9-2013

AS-CRED: Reputation and Alert Service for Inter-Domain Routing

Jian Chang
OpenX, Pasadena, CA

Krishna Venkatasubramanian
Worcester Polytechnic Institute, kven@cs.wpi.edu

Andrew G. West
University of Pennsylvania, westand@cis.upenn.edu

Sampath Kannan
University of Pennsylvania, kannan@cis.upenn.edu

Insup Lee
University of Pennsylvania, lee@cis.upenn.edu

See next page for additional authors

Recommended Citation

Jian Chang, Krishna Venkatasubramanian, Andrew G. West, Sampath Kannan, Insup Lee, Boon Thau Loo, and Oleg Sokolsky, "AS-CRED: Reputation and Alert Service for Inter-Domain Routing", *Systems Journal, IEEE* 7(3), 396-409. September 2013.
http://dx.doi.org/10.1109/JSYST.2012.2221856

AS-CRED: Reputation and Alert Service for Inter-Domain Routing

Abstract
Being the backbone routing system of the Internet, the operational aspect of the inter-domain routing is highly complex. Building a trustworthy ecosystem for inter-domain routing requires the proper maintenance of trust relationships among tens of thousands of peer IP domains called Autonomous Systems (ASes). ASes today implicitly trust any routing information received from other ASes as part of the Border Gateway Protocol (BGP) updates. Such blind trust is problematic given the dramatic rise in the number of anomalous updates being disseminated, which pose grave security consequences for the inter-domain routing operation. In this paper, we present ASCRED, an AS reputation and alert service that not only detects anomalous BGP updates, but also provides a quantitative view of AS' tendencies to perpetrate anomalous behavior.

AS-CRED focuses on detecting two types of anomalous updates (1)*hijacked:* updates where ASes announcing a prefix that they do not own; and (2) *vacillating:* updates that are part of a quick succession of announcements and withdrawals involving a specific prefix, rendering the information practically ineffective for routing. AS-CRED works by analyzing the past updates announced by ASes for the presence of these anomalies. Based on this analysis, it generates AS reputation values that provide an aggregate and quantitative view of the AS' anomalous behavior history. The reputation values are then used in a tiered alert system for tracking any subsequent anomalous updates observed. Analyzing AS-CRED's operation with real-world BGP traffic over six months, we demonstrate the effectiveness and improvement of the proposed approach over similar alert systems.

Keywords
Border Gateway Protocol, autonomous systems, reputation, alert service

Author(s)
Jian Chang, Krishna Venkatasubramanian, Andrew G. West, Sampath Kannan, Insup Lee, Boon Thau Loo, and Oleg Sokolsky

AS-CRED: Reputation and Alert Service for Inter-domain Routing

Jian Chang, Krishna K. Venkatasubramanian, *Member, IEEE,* Andrew G. West, Sampath Kannan, Insup Lee, *Fellow, IEEE,* Boon Thau Loo, and Oleg Sokolsky *Member, IEEE,*

***Abstract*—Being the backbone routing system of the Internet, the operational aspect of the inter-domain routing is highly complex. Building a trustworthy ecosystem for inter-domain routing requires the proper maintenance of trust relationships among tens of thousands of peer IP domains called Autonomous Systems (ASes). ASes today implicitly trust any routing information received from other ASes as part of the Border Gateway Protocol (BGP) updates. Such blind trust is problematic given the dramatic rise in the number of anomalous updates being disseminated, which pose grave security consequences for the inter-domain routing operation. In this paper, we present AS-CRED, an AS reputation and alert service that not only detects anomalous BGP updates, but also provides a quantitative view of AS' tendencies to perpetrate anomalous behavior.**

AS-CRED focuses on detecting two types of anomalous updates (1) *hijacked:* updates where ASes announcing a prefix that they do not own; and (2) *vacillating:* updates that are part of a quick succession of announcements and withdrawals involving a specific prefix, rendering the information practically ineffective for routing. AS-CRED works by analyzing the past updates announced by ASes for the presence of these anomalies. Based on this analysis, it generates AS reputation values that provide an aggregate and quantitative view of the AS' anomalous behavior history. The reputation values are then used in a tiered alert system for tracking any subsequent anomalous updates observed. Analyzing AS-CRED's operation with real-world BGP traffic over six months, we demonstrate the effectiveness and improvement of the proposed approach over similar alert systems.

***Index Terms*—Border Gateway Protocol, autonomous systems, reputation, alert service**

I. Introduction

In the realm of inter-domain routing, large IP domains, called Autonomous Systems (ASes), use the Border Gateway Protocol (BGP) for exchanging reachability information among themselves. Being the backbone routing system of the Internet, the operational aspect of the inter-domain routing is highly complex. It involves the coordination and cooperation of tens of thousands of ASes and millions of routing devices, structured in a fully decentralized and distributed fashion, where each participant AS enforces its own routing policies to achieve various business and traffic engineering goals.

The current design of BGP implicitly assumes complete trust between ASes. This blind trust assumption is problematic as it has been the key vulnerability for a growing number of attacks on the Internet's operation [1]. These attacks are usually carried out by ASes that announce *anomalous BGP updates* containing invalid reachability information (*e.g.,* hijacked IP prefixes). These attacks fundamentally affect the accessibility of the Internet and can have grave consequences to attacks akin to DNS poisoning [2] and phishing [3]. The reasons for these incidents have usually been found to be either malice such as spamming [4] or misconfiguration [5], [6], [7]. There are three major *challenges* in securing the inter domain routing from these attacks: (1) *lack of ground trust:* there is no authoritative source of information to determine the validity of BGP updates; (2) *dynamic and mixed AS behavior:* ASes announce both valid and anomalous updates (often simultaneously, but for different prefixes); and (3) *scale of the Internet:* it is often very expensive (in terms of time and resources) to deploy a security mechanism covering the entire inter-domain routing system.

Two approaches have traditionally been taken for securing inter-domain routing: *prevention* and *detection*. The former requires the use of cryptographic mechanisms and attempts to overcome the first challenge by "building the ground truth", so that only the announcements of the prefixes that an AS can reach directly (*i.e.*, own[1]) would be accepted by its peer ASes. The most famous example of such preventive schemes is S-BGP [8]. However, these approaches often impose a too high deployment and operation cost to be useful [9], [10], thus failing to address the third challenge. Consequently, the principal aim of recent research has been to detect instances of anomalous updates at the control-plane [11], [12], [13] or the data-plane [14], [15], [16] of BGP. These detection approaches propose various heuristics to determine the update validity. However, they do not provide sufficient evidence that the proposed heuristics are robust to the evolving and mixed nature of AS behaviors, thus failing to address the second challenge. Further, some of the detection mechanisms also fail to address the third challenge as they need to be deployed at specific locations (*e.g.,* being the victim of the attacks) to be effective [16].

In this paper we take a different approach. Based on real-world BGP traffic, we observe that ASes that announced an anomalous update in the past are very likely to repeat such behavior in the near future. We therefore focus on developing an AS reputation and alert service that not only detects anomalous BGP updates, but also provides a quantitative view of the tendencies of ASes to perpetrate anomalous behavior.

K.K. Venkatasubramanian is currently with the Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA. E-mail: kven@cs.wpi.edu. J. Chang is currently with OpenX, Pasadena, CA. The other authors are with the Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA. E-mail:{jianchan,westand,kannan,lee,boonloo,sokolsky}@cis.upenn.edu.

This research was supported in part by ONR MURI N00014-07-1-0907. POC: Insup Lee, lee@cis.upenn.edu

[1]We use the term *own* to describe prefixes: (1) allocated to ASes by a Regional Internet Registry (RIR) such as ARIN, RIPE and APNIC; or (2) belonging to the customers of Autonomous Systems whose prefixes the ASes are aggregating.

This service, called *AS-CRED*, is inspired by the notion of *credit score*, which has been used as an effective approach for solving trust problems in the complex world of finance that involves billions of entities and highly uncertain interactions. AS-CRED works by analyzing the past update announcements of all observed ASes in the Internet for the presence of two types of anomalous updates: (1) *hijacked* updates where ASes announcing a prefix that they do not own; and (2) *vacillating* updates that are part of a quick succession of announcements and withdrawals involving a specific prefix, rendering the information practically ineffective for routing. Inspired by previous works [7], [12], [17], the analysis of historical updates is done based on the sustenance of prefix ownership. Out of this stability analysis, AS-CRED generates feedback on the ASes, which is then fed into a reputation function to compute *AS reputation*. Since the reputation is computed based on trustworthy local feedback, AS-CRED is not vulnerable to biased/incorrect feedback deliberately providing by colluding ASes. Additionally, based on the analysis of the historical BGP data, AS-CRED also creates a "white-list": a list of AS-prefix pairs where the prefix is stably owned by the respective AS for long periods of time, proving the legitimacy of the ownership. The reputation and the "white-list" are then used to design a novel tiered alert system, as follows: (1) AS reputation is used as a behavior-predictive metric for generating hijacked or vacillating alerts for updates from ASes that have poor reputation (*i.e.*, announced large number of anomalous updates). (2) The "white-list" is then used to filter out (bound the inaccuracy of the alert generation process) updates with AS-prefix pairs that are considered deemed legitimate. (3) To compensate for sudden behavior pattern changes of reputable ASes, a special alert type, *Potentially Invalid*, is triggered.

