BGP Eye: A New Visualization Tool for Real-time Detection and Analysis of BGP Anomalies

Soon-Tee Teoh[1], Supranamaya Ranjan[2], Antonio Nucci[2], Chen-Nee Chuah[3]
[1] – San Jose State University, San Jose, CA, USA
[2] – Narus Inc, Mountain View, CA, USA
[3] – University of California Davis, Davis, CA, USA

ABSTRACT

Owing to the inter-domain aspects of BGP routing, it is difficult to correlate information across multiple domains in order to analyze the root cause of the routing outages. This paper presents *BGP Eye*, a tool for visualization aided root-cause analysis of BGP anomalies. In contrast to previous approaches, BGP Eye performs real-time analysis of BGP anomalies through a hierarchical deep-dive approach. First BGP updates are clustered to obtain BGP events that are more representative of an anomaly and then these events are correlated across all the border routers to ascertain the extent of the anomaly. Furthermore, BGP Eye provides both the capability to analyze BGP anomalies from an Internet-Centric View through multiple vantage points as well as from a Home-Centric View of a particular Autonomous System (AS). We present the capability for scalable and real-time root-cause analysis provided by BGP Eye through the analysis of two widely contrasting anomalies. First, we provide an Internet-Centric view towards the routing outages caused during the spread of the Slammer Worm on January 25th, 2003 from AS568 and; Second, we provide a Home-Centric view from the perspective of AS6458 into the routing outages caused by the inadvertent prefix hijacking by AS9121 on December 24th, 2004.

1. INTRODUCTION

The Internet is a global, decentralized network comprised of many smaller inter-connected networks. A network under the administrative control of a single organization is called an *Autonomous System* (AS). The process of routing within an AS is called *intra-domain routing* and routing between ASes is called *inter-domain routing*. The dominant inter-domain routing protocol on the Internet is the *Border Gateway Protocol* (BGP). Although BGP's simplicity and resilience have enabled it to play a fundamental role within the global Internet, it has historically provided very limited performance or security guarantees, which often contribute to a global-scale instability and outages.

Due to the inter-domain aspect of BGP, even small routing failures within an AS can sometimes propagate widely into the rest of the Internet causing significant and widespread damage. One such failure occurred on April 25th 1997, when a mis-configured router maintained by a small Internet Service Provider (ISP) in Virginia injected incorrect routing information into the global Internet and claimed to have optimal connectivity to all Internet destinations. Since there are no mechanisms to validate route announcements, most Internet traffic was re-directed to this small ISP, which overwhelmed the mis-configured routers and effectively crippled the Internet for almost two hours. Loss of reachability on the Internet can be consequence of human mistakes (e.g., router misconfigurations) or malicious activities. Several Internet worm outages in the past have indicated the increasing vulnerability of the Internet routing infrastructure to attacks that initially start out in the data plane. For example, the SQL Slammer worm outbreak [21] on Jan 25th, 2003 started by exploiting a vulnerability in the MS SQL server. Although the SQL Slammer worm was not directly targeted towards the Internet routing infrastructure, it resulted in several AS peering links being overloaded on account of the sudden traffic surge. The congestion caused by the port-scanning and worm-payloads resulted in the BGP peering sessions being reset, thus starting the chain of events that led to the huge surge in the number of BGP updates. Such coincidences between worm propagation and Internet routing event surges have been observed in the past as well, such as the Code-Red [22] worm outbreak in July 2001 and the Nimda [23] worm outbreak in September 2001.

As the number of critical applications on the Internet grows, so will the reliance on it to provide reliable and secure services. Because of the increased importance of the Internet, there is much more interest in increasing the security of its underlying infrastructure, including BGP. Although such assertions might seem novel, they are not: the United States government cites BGP security as part of the national strategy for securing the Internet [Department of Homeland Security 2003]. Further, BGP security issues are being looked by both IETF working groups [25] as well as NANOG [26], which is an indication of the importance and relevance of this topic.

1.1 Motivation and Related Work

Several approaches have been proposed recently on root-cause analysis of BGP routing changes [1, 2, 3, 4, 5, 6]. These studies analyze streams of BGP update messages from several vantage points throughout the Internet, with the goal of inferring the most likely cause of the problem and its location. Although these approaches provide "identification of ASes that are involved in the same problem", they do not consider what really matters to a Network Operator, i.e. the view of their specific AS, referred in the paper as *customer-AS*. On the other hand, in [7] the authors proposed a new approach for root-cause analysis that analyzes BGP routing changes from the perspective of a customer-AS such to quantify the effects of these changes on that specific network. *An ideal system would combine both views, e.g. multiple vantage points and customer-AS specific, such to add up the benefits of both approaches.*

The two categories of approaches described above suffer from two serious limitations. First, they produce "large" textual reports that Operators have to parse through to sort out problems related to their networks. This procedure may be time-consuming and inefficient. Second, most of the approaches offer only an off-line capability for data analysis. *An ideal system would establish a real-time interaction between end-users and network traffic such that users can gain insight of both network dynamics and hidden traffic patterns and analyze/react "on-the-fly" to the undergoing problems.*

In order to accomplish the above, several visualization tools were proposed. These visualization tools empower the user to develop a reference model of what is normal on their own network so that they can diagnose problems better and faster. For example, one existing system which maps BGP path attributes to an AS graph is BGPlay [15]. When the user starts BGPlay, a query window will appear, where the user enters the prefix and time interval. The BGPlay server will then query the database for all updates to the specified prefix during the specified time interval. The animation window then displays routing activity of that prefix including (a) a histogram of the number of events over time and (b) an AS graph showing paths that change versus stable paths during the query interval.

Another system that also maps BGP attributes to an AS graph is LinkRank [16]. In a LinkRank graph, the weight of an inter-AS link is determined by the number of prefixes having an AS path that includes that link. In a Rank-Change graph, the weight on each link is the difference between the LinkRank of that link over time. A negative weight indicates routes lost on a link, while positive weight indicates routes gained in that time period. A similar system is the TAMP graph [19], which shows how many prefixes are carried over an AS-AS link. The edge colors indicate how the statistics are changing (e.g., no change, losing prefixes, gaining prefixes, and prefix count flapping too fast to animate).

The Elisha system [20] also contains network visualization of BGP updates. In this system, all the paths from the observation point AS to the origin AS of the IP prefix is plotted. This system also allows animation over time, so that at each frame, all the AS paths used in the time interval are displayed. The color represents the time (less recently or more recently) the path was used within the currently-displayed time window.

