Malicious BGP Hijacks: 
Appearances Can Be Deceiving

Pierre-Antoine Vervier∗† Quinten Jacquemart† Johann Schlamp‡
Olivier Thonnard∗ Georg Carle‡ Guillaume Urvoy-Keller‡ Ernst Biersack‡ Marc Dacier∗

∗Symantec Research Labs, {Pierre-Antoine_Vervier,Olivier_Thonnard,Marc_Dacier}@symantec.com
†Eurecom, {Quentin.Jacquemart,Ernst.Biersack}@eurecom.fr
‡Technische Universität München, {schlamp,carle}@in.tum.de
§Laboratoires I3S Université de Nice Sophia-Antipolis, urvoy@i3s.unice.fr

Abstract—BGP hijacking is a well known threat to the Internet routing infrastructure. There has been considerable interest in developing tools that detect prefix hijacking but such systems usually identify a large number of events, many of them being due to some benign BGP engineering practice or misconfiguration. Ramachandran et al. [1] and later Hu et al. [2] also correlated suspicious routing events with spam and claimed to have found evidence of spammer temporarily stealing prefixes to send spam.

In an effort to study at large scale the existence and the prevalence of malicious BGP hijacks in the Internet we developed a system which (i) identifies hijacks using BGP, traceroute and IRR data and (ii) investigates traffic originating from the reported networks with spam and netflow data. In this paper we present a real case where suspicious BGP announcements coincided with spam and web scam traffic from corresponding networks. Through this case study we show that a correlation of suspicious routing events with malicious activities is insufficient to evidence harmful BGP hijacks. We thus question previously reported cases and conclude that identifying malicious BGP hijacks requires additional data sources as well as feedback from network owners in order to reach decisive conclusions.

I. INTRODUCTION

Considerable effort has been spent on developing techniques to monitor IP prefix hijackings, where an autonomous system (AS) announces BGP routes to network prefixes that it does not own. Traditionally, these techniques have been classified according to the kind of input data they use. Those based on the control plane, such as [3], [4], [5], monitor BGP update messages and/or routing tables (e.g., RouteViews [6], RIPE RIS [7]) of BGP routers. Techniques based on the data plane, such as [2], [8], [9], [10], continuously probe target networks from multiple vantage points, looking for discrepancies that may arise. More recently, [11] proposed a hybrid system that correlates the results from both perspectives to reduce the number of false positives. Unfortunately, even last-generation tools yield output cluttered with alerts corresponding to benign network events.

Meanwhile, Ramachandran et al. [1] reported spectrum agility, where so-called fly-by spammers announce a blackspaced class A prefix for a short time in order to use the IP addresses for spamming. Later, Hu et al. [2] further correlated BGP hijack alerts with spam sources from [1]. In theory, spammers using this technique are able to circumvent backtracking and traditional IP-based blacklisting due to the short-lived nature of the attack. More recently, a validated case of a BGP hijack specifically carried out to send spam from the stolen prefixes was reported in [12], [13]. Unlike the first observations of fly-by spammers, this incident involved a long-term hijack attack of several months and was confirmed by the owner of the victim network.

Despite those anecdotal evidence of spammers hijacking IP space to send spam, little attention has been devoted to study the root cause of BGP hijacks. Some works [13], [14], [15] have carried out forensic analyses of well known hijacks. Others have attempted to validate observed suspicious routing events [2], [11], [16] usually in the context of the evaluation of a hijack detection technique. However such investigations focused more on differentiating benign cases from suspicious, unexplained cases than identifying BGP hijacks performed in preparation of other malicious activities. However, identifying such malicious BGP hijacks is primordial. Indeed if there are attackers hopping between IP addresses in hijacked prefixes as described in [1], this would have a significant impact on IP-based reputation systems like spam blacklists. In addition, hijackers actually steal IP identities, i.e., attacks launched from hijacked prefixes would be wrongly attributed to the owner of the hijacked network. As a consequence, taking legal actions would be highly complicated.

This work aims at presenting a system developed to study at large scale malicious BGP hijacks and showing through the analysis of a case study why the limitations of previous works on fly-by spammers shall question their results. The contribution of our paper is twofold. First, we present our methodology and experimental environment for detecting BGP hijacks and assessing the malicious intent by looking at the traffic generated by the hijacked prefixes extending on the previous work SPAMTRACER [12]. This is achieved by combining several data sources and analysis techniques in a novel way. We look for suspicious MOAS conflicts extracted from BGP update messages, utilize live spam feeds collected at spam traps, issue traceroutes towards spamming networks, look for suspicious traffic in netflow data collected at a scientific network, and analyze historical dumps of Internet Routing Registries. Second, we present a real case study to

version: January 15, 2014. Copyright IEEE. Author’s version for personal use. Not to be offered for sale or otherwise re-printed, re-published or re-used without permission. A version of this paper will appear in IEEE ICC CISS 2014.
show that correlating suspicious routing events with security-related incidents is not sufficient to identify BGP hijack attacks performed with malicious intent. When studying malicious BGP hijacks drawing quick conclusions is tempting and complementary data sources together with feedback from network prefix owners is absolutely necessary in order to identify real malicious BGP hijacks. We thus conclude that past and future cases should be (re-)evaluated using the variety of data sources that are readily available.

