UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

SECURITY ANALYSIS OF NETWORK NEIGHBORS

Sérgio Miguel Geraldes de Oliveira Serrano

MESTRADO EM SEGURANÇA INFORMÁTICA

Novembro 2010

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

SECURITY ANALYSIS OF NETWORK NEIGHBORS

Sérgio Miguel Geraldes de Oliveira Serrano

Orientador
Nicolas Christin

Co-Orientador
Marcelo Pasin

MESTRADO EM SEGURANÇA INFORMÁTICA

Novembro 2010

Resumo

O presente trabalho aborda um problema comum a muitos dos actuais fornecedores de serviços Internet (ISPs): mitigação eficiente de tráfego malicioso na sua rede. Este tráfego indesejado impõe um desperdício de recursos de rede o que leva a uma consequente degradação da qualidade de serviço. Cria também um ambiente inseguro para os clientes, minando o potencial oferecido pela Internet e abrindo caminho para actividades criminosas graves. Algumas das principais condicionantes na criação de sistemas capazes de resolver estes problemas são: a enorme quantidade de tráfego a ser analisado, o facto da Internet ser inerentemente anónima e a falta de incentivo para os operadores de redes de trânsito em bloquear este tipo de tráfego.

No âmbito de um ISP de média escala, este trabalho concentra-se em três áreas principais: origens de tráfego malicioso, classificação de segurança de redes vizinhas ao ISP e políticas de intervenção.

Foram colectados dados de rede considerando, determinados tipos de tráfego malicioso: varrimento de endereços e inundação de fluxos de ligações; assim como informação de acessibilidades rede: mensagens de actualização de BGP disponibilizadas pelo *RIPE Routing Information Service*. Analisámos o tráfego malicioso em busca de padrões de rede, o que nos permitiu compreender que é maioritariamente originário de um subconjunto muito pequeno de ASes na Internet. No âmbito de um ISP e de acordo com um conjunto de métricas de segurança, definimos uma expressão de correlação para quantificar os riscos de segurança associados a conexões com redes vizinhas, a qual denominámos *Risk Score*. Finalmente, propusemos técnicas para concretização das tarefas de rede necessárias à redução de tráfego malicioso de forma eficiente, se possível em cooperação com redes vizinhas / ASes.

Não temos conhecimento de qualquer publicação existente que correlacione as características de tráfego malicioso de varrimento de endereços e inundação de fluxos de ligações, com informação de acessibilidades de rede no âmbito de um ISP, de forma a classificar a segurança das vizinhanças de rede, com o propósito de decidir filtrar o tráfego de prefixos específicos de um AS ou bloquear todo o tráfego proveniente de um AS.

Acreditamos que os resultados apresentados neste trabalho podem ser aplicados imediatamente em cenários reais, permitindo criar ambientes de rede mais seguros e escaláveis, desta forma melhorando as condições de rede necessárias ao desenvolvimento de novos serviços.

Palavras-chave: tráfego malicioso, vizinhança de rede, BGP, regras de especificação de fluxos, segurança de rede

Abstract

This thesis addresses a common issue to many of current Internet Service Providers (ISPs): efficient mitigation of malicious traffic flowing through their network. This unwanted traffic imposes a waste of network resources, leading to a degradation of quality of service. It also creates an unsafe environment for users, therefore mining the Internet potential and opening way for severe criminal activity. Some of the main constraints of creating systems that may tackle these problems are the enormous amount of traffic to be analyzed, the fact that the Internet is inherently untraceable and the lack of incentive for transit networks to block this type of traffic.

Under the scope of a mid scale ISP, this thesis focuses on three main areas: the origins of malicious traffic, security classification of ISP neighbors and intervention policies.

We collected network data from particular types of malicious traffic: address scans and flow floods; and network reachability information: BGP update messages from RIPE Routing Information Service (RIS). We analyzed the malicious traffic looking for network patterns, which allowed us to understand that most of it originates from a very small subset of Internet ASes. We defined a correlation expression to quantify the security risks of neighbor connections within an ISP scope according to a set of security metrics that we named *Risk Score*. We finally proposed techniques to implement the network tasks required to mitigate malicious traffic efficiently, if possible in cooperation with other neighbors/ASes.

We are not aware of any work been done that correlates the malicious traffic characteristics of address scans and flow flood attacks, with network reachability information of an ISP network, to classify the security of neighbor connections in order to decide to filter traffic from specific prefixes of an AS, or to block all traffic from an AS.

It is our belief, the findings presented in this thesis can be immediately applied to real world scenarios, enabling more secure and scalable network environments, therefore opening way for better deployment environments of new services.

Keywords: malicious traffic, network neighbors, peering, BGP, flow specification rules, network security

Acknowledgments

This thesis was only possible with the contributions of many different persons and entities. Their time, opinions, questions and doubts had a crucial impact on this work and helped making it possible.

A special thank you for my advisor Nicolas Christin whose opinions were fundamental for driving this work through with a pragmatic view, it was a true pleasure to work with an advisor that tried to be always present and supportive. To my co-advisor Marcelo Pasin, who gave me very interesting and simple ideas for solving complex problems.

I would like to thank André Cardoso, Pedro Gonçalves and Pedro Mitra for their constant insightful observations and for their effort to provide me all the necessary technical conditions to develop the work presented in this thesis.

Finally I would like to mention the spirit of companionship and discussion from all my MSIT-IS colleagues with whom I shared this intense experience that I will cherish for the rest of my life.

Lisbon, November 2010

Dedicated to Isabel, Helena and Artur for their love, support and unbelievable patience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we describe the main network security challenges faced by Internet Service Providers (ISPs) and explain the contributions provided by this thesis to help solving them.

1.1 Challenges of Internet Service Providers

When dealing with ISP networks several additional issues arise, not only due to the number of hosts, amount of traffic, technology diversity, but mainly due to the lack of control over the sources of malicious traffic. Although this problem is similar to the ones dealt in typical local area networks, one big difference exists, the Internet is managed by a myriad of entities, each with their own agenda and not all overlap. Finally, the fact that ISPs are typically only considered traffic carriers raises a whole new set of challenges, since although the ISP has the responsibility of protecting its network, some of the possible defense mechanisms may be considered too intrusive, with privacy concerns for the end customers (e.g., blocking certain types of traffic, virus detection). A thin line therefore exists between traffic engineering and network operator abuses.

One of the main obstacles to have efficient solutions for dealing with malicious traffic is the lack of incentives for ISPs to mitigate it. First, ISPs of entities that generate malicious traffic do not have any incentive to mitigate this traffic. The malicious users are often paying customers and the incentive is on the opposite direction, i.e., for the ISP to provide the best service, which in those cases means providing Internet connectivity. Second, in case ISPs choose to assume a responsible role on this issue, they need to increase their operational costs for deployment of mitigation mechanisms, i.e., new equipment and new specialized personnel. Also, such mechanisms are not perfect, they add complexity to network architectures and anecdotal evidences of negative impact on well-behaved customers are common.

As mentioned one big obstacle to the mitigation of malicious traffic within an ISP network is its lack of control above the rest of the Internet. Although we can gain a lot from the inherent openness of the Internet, we must be conscious of the problems that derive from that fact. Considering the Internet as an open system means that malicious traffic mitigation in the Internet must be a joint

effort of all parties to be effective. All networks must be called to this effort, or at least the most relevant in current Internet infrastructure, e.g., Tier-1 ASes. This brings many other problems to the table, e.g., business relationships, Internet neutrality, secrecy of traffic policies.

Since we cannot control actions from all providers, one possibility may be to create proper incentives for malicious traffic mitigation best practices.

1.2 Contribution

In this thesis we create a mechanism that enables security classification of network neighbors, allowing the design of security systems that deal with malicious network neighbors and in particular peering relationships.

This thesis provides an analysis of particular types of malicious traffic in a medium size ISP network, providing useful input for the task of designing new tools that deal with malicious traffic within an Internet network. Due to the extremely wide range of issues that building such tools present, this information is extremely useful for focusing on particular ones, to understand the scalability of a specific approach, and to understand if it may even be a viable solution on current Internet.

Simplification in integrating with existing platforms allows capitalization of previous investments, an important factor for new technologies since one of the main problems that currently exist is the increasing operational cost of maintaining a network. Integration is therefore crucial for a sustainable solution.

Based on the observations obtained from the malicious data analysis, we defined Risk Score, a classification mechanism that allows correlation of different metrics to characterize malicious traffic. A specific set of metrics was selected, however the Risk Score was defined with enough flexibility to easily integrate new types of metrics. Having a Risk Score allows a network operator to better interpret its network neighbors, namely its peering relationships, and quantify their security characteristics. Several different possibilities to integrate it with network mechanisms that enable malicious traffic mitigation are also presented.

If we desire the Internet to be a fertile environment for innovation, e-governance and socio-economic changes, we need to deliver the best conditions for doing so. This means security is a fundamental requirement. For this to happen we need proper incentives, and for those we may contribute with technological tools that help us provide them. This work is an effort in that direction, enabling network operators to be better prepared to deal with an environment not completely "friendly". Creating new ways of interaction between different network operators. With the work being presented with this thesis we provide a flexible way to quantify how an ISP relates to other neighbor networks. A continuous perspective of the network neighbors behavior is fundamental and this is the rationale that drives this work.

1.3 Document organization

The rest of this thesis is structured in order to give a contextualization of network environment, current state of the art malicious traffic studies, to support the classification methodology of network neighbors and applicability of proper network mechanisms to address ISP network security issues. Chapter 2 provides the necessary background related to the Internet infrastructure, malicious traffic and proposed traffic control mechanisms. Chapter 3 dives into existing work done within this area of expertise. Chapter 4 presents the analysis done on malicious traffic, BGP reachability information and a description of the tools used in the process. Chapter 4 describes the concept of Risk Score, its metrics, usage and simulation results to understand its applicability. On chapter 6 we propose a set of intervention policies, which have as main input the previously defined Risk Score. Finally on Chapter 7 we present the thoughts derived from the work done and new venues that this thesis opens for the research community.

Chapter 2

Background

Following in this chapter we provide the necessary background to better understand the concepts studied in this thesis, namely the Internet infrastructure, behaviors of malicious traffic and existing network mitigation techniques.

2.1 The Internet

The Internet was born from a United States military and private research program with the purpose of creating highly robust distributed computer networks that could interconnect different technologies in a seamless way. From than to current Internet a lot changed at extremely high speed but not always in the most sustainable way. The particular environment under which the Internet was born is often seen as one of the main reasons for many of its security weaknesses. Nevertheless its flexibility and inherent capability to integrate different technologies is an extremely strong argument for considering the Internet as a success case albeit all its problems.

Current Internet is composed of different interconnected networks administered by distinct entities, some private, some military, some governmental, each one with its particular agenda. The Internet is therefore a global network of networks that allows communication between many different entities, with different intentions and some of which are malicious.

The Internet has two important entities for its infrastructure: ICANN and IETF. The former is responsible for the management of Internet Protocol address space, top-level domain names, Internet protocol identifiers; the latter is responsible for the Internet technology layer decisions. Regarding ICANN it is also responsible for managing another important organization, IANA, which deals with the technical aspects of ICANN operations, being one the most relevant for this thesis, the management of the IP address space.

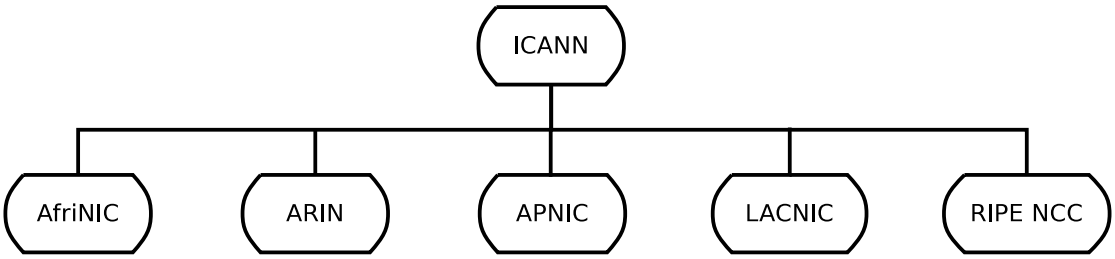

Figure 2.1: ICANN structure.

Depending on the region of the world, IANA delegates the allocation/registration of IP address space and autonomous system numbers (ASNs) to one of five RIRs (Regional Internet Registries):

- AfriNIC - African Network Information Centre for Africa;
- ARIN - American Registry for Internet Numbers for North America;
- APNIC - Asia-Pacific Network Information Centre for Asia/Pacific;
- LACNIC - Latin America and Caribbean Network Information Centre for Latin America and some Caribbean Islands;
- RIPE NCC - Réseaux IP Européens Network Coordination Centre for Europe, the Middle East, and Central Asia.

Considering the scope of this thesis, following are some of the most important aspects of the Internet structure:

Networks and Prefixes In the IP address space the term network usually refers to an IP classful network, i.e., class A, B or C network as specified in RFC 791 [37]. In this context, prefixes are classless hierarchical blocks of IP addresses as defined in RFC 4632 [27], i.e., CIDR blocks that may include subnets of networks or groups of one or more networks.

Autonomous System (AS) According to RFC 1930 [28] an AS is "*(...) a connected group of one or more IP prefixes run by one or more network operators which has a SINGLE and CLEARLY DEFINED routing policy*". An AS is the routing policy entity used in EGP protocols, for example BGP [41], which is the Internet de facto protocol for inter-AS routing. Regarding the routing policy, an AS can have different interior gateway protocols (IGP) and related metrics to make routing decisions, however when interacting with other ASes it should have a consistent routing policy, namely the reachability information of prefixes should be coherent. Usually ASes are classified according to the relationships with their network neighbors. They are considered to be one of the following types:

- Stub: only a connection exists to an upstream AS. If the AS routing policy is equal to the one of the upstream AS, one AS would be enough, in fact since it is considered an unjustifiable waste of resources, this situation is considered to be a bad practice [28]. Albeit a stub, the AS may have peering connections with other ASes;

- Multi-homed: the AS has multiple connections to upstream ASes. This is a common practice if an entity wishes to have different upstream service providers, e.g., to have network redundancy;

- Transit: the AS provides connectivity to other networks. The most common case of this type is an Internet Service Provider (ISP).

These definitions are not standardized, for which reason we can find different classifications in research works [22], they are simply the most common.

Autonomous System Number (ASN) This is currently a 4 byte globally unique number [49] used to identify the AS and to exchange exterior gateway information with other ASes. For BGP messages the ASN can be found in Open Messages, used to negotiate BGP session terms between BGP peers, and in three attributes of a BGP update message: AS_PATH, AGGREGATOR, COMMUNITIES. Till recent times, ASNs only had 2 bytes, however due to resource exhaustion it became necessary to have an update to the standard in order to enable a bigger allocation space. This change is very recent and some concerns still exist in network operators regarding its deployment [33], however since January 1 2009, 4 bytes ASNs were allocated by default for new requests and since January 1 2010 only 4 bytes ASNs are allocated for new requests.

Border Gateway Protocol (BGP) BGP is an inter-domain routing protocol [41] that enables exchange of network reachability information between ASes and is responsible for enabling the interconnection of all ASes in the Internet. BGP standardizes mechanisms to announce and withdrawal prefix routes, enabling forwarding policies solely based on the destination addresses of IP packets. To allow finer grained routing decisions specific attributes may be used, namely:

- MULTI_EXIT_DISC (MED): an optional non-transitive attribute for discriminating among multiple entry or exit points towards a neighbor AS;

- LOCAL_PREF: a well-known attribute that has its value calculated per route and is only sent to internal peers. Due to the importance of this protocol many other mechanisms were added and formally defined in the form of RFCs.

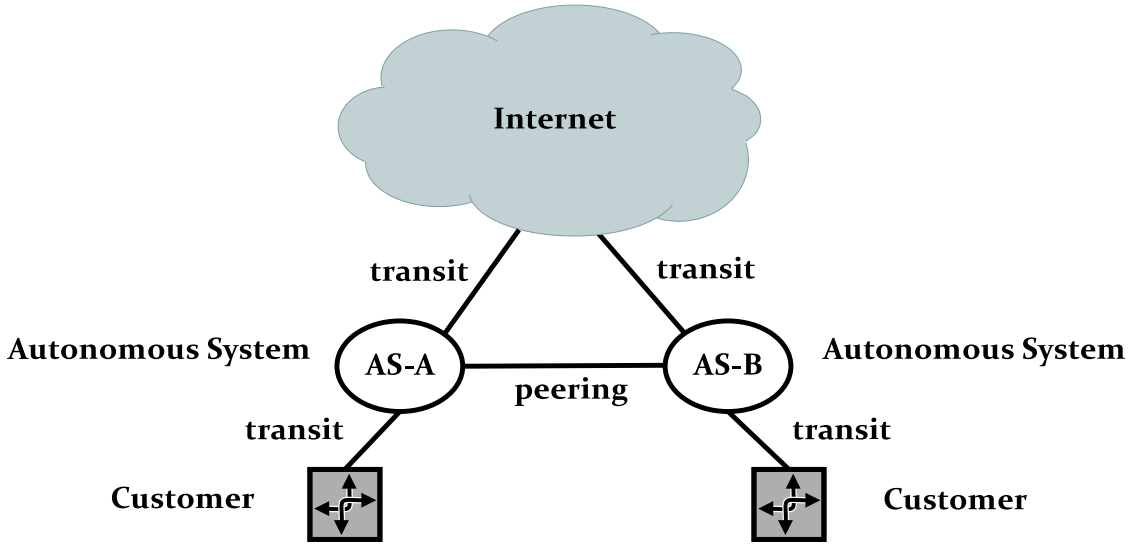

Figure 2.2: Internet private peering topology example.

