

A New Marking Scheme to Defend against Distributed Denial of Service Attacks

Zhiqiang Gao, Nirwan Ansari, and Karunakar Anantharam

Department of Electrical and Computer Engineering

New Jersey Institute of Technology

Newark, NJ 07102, USA

{zg4, nirwan.ansari, ka33} @njit.edu

Abstract— In this paper, we propose a new mechanism to defend against Distributed Denial of Service (DDoS) attacks with path information rather than IP address information. Instead of the complete binary tree model, our proposal is based on the Four Color Theorem. The salient feature of the Theorem is that it allows color *reuse* so that even some portions of the map have more than 4 neighbors, 4 colors are still sufficient to mark all their borders. This idea of *reuse* is very important because some routers have many interfaces and the length of the ID field in the header of an IP packet, where the marking information is embedded, is very limited. Furthermore, our marking scheme takes the Internet hierarchy into account, and greatly relaxes the limitation on the number of interfaces of routers, thus making the scheme more practical. Simulation results have validated our design.

Keywords—network security; DoS/DDoS attack; marking; four color theorem

I. INTRODUCTION

Distributed Denial of Service (DDoS) attacks have become a main threat to the integrity of the Internet. Well-known DDoS attacks against high-profiled websites, such as Yahoo, CNN, and Amazon happened in early 2000. A more recently public-aware malicious DDoS attack occurred in Oct. 2002, which crippled 8 of 13 DNS root servers. If the attack had lasted several more hours, all Internet users would have been affected seriously. Known attacks are only a tip of the iceberg, however. In fact, many DDoS attacks happened without public awareness. The CSI/FBI survey reported that 32% of respondents detected Denial of Service (DoS) attacks against their sites [1]. A recent research [2] conducted by Moore *et al.* found that there were more than 12,000 attacks against more than 5,000 targets in 3 weeks.

Several approaches have been proposed to address issues related to DoS and DDoS attacks. The first approach actively prevents the DoS/DDoS attacks by deploying firewalls. However, the effectiveness of firewalls depends on the attack rate. For example, a specialized firewall designed to resist SYN flood (a kind of DDoS attacks) becomes useless if the flood attack rate exceeds a threshold, say 14,000 packets per second [3]. The design objective of the second class of schemes, referred to as the defensive schemes, is to sustain the services of the attacked sites or network. IP traceback addresses another security issue [4]-[9]. A successful IP traceback can determine

the attack sources and is helpful in finding the hidden attacker and subsequently enforcing some penalty. However, IP traceback cannot mitigate the attack effect of an ongoing attack on the victim [7].

To defend against a DDoS attack, the victim needs to know certain features of the attack traffic. To that end, routers may be employed to perform markings. Similar to IP traceback, two marking strategies can be employed. One method attempts to record the information of the whole path that the attack packets traverse; another focuses only on the information of edge routers from which the attack traffic enters the Internet [8]. While the salient feature of the later scheme is simple and easy to implement, one drawback exists. Since the victim relies *only* on the edge routers to perform markings, in a DoS attack, no additional hint can be exploited about the attack source if the marking information is incorrect (e.g., the edge router is compromised). On the contrary, it is generally impossible for an attacker to sabotage all routers along the attack path. Therefore, the whole path information may be more robust against sophisticated attackers.

Recently, Yaar *et al.* [10] proposed an ingenious scheme, Pi (Path identification). Their idea is to distinguish one path from another with path identification rather than IP addresses. Using the complete binary tree model, the 16-bit ID field can be used to record 16 links, 1 bit for each link. However, a potential assumption of the binary tree model is that each router has only 2 interfaces. In reality, many routers have more than 2 interfaces. Recent Internet measurement shows that the number of interfaces of 99% of the routers in the Internet is no greater than 8 [11]. Therefore, it is reasonable to consider at most 8 interfaces for each router. A naive extension to [10] can use 3 bits rather than 1 bit to distinguish one interface from another, and then the 16-bit ID field can record only 5 links. To reduce the possibility of false positives, we prefer to record as many links as possible in the 16 bits. Specifically, we want to use only 2 bits to distinguish interfaces (up to 8) of a router. This is the motivation to investigate the applicability of the Four Color Theorem [12] to tackle this problem.

