Case Study: Interactive Visualization for Internet Security

Soon Tee Teoh* Kwan-Liu Ma* S. Felix Wu* Xiaoliang Zhao†
Department of Computer Science
University of California, Davis
Department of Computer Science
North Carolina State University

ABSTRACT

Internet connectivity is defined by a set of routing protocols which let the routers that comprise the Internet backbone choose the best route for a packet to reach its destination. One way to improve the security and performance of Internet is to routinely examine the routing data. In this case study, we show how interactive visualization of Border Gateway Protocol (BGP) data helps characterize routing behavior, identify weaknesses in connectivity which could potentially cripple the Internet, as well as detect and explain actual anomalous events.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces (GUI); I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques

Keywords: anomaly detection, graph drawing, information visualization, network security

1 INTRODUCTION

The Internet is crucial to business, government, education and many other facets of society. However, Internet connectivity can be disrupted by unintentional errors or malicious attacks, which have happened on several occasions. Therefore, it is very important to have an effective way of studying various aspects of the Internet to improve its security and stability.

Internet connectivity is maintained by routers using the Border Gateway Protocol (BGP) [7] to communicate. A good way to understand the behaviors and performance of the Internet is thus to collect and analyze BGP data. In the past, monitoring and analyzing network behaviors has been mainly done by looking at some simple visual forms like x-y plots of statistical analysis results. This case study demonstrates how interactive browsing of the visual representations of archived BGP data can help users characterize a specific routing behavior in the Internet, identify weaknesses, as well as detect and explain anomalous activity. We show that visual analysis of network data can make unique and important contributions to Internet security systems.

2 INTERNET ROUTING DATA

We examine Internet routing history to discover traces left behind by configuration errors and malicious attacks. Analyzing Internet routing logs can also help determine the stability of the current routing system.

*{teoh,ma,wu}@cs.ucdavis.edu
†xzhao@unity.ncsu.edu

2.1 BGP Data and AS Routes

Each network within the Internet is identified by its IP address prefix. An example of an IP prefix is 128.120.0.0.0/16, which means every host in the network shares the same first 16 bits: 128.120. One or more networks within a single administrative domain is referred to as an Autonomous System (AS), and is assigned a unique AS number.

Between two ASes, BGP is used to exchange network reachability information so that eventually routers can forward data packets to the correct destination. BGP routers exchange messages in the form of BGP routes. A BGP route lists a particular IP prefix and the path of ASes used to reach that prefix. For example, the BGP route "128.120/16: (7,23,92)" means that the IP prefix 128.120/16 could be reached by first going to AS-7, then to AS-23, and finally to AS-92. Instability of AS routes is a cause of concern, because it impacts end-to-end communication and flow-caching.

2.2 Origin AS Changes (OASCs)

The last AS in an AS path is referred to as the Origin AS of that prefix. In the AS route example of the Section 2.1, AS-92 is the Origin AS of 128.120/16.

The Origin AS of a particular prefix can change if the prefix's ownership has changed, or because of valid network operations or faults like router misconfiguration or intentional attacks. Faults could cause data packets to be delivered to the wrong place.

We obtained archived daily BGP routing data over 480 days from the Oregon Route Views server [6]. We recorded all the instances that the Origin AS of an IP prefix changed. An Origin AS Change (OASC) [10] is an entry in the form (*Prefix,AS,Date,Type*). *Prefix* is the IP prefix whose Origin AS has changed. *AS* is a list of the associated AS(es) of the change. *Date* is the date on which the change occurred. *Type* is the type of the change.

OASCs are classified in eight different types. The details regarding the different OASC types are important to the users but are not relevant in the discussion of the visual aspects of the system, except that in our visual representation, we use different colors to represent the eight types and we allow the user to focus on a subset of the eight types. The users found these features to be useful.

This case study reports our effort in visualizing two aspects of routing data: OASCs and AS route changes. By interacting with a visual representation of the archived BGP data, the user is able to explore the data to discover patterns of behavior, anomalous events, errors, undesirable network characteristics, gaining a better understanding of Internet routing.

