A Heuristic Approach for the Detection of Malicious Activity at Autonomous System

Sowmyashree C.S [*], Prof. Chandrasekhar S[**]

[*] VI semester, M.Tech(CSE), RVCE, Bangalore
[**] Professor, Dept. of CSE, RVCE, Bangalore

Abstract- On the internet, an autonomous system is the unit of router policy which is also sometimes referred to as a routing domain. An autonomous system is assigned a globally unique number called Autonomous System Number (ASN). The investigators and network operators have recently identified that the high profile autonomous systems exhibit disproportionately high malicious behavior since they are attacked by malicious network. In this paper we explore whether some AS's are safe zone for communication through internet. We look for ISPs and ASs that exhibit disproportionately high malicious behavior using 10 popular blacklists, plus local spam data, and extensive DNS resolutions based on the contents of the blacklists. We find that some ASs have over 80% of their routable IP address space blacklisted. Overall, we conclude that examining malicious activity at AS granularity can unearth networks with lax security or those that harbor cybercrime.

Index Terms- Border Gateway Protocol (BGP), security, Internet Service Provider (ISP), Botnet.

I. INTRODUCTION

As we know that Internet Service Provider provides internet service to the customer but the service level agreement made by them and service accomplishment will not be same. This is because of communication media and error in the network. Along with these problems if malicious activity is also takes place in ISP then service will be lesser than that. 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. Furthermore, some networks may exist solely to engage in malicious activity.

The ISP such as the Atrivo has involved in malicious activity so it has been banned by United States government.

In this paper, we [1] examine whether we can find malicious networks in a systematic manner using existing blacklists. The necessary of detecting malicious network is to make the ISP to inform to their customer to limit the amount of malicious activity in their networks to avoid harboring criminals. ISPs could also use the metrics to determine the effectiveness of their efforts to combat abuse and compare themselves to other networks. Also, when receiving traffic, a destination network could prioritize traffic based on the cleanliness of ASs, 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 e-mail messages.

Below figures show how the phishing and malware activity is done by redirecting the user details to attacker sites.

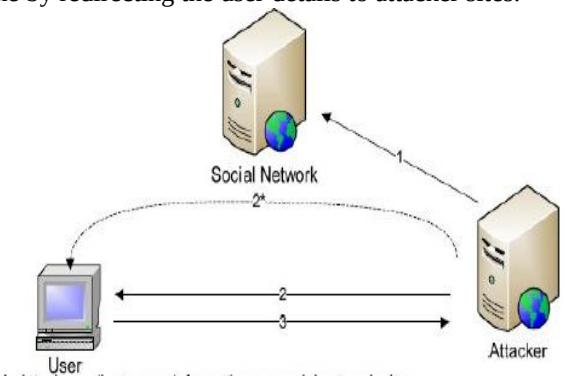


Figure 1.a Attacker leverage the available information on social network

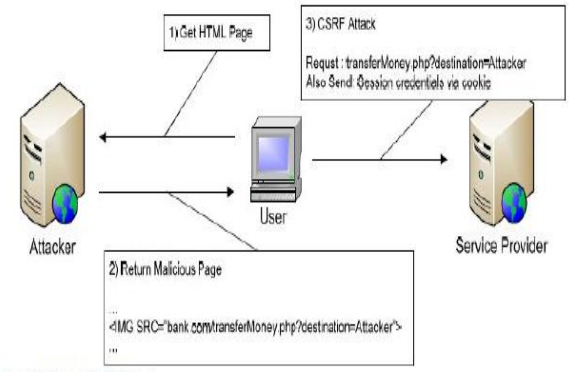


Figure 1.b Typical CSRF scheme

To determine which ASs are malicious, we [1] use 10 of the most commonly used blacklists for spam, phishing, malware, and botnet activities for a period of a month. These blacklists either contain host names or IP addresses to be blacklisted. For host-name-based blacklists, we first determine the IP addresses for each blocked host using real-time DNS queries.

This gives us IP addresses of all blacklisted hosts in our blacklists. We [1] then use BGP routing tables to group these IP addresses into their originating ASs. Upon grouping these addresses by AS, we compare ASs by the percent of infected machines and the rate at which they are cleaned up. We [1] examine other characteristics of the malicious ASs, such as whether their connectivity to other Ass changes more often than

those without malicious activity. The key findings of our study are the following.

• A large fraction of routable space is malicious for some ASs: Four ISPs—two 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 providers regularly peer with malicious ASs: We find 22 provider ISPs with 100%of their customer ASs engaged in significant malicious activity.