The Four Color Theorem states that to color a map so that any adjacent region has a different color, at most 4 colors are required [13] (see Fig. 1). Though there may exist some areas that have many neighbors, the same color can be *reused* as long as their neighbors are not adjacent. Intuitively, this is very similar to our case here. That is, we have many routers, each

This work has been supported in part by the New Jersey Commission on Higher Education via the NJI-TOWER project, and the New Jersey Commission on Science and Technology via NJWINS.

with a different number of interfaces. Interfaces can use the same color as long as they are distinguishable. Based on the Internet hierarchy, we contrive a new color marking method.

The rest of the paper is structured as follows. Related works are reviewed in Section 2. We next discuss two extensions to Pi in Section 3. In Section 4, we outline our marking scheme and show its benefits. The simulation results are shown in Section 5. We finally summarize our findings in Section 6.

II. RELATED WORKS

Related works are discussed in two aspects. One is related to network hierarchy, and the other is related to the defense against DDoS attacks.

Network hierarchy is well known but has rarely been applied in network security. Gao [14] first proposed a new scheme to infer Autonomous System (AS) relationships. Later, Subramanian *et al.* [15] advanced Gao’s work by dividing the Internet into 5 layers. Another similar work can be found in [16]. The main contribution of their works is their observation that the path from one node to another in the Internet first goes “uphill” to the uphill top provider, then from the uphill top provider to the downhill top provider, and finally goes “downhill” from the downhill top provider to the destination. In Fig. 2, node A and T in the uphill part stand for the attack node and the ingress edge router of the uphill top provider, respectively, and node T’ and V in the downhill part stand for the egress edge router of the downhill top provider and the victim, respectively. Note that in this figure, an uphill vertical line stands for the link from a customer to the edge router of its provider, and a downhill vertical line stands for a link from an edge router of a provider to its customer, while a horizontal line represents links within one domain (for example, one AS) where may contain several routers. That is, a vertical line represents only one link, but a horizontal line may represent several links as long as these links are in one domain.

As mentioned earlier, defending against DDoS attacks is a very difficult task because the source IP address can be easily forged. Among many proposals [3]-[10], [17], and [18], Pi [10] is a promising one. Two salient features make it attractive. Since it is deterministic marking rather than probabilistic marking, and all the marking information is imbedded in one packet, it is possible to defend against not only flood-based DDoS attacks, but also attacks induced by a few packets. Another benefit is that the victim can defend against DDoS attacks, without depending on the cooperation of the upstream system administrators. However, one assumption of the binary tree model is that each router has only two interfaces. This assumption imposes a serious limitation on the implementation of the proposal—the very reason that motivates our work.

Our work improves upon [10] in two aspects. First, we use the Four Color Theorem rather than the binary tree model, and thus we greatly relax the limitation on the number of interfaces of routers and make the scheme more practical. Second, they mark packets only at the edge of each Autonomous Systems (ASs) while we mark packets of the first and last portions along the attack path. The first portion of the attack path is from the attack node to (the ingress edge router of) the uphill top

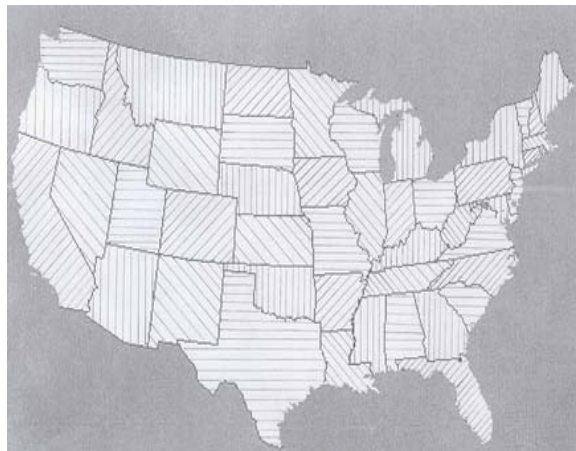


Figure 1. One example of color marking of the US mainland [13]. Four different directions of lines represent four different colors.

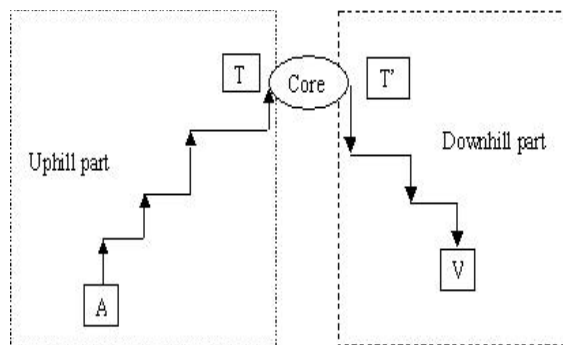


Fig. 2. A schematic representation of one Internet path.

provider and the last portion is from (the egress edge router of) the downhill top provider to the victim.