3 VISUAL DATA EXPLORATION

Recognizing anomaly in network data allows us to take corrective action. Anomaly detection is the process of searching for behavior deviating from normal network use. Most existing anomaly detection methods are based on statistical analysis, where user normal profiles are expressed as sets of statistical measures [5]. In contrast,

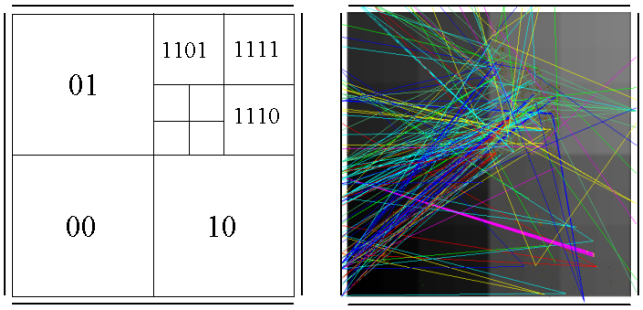

Figure 1: Quadtree coding of IP prefixes. Left: Top levels of the tree, and the most significant bits of the IP prefixes represented by each sub-tree (sub-square). 4 lines representing AS numbers surround the square representing the IP prefix space. Right: Actual data. A line is drawn for every IP-AS pair in an OASC.

our visual-based approach can quickly detect anomaly without a "normal" training set.

Our visual anomaly detection system is based on interactive data exploration. Goldstein et al. [3] describe data exploration as an iterative and interactive process initiated and directed by people. Visual techniques in data mining and/or network intrusion detection include [2] and [4].

Ahlberg and Shneiderman [1] promote visual-based methods as a viable approach to information-seeking due to the ability of humans to recognize features in visual displays and recall related images to identify anomalies. According to Girardin [2], human perception can notice even unexpected features.

Data exploration is an inherently iterative process. It is thus crucial to provide the user with the tools to interactively change parameters, focus on certain details, and animate the data over time. Our user interface design adopts two main principles given in [8] to facilitate intelligent and productive computer-human interaction: 1) rapid, incremental and reversible actions, and 2) immediate and continuous display of results.

4 VISUALIZING ORIGIN AS CHANGES

An OASC visualization system has been created. In this section, we only describe how data values are mapped to graphical values and how the user interacts with the system. The details of other aspects of the system can be found in [9].

4.1 Visual Representation and Interaction

Each IP address is mapped to one pixel on a square. The mapping is done in a quad-tree manner shown in Figure 1. A square is repeatedly subdivided into 4 equal squares. In mapping a 32-bit address to a square, we start with the first two most significant bits of the address to place it in one of the 4 sub-squares. We repeatedly use the next two most significant bits to place the prefix in a sub-square within a sub-square until the prefix is in a square the size of a single pixel or the bits of the prefix are exhausted. We use a 512×512 pixel square to represent the entire IP prefix space. Since IP prefixes have masks at most 24 bits long, at most 64 different IP prefixes map to the same pixel.

Figure 2 shows additional windows offering closeup views of several different portions of the main window, which shows the entire IP prefix space. With the additional level of zooming into a portion of the data, individual IP prefixes can be distinguished. In

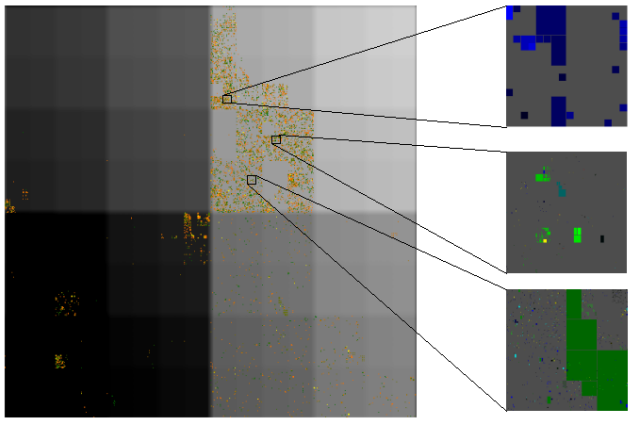

Figure 2: Data from January 2, 2000 to March 3, 2000. Colored pixels in main window show involved prefixes resolved to first 18 bits. Zoom windows resolve prefixes completely.

the main window, a pixel is colored yellow if an OASC occurred on the current day (March 3), brown if a change occurred on a previous day. In the detail windows, a colored rectangle is shown for each OASC. Its position is determined by the IP prefix, the size by the mask, brightness determined by how long ago the change occurred (present day data shown the brightest), and the hue by the type of the OASC. The background of the main window is shaded according to the IP prefix the pixel represents; the brighter the background, the larger the IP prefix represented.

