Improving IP Prefix Hijacking Detection by Tracing Hijack Fingerprints and Verifying Them through RIR Databases

Hussain Alshamrani and Bogdan Ghita
Centre for Security, Communications and Network Research (CSCAN), Plymouth University, Plymouth, U.K.

Keywords: RIPE Database, ASNs and IP Prefix Delegators, Information Correlation, False Positives.

Abstract: In spite of significant on-going research, the Border Gateway Protocol (BGP) still encompasses conceptual vulnerability issues regarding impersonating the ownership of IP prefixes for ASes (Autonomous Systems). In this context, a number of research studies focused on securing BGP through historical-based and statistical-based behavioural models. This paper improves the earlier IP prefix hijack detection method presented in (Alshamrani et al. 2015) by identifying false positives showing up due to the organisations that may use multiple ASNs (Autonomous System Numbers) to advertise their routes. To solve this issue, we link a Verification Database to the previously proposed detection method to improve the accuracy. The method extracts the organisation names (unique code) and associated ASNs from different ASN delegators and RIRs (Regional Internet Registries), more specifically the RIPE (Reseaux IP Europeans) dump database (John Stamatakis 2014) in order to evaluate the method. Since the organisation name is not available in the BGP updates, the data are extracted and processed to produce a structured database (Verification DB). The algorithm excludes false positive IP prefix hijack detection events in the SFL (Suspicious Findings List) introduced in (Alshamrani et al. 2015). Finally, the algorithm is validated using the 2008 YouTube Pakistan hijack event and the Con-Edison hijack (2006); the analysis demonstrates that the improved algorithm qualitatively increases the accuracy of detecting the IP prefix hijacks, specifically reducing the false positives.

1 INTRODUCTION

BGP remains the protocol of choice for core Internet interconnectivity. Although a number of BGP security issues have been identified for almost two decades, the protocol remains vulnerable to IP prefix attack (Goldberg 2014). These weaknesses cause serious attacks and open the door for attacker to perform spam attack (Schlamp et al. 2015), traffic interception and DDoS (Vervier et al. 2015). On Oct, 2014 Sharon Goldberg pointed out that the main reasons why BGP is taking so long to be secured is that, apart from the fact that the BGP security solutions are not deployable, BGP lacks a single centralised authority, each organisation deploys its own routing security solution autonomously, so a complete or mass deployment is unlikely to take place (Goldberg 2014).

Previous studies tried to detect IP prefix hijacks based on monitoring routers' stability, but their methods could not reliably distinguish IP prefix hijacks from normal events, such as power cut-off and submarine cuts (de Urbina Cazenave et al. 2011). In addition, RPKI (Resource Publication Infrastructure) was put forward to detect BGP route hijacking, but the system had several false positives and negatives and need further refinements (Wählisch et al. 2012).

Lastly, some methods propose analysing the routing tables in order to detect IP prefix hijacks, but they are likely to have a limited impact, as organisations may refuse to provide their routing tables or are unable to timely detect a hijack event (Cao et al. 2009). In addition, previous solutions do not support collaboration among routers to detect the IP prefix hijacks, collaboration could limit the attack spreading out and affecting a large number of networks.

This paper aims to address the false positives caused by the limitations of the algorithm in (Alshamrani et al. 2015). After investigating the main reasons we found that route aggregation and AS confederation or reflection BGP operations do not affect the accuracy of the IP prefix detection algorithm proposed previously (Alshamrani et al. 2015) although they have a direct effect on the routes.

Alshamrani, H. and Ghita, B.
Improving IP Prefix Hijacking Detection by Tracing Hijack Fingerprints and Verifying Them through RIR Databases.
DOI: 10.5220/0005934200570063
In *Proceedings of the 13th International Joint Conference on e-Business and Telecommunications (ICETE 2016) - Volume 1: DCNET*, pages 57-63
ISBN: 978-989-758-196-0

One of the factors affecting the algorithm is that big organisation can announce their routes with multiple different ASNs; to counteract this issue, a novel combination of RIRs and ASNs delegation database and BGP updates (Meyer n.d.) is proposed in order to accurately and timely detect IP prefix hijacking events.

