Combining Visual and Automated Data Mining for Near-Real-Time Anomaly Detection and Analysis in BGP

Soon Tee Teoh, Ke Zhang, Shih-Ming Tseng, Kwan-Liu Ma, S. Felix Wu
Department of Computer Science
University of California, Davis
Davis, CA 95616
{teoh,zhangk1, smtseng, ma, wu}@cs.ucdavis.edu

ABSTRACT

The security of Internet routing is a major concern because attacks and errors can result in data packets not reaching their intended destination and/or falling into the wrong hands. A key step in improving routing security is to analyze and understand it. In the past, we and other researchers have presented various visual-based, statistical-based, and signature-based methods of analyzing Internet routing data. In this paper, we describe an integration of visual and automated data mining methods for discovering and investigating anomalies in Internet routing. We show how these different components are combined in such a way as to complement each other, creating a very effective and useful analysis tool. In addition to performing analysis on archived data, our system is able to collect, process and visualize data in near-real-time.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Graphical User Interfaces (GUI)*; I.3.6 [**Computer Graphics**]: Methodology and Techniques—*Interaction Techniques*; D.4.6 [**Opertating Systems**]: Security and Protection

General Terms

Security, Human Factors

Keywords

information visualization, network visualization, internet routing

1. INTRODUCTION

Routing is an important part of the Internet infrastructure. It is the mechanism by which a packet gets from its source to its destination correctly. The distributed and autonomous nature of the current Internet routing protocol, the Border Gateway Protocol (BGP), makes routing susceptible to attacks and errors. An important step in improving the security of Internet routing is to analyze its run-time behavior under BGP, to understand its operational dynamics, as well as to discover potential security problems. In the past, we and other researchers have presented various visual-based, statistical-based, and signature-based methods of analyzing Internet routing data. While these methods have produced significant results, we believe that an adequate coupling of visual data mining, statistical data mining and signature detection can perform even better.

Our methodology is based on the naturally complementary relationship between visualization and automated computation. On one hand, computers are good at processing a large amount of data, and optimizing the search for well-defined patterns. On the other hand, automated methods often have problems finding the right threshold values. This is where interactive visualization can allow the user to examine the data, and use domain knowledge and judgement to adjust and fine-tune the statistical parameters in order to perform more accurate anomaly detection.

The human visual perception system is able to recognize features in visual displays and recall related images to identify anomalies [1]. We use this ability to complement and verify the statistically detected anomalies.

In this paper, we describe the system we have built to combine visual and automated data mining for near-real-time anomaly detection and analysis in BGP. We show examples of the way visual and statistical methods work together to detect and characterize anomalies more efficiently and accurately than the individual components can do on their own. Furthermore, visualization also helps in conducting deeper analysis into the root cause of the problem once the anomalies have been statistically determined.

Previously, we have been able to visualize and analyze a large amount of archived data [2, 3]. This is important because we want to discover and understand long-term routing behavior. However, near-real-time data mining is also extremely important and useful because errors and attacks need to be detected quickly, so that corrective action can be taken promptly to lessen the damage. We therefore describe the new capabilities we implemented in our system to perform near-real-time data collection, processing and analysis.

VizSEC/DMSEC'04, October 29, 2004, Washington, DC, USA.

2. THE BORDER GATEWAY PROTOCOL

The Internet can be considered as a network of Autonomous Systems (ASes). These *autonomous systems* (ASes) entirely manage any traffic within themselves. To communicate between ASes, routers on the edge of an AS use the *border gateway protocol* (BGP) [4].

In BGP, each AS is assigned a unique identifier and owns a set of IP prefixes. An IP prefix consists of an IP address and a mask. This mask identifies which subset of IP addresses corresponds to hosts in the AS. For example, the IP prefix 128.120.0.0/16 means that every machine in the AS shares the same initial 16 bits 128.120. Using BGP, routers communicate network reachability information in order to properly transmit packets. These *BGP routes* consist of a list of the ASes through which data can be routed to reach the destination. The BGP route "128.120.0.0/16: (7, 23, 92)," for example, means that packets for the IP prefix 128.120.0.0/16 need to sequentially pass through AS 7 and AS 23 before reaching the AS representing their destination (AS 92).

Because of the distributed nature of BGP and the lack of verification of the validity of the announcements, Internet routing is susceptible to errors and attacks. Some router can easily be misconfigured to announce a prefix its AS does not own, and also an attacker can also insert invalid announcements to cause disruption in routing.

3. RELATED WORK

3.1 BGP Routing

BGP routing behavior has received a lot of examination in the research literature. Labovitz et al. [5] showed that unstable and pathological routing behaviors dominated the Internet around 1996. Later, they presented potential explanations for these anomalies [6]. Other BGP routing problems, such as slow convergence [7], and persistent MED oscillation [8, 9], have also been well examined. Rexford et al. defined the metric "BGP event" to measure BGP routing instability and concluded that routes to the popular destinations were generally stable [10]. In order to generate realistic BGP traffic for testing, Maennel et al. extensively studied BGP traffic and characterized the statistical properties of BGP traffic [11].

3.2 Visualization

Previous use of visualization for computer security applications includes Erbacher et al.'s [12] work using glyphs to visual intrusion detection data, Yurcik et al.'s [13] tool for visualizing network traffic, Girardin's [14] packet-based visualization, Muniandy's [15] system representing users, machines and other objects as nodes in a network topology graph, and Tudumi [16], a visualization system designed to monitor and audit computer logs to help detect anomalous user activities.

