Probe-response attacks on collaborative intrusion detection systems: effectiveness and countermeasures

Emmanouil Vasilomanolakis[a,b], Michael Stahn[b], Carlos Garcia Cordero[b], Max Mühlhäuser[b]
[a]AGT International, Germany
[b]Telecooperation Group, Technische Universität Darmstadt / CASED, Germany
{manolis, carlos.garcia, max}@cased.de, michael.stahn@stud.tu-darmstadt.de

***Abstract*—Over the last years the number of cyber-attacks has been constantly increasing. Since isolated Intrusion Detection Systems (IDSs) cannot cope with the number and sophistication of attacks, collaboration among the defenders is required. Collaborative IDSs (CIDSs) work by exchanging alert traffic to construct a holistic view of the monitored network. However, an adversary can utilize probe-response attacks to successfully detect CIDS's monitoring sensors. We discuss the practicability of such attacks, suggest improvements, and also propose novel techniques to reduce the effects of such attacks. Moreover, we present preliminary results in the applicability of the attacks and hints on performing such attacks in a well known CIDS.**

I. Introduction

Cyber-attacks are nowadays increasing in terms of numbers but also in their sophistication; malware and botnets are a constant phenomenon that is hard to detect and mitigate. To effectively defend against such attacks, collaboration is required. Isolated IDSs and firewalls are not sufficient and cannot cope with such challenges. Upon this need for collaboration, CIDSs have emerged [1], utilizing a number of IDSs, honeypots, and other defense mechanisms that work together to create a holistic view of the monitored network. Cyber-incident monitors are an example of such systems. They utilize a number of sensors, to collect, correlate, aggregate and present their results to security administrators.

A multitude of such systems exists that publish their results publicly over the Internet [2], [3]. This is important for researchers and security administrators to create datasets, gather statistics, and knowledge that supports research in the cyber-security area. Nevertheless, a class of attacks exists, called *probe-response*, that is able to severely reduce the advantages of such systems. In short, via their usage it is possible to detect the position of collaborative sensors, i.e., their IP address, and thus disrupt them. For instance, one can perform a Distributed Denial of Service (DDoS) attack on them, or create a blacklist of these IP addresses and integrate it into a malware that will propagate without accidentally targeting a sensor (thus remaining undetected for a longer period of time).

II. Probe-Response attacks

Probe-response attacks were introduced independently by researchers in 2005 [4], [5]. The essence of such attacks is that cyber-incident monitors and CIDSs, that publish their results publicly over the Internet, provide a feedback loop that can be exploited by malicious users. By utilizing the responses that the attacker receives from the public output of the CIDS, they can potentially identify the sensors. Figure 1 provides a high level example of a probe-response attack. At a glance, the adversary starts the attack by sending probes to a large group of monitoring sensor candidates (i.e., a large IP address space). These probes contain special crafted watermarks, so called *markers*, that can be subsequently exploited for distinguishing probes from normal alert data in the CIDS's output. The attacker can subsequently reduce the IP space and repeat the *probing steps* until sensors are revealed.

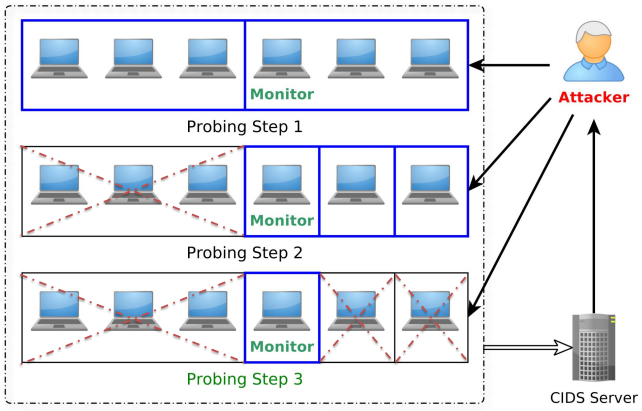

Fig. 1: Probe-response attack example

One of the main limitations of such attacks in the past was the amount of time required to perform a scan in the whole IPv4 address space. However, with the advances of the last years, we argue that this problem has been (at least partially) tackled in a twofold manner. First, the ordinary users' Internet speed and bandwidth has increased significantly. Second, research in the area of rapid Internet-level scanning has resulted in research tools such as ZMap [6] that are able of scanning the entire IPv4 space in less than an hour.

III. Improving attacks and countermeasures

A. Improving probe-response attacks

A *marker* in a probe-response attack can take many forms. For instance, the adversary can make use of a non common source (or destination) port in the probe message to afterwards distinguish it from the responses of the CIDS. Selecting

the proper marker for performing a probe-response attack is important as it influences the time required for performing the attack. In Figure 2 we show the frequency distribution of possible probe markers, i.e., destination ports, source ports, and IP source addresses, in the context of the Dshield CIDS [3]. We plot the frequency of the alert data gathered in a 12 hours period. From the set of all available ports, only a few are ever utilized, and also the IP addresses provide enough space for a marker. Approximately $46,943$ destination, and $4,270$ source ports do not appear in our analysis, which gives enough flexibility to utilize them as markers. This also applies to the (source) IP addresses (in a magnitude of 10^9 available addresses) especially when taking into account that an attacker can spoof IP addresses that have not been seen before. This is the baseline for our future work; we argue that hybrid markers, e.g., via the utilization of source ports and IPs, can be exploited for improving the speed and efficiency of probe-response attacks.

