Reconstruction of Malicious Internet Flows

Omer Demir
Graduate Center, City University of New York (CUNY) New York, NY, U.S.A.
odemir@gc.cuny.edu

Bilal Khan
John Jay College, City University of New York (CUNY) New York, NY, U.S.A.
bkhan@jjay.cuny.edu

Ala Al-Fuqaha
Computer Science Dept. Western Michigan University Kalamazoo, MI, U.S.A.
ala@ieee.org

ABSTRACT

We describe a general-purpose distributed system capable of traceback of malicious flow trajectories in the wide area despite possible source IP spoofing. Our system requires the placement of agents on a subset of the inter-autonomous system (AS) links of the Internet. Agents are instrumented with a uniform notion of *attack criterion*. Deployed, these agents implement a self-organizing, decentralized mechanism that is capable of reconstructing topological and temporal information about malicious flows. For example, when the attack criterion is taken to be based on excessive TCP connection establishment traffic to a destination, the system becomes a traceback service for distributed denial of service (DDoS) attacks. As another special case, when the attack criterion is taken to be based on malicious payload signature match as defined by an intrusion detection system (IDS), the agents provide a service for tracing malware propagation pathways. The main contribution of this paper, is to demonstrate that the proposed system is effective at recovering malicious flow structure even at moderate levels of deployment in large networks, including within the present Internet topology.

Categories and Subject Descriptors

C.2.0 [**Computer Communication Networks**]: General—*Security and protection*

General Terms

Security.

Keywords

Distributed denial of service, flow reconstruction.

1. INTRODUCTION

We define *flow reconstruction* as "determining the true sources and/or routes of packets going to a given destination". There are many obstacles to flow reconstruction arising from the design of Internet architecture itself; here we give only a brief summary – a more detailed treatment is given in [6]. First, in order to gain the most of the Internet, its network link resources are shared among all users. Unfortunately, there is no intrinsic enforcement that the sharing is fair. Second, the network core must deal with very high volumes of traffic which must be processed fast, so core network components do as little as possible per packet, requiring all complex computations to be at the edge computers and thereby leaving a core that is unable to provide enhanced transport services. Finally, the Internet is composed of interconnected networks managed by heterogeneous authorities making widespread deployment of defense mechanisms difficult.

Given the aforementioned inherent obstacles, the notion of "solving" the flow reconstruction problem can have many interpretations. Here we focus on the problem of determining the true origins and mechanics of malicious flows to a given destination, where maliciousness is determined by a well-defined attack criterion. We will consider two special cases of *attack criterion* for concreteness:

- **DDoS**: An agent declares the attack criterion met when it observes excessive TCP connection establishment traffic to a destination. DDoS attacks have been identified as the most critical concern by the Internet Service Providers in Arbor Network's 2008 survey [1].
- **Malware**: An agent declares the attack criterion met when packet payload matches a malware signature.

In general any attack citerion may be used, provided it is applied uniformly to all agents and it satisfies the following condition:

> If an agent A upon observing packets P going to victim v, decides that the attack criterion has been met, then the attack criterion will also be met at all agents further downstream from A (towards v) when they receive packets P.

Identifying the source of malicious packets is difficult because the source address field in the IP packet header of packets is easy to "spoof". The problem is further rendered intractable by the fact that current network standards do not require network devices to maintain information about the paths which packets take. There have been different approaches to this problem and we give a brief synopsis of prominent examples of each of these approaches below.

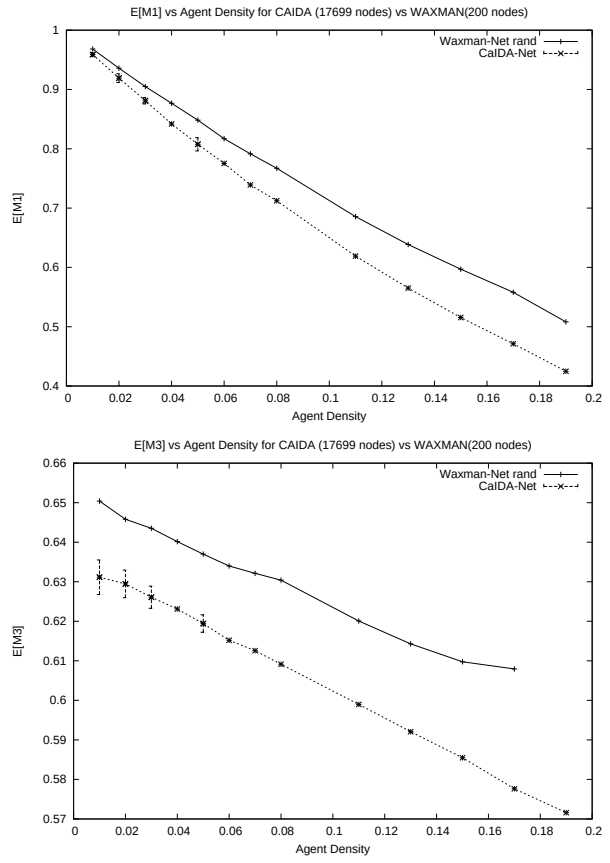

Figure 3: M1 (top) and M3 (below), for agents placed randomly in the Internet.