Our paper describes the tight integration between visualization and automated methods for anomaly detection. An example of a past integration of visualization and automated methods for a different data mining task is Ankerst et al.'s work [17] on classification.

We have also done previous work in visualizing anomalous Origin AS changes in BGP [18]. Of particular relevance to this paper is our routing dynamics visualization system [19], as it is the foundation on which our current system is built. This interactive visualization system allows users to browse through BGP update messages with a slider bar. The update messages are color-coded and drawn according to their timestamp, allowing users to see various patterns. Modules are also provided for visualizing anomalous events matching user-defined signatures, and for visualizing the network topology. Figure 1 shows a screenshot of the visual display.

3.3 Anomaly Detection in BGP

Anomaly detection is widely used in computer security. The basic work-flow of anomaly detection is the following. First, a detection system is fed into the normal model of a subject and is trained to learn the characteristics of the normal behaviors of the subject. Second, the detector examines the testing behaviors and flags an anomaly if the testing behaviors significantly differ from the normal behaviors.

Kruegel et al. [20] derived a model of the autonomous system connectivity from BGP routing table and whois database. Route advertisements are checked to make sure that they are consistent with the network topology; otherwise, they are flagged as anomalous route updates.

In our previous work [21], we applied statistical-based anomaly detection and signature-based detection to examine BGP updates. We compared these two methods and highlight the difficulties in BGP anomaly detection. As our current system integrates these methods, they are described in more detail in Section 4.

4. SYSTEM MODULES

There are four components in our system(Figure 2): (A) Near-real-time Data Collection, (B) Anomaly Detection Engine, (C) Data Server and Communication, and (D) Interactive Visualization Client.

In the following sections, we will describe each component in detail.

4.1 Real-time Data Collection

The main function of *Real-time Data Collection* module is to record raw BGP data for further analysis. There are three different ways to collect these raw BGP data. (i) BGP table dump, (ii) BGP updates, and (iii) Routing table directly from a router. We are able to download (i) and (ii) which are both in MRT format from RouteViews [3] , RIPE [2] and UC Davis. We are able to get (iii) in ASCII format by executing the "show ip bgp" command on a router.

In order to get the real-time data, we need a higher recording rate. Typically, those data provided by RouteViews, RIPE and UC Davis are recorded at a frequency of 15 minutes which is not generally considered to be real-time. Hence, we collect our data in a different way.

First, we download raw data of (i) and (ii) from RouteViews, RIPE and UC Davis. These downloaded data contain the update messages from several different peers. We need to pre-process the data to keep only the updates from the peers that we are interested in examining. Then, we use a BGP replayer which is originally written by Sharad Agarwal to regenerate all of the BGP traffic in DETERLAB [22]. DETERLAB is a testbed which provides a simulation environment with up to 72 nodes. Not only we are able to replay the real BGP traffic into DETERLAB, but also we are able to inject some malicious traffic and see the impact immediately.