The analysis of past BGP data allows AS-CRED to correctly classify historical AS behavior. The idea here is that although there is no complete and accurate ground truth available to determine the validity of BGP updates in real-time, such a task can be effectively performed with the benefit of hindsight, thus addressing the first challenge. To address the second challenge, the reputation function incorporates the notion of time-decay to adapt to the evolution of AS behavior patterns. Moreover, the reputation function is designed to solely consider anomalous AS behavior, thus preventing a misbehaved AS from inflating its reputation by announcing large number of regular updates. As BGP operates by exchanging reachability information about all the active ASes and prefixes in the Internet, local information obtained from a set of well-connected BGP nodes can be used to compute reputation values for the observable portion of the Internet at the inter-domain level, thus addressing the third challenge.

Our implementation of AS-CRED is based on live BGP trace from the RouteViews project with its operation results publicly available[2]. The public availability of the AS reputation and alerts not only incentivizes good behavior from ASes, but also provides an effective diagnostic and forensic tool to debug network connectivity issues at Internet scale. AS-CRED currently is a centralized system, but can easily be implemented by individual ASes in a distributed manner for obtaining their local views of peer ASes reputation and triggering customized alerts. The *contributions* of the paper are: (1) prefix ownership stability heuristics for detecting anomalous BGP updates; (2) an adaptive AS reputation scheme; and (3) a tiered reputation-based alert service that accurately tracks anomalous updates. The analysis of AS-CRED over a six month period indicates its effectiveness and improvement of over similar alert systems.

The paper is organized as follows; Section II presents the background, the anomaly model, the problem statement, and an overview of our approach. Section III presents details of AS-CRED architecture, the feedback and reputation model, and the data source. Section IV and V present the historical AS anomaly detection and the alert generation service of AS-CRED, respectively. Section VI presents mechanisms for tuning the various parameters of AS-CRED. Sections VII and VIII present the security and performance analysis results of AS-CRED. Section IX presents the related work. Finally, Section X concludes the paper.

II. Background

A. The Border Gateway Protocol

The Border Gateway Protocol is a path-vector routing protocol for exchanging information about reaching IP address prefixes [18]. Using BGP, an AS x, which owns a prefix p, announces an *update* notifying its neighboring AS y of its ownership. The AS x is called the announcer or announcing AS. AS y then forwards this update further to its neighbor AS z by adding its own AS number to the path vector, called *AS_PATH*, in the update. This informs AS z that in order to reach the prefix p, the gateway router at AS y is the next hop. When an update is received at an AS, it determines whether the update should be accepted or not. The acceptance of an update implies that the router is willing to add the route to the prefix into its routing information base. Each AS has its own policies that determine whether it accepts a BGP update and whether the update can be forwarded to its neighbors. Routing policies serve an important purpose in BGP and provide an AS with not only the capability to prefer one route over another, but also to filter or tag an update to change the route's relative preference downstream.

B. Anomaly Model

In this section, we summarize the different types of anomalous updates that AS-CRED detects. Critical to understanding these anomalies is the notion of AS-prefix binding. We define the term *AS-prefix binding* $\{a, p\}$ as a claimed ownership of a particular prefix p by AS a. It is *established* when AS a announces prefix p for the first time through a BGP update. An AS-prefix binding may have many *instances*, which refer to an announcement and corresponding withdrawal of prefix p by AS a. AS-CRED considers two types of anomalous updates:

- *Hijacked Updates:* these updates establish AS-prefix bindings with prefixes not belonging to the AS making the announcement [12]. Table I shows a list of well-known hijacked prefix announcements in the past. Hijacking is a persistent threat within the inter-domain world and has been triggered as a result of misconfiguration [5] or for malicious purposes such as spamming [4].

[2] http://rtg.cis.upenn.edu/qtm/ascred/

TABLE I
AS-PREFIX BINDING STABILITY FOR DOCUMENTED PREFIX HIJACKING INSTANCES

Date	Prefix Hijacked	Victim AS	Attacker AS	Duration	# of Instances
Dec. 2004 - Jan. 2005	61.0.0.0/8	Various	4787	$<$ 1 minute	100+
Dec. 2004 - Jan. 2005	82.0.0.0/8	Various	8717	$<$ 1 minute	100+
Jan. 13, 2007	12.0.0.0/8	7018	31604	4 hours 26 minutes	1
Feb. 24, 2008	208.65.153.0/24	36561 (YouTube)	17557	9 hours 45 minutes	1
Mar. 15, 2008	194.9.82.0/24	36915	6461	17 minutes	1
Apr. 8, 2010	29867 Prefixes	Various	23724	$\sim$8 hours	30K+

TABLE II
EXAMPLES OF VACILLATING AS-PREFIX BINDINGS (NAW: NUMBER OF ANNOUNCEMENTS AND WITHDRAWALS)

AS	Prefix	NAW	Duration Observed
145	140.217.157.0/24	1080	Nov. 1 - Nov. 27, 2009
8452	41.235.83.0./24	2088	Nov. 2 - Nov. 10, 2009
8452	41.235.87.0./24	1602	Nov. 2 - Nov. 10, 2009
704	152.63.49.180/30	1628	Dec. 8 - Dec. 31, 2009
2905	41.210.184.0/24	1774	Dec. 23, 2009 - Jan. 06, 2010

- *Vacillating Updates:* these updates establish AS-prefix bindings with a large number of short-lived instances. Such AS-prefix bindings are the result of a quick succession of announcements and withdrawal of prefixes by ASes, rendering the information practically ineffective for routing. For instance, AS37035 was seen announcing and withdrawing the prefix 41.222.179.0/24, which it owns, 4824 times between Dec. 3, 2009 and Dec. 7, 2009 (more examples are shown in Table II). This amounts to announcing and withdrawing the prefix repeatedly, once every 1.5 minutes on average. Vacillating prefixes are an important cause of route-flapping, a behavior which can lead to the propagation of excessive number of updates depleting BGP router resources.

The bindings established by these anomalous updates are called hijacked AS-prefix bindings and vacillating AS-prefix bindings, respectively. AS-prefix bindings established due to such anomalous updates are collectively called *invalid* AS-prefix bindings. Any binding not deemed invalid is considered valid. Note that in this paper we do not consider *AS_PATH* related anomalies (*e.g.*, path spoofing, or violation of valley-free routing). Further, in this paper we operate at the abstraction of AS-prefix bindings and not prefixes or ASes alone. Therefore, a multi-homed prefix p announced by both AS x and AS y, will result in two separate AS-prefix bindings $\{x, p\}$ and $\{y, p\}$. Similarly, a prefix p and its sub-prefix p' announced by AS f and AS g respectively, will be treated as two separate AS-prefix bindings $\{f, p\}$ and $\{g, p'\}$. This allows us to not explicitly distinguish between the cases of prefixes and sub-prefixes when discussing the validity of the AS-prefix bindings. In the rest of the paper, we use the terms *AS-prefix binding*, *prefix binding* and *binding*, interchangeably.

C. Problem Statement and Approach

The principal questions that we want address in this paper are: (1) how to characterize the tendency of an AS announcing anomalous BGP updates? and (2) how to use this AS behavior to generate alerts for any subsequent anomalous updates?

Fundamentally, the answer to the above questions requires effective trust quantifications of AS behavior. In this regard, we take a reputation-based approach. The reputation of an entity (an AS, in this context) is a characterization of its past performance relative to a specific set of behaviors. For entities that are consistent in their behavioral patterns, reputation forms an effective and predictive model. Using reputation, one can trust/distrust entities based on the degree to which they exhibit specific behaviors. The key for successfully using reputation systems is to ensure that gaining high reputation requires a considerable amount of resources and time devoted by an entity. In other words, reputation systems exploit the limitations of the adversary by trading-off resources and time for security.

