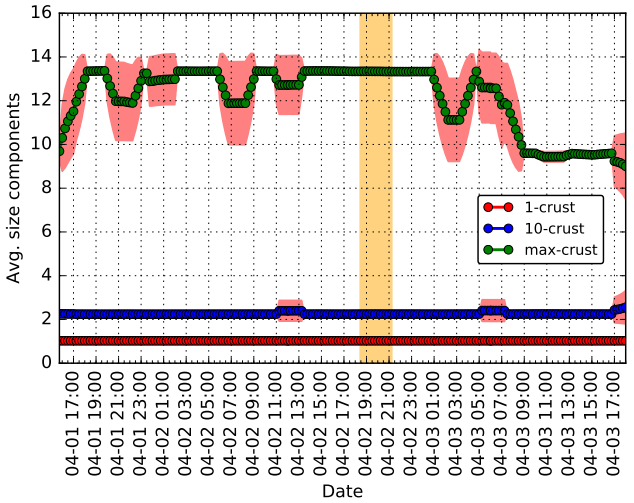

Figure 12: Average size components crust Indonesia event.

Large scale BGP hijack in India. Finally, for the Indian incident, we report similar metrics in the clustering measurements as for previous anomalous events. Figure 30 shows the time series of the average clustering coefficient. In general, the signal has discontinuities in accordance with centrality measures plots. Figures 31 and 15 capture the same property for core and crust subgraphs. It is of interest that for both—core and crust—measurements, there is a significant reduction in the clustering even under the presence of discontinuities as noticed in the case of centrality measures. Finally, the time series in Figure 16 confirms this observation—when it is observed continuous disruption in the average size of components during the observation period.

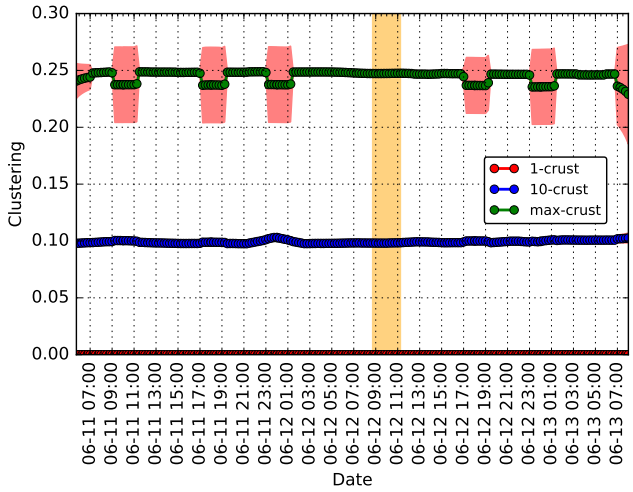

Figure 13: Clustering per crust Malaysia event.

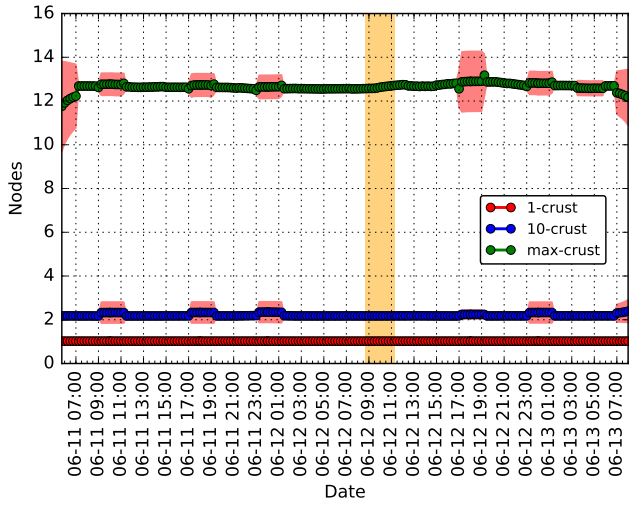

Figure 14: Average size components crust Malaysia event.

4 DISCUSSION

The purpose of this work is to explore the applicability of graph mining to the challenge of identifying BGP anomalies, i.e., considering inputs from the dynamic representation of the AS-level graph. We found some value in examining the robustness of AS-level graph properties in terms of early warning of incidents. A solution that focuses on the monitoring of the dynamic evolution of the AS-level graph may help detect anomalies that are not yet evident using traditional control- and data-plane measurements. In the k-shell decomposition, we have identified a method that is a complement to the current control- and data-plane anomaly detection approaches.

To find the most useful properties to study control-plane anomalies, we used the k-core and k-crust decomposition of the AS-level graphs. Empirically, we noticed that both the core and the crust (for various k-levels) of these representations change, in some cases, more dramatically than in others. This is particularly of interest

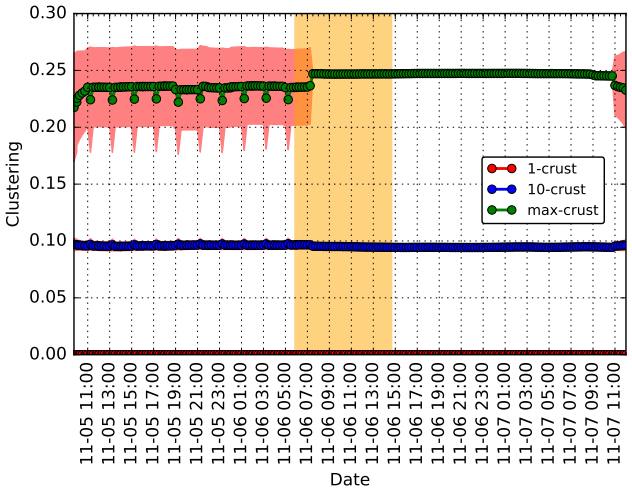

Figure 15: Clustering per crust India event.

