Themis: Accelerating the Detection of Route Origin Hijacking by Distinguishing Legitimate and Illegitimate MOAS

Lancheng Qin, *Tsinghua University;* Dan Li, *Tsinghua University and Zhongguancun Laboratory;* Ruifeng Li, *Tsinghua Shenzhen International Graduate School;* Kang Wang, *Tsinghua University*

https://www.usenix.org/conference/usenixsecurity22/presentation/qin

This paper is included in the Proceedings of the 31st USENIX Security Symposium.

August 10–12, 2022 • Boston, MA, USA

978-1-939133-31-1

Themis: Accelerating the Detection of Route Origin Hijacking by Distinguishing Legitimate and Illegitimate MOAS

Lancheng Qin[1] Dan Li[1,3] Ruifeng Li[2] Kang Wang[1]

[1] *Tsinghua University*
[2] *Tsinghua Shenzhen International Graduate School*
[3] *Zhongguancun Laboratory*

Abstract

Route hijacking is one of the most severe security problems in today's Internet, and route origin hijacking is the most common. While origin hijacking detection systems are already available, they suffer from tremendous pressures brought by frequent legitimate Multiple origin ASes (MOAS) conflicts. They detect MOAS conflicts on the control plane and then identify origin hijackings by data-plane probing or even manual verification. However, legitimate changes in prefix ownership can also cause MOAS conflicts, which are the majority of MOAS conflicts daily. Massive legitimate MOAS conflicts consume many resources for probing and identification, resulting in high verification costs and high verification latency in practice. In this paper, we propose a new origin hijacking system *Themis* to accelerate the detection of origin hijacking. Based on the ground truth dataset we built, we analyze the characteristics of different MOAS conflicts and train a classifier to filter out legitimate MOAS conflicts on the control plane. The accuracy and recall of the MOAS classifier are 95.49% and 99.20%, respectively. Using the MOAS classifier, *Themis* reduces 56.69% of verification costs than Argus, the state-of-the-art, and significantly accelerates the detection when many concurrent MOAS conflicts occur. The overall accuracy of *Themis* is almost the same as Argus.

1 INTRODUCTION

The Internet is composed of more than 70,000 Autonomous Systems (ASes). ASes use Border Gateway Protocol (BGP) to advertise their IP prefixes and exchange routing information with each other. However, BGP lacks authentication and validation of announced routes, which severely compromises the reliability of inter-domain routing. In other words, an AS can easily advertise invalid routing information to redirect normal routes, which is called *route hijacking*. For example, an AS can claim to be the origin for prefixes it does not own to engage in malicious activities such as traffic disruption, sending spam [54], DDoS attacks [21], or stealing crypto currencies [26]. This kind of route hijacking is called *origin hijacking*.

Origin hijacking is the most commonly observed type of route hijacking [2, 10, 16, 49, 54], which accounts for about 70% of route hijacking incidents observed by Oracle Internet Intelligence [16] and BGPmon [2]. Origin hijacking can be caused by accidental misconfigurations [19, 22] or malicious attacks [15, 17, 20], often resulting in serious routing and security problems. For instance, in 2008, Pakistan Telecom hijacked YouTube for two hours due to misconfigurations [22]. In 2017, a Russian government-controlled telecommunication company hijacked more than 20 financial services worldwide for seven minutes [15]. More recently, in Apr. 2020, Rostelecom announced massive prefixes belonging to Akamai, Amazon AWS, Cloudflare, Digital Ocean, and 200 other Internet Service Providers (ISPs), causing widespread service disruptions [20].

In order to improve the trustability and reliability of inter-domain routing, various mechanisms have been proposed to defend against origin hijacking. They can be divided into two main categories: proactive prevention [30, 39, 40, 42, 52] and reactive detection [32, 35, 41, 50, 51, 53, 55]. Proactive prevention mechanisms (*e.g.* RPKI [30]) usually use cryptography to authorize all legitimate origin ASes for the prefix in advance to prevent origin hijacking. However, these mechanisms are fully effective only when deployed by all ASes, which is a long way to go [44, 45].

Therefore, many networks prefer to rely on reactive detection mechanisms [49], which monitor BGP updates from BGP monitors worldwide and raise alarms when detecting route hijackings. Recent proposals (*e.g.* Argus [51]) combine both control-plane analysis and data-plane probing to make the detection. On the control plane, they first detect all multiple origin ASes (MOAS) conflicts based on historical BGP data. MOAS is a special phenomenon in BGP when multiple ASes originate an IP address block. It is an obvious manifestation of origin hijacking since the hijacker and the victim originate the same IP address block in BGP updates. However, although RFC 1930 [34] discourages MOAS in practical operations, it is frequently observed that legitimate changes in prefix ownership caused by business cooperation, IP address transfer/leasing, or DDoS protection services can also result in MOAS [28, 38, 56], *i.e.* legitimate MOAS. Therefore, MOAS conflicts observed on the control plane cannot simply be equated with origin hijackings but also include legitimate MOAS conflicts. Since it is difficult to differentiate origin hijackings from legitimate MOAS conflicts based on MOAS data alone [56], detection mechanisms use traceroutes/pings for each MOAS conflict and then identify origin hijackings based on data-plane reachability information.

However, existing detection approaches suffer from high verification costs and latency. Since their accuracy is sensitive to practical factors such as the location of selected probe points or the choice of the live IP address, manual verification is often required in practice [50]. Note that legitimate MOAS happens much more frequently than origin hijacking in reality. Hence, many legitimate MOAS conflicts often consume considerable resources for data-plane probing and manual verification, which significantly increases the verification costs and latency. Considering that route origin hijackings can pollute 90% of the Internet in less than two minutes [51], even a one-minute latency may cause high financial loss.

In this paper, we provide a systematic empirical analysis of the potential causes of legitimate MOAS conflicts and the behavioral features of hijackers. Based on a ground truth dataset that we construct, we identify six dominant characteristics that can separate legitimate MOAS from origin hijacking. We capture 26 features and train an Extra-Tree classifier. The accuracy and recall of the classifier are 95.49% and 99.20%, respectively. However, we note that some practical legitimate MOAS conflicts are more complicated than legitimate MOAS conflicts of our ground truth dataset, which may be mistakenly classified as origin hijackings. Therefore, we propose a new origin hijacking detection system, *i.e.* $Themis$, by adding the classifier between existing control-plane analysis and data-plane probing. The classifier is responsible for classifying MOAS conflicts into legitimate MOAS conflicts and potential hijackings. After that, $Themis$ only needs to perform data-plane probing for a small number of potential hijackings. Our evaluation shows that $Themis$ reduces verification costs by an average of 56.69% than Argus and significantly accelerates the detection when many concurrent MOAS conflicts occur. Since the classifier rarely has false negatives and $Themis$ uses the same data-plane method as Argus, the overall accuracy of $Themis$ is almost the same as Argus.

To the best of our knowledge, this is the first work that accelerates the real-time origin hijacking detection by putting a control-plane machine-learning (ML) classifier to filter out legitimate MOAS. While there are previous works [28, 37, 38, 56] which observed and analyzed MOAS, they just confirm the prevalence of legitimate MOAS and suggest several possible reasons behind special cases without delving into how legitimate MOAS differs from origin hijacking.

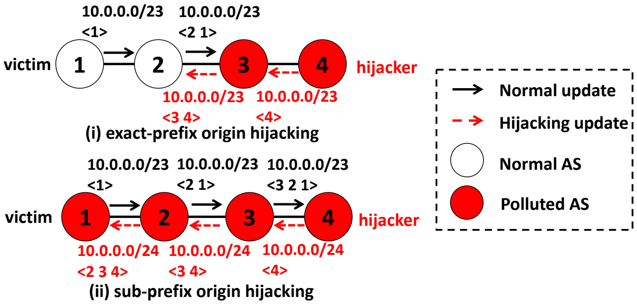


Figure 1: Examples of origin hijacking.

2 BACKGROUND AND MOTIVATION

2.1 Origin Hijacking and Legitimate MOAS

Origin Hijacking. Origin Hijacking, or illegitimate MOAS, is the most commonly observed type of route hijacking. It is usually caused by misconfigurations or malicious attacks. Origin hijacking usually does not last long, but can cause serious routing and security problems such as traffic disruption and financial losses.

Figure 1 shows two examples of origin hijacking. Assume that AS1 owns and legitimately announces prefix 10.0.0.0/23. The BGP advertisement is propagated between neighbors. The AS-path attribute $< x \ldots y >$ sequentially records the passed ASNs, where the last one is the origin AS. In the first example, AS4 is the hijacker and illegitimately announces the exact same prefix 10.0.0.0/23. As a result, AS2 and AS3 learn two routes to prefix 10.0.0.0/23 but with different origin ASes. For AS3, one route is <4> and the other is <2 1>. Since the route with a shorter AS-path is preferred in BGP, AS3 chooses the wrong route <4> and then forwards traffic destined to prefix 10.0.0.0/23 to AS4. While for AS2, it chooses the legitimate route with a shorter AS-path <1> and is not polluted by AS4. This problem, called exact-prefix origin hijacking or exact origin hijacking for short, can pollute parts of the Internet that are close to the hijacker.