Finally, we set up some zebra routers in DETERLAB to

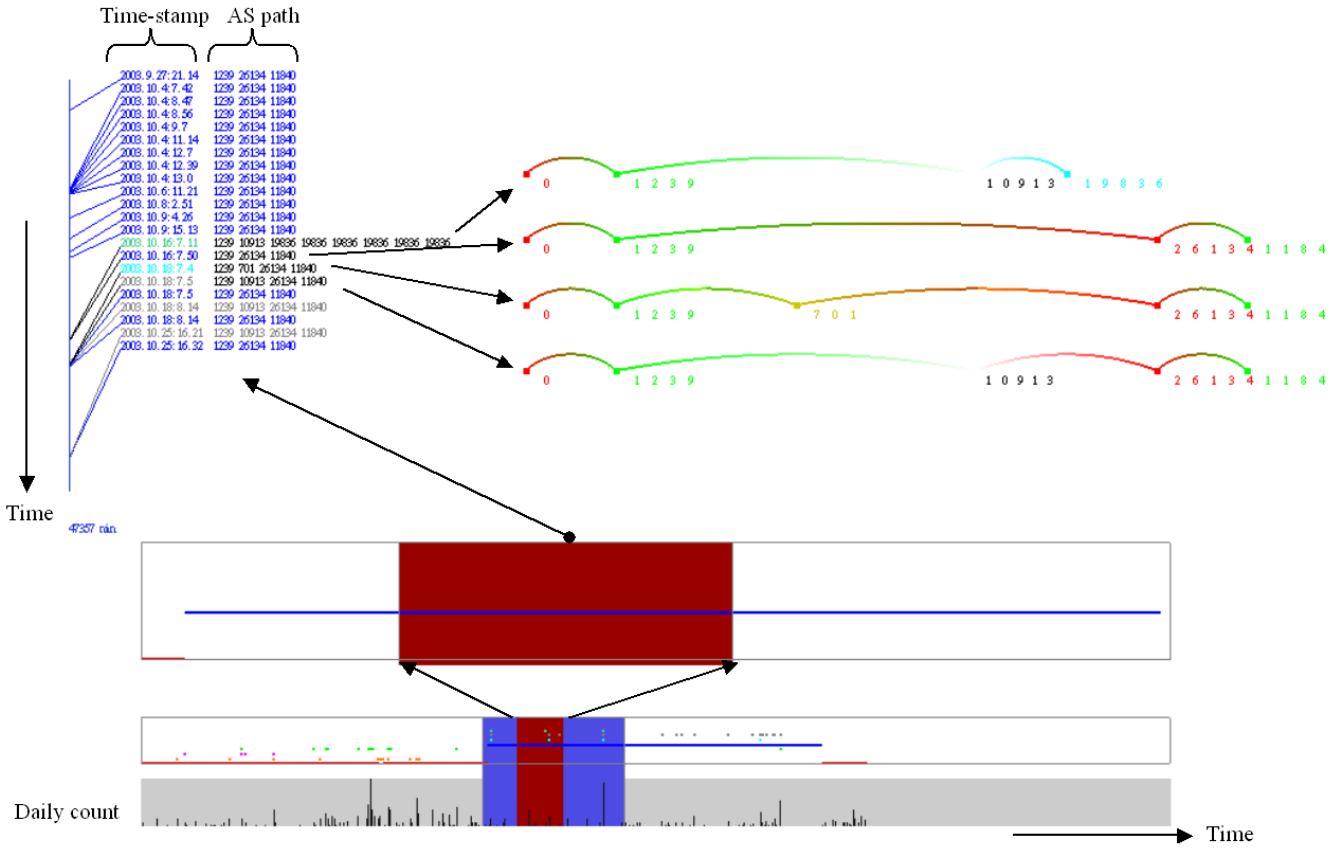

Figure 1: Our previous tool for visualizing BGP update messages. Our current tool extends it to incorporate visualization of statistical measures and signatures, and to enable near-real-time monitoring.

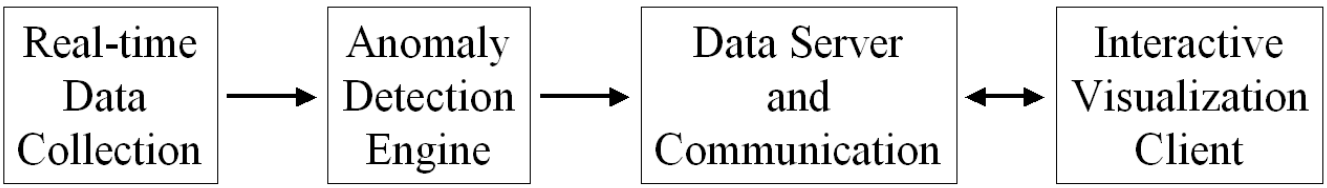

Figure 2: System Overview

serve as observation points. These observation point routers are in charge of dumping and recording raw BGP data every T_i seconds. We are able to set those routers to have a higher recording rate to increase the accuracy.

Assuming there is one event E happened at time T and the recording interval is T_i seconds. We will have E recorded in the next dump file at time $T + x$ (where $0 <= x <= T_i$). Hence, in the worst case, event E can be recorded at time $T + T_i$.

There is one concern that both *Real-time Data Collection* and *Anomaly Detection Engine* are accessing the same record which cause a race condition. One possible solution is that *Real-time Data Collection* pipes the data directly to *Anomaly Detection Engine*. Another possible solution is that *Anomaly Detection Engine* processes the second-to-the-latest dump file. In the worst case, event E is available to be analyzed by *Anomaly Detection Engine* at time $T + T_i + T_i$. In our experiment, T_i equals to 60 seconds. Thus, in the worst case, the event E can be analyzed within 120 seconds. This process delay is feasible for near-real-time traffic analysis.

4.2 Anomaly Detection Engine

4.2.1 Prefix Selection

Currently, in the Internet, there are over 130,000 prefixes in the BGP routing table. A user at the visualization client can select any single prefix that he/she is interested. The anomaly detection engine only examines the BGP updates for the pre-selected prefixes.

4.2.2 Signature-based detection

A route announced by a BGP router is generally the best route at that moment. Comparing the consecutively announced routes, we can infer the route changes in that router's BGP routing table. There are four types of route changes: UP,DOWN,FLAT,WITHDRAWAL [21].

We define BGP update burst as a sequence of updates within a short time window. Formally, BGP update burst is K consecutive updates for the same prefix that space close together. The time interval between update messages is less than T and the average update rate $> \alpha$. The parameters, K, T and α are set by the visualization client.

According to the route changes in BGP update burst, we designed six signatures. For details of the signatures, please refer to our previous paper [21].

4.2.3 Statistics-based detection

We apply a statistics-based anomaly detection method, NIDES/STAT [23]. The NIDES/STAT algorithm monitors a subject's behavior on a computer system,and raises an alarm when the subject's current (short-term) behavior deviates significantly from its expected behavior, which is described by its long-term profile. A subject's behavior is described by a set of detection measures. For each individual measure, there is a corresponding Q statistic. The historical profile records the frequency distribution of Q. For each measure, the corresponding S, derived from Q, is indicative of the degree of abnormality of the behaviors with respect to that measure. T^2 summarizes the abnormality of many measures, reflecting the degree to which recent behavior is similar to the historical profile. Large values indicate abnormal behaviors.

Table 1: Five Measures

Intensity Measure	BGP Updates Message Arrival Frequency (M1) Number of AS paths in a period (M2)
Categorical Measure	BGP Updates Type (M3) AS path Occurrence Frequency (M4)
Counting Measure	AS path Difference (M5)

Like NIDES/STAT, we define 3 types of measures listed in Table 1. M1 measures the inter-arrival time of BGP update messages sent by a router for a single prefix. We devise this measure to detect BGP update burst. During BGP route convergence process, BGP router may receive a number of potential AS paths that are seldom seen in the past. M2 is devised to monitor the variation of the number of AS paths. BGP updates can be roughly classified into 7 types [21]. Different types indicate different BGP events. We apply M3 to monitor the occurrence frequency of these types of updates. According to the observation that only a small number of different AS paths are announced, we define a categorical measure M4 to capture the frequency distribution of AS paths occurrence. Last but not the least, we employ the measure M5 to compare the current AS path with historical dominant AS path.