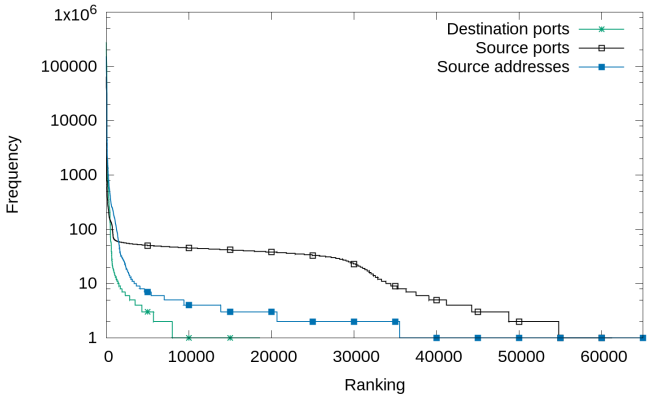

Fig. 2: Distribution of possible probe markers in DShield

B. Improving mitigation

The first step to cope with probe-response attacks is to detect their presence in a CIDS. We propose a simple, yet effective, metric to detect such attacks by utilizing the ratio of generated alerts in relationship to the number of actively reporting sensors. Let A be the set of all generated alerts, S be the set of all sensors, $S_t \subset S$ the set of reporting sensors within time-frame t, and $A_t \subset A$ the set of generated alerts within time-frame t. The ratio r is defined as $r = \frac{|A_t|}{|S_t|}$.

Figure 3 depicts the distribution of r for data gathered by the DShield CIDS within a period of 24 hours. An attacker requires approximately 5 hours (with a 100MBit network connection) to perform one probing step in the entire IPv4 range [6], probing approximately $90,000$ sensor addresses per hour of the total $500,000$ sensors of DShield [3]. We argue that in the presence of a probe-response attack the number of unique reporting sensors within a time-frame $|S_t|$ will increase significantly, while $|A_t|$ will only have a relatively small increase, therefore modifying r. In the presented period we observe the sensors $|S| = 131,344$, the alerts $|A| = 10,934,768$, and an average unique sensor count (per hour) $\sum_t \frac{|S_t|}{24} = 55,000$. We simulate a probing-attack by injecting alarms in the time-frames between 4 and 17 (which enables three complete probing steps) in a 24 hour period. By assuming that the maximum probing rate is 90,000 and that sensors might already be present, the injections are done according to a uniform distribution between 80,000 and 90,000. As shown in Figure 3, it becomes evident that during an attack the ratio r decreases significantly.

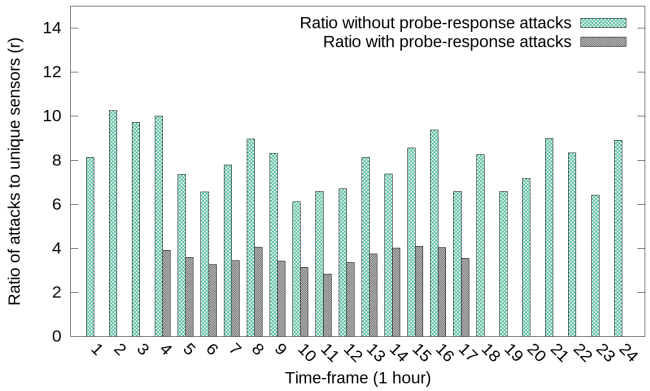

Fig. 3: Attack ratio in the presence of probe-response attacks

We plan to investigate techniques for either reducing the effects of such an attack or making it hard for an adversary to identify probes. For instance, one possible method can be the addition of *noise* data in the output results. This can be further improved with the idea of *adaptive reporting*, i.e., the CIDS adds noise and changes its reports when it detects the presence of a probe-response attack.

IV. Conclusion and Future Work

Probe-response attacks can reduce the benefits of CIDSs and cyber-incident monitors. With the knowledge of the position of sensors, one can either attack them or utilize this knowledge to create malware that is able to avoid detection. Our preliminary results suggest that such attacks do have practical impact on real systems. In our future work we plan to comprehensively study the applicability of probe-response attacks to well known CIDSs, e.g., DShield, along with suggestions for providing feasible defense mechanisms.

References

[1] E. Vasilomanolakis, S. Karuppayah, M. Mühlhäuser, and M. Fischer, "Taxonomy and Survey of Collaborative Intrusion Detection," *ACM Computing Surveys*, vol. 47, no. 4, p. 33, 2015.

[2] E. Vasilomanolakis, S. Karuppayah, P. Kikiras, and M. Mühlhäuser, "A honeypot-driven cyber incident monitor: lessons learned and steps ahead," in *International Conference on Security of Information and Networks*. ACM, 2015.

[3] J. Ullrich, "Dshield," http://dshield.org, 2000.

[4] J. Bethencourt, J. Franklin, and M. Vernon, "Mapping internet sensors with probe response attacks," in *USENIX Security Symposium*, 2005, pp. 193–208.

[5] Y. Shinoda, K. Ikai, and M. Itoh, "Vulnerabilities of passive internet threat monitors," in *USENIX Security Symposium*, 2005, pp. 209–224.

[6] Z. Durumeric, E. Wustrow, and J. A. Halderman, "ZMap: Fast Internet-wide Scanning and Its Security Applications," in *Proceedings of the 22nd USENIX Security Symposium*, 2013, pp. 605–619.