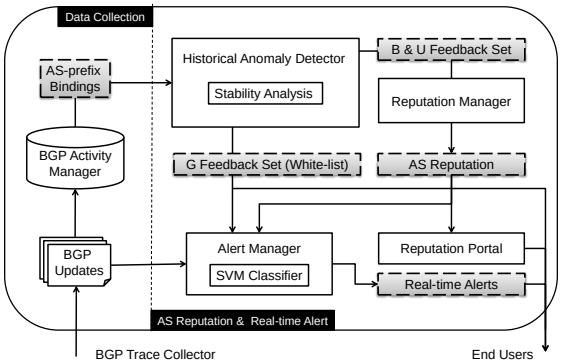


Fig. 1. AS-CRED Architecture

TABLE III
SUMMARY OF USED ACRONYMS

Acronym	Description
ASes	Autonomous Systems
BGP	Border Gateway Protocol
GBU Set	The three feedback sets for AS behavior
Pr (Ps)	Prevalence (Persistence) of AS-prefix bindings
TPr (TPs)	Prevalence (Persistence) threshold for stability analysis of AS behavior
VBL	Valid Binding List
VT (HJ) set	Set of AS-prefix bindings with Hijacked (Vacillating) alert label
IRR	Internet Routing Registeries
IAR	Internet Alert Registry

Reputation systems work by (1) identifying behaviors of interest, (2) monitoring for exhibition of the behaviors, and (3) providing feedback on the experience. Once the feedback being received, the reputation can be computed based on a mathematical function. In this work, the behavior in question is the announcement of anomalous updates that contain invalid AS-prefix bindings. Our approach is a three-step process: (1) *Historical Anomaly Detection:* Evaluate the past updates announced by ASes for establishing hijacked or vacillating bindings, (2) *Reputation Computation:* compute AS reputation based on the identified anomalous behavior, (3) *Alert Generation:* use the reputation to trigger alerts for any invalid bindings in subsequent updates. Table III summarizes the principal acronyms used in the rest of the paper.

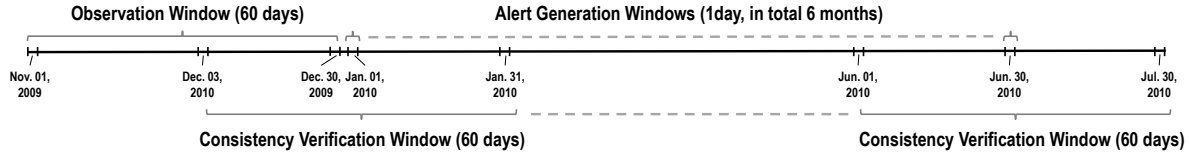


Fig. 2. Data Source Time Windows

III. AS-CRED: Reputation and Alert Service for Inter-domain Routing

In this paper, we present *AS-CRED*, an AS reputation and alert service for inter-domain routing. We begin by describing the architecture of AS-CRED, the feedback and reputation mechanism that forms the core of AS-CRED, and the data source that it uses for reputation computation and later for analysis.

A. AS-CRED Architecture

The AS-CRED service is designed to be a portal for disseminating information about ASes and their anomalous updates announcements. AS-CRED has five main components (see Figure 1):

- *BGP Activity Manager:* This is a database that collects BGP updates from well-connected BGP trace collectors (*e.g.*, RouteViews [19]). The data provides a view of active ASes in the Internet and the prefixes that they announce at different times.
- *Historical Anomaly Detector:* This component analyzes the updates in the BGP Activity Manager, received within a specific time duration called the *observation window*, for the presence of anomalies. The output is a "white-list" of valid AS-prefix bindings and a classification of the past updates announced by ASes. The latter forms the feedback for the Reputation Manager.
- *Reputation Manager:* This computes the reputation of the ASes based on the feedback provided by the Historical Anomaly Detector.
- *Reputation Portal:* The reputation computed for the ASes is made available through a web portal.
- *Alert Manager:* Combining the "white-list" with the reputation values, it triggers real-time alerts.

Note that, AS-CRED *dynamically* manages the reputation of ASes as their behavior changes over time. In this regard, the Historical Anomaly Detector continuously evaluates the updates received over a sliding window, which includes newer updates and excludes older ones, and provides updated feedback to the Reputation Manager.

B. Feedback and Reputation Computation

AS reputation is computed based on the feedback provided by the Historical Anomaly Detector. In this section, we describe the types of feedback that the analysis of historical updates from ASes provides and the reputation function used by the Reputation Manager.

1) Feedback: In AS-CRED, feedback is a triple of the form $\{a, p, t\}$, where a is the AS announcing the prefix p at time-stamp t. Each feedback triple is exclusively classified into one of the three feedback sets, namely, G (good), B (bad), and U (ugly): (i) A feedback $g_i = \{a, p, t\}$ in the *G* set is provided each time an AS announces a valid AS-prefix binding. The AS and prefix involved are said to be exhibiting *good behavior*. (ii) A feedback $b_i = \{a, p, t\}$ in the *B* set is provided each time an AS's behavior is not good but does not subvert the intended BGP operation; and (iii) A feedback $u_i = \{a, p, t\}$ in the *U* set is provided each time an AS does not demonstrate good behavior and subverts the intended BGP operation. We use the term GBU sets to refer to the three feedback sets, collectively. The act of announcing AS-prefix bindings that populate the B or the U set is called *poor behavior*. The GBU sets form the feedback that is provided for AS reputation computation. Note that, an AS may demonstrate good behavior for one prefix but simultaneously demonstrate poor behavior for others. Section IV describes how these feedback sets are populated. Finally, as the feedback is generated locally, we do not have to consider the case of potentially dishonest external feedback affecting our reputation computation outcome.

2) AS Reputation Function: The reputation assigned to ASes by the Reputation Manager is a measure of: (1) how many invalid AS-prefix bindings they establish, and (2) how often they establish such invalid bindings. Computing reputation in this manner allows AS-CRED to protect itself from ASes mounting self-promotion attacks as we shall see in Section VII. Further, valid AS-prefix bindings far outnumber the invalid ones, making the measurement of invalidity far more useful.

The reputation of an AS is computed based on the feedback in the B and the U sets. The values do not have absolute meaning and must be interpreted in a relative manner. The reputation in AS-CRED is calculated using the following function:

$$Rep_X(a) = \sum_t 2^{-(t_{now}-t)/h_X} \quad (1)$$

Here, $Rep_X(a)$ is the reputation of an AS a for exhibiting poor behavior type X, $X \in \{B, U\}$. t_{now} is the current time and t is the time-stamp of when X was observed. h_X is the half-life of the decay function for exhibiting the behavior X. The values of $t_{now} - t$ are in the same units as h_X. It can be seen that the reputation returned for an AS varies between 0 (the best possible reputation) and Ω (the worst possible reputation[3]). Section VIII-A2 describes how we assign the half-life value for computing the reputation.

In AS-CRED, reputation of an AS a is thus a vector of the form $[Rep_B(a), Rep_U(a)]$, where $Rep_B(a)$ is the reputation of an AS a based on each of its entries in the B set. Similarly, $Rep_U(a)$ is the reputation of an AS a based on each of its entries in the U set. The reputation value changes depending upon the addition of associated feedbacks into the B or the U set. Therefore, the reputation essentially quantifies the extent

[3]The absolute worst AS is the one that has an entry in the B or the U set for every possible time-stamp in the observation window and at each time-stamp it has committed a poor behavior for all possible prefixes in the IP address space.

of invalid bindings announced by ASes. AS-CRED reputation has three properties: (1) the initial reputation of ASes is set to the best possible value [0,0] (see Section VI-2 for more details on this choice); (2) the reputation value is updated as incidences of poor behaviors are observed; and (3) more recently observed poor behaviors are weighed more heavily than older ones, as it has been observed that a recent poor behavior is usually a precursor to another one.

C. Data Source and System Setup

We use the RouteViews BGP trace collector [19], maintained by University of Oregon, to populate the BGP Activity Manager. At the time of writing, RouteViews directly received BGP updates from 46 ASes. It has been shown in [20] that RouteViews covers almost all the ASes currently active within the Internet and is therefore a good source for computing reputation of ASes. Consequences of using a different trace collector in the operation of AS-CRED will be discussed in Section VIII-A4. For the purposes of this work, we assume the RouteViews repository is trustworthy and provides accurate information.

In this work we present reputation and analysis results using six months worth of BGP data. The first step in this regard is to determine the length of the sliding window also known as *observation window*. The length of the observation window is a function of the reputation function used, and needs to be chosen with some care. It should be sufficiently long to prevent the reputation values from being biased by transient phenomena such as failures, network outages, BGP update fluctuation, and route-flaps. However, keeping it too long is unnecessary as the reputation value is minimally affected by poor behaviors displayed beyond a certain time in the past. In this work, the chosen value is 60 days. A change in the reputation function used may require a re-calibration of the observation window length. Further details on choosing this value will be discussed in Section VI-1.

Our experiments began with the BGP data from Nov. 1, 2009 - Dec. 30, 2009 (see Figure 2). This 60 day period is the initial *observation window*. AS behavior during this period is used to compute the reputation of the ASes on Jan. 1, 2010, leaving a 24-hour grace-period on Dec. 31, 2009. These reputations are then used to generate alerts for the updates received on Jan. 1, 2010. The observation window is then slid forward by one day (Nov. 2, 2009 to Dec. 31 2009) to recompute AS reputations in order to generate alerts for Jan. 2, 2010, and so on. In this manner, we have analyzed behavior, computed reputation and generated alerts on every day from Jan. 1, 2010 to Jun. 30, 2010. Each day alerts are generated is termed as *alert generation window*. We find that recomputing reputations once a day is computationally feasible and provides sufficient predictive power.