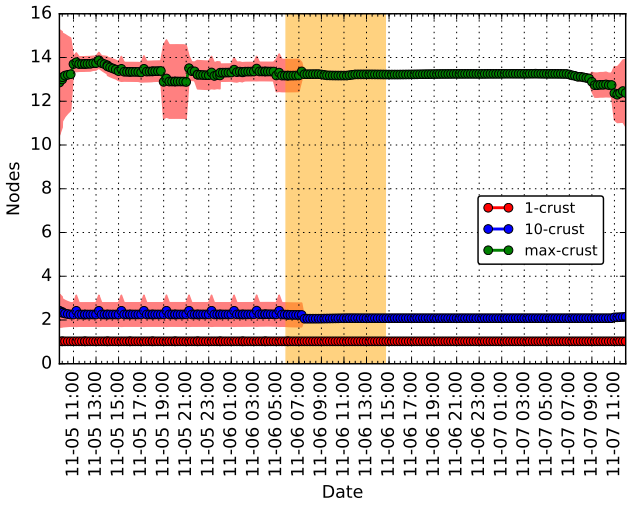

Figure 16: Average size components crust India event.

when considering the origin of the hijacks. Specifically, previous researchers have noted that the majority and more harmful attacks are orchestrated from peripheral ASes [22]. During the anomalous events, the incorrect routing information shifts the AS that is the source of the anomalous information from crust to core, as revealed in the k-shell decomposition.

We have analyzed changes in the topology for three cases of large-scale routing anomalies. We have evaluated the statistical significance of changes in the graph topology before, during, and after the event. This study has a primary focus on an empirical understanding of the robustness of topological properties under the presence of three major disruptions. We have reported a significant shift in the topological properties of the visible Internet some hours before the incidents were reported by others.

We used centrality, average path length, and clustering properties in this dynamic analysis of the AS-level graph. We characterized the network before and after the anomalies (i.e., in the absence and presence of each anomaly). The properties of the decomposed crust and core graphs remain relatively constant without the anomaly. Then there is a time period before each anomaly was detected in which the properties of the crust graph suddenly change.

Several factors can contribute to the difference in time at which we notice changes in the structural properties (with respect to the time announced by other researchers). First, there is a matter of sampling involved in the generation of the graphs. We use publicly available data that may have missing links; in particular, links between customers may be underrepresented [29].

It is worth noting that the presented characterization relies exclusively on the BGP data captured by the RouteViews project. That means that the characterized AS-level graph might not entirely represent the routing infrastructure at the moment of sampling. This is because route information does not necessarily take into account effective changes in the IP network ownership or commercial relationships between network operators. Thus, effective network routing changes are not necessarily captured by the routers. This may cause a lack of precision when mapping the AS-level graph, and thus, the characterization of structural network properties derived from the graph. This may influence the accuracy of our construction of the graph from the AS-level topology. In addition, the study that we perform in this paper is based on the effect of a large-scale events. Therefore, it will be interesting to evaluate if the current approach also applies for other hijack events in which there are fewer number of IP prefixes compromised.

Given that the analysis relies on assumptions concerning the generation of the observed data, the conclusions are also based on that data. Timing and scope are both issues in data compilation. It might be the case that the data is incomplete or delayed which would impinge our analysis. These considerations are important when trying to compare the effectiveness of the proposed approach with the traditional ones discussed above.

5 RELATED WORK

There is a plethora of research attempting to address the problem of Internet routing anomaly identification [13]. The work in [5] proposes a two-tier hierarchy to classify the proposed solutions. The first type of solution encompasses the inclusion of cryptographic, signature-based authentication for route announcements. Cryptographic-signed messages allow the verification of the identity of ASes that claim a certain route. On the one hand, when the authentication only certifies the origin of a certain prefix, the standard is called Resource Public Key Infrastructure (RPKI) [24]. The main idea behind this approach borrows concepts from the Public Key Infrastructure (PKI) framework that is used to authenticate web certificates. RPKI extrapolates this idea to examine the veracity of the origin of routes. On the other hand, another proposed standard is BGPSec [25]. BGPSec uses the RPKI to distribute and manage cryptographic keys that are used to authenticate every AS on the path of a corresponding announcement. In this approach, instead of authenticating the origin of a certain route, the proposal is to authenticate every AS in a certain path.

Although both schemes are able to protect against the announcement of bogus routes, they fail when trying to avoid the adoption of leaked routes (i.e., when an AS announces valid routes to too

many of its neighbors, violating economic agreements). These types of anomalies still can generate blackholes and are even prone to traffic interception [14]. Among other reasons that limit the scope of cryptographic approaches for securing the Internet, researchers have debated the agreement of a trusted Certificate Authority for RPKI [7], the difficulties to correctly configure the RPKI [43], and a general lack of commitment and incentives to lead their implementation [12].

The second type of solution relies on methods that focused on identifying anomalies in control- and data-plane measurements, i.e., monitoring BGP announcements and packet traffic respectively. On one hand, methods at the control-plane level require dealing with BGP routing data measurements to find inconsistencies at the routing level. The basic idea of this approach is to detect prefixes with Multiple Origin AS (MOAS). Once the conflicts are detected, these methods filter false positives by using additional information from the network operators, i.e., checking announcements of similar prefixes from a different ASes that belong to the same organization. For example, the work in [20] illustrates how this type of approach can detect anomalies that have a huge impact, i.e., announcements that pollute a considerable number of paths. On the other hand, data-plane based methods rely on exploring the reachability of routes in specific ASes. For example, the approach in [46] generates an alarm every time the reachability of a predefined prefix is not observable from multiple vantage points, requiring the deployment per AS.

