Computers & Security (2004) 23, 665–678

ELSEVIER

Computers & Security

www.elsevier.com/locate/cose

Characterization of defense mechanisms against distributed denial of service attacks

Li-Chiou Chen[a,b,c,*], Thomas A. Longstaff[c,d], Kathleen M. Carley[a,b,c]

[a]Department of Engineering and Public Policy, Pittsburgh, PA 15213, USA
[b]Institute for Software Research International, Pittsburgh, PA 15213, USA
[c]Carnegie Mellon University, Pittsburgh, PA 15213, USA
[d]Network Survivable Systems, Software Engineering Institute, Pittsburgh, PA 15213, USA

Received 3 October 2003; revised 25 February 2004; accepted 9 June 2004

KEYWORDS
Distributed denial of service attacks;
Characterization;
Defense mechanisms;
Computer network security;
Computer security

Abstract We propose a characterization of distributed denial of service (DDOS) defenses where reaction points are network-based and attack responses are active. The purpose is to provide a framework for comparing the performance and deployment of DDOS defenses. We identify the characteristics in attack detection algorithms and attack responses by reviewing defenses that have appeared in the literature. We expect that this characterization will provide practitioners and academia insights into deploying DDOS defense as network services.

Introduction

Distributed denial of service (DDOS) attacks have emerged as a prevalent way to compromise the availability of networks or servers. Since these attacks have interrupted legitimate access to the networks or servers providing online services, they have imposed financial losses on e-commerce businesses (CERT/CC, 1999; Tran, 2000; Yankee, 2000). To mitigate the impacts of DDOS attacks, it is important to develop defenses that can both detect and react against ongoing attacks. Although many DDOS defenses have been proposed, few of these proposals have been widely deployed at this point. The first step toward the wide deployment of DDOS defenses is to understand the performance tradeoffs and deployment costs of these defenses.

We review and categorize qualitatively current DDOS defense mechanisms that have appeared in the literature. The characterization is based on the attack detection algorithms and attack responses in a defense because the performance tradeoffs and deployment costs of a defense are dependent on them. An attack detection algorithm refers to the procedures which a defense uses to identify attacks based on available network

* Corresponding author. Carnegie Mellon University, Institute for Software Research, 231 Smith Hall, Pittsburgh, PA 15213, USA. Tel.: +1 412 2687527.
E-mail addresses: lichiou@andrew.cmu.edu (L.-C. Chen), tal@sei.cmu.edu (T.A. Longstaff), kathleen.carley@cmu.edu (K.M. Carley).

0167-4048/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2004.06.008

information. An attack response refers to the mitigation strategies that a defense triggers once an attack is identified.

Our purpose is to provide insights to network operators and their managers so that they will know which defenses should be taken under what circumstances. The categories and characteristics listed in the paper will assist Internet Service Providers (ISPs) in considering the provision of DDOS defenses as network services to their subscribers, such as e-commerce companies.

This paper is organized as follows. The next section explains the scope and method of the characterization. Defenses in terms of attack detection algorithms are categorized under section ''Attack detection algorithms''. Defenses in terms of attack responses are categorized under section ''Attack responses''. Conclusions and discussions follow.

Scope and method of the characterization

Both firewall technology (Cheswick and Bellovin, 1994; Zwicky et al., 2000) and intrusion detection systems (Axelsson, 2000; Debar et al., 1999; Mukherjee et al., 1994) have been developed to detect and to respond against various kinds of Internet-based attacks. However, defenses which are specifically designed to respond against large-scale DDOS attacks (CERT/CC, 1999) have not drawn much attention until recent years. In particular, the DDOS attacks in February 2000 against multiple e-commerce web sites (Tran, 2000; Yankee, 2000) highlight the potential risk and the severe impacts of DDOS attacks.

Current literature on the characterization of DDOS defenses is very limited, and each of the current works serves a different purpose than this paper. Most of the available DDOS literature review existing defenses. Among these, Savage et al. (2001) describes the pros and cons of various defenses most extensively, but their purpose is to compare these defenses with a proposed IP traceback method. The most similar work to this paper is the taxonomy of DDOS defense mechanisms (Mirkovic et al., 2002a). This taxonomy reviews current DDOS defense mechanisms and focuses on finding new features in the DDOS attack problems that have not been solved.

We are interested in only the DDOS defenses that can be provided by ISPs as optional network services to their subscribers. Similar security services, such as Virtual Private Networks or firewalls, have been provided to deal with the secrecy of data transportation. In our case, the services that provide DDOS defenses ensure the availability of an online service or a network. Specifically, we characterize only the defenses that have the following two properties:

1) *Reaction points which are network-based*: Reaction points to attacks could be network-based or host-based. We focus on network-based methods. Network-based methods are deployed on the points where packets route through network connections, such as routers or proxy servers. Host-based defenses are deployed on the machines that are potential targets of attacks. Although host-based methods (Spatscheck and Peterson, 1998; Yan et al., 2000) could increase the victims' capability to stay available during attacks, they do not filter out attack traffic before it reaches victims. For bandwidth saturation attacks, these methods need to be used together with network-based methods.
2) *Attack responses which are active*: Attack responses could either react against attack traffic actively or log/trace attack traffic passively. We focus on the former. Tracing back to the real sources of attacks has been an established part of DDOS defense studies (Bellovin, 2000; Burch and Cheswick, 2000; Park and Lee, 2001a; Savage et al., 2001; Snoeren et al., 2001; Snoeren et al., 2002; Song and Perrig, 2001), which have been analyzed previously (Lipson, 2002). These defenses could facilitate future liability assignments but cannot mitigate the impacts of ongoing attack traffic without an associated attack response.

To assist the provision of DDOS defenses as network services, we propose a characterization focusing on identifying factors in defenses that influence their performance and deployment. Two specific questions guide our characterization:

- *Performance*: DDOS defenses are valuable for maintaining a certain availability of victims' network services to their legitimate clients during attacks. What are the characteristics that influence the availability of victims' network/online services during attacks when DDOS defenses are deployed?
- *Deployment*: To provide DDOS defenses to their subscribers, network providers need to consider how and where to deploy DDOS defenses on their current infrastructure. What are the characteristics that influence the

deployment of DDOS defenses as a network service?

