# Detection of Malicious Activity at Autonomous System and Their Internet Connectivity

Tharun Kumar Sarraya
M.Tech Student, Dept. Of CSE,
KMM Institute of Technology & Sciences ,
Tirupati,India.

P. Venkata Maheswara
Asst. Professor, Dept. Of CSE,
KMM Institute of Technology & Sciences ,
Tirupati, India.

***Abstract:*-While many attacks are distributed across botnets, investigators and network operators have recently identified malicious networks through high profile autonomous system (AS) de-peering and network shut-downs. In this paper, we explore whether some ASes indeed are safe havens for malicious activity. We look for ISPs and ASes that exhibit disproportionately high malicious behavior using ten popular blacklists, plus local spam data, and extensive DNS resolutions based on the contents of the blacklists. We find that some ASes have over 80% of their routable IP address space blacklisted. Yet others account for large fractions of blacklisted IP addresses. Several ASes regularly peer with ASes associated with significant malicious activity. We also find that malicious ASes as a whole differ from benign ones in other properties not obviously related to their malicious activities, such as more frequent connectivity changes with their BGP peers. Overall, we conclude that examining malicious activity at AS granularity can unearth networks with lax security or those that harbor cybercrime.**

## I. INTRODUCTION

The Internet is plagued by malicious activity, from spam and phishing to malware and denial-of-service (DoS) attacks. Much of it thrives on armies of compromised hosts, or botnets, which are scattered throughout the Internet. However, malicious activity is not necessarily evenly distributed across the Internet: some networks may employ lax security, resulting in large populations of compromised machines, while others may tightly secure their network and not have any malicious activity. Further, some networks may exist solely to engage in malicious activity. Several recent ISP enforcement actions, such as the Atrivo and McColo autonomous system (AS) de-peering, and the FTC closure of Pricewert networks, highlight that there are networks that exist simply to launch attacks. In this paper, we examine whether we can find malicious networks in a systematic manner using existing blacklists. Also, when receiving traffic, a destination network could prioritize traffic based on the cleanliness of ASes, which the metrics can help estimate. This would allow a network under attack to prioritize traffic that is less likely to be associated with attackers. Finally, such metrics could also aid spam filtering programs in their scoring of email messages.

ISPs, governments and larger organisations often use Border Gateway Protocol (BGP) and Autonomous System numbers to aggregate two or more disparate networks into a single Internet entity for efficiency and to assist in building redundancy.

An Autonomous System (AS) is a collection of networks, or routers, administered as a group and sharing a common set of routing policies. AS numbers, the unique identifiers for these groups, are managed along with IP addresses by the five RIRs (Regional Internet Registries). Autonomous System numbers are a vital part of the Internet's core routing System, Border Gateway Protocol (BGP).

Together with BGP, AS numbers indicate, with a single number, the way Internet traffic moves within and between the multiple networks managed as a single autonomous whole. This greatly improves internetworking efficiency and is a key enabler in multihoming networks to provide redundant Internet connections.

Multihomed networks must have their own public IP address range and an AS number to provide failover between multiple connections to their ISPs. BGP is then used to provide routing services within the Autonomous System and to the ISPs.

On January 2007, as part of a globally coordinated policy to begin the transition to 4-byte AS numbers, the five RIRs - AfriNIC, APNIC, ARIN, LACNIC and RIPE NCC - began assigning 4-bytes AS numbers upon request.

From 1 Jan 2010, APNIC has ceased to make any distinction between two-byte and four-byte when assigning AS Numbers. Network operators may find themselves with a new 4-byte AS number that their upstream provider cannot recognize.

• A large fraction of routable space is malicious for some ASes: Four ISPs, 2 from Ukraine, one from Iran, and one from Belarus, have over 80% of their routable IP addresses blacklisted. This raises concerns regarding the purpose of such ISPs.

• Some ASes account for significantly large fractions of blacklists: Four ASes, three of which are US-based hosting providers and one large broadband service provider in Turkey, account for over 6% of at least one of the blacklists we tested.

• Some providers regularly peer with malicious ASes: We find 22 provider ISPs with 100% of their customer ASes engaged in significant malicious activity.

• Malicious ASes differ from benign ones in other ways: They are more likely to become completely unreachable than those which have less malicious activity, and they are likely to have more peers. However, the duration of unreachability is short for these ASes, which may have 2 implications for orchestrated de-peering attempts.

Overall, these results confirm that examining malicious activity at the AS granularity can help find networks that are disproportionately bad, providing a metric for focusing network clean-up efforts.