In section 2, the paper discusses the previous detection method and the limitations of its algorithm. Section 3 shows the creation of the Verification DB based on the RIPE database. In Section 4 we describe the proposed improvements to the IP prefix detection method based on the information from the Verification DB, together with findings and algorithm challenges. Section 5 describes the collaboration between routers to detect the IP prefix hijacks before it spreads out. Section 6 proposes a general structure of the detection method to be linked with the BGP routers so it can work efficiently. The paper finishes with the conclusion and future work in section 7.

2 PREVIOUS DETECTION METHOD

The detection method presented in (Alshamrani et al. 2015) consists of three main parts: pre-processing, analysis and the algorithm, as shown in Figure 1. This section shows that the algorithm did not have a mechanism to validate the output. It makes decision directly and displays the result either benign or malicious.

The next two subsections explain that by providing an overview of the algorithm functionality and highlight its limitations, specifically the shortcomings that we aim to improve in this paper. BGPdump is a tool used to convert updates from binary data to ASCII data.

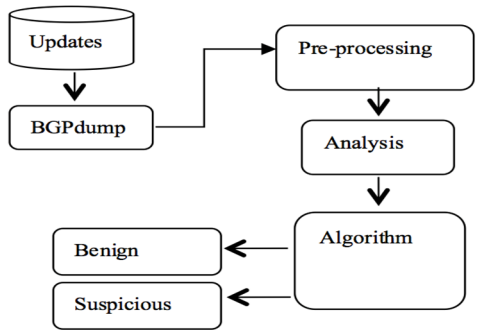

Figure 1: Previous structure of the IP prefix hijack detection method.

2.1 Algorithm

The algorithm has three objectives – firstly, associates the announcer (origin AS) with each advertised IP prefix. Secondly, removes duplications of associated origin ASes and IP prefixes. Finally, identifies any IP prefixes that were announced by more than one AS.

The algorithm receives the origin ASes and their IP prefixes from two different data sources, every fifteen minutes. Origin ASes are mapped on their IP prefixes using a cell array in MatLab (Alshamrani et al. 2015) allowing assignment of one ASN to multiple IP prefixes.

Data reduction is then applied to the dataset by removing duplicated origin ASNs and IP prefixes, which allows the algorithm to categorise faster the input dataset in order to detect suspicious announcements. After computing unique origin ASes and their associated IP prefixes, the algorithm compares the IP prefix of each AS to the IP prefixes of all origin ASes reported during each time interval to find out IP prefixes that were advertised by multiple ASes.

The analysis performs a comparison between individual AS-IP prefix rows in the cell array using the BSA (Binary Search Algorithm) (Dalal 2004) O(log N) due to its ability to execute array comparisons exponentially faster than linear search algorithm (LSA) (Horvath 2012). The algorithm lists the detected incidents (suspicious hijacks) in a new matrix composed of two columns.

Table 1 shows the format of the Mapping Cell Array for Origin ASes and IP prefixes. The comparison part shows the outputs as normal or suspicious routes. However, the algorithm in [1] has some false positives as it is going to be explained in the following subsection.

Table 1: Example of mapping cell array in quarter 82.

Origin ASes	IP prefixes
137	369760021
151	369760021; 369760023; 697600524
174	139438524; 244296124

2.2 Previous Algorithm Limitation

The algorithm proposed in (Alshamrani et al. 2015) has a significant limitation, as it cannot take into account organisations using multiple different ASNs to advertise their own routes. To address this limitation, this paper introduces a Verification database to be included in the detection method in order to enhance the accuracy of the algorithm. Since BGP updates lack the organisation names (codes), we extract data from the RIRs and process them to produce a dataset that links the AS numbers to the unique codes for the organisations that own them.

3 MAPPING OF AS NUMBERS AND ORGANISATIONS

This section discusses the processing of RIR information (specifically the RIPE Whois database (John Stamatakis 2014)) to enhance the BGP update fields used as input and support the algorithm described in (Alshamrani et al. 2015) to reduce the false positives.