III. EXTENSIONS TO PI

As mentioned earlier, 99% of routers in the Internet own no more than 8 interfaces [11] (see Fig. 3). A naïve extension to Pi [10] can use the octary tree with the victim as the root. That is, each node in the tree has no more than 8 children. Since there are 8 children, 3 bits are required to distinguish one from another. As shown in Fig. 4, a path from A to V can be represented by a sequence of bits, e.g., 101110000. Note that this extension is not a “complete” octary tree since some nodes can have less than 8 children.

Given the limitation of the 16-bit ID field, using 3 bits for each link can only record information of 5 links (or hops) out of the possible 32 hops (very few paths in the Internet have more than 30 hops [19]. A maximum path-length of 32 hops is considered as a common practice [4]-[10]). Therefore, further improvements are required.

Note that in the naive extension model, up to 8 children are considered for each node. If a node stands for a router, and a branch represents a link between the interfaces of two different

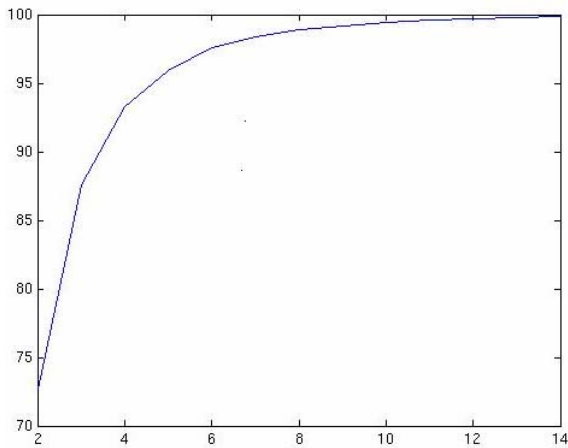


Fig. 3. The cumulative probability distribution (CPD) of the number of interfaces among routers in the Internet.

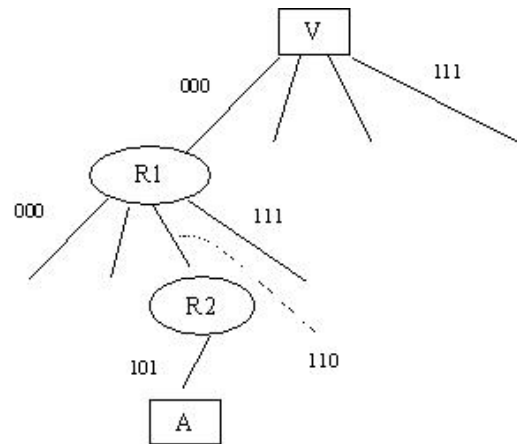


Fig. 4. A schematic representation of an octary tree.

routers, the number of interfaces of a router can be as high as 9. In a k -ary tree, there can be up to k branches from a node to its children. However, the branch from the node to its parent is not taken into account. Since we only consider 8 interfaces (corresponding to 7 children), a 7-ary tree should be used instead of the octary tree. We refer to this as the simple extension model. In fact, a router has 1 link to its “provider”, maybe some to its “peers”, and the rest to its “customers”. Consider a generic case in which a router has a link to its provider via an interface, called *iprv*, and the rest of the interfaces of the router, called *icsts*, are used for its customers. Then, the color used to mark the link via *iprv* can be reused for another link via one of *icsts*, because the two links (reusing the same color) are at different distances (hops) with respect to the victim. A similar reuse example is shown in Fig. 4, where two links of router R_1 using the same marking ‘000’ are at different distance with respect to V , the victim. If ‘000’ stands for one color (e.g., red), the naïve extension may be viewed as a color-marking scheme.