• Malicious ASs differ from benign ones in other ways. They are more likely to become completely unreachable than those that have less malicious activity, and they are likely to have more peers. However, the duration of unreachability is short for these ASs, which may have implications for orchestrated depeering 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 cleanup efforts.

II. RELATED WORK

In this section we mainly concentrate on the data collection and the BGP routing table.

A .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.

TABLE 2.1
Degree to which an IP address appears in multiple blacklists

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

In Table 2.1, we show the number of data sets containing each IP address. The Spamhaus XBL is roughly three orders of magnitude larger than any other data set, so the vast majority of IP addresses appear only in that single data set. It is further unsurprising that some IP addresses appear in two or three data sets. The below Table 2.2 gives the overview of datasets.

Here we have considered top ten websites like APWG, esoft, Spamhaus and so an are going to monitor the ISP's for a month and detect the malicious AS IP addresses.

Table 2.2 Overview of Datasets

Label	Description	Duration (in days)	Unique IP Addresses	Unique ASs	Median IPs Per AS	Std. Dev. IPs per AS
APWG	Phishing URLs from the Anti-Phishing Working Group	30	9,560	1,803	2	18.0
Bot C&C	Botnet command and control IPs from the ShadowServer Foundation	30	1,986	611	1	11.4
CleanMX	Malware serving sites from the CleanMX VirusWatch mailing list	30	2,974	687	1	12.0
eSoft	Malware serving sites from eSoft, Inc.	30	8,000	1,196	2	27.2
Local Spam	URLs from spam messages received by the IU CS Department	30	5,495	1,024	1	16.5
Malware Patrol	MalwarePatrol's block list for malware-serving sites	30	871	368	1	5.3
PhishTank	Phishing URLs from PhishTank	28	7,143	1,580	1	14.2
Spamhaus SBL	Verified spam sources from Spamhaus.org Block List	29	6,422	2,005	1	8.9
Spamhaus XBL	Hijacked machines from Spamhaus.org Exploit Block List	29	29,585,604	13,580	9	31,568.1
SI-Feed	URLs and IP addresses from spam emails from Support Intelligence	30	7,591	1,420	1	20.2
SI-DNS	IP addresses from DNS resolutions on the `SI-Feed` data set	30	4,448	911	1	11.8
SURBL	Host names appearing in spam messages from SURBL	30	29,324	2,739	2	47.2

The below figure 2.a shows the percentage of malicious hosts in AS.

Mmml

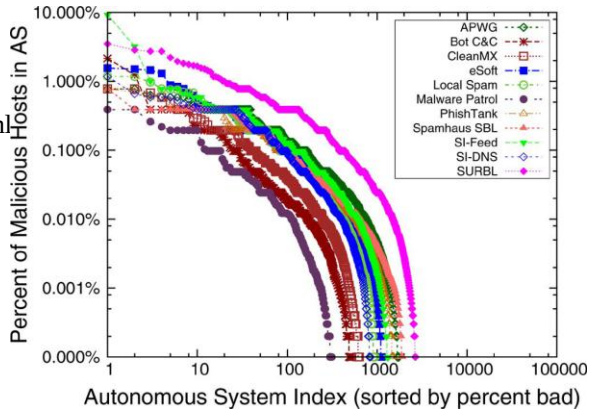


Figure 2.a Percentage of badness for each AS

B.INTERNET ARCHITECTURE

The INTERNET has experienced a tremendous growth in its size and complexity since its commercialization. The Internet connects thousands of autonomous systems (ASs) operated by many different administrative domains such as Internet service providers (ISPs), companies and universities. Since two ISPs might merge into one and each administrative domain can possess several ASs, an administrative domain can operate one or several ASs.

Routing within an AS is controlled by intradomain routing protocols such as static routing, OSPF, IS-IS, and RIP. A pair of ASs interconnects via dedicated links and/or public network access points, and routing between ASs is determined by the interdomain routing protocol such as Border Gateway

Protocol (BGP). One key distinct feature of the interdomain routing protocol is that it allows each AS to choose its own administrative policy in selecting the best route, and announcing and accepting routes. One of the most important factors in determining routing policies is the commercial contractual relationships between administrative domains.

We [2] propose an augmented AS graph representation to capture AS relationships. We classify the relationship between a pair of interconnected ASs into customer–provider, peering, and sibling relationships. There is no publicly available information about inter-AS relationships. ISPs do not register their relationships to the Internet registries. Internet Routing Registries (IRR) was created as a repository of routing policies. However, some ISPs are not willing to reveal their policies, and even if they were, these routing policies might not specify AS relationships.