3.1 Extracting and Numerating Organisations' ASNs and Their Unique Codes

As part of the RIR registration, each organisation has a unique code to uniquely identify it. For instance, in RIPE, ORG-YE1-RIPE field represents Yahoo in Europe but ORG-HBp1-RIPE represents HSBC Bank plc. The Verification DB is processed in three phases.

1. PHASE 1

This phase extracts the ASNs and organisation codes fields from the RIPE dump database and stores data into corresponding fields, *aut-nums* and *orgs*, such as autonomous system number AS20535 and its code ORG-IG12-RIPE.

2. PHASE 2

Since RIPE includes ASNs without an associated organisation code (name), the incomplete records are filtered out, which does inherently limit the capabilities of the presented method because they confuse the order of searching the ownership of specific IP prefix and mix them up.

3. PHASE 3

The organisation code (name) field is structured as an array and was created to include all organisations codes that facing every ASN in RIPE. Each organisation code (organisation name) is divided into three parts (ORG, IG12 and RIPE for example) and saved in an array called ORG. Second and third index in ORG array respectively represent the organisation name and data resource (e.g. RIPE). Currently, the most important part is the second field of the array because it uniquely identifies the organisations.

The third field of the organisation code array represents the database (e.g. RIRs or ASN delegator) that provided the record; this helps to differentiate between multiple database source owners. To optimise the analysis, these two parts are converted to numeric data. Table 2 shows one record of the final structure format of the Verification DB. First column is used as a primary key to be linked to ASNs in the Suspicious Findings List (Alshamrani et al. 2015).

Table 2: Example of the final format to the VerificationDB.

ASNs	ORG codes and sources
200912	18191226

3.2 Filtering Organisations with One ASN

Given the method focuses on organisations with more than one ASN in order to refine the results, all organisations that have only one ASN are filtered out, allowing the algorithm based on (Alshamrani et al. 2015) to parse a significantly smaller dataset in order to determine whether suspicious IP prefix hijacks are real or not.

In the case of RIPE database from February 2015 (Meyer n.d.), the size of the Verification Database before filtering out organisations with only one ASN was 25580 records, reduced to 6283 records through filtering. The improved detection method verifies its results (suspicious hijacks) based on the reduced Verification DB. The general structure of processing the Verification Database is shown in Figure 2.

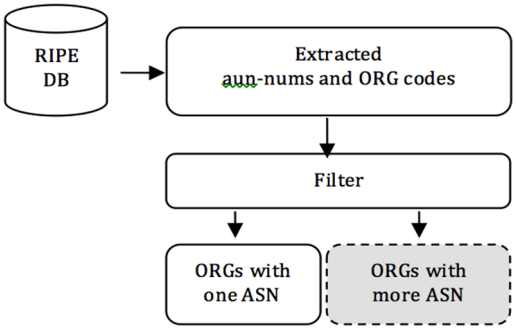

Figure 2: Structure of Verification DB of organisations that have multiple ASNs.

4 IMPROVED DETECTION METHOD

This section discusses the use of the Verification DB in the IP hijack detection method (Alshamrani et al. 2015). The encompassing algorithm validates its outputs based on this database. It also demonstrates the results of the detection method after the improvement.

In the previous work (Alshamrani et al. 2015), the algorithm directly translates the results into two categories, normal and suspicious, but it does not verify the decision against organisations owning multiple ASes. To expand, if an organisation relocates a prefix between two of the ASes it owns, the algorithm would flag the change as a suspicious event; in fact, given both ASes are owned by the same organisation, it is likely that it is due to addressing and logistics ASes and IP prefixes management rather than a hijacking incident. In this paper, we introduce the Verification DB to check against the owners of the ASes involved in the suspicious events.

The Verification DB maps the autonomous system numbers and the corresponding organisations owning them. The extended comparison allows us to verify if a suspicious event is due to an IP prefix being migrated between ASes owned by distinct organisations. If an IP prefix is indeed migrated between ASes owned by different organisations, the event is further flagged as suspicious; if however the migration is between ASes of the same owner, the algorithm concludes that the change is not a suspicious event and continues with the search.

Figure 3 provides a block-diagram overview of the improved detection method, including input from RIRs into the decision process. In the diagram, the Extensional Block provides the required functionality for the RIR information and verification DB processing.