IV. Historical Anomaly Detection

Computing reputation for ASes requires feedback on their historical prefix announcements. In this section, we present the stability property used by the Historical Anomaly Detector component to generate feedback for reputation computation.

In the inter-domain routing world, it has been shown that valid AS-prefix bindings last for long durations and are very stable in nature. On the other hand, shorter binding duration implies greater chances of the binding being invalid [12], [21]. Inspired by this results, we first present two heuristics to compute the level of stability of AS-prefix bindings and can therefore can be used to deduce their validity.

TABLE IV
Prevalence, Persistence and Feedback

Prevalence	Persistence	Feedback
high	high	G
high	low	B (Vacillating)
low	high	G
low	low	U (Hijacked)

A. Prevalence and Persistence

Prevalence (Pr) of an AS-prefix binding is the percentage of time a prefix is claimed to be directly reachable by an AS within a time window (the observation window, in our case). More formally:

$$Pr(a,p) = \sum_{i}^{N} (Tw^i(a,p) - To^i(a,p))/T_{obsv} \qquad (2)$$

Here, N is the number of times the prefix p is claimed to be owned by the AS a within the observation window, i is the index of all the announcements of prefix p by AS a during T_{obsv} (the observation window). $Tw(a,p)$ is the time prefix p is withdrawn by AS a. $To(a,p)$ is the time prefix p is the announced by AS a. If the prevalence is above a threshold then the binding is considered stable. However, the prevalence metric alone is not sufficient, as it will not be able to detect repeated short-duration binding instances. We therefore consider another metric in conjunction with prevalence, called persistence. *Persistence* (Ps) of an AS-prefix binding is defined as the average duration of a binding instance in the observation window. More formally:

$$Ps(a,p) = \sum_{i}^{N} (Tw^i(a,p) - To^i(a,p))/N \qquad (3)$$

The symbols have the same meaning as stated earlier. Given the definition of the two heuristics, it is easy to see that relationship between persistence and prevalence for an AS-prefix binding always follows the relation: $Ps(a,p) \leq Pr(a,p) \times T_{obsv} \leq T_{obsv}$. It is important to note that both prevalence and persistence are applied to AS-prefix bindings received over the observation window. The observation window extends well-beyond the day when the AS-prefix binding was first seen. This gives the historical anomaly detection process the ability to observe how an AS-prefix binding evolves after it was first observed.

B. Feedback

In order to map our observations of AS-prefix binding stability into a reputation, we have to classify them into the GBU sets. Table IV shows the classification based on the prevalence and persistence being above or below two *thresholds*, TPr for prevalence and TPs for persistence. These two thresholds are static in AS-CRED and have been set to $TPr = 1\%$ and $TPs = 10$ hours for prevalence and

persistence, respectively. Section VI-3 provides more details on the approach for choosing the values. A value below the threshold is called *low* and one above is called *high*. We now discuss the types of feedback given to ASes for different prevalence and persistence values.

a) Case 1: High Prevalence and High Persistence: If an AS-prefix binding exhibits high prevalence and high persistence, it is stable and classified into the G set. This G set forms the "white-list" called the *Valid Binding List* (*VBL*), containing the latest set of valid AS-prefix bindings. Such a list is not a new notion in the domain of BGP. Approaches for detecting prefix hijacking such as PGBGP [12] and PHAS [13] also create list of AS-prefix bindings they consider valid. However, the lists are of poorer quality because PGBGP only considers persistence of AS-prefix bindings but not prevalence, while PHAS simply chooses a cut-off time and assumes all the AS-prefix bindings before this time to be valid. AS-CRED, on the other hand, computes average persistence and prevalence of AS-prefix bindings over a sliding window, which results in a more adaptive and accurate list.

b) Case 2: Low Prevalence and High Persistence: If the prevalence is low and persistence is high, it means that the particular AS-prefix binding did not recur many times, and while it lasted it did so for a reasonable amount of time. This is consistent with valid temporary bindings (*e.g.*, backup AS taking over while the main AS serving the prefixes is down for maintenance), as noted in [7], are therefore also classified in the G set and consequently becomes part of the VBL.

c) Case 3: Low Prevalence and Low Persistence: Low persistence generally indicates anomalous updates. Malicious ASes that are trying to hijack a prefix typically announce short AS-prefix binding instances in order to avoid detection and engage in nefarious activities such as mounting targeted denial of service attacks [21]. Therefore, we categorize all such *hijacked AS-prefix bindings*, with low prevalence and low persistence in the U set.

d) Case 4: High Prevalence and Low Persistence: The only remaining case is the one where AS-prefix bindings have high prevalence and low persistence. To classify them into the appropriate GBU sets it is essential to understand the implications of this behavior. When the prevalence is high for an AS-prefix binding it means that the overall time within the observation window for which the AS claimed to have a direct path to the prefix was above an acceptable threshold. However, by the same token, a low average persistence value indicates that each time the AS-prefix binding was announced it was withdrawn after a short time duration. Therefore, for an AS-prefix binding to have prevalence higher than TPr but persistence lower than TPs indicates that such AS-prefix bindings have a large number of instances. This indicates a *vacillating ownership of prefixes*, where the claim for ownership occurs many times during the observation window but does not last as long as the persistence threshold. We classify such AS-prefix bindings in the B set. We do so because the prefix involved, upon further analysis (see Section VIII-A1), is found to be owned by the announcing AS. Consequently, this behavior does not go against the intent of BGP in terms of exchanging correct reachability information between ASes. However, such AS-prefix bindings usually last for a very small time duration that makes them impractical to use for routing data. Table II shows some of the prominent cases of vacillating AS-prefix bindings that were observed.

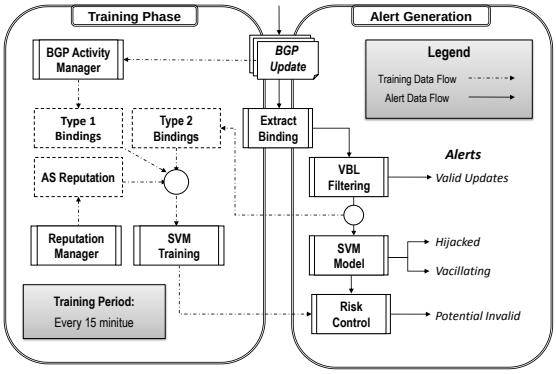


Fig. 3. AS-CRED Alert Generation Process

Note that, vacillating prefixes are independent to the illegal or hijacked nature of prefixes. It is therefore theoretically possible for hijacked prefixes to be vacillating as well. However, we believe that ASes that want to hijack prefixes might rarely want them to vacillate as it will make it difficult for them to attract traffic through a unstable route. Further analysis based on real-world traces also confirms this phenomena (see Section VIII-A1).

C. Feedback Refinement

Given this basic classification of AS-prefix bindings, we now apply a set of refinements to reclassify common mistakes made by ASes while announcing prefixes. Inspired by [12], [7] we use two criteria in this regard:

- *De-aggregation:* An AS y whose binding with prefix p' has been classified in the U set, is reclassified to the G set, if there is an AS-prefix binding $\{y, p\}$ in the G set, such that $p' \subset p$. We do this because the AS in question already has a stable binding with a super-prefix (p). Therefore, there is a high possibility that it owns p' as well. Its announcement does not prevent the expected operation of BGP and merits the re-classification to G set.
- *Stable Owner in the Path:* Suppose $\{n, p, t\}$ is in the G set and an update of the form $\{p, AS\_PATH = \langle a, b, \ldots, n, \ldots, x\rangle\}$ is received at the time t', where x is the announcer. Now if the entry $\{x, p, t'\}$ was originally put in U set, we remove it and ignore the value, the reason being: (1) the short duration of the binding $\{x, p\}$, and (2) the presence of the stable owner n in the *AS_PATH* that can still receive the data traffic directed toward p.

V. Reputation-based Alerts

AS-CRED provides a real-time reputation-based alert service, through the Alert Manager, which flags updates trying to advertise potentially invalid AS-prefix bindings. The alerts along with the reputation values can be used by ASes to make various forms of decisions, from whether to accept updates originating from specific ASes to peering with specific ASes.

We employ a *support vector machine* (SVM) to serve as the alert generation engine of the AS-CRED Alert Manager. We choose SVM because of it is well-understood and has excellent

tool support. Our current implementation uses the libsvm library and chooses the radial basis function kernel [22]. The SVM is trained every fifteen minutes (can be changed if needed) based on two types of AS-prefix bindings: *Type 1*: a set of 5000 AS-prefix bindings sampled from all bindings announced during the last ten days of the observation window, and *Type 2:* the AS-prefix bindings received from the last fifteen minutes that are not contained in the VBL. We chose these sources as they provide a reasonably long historical view of the quotidian BGP operation and capture the behavior pattern of sporadic network events, such as network outage. This makes the alert service adaptive. For every AS-prefix binding in the training set, we use the reputation value of the AS as the feature and its GBU feedback as the label. Type 2 AS-bindings are labeled in the following manner: (1) since such bindings are not contained in VBL, we give them label *U* by default; and (2) some of the bindings are then re-labeled by applying the refinements described in Section IV-C. Once the training is complete, the SVM model is used in the alert generation process. The SVM re-training and the alert generation overhead are minimal — a commodity machine can easily handle the process for real-time BGP updates.