1.2 An Alternate Solution: BGP Eye

These existing visualization tools focus only on raw information, i.e. BGP updates, and do not give any deep insight into the problem. On this perspective, in this paper we introduce a new visualization tool, called *BGP Eye*, that provides a real-time status of BGP activity with easy-to-read layouts. The tool has been designed such to meet criteria like: i) *scalability*, i.e. the ability to process and display a large set of data at very fine time-scales for large-size network deployment; ii) *efficiency*, i.e. variety of different graphical layouts that provide a complete view of the BGP routing behavior; iii) *readability*, i.e. clear and easy-to-read layouts that enable Operators to promptly detect, classify, analyze the under-going anomaly and report rich-enough feedback to Operators in order for them to take the appropriate counter actions. Compared to previous visualization tools, *BGP Eye* is novel in the following ways:

- The scalability of BGP Eye is derived from its use of the metric of *BGP event* which clusters together multiple short-lived somewhat arbitrary BGP updates likely to be originated by the same network problem. This concept allows Operators to identify a *small* number of *important* routing disruptions from a *large volume* of raw BGP updates.
- It provides a deeper and wider view of BGP activity as a whole, e.g. how multiple ASes interact with each other (*Internet-Centric*), as well as AS specific, e.g. how problems involving ASes far away from the customer-AS might affect its normal behavior (*Home-Centric*).
- It provides new information like number of BGP events carried over each AS-AS link, and number of BGP events originated from each AS.
- It is capable of de-noising and profiling BGP events over time using the *Exponential Weighted Moving Average* technique for all/or highlighted ASes and all/or highlighted AS-AS links.
- It can deep-dive into the problem displaying information never shown before, as: (i) total number of BGP events per single and multiple border routers; (ii) classification of BGP events according to family types and per each border router; (iii) prefix status and (iv) prefix instability.

Step	Policy
1	Ignore if the next hop is unreachable
2	Highest Local Preference
3	Shortest AS path
4	Lowest Origin Type
5	Lowest MED among routers from same AS
6	eBGP routes over iBGP routes
7	Lowest IGP cost ("Hot-Potato routing")
8	Lowest router ID

Table 1: BGP Decision Process

The paper is organized as follows. In Section 2 we describe some basic concepts of the BGP routing protocol and how we classify *BGP events* into several family types, as previously defined in [7]. In Section 3 we introduce *BGP Eye* and describe in great details all the layouts supported. In Section 4 we show how *BGP Eye* is able to promptly detect two common problems widely experienced in the Internet: *prefix reachability* and *prefix authenticity*. As an example of prefix reachability, we analyze the impact of the Slammer worm outbreak on January 25th, 2003 from the perspective of *AS568*, which belongs to the *Department of Defense* (DoD). We show how AS568 was severely infected by the Slammer worm and how AS568 started to actively spread the worm deeply and widely across the Internet. As an example of prefix authenticity, we analyze the inadvertent hijacking of 100,000+ prefixes by Turkey Net (AS9121) on December 24th, 2004. Specifically, from the view point of a few ASes (AS6453, AS3257), we study the impact of this colossal hijacking event, and conclude that the impact of a colossal event as this could vary widely at different points in the Internet, depending on the peering topology. Section 5 concludes the paper.

2. FROM BGP UPDATES TO BGP EVENTS

We give a quick overview of BGP routing protocols in Section 2.1 and discuss the clustering process of BGP updates into BGP events and their family-types in Section 2.2

2.1 BGP Overview

BGP is the routing protocol that ASes use to exchange information about how to reach destination *address blocks* (or *prefixes*). Three important aspects of BGP are:

- **Path-vector protocol:** Each BGP advertisement includes the list of ASes along the path and other attributes such as next-hop IP address. By representing the path at the AS level, BGP hides the details of the topology and routing inside each network.
- **Incremental protocol:** Every BGP update message is indicative of a routing change, such as the old route disappearing or

the new route becoming available.

- **Policy-oriented protocol:** Routers can apply complex policies to influence the selection of the best route for each prefix and to decide whether to propagate this route to neighbors. A router applies the decision process shown in Table 1 to compare the routes learned from BGP neighbors and select the best route. In the backbone networks, BGP route selection depends on the interaction between three routing protocols: *External BGP (eBGP)*, *Internal BGP (iBGP)* and *Interior Gateway Protocol (IGP)*. IGP determines the routing paths between two routers within the same AS. The routers use the IGP path costs in the seventh step in Table 1 to select the *closest* egress point. eBGP is used to exchange routing information with external ASes, while iBGP is used to re-advertise the routes learned from external peers to other routers in the same AS.

A single network disruption, such as a link failure or policy change can trigger multiple BGP update messages as part of the convergence process. The intermediate routes are short-lived somewhat arbitrary, since they depend on subtle timing details that drive how the routers explore alternate paths. Operators may be interested in removing the transitory behavior associated with BGP updates and analyze BGP behavior in a more stationary regime. For this matter, we introduce a level of aggregation into our analysis, by borrowing the concept of *BGP event* from [7]. A BGP event is defined as a sequence of BGP updates for the same prefix from any border router where the inter-arrival time is less than a predefined *event-timeout*. Since there is the risk that certain prefixes never converge to a stable path due to persistent routing instabilities, it is important to upper bound the maximum duration allowed for any route to converge. This is defined as *convergence-timeout*. Appropriate event-timeout and convergence-timeout values are analyzed and properly set up by *BGP Eye*.

2.2 Classification of BGP Routing events: Single and Multiple Border Router(s) View

Another way to look at the data is how each border router reaches a specific prefix over time. Indeed, although a border router might generate a BGP update or BGP event for a specific prefix at a specific point in time, it may either undergo a transient routing change only to return to the same stable best route or change to a new route. Operators may be interested in tracking the status of specific prefixes or routers over time. We use the same concept of *routing vector* proposed by [7]. Let $B = \{B_1, B_2, ..., B_N\}$ be the set of border router of the specific AS that deployed our system. Let $P = \{P_1, P_2, ..., P_M\}$ the set of all admissible prefix. Let $R_i^j(t)$ be the best route selected by the border router B_j to reach destination prefix P_i at time t. We represent each route $R_i^j(t)$ with a two entries vector $< nhop_i^j, flag_i^j > (t)$, where the entry $[nhop_i^j](t)$ corresponds to the next-hop address of the eBGP neighbor router. Since a border router B_j may select as $R_i^j(t)$ a route learned via iBGP from another border router, we use the entry $[flag_i^j](t)$ to capture this information, with value **i** in case the best route is learned via iBGP or **e** in case the best route is learned via eBGP. With this definition, we can track in time the evolution of the *routing vector* $RV_i(t) =< R_i^1(t), R_i^2(t), ..., R_i^N(t) >$ for any specific prefix $P_i \in P$. With this definition, Operators gain insight of the routing status, i.e. how each border router reaches a specific prefix over time.

From a **single border router** perspective, we can classify BGP events as:

- **No Change (NC):** For this scenario, traffic entering the network at border router B_j destined to the prefix P_i would continue to flow through the AS in the same way ($nhop_i^{j,old} = nhop_i^{j,new}$ and $flag_i^{j,old} = flag_i^{j,new}$).
- **Internal/External Path Change (IPC/EPC):** An Internal path event may cause a router to switch from one egress point to another (IPC). In this case, router B_j uses iBGP-learned route before and after the routing change ($flag_i^{j,old} = flag_i^{j,new} =$**i**) but with a different next-hop router ($nhop_i^{j,old} \neq nhop_i^{j,new}$). An external path event may cause a router to switch between eBGP-learned routes with different next-hop ASes (EPC). In this case, the $flag_i^{j,old} = flag_i^{j,new} =$**e** while the next hop changes, i.e. ($nhop_i^{j,old} \neq nhop_i^{j,new}$).
- **Loss/Gain of Egress Point (LEP/GEP):** An external event may cause a route to disappear, or be replaced by a less attractive alternative, forcing a border router to select an iBGP-learned route (LEP). In this case, a router B_j has $nhop_i^{j,old} = nhop_i^{j,new}$ but $flag_i^{j,old} =$**e** while $flag_i^{j,new} =$**i**. On the contrary, an external event may cause a router to switch between eBGP-learned routes with different next-hop ASes (GEP). In this case, a router B_j has $nhop_i^{j,old} = nhop_i^{j,new}$ but $flag_i^{j,old} =$**i** while $flag_i^{j,new} =$**e**.