Hybrid approaches have been developed to address the limitations of exclusively control- and data-plane methods. The main idea behind hybrid approaches is to use control-plane inconsistencies to inform data-plane measurements that explore the reachability of packets for a particular network. The work in [38] explores this idea by introducing a system that creates announcements for BGP anomalies and alarms for potential hijacks. The system also exports the metadata related to the events. Although most of the events identified are not confirmed by operators, the data provided by this mechanism is useful for understanding the nature of the attacks and to identify possible attacks.

Finally, it is also possible to find anomaly detection schemes based on the understanding of the topological structure of the AS-level graph. In particular, the work in [21] explores the structure of the AS-level graph to point out that its topological structure is hierarchical in terms of being composed of multiple shells interconnected between them. To do so, the authors classify the ASes based on their positions on the graph as core (large ISPs) and periphery nodes (local providers). They observed that in most of the cases, periphery nodes are closely connected between them (which is also noticeable by their geographical proximity). The authors explore this fact to note that periphery nodes are not used to be connected with many core nodes. By relying in this observation, they propose a technique to infer whether routing updates are malicious, i.e., those that do not follow the observed pattern. Another work along the same lines is [22]. In this work, the authors build a synthetic graph topology based on BGP measurements to run experiments on how the position of an AS in the graph makes it more resilient to hijack attacks. The study concludes with the observation that ASes that are directly connected with core ASes are the more resilient (even more than the core ASes) but at the same time are more effective at launching these types of attacks.

6 CONCLUSIONS

When BGP was originally implemented, the operators of the control plane were part of a smaller community than is the case today, with higher levels of both trust and technical expertise. The vulnerability of the BGP trust model has since been proven by mistake or malfeasance. The solutions to this have included cryptographic protocol for ensuring trustworthy information from trustworthy sources, as well as methods for identification and remediation of anomalies when they occur.

In this paper, we have characterized BGP anomalies from a different perspective, one derived by mining Internet graphs at the AS-level. As a complement to current anomaly identification approaches, we have implemented a set of passive measurements to better understand malicious real hijack events, and show their efficacy in three large-scale examples. The proposed characterization relies on the construction of graphs using the set of BGP measurements provided by the RouteViews project. Specifically, we used approximately a day of BGP observations before and after three major events to reconstruct an AS-level graph. We then quantified the topological properties of the AS-level graph. The proposed set of measurements allowed us to identify the more relevant network features, i.e., affected metrics, when an event of such scale happens.

A natural extension of the proposed characterization is the study of similar graph properties around other malicious BGP anomalous events, ones at smaller scale. It is possible that the AS-level graph topology can be more robust against less severe attacks, i.e., when a lower number of networks are hijacked. In this work, we have explored our hypothesis under the conditions of large service disruptions (including an event that compromised roughly two-thirds of the Internet). Future work includes examining how these topological features change with smaller events. Such future analysis relies on the proper identification and labeling of past events.

Additional future work includes examining the necessary scope of data required in addition to examining the scale of the event. Is it necessary to have data on the entire network, or would regional data be adequate? Given that the proposed characterization relies on existing measurement methods and does not interfere with the existing infrastructure, partnering for such investigations is the focus of our future work.

Another consideration is to what degree higher-frequency measurements would impinge the efficacy of this approach. The approach in this paper can be also tested with BGP collectors that provide a different frequency sampling, such as with the OpenBMP protocol [36], i.e., a protocol for monitoring purposes. OpenBMP is able to get updates every minute instead of every 15 minutes as analyzed here. An implementation of a prototype for anomaly detection based on the principles of this paper seems feasible with the availability of data from projects such as BGPStream.

REFERENCES

[1] J. Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. 2006. Large scale networks fingerprinting and visualization using the k-core decomposition. In *Advances in Neural Information Processing Systems 18*. 41–50.

[2] J. Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. 2008. K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases. *Networks and Heterogeneous Media* 3, 2 (2008), 371–393.

[3] Axel Arnbak and Sharon Goldberg. 2015. Loopholes for circumventing the Constitution: Unrestrained bulk surveillance on Americans by collecting network traffic abroad. *Michigan Telecommunications and Technology Law Review* 317 (2015).

[4] M. Brown. 2008. Pakistan hijacks YouTube. http://research.dyn.com/2008/02/pakistan-hijacks-youtube-1/. (February 2008).

[5] Kevin Butler, Toni R Farley, Patrick McDaniel, and Jennifer Rexford. 2010. A Survey of BGP Security Issues and Solutions. *Proc. IEEE* 98, 1 (2010), 100–122.

[6] Kai Chen, Chengchen Hu, Wenwen Zhang, Yan Chen, and Bin Liu. 2009. On the Eyeshots of BGP Vantage Points. In *Proceedings of the 28th IEEE Conference on Global Telecommunications (GLOBECOM'09)*. Piscataway, NJ, USA, 3558–3563.

[7] Danny Cooper, Ethan Heilman, Kyle Brogle, Leonid Reyzin, and Sharon Goldberg. 2013. On the Risk of Misbehaving RPKI Authorities. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks (HotNets-XII)*. College Park, MD, USA, 16:1–16:7.

[8] J. Cowie. 2013. The New Threat: Targeted Internet Traffic Misdirection. http://research.dyn.com/2013/11/mitm-internet-hijacking/. (November 2013).

[9] John C Doyle, David L Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger. 2005. The "robust yet fragile" nature of the Internet. *Proceedings of the National Academy of Sciences of the United States of America* 102, 41 (2005), 14497–14502.

[10] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. 2012. Internet Topology over Time. (02 2012). arXiv:1202.3993v1.