We [2] present heuristic algorithms that infer the augmented AS graph from BGP routing tables. BGP routing tables are retrieved from the Route Views server in Oregon, which is publicly available and has the most complete view currently available. The RouteViews server establishes BGP peering sessions with many tier-1 and tier-2 ISPs. Among the connected AS pairs, the algorithms infer that more than 90.5% of the AS pairs have customer– provider relationships, less than 1.5% of the AS pairs have sibling relationships, and less than 8% of the AS pairs have peering relationships.

We can model the connectivity between ASs in the Internet using an AS graph G=(V, E) where the node set V consists of ASs and the edge set E consists of AS pairs that exchange traffic with each other Fig.2.b shows an example of an AS graph. The *degree* of an AS is the number of ASs that are its neighbors. Formally, the degree of AS u, D(u)={v|(u,v) $\in$ E}

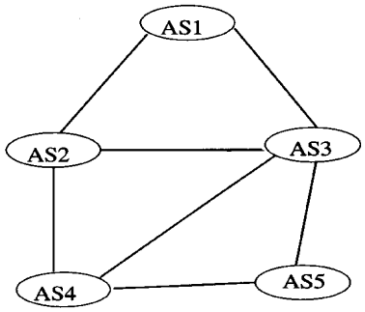


Figure 2.b AS graph example

III. PROPOSED WORK

In this section we [2] describe a heuristic algorithm for inferring AS relationships and method to overcome BGP vulnerability.

A. HEURISTIC ALGORITHMS FOR INFERRING AS RELATIONSHIPS

The below algorithm will be treated as a final heuristic algorithm because identify peering relationships from the rest of connected AS pairs by using the heuristic that two peering ASs' degrees do not differ by more than R times where R is some constant that has to fine tuned.

Input: BGP routing tables
Output: Annotated AS graph G

Phase 1: Use either Basic or Refined algorithm to coarsely classify AS pairs into provider-customer or sibling relationships

Phase 2: Identify AS pairs that can not have a peering relationship

1. For each AS path $(u_1, u_2, \ldots, u_n)$,
2. find the AS u_j such that degree$[u_j]$=max$_{1\le i\le n}$ degree$[u_i]$
3. for $i = 1, \ldots, j-2$,
4. notpeering$[u_i, u_{i+1}]$=1
5. for $i = j+1, \ldots, n-1$,
6. notpeering$[u_i, u_{i+1}]$=1
7. if edge$[u_{j-1}, u_j] \neq$ sibling-to-sibling and edge$[u_j, u_{j+1}] \neq$ sibling-to-sibling
8. if degree$[u_{j-1}]$>degree$[u_{j+1}]$
9. notpeering$[u_j, u_{j+1}]$= 1
10. else
11. notpeering$[u_{j-1}, u_j]$= 1

Phase 3: Assign peering relationships to AS pairs

1. For each AS path $(u_1, u_2, \ldots, u_n)$,
2. for j=1, ..., n-1,
3. if notpeering$[u_j, u_{j+1}] \neq$ 1 and notpeering$[u_{j+1}, u_j] \neq$ 1 and degree$[u_j]$/degree$[u_{j+1}]$< R and degree$[u_j]$/degree$[u_{j+1}]$> $1/R$
4. edge$[u_j, u_{j+1}]$ = peer-to-peer

B. SECURE BGP ROUTE DISTRIBUTION

The approach we[3] adopted to securing BGP route distribution involves two Public Key Infrastructures3 (PKI's), a new path attribute containing "attestations," and the use of IPsec. These components are used by a BGP speaker to validate the authenticity and data integrity of BGP UPDATE's that it receives, and to verify the identity and authorization of the senders

Public Key Infrastructures (PKI's) and Certificates

S-BGP uses two PKI's, based on X.509 (v3) certificates, to enable BGP speakers to validate the identities and authorization of BGP speakers and of owners of ASes and of portions of the IP address space. These PKI's parallel the existing IP address and AS number assignment delegation system and take advantage of this extant infrastructure.

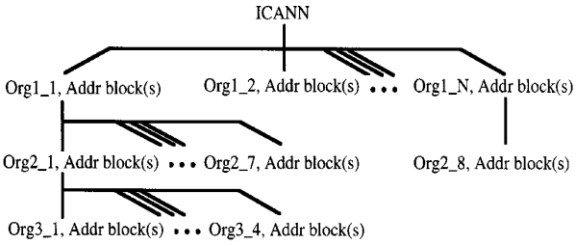
The two PKI's involve four types of certificates, as illustrated below (in the diagrams).

• The higher node is the issuer for the certificates defined in the tier below it.

• The name of the current tree node (organization, AS, router, etc.) is the subject of the certificate.

• Any additional fields shown in the node, e.g., address block(s), are in an extension in the certificate.