For each measure, NIDES/STAT algorithm defines another variable S which is "normalizing" transformation of Q statistics so that the degree of abnormality for different measures can be added on a comparable basis. S has a half-normal distribution. The larger S, the more abnormal. Since each individual measure has a S value for each BGP update message, the anomaly detector can generate a single score value T^2 by the following formula:

$$T^2 = (S_1^2 + S_2^2 + \cdots + S_n^2)/n$$

Users can define the T^2 threshold on the Visualization client. Any update with T^2 above the threshold is considered to be an anomaly. The T^2 value, and the Q and S values for all five measures M1 through M5 for each update message are appended to the message and stored at the Data Server, ready for transmission to the Visualization client.

4.3 Data Server and Communication

The main function of the Data Server is to transfer data which has been processed by the Anomaly Detection Engine to the Interactive Visualization Client. We use the client-server concept to implement the module. A non-blocking server takes request from the Interactive Visualization Client. A request message contains a time frame , a prefix and a AS number. The server sends back the desired data back to the client. We use TCP/IP communication protocol in this client-server module to achieve the minimal level of transmission reliability. In addition, IP/SEC can be applied to increase the level of security.

4.4 Interactive Visualization Client

The interactive visualization client is based on our previous system described in Section 3, with four new features. First, it is able to request and receive data from the data server. This allows visualization of routing data in near-real-time, that is, BGP update messages that just happened can be collected and transmitted to the visualization within a few minutes. The communication with the server also allows the user to interactively select different IP prefixes, peers and time periods to display.

Second, the visualization client can display the statistical measures calculated by the Anomaly Detection Engine. These statistical measures are used for anomaly detection. Visualizing the measures together with the update messages enables the user to compare visual anomaly with statistical anomaly. An example is shown in Figure 3.

Third, in our current system, *anomalous events* are generated from statistical anomaly detection, instead of signature detection, to give a more consistent definition of an anomalous event.

The way an anomaly event is defined, classified and visualized is as follows. First, a T^2 value has been generated for each BGP update message in the Anomaly Detection Engine. Next, the user interactively sets the threshold T^2 value. Any BGP update message that exceeds this threshold T^2 value is considered an anomaly. Next, three parameters, K, T and α are used to group consecutive BGP update messages together to form an anomaly event. To be considered one single anomaly event, the sequence of messages must contain at least K messages, with time interval between each message less than T, and have an average update rate α. The user is allowed to set different values for the parameters, K, T and α for signature detection.

Next, signature detection is used to assign a class to each anomaly event, according to the signature(s) the event matches. As before, EventShrubs are used to visualize anomalous events. Each EventShrub's size is determined by the T^2 value of the event it represents, and colored by the signatures matching the event. If the event matches more than one signature, the circle will have a segment to represent each signature. The base of the EventShrub covers the time-period of the instability event. Figure 4 shows an example of the user interactively changing the statistcal measures defining an anomaly event, and visualizing the EventShrubs resulting from this change.

The fourth addition to our system is the ability to visualize the EventShrubs together with the update messages. This makes it tremendously more convenient to visually correlate statistical and signature anomalies with the actual update messages. Figure 5 shows an example.

5. DISCUSSION

The figures that have been presented show that our visualization system is able to display the results of statistical data mining and signature detection. This itself is useful as a presentation. However, we would like to consider the following question: "Does this system enable visual and statistical data mining to help each other achieve better results?" We use three examples to show that the above question can be answered in the affirmative.

5.1 Verification of statistical values

The joint visualization of update messages with statistical values can help the user verify both the validity of the statistical measures, and the impression created by the visualization. This is illustrated in Figure 3. In this figure, there is a sequence of update messages that has unusually large T^2 values. Visual inspection of the five statistical components shows that all five measures are abnormally large. With the adjacent visualization of the update messages, the user can see that indeed there is a cluster of different BGP route paths corresponding to this statistical anomaly, thus verifying the results of statistical analysis.

Conversely, when the user only sees the visualization of the BGP update messages, the sequence of five different BGP route paths immediately draws attention. However, the user unsure whether this is truly an anomaly as compared with historical data. This is because the display only shows a short period of recent data. With the adjacent visualization of statistical measures, the user gains confidence that this is indeed an anomaly. The user can then focus more investigation to find the root cause for this anomaly.

5.2 Adjustment of clustering parameters

The visualization can be used to fine-tune the parameters used for statistical anomaly detection and for clustering consecutive update messages into a single event. In statistical anomaly detection, a T^2 value is chosen as the threshold, such that whenever a BGP update message exceeds this threshold, it is considered to be an anomaly. As is typical of statistical data mining methods, there is no good methodology of choosing the threshold, so often it is arbitrarily chosen. Similarly, the parameter values for using signature detection to cluster update messages into a single event are often arbitrarily chosen.

Figure 6 shows how visualization can help select a better set of threshold and parameter values. In this figure, a cluster of update messages highlighted in grey exhibits high T^2 values. However, with the default value of K set at 3, signature detection produced two separate events. Changing the value of K to 5 resulted in one event covering the entire anomalous sequence. This shows that K=5 is the more appropriate value for characterizing this data.

5.3 Investigating root causes

The purpose of anomaly detection in BGP data is not simply to find cases of anomaly, but to investigate the anomaly once it is found. In fact, anomaly detection is only a means to an end. The final goal is to discover the cause of abnormal events in BGP updates so that a better understanding of BGP dynamics can be gained. Understanding BGP dynamics can help discover problems of two kinds: first, we would like to identify malicious attackers or misbehaving routers; second, we would like to find out if there are any vulnerabilities inherent in the current Internet routing architecture.