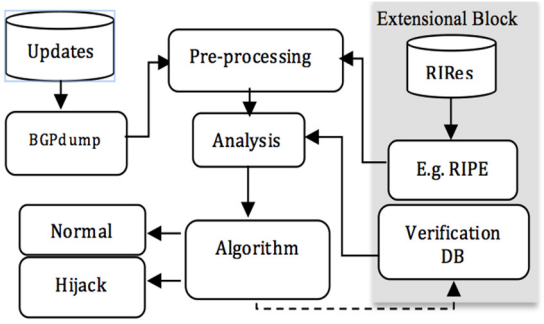

Figure 3: Improved structure of the IP prefix hijack detection method.

5 VALIDATION OF IMPROVED DETECTION METHOD

The algorithm proposed in the previous section was applied to two incidents: the whole day of Pakistan and YouTube hijacking day (24/02/2008) and the day of the Con-Edison hijack (22/01/2006). In other words, before the algorithm takes a decision with the suspicious routes, it checks out if two suspicious routers that were impersonating the same IP prefix exist in the Verification DB with one organisation name; if so, they are ignored otherwise the advertisement will be flagged as a hijack. Pseudo code below explains the steps of the validation in details. The algorithm develops the accuracy of the suspicious results that were already caught by searching for the signature attack of IP prefix hijackings. It takes each two suspicious ASes in the list of Suspicious Finding List and searches for them in VerificationDB which contains organisations that have more than one ASN; if they exist in the SFL, the ASes will be removed from suspicious list as they are not a real signature for the IP prefix hijackings.

```
Suspicious = dlmread
(Suspicious_Finding_List);
suspiciouslen = length (suspicious);
VerifDBLen = length(ORGsWithMultiASN);

CASE = 1;
ORGCODE = [1 0; 2 1];

WHILE CASE <= suspiciouslen
    ASN1 = suspicious (CASE, 4);
    ASN2 = suspicious (CASE+1,4);
    CHECK = 1;
    WHILE CHECK < VerifDBLen
```

```
        ASN3=ORGsWithMultiASN (CHECK,
1);
        IF (ASN1 == ASN3 OR ASN2 ==
ASN3)
            IF (ASN1 == ASN3)
              ORGCODE (1,1)=CASE;
              ORGCODE (1,2)= . . .
              ORGsWithMultiASN
(CHECK,2);
            ELSEIF (ASN2 == ASN3)
              ORGCODE (2,1)= CASE;
              ORGCODE (2,2)= . . .
              ORGsWithMultiASN
(CHECK,2);
            END
        END
        IF (ORGCODE (1,2) ==
ORGCODE(2,2)
              ORGCODE (1,1) == ORGCODE
(2,1))
       suspicious(CASE-1: CASE+1,:)=[];
              suspiciouslen= . . .
              length(suspicious);
           ORGCODE (1,2)=0;
           ORGCODE (2,2)=1;
        END
        CHECK= CHECK+1;
    END
    CASE= CASE+3;
END
```

5.1 Findings

The improved algorithm added functionality has two advantages: it can detect multiple occurrences of the same incident and allows the algorithm to identify organisations that announce their routes with more than one ASN. In the specific example of the YouTube hijack, the algorithm from (Alshamrani et al. 2015) identifies 1767 incidents; following the analysis of repeated incidents, 975 unique incidents can be identified. Parsing the analysis through the Verification DB, the number of Suspicious hijacks drops to 969, due to the SLF suspicious hijack exclusions. Following a similar processing, the events from (the 22nd Jan 2006) do not show any improvement because the incidents took place outside RIPE, so the Suspicious Findings List from RIPE is empty (none of suspicious results in the findings list is in the RIPE database). Thus, the algorithm needs several sources such as AFRINIC (Africa Region), APNIC (Asia/Pacific Region) and ARIN (North America Region) to improve its accuracy.

5.2 Algorithm Challenges and Solutions

Since the Verification DB uses only the RIPE database as a case study, the results would still include false positives but with lower percentage. The algorithm would be more accurate if the number of different sources (RIRs' and ASN delegators' database) used for the Verification DB increases. This challenge can be addressed as described at the end of the previous sub-section.