To address these two questions, we characterize the network-based active defenses in terms of attack detection algorithms and attack responses. Attack detection algorithms identify attack traffic from the network traffic monitored and attack responses act against attack traffic by triggering filtering. The characterization is based only on information from the current literature that documents both aspects in enough detail. We exclude some commercial products (Arbor, 2002; Asta, 2002; Recourse, 2002) that do not provide enough public available information to create the characterization.

Attack detection algorithms

To address the performance question, we investigate how an attack detection algorithm results in false positives (legitimate traffic is regarded as attack traffic) and false negatives (attack traffic is regarded as legitimate traffic). False positives occur when the specific patterns, network statistics or network information of legitimate traffic match those of attack traffic. Similarly, false negatives occur when the specific patterns, network statistics or network information of attack traffic do not match those of attack traffic. We derived the following characteristics to identify false positives and false negatives:

- *Granularity of detection*: the basic unit of network traffic examined. Network traffic can be examined in terms of packets, flows, network-level connections or application-level connections. The impact of false positives/false negatives is applied on this basic unit.
- *Network information needed to monitor attacks*: statistics regarding the network traffic monitored or contents of the monitored traffic. The information may include network packet headers, packet rates of network flows/connections, or information on dropped packets.
- *Specific characteristics of attack traffic*: the specific values in network information used to identify attacks. False positives occur when legitimate traffic has these specific characteristics. False negatives occur when attack traffic does not have these characteristics.
- *Sources of false positives/false negative*: the reasons that the legitimate/attack traffic has the same characteristics as the attack/legitimate traffic.

Table 1 summarizes the above characteristics of the surveyed defenses along with their limitations. We also list each defense in either one of the three categories: congestion-based, anomaly-based and source-based in order to distinguish their sources of false positives/false negatives. Congestion-based defenses detect attack traffic when the network links are congested to a certain level, anomaly-based defenses detect attack traffic when anomalous patterns are discovered in the network traffic and source-based defenses detect attack traffic when the sources of the network traffic are not valid or blacklisted.

Congestion-based

Once the monitored network links are congested, the attack detection algorithm identifies the type of network flows/connections that contribute to the congestion. These methods identify attack traffic effectively only when attack traffic induces congestion of the monitored links, and the congestion can be observed. False positives occur when the attack detection algorithm cannot single out the legitimate traffic that contributes to congestion and false negatives occur when attack traffic does not result in congestion.

Aggregate-based congestion control (ACC) (Ioannidis and Bellovin, 2002; Mahajan et al., 2001) has been proposed to reduce DDOS attack traffic and flash crowds based on congestion level. DDOS attack traffic is defined as a high-bandwidth aggregate, which is a collection of packets from one or more flows that have the same destination address prefix. The detection algorithm in ACC determines the destination addresses of the victim machines based on the destination network prefix of packets dropped at the observed router during a very short period. If the number of dropped packets of a certain destination address is larger than average, ACC puts the destination address on a list. The destination addresses in this list are then clustered into 24-bit or longer network prefixes. The arrival rate of each network prefix is estimated from the number of dropped packets. If the arrival rate of a network prefix exceeds a threshold, ACC regards all traffic to this network prefix as DDOS attack traffic and responds to all incoming traffic sent to this network prefix. The setting of the threshold and the responses will be discussed later in section ''Attack responses''.

Many other studies (Huang and Pullen, 2001; Sterne et al., 2001; Xiong et al., 2001) have suggested network congestion level as an indicator of DDOS attacks. These studies focus on attack

Table 1 Characterization of DDOS defenses in terms of attack detection algorithms

Category	DDOS defense mechanisms	Granularity of detection	Network information monitored	Specific characteristics of attack traffic	Sources of false positives	Limitations
Congestion-based	ACC & pushback (Internet draft expired) (Ioannidis and Bellovin, 2002; Mahajan et al., 2001)	Flow	Destination IP prefix, transmission rate of network traffic	Network flows that cause link congestion	Legitimate traffic that contributes to the congestion	1. False positive increases when the enforcement locations of responses closer to the victims.
	Automatic responses & IDIP (Sterne et al., 2001; Sterne et al., 2002)	Not specified	Not specified			2. Can only identify attack traffic when congestion occurs.
Anomaly-based	TCP SYN anomaly (Schuba et al., 1997)	Connection	IP protocol type (TCP SYN), source IP address	Expired TCP SYN half-open connections	Connections with longer transmission time will not be served	Can only apply on TCP SYN attacks.
	MULTOPS (Gil and Poletto, 2001)	Connection	IP protocol type (TCP), TCP packet rate, source IP address or destination IP address	Asymmetric number of TCP packets to and from one source or destination	IP routing is not necessary symmetric (inbound and outbound traffic may from different border routers)	Can only apply on TCP SYN attacks.
	D-WARD (Mirkovic et al., 2002a)	Flow or connection	IP protocol type, packet rate, source IP address, destination IP address	Packet rates to and from one source (TCP and ICMP) or a maximum sending rate (UDP)		Has to determine the threshold of packet rates for TCP and ICMP, and the maximum sending rate for UDP.
	MIB variables correlation (Cabrera et al., 2001; May et al., 2001)	Packet	Source IP address, destination IP address, MIB variables	Specific values in MIB variables	Some legitimate traffic has the same correlation	Can only apply within a network that is administrated by SNMP and MIB database.
Source-based	Egress/ingress filtering (RFC 2267) (Ferguson and Senie, 1998)	Packet	Source IP address, valid source IP range	Spoofed source IP address	Traffic from an mobile IP that is not tunneled	1. Cannot identify the attack traffic that does not utilize spoofed source IP. 2. Need wide deployment.
	Route-based filtering (Park and Lee, 2001b)	Packet	Source IP address, valid source IP range	Spoofed source IP address	Forwarding tables in core routers do not provide enough information	1. Not applicable to attacks that do not utilize spoofed source IP. 2. Global information about incoming network devices is currently underdeveloped.
	Preferential filtering (Sung and Xu, 2002), threshold filtering (Yaar et al., 2003)	Packet	IP identification (marks by intermediate routers)	Packets with marks considered as attack paths	Legitimate packets may contain the same marks as attack packets	Intermediate routers have to be reconfigured to insert marks.

responses with an implicit assumption that the responses are triggered when link congestion is observed. However, the methods used to determine congestion has not been specified in these studies.