The other type of origin hijacking is sub-prefix origin hijacking or sub origin hijacking for short. As shown in the second example of Figure 1, the hijacker AS4 announces a sub prefix 10.0.0.0/24 of prefix 10.0.0.0/23. Since the route with a more specific prefix is preferred, the entire Internet may be polluted. Therefore, sub origin hijacking has a more significant impact than exact origin hijacking.

Legitimate MOAS. Although RFC 1930 recommends that a prefix should originate from only one AS, a series of previous works [28, 38, 56] have observed that legitimate network engineering practices can also lead to MOAS conflicts, *i.e.* legitimate MOAS. Zhao *et al.* [56] pioneer the analysis of MOAS conflicts and discover that not all MOAS conflicts are origin hijackings but can also occur for legitimate reasons such as announcing exchange point addresses and multi-homing. They mainly focus on the duration of MOAS conflicts and conclude that most MOAS conflicts are short-lived, especially origin hijackings caused by misconfigurations, which often last less than one day. They believe that the duration can be a useful heuristic to distinguish between origin hijacking and legitimate MOAS. However, this characteristic does not apply to real-time hijacking detection. Inspired by Zhao *et al.*, Chin [28] and Jacquemart *et al.* [38] further study the prevalence and temporal characteristics of MOAS conflicts. They revisit the work of Zhao *et al.* and further confirm the prevalence of legitimate MOAS. Besides, Chin observes that multinational companies or companies hosting servers in multiple data centers may advertise the same prefixes from different origin ASes. For example, the subsidiary of Glenayre Technologies in China and its parent office in the US announce the same prefix using individual ASNs. Through this operation, traffic from China can be forwarded directly to the subsidiary rather than to the United States, and thus the latency is greatly reduced. This finding inspires us to study potential commercial relationships between origin ASes announcing the same IP address block. It is worth noting, that the use of commercial relationship information is not novel (*e.g.* Schlamp *et al.* [48] uses IRR database to build an organization graph). In this work, this information is only a part of our characteristics and is put to better use.

Previous works have proposed a few reasons (*e.g.* commercial relationships) behind legitimate MOAS conflicts, but they mainly focus on the duration of MOAS conflicts and do not systematically study the general characteristics that can accurately differentiate origin hijackings from legitimate MOAS conflicts. It is difficult to distinguish legitimate MOAS and origin hijacking in real time based on existing MOAS characteristics. Besides, they mainly study exact-prefix legitimate MOAS. Our analysis in Section 4 shows that there are significant differences between exact-prefix legitimate MOAS and sub-prefix legitimate MOAS in some characteristics. We abbreviate the two types of legitimate MOAS as exact legitimate MOAS and sub legitimate MOAS.

2.2 Motivations and Goals

Since it is hard to distinguish between origin hijackings and legitimate MOAS conflicts solely based on control-plane data, Argus performs data-plane probing on all MOAS conflicts to identify real origin hijackings. Since legitimate MOAS conflicts account for the majority of MOAS conflicts, most resources are wasted on identifying legitimate MOAS conflicts, which greatly increases verification costs and latency. To address this problem, Artemis [50] proposes a self-operated hijacking detection system. Artemis identifies origin hijackings with private information, *i.e.* all legitimate origin ASes of its prefixes. However, it only works when its prefixes are hijacked. For example, in Figure 1, even if AS3 operates Artemis, it is still unable to determine whether AS4 announcing prefix 10.0.0.0/23 is an origin hijacking or a legitimate MOAS conflict.

In this paper, we try to study general characteristics that can differentiate origin hijacking from legitimate MOAS and train a machine-learning (ML) classifier to evaluate the applicability of these characteristics. Furthermore, we propose a new origin hijacking detection system, called *Themis*, based on the classifier. For each MOAS conflict, *Themis* first captures its characteristics and then conducts data-plane probing only if it is classified as a potential hijacking by the classifier. Although it is unable to make final decisions solely based on the classifier, *Themis* significantly reduces the workload for data-plane probing and accelerates the detection of origin hijacking compared to current practices.

Low false negative rate and low false positive rate are two important targets of the classifier. False positive means that a legitimate MOAS conflict is identified as a potential hijacking. Hence, the false positive rate is proportional to the verification costs and latency of *Themis*. Compared to false positive rate, we are more concerned about the false negative rate. False negative means that a real hijacking is mistakenly considered as a legitimate MOAS conflict. In this case, the hijacker can successfully hijack traffic destined to the victim without being detected. Therefore, we should not compromise false negatives to reduce false positives.

3 DATASETS

In this section, we first introduce the ground truth used to analyze the difference between origin hijacking and legitimate MOAS. We then introduce the historical BGP data used to capture the behavior characteristics for each MOAS.

3.1 Ground truth Dataset

As described in Section 2, due to the lack of ground truth, previous works simply summarize several reasons behind limited cases. In this work, we first build a ground truth dataset including reliable origin hijackings and legitimate MOAS conflicts. In order to facilitate the analysis of the relationships between multiple origin ASes, each ground truth MOAS only contains two ASes. Table 1 describes the size of the ground truth dataset.

Origin hijackings: BGPmon [2] is a popular commercial company that provides hijacking detection services and reports observed hijacking events daily. By collecting records from BGPmon, we first extract 2,223 reported origin hijacking events occurred between February 1, 2020 and April 30, 2021. Each event has four attributes: the hijacker ASN, the victim ASN, the prefix announced by the hijacker, and the affected prefix owned by the victim.

Since BGPmon is known to feature false positives [48], we manually validate the 2,223 events using four filters, *i.e.* Route Origin Validation (ROV) filter [36], Internet Routing

Table 1: Ground truth Dataset (From February 1, 2020 to April 30, 2021).

	legitimate exact MOAS	legitimate subMOAS	exact prefix hijacking	subprefix hijacking
Number	499	1866	867	476

Registry (IRR) filter [8], topology filter [48], and MOAS duration filter. ROV is executed using Route Origin Authorizations (ROAs) which specify the prefixes that each AS is authorized to announce. IRR has recorded much prefix ownership information provided by 25 routing registries including five Regional Internet Registries (RIRs). By using the latest ROAs and IRR information, we filter out 281 events of the 2,223 events. For example, on December 21, 2020, BGPmon reported that AS 133929 hijacked two sub prefixes of prefix 45.200.0.0/16 which was normally announced by AS 134548. By looking up AS 133929 in ROAs, we find that AS 133929 has been authorized to be the valid origin of the two sub prefixes since December 29, 2020, *i.e.* eight days after the event. This fact indicates that this event is not an origin hijacking but a legitimate MOAS conflict.

Then, we use the topology filter proposed by Schlamp *et al.* and filter out other 113 events. For example, on February 20, 2020, BGPmon reported that AS 207952 hijacked a same prefix 176.100.40.0/22 of AS57439. By checking the ASN path of the suspicious BGP announcement, we find AS 57439 was located at upstream of AS 207952, indicating that AS 57439 allowed this BGP announcement to pass and approved this behavior.

Furthermore, since most origin hijackings are short-lived [28, 38, 56], we design an empirical MOAS duration filter. The filtering rule is that, within ten days after each event, if the hijacker continues to originate the affected prefix for more than three days or for a longer time than the victim does, it is considered a legitimate event. Finally, we filter out other 485 events and remain 1,344 events. For example, on October 2, 2020, BGPmon observed that AS 39404 hijacked a sub prefix 193.221.130.6/24 of prefix 193.221.128.0/19 which was normally announced by AS 35201. Although these two prefix-origin pairs are not present in the most recent global routing tables, AS 39404 had been continuously announcing prefix 193.221.130.6/24 for more than five months since October 2, 2020.

To evaluate the reliability of the 1,344 events, we try to email each victim to confirm whether it was hijacked. We find the contact of each victim through Whois [24] and send an email to each victim. The note[1] contains the time, the affected prefix, and the hijacker AS of the event. Since most of the victims' contact information is out of date or even not available, we only receive 37 replies. Six of the respondents are from the United States, and the rest are almost evenly distributed across 17 different countries including Brazil, the United Kingdom, Canada, Germany, Singapore, Australia, Seychelles and Russia. They are mainly cloud service companies (such as cloud computing services, CDN services, etc.), ISP operators, equipment vendors and research institutions.

Of the 37 replies, 36 confirm the hijacking, while one suggests it was a "probably legitimate" anycast test. We note that of the 36 confirmed hijacking events, several hijacking events were carried out by the same hijacker at the same time, which means that the hijacker was hijacking multiple victims simultaneously. We further find that the hijacker of the confirmed hijacking event was also responsible for some events of the unconfirmed 1307 events at the same time[2]. So, we consider these events as highly reliable origin hijacking events. In this way, we additionally identify 255 origin hijacking events. However, the “hijacker” of the confirmed legitimate MOAS event did not "hijack" more "victims" at the same time. We do not identify more legitimate MOAS events. We use Clopper-Pearson method to calculate the confidence intervals for the two experiments with a 95% confidence level [4]. For the first experiment (36 positive samples and 1 negative sample), the confidence interval is [0.8584, 0.9993]. For the second experiment (291 positive samples and 1 negative sample), the confidence interval is [0.9811, 0.9999].