The paper is organised as follows. In Section II we present the methodology and the experimental environment we have setup to study at large scale the existence and prevalence of malicious BGP hijacking events. It extends on the previous work [12] with a control/routing plane based detection of prefix ownership attacks, a routing consistency analysis using Internet Routing Registries and an analysis of the network traffic from suspect hijacked networks using NetFlow data. In Section III we perform an in-depth analysis of a real case where suspicious routing events were correlated with spam and web scam traffic strongly suggesting a BGP hijack performed for malicious purposes. In Section IV we conduct a second examination of the case revealing evidence against a hijack leading us to question results from the previous works [1], [2] on fly-by spammers. Section V discusses the interpretations of the case study and the lessons learned on the study of the root cause of BGP hijacks. Finally, Section VI concludes the paper and presents future work.

II. EXPERIMENTAL ENVIRONMENT

To study malicious BGP hijacks we have developed a system that combines several data sources and measurement techniques, in particular data passively collected from spam traps, BGP feeds, netflow data collected at a large academic/research network and active traceroute probing. The overall system architecture is depicted in Figure 1.

In order to obtain suspicious BGP events that pose a security threat we correlate the output of a BGP control-plane monitoring scheme (1) with spam data collected from spam traps (2), and with traceroute results performed to suspicious, spam-emitting networks (3). We further correlate those prefixes with network footprints using netflow data collected at a large academic/research network (4). Finally, we search IRR databases for evidence to (in)validate suspicious BGP announcements (5).

A. MOAS Detection and Filtering

We monitor BGP’s control plane and look for situations in which a prefix is simultaneously originated by multiple ASes, otherwise known as MOAS conflicts (Multiple Origin AS conflicts). Their occurrence is very easy to detect, and techniques such as PHAS [4] do a fine job. However, these techniques usually fail to filter out the many valid reasons why a network would be a MOAS [17]. To name a few: anycasting, multihoming with a private ASN or with static links. In order to focus on malicious hijacks, benign MOAS occurrences must completely be removed. Moreover, the time needed to complete data plane analyses, e.g. active traceroute measurements to suspect networks, warrant us to remove cases that will eventually turn out to be false positives.

A deep analysis of MOAS events led to the discovery of shared patterns embedded in their AS paths; some of which can be used to reduce the number of MOAS monitoring alerts. Using data from RIPE RIS’ Amsterdam collector (rrc00) [7] in 2012, we found out that roughly 75% of MOAS events were the result of a provider-customer relationship. This indicates that the multiple origins (e.g. a1 and a2) of a prefix p share a direct AS-level link. There seems to be no operational reason why a1 would keep announcing p in the long term. Indeed, while the average duration of such MOAS events is about 48d, the median duration is only about 7h. Also, 35% of the provider-customer MOAS events last less than 6 minutes, and 55% of them last less than a day. More importantly, the provider does not gain anything from hijacking p in such cases, since it is in the path-to-destination anyway.

Another benign type of MOAS are long-lived events, where the uptime of multiple origins for a prefix is larger than a threshold T. While [18] uses T = 24h, we set ours to T = 48h in order not to limit active measurements on suspect networks too early. Using rrc00’s 2012 data, we saw that 14% of non-provider-customer MOAS lasted more than 48h.

From our experience, we argue it is difficult to assess the security threat posed by the remaining MOAS conflicts solely from BGP data. However, the number of daily filtered alerts is within our achievable number of traceroute measurements.