Figure 3 illustrates the alert generation process. In AS-CRED, the alerts are generated based on a combination of the *VBL* filtering and reputation-based labeling. The alert service is tiered in the types of alerts generated. This is specifically designed to tackle the complex dynamics of BGP operation. We believe that existing alert systems that produce binary alerts of goodness or badness of updates are inherently incapable of capturing this complexity [12]. Overall, the alert generation process works as follows:

- *VBL Filtering:* When a new update is received, the Alert Manager first check to see if its corresponding AS-prefix binding $\{a, p\}$ is in the *VBL*. If so, it is considered to be valid and no alerts are generated. Otherwise, the binding is called *non-VBL* binding.
- *Invalidity Labeling:* For a non-VBL binding, the Alert Manager then fetches $Rep_B(a)$ and $Rep_U(a)$ for the announcing AS a as the feature and feeds it into the trained SVM model. The model then predicts the alert label "Vacillating" or "Hijacked" for the bindings within the update. (The bindings which are labeled "Vacillating" are added to a set called VT and the ones labeled "Hijacked" are added to the HJ for accuracy analysis. The results of the analysis are discussed in Section VIII-B).
- *Potential Invalidity Labeling:* Within an alert generation window, if an reputable AS announces more than $T_{TrustLimit}$ number of non-VBL bindings without triggering "Vacillating" or "Hijacked" alerts, the Alert Manager generates "Potentially Invalid" alerts for all the updates that contain such non-VBL bindings (including the ones previously deemed valid).

The *Potential Invalidity Labeling* is designed to tackle the dynamic nature of AS behavior, where highly reputable ASes may start exhibiting poor behaviors. In AS-CRED, the risk associated with blindly trusting reputable ASes is controlled by introducing a trust upper bound: $T_{TrustLimit}$. $T_{TrustLimit}$ essentially specifies the maximum number of false negatives that can be tolerated within an alert generation window (*i.e.,* 24 hours). Therefore, the value of $T_{TrustLimit}$ can be conservatively set to *zero* to eliminate such risk. However, this will prevent benign ASes from announcing valid new AS-prefix bindings. The value needs to be set low enough to allow ASes to announce new prefix bindings while minimizing the false negatives. Further discussions selecting the threshold value are presented in Section VI-4.

VI. AS-CRED Parameter Selection

As AS-CRED deals with the complex inter-domain routing infrastructure whose dynamics change over time. We have therefore designed it to be tunable. One can adaptively select new values for its various parameters as the AS behavior evolves. In this section, we share our experiences in choosing appropriate values for a list of parameters used in AS-CRED, namely: the half-life values in the reputation function, the length of observation window, the choice of default reputation values, the stability threshold, and the $T_{TrustLimit}$ threshold for alert generation. Note that, in this discussion, we provide one possible approach for selecting these parameters, which has yielded satisfying results as seen in Section VIII. Other, more involved, approaches can easily be utilized for tuning AS-CRED as required.

1) Half-life Values & Observation Window Length: Based on the historical anomaly detection we found that over 75% of the ASes within the B and the U sets reappear within 3 and 6 days, respectively. This demonstrates that AS behavior is repetitive. This observation allows us to set the half-life values, used in Section III-B2, to be $h_U = 6$ days for Rep_U and $h_B = 3$ days for the Rep_B. The half-life values enable us to worsen the reputation of the ASes that repeat their poor behavior frequently. Conversely, ASes that seldom repeat their poor behavior will not be penalized as much.

The length of the observation window is determined by the time-decay property of the reputation function and the repetitiveness of the AS behavior pattern. Given these half-life values, after 60 days an instance of invalid AS-prefix binding will contribute only 2^{-10} (for Rep_B) and 2^{-20} (for Rep_U) to the reputation. Therefore, our 60 day observation window is adequate for this work.

2) Default Reputation Value: In typical reputation systems, one would prefer to assign newcomers with a rather low reputation value. This design choice is often made to mitigate the possibility of Sybil attacks (*i.e.,* malicious entities creating multiple new identities). However for BGP, the threshold to entry is sufficiently high to prevent this situation. Moreover, the creation of new ASes are relatively rare events, and sufficient historical information is often available for the AS reputation computation. As a result, we currently choose to use [0,0] (*i.e.,* the best reputation) as the initial reputation for ASes. However, this design choice can be changed if the above observations change.

3) Stability Threshold: To select the thresholds for the stability analysis of AS-prefix bindings, we compute the prevalence and persistence for each AS-prefix binding observed within the observation window. We then assign the TPr (*i.e.,* threshold for prevalence) and TPs (*i.e.,* threshold for persistence) to specific values and classify the bindings into the

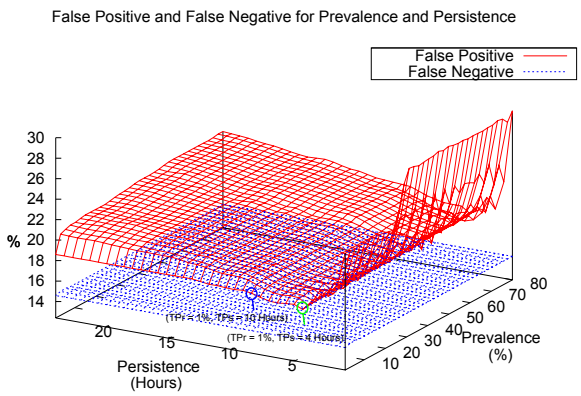


Fig. 4. Setting TPs and TPr Threshold Values

GBU sets. For the TPr and TPs pair, we compare the entries in the GBU sets with Internet Routing Registries (IRR), and compute the false positives (FPs). We repeated the previous two steps, by varying the TPr and TPs pair, until we find the values that minimize FPs. We choose IRR because: (1) an average of 87% of the IP prefixes announced through updates were found in the IRR, which makes IRR largely complete; (2) more than 70% of AS-prefix bindings on average had a matching record in IRR, and (3) there is a lack of other authoritative source of the same nature.

Figure 4 shows our analysis using IRR for different TPr and TPs for updates received between Nov. 1, 2009 to Dec. 30, 2009. Notice that we do not have to consider false negatives (FNs) in identifying the threshold values as it falls entirely under the FP surface. We find that the lowest FP value is obtained at $TPr = 1\%$ and $TPs = 4$ hours. However, for this work, we chose the values $TPr = 1\%$ and $TPs = 10$ hours as the thresholds. The decision is based on three factors: (1) the value of 10 hours allows us to capture 95% of the poor behaviors as suggested in [7], (2) the difference between the FPs at the two points was less than one percent (17.7% to 18.4%), and (3) a TPs of 10 hours prevents an AS from sustaining an unowned prefix announcement long enough to avoid detection. Note that, this FP value should not be seen as a true representation of the Historical Anomaly Detector's capabilities. This is because IRR, which forms the basis of this value, is an imperfect ground truth.

4) Potential Invalidity Threshold: Threshold $T_{TrustLimit}$ is used for triggering "Potential Invalid" alerts (see Section V). Our strategy for selecting this threshold value is to find a value: (1) that is low enough to bound the potential risk associated with accepting anomalous updates from reputable ASes; and (2) that prevents a considerable portion of new and valid bindings announced by reputable ASes from raising alarms. Please note that alternative strategies are also possible depending upon the risk objectives. To this end, we studied the AS-prefix bindings announced by reputable ASes, which were not in the VBL. Within an alert generation window (*i.e.,* 24 hour period), around 400 ASes were found to be involved in such behavior and each AS announces four such bindings, on average. Moreover, about 50% of these ASes announce one or less of such bindings per day, 70% announce two or less, 90% announce ten or less. According to this observation, we choose the $T_{TrustLimit}$ value to be 2 bindings/24 hours in the current operation of AS-CRED.

VII. Security Analysis of AS-CRED

AS-CRED provides reputation values that can be used by ASes to be cognizant of the behavior of others. Consequently, it exposes itself to attacks that try to: (1) promote ASes, or (2) defame ASes. In this section, we analyze AS-CRED's resilience to such attacks.

Self Promotion: An AS may want to improve its current reputation in order to minimize the chances of triggering an alert while announcing invalid bindings.

Self-promotion is not possible in AS-CRED. The underlying assumption of AS-CRED is that no amount of good deeds can redeem poor behavior. As a result, the reputation function is designed to only consider poor behavior. The only way reputation can be improved is to wait and let the time-decay function heal the reputation (see Section III-B2). Furthermore, a "healed" reputation value will not give an AS any substantial benefit since: (1) reputable ASes are given only limited trust (see Section V); and (2) AS-CRED provides not just current AS reputation values but also the past reputation trends. Users of AS-CRED can take this information into account and make an informed decision about the trustworthiness of an AS, irrespective of its current reputation value.