Finally, we drop the "probably legitimate" event and take the remaining 1,343 events as ground truth origin hijackings. In addition, we also send emails for those events filtered out by the MOAS duration filter and receive 12 replies. Each confirms that it was legitimate, indicating that the MOAS duration filter is valid.

Legitimate MOAS conflicts: Unlike origin hijackings, less ground truth on legitimate MOAS conflicts is available except for limited cases described in previous works. Resource Public Key Infrastructure (RPKI) [27] is a public key infrastructure framework designed to associate prefixes with valid origin ASes, *i.e.* ROA. Although the deployment rate of RPKI is only about 30%, we find that MOAS is quite common in ROAs. We extract a total of 8,477 legitimate MOAS conflicts from ROAs, but keep only those legitimate MOAS conflicts that also appear in global routing tables to avoid misconfigurations [11]. Eventually, we extract 2,365 legitimate MOAS conflicts, including 499 exact legitimate MOAS conflicts and 1,866 sub legitimate MOAS conflicts. Each legitimate MOAS conflict has the same four attributes as the origin hijacking. In particular, in each exact legitimate MOAS conflict, we select the AS announcing the prefix later as the "hijacker" and the

[1] We show the note content in appendix.

[2] In most cases, multiple events occurred in the same second. But when a hijacker hijacked a large number of prefixes, different events might occur in different seconds, but mostly in the same minute. In this case, we consider events that occurred consecutively within a minute to be carried out at the same time.

AS announcing the prefix earlier as the "victim", which is consistent with current detection mechanisms. While in each sub legitimate MOAS conflict, the AS announcing a more specific prefix is considered the "hijacker" and the AS announcing a less specific prefix is considered the "victim".

3.2 Historical BGP Data

To capture transient characteristics when each MOAS conflict occurred, we use RIPE RIS [12] and Route Views [14] to download historical BGP data from February 1, 2020 to April 30, 2021. RIPE RIS and Route Views are two publicly available services that collect BGP data from route collectors (RCs) worldwide. Each RC collects BGP routing tables and BGP updates from its peering ASes, which are also called vantage points (VPs). We use BGPstream [3], a tool proposed by CAIDA [6], to access these BGP data. In addition, when evaluating the practical recall of MOAS classifier and the cost of *Themis*, we download historical BGP data from May 1, 2021 to September 30, 2021.

4 MOAS CHARACTERISTICS

In this section, we first study the potential relationships between origin ASes in legitimate MOAS conflicts and then develop hypotheses on hijackers' behavioral characteristics when hijacking. Based on the analysis, we identify six dominant characteristics: exact prefix or sub prefix, rank difference, business relationship, geographical relationship, announcement stability and hijacking activity.

4.1 Exact Prefix or Sub Prefix

The definitions of exact prefix and sub prefix are described in Section 2. This characteristic can be obtained naturally when an MOAS conflict is detected. Although our purpose is to distinguish origin hijacking and legitimate MOAS regardless of the affected prefix, we note that exact legitimate MOAS and sub legitimate MOAS are significantly different in some characteristics. Furthermore, we are surprised to find that sub legitimate MOAS even resembles origin hijacking in some characteristics. We intuitively assume that exact legitimate MOAS and sub legitimate MOAS are caused by different operations. In order to understand the difference between origin hijacking and legitimate MOAS, we divide MOAS into four types, *i.e.* exact legitimate MOAS, sub legitimate MOAS, exact origin hijacking, and sub origin hijacking.

4.2 Rank Difference

We collect customer cone information for each AS from CAIDA AS rank [1]. The customer cone size of an AS is the number of ASes that can be reached from the AS following only provider-to-customer (P2C) links. The AS with a

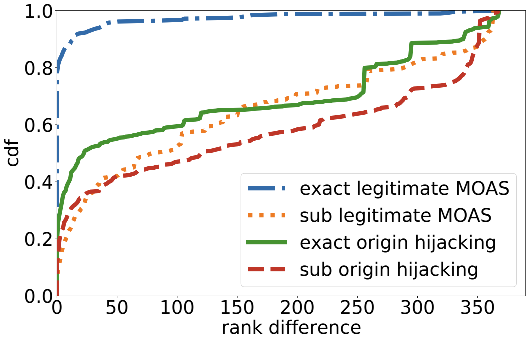


Figure 2: Rank difference of individual MOAS types.

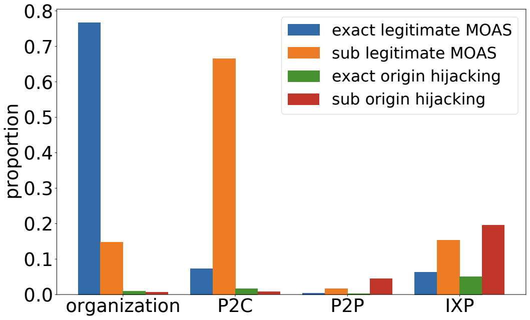


Figure 3: Proportion of different business relationships in individual MOAS types.

larger customer cone size gets a higher rank and ASes with equal customer cone sizes get the same rank. To some extent, an AS's rank reflects its influence. We are interested in the rank difference between two origin ASes in the MOAS conflict. To this end, we rank 71,665 ASes based on their customer cone sizes. We find that most Tier 1 ASes are in the top 20, where AS 3356 (a large Tier 1 AS owned by Level 3) ranks the first because it owns the largest customer cone size. While all stub ASes rank at the bottom with a rank of 366. Figure 2 shows the distribution of rank difference of the four MOAS types. We find the overall rank difference of exact legitimate MOAS is much smaller than the other three MOAS types, and 80% of exact legitimate MOAS conflicts have a rank difference of 0. We speculate that an AS is more willing to cooperate with another AS with similar influence, and they may announce the same prefixes for commercial purposes. However, this phenomenon is not prevalent in the other three types of MOAS.

Another important finding is that sub legitimate MOAS has a similar distribution to origin hijacking. Moreover, only about 10% of sub legitimate MOAS have a rank difference of 0, even lower than exact origin hijacking and sub origin hijacking. We intuitively assume that exact legitimate MOAS and sub legitimate MOAS are two different MOAS patterns.

Sub legitimate MOAS may happen when a smaller AS buys or borrows a more specific prefix from a larger AS but the larger AS still announces the less-specific prefix for convenience. Besides, we find influential ASes rarely have misconfigurations or participate in malicious attacks to maintain public credibility. In more than 65% of origin hijackings, the hijackers are stub ASes.

We also calculate the difference in customer cone sizes for each type of MOAS. The distribution is similar to Figure 2.

4.3 Business Relationship

Inspired by the results in Section 4.2 and the case in Section 2, we assume that there must be potential business relationships between ASes in legitimate MOAS conflicts. We measure the proportion of different business relationships, *i.e.* organization relationship, provider-to-customer relationship (P2C), peer-to-peer relationship (P2P), and Internet exchange Point relationship (IXP), in individual MOAS types. We use P2C and P2P information inferred by AS-Rank [1,43], and use AS organization information and IXP information provided by CAIDA [5,7].

As shown in Figure 3, there is a clear distinction between legitimate MOAS and origin hijacking in terms of the proportion of P2C relationship and organizational relationship. More specifically, the exact legitimate MOAS prefers to happen within the same organization while the sub legitimate MOAS is more likely to occur between a provider and its customer. However, P2P and IXP relationships are not as highly correlated with legitimate MOAS as expected.

It's noted that using business relationships to investigate MOAS is not novel. Schlamp *et al.* have used organization data to build a rule-based filter. The filter treats MOAS conflicts that satisfy organization relationship as legitimate MOAS conflicts and others as origin hijackings. However, the filtering rule has some limitations in accuracy. In Figure 3, 85.26% of sub legitimate MOAS conflicts and 0.78% origin hijackings do not obey this filtering rule. These origin hijackings may be caused by misconfigurations or wrong operations and can lead to serious routing problems as well. For example, on February 20, 2020, BGPmon raised an alarm that AS 20773 hijacked the prefix 173.201.64.0/23 of AS 44273. AS 20773 and AS 44273 are owned by a same German organization but are located in different regions. Particularly, AS 20773 has never announced prefix 173.201.64.0/23 before and after that day. Finally, the overall accuracy of the filter from our ground truth dataset is just 52.86%. To make better use of organization information, we feed organization relationship as one feature to an ML classifier, achieving better accuracy. The performance of the ML classifier will be described in Section 5.