[11] Marco Gaertler and Maurizio Patrignani. 2004. Dynamic analysis of the Autonomous System graph. In *International Workshop on Inter-domain Performance and Simulation (IPS 2004)*. Budapest, Hungary, 13–24.

[12] Phillipa Gill, Michael Schapira, and Sharon Goldberg. 2011. Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security. In *Proceedings of the ACM SIGCOMM 2011 Conference (SIGCOMM '11)*. Toronto, Ontario, Canada, 14–25.

[13] Sharon Goldberg. 2014. Why Is It Taking So Long to Secure Internet Routing? *Commun. ACM* 57, 10 (2014), 56–63.

[14] Sharon Goldberg, Michael Schapira, Peter Hummon, and Jennifer Rexford. 2010. How Secure Are Secure Interdomain Routing Protocols. In *Proceedings of the ACM SIGCOMM 2010 Conference (SIGCOMM '10)*. New Delhi, India, 87–98.

[15] Enrico Gregori, Alessandro Improta, Luciano Lenzini, Lorenzo Rossi, and Luca Sani. 2012. On the Incompleteness of the AS-level Graph: A Novel Methodology for BGP Route Collector Placement. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC '12)*. Boston, Massachusetts, USA, 253–264.

[16] Hamed Haddadi, Damien Fay, Almerima Jamakovic, Olaf Maennel, Andrew W Moore, Richard Mortier, Miguel Rio, and Steve Uhlig. 2008. Beyond node degree: evaluating AS topology models. *arXiv preprint arXiv:0807.2023* (2008).

[17] J. Hawkinson and T. Bates. 1996. *Guidelines for creation, selection, and registration of an Autonomous System (AS)*. RFC 1930. RFC Editor. 1–10 pages. https://tools.ietf.org/html/rfc1930

[18] Rahul Hiran, Niklas Carlsson, and Phillipa Gill. 2013. Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident. In *Proceedings of the 14th International Conference on Passive and Active Measurement (PAM'13)*. Hong Kong, China, 229–238.

[19] Steven Hofmeyr, Tyler Moore, Stephanie Forrest, Benjamin Edwards, and George Stelle. 2013. Modeling internet-scale policies for cleaning up malware. In *Economics of Information Security and Privacy III*. Springer, 149–170.

[20] Varun Khare, Qing Ju, and Beichuan Zhang. 2012. Concurrent prefix hijacks: Occurrence and impacts. In *Proceedings of the 2012 Internet Measurement Conference (IMC '12)*. Boston, MA, USA, 29–35.

[21] Christopher Kruegel, Darren Mutz, William Robertson, and Fredrik Valeur. 2003. *Recent Advances in Intrusion Detection*. Vol. 2820. Springer Berlin Heidelberg, Chapter Topology-Based Detection of Anomalous BGP Messages, 17–35.

[22] M. Lad, R. Oliveira, B. Zhang, and L. Zhang. 2007. Understanding Resiliency of Internet Topology against Prefix Hijack Attacks. In *37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07)*. Washington, DC, USA, 368–377.

[23] Nir Lahav, Baruch Ksherim, Eti Ben-Simon, Adi Maron-Katz, Reuven Cohen, and Shlomo Havlin. 2016. K-shell decomposition reveals hierarchical cortical organization of the human brain. *New Journal of Physics* 18, 8 (2016), 083013.

[24] M. Lepinski and M. Kent. 2012. *An Infrastructure to Support Secure Internet Routing*. RFC 6480. RFC Editor. 1–24 pages. https://tools.ietf.org/html/rfc6480

[25] M. Lepinski and K. Sriram. 2017. *BGPsec Protocol Specification*. RFC 18. RFC Editor. 1–25 pages. https://tools.ietf.org/html/draft-lepinski-bgpsec-protocol-00

[26] Riad Mazloum, Marc-Olivier Buob, Jordan Auge, Bruno Baynat, Dario Rossi, and Timur Friedman. 2014. Violation of Interdomain Routing Assumptions. In *Proceedings of the 15th International Conference on Passive and Active Measurement*. Los Angeles, CA, USA, 173–182.

[27] D. Meyer. 2004. University of Oregon Route Views Archive Project. http://archive.routeviews.org. (June 2004).

[28] RIPE NCC. 2011. RIS Raw Data. https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data. (February 3 2011).

[29] Ricardo Oliveira, Dan Pei, Walter Willinger, Beichuan Zhang, and Lixia Zhang. 2010. The (in)Completeness of the Observed Internet AS-level Structure. *IEEE/ACM Trans. Netw.* 18, 1 (2010), 109–122.

[30] Chiara Orsini, Alistair King, Danilo Giordano, Vasileios Giotsas, and Alberto Dainotti. 2016. BGPStream: A software framework for live and historical BGP data analysis. In *Proceedings of the 2016 Internet Measurement Conference (IMC '16)*. Santa Monica, CA, USA, 429–444.

[31] Romualdo Pastor-Satorras and Alessandro Vespignani. 2007. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press.

[32] A. Peterson. 2013. Researchers say U.S. Internet traffic was re-routed through Belarus. That's a problem. https://www.washingtonpost.com/news/the-switch/wp/2013/11/20/researchers-say-u-s-internet-traffic-was-re-routed-through-belarus-thats-a-problem/. (November 2013).

[33] Boris Pittel, Joel Spencer, and Nicholas Wormald. 1996. Sudden emergence of a giant k-core in a random graph. *Journal of Combinatorial Theory, Series B* 67, 1 (1996), 111–151.

