On Knowledge-Based Classification of Abnormal BGP Events

Jun Li, Dejing Dou, Shiwoong Kim, Han Qin, and Yibo Wang*

University of Oregon
{lijun, dou, shkim, qinhan, wangyibo}@cs.uoregon.edu

1 Introduction

One key factor that ensures smooth data delivery over the Internet and keeps the Internet healthy is the well-being of the Internet's inter-domain routing. In today's Internet, the *de facto* standard inter-domain routing protocol is the Border Gateway Protocol, or BGP, that keeps every BGP router updated about which BGP router is the next hop in reaching a particular network and which autonomous systems (AS), in order, it has to cross. Unfortunately, various abnormal events—such as fast-spreading worms or large-scale power outages—can affect the normal operation of BGP. Not only can these events cause routers or BGP sessions between routers to go down—a denial-of-service attack, but they can also create havoc as the scale of damage rises.

It is therefore critical to investigate how such events may impact BGP and whether or not different events can be classified into different types so that proper actions can be taken. Some may argue that the occurrence of such events is uncommon, and once they occur, people will easily know them anyway because of their large-scale damage. However, even if BGP anomalies may be uncommon today, they can have disastrous results once they occur tomorrow. It is also likely that the increased Internet complexity and the continuing challenges to make BGP secure and robust will cause future BGP anomalies both more common and more damaging.

We have designed an *Internet Routing Forensics* framework to provide a new, systematic approach to detecting the occurrence of abnormal events that impact BGP [1]. Basically, we are able to apply data mining techniques to BGP data corresponding to already-known abnormal events, discover rules about how BGP data may differ from the norm during those events, and then further use those rules to detect the occurrence of abnormal events from the past or in the future.

What remains unclear, then, is whether different abnormal events can be further differentiated from each other, and if so, how. In addition to obtaining rules to effectively capture the existence of anomalies in BGP data (BGP updates in particular), it is important to learn whether we can also obtain rules to indicate the disparity—as well as commonality—between, say, a large-scale power outage and a fast-spreading worm, or between different worms.

* This material is based upon work supported by the National Science Foundation under Grant No. 0520326.

P. McDaniel and S.K. Gupta (Eds.): ICISS 2007, LNCS 4812, pp. 267–271, 2007.

2 Approach

In [2], we have studied a *data-driven* approach to identifying the specific type of an abnormal event without knowledge of BGP. In this paper, we devise an approach that relies on BGP knowledge to classify different abnormal events that impact BGP, i.e., a *knowledge-driven* approach. As events at the global level tend to affect the largest number of networks over the Internet, in this paper we focus on these events, and study how to develop accurate classification rules to describe each individual class of them. In order to support real-time applicability, our basis for classification is the observable impact on BGP from abnormal events that can be measured in real time.

Knowledge-based classification requires knowledge of abnormal BGP events before we try to obtain rules of different classes of these events. The knowledge can be simply the class name of a particular type of events. In this case, we can treat all classes of abnormal events at the same level and conduct i.e., **flat classification**. Or, our knowledge about abnormal BGP events can be enriched by knowing the hierarchical relationship of different classes of abnormal BGP events, allowing us to obtain and test rules for a hierarchy of abnormal event classes, i.e., **hierarchical classification**.

Our BGP data are BGP updates from the periods of the events as well as normal periods, archived by RouteViews [3] or RIPE [4]. We calculate the per-minute values of the most relevant attributes (selected through information gain measure) about these BGP updates, and arrange these values in a chronological sequence of 1-minute bins. If a 1-minute bin is known to correspond to a specific class of abnormal event, we label it with the name of that class.

We then conduct a training process to obtain rules for different classes of abnormal events, using the *C4.5* classification algorithm [5].

In applying these rules against testing bins from a certain event period, we use a probabilistic approach. As a rule is not typically 100% accurate, and a testing bin may match to more than one rule for different classes, or match no rule at all, we design an alert algorithm as follows: If more than Γ percentage of testing bins have a probability matching class C higher than ϵ, we raise an alert than an event of class C occurs. We use 40% for Γ and 0.5 for ϵ in this paper.

3 Case Studies

We conduct case studies on six abnormal events: Code Red worm, Nimda worm, Slammer worm, East Coast blackout, Florida blackout, and Katrina blackout.

With flat classification, we obtain rules for seven classes at the same level: CODERED, NIMDA, SLAMMER, EAST-COAST, FLORIDA, KATRINA, and NORMAL. Table 1 shows the percentage of "hits" in a test set for each of the seven classes, i.e., the γ values (Section 2). Here, the flat classification is effective in distinguishing the three worm-related classes—CODERED, NIMDA, SLAMMER—as well as the NORMAL class. However, it is not effective in telling the three blackout-related classes apart (we explain this toward the end of this section).