Second challenge is that the RIRs and ASNs delegators' databases need to be updated regularly and concurrently with the changes to ASNs and organisation names. Third difficulty is the algorithm detects IP prefix hijacks based on off-line analysis.

Furthermore, some organisations do not include their code in their associated record in the RIPE database. In addition, some RIRs do not keep historical records of old Whois registrations details. Once a record is updated or deleted, the old record is not stored in an archived database. As a result, the algorithm cannot evaluate organisation names and ASNs changes when it compares past suspicious hijacks to the current Verification DB.

Finally, prefix hijacks may be transparent for the algorithm on a subset of routers due to partially propagated prefix updates; therefore routers need to work collaboratively to compare and aggregate update information with their neighbours. The following section discusses the steps of this collaboration.

6 PROPOSED DETECTION ARCHITECTURE

This section describes a possible architecture that allows aggregating data collaboratively on several different routers. The aim of the architecture is to allow BGP routers to jointly benefit from the independently identified events on each router and, subsequently, lead to higher accuracy when detecting anomalous behaviour.

6.1 Architecture Method and the Advantages

Routers that run the hijack detection algorithm should work together in order to improve the reliability and timeliness of the information derived from the UPDATE messages. An IP prefix hijack might not significantly affect traffic exchanged with the impersonated AS until it spreads to

multiple/different ASes; to alleviate the effect of the hijack, the algorithm has to work collaboratively to prevent the propagation of invalid routes. The detection algorithm operates independently from BGP and categorises network events, but may benefit from sharing and receiving data from other neighbouring routers in order to detect the effect of the attack rapidly. The BGP updates may be collected and aggregated by a router over a specific operational timeslot, while bearing in mind that anomaly detection becomes stale with higher aggregation slots. In case of detecting a suspicious route, an alarm of the invalid route would be sent to all neighbours.

The algorithm should run in each router, based on the different information received. In addition to the use BSA (of Binary search algorithm), making the routers work collaboratively and independently would increase the detection speed and would not require any modifications of the infrastructure of the BGP routers. Figure 4 shows the general structure of the improved detection method when linked to the BGP routers.

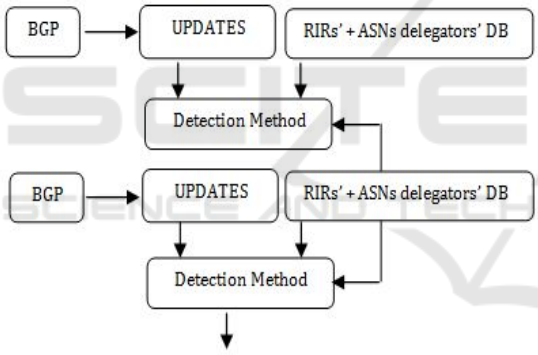

Figure 4: Architectural method of linking BGP to the detection method.

Moreover, if some routers do not actively run the detection system, the other routers may identify and publicise the anomaly. By doing so, each BGP router will have a chance to suppress any suspicious routes to prevent itself from further propagating the hijacked routes.

6.2 The Effectiveness of the Architecture over the Algorithm

The advantage of a collaborative architecture in the BGP context is that each router can only check its own received update packets so there is no load to the algorithm to find out the hijack. Another advantage of this collaboration is that the check will be periodic, with timeslot starting times distributed over time. For example, if Router A cannot detect the hijack at 1:15 AM because it is not the time slot to do the check, there may be another copy of the algorithm in the neighbouring routers doing the check and detecting the hijack faster.

7 CONCLUSION AND FUTURE WORK

A new framework was proposed to enhance the accuracy of a previously proposed method for IP prefix hijack detection. The framework extracts the unique code and associated ASNs of organisations from different RIRs; the algorithm then excludes previously detected IP prefix hijacks that are likely to be false positives. After proposing the framework, its efficiency is validated on the Pakistan IP hijacking event from 24th Feb 2008 and the Con-Edison hijack (22nd Jan 2006). The analysis used the RIPE dump database from the two respective dates as a case study to evaluate the proposed framework. In the evaluation, the algorithm was able to improve the accuracy of the IP prefix hijacks, reducing the false positives by 0.61% (18 suspicious hijack) for the two events.