[34] Yakov Rekhter and Tony Li. 1995. *A Border Gateway Protocol 4 (BGP-4)*. RFC 1654. RFC Editor. 1–56 pages. http://www.rfc-editor.org/rfc/rfc1654.txt

[35] Yakov Rekhter, Tony Li, and S. Hares. 2006. *A Border Gateway Protocol 4 (BGP-4)*. RFC 4271. RFC Editor. 1–90 pages. https://tools.ietf.org/html/rfc4271

[36] J. Scudder, R. Fernando, and S. Stuart. 2016. *BGP Monitoring Protocol*. RFC. RFC Editor. 1–26 pages. https://tools.ietf.org/html/draft-ietf-grow-bmp-17

[37] A. Shaw. 2013. Spam? Not Spam? Tracking a hijacked Spamhaus IP. https://greenhost.nl/2013/03/21/spam-not-spam-tracking-hijacked-spamhaus-ip/. (March 2013).

[38] Xingang Shi, Yang Xiang, Zhiliang Wang, Xia Yin, and Jianping Wu. 2012. Detecting Prefix Hijackings in the Internet with Argus. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference (IMC '12)*. Boston, MA, USA, 15–28.

[39] Yixin Sun, Anne Edmundson, Laurent Vanbever, Oscar Li, Jennifer Rexford, Mung Chiang, and Prateek Mittal. 2015. RAPTOR: Routing Attacks on Privacy in Tor. In *Proceedings of the 25th Conference on USENIX Security Symposium*. Washington, DC, USA, 271–286.

[40] A. Toonk. 2014. Hijack event today by Indosat. http://bgpmon.net/hijack-event-today-by-indosat/. (April 2014).

[41] A. Toonk. 2015. Large scale BGP hijack out of India. https://bgpmon.net/large-scale-bgp-hijack-out-of-india/. (November 2015).

[42] A. Toonk. 2015. Massive route leak causes Internet slowdown. https://bgpmon.net/massive-route-leak-cause-internet-slowdown/. (June 2015).

[43] Matthias Wählisch, Olaf Maennel, and Thomas C. Schmidt. 2012. Towards Detecting BGP Route Hijacking Using the RPKI. *SIGCOMM Comput. Commun. Rev.* 42, 4 (2012), 103–104.

[44] Jaewon Yang and Jure Leskovec. 2014. Overlapping communities explain core–periphery organization of networks. *Proc. IEEE* 102, 12 (2014), 1892–1902.

[45] Guo-Qing Zhang, Guo-Qiang Zhang, Qing-Feng Yang, Su-Qi Cheng, and Tao Zhou. 2008. Evolution of the Internet and its cores. *New Journal of Physics* 10, 12 (2008), 123027.

[46] Zheng Zhang, Ying Zhang, Y.C. Hu, Z.M. Mao, and R. Bush. 2010. iSPY: Detecting IP Prefix Hijacking on My Own. *IEEE/ACM Transactions on Networking* 18, 6 (2010), 1815–1828.

[47] E. Zmijewski. 2014. Indonesia Hijacks the World. http://research.dyn.com/2014/04/indonesia-hijacks-world/. (April 2014).

7 APPENDIX

In this section, we provide more details on the additional empirical measurements we computed to test the proposed hypothesis.

7.1 Centrality measures

An Indonesian ISP hijacking the world. Here, we report on the results of centrality measures which illustrate the prominence of ASes. Figure 17 shows the number of nodes over time. From this plot, it is possible to infer that the only significant change in this measure is for the graph that is captured at April 2, 2014, at 12:00 and April 3, 2014, at 6:00.

To better understand this behavior, we also study the dynamic transition of the number of edges in Figure 18. We observe that

there is a considerable decrease in the number of edges for the graphs that are built on the same snapshots—in accordance with the measure of the number of nodes.

Figure 19 shows the number of nodes at various k-levels of cores graphs, i.e., $k = 1, 10$, and the maximum k possible—the one that encloses the nucleus of the Internet. As we might expect, it is not possible to observe significant changes regarding the total number of nodes in this time series. This suggests that the core remains almost the same with respect to the number of ASes.

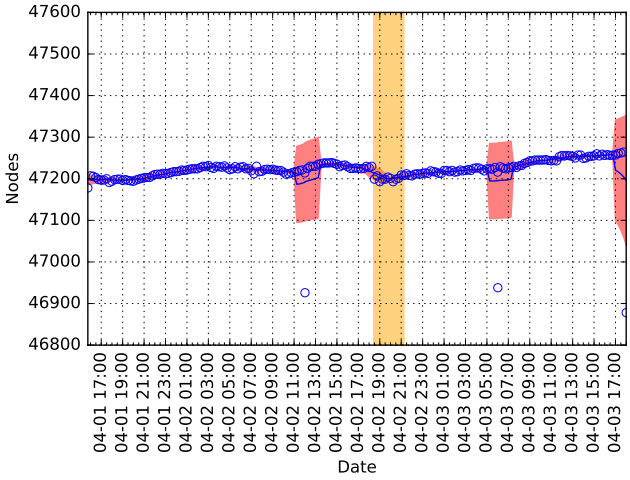

Figure 17: Number of nodes Indonesia event.

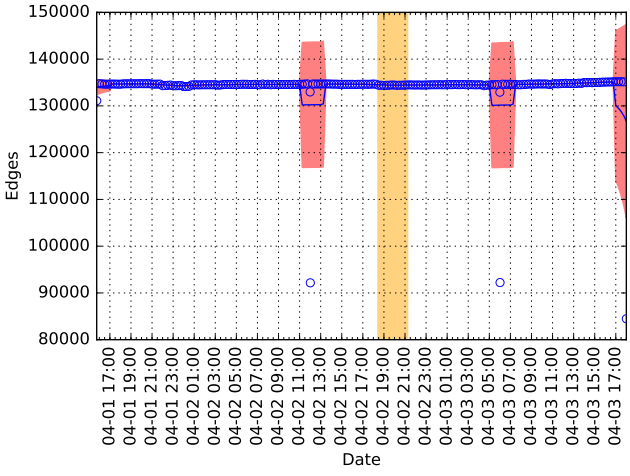

Figure 18: Number of edges Indonesia event.