random network. Figure 3 (bottom) show M3 curve for the same two networks. Again, both curves indicate similar behavior, but the CAIDA network is lower than the curve of the Waxman network, indicating that our scheme scales to large networks and beyond the Waxman model. At a 20% deployment within the Internet, half the attacker flows are represented within the agent tree an we are able to traceback on average halfway from the victim to the attacker.

6. CONCLUSION AND FUTURE WORK

We presented a scalable agent-based system for collecting information about the structure and dynamics of malicious attack flows. The agents in our system are placed on inter-autonomous system (AS) links and implement a self-organizing decentralized mechanism capable of reconstructing topological information about the spatial and temporal structure of attacks. We showed that our system is effective at recovering malicious flow structure, even at moderate levels of agent deployment. Finally, and most significantly, we report that the effectiveness of the system *improves* as network size increases and extends beyond artificial Waxman networks to real-world data Internet topologies. Taken together, these results point to the viability of the proposed system as a scalable solution to the flow reconstruction problem.

Future work. The next stages of our research will involve the development of more sophisticated algorithms for placing agents. While the variance in M1 and M3 measures obtained are quite small, we are hopeful that by carefully placing the agents within the network, we will be able to deploy a system that performs well at even sparser agent densities.

Acknowledgements. The first author would like to thank the Turkish National Police for funding these research efforts.

7. REFERENCES

[1] Arbor Networks, `http://www.arbornetworks.com`.

[2] Bellovin, ICMP traceback messages,RFC draft, September ‘`http://tools.ietf.org/draft/draft-bellovin-itrace/draft-bellovin-itrace-00.txt` (2000).

[3] Bellovin, Cert advisory ca-1996-26, Cert Advisory, ‘`http://www.cert.org/advisories/CA-1996-26.html` (1996).

[4] Bloom, B. H.: Space time trade-offs in hash coding with allowable errors, Commun. ACM, vol. 13, no. 7, pp. 422–426, (1970).

[5] Burch and Hal: Tracing anonymous packets to their approximate source, Proceedings of the 14th USENIX conference on System administration. Berkeley, CA, USA: USENIX Association, 319–328 (2000).

[6] Demir O. : A Survey of Network Denial of Service Attacks and Countermeasures. City University of New York, Computer Science Department. (2009).

[7] Demir, O., Khan, B. : An Agent-based Architecture for Flow Reconstruction of DDoS Attacks. Submitted to International Communications Conference (ICC) 2010, Cape Town, South Africa, 23-27 May 2010.

[8] Demir, O., Khan, B.: Quantifying Distributed System Stability through Simulation A Case Study of an Agent-based System for Flow Reconstruction of DDoS Attacks. In: Proceedings of the 1st Intelligent Systems, Modelling and Simulation Conference, Liverpool, England, 27-29 January 2010.

[9] Gemberling B., Morrow, C., and Greene, B.:ISP security-real world techniques. presentation, nanog. NANOG, `www.nanog.org` (2001).

[10] Gligor V.D.: A Note on Denial-of-Service in Operating Systems. IEEE Trans. Softw. Eng. 10, 320–324 (1984).

[11] Savage, S, Wetherall, D. Karlin, A. and Anderson, T.: Practical network support for IP traceback, SIGCOMM Comput. Commun. Rev., vol. 30, no. 4, pp. 295–306, (2000).

[12] Snoeren, A. C.: Hash-based IP traceback, in SIGCOMM ’01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications. New York, NY, USA: ACM, pp. 3–14, (2001).

[13] The IPv4 Routed /24 AS Links Dataset, Young Hyun, Bradley Huffaker, Dan Andersen, Emile Aben, Matthew Luckie, K. C. Claffy, and Colleen Shannon, 11/15/2009, `http://www.caida.org`.

[14] Waxman, B. M.: Routing of Multipoint Connections.: Broadband Switching: Architectures, Protocols, Design, and Analysis. IEEE Computer Society Press, Los Alamitos, CA, USA (1991).