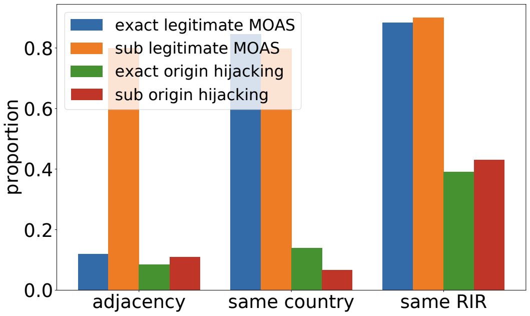


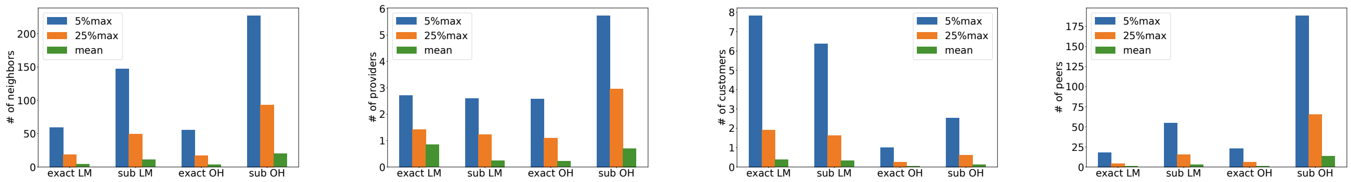
Figure 4: Proportion of different geographical relationships in individual MOAS types.

4.4 Geographical Relationship

Considering victims of origin hijackings could be located anywhere globally, we expect ASes in the legitimate MOAS conflict to be geographically close to each other. We focus on three geographical relationships: adjacency, located in the same country, and located in the same RIR. We extract adjacency relationship information from global routing tables and collect the country and RIR information of each AS from the websites of the five RIRs.

Figure 4 demonstrates our hypothesis. Most legitimate MOAS conflicts happen within the same country or the same RIR. Other legitimate MOAS conflicts may be caused by multinational companies. In particular, the proportion of adjacency relationship in sub legitimate MOAS is greatly higher than the other three MOAS types, which further reinforces our view that most sub legitimate MOAS conflicts are due to IP address transfer between providers and customers. However, some hijackers are also located near the victims, if, *e.g.*, the hijacker has a route leak or misconfiguration problem. Nonetheless, these characteristics separate origin hijackings and legitimate MOAS conflicts well.

Using business relationship information from AS-Rank and adjacency relationship information from global routing tables, we excavate more specific geographical relationships between every two origin ASes in the MOAS conflict, including the number of common neighbors, the number of common providers, the number of common customers and the number of common peers. Figure 5 shows the results. We find that even though only about 10% of hijackers are directly connected to the victims in Figure 4, some hijackers and victims may have hundreds of neighbors in common. For example, AS 3549 hijacked a sub prefix of its customer AS 14537 on March 24, 2020. Although AS 3549 ranks in the top 10 and AS 14537 just ranks 351, AS 14537 has up to 1,210 neighbors (11 providers, 1,185 peers, and 14 customers) of which 517 are also neighbors of AS 3549. On the contrary, we find ASes in legitimate MOAS conflicts generally rank lower and have



(a) number of common neighbors (b) number of common providers (c) number of common customers (d) number of common peers

Figure 5: Number of common neighbors, providers, customers, and peers in individual MOAS types (LM is the short for legitimate MOAS and OH is the short for origin hijacking).

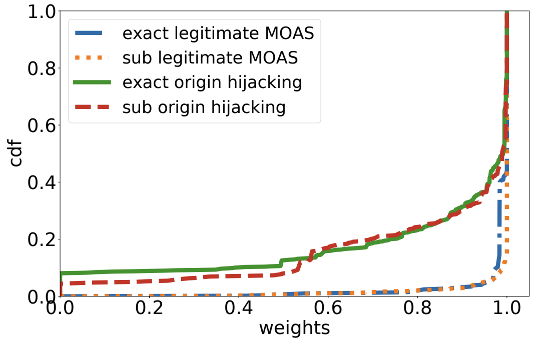


Figure 6: Average announcement activity of hijackers and legitimate ASes.

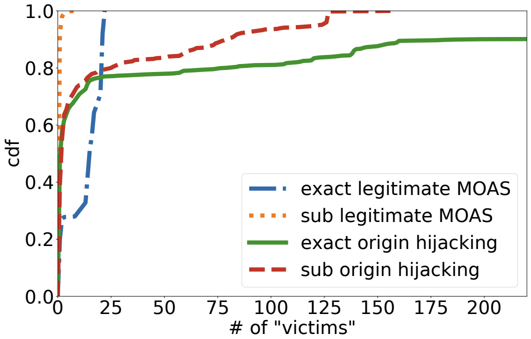


Figure 7: Hijacking activity of hijackers and legitimate ASes.

fewer neighbors in common. We also note a large proportion of common neighbors are peers as shown in Figure 5(d). This may be because P2P relationships make up the majority in AS-Rank. Besides, we cannot determine the business relationships for parts of common neighbors because our business relationship information is limited. We do not show them in Figure 5.

4.5 Announcement Activity

We note that some hijackers do not appear in global routing tables for a long time before hijacking, while legitimate ASes usually announce prefixes in daily BGP updates. Therefore, we do not expect hijackers to continuously announce a prefix for long periods, especially those who frequently launch malicious attacks. To investigate the announcement activity, we download long-term historical BGP data and locate each MOAS conflict according to the time it occurred. Unlike the origin hijacking in which the identities of the hijacker and the victim are clear, there are no hijacker and victim in a legitimate MOAS conflict. So we artificially designate the "hijacker" and the "victim" in each legitimate MOAS conflict, as described in Section 3.

For each MOAS conflict, we capture all prefixes announced by the "hijacker" in the last few days before the conflict and determine the days on which each prefix was originated from the "hijacker". We set an increasing weight for each day in chronological order, *i.e.* the weight for the Nth day is N^2, because we expect the AS announcing prefixes in more recent days to be more like a legitimate AS than an AS announcing prefixes only in earlier days. The announcement activity of each prefix is the sum of weights of the days on which it was originated from the "hijacker". The average announcement activity of the "hijacker" is the average of the announcement activity of each prefix.

For example, we calculate the average announcement activity of AS x in a 10-day historical time window. Assume AS x announced two prefixes in the past 10 days. It announced prefix 10.0.0.0/23 only in the first two days and announced prefix 11.0.0.0/23 only in the last two days. Therefore, the announcement activity of prefix 10.0.0.0/23 is 5, *i.e.* 1^2 plus 2^2. The announcement activity of prefix 11.0.0.0/23 is 181, *i.e.* 9^2 plus 10^2. Finally, the average announcement activity of AS x is an average of 5 and 181, *i.e.* 93.

Figure 6 shows the distribution of the average announcement activity of "hijackers" in a 10-day historical time window. We find that legitimate ASes almost continuously announce prefixes while the overall announcement activity of real hijackers is much lower. In addition, we also calculate announcement activity in diverse historical time windows.

4.6 Hijacking Activity

Since the hijacker in a large hijacking incident may hijack hundreds of victims at the same time, we try to measure the hijacking activity for each "hijacker" in the ground truth dataset. For each MOAS conflict, we go back to the time when it occurred and count how many different possible "victims" the "hijacker" hijacked simultaneously. Given that announcing multiple BGP updates takes time, we treat events that occurred in five seconds to be simultaneous.

Figure 7 shows the distribution of hijacking activity for each type of MOAS. A hijacker may attack hundreds of victims simultaneously while a legitimate AS is generally present in limited MOAS conflicts. Particularly, in about 10% of exact origin hijackings, hijackers attack more than 200 victims simultaneously. In the largest hijacking, on April 16, 2021, a hijacker AS 55410 advertised a large number of prefixes belonging to 1,285 different ASes simultaneously, severely affecting Internet routing. On the contrary, legitimate ASes rarely participate in more than 20 legitimate MOAS conflicts simultaneously. Similarly, we also measure how many prefixes of the "victim" are hijacked by the “hijacker” simultaneously.

Besides, we find that the victims of large hijackings may come from different countries and even different RIRs. To quantify the geographic distribution of the victims, we calculate the Gini coefficient of RIR distribution of "victims". A Gini of 1 means all victims come from one RIR. The closer the Gini coefficient is to 0, the more evenly the victims are distributed among the 5 RIRs. Figure 8 shows the distribution of Gini coefficient for each type of MOAS. We observe that origin hijacking shows a higher Gini coefficient than legitimate MOAS.

5 MOAS CLASSIFIER

Choice of Classifier: We train an ML model named MOAS classifier to evaluate the applicability of characteristics above. In particular, we choose the Extremely Randomized Trees (Extra-Trees) classifier [31] because decision trees do not require normalized data and perform well with heavy-tailed data. The Extra-Trees classifier is an ensemble ML algorithm that combines the predictions from multiple decision trees. Unlike the traditional random forest classifier which always chooses the optimum split, the Extra-Trees classifier splits nodes at random and thus greatly reduces overfitting.