Other relevant BGP attributes that will be addressed in this thesis are:

- AS_PATH: a well known attribute composed by all the autonomous systems through which a BGP update message has passed by. From this attribute we know the ASes that a packet will traverse when sent towards a certain destination;
- COMMUNITIES [17]: an optional non-transitive attribute to enable definition of a common goal for a set of prefixes, e.g., control how to distribute routing information among a BGP speaker neighbors.

Peering Based on an agreement between network operators (i.e., ASes) named peers, for the exchange of data traffic at an interconnection point. This strategy allows direct exchange of data without the need to go through the Internet cloud. Peering is done when found beneficial for the involved parties, therefore peering agreements often imply definition of contracts with specific SLAs agreed between parties, where peering relationship terms are defined, e.g., allowed network activities. Two main types of peering exist:

- Private: when done between only two networks that create a direct interconnection point. Often this type of peering is not publicly disclosed, which makes the understanding of Internet data flows more complex to understand [18]. Depending on the type of contract agreed it might imply a fee, or be simply free, it is a business decision dependant on the mutual benefits for both networks;
- Public: when done at an interconnection point where more than two network entities exist e.g. Amsterdam Internet Exchange (AMS-IX), Seattle Internet Exchange (SIX). In most cases this type of peering is economically more viable for networks to connect since many network entities are reachable from a single interconnection point.

Depeering When peering ceases to be considered beneficial for the parties involved the peering infrastructure is dismantled, which means traffic between both networks will start traversing the Internet cloud. This means reachability between both networks should still be possible, however without taking advantage of a direct connection, e.g., lower round trip times, higher dedicated bandwidth. The technical implication for a depeering action is the tier down of the BGP session between both ASes.

Security Current Internet has many security limitations, some with higher severity than others. At the top of the most serious are the ones related to BGP, which have at prefix hijacking the most frightening attack towards network operators and that apparently is still very common nowadays [33] despite all the research efforts that have been done to mitigate this problem [11, 39, 29].

From the background knowledge described in this sub-chapter it becomes easier to understand that the Internet is a set of unregulated interconnected networks that behave as open systems, i.e., the actions performed within a network at one side of the world, may affect networks on the opposite side of the world. The traffic flows through a myriad of ASes in the Internet, often without any security mechanisms in place.

2.2 Malicious Traffic

The Internet provides a fertile ground for malicious activities to occur. Analyzing the Symantec Global Internet Security Threat Report for 2009 [25] we can observe that malicious agents in the Internet have been highly active, and although threat trends change, e.g., decrease in data breaches, increase in botnets activity; it is not clear for what reason exactly, if due to intended actions or simply due to cyber criminals moving to more profitable areas. Overall still much has to be done to create a safer environment.

Malicious traffic can present itself through different attack vectors, different types of threats and different agendas from the part of the malicious agents. According to a senior responsible for security technology and response at Symantec: *"The scale of these attacks and the fact that they originate from across the world makes this a truly international problem requiring the cooperation of both the private sector and world governments."*. As more enterprises move to the Internet their communications infrastructure, malicious activity also increases since it becomes a more profitable business for cyber criminals.

According to the Arbor report [33] the most relevant attack vectors in 2009 were flood based attacks, e.g. UDP, ICMP, which indicates special attention must be given to address this particular type of attack vector.

Distributed Denial of Service (DDoS) attacks bandwidth growth slowed down, however it is still growing, for example in 2009 the largest DDoS attack had a peak of 49 Gigabits per second. Besides bandwidth denials of service, an increase in other types of resource exhaustion attacks were detected, e.g. CPU exhaustion, targeted at services facing the Internet, e.g., DNS. According to the Arbor report [33] the biggest threat was in fact DDoS to links, hosts or services.

Although all the evolutions seen for mitigating flood attacks, whether TCP or UDP, it is still an extremely common attack vector used in current Internet, approximately 45% of all registered attack vectors. In fact, this type of attacks is not only toward typical public services, e.g., HTTP, but also towards BGP peering addresses, which if successful, may have a devastating impact in networks connectivity.

Spam is a malicious activity responsible for much effort allocation by the part of corporations and one of its usual side effects is port/host scanning. For this reason, the latter is often used as a correlation factor when understanding this particular malicious traffic behavior.

In current Internet, many network operators already have some form of detecting and mitigating DDoS attacks [36]. Some of the most popular mitigation techniques for malicious traffic include: source/destination based access control lists (ACLs), source/destination based BGP remotely triggered blackholing (RTBH) [48], intelligent filtering, rate-limiting and BGP flow specification. Since some of these techniques may have negative impact in legitimate traffic and in the service provided to well-behaved paying customers, less harsh approaches are being used not to completely block traffic from entire networks but placing customers in wall-gardens or in quarantine. These more "sympathetic" solutions are of course done for the internal customers, relationships with external networks are usually addressed differently. This thesis focuses on the latter.

All these observations are obviously considering that malicious traffic is an unwanted type of traffic for a network operator. This is, unfortunately, not always true, placing this area of research in considerably fuzzier grounds. The following question is therefore very relevant:

> Why may a network operator want malicious traffic?

Some possible answers include:

- The malicious traffic originator is a paying customer. Blocking its traffic leads to an unsatisfied customer;
- The network operator has malicious intentions, e.g. organized crime, and is colluding with the originator of malicious traffic;
- Some network operators base their business models on accounted traffic, this means if a customer exchanges more traffic it also pays more.

2.3 Malicious Traffic Mitigation

Two main areas exist for the development of defense mechanisms in regard to malicious traffic, host-based and network-based. Host-based approaches (e.g., rootkit scanners, patching policies) are confined to hosts and although they have an extremely important role in the deployment of efficient security measures, network-based approaches (e.g., Intrusion Detection System) provide a wider defense barrier. The work done in this thesis focuses on network-based approaches, for which reason the defense mechanisms that will be described within this section belong to that class of defenses.

Typical network mechanisms used for mitigation of malicious traffic include:

Source/destination based access control lists (ACLs) A simple solution for malicious traffic mitigation is to have ACLs deployed in routers found in the path to the target prefix within the target network. These ACLs may be defined to drop/rate-limit traffic destined to the target network (destination based ACLs), or have traffic from a specific origin dropped/rate-limited (source based ACLs). This technique has the advantages of being simple and enabling good filter granularity, e.g., traffic type, source/destination prefix, TCP port, packet size. The main disadvantages are the inherent complexity of managing ACLs, the fact they must be configured router-by-router in their ingress or egress interfaces, and they do not distinguish malicious from legitimate traffic.

Destination based black-holing One technique that uses BGP to mitigate malicious traffic sent to a particular target is simple black-holing. When an attack is detected within an AS, an iBGP custom update message is created, where the targeted prefix (host or network) has the next hop selected from the private internet networks defined in RFC 1918 [42]. This advertisement is than exchanged within the iBGP domain. This technique considers that most routers in the Internet, in particular edge-routers, have routes configured to send networks from RFC 1918 to the null interface. Since the null interface route will generate a host unreachable ICMP message, if the network operator does not want for the attacker to realize the attack is being manipulated, workarounds must be but in place for the ICMP message to not reach the attacker machine(s). This approach has the advantage of avoiding the rest of the network from being affected, however for this to be achieved the network operator is shutting down the part of its network that the attack is targeting, affecting malicious as well as legitimate traffic. This is consequence of another disadvantage of this technique, the low granularity available to characterize the malicious traffic.

Source/destination based BGP remotely triggered black-holing *(*RTBH) RTBH is a more advanced technique that tries to overcome the limitations of simple black-holing by defining special BGP communities within the target AS. BGP is therefore used as a signaling protocol, which can be deployed from a unique management point. For each edge-router a special community is defined and another for particular groups of edge-routers. When an attack is detected, an iBGP advertisement is propagated in the victim AS for the destination prefix, with the communities associated with the routers through which the attack is passing through and with a no-export policy. A pre-configured policy exists in every router to filter advertisements with their particular community identification strings. That policy will then state the injection of routes with specific next hops, which can be for example the null interface. With the RTBH mechanism we can choose only the routers through which the malicious traffic is flowing and signal them to manipulate the malicious traffic. This has the advantage that all other routers will continue to provide the correct service for legitimate traffic. Details for this technique can be found in RFC 3882 [48].

Sinkhole tunnels [48] The RTBH technique implicitly states black-holing traffic, however, instead of simply dropping packets by specifying the next hop as the null interface, it can be configured to maintain the next hop but to forward the packets through a particular path/tunnel, which can have a packet sniffer for malicious traffic analysis, or be manipulated in more intelligent ways, e.g., using traffic engineering techniques for:

- QoS policies deployment;
- Rate-limiting;
- ACLs for dropping traffic.

One of the big security issues of this approach is the eBGP messages exchanged with the network peers, which may be crafted with specific communities used for this technique and that may lead the network operator to wrongly act on traffic otherwise considered legitimate.

BGP flow specification To tackle one of the limitations of RTBH techniques, i.e., low granularity, another technique named *BGP flow specification* has been proposed, and is now defined under the proposed standard RFC 5575 [32]. Due to its high communication efficiency, this technique still uses BGP as a signaling protocol, however it also enables a much finer granularity level for characterizing traffic, including source/destination ports, protocol type. The same traffic engineering techniques mentioned for sinkhole tunnels, can also be applied for the BGP flow specification.

Chapter 3

Related Work

The problems originated from malicious traffic do not seem to have an end in the near future; nevertheless a lot of effort has been put into this field of research. Several improvements have been made but still no perfect solution exists which leads to a continuous effort among research community to find better mechanisms to deal with this problem. Some believe that the only solution is to start from a clean slate [12], others believe it has been proved through the years that although changes and new mechanisms may be required, an evolutionary approach is the best solution [23]. We believe that the latter approach is not just more pragmatic but can indeed provide the better results for designing efficient countermeasures. The power provided by years of use of the Internet infrastructure is not only valuable for checking the quality of network protocols but also not duplicable in any test bed. For this reason, the work developed in this thesis follows the evolutionary approach.

Some of the biggest security shortcomings of the Internet infrastructure exist at its core, the nature of TCP/IP itself [13], allowing the user to access the data and control planes, namely the routing facilities. From the latter, BGP is one of the most influential protocols, however it is considered too naive to deal with current threats. Several alternative solutions have been tried out [16], however none of which have found its way to total acceptance, several reasons are cited: too cumbersome mechanisms (e.g., S-BGP), lack of incentives or lack of liability. Although this work does not solve these basilar problems, they must be referenced for the reader to understand the challenge they present.

An infinite number of research works has been done focusing on the BGP problems. Nordstrom and Dovrolis [34] raised awareness in the networking community for BGP and interdomain routing vulnerabilities. They believed that since no relevant attacks had been performed till that time, proper care was not being given to such a fundamental component of current Internet. A description of the main attacks existing at the time (objectives and mechanisms) was done, which included loss of connectivity, traffic subversion and data interception. For the attacks described, it was assumed one or more BGP peers had been compromised and were being maliciously controlled, therefore entering in the study field of Byzantine faults. These observations make us aware of, not only the importance that network neighbors represent as means for malicious activity but that may also have as blocks for the construction of defense mechanisms. This thesis does not deal with the

problems of malicious BGP neighbors as malicious entities of the BGP protocol, instead it presents a mechanism that considers network neighbors as important parts of security classification and that may be used for reducing the amount of malicious traffic received by a network.

One of the most feared threats related to BGP is prefix hijacking - currently without a global solution deployed - for which we had in the Pakistan Telecom/Youtube hijack event [10] one of the most well known. For better prepare network operators to deal with this problem, the research community has created some helpful tools. Lad, Massey et al. presented PHAS [29], a tool that enables prefix owners to be notified in case their prefixes' BGP origin is modified. Qiu et al. have recently developed LOCK [39], a monitoring system that enables locating prefix hijackers and therefore improving the efficiency of mitigation actions. This particular issue, although not directly approached in this thesis, serves as reference, not only to understand the dependability of the Internet infrastructure on BGP, but also to understand how the publicly available reachability information, namely BGP update messages, can be used as input in a security tool.

To better understand the network security problems of the Internet, one of the obvious steps would be to understand how exactly the Internet is laid out, what exactly is the topology of the Internet. Although apparently easy this has been a herculean task and till this moment without an exact answer. It is hard if not impossible to determine which network relationships currently exist in the Internet [35]. Some authors [18] argue it is due to limited reach of monitoring systems based in probe networks, for example RouteViews [8], RIS [9] (commonly used by the research community) and propose use of other data inputs, e.g., looking glass sites or routing policy information available in Internet Routing Registry (IRR) databases. Dimitropoulos et al. [22] observe that information only gathered from BGP tables is far from providing enough information to define complete adjacency tables for ASes in the Internet. Another difficulty is the private character that private peering Internet connections usually have, which although difficult to maintain secret for Tier-1 ASes due to their visibility, it is a lot easier for all the other ones. Other interesting data sources for AS connectivity information are the Internet Routing Registries (RIRs) that manage databases for ISPs, which can register routes and routing policies. The main problem with these databases is that their updates are not mandatory for network operators, this leads to lack of confidence in their own community regarding their accuracy, unfulfilling the whole purpose of their existence. Although all these challenges exist, to build efficient network defense mechanisms in the Internet, being aware of the network topology provides a great advantage and that was a principle used for creating the Risk Score algorithm we present in this thesis. We propose a pragmatic method to define the network topology for specific networks, in order to use it as an input for the Risk Score.

Knowing the AS topology based on the exchange of public data can therefore be of considerable value. Several tools [30, 51, 38] have been developed for monitoring BGP messages through the Internet, whether for analysis of BGP update messages patterns or associated network behaviors. One particular example is the work done by Chi, Oliveira and Zhang, whom developed Cyclops [21], a tool that provides a graphical interface for AS-level connectivity in the Internet, based on a number of data inputs, e.g., RouteViews, looking glasses. In our thesis we chose to use the RIS platform from RIPE [9] as the data input.

Studies of malicious traffic behaviors and patterns within the Internet environment have been done and some interesting conclusions have been reached. Chen, Ji and Barford [19] present an inter-

esting study on network behavior for malicious sources in which they consider the main problem of the Internet to be the lack of built-in security mechanisms, leading to the need of having several add-on mechanisms deployed. According to this work, malicious sources are mainly constant in terms of:

- Space: regarding the IP address space usage;
- Time: regarding the lifetime of malicious source addresses.

Most of the sources, which are responsible for a reduced number of attacks, also have small lifetimes, and the source prefixes responsible for the highest number of attacks, belong to the same AS. This may be due to the infection of neighboring machines and to the network effect. Authors observe the rule of 80/20 "about 80% sources locate in the same 20% IP address space over time.", for which reason they believe 20% of the IP address space should be the focus for both attackers and defenders. Another interesting observation is that a considerable amount of traffic originates from unrouted prefixes, i.e., private addresses and IANA unallocated prefixes, also known as "bogons". The existence of "bogon" routes is in itself a management problem for network operators and not always dealt in the best manner [24]. Most of these observations are corroborated by our own work in this thesis, providing strong confidence that defense mechanisms based in network behaviors are not only viable but may provide high efficiency values. The authors however do not use the results of their study for building any system, which is one of the contributions from the current work.

Through the years, most of the security research has focused on spam, the main reason for this investment is because spam has been one of the major concerns for service providers [33, 25]. Ramachandran and Feamster [40] studied this particular type of malicious traffic and tried to correlate it with network behaviors, in particular an interesting correlation between BGP prefix hijacking and spam. From their conclusions we may say that short-lived BGP routes provide a good metric for characterizing an AS as malicious. The authors believe that the network properties of malicious traffic may provide a more robust form of filtering than common content rules (e.g. signatures). The work presents an analysis of blacklists and concludes that the number of false positives is high and they are only effective if several different types of blacklists are used in conjunction. Although many problems in characterizing malicious traffic through network properties exist, it is believed that it may be the one less prone to be manipulated. Besides the advertisement of routes for more specific prefixes by spammers, which is already common knowledge, one interesting discovery is that spammers sometimes advertise large address blocks to workaround network filtering usually done for smaller network blocks e.g. less then a Class C. Authors introduce a new concept named BGP *spectrum agility* where spammers announce IP address space (usually hijacked) for short periods of time, from which they send spam. From the point of view of the spammer this technique has also the problem of circumventing black lists. For this reason we consider that defense mechanisms based on lists have high manageability costs, scalability constraints and efficiency problems, reason why we present alternatives exactly based on network behaviors to deal with malicious traffic.

Some research efforts have tried to have in consideration not only the routers near the target of the malicious traffic but the whole path, namely through different ASes. Ioannidis et al. [31] present

a protocol named Aggregate-based Congestion Control (ACC), intended to detect and limit high bandwidth aggregates in network infrastructures. Authors consider one of the main challenges to be how to distinguish between legitimate and malicious traffic, even because it is not a simple case of applying policies to one particular flow, since in both cases many undifferentiated flows may exist, hence the term aggregate. Authors describe particular mechanisms of the protocol as the calculation of rate-limit values for creating traffic filters, and explain how congestion control messages are communicated upstream to all nodes in the path (i.e. pushback mechanism) to the source of malicious traffic. We follow the rationale of this work and propose intervention policies that consider a joint effort among network neighbors to deal with malicious traffic through the use of known technologies within the network operators context (e.g., BGP).

Other approaches not so ambitious in terms of network relationships, take more pragmatic approaches. Borremans and Valke [15] embrace the inevitability of attacks affecting services and the only solution being the damage minimization through traffic diversion. Depending on the area of the network where malicious traffic is diverted, three different types of traffic diversion classifications are defined: early (at the carrier level), near (at the network neighborhood near the ISP's upstream providers and peers) and late (as near as possible to the ISP). Their work describes some DDoS defense techniques namely: Rate Limiting, Oversizing, Firewalling (TCP/UDP blocking), External ISP diversion (including BGP community dropping), Stop Announcing, Isolation (ranking of malicious sources and different announcements through different links) and some commercial implementations. The authors also take particular attention to relationships between neighbor ISPs and mention that contracts between ISPs may include network specific details, e.g., minimum size of networks to be advertised, re-advertisement of communities. The use of defense mechanisms at a network level is also used in this thesis and, regarding peering agreements, that idea can also be used as an enhancement for the work presented in this thesis. Although Borremans and Valke do not perform a thorough study of this issue, it is an interesting idea to apply.