• Other X.509 certificate fields are assumed, but not Shown—sequence number, subject public key, signature, validity period, etc.



Subject	Description
Org1_x	a 1st tier organization (usually a registry)
Org2_x	a 2nd tier organization (usually an ISP or DSP)
Org3_x	a 3rd tier organization (usually a DSP or user organization)

Extension	Description
Addr blocks(s)	one or more IP address blocks assigned to the organization

Figure 3.a Address allocation for PKI structure

Route Validation

Attestations and certificates are used by BGP speakers to validate routes asserted in UPDATE messages, i.e., to verify that the first AS in the route has been authorized to advertise the address block(s) by the address block owner(s), and that each subsequent AS has been authorized to advertise the route for theaddress block(s) by the preceding AS in the route.

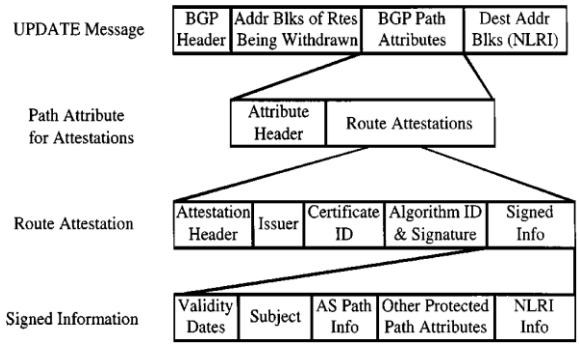


Figure 3.b UPDATE format with route attestation

IV. CONCLUSION

This paper has examined whether some networks are safe harbors for malicious activity. This paper found that several ASs have high concentrations of malicious IP addresses, while others represent disproportionately higher malicious activity than their equivalently sized peers. Our analysis can be used to help increase ISP accountability and can become a mechanism to combat malicious activity. This paper provides efficient methods for AS path selection and also provides methods for overcoming from BGP vulnerability. The limitations of the proposed system are changing the black listed IP address by administrator, IP address of the black listed AS is not properly converted to host name and inaccuracy of the heuristic algorithm.

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.

ACKNOWLEDGMENT

I would like to express sincere thank to my guide Prof. Chandrasekhar S for his guidance and support to publish this paper.

I would also like to thank all my friends and family members who helped me indirectly to present the paper.

REFERENCES

[1] Craig A Shue, Andrew J.Kalafut, and Minaxi Gupta , " Abnormally malicious autonomous systems and their internet connectivity", 2012.

[2] L.Gao, "On inferring autonomous system relationships in the internet," IEEE/ACM Trans. Netw., vol. 9, no. 6, pp. 733–745, Dec. 2001.

[3] S. Kent, C. Lynn, and K. Seo, "Secure border gateway protocol (S-BGP)," IEEE J. Sel. Areas Commun., vol. 18, no. 4, pp. 582–592, Apr. 2000

[4] R. White, "Securing BGP through secure origin BGP (soBGP)," Internet Protocol J., vol. 6, no. 3, pp. 15–22, 2003.

[5] J. Rexford, J. Wang, Z. Xiao, and Y. Zhang, "BGP routing stability of popular destinations," in Proc. ACM SIGCOMM IMW, 2002, pp. 197–202.

[6] J. Hruska, "Bad seed ISP Atrivo cut off from rest of the Internet,"2008.

[7] B. Krebs, "Major source of online scams and spams knocked offline," 2008 .

[8] "Route Views project," University of Oregon Advanced Network Technology Center, Eugene, OR [Online]. Available:http://www.routeviews.org/

[9] "APWG," Anti-Phishing Working Group [Online]. Available: http://www.antiphishing.org/

[10] "PhishTank," OpenDNS, San Francisco, CA [Online]. Available:http://www.phishtank.com/

[11] "SURBL," [Online]. Available: http://www.surbl.org/

[12] "Spamhaus block list (SBL)," Spamhaus Project [Online]. Available:http://www.spamhaus.org/sbl/index.lasso

[13] "Exploits block list (XBL)," Spamhaus Project [Online]. Available:http://www.spamhaus.org/xbl/index.lasso

[14] "eSoft Inc.," eSoft Inc., Broomfield, CO [Online]. Available: http://www.esoft.com/

[15] "Malware block list,"Malware Patrol [Online]. Available: http://www.malwarepatrol.net/lists.shtml

AUTHORS

First Author – Sowmyashree C.S, B.E, [M.TECH], RV college of Engineering, Bangalore, Sowmyacs9@gmail.com.
Second Author – Prof. Chandrasekhar S, Professor, Dept. of CSE, RVCE, Bangalore, Email: chandrasekhar0708@rediffmail.com