Model accuracy for parameter selection: To make the most usage of the ground truth dataset, we use bootstrapping samples in the training phase of individual trees and compute the Out-Of-Bag (OOB) score. The OOB score is an established method of measuring the prediction accuracy of random forests. It is the mean prediction accuracy for each training sample t computed by the trees that do not have t in their bootstrap samples. It has been proved to be the generalization accuracy of random trees and its result is approximate to n-fold cross-validation test accuracy [33].

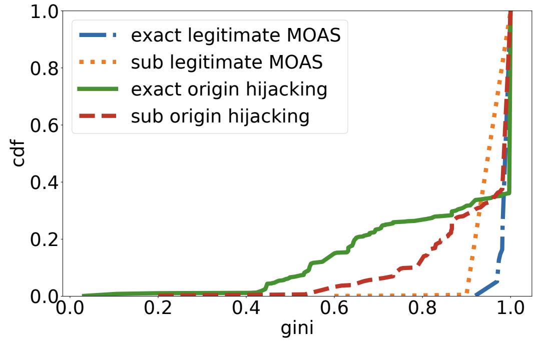


Figure 8: Gini coefficient of victims’ RIR distribution.

Feature selection and importance: We initially select 28 features that capture the six characteristics: exact prefix or sub prefix, rank difference, business relationship, geographical relationship, announcement activity and hijacking activity. For example, we check whether "hijacker" and "victim" of each MOAS conflict are located in a same country and set the result as a feature. We program our classifier using the sklearn [47] library in Python and calculate the importance of each feature using a drop-importance function. Specifically, we drop each feature and calculate the OOB accuracy score based on other features. If a feature is removed but OOB accuracy is improved, it proves that this feature compromises OOB accuracy; otherwise, it proves that this feature makes positive contributions to OOB accuracy. We find that features related to geographical relationship are generally more important than others. Besides, P2P relationship and IXP relationship also make positive contributions to OOB accuracy. Instead, the minimum announcement activity with a historical time window of 10 days and the hijacking activity with a time window of 5 seconds compromise the OOB accuracy. Finally, we select 26 features, all of which add positive OOB accuracy to our classifier. The 26 features will be detailed in the appendix and we plan to make the feature dataset and the results public to allow for reproduction.

The trained classifier: Since the number of estimators and max length are the two main parameters for decision trees, we use a grid search to find the best parameters according to the OOB accuracy. The final Extra-Trees model combines 350 estimators with a max length of 20 and the OOB accuracy score is 95.49%, much higher than the score of organization graph filter, *i.e.* 52.86%, computed in Section 4.3. We also try to use different sampling technologies on our ground truth dataset to train a better classifier but it has little improvement.

Practical performance: Due to the lack of ground truth on legitimate MOAS conflicts, we can only find limited authoritative information from RPKI. In practice, some legitimate

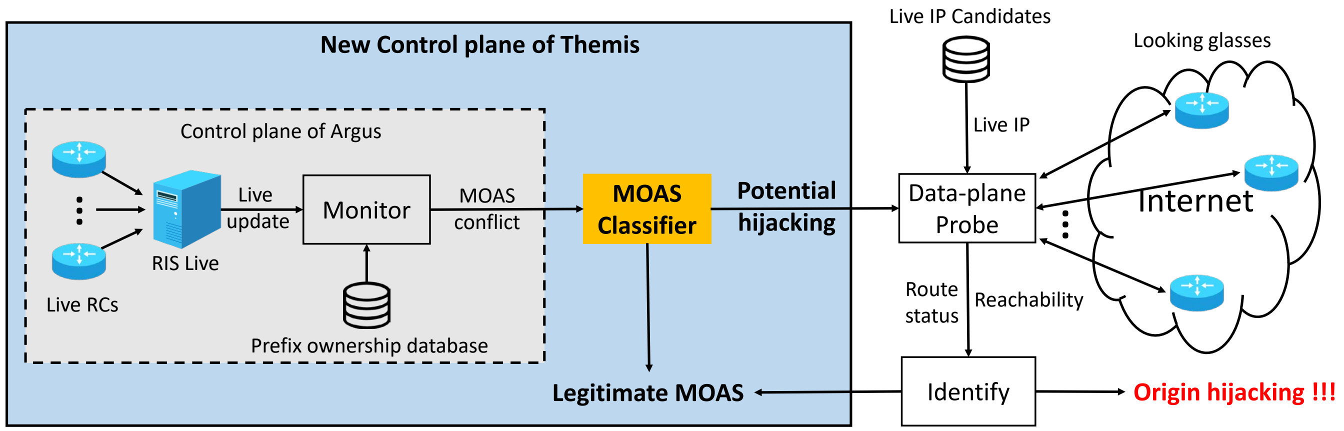


Figure 9: The architecture of Themis.

MOAS conflicts may be more complicated and MOAS classifier does not learn their characteristics. This may compromise the practical performance of MOAS classifier. Therefore, we further test it based on historical BGP data from May 1, 2021 to September 30, 2021. To evaluate the practical recall of MOAS classifier, we collect reports from BGPmon during this time and validate them using the four filters introduced in Section 3, and eventually get 376 origin hijackings. In the offline experiment, the practical recall of MOAS classifier is 99.20%. Given the confidence interval calculated in Section 3, it can be considered that MOAS classifier rarely shows false negatives in practice.

However, we find that MOAS classifier shows a higher false positive rate than the result tested from the ground truth dataset. Although it is still impractical to identify origin hijacking solely based on MOAS classifier, MOAS classifier can reduce the false positives of existing control-plane methods by 56.69%. These details will be described in Section 7. Inspired by this result, in Section 6, we design a new origin hijacking detection system *Themis* by combing MOAS classifier and data-plane probing. By using MOAS classifier, *Themis* filters out as many legitimate MOAS conflicts as possible before data-plane probing and thus significantly accelerates the detection.

6 Themis: A NEW ORIGIN HIJACKING DETECTION SYSTEM

6.1 Overview

Figure 9 shows the architecture of *Themis*. It receives live BGP updates from RIS Live [13] and detects MOAS conflicts based on the local prefix ownership database. After that, instead of directly probing all MOAS conflicts like Argus, *Themis* uses MOAS classifier to filter out as many legitimate MOAS conflicts as possible. Since MOAS classifier rarely shows false negatives, *Themis* only needs to probe and verify the rest events *i.e.* potential hijackings, and thus dramatically reduces verification costs and latency. Eventually, *Themis* combines reachability information and route status collected from looking glasses to identify origin hijackings from potential hijackings.

The biggest problem of advanced detection mechanisms is that the high false positive rate of the existing control plane greatly increases the verification cost and latency. To address this problem, *Themis* makes two main improvements on the control plane of Argus. The first is that we collect additional prefix ownership information from RPKI and IRR to optimize the monitor of Argus. The second is that we use MOAS classifier to filter out a large part of legitimate MOAS conflicts before data-plane probing, which is the main contribution of *Themis*. *Themis* uses the same data-plane probing and identification method as Argus to ensure detection accuracy.

6.2 Building Prefix Ownership Database

Prefix ownership refers to the association between a prefix and its legitimate origin ASes. Argus builds the prefix ownership database by extracting active prefix-origin pairs from the last two months of historical BGP data. Suppose we need to generate a prefix ownership database for May 1, 2021. In the same way, we extract 1,077,196 IPv4 prefixes from historical BGP data between March 1, 2021 and April 30, 2021. Although these prefixes account for 84.53% of the allocated IPv4 address space, historical BGP data does not contain allocated but unannounced prefixes. Hijackers may squat these prefixes to send spam or implement DDoS attacks. Moreover, historical prefix ownership can be outdated since current operations may change the origin ASes of prefixes. Therefore, we try to collect more prefix ownership information from RPKI and IRR.

Table 2: Statistics of different prefix ownership databases (calculated on May 1, 2021).

Database	Prefixes	IPv4 Coverage	Exclusive Prefixes
Historical BGP	1,077,196	84.53%	593,283
RPKI	186,955	26.20%	16,651
IRR (since 2011)	1,422,389	44.80%	955,758

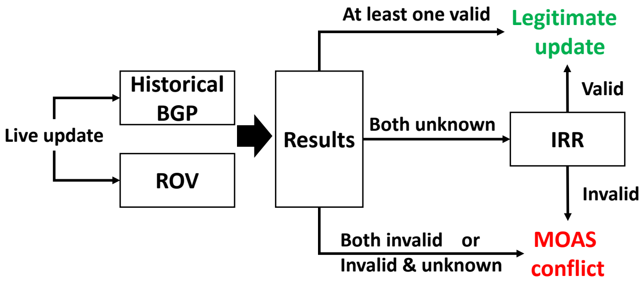


Figure 10: The work flow of Monitor.