Anomaly-based

TCP SYN anomaly detection

TCP SYN flood attacks are one type of DDOS attacks that exploit half-open TCP connections to deplete the memory of receiver machines. To initiate a normal TCP connection, a sender first sends a ''SYN'' packet and the receiver then sends back a ''SYN ACK'' packet to acknowledge the sender. The sender replies with an ''ACK'' packet to complete the initialization. In a TCP SYN flood attack, the senders do not reply with the ''SYN ACK'' packets. A TCP connection to which the sender has not responded is called a ''half-open'' TCP connection. The receiver stores the connections in system memory and waits for replies. Since the replies never come, the ''half-open'' TCP connections eventually deplete the memory of the receiver and the receiver can no longer serve further connections.

An active monitoring tool has been developed to monitor and to reduce TCP SYN flood attacks (Schuba et al., 1997). The active monitoring method monitors TCP traffic at several points on a local network and utilizes a state machine to determine attack traffic. A new source address that sends TCP SYN is recorded and is assigned to a ''new'' state. The source addresses that do not reply with SYN ACK are assigned to a ''bad'' state. Any SYN packets from the source addresses in the ''bad'' state are regarded as attack traffic. However, if attackers forge and randomize the source addresses of attack packets even if they are sent out from the same machine, the memory of the receiver machine can still be depleted by a large amount of TCP SYN packets.

Asymmetric TCP communications

MULTOPS (Gil and Poletto, 2001) has proposed to detect TCP SYN floods at network routers based on TCP packet rates. In a normal TCP connection, receivers acknowledge packets from senders at a constant rate so that the number of the packets received is proportional to the number of packets sent between the two parties of a connection. In TCP SYN flood attacks, attack sources send out a large amount of SYN packets but receivers will probably not be able to reply to the SYN packets. Based on this pattern, Gil and Poletto assume that the packet rate for the traffic to a network prefix is proportional to the packet rate from the same network prefix. If the proportional pattern changes, the network prefix is either the source of an attack or the destination of an attack.

Normal models of network flows

D-WARD (Mirkovic et al., 2002b) proposes to detect DDOS attack traffic by matching network traffic information with predefined normal flow models. This approach monitors both inbound and outbound traffic of a source network, and is intended to stop attack traffic originating from a network at the border of the source network. Attack flows are identified if they mismatch the normal flow models. Since TCP peer acknowledges every packet it receives, the proposed TCP normal model is defined by a maximum allowed ratio of the number of packets sent and received in the aggregate TCP flow to the peer. The proposed ICMP normal model is defined by a maximum allowed ratio of the number of ICMP request and reply packets, since each normal ICMP message should be paired with a corresponding reply. Since UDP peer is not required to reply to a UDP message, the normal UDP flow model can only be defined by a set of thresholds on UDP packets sent. Although the system is currently underdeveloped, the D-WARD proposal illuminates a new way to detect DDOS attacks at their sources. False positives depend on the calibration of the proposed normal flow models.

MIB variable correlation

Network management information (Cabrera et al., 2001; May et al., 2001) is used to detect DDOS attacks in this method. SNMP is a network management protocol that stores information about network devices in local databases each of which is called a Management Information Base (MIB). Local SNMP agents update variables in MIB periodically. Network administrators can view MIB variables for the traffic sent to local network devices. The assumption is that some MIB variables may indicate attacks if these variables from receiver machines and from sender machines have some correlation on a sequential time line. For example, in ICMP ping flood, attackers send out ICMP Echo requests in which the IP variable in MIB is ''ipOutRequest'', and later the receivers will reply with an ICMP Echo in which the same set of variables contains ''icmpInEchos.'' The detection algorithm queries the values of several specific MIB variables from local network devices periodically and correlates the relationship of these values. The purpose of

the correlation is to reduce the false positives of identifying attack traffic.

Source-based

Since the current IP protocol permits source hosts to alter source addresses in IP packets, attackers are able to send out IP packets with empty or false source addresses. Although IPSEC (Kent and Atkinson, 1998a, 1998b) can be used to authenticate the source addresses of IP packets, this method is not widely adopted at this point. False source addresses will still be a big problem in detecting and filtering DDOS attacks in the short term. Attack victims cannot rely on the source addresses in attack packets to distinguish them from legitimate packets. Various methods have been designed to validate the sources of IP packets.

Egress filtering

Egress filtering[1] (Ferguson and Senie, 1998) determines false source addresses at edge routers based on the valid IP address range internal to the network. However, a false source address on the victim's network will not be detected by this method. For example, if a packet is sent out from host A with the source address of host B, the filtering will not regard it as a false source address if B is valid in this network. In addition, network traffic from a legitimate mobile IP address has to be tunneled to avoid filtering.

Route-based

Route-based filtering proposes filtering packets of spoofed source IP addresses based on routing information on backbone border routers (Park and Lee, 2001b). A border router maintains a routing table that contains fixed routes to all other domains by exchanging routing information with its neighboring routers in Border Gateway Protocol (BGP). The proposal suggests using routing information to determine if a packet comes from the forwarding network device from which it is sent. If it is not, the packet is regarded as an attack packet and should be filtered out. However, current core routers maintain only a forwarding table (a list of destination network prefixes and the corresponding forwarding network interface) but not an incoming table (a list of source network prefixes and the corresponding incoming interface). Although the forwarding table in a router may indicate the routes that a packet will be forwarded to, the routes are not necessarily reversible because routing on the Internet is not completely symmetric (Paxson, 1996). In addition, there is no way to determine where the packet comes from when multiple routes are present. SAVE is a protocol being proposed to build up incoming tables in routers (Li et al., 2001). This protocol proposes that routers propagate their incoming address space to their forwarding destinations.

Web connection authentication

A cryptographic method has been proposed to protect a web server from TCP SYN attacks with spoofed source addresses (Xu and Lee, 2003). This method drops the first TCP SYN packet from the sender and sends back an HTTP redirection with two Message Authentication Code (MAC) keys. The first MAC is encoded with the pseudo-IP address of the redirected web site and the port number pair. The second MAC is encoded with the source IP address of the client and the port number pair. The second MAC is sent in the TCP sequence number of TCP SYN cookie. Future packets with the correct MAC keys will pass through perimeter routers and the ones without will be filtered out. This system is an extension of TCP SYN cookie technology (Karn and Simpson, 1999), which helps an end-system avoid SYN floods by eliminating the half-open connection queue, using a cryptographic value stored in the sequence number of a SYN ACK response. This technology is implemented in the Linux 2.4 kernel.