Table 1. γ values (percentages) for test sets in the case study using flat classification

Test set	CODERED	NIMDA	SLAMMER	EAST-COAST	FLORIDA	KATRINA	NORMAL
Code Red worm	**82.3**	5.4	0.0	0.8	0.0	0.8	13.1
Nimda worm	0.8	**84.6**	10.8	0.0	0.0	0.0	3.8
Slammer worm	0.0	13.8	**86.2**	0.8	0.0	0.8	0.8
East Coast blackout	0.0	0.0	0.0	**61.5**	0.0	**47.7**	34.6
Florida blackout	0.0	0.0	0.8	0.8	**36.9**	0.8	49.2
Katrina blackout	0.0	0.0	0.8	0.0	7.7	**0.0**	40.8
Normal	0.0	0.0	0.8	4.5	7.6	4.5	**51.5**
Alert Threshold $\Gamma = 25\%$							

Table 2. γ values (percentages) for test sets in the case study using hierarchical classification at a *high* level

Test set	WORM	BLACKOUT	NORMAL
Code Red worm	**85.4**	1.5	14.6
Nimda worm	**96.2**	0.8	4.6
Slammer worm	**99.2**	1.5	0.0
East Coast blackout	0	**75.4**	27.7
Florida blackout	0.77	**68.5**	25.4
Katrina blackout	2.31	**66.9**	26.9
Normal	0.0	22.0	**45.5**
Alert Threshold $\Gamma = 25\%$			

With hierarchical classification, we have two high-level classes—**WORM** and **BLACKOUT**, three sub-classes of the WORM class—**WORM.CODERED**, **WORM.NIMDA** and **WORM.SLAMMER**, and three sub-classes of the BLACKOUT class—**BLACKOUT.EAST-COAST**, **BLACKOUT.FLORIDA**, and **BLACKOUT.KATRINA**. Table 2 shows that the hierarchical classification case study can distinguish between WORM and BLACKOUT (and also as opposed to the NORMAL class). Moreover, the three WORM subclasses can be distinguished (Table 3), and so can the three BLACKOUT subclasses (Table 4).

As our results above show, the hierarchical classification is more accurate than the flat classification. It does not need to train many classes altogether, an advantage when the difference between different classes are small. In our case studies, as opposed to seven classes in flat classification, the hierarchical classification only needs to train two or three each time. The hierarchical structure of classes also helps incorporate a new class more efficiently: We only need to regenerate rules for classes at the level of the new class on a hierarchy, as opposed to all classes in the flat classification.

The hierarchical classification is also more efficient as it checks less number of classes. A simplified comparison is as follows: Assume that the cost of verifying rules associated with every class is the same. In hierarchical classification, every non-leaf class has m sub-classes, level i has m^i classes, and there are a total of

Table 3. γ values (percentages) for each test set in the case study using hierarchical classification at a *specialized* level for worm-related classes

Test set	WORM.CODERED	WORM.NIMDA	WORM.SLAMMER	NORMAL
Code Red worm	**66**	1.2	0.6	31.5
Nimda worm	0.6	**69.1**	14.2	6.8
Slammer worm	0	2.3	**95.4**	0.8
Normal	0	2.1	0	**95.1**
Alert Threshold $\Gamma = 25\%$				

Table 4. γ values (percentages) for each test set in the case study using hierarchical classification at a *specialized* level for blackout-related classes

Test set	BLACKOUT.EAST-COAST	BLACKOUT.FLORIDA	BLACKOUT.KATRINA	NORMAL
East Coast blackout	**54.6**	0.8	0.0	39.2
Florida blackout	0.0	**30.8**	0.0	58.5
Katrina blackout	0.0	4.6	**40.0**	52.3
Normal	13.1	4.6	6.2	**64.6**
Alert Threshold $\Gamma = 25\%$				

L levels. In flat classification, there are, in total, m^L classes (equivalent to the number of leaf classes in hierarchical classification). During hierarchical classification, we need to check rules of all m classes from level 1, find the matching class, check rules of all its m sub-classes, and repeat until we find out which leaf class matches the testing data. We thus need to check $m \times L$ classes. On the other hand, during flat classification, we need to check against the rules of all m^L classes. Clearly, in most cases, $m \times L \ll m^L$.

4 Summary

In this paper, we proposed a knowledge-based classification approach to distinguishing abnormal events that affect BGP. We demonstrated that we can obtain classification rules about every different abnormal event class, and use the rules to report the occurrence of an abnormal event of a certain class. Our approach further encompasses two classification methodologies: flat classification and hierarchical classification, and our case studies show that the hierarchical classification, in general, is more accurate, efficient, and scalable.

A direct implication of this work is the real-time application in detecting BGP anomalies caused by certain events, an important but missing component in today's Internet. In the future, we will investigate how our studies can complement other work on BGP anomalies and BGP dynamics root cause analysis, and further explore how to quantify the impact on BGP by abnormal events.

References

1. Li, J., Dou, D., Wu, Z., Kim, S., Agarwal, V.: An Internet routing forensics framework for discovering rules of abnormal BGP events. ACM SIGCOMM Computer Communication Review 35(5), 55–66 (2005)
2. Dou, D., Li, J., Qin, H., Kim, S., Zhong, S.: Understanding and utilizing the hierarchy of abnormal BGP events. In: SIAM International Conference on Data Mining, Minneapolis, Minnesota, pp. 457–462 (April 2007) (short paper)
3. University of Oregon Route Views Project, `http://antc.uoregon.edu/route-views/`
4. RIPE NCC, RIPE routing information service raw data, `http://data.ris.ripe.net/`
5. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco (1993)