The work done in this thesis shares the same approaches of part of the works here described, for example the importance of network behaviors of malicious traffic, the use of an evolutionary approach, avoids others, for example the use of pre-computed lists . Nevertheless, we used all the research work here presented to define a strategy on building efficient security defense mechanisms, contributing with new security classification mechanisms integrated with viable deployment scenarios.

Chapter 4

Data Analysis

In this chapter we analyze the data corpus of malicious data sent towards a mid scale ISP network. Also, we perform an analysis of network reachability information based on raw BGP data from RIS [9]. Finally we explain the software tools used to perform the analysis.

4.1 Malicious Traffic Identification

This thesis contains an analysis of malicious traffic sent towards a medium scale ISP network infrastructure for the time window from 2010-04-13 to 2010-07-12. The malicious traffic identification and classification is performed by a closed source detection platform. Two types of malicious traffic were selected: flow floods and address scans; to be described on sections 4.1.2 and 4.1.3. These classifications are not based on static signatures but on traffic behavior. For both types of traffic it is possible to refine what exactly is considered malicious or legitimate. This can be done by defining thresholds for time or for number of events after which it can be considered malicious traffic.

For the particular types of malicious events under analysis we have the following definitions:

> DEFINITION 1 *Let N and T be configurable system parameters, a flow flood event is defined by a host trying to open N connections towards a single destination host and port within a T seconds time window.*
>
> DEFINITION 2 *Let N and T be configurable system parameters, an address scan event is defined by a host sending messages towards N ports at a single destination host within a T seconds time window.*

For the particular types of malicious traffic under analysis we have the following definitions:

> DEFINITION 3 *Let M and T be configurable system parameters, flow flood malicious traffic is defined by M flow flood events detected within a T seconds time window.*
>
> DEFINITION 4 *Let M and T be configurable system parameters, address scan malicious traffic is defined by M address scan events detected within a T seconds time window.*

For the mentioned time window a total of 8453 events were detected:

- 3359 flow floods;
- 5094 address scans.

4.1.1 Traffic Anonymization

To allow the traffic analysis it was required by the ISP that the malicious traffic information be anonymized. This was achieved with the help of CryptoPAn (see chapter 4.3.4) for the IP addresses and a simple conversion mechanism for the AS numbers. We used the real source IPs of the malicious traffic as input for a whois service (see chapter 4.3.2) from which we received the real AS numbers. At this point we could anonymize the IP addresses by running a tool based on the Crypto-PAn library. Besides the IP addresses we also used the same tool to anonymize the IP prefixes. The Crypto-PAn has as one of its best features the capability to preserve the prefixes after anonymization, which means that two real IP addresses included in a particular prefix will remain in the same prefix after anonymization. This ensures that we can perform coherent analysis based on IP addresses or prefixes.

Regarding the AS numbers we used a much simpler mechanism based in the private AS numbers reserved by IANA [28] - from 64512 to 65535 - creating a translation between the real and private AS numbers, therefore ensuring the anonymity of the real ASes.

Using these mechanisms we ensure that the data analysis will still be valid, while preserving the data anonymity.

4.1.2 Flow floods

As mentioned in several research works [47, 50] flow flood attacks are considerably difficult to detect and to defense mechanisms that do not affect legitimate traffic are almost impossible to develop. The reason for this is related to the extreme difficulty in distinguishing an attack from a perfectly legitimate access to a service, e.g., flash crowd. Due to these reasons they are often used by malicious entities for inflicting distributed denial of service (DDoS) attacks. Another important factor that potentiates this type of attacks is the current availability of botnets on the black market, rented as a commercial service. Botnets not only provide anonymity and big attack power for the attacker but also, and more relevant for the DDoS attacks, diffusion.

Typical use cases include infected hosts initiating many simultaneous connections towards a specific victim. Currently, there are even certain websites that provide the infrastructure for this type of attacks from more radical activists campaigns. The impact of such attack is dependent on the network size of infected machines, higher network sizes enable more powerful attacks.

For this type of malicious traffic the following data was analyzed:

- Number of events;

- Number of bytes;
- Number of packets;
- Source IP address;
- Start timestamp.

Commonly TCP SYN flood attacks are relatively small in actual bits per second but they often exhaust other resources through an increase of number of connection requests since the end point (e.g., router) must save the connections state.

4.1.3 Address scans

Address scans are usually used for acquiring information regarding services that are open to the network. This is often a way to determine the type of attack that is more likely to succeed. Although not mandatory, some of the typical originators of such traffic are hosts infected by worms or viruses when trying to propagate to other devices. Although not all attacks are preceded by address scans (many malicious code simply searches for specific vulnerabilities in specific service ports and tries to exploit them), existing address scans usually precede attacks, for which reason they should be considered as an interesting metric for understanding the behavior of malicious traffic. Knowing which ports are open restricts the vulnerabilities that the attacker may try to exploit. For example knowing that port 80 is opened may tell the attacker that an exploit to the HTTP service may be possible.

For this type of malicious traffic the following data was analyzed:

- Number of events;
- Number of bytes;
- Number of packets;
- Source IP address;
- Start timestamp.

4.1.4 Analysis

In this chapter we analyze the characteristics of malicious traffic as a whole and a more detailed analysis of the top 5 higher originators of malicious traffic. The analysis will be performed for the two types of malicious traffic previously described in 4.1.2 and 4.1.3. Due to the considerable amount of different traffic sources and since most of it was from a small subset of origins, selecting the top 5 most significant origins reduced the analysis scope allowing a more thorough analysis of the relevant sources.

Cumulative

When considering the whole data set from which this analysis was done, we get 474 different AS origins. However, looking at Figures 4.1 and 4.2 we can conclude that only a few are responsible for most of the malicious traffic.

The top 5 ASes for address scans are described by the ordered set $\{64519, 64524, 64521, 64517, 64528\}$ and for flow floods by the ordered set $\{64549, 64557, 64542, 64562, Reserved\}$. We can observe that although address scan events have approximately 50% more events than flow floods, both figures have similar curves. This means although the magnitude of values differs, trends for both types of events are similar.

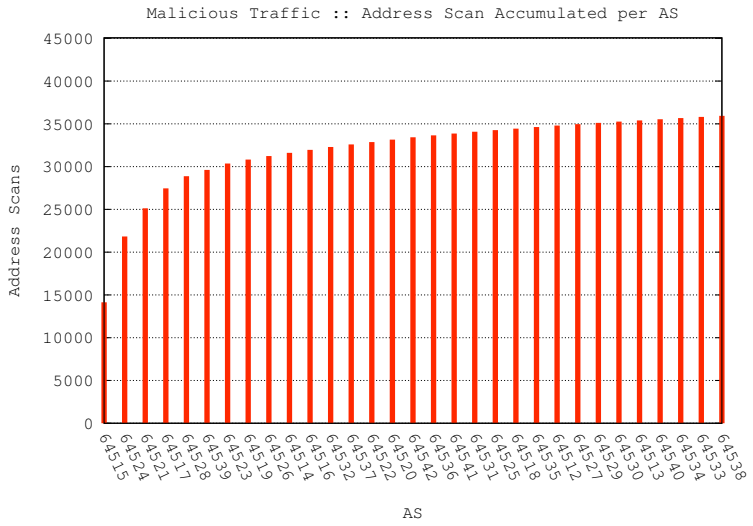

Figure 4.1: Address scans cumulative source addresses.

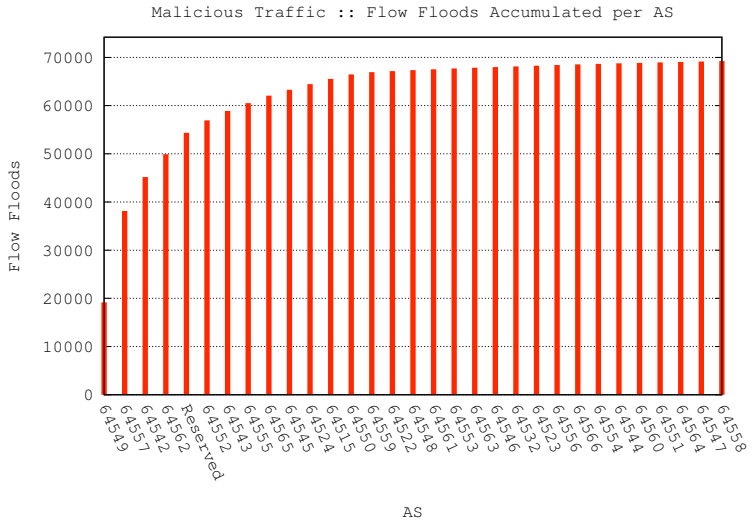

Figure 4.2: Flow floods cumulative source addresses.

From the charts we can see that flow flood events reach a ceiling faster than the address scans. Although the behavior is similar, the ASes responsible for contributing to the cumulative value of flow floods become almost irrelevant for other than the first fifteen. This observation is more visible in the ranking charts present in figures 4.3 and 4.4, where the number of flow flood events is approximately zero.

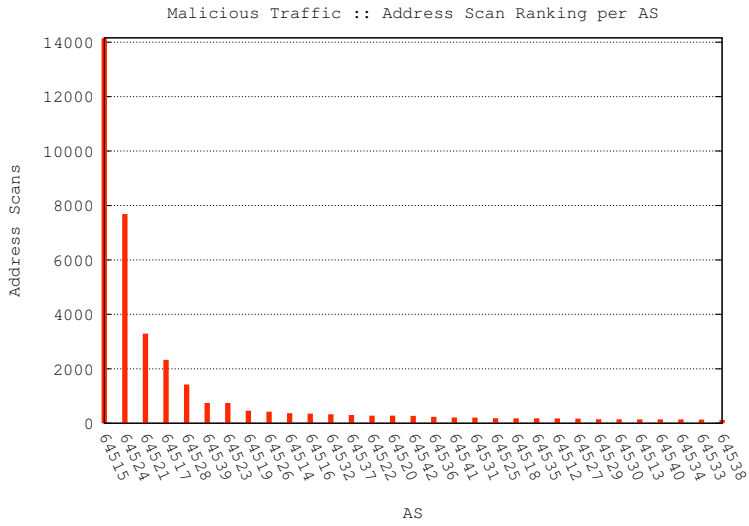

Figure 4.3: Address scans ranking source addresses.

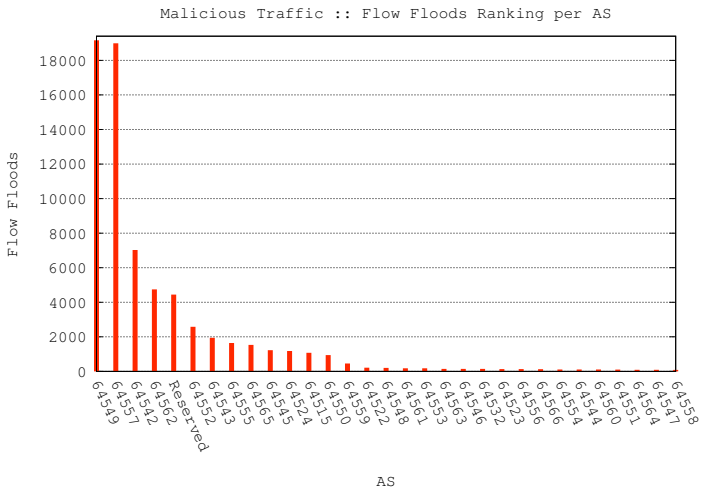

Figure 4.4: Flow floods ranking source addresses.

For address scans however, we can see that this difference is not so abrupt. Although, as mentioned before, the first ASes are the main originators of malicious events, after the first ten the trend toward value zero is also visible but with a softer slope. From this observation we may also conclude that address scans are more common than flow floods.

Prefix distribution

Although it is interesting and relevant to know the amount of malicious traffic originating from a particular AS it is also relevant to know if sending malicious traffic is a generalized behavior of the AS or simply the action from a particular segment of the network, namely a particular prefix. For this reason, an analysis of the prefixes originating the traffic was also done. From the total number of ASes we chose a subset to understand what exactly was the impact of particular prefixes versus the total malicious traffic originating from the AS. We chose 8 ASes for address scans and 10 ASes for flow floods since, as previously explained, the contribution of the first ranked ASes for flow floods is more relevant than for address scans. The choice of prefixes was done based on their ranking regarding all the detected prefixes, the ones responsible for originating the higher number of events were selected.

Figure 4.5: Prefixes versus AS address scans.

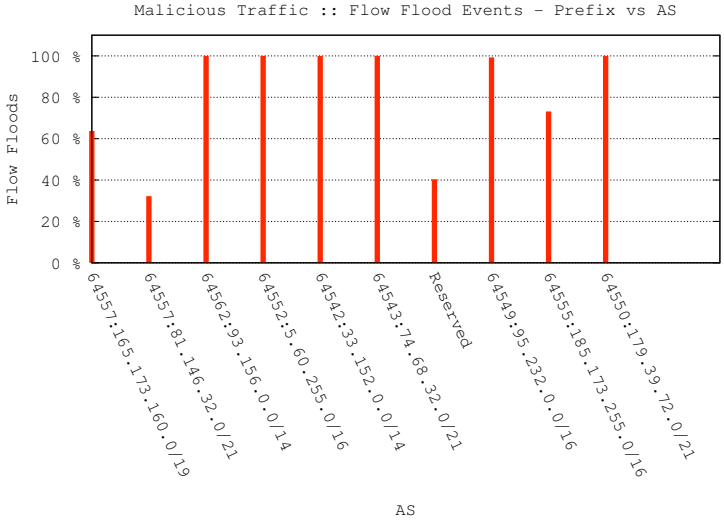

Figure 4.6: Prefixes versus AS flow floods.

From the charts in figures 4.5 and 4.6 we observe that two main types of situations exist:

1. Malicious traffic from an AS is originating mostly from a single prefix, sometimes completely, e.g., 64562:93.156.0.0/14 (flow floods);

2. Malicious traffic from an AS is originating from more than a single prefix, e.g., 64524:67.248.255.0/18 (address scans).

Considering the first situation, if defense mechanisms are put in place to filter only the malicious prefixes, they will have high efficiency values. For the second situation, it may be possible to apply more interesting defense mechanisms, for example depeering from ASes that provide transit for traffic originating from that particular AS identified as malicious. These defense mechanisms are based on the detection and classification platform, which accuracy is dependent on several factors, namely thresholds definition. Knowing that such platforms are not 100% accurate, we must always be careful when introducing harsh measures into the defense mechanisms and we must have in mind that an evolutionary approach to the detection platform itself is required, with the thresholds being constantly validated and fine-tuned. Another issue to be considered regarding the detection mechanism is the possibility of packet origins being spoofed with addresses from a legitimate AS and that AS being wrongly classified as malicious. The top malicious ASes are not required to have direct links to the ISP, which greatly increases the success potential of such spoofing mechanisms. This scenario could therefore become an attack strategy from a malicious entity with the purpose of forcing a legitimate AS to be considered malicious.

For flow floods in particular, a considerable amount of traffic is originated from reserved IP address ranges, which are enumerated in Table 4.1. It is interesting to notice that most of these detected malicious events is from private IP addresses [42] or, typically, misconfigured machines [20]. This type of traffic can be simply ingress filtered using access control lists (ACLs), however it imposes allocating resources in segments of the network that may be under extreme stress, for which reason having more lines inserted in ACLs is sometimes avoided by network operators.

IP Network	Description	Number of Events
5.0.0.0/8	IANA RESERVED-5	933
169.254.0.0/16	Link Local RFC3927 IANA Reserved[20]	2173
192.168.0.0/16	Private Address CBLK RFC1918 IANA Reserved[42]	1383

Table 4.1: Reserved IP addresses.

In the case of address scans, five of the first eight prefixes are originating less than 21% of the total number of detected events. As for flow floods, eight of the first ten prefixes are responsible for more than 60% of the total number of detected events, the prefix with less number of events (64557:81.146.32.0/21) has 32%, which is still a large percentage from individual prefixes. From these results we can understand that depending on the type of malicious traffic being analyzed, the most efficient defense mechanisms changes. A simpler and more flexible approach for supporting different types of malicious traffic may be to consider a threshold value to define which defense mechanisms to apply.

Temporal distribution

For the top 5 ASes we obtain the charts presented in Figures 4.7 and 4.8 from which we can directly compare the traffic originated from each of the ASes through the time window analyzed.