IP traceback-based

Methods in this category mitigate DDOS attack traffic by using IP traceback and packet filtering. Packet marking (Park and Lee, 2001a; Savage et al., 2001; Song and Perrig, 2001; Sung and Xu, 2002; Yaar et al., 2003) identifies the paths that attack traffic comes from by inserting marks in packets. Among these methods, currently only IP traceback-based intelligent packet filtering (Sung and Xu, 2002) and Pi (Yaar et al., 2003) have been designed to filter out ongoing attack traffic.

The basic idea of packet marking is that the routers on the path from attack sources to victims insert marks in the IP identification field of ongoing packets, and the victims distinguish the attack packets from legitimate packets based on the marks in the packets. The problem is that the IP identification field is only 16 bits, which is not enough for storing the entire path (the average path length is roughly 15) (Yaar et al., 2003). Certain coding schemes have to be applied to

[1] The term here is from an end-organization view and not from an ISP view. From an ISP-centric point of view, this exact same concept is called "ingress filtering". It's all a matter of where you stand in the network and apply the filters.

shorten the length of marks. Since the current coding schemes are not able to assign each mark to an unique path, legitimate packets would be treated as attack packets if they have traversed the path coded as the same mark as the path traversed by the attack packets. IP traceback-based intelligent packet filtering (Sung and Xu, 2002) proposes a preferential filtering to filter out packets with different types of marks with different probabilities. Pi (Yaar et al., 2003) proposes to filter packets at edge routers at a certain threshold if the packets have marks that indicate they are from attack sources. Since the mark under this scheme is not unique to every path, the threshold filtering allows the victim to lower the false positives at the expense of raising the false negatives. Both methods allow attack victims to categorize network packets based on marks in packets but they need to be combined with other methods for identifying the marks that represent attack traffic.

Attack responses

After the attack detection algorithm identifies attack traffic, attack responses decide where, when and how network routers or proxy servers drop the attack traffic. To address how the responses are deployed on the Internet infrastructure, we derived characteristics of attack responses to identify how the responses are generated (response generation), what actions the responses take (response mechanism), where the responses collect information to decide their actions (decision locations) and where the responses are applied (enforcement locations). These characteristics determine the number of locations needed to deploy defenses, and thus influence the deployment costs.

In addition, to consider the feasibility of defenses to tolerate future changes in an infrastructure, we discuss if the defense is topology dependent, how the responses are communicated in a distributed network and the overhead of the responses. Table 2 summarizes the categorization of attack responses. Since network providers could provide DDOS defenses to either attack victims or potential attack sources, we further categorize attack responses into destination filtering (for attack victims) and source filtering (for attack sources).

1) Destination filtering are attack responses that are triggered when attacks are detected in the inbound traffic of some destination networks such as potential attack victims. Defenses in this category monitor the network traffic received by some destination networks, and mitigate the impacts of ongoing attack traffic to these destinations. As in Figs. 1 and 2, when subscriber 1 (in ISP 1's network) originates attacks on subscriber 2 (in ISP 2's network), the attack responses are deployed in ISP 2's network. In this case, ISP 2 (the downstream ISP) can only trace back the sources of attacks within the administrative boundary of its network, such as the access router connecting to the subscriber as in Fig. 1 or the border of its network as in Fig. 2. Proposed responses that fall in this category include Pushback (Ioannidis and Bellovin, 2002; Mahajan et al., 2001), Active Responses (Sterne et al., 2001; Sterne et al., 2002), TCP anomaly detection (Schuba et al., 1997), MIB correlation (Cabrera et al., 2001; May et al., 2001), preferential filtering (Sung and Xu, 2002) and threshold filtering (Yaar et al., 2003).
2) Source filtering occurs when attack responses are triggered when attacks are detected in the outbound traffic of some destination networks such as potential attack sources. Defenses in this category monitor the network traffic sent from some source networks, and mitigate the impacts of ongoing attack traffic originating from these sources. Since the attacks are filtered out at the sources before they are sent to the downstream subscribers, this method decreases the observable number of attacks at downstream ISPs. Fig. 3 illustrates an example where ISP 1 places filters at the upstream routers of subscribers 1 so that the attack traffic is filtered out before it is sent to subscriber 2. Defenses in this category are egress filtering (Ferguson and Senie, 1998) and D-WARD (Mirkovic et al., 2002b). Both MULTOPS (Gil and Poletto, 2001) and route-based filtering (Park and Lee, 2001b) can be either implemented as destination filtering or source filtering.

Attack response generation

Attack response generation refers to the process of generating rules to filter out attack traffic. We distinguish attack responses between static filters and dynamic filters. Static filters are attack responses in which the filtering rules are set manually by network administrators. Dynamic filters are the ones in which the filtering rules are set

Table 2 Characterization of DDOS defenses in terms of attack responses (both MULTOPS and route-based filtering can be applied on either inbound or outbound traffic)

Category	DDOS defense mechanisms	Response generation	Response mechanism	Decision locations	Enforcement locations	Topology dependent	Com. protocol	Overhead
Destination filtering (monitor inbound traffic of subscribes)	ACC & pushback (Ioannidis and Bellovin, 2002; Mahajan et al., 2001)	Dynamic	Rate limiting	Edge routers of destinations or upstream access routers (L4, L5)	All locations (L1—L5)	Yes	Yes. Network layer protocol	Controls messages to push responses
	Automatic responses & IDIP (Sterne et al., 2001; Sterne et al., 2002)	Dynamic	Rate limiting & packet filtering	The discovery coordinator (single point on a network)	All locations (L1—L5)	Yes	Yes. Application layer protocol	Control messages for coordination
	TCP SYN anomaly (Schuba et al., 1997)	Dynamic	Packet filtering	Edge routers of destinations (L4)	Edge routers of destinations (L4)	Not specified	No	States of connections
	MIB variables correlation (Cabrera et al., 2001; May et al., 2001)	NA	NA	NA	NA	Not specified	SNMP for retrieving MIB variables	SNMP messages
	MULTOPS (Gil and Poletto, 2001)	NA	NA	Edge routers of destinations or upstream access routers (L4, L5)	Edge routers of destinations or access routers (L4, L5)	Not specified	No	Hash table to store TCP connection info.
	Route-based filtering (Park and Lee, 2001b)	Static	Packet filtering	Core routers (L3)	Vertex cover set of core routers (L3)	Yes	No	Route information
	Preferential filtering (Sung and Xu, 2002), threshold filtering (Yaar et al., 2003)	Dynamic	Packet filtering	Edge routers of destinations (L4)	Edge routers of destinations or access routers (L4, L5)	Yes	No	Mark insertion in intermediate routers
Source filtering (monitor outbound traffic of subscribers)	MULTOPS (Gil and Poletto, 2001)	NA	NA	Edge routers of destinations or upstream access routers (L1, L2)	Edge routers of sources (L1) or upstream access router of sources (L2)	Not specified	No	Hash table to store TCP connection info.
	Egress/ingress filtering (Ferguson and Senie, 1998)	Static	Packet filtering			Not specified	No	Access lists
	D-WARD (Mirkovic et al., 2002a)	Dynamic	Rate limiting			Not specified	No	Hash table to compute flow measures
	Route-based filtering (Park and Lee, 2001b)	Static	Packet filtering	Core routers (L3)	Core routers (L3)	Yes	No	Route information