As previously proposed by [7], it might be important to correlate different views from multiple BGP routers to identify which subset of the border routers have similar views of the problem.

The **multi-border router** views leads to the classification of BGP events into the following six family-types:

- **Distant/Transient Disruption (TD):** A BGP event is classified as belonging to this family if and only if *each element of its routing vector has "NC"*. These events do not have any influence on the flow of traffic through the AS. A transient disruption may cause temporary routing changes before the border routers converge back to the original BGP routes.
- **Internal Disruption (ID):** A BGP event is classified as belonging to this family if and only if *the change of each of the elements in its routing vector is either of type "NC" or type "IPC" with at least one element undergoing an "IPC"*. These events are important because they may cause a large shift in traffic as routers switch from an egress point to another.
- **Single External Disruption (SED):** A BGP event is classified as belonging to this family if and only if *only one element of its routing vector has a change of type "LEP", "GEP" or"EPC"*. Typically, an ISP has eBGP sessions with a neighboring AS at multiple geographical locations, making it interesting to highlight routing changes that affect just one of these peering points.
- **Multiple External Disruptions (MED):** A BGP event is classified as belonging to this family if and only if *multiple elements of its routing vector have a change of type "LEP", "GEP" or"EPC"*.
- **Loss/Gain of Reachability (LR/GR):** A BGP event is classified as belonging to the family *Loss of Reachability* if and only if *every element of its routing vector with an external route experiences a "LEP"*. Similarly, a BGP event is classified as belonging to the family *Gain of Reachability* if and only if *every element of its routing vector with an external route experiences a "GEP"*. A loss of reachability may signify complete loss of connectivity to the destination addresses, especially if the routers have no route for other networks that contain these addresses. A gain of reachability might also indicate a problem, if there are no existing routes for that prefix (e.g., an AS that mistakenly starts advertising a large number of small subnets).

With this type of information, Operators can cluster BGP events of the same type across prefixes to infer the real cause of a network problem, i.e. correlation across destination prefix space.

3. BGP EYE LAYOUTS

Goal of *BGP Eye* is to track the healthiness of BGP activity, raise an alert when an anomaly is detected, and indicate its most likely cause. *BGP Eye* offers two different views of BGP Dynamics: *Internet-Centric View* and *Home-Centric View*.

The *Internet-Centric View* studies the activity among ASes in terms of BGP events exchanged. We have created a graph-based visualization of BGP routing changes that displays information like: (a) moving average of the total number of BGP events originated by ASes and traversing AS-AS links; (b) instantaneous deviation from historical trends of BGP events originated by ASes and traversing AS-AS links. Operators can use this view to: (i) *monitor* the Internet stability over time, i.e. number of BGP events generated across different ASes; (ii) promptly *detect* abrupt changes in the routing activity and which ASes are experiencing the observed problem; (iii) *analyze* the propagation of the problem through the entire Internet, e.g. growing rate and spreading factor, and *forecast* the time at which their network will be hit. Moreover, this view can also help the Operator in identifying which ASes are the most unstable over time in order for them to select appropriate peers for the future or revise current peering agreements.

The *Home-Centric View* has been designed to understand the BGP behavior from the perspective of a specific AS, e.g. customer-AS. BGP updates originated and received by the customer-AS are clustered into different types of BGP events. We have created several layouts that will help Operators to: (i) *monitor* the routing dynamics of their AS and its interaction with its peers; (ii) promptly *detect* routing instabilities, prefixes involved and which border routers are processing such routes; (iii) promptly *diagnose* the causes of the problem whether is in/near/far away their AS; (iv) *predict* the potential damage associated to the undergoing routing instability by incrementally gaining knowledge of the type of instability, prefixes and border routers involved and how much traffic is associated with [1].

3.1 Internet-Centric View

The Internet-Centric view provides the Operator a real-time view of the routing activity from an Internet perspective. BGP updates are collected from several vantage-points, clustered together into BGP events and processed such to provide information associated with their AS routing paths, e.g. chain of ASes traversed, from the Origin AS, as the root of propagation tree, to all its Destination ASes, e.g. leaves of the tree. In this tree, a generic AS_a-AS_b link is drawn if a peering session was observed between the two adjacent ASes, e.g. AS_a and AS_b, when the snapshot was taken.

Due to the bursty nature of BGP updates and events, it is imperative for the tool to provide a first level de-noising of the data as well as a trend analysis. In order to achieve a good trade-off between accuracy and the tight constraint of being real-time, we use a simple but efficient learning algorithm known as *EWMA*. In each generic timeslot k, BGP Eye does the following steps: (i) collects the BGP updates originated by each AS and traversing each AS-AS link, and generates the associated BGP events. Let's $a(k)$ represent the sample of BGP events generated in the time slot k for a generic time series, e.g. generic AS or AS-AS link; (ii) computes its moving average $a_N(k)$ using the last N "good" samples stored in memory, such to smooth out the large variance present into the data and extract the major trend; (iii) uses the EWMA to predict the value for the current timeslot, e.g. $a(\hat{k})$, that would obey the historical trend. $a(\hat{k})$ is computed as $a(\hat{k}) = \beta \times a(k) + (1 - \beta) \times a_N(k)$, where β represents a decay factor chosen by the user, $0 \leq \beta \leq 1$, and strictly related to the number of samples N used to calculate $a_N(k)$, e.g. $\beta = 2/(N+1)$. Note that, a large value of β means more importance to the present, while a small value of β gives more importance to the historical trend, i.e. the past. Usually a good recommendation for β is to be equal to 0.2. (iv) evaluates the deviation of the sample $a(k)$ and its predicted value $a(\hat{k})$, e.g. $\delta(k) = |a(\hat{k}) - a(k)|$. (v) generates an alert if $\delta(k)$ exceeds a pre-specified threshold, like for example 0.3. Samples for which an alert is generated are discarded and not used to compute future running averages, thus avoiding to compromise the historical trend with bad samples.

The AS graph layout uses size and color of the objects, e.g. nodes and links, as two different dimensions to report more information to the operator. The size of an AS represents the moving average of the number of BGP events originated by the AS, e.g. $a_N(k)$, while its color represents the instantaneous deviation of the current sample from its historical trend, e.g. $\delta(k)$. The more events originating from the AS, the larger its size. The color map is shown at the lower left of Figure 1, with the minimum value mapped to blue, and the maximum value mapped to red.