Additional prefix ownership data: RPKI uses cryptography to ensure the authenticity of prefix ownership information in Route Origin Authorizations (ROAs) and encourages participants to update their changes in prefix ownership in advance. IRR has recorded a great deal of prefix ownership information since 1995. However, IRR is widely criticized for its outdated and inaccurate information, as it does not force providers to update their information on time. In order to avoid as much outdated data as possible, we only consider IRR data registered in the last 10 years and select the most updated AS as the unique origin of each prefix.

Table 2 shows the statistics of historical BGP data, RPKI, and IRR. As of May 1, 2021, RPKI recorded 186,955 IPv4 prefixes and IRR recorded 1,422,389 IPv4 prefixes, accounting for 26.20% and 44.80% of the allocated IPv4 address space, respectively. Although the deployment rate of RPKI was just 26.20%, it still had 16,651 exclusive prefixes. For instance, prefix 192.188.82.0/23 is associated with four valid origin ASes (AS 17444, AS 9269, AS 10103 and AS 9381) in RPKI, but prefix 192.188.82.0/23 never appears in BGP and IRR (since 2011).

6.3 Monitoring MOAS Conflicts

We use RIPE RIS Live [13] which normally provides live BGP updates except for some abnormal cases to monitor BGP updates. To drop outdated BGP updates, we only accept BGP updates originated within 120 seconds since the Internet often converges in less than 2 minutes. When receiving a live BGP update, the monitor checks whether the prefix and the origin AS of the received BGP update are consistent with local prefix ownership databases. *Themis* optimizes the monitor of Argus by combing historical BGP validation, ROV, and IRR validation to identify as many legitimate BGP updates as possible. It is worth noting that although ROV is considered to be highly reliable, recent researches [11, 30] have shown that it also has serious false positives due to misconfiguration or partial deployment.

Figure 10 illustrates the workflow of the monitor. The "valid", "invalid", and "unknown" validity states of historical BGP validation and IRR validation are the same as those of ROV [27], except that the max length attribute of each IPv4 prefix in historical BGP data and IRR data is considered as 32. To avoid outdated data from IRR, the monitor only uses IRR's exclusive prefixes. Upon receiving a live BGP update, the monitor first checks the validity of its prefix and origin AS using historical BGP validation and ROV, respectively. If their decisions have at least one "valid" state, the BGP update is legitimate. If there are two "invalid" states, or one "invalid" state and one "unknown" state, the monitor outputs an MOAS conflict. Only when both states are "unknown" will the monitor use IRR validation to check the validity and make the final decision. Our evaluation shows that the optimized monitor can effectively reduce 15.27% false positives of the traditional monitor which only uses historical BGP data.

Noted that an MOAS conflict can be composed of one hijacker and multiple victims if the announced prefix is owned by multiple ASes in the prefix ownership database. Each MOAS conflict will be further verified by MOAS classifier. It is also possible that the prefix and its less-specific prefixes are not in any prefix ownership database, *i.e.* the final result of IRR validation is still "unknown" state. In this case, *Themis* will directly probe and identify the authenticity of this BGP update. In practice, this case seldom happens because the three prefix ownership databases have contained more than two million prefixes, covering more than 90% of allocated IPv4 address space.

6.4 Filtering Legitimate MOAS

This module is the core and the main contribution of *Themis*. Upon receiving MOAS conflicts from the monitor, it classifies them into potential hijackings and legitimate MOAS conflicts using the MOAS classifier trained in Section 5. If there are multiple victims in one MOAS conflict, *Themis* will separately determine whether the hijacker and each victim are legitimate MOAS. The MOAS conflict will be considered a legitimate MOAS conflict if the hijacker and at least one victim are deemed to announce the prefix legitimately. Otherwise, it will be considered a potential hijacking. Since MOAS classifier rarely shows false negatives in practice, *Themis* only needs to further probe these potential hijackings, a small subset of all MOAS conflicts, to identify real origin hijackings.

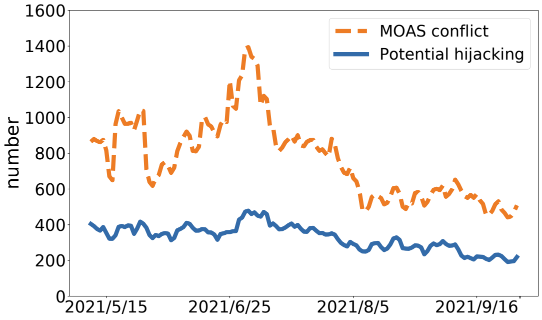


Figure 11: The 10-day moving average of MOAS conflicts and potential hijackings (from May 1, 2021 to September 30, 2021).

6.5 Probing and Identifying Origin Hijacking

This module reproduces the data plane and identification method of Argus. For each potential hijacking, *Themis* first selects the most appropriate IP address of the prefix from a list of candidate live IP addresses which are collected by Zmap [25] offline. *Themis* then uses all available looking glasses [9] which provide public service for probing to test whether the IP address is reachable. Meanwhile, it also checks whether the looking glasses are affected by the suspicious BGP announcement. Finally, *Themis* calculates a fingerprint based on control-plane route status and data-plane reachability to determine whether the potential hijacking is a real origin hijacking. We do not go into the details of this part of work since they are explicitly described in the Argus paper. Since MOAS classifier rarely has false negatives and *Themis* uses the same data plane and identification method as Argus, the overall accuracy of *Themis* is almost the same as Argus.

7 EVALUATION

In this section, we evaluate the cost and latency of *Themis* and compare them with those of Argus. Since *Themis* uses the same data plane as Argus, the only thing that can compromise detection accuracy is the false negative rate of *Themis*'s MOAS classifier. Although the accuracy and recall of MOAS classifier have been evaluated in Section 5 and it rarely has false negative problems, we further evaluate the false negative rate of *Themis* and propose a priority *Themis* which can achieve 0 false negative.

7.1 Cost of Themis

To test the cost of Themis, we download historical BGP updates between May 1, 2021 and September 30, 2021 using 95 full VPs of RIPE RIS. The 95 full VPs are distributed across the five RIRs, thus forming a global perspective. During this time, *Themis* detects 112,655 MOAS conflicts and classifies 48,794 as potential hijackings. Figure 11 shows the 10-day moving average of MOAS conflicts and potential hijackings. We find that MOAS classifier reduces MOAS conflicts by an average of 56.69%. Especially, it reduces MOAS conflicts by up to 88.24% on June 25, 2021. Note that the 56.69% reduction is separate from the 15.27% reduction mentioned in Section 6.

Table 3: Distribution of the number of concurrent MOAS conflicts in alerts (from November 1, 2021 to December 31, 2021).

# of concurrent MOAS conflicts	Proportion
1	74.33%
2	11.18%
3	4.56%
4	2.94%
≥ 5	6.99%

Table 4: Distribution of the number of concurrent potential hijackings in alerts (from November 1, 2021 to December 31, 2021).

# of concurrent potential hijackings	Proportion
0	32.59%
1	51.05%
2	7.28%
3	2.84%%
4	1.80%
≥ 5	4.44%

Moreover, we note that MOAS conflicts do not occur evenly. It is frequently observed that multiple MOAS conflicts occur simultaneously and existing detection approaches have to verify every conflict using data-plane probing, which is the main cause of high latency. Therefore, we online run *Themis* for 60 days from November 1, 2021 to December 30, 2021 and focus on the real-time verification cost and latency, especially when dealing with concurrent MOAS conflicts. Here, we give the definition of an *alert*: an alert contains all MOAS conflicts that occurred at the same time.

We find *Themis* still works well when handling concurrent MOAS conflicts, since MOAS classifier filters out most concurrent legitimate MOAS conflicts. Table 3 shows the distribution of the number of concurrent MOAS conflicts in alerts. Most alerts contain less than 5 concurrent MOAS conflicts, but 6.99% of alerts contain more concurrent MOAS conflicts with a max peak of 59. We call those alerts containing more than 4 concurrent MOAS conflicts *crucial alerts*, since they usually cause high verification latency in existing hijacking detection systems. As a contrast, Table 4 shows the distribution of the number of concurrent potential hijackings in the same alerts. The comparison shows that MOAS classi-

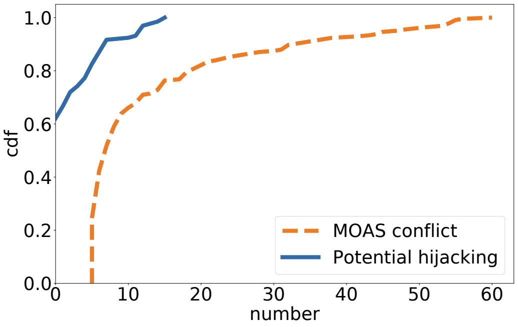


Figure 12: The number of concurrent MOAS conflicts and potential hijackings in *crucial alerts* (from November 1, 2021 to December 31, 2021).

fier significantly reduces the number of concurrent conflicts that should be probed and identified in most cases. What's more, in 32.59% of alerts, the number of potential hijackings is 0. More specifically, we find 93.8% of these alerts come from alerts containing less than 5 MOAS conflicts and the rest 6.2% come from *crucial alerts*.