To represent the relationship between a prefix and its associated AS number, we place 4 lines surrounding the IP square. An AS number is mapped to a pixel on one of the 4 lines (see Figure 1). A line is then drawn from an IP address to an AS number if there is an OASC involving that IP address and that AS number. Since there are more AS numbers than pixels, more than one AS number maps to a pixel. Again, we provide zooming features for the user to differentiate between AS numbers which map to the same pixel in the main display. The lines representing changes for the AS(es) in focus are shown with brighter colors to highlight the AS.

Our design shows one day's data at a time, allowing the user to animate the visualization, each frame showing consecutive day's data. With this "movie" display, the user can detect temporal patterns. To assist our memory of patterns from previous days, we allow a user-defined window of a certain number of days prior to the currently shown date. Data from these previous days are displayed, but with darker colors, so that the current day's data stand out. A slider bar gives the user control of the date shown.

4.2 Anomaly Detection and Analysis

Two examples of anomaly detection and analysis using our OASC visualization system are presented here.

The first is the discovery of event correlation via animation. Figure 3 shows a large number of CSM changes due to AS-15412 erroneously announcing prefixes belonging to many different ASes on Apr 6, 2001, and the CMS changes over the next 6 days to correct the error, before returning to normal on Apr 13. Figure 4 shows AS-15412 making similar errors again on Apr 18, and the corresponding CMS changes on the next day. A virtually identical pattern is easily observed. From this pattern, the user infers that the events on April 19 are corrective actions and not a new fault, demonstrating the adeptness of humans in recognizing patterns.

The second example is an anomaly on Aug 14, 2000. We used an alternative representation of the data, presented in Figure 5, In

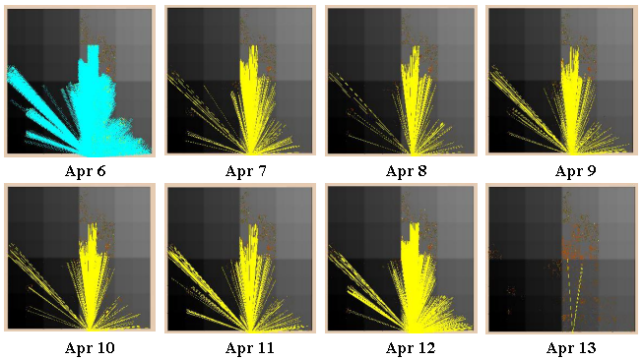

Figure 3: CSM activity on April 6, 2001 involving AS-15412 followed by 6 days of corrective CMS activity, before returning to normal.

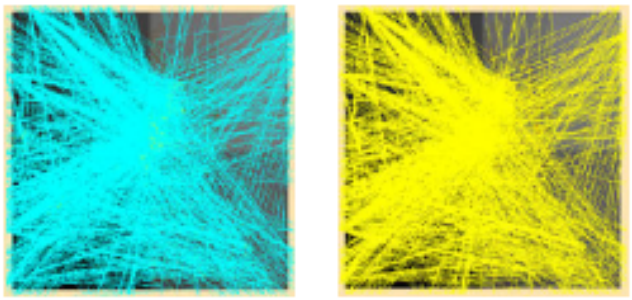

Figure 4: CSM activity on April 18, 2001 (left) involving many different ASes followed by virtually identical CMS activity (right) the next day. This indicates misconfiguration by a router, followed by correction on the next day.

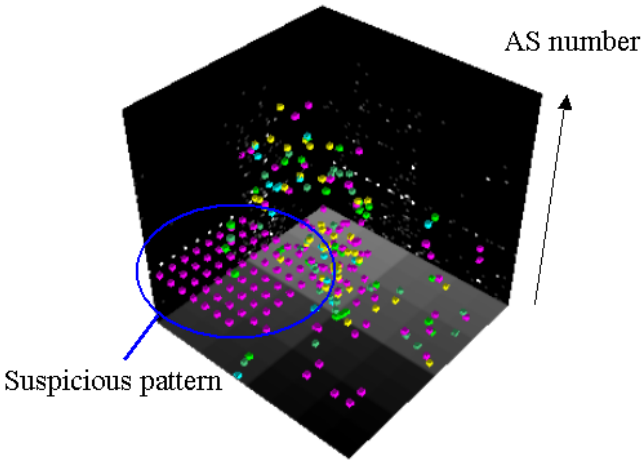

Figure 5: The regular pattern of OS-type (pink) changes on August 14, 2000 is highly suspicious. This pattern appeared because AS-7777 erroneously announced ownership of consecutive prefixes.