Slandering: In contrast to the self promotion attack, an attacker AS may attempt to slander other innocent ASes, by announcing low persistence AS-prefix bindings in their name, and try to damage their reputation.

This is equivalent to performing "identity theft" in our setting. Such attacks can usually be mitigated by cryptographic approaches such as S-BGP [8]. However, given their complexity, such adoption has been rather slow. Interestingly, this is a problem faced by all schemes that deal with BGP update semantics such as PGBGP [12] and PHAS [13]. In the future, we plan to enhance AS-CRED with other data-plane probing techniques such as [14], which can potentially build fingerprints of ASes. Such techniques can aid enormously in automated, real-time slander mitigation.

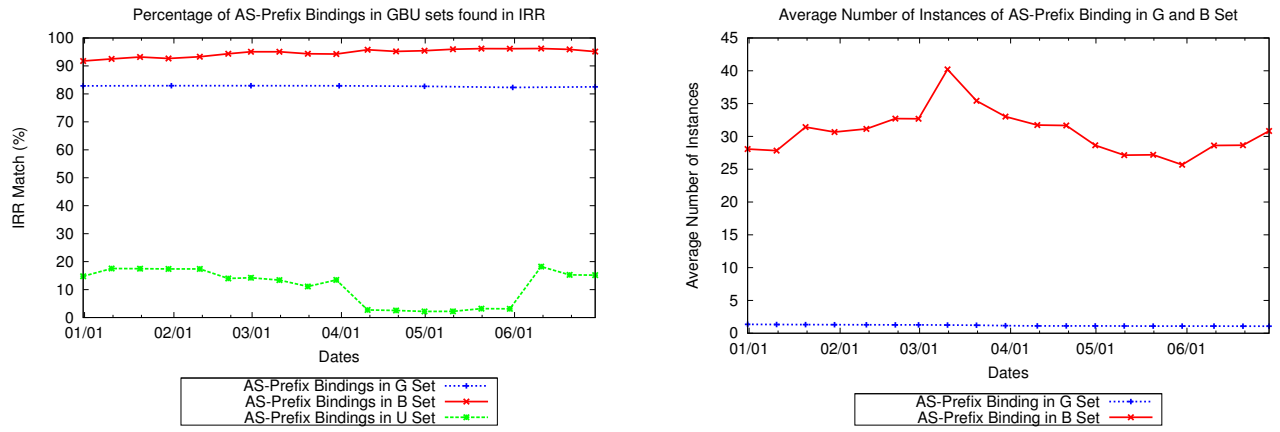
VIII. Performance Results

In this section, we present the performance results of the AS-CRED service in terms of both its historical anomaly detection capabilities, and the effectiveness of its alert system. The results have been obtained by setting the AS-CRED parameters to the values described in Section VI.

A. Historical Anomaly Detection Analysis

Here, we demonstrate: (1) AS-CRED's historical anomaly detection is accurate, (2) ASes repeat their poor behavior, and (3) the reputation values of ASes are representative of their anomalous behavior. Together, they illustrate that AS reputation is both past-representative and future-predictive, which forms an ideal metric for triggering alerts for any subsequent anomalous updates.

1) Accuracy of Anomaly Detection: The reputation of an AS depends upon the GBU feedback provided by the historical anomaly detection mechanism used to identify invalid AS-prefix bindings of the past. It is therefore necessary to ensure that AS-prefix bindings in the GBU sets are there for the correct reason. We demonstrate this based on the satisfaction



(a) (b)

Fig. 5. (a) Validity of AS-prefix bindings in GBU sets, (b) Number of Announcements and Withdrawals of vacillating AS-prefix bindings in B set compared to AS-prefix bindings in G set

of the following two tests: (1) an evaluation of the *validity* of the AS-prefix bindings in the G and the B sets compared to the U set, and (2) an evaluation of the *stability* of the AS-prefix bindings in the B set compared to the G set. Together the two tests demonstrate that the historical anomaly detection quality of AS-CRED is satisfactory.

a) Validity: The semantics of the bindings in the U set is that the AS does not own the prefix. On the other hand, bindings in the G and the B sets, contain prefixes owned by the corresponding AS. To validate this, we use the IRR to check whether the AS-prefix binding in the GBU sets match the documented prefix ownership information. Figure 5(a) shows the percentage of AS-prefix bindings in the G set, vacillating bindings in the B set and hijacked ones in the U set that have a match in IRR. It can be seen that AS-prefix bindings in the G and the B sets can overwhelmingly be found in IRR, compared to those in the U set. This shows that AS-prefix bindings that are classified as hijacked are usually accurate. In Figure 5(a), the percentage of IRR matches for B is higher G. This is because $|B| << |G|$, which makes the difference in the match percentages less statistically significant.

b) Stability: The semantics of an AS-prefix binding in the B set is that they are vacillating[4]. Figure 5(b) charts the average number of instances of binding establishment and withdrawal seen for entries in the G set and the vacillating entries in the B set on a selected set of dates. Overall, the vacillating AS-prefix bindings were established and terminated on average 30 times more often (with a maximum of 4492) than AS-prefix bindings classified in the G set, where the average number was close to one. The results demonstrate that vacillating AS-prefix bindings are distinct from those in the G set, given that their quantity is an order of magnitude larger.

2) Anomalous Behavior Trends and Repetitiveness: In this subsection, we summarize the results of the historical anomaly detection and show that ASes repeat their behaviors. Figure 6(a) shows the summary of the historical anomaly detection over the six months of AS-CRED's operation. We find that on average over 35K unique ASes were observed, out of which only 5% (about 1740) of the ASes were found to display poor behaviors. Only about 0.2% (about 70) of the ASes displayed exclusively poor behaviors for all prefixes they announce. Overall, 421K AS-prefix bindings were observed, out of which about 10.9% were classified as displaying poor behaviors.

An interesting piece of information that can be discerned from the historical anomaly detection summary is the extent to which poor behaviors afflict the inter-domain routing world. Figure 6(b) shows, with the benefit of hindsight, how many of the AS-prefix bindings seen every day from Jan. 1, 2010 to Jun. 30, 2010 eventually turned out to be hijacked or vacillating. It can be seen that AS-prefix bindings that eventually turn out to be vacillating are an order of magnitude greater in number than hijacked AS-prefix bindings or those with illegal AS numbers. However, there are some clear spikes in the case of the latter. For example, the spike on April 8th, 2010 is the due to AS23734's Internet-scale hijacking attempt [23]. Overall, announcement of poor AS-prefix bindings seems to be consistently present and their magnitude, barring occasional jitters, is largely even.

3) AS Reputation Trends: With the data analyzed and feedback obtained in the form of the GBU sets, we can now compute the reputation of the ASes. Figure 7 shows the count of ASes that exhibited poor behaviors over the six months of AS-CRED operation. It can be seen that number of such ASes remains more or less the same over the entire period. Further, around 90% of such ASes have a reputation value between zero and one. This is significant because it demonstrates three things: (1) the reputation value properly characterizes the ASes in terms of the historical anomalies they exhibit, (2) even among the ASes that have exhibit anomalies, an overwhelming majority do so rarely[5], and (3) AS-CRED is sensitive enough to capture even those ASes that rarely exhibit anomalies.

4) Reputation and Alternate Data Sources: Due to the distributed nature of the Internet, it is very difficult to obtain a complete knowledge of it. The RouteViews data provides only a partial view of the information exchanged at the inter-domain level of the Internet. We therefore investigate the consistency

[4]We exclude updates announced by BGP Beacons (used for studying BGP dynamics (http://www.psg.com/?zmao/BGPBeacon.html)), as they often display similar characteristics.

[5]Repeat offenders sometimes have reputation in the thousands.