B. Spam Network Monitoring

Data plane measurements can be leveraged to determine the impact of a routing change on the forwarding paths towards a monitored network. We have developed a tool called SPAMTRACER [12] to monitor the routing behavior of spamming networks by performing traceroute measurements towards networks that have sent spam to Symantece.cloud spam traps. These measurements are performed on a daily basis and repeatedly for a certain period of time after spam is received. We focus on short-lived hijacks by fly-by spammers as observed in [1], i.e. hijacks lasting no longer than one day, thus we set the monitoring period to one week. Currently the system is able to monitor up to ~8,000 network prefixes everyday with one IP address traced per prefix. By performing measurements on consecutive days for one week, data plane paths and BGP routes towards a given network can be compared and analysed in depth to find indications for an ongoing hijack. Because we monitor networks just after spam is received, we expect to observe a routing change as soon as the hijack ends, provided the network was indeed hijacked.
C. Netflow Data Analysis

In addition to the techniques described above, we look at netflow data to analyze changes in traffic patterns before, during and after a suspected hijack. Such changes can range from simple outages in monitored networks, where outgoing connection attempts are unanswered, to changes in traffic volume or even to a significant amount of new connections from and to different sets of ports.

We utilize archived netflow data of the Münchner Wissenschaftsnetz (MWN) – Munich’s scientific network – which comprises more than 80,000 end hosts. It is used by researchers, students, and administrative personnel, who generate monthly upstream and downstream traffic volumes of more than 300 and 600 Terabyte, respectively. We consider the MWN large enough to be affected by large-scale spam campaigns, and expect to observe at least some portions of spam that originate from hijacked networks. The netflow data is collected according to RFC5103.

D. Internet Routing Registries Analysis

In order to take an administrative point of view for obtaining further conclusive information on suspected hijack events, we search the IRR databases for relationships between the involved ASes and IP prefixes. This is achieved by extracting route objects from archived daily IRR dumps [19] provided by the different Regional Internet Registries (RIR).

These route objects are maintained by ISPs or end-users that are responsible for the IP space and allow them to specify which AS(es) should announce a given prefix. Although ISPs and end-users are not forced to keep those records complete and up-to-date, when available, they still provide valuable forensic information on past and present relationships between the holder of an AS and a prefix. Such an ordinary relationship can thus cast a malicious hijack event into doubt. IRR records may also contain meaningful information about prefix and AS holders, e.g., a description of the holder’s business or contact details that can further be used in the analysis of hijack events.

III. THE BULGARIAN CASE

For the month of February 2013, 2,331 distinct prefixes were involved in control plane alerts, i.e., MOAS conflicts. We use a time window of 15 days to correlate these events with spam from IP addresses observed at spam traps and on blacklists to identify malicious hijacks. In the following we present an in-depth analysis for one of the matching events. Note that all results are anonymized with good cause; we are nevertheless willing to share details upon request.

Based on several alarms raised by our detection system on February 3rd, 2013, we became aware of an incident taking place in Bulgaria. Several MOAS conflicts were observed for networks that correlated with emerging spamming activities. We carried out a detailed analysis of these events, and present our results in chronological order below.

Phase 1: Normal Situation

Since 2008, the prefix A.B.0.0/16 has been announced in BGP by a Tier-3 ISP Alice. This ISP is known to provide hosting services for a variety of customers. We did not observe announcements of more specific prefixes during the whole time of phase 1 (Figure 2).

Phase 2: Hijack and Spam

On December 4, 2012, Mallory started announcing a set of nine more specific (/24) prefixes of Alice, who carry on with the original /16 announcement (Figure 2). By using online whois queries, we learned that Mallory supposedly is a VPS service provider also located in Bulgaria. A thorough web-search however returned no result for this specific company.

a) Spam: Figure 3 shows spam 1 received by Symantec.cloud spam traps from IP addresses belonging to the nine prefixes announced by Mallory. The figure also presents blacklisted IP addresses from Uceprotect Level-1 [20] related to these prefixes. This figure shows a strong correlation between the BGP routing announcements, spam, and blacklisted IP addresses. On some days, up to 80 spam emails were sent to our spam honeypots. Many prefixes also had around 100 blacklisted IP addresses for several days. On Figure 3 we still observe some blacklisted IP addresses after the end of phase 2 but we attribute them to the one-week expiration period of Uceprotect records. Symantec.cloud spam dataset may provide the spam botnet name responsible for the spam based on spam bot signatures. Because spam bots are usually compromised machines, they should not be observed on hijacked IP space. And indeed no such botnet could be inferred from Symantec.cloud’s reports for spam hosts in the suspicious prefixes. This indicates that those machines were likely set up by the spammers themselves.

b) Scam Hosting Infrastructures: We further analyzed the spam mails and were able to identify several URLs within these messages. Out of 118 extracted domain names, 89 resolved to an IP address within six of the obtrusive prefixes. We conclude that the spam was also used as a platform to promote a scam infrastructure hosted within these prefixes. About 90% of all spam hosts in the nine A.B.x.0/24 networks coincided with IP addresses of spam hosts, which

1Live spam feed of ~4M spam per day starting in January 2012
indicates that the spammers took full advantage of the prefixes under their control.