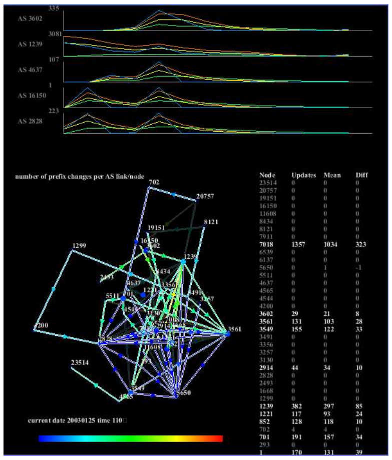
Similarly, for each AS-AS link shown in the AS graph, we represent the number of BGP events traversing link by its thickness. Specifically, the thickness of the link represents the moving average of the number of BGP events traversing the link, while its color represents the deviation of the current sample from its historical trend. In addition, we allow the user to select a few ASes to show in detail. For each selected AS, its data will be plotted over time. The data plotted are (1) the number of BGP events per time window, and (2) the moving average over time. Our system also allows highlighting of selected AS nodes and AS-AS links. There are several ways of highlighting a node: (1) the user can explicitly click on a node to toggle whether it's highlighted, (2) the user can enter the AS number of the AS to highlight, (3) the user can set a threshold such that every AS that has originated BGP events more than the threshold will be automatically highlighted, or (4) the user can set a number n such that the top n ASes/AS-AS links will be highlighted. The number n can be set by the Operator and can represent, for example, i) top n ASes/AS-AS links with the largest number of BGP events, or ii) top n ASes/AS-AS links that are deviating the most from their historical trends. Each highlighted node or link is shown in bright colors in the foreground, while nodes and links that are not highlighted are shown in dull colors in the background as context. An example is shown in Figure 1.

To visualize the Internet-AS network, we allow the user to choose among three different ways of laying out the network graph.

Force-based layout: Various force-based methods exist; a recent example is Lin and Yen's work [17]. In this method, the system starts with an initial placement of the nodes. Each link is assigned a rest length, and when it deviates from its rest length, it exerts a force on the nodes at its end-point. At each iteration, the resultant force on each node is calculated and the node is moved in the direction of the force. The AS network shown in Figure 1 has been laid out using the force-based method.

Path distance based layout: The second layout method starts by detecting all the source and sink nodes. The source nodes (with only outgoing edges) are placed at the bottom of the display, and the sink nodes are placed at the top. The system then uses a breadth-first-search algorithm to assign a distance number to each remaining node. The distance number of a node is defined as its distance from the source. All the nodes with the same distance are placed in the same horizontal layer in the display. The method then uses a heuristic to place the nodes within each layer in a way that reduces

9[1]To quantify the impact of problems detected by BGP Eye in terms of Traffic Engineering, Netflow data collected on the outgoing links can be aggregated to compute prefix-level traffic statistics. For each destination prefix involved, the Operator can generate a traffic weight that corresponds to the percentage of traffic destined to that prefix across the overall traffic volume in the network. The weights allow the Operator to estimate the potential impact of occurrence of routing events we have discussed previously.



Node	Updates	Mean	Diff
23514	0	0	0
20757	0	0	0
19151	0	0	0
16150	0	0	0
11608	0	0	0
8434	0	0	0
8121	0	0	0
7911	0	0	0
7018	1357	1034	323
6539	0	0	0
6137	0	0	0
5650	0	1	-1
5511	0	0	0
4637	0	0	0
4565	0	0	0
4544	0	0	0
4200	0	0	0
3602	29	21	8
3561	131	103	28
3549	155	122	33
3491	0	0	0
3356	0	0	0
3257	0	0	0
3130	0	0	0
2914	44	34	10
2828	0	0	0
2493	0	0	0
1668	0	0	0
1299	0	0	0
1239	382	297	85
1221	117	93	24
852	128	118	10
702	4	4	0
701	191	157	34
293	0	0	0
1	170	131	39

Figure 1: Visualization of number of BGP events originating from selected ASes and carried over selected AS-AS links

the number of crossings of the links. Figure 2 shows an example of the distance-layer layout.

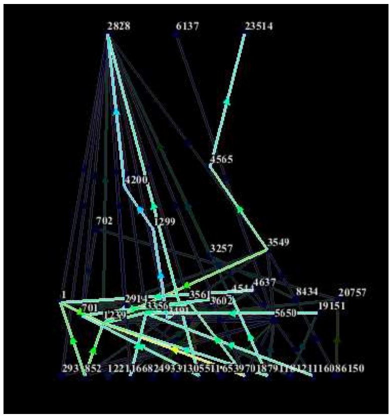


Figure 2: Layout of AS nodes according to their distance from the observation point

Manual layout: In this method, we allow the user to manually adjust the position of each node by clicking and dragging the node. This method can be used in conjunction with either of the two above methods.

3.2 Home-Centric View

Another view provided by our system is the "Home-centric" view, where we focus on how the BGP events are observed from the customer-AS, much like how this term is used in Ball et. al's work [18]. The first panel is shown in Figure 3. In this view, we place the routers in the customer-AS in the inner ring, and their peer routers, belonging to other ASes, in the outer ring. In the outer layer, the method groups routers belonging to the same AS together. We use a node placement algorithm to obtain a placement of the nodes that reduces the distance between connected nodes.

After we have placed the nodes, we draw lines between the inner routers and outer routers which are connected. The size/thickness and color of the nodes/links are assigned using the same principles described for Internet-Centric layouts represented in Figures 1,2. The links are drawn as curved lines to avoid cutting across the inner circle. If the internal radius is r_1, the external radius is r_2, the angle of an internal node is θ, and the angle of an external node is ϕ, then, the curved line between the two nodes is ($(p \times r_1 + (1-p) \times r_2) \times cos(p \times \theta + (1-p) \times \phi)$, $(p \times r_1 + (1-p) \times r_2) \times sin(p \times \theta + (1-p) \times \phi)$), $0 \leq p \leq 1$. An example is shown in Figure 3.

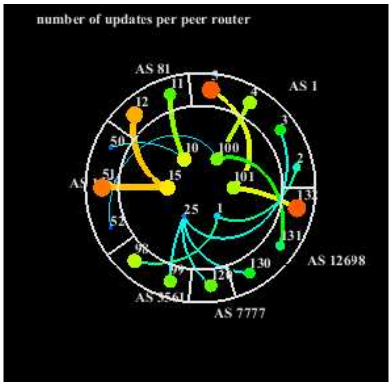


Figure 3: Circular layout of observation routers in the inner ring, and peer routers in the outer ring

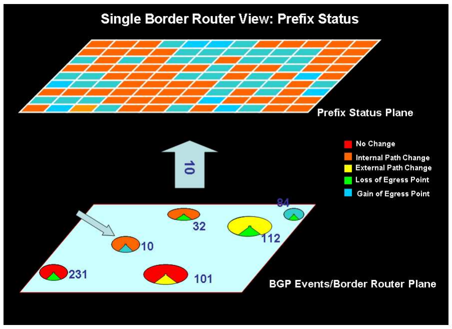


Figure 4: Visualization of Prefixes Status from a Single Border Router.