Historical Anomaly Detection Statistics from Jan. 1, 2010 to Jun. 30, 2010

(For each day, the analysis considers BGP updates from the past 60 days)

Property	Value
Avg. # of Prefixes	400410.7
Avg. # of SOAS Prefixes	379937
Avg. # of MOAS Prefixes	20473.6

Prefix Statistics

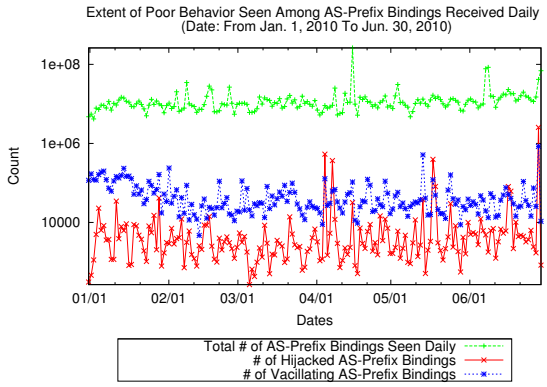
Property	Value
Avg. # of AS-Prefix Bindings	421704.1
Avg. # of Hijacked Bindings	16882.1
Avg. # of Vacillating Bindings	29256.6

AS-Prefix Binding Statistics

Property	Value
Avg. # of AS Observed	35448.2
Avg. # of AS Announcing Vacillating AS-Prefix Bindings	1132.8
Avg. # of AS Announcing Hijacked AS-Prefix Bindings	605.4
Avg. # of AS Exclusively Announcing Vacillating AS-Prefix Bindings	17.81
Avg. # of AS Exclusively Announcing Hijacked AS-Prefix Bindings	54.3

AS Statistics

(a)



(b)

Fig. 6. (a) Historical Anomaly Detection Statistics, (b) Extent of Poor Behaviors displayed

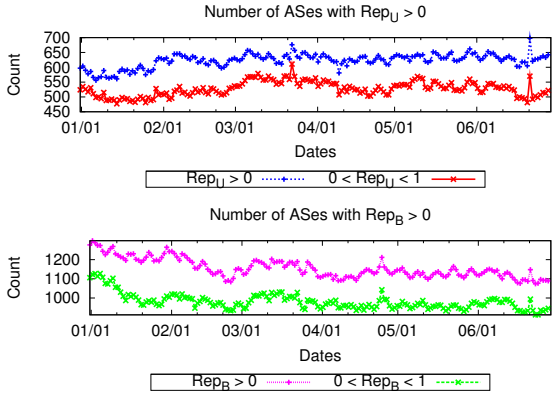


Fig. 7. Trend demonstrating number of ASes with $0 < Rep_U \leq 1$ and $0 < Rep_B \leq 1$, from Jan. 1, 2010 - Jun. 30, 2010

of AS reputation computed using BGP updates collected from different locations. To this end, we compared the AS reputation computed based on RouteViews with another trace collector maintained by Reseaux IP Europeens (RIPE) called RIS [24]. We observed that almost all ASes have identical reputation values, except only about 0.09% of the ASes have an absolute difference of $Rep_U \geq 1$. For Rep_B the percentage is 0.50%. Such differences are mainly due to the fact that certain AS-prefix bindings are observed at one data source but not the others. In the future, we plan to further improve the quantity and location diversity of the data sources used by AS-CRED to improve its coverage.

B. AS-CRED Alert Analysis

In order to evaluate the correctness of alert generation process, each time the alert type is re-labeled to "Hijacked" or "Vacillating", the associated AS and prefix are added to either HJ or VT sets, respectively. Figure 8 shows the percentage of updates triggering alerts during the six months of alert generation. As seen during historical anomaly detection, we find that the number of alerts generated for updates with vacillating bindings in VT set are an order of magnitude greater than those in the HJ set or updates having illegal AS numbers. In the rest of the section, we analyze the correctness and errors of the alerts generated.

1) Alert Accuracy Analysis: To evaluate the correctness of the "Hijacked" alerts generated by AS-CRED, we compare it with an alternative alert system called the Internet Alert Registry (IAR). IAR is a well-known historical information-based prefix hijack alert system. It is based on Pretty Good BGP (PGBGP) [12]. IAR identifies suspicious AS-prefix bindings by consulting a trusted list, learned from the recent history of BGP updates. Initially, the trusted list is empty. All bindings received during the next 10 days are added to the trusted list. After this initial phase, any new bindings not present in the trusted list are quarantined for 24 hours. If the bindings have not been withdrawn at this time, they are added to the trusted list. IAR triggers alert for all newly observed AS-prefix bindings not in the trusted list [12]. We use IAR for our comparison study because: (1) it is one of the few systems that provides the latest prefix hijacking alerts, and (2) it has been operational during the time-frame when we collected our data.

For the purposes of this study, we use the IRR to provide a common basis for comparison. The metric for comparing the AS-CRED and IAR is *error* — the percentage of AS-prefix bindings with a matching record in IRR. We do not perform a more elaborate false positive, false negative based analysis because: (1) the IAR database only provides information about the AS-prefix bindings it considers hijacks, and (2) the BGP updates seen by the IAR system might be different from AS-CRED. For AS-CRED we find that the average error rate (12.8%) is about fives times smaller than IAR (66.3%). This result shows that the percentage of false alerts generated by AS-CRED are much lower than IAR. The false negative rates of the two systems are hard to compare because of the lack of availability of associated IAR data. However, we reiterate that risk of potential false negatives can be controlled in AS-CRED by choosing appropriate $T_{TrustLimit}$ threshold. We plan to conduct more extensive studies in this regard, as the necessary dataset becomes available.

Evaluating the correctness of the updates classified as "Vacillating" in the VT set was slightly different. As we do not have a ground truth available to check for the correctness of the classification, we depend upon behavior analysis that considers "future" BGP updates. In this regard, we make use of a *consistency verification window*. The idea is to allow sufficient time for the alert-triggering AS-prefix bindings to evolve in order to be analyzable with the benefit of hindsight. The consistency verification window is 60 days long and centered around each alert generation window. For example,

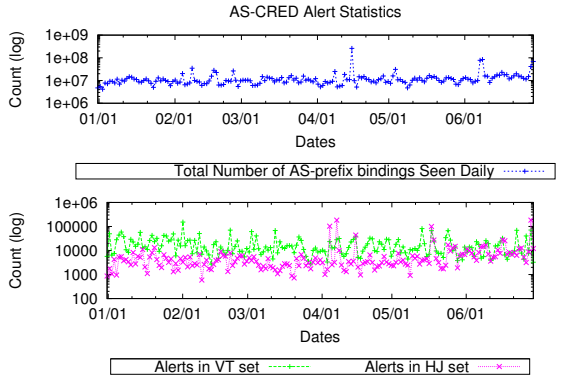


Fig. 8. Alert Generation Statistics

the consistency verification window for alerts generated on Jan. 1, 2010 will span from Dec. 2, 2009 to Jan. 30, 2010 (see Figure 2). The bottom graph of Figure 8(b) shows the *error* results for entries in the VT set. We find that our prediction of an AS-prefix binding to be vacillating is correct around 95% of the time.

2) Alert Error Analysis: The alerts generated by AS-CRED can not be absolutely accurate as we saw in the previous section. Therefore, for each AS-prefix binding in the VT and HJ set we determine its true designation with hindsight, and investigate the reasons for the discrepancy, if any. In this regard, we analyze the behavior of the AS-prefix bindings in the VT and the HJ sets over the *consistency verification window* (see Table V). We find 94.7% of AS-prefix bindings in the VT set eventually turned out to be classified correctly. Out of the remaining AS-prefix bindings, which erroneously triggered an alert for being vacillating, 3.9% turned out to be eventually in the G set with only 1.4% percent being in the U set (*i.e.*, hijacked). Such a small discrepancy is because of mixed behaviors of ASes. That is, ASes that consistently announce vacillating prefixes, do announce valid or hijacked bindings, once in a while. However, given the very small percentage of the misclassification, we believe that AS behavior remains largely repetitive, allowing reputation to be a good metric for triggering alerts. Similarly, 87.4% AS-prefix bindings in the HJ set eventually turned out to be classified correctly. Of the remaining incorrectly classified AS-prefix bindings, 3.7% turned to be in the B set (*i.e.*, vacillating), with 8.9% turning out to be eventually in the G set. This again demonstrates the largely stable nature of AS behavior, with occasional discrepancies. The AS-prefix bindings that trigger alerts for being hijacked are more error prone because we err on the side of caution and tune the SVM to generate higher number of alerts, sacrificing some correctness in the process. We take such a punitive stance because hijacked AS-prefix bindings, if not detected, have the potential to be disruptive.

IX. Related Work

Recent years have seen considerable number of works in anomaly detection and prevention for the inter-domain routing system. In this section we describe the prominent research in this area.

Anomaly Prevention Mechanisms. S-BGP [8] is one of the earliest and the most concrete security mechanism to address BGP vulnerabilities. It constructs PKIs rooted at RIRs for authenticating IP prefix ownership. *AS_PATH* information is also protected using nested digital signatures as the BGP updates propagate through the network. However, the deployment difficulties and computational overhead of these schemes have made their adoption cumbersome in the inter-domain world. To overcome some of these issues, a more incrementally deployable scheme called So-BGP [25] has been proposed. So-BGP adoptes a more flexible trust model to achieve AS public key authentication. However, it has limited capability to ensure the correctness of *AS_PATH*. Further trade-offs have been proposed in psBGP [26], which a low cost scheme to verify the validity of the prefix origins with neighbor ASes using a prefix assertion list. A prefix ownership assertion made by an AS is valid if it is consistent with assertions made by one of its peers. However, colluding ASes can still forge origin information under this scheme. In [27], the authors formalize the semantics of address delegation and design strategies for reducing resource costs associated with existing origin authentication schemes.