## II. AUTONOMOUS SYSTEM CHARACTERISTICS

The classic definition of an Autonomous System is a set of routers under a single technical administration, using an interior gateway protocol (IGP) and common metrics to determine how to route packets within the AS, and using an inter-AS routing protocol to determine how to route packets to other ASs. Since this classic definition was developed, it has become common for a single AS to use several IGPs and sometimes several sets of metrics within an AS. The use of the term Autonomous System here stresses the fact that, even when multiple IGPs and metrics are used, the administration of an AS appears to other ASs to have a single coherent interior routing plan and presents a consistent picture of what destinations are reachable through it._InternetTherefore, we do not expect all malicious ASes to have the same properties as each other or for there to be no overlap with good ASes. However, we do hope to see trends in the characteristics of malicious ASes.

Within the Internet, an **Autonomous System** (**AS**) is a collection of connected Internet Protocol (IP) routing prefixes under the control of one or more network operators that presents a common, clearly defined routing policy to the Internet.[1]

Originally the definition required control by a single entity, typically an Internet service provider or a very large organization with independent connections to multiple networks, that adhere to a single and clearly defined routing policy, as originally defined in RFC 1771.[2] The newer definition in RFC 1930 came into use because multiple organizations can run BGP using private AS numbers to an ISP that connects all those organizations to the Internet. Even though there may be multiple Autonomous Systems supported by the ISP, the Internet only sees the routing policy of the ISP. That ISP must have an officially registered **Autonomous System Number** (**ASN**).

A unique ASN is allocated to each AS for use in BGP routing. AS numbers are important because the ASN uniquely identifies each network on the Internet.

### *A. BGP Behavior*

BGP is a very robust and scalable routing protocol, as evidenced by the fact that BGP is the routing protocol employed on the Internet. At the time of this writing, the Internet BGP routing tables number more than 90,000 routes. To achieve scalability at this level, BGP uses many route parameters, called attributes, to define routing policies and maintain a stable routing environment.

In addition to BGP attributes, classless interdomain routing (CIDR) is used by BGP to reduce the size of the Internet routing tables. For example, assume that an ISP owns the IP address block 195.10.x.x from the traditional Class C address space. This block consists of 256 Class C address blocks, 195.10.0.x through 195.10.255.x. Assume that the ISP assigns a Class C block to each of its customers. Without CIDR, the ISP would advertise 256 Class C address blocks to its BGP peers. With CIDR, BGP can supernet the address space and advertise one block, 195.10.x.x. This block is the same size as a traditional Class B address block. The class distinctions are rendered obsolete by CIDR, allowing a significant reduction in the BGP routing tables.

BGP neighbors exchange full routing information when the TCP connection between neighbors is first established. When changes to the routing table are detected, the BGP routers send to their neighbors only those routes that have changed. BGP routers do not send periodic routing updates, and BGP routing updates advertise only the optimal path to a destination network.

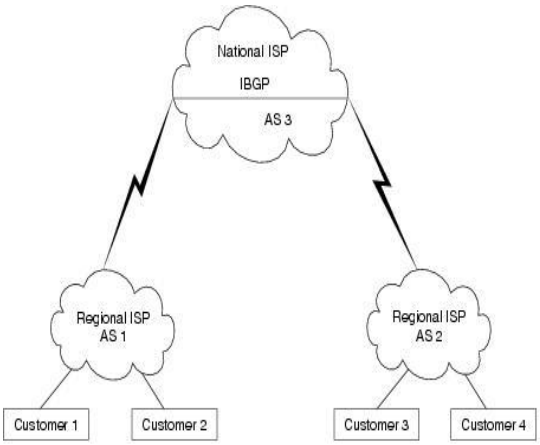


Fig1: External and Interior BGP

However, looking at just those ASes where 1% of their IP addresses have been marked as bad, we see that 24.4% become unreachable. *ASes with the most malicious activity appearto be disconnected more often than others*.

However, among the ASes which make up at least 0.25% of the malicious IP addresses in their data sets, only 8 (3.0%) ever become unreachable. Many of the ASes which become unreachable do not stay that way for long. We now look at if how long they are unreachable is dependent on the degree of maliciousness of the AS. Figure 3 shows the duration of time ASes in each category become unreachable, except for those making up at least 0.25% of malicious IP addresses in a data set, which we exclude from this figure due to the low number of data points. Some become unreachable multiple times for short durations; however, the time plotted in this figure represents the aggregate for each AS. Timestamps on the BGP updates are at a resolution of one second, so when an AS becomes unreachable for less than one second, we count it as becoming unreachable but do not add time for this period.