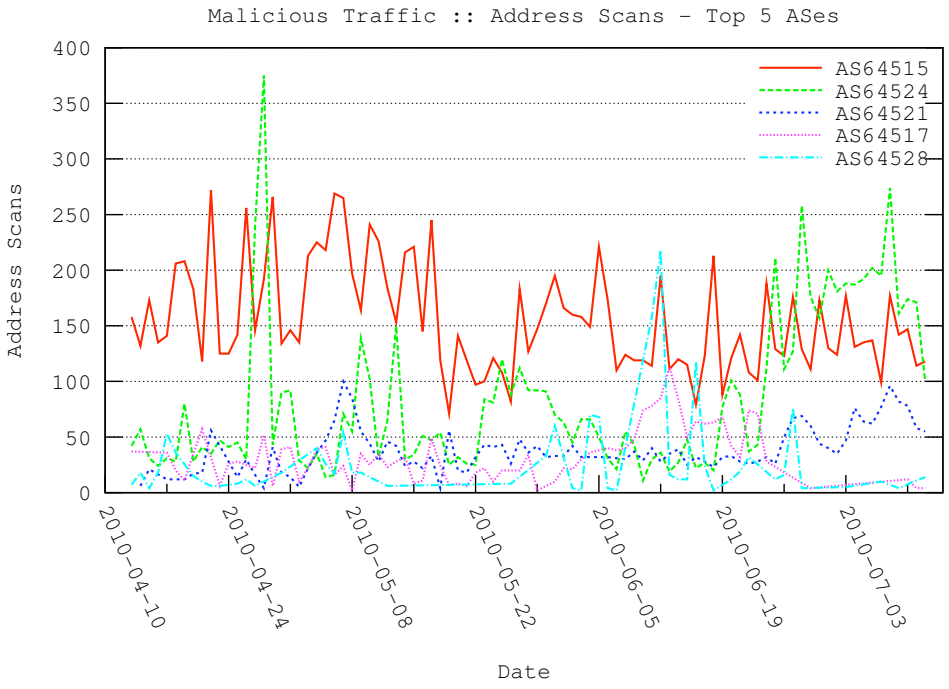

Figure 4.7: Time series for top 5 address scan origins.

For address scans there are two ASes responsible for most of the traffic, AS 64515 and AS 64524, however the former has the most consistent behavior. If the malicious traffic from these two ASes could be mitigated it would have a high impact in the whole network infrastructure of the ISP. An interesting observation that can be made from the chart is the sawtooth pattern of the time series events. This is true for all five ASes depicted in the chart. One possible reason for this could be the periodic behavior that characterizes many of the propagation mechanisms of worms. From the chart we can also observe an increase in the total number of detected events from the month of June onwards. Although AS 64515 is the most constant and therefore the one contributing with the highest number of events overall, in this time window it had a reduction while the other ASes had an increase in their contribution of address scan events.

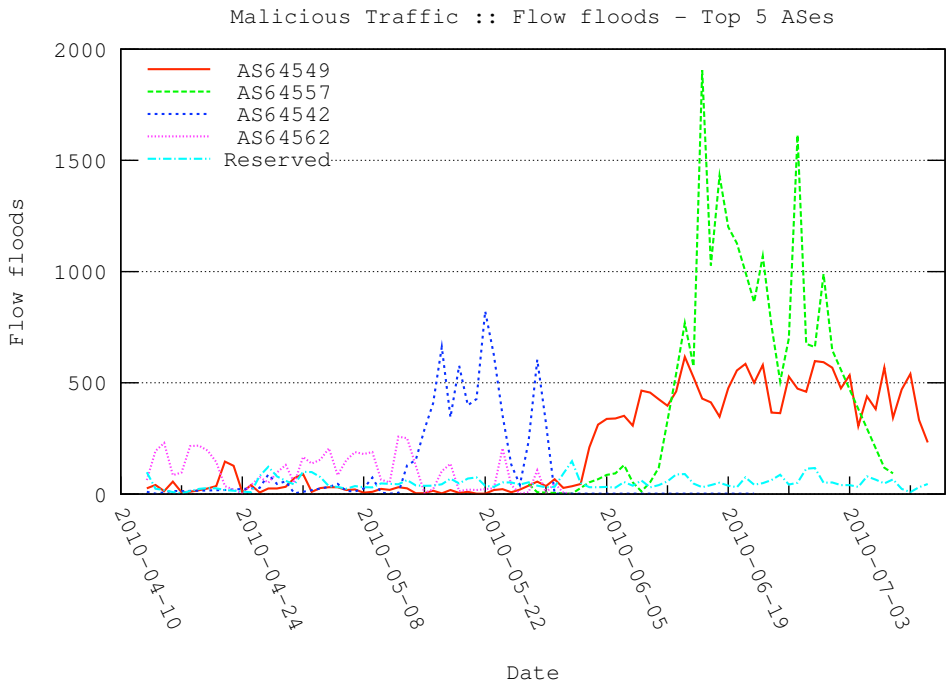

Figure 4.8: Time series for top 5 flow flood origins.

For flow floods there are also two ASes responsible for most of the traffic, AS 64549 and AS 64557, however, although they have the highest volumes of traffic, that traffic is not constant through all the 4 months of data, most of it is concentrated on the months of June and July. This seemed to be a very active period for malicious activity from these two ASes but a more peaceful one for all the other ASes, considering that the majority almost ceased activity. Similar to address scans, mitigating flow floods from these two ASes would have a high impact in the whole network infrastructure of the ISP. As mentioned for address scans, also with flow floods we can observe a sawtooth pattern, which may be due to the same reasons as the ones described for the address scans case.

AS distribution

Another interesting perspective over the data is given by the way malicious activity is distributed across all ASes, and how each AS contributes to the total malicious traffic universe.

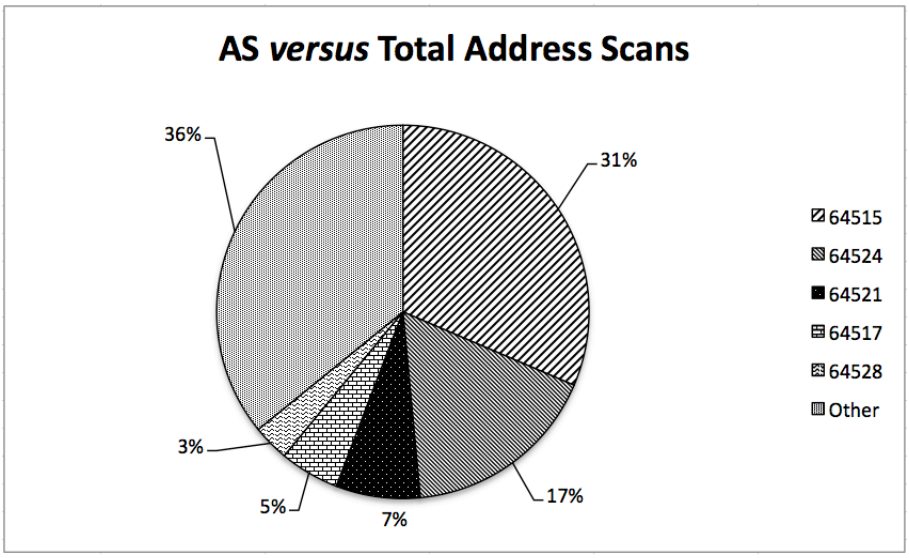

Figure 4.9: AS versus Total address scan events.

Considering the address scans (figure 4.9), most of the malicious traffic (36%) is spread ac cross the AS space, however a significant amount is originating in a small subset of ASes, in particular AS 64515 with 31% and AS 64524 with 17% of all malicious traffic.

The same conclusions can be obtained for flow floods (figure 4.10), in which 27% of the malicious traffic is spread through different ASes but three of them, namely ASes 64557, 64549 and 64562, are contributing with significant volumes of malicious traffic, 26%, 26% and 6% respectively.

For both types of malicious traffic it is perfectly viable to focus our mitigation efforts only in a small subset of ASes.

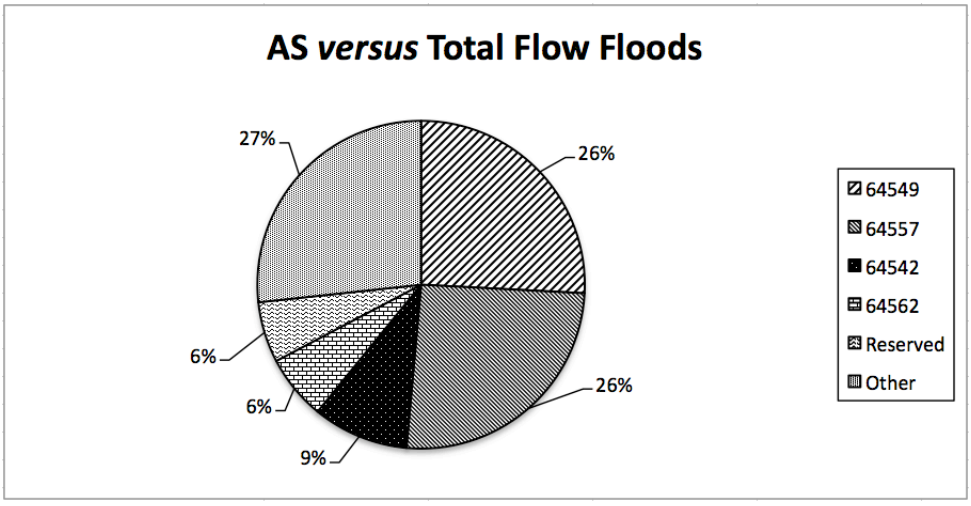

Figure 4.10: AS versus Total flow flood events.

Given the data set provided by the monitoring system, we analyzed how all the malicious traffic is distributed through the AS space in order to better understand how a real ISP is targeted. We could

conclude that although traffic is sent from many different sources, most of the traffic can be contained within a small subset, in our case we chose 5 ASes which are responsible for approximately 64% of all address scans and 73% of all flow floods. This conclusion gives hope that protecting the network from a particular AS can have a great impact in the overall security of the network. To consider this information when designing defense mechanisms can be extremely useful not only for estimating their efficiency but also their scalability.

4.2 Network Reachability Information

4.2.1 BGP update messages

Considering that our thesis is under the scope of a particular ISP (i.e., AS), all its neighbors may be defined using the BGP information available from the core routers of the ISP. The network topology issues only arise regarding the topologies of other ASes, since one of the main challenges of defining the Internet AS topology is exactly the lack of complete BGP public information, which can be due to different reasons e.g. business secrecy, traffic aggregation [18].

From the BGP update messages we are only interested in a subset of fields:

- Prefix - IP prefix;
- Origin - origin AS;
- Type - announcement, withdraw;
- AS Path (see section 2.1 for details).

With the above information we are able to understand what ASes an IP packet will traverse to reach a particular prefix. First we need to understand how the Internet sees the ISP AS, what known paths currently exist, in order to create the ISP AS reachability map. Second we need to understand how we can classify an AS in terms of neighborhood, i.e., which ASes are directly connected. Since the Internet is an extremely vast and dynamic environment we must restrict our scope of analysis. For achieving this purpose we defined the following sets of ASes we considered interesting:

1. AS paths related to the ISP AS prefixes;
2. AS paths that include the top 5 sources of malicious traffic previously determined.

4.2.2 Routing Information Service from RIPE

Different platforms for collecting BGP routing tables and update messages information exist, namely:

- Route Views;
- Looking Glasses;

- RIPE Routing Information Service (RIS);
- CAIDA.

For this work we chose to use RIS [9]. RIS has a network of 14 active probes, which collect BGP update messages from several locations around the world (see table 4.2). This information enables us to know, approximately, how other ASes in the Internet reach a particular AS.

Probe	Location
RRC00	RIPE-NCC Multihop, Amsterdam
RRC01	LINX, London
RRC03	AMS-IX / NL-IX / GN-IX, Amsterdam
RRC04	CIXP, Geneva
RRC05	VIX, Vienna
RRC06	DIX-IE, Tokyo
RRC07	Netnod, Stockholm
RRC10	MIX, Milan
RRC11	NYIIX, New York
RRC12	DE-CIX, Frankfurt
RRC13	MSK-IX, Moscow
RRC14	PAIX, Palo Alto
RRC15	PTTMetro, Sao Paulo
RRC16	Terremark - NOTA, Miami

Table 4.2: RIS probes.

We can parse the BGP update messages, searching for announcements of prefixes belonging to the ISP AS, in order to collect all the published AS paths. This defines the first set of interesting ASes too look for in BGP update messages. For the second set of ASes we must filter all announcements/withdraws that include the malicious top 5 ASes.

Given these two sets of ASes, we are now able to know all paths from the Internet towards the ISP AS and an adjacency list for all interesting ASes with their neighbors, which may be malicious or benign. This approach to the total universe of existing ASes provides scalability to our approach and therefore to the classification mechanism being presented in this work. The goal of this step is not to create an exact topology of data flows but to have the most probable paths that data may take towards the network operator AS, since it is not possible to determine - in useful time - the exact path that a data flow will take.

4.2.3 Analysis

Besides the reachability perspective given by the RIS service we could have yet another one, the ISP network that could be given by its BGP speaking routers. The former one gives the perspective of Internet-to-ISP and the latter one ISP-to-Internet. To understand how malicious traffic is sent towards the ISP network we should focus on the Internet-to-ISP perspective, i.e., the probes network.

For analyzing the BGP update messages we chose a period of 8 days, from June 1 to June 8 2010. For that period we retrieved the raw update messages available at RIS and parsed them with the goals previously explained (i.e., ISP reachability map, AS adjacency list).

To understand the dynamics of BGP update messages, we present the amount of announcements and withdraws observed for only this period:

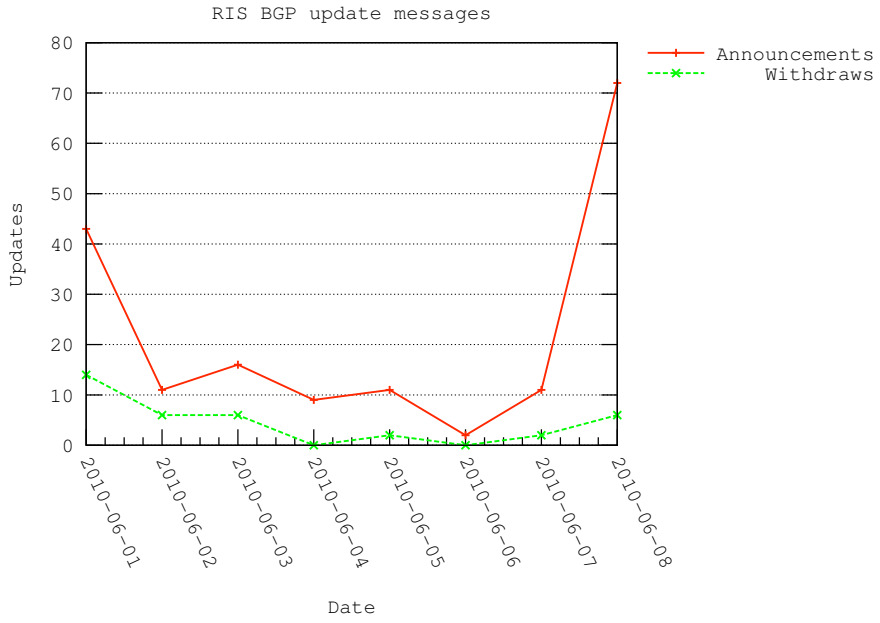

Figure 4.11: BGP update messages dynamics.

As reference for parsing the BGP update messages, we considered the ISP AS with ASN 3243. Filtering reachability information for its prefixes, a total of 49 unique paths were detected, from which information we could create the network topology present in figure 4.12.

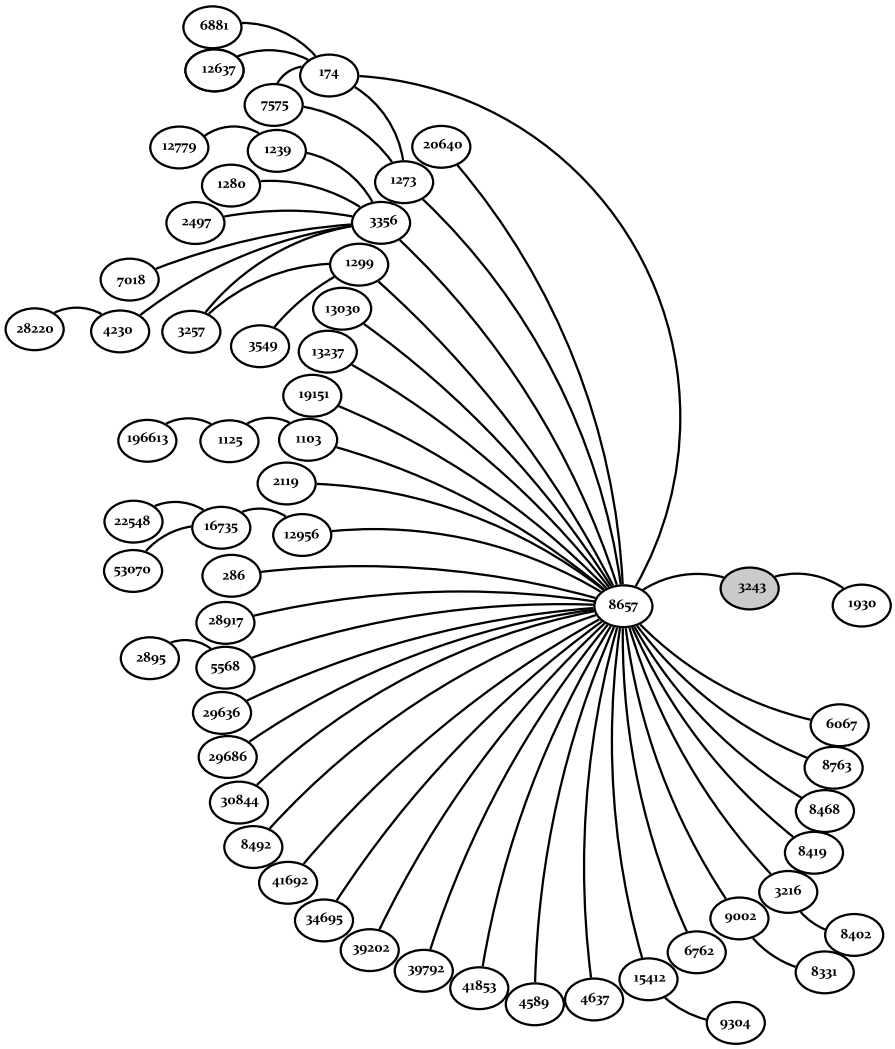

Figure 4.12: AS 3243 public reachability paths.