It is interesting to see that almost all scam hosts’ IP addresses shared the same last byte while being spread over all abused networks (e.g. A.B.95.96.114.5, A.B.95.114.9, A.B.95.96.97.114.14, etc). Similar characteristics appear for the resolution of domain names to IP addresses within the nine prefixes. All 89 resolvable domains were created at nearly the same time as the prefixes were first announced in BGP by Mallory. All pieces of evidence suggest a single administrator behind the domains and network infrastructure.

c) Netflow Traffic Analysis: We analyzed netflow data for the period of December 2012 to March 2013, and were able to collect 13,001 inbound flows from the suspicious prefixes. The majority of these flows accounted for SMTP requests (71.0%), DNS replies (25.2%), HTTP replies (1.6%) and SMTP replies (1.4%). The remaining 1.8% of flows indicated traffic to an IRC server in our networks, and to ephemeral UDP ports. For 97.4% of all incoming flows, we observed corresponding outgoing flows. An analysis of the IRC traffic revealed that these flows originated from 1,381 hosts spread over 254 different /24 subnets within the /16 prefix announced by Alice. Such orchestrated IRC traffic across all networks of Alice’s customers seems to be implausible: we thus assume that these flows attribute to IP spoofing activities unrelated to the Bulgarian case, and exclude them from our analysis.

All connection requests (incoming for SMTP and outgoing for DNS and HTTP) are depicted in Figure 4. We observe a strong correlation in phase 2 between the BGP announcements (Figure 2), the observed spam (Figure 3), and the blacklist records (Figure 3). We observed a total of 925 IP addresses for the delinquent’s activities, of which 850 IP addresses were used to send spam mail. Less than 10% of these addresses were re-used for the DNS and HTTP activities. We further found 30 distinct DNS servers mostly hosted in the prefixes A.B.96.0/24 and A.B.114.0/24, which were queried over 3,000 times by clients in our networks. The flow data also shows 200 bidirectional HTTP connections to more than 100 web servers in the reported prefixes.

This analysis confirms that the prefixes were used in order to massively send spam from several hundred clients. Furthermore, it clearly shows that the person in charge hosted more than 100 live services (DNS and HTTP), presumably to do phishing or similar fraudulent activities.

Phase 3: Legitimate Use of Assigned IP Space

On February 3, 2013, Bob started announcing five of the nine prefixes announced by Mallory, resulting in MOAS conflicts during a few hours before Mallory withdrew all of its announcements. Alice, once more, kept on announcing the original /16 prefix (Figure 2). Several spam hosts that used to reply to traceroute probes on consecutive days during phase 2 also suddenly became unreachable suggesting a real change in network topology.

Bob is a business-to-business IT service provider located in the same country as Mallory and Alice, according to their website. Its ASN first appeared in BGP in November 2008. All five /24 prefixes were announced via Alice acting as legitimate upstream provider. Figure 5 depicts the overall topology from a BGP’s point of view.

With beginning of phase 3, all malicious activities suddenly stopped. This indicates that Bob was regularly assigned the five prefixes by Alice in the context of a provider-to-customer business relationship.
Phase 4: Back to Normal

On March 20, 2013, Bob withdrew its announcements of Alice’s five prefixes, resulting in the same initial situation as described for phase 1, where the whole prefix A.B.0.0/16 was announced by Alice only.

Given these findings, approaches presented in [1], [2] would conclude the existence of a malicious BGP hijack. All evidence presented so far, especially the strong correlation for both the control plane and the data plane, lead us to the conclusion that we indeed observed a malicious hijacking event for this Bulgarian Case.

IV. A SECOND EXAMINATION

Despite the evidence for a malicious hijack incident described so far, we decided to further investigate the case and found significant evidence against a hijacking event. We analyzed more than one year of archived RIPE IRR database dumps in order to infer the legitimate owners of the suspected prefixes by searching for route objects and looking into the corresponding origin (AS) attributes. We found that Alice carefully maintained such route objects in the RIPE IRR database throughout all four phases. We obtained the first three objects related to the prefixes in question on December 4th, 2012 (Figure 6). Their origin attributes were set to Mallory, and the creation time corresponded to her first BGP announcements. This clearly indicates that – at least according to the RIPE IRR database – Mallory was authorized to use these prefixes.

Figure 6 gives an overview for all relevant route objects that we found in the RIPE IRR database. We learned that the dates of appearance fully match all BGP announcements of Mallory and Bob (see Figure 2), and all objects were maintained by Alice. If we assume that an attacker is incapable to alter the RIPE IRR database at will (and that he had no access to Alice’s maintainer account), we must conclude that Alice delegated all nine prefixes to Mallory by choice, and reassigned some of them around February 3rd, 2013 to Bob.