Figure 7 shows an example of how visualization helps in investigating the root cause of an anomaly event. This example illustrates the collaboration between visualization and automated anomaly detection. Automated anomaly detection is responsible for quickly identifying an anomaly, since it can do so much faster than a human browsing through the data, even with visual aids. Then, the user looks at the visual display of the data, and uses his/her cognitive skills and domain knowledge to draw inferences. This is the advantage offered by visualization.

In this example, statistical anomaly detection first flags the highlighted sequence as anomaly. Visual browsing of the color-coded update messages indicate that there is no cause for alarm. These messages occur whenever the network is unstable, for example, the AS-paths (1237, 701, 26134, 11840) and (1239, 10913, 26134, 11840) occur again in Figure 6, and there is nothing unusual about them. The AS-path (1239, 10913, 19836, 19836, 19836, 19836, 19836, 19836) is interesting because it has a different Origin AS from the usual 11840, and also 19836 is repeated six times,

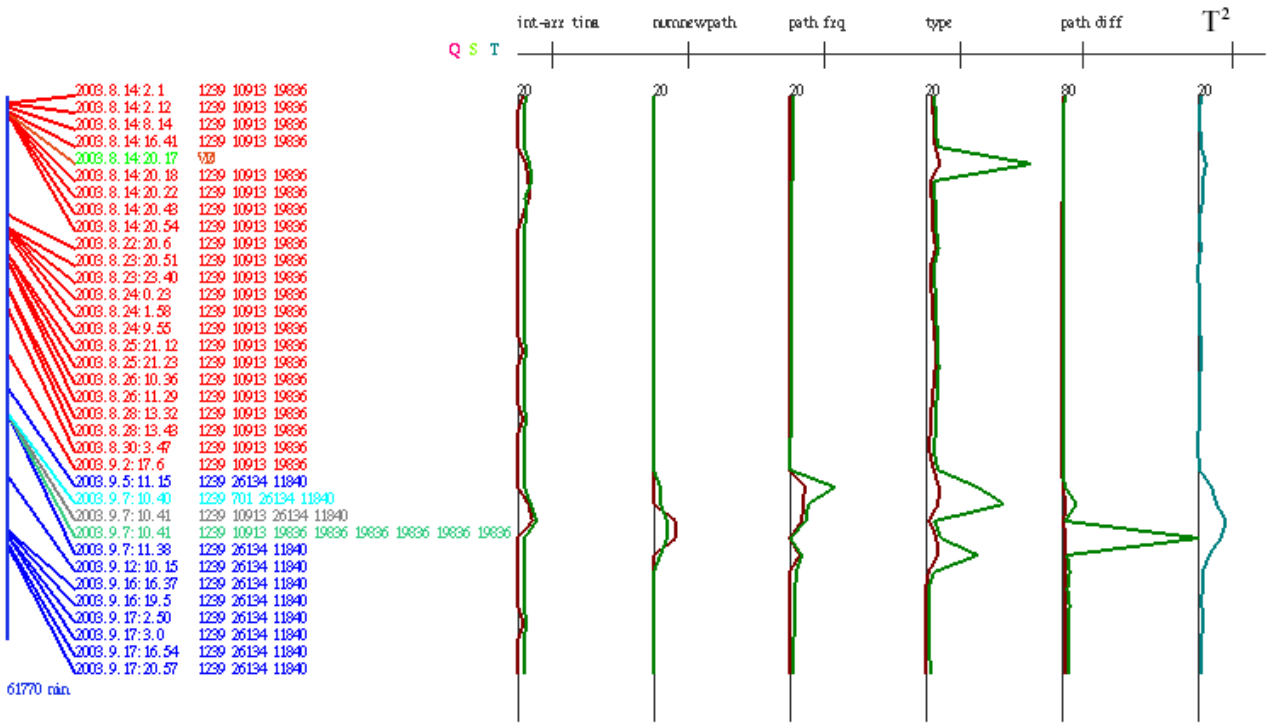

Figure 3: Visualizing statistical values with update messages. The aggregate T^2 value gives an indication of the anomaly of each update message. This joint visualization allows verification between visual and statistical anomaly detection.

indicating that this is not a preferred route. Information from external sources tell us that the Root A IP prefix (which is what we are visualizing) is multi-homed at 11840 and 19836. From these information, we infer that the AS-path (1239, 10913, 19836, 19836, 19836, 19836, 19836, 19836) is only used as a last resort, when the connection from the preferred AS 11840 is down. The presence of the other two paths is most likely due to slow convergence. The coupling of statistical anomaly detection and visualization makes this analysis very simple and fast.

6. CONCLUSION

We have presented a near-real-time system to perform anomaly detection and analysis on BGP update messages. This system tightly integrates visual, statistical, and signature detection components.

We have explained how data is collected from BGP routers. The data is filtered and processed to obtain statistical measures of anomaly for each BGP update message. The measures are appended to the update messages, and kept at a server. At the visualization client, the user can select the prefix and time period to display. Upon receiving the request, the server transmits the data over the network to the client. With this communication mechanism, the client can receive a BGP update message within minutes from their its first appearance at the BGP router. This allows near-real-time monitoring and visualization of Internet routing dynamics from anywhere in the world, as long as a network connection to the data server can be established. The client can request data from any time period from the server, so that both near-real-time data and data collected in the past can be visualized together. The visualization client allows the user to browse through the data to see clusters of BGP update messages. With each update message, the user can also see its associated statistical anomaly measures. This allows the user to correlate the statistical anomalies with the visualization. This helps the user to verify the correctness of the results.