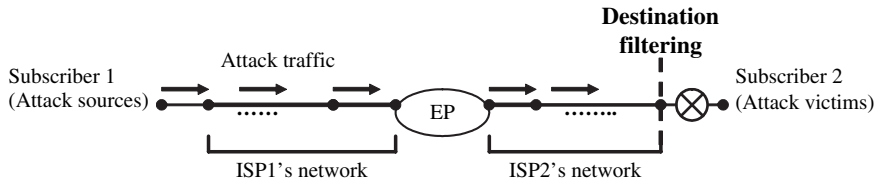

Figure 1 An illustration of destination filtering[2] (at victim upstream).

automatically by attack detection algorithms. Ingress filtering is an example of static filters. Rules in egress filtering are set manually by network administrators in against spoofed source IP addresses (Ferguson and Senie, 1998). For example, the network prefix of a local network is 204.69.207.0/24. In order to prevent IP addresses that are not within the legitimate address range to originate network traffic, the rules that trigger attack responses are defined to drop all packets in which source IP addresses is outside 204.69.207.0/24.

If a defense consists of an attack detection algorithm, rules can be set automatically when attack traffic is detected. For example, the outbound link of the above network is 2 Mpbs. An attack is detected when the attack source sends 5 Mbps TCP SYN packets to port 80 of the host 204.69.207.9. An attack response to limit the transmission rate of TCP packets to this machine can be generated automatically to limit the packet rate of the network traffic sent to the host 204.69.207.9 to be much lower than 2 Mbps.

Response mechanisms

To implement attack responses, contemporary routers usually have the functionalities to process network traffic flows based on a set of access rules that defines the characteristics of attack traffic (CISCO, 2000). Packet filtering and rate limiting are two mechanisms to implement responses in access rules of routers. Either one of these methods is used in the defenses described earlier to implement filtering. Packet filtering either drops or accepts the packet being examined. The granularity of attacks in these two mechanisms is different. Packet filtering detects attacks based on per-packet information while rate limiting limits the transmission rate of the traffic flows to which the packet belongs.

Packet filtering is the action that a device takes to selectively control the flow of data to and from a network. Packet filters allow or block packets, usually while routing them from one network to another. To accomplish packet filtering, network administrators have to establish a set of rules that specify what types of packets are to be allowed and what types are to be blocked. Packet filtering may occur in a router, in a bridge, or in an individual host (Zwicky et al., 2000).

Rate limiting is the function that allows a router to control the transmission rate of a specific traffic flow. Rate limiting is a traffic-policing tool used to control network congestion. In the case of protecting against DDOS attacks, an attack detection algorithm identifies the characteristics of the traffic flow that will be policed. Once the characteristics are determined, the rate limiting function will guarantee that the transmission rate of the traffic flow will be lower than a certain rate, which means packets that arrive at a higher rate will be queued or dropped at the router.

Both packet filtering and rate limiting are mechanisms to respond against the DDOS attack traffic; however, they control the attack traffic in different ways. Packet filtering discards all packets that match the characteristics of attack traffic. In contrast, rate limiting allows some network traffic regarded as attack traffic to pass through, but it is limited by a transmission rate. Because of the difference, packet filtering is usually used with an attack detection algorithm that can detect attacks by packet headers, such as anomaly-based and source-based, and rate limiting is used with congestion-based attack detection algorithms in which the attack traffic cannot be distinguished from legitimate traffic sent to the same destination.

Decision locations

Decision locations refer to where the filtering rules are generated if the attack responses are dynamic filters. In order to generate the filtering rules, an attack detection algorithm needs to collect network traffic information from the decision locations. Theoretically, attack response generation can take place at either one of the following locations (Fig. 4):

- Attack sources (L1): edge routers of the local network from where the hosts send out packets.

[2] EP refers to the exchange point that exchanges the network traffic between two backbone networks.

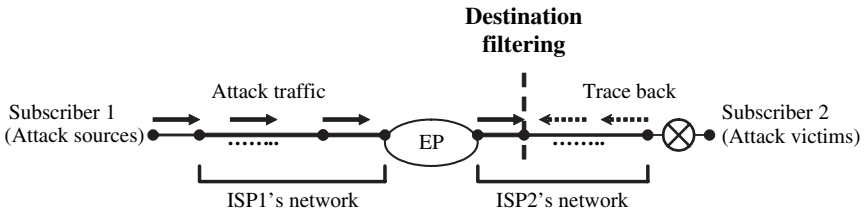

Figure 2 An illustration of destination filtering.

- Source upstream (L2): access routers of an ISP that connect to subscribers' edge routers.
- Backbone routers (L3): core routers that transport network traffic.
- Victims (L4): edge routers of the local network where hosts will receive packets.
- Victim upstream (L5): access routers of an ISP that connect to edge routers of the victims' network.

In practice, attack response generation rarely takes place at backbone routers (L3) since it is difficult under current technology to monitor high-speed backbone peering links and to analyze the information from these links for attack detection. Studies have been done on monitoring OC-48 peering links (Claffy et al., 1998; Fraleigh et al., 2001). No current published study has monitored links higher than OC-48.