We further extracted the database objects’ descr attributes, and even found some weak evidence for a relationship between Mallory and Bob. Those free text fields can be set to any value. For Mallory, all fields were set to BG-XX-N. BG indicates Bulgaria, whereas N corresponds to each of the prefixes’ third byte. More importantly, XX represented the initial letters of Bob’s company name. After reassignment, the description changed to Bob’s full company name.

Finally, we contacted Mallory’s upstream provider and learned that Mallory requested to announce rented prefixes. After receiving complaints, the upstream provider cancelled Mallory’s contract.

Given all circumstances, we must conclude that Mallory acted maliciously by sending spam. However, we cannot decide if Mallory really hijacked prefixes, or if Mallory just rented the networks for abuse.

V. DISCUSSION

Even though we have accumulated a series of converging indices incriminating one of the actors, namely Mallory, involved in performing BGP hijacks with malicious intent, we still cannot reach a decisive conclusion.

As presented in Section III, the strong correlation between the BGP announcements of Alice’s sub-prefixes by Mallory, the spam received by Symantec.cloud and the evidence of scam hosting infrastructures during phase 2 initially led us to believe that Mallory had indeed hijacked these prefixes to emit spam. This result is supported by the following observations: (i) The temporal correlation between the BGP announcements and the emerging spam during phase 2 strongly suggests that machines in Mallory’s network are the spam sources. (ii) Mallory’s first appearance in BGP as well as the registration date of the domain names advertised in the received spam mails directly coincident with phase 2 of the incident. (iii) Alice provided upstream connectivity for Bob, while Mallory hired an independent upstream provider, although Alice continuously announced the full enclosing /16 prefix. (iv) As soon as Bob started to announce his assigned prefixes in phase 3, Mallory’s announcements and the emission of spam stopped, and no more traffic flows were observed.

Our findings in Section IV validate prefix ownership based on the RIPE IRR database for all involved parties during all phases of the incident. However, this fact does not exclude a malicious BGP hijack: it is possible that an attacker covered up his traces by altering objects in the RIPE IRR database. According to RIPE, 86% of database maintainers were using
password-only authentication in 2011 [21]. However, password protection may not be enough since an attacker could use information leaked from the IRR database [22] and/or phishing e-mails [23] to gain privileged access to the database.

Our system to detect malicious BGP hijacks was partly designed upon findings of previous studies on the root causes of BGP hijack events, like Ramachandran et al.’s study [1] on short-lived BGP announcements, the correlation between BGP hijack alerts and spam by Hu et al. [2] and a validated hijack case performed by a spammer described by Vervier et al. in [12] and by Schlamp et al. in [13]. Comparing our findings presented in Section III with those reported in previous work quickly led us to the conclusion that the Bulgarian Case was indeed a malicious BGP hijack. However, the novel forensic analysis of an IRR database described in Section IV at least opened our mind that we possibly have not found a real hijack event, but rather a plain abuse of rented IP space. In the end, although we remain indecisive, we learned that it is crucial to consider complementary data sources, preferably as independent as possible (e.g., IRRs) as well as feedback from network owners (e.g., via mailing lists like NANOG as in [12],[13]) in order to avoid drawing conclusions too quickly based on a limited set of evidence skewed towards one verdict or the other. This fact is of particular interest to avoid misattributing attacks launched from hijacked IP space when responding with possibly legal actions.

VI. Conclusion and Future Work

We have developed a system to study the root cause of BGP hijacks to identify hijack events performed in preparation of other malicious activities. We are able to detect network anomalies from BGP and traceroute data and uncover malicious intents by combining a variety of data sources, including spam traps and netflow data. We balance our results with a forensic IRR databases analysis on relationships between involved parties.

We closely studied one suspicious hijack incident raised by our system, in which routing anomalies coincided with spam originated from the affected prefixes. We further observed a variety of scam activities hosted on these prefixes and we finally put together conclusive evidence for an ongoing hijack attack. With similar findings, previous work would have concluded the existence of a malicious hijack case. We decided to question our results and learned that the presumed delinquent might have legitimately rented IP space to carry out his malicious activities. We thus conclude that considering multiple and independent data sources as well as feedback from network owners is primordial to avoid drawing conclusions biased by a limited set of evidence possibly skewed towards one verdict or the other. We consequently suggest that previous cases should again be put to test, and conclude that state-of-the-art detection systems have still great room for improvement for the study of malicious BGP hijacks.

We are currently working further on the integration of our system so as to perform a large scale validation of similar hijack cases.