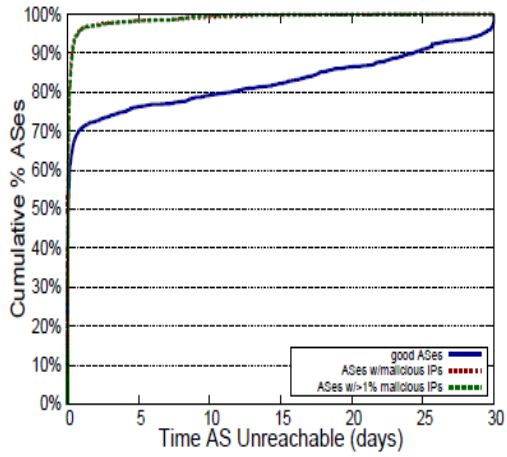


Fig 2: Unreachability duration for good and bad ASes which becomeunreachable during our data period.

We see a significant difference among our categories here. 96% of malicious ASes are disconnected for less than a single day, with a similar number for ASes with 1% bad IP addresses. Correspondingly, only 71% for the ASes not identified as malicious become disconnected for less than a day. On the high end, while 45.7% of ASes which become unreachable have malicious behaviors, just 1% of those unreachable for more than 2 weeks have malicious behaviors. *When maliciousASes become unreachable, they do not tend to stay that way for long*. If these disconnections are intentional de-peerings, the approach is not effective at isolating the AS for long. The results for the length of time an AS becomes unreachable were opposite of what we initially expected. To examine routing behavior in further detail, we now consider all connectivity changes to ASes which originate a route (gaining or losing a peer), not just those which change its overall reachability. Of all ASes originating a prefix, 17,286 (53.7%) have some change during our data period. For malicious ASes, this is 8,695 (58.7%), and for those with at least 1% malicious IP addresses, this is 2,036 (66.1%). For those making up at least .25% of one of our data sets, this is 166 (60.9%). Malicious behavior in an AS is clearly associated with routing instability; however, this could be the result of other factors and not simply the malicious activity.

## III. RELATED WORK

This work focuses on finding the ASes harboring malicious activity on the Internet, and investigating the behaviors of and connections between those ASes. Accordingly, the related work falls into two broad categories, work which examines AS topology, and work which attempts to characterize the location of malicious behaviors.

### *A. AS Topology*

Numerous studies have focused on accurately determining types of AS relationships, including those by Di Battista *etal.*, Dimitropoulos *et al.*, Gao, and Subramanian *et al*. Where we deal with connections between ASes, we are most concerned just with if a malicious AS is related to other malicious ones. Therefore to infer the type of relationship, we use a simple algorithm similar to the one Gao describes as her basic algorithm. Rexford *et al.* examined BGP routing stability for the ASes of popular destinations on the Internet. They found that most BGP instability was from unpopular destinations and that popular destinations had more stable routes. Work by Feldmann *et al* identifies ASes which cause routing changes. In our work, we find that ASes containing malicious IP addresses have disproportionately high routing changes

### B. *DATA COLLECTION*

To create a comprehensive evaluation of an AS, we [1] use a diverse set of data sources. Each of our data sources lists machines reported as engaging in some form of malicious activity. *1) Phishing Sites:* Phishing sites attempt to collect sensitive data, such as login credentials, credit card numbers, account numbers, and social security numbers, from users by impersonating legitimate organizations or brands. The Anti- Phishing Working Group and Phish Tank have among the largest data feeds listing such phishing sites. *2) Spam Senders:* A mail server can use IP blacklisting to prevent compromised machines from sending mail directly. Spamhaus runs the most widely used blacklist in this context.

*3) Exploited Hosts:* Spamhaus also maintains a second blacklist, known as the XBL. This list contains prefixes (often individual IP addresses) of hosts infected with exploits often used to send spam. This includes open proxies, computers infected with viruses that are known to send spam, and other exploits. This data is updated every half hour and is labeled Spamhaus XBL *4) Malware Downloads:* Malicious software, or *malware*, including viruses, worms, and trojans, have harmful effects on the computers they infect. Three of our data sets list Web sites that host malware downloads. The Clean-MX Viruswatch mailing list, eSoft, and Malware Patrol all independently collect URLs that host malware. *6) Bot Command and Control:* Botnets consist of groups of compromised

machines used for malicious purposes on the Internet.. Bots must get their instructions from their bot masters, often through command and control servers. The ShadowServer Foundation provides lists of botnet command and control servers along with their IP addresses. Fig

| Number of Blacklists with Given IP Address | Number of IP Addresses |
|---|---|
| 1 | 29,631,573 |
| 2 | 9,566 |
| 3 | 3,650 |
| 4 | 1,290 |
| 5 | 320 |
| 6 | 112 |
| 7 | 29 |
| 8 | 7 |
| 9 | 8 |