Instead of deploying a defense at backbone routers, edge routers are another choice. In order to protect a local network against attack traffic from other networks, network administrators have an incentive to deploy attack detection tools at edge routers to examine inbound network traffic. All anomaly-based detection algorithms described in section "Anomaly-based" generate filtering rules either at the victims (L4) or at the victim upstream (L5) to examine inbound network traffic.

Generating filtering rules at the attack sources or the source upstream (L1 or L2) is hard due to three reasons. First, the sources of attack traffic can be spoofed so that victims cannot identify the real sources of attacks. Secondly, even if the genuine sources of attacks can be identified, these sources can be located at many different administrative network domains. In this case, cooperative attack detection and response are necessary. Thirdly, technical difficulties occur for generating filtering rules at the sources of attacks. In particular, it is hard to distinguish DDOS attack traffic from legitimate traffic at the sources of attacks since the volume of attack traffic is usually small and only aggregates at certain points close to destinations. Congestion-based attack detection algorithms are not effective in this case since attack tools usually do not cause congestion at the sources. Anomaly-based algorithms, such as D-WARD, and source-based algorithms are able to generate filtering rules at attack sources.

Enforcement locations

Enforcement locations refer to where on a network the attack responses will be applied. Once an attack response is enforced on a certain network router, all network packets that pass through the router/links will be examined. If network packets are determined to be attack traffic, the responses will be applied to these packets.

Possible enforcement locations are the same as decision locations (L1—L5 in section "Decision locations"). The difference is that enforcement locations in practice are not as restrictive as decision locations because defenses do not analyze network traffic at enforcement locations. Once filtering rules are generated at decision locations, they can be sent to enforcement locations. Attack responses at enforcement locations only impose filtering overhead. In DDOS attacks, appropriate allocation of enforcement locations

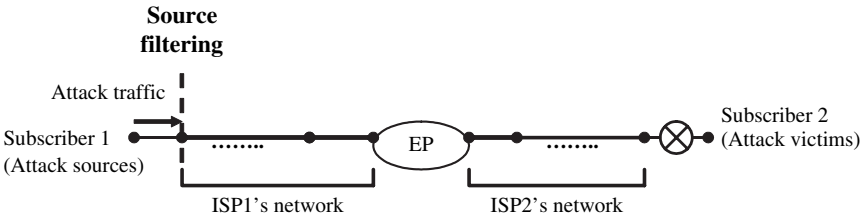

Figure 3 An illustration of source filtering.

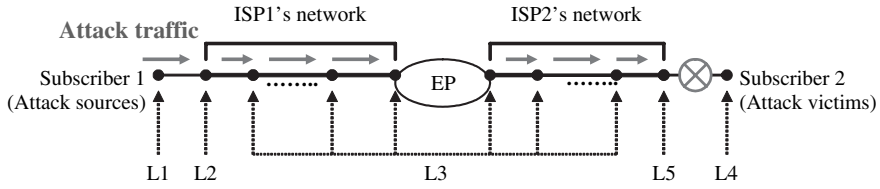

Figure 4 Possible locations for attack response generation in an example with only one attack source and attack victim.

may enhance the performance of defenses and reduce the overhead imposed. Both Pushback (Ioannidis and Bellovin, 2002; Mahajan et al., 2001) and Active Responses (Sterne et al., 2001; Sterne et al., 2002) can be enforced at all locations discussed above (L1—L5).

Impact of topology

The performance of some defenses depends on where the attack responses are deployed. As a result, a change in the network topology due to changes in ISPs' network infrastructure or in routing protocols will have an impact on how well the defense can filter out attack traffic. For example, congestion-based Pushback (Ioannidis and Bellovin, 2002; Mahajan et al., 2001) method reduces false positives by pushing the attack responses closer to the attack sources. Route-based filtering (Park and Lee, 2001b) requires topology information to determine if the network traffic is sent from a correct forwarding network device. When duplicate marks on different routes increase, false positives in IP traceback-based defenses (Sung and Xu, 2002; Yaar et al., 2003) increase.

Communicating protocols

Communicating protocols refer to the protocols used to send control messages between various nodes of a network to coordinate attack detection or attack responses. These control messages are either attack patterns sent from attack detectors to attack response decision locations or filtering rules sent from decision locations to enforcement locations. Sending control messages has been done manually which imposes high managerial overhead and has a longer lag time. To reduce the managerial overhead and lag time, communicating protocols have been studied to manage the generation and the distribution of attack responses in distributed locations. Three communicating protocols are explained below.

First, pushback messages (Ioannidis and Bellovin, 2002; Mahajan et al., 2001) are used to distribute congestion patterns observed at congested links to trigger rate limiting in routers along the path that attack packets have traveled. The "pushback-request" message used to trigger rate limiting includes congestion signature, bandwidth limit, expiration time, depth (how many hops away from congested links), and message type. Second, the Intruder Detection and Isolation Protocol (IDIP) is an application layer protocol that coordinates attack detection and response at distributed locations. In IDIP, attack detectors send descriptions of suspicious attack events to the Discovery Coordinator, which determines responses and sends out its decisions to nodes that will enforce the decisions (Schnackenberg and Djahandari, 2000; Sterne et al., 2001; Sterne et al., 2002). Third, the Common Intrusion Detection Framework (CIDF) proposes a language called Common Intrusion Specification Language (CISL) for intrusion detection systems to communicate attack responses (Staniford-Chen et al., 1998). CISL provides a common platform for communicating filter policy and attack detection patterns between heterogeneous intrusion detection systems located at distributed locations.

Finally, not all defenses require additional control messages. If each network node can detect attacks autonomously based on the information that a network node collects periodically, attack detection can be implemented without additional communicating protocols. In addition, the lack of bandwidth during DDOS attacks may have impacts on in-band control protocols for communications. Especially on downstream systems, a DDOS flood could overwhelm systems and limit the use of in-band control protocols to detect and respond to the trouble.

Additional overhead of responses

Defenses mitigate the impact of the attack traffic on the victim network but may impose an additional overhead on the networks that implements them. The additional overhead includes computational overhead imposed by attack detection and attack response enforcement; storage

requirement to save logs for attack detection; and communications overhead used to send control messages to distributed locations of a network. The overhead is described below in detail.