IV. OUR PROPOSED SCHEME

The fundamental problem is that the 16-bit ID field, where the marking is stored, is far less than required. As shown in [10], even if we record only one bit of each router’s IP address information (not the IP address itself, but rather the mapping function of the IP address, such as using MD5 digest), 16 bits are insufficient. To overcome this problem, Yaar *et al.* [10] proposed to record the first edge router information of one AS rather than all routers along the attack path. Since our scheme is based on the Four Color Theorem, we need to use 2 bits to represent one color. As one link is represented by one color, we can record 8 links at most in the ID field. Obviously, this is not enough; we have to create an innovative scheme to cope with the case of having path lengths with more than 8 hops.

A. Marking Algorithm

We propose to record information of the first and last 4 links. Our goal is to distinguish one path from another, and therefore packets launched from a different source shall have a different marking; packets from the same source but traversing

different paths shall have a different marking; only packets from the same source and taking the same path own the same marking. Our marking method can satisfy the above requirements. Consider the directed acyclic graph (DAG) rooted at the victim as shown in Fig. 5. The network as viewed from the victim is a DAG [4], [6]. Note that paths with the same first several links and the same last several links usually traverse the same route. In doing so, we can distinguish different paths with information of only 8 links.

Denote n as the hop count of the attack path. Before we proceed to the marking algorithm, let’s consider the following two cases.

1) $n \leq 8$. In this case, the whole path information can be completely recorded.

2) $n > 8$. Under this circumstance, we only want to record the first 4 links and the last 4 links. However, routers have no idea in advance how many hops there are between itself and the victim. That is, routers do not know whether they lie in the last 4 links or not. To address this problem, we divide the ID field into 2 parts, the first 4 links are marked in the lower byte of the 16 bits, from bit 0 to bit 7. The last 4 links are recorded in bits 8 to 15. Thus, our marking algorithm can be

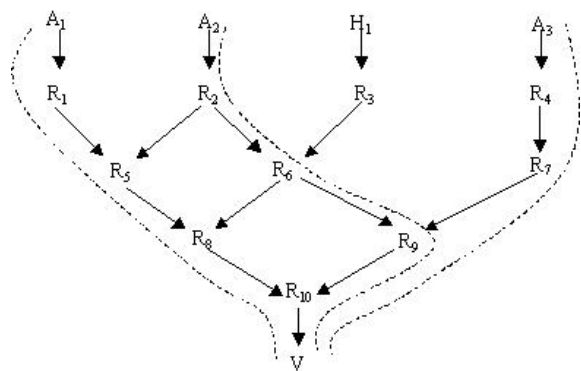


Fig. 5. Network as seen from the victim, V , of an attack. The dotted lines stand for attack paths.

summarized in Fig. 6.

The index field is 2-bit long (two reserved bits in the IP packet header), which tells the current router the position of the marking. For example, assume the total hop count of a path is 11. Then, the value of the ID field changes in the following order (see Table I).

In Table I, each position is composed of two bits, and C_i stands for the color of link i . When one packet arrives at the first router, the color of this link is recorded in bit 0 and bit 1 of the ID field; then, the color of the second link is saved in bits 2 and 3. Finally, when the packet reaches the victim, the ID field has the value $\{C_8, C_{11}, C_{10}, C_9, C_4, C_3, C_2, C_1\}$. This information can be used to actively defend against DDoS attacks. Note the victim can reorder the sequence to $\{C_{11}, C_{10}, C_9, C_8, C_4, C_3, C_2, C_1\}$ from information provided by the index field (index=2, in this example), thus yielding the marking of the color of the first and last 4 links of the path in the reverse order.

B. A Related Issue

One problem remains to be addressed is how to assign color to each link. Assume that N stands for the total number of interfaces of a router. Let us consider the following cases.

Case 1: $N \leq 4$. In this case, one color is used for one interface. 4 colors are sufficient. Note that, as high as 93.30% of all routers in the Internet own no more than 4 interfaces (see Fig. 3).

Case 2: $N=5$. In this case, one color has to be reused. Considering the Internet hierarchy, we can reuse the color for the uplink (the link to its providers). Note that 95.93% of routers own less than 6 interfaces.

Case 3: $N > 5$. To cope with this case, we take advantage of the Internet hierarchy. According to [14]-[16], 3 types of links exist between two routers. That is “uphill” (from a router of a customer to a router of its provider); “downhill” (from a router of a provider to a router of its customer); and “peer” (from a router in a domain to a router of its peer). Assume that the number of links does not exceed 4 for the same type of links.