Next, the user can choose to see a visualization of the anomaly events, represented as EventShrubs. With our current system, the user can interactively select the threshold of what is considered an anomaly. As choosing the right threshold and values is a very difficult task, visualization helps the user in determining the right threshold according to his/her preferences.

Finally, our new system also allows the user to visualize these EventShrubs together with the update messages. As we have shown with an example, this visualization enables the user to fine-tune the parameter values further so that a cluster of anomalies can be properly grouped together.

With examples, we have shown how visual and statistical data mining methods can work together to achieve effective anomaly detection and analysis on BGP data. This works because of the combination of human judgement with algorithmic precision.

The purpose of performing anomaly detection is to focus analysis on a manageable subset of all BGP update messages. The identification of such anomalous sequences of BGP update messages is important so that investigation can be conducted to discover the root causes of such events, which may lead to the discovery of new attacks or errors

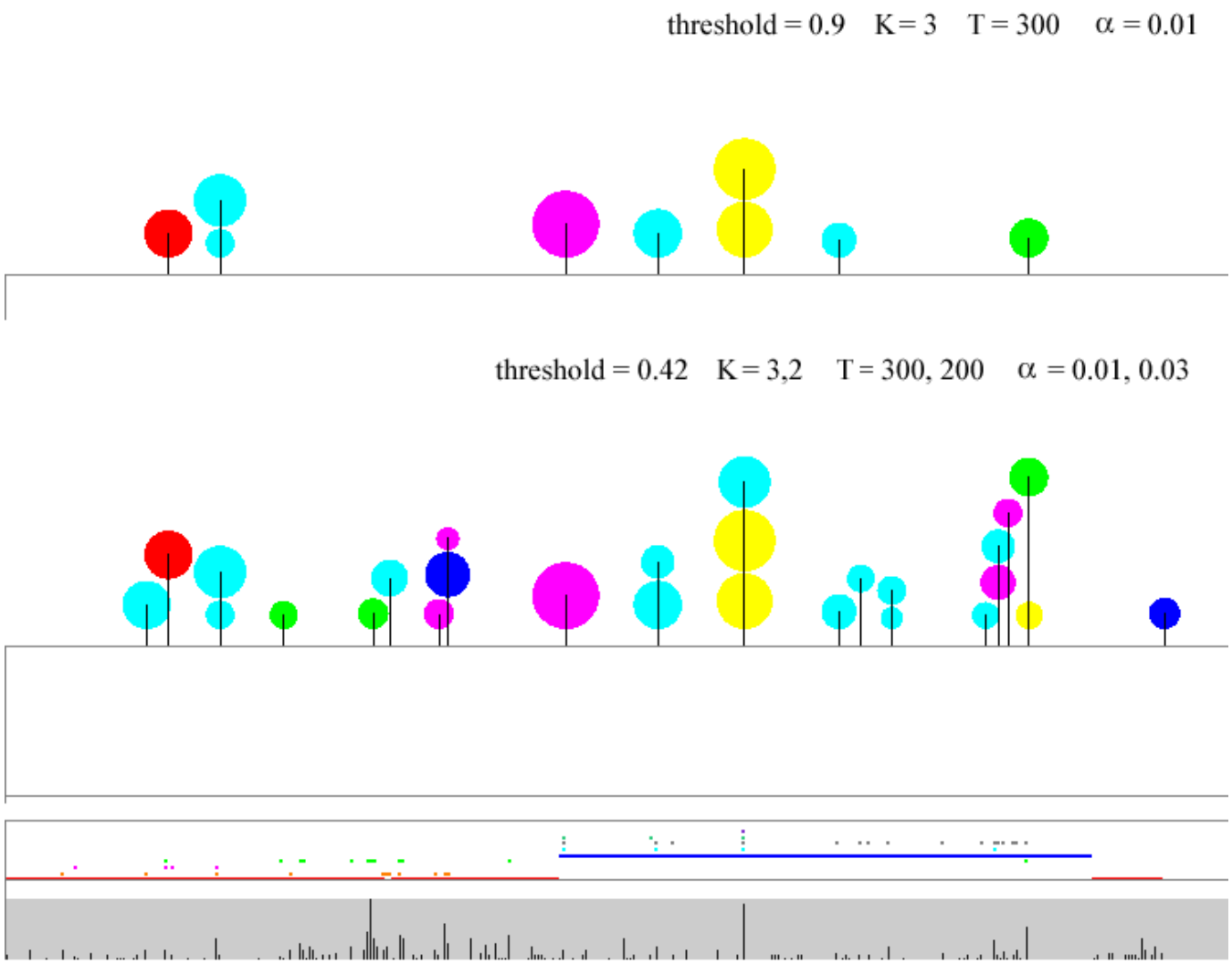

Figure 4: The effect of interactively changing the statistical parameters defining an anomaly event. These parameters are set according to the nature of the dataset and user preferences. Visualization simplifies the otherwise challenging task of selecting an appropriate set of parameter values.