First, attack responses may impose computational overhead on network devices. Once filtering rules are enforced to examine network packets, a per-packet delay will occur for matching filter rules. Minimizing the per-packet delay is a packet classification problem in router performance optimization. Although most commercial routers are optimized for routing, the per-packet delay of matching filtering rules depends on the number of filtering rules, the number of characteristics used to identify attacks, and the updating frequencies of the filtering rules (Feldmann and Muthukrishnan, 2000).

Second, attack detection algorithms impose a storage requirement of saving network information to determine attack characteristics. To monitor high-speed network links, the storage requirement is usually very large. Current technology can scale up to 10 Gbps link speed without losing much information on IP packets. To reduce the storage requirement and to catch network packets from high throughput routers, sampling and processing of packet data dynamically will be needed in the future (Iannaccone et al., 2001).

Third, control messages to coordinate attack detection are an additional overhead to network transmission. If communication occurs between network routers, it is important to know if such communication will result in abnormal behavior of routers. Since most commercial routers are optimized for routing, it is not certain if additional communications among routers will impose additional delay on routers or not. Future work should explore how communication overhead impacts system performance. For example, in downstream systems, a DDOS flood could overwhelm systems and limit the use of in-band control protocols to detect and respond to the trouble. This is a limitation of such technology, and lends credence to more local intelligence for throttling attacks. However, future research on this topic is needed.

Conclusions

We presented a categorization of DDOS defenses that are network-based and actively filter out ongoing attack traffic. DDOS defenses are categorized based on both attack detection algorithms and attack responses.

Categorizing DDOS defenses based on attack detection algorithms helps to identify the factors that influence the performance tradeoff of defenses. In the congestion-based defenses, attack detection is based on link congestion and rate limiting is used to respond against attacks. False positives for these defenses occur when both attack traffic and legitimate traffic happen to have the same destination IP prefix. In the anomaly-based defenses, attack detection is based on the anomaly patterns of network traffic, and packet filtering is used to drop attack packets. False positives occur when legitimate traffic shows anomaly patterns in some rare cases. In the source-based defenses, attack detection is based on false source IP addresses. False positives occur only when the criteria for determining false IP addresses cannot distinguish it from true source addresses. However, false negatives depend on how many attack packets contain true source addresses.

Categorizing defenses based on attack responses can help the deployment of a defense. ISPs can utilize the distinction between destination filtering and source filtering to design the service provision for either potential attack sources or attack victims. The locations where attack responses are generated and enforced determine the number of locations needed to deploy defenses, and thus influence the deployment costs.

This characterization can serve as a foundation for quantitative analyses that compare the performance and deployment costs of DDOS defenses. We hope that this categorization will provide Internet Service Providers (ISPs) some insights when they consider DDOS defenses as network services to their subscribers.

Acknowledgments

This work was supported in part by the NSF/ITR 0218466 and the Pennsylvania Infrastructure Technology Alliance, a partnership of Carnegie Mellon, Lehigh University, and the Commonwealth of Pennsylvania's Department of Economic and Community Development. Additional support was provided by ICES (the Institute for Complex Engineered Systems) and CASOS — the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University (Available from: http://www.casos.cs.cmu.edu). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, the Commonwealth of Pennsylvania or the U.S. government.

References

Arbor. PeakFlow, Waltham, MA: Arbor Networks, Inc.; 2002.

Asta. Vantage System, Seattle, WA: Asta Networks, Inc.; 2002.

Axelsson S. Intrusion detection systems: a survey and taxonomy. Department of Computer Engineering, Chalmers University, Goteborg, Sweden Technical Report 99-15; March 2000.

Bellovin SM. ICMP traceback message. Internet Draft: draft-bellovin-itrace-00.txt; March 2000.

Burch H, Cheswick B. Tracing anonymous packets to their approximate source. Presented at LINUX System Administration Conference, New Orleans, LA; 2000.

Cabrera JBD, Lewis L, Qin X, Lee W, Prasanth RK, Ravichandran B, et al. Proactive detection of distributed denial of service attacks using MIB traffic variables — a feasibility study. Presented at IEEE/IFIP International Symposium on Integrated Network Management; 2001.

CERT/CC. Results of the distributed-systems Intruder tools workshop, Pittsburgh, Pennsylvania, USA: CERT Coordination Center, Software Engineering Institute, Carnegie Mellon University; November 2—4. 1999.

Cheswick WR, Bellovin SM. Firewalls and internet security: repelling the Wily Hacker: Addison-Wesley Pub Co; 1994.

CISCO. Strategies to protect against distributed denial of service. CISCO 2000.

Claffy KC, Miller G, Thompson K. The nature of the beast: recent traffic measurements from an Internet backbone. Presented at INET, Geneva, Switzerland; 1998.

Debar H, Dacier M, Wespi A. Towards a taxonomy of intrusion detection systems. Computer Networks, vol. 31, 1999.

Feldmann A, Muthukrishnan S. Tradeoffs for packet classification. Presented at IEEE Infocom; 2000.

Ferguson P, Senie D. Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing. IETF RFC2267 RFC 2267; January 1998.

Fraleigh C, Moon S, Diot C, Lyles B, Tobagi F. Packet-level traffic measurement from a Tier-1 IP backbone. Sprint ATL, Burlingame, CA Sprint ATL Technical Report TR01-ATL-110101; 2001.

Gil TM, Poletto M. MULTOPS: a data-structure for bandwidth attack detection. Presented at USENIX Security Symposium. Washington, DC; 2001.

Huang Y, Pullen JM. Countering denial-of-service attacks using congestion triggered packet sampling and filtering. Presented at 10th International Conference on Computer Communications and Networks; 2001.

Iannaccone G, Diot C, Graham I, McKeown N. Monitoring very high speed links. Presented at ACM Internet Measurement Workshop. San Francisco; 2001.

Ioannidis J, Bellovin SM. Implementing pushback: router defense against DDoS attacks. Presented at Network and Distributed System Security Symposium; 2002.

Karn P, Simpson W. Photuris: session-key management protocol. IETF RFC 2522; March 1999.

Kent S, Atkinson R. Security architecture for the Internet protocol. The IP Security Protocol Working Group, Internet Engineering Task Force; 1998a.

Kent S, Atkinson R. IP authentication header. The IP Security Protocol Working Group, Internet Engineering Task Force; 1998b.

Li J, Mirkovic J, Wang M, Reiher P, Zhang L. SAVE: source address validity enforcement protocol. Presented at IEEE INFOCOM; 2001.