Global collateral damage of Telecom Malaysia leak. For this incident, Figures 20, 21 illustrate general centrality measures for the number of nodes and edges, respectively. As can be seen, the only significant variations for these properties occur for observations derived at June 11, 2015, at 10:00, and 18:00; June 15, 2014, at 00:00, and 18:00; and June 13, 2015, at 00:00, and 8:00. Similarly, for the core graphs, we computed the number of nodes as is shown in Figure 22. We did not observe significant variations for these properties.

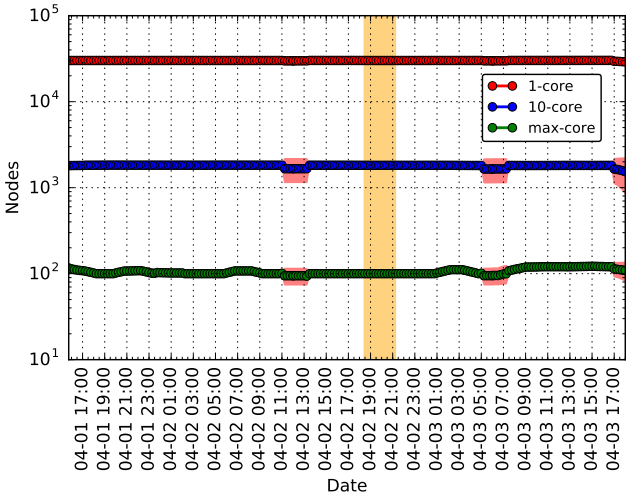

Figure 19: Nodes per core Indonesia event.

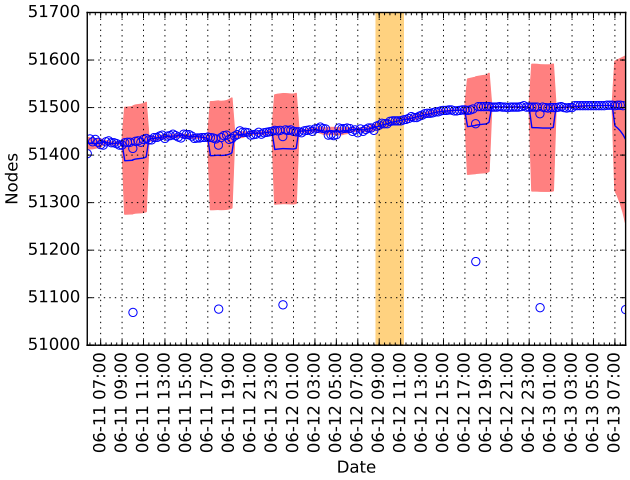

Figure 20: Number of nodes Malaysia event.

Large scale BGP hijack in India. For the Indian incident, we tracked of the same properties we did for the Indonesian and Malaysian incidents. In particular, Figures 23, 24 show the number of nodes and edges, for each graph snapshot during the observation period. As in the previous cases, there are some discontinuities in each of these time series. The discontinuities are evident in November 5, 2015, at 10:00, 12:00, 14:00, 16:00, 18:00, 20:00, 22:00; and November 6, 2015, at 00:00, 02:00, 04:00, and 06:00. Figure 25 shows the number of nodes in the core subgraphs. This measure does not reveal significant changes during the observation period.

7.2 Community structure

An Indonesian ISP hijacking the world. Figure 26 shows the average clustering coefficient of the graph snapshots during the period of study. This measure seems to be stable during the observation period except for the discontinuities around the same time as we

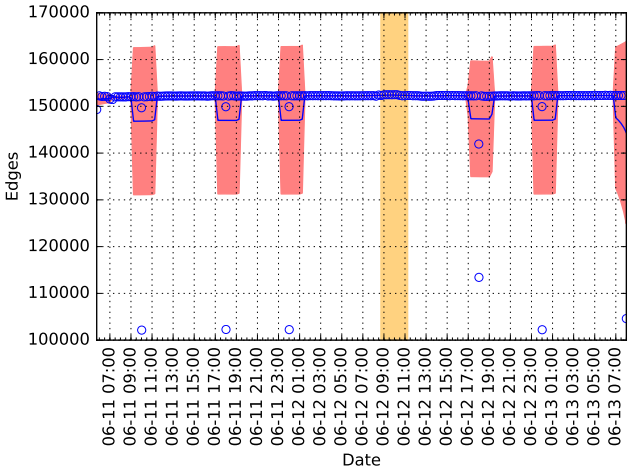

Figure 21: Number of edges Malaysia event.

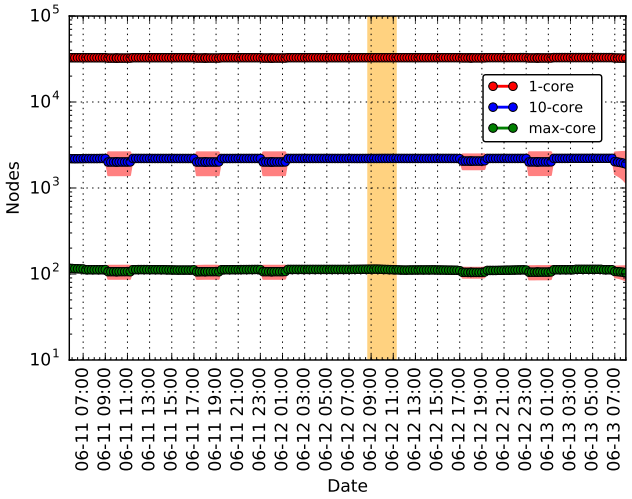

Figure 22: Nodes per core Malaysia event.

observed before in the centrality and average path length measurements.