From the above chart we conclude that AS 3243 has only two different access points **from** the Internet **to** its network: AS 8657 and AS 1930. This however is not a full characterization of the AS topology since it does not consider its role as a transit AS. From the BGP update messages we could also discover other directly connected ASes, not included in BGP update messages for prefixes from the ISP AS, but for which the ISP provides transit: ASes 8426, 42863, 12527, 28672, 41159, 65001, 9118, 31497, 39088, 28998, 15525, 47784, 29673, 13200, 6773, 25253, 34873 and 6939. As mentioned before we must not forget that we are interested in collecting all the neighbors of ASes:

- in the path to the ISP AS, which can be done by filtering BGP update messages that include the ISP AS in their AS PATH attribute;
- neighbors of the malicious ASes.

After collecting this information we have fulfilled the minimal requirements for creating a topology snapshot of the ISP AS and its neighborhood. Regarding the top 5 ASes responsible for originating

malicious traffic, the same rationale can be used to create their topology and neighborhood characterization. At this point it is not possible to analyze this information since we do not know the public ASNs of the malicious ASes, due to the anonymization process explained in 4.1.1.

4.3 Software Tools

To perform the data analysis a set of tools and processes were used. When choosing the best tools for this task, we were looking for tools simple to use, easy to integrate, fast to deploy and with enough background to provide confidence in their quality. Although not all tools share the same licensing scheme (e.g., GPL, Google CLA) they are all open-source, with good documentation and with a good community support.

4.3.1 PyBGPdump

To perform the analysis of the RIS raw files a support software was required. We chose to use PyBGPdump, a Python library that enables parsing of BGP dump files produced by Zebra/Quagga or Multi-threaded Routing Toolkit (MRT)[14], a routing information export format. PyBGPdump is based on the dpkt Python library that provides packet manipulation functionalities for some TCP/IP protocols, namely BGP, and MRT. The current version of dpkt (1.7) does not support ASNs with 4 bytes nor MRT TableDumpV2 RIB dumps, for which reasons the patch provided with MRT dump file manipulation toolkit [43] from CAIA was applied. Since some of the MRT files can be of considerable size, one interesting feature of this tool is its capability to support gzip'ed and bzip2'ed files. Following is a simple source code snippet written in Python that parses a MRT file and goes through its attributes, in order to understand the ease of use of this tool:

```
# Import an MRT file
dump = pybgpdump.BGPDump(mrt_file)
# Access the BGP attributes
for mrt_h, bgp_h, bgp_m in dump:
        for attr in bgp_m.update.attributes:
            if attr.type == bgp.AS_PATH:
                print "AS_PATH"
            elif attr.type == bgp.ORIGIN:
                print "ORIGIN"
            elif attr.type == bgp.NEXT_HOP:
                print "NEXT_HOP"
            elif attr.type == bgp.LOCAL_PREF:
                print "LOCAL_PREF"
            elif attr.type == bgp.MULTI_EXIT_DISC:
                print "MULTI_EXIT_DISC"
            elif attr.type == bgp.COMMUNITIES:
                print "COMMUNITIES"
```

```
        elif attr.type == bgp.ATOMIC_AGGREGATE:
            print "ATOMIC_AGGREGATE"
    # Print the number of routes in the update message
    print "BGP total announced routes %s" % bgp_m.update.announced
    print "BGP total withdrawn routes %s" % bgp_m.update.withdrawn
```

4.3.2 Whois service

To perform the traffic analysis it was necessary to map the source IP addresses of the malicious traffic detected to the ASes to which they belonged. Typical whois services were not applicable due to the large amount of queries required, which is reason enough for some of the publicly available whois services to black list the host executing the queries. An online free service was therefore used that allows bulk queries given a set of IP addresses: cymru-ip2asn [5]. The service syntax is extremely simple and following is a quick example:

```
$ netcat whois.cymru.com 43 < malicious_traffic_ip_addresses
```

As we can see it uses the GNU netcat command to send the IP addresses from a text file to the whois service available at port 43 of host whois.cymru.com and it returns their ASes. For example if we create a file with the IPs of the following hosts:

- ris.ripe.net
- www.iana.net

we would get the following answer from the service:

```
AS      | IP               | AS Name
3333    | 193.0.19.19      | RIPE-NCC-AS RIPE Network Coordination Centre
AS      | IP               | AS Name
40528   | 192.0.32.8       | ICANN-LAX - ICANN
```

To parse the malicious traffic data and execute the queries, we used mainly three UNIX utilities: `cat`, `awk` and `netcat`.

4.3.3 ipaddr-py

To contextualize the analysis of the BGP update messages received from RIS with the IP addresses of the top sources of malicious traffic, ipaddr-py [6] was used. It is a Python library developed at Google for manipulation of IPv4/IPv6 addresses and prefixes. Some of the functionalities used for the analysis were:

- Check if IP addresses are contained within a specific prefix;

- Check if an IP prefix is included within another prefix.

Using this library was extremely easy and the integration with the PyBGPdump library (see section 4.3.1) was straightforward. Following is a small Python code snippet used for checking if an IP prefix is included in a set of prefixes:

```
prefixes.append(IPNetwork('10.0.0.0/13'))
prefixes.append(IPNetwork('192.168.0.0/16'))
prefix_object = IPNetwork('10.0.1.0/24')
if (prefix_object in prefixes):
    print "Prefix %s is included in prefixes set."
else:
    print "Prefix %s is not included in prefixes set."
```

4.3.4 Crypto-PAn

For the analysis performed in this thesis it was not important to have the raw data with the original addresses, for example we did not intended to make a correlation of the malicious activity with geographical locations. To analyze the data ensuring that it could be kept anonymous, Crypto-PAn [3], a cryptographic sanitization tool was used. Some of the main characteristics of Crypto-PAN are:

- Based on the Rijndael block cipher;
- The cryptographic algorithm receives a secret key for the IP addresses anonymization;
- Using the same secret key for different traffic traces ensures consistency, i.e., the same IP address will be anonymized to the same address in different traces.

Crypto-PAn is a well established and mature tool, integrated with other tools from The Cooperative Association of Internet Analysis (CAIDA), e.g., nfdump [7], CoralReef [2], and used in several research works that required trace anonymization features [45, 44, 46, 52, 26].

Chapter 5

Risk Score

In this chapter we present a method for classifying a network neighbor in terms of security under the scope of an ISP, we explain the goal of the classification method, the metrics, the algorithm used and simulations for understanding the behavior of the algorithm in face of malicious traffic dynamics.

5.1 Concept

With the data analysis performed for a particular ISP we were able to better understand how certain types of malicious data (i.e., address scans and flow floods) behave. Although each ISP in the Internet has its own specificities we think extrapolating this analysis to common mid scale ISP can be done. We intend to use the observations reached from the data analysis into the design of defense mechanisms. Different approaches exist to deal with malicious traffic, some are host centric, others are network centric, given the observations presented in chapter 4, we are presenting in this thesis an approach from the latter.

To enable defense mechanisms based on network behaviors we propose a method named Risk Score, which quantifies each network neighbor in terms of the security risk it represents to the ISP. The goal of the Risk Score is to enable decision processes for choosing the more appropriate defense mechanism: depeer; filter specific prefixes from malicious ASes; no action, i.e., allow malicious traffic to flow.

Since security classification is a rather complex and dynamic field, dependent of different events, the classification method should have the flexibility to support the contribution of several metrics and the inclusion of new ones. This was achieved by creating a metric correlation algorithm with custom parameters.

As explained by other authors [22, 18] different types of relationships exist for directly connected ASes. In the context of malicious traffic origins, the AS may be the originator of malicious traffic or be a transit AS for malicious traffic. The latter is the most common situation and the classification mechanism was required to address this issue. This means that although a network operator of a

directly connected AS may not be responsible for the malicious traffic, it can use the same rationale to its directly connected ASes, i.e., using a recursive strategy in order to restrict the actions of the malicious AS, mitigating its malicious traffic. Having this definition in mind the Risk Score should be defined with focus on providing a characterization of the ISP neighbors and not only on the malicious AS itself.

Since one of the goals of our work is to decide if neighborhood relationships (e.g., peering) are appropriate, this property was fundamental. In the particular case of peering, the rationale behind depeering is to reduce the level of malicious traffic received in a network operator using network mechanisms but also creating disincentives for providers to transport this type of traffic, which is usually extremely complicated if not impossible due to the way Internet is organized 2.1. Besides technical mechanisms we believed that business logic should also be addressed by this work. When announcing the depeering decision to the neighbor AS, the latter is forced to perform a thorough analysis of pros and cons of what is more beneficial. Whether in a peer or transit relationship, Tier-1 ASes are in a strong position to implement this type of mechanisms, however stub ASes are in an extremely fragile situation since whatever measure they implement it will only have local effect.

Considering that individual initiatives from ASes are not as efficient as congregated actions from the Internet community, we created the algorithm to facilitate this interaction. If more than only an AS uses this approach, ASes will have the incentive to mitigate the malicious traffic in their own network.

Following a description of the main entities used in the construction of algorithm will be provided. To better understand the purpose of the metrics and parameters used in the construction of the algorithm, we must have in consideration the ultimate goal of the Risk Score: risk classification of the network neighbor. It is not intended to classify the malicious ASes themselves, they are simply part of the classification process of the neighbor.

5.2 Metrics

Malicious AS We define a malicious AS as a network entity from which malicious traffic originates and its ranking in the universe of ASes that send malicious traffic to the operator network, exceed a certain upper bound. Considering the analysis previously made to the corpus of malicious data, we can state that most of the malicious traffic has its origin in a small subset of the total ASes with which the network exchanges traffic. To quantify the upper bound, we use the values observed in the real network data. The concept of malicious AS is therefore tied to the network under study and takes in consideration the detected values for all ASes.

Address scans traffic ratio Quantifies the amount of address scans originated from an AS *i* in comparison to the total amount of address scans received from all ASes, i.e., $\frac{AS_i\ address\ scans\ detections}{Total\ taddress\ scans\ detections}$.

Flow floods traffic ratio Quantifies the amount of flow floods originated from an AS *i* in comparison to the total amount of flow floods received from all ASes, i.e., $\frac{AS_i\ flow\ floods\ detections}{Total\ flow\ floods\ detections}$.

5.3 Custom Parameters

We now present all the customizable parameters used for calculating the Risk Score for a particular AS. This construction assumes to be biased by the ISP network operator security policies, which is not considered to be a problem since it is assumed that the Risk Score is used in the context of a particular ISP and not to be used as a comparison mechanism between different ISPs. However, we still propose default values for each of the parameters, not only to allow comparisons if required but also to simplify operational actions, namely configuration.

Weight Considering that the relevance of each type of malicious traffic should depend on the network operator security policy, a weight value is given for each of the different types of malicious traffic. In the particular case of the work done in this thesis it means address scans and flow floods. Weight can have values from 1 to 5, depending on the relevance a metric has to the network operator, they have to the operator, the higher the value, the higher the importance. For weight values associated with all types of malicious traffic we propose 3 as the default value.

Karma Although an AS is responsible for originating malicious traffic, the network operator of the target network, i.e., the ISP, may find strategic value in that AS. Whether this is due to: business agreements with the malicious AS (e.g. content provider for the ISP, peering contracts); volume of legitimate traffic much higher than the volume of malicious traffic. Although these reasons are often responsible for malicious traffic not being mitigated efficiently, it was important for the Risk Score expression to reflect this type of particularities. In this sense, the Karma parameter could also be named pragmatic parameter, since it is exactly what introduces to the expression, pragmatism. The use of the Karma parameter is strictly dependent on the policy of the network operator - in the same sense as the LOCAL_PREF attribute is used in BGP, according to which an internal routing policy imposes that certain routes be preferred in detriment of others - the value of Karma follows that same rationale, it exists as a mean to impose a policy and therefore strictly dependent on the network operator decision. The Karma is a custom parameter defined per AS, which may have values from 1 to 5, depending on the relevance they have to the operator, the higher the value, the higher the importance. For Karma and regarding all possible ASes, we propose 1 as the default value. This value indicates that by default we do not trust ASes originating malicious traffic. It is the responsibility of the network operator to attribute values to Karma and it must be applied to all ASes considered malicious independently of their distance towards the ISP.

Neighborhood Level (NL) Defines the level of neighbors to check for within the adjacency table. The adjacency table has the mappings of directly connected ASes per AS. With NL it is assumed that we do not require a complete mapping of the Internet AS topology, we only want to consider ASes within a certain scope, that scope is given by the NL and the BGP paths for the network operator prefixes.

Proximity Level (PL) Defines the malicious AS proximity to the ISP network, i.e., if it is a directly connected neighbor or not. It is the only parameter that can be fixed, it only has 2 levels: directly

connected; not directly connected. In the former case PL is 2 and in the latter it is 1. Since directly connected ASes have higher control of the malicious traffic originating in their networks, their responsibility is higher and that is reflected in the Risk Score.

5.4 Malicious AS in Path

To define which malicious ASes exist in all the *known/possible* paths towards the ISP we use the BGP data available from RIS (see section 4.2.2) and we parse it to obtain all the AS paths (see section 2.1) that are announced for the prefixes belonging to two entities:

1. The ISP AS;
2. The malicious ASes.

With this information we build an adjacency table composed of all existing ASes included in paths announced for prefixes from the ISP and the malicious ASes.

Due to the complexity of the Internet topology, this metric is the most complicated one to obtain and some assumptions are required:

- From the AS paths present in the BGP update messages, every direct link connecting two neighbors can be used in a path;
- The values present in the AS paths are correct, i.e., we do not consider malicious manipulation of BGP update messages.

OBSERVATION 1 *Every possible path should be used since we do not control the way operators change their routing policies. For this reason we take a pessimistic approach and consider that every possible path, which includes malicious ASes, should be accounted for in the Risk Score.*

OBSERVATION 2 *Since we only consider the BGP messages related to ASes previously classified as malicious, the possible paths of interest are a small subset of the whole Internet. This helps providing scalability to the system without reducing its efficiency by giving it focus.*

For a simple example let us consider figure 5.1, where M1, M2, M3 are malicious ASes and A is the ISP AS.

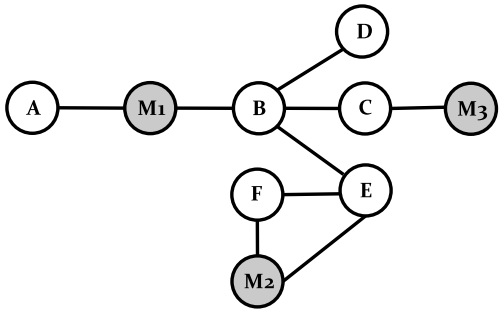

Figure 5.1: AS topology example.

From RIS we collect the reachability information for AS A and for malicious ASes M1, M2, M3, by parsing the BGP update messages and collecting all the AS paths from their prefixes. The parsed AS paths are presented in Table 5.1. An important observation is that the efficiency of RIS data is constrained by their network of probes and we should have present that most of the BGP messages exchanged by ASes in the Internet are not available in RIS.

AS Path
{C, B, M1, A}
{D, B, M1, A}
{M2, F, E, B, M1, A}
{F, E, B, M1}
{F, M2}
{F, E, B, C, M3}

Table 5.1: RIS reachability map for ASes A, M1, M2, M3.

With this information we are now able to build the adjacency table 5.2

AS	Directly connected ASes
A	M1
B	M1, C, D, E
C	B, M3
D	B
E	F, B
F	M2, E
M1	B, A
M2	F
M3	C

Table 5.2: Adjacency table example.

We can now select the ASes that will contribute for calculating the Risk Score based on the following four elements:

1. The neighbor;
2. The NL;
3. The ISP network reachability information (table 5.1);
4. The adjacency table (table 5.2).

An important observation is that, although in this example we are assuming all prefixes from a particular AS have the same reachability information, i.e., same AS path, in a real case scenario this may not occur. The consequence of this observation is a bigger adjacency table. When defining the adjacency table, we intentionally ignore which prefixes routing information were parsed, we are only interested from which AS they are originated. Although more disperse, this choice provides greater flexibility regarding the paths malicious traffic may flow through.

5.5 Correlating Metrics

Continuing with our example let us consider that we want to calculate the Risk Score of the neighbor M1. We define the NL for the neighbor (M1); with the information from tables 5.1 and 5.2 we create the AS table for that particular neighbor (table 5.3). To better understand the impact of defining NL we present two different sets of ASes for different values of NL.

Neighbor	Reachable ASes	NL
M1	B, C, D, E, F, M1, M2	1
M1	B, C, D, E, F, M1, M2, M3	2

Table 5.3: Neighbor AS table example.

To build the AS table we use as reference the different AS paths from table 5.1, we follow each path and add each of the ASes present in the path to the table. If NL is 1 then we only put into the table the ASes existing in the AS paths associated with prefixes from the ISP AS. If NL has a higher value, e.g., 2, then we use ASes from previous NL, e.g., 1, and consulting the adjacency table we add the direct neighbors of those, meaning that another hop is considered. Using higher values of NL ensures a wider set of ASes will be analyzed for the Risk Score and greater sets are preferable since they provide better accuracy by using more ASes for the neighbor classification, however, given the size of current Internet, is also a mean to limit the scope of the algorithm, providing scalability. The values used in table 5.3 were selected as **example values** but higher values could be used, e.g., 4.

Now that we know which ASes can reach the ISP AS through a particular neighbor, we can start calculating the Risk Score. For each day i, the following expression is calculated for each of the N malicious ASes, considering a total of Q metrics M, each with a given weight W. We name it α:

$$\alpha_i = \frac{\sum_{j=1}^{Q} M_j \times W_j}{karma_i} \quad , \begin{matrix} \{W_j \in \mathbb{N} : \ 1 \ \leq \ W \ \leq \ 5\} \\ \{karma_i \in \mathbb{N} : \ 1 \ \leq \ W \ \leq \ 5\} \end{matrix} \tag{5.1}$$

Now we calculate the sum of α values from N malicious ASes, which is in fact the instant value of the Risk Score. We name it β.