From the results, it is clear that the algorithm can work accurately but also could omit some events; more specifically, several incidents from 22nd Jan 2006 were still false positives, since the analysis was based only on the RIPE database. Additionally, if an AS announces an IP prefix in the absence of the real origin AS, the algorithm will not be able to detect the impersonation when it works independently (non-collaboratively).

In terms of router interconnectivity, some routers do not have a direct connection to the hijacker. In other words, the detection method ought to be decentralised in order to collect direct information regarding the hijacker and detect the hijack faster. Another advantage of the decentralisation is that detection of anomalies can be done for various, partially overlapping timeslots. Another challenge of the algorithm is that the hijacker could impersonate one of the net-range IP prefixes (sub-prefixes), event that may be transparent for the algorithm. Last, the period gap (synchronisation) between fetching BGP updates and the current status of the ASN of an organisation, together with the IP prefixes changes, could have a negative impact on the accuracy of the algorithm.

In future, the proposed approach may provide further insight into and refine the rationale behind organisations announcing the same IP prefix with different ASN. This is needed in order to distinguish between normal BGP operations and malicious ones, and then address the false positive errors.

REFERENCES

Alshamrani, H., Ghita, B. & Lancaster, D., 2015. Detecting IP prefix hijacking using data reduction-based and Binary Search Algorithm. In *2015 Internet Technologies and Applications (ITA)*. Wrexham: IEEE, pp. 78–84. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7317374.

Cao, H. et al., 2009. A Packet-Based Anomaly Detection Model for Inter-domain Routing. In *2009 IEEE International Conference on Networking, Architecture, and Storage*. Hunan: IEEE, pp. 192–195. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5197320 [Accessed December 6, 2013].

Dalal, A., 2004. Searching and Sorting Algorithms. , (100), pp.1–13. Available at: http://www.cs.carleton.edu/faculty/adalal/teaching/f04/117/notes/searchSort.pdf [Accessed December 8, 2014].

Goldberg, S., 2014. Why is it taking so long to secure internet routing? *Communications of the ACM*, 57(10), pp.56–63. Available at: http://dl.acm.org/citation.cfm?doid=2661061.2659899.

Horvath, A., 2012. Quicksort, binary search and linear search performance - far from what you believe. , p.6. Available at: http://blog.teamleadnet.com/2012/02/quicksort-binary-search-and-linear.html [Accessed August 5, 2013].

John Stamatakis, 2014. Pen Test Live: Download Database. Available at: http://www.pentestlive.com [Accessed January 9, 2014].

Meyer, D., University of Oregon Route Views Archive Project. *University of Oregon*. Available at: http://archive.routeviews.org/bgpdata/2008.02/UPDATES/ [Accessed October 5, 2013].

Schlamp, J. et al., 2015. The Abandoned Side of the Internet: Hijacking Internet Resources When Domain Names Expire. In *arXiv preprint arXiv: ….* pp. 188–201. Available at: http://link.springer.com/10.1007/978-3-319-17172-2_13.

de Urbina Cazenave, I.O., Kosluk, E. & Ganiz, M.C., 2011. An anomaly detection framework for BGP. In *2011 International Symposium on Innovations in Intelligent Systems and Applications*. Istanbul: IEEE, pp. 107–111. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5946083.

Vervier, P., Thonnard, O. & Dacier, M., 2015. Mind Your Blocks: On the Stealthiness of Malicious BGP Hijacks. In *Proceedings 2015 Network and Distributed System Security Symposium*. Reston, VA: Internet Society, pp. 8–11. Available at: http://www.internetsociety.org/doc/mind-your-blocks-stealthiness-malicious-bgp-hijacks.

Wählisch, M., Maennel, O. & Schmidt, T.C., 2012. Towards detecting BGP route hijacking using the RPKI. *ACM SIGCOMM Computer Communication Review*, 42(4), p.103. Available at: http://dl.acm.org/citation.cfm?doid=2377677.2377702.