We then looked at the average clustering coefficient in the core subgraphs for different values of k in Figure 27. It is worth noting that—in advance—of the reported times of the incidents, it is possible to observe some disruptions in the clustering measure for the core subgraphs.

Global collateral damage of Telecom Malaysia leak. For the Malaysian incident, Figure 28, shows the average clustering coefficient for the whole graph—with no k-shell decomposition applied yet. Discontinuities in the signal are observed in correspondence with the same behavior exhibited for other structural properties measured at the general graph, e.g., Figure 20. We also studied the patterns in the number of nodes in the core subgraphs. Figure 29 shows the variability in this pattern. It seems to coincide with previous illustrated discontinuities for the whole graph snapshots.

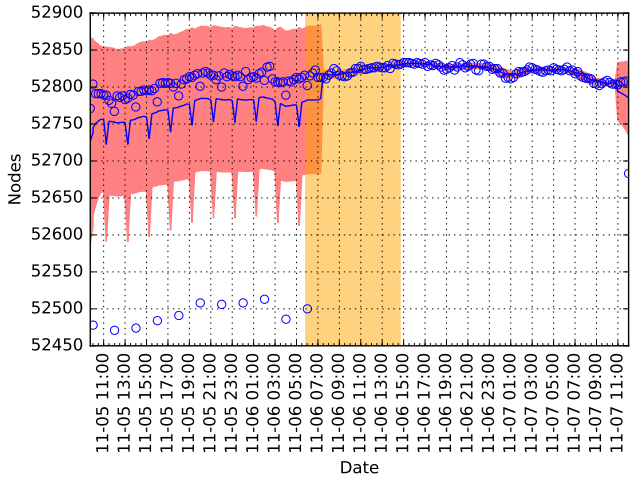

Figure 23: Number of nodes India event.

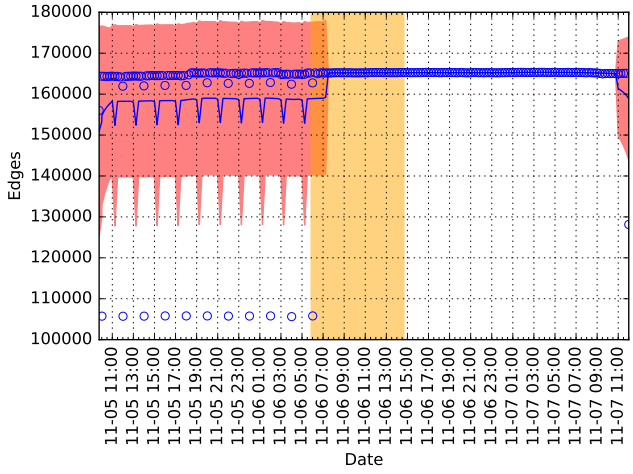

Figure 24: Number of edges India event.

Large scale BGP hijack in India. Finally, for the Indian incident, we report similar metrics in the clustering measurements as for previous anomalous events. Figure 30 shows the time series of the average clustering coefficient. In general, the signal has discontinuities in accordance with centrality measures plots. Figure 31 captures the same property for core and crust subgraphs. It is of interest that for core measurements, there is a significant reduction in the clustering even under the presence of discontinuities as noticed in the case of centrality measures.

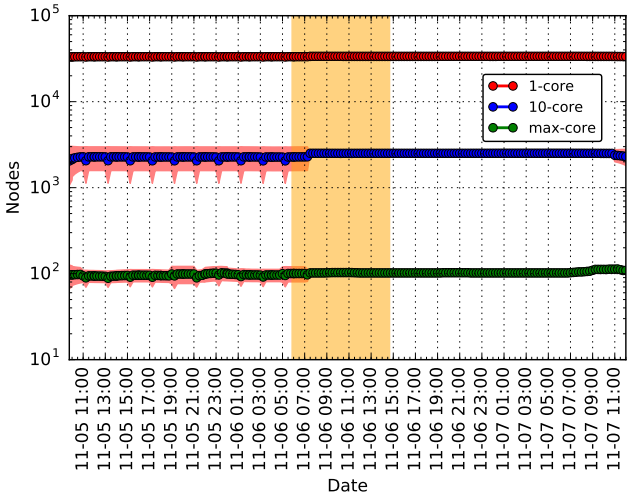

Figure 25: Nodes per core India event.

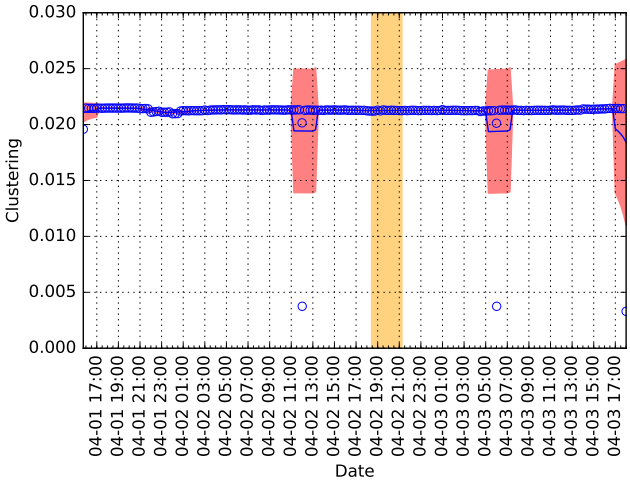

Figure 26: Clustering coefficient Indonesia event.