Although the panel shown in Figure 3 monitors the number of BGP events processed by each border router over time to highlight instantaneous shifts from normal trends, it does provide a deeper understanding of neither what kind of BGP events are processed by each border router nor which prefixes show the most unstable behavior. For this purpose, we provide a new panel shown in Figure 4. The panel displays the prefix status from a single border router perspective. The goal of this layout is to show the: i) total number of BGP events processed by any border router, ii) the repartition of BGP events in types (using the same taxonomy introduced in Section 2.2), and iii) the routing status for any destination prefix. The layout is composed by two planes, 'BGP events/Border Router Plane" (on the bottom) and "Prefix Status Plane" (on the top). We place border routers into the "BGP events/Border Router Plane" and the destination prefixes into the "Prefix Status Plane". Each border router and prefix are unequivocally identified by their own id. For each border router we display i) the total number of BGP events processed (represented by the size of the associated pie-chart, i.e. the larger the more BGP events the border router has processed) and ii) the number of BGP events for each of the five families, accordingly to the taxonomy introduced in Section 2.2. Each family is displayed using a different color. The size of each sector (displayed with a different color) in the pie-chart reflects the number of BGP event observed for that specific family of BGP events. The Operator interested in learning how a specific border

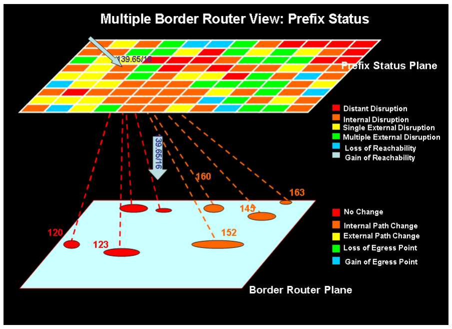


Figure 5: Visualization of Prefixes Status from Multiple Border Routers.

router reaches a specific destination prefix, can click on the router of interest and consult the plane on the top. On this plane we report the status of all destination prefixes for the router selected. For example, a destination prefix appears red if no routing change has been experienced by the router ad the router is able to reach the prefix by using the same route learned previously. On the contrary, a prefix that appears orange implies that a routing change has been generated by the router and a new route has been learned via an iBGP session.

In the following we describe two algorithms to place border routers and destination prefixes onto the two planes.

- To place border routers onto the "BGP events/Border Router Plane" we first divide the plane into a regular grid, such that the size of each grid square is equal to the maximum node size, i.e. border router with the largest number of BGP events. Assume that the total number of grid squares is greater or equal to the total number of nodes to be placed. Next, each node is positioned according to its coordinates x and y. The user can decide between two options: i) place the routers according to their geographical location, i.e. x represents its latitude while y its longitude; ii)place the routers according to the ASes they have established a BGP session with. The two placement methods help the Operator to localize a possible problem both geographically, i.e. routers geographically close each other, and logically, routers establishing BGP sessions with same ASes. Then, we search starting from this grid square to find the nearest empty grid square. Place the node in the first empty grid square. Same steps are followed for each node until all nodes are placed.
- The placement of destination prefixes is slightly more complicated due its larger cardinality. To efficiently display the prefix space we propose the following algorithm. The prefixes are ordered by first considering the mask. A prefix with a smaller mask is considered smaller. For all prefixes with the same mask, a prefix with a smaller address is considered smaller. The ordered prefixes are then arranged from the to the highest row. Within each row, the prefixes are arranged from left to right. The visualization provided by this mapping will show patterns of the distribution of prefixes affected by the routing updates, and can provide some interesting insights.

An Operator that is interested in learning how its AS reaches a specific prefix over time may want to consult the panel presented in Figure 5. This panel shows the prefix status from a multiple border router perspective. Goal of this layout is to show i) the status of all prefixes from a multiple border router perspective and ii) which border router was involved in a routing change, if any. The layout is composed by two planes: "Prefix Status Plane" (on the top) and "Border Router Plane" (on the bottom). The placement of both routers and prefixes is done using the same layout algorithms described for Figure 4. The Prefix Border Router Plane shows the status of all prefixes when the routing information from all border routers is collected and analyzed together. Prefixes mapped to the same color are likely to be associated to the same network problem. By clicking on a specific destination prefix, information of the routing status of all border routers is displayed on the bottom plane. The size of the the pie-chart associated to each border router is an indicator of how many BGP events of that family the router has processed.

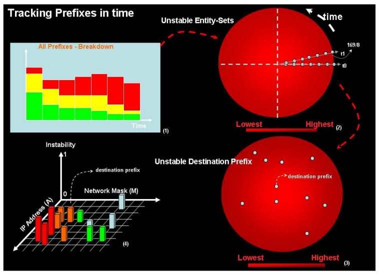


Figure 6: Tracking Noisy Prefixes over time.

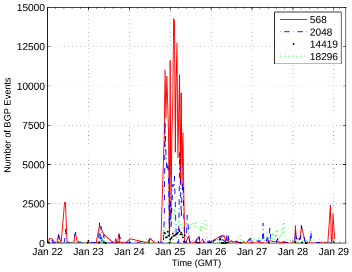
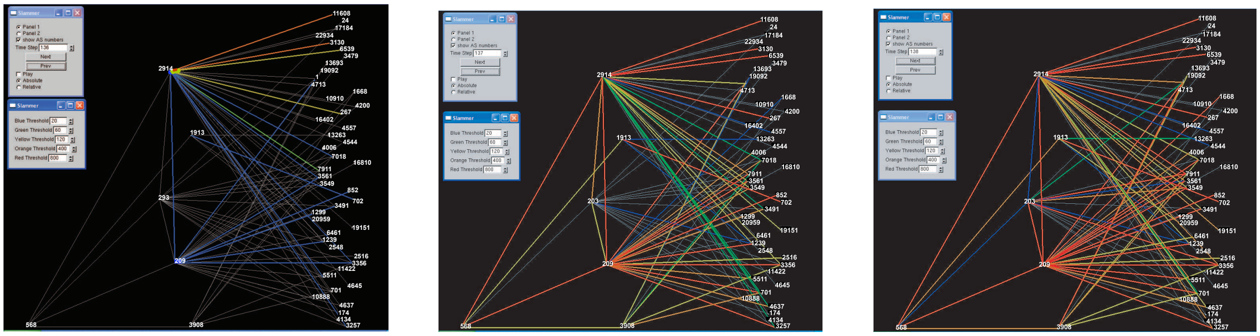


Figure 7: Number of BGP events originated from AS568, AS2048, AS14419 and AS18296.

The last panel is presented in Figure 6. The panel shows the status of prefixes over time, highlighting the prefixes that exhibit the most unstable behavior, i.e. large number of BGP events over time. The layout is composed by three plots. On the left we show the number of prefixes belonging to different *prefixes states* over time. Three prefix states are defined: i)*green* indicates a stable set of prefixes, ii) *red* indicates an highly unstable set of prefixes that an Operator must constantly monitor, while iii)*yellow* represents a transitory state that falls in the between. The mapping of prefixes in states is done accordingly to customizable thresholds whose values can be changed manually by the Operator. The thresholds can be either be i) static, i.e. based on hard-coded numbers, for example, all prefixes for which we see a number of BGP events larger than $T = 1000$ per time-window fall into the red-state, or ii) dynamic, i.e. learned over time, for example, all prefixes for which we see a number of BGP events that is 10 percent lower than the maximum fall into the red-state. An Operator interested in monitoring over time a specific prefix state, can then consult the panel at the top-right. The layout is represented by a (0,1) circle, that is colored according to the prefix state selected. In our case we assumed that the Operator has selected the unstable prefix state, i.e. red state. The time dimension is captured by the angular coordinate θ, while the instability of a prefix is captured by the radial coordinate r. For ex-



Slammer

○ Panel 1
○ Panel 2
☑ show AS numbers
Time Step 136
Next
Prev
□ Play
○ Absolute
○ Relative

Slammer

Blue Threshold 20
Green Threshold 60
Yellow Threshold 120
Orange Threshold 400
Red Threshold 800

(a) Before.