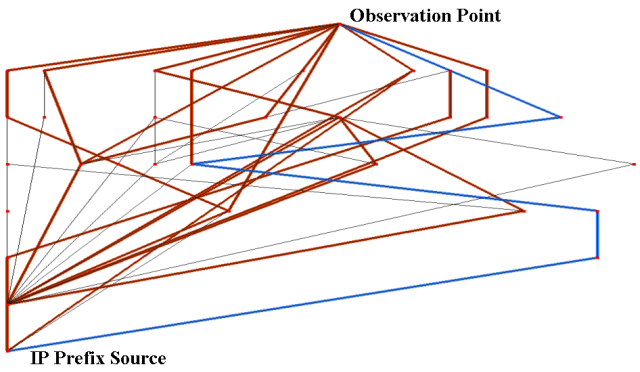

Figure 6: Each node in the graph represents an AS, and each edge represents a link between ASes. The brown lines show AS routes within one minute during Nimda worm attack. These short-lived routes are believed to be invalid. Normally, only one or two different routes should be used in a one-minute period.

this 3D representation, each Origin AS change is mapped to a cube on a horizontal plane in the same quad-tree manner. The vertical position of the horizontal plane is based on its AS number. Each cube is colored according to the type of change, as before. The OS-type (pink) changes arranged in a regular pattern on one horizontal plane are highly suspicious; they correspond to AS-7777 making erroneous announcements of these IP prefixes. Noticing the grid of pink cubes is trivial, but the same task would be very difficult for a fully-automated system to detect unless this particular pattern has been explicitly programmed into its pattern-matching mechanism.

5 VISUALIZING AS ROUTE CHANGES

Visualization can also help characterize routing behavior. The brown lines in Figure 6 show all the routes to IP prefix 55/8 within a minute during the Nimda worm attack in September 2001. Animation of the routes show many different routes involving various ASes used in a very short period of time, indicating severe instability.

To determine the stability of AS routes, it is necessary to analyze how they change over time. Statistics regarding the frequency of such changes can be easily gathered and collated. However, even when statistics such as the mean and deviation of AS route changes are known, several key questions remain: How many of these changes reflect true route changes and how many are just due to temporary link failures? What are the patterns of route changes, for example do the changes occur between two routes or many different routes? Are the different routes completely different or are there ASes that are common all all the routes? Are any of the route changes indicative of bad router implementation?

Figure 7 shows how these questions can be answered with appropriate visualization of the data. At the bottom of the figure, the number of announcements of AS route changes is plotted against time. From this plot, it is observed that there is a two-month period with significantly more announcements than usual. A 2-year window shows unique AS routes used to reach a DNS server (198.41.0.0). Each horizontal level represents a unique route. A moving 2-hour window reveals very undesirable frequent switching between AS-4200 and AS-1 during a 2-month period. The user was able to identify primary routes, secondary routes, periods of instability, two instances of change in primary routes and overlap between the routes.