TABLE I. HOW THE ID FIELD IS MARKED ALONG THE ATTACK PATH

Curr. link	index	Upper byte				Lower byte			
		Pos. 3	Pos. 2	Pos. 1	Pos. 0	Pos. 3	Pos. 2	Pos. 1	Pos. 0
1st	0								C1
2nd	1							C2	C1
3rd	2					C3	C2	C1	
4th	3					C4	C3	C2	C1
5th	0				C5	C4	C3	C2	C1
6th	1			C6	C5	C4	C3	C2	C1
7th	2		C7	C6	C5	C4	C3	C2	C1
8th	3	C8	C7	C6	C5	C4	C3	C2	C1
9th	0	C8	C7	C6	C9	C4	C3	C2	C1
10th	1	C8	C7	C10	C9	C4	C3	C2	C1
11th	2	C8	C11	C10	C9	C4	C3	C2	C1

This is reasonable because we only record the first and last 4 links of an Internet path, that are normally located in the access part of the Internet, and therefore the link number of the same type in a router is not expected to be too large. In this way, the same color can be reused in different types of links of the same router. When we need to distinguish the links with the same color (e.g., for IP traceback), say red, we may assign priority to each type of link according to the portions of the path. For instance, in the uphill part of Fig. 2, for a router R, we first examine whether there exists one “uphill” link marked with red. If this is the case, we view this link as the one we are looking for. Otherwise, we check the “peer” links of router R. If no “peer” link is marked with red, we finally check the router R’s “downhill” links. In the downhill part, the order to check whether there is one link marked with red shall be from the “downhill” link, to the “peer” link, and finally to the “uphill” link. In this way, we can reuse the same color.

C. Benefits

Several benefits can be achieved with this marking scheme. First, the marking scheme is more practical because the average number of interfaces of routers is 3 to 4. Second, we believe that with the first and last portion of the path information, we can reach the same defending effect owing to the DAG model. Third, all required path information is embedded in one packet, and all packets traversing the same path have the same marking. By counting the number of packets with the same marking, the victim can distinguish malicious attack packets from normal packets. Once identified, malicious packets may be dropped or filtered by the victim. This empowers the victim to actively defend against DDoS attacks and facilitate faster response time.

V. SIMULATIONS

To test the accuracy of our scheme, several simulations on different configurations have been conducted using ns-2. Owing to space limitation, we only present one example. Fig. 7 shows the network topology used in our simulations. Two scenarios are considered. First, we simulate attack packets launched from different paths. Attackers 1, 2, and 3 traverse nodes 3-15-19-16-17-18 to the victim, node 14. Attackers 4 and 5 take the path 4-15-19-16-17-18, called path 1, to the victim. Attacker 7 goes through nodes 6-18 to node 14. Other nodes send normal traffic. At node 14, we verify and compute

For the lower byte (bit 7 to 0)

- 0) index=0,
- 1) record the color of the current link in the position pointed by the index,
- 2) index=mod(index+1,4),
- 3) If index=0, then goto step 4; else, goto step 1.

For the upper byte (bit 15 to 8)

- 4) record the color of the current link in the position pointed by the index,
- 5) index=mod(index+1, 4),
- 6) goto step 4.

Fig. 6. The proposed marking algorithm.

the percentage of packets that have been correctly marked by our scheme. The result shows that the marking information is 100% correct. Second, we consider the case that attack packets are launched from the same source but take different paths. This scenario happens for reasons such as load balancing and routing instability. We cut the link between node 16 and 17, and thus traffic from attackers 4 and 5 go through node 4-15-19-16-5-17-18, called path 2, to reach the victim. Note that these packets traversing path 2 have different markings from packets traversing path 1 even though they are from the same sources (attackers 4 and 5). Under this scenario, the marking information of all packets is still 100% correct. The results confirm the correctness of our scheme.

VI. CONCLUSIONS

In this paper, we have proposed a new marking method based on the Four Color Theorem that takes advantage of the Internet hierarchy. With this scheme, one serious limitation on the number of interfaces of routers is relaxed, thus rendering the method more practical. Furthermore, the victim can actively protect itself rather than passively depend on others. Both are very desirable features to defend against DDoS attacks.