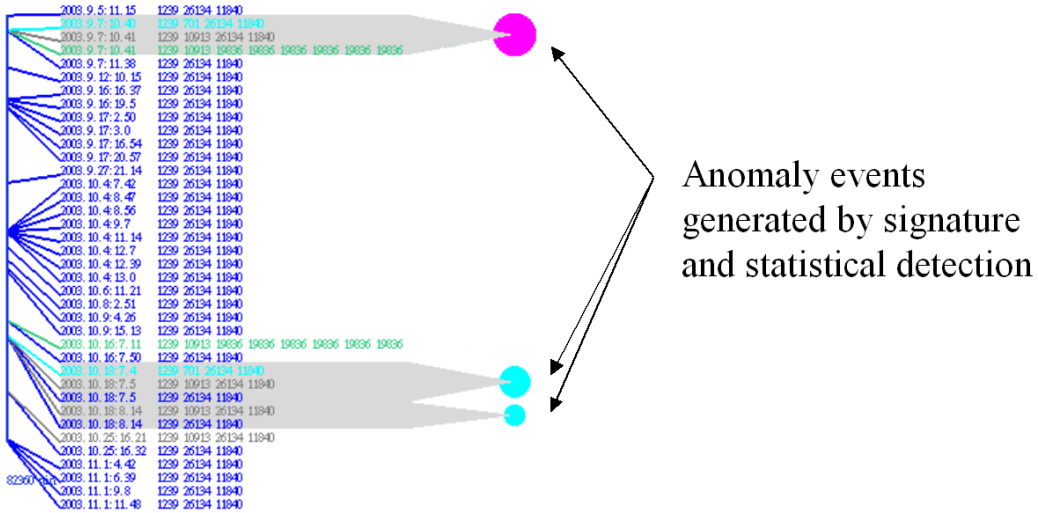

Figure 5: EventShrub anomaly event visualization together with update messages. This anomaly event is generated by the statistical measures shown in Figure 3. With this visualization, the user can quickly focus attention on a selected sequence of BGP update messages highlighted by the EventShrub anomaly event.

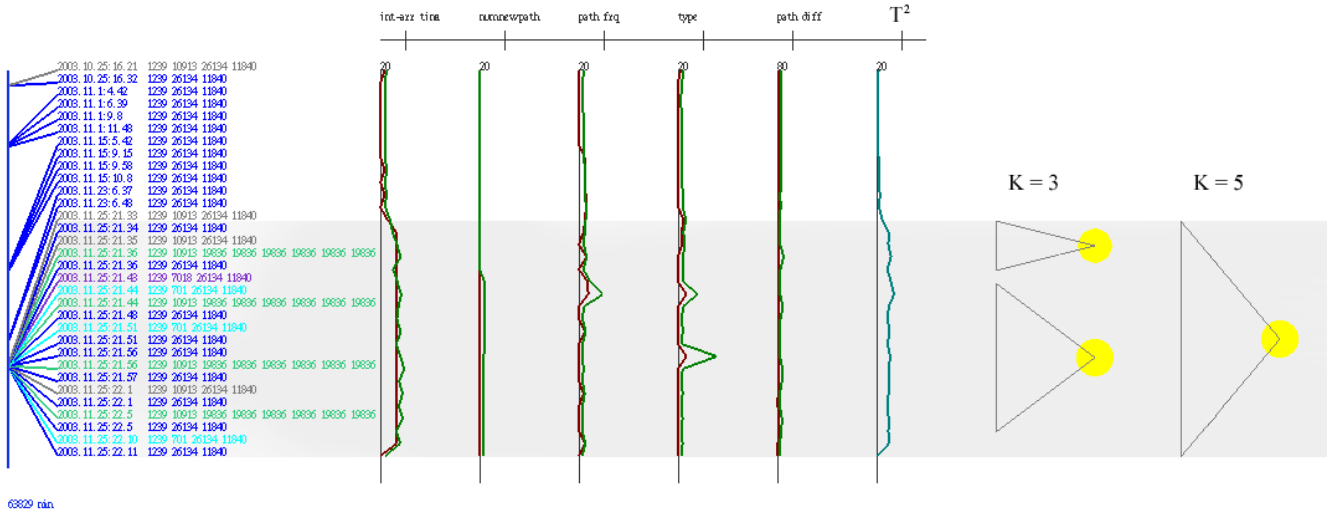

Figure 6: A cluster of anomalous update messages is highlighted in grey. With the default K value set at 3, signature detection classifies this as two separate events. Visualization enables the user to change the the K parameter value from 3 to 5 so that the entire anomalous event can be more accurately grouped as one single event.