TABLE V
AS-CRED Alert Error Classification Summary (with benefit of hindsight)

Classification	Percentage
HJ set entries classified as *Good*	8.9%
HJ set entries classified as *Vacillating*	3.7%
HJ set entries classified as *Hijacked*	87.4%
VT set entries classified as *Good*	3.9%
VT set entries classified as *Vacillating*	94.7%
VT set entries classified as *Hijacked*	1.4%

Anomaly Detection. Detecting attacks on the BGP routing infrastructure has received its own share of attention. Many of these schemes use data-plane probing where an AS, on suspecting an update to be an attempted hijack, probes the announcer to verify its suspicion [14], [16], [28]. Although they achieve reasonably high detection accuracy, some of these approaches can only be leveraged by the victim originator AS during the attack phase. Therefore, such approach will have limited global impacts without a full network deployment. Another approach is to analyze historical control-plane information for detecting any subsequent problematic updates [13]. The recent proposal of Pretty Good BGP (PGBGP) [12] uses this approach to delay the selection of suspicious routes. However, as demonstrated in our evaluation with real world traces, it suffers from high error rates. Moreover, the focus of all these approaches is limited to detecting instances of prefix hijacking. None of these approaches study vacillating bindings as AS-CRED does nor provide a quantitative way to analyze and understand AS behavior itself. Instead of focusing on proposing concrete detection mechanism, [29] focuses on accurately locating the attacker for a prefix hijacking incident through the active monitoring of routes changes. This work compliments existing detection mechanisms by pin-pointing the root-cause of anomalous route changes. In [30], the authors study the strategies of utilizing existing protection and detection mechanisms to achieve effective and feasible solutions for dealing with prefix hijacking in the real-world. However, the solutions used require the presence of detection agents in impacted ASes which is an assumption AS-CRED does not make.

Reputation Schemes. In [31], the authors use the notion of reputation for accepting or rejecting updates based on a trusted overlay network over the existing AS topology. Once such an overlay is set up, if a node wants to determine the accuracy of an update with respect to prefix hijacking and AS path spoofing, then it can simply query its neighbors in the overlay network. Similarly, in [32], the authors present a reputation system for ASes with a focus on preventing propagation of bogus routing information. However, their mechanism also depends on computing reputation based on an alliance of ASes. As AS-CRED does not depend on inputs from other ASes to compute reputation, it avoids complications or inaccuracies relating to possibly biased feedback. In [1], the authors present an AS reputation scheme that has probabilistic interpretation. Unlike [1], the reputation value computed by AS-CRED is independent of the good behavior an AS exhibits. In other words, [1] presents a complementary view to the reputation scheme used in AS-CRED.

X. Conclusions

In this paper we presented *AS-CRED*, an AS reputation and alert service that not only detects anomalous BGP updates but also provides a quantitative view of AS behavior. AS-CRED works by computing AS reputation based on feedback provided by analyzing the historical BGP data for the presence of anomalies (*i.e.,* hijacked or vacillating). Based on this analysis, AS-CRED also creates a "white-list" of valid AS-prefix bindings. The reputation and "white-list" are combined to design a novel tiered alert system for tracking subsequent anomalous updates. We publish the AS reputation information on a publicly available portal website (`http://rtg.cis.upenn.edu/qtm/ascred/`). The analysis of AS-CRED over a six month period indicates its effectiveness and improvement of over similar alert systems, a fact also demonstrated by its ability to successfully detect large scale hijack events [23]. In the future, we would like to construct more descriptive AS behaviors, and use the resulting AS reputation information to predict the likely amount of invalid BGP behaviors that are going to be exhibited at any given time in the future.

References

[1] J. Chang, K. Venkatasubramanian, A. G. West, S. Kannan, B. T. Loo, O. Sokolsky, and I. Lee, "AS-TRUST :A trust characterization scheme for autonomous systems BGP," in *4th International Conference on Trust and Trustworthy Computing (TRUST 2011)*, Pittsburgh, PA, June 2011.

[2] H. Kim and J. Huh, "Detecting dns-poisoning-based phishing attacks from their network performance characteristics," *Electronics Letters*, vol. 47, no. 11, pp. 656 –658, 26 2011.

[3] S. Abu-Nimeh and S. Nair, "Circumventing security toolbars and phishing filters via rogue wireless access points," *Wireless Communications and Mobile Computing*, vol. 10, no. 8, pp. 1128–1139, 2010. [Online]. Available: http://dx.doi.org/10.1002/wcm.829

[4] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," *SIGCOMM Computation and Communication Review*, vol. 36, no. 4, pp. 291–302, 2006.

[5] "Pakistan hijacks YouTube," http://www.renesys.com/blog/2008/02/pakistan_hijacks_youtube_1.shtml/.

[6] "7007 Explanation and Apology," http://www.merit.edu/mail.archives/nanog/1997-04/msg00444.html/.

[7] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," in *Proc. of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, 2002, pp. 3–16.

[8] S. Kent, C. Lynn, and K. Seo, "Secure Border Gateway Protocol (S-BGP)," *IEEE Journal On Selected Areas Communications*, vol. 18, no. 4, pp. 582–592, April 2000.

[9] M. Zhao, S. Smith, and D. Nicol, "The performance impact of BGP security," *Network, IEEE*, vol. 19, no. 6, pp. 42 – 48, nov.-dec. 2005.

[10] N. K. Khan and G. K. Gupta, "Deployment issues of S-BGP, soBGP and psBGP: A comparative analysis," *International Journal of Advances in Engineering and Technology*, vol. 1, no. 4, pp. 236–243, 2011.

[11] G. Goodell, W. Aiello, T. Griffin, J. Ioannidis, P. McDaniel, and A. Rubin, "Working around BGP: An incremental approach to improving security and accuracy of interdomain routing," in *Proc. of theNetwork and Distributed Systems Security 2003*. San Diego, CA, USA: Internet Society, February 2003.

[12] J. Karlin, S. Forrest, and J. Rexford, "Autonomous security for autonomous systems," *Comput. Netw.*, vol. 52, no. 15, pp. 2908–2923, 2008.

[13] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in *Proc. of the 15th conference on USENIX Security Symposium*. Berkeley, CA, USA: USENIX Association, 2006.

[14] X. Hu and Z. M. Mao, "Accurate Real-time identification of IP prefix hijacking," in *SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 3–17.

[15] Z. Zhang, Y. Zhang, Y. C. Hu, and Z. M. Mao, "Practical defenses against BGP prefix hijacking," in *CoNEXT '07: Proceedings of the 2007 ACM CoNEXT conference*, 2007, pp. 1–12.

[16] Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush, "Ispy: detecting IP prefix hijacking on my own," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 327–338, 2008.

[17] D. Pei, W. Aiello, A. Gilbert, and P. McDaniel, "Origin disturbances in bgp," AT&T Labs - Research, Florham Park, NJ, Tech. Rep. TD-62TJJF8, July 2004.

[18] "A Border Gateway Protocol 4 (BGP-4) RFC," http://www.rfc-editor.org/rfc/rfc4271.txt.

[19] "RouteViews," http://www.routeviews.org/.

[20] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level topology," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 1, pp. 53–61, 2005.

[21] P. Boothe, J. Hiebert, and R. Bush, "Short-lived prefix hijacking on the Internet," in *Proc. of the NANOG 36*, February 2006.

[22] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[23] "Chinese ISP hijacks the Internet," http://bgpmon.net/blog/?p=282.

[24] "RIPE RIS," http://www.ripe.net/ris/.

[25] J. Ng, "Extensions to BGP to support secure origin BGP (soBGP)," in *Expired Internet draft draft-ng-sobgp-bgp-extensions-02*, April 2004.

[26] T. Wan, E. Kranakis, and P. C. Oorschot, "Pretty secure bgp (psBGP)," in *In The 12th Annual Network and Distributed System Security Symposium (NDSSI 05*, 2005.

[27] W. Aiello, J. Ioannidis, and P. McDaniel, "Origin authentication in interdomain routing," in *Proceedings of the 10th ACM conference on Computer and communications security*, ser. CCS '03. New York, NY, USA: ACM, 2003, pp. 165–178. [Online]. Available: http://doi.acm.org/10.1145/948109.948133

[28] C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis, "A light-weight distributed scheme for detecting IP prefix hijacks real-time," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 277–288, 2007.

[29] T. Qiu, L. Ji, D. Pei, J. Wang, J. Xu, and H. Ballani, "Locating prefix hijackers using LOCK," in *18th USENIX Security Symposium*, Aug. 2009.

[30] T. Qiu, L. Ji, D. Pei, J. Wang, and J. Xu, "Towerdefense: Deployment strategies for battling against IP prefix hijacking," in *Proceedings of the The 18th IEEE International Conference on Network Protocols*, ser. ICNP '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 134–143. [Online]. Available: http://dx.doi.org/10.1109/ICNP.2010.5762762

[31] H. Yu, J. Rexford, and E. Felten, "A distributed reputation approach to cooperative Internet routing protection," in *Secure Network Protocols, 2005. (NPSec). 1st IEEE ICNP Workshop on*, Nov. 2005, pp. 73–78.

[32] N. Hu, P. Zhu, and P. Zou, "Reputation mechanism for inter-domain routing security management," in *Proc. of the 9th International Conference on Computer and Information Technology*, October 2009, pp. 98–103.