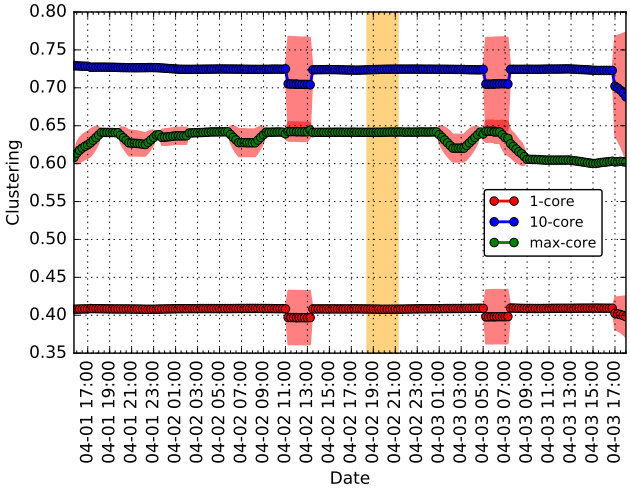

Figure 27: Clustering per core Indonesia event.

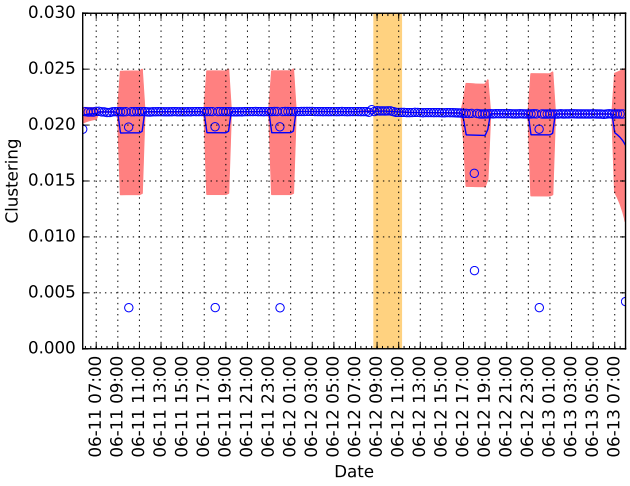

Figure 28: Clustering coefficient Malaysia event.

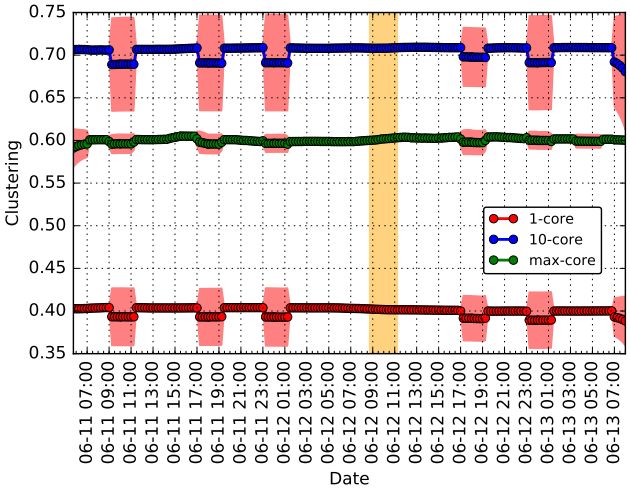

Figure 29: Clustering per core Malaysia event.

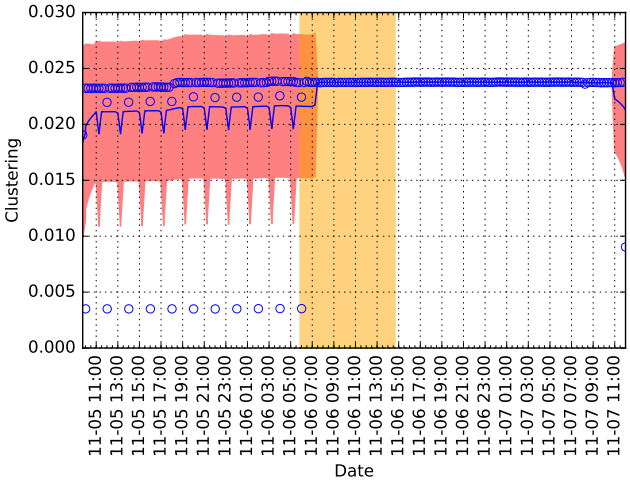

Figure 30: Clustering coefficient India event.

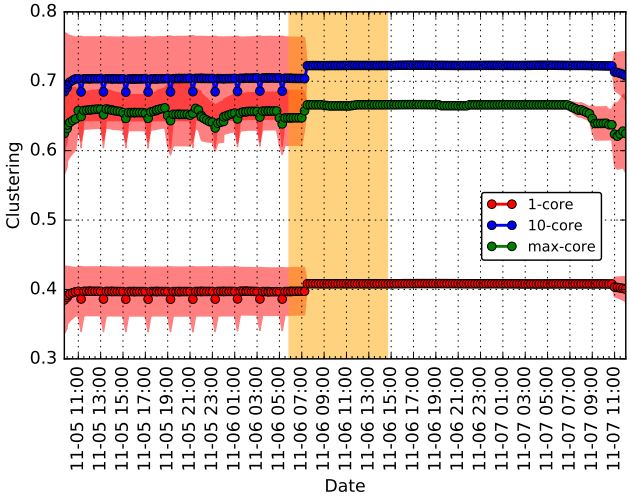

Figure 31: Clustering per core India event.