Let β_i be the value β for i days, N be the total number of malicious ASes, α_{ij} the value α_i for AS j and PL_j the PL value for AS j:

$$\beta_i = \sum_{j=1}^{N} \alpha_{ij} \cdot PL_j \tag{5.2}$$

Since the risk should depend on events that occur through time, it makes sense to have that rationale reflected in the final expression through a feedback mechanism. We chose to implement it in a simple form, the expression considers the previous values of β within a 30 days time window, according to the following expression:

$$I_i = I_{i-1} + \beta_i \qquad \text{, let } i \text{ be the number of days and } Max(i) = 30 \tag{5.3}$$

By choosing a 30 days time window we ensure scalability to the algorithm since it is not required to keep the total of previous Risk Score values, only the last 30 days, a time window of this length provides enough information for classifying a particular neighbor.

The Risk Score expression R for a period of i days for a particular neighbor is described by:

$$R_i = \begin{cases} \beta_1 & , i = 1 \\ \frac{I_i}{i} & , 1 < i \leq 30 \\ \frac{I_i}{30} & , i > 30 \end{cases} \tag{5.4}$$

Following we summarize the steps executed to calculate the Risk Score of an ISP neighbor:

1. From the malicious traffic AS distribution explained in section 4.1.4, determine the top 5 malicious ASes per malicious traffic type. This step defines the universe of malicious ASes;
2. Define the NL to be used and the weights to be used per metric;
3. Considering the BGP reachability information, get the AS paths for the ISP, for the malicious ASes and create the adjacency list;
4. Choose the network neighbor, its karma value according to the network policy and the NL parameter;
5. With the adjacency list and the NL parameter, discover which malicious ASes (determined in step 1) are contributors for the Risk Score of that particular neighbor;
6. Given the metrics presented in section 5.2 we calculate the Risk Score according to the expressions for R (5.4), α (5.1) and β (5.2) considering the contributing malicious ASes calculated in step 5.

5.6 Comparing Risks

A common way of calculating risk in security applications is to have a score that has its values within a limited interval, e.g., from 1 to 5. Although this approach simplifies the analysis of security status for network operators, it has the disadvantage of loosing granularity. For this work we intend to make possible a granular comparison of risk from the different neighbors, for which reason this method will not be used, we will instead accumulate the values of the different malicious ASes.

For calculating the Risk Score we should not make an average of all malicious ASes but instead add all the individual values of α. This choice has two main reasons. The first reason is the one stated above, since the existence of several malicious ASes would lead to a dissipation of the values of α and a misleading β. Imagine that we have neighbor A that has three malicious ASes with Risk Scores of 90, 90 and 70, and neighbor B that has only one malicious AS with Risk Score

85. Although AS A provides transit for much more malicious traffic, it would have a lower Risk Score than B. The second reason refers to malicious users trying to manipulate the Risk Score values. The rationale is the same, averages soften the differences of β and a malicious user could have one, or more, metrics with very low values and one with high value. Although the malicious traffic could be dangerous, the final result would be a low value due to the average calculation, and it could become even more dangerous if facing a collusion situation among different malicious ASes. This means that the Risk Score should not have an upper bound value and averages should be avoided.

To properly compare risks we should consider two main types of neighbors, peers and all the remaining. The rationale for these two types is due to the fact that the typical concept of peer is used for two ASes exchanging traffic between them but not providing transit for each other traffic. In this scenario it only makes sense to compare peers with peers since the traffic is only local and the AS topology, i.e., its connections to other ASes is irrelevant for the Risk Score calculation since that traffic will not reach the ISP AS, only traffic originating from that peer.

5.7 Simulation

For simulating the Risk Score behavior we are required to have two main inputs:

1. Malicious traffic information - we used the one analyzed in section 4.1;
2. BGP reachability information - since we do not have the public ASNs for the malicious traffic, the public data from RIS (see section 4.2 for details) is not useful. We therefore defined a scenario for a neighbor with three malicious ASes that may use it as transit: ASes 64557, 64562 and 64549. The neighbor itself is not originating malicious traffic. The reason for choosing these malicious ASes is because they are characterized by different types of malicious traffic, forming a good sample for the simulation.

Given this scenario, we could start with the first simulation. Table 5.4 provides the values required for expressions 5.1, 5.2 and 5.4.

AS	AS/Total Malicious Traffic (address scans)	AS/Total Malicious Traffic (flow floods)
64515	32	0
64562	0	13
64549	0	26

Table 5.4: Malicious traffic metrics (real data).

With these inputs and with the default values for the custom parameters, i.e. $karma = 1$, $weight = 3$, we get the following results for α and β for the first day:

- $\alpha_{64515} = 160,\ \alpha_{64562} = 30,\ \alpha_{64549} = 130$
- $\beta_1 = R_1 = 320$

In this case we are not simulating a time series and for that reason, according to expression 5.4 the Risk Score value is given by β, i.e., expression 5.2.

Now we wanted to understand how the Risk Score is affected by varying the karma parameter. We changed the value of karma for the three ASes in concordance, i.e., we varied the karma value from 1 to 5 simultaneously, obtaining five different values for the Risk Score. With this simulation we were not analyzing the Risk Score behavior within a time window, therefore we calculated the Risk Score for a single day without previous values, i.e., $i = 1$. Table 5.5 and figure 5.2 present the values for that simulation:

$karma_{64515}$	α_{64515}	$karma_{64562}$	α_{64562}	$karma_{64549}$	α_{64549}	R_1
1	160	1	30	1	130	320
2	80	2	15	2	65	160
3	53	3	10	3	43	107
4	40	4	8	4	33	80
5	32	5	6	5	26	64

Table 5.5: Karma impact on Risk Score.

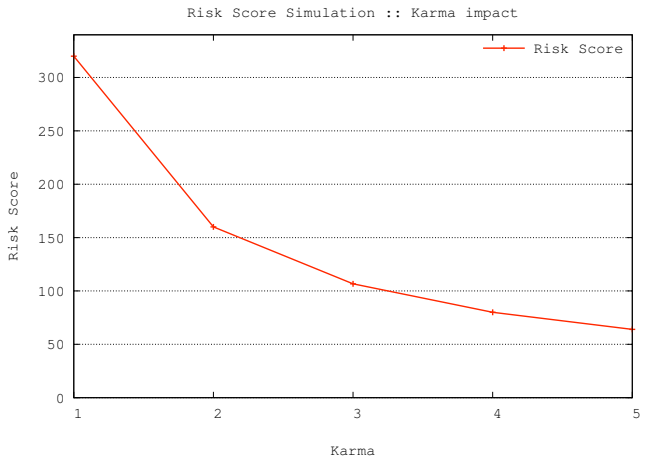

Figure 5.2: Karma impact on Risk Score.

From these results we can understand that the Risk Score decreases with the value of karma, which is the wanted result. As explained in section 5.2 the operator should use higher values of karma for ASes with which it has interest in its traffic. Now it becomes clearer why to use value 1 as the default value for karma, by default we do not trust ASes responsible for sending malicious traffic.

Concerning the weight parameters, since they enable tweaking of proportionality between different metrics, when we vary the values of weight we should only see a linear impact in the Risk Score. To prove this assumption, another simulation was performed with the same values from table 5.4 but varying the values of weight for the metric related to flow floods, and the results are presented in the following table:

$Flow\ floods\ Weight$	α_{64515}	α_{64562}	α_{64549}	R_1
1	32	1	5	38
2	32	2	10	45
3	32	4	16	51
4	32	5	21	58
5	32	6	26	64

Table 5.6: Weight impact on Risk Score.

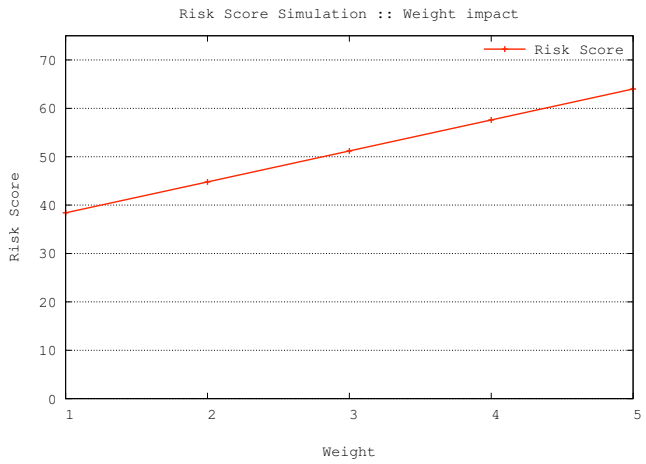

Figure 5.3: Weight impact on Risk Score.

As we can observe, the value of α_{64515} remains constant although the values of weight change. This simulation only changes the weight for a particular metric, flow floods, since AS 64515 only has values for address scans, changing the weight for flow floods does not impact on α_{64515}. From this simulation we also understand that the value of weight has a linear impact on the Risk Score, and we can influence the values of Risk Score depending on the importance we attribute to each metric, value 1 for lower importance and value 5 for higher importance. For this reason we believe the best default value should be 3. The reason for this choice is because by default the metric has a medium value of importance, allowing the operator to increase or decrease it as he/she thinks is appropriate.

Another question that required an answer, was the behavior of Risk Score for a period of time with variations in its inputs, the metrics. For this purpose we created a simulation for calculating the Risk Score of a neighbor, with the following scenario:

- Phase 1
  - Time Period: From day 1 to 60;
  - Description: three malicious ASes start sending malicious traffic towards the ISP AS, the values are incremental with an increase of 1% of malicious traffic per AS, per day.
- Phase 2

- Time Period: From day 61 to 75;
- Description: One of the malicious ASes stops sending malicious traffic, the other two continue sending malicious traffic.

- Phase 3
  - Time Period: From day 76 to 154;
  - Description: One of the malicious ASes starts decreasing the malicious traffic with a decrease of 1% of malicious traffic per AS, per day until it completely stops.

Until the end of the simulation only one AS continues to send malicious data. From this scenario we calculated the Risk Score and created a chart to compare how the Risk Score behaves in comparison to the β parameter, since β gives the instantaneous value.

Figure 5.4: Risk score simulation.

The purpose of this last simulation is to understand how the algorithm behaves with the dynamics of malicious ASes. As we can observe from the chart in figure 5.4 as the system changes its inputs, the Risk Score accompanies those changes and when the input flattens, i.e., remains constant, the algorithm starts converging, a behavior that deals well with the dynamics of the Internet. We can observe that a possible problem for this algorithm is the time it consumes for reacting to malicious traffic, since the β parameter was constant for approximately 290 days and the Risk Score is still approaching that value.

Chapter 6

Intervention Policies

Beyond the classification of network neighbors, we can integrate this concept with active security policies, namely the deployment of defense mechanisms to mitigate malicious traffic. These are two facets of the study provided by this thesis.

In this chapter we analyze different mechanisms to mitigate malicious traffic received at the ISP network, using the Risk Score (RS) previously calculated for a network neighbor and the ratio of prefix versus total AS malicious traffic.

A possibility would be to simply depeer or block all the traffic from the AS with highest Risk Score. Although a possible approach, this may be too harsh or even - in a worst-case scenario - counter-productive. A feasible scenario would be an AS originating high volumes of malicious traffic from one particular prefix. In this case the malicious traffic would be contained in one particular prefix and filtering just that prefix, while notifying the AS network operator could solve the problem.

As mentioned, it is the responsibility of the network operator to choose which class of intervention policy should be used, also, in order to provide coherence to the intervention policy the class must be global, i.e., the same policy is used for all network neighbors of the same type.

For these reasons following we will now present possible solutions for different scenarios.

6.1 Network Mitigation Techniques

6.1.1 Depeering

Choosing to peer with a certain AS is done under thorough scrutiny and depends on a number of different reasons, as explained in section 2.1. Although many of these reasons differ from AS to AS there is one that is common: only when mutually beneficial to the parties will it be implemented. Regarding the beneficial concept, it is also dependent on the specificities of each AS view (e.g., business revenue, technology choices). To determine when a certain agreement is beneficial can be extremely hard to quantify and the Risk Score intends to provide another argument to support that choice. We must understand when it stops being beneficial to peer with another AS, which can

occur when considerable amounts of malicious traffic are received from that particular peer, and we may be required to take the drastic approach of depeering from that AS.

When we choose to depeer from a particular AS, it does not mean that malicious traffic previously detected will suddenly stop arriving to the ISP network since there are other routes that it may use to reach the ISP network (that is one of the marvels of the Internet - its implicit redundancy). This being said, the Risk Score algorithm also classifies all the other neighbors, peers or not, and if the malicious traffic is still being detected, the same process will be run again.

6.1.2 Prefix Filtering

A possible and widely used solution to mitigate malicious traffic is prefix filtering based in deployment of ACLs (see section 2.3). Prefix filtering through ACLs allows dropping traffic from certain specific prefixes, however it is not a very flexible solution in particular in terms of scalability, since new prefixes responsible for originating malicious traffic appear with some regularity and the deployment of ACLs through all the ingress routers is extremely difficult. In practical terms this means the filters need to also include the new prefixes, which means considerable maintenance efforts are required, increasing the complexity of the filters, increasing the space for configuration errors and also increasing the processors load for that particular task - something that may introduce delays into the network.

Prefix filtering through ACLs, although a simple technique implies many disadvantages, as previously stated, for which reason should be avoided as a long term deployment solution.

6.1.3 Route Injection and Flow Spec

The use of route injection and Flow Spec takes advantage of a trusted signaling infrastructure already in place and specialized personnel already accustomed with the technology, BGP. The use of Flow Spec provides a defense mechanism of high granularity since it enables definition of protocols, source/destination addresses, source/destination ports, packet size, fragmentation, etc. Flow Spec distributes flow specifications through BGP and delegates on the network routers that implement BGP, the task of filtering the identified flows, therefore enabling a triggering mechanism fast and easy to deploy[1]. This technology uses extended communities and a new Network Layer Reachability Information (NLRI) address family. Some of the actions supported by Flow Spec are traffic discard, traffic rate limitation or traffic redirection into an MPLS tunnel, meaning that traffic engineering techniques can be applied to particular traffic flows.

This approach could have even more potential since it can be used for communication between network neighbors. If previously defined rules were agreed between the involved network neighbors, the Flow Spec rules could be propagated upstream till the source of the malicious traffic. A cooperative approach could therefore increase the power of the defense mechanisms and to deploy it using BGP is a simple and ubiquitous network mechanism already widely used in current Internet, and with which network operators are familiar.

[1] Vendor support is currently still an issue.

6.2 Making a Choice

From an operator point of view, making the choice of the most suitable defense mechanism is dependent on two types of values: the value of the Risk Score after which mitigation actions are in deployed; the values of prefixes versus AS total malicious traffic (see section 4.1.4). In case of a source of malicious traffic being confined to a particular prefix, for example AS64542 has 100% of its malicious traffic from prefix 33.152.0.0/14 (see figure 4.6 for details). In these cases it makes sense to affect only particular prefixes, since it will mitigate all the malicious traffic from that AS, otherwise, if depeering was applied, it would affect all the customers from AS64542 even if being perfectly legitimate users.

The best solution for this choice would be to define a threshold to the ratio of prefixes versus AS malicious traffic, after which we should choose to redirect particular source prefixes through Route Injection or Flow Spec mechanisms, i.e., most of the malicious traffic originates from a particular prefix. Using these mechanisms it would be possible to mitigate the malicious traffic in an efficient and automatic way. If however the source of malicious traffic originated in a particular AS is more spread through its address space than we should not use this type of mechanisms but instead consider it as an argument for legitimate depeering actions.

As we can observe from the figure 4.5 some of the sources of malicious traffic in terms of prefixes are below 20%. This implies the diffusion of malicious traffic through the address space explained above. For this reason it is not efficient to start deploying a set of mitigation rules for each of the prefixes and a wider approach is preferable, in particular depeering or blocking all traffic from malicious ASes.

When classifying network neighbors, the maximum acceptable value for RS is given by the maximum value of α for one malicious AS. Considering the expressions for α 5.1, β 5.2, and Risk Score 5.4 we obtain the following statement:

> STATEMENT 1 *Let Q be the total number of metrics and Karma have its default value, i.e., 1 we have* $Maximum\ Acceptable\ RS\ Value = \sum_{j=1}^{Q} 100 \times W_j$.

Since we can have network operators with specific concerns, we propose the use of three classes of intervention policies: strict, moderate and soft.

Strict A network operator with high security concerns and with low tolerance for malicious traffic.

Moderate A network operator with high security concerns that may tolerate some malicious traffic.

Soft A network operator with security concerns that tolerates malicious traffic.

For determining which intervention policy to apply, the network operator must define two important thresholds:

Risk Score Threshold (T_{RS}) Let T_{RS} be the maximum value of β for a given set of weight parameter values. T_{RS} is the upper bound value after which an intervention policy should be applied.

From statement 1 we know the maximum value RS may be, therefore this threshold value depends on how strict the network operator is regarding security concerns. The relationships between the different T_{RS} values for the three classes of policies is given by: T_{RS}(strict) < T_{RS} (moderate) < T_{RS}(soft).

Prefix Threshold (T_P) Let T_P be the maximum value of malicious traffic a prefix may originate, after which the prefix is considered malicious. T_P is the upper bound value of malicious traffic after which we should only filter that particular prefix. The relationships between the different T_P values for the three classes of policies is given by: T_P (strict) > T_P (moderate) > T_P (soft).

Now we are able to define the specific values for the three policy classes previously mentioned. The values present in table 6.1 were chosen based on the results of the data analysis from chapter 4.