(b) 10 minutes after.

(c) 60 minutes after.

Figure 8: Topological Map with AS568 as the root of the propagation tree. Snapshot of BGP activity during the Slammer worm

ample, if the Operator wants to track the last 6 hours worth of data, then each incremental degree will be equal to 360 degrees divided by 360 minutes, i.e. $\delta\theta = 1$ degree/minute. In order to have a readable and comprehensible visualization of the unstable prefixes, we group the prefixes into block of IP addresses, called *Entity-Set* that each network entity owns, for example, University of California, San Jose owns 169/8. The number N_{SET} of unstable *Entity-Set* is customizable and can be changed by the Operator at any time. For example, if the Operator is interested in the top $N_{SET} = 10$ unstable sets, then the radius r is divided in $N_{SET} + 1$ intervals and each set is placed along the radial coordinate in each interval, with the most unstable set placed at $r = 1$ and the lowest unstable placed at $r = 2/(N_{SET} + 1)$. We point out that the repartition in $N_{SET} + 1$ is done to avoid the overlapping of the lowest unstable sets at different time into the center $r = 0$.

An Operator interested in understanding the distribution of IP Addresses inside a specific set *Entity-Set*, can select the node from this layout and can consult the layout on the bottom-right. Two different layouts are provided to the user. The first layout is a 2D-circular layout represented on the bottom-right of Figure 6. This layout displays all destination prefixes belonging to the *Entity-Set* selected into a (0,1) circle, mapping the most stable prefixes close to the center $r = 0$ and the most unstable ones close to the radius $r = 1$. The placement of the prefixes in the circle is done according to the following algorithm. Let's assume we want to visualize a specific set Entity-Set S, composed by N_S prefixes. First, we classify each prefix according to their (IP Address, Network Mask) and we order the prefix according to their Network Mask, as described for Figure 4. Let's assume that the largest number of BGP events observed in this family is Max_S. Then, we place each prefix $p \in S$ in the (0,1) circle according to the pair (θ_p, r_p). The angular coordinate θ_p unequivocally identifies the prefix p in the set S, while the radial coordinate r_p represents the number of BGP events associated to the prefix p. Thus, each prefix p is placed in the (0,1) circle with polar coordinates $\theta_p = \theta_{p-1} + \delta\theta$ and $r_p = n_p/Max_S$, where θ_{p-1} represents the angular coordinate of the previous prefix placed (starting from $\theta_0 = 0$), $\delta\theta = 360/N_S$ and n_p represents the number of BGP events associated to the prefix p processed. We point out, that the Operator can choose Max_S to be the maximum number of BGP events observed across several sets S in case he wants to have a normalized view of multiple sets at the same time. The second layout is a planar 3D plot, where each destination prefix is identified by its IP Address A and Network Mask M. Each prefix is placed according to (A,M) into the plane, where the x-axis represents the IP address (A), while the y-axis its Network Mask (M). The instability of the prefix is then represented by a bar between 0 and 1, where 0 represents a very stable prefix (green color) while 1 a very unstable prefix (red color). All prefixes falling between these two extreme case are then mapped using colors from green to red. The Operator can choose between the two layouts according to the size of the set he is interested in analyzing.

4. RESULTS

In this section we show *BGP Eye* in action for two wide spread Internet anomalies representative of a typical worm outbreak and prefix hijacking attack.

4.1 Outage during Slammer Worm

In this section we use *BGP Eye* to identify the role played by AS568, corresponding to the *Department of Defense* (DoD), during the spreading of the SQL Slammer worm. We analyzed one week worth of BGP data collected from January 22nd to January 29th 2003 [2]. We found three major results: (i) AS568, after being infected by the Slammer worm, played an active role during the contamination, spreading the epidemic widely and deeply through the entire Internet; (ii) AS568 spread the infection heavily using peering links with four out of five of its peers AS1913, AS209, AS2914 and AS3908 during the first 10 minutes; (iii) AS658 reached more than 800 ASes in the first 60 minutes, 100 of which were successfully infected.

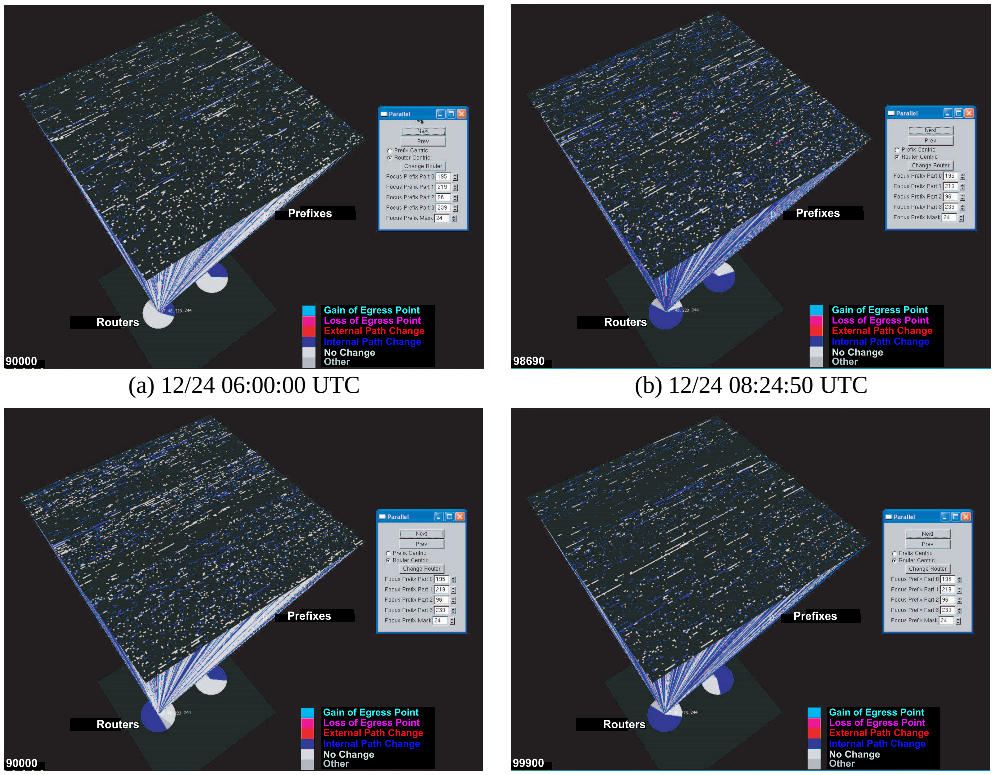
4.1.1 AS568 as an active propagator of the epidemic

BGP Eye analyzed the behavior of the top 4 edge customer ASes that generated the largest number of BGP events during the one week observation period: AS568, AS2048, AS14419 and AS18296 (see Figure 7). During this analysis, *BGP Eye* identified AS568 as the one contributing the most to the spread of the infection across Internet. The AS568 suddenly generated up to 15,000 BGP events on January 25^{th}, 2006 while never generated more than 2,500 BGP events under normal conditions.