Table 1: Degree to which an IP address appears in multiple blacklists

### *C. Applications of Our Research*

Comparing ASes and their degree of maliciousness can be used in several applications, including public policy, peering preference, and destination prioritization. Governments have increasingly recognized that critical national assets are exposed to the Internet and that cyber-attackscan have profound implications on national operation. Due to the distribution of compromised machines, these nations must address attacks coming from within their borders. Accordingly, regulators may seek to curtail computer attacks; however, mechanisms to evaluate ASes, regulators would be unable to establish baselines for compliance and what constitutes responsible network management. Our approach can provide these metrics. Alternatively, ISPs may choose to self-regulate to ward off government intervention. To influence others to adopt better security practices, peers may place requirements for controlling the spread of malicious machines in their peering agreements in exchange for lower peering costs. Larger ISPs could pressure their customers to practice better security. Finally, destination networks can leverage information on AS maliciousness to determine how to prioritize traffic. In the case of bandwidth contention, a destination may prefer traffic from an AS with low malicious activity over a highly malicious AS since doing so would be more likely to service a legitimate user. Maliciousness scores could also be used in spam filtering; however, this cannot be a sole discriminator since there may be legitimate machines in many highly malicious networks.

## IV. CONCLUSION

In this study, we examined whether some networks are safe harbors for malicious activity.We found that several ASes have high concentrations of malicious IP addresses while others represent disproportionately higher malicious activity than their equivalently sized peers. The above limitations can be overcome by collecting the attack history from the destination or by network routing infrastructure. Network and host-based intrusion detection services may collect and aggregate data on attacks and provide them to the security service vendors to analyze. Instead of using Heuristic algorithm for BGP path selection we can make use of BGP Decision algorithm and decision will be done using attributes like AS_PATH, origin, next hop attribute and many more. This information canalso be used in peering agreements to place pressure on ISPsto respond to malicious activity.

## REFERENCES

[1] J. Hruska, “Bad seed ISP Atrivo cut off from rest of the Internet,” 2008, http://arstechnica.com/security/news/2008/09/bad-seed-isp-atrivo-cut-off-from-rest-of-the internet.ars.
[2] B. Krebs, “Major source of online scams and spams knocked offline,” 2008, http://voices.washingtonpost.com/securityfix/2008/11/ major source of online scams a.html.11
[3] J. Cheng, “FTC forces hive of scum and villainy ISP offline,” 2009, http://arstechnica.com/tech-policy/news/2009/06/ ftc-forces-hive-of-scum-and-villainy-isp offline.ars.
[4] U. of Oregon Advanced Network Technology Center, “Route Views project,” http://www.routeviews.org/.
[5] Spamhaus Project, “Spamhaus block list (SBL),” http://www.spamhaus. org/sbl/index.lasso.
[6] NETpilot GmbH, “Viruswatch mailing list,” http://lists.clean-mx.com/cgi-bin/mailman/listinfo/viruswatch.
[7] eSoft Inc., http://www.esoft.com/.
[8] Malware Patrol, “Malware block list,” http://www.malwarepatrol.net/lists.shtml.
[9] ShadowServer Foundation, http://www.shadowserver.org/wiki/.
[10] L. Gao, “On inferring autonomous system relationships in the Internet,” *IEEE/ACM Transactions Of Networking*, vol. 9, no. 6, pp. 733–745, Dec. 2001.
[11] G. D. Battista, M. Patrignani, and M. Pizzonia, “Computing the types of the relationships between autonomous systems,” in *IEEE Conferenceon Computer Communications (INFOCOM)*, 2003.
[12] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, kc claffy, and G. Riley, “AS relationships: Inference and validation,” *ACM SIGCOMM Computer Communications Review (CCR)*, vol. 37, no. 1, pp. 29–40, Jan. 2007.
[13] A. Feldmann, O. Maennel, Z. M. Mao, A. Berger, and B. Maggs, “Locating Internet routing instabilities,” in *ACM SIGCOMM*, 2004.
[14] Z. Zhang, Y. Zhang, Y. C. Hu, and Z. M. Mao, “iSPY: detecting IP prefix hijacking on my own,” in *ACM SIGCOMM*, 2008.
[15] “Practical defenses against BGP prefix hijacking,” in *Conference on Future Networking Technologies (CoNext)*, 2007.