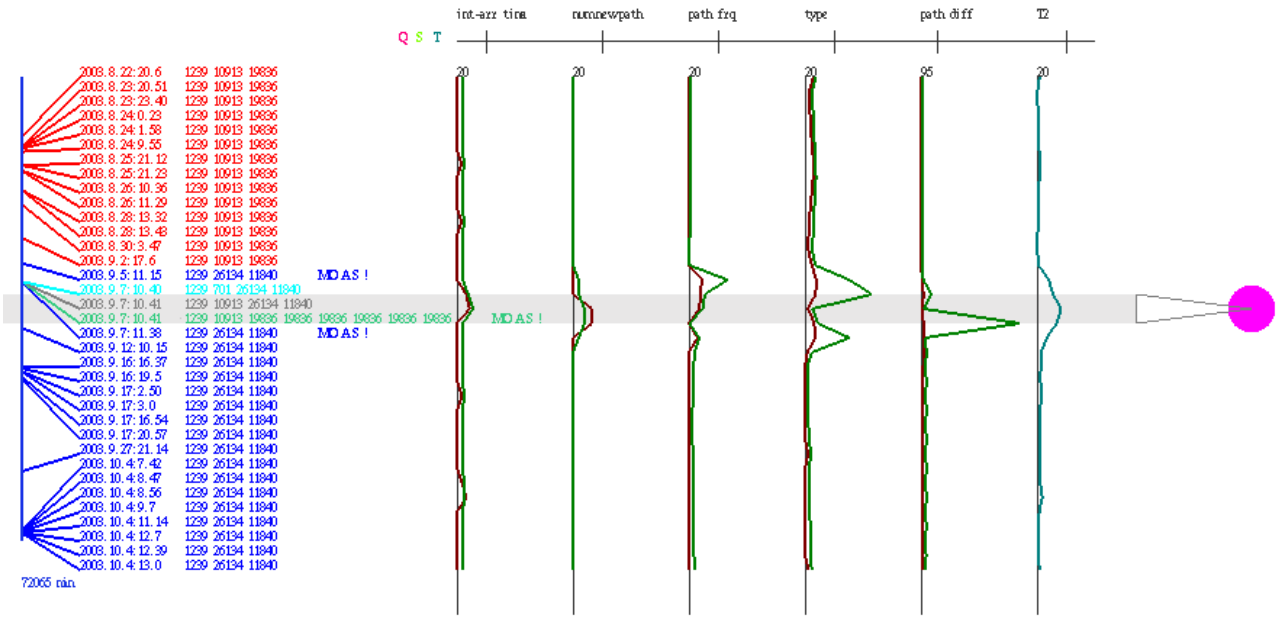

Figure 7: Anomaly detection flags the highlighted sequence as anomaly. Visual browsing of the color-coded update messages indicate that there is no cause for alarm. These messages occur whenever the network is unstable (see Figure 6).

in Internet routing. For our future work, we plan to use our visualization tool to help in this analysis, and also to gather information from other sources to help identify the root causes of the anomaly events. We hope that the use of our system will lead to a more stable and secure Internet routing system.

7. REFERENCES

[1] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings CHI'94: Human Factors in Computing Systems*, pages 313–317, 1994.

[2] The RIPE Routing Information Services. http://www.ris.ripe.net.

[3] The Route Views Project. http://www.antc.uoregon.edu/route-views/.

[4] Y. Rekhter and T. Li. A border gateway protocol 4 (BGP-4). Technical Report RFC 1771, The Internet Engineering Task Force, 1995.

[5] C. Labovitz, G. Malan, and F. Jahanian. Internet Routing Instability. In *Proceedings of ACM SIGCOMM*, September 1997.

[6] C. Labovitz, F. Jahanian, and G.R.Manlan. Origin of Internet Routing Stability. In *Proceedings of the IEEE INFOCOM*, June 1999.

[7] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet Routing Convergence. In *Proceedings of ACM SIGCOMM*, August 2000.

[8] A. Basu, C. Ong, A. Rasala, F. Shepherd, and G. Wilfong. Route Oscillations in I-BGP with Route Reflection. In *Proceedings of ACM SIGCOMM*, August 2002.

[9] T. Griffin and G. Wilfong. Analysis of the MED Oscillation Problem in BGP. In *Proceedings of ICNP*, November 2002.

[10] J. Rexford, J. Wang, Z. Xiao, and Y. Zhang. BGP Routing Stability of Popular Destinations. In *Proceeds of Internet Measurement Workshop*, November 2002.

[11] O. Maennel and A. Feldmann. Realistic BGP Traffic for Test Labs. In *Proceedings of the ACM SIGCOMM '02*, August 2002.

[12] R. F. Erbacher, K. L. Walker, and D. A. Fincke. Intrusion and misuse detection in large-scale systems. *IEEE Computer Graphics and Applications*, 22(1):38–48, January/February 2002.

[13] W. Yurcik, K. Lakkaraju, J. Barlow, and J. Rosendale. A prototype tool for visual data mining of network traffic for intrusion detection. In *Proceedings of the ICDM Workshop on Data Mining for Computer Security (DMSEC'03)*, 2003.

[14] L. Girardin. An eye on network intruder-administrator shootouts. In *Proceedings of the Workshop on Intrusion Detection and Network Monitoring (ID'99)*, Berkeley, CA, USA, 1999. USENIX Assoc.

[15] K. Muniandy. Case study: Visualizing time related events for intrusion detection. In *Proceedings of the IEEE Symposium on Information Visualization*, 2000.

[16] T. Takada and H. Koike. Tudumi: Information visualization system for monitoring and auditing computer logs. In *Proceedings of the 6th International Conference on Information Visualization*, 2002.

[17] M. Ankerst, M. Ester, and H.-P. Kriegel. Towards an effective cooperation of the user and the computer for classification. In *Proceedings of the 6th International*

Conference on Knowledge Discovery and Data Mining (KDD '00), 2000.

[18] 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.

[19] S. T. Teoh, K.-L. Ma, and S. F. Wu. Visual exploration process for the analysis of internet routing data. In *Proceedings of the IEEE Conference on Visualization 2003*, pages 523–530, 2003.

[20] C. Kruegel, D.Mutz, W. Robertson, and F. Valeur. Topology-based Detection of Anomalous BGP Messages. In *RAID*, 2003.

[21] K. Zhang, A. Yen, X. Zhao, D. Massey, S.F. Wu, and L. Zhang. On Detection of Anomalous Routing Dynamics in BGP. In *Proceedings of Networking*, 2004.

[22] Network Security Testbed based on Emulab. http://www.isi.deterlab.net.

[23] H.S. Javitz and A. Valdes. The NIDES Statistical Components: Description and Justification. Technical report, SRI Network Information Center, March 1993.