Class	T_{RS}	T_P
Strict	30% of Max(α)	45%
Moderate	50% of Max(α)	30%
Soft	70% of Max(α)	15%

Table 6.1: Policy Class Threshold Values.

To decide which policy to apply, we can use these thresholds as references for the values collected by the detection platform and in particular RS. For this process we propose the workflow defined in figure 6.1. This workflow intends to go through every ISP neighbor, calculate the RS for each one, compare it with the threshold previously defined for RS (T_{RS}) and choose the appropriate policy to apply.

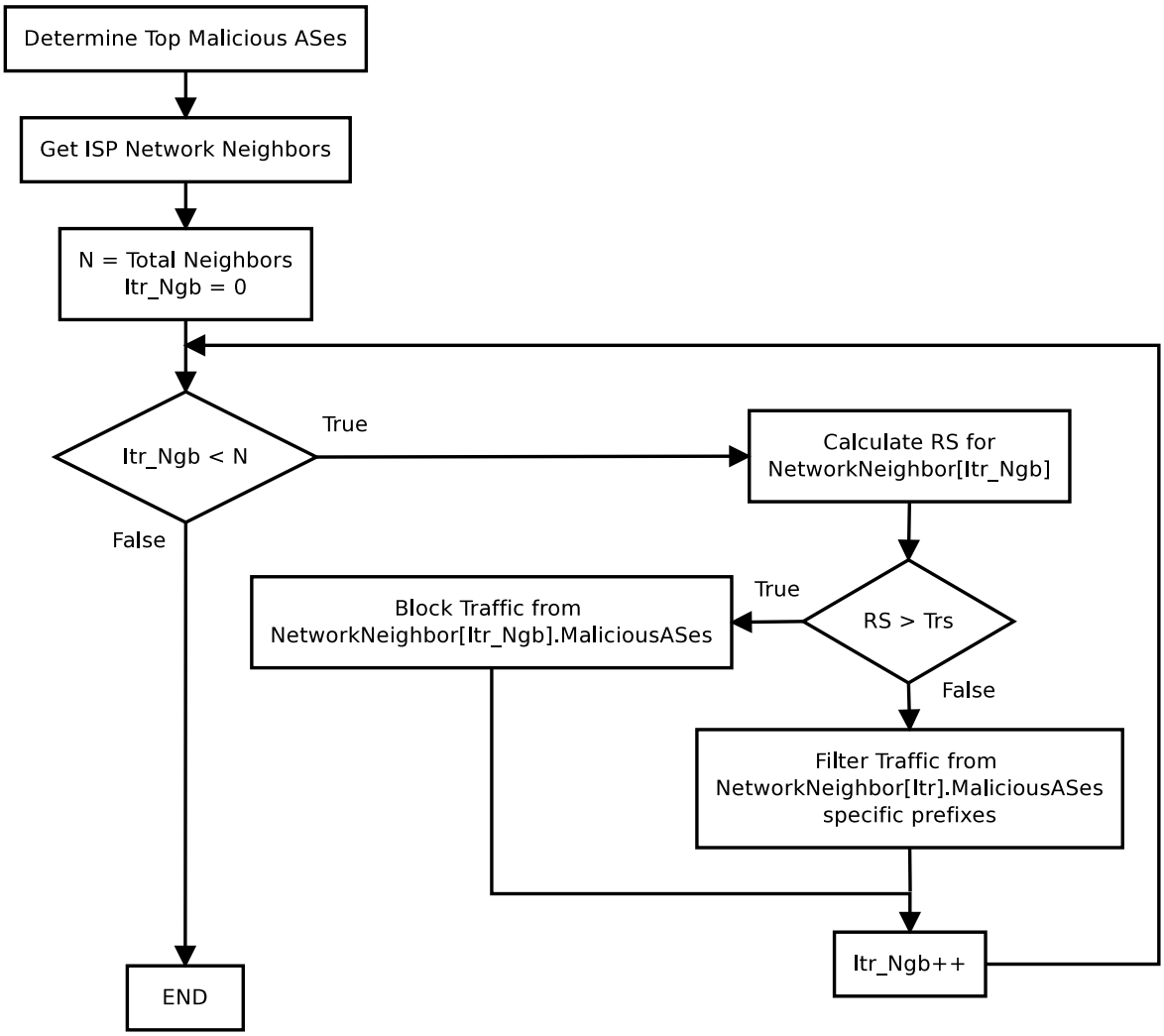

Figure 6.1: Risk Score Intervention Policy Selection.

For the processes of "Block Traffic From NetworkNeighbor[Itr_Ngb].MaliciousASes" and "Filter Traffic From NetworkNeighbor[Itr_Ngb].MaliciousASes specific prefixes" we present figures 6.2 and 6.3.

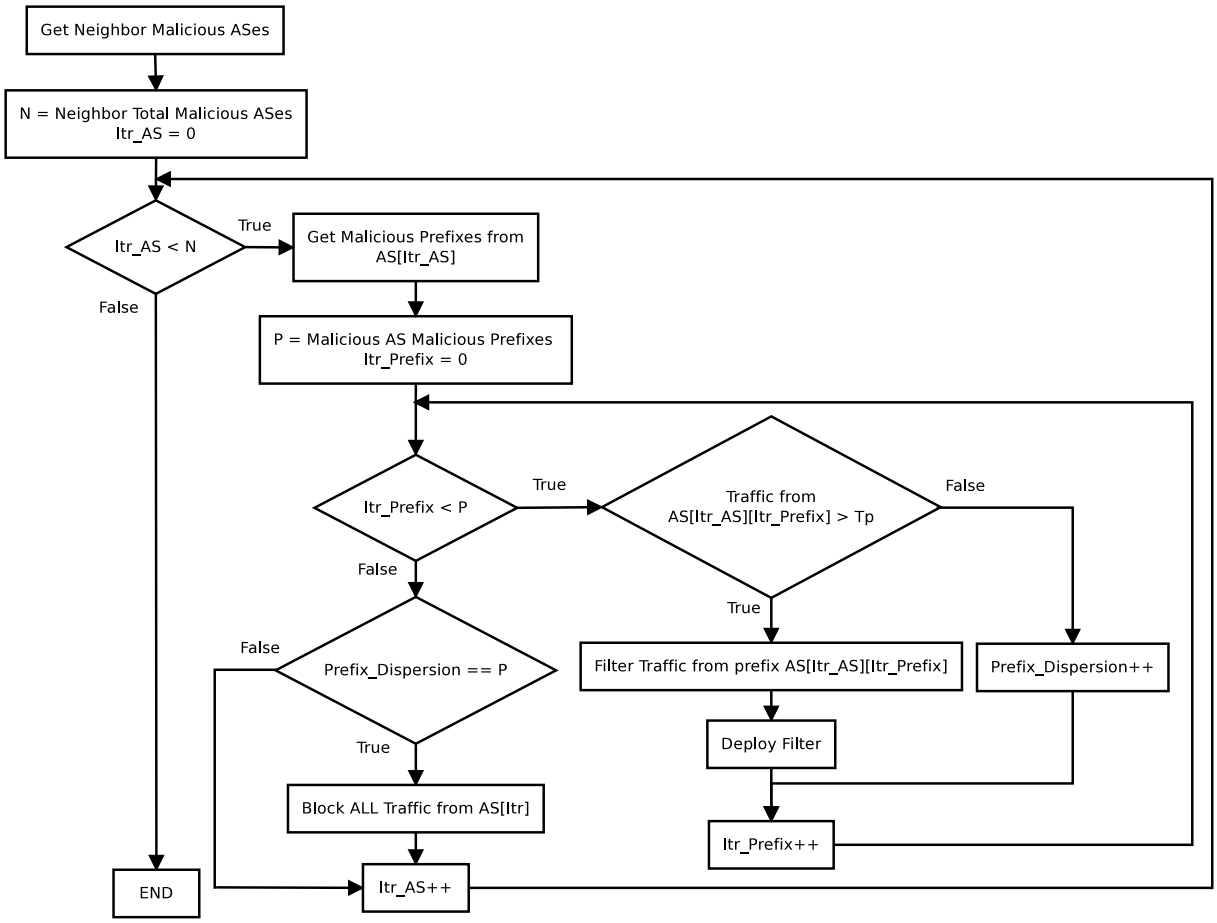

Figure 6.2: Risk Score Intervention Policy Selection - Block All Traffic.

Regarding figure 6.2, it refers to the use case where RS surpasses T_{RS}. In this use case, the network neighbor is considered to be a threat and special attention must be given. The workflow proposes to go through every malicious AS associated with this neighbor; for each of the malicious ASes we iterate through their prefixes responsible for generating malicious traffic. In case the malicious traffic from these prefixes surpasses the threshold defined for prefix malicious traffic ratio (T_P), it means that particular prefix is a high contributor for the total malicious traffic from that AS, in which case we should only filter the prefix. In case the traffic for that prefix is below T_P , we only increment a counter. This counter quantifies the degree of malicious traffic dispersion through the address space for that malicious AS. From that point onward, the process is similar for all prefixes. As can be observed, till this moment we do not block all traffic from an AS, that will only happen in case we do not have prefixes that surpass T_P , which means the malicious traffic is disperse through the malicious AS and therefore it is more efficient to block all the AS traffic. For the particular case of the neighbor being a peer, the blocking of traffic is done in parallel with a depeering process. The depeering process should be viewed not only as a technical issue but also as a more complex one as explained in section 6.1.1.

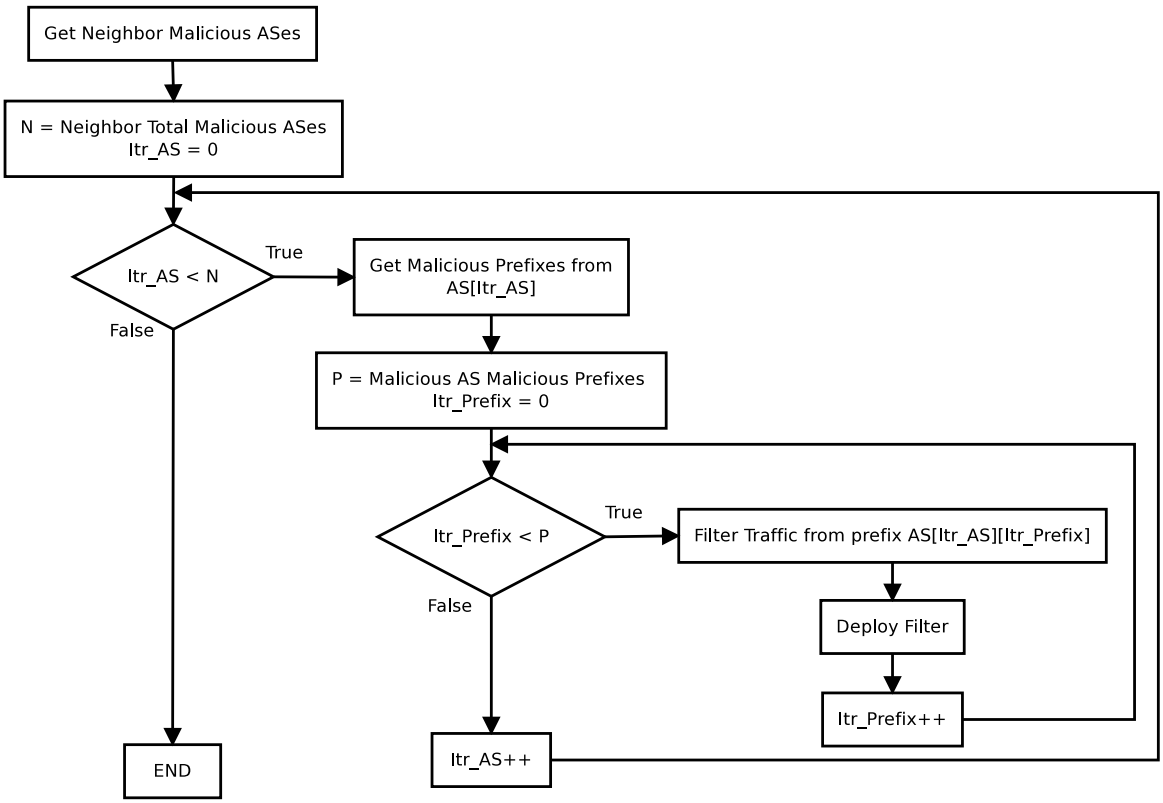

Figure 6.3: Risk Score Intervention Policy Selection - Block Specific Traffic.

The workflow present in figure 6.3 has a similar approach but we do not consider the possibility of blocking all traffic from the malicious ASes. Since this workflow is executed in case RS is below the threshold T_{RS}, i.e., the network neighbor is not considered to be a high security risk, we are not as harsh in the policy applied and only consider filtering the malicious prefixes. In this use case we go through all the malicious ASes associated with the network neighbor; for each of the malicious ASes we iterate through their prefixes generating malicious traffic, and in case they surpass the threshold T_P we filter the prefix traffic, otherwise we let the traffic flow.

6.3 Deployment

In case the depeering solution is chosen, the technical details are simple since the main requirement is to shutdown the BGP peer. After that point onward the reachability information exchanged with that particular peer stops being exchanged.

For specific prefixes, the network defense mechanisms are different and the deployment of Flow Spec is considered the best choice. It is however more complex to deploy than depeering and currently does not have complete support of network equipment vendors. Concerning the routers support, the specifications are very recent and although some vendors already support them (e.g., Juniper), others (e.g., Cisco) do not provide available support at this moment.

Concerning Flow Spec rules injection, a viable possibility is a tool named exabgp [4], which besides Flow Spec can also inject generic routes into BGP communications and supports IPv4 and IPv6. The tool implements the following RFCs:

- RFC 1997 [17] - BGP Communities Attribute;
- RFC 2545 - Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing;
- RFC 4271 [41] - A Border Gateway Protocol 4 (BGP-4);
- RFC 4360 - BGP Extended Communities Attribute;
- RFC 4724 - Graceful Restart Mechanism for BGP;
- RFC 4760 - Multiprotocol Extensions for BGP-4;
- RFC 4893 [49] - BGP Support for Four-octet AS Number Space;
- RFC 5492 - Capabilities Advertisement with BGP-4;
- RFC 5575 [32] - Dissemination of Flow Specification Rules.

With a deployment scenario of Flow Spec we need to deploy a control peer. It can be a Network Operations Center (NOC) server, internally peering with other BGP speaking neighbors, which would be responsible for injecting the Flow Spec rules whenever considered necessary, e.g., a prefix originating a considerable amount of malicious traffic.

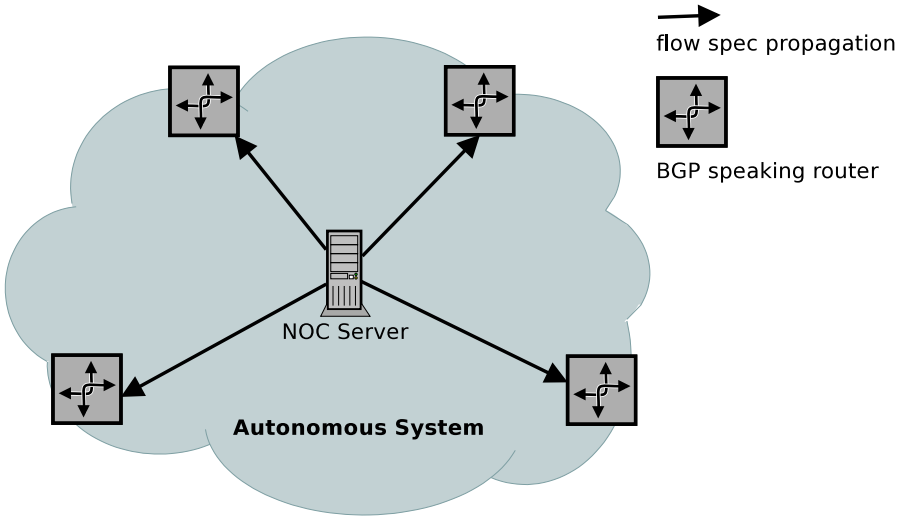

Figure 6.4: Flow Spec network diagram.

Considering the alternatives, we believe the best choice would be to use BGP as a signaling mechanism and exabgp as the tool of choice for implementing Flow Spec. This mechanism provides the necessary flexibility to filter specific prefixes; to filter/block traffic from a complete AS; to quickly

communicate decisions to network equipments and, also of high relevance, enables a communication mechanism to network neighbors in case of a pre-agreement between network operators to implement joint security countermeasures.

To present a possible deployment scenario for implementing the security policies described in this chapter, we will consider the same values used for the Karma simulation in section 5.7, in particular for the default value of Karma (see table 5.5), which assume default weights for the two metrics used. In the simulation we considered as malicious ASes the set $\{64515, 64562, 64549\}$. Regarding the main contributing prefixes for each of the malicious ASes from this set (see section 4.1.4), we get the following distribution:

	Malicious ASes		
	64515	64562	64549
Prefixes [Malicious Traffic]	133.191.255.0/13 [21%] 249.196.0.0/14 [6.2%] Other [72.8%]	93.156.0.0/14 [100%] Other [0%]	95.232.0.0/16 [99.3%] Other [0.7%]

Table 6.2: Malicious ASes main prefixes malicious traffic distribution.

In the simulation scenario used, the calculated value for RS was 320. Considering this scenario, we have two metrics, the default values for weights and Karma, three and one respectively, and Statement 1 described in section 6.2, we can calculate the maximum value of RS, which is 600. Given this value, and considering table 6.1, we obtain the values of T_{RS} for the different policy classes presented in table 6.3.

Policy Class	T_{RS} (%)
Strict	180
Moderate	300
Soft	420

Table 6.3: Risk Score Threshold (T_{RS}) Calculation.