We particularly measure the distribution of the number of concurrent MOAS conflicts and potential hijackings in all *crucial alerts* during the 60 days. As shown in Figure 12, the classifier greatly reduces the costs for probing when dealing with *crucial alerts*. In more than 60% of *crucial alerts*, all concurrent MOAS conflicts are classified as legitimate MOAS conflicts, so *Themis* does not need to probe in these cases. Besides, *crucial alerts* with less than 5 potential hijackings account for more than 80%. Therefore, *Themis* not only reduces verification costs by an average of 56.69%, but also significantly reduces verification costs when dealing with *crucial alerts*.

Combination of MOAS classifier and topology filter: We are interested in whether adding a topology filter to the control plane of *Themis* can improve the performance. We first replace MOAS classifier with a topology filter and find that topology filter can reduce verification costs by 27.75% on average. Then, we combine MOAS classifier with topology filter. Our evaluation shows that this combination reduces verification costs by an additional 14.10% on the basis of MOAS classifier, indicating that using topology filter can filter out more legitimate MOAS conflicts. It is highly recommended that the topology filter be introduced to *Themis*. However, to evaluate the performance of MOAS classifier separately, topology filter is not considered in this section.

7.2 Latency of Themis

We also run Argus during the same 60 days to compare the detection latency of *Themis* with Argus. *Themis* only probes and verifies potential hijackings, while Argus needs to probe and verify all MOAS conflicts. In the real world, probing always has high costs in time because of the limitations of publicly available looking glasses and the time-consuming nature of ping operation. We cannot probe multiple IP addresses simultaneously by multi-threading since existing looking glass services limit the frequency of access. Therefore, we run *Themis* and Argus in a practical way that a conflict should wait to be probed until the previous one is fully identified. The time used for logging in to existing looking glasses and probing an IP address ranges from 15 to 25 seconds. Figure 13 shows the distribution of detection latency (*i.e.* from the time when the BGP update is observed to the time when it is identified as an origin hijacking) of *Themis* and Argus. Overall, the detection latency of *Themis* is much shorter than Argus. It is because *Themis* generally probes much fewer events than Argus, especially when dealing with *crucial alerts*.

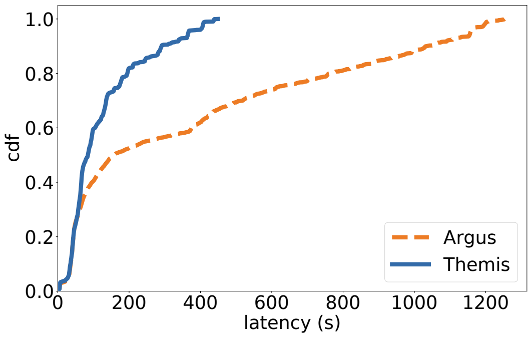


Figure 13: Detection latency of Themis and Argus (from November 1, 2021 to December 31, 2021).

7.3 False negative of Themis

Although MOAS classifier rarely shows false negatives in the offline experiment, we further evaluate the false negative rate of *Themis* in the 60-day online experiment.

To this end, we verify all MOAS conflicts including the legitimate MOAS conflicts classified by MOAS classifier. In the 60-day online experiment, we finally identify 146 exact origin hijackings and 10 sub origin hijackings. By comparing with the results of *Themis*, we find that only 1 origin hijacking is mistakenly considered as a legitimate MOAS conflict by *Themis*. The false negative rate of *Themis* is extremely low, *i.e.* 0.65%.

Trade-offs between false negative and probing cost: Although it misses 1 origin hijacking during the 60 days, we argue that, in practice the real-time performance of origin hijacking detector outweighs the little false negative. To make up for the loss, we also design an alternative implementation of *Themis*, namely priority *Themis*. In priority *Themis*, the legitimate MOAS conflicts determined by MOAS classifier should also be probed with a lower priority than potential hijackings. In this way, priority *Themis* can guarantee 0 false negative and achieves almost the same detection latency as the ordinary *Themis*. As the first effort to distinguish legitimate MOAS and potential hijackings in real time based on control-plane MOAS characteristics, *Themis* meets the expectations in Section 2.2.

8 RELATED WORK

Existing hijacking detection approaches can be classified into three categories based on the type of information they use.

Control-plane approaches [32, 41, 50] passively monitor BGP updates and detect changes in the origin of a prefix. They simply consider all MOAS conflicts as origin hijackings, resulting in high false positives. To address this problem, a self-operated control-plane approach Artemis [50] is proposed. Artemis can accurately differentiate origin hijackings from legitimate MOAS conflicts with private knowledge. However, it is unable to verify the authenticity of routes for prefixes it does not own. Imai *et al.* [37] try to determine the risk for each MOAS conflict based on fixed rules. However, their division of MOAS conflicts is coarse-grained. Papadopoulos *et al.* [46] and Cho *et al.* [29] also try to train a ML classifier to distinguish legitimate MOAS from origin hijacking. However, they are either not suitable for real-time detection or do not capture the general characteristics of legitimate MOAS.

Data-plane approaches [53, 55] continuously use traceroutes or pings to probe fixed IP addresses and raise alarms when they detect changes in the reachability or data paths. By analysing the reachability information, these approaches achieve high accuracy. However, they suffer from poor scalability, since they require a large number of active measurements. Moreover, they cannot detect sub origin hijackings because they only probe a few specific IP addresses for each prefix.

Hybrid approaches [35, 51] combine control-plane approaches and data-plane approaches. The state-of-the-art, Argus [51], first detects all MOAS conflicts on the control plane. Then, Argus probes each MOAS conflict to identify real origin hijackings. Due to the high positives on control plane, Argus wastes lots of resources to probe and identify the large number of legitimate MOAS conflicts, which greatly increases verification costs and detection latency. Schlamp *et al.* [48] propose three rule-based filters to identify legitimate MOAS, *i.e.* relationship graph filter, topology filter, and SSL/TLS filter. As described above, the performance of the first two filters is not as good as our MOAS classifier. The latter filter requires additional data-plane scans, increasing verification cost and latency.

9 DISCUSSION

Limitations: Legitimate MOAS conflicts in ground truth are not generalized enough, so data-plane is still needed in Themis. It is because less ground truth on legitimate MOAS conflicts is available except for RPKI. MOAS classifier distinguishes legitimate MOAS and potential hijacking solely based on the explicit characteristics described in Section 4. A sophisticated origin hijacking may evade detection by masquerading as a legitimate MOAS conflict. Even so, these characteristics can greatly limit the target scope and hijacking activity of the hijacker. Another limitation of this work is that Themis only focuses on origin hijackings. The detection for other types of route hijacking will be considered in future work. In addition, we use the same data plane and identification method from Argus, which means that *Themis* is unable to detect the hijacking if the attracted traffic is manipulated or eavesdropped and then sent to the victim. Given this situation, manual verification is still necessary. However, it is worth noting that *Themis* can greatly accelerate the detection compared to Argus.

Future work: In the future, we plan to collect more ground truth data to improve the practical performance of MOAS classifier. Our ultimate goal is to completely remove the data-planing probing and identify origin hijacking by MOAS classifier only. Besides, we plan to establish a comprehensive hijacking detection system for all types of route hijacking, including AS-path hijacking, policy hijacking and hybrid hijacking. We also plan to monitor BGP updates with IPv6 prefixes since IPv6 has become more and more popular on the Internet. Furthermore, we plan to use additional valuable information to better identify origin hijacking. For example, the Don't Route Or Peer List (DROP) [23] contains a set of prefixes and ASNs that are controlled by spammers or cyber criminals. BGP updates with these prefixes and ASNs are most likely malicious attacks. The Mutually Agreed Norms for Routing Security (MANRS) [10] is a global initiative to secure the Internet, so its participants are less likely to participate in origin hijackings. In addition, the transfer information of IP and ASNs [18] published by five RIRs can be greatly useful as well.

10 CONCLUSION

In this work, we investigate the difference between legitimate MOAS and origin hijacking. We use six dominant characteristics that can accurately separate them. Based on the characteristics, we train the MOAS classifier to distinguish legitimate MOAS and potential hijacking automatically. We propose a new origin hijacking detection system, *Themis*, based on MOAS classifier. The accuracy and recall of MOAS classifier are 95.49% and 99.20%, respectively. By using the classifier, *Themis* greatly accelerates the detection of origin hijackings and achieves almost the same accuracy as Argus.

Acknowledgements

We thank our shepherd Ralph Holz and the anonymous reviewers for their thoughtful comments. Dan Li is the corresponding author. This work is supported by the National Key Research and Development Program of China under Grant 2018YFB1800600.

References

[1] As rank. https://asrank.caida.org/.

[2] Bgpmon (commercial). https://bgpstream.com/.

[3] Bgpstream. https://bgpstream.caida.org/.

[4] Binomial proportion confidence interval. https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval.