Lipson H. Tracking and tracing cyber-attacks: technical challenges and global policy issues, Pittsburgh: CERT Coordination Center, Software Engineering Institute; November 2002. Available from: <http://www.cert.org/archive/pdf/02sr009.pdf>.

Mahajan R, Bellovin SM, Floyd S, Ioannidis J, Paxson V, Shenker S. Controlling high bandwidth aggregate in the network. Computer Communications Review 2001.

May J, Peterson J, Bauman J. Attack detection in large networks. Presented at DARPA Information Survivability Conference and Exposition. Anaheim, CA; 2001.

Mirkovic J, Martin J, Reiher P. A taxonomy of DDoS attacks and DDoS defense mechanisms. Computer Science Department, University of California, Los Angeles Technical report #020018; 2002a.

Mirkovic J, Prier G, Reiher P. Attacking DDoS at the source. Presented at Proceedings of ICNP. Paris, France; 2002b.

Mukherjee B, Heberlein LT, Levitt KN. Network intrusion detection. IEEE Network 1994;8:26—41.

Park K, Lee H. On the effectiveness of probabilistic packet marking for IP traceback under denial of service attack. Presented at Proceedings of IEEE INFOCOM; 2001a.

Park K, Lee H. On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law internet. Presented at ACM SIGCOMM'01. San Diego, CA; 2001b.

Paxson V. End-to-end routing behavior in the Internet. Presented at ACM SIGCOMM '96; 1996.

Recourse. ManHunt, Redwood City, CA: Recourse Technologies, Inc; 2002.

Savage S, Wetherall D, Karlin A, Anderson T. Practical network support for IP traceback. ACM/IEEE Transactions on Networking 2001;9:226—37.

Schnackenberg D, Djahandari K. Infrastructure for intrusion detection and response. Presented at DARPA Information Survivability Conference and Exposition (DISCEX); 2000.

Schuba CL, Krsul IV, Kuhn MG, Spafford EH, Sundaram A, Zamboni D. Analysis of a denial of service attack on TCP. Presented at IEEE Symposium on Security and Privacy; 1997.

Snoeren AC, Partridge C, Sanchez LA, Jones CE, Tchakountio F, Kent ST, et al. Hash-based IP traceback. Presented at ACM SIGCOMM; 2001.

Snoeren AC, Patridge C, Sanchez LA, Jones CE, Tchakountio F, Schwartz B, et al. Single-packet IP traceback. IEEE Transaction on Networking 2002;10:721—34.

Song DX, Perrig A. Advanced and authenticated marking schemes for IP traceback. Presented at IEEE Inforcom; 2001.

Spatscheck O, Peterson LL. Defending against denial of service in Scout. Operating Systems Review 1998.

Staniford-Chen S, Tung B, Schnackenberg D. The common intrusion detection framework (CIDF). Presented at DARPA Information Survivability Workshop. Orlando FL; 1998.

Sterne D, Schnackenberg D, Babson B, Holliday H, Djahandari K, Reid T, et al. Autonomic response to distributed denial of service attacks. Presented at Recent Advances in Intrusion Detection conference; 2001.

Sterne D, Schnackenberg D, Balupari R, Cholter WL, Babson B, Wilson B, et al. Active network based DDoS defense. Presented at DARPA Active Networks Conference and Exposition; 2002.

Sung M, Xu J. IP traceback-based intelligent packet filtering: a novel technique for detecting against Internet DDos attacks. Presented at IEEE International Conference on Network Protocols; 2002.

Tran KTL. Hackers attack major Internet sites, temporarily shutting Buy.com, Ebay. Wall Street Journal 2000:3.

Xiong Y, Liu S, Sun P. On the defense of the distributed denial of service attacks: an on—off feedback control approach. IEEE Transaction on Systems, Man, and Cybernetics — Part A: Systems and Humans 2001;31:282—93.

Xu J, Lee W. Sustaining availability of web servers under server denial of service attacks. IEEE Transaction on Computers,

special issue on Reliable Distributed Systems 2003;52(2): 195–207.
Yaar A, Perrig A, Song D. Pi: a path identification mechanism to defend against DDos attack. Presented at IEEE conference on security and privacy; 2003.
Yan J, Early S, Anderson R. The XenoService — a distributed defeat for distributed denial of service. Presented at Information Survivability Workshop; 2000.
Yankee. $1.2 billion impact seen as a result of recent attacks launched by Internet hackers. The Yankee Group, Research Notes February 2000.
Zwicky ED, Cooper S, Chapman DB, Russell D. Building internet firewalls. 2nd ed.: O'Reilly & Associates; 2000.

Li-Chiou Chen received her PhD in Engineering and Public Policy in 2003 from Carnegie Mellon University in Pittsburgh, PA. She is currently a post-doctorate researcher at the Institute for Software Research International in the School of Computer Science, Carnegie Mellon University. Since August 2004, she will be an assistant professor at the School of Computer Science and Information Systems, Pace University. Her dissertation entitled "Computational Models for Defenses against Internet-based Attacks," utilizes a network-based simulation tool to analyze the policy and economic issues in the provision of defenses against Distributed Denial of Service attacks. Her current research interests are focused on combining artificial intelligence and agent-based modeling to conduct technological and policy analysis in the area of information security.

Thomas A. Longstaff received his PhD in 1991 at the University of California, Davis, in software environments. He is a technical manager in the Network Situational Awareness Program at the Software Engineering Institute (SEI), Carnegie Mellon University. He is currently managing research and development in network security for this Program. His publication areas include information survivability, insider threat, intruder modeling, intrusion detection, and infrastructure security. Prior to coming to SEI, Dr. Longstaff was the technical director at the Computer Incident Advisory Capability (CIAC) at Lawrence Livermore National Laboratory in Livermore, California.

Kathleen M. Carley received her PhD from Harvard. She is a professor at the Institute for Software Research International, Carnegie Mellon University. Her research combines cognitive science, social networks and computer science. Specific research areas are dynamic network analysis, computational social and organization theory, adaptation and evolution, computational text analysis, and the impact of telecommunication technologies and policy on behavior and disease contagion within and among groups. Her models meld multi-agent technology with network dynamics and empirical data. Illustrative large-scale multi-agent network models she and the CASOS team have developed are: BioWar — city, scale model of weaponized biological attacks; OrgAhead — a strategic and natural organizational adaptation model; and DyNet — a change in covert networks model.