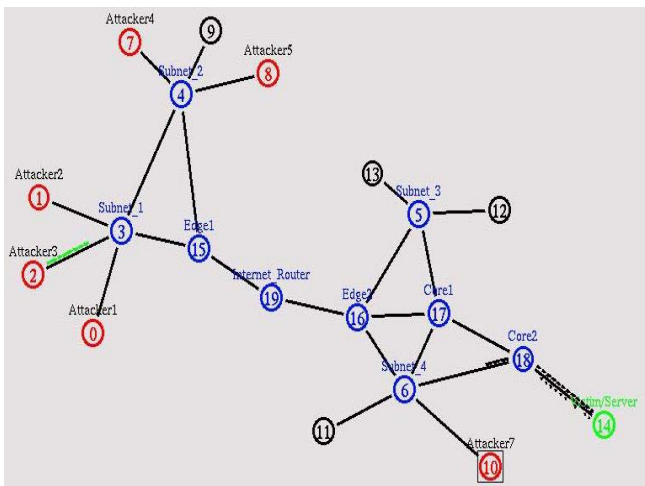


FIG. 7. THE NETWORK TOPOLOGY USED FOR SIMULATION.

REFERENCES

- [1] Computer Security Institute and Federal Bureau of Investigation, "2003 CSI/FBI Computer Crime and Security Survey," Sep. 2003. http://www.securitymanagement.com/library/CSI_Fbi0903.pdf.
- [2] D Moore, G. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity," *USENIX Security Symposium*, Washington, D.C., Aug. 2001.
- [3] H. Wang, D. Zhang, and K. Shin, "Detecting SYN flooding Attacks," *IEEE INFOCOM 2002*, Jun. 2002, pp. 1530-1539.
- [4] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network Support for IP Traceback," *IEEE/ACM Trans. Networking*, vol. 9, pp. 226-237, Jun. 2001.
- [5] A. C. Snoeren, C. Partridge, et al., "Single Packet IP Traceback," *IEEE/ACM Trans. Networking*, vol. 10, pp. 721-734, Dec. 2002.
- [6] D. Song and A. Perriag, "Advanced and Authenticated Marking Schemes for IP traceback," *INFOCOM 2001*, pp. 878-886.
- [7] M Sung and J. Xu, "IP traceback-based intelligent packet filtering: a novel technique for defending against Internet DDoS attacks," *IEEE Trans. Parallel & Distributed Systems*, vol. 14, Sep. 2003, pp. 861-872.
- [8] A. Belenky and N. Ansari, "IP Traceback with Deterministic Packet Marking," *IEEE Comm. Letters*, vol. 7, no. 4, pp. 162-164, Apr. 2003.
- [9] S. Bellovin, "ICMP Traceback Messages," IETF Draft, Mar. 2000. [Online]. Available: <http://www.research.att.com/smb/papers/draft-bellovin-itrace-00.txt>.
- [10] A. Yaar, A. Perriag, and D. Song, "Pi: A Path Identification Mechanism to Defend against DDoS Attacks," *IEEE Symposium Privacy and Security*, May 2003.
- [11] CAIDA. <http://www.caida.org/tools/measurement/iffinder>.
- [12] T. Satty, "The four-color problem : assaults and conquest", New York : Dover Publications, 1986.
- [13] Galtech, Four Color Theorem. [Online]. Available: <http://www.math.gatech.edu/~thomas/FC/fourcolor.html>.
- [14] L. Gao, "On inferring autonomous system relationships in the Internet," *IEEE/ACM Trans. Networking*, vol. 9, pp. 733-745, Dec. 2001.
- [15] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz, "Characterizing the Internet hierarchy from multiple vantage points," *IEEE INFOCOM 2002*, vol. 2, Jun. 2002, pp. 618-627.
- [16] L. Gao and J. Rexford, "Stable Internet Routing without Global Coordination," *IEEE/ACM Trans. Networking*, vol. 9, pp. 681-692, Dec. 2001.
- [17] J. Ioannidis and S. Bellovin, "Implementing Pushback: Router-based Defense Against DDoS Attacks," *Symposium Network & Distributed System (NDSS'02)*, San Diego, CA, Feb. 2002.
- [18] Y. Xiong, S. Liu and P. Sun, "On the defense of the distributed denial of service attacks: an on-off feedback control approach," *IEEE Trans. Systems, Man and Cybernetics*, vol. 31, Jul. 2001, pp. 282-293.
- [19] W. Theilmann and K. Rothermel, "Dynamic distance maps of the Internet," *Proc. IEEE INFOCOM*, vol. 1, Mar. 2000, pp. 275-284.