4.1.2 AS568's peers infected instantaneously by AS568

Next, *BGP Eye* analyzed the propagation of the BGP anomalies that originated from AS568 to the Internet with the final goal to quantify the growing rate of the infection over time and identify *when* and *which* ASes were successfully infected by the worm. For the analysis of this specific problem, we use only the first panel of the tool, e.g. *Internet-Centric View*. Figure 8(a) provides a topological map of the customer AS568 before the anomaly event, shown in the map as the root of the tree, and its activity with other ASes. *BGP Eye* monitors in real-time the total number of BGP events observed on each AS-AS link and profiles the evolution of this metric over time as explained before. The tool provides four different colors to represent four different hidden BGP instability states: the

9[2] The Slammer worm was released on January 25 2003.



(a) 12/24 06:00:00 UTC

(b) 12/24 08:24:50 UTC

(c) 12/24 08:45:00 UTC

(d) 12/24 09:25:38 UTC

Figure 9: Router centric view of prefix hijacking incident

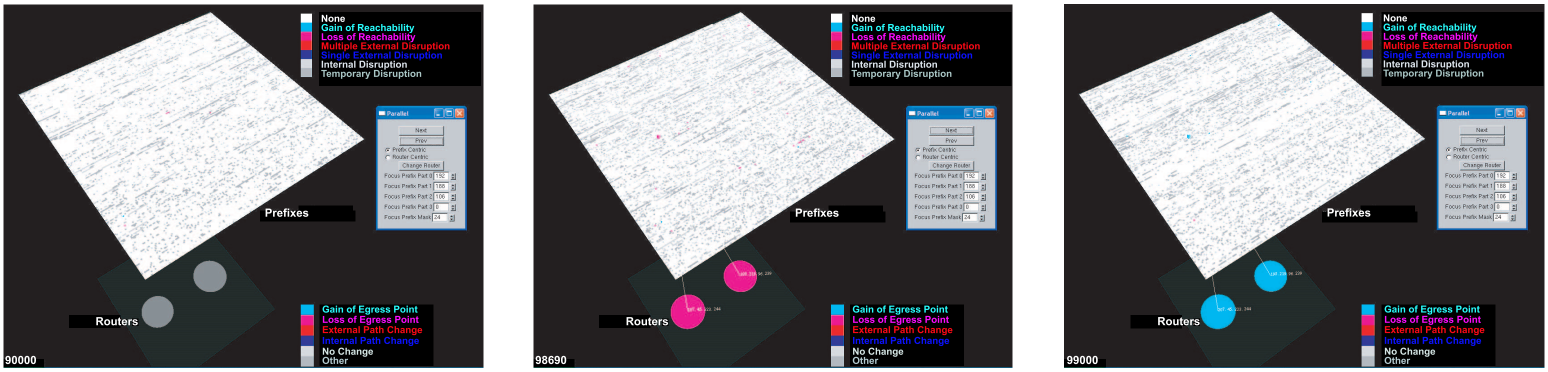
color *green* defines a *very stable* behavior, e.g. instantaneous deviation less than 5%; the color *blue* defines a *stable* behavior, e.g. instantaneous deviation less than 10%; the color *yellow* defines an *unstable* behavior, e.g. instantaneous deviation less than 15%; the color *red* defines a *very unstable* behavior, e.g instantaneous deviation greater than 15%. Figures 8(a), 8(b) and 8(c) show three snapshots of the BGP activity associated to AS568, respectively before the worm, within first 10 minutes and 60 minutes after the Slammer worm outbreak. As you can see, it is very evident and crystal clear how the network behavior suddenly changed and how severe was the damage caused by the worm. AS568 was infected and used its peers as vehicles to spread the anomaly faster, e.g. AS1913, AS209, AS2914 and AS3908. Its peers got infected in the first 10 minutes and spread further along the infection to their peers. After a rigorous analysis we have counted around 100 ASes and 350 AS-AS links infected in the first 60 minutes due to the activity played by AS568 in this process.

4.2 TTNet Prefix Hijacking

In this section, we analyze the inadvertent prefix hijacking by Turkey Net or TTNet (AS9121) on the Christmas eve, December 24th, 2004. This day TTNet started advertising routes towards over 100,000 prefixes that were not owned by it. Some of the peering neighbors of TTNet, both immediate and those that were multiple hops away, updated their routing tables in response to the illicit BGP advertisements. Owing to the path vector based routing protocol followed by BGP, the neighbors of TTNet in turn advertised these new routes to their other neighbors, thereby cascading the effects of these illicit announcements. While this incident did not have a malicious intent on the part of TTNet, it caused a catastrophe of sorts, that revealed an integral component to routing security that is missing in BGP. Some of these prefixes that were illegally advertised by TTNet belonged to web sites such as Amazon, Yahoo, Microsoft, CNN, BBC, etc., and the neighbors of TTNet shifted their outbound traffic away from the legitimate ASes corresponding to these sites and towards TTNet.

Using data collected from RouteViews, *BGP Eye* analyzed the impact of these illicit prefix announcements on various ASes: (1) AS1239, which is direct neighbor of TTNet; (2) AS6453, which is two hops away from TTNet. An important conclusion derived out of this study is that the neighbors of TTNet were affected differently depending on the routing topology between TTNet and the neighbor. In the rest of this section, we present how the *Home-Centric View* described in Section 3.2 allows AS6453 to analyze the sudden onslaught of BGP announcements originating from TTNet. In particular, we analyze the route advertisements that were received at AS6453 over a 3-day period surrounding the incident.

Figure 9 is an instantiation of the "Single Router View" panel presented in Figure 4 and shows the two border routers, 207.45.223.244 and 195.219.96.239 used by AS6453. The figure presents the status of prefixes as viewed by one of the border routers (195.219.96.239) over different timestamps through the prefix hijacking incident. Notice that around the time that the prefix hijacking incident started, 12/24 08:24:50 UTC, several prefixes change to non-greyscale colors. A majority of routes that were obtained through an IGP session, are now obtained through an IGP session although through a different next-hop AS, thereby being classified as Internal Path Change events. Accordingly, the bottom plane which presents the proportion of events experienced per router, shows a greater proportion of Internal Path Change events. This onslaught of illicit BGP advertisements lasts until 12/24 09:25:38 UTC, when most of the prefixes have recovered their original routes and the routing tables at AS6453 have stabilized to the correct routes. The "Multiple Router View" as shown in Figure 10 presents more insights by al-



● Prefix Centric
○ Router Centric
Focus Prefix Part 0 192
Focus Prefix Part 1 188
Focus Prefix Part 2 106
Focus Prefix Part 3 0
Focus Prefix Mask 24

(a) 12/24 06:00:00 UTC (b) 12/24 08:24:50 UTC (c) 12/24 08:30:00 UTC

Figure 10: Prefix centric view of prefix hijacking incident

lowing the Operator to analyze the routing changes as experienced by a prefix after correlating both the routers. When configured to present the changes on a particular prefix, 192.188.106.0/24, this panel reveals an insight not observable in the Single Router View panel. The route to this prefix was earlier obtained through an EGP session and during the incident it is obtained through an IGP session by both the border routers. Since, both the routers lose an EGP route, the incident is interpreted as a Loss in Egress Point event for the prefix, which could indicate a significant shift in traffic between AS6453 and prefix 192.188.106.0/24. This Loss in Egress Point event can be explained by a deeper analysis of the routing topology around AS6453. Several of the illicit prefixes, which were normally reached from AS6453 via an EGP session with AS3561, were now being reached via an IGP session with AS6762, which in turn was propagating announcements from AS9121. This prefix gets restored to its original EGP obtained route after 6 minutes, and hence we observe a Gain in Egress Point event for it.