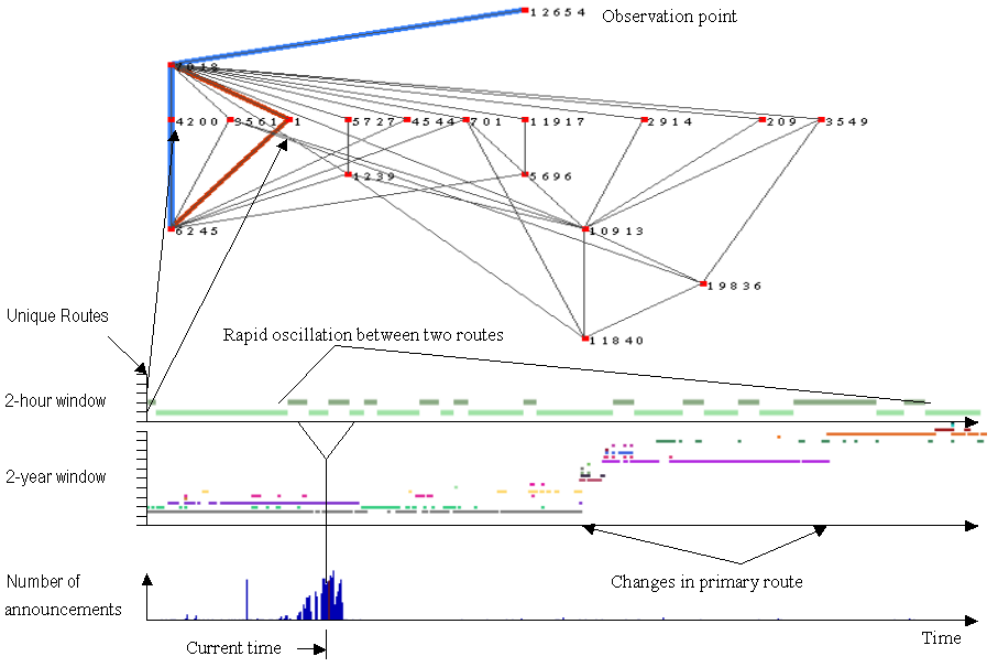

Figure 7: Visualization of AS Routes to a DNS Server. In the 2-year window, each horizontal level represents one unique AS route. A pixel is colored if in that time period, the AS route is used (note that in any period, more than one AS route may be used). Two changes in primary routes are observed and labeled. A vertical line marks the current time, which is expanded into the 2-hour window. This window shows rapid oscillation between two AS routes, which are highlighted in the network diagram above.

6 CONCLUSIONS

We have demonstrated how visualization made it easier to detect and analyze anomalies, as well as to characterize Internet routing behavior and instability. Non-visual Internet security methods would have much greater difficulty in correlating event, discovering announcement of consecutive prefixes, and distinguishing between different types of AS route changes. Animation of AS routes superimposed on the network graph showed users the differences in the routes, giving them hints regarding whether a particular node or link may be a point of failure. With statistical analysis of AS route changes, the user gets general information such as the average persistence or prevalence of a route, whereas our case study shows that with visualization of the data, the user was able to obtain much richer information.

One lesson we have learned from this study is that multiple representations are sometimes needed for understanding different aspects of the data. We also found that visual metaphors often have to be customized for particular applications. From user feedbacks, we confirm that interactivity and an intuitive interface are crucial to successful data exploration.

In our continuing work, we will apply the visualization tool to analyze more data from different IP prefix sources and observation points, with the goal of discovering as yet unknown problems. Understanding routing behavior and recognizing weaknesses, attacks and faults from BGP data would help make the Internet more stable and secure.

REFERENCES

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

[2] L. Girardin. An eye on network intruder-administrator shootouts. *Proceedings of the Workshop on Intrusion Detection and Network Monitoring (ID'99)*, 1999.

[3] J. Goldstein, S.F. Roth, and J. Mattis. A framework for knowledge-based, interactive data exploration. *Journal of Visual Languages and Computing*, pages 339–363, December 1994.

[4] D. Keim and H-P Kriege. Visualization techniques for mining large databases: A comparison. *Transactions on Knowledge and Data Engineering, Special Issue on Data Mining*, 1996.

[5] T. Lunt. Detecting intruders in computer systems. *Proceedings of the 1993 Conference on Auditing and Computer Technology*, 1993.

[6] D. Myer. University of oregon route views project. *http://www.antc.uoregon.edu/route-views/*.

[7] Y. Rekhter and T. Li. A border gateway protocol 4 (bgp-4). *RFC 1771*, 1995.

[8] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Second Edition*. Addison-Wesley Publ. Co, Reading, Massachusetts, 1992.

[9] S.T. Teoh, K.-L. Ma, S.F. Wu, and X. Zhao. A visual technique for internet anomaly detection. *IASTED International Conference on Computer Graphics and Imaging*, 2002.

[10] X. Zhao, D. Pei, L. Wang, D. Massey, A. Mankin, S.F. Wu, and L. Zhang. An analysis of bgp multiple origin as (moas) conflicts. *SIGCOMM Internet Measurement Workshop*, 2001.