At this point we can calculate the values for the total malicious traffic after applying the different policy classes, through the use of the workflows previously explained (figures 6.1, 6.2 and 6.3). Depending on the value of RS for the neighbor and the policy class selected, different intervention policies may be applied. We now explain how this is achieved for each of the ASes:

64515

- *strict*: since RS is 320 and T_{RS} is 180, we choose the use case in which we may block all traffic from AS 64515. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (45%). We increment a counter responsible for accounting the dispersion of malicious prefixes (Prefix_Dispersion), i.e., only if the malicious traffic from all the main prefixes is below the T_P threshold will we block all traffic from an AS. Since none of the prefixes contributes with more malicious traffic than T_P , the policy blocks all traffic from AS 64515.

- *moderate*: since RS is 320 and T_{RS} is 300, we choose the use case in which we may block all traffic from AS 64515. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (30%). We increment a counter responsible for accounting the dispersion of malicious prefixes (Prefix_Dispersion), i.e., only if all the main prefixes malicious traffic is below the T_P threshold will we block all traffic from an AS. Since none of the prefixes contribute with more malicious traffic than this threshold, the policy blocks all traffic from AS 64515.
- *soft*: since RS is 320 and T_{RS} is 420, we choose the use case of possibly filter the traffic from specific prefixes of AS 64515. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (15%). Prefix 133.191.255.0/13 contributes with 21% of malicious traffic and is therefore filtered. None of the remaining prefixes are filtered, for which reason we still have 79% of the initial malicious traffic.

64562

- *strict*: since RS is 320 and T_{RS} is 180, we choose the use case in which we may block all traffic from AS 64562. Next, we go through all prefixes which were mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (45%). Prefix 93.156.0.0/14 contributes with 100% of malicious traffic and is therefore filtered. Since there is no other prefix generating malicious traffic, all the malicious traffic from 64562 is mitigated without being necessary to block the whole AS.
- *moderate*: since RS is 320 and T_{RS} is 300, we choose the use case in which we may block all traffic from AS 64562. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (30%). Prefix 93.156.0.0/14 contributes with 100% of malicious traffic and is therefore filtered. Since there is no other prefix generating malicious traffic, all the malicious traffic from 64562 is mitigated without being necessary to block the whole AS.
- *soft*: since RS is 320 and T_{RS} is 420, we choose the use case of possibly filter the traffic from specific prefixes of AS 64562. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (15%). Prefix 93.156.0.0/14 contributes with 100% of malicious traffic and is therefore filtered. Since there is no other prefix generating malicious traffic, all the malicious traffic from 64562 is mitigated without being necessary to block the whole AS.

64549

- *strict*: since RS is 320 and T_{RS} is 180, we choose the use case in which we may block all traffic from AS 64549. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (45%). Prefix 95.232.0.0/16 contributes with 99.3% of malicious traffic and is therefore filtered. None of the remaining prefixes are filtered, for which reason we still have 0.7% of the initial malicious traffic. We increment a counter responsible for accounting the

dispersion of malicious prefixes (Prefix_Dispersion), i.e., only if all the main prefixes malicious traffic is below the T_P threshold will we block all traffic from an AS. This means that blocking a malicious AS is a countermeasure only used as a last resort and in this particular case it is not considered to be required, for which reason we still have 0.7% of the initial malicious traffic.

- *moderate*: since RS is 320 and T_{RS} is 300, we choose the use case in which we may block all traffic from AS 64549. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (30%). Prefix 95.232.0.0/16 contributes with 99.3% of malicious traffic and is therefore filtered. None of the remaining prefixes are filtered, for which reason we still have 0.7% of the initial malicious traffic.
- *soft*: since RS is 320 and T_{RS} is 420, we choose the use case of possibly filter the traffic from specific prefixes of AS 64549. Next, we go through all prefixes mainly responsible for generating malicious traffic and check if they surpass the threshold value T_P for this policy class (15%). Prefix 95.232.0.0/16 contributes with 99.3% of malicious traffic and is therefore filtered. None of the remaining prefixes are filtered, for which reason we still have 0.7% of the initial malicious traffic.

Table 6.4 summarizes these calculations.

Malicious AS	Policy Class	Main Prefixes (%)	Other Prefixes (%)	Total Malicious Traffic (%)
64515	Strict	0	0	0
	Moderate	0	0	0
	Soft	6.2	72.8	79
64562	Strict	0	0	0
	Moderate	0	0	0
	Soft	0	0	0
64549	Strict	0	0.7	0.7
	Moderate	0	0.7	0.7
	Soft	0	0.7	0.7

Table 6.4: Intervention Policy Effect On Malicious Traffic.

The question that arises at this moment is "*What if the malicious traffic ingress route changes to a different network neighbor?*". Since the intervention policy is coherently propagated through BGP across the ISP network core, the malicious traffic would still be filtered.

It should be mentioned that the mechanism here proposed intends to provide a scalable method to deal with malicious traffic. The selection mechanism based on the Risk Score and the volume of malicious traffic from specific prefixes is used to present to the network operator the most efficient defense option: filter specific prefixes, block all traffic from an AS or let traffic flow.

Chapter 7

Conclusions

In the current chapter we propose possible new directions that can be taken based on the work presented in this thesis, and explain the main conclusions reached with it.

7.1 Future Work

The scope of this thesis is very wide, which creates several opportunities to enhance it, from different perspectives.

The definition of Risk Score only uses ratios. One possible outcome of such choice is that although the score is high, the absolute input values are low. This can be overcome by two parallel approaches. First, besides only the ratio, it should also consider a threshold value after which we consider the traffic interesting. Second, using another metric that also accounts for the values of total traffic received per origin AS. For example, spam is a type of malicious traffic easier to quantify when compared to the total exchanged traffic. The goal of this metric would be to better understand the weight each origin AS has in the total traffic received by the ISP versus the total malicious traffic. As explained in this thesis, depending on the type of malicious traffic, the volume of traffic, i.e., the bandwidth, may not be the most relevant metric if compared to the ISP traffic as a whole. Nevertheless it is important to understand the importance an AS has to an ISP and that metric may be a possible answer.

The metrics analyzed in this work only consider specific traffic between end points. A more broad view of the network is important, for which reason integrating concepts of BGP malicious traffic into the expression could enhance the classification quality of a neighbor, e.g., according to Ramachandran and Feamster [40] the existence of short lived BGP routes may indicate an existence of a malicious AS.

Besides the BGP update messages analyzed in this thesis, other public services in the Internet could also provide similar information, e.g., looking glasses, increasing the BGP input data and therefore enabling a more accurate characterization of BGP reachability information, from the Internet towards the ISP. However, another input that was not regarded, were BGP update messages

exchanged in the core network of the ISP. Currently available software already exists for this purpose [1], allowing an insightful view of the operator towards the Internet, namely what routes are available for certain interesting destinations from a security point of view.

In this thesis we exclude the scenario of malicious routers manipulating the BGP update messages. One possible attack to the system may be implemented by manipulating the BGP update messages. A malicious router may inject bogus BGP update messages with manipulated AS paths in order to remove malicious ASes from the adjacency table of a certain neighbor. This attack could, in a worst case scenario, lead a network provider to block traffic from all its neighbors. To deal with this type of issues some research work has been done, enabling detection of these malicious actions [39] that can also be applied to the work here presented.

Given the concepts proposed in this thesis, the design and implementation of an application that integrates them is still to be done. The possibility of having a tool with such features and tested in a real network environment, is in our opinion a good enhancement to a network operator security toolkit.

7.2 Conclusions

The work presented in this thesis enables a network operator to classify a network neighbor in terms of the security risk it poses to the ISP and provides possible actions to deal with it.

Real data was collected from a mid scale ISP network that allowed to study the behavior of particular types of malicious traffic, i.e., address scans and flow floods, sent towards the ISP network. From this analysis we understood that although the origins of such traffic are many, most of the traffic is sent from only a small subset. We could also observe that some of the malicious traffic activity appeared to have a periodic behavior.

To better understand how traffic could flow towards a network, we used the BGP update messages available from the RIPE's RIS repository. This service is based on a network of probes spread through different locations worldwide with several peering connections, in order to capture BGP update messages and therefore allowing a mapping of Internet reachability information dynamics. In particular, we focused on how a particular AS was seen from the Internet, i.e., how it could be reached from other networks in the Internet.

To enable an ISP to classify a network neighbor we developed an algorithm to determine the security risk it poses to the ISP and we named it Risk Score. For achieving this goal, we used the results from the data analysis previously performed on the malicious traffic, and defined metrics to be used as inputs in the algorithm. We propose the algorithm to be used continuously by an ISP in order to understand the entrance points of malicious traffic into its network. Therefore enabling not only a detection and recovery strategy but also a deterrence one, since it is intended for ASes that provide transit facilities for malicious traffic be lead to react upstream towards the source of the malicious traffic. For understanding how the algorithm would behave under the dynamics of malicious traffic, we simulated the input from three different ASes and from the results we could observe the algorithm converging.

Based on the Risk Score algorithm and on particular observations done when analyzing the malicious data, e.g., some ASes originate all their malicious traffic from a single prefix; we propose a set of intervention policies that are sensitive to network behaviors and can be applied according to a predefined flow. To understand the effect of real data with those intervention policies, we used the same data from the Data Analysis in chapter 4, and we were able to observe malicious traffic being considerably reduced in some cases and in others completely mitigated.

Bibliography

[1] "Bird project. the BIRD Internet Routing Daemon." http://bird.network.cz/. 7.1

[2] "CoralReef," http://www.caida.org/tools/measurement/coralreef/. 4.3.4

[3] "Cryptography-based prefix-preserving anonymization," http://www.cc.gatech.edu/computing/Telecomm/projects/cryptopan/. 4.3.4

[4] "Exabgp: A BGP route injector," http://code.google.com/p/exabgp/. 6.3

[5] "IP to ASN mapping," http://www.team-cymru.org/Services/ip-to-asn.html. 4.3.2

[6] "IPv4/IPv6 manipulation library in Python," http://code.google.com/p/ipaddr-py/. 4.3.3

[7] "NFDUMP tools," http://nfdump.sourceforge.net/. 4.3.4

[8] "Routeviews routing table archive," http://www.routeviews.org/. 3

[9] "Routing information service (RIS)," http://www.ripe.net/projects/ris/. 3, 4, 4.2.2

[10] "Youtube hijacking: A RIPE NCC RIS case study," http://www.ripe.net/news/study-youtube-hijacking.html, 2008. 3

[11] H. Ballani, P. Francis, and X. Zhang, "A study of prefix hijacking and interception in the internet," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, p. 276, 2007. 2.1

[12] S. Bellovin, D. Clark, A. Perrig, and D. Song, "A clean-slate design for the next-generation secure internet," in *Technical report, Pittsburgh, PA: Report for NSF Global Environment for Network Innovations (GENI) Workshop*, 2005. 3

[13] S. Bellovin, "Security problems in the TCP/IP protocol suite," *ACM SIGCOMM Computer Communication Review*, vol. 19, no. 2, pp. 32–48, 1989. 3

[14] L. Blunk, M. Karir, and C. Labovitz, "MRT routing information export format," *draft-ietf-grow-mrt-08. txt (Internet Draft)*, 2008. 4.3.1

[15] W. Borremans and R. Valke, "BGP (D) DoS Diversion," 2005. 3

[16] K. Butler, T. Farley, P. Mcdaniel, and J. Rexford, "A survey of BGP security issues and solutions," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 100–122, 2010. 3

[17] R. Chandra, P. Traina, and T. Li, "RFC 1997: BGP communities attribute," 1996. 2.1, 6.3

[18] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Towards capturing representative AS-level internet topologies," *Computer Networks*, vol. 44, no. 6, pp. 737–755, 2004. 2.1, 3, 4.2.1, 5.1

[19] Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of internet malicious sources," in *Infocomm Mini-Conference*. Citeseer, 2008. 3

[20] S. Cheshire, B. Aboba, and E. Guttman, "RFC 3927 (proposed standard): Dynamic configuration of IPv4 link-local addresses," *IETF*, 2005. 4.1.4

[21] Y. Chi, R. Oliveira, and L. Zhang, "Cyclops: The AS-level connectivity observatory," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 5, pp. 5–16, 2008. 3

[22] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun *et al.*, "AS relationships: inference and validation," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 29–40, 2007. 2.1, 3, 5.1

[23] C. Dovrolis, "What would Darwin think about clean-slate architectures?" *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 1, pp. 29–34, 2008. 3

[24] N. Feamster, J. Jung, and H. Balakrishnan, "An empirical study of bogon route advertisements," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 1, p. 70, 2005. 3

[25] M. Fossi, E. Johnson, T. Mack, D. Turner, J. Blackbird, T. Adams, D. McKinney, S. Entwisle, B. Graveland, J. Mulcahy, and C. Wueest, "Symantec global internet security threat report: Trends for 2009," *Volume XV, Published April*, 2010. 2.2, 3

[26] M. Foukarakis, D. Antoniades, S. Antonatos, and E. Markatos, "Flexible and high-performance anonymization of NetFlow records using anontool," in *Third International Conference on Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007*, 2007, pp. 33–38. 4.3.4

[27] V. Fuller and T. Li, "RFC 4632: Classless inter-domain routing (CIDR): The internet address assignment and aggregation plan," *IETF*, 2006. 2.1

[28] J. Hawkinson and T. Bates, "RFC 1930: Guidelines for creation, selection, and registration of an autonomous system (AS)," *IETF*, 1996. 2.1, 4.1.1

[29] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in *Proc. USENIX Security Symposium*, 2006. 2.1, 3

[30] Y. Liao and K. Zhang, "BGP behavior monitoring and analysis," *ECS 289M (Advanced Topics in Computer Security) project*, 2002. 3

[31] R. Mahajan, S. Bellovin, S. Floyd, J. Ioannidis, V. Paxson, and S. Shenker, "Controlling high bandwidth aggregates in the network," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 3, pp. 62–73, 2002. 3

[32] P. Marques, N. Sheth, R. Raszuk, B. Greene, J. Mauch, and D. McPherson, "RFC 5575: Dissemination of flow specification rules," August 2009. 2.3, 6.3

[33] D. McPherson, R. Dobbins, M. Hollyman, C. Labovitzh, and J. Nazario, “Worldwide infrastructure security report, volume v, arbor networks,” 2010. 2.1, 2.1, 2.2, 3

[34] O. Nordstrom and C. Dovrolis, “Beware of BGP attacks,” *Computer Communications Review*, vol. 34, no. 2, pp. 1–8, 2004. 3

[35] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, “In search of the elusive ground truth: the Internet’s AS-level connectivity structure,” in *Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*. ACM, 2008, pp. 217–228. 3

[36] T. Peng, C. Leckie, and K. Ramamohanarao, “Survey of network-based defense mechanisms countering the dos and ddos problems,” *ACM Computing Surveys (CSUR)*, vol. 39, no. 1, p. 3, 2007. 2.2

[37] J. Postel, “RFC 791: Internet protocol: DARPA internet program protocol specification,” *Information Sciences Institute*, 1981. 2.1

[38] B. Prakash, N. Valler, D. Andersen, M. Faloutsos, and C. Faloutsos, “BGP-lens: Patterns and anomalies in internet routing updates,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 1315–1324. 3

[39] T. Qiu, J. Wang, L. Ji, D. Pei, and H. Ballani, “Locating prefix hijackers using LOCK,” 2010. 2.1, 3, 7.1

[40] A. Ramachandran and N. Feamster, “Understanding the network-level behavior of spammers,” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, p. 302, 2006. 3, 7.1

[41] Y. Rekhter, T. Li, and S. Hares, “RFC 4271: A border gateway protocol 4 (BGP-4),” *IETF*, 2006. 2.1, 2.1, 6.3

[42] Y. Rekhter, B. Moskowitz, D. Karrenberg, G. J. d. Groot, and E. Lear, “RFC 1918: Address allocation for private internets,” *IETF*, 1996. 2.3, 4.1.4

[43] M. Rossi, “MRT dump file manipulation toolkit (MDFMT)-version 0.2,” *Centre for Advanced Internet Architectures (CAIA)-Swinburne University of Technology, Tech. Rep., July*, 2009. 4.3.1

[44] A. Slagell, K. Lakkaraju, and K. Luo, “Flaim: A multi-level anonymization framework for computer and network logs,” in *Proceedings of the 20th USENIX Large Installation System Administration Conference*, 2006, pp. 63–77. 4.3.4

[45] A. Slagell, J. Wang, and W. Yurcik, “Network log anonymization: Application of crypto-pan to cisco netflows,” in *Proceedings of the Workshop on Secure Knowledge Management 2004*. Citeseer, 2004. 4.3.4

[46] A. Slagell and W. Yurcik, “Sharing computer network logs for security and privacy: A motivation for new methodologies of anonymization.” 4.3.4

[47] A. Studer and A. Perrig, “The coremelt attack,” *Computer Security–ESORICS 2009*, pp. 37–52, 2010. 4.1.2

[48] D. Turk, “RFC 3882: Configuring BGP to block denial-of-service attacks,” *IETF*, 2004. 2.2, 2.3, 2.3

[49] Q. Vohra and E. Chen, “RFC 4893 (proposed standard): BGP support for four-octet AS number space,” *IETF*. 2.1, 6.3

[50] A. Yaar, A. Perrig, and D. Song, “SIFF: A stateless internet flow filter to mitigate ddos flooding attacks,” 2004. 4.1.2

[51] H. Yan, R. Oliveira, K. Burnett, D. Matthews, L. Zhang, and D. Massey, “BGPmon: A real-time, scalable, extensible monitoring system,” in *Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH)*. Citeseer, 2009. 3

[52] Q. Zhang and X. Li, “An IP address anonymization scheme with multiple access levels,” *Information Networking. Advances in Data Communications and Wireless Networks*, pp. 793–802, 2006. 4.3.4