[5] Caida internet exchange points (ixps) dataset. https://www.caida.org/catalog/datasets/ixps/.

[6] The center for applied internet data analysis (caida). https://www.caida.org/.

[7] Inferred as to organization mapping dataset. https://www.caida.org/catalog/datasets/as-organizations/.

[8] Internet routing registry (irr). http://www.irr.net/.

[9] Looking glass. https://lg.he.net/.

[10] Mutually agreed norms for routing security (manrs). https://www.manrs.org/.

[11] Resource public key infrastructure (rpki) technical analysis. https://www.icann.org/en/system/files/files/octo-014-02sep20-en.pdf.

[12] Ripe ris (routing information service). https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data.

[13] Ris live. https://ris-live.ripe.net/.

[14] The route views project. http://www.routeviews.org/routeviews/.

[15] Russian-controlled telecom hijacks financial services' internet traffic. https://arstechnica.com/information-technology/2017/04/.

[16] https://blogs.oracle.com/internetintelligence/.

[17] https://radar.qrator.net/blog/as1221-hijacking-266asns.

[18] https://www.apnic.net/manage-ip/manage-resources/transfer-resources/.

[19] https://www.bgpmon.net/chinese-isp-hijacked-10-of-the-internet/.

[20] https://www.manrs.org/2020/04/not-just-another-bgp-hijack/.

[21] https://www.netscout.com/what-is-ddos/bgp-hijacking.

[22] https://www.ripe.net/publications/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study.

[23] https://www.spamhaus.org/drop/.

[24] Whois. https://who.is/.

[25] Zmap. https://zmap.io/.

[26] Maria Apostolaki, Aviv Zohar, and Laurent Vanbever. Hijacking bitcoin: Routing attacks on cryptocurrencies. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 375–392. IEEE, 2017.

[27] R Bush. Origin validation operation based on the resource public key infrastructure (rpki). *IETF RFC7115 (January 2014)*, 2014.

[28] Kwan-Wu Chin. On the characteristics of bgp multiple origin as conflicts. In *2007 Australasian Telecommunication Networks and Applications Conference*, pages 157–162. IEEE, 2007.

[29] Shinyoung Cho, Romain Fontugne, Kenjiro Cho, Alberto Dainotti, and Phillipa Gill. Bgp hijacking classification. In *2019 Network Traffic Measurement and Analysis Conference (TMA)*, pages 25–32. IEEE, 2019.

[30] Taejoong Chung, Emile Aben, Tim Bruijnzeels, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, Roland van Rijswijk-Deij, John Rula, et al. Rpki is coming of age: a longitudinal study of rpki deployment and invalid route origins. In *Proceedings of the Internet Measurement Conference*, pages 406–419, 2019.

[31] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[32] Andreas Haeberlen, Ioannis C Avramopoulos, Jennifer Rexford, and Peter Druschel. Netreview: Detecting when interdomain routing goes wrong. In *NSDI*, volume 2009, pages 437–452, 2009.

[33] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[34] John Hawkinson and Tony Bates. Rfc1930: Guidelines for creation, selection, and registration of an autonomous system (as), 1996.

[35] Xin Hu and Z Morley Mao. Accurate real-time identification of ip prefix hijacking. In *2007 IEEE Symposium on Security and Privacy (SP'07)*, pages 3–17. IEEE, 2007.

[36] Geoff Huston and George Michaelson. Validation of route origination using the resource certificate public key infrastructure (pki) and route origin authorizations (roas), 2012.

[37] Hironori Imai, Masayuki Okada, and Akira Kanaoka. Poster: Moai: Multiple origin ases identification for ip prefix hijacking and mis-origination.

[38] Quentin Jacquemart, Guillaume Urvoy-Keller, and Ernst Biersack. A longitudinal study of bgp moas prefixes. In *International Workshop on Traffic Monitoring and Analysis*, pages 127–138. Springer, 2014.

[39] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Pretty good bgp: Improving bgp by cautiously adopting routes. In *Proceedings of the 2006 IEEE International Conference on Network Protocols*, pages 290–299. IEEE, 2006.

[40] Stephen Kent, Charles Lynn, and Karen Seo. Secure border gateway protocol (s-bgp). *IEEE Journal on Selected areas in Communications*, 18(4):582–592, 2000.

[41] Mohit Lad, Daniel Massey, Dan Pei, Yiguo Wu, Beichuan Zhang, and Lixia Zhang. Phas: A prefix hijack alert system. In *USENIX Security symposium*, volume 1, page 3, 2006.

[42] Matt Lepinski and Kotikalapudi Sriram. Bgpsec protocol specification. *Internet Engineering Task Force (IETF)*, 2017.

[43] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and KC Claffy. As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256, 2013.

[44] Robert Lychev, Sharon Goldberg, and Michael Schapira. Bgp security in partial deployment: Is the juice worth the squeeze? In *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, pages 171–182, 2013.

[45] Stephanos Matsumoto, Raphael M Reischuk, Pawel Szalachowski, Tiffany Hyun-Jin Kim, and Adrian Perrig. Authentication challenges in a global environment. *ACM Transactions on Privacy and Security (TOPS)*, 20(1):1–34, 2017.

[46] Stavros Papadopoulos, Konstantinos Votis, Christos Alexakos, and Dimitrios Tzovaras. Feature extraction and visual feature fusion for the detection of concurrent prefix hijacks. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 310–319. Springer, 2014.

[47] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[48] Johann Schlamp, Ralph Holz, Quentin Jacquemart, Georg Carle, and Ernst W Biersack. Heap: reliable assessment of bgp hijacking attacks. *IEEE Journal on Selected Areas in Communications*, 34(6):1849–1861, 2016.

[49] Pavlos Sermpezis, Vasileios Kotronis, Alberto Dainotti, and Xenofontas Dimitropoulos. A survey among network operators on bgp prefix hijacking. *ACM SIGCOMM Computer Communication Review*, 48(1):64–69, 2018.

[50] Pavlos Sermpezis, Vasileios Kotronis, Petros Gigis, Xenofontas Dimitropoulos, Danilo Cicalese, Alistair King, and Alberto Dainotti. Artemis: Neutralizing bgp hijacking within a minute. *IEEE/ACM Transactions on Networking*, 26(6):2471–2486, 2018.

[51] Xingang Shi, Yang Xiang, Zhiliang Wang, Xia Yin, and Jianping Wu. Detecting prefix hijackings in the internet with argus. In *Proceedings of the 2012 Internet Measurement Conference*, pages 15–28, 2012.

[52] Lakshminarayanan Subramanian, Volker Roth, Ion Stoica, Scott Shenker, and Randy Katz. Listen and whisper: Security mechanisms for bgp. In *Proc. NSDI*, volume 4. Citeseer, 2004.

[53] Mitsuho Tahara, Naoki Tateishi, Toshio Oimatsu, and Souhei Majima. A method to detect prefix hijacking by using ping tests. In *Asia-Pacific Network Operations and Management Symposium*, pages 390–398. Springer, 2008.

[54] Pierre-Antoine Vervier, Olivier Thonnard, and Marc Dacier. Mind your blocks: On the stealthiness of malicious bgp hijacks. In *NDSS*, 2015.

[55] Zheng Zhang, Ying Zhang, Y Charlie Hu, Z Morley Mao, and Randy Bush. ispy: Detecting ip prefix hijacking on my own. In *Proceedings of the ACM SIGCOMM 2008 conference on Data Communication*, pages 327–338, 2008.

[56] Xiaoliang Zhao, Dan Pei, Lan Wang, Dan Massey, Allison Mankin, S Felix Wu, and Lixia Zhang. An analysis of bgp multiple origin as (moas) conflicts. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pages 31–35, 2001.

Appendix A FEATURE SET

Table 5: The feature set (sorted by feature importance).

same country
adjacency relationship
P2C relationship
exact prefix or sub prefix
same RIR
organization relationship
average announcement activity in a 10-day historical time window
rank of "hijacker"
hijacking activity in 2 seconds
customer cone size of "hijacker"
customer cone size of "victim"
hijacking activity in 1 second
customer cone size difference
Gini coefficient
number of common neighbors
number of common providers
max announcement activity in a 10-day historical time window
rank of "victim"
rank difference
number of common peers
IXP relationship
number of affected prefixes of the "victim" in 2 seconds
number of affected prefixes of the "victim" in 1 seconds
number of affected prefixes of the "victim" in 5 seconds
number of common customers
P2P relationship

Appendix B THE NOTE CONTENT

Hello, I am a researcher from Tsinghua University. Our research focuses on monitoring BGP hijacking and maintaining Internet routing security. We observed a possible BGP hijacking event. <prefix a> was normally announced by <AS x>. However, at <time>, the same prefix (or a more specific prefix) <prefix b> was announced by <AS y>.

We find this contact of the victim <AS x> through Whois and we would like to confirm whether this was a legitimate operation or BGP hijacking. We look forward to your reply, as it is very important for both your routing security and our research.