The capability to succinctly present events and correlate them across routers in the Multiple Router Panel 10 allows us to study the entire sequence of events for these 10 prefixes. The snapshots for these prefixes are not shown here due to lack of space, however, the panels allowed us to analyze them as follows. Prefixes such as 193.151.108.0/24 suffered the following sequence of events: (a) 12/24 08:24:50 UTC: Original EGP route is lost and changed to an IGP route leading to a Loss in Egress Point event for the router and Loss in Reachability for the prefix; (b) 12/24 08:30:00 UTC: The original route is obtained, however not through the EGP session but through an IGP session, thereby the event is classified as Internal Path Change for the router and Internal Disruption for the prefix.

To summarize, through anomaly classification algorithms, *BGP Eye* provides a customer AS the unique ability to monitor the authenticity of the BGP announcements that are being received by it from the rest of the Internet. Moreover, through novel visualization techniques, the customer AS can quickly analyze the impact of the problem as well as it provides pointers towards the possible root causes of the incident.

5. CONCLUSION

In this paper, we presented BGP Eye, a scalable and real-time tool for root-cause analysis of BGP anomalies. BGP Eye takes a hierarchical approach towards analyzing BGP updates by first clustering the updates as received by a router into BGP events that are more representative of a problem. Next, BGP Eye correlates the BGP events across all the routers belonging to an AS to also obtain an insight in to the extent and impact of the anomaly. Moreover, BGP Eye is the first visualization tool that can analyze BGP anomalies from two different perspectives: (a) Internet Centric, which can track the spread of an anomaly through analysis of the AS-AS interactions and; (b) Home Centric, which provides the insight in to how an AS is affected by anomalies that originate from external ASes several hops away. In this regard, we analyzed two separate incidents to establish the scalability and efficacy of BGP Eye towards their root-cause analysis. Specifically, we presented an Internet-Centric view of how AS568 when infected by the Slammer Worm was the origin for several of the routing changes seen on January 25, 2003. Furthermore, we presented how the inadvertent hijacking of 100,000+ prefixes by AS9121 contributed to the sudden onslaught of routing changes seen by AS6453 through our Home-Centric panels.

6. REFERENCES

[1] M. Caesar, L. Subramanian. and R. Katz. Towards localizing root causes of BGP dynamics. In *Tech. Rep. CSD-03-1292, UC Berkeley, November 2003*

[2] D. Chang, R. Govindan and J. Heidemann. The temporal and topological characteristics of BGP path changes. In *Proceedings of IEEE ICNP (November 2003)*.

[3] A. Feldmann, O. Maennel, Z. Mao, A. Berger and B. Maggs. Locating Internet routing instabilities. In *Proceedings of ACM Sigcomm (August 2004)*.

[4] M. Lad, A. Nanavati, D. Massey and L. Zhang. An algorithmic approach to identifying link failures. In *Proceedings of Pacific Rim Dependable Computing (2004)*.

[5] T. Wong, V. Jacobson and C. Alaettinoglu. Making sense of BGP. In *Nanog presentation (February 2004)*.

[6] K.Xu, J.Chandrashekar, Z.L. Zhang. A First Step Toward Understanding Inter-Domain Routing Dynamics. In *ACM Sigcomm 2005 Workshop on Mining Network Data (August 2005)*.

[7] J. Wu, Z.M. Mao, J. Rexford, J. Wang. Finding a Needle in a Haystack: Pinpointing Significant BGP Routing Changes in an IP Network. In *NSDI (2005)*.

[8] S. Agarwal, C. Chuah, S. Bhattacharyya, C. Diot. Impact of BGP Dynamics on Intra-Domain Traffic. In *ACM Sigmetrics (June 2004)*
J. Wu, Z.M. Mao, J. Rexford, J. Wang. Finding a Needle in a Haystack: Pinpointing Significant BGP Routing Changes in an IP Network. In *NSDI (2005)*.

[9] Y. Rekhter, T. Li, S. Hares. A Border Gateway Protocol 4 (BGP-40. Internet Draft draft-ietf-idr-bgp4-26.txt. In *Work In Progress (October 2004)*.

[10] C. Labovitz, R. Malan and F. Jahanian. Internet routing instability. In *Proceedings of ACM Sigcomm (1997)*.

[11] C. Labovitz, R. Malan and F. Jahanian. Origin of Internet routing instability. In *Proceedings of IEEE Infocom (1999)*.

[12] C. Labovitz, A. Ahuja, A. Bose and F. Jahanian. Delayed Internet routing convergence. In *Proceedings of IEEE/ACM Transaction on Networking 9, 3 (June 2001), 293-306*.

[13] O. Maennel and A. Feldmann. Realistic BGP traffic for test labs. In *Proceedings of ACM Sigcomm (2002)*.

[14] J. Rexford, J. Wang, Z. Xiao and Y. Zhang. BGP routing stability of popular destinations. In *Proceedings of Internet Measurement Workshop (2002)*.

[15] L. Colitti, G. Di Battista, I. De Marinis, F. Mariani, M. Pizzonia, and M. Patrignani. Bgplay. http://bgplay.routeviews.org/bgplay/.

[16] M. Lad, D. Massey, and L.Zhang. Linkrank: A graphical tool for capturing bgp routing dynamics. In *Proceedings of the IEEE/IPIF Network Operations and Management Symposium (NOMS)*, April 2004.

[17] C.-C., Lin and H.-C. Yen. A New Force-Directed Graph Drawing Method Based on Edge-Edge Repulsion. In *Proceedings of the Ninth International Conference on Information Vizualisation (IV)*, 2005.

[18] R. Ball, G.A. Fink and C. North. Home-centric visualization of network traffic for security administration. In *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*, 2004.

[19] T. Wong, V. Jacobson, and C. Alaettinoglu. Internet routing anomaly detection and visualization. In *Proceeedings of the 2005 International Conference on Dependable Systems and Networks (DSN'05)*, pages 172–181, 2005.

[20] S. T. Teoh, K.-L. Ma, S. F. Wu, and X. Zhao. Case study: Interactive visualization for internet security. In *Proceedings of the IEEE Visualization Conference 2002*, pages 505–508, 2002.

[21] CERT Advisory CA-2003-04, "SQL Slammer", http://www.cert.org/advisories/CA-2003-04.html

[22] CERT Advisory CA-2001-19, "Code Red" Worm Exploiting Buffer Overflow in IIS Indexing Service DLL", http://www.cert.org/advisories/CA-2001-19.html

[23] CERT Advisory CA-2001-26, "Nimda Worm", http://www.cert.org/advisories/CA-2001-26.html

[24] D. Moore, V. Paxon, S. Savage, C. Shannon, S. Staniford, and N. Weaver "The Spread of the Sapphire/Slammer Worm", *IEEE Security and Privacy, 1(4)* July 2003.

[25] IETF RFC 4272, "BGP Security Vulnerabilities Analysis" http://www.ietf.org/rfc/rfc4272.txt

[26] Nanog, "Nanog Subject List" http://www.nanog.org/subjects.html